# The *Drosophila* 7SK snRNP complex is required for synaptic growth and function of motoneurons

## Dissertation

Zur Erlangung des Grades
Doktor der Naturwissenschaften
Am Fachbereich Biologie
Der Johannes Gutenberg-Universität Mainz,
Deutschland

(With the aim of achieving a doctoral of science at the
Department of Biology, Johannes Gutenberg University,
Mainz, Germany)

Submitted by

**Giriram Kumar Mohana Sundara Shanmugam**

Date of Birth: 16 October 1987, Coimbatore, India.

# Evaluation Committee

Dean:

1st Reviewer:

2nd Reviewer:

Date of Oral Defense: **26th September 2019**

# Table of Contents

# Abstract

# 1 ABSTRACT

The 7SK snRNP is a ribonucleoprotein complex composed of the abundant non-coding nuclear RNA *7SK*, the RNA binding proteins Methylphosphate Capping Enzyme (MePCE), La-related protein 7 (Larp7) and Hexamethylene bis-acetamide-inducible (HEXIM). In higher eukaryotes, the 7SK snRNP complex plays a major role in preventing the premature entry of paused RNA Pol II into the elongation phase by sequestering the positive transcription elongation factor (P-TEFb). Intriguingly, despite this general function characterized essentially from cell culture studies, the *LARP7* loss of function in human is viable. Nevertheless, the patients suffer from several defects including restricted growth and intellectual disability, also known as the Alazami syndrome. Currently, it is unclear how the absence of LARP7 specifically gives rise to this syndrome. In order to gain insights into this question and to more globally assess the function of promoter-proximal pausing in a developmental context, I used *Drosophila* as a model organism to generate mutants of several 7SK snRNP complex subunits. I found that the knockout of *Larp7* or *7SK* RNA in *Drosophila* does not affect viability but alters fly locomotion. Consistently, alteration of the 7SK snRNP complex specifically reduces axonal growth at neuromuscular junctions (NMJ) of developing larvae. I showed that Larp7 is enriched in a few subtypes of motoneurons and acts autonomously in these cells to promote axonal growth. In addition, electrophysiology analysis of synaptic activity shows mild alteration of synaptic transmission. Interestingly, decreasing the level of P-TEFb fully restores the axonal growth and partially the locomotion, indicating that the 7SK snRNP complex regulates growth via a transcriptional function. Our transcriptomic analysis of mutant motoneurons revealed that the 7SK snRNP complex regulates genes that contain high GC content at their promoter as well as long introns. Altogether, our work adds new insights into the specificity of the 7SK snRNP complex during the development of a multicellular organism and highlights the importance of promoter-proximal pausing in the development of motoneurons.

# 2 ZUSAMMENFASSUNG


**Der *Drosophila* 7SK snRNP-Komplex wird für synaptisches Wachstum und die Funktion von Motoneuronen benötigt.**

Das 7SK snRNP ist ein Ribonukleoprotein-Komplex, der sich aus der reichlich vorhandenen nicht-kodierenden nuklearen RNA *7SK*, den RNA-Bindungsproteinen Methylphosphat-Capping-Enzym (MePCE), La-related Protein 7 (Larp7) und Hexamethylene bis-acetamide-inducible (HEXIM) zusammensetzt. In höheren Eukaryoten spielt der 7SK-snRNP-Komplex eine wichtige Rolle bei der Verhinderung des vorzeitigen Eintritts von pausierter RNA Pol II in die Elongationsphase durch Sequestrieren des positive transcription elongation factor (P-TEFb). Interessanterweise ist trotz dieser allgemeinen Funktion, die im Wesentlichen aus Zellkulturstudien hervorgeht, der *LARP7*-Funktionsverlust beim Menschen lebensfähig. Nichtsdestotrotz leiden die Patienten an mehreren Defekten, einschließlich eingeschränktem Wachstum und geistiger Behinderung, auch bekannt als Alazami-Syndrom. Derzeit ist unklar wie das Fehlen von LARP7 zu diesem Syndrom führt. Um Einsichten in diese Frage zu gewinnen und die Funktion des promoter proximal pausings globaler im Kontext der Entwicklung zu bewerten, generierte ich Mutanten mehrerer 7SK snRNP-Komplexuntereinheiten im Modellorganismus *Drosophila*. Ich fand heraus, dass der Funktionsverlust von *Larp7* oder *7SK*-RNA in *Drosophila* nicht die Lebensfähigkeit beeinflusst, sondern die Fortbewegung der Fliegen verändert. In Übereinstimmung damit reduziert die Veränderung des 7SK snRNP-Komplexes spezifisch das axonale Wachstum an neuromuskulären Übergängen (NMJ) von sich entwickelnden Larven. Ich habe gezeigt, dass Larp7 in einigen Motoneuronsubtypen angereichert ist und in diesen Zellen autonom wirkt, um das axonale Wachstum zu fördern. Darüber hinaus zeigt die elektrophysiologische Analyse der synaptischen Aktivität eine leichte Veränderung der synaptischen Übertragung. Interessanterweise stellt eine Verringerung des P-TEFb-Spiegels vollständig das axonale Wachstum und partiell die Fortbewegung wieder her,

was darauf hinweist, dass der 7SK-snRNP-Komplex das Wachstum über eine Transkriptionsfunktion reguliert. Unsere transkriptomische Analyse von mutierten Motoneuronen ergab, dass der 7SK-snRNP-Komplex Gene reguliert, die an ihrem Promotor einen hohen GC-Gehalt sowie lange Introns enthalten. Zusammengefasst liefert unsere Arbeit neue Einblicke in die Spezifität des 7SK-snRNP-Komplexes während der Entwicklung eines mehrzelligen Organismus und unterstreicht die Bedeutung des promoter proximal pausings bei der Entwicklung von Motoneuronen.

# Introduction

# 3 INTRODUCTION

## 3.1 Gene Regulation

Nucleic acids are the fundamental genetic element present in all living organisms on earth. They are known for more than a century. Friedrich Miescher (1868-69) in Tubingen, first isolated DNA in 1871 and named it nuclein (Dahm, 2005). The discovery of nuclein was the first breakthrough in the history of DNA research. Following this, the chemical nature of the nuclein was identified by P.A Levene in 1912 (Levene and Jacobe, 1911; Levene, P A La Froge, 1912; Levene, 1917). In 1953, Rosalind Franklin and Maurice Wilkins used X-ray diffraction to observe the helical structure of DNA, and in the same year, Francis Crick and James Watson discovered the double helical structure. Later in 1977, Frederick Sanger developed the DNA sequencing method, and the first genome of a living organism to be sequenced was *Haemophilus influenzae*. As of today, the genome of thousands of organisms has been sequenced. The discovery of the DNA structure and DNA modifying enzymes leads to significant advancement in molecular biology. DNA is the genetic element, which is transferred from one generation to other. The genome preserves the information of an organism, and this information is further decoded in germ cells to ensure the development of the whole animal.

In a cell, the information from DNA is further transcribed into RNA. This process is spatiotemporally regulated. A coding gene contains certain basic elements as follows: promoter, 5´untranslated region (UTR), exon, intron, 3´UTR and terminator. Before synthesizing the functional protein, the information from DNA should be processed properly. The coding region of the DNA is first transcribed into pre-messenger RNA. The pre-mRNA is synthesised by the protein complex DNA dependent RNA polymerase II. The introns of the pre-mRNA disrupt the coding message and need to be removed. The spliceosome, one of the largest RNA-protein complex in a cell is responsible for this function. The spliced mature RNA is further transported into the cytoplasm and translated into proteins by another large RNA-protein complex, the ribosome. Proteins are folded properly and transported to specific functional locations in the cell.

Production of functional proteins from genes is a multistep process with many regulations at each step by different RNA-protein complexes. This is to ensure that the information from the DNA is properly and faithfully transcribed and translated into protein without any error. Mistakes in any of these steps lead to non-functional proteins, which can cause detrimental effects for the organism. Here I focus on transcription, the very first step of the central dogma of life.

### 3.1.1 Transcription

The RNA polymerase (RNA pol) complex transcribes DNA into RNA. Three types of RNA polymerases were identified by chromatographic separation (Roeder and Rutter, 1969) and further classified according to the sensitivity of RNA pol to alpha-amanitin. RNA Pol I transcribes large rRNA like 25s rRNA. RNA Pol II is responsible for messenger RNA and some small RNA while RNA pol III transcribes tRNA and small rRNAs (Wieland *et al.*, 1968; Seifart and Sekeris, 1969; Kedinger *et al.*, 1970; Zylber and Penman, 1971; Weinmann and Roeder, 1974).

RNA Pol II is responsible for the synthesis of a variety of RNA that codes for proteins. Here I focus on this polymerase and on the known mechanisms of transcription regulation. The transcription cycle consists of four major steps: assembly, initiation, elongation and termination (Figure 1).

### 3.1.2 The Assembly

The very first step of transcription is the assembly of different transcription factors (TF) on promoters. Decades of research in different organisms lead to the discovery of general transcription factors (GTFs). GTFs are named as TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH according to the chromatographic profiles (Matsuis *et al.*, 1980). The GTFs form the preinitiation complex and recruit RNA Pol II. Activators and repressors regulate the assembly of the transcription factors to regulate gene expression.

Promoters contain different elements including the TATA box, Initiator (Inr), downstream promoter element (DPE), downstream core element (DCE), TFIIB

recognition element (BRE) and GAGA motifs. These elements are not always present, and gene promoters can contain different element combinations.

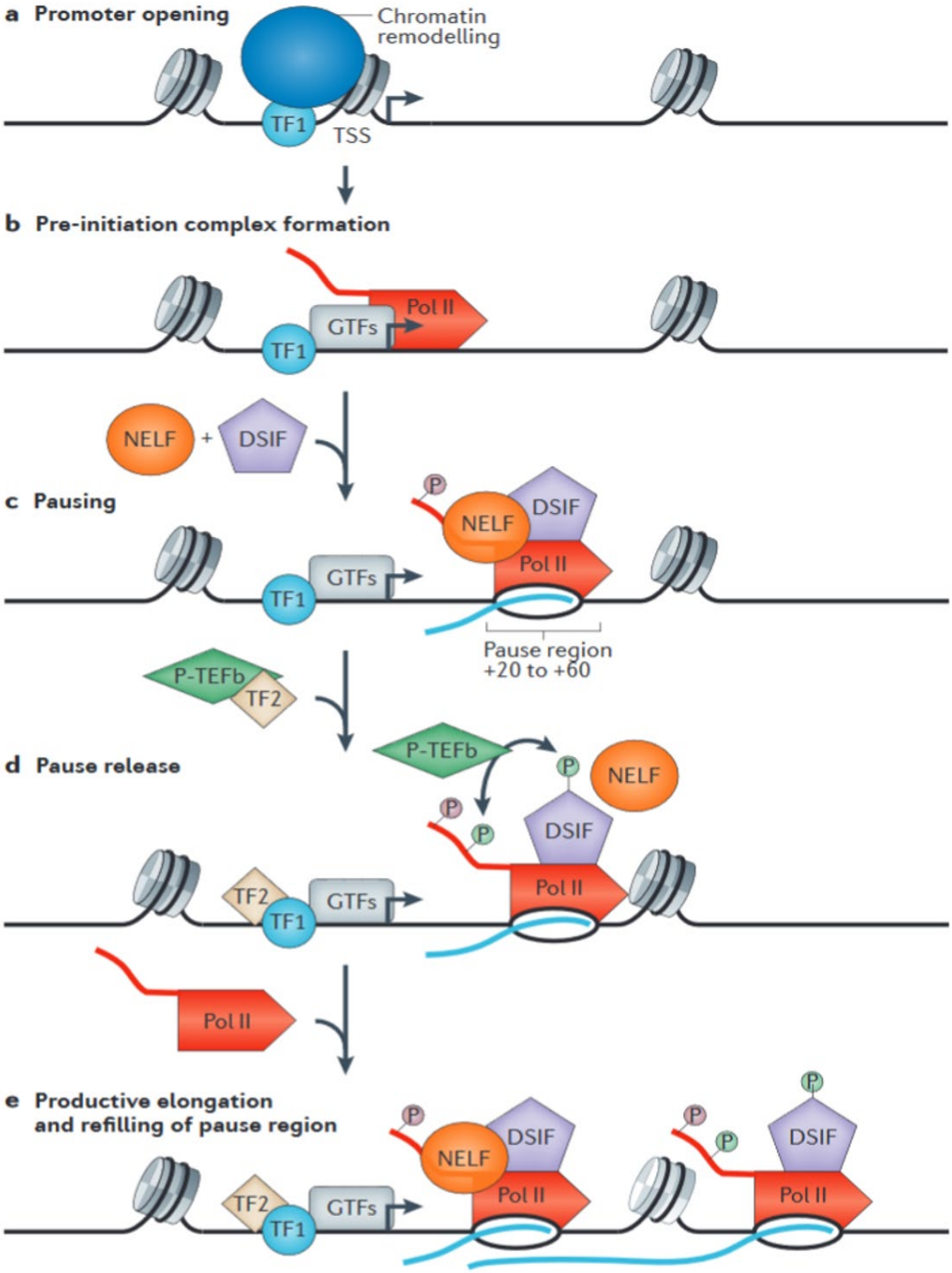



**Figure 1:** Transcription cycle and Pol II pausing. GTFs and RNA Pol II assemble at gene promoter and initiate transcription(a,b). NELF and DSIF associate with initiated RNA Pol II and pause transcription (c). P-TEFb phosphorylates RNA Pol II and negative elongation factors (d). This results in the release of Pol II into productive elongation (e). Adapted from Karen et al., (Karen Adelman & Lis, 2012)

On a promoter, the GTFs assemble sequentially. First, the subunit of TFIID, TATA-binding protein (TBP), specifically binds TATA motif. Following this binding, TFIIA and TFIIB bind TBP and helps in bending the DNA to facilitate melting of the double strand. Then TFIIF brings RNA pol II and stabilises the preassembled complex. Finally, TFIIE recruits an essential cyclin-dependent kinase (CDK7) containing complex TFIIH to the promoter to form the final preinitiation complex (PIC) (Matsuis *et al.*, 1980; Orphanides, Lagrange and Reinberg, 1996; Pugh, 1996; Juven-Gershon *et al.*, 2008; Jennings, 2013; Roy and Singer, 2015; Zhang *et al.*, 2016)

### 3.1.3 Initiation

After the formation of the preinitiation complex, CDK7 phosphorylates Ser5 at the carboxy-terminal domain (CTD) of RNA pol II (Watanabe *et al.*, 2000). The CTD of RNA pol II consists of heptad (YSPTSPS) repeats, and the number of repeats varies depending on the organism. The phosphorylation of Ser5 of RNA Pol II is the hallmark signature of transcription initiation. Chromatin immunoprecipitation sequencing (ChIP-Seq) using Ser5P specific antibody revealed that this modified RNA Pol II is localised around the transcription start site (TSS).

Next, by an adenosine triphosphate (ATP) dependent process, the promoter element melts at the region where the RNA pol II – GTFs bound, forming a "Transcription Bubble", and this forms the open promoter complex (Orphanides, Lagrange and Reinberg, 1996). Then RNA pol II at the active site initiates the DNA dependent RNA synthesis (Sainsbury, Bernecky and Cramer, 2015). As soon as a few nucleotides of RNA are synthesized, RNA pol II is released from the GTFs, which results in "promoter clearance" and engagement into elongation.

Another level of regulation occurs by 5' RNA capping enzyme. Newly synthesized RNA is capped at the 5'end by three enzymatic activities. First, digestion of 5'triphosphate of the nascent RNA, then the addition of guanine base in 5' to 5' linkage and finally methylation of the added guanine base (Perales and Bentley, 2009). The guanylyltransferase activity of capping enzyme is stimulated when it binds to the CTD of RNA pol II Ser5P but not to Ser2P (Ho and Shuman, 1999).

This binding facilitates promoter clearance and RNA pol II elongation (Schroeder *et al.*, 2004).

### 3.1.4 Pausing

Earlier it was thought that transcription initiation was the rate-limiting step, and once the RNA pol II was released into elongation, RNA synthesis continues until termination. However, research from the last 20 years showed that most of the genes that have initiated elongation paused after transcription of 80-120 nt. Initially, paused Pol II was observed in stress response and developmental genes, but recent studies demonstrated that pausing is more prevalent. Paused RNA pol II is important for synchronised expression of genes during development and for genes that encode products of the same molecular complex. Not all genes which burst expression upon stress are occupied with paused Pol II, but all stress response genes with paused Pol II burst into transcription elongation (Gilmour and Lis, 1986a, 1986b; Rougvie and Lis, 1988; Nechaev, David C Fargo, *et al.*, 2010).

Recent reports suggested that pausing of RNA Pol II can be categorized into 'regulated pausing' and 'intrinsic pausing'. In regulated pausing, factors like P-TEFb, NELF and DSIF fine-tune the duration of the pausing, while in intrinsic pausing, pausing occurs due to the nucleosome barriers and nucleic acid sequence (Adelman and Henriques, 2018).

### 3.1.5 Elongation

The major complex involved in productive elongation is positive transcription elongation factor b (P-TEFb). It is a heterodimer of cyclin-dependent kinase (CDK9) and cyclin T (CYC-T). The kinase phosphorylates Ser2 of RNA Pol II CTD (Ramanathan *et al.*, 2001), as well as pausing factors DRB sensitivity-inducing factor (DSIF) (Kim and Sharp, 2001; Yamada *et al.*, 2006) and Negative elongation factor NELF (Fujinaga *et al.*, 2004). Once RNA pol II is released into elongation, other factors assemble with RNA Pol II to overcome the barriers like nucleosomes. For instance, Elongin is shown to enhance the elongation of RNA pol II (Shilatifard *et al.*, 1996). Similarly, the super elongation complex (SEC) enhances

elongation and is formed by interaction of RNA pol II with AF4/FMR2 family member 1 and 4 (AFF1 or AFF4), eleven-nineteen leukemia (ENL) or ALL fused gene from chromosome 9 (AF9), Eleven-nineteen lysine-rich leukemia 1 or 2 (ELL1 or ELL2) and P-TEFb complex (Lin *et al.*, 2010; Luo, Lin and Shilatifard, 2012).

### 3.1.6 Termination

Termination of RNA Pol II at the protein-coding genes does not occur at a constant distance from the 3' end of RNAs or at a conserved site (Richard and Manley, 2009). In mammals, the distance can be a few base pairs to kilobase pairs downstream of pre-mRNA (Proudfoot, 1989). Transcription termination is coupled with the processing of RNA at the 3'end (Proudfoot, 1989; Richard and Manley, 2009; Porrua and Libri, 2015). The polyadenylation site at the end of pre-mRNA is the signal for termination (Connelly and Manley, 1988).

## 3.2 Co-transcriptional Splicing.

### 3.2.1 The Spliceosome

RNA Pol II synthesizes pre-mRNA from protein-coding genes. pre-mRNA are further processed to remove introns which are intervening the coding sequence. The removal of introns is carried out by a macromolecular machine called spliceosome. This results in the formation of mature mRNA. The core spliceosome is composed of five small nuclear ribonucleoproteins particles (snRNPs) U1, U2, U4, U5 and U6. Along with these core components, many splicing factors associate with the spliceosome at different steps of splicing (Kramer, 1996).

Splicing is a multistep process, which highly depends on the base pairing of RNA in snRNP with the pre-mRNA at the junction of exon-intron. Introns contain highly conserved motifs such as GU at the 5'splice site (5' ss), AG at the 3'ss, and branch point adenosine A positioned approximately 15-50 nucleotides upstream of the 3'ss (Figure 2A).

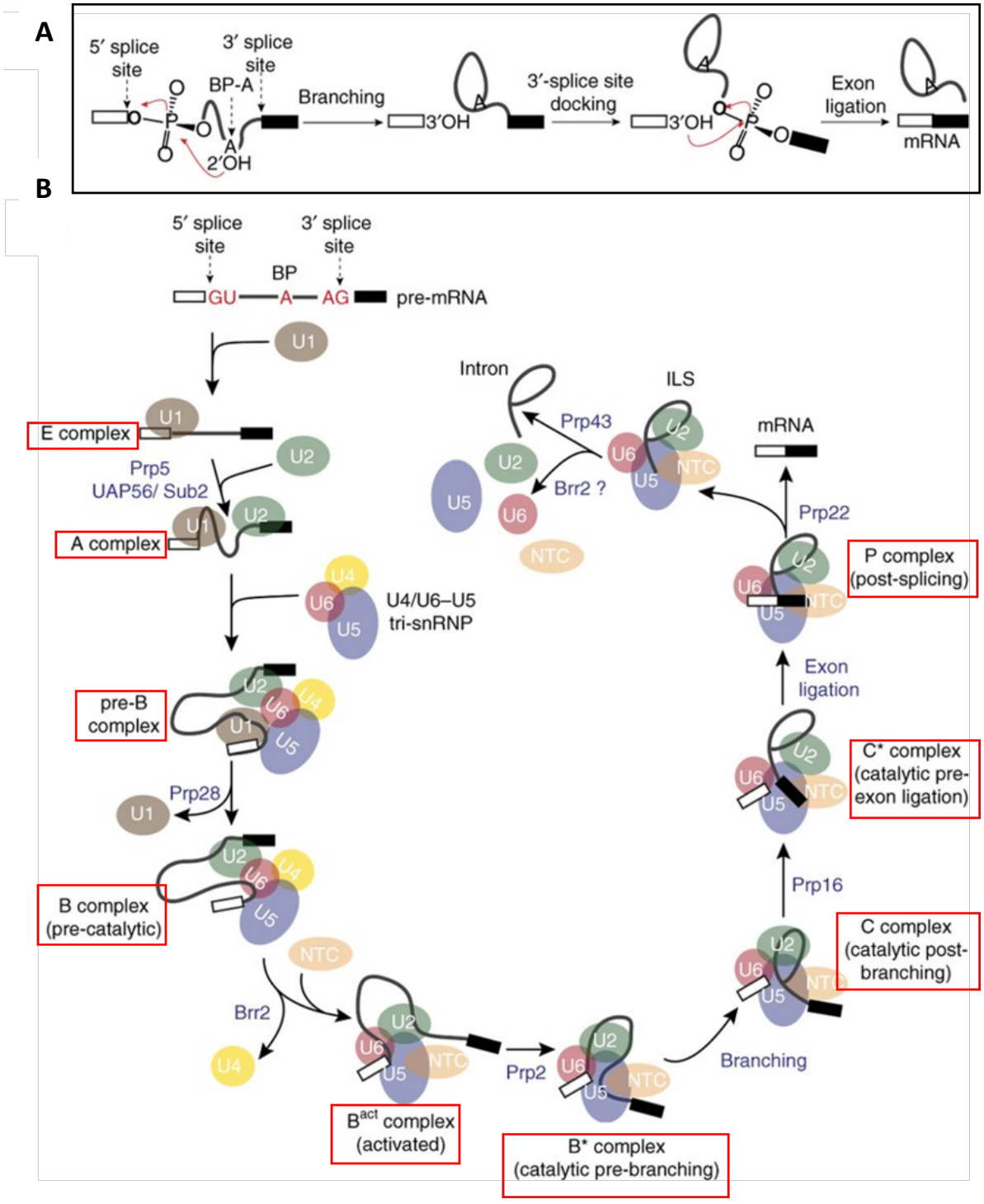



**Figure 2:** (A) Two step mechanism of pre-mRNA splicing. (B) Assembly and execution of splicing by U snRNPs. Each complex is highlighted by red box. Adapted from Sebastin et al., (Fica & Nagai, 2017).

U1 snRNP associates with the CTD of transcribing RNA Pol II (Morris and Greenleaf, 2000; Wiesner *et al.*, 2002). U1 snRNA then base paired with the 5'ss of introns and form the complex E. The interaction of U1 snRNP and 5'ss is stabilized by cap-binding proteins (Pabis *et al.*, 2013) and serine/arginine (SR) rich proteins (Staknis and Reed, 1994; Cho *et al.*, 2011). U2 snRNP interacts with the 3'ss along with associated factors, such as splicing factor 1 (SF1) and U2AFs. This U1 snRNP and U2 snRNP interaction with the specific regions of introns complete the complex E.

U2 snRNP further identifies the branch point region and interacts with U1 snRNP. The recognition of the branch point is ATP-dependent and requires Prp5 (yeast pre-mRNA-processing 5), a DExD/H helicase, and Sub2. This pre-spliceosomal complex is called complex A. Then the pre-assembled U4-U6 and U5 snRNP (tri-snRNP) is recruited to complex A via Prp28 catalytic function. This interaction of the complex A with tri-snRNP results in catalytically inactive complex B. The complex B is activated by interaction of RNA helicases like Brr2, Snu114 and Prp2. This results in the release of U4 (Sun and Manley, 1995; Raghunathan and Guthrie, 1998). Then the active complex B achieves the first catalytic step of splicing. This results in the free 5'end exon and intron-3'end exon lariat intermediate forming the complex C. This complex further undergoes ATP dependent rearrangements with the help of Prp8, Prp16 and Slu7 (Schwer and Guthrie, 1991; Frank and Guthrie, 1992; Grainger and Beggs, 2003). The second catalytic step of splicing is carried out by this complex C and results in the formation of the post-spliceosomal complex. The intron is then removed, and the exons are ligated to form the matured mRNA. Prp22 facilitates the release of spliced products from the spliceosome, and Brr2, Snu114, Prp43 disassemble the post catalytic spliceosome in an ATP-dependent manner (Schwer and Gross, 1998; Christian *et al.*, 2013; Ilagan *et al.*, 2013). The released U snRNP undergoes recycling and forms a new spliceosome (Figure 2B).

### 3.2.2 RNA Pol II Kinetics and Splicing.

The different steps of pre-mRNA maturation initiate while the transcript is still being transcribed by RNA pol II. The first evidence to support co-transcriptional

splicing came from elegant studies showing electron microscopy images of *Drosophila* chorion genes. We can see that the nascent RNA forms loops while it is still attached to the chromosome. These loops are in fact the introns that are being removed by the spliceosome. (Osheim, O.L. Miller and Beyer, 1985). (Figure 3A). This was also found in the lampbrush chromosome (Wu *et al.*, 1991). Despite this observation, it remained unclear whether transcription and splicing machinery were functionally coupled. Today ample evidence indicates that transcription by RNA pol II and maturation of the nascent RNA is kinetically and spatially coupled. The key factor that integrates the two processes is the CTD of RNA pol II. The CTD recruits mRNA processing factors to the RNA (McCracken *et al.*, 1996, 1997; Hirose, Tacke and Manley, 1999; Spector and Misteli, 1999). Phosphorylation of

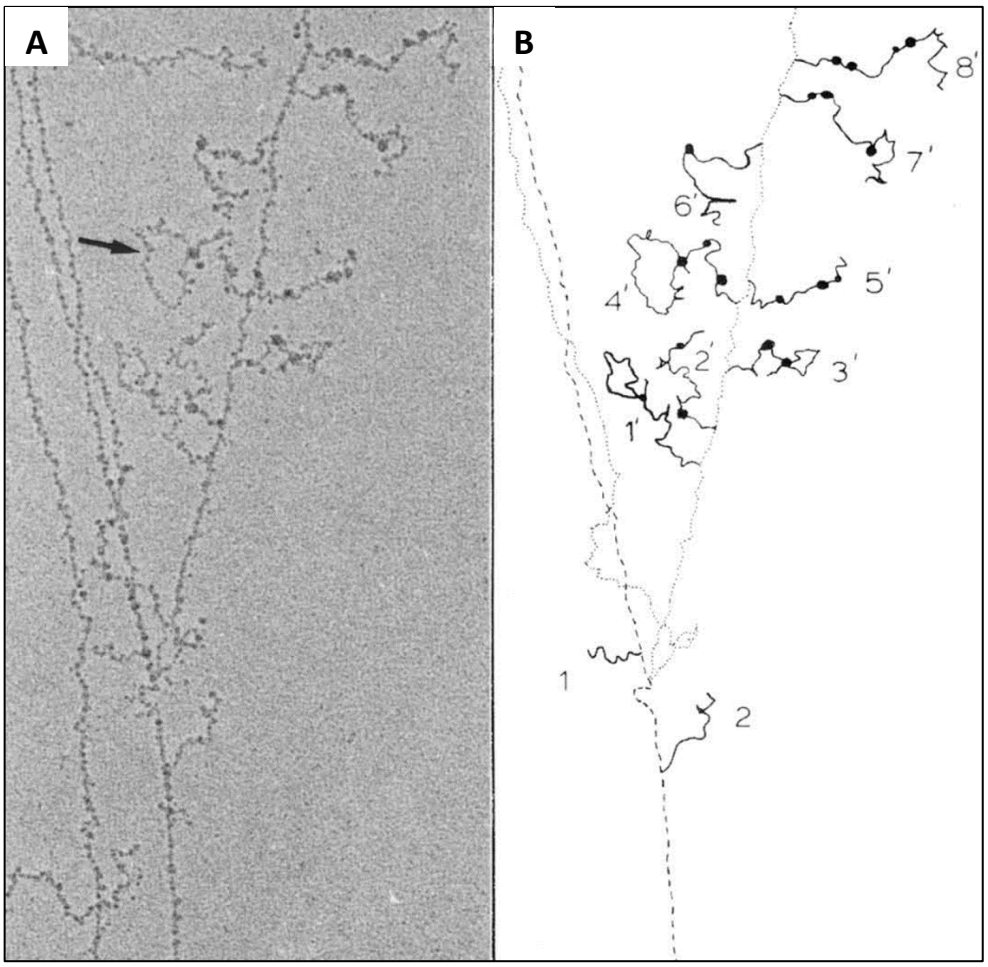



**Figure 3:** (A) Electron micrograph of *Drosophila* embryo sister chromatid transcription units. Black solid arrow indicates intron loops. (B) Tracing of EM. Numbers indicate individual transcripts from the DNA. Adapted from Beyer et al., (Beyer & Osheim, 1988)

Ser5 at the CTD is essential for the recruitment of nascent RNA capping enzymes, while Ser2 phosphorylation brings in the cleavage/polyadenylation factors at the 3'end of transcription (Spector and Misteli, 1999; Suzuki and Arita, 2012).An interesting example is the FUS protein, which is involved in the pathogenesis of amyotrophic lateral sclerosis (ALS). FUS interacts with the CTD of RNA pol II and maintains Ser2 phosphorylation. FUS also interacts with U1 snRNP and link it to the CTD (J.C. *et al.*, 2012; Yu and Reed, 2015).

Apart from the CTD of RNA Pol II, the mediator complex can also recruit splicing factors to pre-mRNA. Interestingly, this observation suggests that the mediator-promoter axis can determine the splicing outcome of a pre-mRNA (Cramer *et al.*, 1997; Huang *et al.*, 2012).

The kinetic coupling is based on "first come, first served" model. According to this model, the upstream element in the transcript gets recognised before the downstream one when RNA pol II is transcribing slow. When Pol II is transcribing fast, both upstream and downstream elements are concomitantly available to the spliceosome machinery and compete with other for their recognition by the spliceosome (Aebi and Weissman, 1987; Howe, Kane and Ares, 2003). In this case, stronger elements, whether situated upstream or downstream, will be favoured by the splicing machinery. This model suggests that the elongation rate of RNA pol II influence the alternative splicing by modulating the availability of upstream and downstream 3'splice sites (de la Mata *et al.*, 2003; Schor *et al.*, 2009; Braberg *et al.*, 2013). The kinetics of Pol II can also alter the recruitment of positive and negative effectors of splicing to their cis-acting elements, and this might result in exon skipping. One such example is ETR-9, a negative splicing factor that competes with U2AF65, and slow elongation enhances CFTR exon nine skipping by favouring ETR-9 binding (Dujardin *et al.*, 2014).

### 3.2.3 CDK9/CYCT

The P-TEFb complex contains cyclin-dependent kinase 9 (CDK9) and Cyclin-T. CDK9 is stable only when it forms a heterodimer with cyclin T. The newly synthesized CDK9 is folded properly with the help of heat shock proteins Hsp70

and Hsp90/Cdc37. Then CDK9 is transferred to cyclin to form a stable P-TEFb complex (O'Keeffe *et al.*, 2000). CDK9 undergoes autophosphorylation at Thr186 to form enzymatically active P-TEFb. The active P-TEFb must undergo further post-translational modifications and recruited to the transcription complexes to induce productive elongation (Figure 4). However, P-TEFb which is not recruited neither to the transcription machinery nor 7SK, can actively phosphorylate targets like pRb and p53 in the nucleus (Graña *et al.*, 1994; Radhakrishnan and Gartel, 2006; Lew *et al.*, 2013). The P-TEFb Thr186-P must be dephosphorylated to be recruited to the transcription complex (Zhou *et al.*, 2001). The dephosphorylation of Thr186-P in CDK9 is catalysed by protein phosphatase PP1 (Ammosova *et al.*, 2005; Wang *et al.*, 2008). CDK9 is further modified at different residues such as Ser175-P (Ammosova *et al.*, 2011; Mbonye *et al.*, 2013), Lys44/48 acetylation (Sabo *et al.*, 2008) and autophosphorylation at Thr29 induced by Brd4 interaction (Zhou *et al.*, 2009). Importantly, during all these steps, CDK9 is enzymatically inactive, i.e. the P-TEFb complex cannot phosphorylate the CTD of RNA pol II.

The most important regulatory step of P-TEFb activity is its recruitment to the 7SK snRNP complex (Nguyen *et al.*, 2001; Yang *et al.*, 2001; Michels *et al.*, 2003; Yik *et al.*, 2003). In a cell, P-TEFb is found in 2 configurations, either inactive when associated with 7SK snRNP or active when bound to Brd4. (Yang *et al.*, 2005) Almost half of P-TEFb is associated with the 7SK snRNP complex and the other half with Brd4 (Zhou *et al.*, 2009).

### 3.2.4 Pause Release by CDK9 Activity.

The role of P-TEFb in transcription regulation was first discovered by the virtue of its sensitivity to 5,6-dichloro-1-β-d-ribofuranosyl benzimidazole (DRB). DRB is a kinase inhibitor which specifically reduces the RNA Pol II Ser2 phosphorylation (Bensaude, 2011). Later, negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF) were identified as factors which are responsible for DRB induced pausing of RNA Pol II (Wada *et al.*, 1996; Zhu *et al.*, 1997; Yamaguchi *et al.*, 1999). NELF consist of four subunits (A, B, C/D and E) and DSIF has two subunits (Spt4 and Spt5). As mentioned before, DSIF and NELF initiate pausing of RNA Pol II 50 to 80nt after TSS and stably maintain it (Yamaguchi *et al.*, 1999).

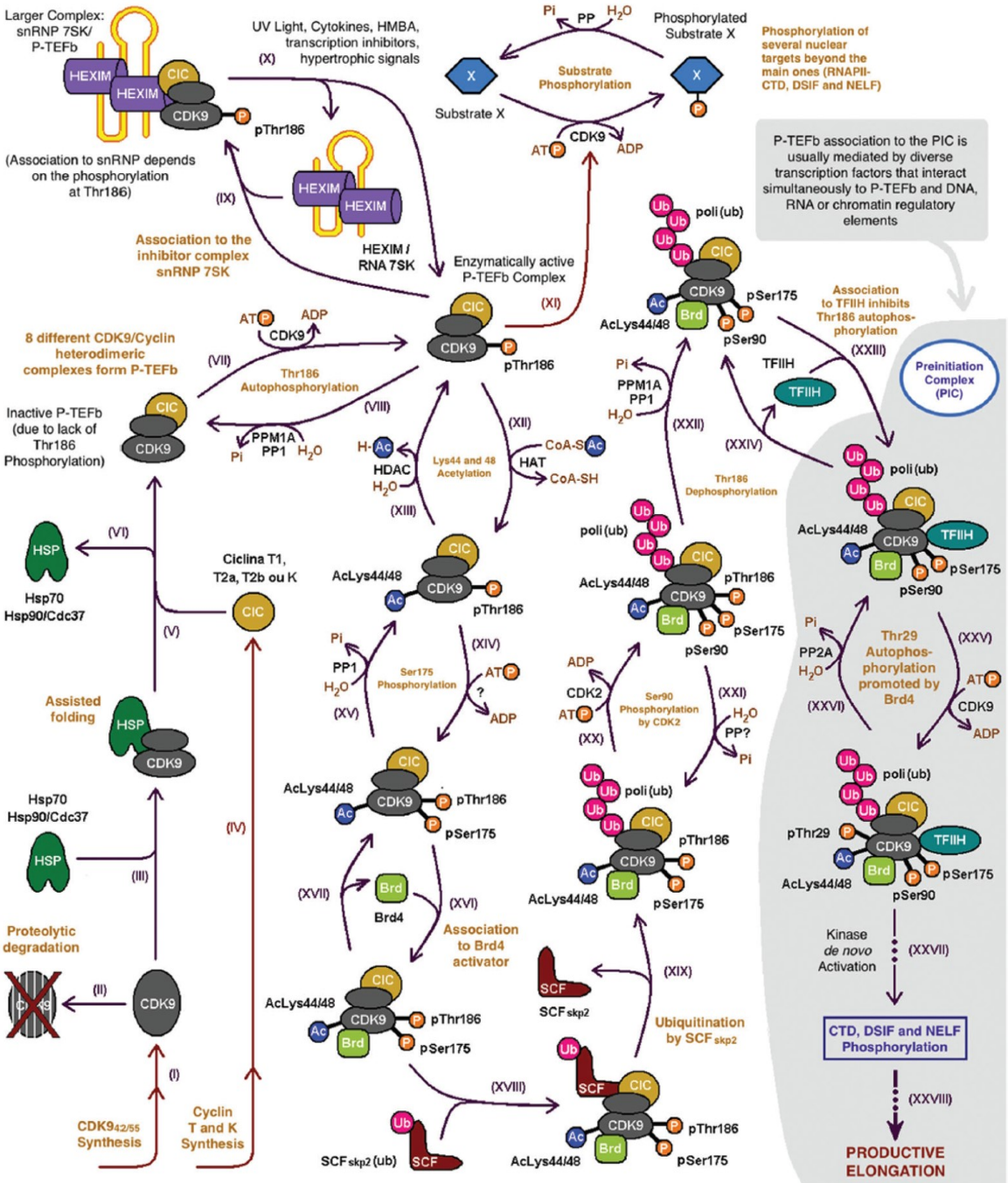



**Figure 4:** Maturation and activity of P-TEFb. The scheme shows the synthesis and maturation of P-TEFb complex and the effect of phosphorylation on the enzymatic activity of the complex. Pathways shaded in gray highlight the activity of P-TEFb in RNA Pol II pause release. Adapted from (Paparidis, Durvale, & Canduri, 2017).

NELF in particular suppresses the reactivation of paused RNA Pol II by inhibiting the entry of nucleotides into the RNA Pol II (Adelman and Henriques, 2018; Vos, Farnung, Urlaub, *et al.*, 2018). The paused RNA Pol II is released into productive elongation by the phosphorylation activity of P-TEFb. P-TEFb is recruited to the promoter via interacting with transcription factors, mediators and cofactors (Peterlin and Price, 2006; Takahashi *et al.*, 2011), including Brd4 and NF-kB. Interestingly, Human Immunodeficiency Virus (HIV) was shown to hijack this machinery to enhance transcription of its genome and its replication. HIV encodes for trans-activator transcription (Tat), which interacts with stem-loop-1 of *7SK* RNA to release HEXIM and hijack P-TEFb to enhance the viral gene transcription (Krueger *et al.*, 2010). P-TEFb complex phosphorylates Ser2 in CTD of RNA pol II, which constitutes the hallmark of active elongation. P-TEFb also phosphorylates the NELF-E subunit, and interestingly, phosphorylation of Spt5 converts the repressive DSIF complex to an activator of transcription elongation (Guo *et al.*, 2000; Yamada *et al.*, 2006; Cheng and Price, 2007; Barboric *et al.*, 2009). Apart from these inhibitory factors, a few other factors are shown to regulate pausing such as Pol II-associated factor 1 (PAF1), which enhances productive elongation by promoting the recruitment of super elongation complex (SEC) (Chen *et al.*, 2015). Recent structural studied revealed that the release of NELF from the pause site induces the interaction between PAF1 and the elongation complex (Shi *et al.*, 2015; Vos, Farnung, Boehning, *et al.*, 2018) by allowing Paf1 to interact with RNA Pol II at the first nucleosome position (Adelman *et al.*, 2005; Van Oss *et al.*, 2016). Recently, we showed that the exon-junction complex (EJC) stabilizes RNA pol II pausing likely by competing with P-TEFb for binding to the CTD of Pol II (Akhtar *et al.*, 2019).

## 3.3 The 7SK snRNP Complex Regulates RNA Pol II Pause Release.

### 3.3.1 *7SK* RNA

The *7SK* RNA was first identified among other small RNA in HeLa cells (Zieve and Penman, 1976) and later, it was cloned and sequenced from these same cells (Ullu and Melli, 1982). *7SK* accounts for almost 0.4% of total RNA (Gunning *et al.*, 1981) and is transcribed by RNA pol III (Murphy, Tripodi and Melli, 1986;

Murphy and Liegro, 1987). Earlier, *7SK* RNA was also known as K-RNA and 7-3 RNA. *7SK* RNA is found across nematodes, insects and vertebrates, but has not been identified in yeast (Marz *et al.*, 2008). The promoter of a human *7SK* gene consists of a distal sequence element (DSE), a proximal sequence element (PSE), and a TATA box (Boyd *et al.*, 1995; Kleinert, Bredow and Benecke, 2018). The octamer binding protein 1 (Oct1) binds to the DSE while selenocysteine tRNA gene transcription activating factor 1 (Staf1) and snRNA-activating protein complex (SNAPc) interact with PSE to promote *7SK* transcription (Murphy *et al.*, 1989; Jong-bok *et al.*, 1995; Schaub *et al.*, 1997). The *7SK* RNA sequence is not highly conserved between species, so its discovery in invertebrates only occurred recently by the Benecke group, which used the conserved promoter domain as a mean of identification (Marz *et al.*, 2008). Soon after the transcription of *7SK* RNA by RNA Pol III, Lupus antigen (La) protein binds to the 3'end of the RNA and protects it from nucleolytic cleavage (Maraia and Intine, 2002; Bayfield, Yang and Maraia, 2010). The newly synthesized *7SK* RNA has 4-5 U residues at the 3'end, and up to 3 U are removed, followed by addition of an A residue by the action of an adenylating enzyme (Sinha *et al.*, 2002). Concomitantly, the La protein bound to the 3'end of *7SK* RNA is replaced by La-Related Protein 7 (LARP7) (Figure 5). The secondary structure of the mature *7SK* RNA in the complex was predicted and verified experimentally. There are four stem-loop structures present in the human *7SK* RNA, and different proteins were shown to interact with

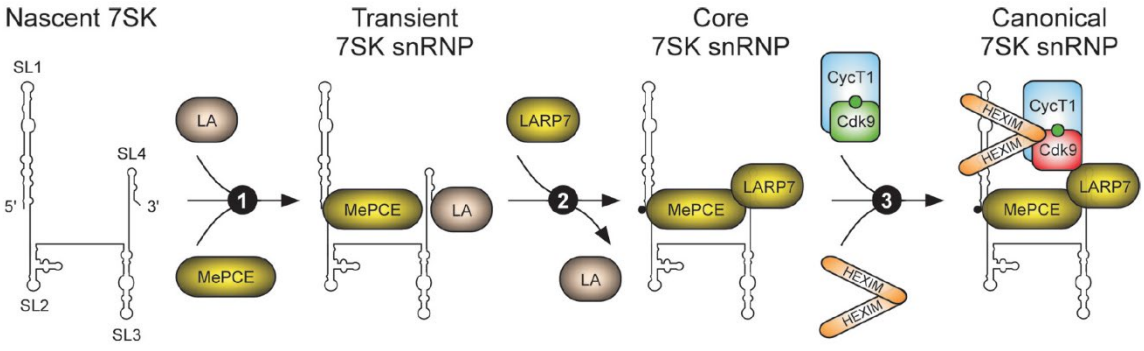



**Figure 5:** Synthesis and maturation of 7SK snRNP. The three-step process showing the folding of *7SK* RNA, association of LA and MePCE with the RNA. LARP7 replacing LA and formation of the final complex with P-TEFb and HEXIM. Adapted from (Quaresma, Bugai, & Barboric, 2016)

these structures (Reddy *et al.*, 1984; Diribarne and Bensaude, 2009; Peterlin, Brogie and Price, 2012).

### 3.3.2 LARP7

LARP7 is closest to La protein in the LARP family of proteins. LARP7 contains three domains La Motif (LaM), RNA recognition motif 1 (RRM1) and RRM2. The La module includes LaM, and the RRM1 in the N-terminus of the protein and this module is conserved in LARPs (1,4,6 and 7) (Figure 6) (Deragon and Bousquet-Antonelli, 2015). The specificity of LARP7 binding arises from the unique C-terminal RRM2 (Bousquet-Antonelli and Deragon, 2009; Bayfield, Yang and Maraia, 2010; Conte *et al.*, 2012). Indeed, the structural studies showed that RRM2 interacts with stem-loop four at the 3'end of *7SK* RNA (Eichhorn, Chug and Feigon, 2016). The structural studies of the La module also confirms that the interaction with 7SK RNA 3'end is essential for the stability of the 7SK snRNP (Uchikawa *et al.*, 2015a). Interestingly, in contrast to the full LARP7, a C-terminally truncated version lacking RRM2, partially rescues the tRNA maturation defect of La mutant, confirming that the RRM2 domain switch the specificity of LARP7 towards *7SK* binding (He *et al.*, 2008).

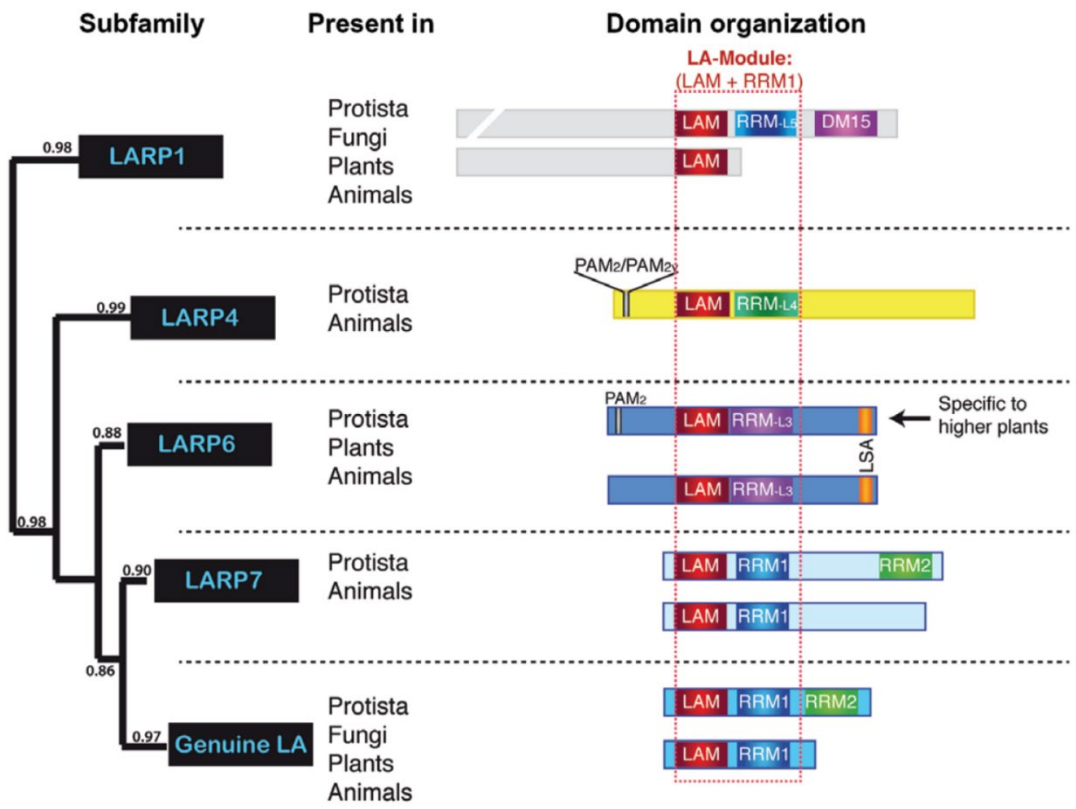



**Figure 6:** Phylogenetic and protein domain organization of La-Motif (LAM) containing proteins. LA and LARP 1,4,6 contain RNA recognition motif (RRM) close to LAM. Only in LA and LARP7 a second RRM2 is found towards the C-terminal end (Deragon & Bousquet-Antonelli, 2015).

### 3.3.3 MePCE

Methyl Phosphate Capping Enzyme (*MePCE*) was first identified in *Drosophila* as Bicoid-interacting protein 3 (*Bin3*). MePCE is a highly conserved protein, and it is essential for axis formation during *Drosophila* embryonic development (Singh, Morlock and Hanes, 2011). MePCE methylates a specific set of RNA in the cell, which include 7SK and U6 snRNA (Jeronimo *et al.*, 2007). It has a highly conserved AdoMet-binding domain which interacts with the 5'end of *7SK* RNA and mono-methylated 5'γ-phosphate (Shimba and Reddy, 1994). The capping by MePCE protects the *7SK* RNA from endonucleolytic degradation (Hamm *et al.*, 1990). While MePCE binds both *7SK* and U6 RNA which are transcribed by RNA pol III, its depletion only affects the stability of *7SK* RNA (Jeronimo *et al.*, 2007). Apart from capping *7SK* RNA, MePCE interacts with LARP7 to stabilize the complex (Xue *et al.*, 2009). Nevertheless, MePCE can interact with *7SK* RNA

independently of its interaction with LARP7, as it is already bound when the La protein is still bound to *7SK* (Muniz, Egloff and Kiss, 2013) (Figure 5). From these studies, we can hypothesise a mechanism that the newly synthesized *7SK* is bound by La at its 3'end and by MePCE at the methylated 5'end. Soon after, La is exchanged with LARP7, which can make direct contact with MePCE (Figure 5). The mechanism of La -*7SK*- LARP7 assembly is not understood clearly. Also, MePCE losses its capping activity when it interacts with LARP7. Interestingly, MePCE is the only known component of the 7SK snRNP, which can also interact with P-TEFb in RNA independent manner (Yik *et al.*, 2003).

### 3.3.4 HEXIM

HEXamethylene-bis-acetamide-Inducible protein in vascular smooth Muscle cells (HEXIM) was first identified as a factor involved in differentiation and development (Cai *et al.*, 2002; Aikawa *et al.*, 2014). HEXIM involvement in the regulation of transcription via P-TEFb was revealed later (Michels *et al.*, 2003; Yik *et al.*, 2003). Most of the mammalian genomes contain two paralogs, HEXIM1 and HEXIM2. It is HEXIM1 that binds to the 5'stem-loop of *7SK* RNA as a homodimer. HEXIM2 can compensate the function in HEXIM1 knockdown cells (Yik *et al.*, 2005). HEXIM consists of six regions, which include, an unstructured N-terminal domain, an Arginine-Rich Motif (ARM), a central region, a basic rich region (BR), a conserved PYNT sequence, and an acidic region (AR1 and AR2) (Michels *et al.*, 2003, 2004; Zhou *et al.*, 2004; Barboric *et al.*, 2005; Marz *et al.*, 2009). The PYNT sequence is essential for HEXIM to bind and inactivate P-TEFb (Michels *et al.*, 2004; Kolesnikova *et al.*, 2016). The Thr186 phosphorylation of CDK9 is essential for the interaction of HEXIM with the 7SK complex. Interestingly, the phosphorylation state of certain residues in the helix α1 of HEXIM determines the rate of activity of P-TEFb. For instance, it was shown that Y271F mutation in the α1 helix leads to a reduced level of active P-TEFb (Contreras *et al.*, 2007; Chance *et al.*, 2015). Recently, it was shown that HEXIM could also interact with NEAT1 nuclear non-coding RNA and participate in innate immunity signalling. In this case, the core 7SK snRNP components are not associated with HEXIM-NEAT1 complex (Morchikh *et al.*, 2017), indicating that HEXIM can interact with different ncRNA

in the nucleus and carry out multiple cellular functions (Michels and Bensaude, 2018).

## 3.4 Summary of 7SK snRNP Function.

The *7SK* RNA is a scaffold, which requires MePCE and LARP7 for its stability. The matured 7SK snRNP consists of *7SK* RNA, MePCE (mono-methylate 5'γ-phosphate) and LARP7 (3'end). HEXIM binds to this complex via *7SK* RNA and brings in P-TEFb. The 7SK snRNP/HEXIM/P-TEFb complex is inactive and is the most prevalent form of 7SK snRNP. P-TEFB can be released by the action of multiple factors, such as BRD4, NF-kB or Tat. It was recently demonstrated that the core 7SK snRNP could exist in other complexes. For instance, it can interact with Little Elongation Complex (LEC) and regulates transcription of small nuclear (sn) and small nucleolar (sno) RNA, independently of P-TEFb. Thus components of the 7SK snRNP complex have separate functions apart from the regulatory activity of P-TEFb. As discussed above, HEXIM can also form an independent complex with NEAT1 RNA in the nucleus, and MePCE can methylate other RNAs. Lastly, LARP7 is involved in ribosome biogenesis in the nucleolus, independently of its function within the 7SK complex (see also below) (Slomnicki *et al.*, 2016).

## 3.5 The *Drosophila* 7SK snRNP Complex

For a long time, it was thought that the 7SK module was specific to vertebrates. *7SK* gene promoter motif similarity search provided evidence that the *7SK* RNA also exists in invertebrates (Gruber *et al.*, 2008; Marz *et al.*, 2008). The *7SK* RNA and its associated proteins in *Drosophila melanogaster* were characterised recently (Nguyen *et al.*, 2012a). Nguyen et al. identified a 444 nt RNA, which behaves like *7SK* RNA. Also, *Larp7* was identified by homology search, and its interaction with *7SK* RNA was validated. *MePCE* was already known to be present in *Drosophila* as *Bin3*, and indeed it interacts with *7SK* RNA.

Interestingly, only one copy of *Hexim* is present in *Drosophila,* and it produces two isoforms. Nguyen et al., produced antibodies against *Larp7* and *Hexim*, and by

immunostaining, showed that these proteins were highly expressed in all developmental stages. Before our study, this was the only report that characterized the 7SK snRNP complex in *Drosophila*. The authors showed that the knockdown of *Hexim* leads to severe developmental defects. As *HEXIM* in human cells has multiple functions, this phenotype may be unrelated to the 7SK complex function. Therefore to understand the function of the 7SK snRNP complex *in vivo* and more globally the function of pausing in gene regulation during development, it is essential to delineate the role of *7SK* RNA and other individual components.

### 3.5.1 The 7SK snRNP Complex is Essential for the Development in Mouse and Zebrafish.

The 7SK snRNP complex regulates the activity of P-TEFb, which in turn regulates release of RNA Pol II pausing. The knock out of *Larp7* in mouse is lethal during embryogenesis and animals shows severe phenotypes. Head development is particularly impaired compared to other organs. The loss of 7SK snRNP in these mutants also lead to growth arrest of primordial germ cells (PGCs) (Okamura *et al.*, 2012a). Knockdown of *Larp7* in zebrafish using morpholinos lead to defect in head development. Similar to the mouse *Larp7* knockout, its depletion leads to lethality in embryogenesis (Barboric *et al.*, 2009). Thus observations from *Drosophila*, zebrafish and mouse strongly suggest that *Larp7* is essential for development.

### 3.5.2 The 7SK snRNP Complex Role in Cancer.

Based on studies in different cell lines, the 7SK snRNP complex is considered to be a tumour suppressor factor. *LARP7* is downregulated in breast cancer cells, resulting in increased active P-TEFb level, which further promotes epithelial-mesenchymal transition and metastasis (Ji *et al.*, 2014). In HEK 293T cells, overexpression of *7SK* RNA is sufficient to induce apoptosis (Keramati *et al.*, 2015). Interestingly the comparison of *7SK* RNA expression level in different cancer cell lines and stem cells suggest that the level of *7SK* RNA is down regulated in these cells compared to differentiated cells. (Abasi *et al.*, 2016). Decreased level of LARP7 is also observed during early gastric tumorigenesis, and it is hypothesized that the hyper activation of P-TEFb promotes tumorgenesis (Cheng

*et al.*, 2012). Lastly, the increased level of HEXIM was observed in erythroleukemia and neuroblastoma cell differentiation. (Napolitano *et al.*, 2005). All these observations support a role for Larp7 as a tumour suppressor via attenuating transcription.

### 3.5.3 Alazami Syndrome (AS).

In 2012, Alazami et al., first identified *LARP7* loss of function mutation in a Saudi family. In contrast to the essential role of Larp7 in mouse and zebrafish, the lack of LArp7 is viable in humans. However they display symptoms of primordial dwarfism (PD). PD is a genetic and a heterogeneous condition, and in these patients, they found severe intellectual disability and distinct facial features (Alazami *et al.*, 2012) (Figure 6). The ensemble of these phenotypes was named Alazami syndrome (AS). A few years later, the second case of AS was reported in a two-year-old Northern European/Caucasian female. A novel mutation in the *LARP7* gene was discovered by whole exome sequencing of this patient. (Ling and Sorrentino, 2016).

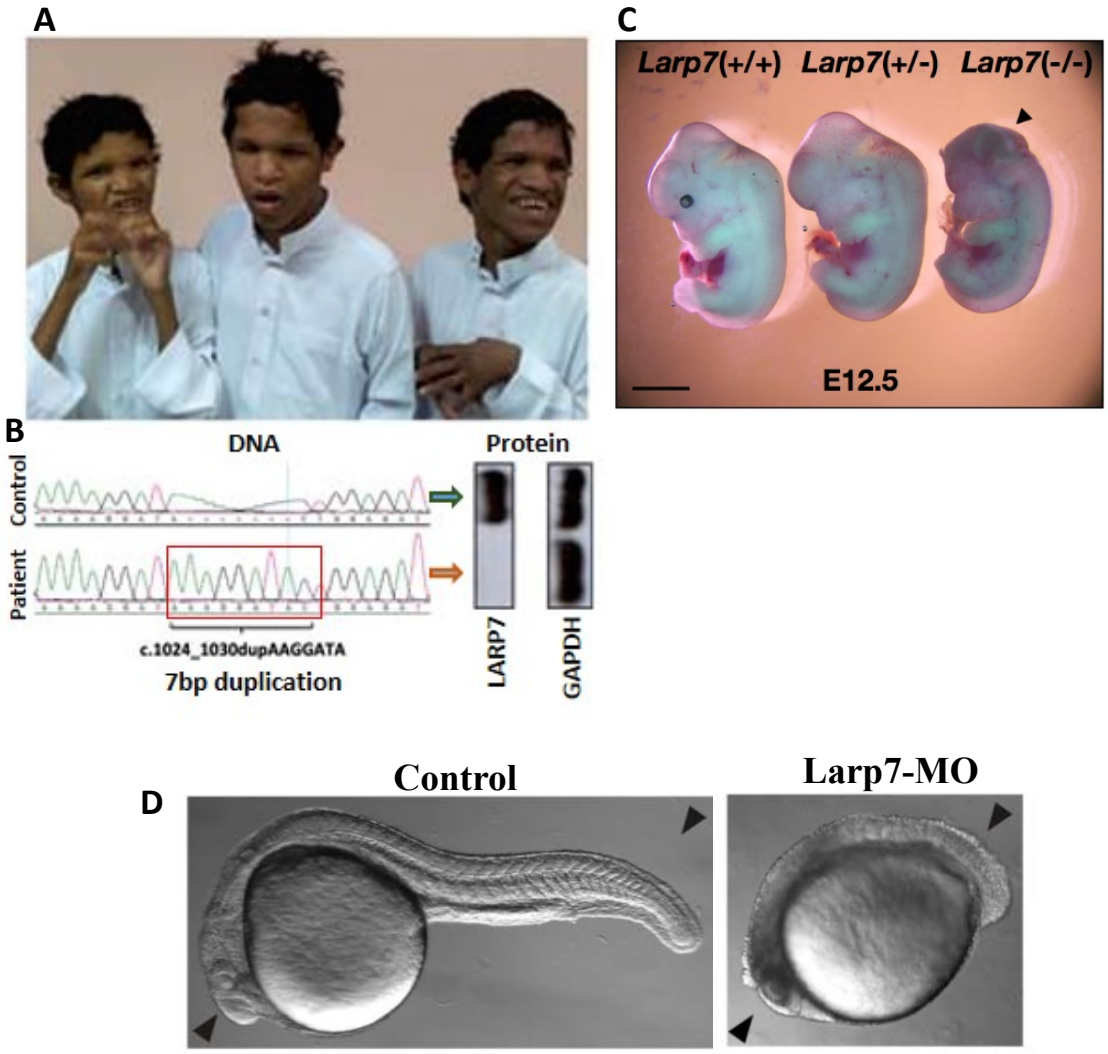



**Figure 7:** (A) Representative images of AS patients from Saudi family. (B) DNA chromatogram of 7bp duplicated region in *LARP7* locus. Immunoblot showing loss of LARP7 protein (Alazami et al 2012). (C) *Larp7* KO mouse showing impaired head development (black arrow head) and reduced overall development of the embryo (Okamura et al, 2012). (D) Knock down of *Larp7* in zebrafish embryo lead to head and vertebrate developmental defect (black arrow head) (Barboric et al., 2009).

In the same year, another case report confirms *LARP7* mutation in a six-year-old Dutch boy and a two and a half-year-old Saudi boy (Hollink *et al.*, 2016). Other cases of Alazami syndrome were reported in a two-year-old Japanese boy (Kitajima *et al.*, 2018) and two Algerian sisters of age 22 and 26. In all these cases, the common observation is that their motor development was delayed and impaired. They all started walking at the age of 2-3 years, which is significantly delayed from the average duration of 12-15 months.

In summary, AS is an autosomal recessive disease caused by loss of function mutation in the *LARP7* gene. The mechanism underlying this syndrome is currently poorly understood. A recent report showed that AS cases have shortened telomere (Holohan *et al.*, 2016), but it is not known whether this a cause or a consequence of the disease. The wide range of clinical studies from different geographical locations suggests that there could be more cases of AS that was perhaps mistakenly categorized for some other syndromes, such as Silver-Russel Syndrome, Meier-Gorlin Syndrome and Seckel syndrome (Deleuze *et al.*, 2018). Thorough understanding of the mechanism underlying this syndrome is still awaiting and would be essential to provide new treatment avenues.

### 3.5.4 *Drosophila* as a Model to Understand the Role of 7SK snRNP Complex *in vivo*.

In order to get new insight into the poorly characterized role of the 7SK snRNP complex during development, I study its function in *Drosophila melanogaster*. *Drosophila* has been used as a model organism for over a century to study a range of diverse biological processes, including embryonic development, learning, behaviour, and ageing. As most of the fundamental biological mechanisms and signalling pathways that control development and survival are conserved throughout evolution, there are many examples that demonstrate the utility of using *Drosophila* as a model system. Mechanistic details of the genetic and molecular regulation of cellular processes can be thus established in the fruit fly and then transferred to other organisms.

## 3.6 Development of *Drosophila*

*Drosophila melanogaster* undergoes four stages of the life cycle as follows: egg, larvae, pupae and fly. After fertilization of an egg, the development starts with cell division. The embryo undergoes 17 stages over 24 hours. Then the embryo hatch into larvae. There are three stages of larvae, L1, L2 and L3. At the L1 stage, larvae start feeding for the first time and thrive until the L3 stage. This is the only growth phase in the whole lifecycle of *Drosophila* (Figure 8). Then L3 larvae pupate and undergo many changes by apoptosis and regeneration to form the adult fly. The whole life cycle from embryo to adult fly takes about ten days at 25˚C with 60% humidity, and the adult fly can live up to 80 days at this temperature. An increase of temperature to 29˚C decreases the lifespan, while a decrease at 18˚C prolongs it.

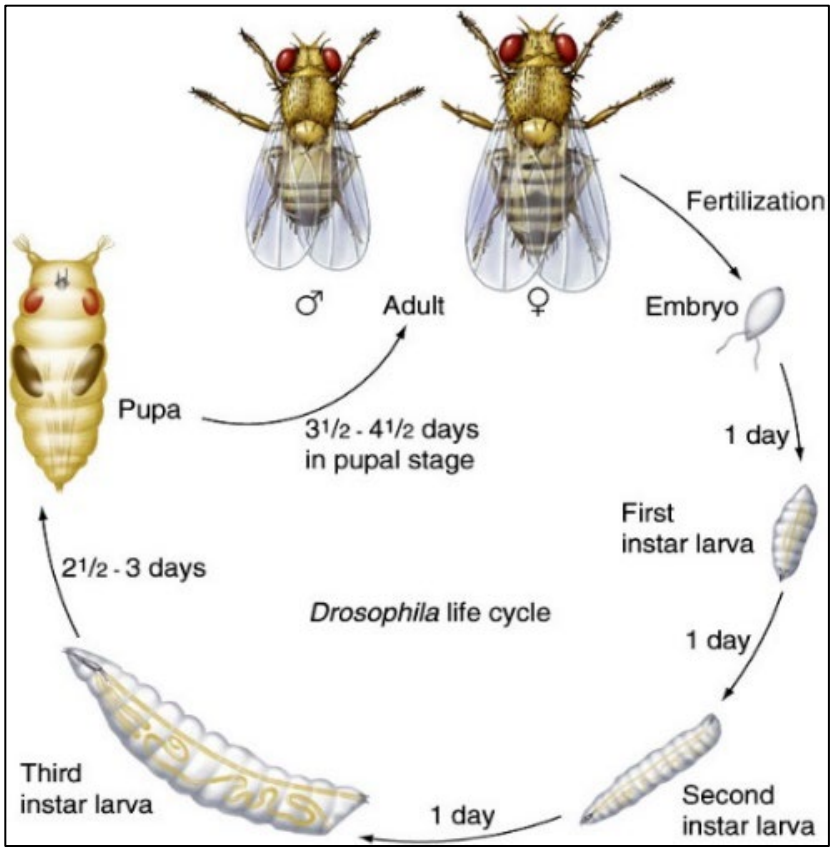



**Figure 8:** Life cycle of *Drosophila melanogaster*

### 3.6.1 Locomotion of *Drosophila* larvae

Larvae are about 7-8 mm in length at the L3 stage. The locomotion of the larva is one of the well-studied behaviours. I will first introduce the different locomotion pattern of L3 larva and then the neuronal circuit necessary for the locomotion.

*Drosophila melanogaster* larvae display different motion named as feeding, bending, turning and peristalsis (Green, Burnet and Connolly, 1983). The pattern of larval locomotion depends on different environmental cues. The primary and essential need for the development of the larvae is food, and it is generally attached to it (Gomez-Marin and Louis, 2012). For survival, the larva also responds to the predators by the increased exploration of the environment (De La Flor *et al.*, 2017). The larva also exhibits a response to light (Keene and Sprecher, 2012). Interestingly, larvae in the crowded laboratory condition also show cannibalism (Vijendravarma, Narasimha and Kawecki, 2013). Another recent report suggests that the pheromone 7,11-heptacosadiene (7,11-HD) in the eggshell's wax protect embryos from feeding by larvae (Narasimha *et al.*, 2019), indicating the larval response to pheromones.

A larva has ten body segments, which include seven abdominal (A1-A7) and three thoracics (T1-T3) segments. Also, it has an anterior mouth or head and posterior tail-end (Keshishian, 2002). Each segment is further divided into hemi-segments. There are 30 muscles in hemisegment. The muscle number varies at the terminals and a few segments of the larva. The muscles are further classified based on their orientation to the body of larva. The exterior most muscles run transverse (dorso-ventral axis) and the interior muscles run longitudinal (anteroposterior axis). During the peristaltic movement of larvae, a wave of muscle contraction runs from posterior to the anterior (Kohsaka *et al.*, 2012) (Fig X a). The contraction of longitudinal and transverse muscles results in longitudinal contraction and circumferential contraction. This wave passes to the next segment until it reaches the head where the mouth hook is used to anchor to the surface to move forward. This one cycle of contraction and relaxation of muscles is called peristalsis. In a normal condition, one peristalsis takes approximately one second on 1% agar surface — the rhythmic cycle of peristalsis results in the forward movement of the larva.

### 3.6.2 Motor Circuits and Larval Locomotion.

The locomotion of larvae is primarily regulated by motor and interneurons. During the development of *Drosophila* in the late embryonic stage, all the neurons required in larva are formed (Landgraf and Thor, 2006b). The larval central nervous system (CNS) is made of a pair of brain lobes, and a ventral nerve cord (VNC) consisting of thoracic and abdominal ganglia. The brain lobe and VNC are interconnected by different neurons (Cardona, Larsen and Hartenstein, 2009). Like the muscle segments, three thoracic and eight abdominal neuromeres are present. The motor neurons are located in the dorsal and ventral region of the VNC neuromere segment and target corresponding muscle segments (Figure 9). There are different types of motoneurons targeting different muscles.

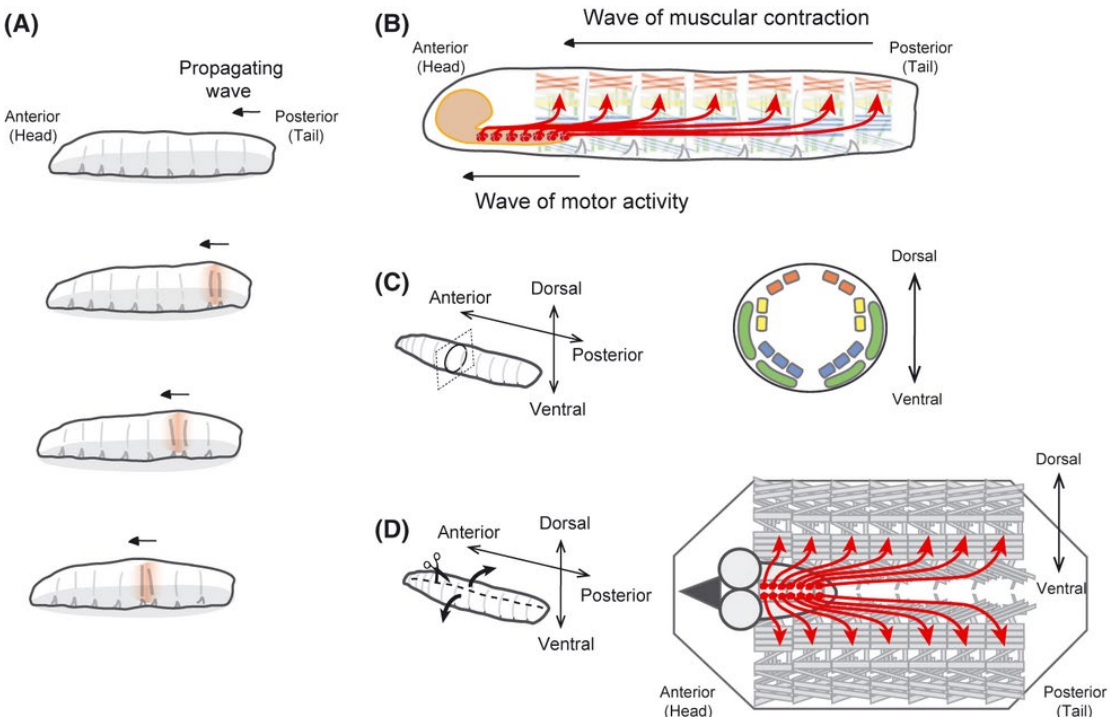



**Figure 9:** Larval locomotion and motor circuit. (A) Cartoon depicting the forward peristaltic movement of a larva. (B) Larval motor circuit, red line represent motor neuron targeting different muscles. (C-D) Neuromuscular junction (NMJ) preparation to study the motor neurons. Adapted from Kohsaka et al., (Kohsaka et al., 2012)

At the end of embryogenesis, all the different types of motoneurons are generated. Approximately around 38 motor neurons innervate the muscles. Motor axons from the VNC neuromere project into six nerves: (1) transverse nerve (TN), (2) intersegmental nerve (ISN), and (3) segmental nerves SN (SNa, SNb, SNc, and SNd) that project dorsally to ventral. (Landgraf *et al.*, 2018). The ISN further give rise to three nerve branched ISN (inverted, ISNb, and ISNd, which target the internal muscles). The SNa and SNc of segmental nerve targets external muscles. The motoneurons are first classified into ventral projecting motoneurons (vMN) and dorsally projecting motoneurons (dMN). Table 1 shows the classification of the larval motor neurons with its target muscles and its embryonic origin (Kim, Wen and Jan, 2009). During the late stages of embryogenesis, the motor neurons are formed, and the growth cone of axon terminals forms specific contact with muscles and develop into presynaptic terminals (Landgraf and Thor, 2006a). By the end of embryogenesis, the presynaptic terminal is matured into functional neuromuscular junction (NMJ). At this point, the NMJ is made of a few synaptic boutons.

The boutons are a spherical structure, which is made of synapses. The bouton contains many neurotransmitter release site called active zones, and each active zones are opposed by glutamate receptors present on the muscle (Menon, Carrillo and Zinn, 2013). From L1 larval stage to L3 larval stage, the muscles develop up to 100 fold, and the boutons also develop to maintain the synaptic drive. (Schuster *et al.*, 1996; Luo, Lin and Shilatifard, 2012).

As mentioned before there are different types of motoneurons targeting different muscles. In my thesis, I focused on RP2 motoneurons. This motoneuron targets dorsal muscles and NMJ of dorsal muscle 4.

Embryonic origin of larval motor neurons.

| Nerve | Embryonic motor neuron/neuroblast[a] | Target muscle(s) | Larval motor neuron[b] (*n*) | Target muscle(s) |
|---|---|---|---|---|
| ISN | aCC | 1 | MN1-Ib (12) | 1 |
| | U (NB7-1) | 2 | MN2-Ib (6) | 2 |
| | U (NB7-1) | 3 | MN3-Ib (1) | 3 |
| | U (NB7-1) | 4 | MN4-Ib (1) | 4 |
| | U (NB7-1) | 9 | MN9-Ib (24) | 9 |
| | U (NB7-1) | 10 | MN10-Ib (3) | 10 |
| | NB3-2 | 11 | MN11-Ib (6) | 11 |
| | NB3-2 | 18 | MN18-Ib (7) | 18 |
| | NB3-2 | 19 | MN19-Ib (2) | 19 |
| | NB3-2 | 20 | MN20-Ib (2) | 20 |
| | RP2 | Dorsal | MNISN-Is (6) | 1, 2, 3, 4, 9, 10, 18, 19, 20 |
| | VUM-dorsal | Dorsal | VUM (3) | Dorsal |
| ISNb /ISNd | RP3 | 6, 7 | MN6/7-Ib (4) | 6, 7 |
| | RP5 (NB3-1) | 12 | MN12-Ib (3) | 12 |
| | V (NB5-2) | 12 | MN12-III[c] (0) | 12 |
| | RP1 (NB3-1) | 13 | MN13-Ib (3) | 13 |
| | RP4 (NB3-1) | 13 | Not identified | |
| | NB4-2 | 14, 30 | Not identified | |
| | Unknown | | MN14-Ib (3) | 14 |
| | NB4-2 | 28 | MN28-Ib (1) | 28 |
| | Unknown | | MN30-Ib (6) | 30 |
| | NB7-1 | 15, 16 | MN15/16-I (6) | 15, 16 |
| | NB7-1 | 17 | Not identified | |
| | Unknown | | MN15/16/17-Ib (3) | 15, 16, 17 |
| | Unknown | | MNISNb/d-Is (4) | 6, 7, 12, 13, 14, 15, 16, 20 |
| | Unknown | | MNISNb/d-II (4) | 12, 13, 14, 15, 16, 17, 30 |
| | VUM-ventral | Ventral | VUM (2) | Ventral |
| SNa | NB2-2 | 21 | Not identified | |
| | NB2-2 | 22 | Not identified | |
| | NB3-2 | 23 | Not identified | |
| | NB3-2 | 24 | Not identified | |
| | Unknown | 5 and/or 8 | MN5/8 (?) | |
| | Unknown | | MN21/22-Ib (1) | 21, 22 |
| | Unknown | | MN22/23-Ib (2) | 22, 23 |
| | Unknown | | MN23/24-Ib[c] (0) | 23, 24 |
| | VUM-lateral | Lateral | VUM (1) | Lateral |
| SNc | NB4-2 | 26 | Not identified | |
| | NB4-2 | 27 | Not identified | |
| | NB4-2 | 29 | Not identified | |
| | Unknown | | MNSNc (5) | 26, 27, 29 |

[a] Based on studies from Landgraf et al., 1997 and Schmid et al., 1999.
[b] Based on nomenclature from Hoang and Chiba, 2001.
[c] Not identified in this study.

**Table 1:** Classification of motor neuron based on the target muscles in embryo and larval stage. Adapted from Kim et al., (M. D. Kim et al., 2009).

# 4 Aim of This Study

The RNA Pol II promoter-proximal pausing is one of the key regulatory steps in the transcription of developmental and stress response genes. The 7SK snRNP complex is a major regulator of promoter-proximal pausing by interacting with and sequestering P-TEFb. Despite the understanding of the biochemical function of the components of 7SK snRNP complex and P-TEFb, very little is known about the function of this complex *in vivo*. Previous studies in *Drosophila*, zebrafish and mouse attempted to address its function but often by knocking down/out only one component of the core complex, and never the *7SK* RNA itself, which is complicated in vertebrates, as it is encoded by multiple gene copies. Each component of the 7SK complex has individual biological function apart from this complex. Thus it is difficult to ascertain specific phenotypic defects to the 7SK complex itself. Therefore, currently, there is no definitive answer about the biological role of the 7SK snRNP complex. To address this, I decided to investigate its function in *Drosophila melanogaster*. This model presents the advantage of having established genetic tools as well as well-known development and behaviour. Furthermore, only one gene copy of *7SK* is present in the genome, which renders possible its inactivation. By studying its function, we aimed to address several fundamental questions:

Is the 7SK snRNP complex essential during development?

Are all cell types dependent on this complex?

More generally, what is the role of promoter-proximal pausing in vivo?

Can we explain the phenotypic specificity of the Alazami syndrome using this model organism?

# Results

# 5 RESULTS

## 5.1 *Larp7* is Differentially Expressed during *Drosophila* Development

### 5.1.1 *Larp7* Expression during Development.

The 7SK snRNP complex regulates transcription via sequestration of P-TEFb. The *in vivo* dynamics and specificity of this complex is still unclear. To get more insights, we first examined the expression of *larp7* mRNA in various developmental stages (Figure 10A). *Larp7* mRNA is expressed in all developmental stages but shows differences in its abundance. It is maternally deposited in the embryos. The expression level is relatively enriched in the ovaries and embryos until 2 hours (hrs) (Figure 10A). Then its level drops dramatically. The lowest expression during all stages of development is observed between 10hrs and 24hrs of embryogenesis. At the L1 stage, the level gradually increases and reaches a stable expression until eclosion. Overall, expression of *Larp7* in whole female flies is higher than in the male flies, likely due to the high expression in ovaries. Interestingly, the adult male head has higher expression of *Larp7* than the female head. Possibly, the *Larp7* expression is regulated in an organ-specific and sex-specific manner.

We next used the CRISPR/Cas9 system to insert enhanced Green Fluorescent Protein (eGFP) at the 5'end of the *Larp7* gene (Figure 10B). The insertion was validated by qPCR on genomic DNA and by western blot with an anti-GFP antibody (Figure 10C). The recombinant flies did not show developmental or behavioural changes compared to wild-type flies, suggesting that the Tag does not interfere with Larp7 function. Next, we mapped the expression of eGFP-Larp7 in different developmental stages. Irrespective of the cell types, eGFP-Larp7 localizes in the nucleus and is highly enriched in the nucleolus (Figure 10D). The eGFP-Larp7 is ubiquitously expressed in all developmental stages (Figure 10E). However, we observed a differential expression level between cell types. Particularly in the late embryonic stages, eGFP-Larp7 is enriched in a group of midline cells in each segment (Figure 11A). The neuroblasts in L3 larvae also have enriched level

(Figure 10D). Thus, these experiments reveal that Larp7 is ubiquitously expressed but exhibits differential expression between cell types and developmental stages.

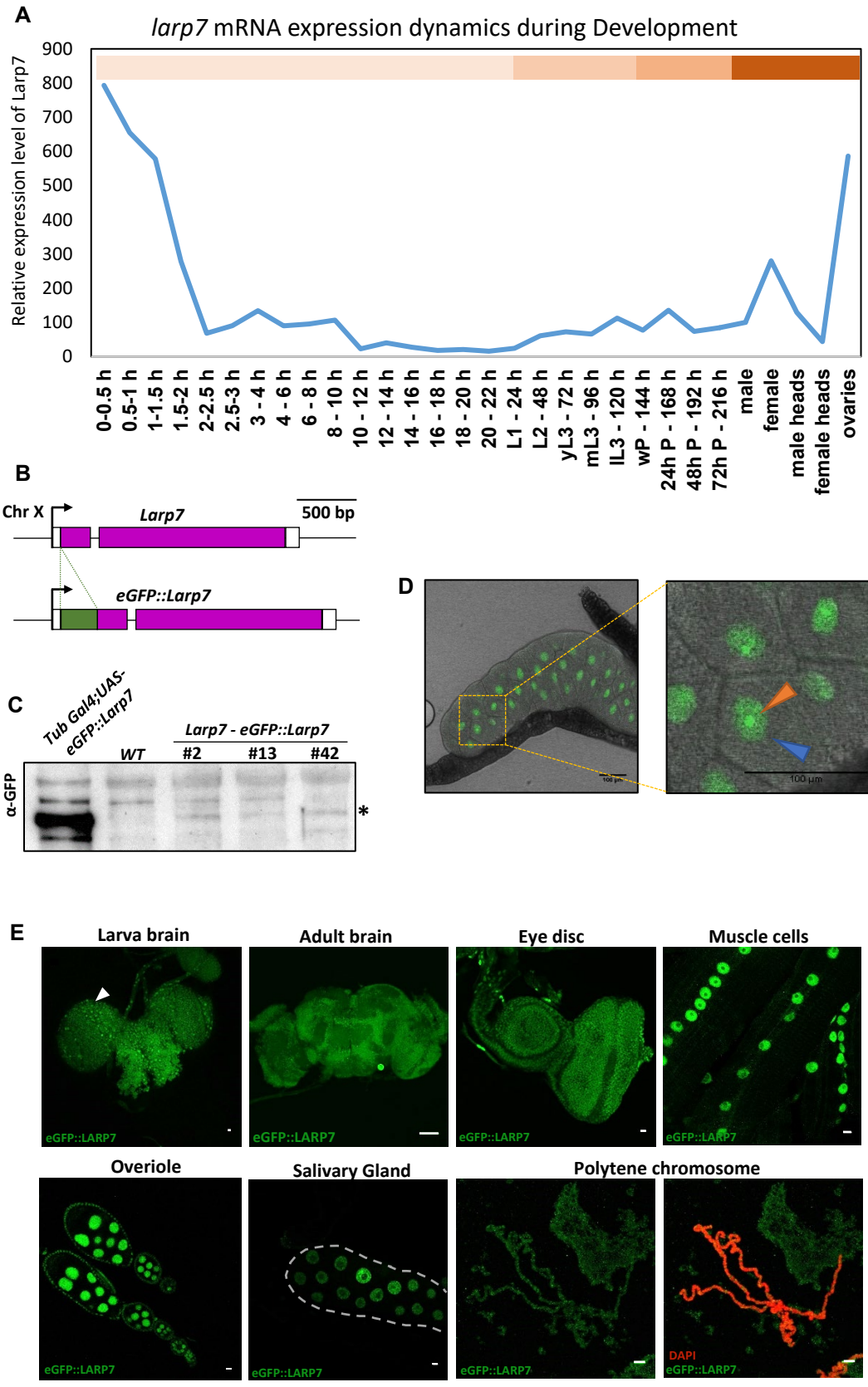


**Figure 10**: CRISPR/Cas9 mediated endogenous tagging of larp7. (A) RT-qPCR of larp7 mRNA at different developmental stages. The X-axis showing the sample collection time after egg laying. (B) Scheme of eGFP knock-in at translation start site of larp7 gene. (C) Immunoblot using GFP antibody showing expression of eGFP tagged Larp7 in whole larvae extract. *indicates the size of eGFP-Larp7. (D) Expression of eGFP-Larp7in salivary gland. Enlarged image showing eGFP-Larp7 localized in the nucleus (blue arrow head) and enriched in the nucleolus (orange arrow head). (E) confocal image of various tissues from larvae and adult showing expression of eGFP-larp7.

### 5.1.2 Identification of eGFP-Larp7 Enriched Midline Cells.

eGFP-Larp7 is highly enriched in a subset of cells in the late embryonic stages. To identify the nature of these cells, we used a set of midline drivers expressing Gal4 under the control of different promoters expressed in subtypes of cells at the midline. Using this approach, we find that enriched eGFP-Larp7 cells overlaps strongly with *Tbh, per, MzVUM,* and *engrailed* positive cells (Figure 11C). The common feature of these cells is that they are all motoneurons while engrailed positive cells also mark some interneurons. Therefore, this co-localization study clearly shows that Larp7 is differentially and specifically enriched in motoneurons and interneurons at the midline of embryos (Figure 11D).

## 5.2 Generation of *7SK* and *Larp7* knock-outs

### 5.2.1 The 7SK snRNP Complex is Required for Locomotion of *Drosophila* Larvae.

To get more insights into the role of the 7SK complex during organismal development, we created *Drosophila* mutants of *7SK* RNA and *Larp7* using the CRISPR/Cas9 methodology. Using specific guide RNAs directed against *Larp7* and *7SK* loci, mutants were generated and validated by sequencing (Figure 12A-B). The deleted regions are highlighted in the figure (Figure 12A-B). To further validate the mutants, we examined the level of the *7SK* RNA by qPCR in adult flies. As expected, the *7SK* mutant shows a significant reduction in the level of *7SK* RNA

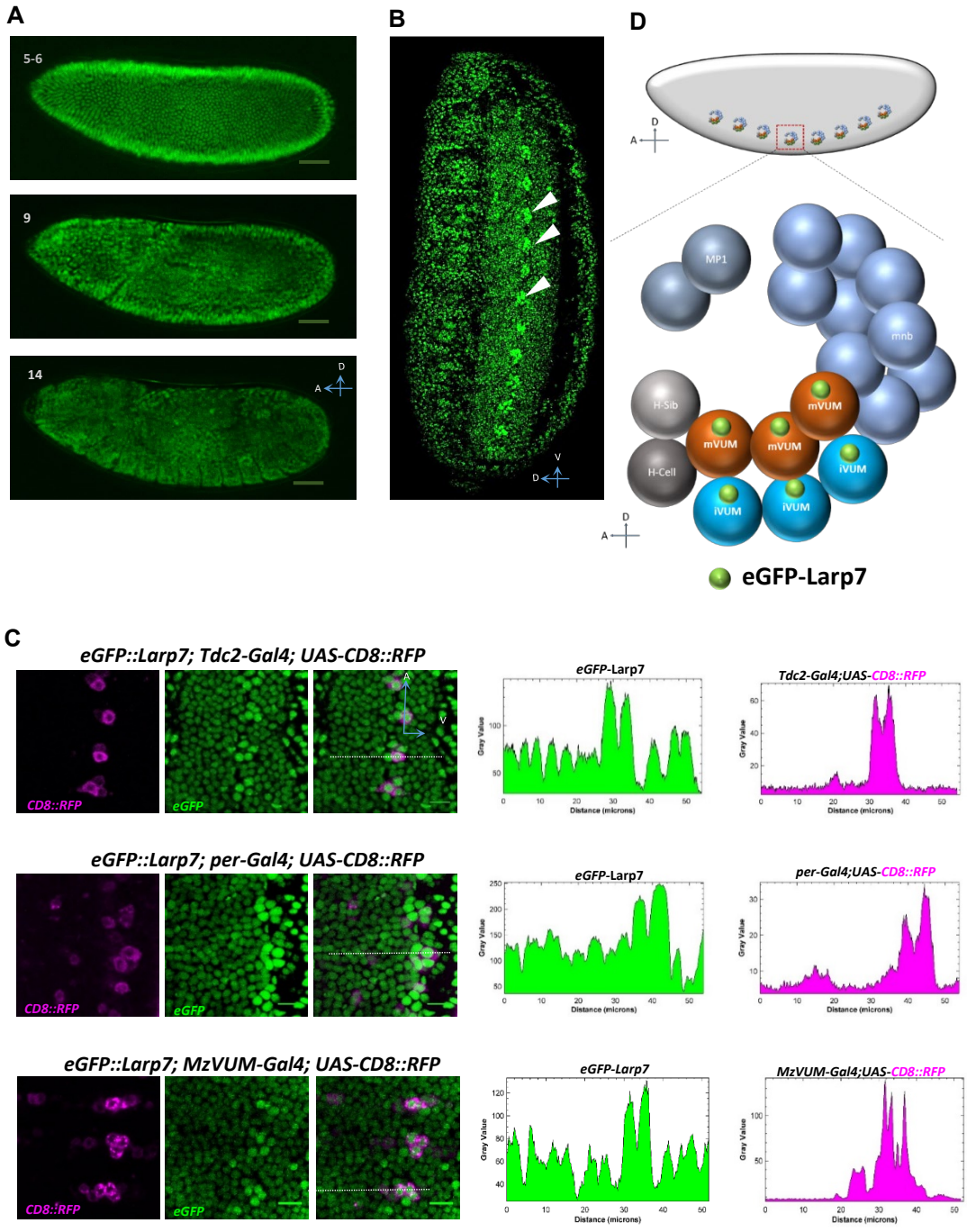



**Figure 11**: Identification of midline cells expressing eGFP-Larp7. (A) Representative time-lapse images of live embryo stages 5,9 and 14 showing expression pattern of eGFP-Larp7. (B) Stage 16 embryo, white arrowhead showing enriched expression of eGFP-Larp7 in the midline cells of segment 4,5 and 7. (D) Cartoon showing the types of cells in the midline of neuroectoderm. (C). Confocal images showing colocalization of eGFP-Larp7 with CD8-RFP driven by indicated Gal4 driver lines. White dotted line indicates the area analyzed in ImageJ. Graph showing gary value profile of eGFP-Larp7 and CD8-RFP, Plot-Profile application form ImageJ.

(Figure 12C-D). Similarly, and in light with the observation in human cells, the level of *7SK* RNA was strongly decreased in *Larp7* mutant. Importantly, the stability of the *7SK* RNA was restored when *Larp7* and *7SK* were ectopically expressed in *Larp7* and *7SK* mutants, respectively (Figure 12C-D).

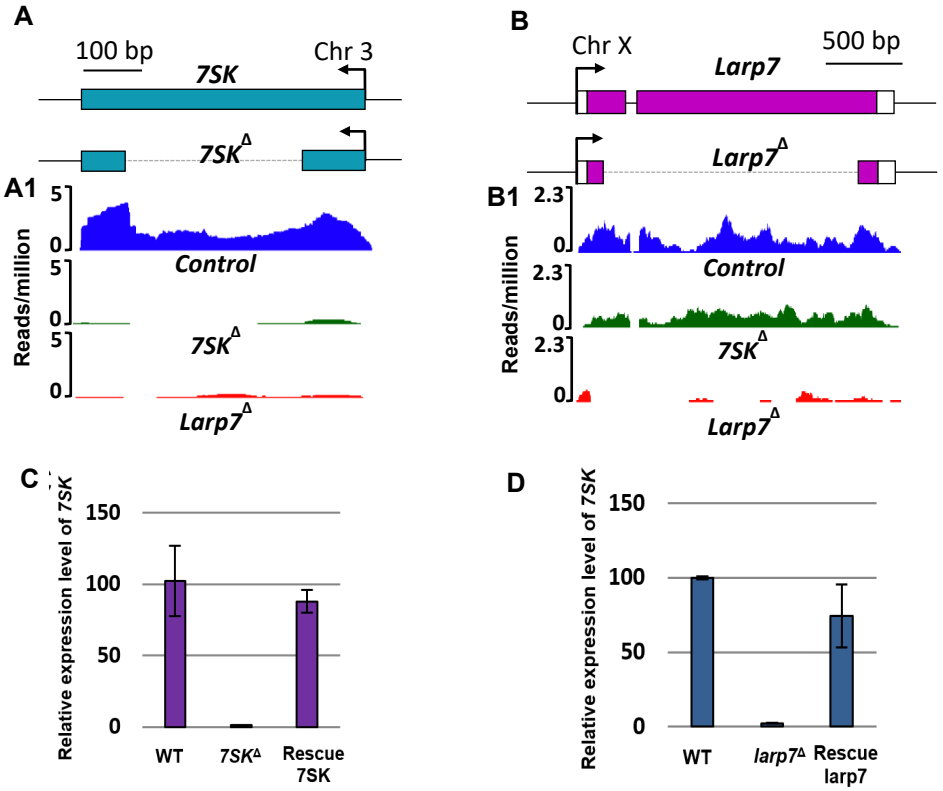



**Figure 12:** CRISPR/Cas9 mediated knock out of 7SK snRNP. (A-B) Scheme of 7SK and Larp7 locus showing the deleted region in dotted lines. (A1-B1) Larvae polyA RNA seq track from UCSC genome browser showing control, Larp7 KO and 7SK KO conditions. (C) RT-qPCR showing expression stability of 7SK RNA in control, 7SK KO and 7SK rescue adult flies. (D) Stability of 7SK RNA measured by RT-qPCR in the control, larp7 KO and eGFP-larp7 rescue adult flies.

To our surprise, both *Larp7* and *7SK* mutants were homozygous viable and fertile. This is in contrast to the previous study suggesting that the 7SK complex was essential in flies (Nguyen *et al.*, 2012b). We further examined each developmental stage of the mutants for any physiological or behavioural defects. Even though the mutants did not show developmental delay until eclosion, the lifespan of adult flies was significantly reduced (Figure 13A). This defect was rescued upon ectopic expression of *Larp7* and *7SK* (Figure 13B-C).

Furthermore, I noticed that *Larp7* KO flies live shorter compared to *7SK* KO flies, suggesting additional functions of Larp7 (Figure 13D-E). Apart from the reduced lifespan in adults, we uncovered a specific locomotion phenotype in *Larp7* and *7SK* mutant larvae (Figure 13F). The locomotion of third instar larvae was quantified by calculating the number of peristaltic movements for two minutes. Both mutants show significantly reduced a number of these movements (Figure 13G) compared to control larvae, indicating impaired locomotion.

### 5.2.2 *Larp7* is Essential in Post Mitotic Neurons.

To address the underlying mechanism of the 7SK RNP complex in locomotion, we attempted to identify the relevant cells pertaining to this defect. For this purpose, we ectopically expressed eGFP-tagged Larp7 (different from the endogenous tagged eGFP-Larp7) using different cell-type specific Gal4 drivers (Brand and Perrimon, 1993), aiming to rescue the *Larp7* mutant. Beforehand, I validated the functionality of the eGFP-tagged-Larp7. I showed by immunoprecipitation that eGFP-Larp7 could bind *7SK* RNA (Figure 13H). Furthermore, ubiquitous expression of eGFP-Larp7 rescued the stability of *7SK* RNA as well as the locomotion defect of Larp7 mutants. These results demonstrate that the eGFP-Larp7construct is fully competent to substitute for the loss of Larp7 and that the locomotion defect is not due to off-target activity.

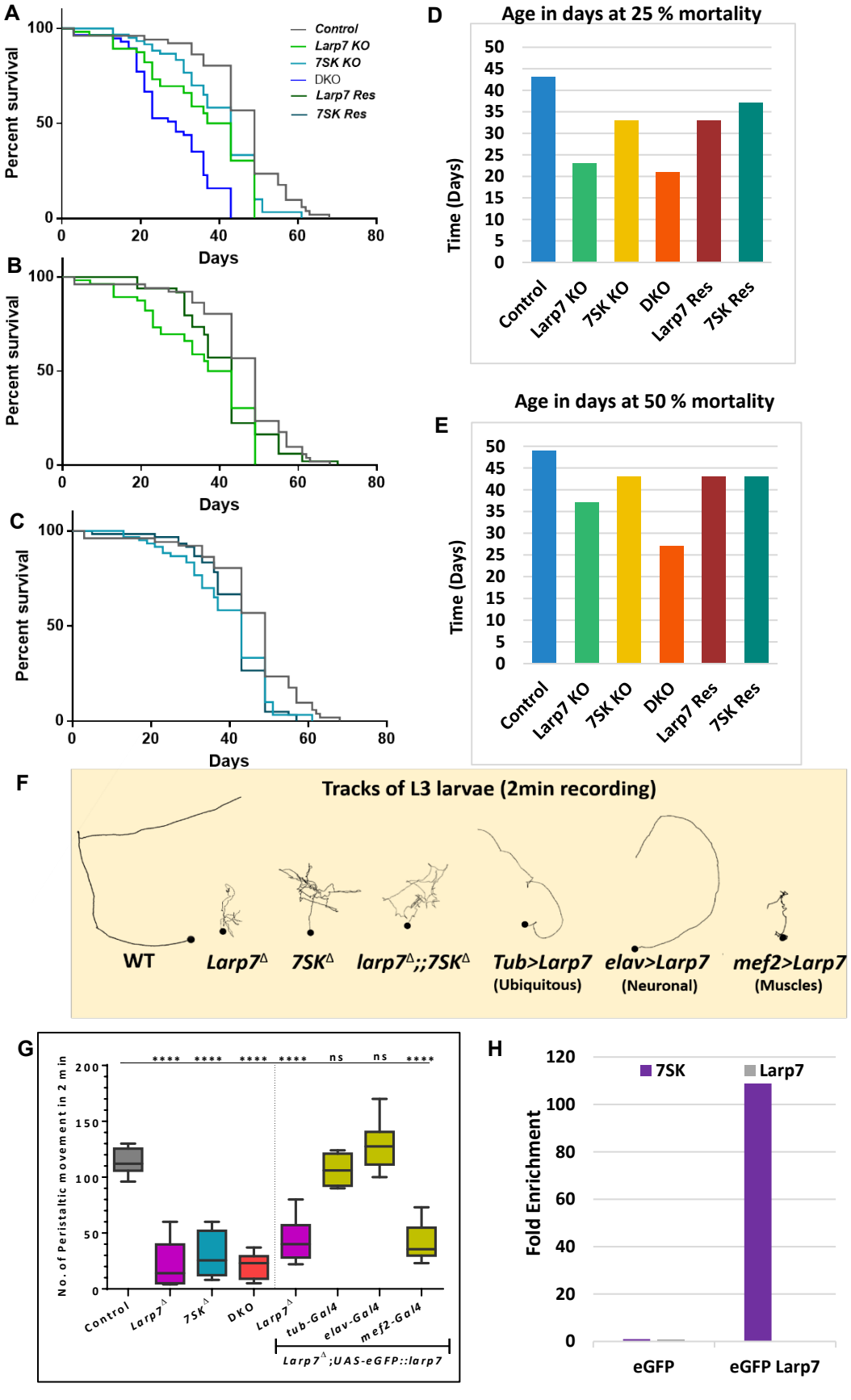

A
Percent survival
Days
Control
Larp7 KO
7SK KO
DKO
Larp7 Res
7SK Res
B
C
D
Age in days at 25 % mortality
Time (Days)
Control
Larp7 KO
7SK KO
DKO
Larp7 Res
7SK Res
E
Age in days at 50 % mortality
F
Tracks of L3 larvae (2min recording)
WT
Larp7Δ
7SKΔ
larp7Δ;;7SKΔ
Tub>Larp7
(Ubiquitous)
elav>Larp7
(Neuronal)
mef2>Larp7
(Muscles)
G
No. of Peristaltic movement in 2 min
****
ns
Control
Larp7Δ
7SKΔ
DKO
tub-Gal4
elav-Gal4
mef2-Gal4
Larp7Δ;UAS-eGFP::larp7
H
7SK
Larp7
Fold Enrichment
eGFP
eGFP Larp7

**Figure 13:** Behavioral characterization of 7SK snRNP mutants. (A) Lifespan of control 7SK KO, larp7 KO and DKO adult flies. Percentage survival is plotted. (B-C) including larp7 rescue and 7SK rescue. (D-E) Percentage of mortality with respect to days showing larp7 mutants die faster than other mutants. (F) Track of L3 larvae movement recorded for 2 minutes of indicated genotypes. Black dot indicates the position of the larvae at time zero (start point). (G) Box-plot showing the number of peristaltic movement counted per two minutes of indicates genotypes. (H) RNA immunoprecipitation –qPCR showing the binding of 7SK to the ectopically expressed eGFP tagged larp7. Larp7 mRNA is used as negative control.

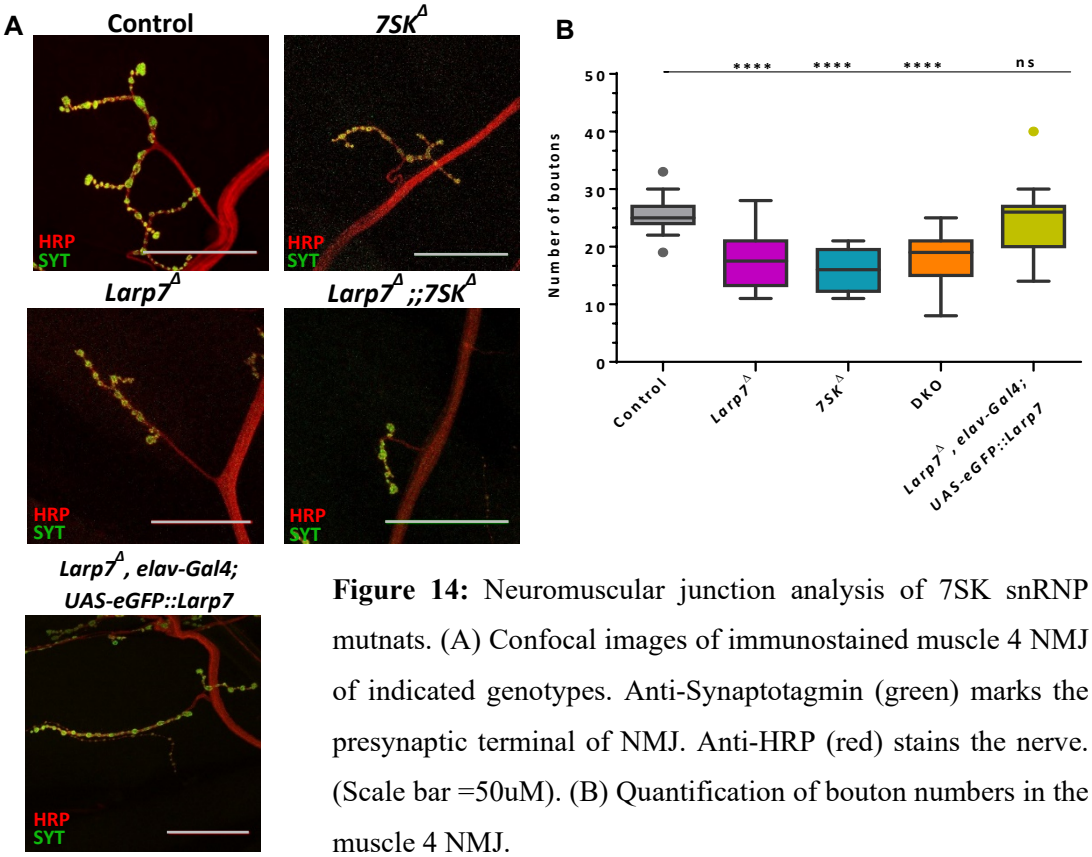



**Figure 14:** Neuromuscular junction analysis of 7SK snRNP mutnats. (A) Confocal images of immunostained muscle 4 NMJ of indicated genotypes. Anti-Synaptotagmin (green) marks the presynaptic terminal of NMJ. Anti-HRP (red) stains the nerve. (Scale bar =50uM). (B) Quantification of bouton numbers in the muscle 4 NMJ.

In *Drosophila* larvae, a locomotion phenotype can arise due to improper function of neurons, muscles or both tissues. To address which was the relevant tissue I used the *elav*-Gal4 line, which expresses Gal4 is all postmitotic neurons and *mef2*-Gal4, which expresses Gal4 in muscles. Inspection of the locomotion in both conditions revealed that expression in postmitotic neurons was sufficient to rescue the defect (Figure 13F-G). In contrast, expression of Larp7 in mutant muscles did not show improvement of the defect (Figure 13F-G). Therefore, this analysis underscores the importance of *Larp7* in neuronal tissues.

## 5.3 The 7SK RNP complex is required in motoneurons

### 5.3.1 Neuromuscular junction analysis of 7SK snRNP mutants.

The fact that *Larp7* controls locomotion via a neuronal function prompted us to examine the neuromuscular junctions (NMJ). NMJ forms the interface between axon terminals of the motoneurons and the muscles. This interface regulates the excitation of the muscles, and consequently, the locomotion of larvae. In *Drosophila*, NMJ at muscle 4 and 6-7 is widely studied and is used as a paradigm to study axonal growth and function. The NMJ at muscle 4 is easily accessible, and the number of synapses in respective body segments is similar among animals of the same developmental stage. We examined the NMJs of muscle 4, at segment A2 and A3 in wandering L3 larvae. Interestingly, we found that the length of the innervated nerve was reduced both in *7SK* and *Larp7* mutants (Figure 14A), suggesting that the development of the NMJ is affected in the absence of the 7SK complex. We calculated the number of synaptic boutons after immunostaining with the presynaptic marker synaptotagmin. We found a drastic reduction of boutons in both *Larp7* and *7SK* mutants (Figure 14A). We observed that the development of the NMJ was also affected at muscles 6/7 (Figure 14C). These defects at the NMJ were completely rescued when eGFP-Larp7 was expressed in postmitotic neurons of the *Larp7* mutant. Collectively, our results indicate that *Larp7* is required for axonal growth of motoneurons at the NMJ.

### 5.3.2 Cell specific expression of *Larp7* in motoneurons restores locomotion

To confirm the requirement of *Larp7* in motoneurons, we ectopically expressed eGFP-Larp7 in different neuronal populations and examined the locomotion of *Larp7* mutants. Interestingly, we found that only the drivers that express Gal4 in motoneurons rescued the locomotion defect of *Larp7* mutants. For instance, neuroblast specific driver (*Insc*-Gal4) did not rescue the phenotype, suggesting that *Larp7* is dispensable in the undifferentiated state. Similarly, expressing Lar*p7* in glial cells (*Repo*-Gal4) had no rescuing effect. These data strongly suggest that *Larp7* is specifically required for the development of motoneurons, and is dispensable in other neuronal cell types.

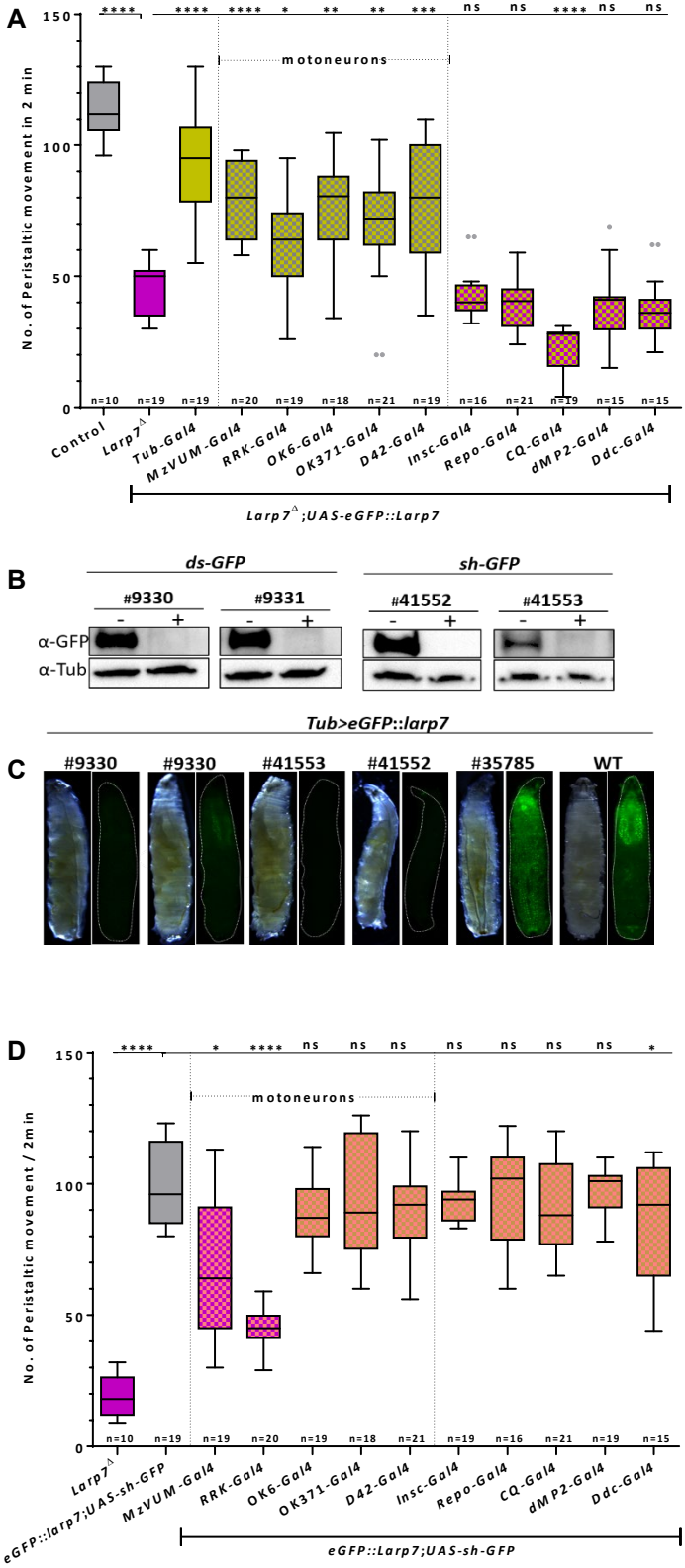



**Figure 15:** Motoneurons require 7SK snRNP complex for proper locomotion. (A) Crawling assay showing rescue of larp7 mutant defect in the larvae by expressing cDNA of larp7 in the motoneurons. (B) Validation of GFP KD efficiency in adult head by immunoblot using anti-GFP antibody. ds-GFP and sh-GFP shows strong reduction in level of eGFP-Larp7. Anti-Tubulin antibody is used as loading control. (C) Bright field and fluorescent images of L3 larvae (Tub-Gal4>UAS-eGFP::Larp7) ectopically expressing eGFP-Larp7. iGFPi lines are indicated. #35785 is ds-mCherry which is used as negative control. (D) Crawling assay showing knockdown of Larp7 by iGFPi using various neuronal drivers.

### 5.3.3 Depletion of *Larp7* in motoneuron recapitulate mutant phenotype.

To confirm the cell autonomous role of *Larp7* in motoneurons, we specifically depleted its product in different neuronal cell types. Since no specific knock down line was available for Larp7, we used the in vivo GFP interference (iGFPi) method (Mugat *et al.*, 2003; Pastor-Pareja and Xu, 2011) . In iGFPi, double-stranded RNA against GFP is used to target a specific fusion protein containing GFP. In our case, we used this approach to deplete eGFP-Larp7. We first checked the knockdown the efficiency of diverse available GFP dsRNA fly lines. By western blot against GFP and fluorescent imaging, we determine that the dsRNA lines 41552 and 41553 were the most efficient in depleting eGFP-Larp7 (Figure 15B-C). We decided to carry out our experiments using with the 41552 line because 41553 was not healthy. Consistent with our rescue experiment, we found that depleting *Larp7* in motoneurons results in mild to strong locomotion defect in L3 larvae (Figure 15D). Depletion with the motoneuron driver RRK-Gal4 (with expresses Gal4 in aCC and RP2 cells) displayed the strongest defect. RP2 cells are among the motoneurons that innervate muscle 4 in each hemisegment. Other motoneuron drivers like OK6-GAL4 and OK371-GAL4 are specific to all motoneurons, but we did not see strong defect as with RRK-Gal4 (Figure 15D). This might be due to the expression strength of individual Gal4 drivers, which can impact the knockdown efficiency.

We next asked whether the same motoneurons impact the growth of NMJ at muscle 4. For this, we used both overexpression and depletion of *Larp7*, as shown above. Consistent with the effect on locomotion, we found that expressing *Larp7* with RRK-Gal4 line can rescue the growth defect at NMJ while depleting *Larp7* with the same driver leads to growth defect (Figure 16A-B). Together these experiments demonstrate that *Larp7* is required in motoneurons, in particular in aCC and RP2 cells, for regulating their growth and ultimately the locomotion of L3 larvae.

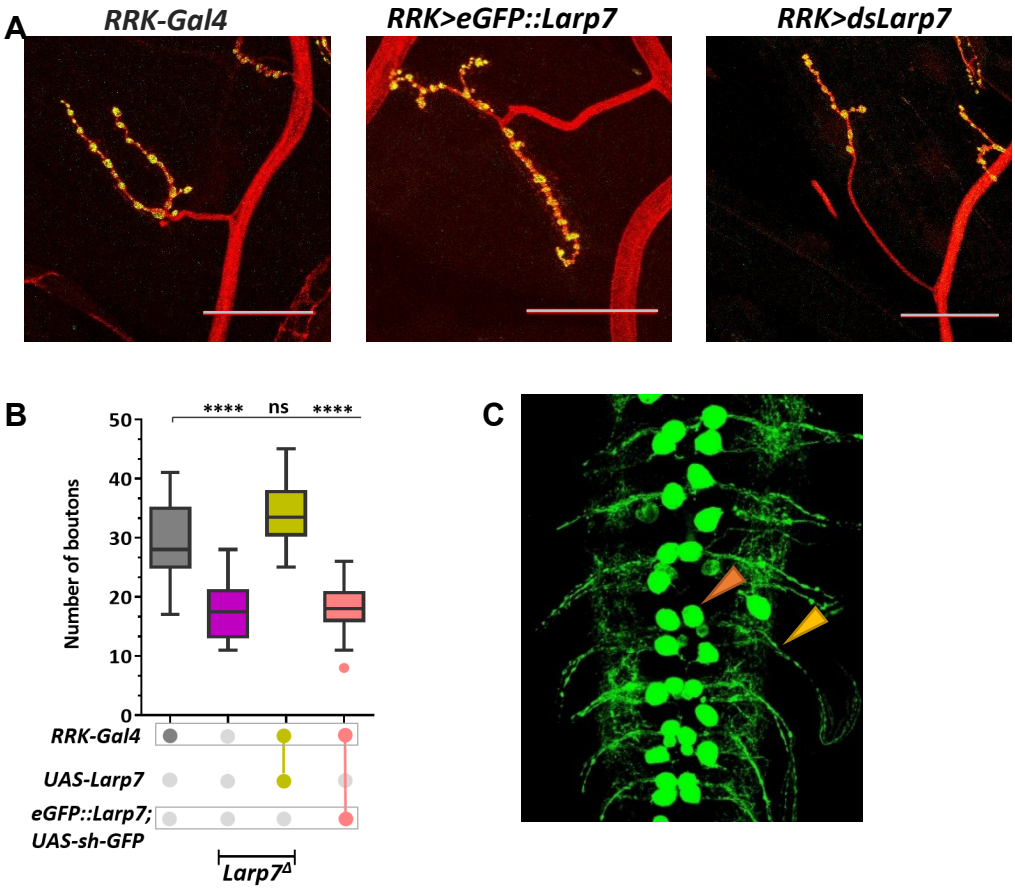



**Figure 16:** NMJ development regulated by level of 7SK snRNP in motoneurons. (A) Representative images of NMJ at muscle 4 of the indicated genotypes. (Scale= 50uM). (B) Quantification of bouton numbers in the muscle 4 NMJ. (C) Projected confocal images of RRK-Gal4 positive cells expressing membrane marker CD8-GFP in L3 larvae ventral nerve chord. Cell body (red arrow head) and axon projecting out of ventral nerve chord (orange arrow head) indicated.

## 5.4 The locomotion and axonal growth defects result from P-TEFb hyperactivation.

The best-described function of the 7SK snRNP is the regulation of transcriptional pausing by sequestration of P-TEFb. We thus hypothesized that the locomotion phenotype might derive from transcriptional misregulation via hyperactivation of P-TEFb. If this hypothesis is correct, we expected that reducing P-TEFb dosage could rescue the defect associated with the loss of 7SK RNP function. To test this, we knocked down *Cdk9* in the *Larp7* mutant using the pan-neuronal driver, *elav*-Gal4, as well as RRK-Gal4. We first examined the knockdown efficiency of two

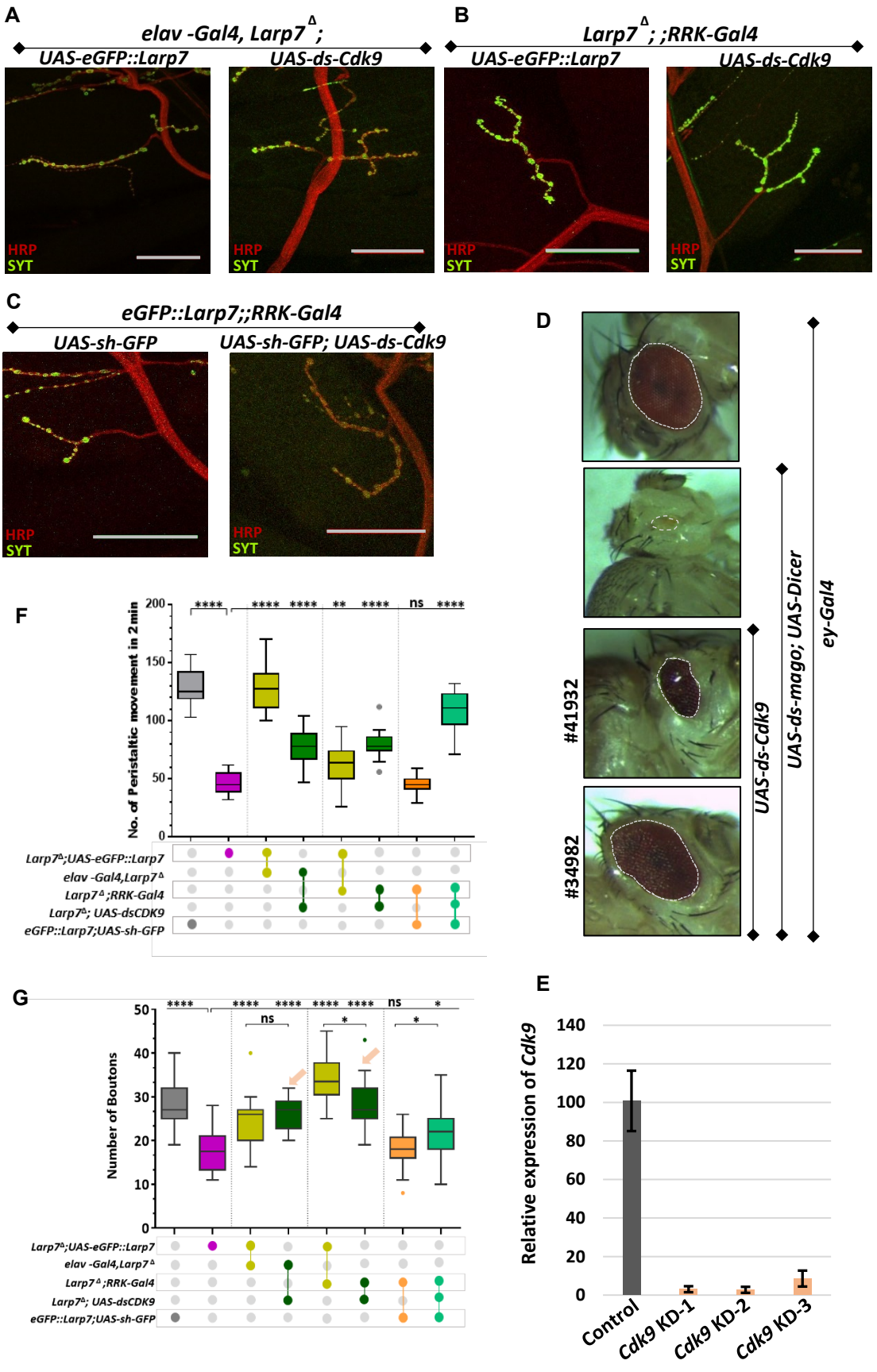

A
elav -Gal4, Larp7Δ;
UAS-eGFP::Larp7
UAS-ds-Cdk9
HRP
SYT
B
Larp7Δ; ;RRK-Gal4
UAS-eGFP::Larp7
UAS-ds-Cdk9
HRP
SYT
C
eGFP::Larp7;;RRK-Gal4
UAS-sh-GFP
UAS-sh-GFP; UAS-ds-Cdk9
HRP
SYT
D
#41932
#34982
UAS-ds-Cdk9
UAS-ds-mago; UAS-Dicer
ey-Gal4
F
No. of Peristaltic movement in 2 min
200
150
100
50
0
**** **** **** ** **** ns ****
Larp7Δ;UAS-eGFP::Larp7
elav -Gal4,Larp7Δ
Larp7Δ;RRK-Gal4
Larp7Δ; UAS-dsCDK9
eGFP::Larp7;UAS-sh-GFP
G
Number of Boutons
50
40
30
20
10
0
**** **** **** **** **** ns *
ns
*
*
Larp7Δ;UAS-eGFP::Larp7
elav -Gal4,Larp7Δ
Larp7Δ;RRK-Gal4
Larp7Δ; UAS-dsCDK9
eGFP::Larp7;UAS-sh-GFP
E
Relative expression of Cdk9
140
120
100
80
60
40
20
0
Control
Cdk9 KD-1
Cdk9 KD-2
Cdk9 KD-3

**Figure 17:** Transcriptional misregulation in 7SK snRNP mutants results in locomotion defect. (A-C) Representative confocal images of muscle 4 NMJ of the indicated genotypes. (D) validation of *Cdk9* knockdown efficiency. The rescue of photoreceptor differentiation (size of adult eye) in mago depleted and *Cdk9* KD lines. (E) Validation of *Cdk9* KD (#34982) in RRK-Gal4 cells. RT-qPCR showing level of *Cdk9* mRNA in sorted embryo RRK-Gal4 positive cells of three biological samples. (F) Crawling assay showing partial rescue of larp7 mutant defect upon knockdown of *Cdk9* in all neurons (Elav-Gal4) and RRK positive cells. (G) NMJ analysis showing rescue of number of boutons in muscle 4 NMJ upon *Cdk9* KD in larp7 mutants.

dsRNA *Cdk9* lines. We chose line 34982 because it showed the strongest reduction of *Cdk9* mRNA level (Figure 17D-E). Strikingly, we found that the *Cdk9* KD partially rescued the crawling defect of *Larp7* mutant (Figure 17F). Furthermore, the number of synaptic boutons and the axonal length were completed rescued in these larvae (Figure 17A). Similar observations were made in the double knockdown for *Larp7* and *Cdk9*. (Figure 17C). Importantly, the knock down of *Cdk9* alone in wild type neurons was embryonic lethal, indicating that *Cdk9* level needs to be tightly regulated in neurons and that *Larp7* and the *7SK* play an essential role in this regulation. Overall our findings suggest that the locomotion defect and the growth defect at the NMJ observed in the 7SK snRNP loss of function is the result of P-TEFb hyperactivation.

## 5.5 Transcriptomic analysis of 7SK snRNP mutant in aCC and RP2 cells.

### 5.5.1 Specific set of genes are regulated by 7SK snRNP complex.

Our results point towards a specific transcriptional defect in motoneurons upon 7SK snRNP ablation. Therefore, we next investigated the transcriptome of aCC and RP2 cells, which are the motoneurons that appear to be more sensitive to the loss of the 7SK complex. For this purpose, we used FACS to sort cells from control, *Larp7*, *Larp7* rescue, and *7SK* KOs of late embryos (stage 15-16). The identity of the isolated cells was validated by RT-qPCR (Figure 18A-B). Overall,

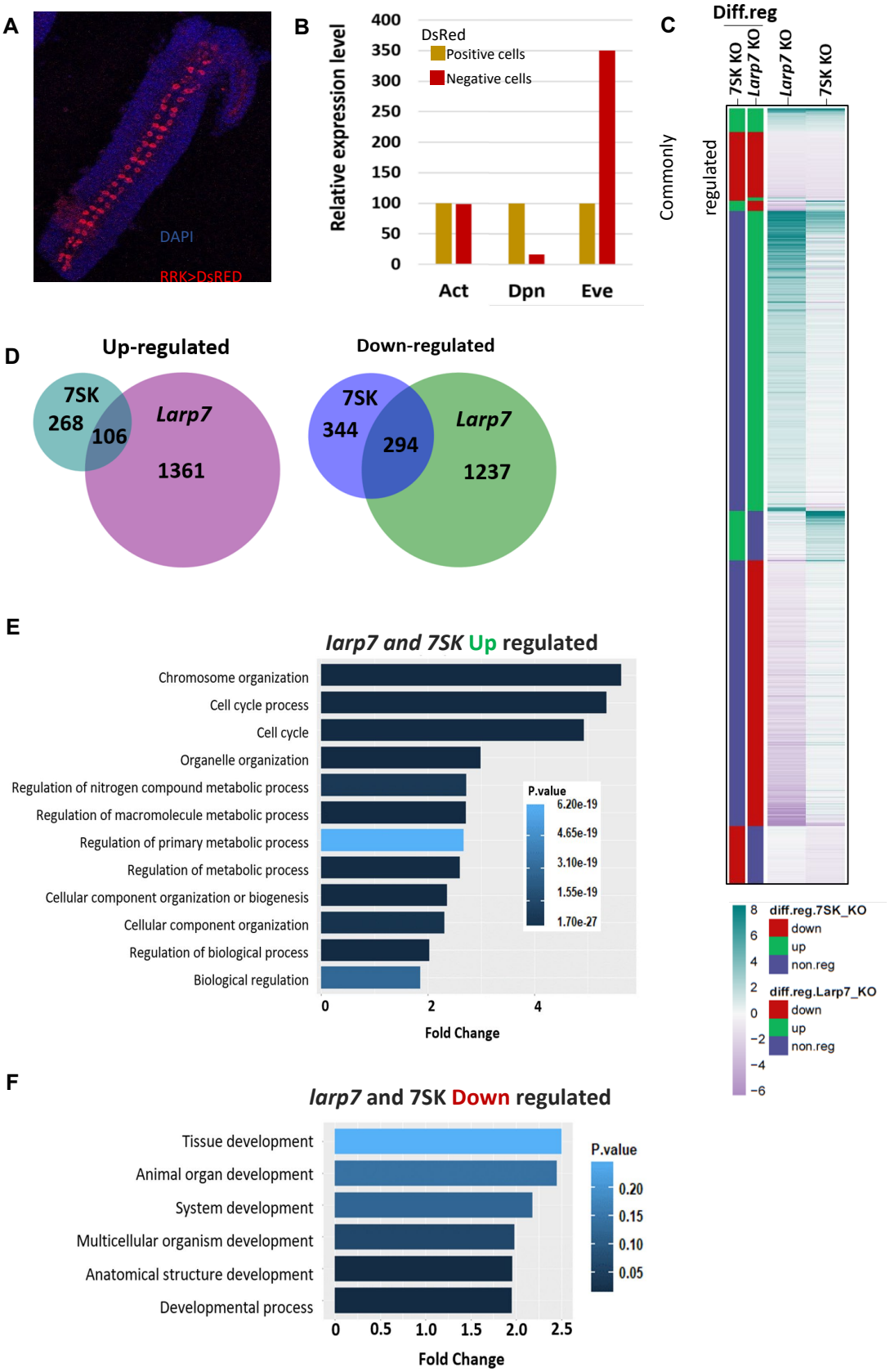

A
DAPI
RRK>DsRED
B
Relative expression level
DsRed
Positive cells
Negative cells
Act
Dpn
Eve
C
Diff.reg
7SK KO
Larp7 KO
Larp7 KO
7SK KO
Commonly regulated
diff.reg.7SK_KO
down
up
non.reg
diff.reg.Larp7_KO
down
up
non.reg
D
Up-regulated
7SK
268
106
Larp7
1361
Down-regulated
7SK
344
294
Larp7
1237
E
larp7 and 7SK Up regulated
Chromosome organization
Cell cycle process
Cell cycle
Organelle organization
Regulation of nitrogen compound metabolic process
Regulation of macromolecule metabolic process
Regulation of primary metabolic process
Regulation of metabolic process
Cellular component organization or biogenesis
Cellular component organization
Regulation of biological process
Biological regulation
P.value
Fold Change
F
larp7 and 7SK Down regulated
Tissue development
Animal organ development
System development
Multicellular organism development
Anatomical structure development
Developmental process
P.value
Fold Change

**Figure 18:** Transcriptomic analysis of 7SK snRNP mutants.(A) Projected confocal images of dissected embryo neuroectoderm expressing nuclear DsRed in RRK positive cells which is used as marker for sorting cells. (B) RT-qPCR of actin, deadpan (*dpn*) and even skipped (*eve*) in sorted DsRed positive and negative cells. DsRed positive cells are enriched with *eve* confirming the identity of the isolated cells. (C) Heat map of all transcripts in larp7 and 7SK mutant motoneurons. Genes commonly regulated in both mutants are clustered on the top. Diff.reg – indicated the differentially regulated gene with up regulation in green, down regulation in red, and nonregulated in blue. (D) Eulers diagram showing the overlap of differentially regulated gene between mutants. (E-F) Gene ontology of commonly regulated genes

*Larp7* KO has more differentially regulated genes (2,998) compared to *7SK* KO embryos (1,012) (Figure 18C-D). 106 up-regulated and 294 down-regulated genes are common within both mutants. We focused on these common genes for further analysis as they likely represent regulation by the whole 7SK complex. Gene-ontology analysis indicates that down-regulated genes are highly enriched for the developmental process, (Figure 18 E) while up-regulated genes are enriched for chromosome organization and metabolic processes (Figure 18F).

It was previously shown that the rate of RNA pol II elongation can impact alternative splicing (de la Mata *et al.*, 2003; Khodor *et al.*, 2011; Moehle *et al.*, 2014). Since 7SK snRNP complex regulates recruitment of P-TEFb to promoters, we wondered whether its loss would affect splicing. However, our analysis of splicing events did not uncover many changes. This suggests that the release of pausing does not necessarily impact on the rate of transcriptional elongation but more on the number of active RNA Pol II into elongation.

### 5.5.2 GC distribution at the promoter is a signature of pausing.

Previous reports on analyses of RNA Pol II chromatin immunoprecipitation (ChIP) experiments in early stage *Drosophila* embryos demonstrated that genes that have strong paused Pol II have higher GC content at promoters (Hendrix *et al.*, 2008). Therefore, we analyzed the GC content of the genes misregulated in the 7SK snRNP mutants (Figure 19).

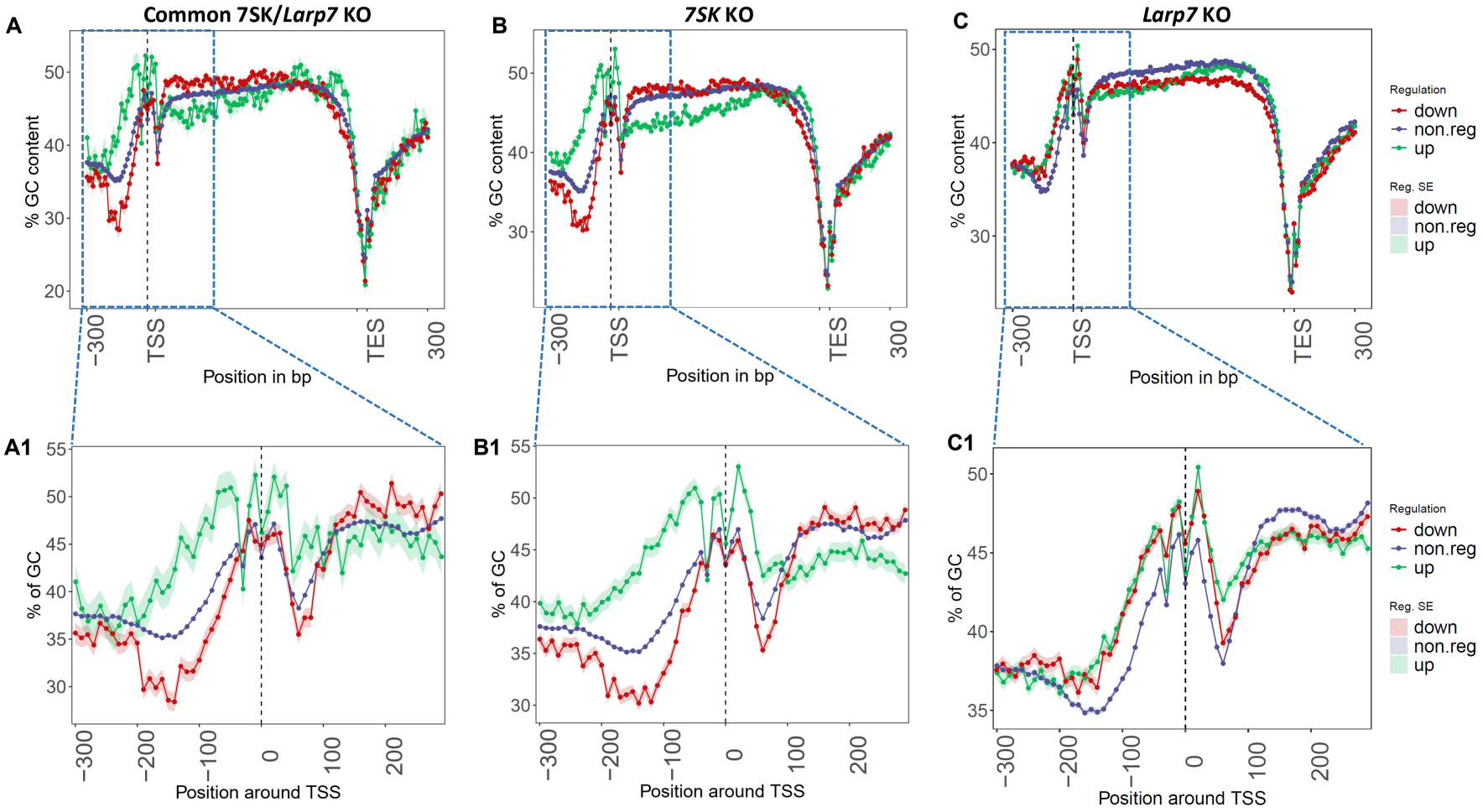

A
Common 7SK/Larp7 KO
B
7SK KO
C
Larp7 KO
% GC content
Position in bp
−300
TSS
TES
300
A1
B1
C1
% of GC
Position around TSS
−300
−200
−100
0
100
200
Regulation
down
non.reg
up
Reg. SE
down
non.reg
up

**Figure 19:** GC content of the 7SK snRNP regulated genes inmotoneurons. (A-C) Metagene GC profile of differentially regulated and nonregulated genes in commonly misregulated, 7SK mutant and larp7 mutants. 300bp upstream of transcription start site (TSS) and 300bp downstream of transcription end site (TES) was included in the analysis. (A1-C1) enlarged region of metagene profile A-C, 300bp around TSS. Upregulated genes in green, downregulated genes in red and non regulated genes in blue.

Interestingly, commonly up-regulated genes show higher GC content at their promoter while down-regulated ones display lower content (Figure 19A-A1). This suggests that upregulated genes have strong paused RNA Pol II. We could speculate that in the absence of the 7SK complex, these genes would be the most affected as they would be more impacted by the release of P-TEFb. In our analysis, another striking observation is the low GC content at the promoter of the downregulated genes. A similar tendency is observed in the *7SK* regulated genes, while *Larp7* regulated genes did not show this trend (Figure 19B-B1). We also checked the GC content in the gene body of misregulated genes. We found that commonly up-regulated genes have high GC content until 100bp downstream of TSS, while in the gene body, it was significantly reduced compared to non-regulated genes (Figure 19A1). A similar pattern is observed in *7SK* regulated genes (Figure 19B1). There is no significant difference in the *larp7* regulated genes (Figure 19C-C1).

In conclusion, the 7SK snRNP complex affects more specifically genes that exhibit differential GC content at their promoters and within the gene body. It restricts the expression of genes containing high GC content at their promoter and conversely, it facilitates the expression of genes with lower GC content. The potential underlying mechanism will be discussed below.

### 5.5.3 Long intron-containing genes are sensitive to 7SK snRNP complex depletion.

To further get insight into the specificity of the 7SK complex, we next examined the size of regulated genes (Figure 20). We analyzed the full gene length, the length of the exons and introns. In the commonly up-regulated genes, we found that the length of introns was relatively longer in comparison to the non-regulated and

down-regulated ones (Figure 20C). A similar pattern was observed in *7SK* KO up-regulated genes but not in the *Larp7* KO (Figure 20C). Consistently, the total gene

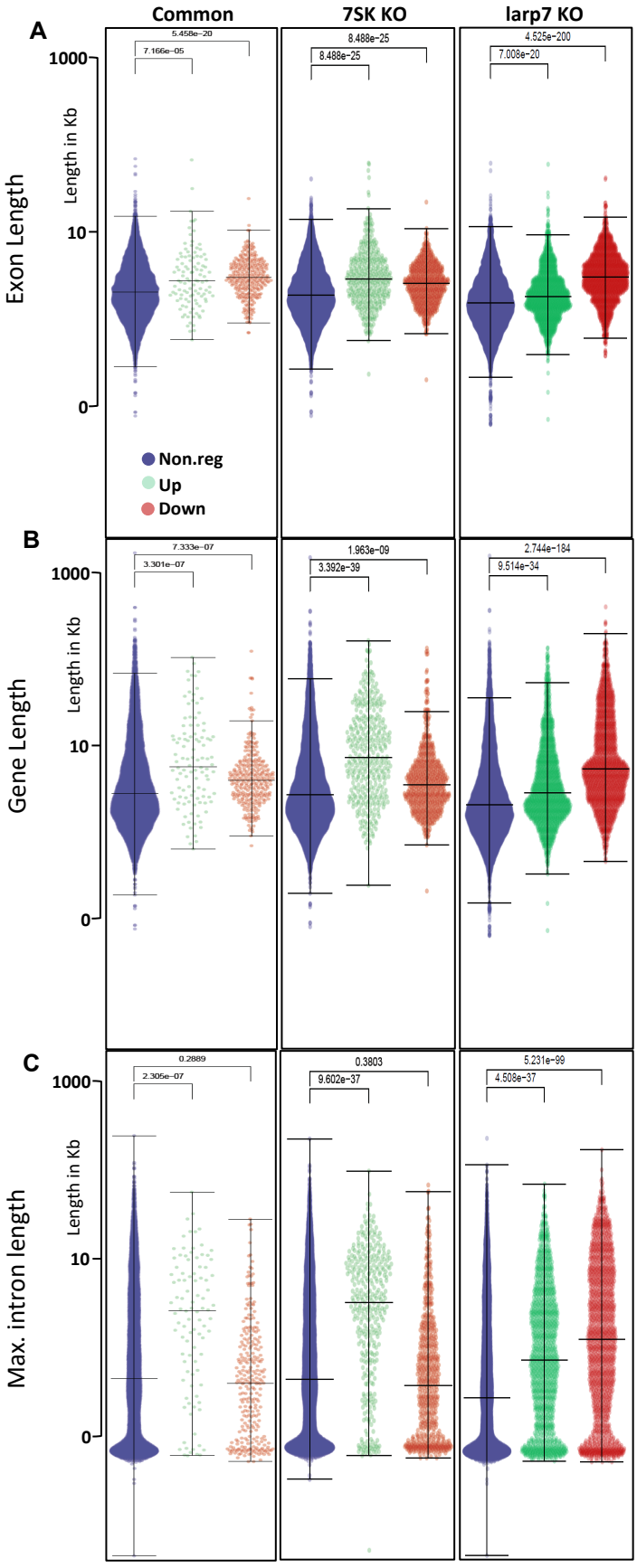



**Figure 20:** Gene length analysis of differentially regulated gene in 7SK snRNP mutant motoneurons. (A) Exon length of differentially regulated genes in indicated conditions. Upregulated genes are green, downregulated genes are in red and nonregulated genes are in blue. (B) Total gene length in kilobases of above mentioned conditions. (C) Maximum intron length in kilobases.

length was also significantly increased in the commonly, and *7SK* KO up-regulated genes (Figure 20B), while the exon length was unchanged (Figure 20A). Together, our transcriptomic analysis in motoneurons thus indicates that the 7SK snRNP complex specifically restricts the expression of long genes containing high GC content at their promoters.

### 5.5.4 The 7SK snRNP mutants regulate genes containing set of motifs at around TSS.

Pausing was shown to be enhanced by certain motifs enriched around the TSS (Hendrix *et al.*, 2008). We thus examined whether differentially regulated genes bear some of these motifs. For this analysis, we considered the region spanning +300bp to -300bp around TSS (Figure 21A-C). Consistent with the possibility that commonly up-regulated genes are strongly paused, we found that their promoters contain motifs characteristic of paused genes, including GAGA, Initiator, DPE, and pause button (Figure 21A-C,E) (Hendrix *et al.*, 2008). We also found that the TATA motif was enriched in differentially regulated genes, but the distribution of the motif was distinct from up and down-regulated genes (Figure 21D). Also, in the down-regulated genes, the DRE motif was found enriched at promoters (Figure 21F).

In conclusion the computational analysis of the differently regulated genes in the 7SK snRNP mutants indicates that they have specific signatures at their promoter and within the gene body, which likely instruct their regulation by the 7SK complex.

Our work so far suggests that genes that are highly paused are more dependent on the 7SK complex for their expression. However, a majority of genes are paused, and yet only a subset of genes is regulated by 7SK RNP. Therefore it remains unclear what drives the specificity of the complex towards specific promoters. In order to get more insights, we performed a de novo motif analysis around TSS, and we found that SMAD motif is enriched only in the commonly and *7SK* KO upregulated genes (Figure 22). Suggesting 7SK snRNP complex regulated the development of NMJ through SMAD transcription factors.

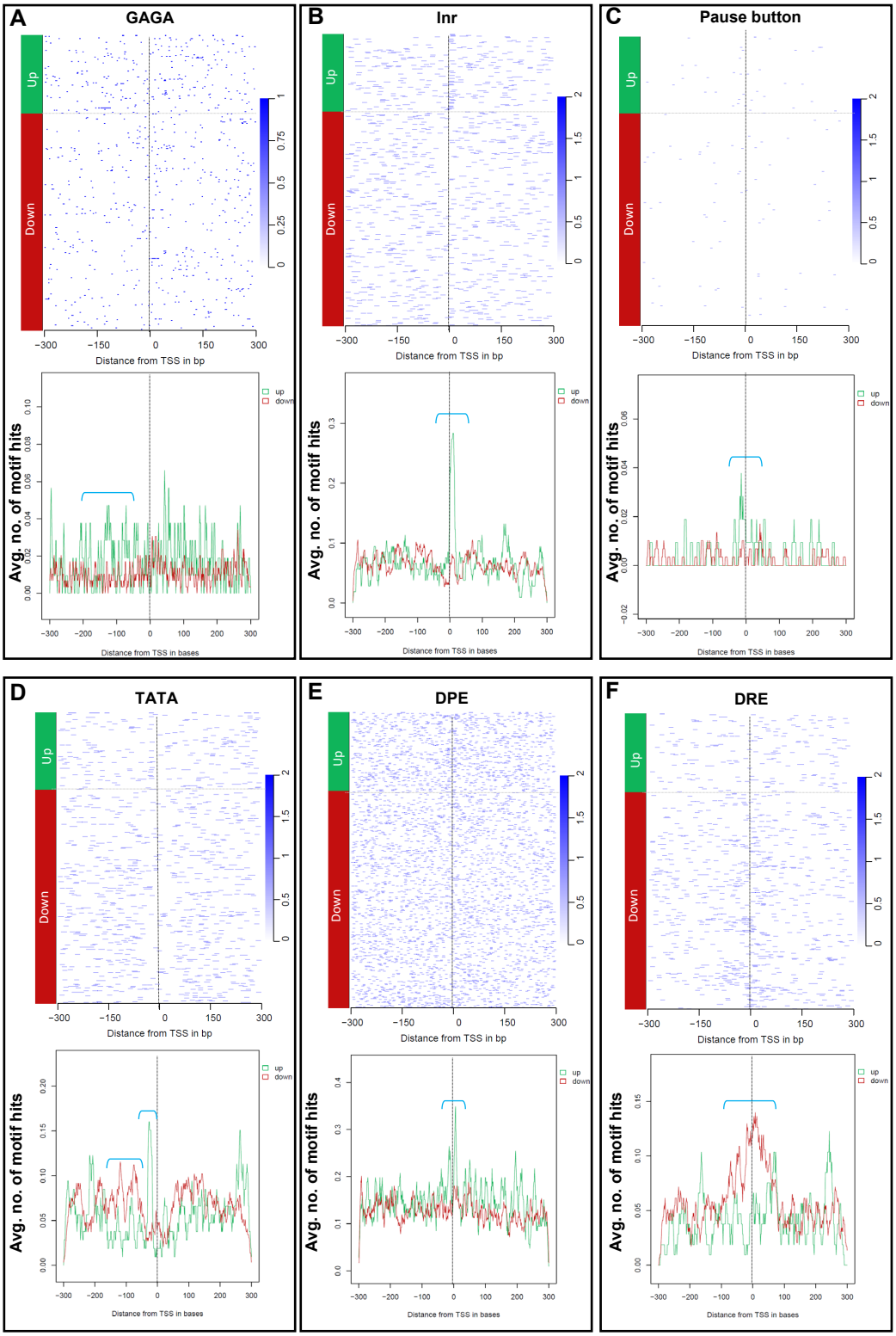



**Figure 21:** Motif analysis of differentially regulated genes in 7SK snRNP mutant motoneurons. (A-F) Heat map and enrichment of indicated motifs were checked in the commonly regulated genes. Upregulated genes in green, downregulated

genes in red and nonregulated genes in blue. 300bp around TSS was used for this motif enrichment analysis. Downward brackets (turquoise) highlight the enriched region for indicated motifs either in up or down regulated genes.

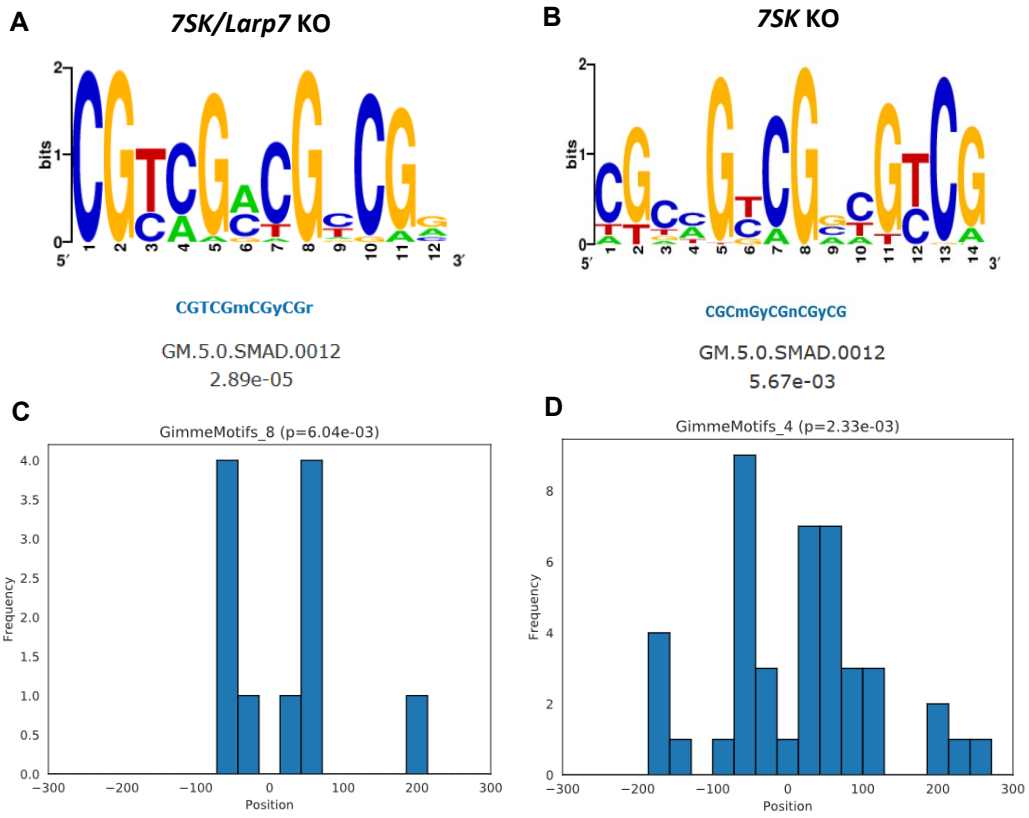



**Figure 22:** De novo motif search at the promoter region of upregulated genes in *7SK/Larp7* KO (A,C) and *7SK* KO (B,D). A and B showing GimmeMotif search result. C and D histogram identifies the position of the SMAD motif in -300bp to +300bp window around TSS of upregulated genes in the KO

## 5.6 Larp7 function is evolutionarily conserved.

The 7SK snRNP complex is conserved among metazoans. The *Drosophila 7SK* RNA is about 100 nt longer than the vertebrate *7SK* RNA and bears low sequence similarity with its vertebrate counterpart (less than 50%). However, the proteins associated with the *7SK* RNA are highly conserved in metazoans. Larp7 has evolutionarily conserved 2X RRM domains, and it was shown that the RRM interaction with the 3′end of *7SK* RNA is essential for the stability of the RNA (Uchikawa *et al.*, 2015b). We wondered whether vertebrates Larp7 could rescue the *Drosophila* one despite the low sequence similarity of *7SK* RNA. To test this, we cloned the human (hLarp7) and *Xenopus Larp7* (*xLarp7*) tagged with GFP and expressed the fusion constructs in *Drosophila* S2R+ cells that were previously treated with dsRNA targeting the 5′UTR of *Drosophila Larp7* (this dsRNA does not recognize the exogenous Human and Xenopus *Larp7*) (Figure 23C and 14F). As readout, we examined the level of *7SK* RNA by RT-qPCR. Surprisingly, we found that *Drosophila 7SK* RNA level was rescued by the Human and *Xenopus* Larp7 (Figure 23A), to the same extent as with the *Drosophila* Larp7. This implies that the molecular function of Larp7 is conserved throughout evolution. To next address the functional conservation *in vivo,* we generated transgenic flies expressing human *Larp7* under the control of UAS promoter. Strikingly, expressing *hLarp7* in all tissues using the *tubulin*-Gal4 driver completely rescued the locomotion and axonal growth defect of *Larp7* mutant L3 larvae. Therefore, these data indicate that Larp7 function within the 7SK complex is evolutionarily conserved from insects to vertebrates.

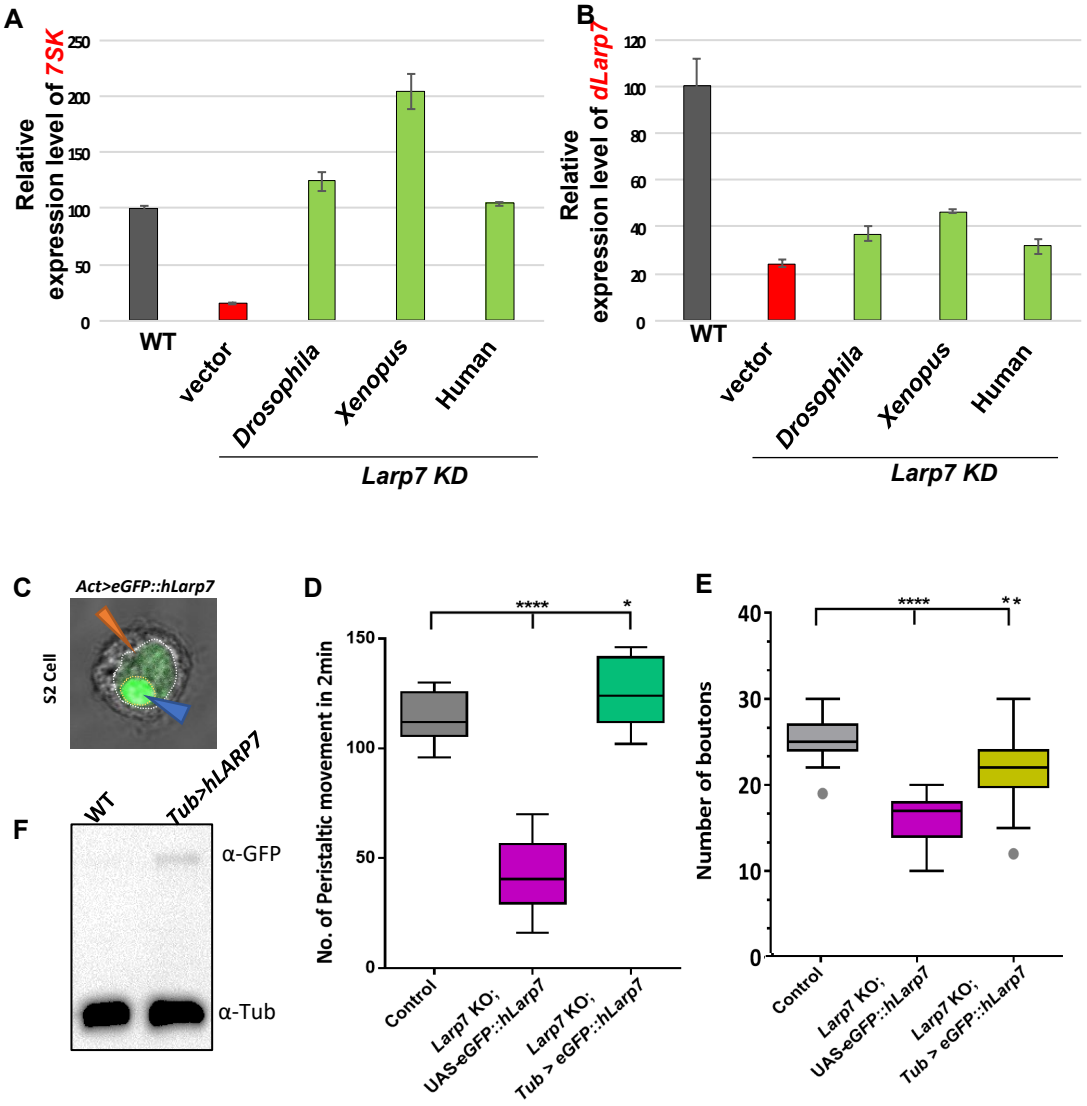



**Figure 23:** Human Larp7 is functionally conserved and rescue *Drosophila* larp7 mutant. (A) RT-qPCR showing the stability of 7SK RNA in dLarp7 knockdown S2R+ cells. cDNA of dLarp7, hLarp7 and xLarp7 were expressed ectopically with actin promoter. (B) RT-qPCR showing dLarp7 knockdown efficiency. (C) Bright field and fluorescent image of S2R+ cells expressing eGFP-hLarp7. hLarp7 localize strictly in the nucleus (orange arrow head) and enriched in the nucleolus (blue arrow head) as seen in dlarp7 (figure 2D). (F) Immunoblot showing expression of eGFP-hLarp7 in transgenic adult fly head extract. Tubulin is used as loading control. (D) Crawling assay showing the rescue of Larp7 mutant locomotion defect by hLarp7. (E) Bouton numbers in muscle 4 NMJ of indicated genotype. Human Larp7 expression in *Drosophila* larp7 KO larvae rescues the bouton numbers.

# Discussion

# 6 DISCUSSION

## 6.1 The *Drosophila* 7SK snRNP complex is structurally similar to the vertebrate counterpart.

The 7SK snRNP complex was first identified in vertebrates, and it was thought to be present only in vertebrates because of the failure to identify *7SK* RNA in lower organisms. In 2012, the *Drosophila* 7SK snRNP complex was identified. The core proteins LARP7, HEXIM and MePCE/Bin3 are highly conserved and were shown to be present in a complex with *7SK* RNA. In the budding yeast the ortholog of LARP7, Pof8 was shown to be essential for the maintenance of telomere length (Mennie, Moser and Nakamura, 2018; Páez-Moscoso *et al.*, 2018). Pof8 carries only the LAM domain and lacks the C-terminal RRM, which is essential for the stability of the *7SK* RNA. This suggests that Pof8 exerts its function independently of the 7SK RNP. Accordingly, the *7SK* RNA was not found in this species. Interestingly, the telomere length of Alazami syndrome patients is shortened both in homozygous and heterozygous conditions. This indicates that the vertebrate LARP7 can influence telomere length. Whether the *7SK* RNA is involved in this function remains to be determined. In *Drosophila*, retrotransposons are used to maintain the telomere length, which is mechanistically different from vertebrates. It would be interesting to examine whether the telomere length in *Drosophila Larp7* mutants, as well as in the *7SK* RNA mutant, is also affected. This can help to dissect the independent function of *Larp7* and related phenotypes in Alazami syndrome.

### 6.1.1 Sequence alignment comparison between species.

The *7SK* RNA is highly conserved in vertebrates. The *Drosophila 7SK* RNA was identified by sequence similarities to the promoter region. The total length of the *Drosophila 7SK* RNA is around 100 nt longer than the vertebrate counterpart and has less than 50% similarity. The *7SK* RNA act as a scaffold for the assembly of the core proteins and other hnRNPs. The secondary structure of vertebrate *7SK* RNA was predicted using computational and chemical methods like i-shape. The excess 100 nt in the *Drosophila 7SK* RNA is distributed in the middle region of the

RNA. This sequence might increase the complexity of the secondary structure of the *7SK* RNA by inducing the interaction of novel proteins and regulating different molecular functions. It would be interesting to determine the structure of *Drosophila 7SK* RNA and identify novel potential interactors. Furthermore, it would be exciting to address whether human *7SK* RNA can rescue the lack of *Drosophila 7SK*. If so, this would indicate that the structure of the RNA is likely more important than the sequence per se.

### 6.1.2 Subcellular localization of the 7SK snRNP complex.

The endogenously tagged eGFP-Larp7 in *Drosophila* is strictly nuclear and relatively enriched in the nucleolus as well as in the nucleoplasm. Our finding is partly consistent with earlier observations from Larp7 staining in human and *Drosophila* using specific antibodies. While Larp7 was found in the nucleus of several tissues in *Drosophila* using a specific antibody the nucleolar localization was however not detected (Nguyen *et al.*, 2012a) Nevertheless, Slomnicki et al., 2016 show that Larp7 is one of the nucleolar enriched protein in rat cerebral cortex, suggesting that at least in Vertebrates LARP7 localizes in the nucleolus. The ectopic expression of Human LARP7 and Xenopus LARP7 in *Drosophila* S2 cells also show localization both in the nucleolus and in the nucleoplasm. Based on these findings, we favour the interpretation that Larp7 is a strictly nuclear protein in different organisms with some expression in the nucleolus.

The role that Larp7 may play in the nucleolus is currently unclear. It seems that the localization of Larp7 is independent of the 7SK snRNP complex as this was not changed when *7SK* RNA was depleted in flies. This observation suggests that probably Larp7 has a function unrelated to 7SK in this compartment. Slomnicki et al., showed that the loss of *Larp7* in the rat hippocampal neurons lead to decreased perikaryal ribosome content and reduced protein synthesis. In contrast, we did not observe profound changes in protein level in L3 larvae depleted for *7SK* RNA or *Larp7*. To detect more subtle protein synthesis defect, it would be interesting to examine nascent translation in vivo, in cell type-specific manner, using for instance fluorescent non-canonical amino-acid-tagging techniques (Erdmann *et al.*, 2015).

Recently, Larp7 was shown to bound U6 snRNA and be required for its 2'O methylation in human cells (Gunther Meister, personal communication). As U6 is synthetized in the nucleolus, this could be an important function for Larp7 in this compartment. However, the physiological function of 2'O methylation of U6 snRNA is currently unknown.

Regarding the subcellular localization of the *7SK* RNA, it was found to be enriched in nuclear speckles or interchromatin granule clusters (IGCs) in HeLa cells. With the help of reporter system, it was shown that the *7SK* RNA transiently interact with the reporter locus and displace P-TEFb (Prasanth *et al.*, 2010). In mouse primary motoneurons grown in a microfluidic chamber, *7SK* is sparsely localized in the axons. It was proposed that the heterogeneous nuclear ribonucleoprotein R (hnRNP R) binds to the *7SK* RNA in the axons and the loss of either hnRNP R or *7SK* in this cell culture system leads to defective axonal growth, suggesting a cytoplasmic function for the *7SK*-hnRNP R complex (Briese *et al.*, 2018). In *Drosophila*, we did not address the localization of *7SK* in axons. However, we did not detect Larp7 in this compartment, which is incompatible with the localization of *7SK* in the axons as Larp7 is essential for *7SK* stability. Furthermore, our results point towards a transcriptional function for the axonal growth rather than a cytoplasmic one.

### 6.1.3 Vertebrate Larp7 can stabilize the *Drosophila 7SK* RNA.

Protein components of 7SK snRNP are highly conserved in vertebrates and *Drosophila*. Despite the fact that *7SK* RNA sequence is only partially conserved, it likely provides the same structural scaffold for the assembly of the core proteins. This is supported by the rescue of *7SK* RNA level and the locomotion/NMJ defects of *Larp7* mutant by overexpression of vertebrates Larp7. This suggests that the structural information of the RNA is more important than the sequence per se.

## 6.2 Ubiquitous and differential expression of 7SK snRNP components during development of *Drosophila.*

*Larp7* mRNA is maternally deposited, and LARP7 protein is readily visible as soon as the cellularization happens during embryogenesis. Time-lapse imaging of developing embryo revealed that *Larp7* is expressed in all embryonic developmental stages, albeit at a different level. Similarly, *Larp7* is expressed in all tissues in larva and adult. An earlier report using Larp7 antibody claimed that Larp7 was expressed differentially in the eye-antennal and wing imaginal discs of L3 larvae (Nguyen *et al.*, 2012a). However, we could not reproduce this observation using our endogenously GFP-tagged Larp7. This difference is likely due to antibody specificity and processing of the sample for immunostaining. Our live-imaging analysis of eGFP-Larp7 is expected to reduce these potential biases. We cannot completely rule out; however, that insertion of GFP alters the distribution and abundance of Larp7. Nevertheless, the fact that we found enrichment of Larp7 in motoneurons of late embryos is more consistent with our phenotypic observations.

## 6.3 Depletion of the 7SK snRNP complex leads to a developmental defect.

The discovery that 7SK snRNP regulates promoter Pol II pausing via sequestration of P-TEFb complex was first demonstrated in human cells (Nguyen *et al.*, 2001; Yang *et al.*, 2001). In cancer cells it was shown that LARP7 and *7SK* levels were reduced, suggesting that they act as potential tumour suppressors (Cheng *et al.*, 2012; Ji *et al.*, 2014; Abasi *et al.*, 2016). In a human gastric tumour, *LARP7* was shown to carry microsatellite, resulting in a frameshift mutation and the loss of the C-terminus region, which is essential for *7SK* RNA interaction (He *et al.*, 2008). Even though there is evidence that the level of *7SK* RNA and Larp7 is reduced in these cancer cells, the underlying mechanism driving tumorigenesis is unknown.

Interestingly, mice depleted for Larp7 are embryonic lethal. This is in sharp contrast with the viability observed in *Drosophila* and human patients (see also below). Furthermore, tumorigenesis was not reported in this mouse model. Likewise, patients lacking LARP7 show no sign of tumour formation. Therefore the *ex-vivo*

observations contradict the defects observed *in vivo* (Alazami et al., 2012; Okamura et al., 2012). It is possible that the correlation observed between tumorgenesis and the decreased level of LARP7 in these cancer cells is only an indirect consequence of additional changes promoting tumorigeneses such as a mutation in other oncogenes or tumour suppressor genes.

Why does the *Larp7* mutant is embryonic lethal in mouse? The authors showed that the lethality occurs due to growth arrest in primordial germ cells (Okamura et al., 2012) (Figure 24A). There is however, a possibility that the lethality is independent of *Larp7* loss of function. In fact, the mouse *Larp7* gene contains nine introns and a miRNA cluster (Mirc20) consisting of Mir302a/b/c/d, and Mir367 that is located in intron 8 (fig). The miR-302 is essential for the development of the embryo, and its absence leads to neuronal defects (Parchem *et al.*, 2015) (Figure 24B). Both *Larp7* KO and miRC20 KO mouse survival rate, as well as the time of death, are identical. Since in the *Larp7* KO mouse (Okamura et al., 2012), the level of miRC20 expression was not addressed, it is possible (Parchem *et al.*, 2015) that these phenotypic similarities arise from a loss function of miRC20 rather than the loss of *Larp7*. Therefore, more experiments will be necessary to confirm that the embryonic lethality is due to Larp7 deficiency. The localization of miRC20 in the intron of *Larp7* is conserved in human and mouse, so to fully understand the phenotype of the Alazami syndrome, one should consider addressing the expression of the miRNA as well. In *Drosophila*, the *Larp7* gene contains one intron and does not contain any miRNA. This makes *Drosophila* a better system to avoid confounding effect by possible misregulation of this miRNA.

## 6.4 *Drosophila* 7SK snRNP mutants have strong locomotion defect

We generated a loss of function mutations for components of the *Drosophila* 7SK snRNP complex using the CRISPR/Cas approaches. Even though the mouse mutant and the cell culture studies suggest that the 7SK snRNP complex is essential, the *Drosophila* 7SK snRNP mutants are viable. This observation is similar to human patients lacking *LARP7* (Alazami syndrome). Currently, there is no report about the lifespan expectancy and fertility of these patients. We show that *Drosophila Larp7* and *7SK* mutants are fertile and homozygote flies can be maintained over a

generation. Since the studies on AS patients are relatively new, and none of them report about the fertility of the patients, it is difficult to compare with the *Drosophila* mutant observation (Alazami *et al.*, 2012; Holohan *et al.*, 2016).

We further analyzed the phenotypes of 7SK snRNP mutants at all developmental stages. As for oogenesis the embryonic development of these mutants is normal, and until adult eclosion, the developmental time is similar to wild type.

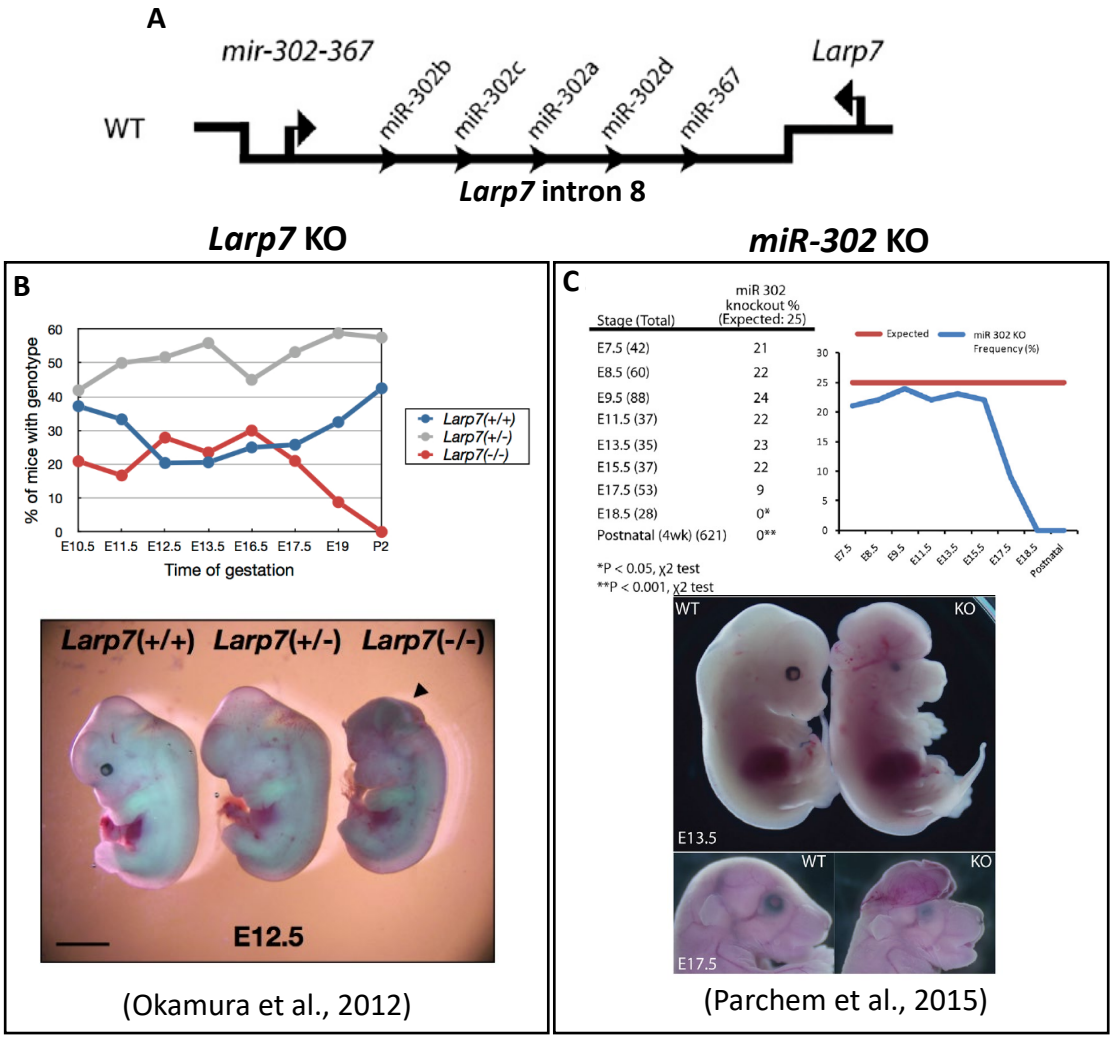



**Figure 24:** (A) *Larp7* KO mouse embryos. Developmental stages of wild type and *Larp7* mutants as indicated. Picture of embryo at E12.5 (B) *miR-302* KO mouse. Developmental stage of mutant with respect to expected. Pictures of embryos at E13.5 and E17.5. Data adapted form Okamura et al, 2012 and Parchem et al., 2015.

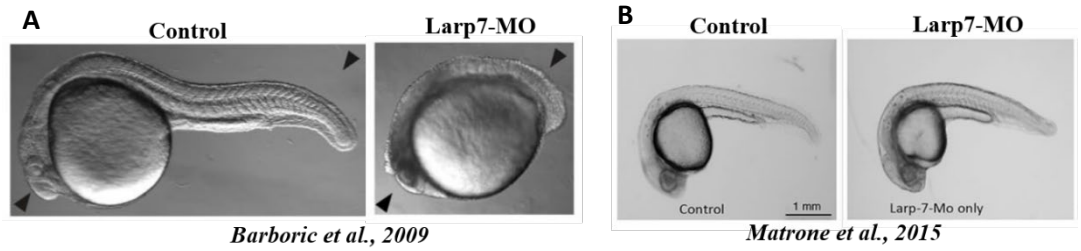



**Figure 25:** (A) Barboric et al., showing vertebrate developmental defect in *Larp7*-morpholino treated zebrafish embryos. (B) Matrone et al., zebrafish embryos did not show any developmental defect in *Larp7* knockdown condition.

However, the *7SK* and *Larp7* mutant flies have a shorter lifespan, and this phenotype is enhanced in the double mutant. We further found that mutant L3 larvae and adult flies display strong locomotion defects. This defect is not seen in the early larval stage, suggesting that the 7SK snRNP complex is dispensable during embryogenesis. We further show that the locomotion defect is caused by the loss of function of the 7SK snRNP complex in neurons and not in muscles.

In *Drosophila* the actual development of the body size and most of the nervous system happens at the late embryonic and larval stage. In human, the early development of the nervous system happens in utero, which is analogous to the *Drosophila* larval stage. This suggests that the 7SK snRNP is essential for proper neuronal development of the organism, probably by regulating transcription. In all reported AS patients, there is a significant delay in motor development, suggesting that the 7SK snRNP complex may have an evolutionarily conserved function in the axonal growth of motor neurons.

## 6.5 The NMJ development is highly affected in 7SK snRNP mutants

7SK snRNP mutants show strong locomotion defect that can be rescued by ectopically expressing *Larp7* in motoneurons. To identify the cause for the defect, we studied the morphology and electrophysiology of the motoneurons. The loss of 7SK snRNP did not alter the morphology of the motoneuron and their pathfinding in the late embryonic stages, supporting that the main time of action is during the larval stage. In L3 larvae, we found that the neuromuscular junctions of motoneurons targeting muscle 6/7 and muscle 4 were affected. These NMJ had shorter terminal and fewer synaptic boutons. In collaboration with Pr. Carsten Duch at Mainz University, we also examined the electrophysiological properties of the NMJ in *Larp7* mutants. Two-Electrode Voltage Clamp (TEVC) recordings from L3 larvae muscles 6&7 in abdominal segments 2&3 show mild defects. *Larp7* mutants show an increased peak amplitude upon evoked synaptic transmission compared to control animals. This phenotype can be fully rescued by *Larp7* overexpression in a *Larp7* mutant background. This mild alteration in synaptic transmission does show a statistically significant difference but is not sufficient to explain the changes in crawling behaviour. In this experiment, the current was injected into the nerve, and

maybe in the mutants, there is a deficit of action potential generated by the mutant motoneurons. To further understand the discrepancy between strong locomotion defect and mild synaptic impairment, one can use calcium imaging of motoneurons and combine with muscle recordings. This will tell us whether there is any defect in the axons of the motoneurons in their ability to excite the muscles.

## 6.6 The axonal growth defect in the 7SK snRNP mutants is due to transcriptional changes.

In *Drosophila*, we found that the loss of the 7SK snRNP complex leads to locomotion defect, which is the result of impaired NMJ development and function. The knockdown of *Larp7* in zebrafish was found in one study to give rise to developmental defect (Barboric *et al.*, 2009), while in another one, the KD had no consequence. This might be due to the efficiency of KD by the different morpholino or by of target activity. Nevertheless, in the later study, the KD of *Larp7* and *Cdk9* rescued the phenotype observed in the KD of *Cdk9* only (Matrone *et al.*, 2015) (Figure 25). Our findings also support the fact that the locomotion phenotype in L3 larvae is caused by specific transcriptional changes in motoneurons. Indeed, the knockdown of *Cdk9* using the pan-neuronal or motoneuron driver rescue the locomotion phenotype of *Larp7* mutant. Why only motoneurons appear specifically affected by the loss of Larp7 or *7SK*? What makes them particularly susceptible to changes in the level of 7SK snRNP complex? Both *Drosophila* and human motoneurons are among the cells having the most extended structures. Therefore one can speculate that owing to their extremely specialized morphology, it is possible that their formation and function require more precise and robust transcription regulation in comparison to other cells in the organism.

## 6.7 Misregulated genes in the 7SK snRNP mutants have unique architecture.

The 7SK snRNP complex regulates the activity of P-TEFb complex, which is a generic mechanism, but in *Drosophila*, its function is more apparent in motoneurons. We analyzed the transcriptional changes in the *7SK* KO and *Larp7*

KO motoneurons that are *even-skipped* positive (aCC and RP2) because the KD of *Larp7* in this subtype of motoneurons alone phenocopy the 7SK snRNP mutant phenotype. A second reason was to have a homogenous cell population, which is important to identify cell-specific transcriptional changes. The differential gene expression analysis revealed that the *Larp7* KO results in larger changes in RNA expression than the *7SK* KO. This observation suggests that Larp7 itself may regulate transcription or RNA processing independently of the 7SK snRNP complex. This might be related to its expression in the nucleolus and its function in regulating methylation of U6 snRNA. Another possibility is that Larp7 interacts with other non-coding RNA and regulate their fate. To decipher this, mapping cell type-specific Larp7 interacting RNAs by iCLIP followed by sequencing of associated RNA could reveal potential new targets. Another observation from human patients is the reduced telomere length with can directly impact gene expression, a function that might be conserved in flies. Our results thus indicate that also in *Drosophila* Larp7 likely have additional functions unrelated to the 7SK RNP complex.

Since we were interested in the function of the 7SK snRNP complex itself, we focused on the genes commonly regulated in both mutants or regulated by the loss of *7SK* RNA only. Transcription of a gene can be regulated at different steps. The specificity of gene activation is mainly regulated by sequence-specific binding of transcription factors along with general transcription machinery. In Jurkat cells, it was shown that KAP1 also known as TRIM28, is required for the recruitment of 7SK snRNP complex to the promoter of paused genes. Upon stimulation by the action of NF-kB, the active P-TEFb is released from the 7SK snRNP and initiates transcription elongation (McNamara *et al.*, 2016). An ideal experiment to identify the 7SK snRNP associated transcription factors in motoneurons would be to pull down chromatin-associated 7SK snRNP and perform mass spectrometry. However, our initial attempts to detect Larp7 on chromatin failed, perhaps due to the weak or transient nature of this interaction.

We used a computational approach to identify how the 7SK snRNP specifically targets the set of genes in motoneurons. First, we analyzed the GC content of the promoter of the differentially regulated genes. The genes which are commonly upregulated in *LARP7 / 7SK* KO and *7SK* KO have high GC content between -

100bp and +200bp distance around TSS. Putting our observations together with earlier reports from *Drosophila* embryo showing that highly paused genes have high GC content (Hendrix *et al.*, 2008), suggest that the 7SK snRNP complex inhibit more strongly the expression of highly paused genes. This could be addressed by assessing the pausing state of RNA Pol II in motoneurons using ChIP-sequencing. This experiment became feasible in this last year owing to the development of a new protocol allowing to perform ChIP-seq with low input material (Skene and Henikoff, 2017).

Surprisingly, the commonly downregulated genes in *Larp7* / *7SK* KO and *7SK* KO have low GC content between -200 and -100bp. This specific change may be due to the indirect effect of the upregulated genes or any unknown mechanism by which the 7SK snRNP enhances transcription. It is possible that the 7SK snRNP bring in specific transcription factors to the promoter of specific genes and enhance their transcription.

In contrast to promoter regions, we found an inverse correlation with regards to GC content in the gene body. Metagene analysis of the gene body of the differentially regulated genes showed that up and down-regulated genes in *Larp7 / 7SK* KO and *7SK* KO has low and high GC content from -100bp from TSS until TES, respectively. This unique feature of GC content at the promoter and in the gene body somehow distinguishes genes which are regulated by 7SK snRNP.

The differentially regulated genes in *Larp7* KO condition alone did not show any change in GC content distribution, neither at the promoter nor in the gene body. As discussed above, this suggests that *Larp7* has additional functions which presumably masked the gene architecture requirement with respect to its role within the 7SK snRNP complex.

Finally, we noticed that in addition to the GC content, the size of the differentially regulated genes was also distinct from the average size. We found that the average intron length of upregulated genes was significantly larger than the intron size of non-regulated or downregulated genes.

Gabriel et al. revealed that the position of long introns affects the gene expression level. Computational analysis revealed that in *Drosophila*, genes with the first longest intron have a positive correlation with expression (Marais *et al.*, 2005). It is

possible that the longest intron around the transcription start site contains a regulatory element which is fine-tuning the gene expression. It would be interesting to validate the effect of intron size on gene expression by manipulating the size of the first intron. 7SK snRNP regulated genes may have some regulatory elements in these introns, which can be bound by P-TEFb::7SK snRNP associated factors, thereby releasing paused RNA Pol II, and enhancing transcription.

The promoters of many paused genes are enriched with certain motifs like GAGA, initiator (Inr) and Pause Button (PB) motif (Hendrix *et al.*, 2008). Our study also revealed that the upregulated genes are enriched for GAGA motif (-150bp to TSS), Inr (TSS to +50bp) and PB (-50bp to +50 bp). Interestingly, the TATA motif is enriched in both up and down-regulated genes, but the motif distribution is not similar. In up-regulated genes TATA motifs span within -50bp to TSS around TSS and in down-regulated genes they are present within -200bp to 150bp. We also found enrichment of DPE motifs in up-regulated genes around TSS while DRE motifs are enriched only in down-regulated ones. In summary, upregulated genes are enriched for GAGA, Inr, PB, TATA and DPE motifs around the TSS, while downregulated genes are enriched for TATA and DRE.

Even though we observed specific regulatory elements within the promoter of regulated genes, it is still unclear how these genes are specifically targeted by the 7SK snRNP complex. We hypothesized that 7SK or P-TEFb interact with specific transcription factors to regulate specific gene expression in motoneurons. To identify these potential factors, we performed a *de novo* motif search at the promoter. We identified the SMAD motif as being enriched in the upregulated genes. In *Drosophila* there are four SMAD transcription factors Mother against dpp (Mad), Medea (Med), Nuclear factor I (Nfl) and Smad on X (Smox). Particularly, phosphorylated Mad accumulate at the active zone after activation of type-A glutamate receptor (Sulkowski *et al.*, 2016) (Figure 26). Importantly, Bone morphogenetic protein (BMPs) stimulate type I/II serine/threonine receptors to regulate Smad mediated transcription in the motoneurons (Wu, Xiong and Mei, 2010). Mutation in TGFβ ligands (Smad transcription factors) lead to the reduced number of NMJs and impaired neurotransmitter release which is similar to our observation in the 7SK snRNP mutants (Wu, Xiong and Mei, 2010). This observation raises interesting questions like how the 7SK snRNP complex

specifically regulate SMAD target genes? Does the SMAD protein interact with 7SK snRNP complex directly or through any other partners? How is this specific to the motoneurons?

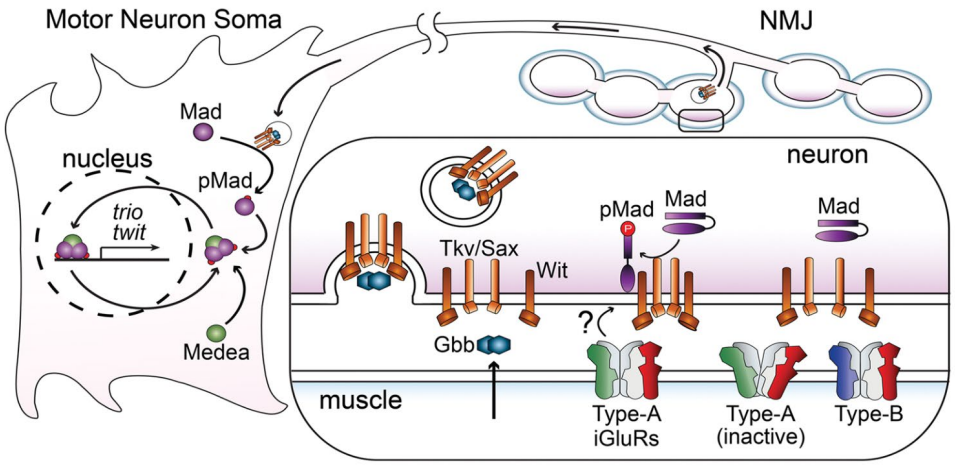



**Figure 26:** Cartoon from Slukowski et al., 2016 showing BMP signaling complexes that control the accumulation of nuclear and synaptic phosphorylated Mad (pMad) and Mad.

## 6.8 The exon junction complex regulates RNA Pol II pausing and splicing of large-intron containing genes.

In addition to my main project studying the function of the 7SK snRNP complex *in vivo,* I contributed to investigating the role of the exon junction complex (EJC) in promoter-proximal pausing of RNA Pol II. This has led to a publication in *Nature Communication*, in which I am a co-author (see attached).

The EJC assembles on RNA upstream of exon-exon junction during the splicing process. The EJC consists of three core proteins: Mago, eIF4AIII and Y14. In *Drosophila*, loss of Mago in the developing eye leads to splicing defect in *MAPK* transcripts, resulting in photoreceptor differentiation defect (Ashton-Beaucage *et al.*, 2010; Roignant and Treisman, 2010). The EJC-dependent splicing changes are seen in large transcripts that are mainly expressed from heterochromatic regions. We also observed that the depletion of core EJC proteins in S2R+ cells lead to decreased promoter-proximal pausing of RNA Pol II, increased Ser2P at CTD of RNA Pol II and a premature entry into elongation. Interestingly, the EJC-dependent splicing events could be rescued by knocking down *Cdk9*, indicating that the splicing defects are a consequence of the premature entry of Pol II into elongation. I could show that the double knock down of Mago and Cdk9 *in vivo* rescued *MAPK* splicing and the associated photoreceptor differentiation defect. I further performed a candidate RNAi screen using photoreceptor differentiation as a readout to identify additional genes that interact with Mago to control *MAPK* splicing. I found that depletion of transcription initiation (*Cdk7*) and elongation factors (SEC, PAF complex and FACT complex) did not rescue the photoreceptor differentiation defect. However, depletion of *Cdk9* and *Cdk12*, as well as the NURF chromatin remodeler, fully rescued photoreceptor differentiation. These results, therefore, suggest that Cdk12 and NURF also play a role in promoter proximal pausing. A part of these data has been included in the published manuscript.

# Materials & Methods

# 7 MATERIALS AND METHODS

## 7.1 *Drosophila* Stocks.

All the flies were maintained at 18˚C and 25˚C in 60% humidity incubator with 12hrs day – 12hrs light cycle and the flies were fed with standard cornmeal food. The genotypes used in this study are listed in table 2.

## 7.2 CRISPR/Cas9 Gene editing in *Drosophila*.

Tools and protocol from the National Institute of Genetics (NIG-Fly) FlyCas9 were used to generate *larp7* and *7SK* gene deletion and eGFP-*Larp7* endogenous tag. Briefly, for gene deletion, a pair of guide RNAs (gRNA) flanking the gene of interest and for endogenous tagging single gRNA close to the transcription start site (TSS) of *Larp7* was designed using "*Cas9 target finder*". The gRNA was synthesised (Merck, Germany) and cloned into a pBFv-u6.2/B vector. For gene deletion, the vector was integrated (PhiC31 integrase) into TBX-0002 ($y^1$ $v^1$ *P{nos-phiC31\int.NLS}X; attP40 (II)* ) and crossed with CAS-0001 ($y^2$ $cho^2$ $v^1$*; attP40{nos-Cas9}/CyO* ). The offspring were individually crossed with respective balancer fly lines, and the parents were screened by PCR (primer sequence in Table 3) for deletion. For endogenous tagging, the donor plasmid was created as follows. Approximately 1000 base pairs (bp) upstream of *Larp7* TSS and 1000 bp downstream of *Larp7* TSS were cloned upstream (BglII) and downstream (KpnI) of eGFP in the pUAST eGFP attB vector (eGFP sequence was cut out). The donor plasmid and gRNA vector pBFv-U6.2 were injected into CAS-0001 embryos. The survived flies were individually crossed with *X* chromosome balancer (*Fm6*,*w)*, after five days or once the offspring L1 larvae hatch, the parent flies were screened by PCR (primer sequence in Table 3) followed by agarose gel electrophoresis. The positive candidates were further verified by PCR-sequencing.

## 7.3 Gene cloning and generation of transgenic flies.

The *Drosophila Larp7* cDNA was synthesised from total RNA of S2 R+ cells and cloned in KpnI site of the pUAST eGFP attB vector, eGFP as N-terminal tag. The Human *Larp7* (*hLarp7*) and *Xenopus laevis Larp7* (*xLarp7*) cDNA were synthesised from HeLa cells and *X.laevis* embryos, respectively. Further, *hLarp7* and *xLarp7* were cloned in KpnI and NotI site of the pUAST-eGFP-attB vector.

The final clones were verified by sequencing and expression of the eGFP-*hLARP7* and eGFP-*xLarp7* was first validated in S2R+ cells by microscopy and immunoblot. For expression in S2R+ cells, the pACT-Gal4 vector was co-transfected to induce the expression of eGFP-*Larp7* under UAS promoter (Brand and Perrimon, 1993). For the *7SK* rescue clone, *7SK* gene and upstream 1500 bps (including transfer RNA: Valine) was amplified by PCR from fly genomic DNA. The amplicon was further cloned into HindIII-EcoRI digested pUAST eGFP attB vector. The cDNA clones were injected into embryos expressing PhiC31 integrase (BestGene Inc., U.S.A) in the genetic background of BDSC #9723 [$y^1$ w1118; PBac{y+-attP-3B}VK00002]. The transgenic flies were selected based on *white* gene expression form the pUAST vector.

## 7.4 Single fly genomic DNA

The single fly protocol (Kumar, 2016) (minor modifications) was used for screening the transgenic flies. Individual flies were collected in 1.5mL centrifuge tubes on ice. Then 50uL of squash buffer (10mM Tris pH8.0, 1mM EDTA, 25mM NaCl) with freshly added 1uL of 10mg/mL Proteinase K was pipetted using a 200uL pipette, and the flies were crushed using the same pipette tip for a few times. Then the sample was incubated at 37°C for 15 min followed by 5 min at 95°C to inactivate Proteinase K. Then spin down the derbies at 14,000 RPM for 5 min, room temperature or 4°C. Then 5uL of the supernatant was used as a template in a 20uL polymerase chain reaction (PCR).

## 7.5 Reverse transcription and quantitative real-time PCR.

Total RNA from required samples were isolated by TRIzol™ reagent. The required amount of total RNA was treated with two units of DNase I (New England Biolabs). For reverse transcription, random oligonucleotides (Sigma, custom made), 200 units of M-MLV reverse transcriptase (Promega), and 18 units of Murine RNase Inhibitor were used according to the manufacturer's protocol. SYBR Green PCR Master Mix and gene-specific primers (Table 3) and normalisation control primers (Rpl15, Rp49, and Act5C) were used to perform quantitative real-time PCR in ViiA 7 Real-Time PCR system (Applied Biosystems). The relative expression level was calculated using Microsoft Excel, and the student’s t-test was performed for biological triplicates.

## 7.6 S2 cells knockdown and transfection.

Knockdown of genes in S2 R+ cells were done according to the protocol (Rogers and Rogers, 2008) with minor modifications. Genomic DNA or cDNA from flies were used as PCR (OneTaq, NEB) template to amplify the exonic region of genes. The dsRNA was synthesised using MEGAscript™ T7 Transcription Kit as manufactures protocol. The knockdown treatments were done either once or three cycles for 7days according to the knockdown efficiency of individual dsRNA, which was first optimised by titration. Knockdown efficiency was analysed by RT-qPCR (primer table 3). For *hLarp7* rescue experiment in S2R+ cells, the cells in six-well plates were treated with dsRNA targeting 3'UTR of *dLarp7* for five days, three treatments and on the 3rd day, the pUAST rescue contract cotransfected with pAct5C-Gal4 in the ratio 4:1 using Effectene reagent as manufactures protocol. The expression of rescue contracts were validated by eGFP fluorescence with a stereo microscope and immunoblot. On Day 5, the eGFP positive cells were sorted (BD FACSAria™ II SORP) into TriZol and performed RT-qPCR to check the level of *7SK* RNA.

## 7.7 RNA sequencing.

The wild-type control, *Larp7* KO and *7SK* KO flies were recombined with UAS-nuclear DsRed, BDSC #8546 (w[1118]; P{w[+mC]=UAS-RedStinger}4/CyO). Then these flies were crossed with RN2-Gal4 (BDSC #7470) or RRK-Gal4 (BDSC #42739), and expression pattern of DsRed was verified by confocal imaging of embryo and L3 larvae. Then *Larp7* KO and *7SK* KO flies were recombined with RN2-Gal4 line. We could not successfully recombine RRK-Gal4 line with *7SK* mutants. Also, we did not see any difference in expression pattern between these two lines. So, RN2-Gal4 line was used in this experiment. The crosses were maintained in a plastic flask with apple agar plate coated with yeast. After two pre-laying every two hours, embryos were collected for 2 hours on an apple agar plate. The staged (stage 15-16) embryos were collected and processed as described (Perrimon, Zirin and Bai, 2011) with minor modifications. The embryos were decorated with 50% bleach (sodium hypochlorite) for 2 min and washed thoroughly with deionised water in 100µM cell strainer. Then transferred to 250 uL Schneider's

*Drosophila* Medium (Gibco) and dissociated using Dounce homogenizer with loose pestle 15 times on the ice. Then the dissociated cells were filtered through the 35µM cell strainer and collected in FACS compatible 5mL tube on ice. To the dissociated cells, DAPI was added to eliminate dead cells during sorting and immediately analyzed in FACS. The purity of the sorting was analyses visually by microscope. Further, RNA was isolated from about 250 DsRed cells, and RT-qPCR was performed to check the marker DsRed and *eve* transcripts. For RNA sequencing, the DsRed positive live single cells were sorted in 250uL TriZol reagent at 4°C. The sorted cells were frozen on dry ice and stored at -80°C. Then the total RNA was isolated according to to the TriZol manufacture's protocol. The quality of the RNA was analysed using Agilent RNA 6000 Pico assay in Agilent 2100 Bioanalyzer and quantified using NanoDrop 3300. Equal concentration of RNA from each biological triplicate were used in Smart-Seq2 Illumina protocol to generate next-generation sequencing libraries.

## 7.8 Raw-data Processing

Libraries were sequenced on an Illumina NextSeq 500 in single read mode with a read length of 84 bp on a single flowcell and sorted according to their index using bcl2fastq (v2.19). Individual samples were mapped first mapped against an rRNA reference to remove ribosomal reads and then against BDGP6 using the annotation of Ensembl release 90. The mapper used was STAR (Dobin et al. 2013, v2.5.2b)). Gene counts were derived using featureCounts (Liao, Smyth, and Shi 2013, v. 1.5.1)).

## 7.9 RNA-Seq Analysis

Differential expression analysis was performed using Bioconductor/DESeq2 (Huber et al. 2015, v2.38; Love, Huber, and Anders 2014, v1.18.1). Genes were deemed sign. Diff. regulated with and FDR below 5%.

## 7.10 GC content

Figure 19 displaying the GC content around the TSS/Metagene are based on the differentially regulated genes derived from the DESeq2 analysis. For each gene, the TSS with the highest coverage and the associated transcript was chosen for display.

• GC content around the TSS was calculated using R/Bioconductor packages (Huber et al. 2015; R Core Team 2018; Lawrence et al. 2013)) and plotted using ggplot2 (Wickham 2016).

• GC content on the metagene profiles were based on GC content per bp retrieved from the genome using bedtools (Quinlan and Hall 2010, v2.27.1) and kentUtils (Kent, n.d.). Then the data were processed further using deeptools (Ramirez et al. 2016, v3.1) and plotted using ggplot2 (Wickham 2016).

## 7.11 Gene Length

Length of genes and introns were derived from Ensembl release 90 using the GenomicFeatures and GenomicRanges packages (Lawrence et al. 2013) and the plotted using ggplot2 (Wickham 2016).

## 7.12 Staging of development.

To analyse the relative expression level of *Larp7* during developmental stages, the total RNA was isolated and performed RT-qPCR as described in (Lence *et al.*, 2016).

## 7.13 Microscopy

### 7.13.1 Live imaging:

Freshly laid embryos on an apple agar plate were collected on adhesive tape, and chorion membrane was removed using forceps. The ventral side of the embryo was stick to glass bottom of the dish (IBIDI, 35mm high glass bottom dish Cat. No#81158) coated with an adhesive (embryo glue). Then the embryos were covered with halocarbon oil (Voltalef, H3S). Live images were captured for every 15-20 min until the last stage of embryo development using VisiScope (Visitron Spinning disc microscope, Germany), GFP channel. For GFP and mCherry co-localization experiment, stage 15-17 embryos were collected and mounted as described before and imaged using Leica SP5 confocal microscope. The images were visualised and analysed in IMARIS 9.2 Bitplane and ImageJ (Schindelin *et al.*, 2012).

### 7.13.2 Expression pattern of Larp7:

Larvae or adult flies were dissected and fixed with 4% paraformaldehyde in PBS (137mM NaCl, 2.7mM KCl, 10mM $Na_2HPO_4$, 1.8mM $KH_2PO_4$, pH 7.4) for 20 min

and washed with PBS 10 min three times. Then the samples were mounted in Vectashield with DAPI. Polytene chromosome squashes were prepared as described (James, 2005) and imaged using SP5 microscope.

### 7.13.3 NMJ immunohistochemistry:

Crosses of the desired phenotype were set up with a maximum of 20 flies per vial in 25°C and 50% humidity. Wandering L3 larvae were collected and segregated according to the genotype. The larva was dissected on a silicone plate (15cm dish) with cold PBS, and internal organs were removed. Then fixed with 4% Paraformaldehyde (PFA) in PBS for 45 min and washed in PBS for one hour. Then the preparations were transferred to 35mm dish with silicone surface and washed six times with PBST (PBS+0.5% Triton-X) for 30 min each wash. Then primary antibody anti-Syt 1:200 (3H2 2D7, Developmental Studies Hybridoma Bank, DSHB) in 0.3% PBST was added and incubated overnight at 4°C. The antibody was removed and washed eight times with PBS for each 30 min. The secondary antibody, anti-mouse-Alexa 488 or anti-mouse-Cy3 (1:500) and Horseradish Peroxidase (HRP)-conjugated TRITC [Tetramethylrhodamine-5-(and 6)-isothiocyanate] (1:1000) was added and incubated in 4°C, overnight. Finally, washed six times with PBS and mounted with Vectashield. NMJ from muscle four was imaged from segment A2 and A3 for NMJ analysis. Nerves were identified using anti-HRP-TRITC staining and anti-Syt (DSHB) to mark the synapses. Z stack for the whole NMJ was imaged with 0.4uM thickness. The images were further analysed using IMARIS 8 image analysis software.

## 7.14 Immunostaining.

Embryos laid overnight by the desired genotype flies were collected on an apple agar plate. Then dechorinate with 50% bleach (sodium hypochlorite) for 2 min and washed with deionised water on 100µM cell strainer. Further, the embryos were transferred to 1.5mL centrifuge tubes containing (1:1) 4% Paraformaldehyde: n-Heptane and placed on a shaker for 20 min at 500 RPM in room temperature. Then spun down at 500 G and the lower phase (fix) was removed, and an equal volume of methanol was added and mixed for one minute. Then washed with methanol for two more times and then stored in -80°C. The fixed embryos were rehydrated with

PBS and incubated in primary antibody in PBSB (PBS + 5% Donkey serum) overnight at 4°C. The embryos were washed with PBT, 30 min, four times and secondary antibody [Alexa 488, Cy5 and Cy3 (1:1000)] was added and incubated overnight at 4°C. Finally, the embryos were washed with PBT and mounted with Vectashield containing DAPI. The embryos were imaged using SP5 confocal microscope.

## 7.15 Crawling assay

Properly staged wandering L3 larvae were collected and briefly washed with water. Then the larva was placed on the centre of the 15cm agar plate (1%). The locomotion was observed for 2 minutes after the recovery time for up to 1 min. The number of peristaltic movement was counted manually by observing the larvae or analysed recorded videos in IMARIS software.

## 7.16 Life span assay

Five days old, 20 flies were collected for each conditions and placed in a vial containing apple agar and a layer of dry yeast and maintained at 25° C, 55% humidity with 12hrs dark-light cycle. Flies were transferred to new food every third day and recorded the number of dead flies. The observation was done until all the flies died in all conditions. Survival graph analysis form PRISM 8 was used to perform statistical analysis. Oasis online tool (Yang *et al.*, 2011) was used to plot the mortality rate between the samples.

## 7.17 Bouton calculation

The NMJ images were acquired as described before. The images were visualized and analyzed in IMARS 8 as follows. The surface application was used to mark all the boutons in the NMJ muscle 4. Then the number of boutons were counted manually and recorded in GraphPad (PRISM 8), and statistical tests were performed according to the distribution of the data.

## 7.18 Immunoblot

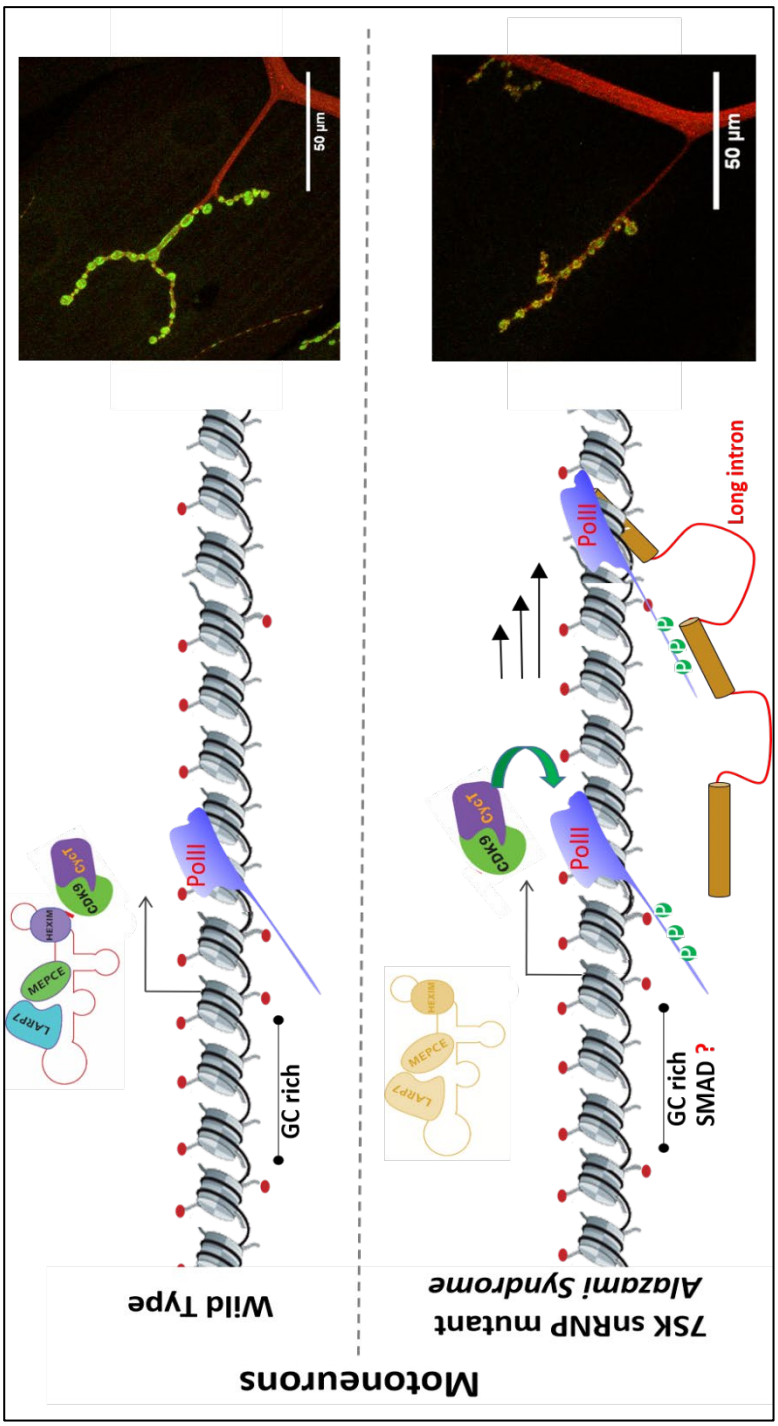
About five heads of adult flies were squashed in 50ul of 2xLDS buffer (NuPAGE) then incubated at 90°C for 10 min, cool down to room temperature and spun at 14,000 RPM for 10 min. 20ul of the sample was loaded in 10% SDS gel and resolved. Further, transferred the proteins onto PVDF membrane and blocked with

5% milk. Then incubate membrane with primary antibody overnight and with secondary antibody for 2 hours and imaged using Pico or Femto reagents (ThermoFisher). The images were captured using a BioRad GelDoc system and quantified using ImageJ software.

## 7.19 RNAi in larvae

To knockdown *Larp7* in flies, we used endogenously tagged eGFP-*Larp7* flies. First eGFP-*Larp7* flies were recombined with UAS-*Dicer* (*X* chromosome) to increase the efficiency of knockdown. Then sh-*GFP* (#41552) in the second chromosome was recombined to get eGFP-*Larp7*,UAS-*Dicer*; sh-*GFP*. The knockdown efficiency of all available GFP RNAi lines was analysed by immunoblot and imaging. sh-RNA line #41552 and #41553 showed the strongest depletion of GFP, but #41553 was less fertile than #41552. Due to this, we used #41552 to for all the knockdown experiments. Then virgin female of eGFP-*Larp7*, UAS-*Dicer*; sh-GFP line crossed with different Gal4 drivers and maintained the crosses in 25°C. The crawling assay was performed with F1 male L3 larvae, for *Cdk9* knockdown #34982 line was used since it shows the highest knockdown efficiency (Akhtar *et al.*, 2019). *dsCdk9* ($3^{rd}$ Chr) was recombined with RRK-Gal4 line and obtained RRK-Gal4, UAS- *dsCdk9* and verified presence of *dsCdk9* by PCR (primer table 3) and sequencing.

# 8. Graphical Abstract



# 9. Article

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-08381-0 OPEN

# Promoter-proximal pausing mediated by the exon junction complex regulates splicing

Junaid Akhtar[1,5], Nastasja Kreim[2], Federico Marini[3], Giriram Mohana[1], Daniel Brüne[1], Harald Binder[4] & Jean-Yves Roignant[1]

Promoter-proximal pausing of RNA polymerase II (Pol II) is a widespread transcriptional regulatory step across metazoans. Here we find that the nuclear exon junction complex (pre-EJC) is a critical and conserved regulator of this process. Depletion of pre-EJC subunits leads to a global decrease in Pol II pausing and to premature entry into elongation. This effect occurs, at least in part, via non-canonical recruitment of pre-EJC components at promoters. Failure to recruit the pre-EJC at promoters results in increased binding of the positive transcription elongation complex (P-TEFb) and in enhanced Pol II release. Notably, restoring pausing is sufficient to rescue exon skipping and the photoreceptor differentiation defect associated with depletion of pre-EJC components in vivo. We propose that the pre-EJC serves as an early transcriptional checkpoint to prevent premature entry into elongation, ensuring proper recruitment of RNA processing components that are necessary for exon definition.

[1] Laboratory of RNA Epigenetics, Institute of Molecular Biology (IMB), 55128 Mainz, Germany. [2] Bioinformatics Core Facility, Institute of Molecular Biology (IMB), 55128 Mainz, Germany. [3] Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), 55101 Mainz, Germany. [4] Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, 79104 Freiburg, Germany. [5] Present address: Institute of Neurobiology and Developmental Biology, JGU, 55128 Mainz, Germany. Correspondence and requests for materials should be addressed to J.-Y.R. (email: j.roignant@imb-mainz.de)

Transcripts produced by RNA polymerase II (Pol II) undergo several modifications before being translated, including 5′-end capping, intron removal, 3′-end cleavage and polyadenylation. These events usually initiate co-transcriptionally while the nascent transcript is still tethered to the DNA by Pol II[1–4]. This temporal overlap is important for the coupling between these processes[5–9]. Initially, Pol II is found in a hypophosphorylated form at promoters. At the onset of initiation, the CTD of Pol II becomes phosphorylated at the Ser5 position. Pol II subsequently elongates and often stalls 20–60 nucleotides downstream of transcription start sites (TSS), an event commonly referred to promoter proximal pausing[10,11]. Promoter proximal pausing of Pol II is widely seen at developmentally regulated genes, and is thought to play critical roles in facilitating rapid and synchronous transcriptional activity upon stimulation[12–17]. Pol II pausing is also suggested to act as a checkpoint influencing downstream RNA processing events such as capping and splicing, but evidence for this function is still limited. The transition from the paused state to elongation is promoted by the positive transcription elongation factor (P-TEFb) complex, which includes the cyclin-dependent kinase 9 (Cdk9) and cyclin T[18–21]. P-TEFb phosphorylates Ser2 of the CTD as well as the negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF), leading to the release of Pol II from promoter[22–24]. Another related kinase, Cdk12, was also recently suggested to affect Pol II pausing after its recruitment through Pol II-associated factor 1 (PAF1)[25,26].

The exon junction complex (EJC) is a ribonucleoprotein complex, which assembles on RNA upstream of exon-exon boundaries as a consequence of pre-mRNA splicing[27,28]. The spliceosome-associated factor CWC22 is essential to initiate this recruitment[29–32]. The nuclear EJC core complex, also called pre-EJC, is composed of the DEAD box RNA helicase eIF4AIII[33], the heterodimer Mago nashi (Mago)[34] and Tsunagi (Tsu/Y14)[35,36]. The last core component, Barentsz (Btz), joins and stabilizes the complex during or after export of the RNA to the cytoplasm[37]. Non-canonical association of Y14 at promoters has also been previously reported, although the significance of this binding remains unknown[38]. The EJC has been shown to play crucial roles in post-transcriptional events such as RNA localization, translation and nonsense-mediated decay[39–41]. These functions are mediated by transient interactions of the core complex with effector proteins[42].

The pre-EJC, along with the accessory factors RnpS1 and Acinus, participate in intron definition[43,44]. In absence of the pre-EJC, many introns containing weak splice sites are retained. The pre-EJC facilitates removal of weak introns by a mechanism involving its prior deposition to adjacent exon junctions. In addition, the depletion of pre-EJC components results in frequent exon-skipping events, particularly at large intron-containing transcripts, although the mechanism is poorly understood[45–47]. In *Drosophila*, loss of Mago in the eye leads to several exon skipping in *MAPK*, resulting in photoreceptor differentiation defects. Other large transcripts, often expressed from heterochromatic regions, show the same Mago-splicing dependency. Similarly, in human, exons flanked by longer introns are more dependent on the EJC for their splicing[47].

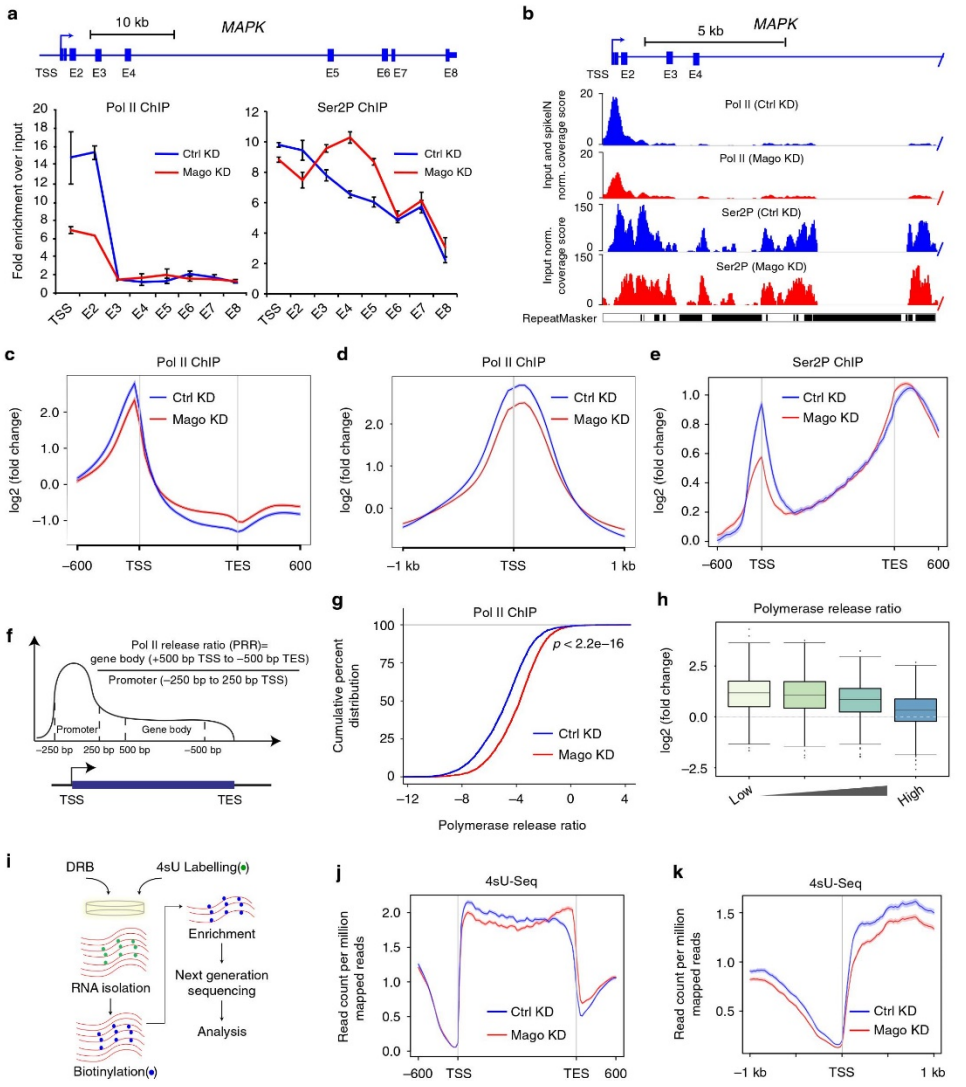
Here, we investigated the mechanism underlying the role of the pre-EJC in exon definition in *Drosophila*. We observed that depletion of pre-EJC components, but not of the EJC splicing subunit RnpS1, lead to a global decrease in promoter proximal pausing, altered Pol II phosphorylation state and premature entry into elongation. These changes are concomitant with underlying changes in chromatin architecture and correlate strongly with exon skipping events. These effects are driven by non-canonical recruitment of pre-EJC components at promoters. Co-immunoprecipitation experiments indicated that Mago associates with Pol II but this association is largely dependent on nascent RNA. Upon knockdown (KD) of pre-EJC components, Cdk9 binding to Pol II is increased, partly accounting for the premature Pol II release. Remarkably, genetically increasing Pol II pausing rescues exon skipping events and the eye phenotype associated with KD of pre-EJC components, indicating that restraining Pol II release into gene bodies is sufficient to complement the loss of pre-EJC components in exon definition. Altogether, our results demonstrate a direct role of the pre-EJC in exon definition via the control of promoter proximal pausing.

## Results

**The pre-EJC regulates expression of long genes**. To investigate the role of the EJC in exon definition, we performed RNAi in *Drosophila* S2R+ cells. As expected, Mago depletion triggered exon skipping in *MAPK* in *Drosophila* cells (Supplementary Figure 1a-c)[45,46]. Further, we found that depletion of other pre-EJC components (eIF4AIII and Y14), but not of the cytoplasmic EJC subunit Btz or the accessory factor RnpS1, strongly impaired *MAPK* splicing and expression of large-intron containing transcripts (Supplementary Figure 1a–c, f, g). In particular, depletion of pre-EJC components led to a higher number of exon skipping events than depletion of Btz or RnpS1 (Supplementary Figure 1h and data not shown). This effect requires pre-EJC assembly as a mutant version of Mago, which is unable to bind Y14, failed to rescue the *MAPK* splicing defect (Supplementary Figure 1d, e). Thus, the pre-EJC is required for proper expression and splicing of large intron-containing genes. In contrast to intron definition, this exon definition activity only slightly required the EJC splicing subunit RnpS1, suggesting a distinct mechanism.

**Lack of pre-EJC alters Pol II phosphorylation**. Introns are spliced while nascent RNA is still tethered to Pol II, allowing coupling between splicing and transcription machineries[6,7,9,48–50]. To address whether the pre-EJC regulates splicing via modulation of transcription, we performed chromatin immunoprecipitation (ChIP) experiments for the different forms of Pol II. We found that Mago KD results in decrease of total Pol II occupancy at the 5′ end of *MAPK* while the distribution in the rest of the gene body was comparable to the control (Fig. 1a). In addition, the elongating Ser2-phosphorylated (Ser2P) form of Pol II was mildly decreased at the TSS, but significantly enriched along the gene body (Fig. 1a). This was specific to Mago depletion and on pre-EJC assembly, as reintroducing WT Mago cDNA, unlike the mutant version, restores the wild type profiles (Supplementary Figure 2a, b). Examining Pol II and Ser2P profiles in a genome-wide manner show extensive changes with decrease at the TSS and increase towards transcription end sites (TES) (Fig. 1b–e and supplementary Figure 2c). Similar changes in Pol II occupancy were observed upon depletion of Y14 and eIF4AIII, especially at the TSS (Supplementary Figure 2d–f), but neither on depletion of RnPS1 nor of Btz (Supplementary Figure 2d–h). Thus, pre-EJC components regulate Pol II distribution genome-wide.

**The pre-EJC facilitates pol II pausing**. To further investigate transcriptional changes in pre-EJC-depleted cells, we analyzed the Pol II release ratio (PRR), which is the ratio of Pol II occupancy between gene bodies and promoter regions (Fig. 1f). Notably, depletion of pre-EJC components, but not of RnpS1, significantly increased the PRR (Fig. 1g and Supplementary Figure 2i,j). Together, these results indicate an unanticipated and specific role for pre-EJC components in promoting promoter-proximal pausing of RNA Pol II. We next divided the changes in PRR into four equal size quartiles, from low to high PRR derived from




Pol II occupancy in WT condition. When classified accordingly, the quartile with the lowest PRRs showed largest increase in PRR upon Mago KD (Fig. 1h), suggesting that strongly paused genes are more affected upon the loss of the pre-EJC.

Next, we performed 4sU-seq to detect nascent transcripts (a modified approach of[51], see Material and methods), (Fig. 1i). The 4sU-Seq metagene profile of Mago KD cells revealed lower read counts at the TSS and an increase towards the 3′ end of transcripts, consistent with the Pol II ChIP-Seq and reduced pausing (Fig. 1j, k). To monitor nascent transcription overtime, we coupled this approach to treatment with the pausing inhibitor 5,6- dichlorobenzimidazole 1-β-D-ribofuranoide (DRB), (4sU-DRB-seq)[52,53]. Our analysis revealed an average elongation rate of 1 kb per minute in *Drosophila* S2 cells, in agreement with previous reports but slower than in human cells (Supplementary Figure 3a–c)[54–57]. Importantly, in contrast to the widespread change in promoter-proximal pausing, the average elongation rate was unaffected in Mago-depleted cells (Supplementary Fig. 3c), and the moderate gene-to-gene variation in elongation rate did not correlate with changes in exon inclusion

**Fig. 1** Mago prevents premature release into transcription elongation. **a** ChIP-qPCR analysis of Pol II and Ser2P occupancies at *MAPK* locus. The tested regions for enrichment are shown in the scheme. Error bars indicate the standard deviation of three biological replicates. **b** Track examples of total Pol II and Ser2P ChIP-Seq from S2R+ cells extracts, after either control or Mago knockdown. The tracks are average of two independent biological replicates after input and "spike-in" normalization. Shown here are the profiles on *MAPK*, a well-described pre-EJC target gene. **c**, **d** Metagene profiles of averaged total Pol II occupancies from two independent biological replicates after "spike-in" normalization in control and Mago-depleted cells along with standard error of mean for all the expressed genes, −600 bp upstream of transcription start sites (TSS) and +600 bp downstream of transcription end sites (TES) (**c**); or centered at the TSS in a ±1 Kb window (**d**). Log2 fold changes against input control are shown on *Y*-axis, while *X*-axis depicts genomic coordinates. **e** Metagene profiles of averaged Ser2P occupancies in control and Mago-depleted cells of two independent biological replicates also displaying standard error of mean for all expressed genes. Log2 fold changes against input control are shown on *Y*-axis, while *X*-axis depicts scaled genomic coordinates. **f** Schematic representation of the calculation of the Pol II release ratio (PRR). The promoter is defined as 250 bp upstream and downstream of TSS, while the gene body is 500 bp downstream of TSS to 500 bp upstream of transcription end site (TES). **g** The empirical cumulative distribution function (ECDF) plot of computed PRR in control and Mago knockdown conditions, after "spike-in" normalization. *p*-value is derived from two-sample Kolmogorov-Smirnov test. **h** Box plots showing changes in PRRs upon Mago depletion when compared to control, separated into different PRR quartiles. The quartiles were generated for genes that are Pol II bound in both control and Mago knockdown conditions. **i** Schematic depiction of the DRB-4sU-Seq approach. **j** Metagene profile of nascent RNA from non-DRB treated 4sU-Seq data in control and Mago-depleted cells, with standard error of mean for all the expressed genes. Averaged read counts per million of mapped reads of two independent biological replicates from 4sU-Seq are shown on *Y*-axis while *X*-axis depicts scaled genomic coordinates. **k** Metagene profile of nascent RNA from non-DRB treated 4sU-Seq data in control and Mago-depleted cells with standard error of mean for all the expressed genes based on the average of two independent biological replicates, centered at the TSS in a ±1 Kb window. Nascent RNA was fragmented to ≤100 bp during enrichment. Averaged read counts per million of mapped reads of two independent biological replicates from 4sU-Seq are shown on *Y*-axis while *X*-axis depicts genomic coordinates

(Supplementary Figure 3d). Altogether, our data suggest that the pre-EJC controls Pol II pausing but does not significantly affect the elongation rate.

To better dissect the role of pre-EJC in Pol II pausing we examined *heat shock* (*hsp*) genes, which possess a promoter-proximal Pol II that has been extensively characterized[58]. We performed Pol II ChIP-qPCR to monitor Pol II occupancy before, during and after HS on the *Hsp70Aa* gene. We found that Pol II occupancy at the 5′ end of the gene was higher in control cells compared to Mago-depleted cells before HS (Supplementary Figure 4). During HS, Pol II occupancy rose dramatically and the extent of this increase was similar in control versus Mago KD, suggesting that Mago has no impact on transcription initiation. However, during recovery after HS, Pol II occupancy remained high at the 5′ end of the gene in control condition but was significantly lower in the Mago KD. These results thus further suggest that the pre-EJC is specifically involved in the control of Pol II pausing rather than in transcription initiation.

**Pre-EJC components associate at promoter regions**. To investigate how the pre-EJC controls Pol II pausing, first we evaluated the expression of known pausing factors. Depletion of pre-EJC components did not affect the expression of Cdk9, Spt5, subunits of the NELF complex, GAGA, Med26, TFIID (Supplementary Figure 5a–d). To test whether the pre-EJC might itself associate with chromatin, as suggested by previous immunostaining of pre-EJC components on polytene chromosomes of *Drosophila* salivary glands[38], we performed ChIP-Seq experiments. We observed genome-wide enrichment of HA-tagged pre-EJC components, but not of RnpS1-HA, primarily at promoters of expressed genes (Fig. 2a–c and Supplementary Figure 6a, c). Mago depletion reduced the Mago-HA enrichment, demonstrating the specificity of the signal (Supplementary Figure 6c). The degree of overlap between the bound targets of pre-EJC components was 34%, corresponding to 816 genes (Supplementary Figure 6d).

Next, we tested whether Mago might associate with promoters via an interaction with RNA Pol II. We found that Flag-tagged Mago bound to Pol II by co-IP (Fig. 2d). Importantly, FLAG-Mago interacted with Pol II Ser5P but not with elongating Ser2P, potentially explaining the enrichment of pre-EJC binding at promoter regions. In addition, interaction with Pol II was reduced after treatment with RNase T1, indicating that a RNA intermediate facilitates this association (Fig. 2d). Accordingly, most of Mago binding to promoters was lost when the chromatin was treated with RNAse T1 prior to immunoprecipitation (Fig. 2e and Supplementary Figure 6a, 6c). Furthermore, only mRNAs whose corresponding genes were bound by pre-EJC co-immunoprecipitated with Mago-HA, including intronless transcripts, indicating that in contrast to canonical EJC deposition, association of Mago at promoters can occur independently of pre-mRNA splicing (Supplementary Figure 6e). Further, the association of Mago-HA with the TSS was not substantially affected by the depletion of the spliceosome-associated factor CWC22 (Supplementary Figure 6f, g) or by treatment with a splicing inhibitor on intronless genes. Nevertheless, applying the same condition on intron-containing genes significantly reduced Mago binding, suggesting that splicing can contribute to Mago enrichment at promoters (Supplementary Figure 6g). Finally, pre-treatment with a general inhibitor of RNA Pol II such as α-amanitin reduced Mago binding at TSS (Supplementary Figure 6h). A similar result was obtained when Pol II initiation was blocked using Triptolide treatment, while preventing Pol II elongation after pausing using DRB did not alter Mago binding (Supplementary Figure 6i, j). Collectively our data suggest that pre-EJC components bind to promoters via Ser5P Pol II, and that nascent RNA is required to stabilize this interaction.

**Pre-EJC binding to nascent RNA increases Pol II pausing**. To define the relationship between pre-EJC-binding and promoter proximal pausing, we evaluated all genes bound by pre-EJC components ($n = 816$) by several criteria. First, heatmaps show a positive correlation between Pol II occupancy and pre-EJC binding at TSS (Fig. 2e). Consistently, the proportion of Mago or pre-EJC-bound genes was higher at highly expressed genes (Fig. 2f and Supplementary Figure 6k). Second, we noticed that Mago was highly enriched at the TSS of strongly paused genes, which have a low PRR (Supplementary Figure 6l). Lastly, we found a positive correlation between pre-EJC binding and changes in Ser2P levels upon Mago KD (Fig. 2g, $p < 2.2 \times 10^{-16}$). Altogether these results suggest that pre-EJC binding to promoters might modulate promoter-proximal pausing.

To examine whether the pre-EJC is sufficient to promote Pol II pausing, we tethered Mago to the 5' end of a nascent RNA via the λN-boxB heterologous system[59]. Compared to λN alone, ectopic expression of λN-Mago ($p = 9.5 \times 10^{-5}$) led to increased enrichment of Pol II at the luciferase promoter and a slight depletion of

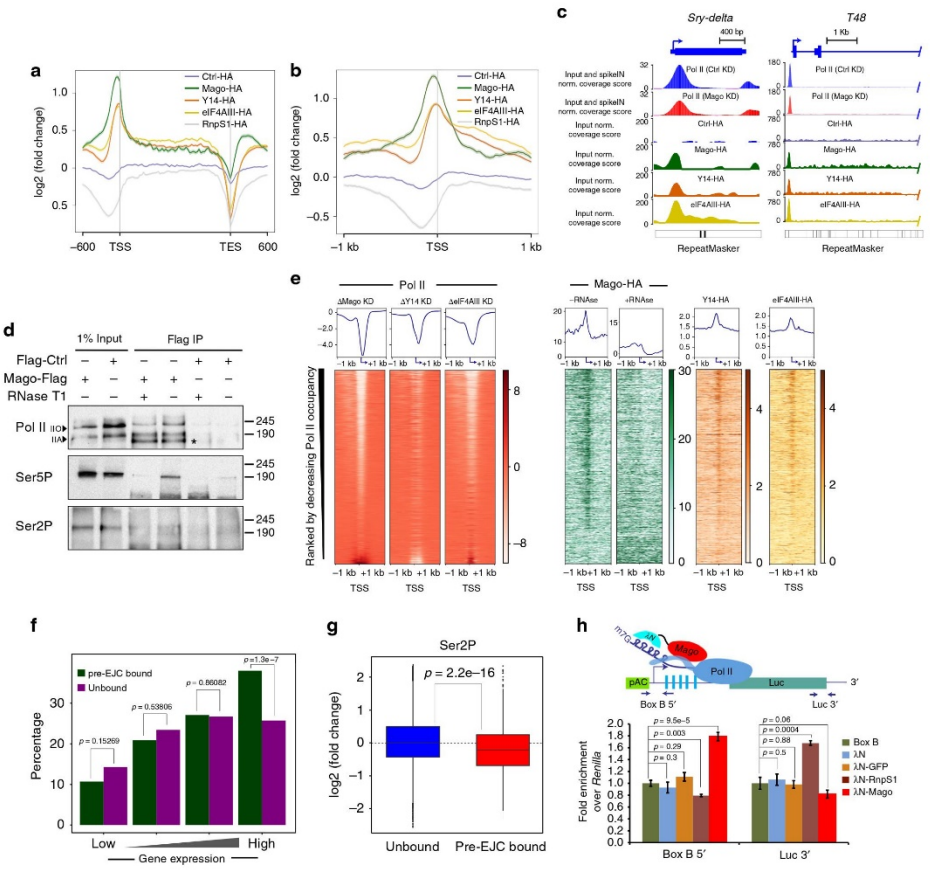



**Fig. 2** Mago binding to promoter regions modulates Pol II pausing. **a** Metagene profiles of ChIP-Seq performed with HA-tagged Mago, Y14, eIF4AIII, and Ctrl, with standard error of the mean for all the expressed genes based on averaged enrichment over input for two independent biological replicates. Log2 fold changes against input control are shown on *Y*-axis while *X*-axis depicts scaled genomic coordinates. **b** Metagene profiles with standard error of mean based on average enrichment over input for two independent biological replicates of ChIP-Seq performed with HA-tagged Mago, Y14, eIF4AIII, and Ctrl, centered at the TSS in a ±1 Kb window. Log2 fold changes against input control are shown on *Y*-axis while *X*-axis depicts genomic coordinates. **c** Input normalized and replicate averaged track examples of ChIP-Seq experiments from S2R+ cell extracts transfected with HA-tagged Mago, Y14, eIF4AIII, or Ctrl. Shown here is recruitment of pre-EJC components to an intronless (*Sry-delta*) and intron-containing (*T48*) genes. **d** Co-immunoprecipitation using anti-Flag antibody from cell extracts expressing either Flag-Mago or Flag alone (Flag-Ctrl), revealed with total Pol II, Ser5P and Ser2P antibodies. Note that Mago interacts with total Pol II, both hypo (IIA) and hyper phosphorylated (IIO) forms (indicated with the arrowheads). Mago also interacts with Ser5P but not with Ser2P, and this interaction with Pol II is partially dependent on RNA. **e** Heatmaps of HA-tagged pre-EJC components and change in total Pol II, centered at the TSS (−1 kb to +1 kb). Rows indicate all the genes bound by Pol II and are sorted by decreasing Pol II occupancy. The color labels to the right indicate the levels of enrichment. **f** Histogram showing percentage of pre-EJC bound genes amongst different quartiles of genes expressed in control condition. For quartile classification, all of the expressed genes in S2R+ cells were divided into four equal sized quartiles according to the level of expression, from low to high level. *p*-values for significance of the associations, derived from Fisher's exact test, are shown on top of the histogram. **g** Log2 fold changes in Ser2P level at the TSS upon control and Mago knockdowns, when separated according to pre-EJC binding. *p*-value is derived from a two-sample *t*-test. **h** Recruitment of Mago at the 5′ end of RNA is sufficient to induce pausing. (Top) Schematic of the BoxB-λN tethering assay. BoxB sequences (blue rectangles) were inserted upstream of the CDS of Firefly luciferase (green rectangle). The λN peptide (blue) was fused to Mago (shown in red), GFP or RnpS1, and transfected into S2R+ cells along with the modified Firefly luciferase plasmid as well as with a *Renilla* luciferase construct. (Bottom) Quantification of the ChIP experiment. Chromatin was prepared for the different conditions and followed by immunoprecipitation using antibody directed against total Pol II. The enrichment of Pol II at the promoter and at the 3′ end of Firefly luciferase was calculated after normalizing against a negative loci and *Renilla*. The enrichment for three independent biological replicates is shown along with *p*-values for tested conditions

Pol II at the 3′ end of luciferase (Fig. 2h). In contrast, ectopic expression of λN-GFP had no effect, whereas expression of λN-RnpS1 had an opposite effect regarding Pol II occupancy (Fig. 2h). These tethering experiments were repeated using additional genes selected on specific criteria: *piwi*, whose splicing is dependent on pre-EJC[43,44]; BBS8, which is unbound by the pre-EJC but has a paused Pol II; Crk, which is unbound by the pre-EJC and does not have a paused Pol II. In all conditions, tethering Mago to their 5′ UTR increased Pol II occupancy at 5′ end of the corresponding locus. Moreover, the decrease in Pol II occupancy observed upon Mago KD could be rescued by tethering Mago but not the GFP control (Supplementary Figure 7a). Thus, Mago recruitment to the 5′ end of nascent RNA is sufficient to increase promoter-proximal pausing of RNA Pol II at the corresponding locus, irrespective of whether the endogenous gene is bound by the pre-EJC.

**Loss of Mago results in changes in chromatin accessibility.** Transcription is tightly coupled to chromatin architecture[60]. To address whether KD of pre-EJC components affects chromatin organization, we performed MNase-Seq. We observed an increase in nucleosomal occupancy at the TSS upon depletion of Mago (Fig. 3a, b), consistent with elevated promoter-proximal pausing and with previous reports that paused Pol II competes with nucleosomes at TSS[60]. Furthermore, the phasing of nucleosomes within the gene body was strongly altered upon Mago depletion (Fig. 3a, b). Pre-EJC-bound promoters showed the most significant changes, consistent with a direct effect (Fig. 3c, $p < 2.2 \times 10^{-16}$). We also detected a mild but significant negative correlation between changes in Ser2P enrichment and chromatin accessibility (Fig. 3d, coefficient of determination $R^2 = -0.2274$). Mago KD also led to depletion of the activating histone mark H3K4me3, in particular at pre-EJC-bound genes (Fig. 3e, $p < 2.2 \times 10^{-16}$). Therefore, Mago modulates histone marks and chromatin accessibility, likely via its promoter-proximal pausing activity.

**Pre-EJC gene size dependency is mediated transcriptionally.** Depletion of pre-EJC components primarily affected the expression of genes containing larger introns (Supplementary Figure 1). We hypothesized that the underlying transcriptional changes upon Mago depletion might drive this size dependency. Indeed, Mago depletion led to an intron-size dependent increase in nucleosomal occupancy at promoters, and decrease in nucleosome occupancy along the gene body and at the TES (Fig. 4a). In contrast, Ser2P enrichment displayed anti-correlative changes with respect to the nucleosomal occupancy (Fig. 4b). Further, the increase in PRR upon Mago depletion also correlated with intron size (Fig. 4c). Thus, Mago has a stronger impact on the transcriptional regulation of genes with longer introns than genes with shorter introns. To determine whether these changes in nucleosomal and Ser2P occupancies result from pre-EJC binding, we calculated the percentage of genes bound by the pre-EJC in different classes relative to their representation in the total number of expressed genes. Interestingly, we found that pre-EJC binding was significantly over represented at genes containing longer introns (Fig. 4d, $p < 2.2 \times 10^{-16}$). Consistent with a direct control of gene expression by pre-EJC components on long intron-containing genes, we found that expression of pre-EJC-bound genes was significantly decreased upon KD of pre-EJC components (Fig. 4e–g, $p < 2.2 \times 10^{-16}$) and this decrease was also largely observed at nascent RNA (Fig. 4h, $p < 2.2 \times 10^{-16}$). Collectively, our results suggest that pre-EJC components preferentially bind and regulate the expression of large intron-containing genes via a direct transcriptional effect.

**Mago restricts P-TEFb binding to Pol II.** The P-TEFb complex induces Pol II release by promoting NELF and Ser2 phosphorylation of Pol II. To determine whether pre-EJC components influence pausing through an interplay with NELF we re-analyzed the publicly dataset available from the ref. [60]. We first noticed that NELF binds substantially more genes in comparison to the pre-EJC (3796 vs. 816), and that 45% of pre-EJC-bound genes does not overlap with NELF binding (Supplementary Figure 8a). Furthermore, like Mago, NELF-bound genes are over-represented on highly expressed genes (Supplementary Fig. 8b, c) and on promoters that are strongly paused (Supplementary Figure 8d, e). Accordingly, NELF KD affects PRR more strongly on highly paused genes (Supplementary Figure 8f, g). Given the similarity of NELF and Mago on pausing we tested whether their binding to promoters was dependent on each other. However, we found only minor effect on their binding upon the respective KD (Supplementary Figure 8h, j). Furthermore, NELF does not bind *MAPK* and its depletion had no effect on *MAPK* splicing (Supplementary Figure 8k), strongly suggesting independent mode of actions.

To determine whether the pre-EJC influences pausing by regulating P-TEFb occupancy, we monitored occupancy of Cdk9 via DamID[61,62]. We expressed N-terminally tagged Dam-Cdk9 in control and Mago-depleted S2R+ cells and observed increased Cdk9 enrichment at the TSS upon Mago depletion (Fig. 5a, b). Furthermore, the increase in Cdk9 enrichment correlated with Pol II occupancy (Fig. 5c). To validate the increased occupancy of Cdk9 in the absence of Mago we also performed ChIP-qPCR. In agreement with the DamID result, Cdk9 occupancy was increased at the 5′ end of *MAPK* (Fig. 5d). Importantly, Mago KD did not alter Cdk9 levels, indicating that this increased enrichment was not due to changes in protein expression (Supplementary Figure 9a). To evaluate whether the change in Cdk9 occupancy was directly driven by Mago occupancy, we analyzed pre-EJC-bound and unbound genes. The increase in Cdk9 enrichment for pre-EJC-bound class was mild albeit significantly higher than the unbound class (Fig. 5e, $p = 0.02508$). These data suggest that pre-EJC binding at the TSS controls Ser2 phosphorylation and Pol II pausing by restricting Cdk9 recruitment.

To address whether Mago inhibits P-TEFb recruitment by restricting its binding to Pol II, we evaluated the association of Cdk9 with Ser5P Pol II. We immunoprecipitated HA-SBP-tagged Cdk9 from control and pre-EJC KD cells and observed a substantial increase in the interaction between Cdk9 and Ser5P Pol II upon Mago depletion (Fig. 5f). Similar results were obtained upon KD of other pre-EJC components. Thus, these data strongly suggest that the pre-EJC restricts binding of P-TEFb to Pol II, which in turn reduces Ser2P levels and the entry of Pol II into elongation.

**Reducing Pol II release rescues Mago defects in vivo.** We hypothesized that reduced Pol II pausing upon Mago KD accounts for some of the increased exon skipping. To test this hypothesis, we attempted to rescue the splicing defects by decreasing the release of Pol II into gene bodies via simultaneously depleting Cdk9. We found that Cdk9 KD restored Ser2P levels upon Mago depletion (Fig. 6a) and partially rescued Ser2P occupancy at the *MAPK* gene (Fig. 6b). Further, the dependence of gene expression on intron size observed upon Mago KD was lost in the double KD (Supplementary Figure 9b). These data suggest that Cdk9 and Mago antagonistically regulate transcription. Importantly, reducing Cdk9 levels almost fully rescued *MAPK* splicing (Fig. 6c–e) as well as other Mago-dependent exon skipping events (Fig. 6f, $p = 8 \times 10^{-8}$) in Mago KD cells. Consistent with the pre-EJC influencing splicing via modulation of

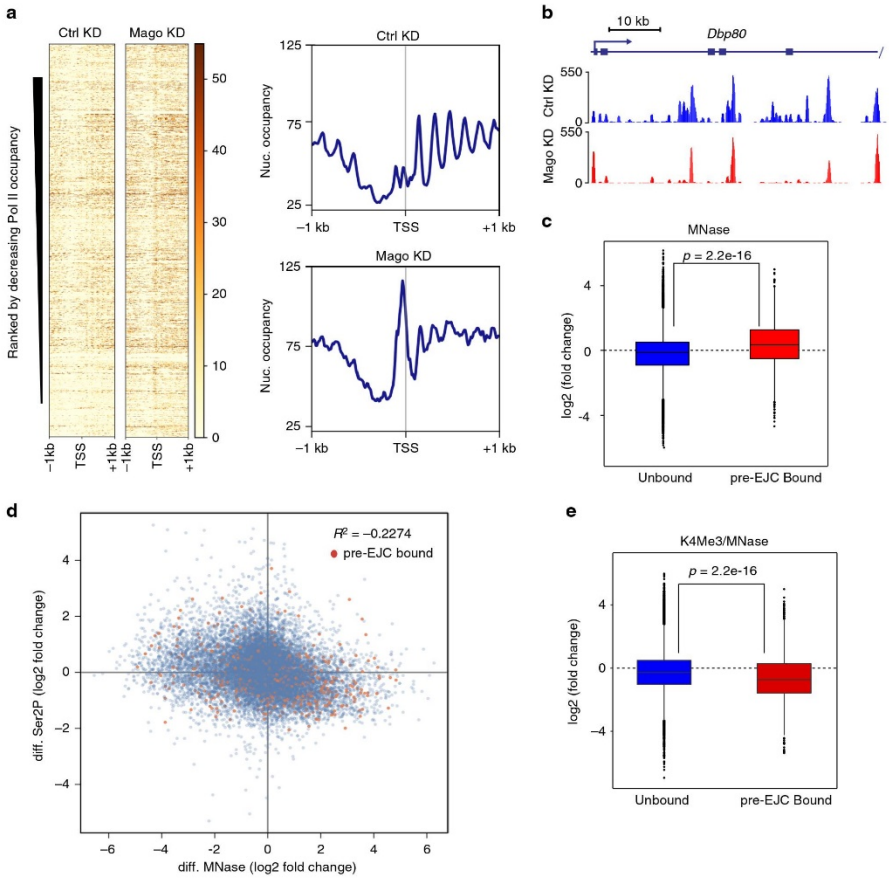
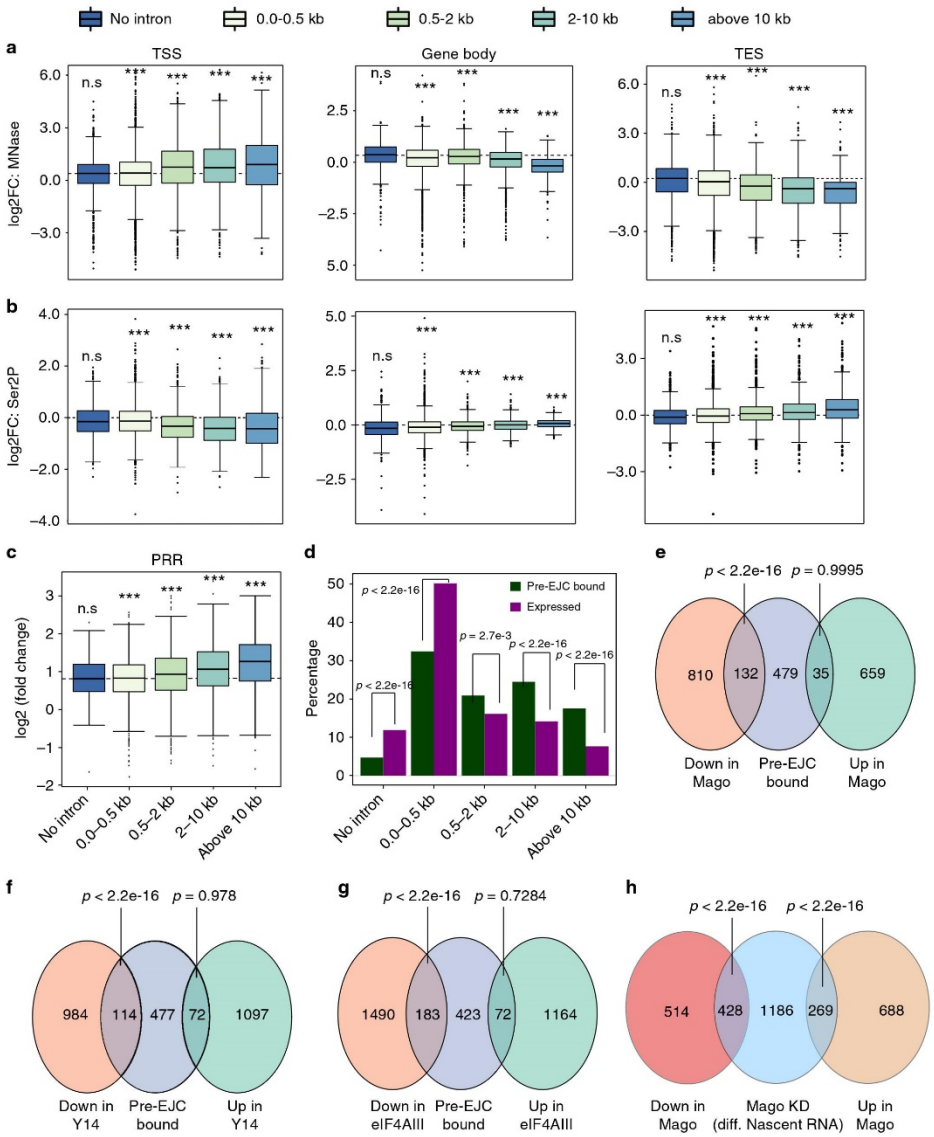



**Fig. 3** Mago controls chromatin accessibility. **a** Heatmaps of nucleosomal occupancy from MNase-seq experiments performed in biological duplicates from S2R+ cells in control or Mago knockdown conditions centered at the TSS in a ±1 Kb window. Rows indicate all the genes bound by Pol II and are sorted by decreasing Pol II occupancy. The color labels to the right indicate the levels of nucleosomal occupancy. Composite metagene profiles are also shown, with nucleosomal occupancy level on *Y*-axis and genomic coordinates on *X*-axis. **b** Genome browser view of averaged enrichment MNase-seq data from two independent biological duplicates in S2R+ cells treated with control or Mago double stranded RNA. The example shown here is *Dbp80* gene, a well-defined pre-EJC target. **c** Log2 fold changes in nucleosomal occupancy at the TSS (250 bp upstream and downstream of TSS) after control and Mago knockdowns. The changes were separated according to pre-EJC binding and a two-sample *t*-test was performed. **d** Scatterplot between changes in nucleosome and Ser2P occupancy after either control or Mago knockdown. Pre-EJC-bound promoters are highlighted by orange color. A mild negative correlation, as shown in the indicated pearson coefficient of correlation, between nucleosome and Ser2P occupancies was found. **e** Log2 fold changes in K4Me3 levels normalized to the nucleosomal occupancy (MNase data) at the TSS (250 bp upstream and downstream of TSS), after control and Mago knockdowns. The changes were separated according to pre-EJC binding and a two-sample *t*-test was performed

promoter-proximal pausing, we found that genes that display differential splicing upon depletion of pre-EJC components were significantly enriched for pre-EJC binding (Supplementary Figure 9c–e, Fisher's test $p < 2.2 \times 10^{-16}$). Futhermore, the tethering of Mago to the 5′ end of piwi or Crk that increased Pol II pausing was sufficient to rescue splicing defects associated with the Mago KD (Supplementary Figure 7b). Lastly, depletion of Cdk12, another kinase involved in the release of promoter-proximal pausing[26], also rescued *MAPK* splicing of Mago-depleted cells (Supplementary Figure 9f, g). Altogether, these results strongly suggest that Mago regulates gene expression and exon definition via regulation of Pol II promoter-proximal pausing.

*MAPK* is the main target of the EJC during *Drosophila* eye development[46]. As shown previously, eye-specific depletion of Mago strongly impairs photoreceptor differentiation[45,46]. Strikingly, decreasing Cdk9 function in a similar background rescued eye development (Fig. 6g). Notably, the number of differentiated




photoreceptors in larvae and adults was substantially increased. We observed a similar rescue of photoreceptor differentiation in a double KD for Mago and Cdk12 (Supplementary Figure 10a). Further, depletion of Cdk9 rescued the lethality and the eye defects associated with eIF4AIII KD (Supplementary Figure 10b). In contrast, reducing the speed of Pol II or depleting several transcription elongation factors failed to substantially rescue eye development in the absence of Mago (Supplementary Figure 10c–f), providing additional evidence that Mago transcriptional function occurs at the level of promoter-proximal pausing rather than at the transcription elongation stage. Thus, despite the numerous post-transcriptional functions of the EJC, modulating Pol II release is sufficient to rescue the eye defect associated with pre-EJC depletion.

**Fig. 4** Mago controls chromatin accessibility and Pol II pausing in a gene size dependent manner. **a**, **b** Changes in nucleosome occupancy (**a**) and Ser2P levels (**b**) in control and Mago knockdown in S2R+ cells at the promoter, gene body, and TES, separated according to the size of the largest intron (ANOVA, $p < 2.2 \times 10^{-16}$). **c** Box plots showing changes in PRRs upon Mago depletion when compared to control, separated into different intron classes. **d** Percentage of pre-EJC-bound genes in different intron classes, along with the percentage of each class amongst all the expressed genes. The proportion of pre-EJC-bound genes is increased for genes containing larger introns, relative to their abundance. *p*-values are derived from Fisher's exact test. **e**–**g** Venn diagrams showing the overlap between Mago-bound, Y14-bound, and eIF4A3-bound genes and genes with differential expression upon respective knockdowns. The overlap between genes that are downregulated upon each pre-EJC component knockdown, and their respective target genes is significant. *p*-values are derived from Fisher's exact test. **h** Venn diagram showing the overlap between genes identified as differentially expressed in mRNA sequencing and nascent RNA sequencing (4sU-Seq) upon Mago KD. *p*-values are derived from Fisher's exact test, separated for either upregulated or downregulated genes

**The function of Mago in Pol II pausing is conserved**. To determine if EJC-mediated promoter-proximal pausing is conserved in vertebrates, we investigated the function of Magoh, the human ortholog of *Drosophila* Mago. We found that depletion of Magoh in HeLa cells led to an increased release of Pol II from the promoter to the gene body, and in turn to a higher PRR (Fig. 7a–d), as well as higher level of Ser2P, but not of Ser5P (Fig. 7e). Additionally, Magoh specifically interacted with Pol II and Ser5P, but not Ser2P (Fig. 7f–h). Finally, we immunoprecipitated Cdk9 from control and Magoh KD cells, and observed a stronger interaction between Ser5P and Cdk9 upon depletion of Magoh (Fig. 7i). Thus, the function and mechanism of the pre-EJC in the control of promoter proximal pausing is conserved in human cells.

## Discussion

Our work uncovers an unexpected connection between the nuclear EJC and the transcription machinery via the regulation of Pol II pausing, which is conserved from flies to human. The pre-EJC stabilizes Pol II in a paused state, at least in part, by restricting the association of P-TEFb with Pol II via non-canonical binding to promoter regions. The premature release of Pol II into elongation in absence of the EJC results in splicing defects, highlighting the importance of this regulatory step in controlling downstream RNA processing events (Fig. 7j).

Promoter proximal pausing is a widespread transcriptional checkpoint, whose functions and mechanisms have been extensively studied. Several regulators have been identified, which includes P-TEFb, NELF and DSIF. Our data reveal that the pre-EJC plays a similar role as the previously described negative factors by preventing premature Pol II release into elongation. How does the pre-EJC control Pol II pausing and how does it interplay with other pausing regulators? Our study provides some answers to these questions. In absence of pre-EJC components, P-TEFb associates more strongly with Pol II, which results in increased Ser2 phosphorylation, demonstrating that one of the activities of the pre-EJC is to restrain P-TEFb function by diminishing its association with chromatin. While it is not clear yet how the pre-EJC exerts this function, a simple mechanism would be by steric interference for Pol II binding, although more indirect mechanisms might also exist. This mechanism infers that both the pre-EJC and Cdk9 bind similar sites on the CTD on Pol II, which fits with the association of the pre-EJC with the Ser5 phosphorylated form of Pol II and not with Ser2P. However, we also found elevated Cdk9 binding and premature release of Pol II at Mago-unbound genes, albeit to a lesser extent compared to Mago-bound genes, suggesting that additional mechanisms must be involved.

It is interesting to note that the binding of the pre-EJC to Pol II requires the presence of nascent RNA. A recent study also supports these findings showing specific association of pre-EJC components on polytene chromosomes that depends on nascent transcription but is independent of splicing[38]. This is reminiscent to the binding of DSIF and NELF[63–68], suggesting that interaction with Pol II and stabilization via nascent RNA is a general mechanism to ensure that pausing regulators exert their function at the right time and at the right location. Upon external cues, P-TEFb modifies the activities of both NELF and DSIF through phosphorylation, promoting Pol II release. It would be of interest to address whether P-TEFb also regulates the EJC in a similar manner. Intriguingly, previous studies revealed that eIF4AIII is present in the nuclear cap-binding complex[69], while Y14 directly recognizes and binds the mRNA cap structure[70,71]. It is therefore possible that this cap-binding activity confers the ability of the EJC to bind nascent RNA. Consistent with this hypothesis, the KD of Cap binding protein (Cbp) strongly reduced association of Mago to chromatin (Supplementary Figure 8h). Nevertheless, since Cbp is also required to stabilize transcripts, the reduced Mago binding might result from this confounding effect. Moreover, other factors must be clearly involved as only a subset of genes is bound by the pre-EJC.

SRSF2 is another splicing regulator that was previously demonstrated to modulate Pol II pausing via binding to nascent RNAs[72]. In this case, SRSF2 exerts an opposite effect by facilitating Pol II release into the elongation phase. This effect occurs via increased P-TEFb recruitment to gene promoters. Although we have not found convincing evidence for a conserved role of the *Drosophila* SRSF2 homolog in this process (unpublished data), one may envision that the pre-EJC counteracts the effect of SRSF2 to stabilize Pol II pausing. Consistent with this possibility, EJC binding sites are often associated with RNA motifs that resemble the binding sites for SR proteins[73]. It is therefore possible that SR proteins influence pre-EJC loading to mRNA and vice versa.

Our previous work along with studies from other groups suggested that the EJC modulates splicing by two distinct mechanisms[43–46]. On one hand, the EJC facilitates the recognition and removal of weak introns after prior deposition to flanking exon-exon boundaries. We proposed that EJC deposition occurs in a splicing dependent manner after rapid removal of bona fide introns, which are present in the same transcript. Thus, a mixture of "strong" and "weak" introns ensures EJC's requirement in helping intron definition. This function requires the activity of the EJC splicing subunits Acinus and RnpS1, which are likely involved in the subsequent recruitment of the splicing machinery near the weak introns. While this model is attractive, it does not however explain every EJC-regulated splicing event. In particular, depletion of pre-EJC components results in a myriad of exon-skipping events, which occur frequently on large intron-containing transcripts (this study and refs. [45,46]). In contrast to intron definition, this exon definition activity only slightly required the EJC splicing subunits, suggesting an additional mechanism. We now show that the pre-EJC controls exon definition at least in part by preventing premature release of Pol II into transcription elongation. Our results shed light on a recent

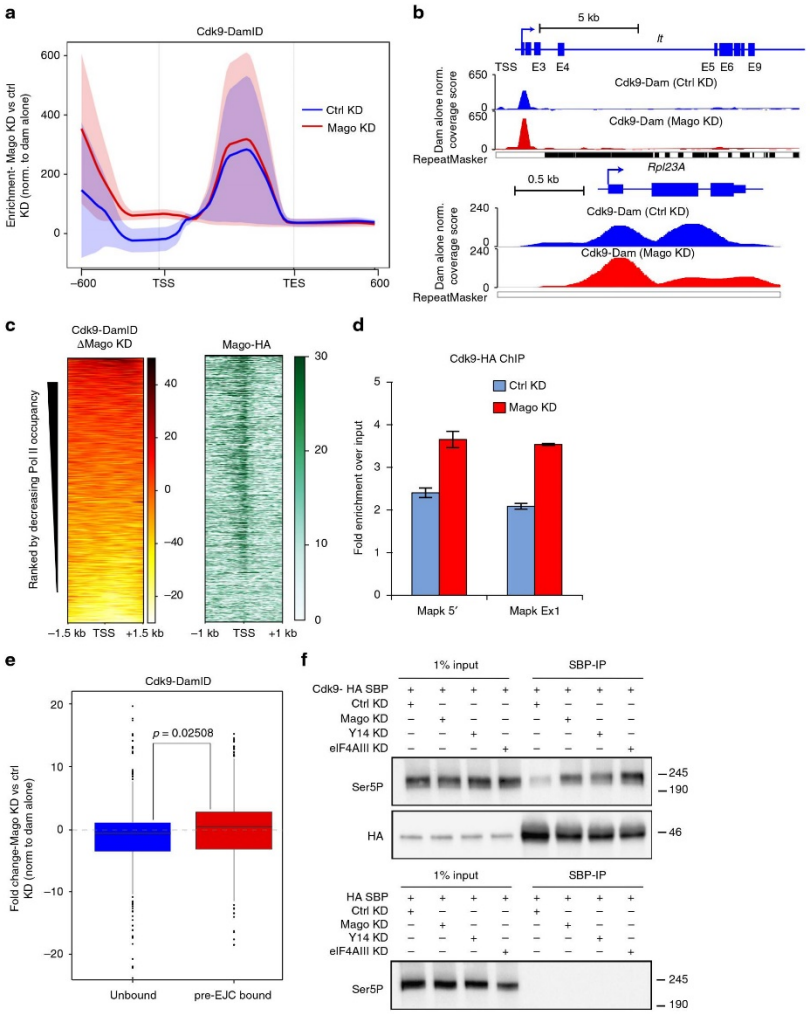



**Fig. 5** Loss of pre-EJC core components results in increased Cdk9 binding to Pol II. **a** Metagene profiles of Cdk9-DamID in control and Mago knockdown cells as averaged enrichment over input from two independent biological replicates with standard error of the mean for all the expressed genes. Fold enrichment represented on *Y*-axis was calculated against Dam alone control in respective conditions (using damidseq_pipeline), while *X*-axis depicts scaled genomic coordinates. **b** Genome browser view of Dam alone normalized and averaged tracks of Cdk9-DamID for *light* and *Rpl23A*. **c** Heatmaps of changes in normalized Cdk9-DamID enrichment in Mago depleted S2R+ cells compared to control, centered at the TSS in a ±1.5 Kb window. Rows indicate all the genes bound by Pol II and are sorted by decreasing Pol II occupancy, and the color labels to the right indicate the level of enrichment. Heatmap of Mago-HA centered at the TSS (−1 kb to +1 kb) is also shown. **d** ChIP-qPCR analysis of Cdk9-HA occupancies at indicated *MAPK* locus. Error bars indicate the standard deviation of three biological replicates. **e** Changes in Cdk9 occupancy at the TSS (250 bp upstream and downstream of TSS) after control and Mago knockdowns, calculated from Cdk9-DamID experiment, after normalizing to the Dam alone control. The changes were separated according to pre-EJC binding and a two-sample *t*-test was performed. **f** Co-immunoprecipitation experiments from S2R+ cells extracts expressing either HA-SBP-Cdk9 or HA-SBP alone, revealed with Pol II Ser5P antibody. Immunoprecipitations were performed from control cells or cells depleted for pre-EJC core components (Mago, Y14, and eIF4AIII). Shown also is the western blot against HA tag to assess the efficiency of the pull-down of HA-SBP-Cdk9 in each knockdown condition

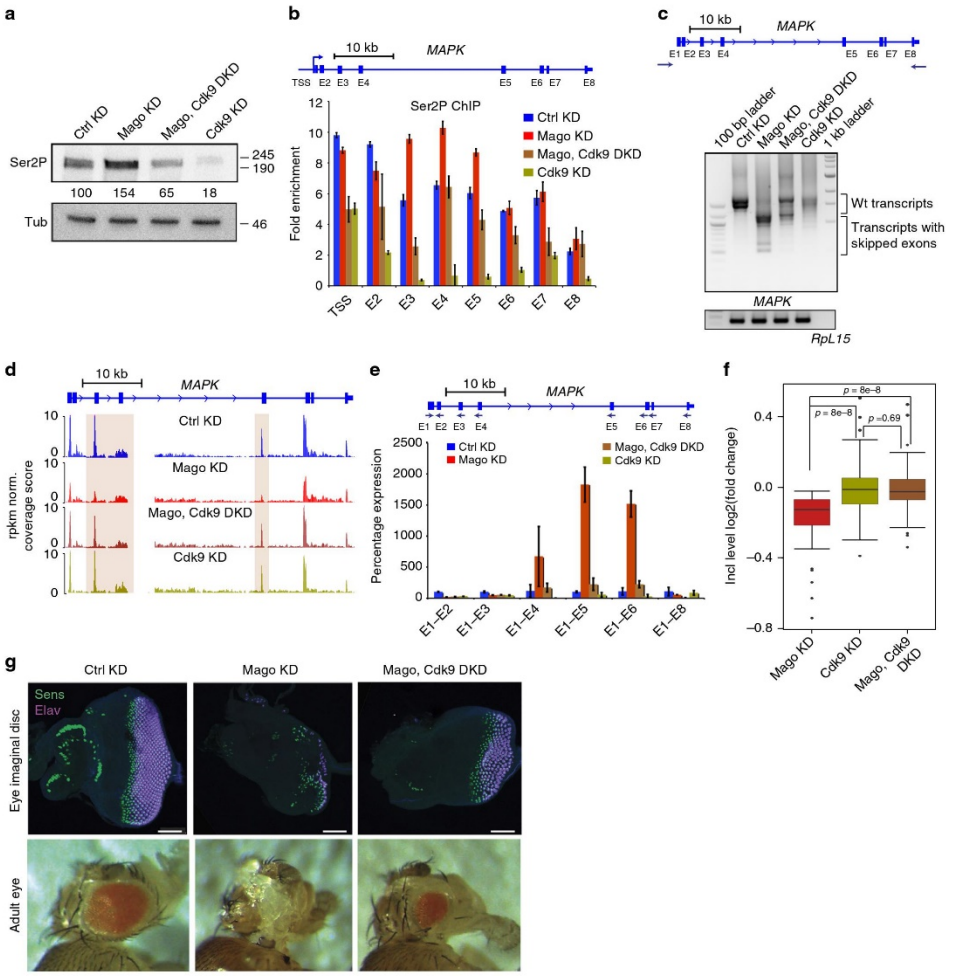
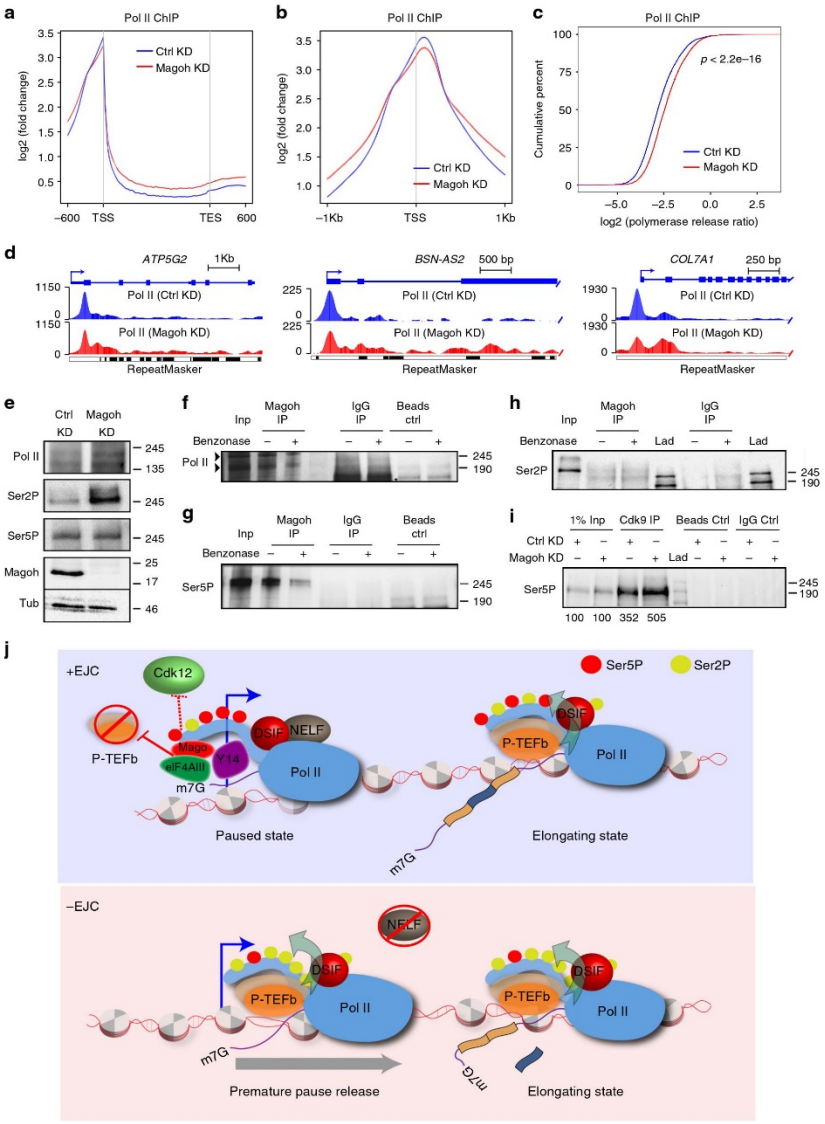



**Fig. 6** Restoring pausing is sufficient to rescue Mago-associated exon skipping defects. **a** Mago knockdown results in elevated level of Ser2P phosphorylation of Pol II. Western blots using antibodies against Pol II Ser2P and Tubulin, using S2R+ cell extracts with indicated knockdowns. Signal in the knockdown conditions was normalized to the control condition using Tubulin as loading control and quantification of the intensity was performed with ImageJ. **b** ChIP-qPCR analysis of Ser2P occupancy level at *MAPK* locus in the indicated knockdowns. The primers used for the analysis are indicated in the scheme above. Bars indicate the 95% confidence interval from the mean of two biological replicates. **c** Agarose gels showing RT-PCR products for *MAPK* and *RpL15* using RNA extracted from S2R+ cells with indicated knockdowns as template for cDNA synthesis. The primers used for the PCR 5′ and 3′ UTR of *MAPK* are shown in the scheme above. RT-PCR products for *RpL15* from respective knockdown condition were used as loading control. **d** Replicate averaged RNA-Seq track examples of *MAPK* in several knockdown conditions. Mago KD results in several exon skipping events, which are rescued upon simultaneous knockdown of Cdk9. The exons skipped in Mago knockdown condition are highlighted by colored rectangles **e** Quantitative RT-PCR using RNA extracts derived from S2R+ cells with the indicated knockdowns. The amplicons were obtained using the same 5′ forward primer (E1) together with the reverse primers on respective exons, as shown in the scheme above. The level of exon skipping is compared to the control treatment, with *RpL49* used for normalization. Bars indicate the 95% confidence interval from the mean of two biological replicates. **f** Box plots showing the log2 fold change in inclusion level of alternatively spliced exons in the indicated knockdowns (rMATS was used for the analysis). **g** Upper panel: Staining of eye imaginal discs from third instar larvae with indicated dsRNAs specifically expressed in the eye (using the *ey*-GAL4 driver). All photoreceptors are stained with anti-Elav antibody (purple), and R8 (the first class of photoreceptor to be specified) with anti-Senseless (green). Scale bar 50 μM. Lower panel: Adult eyes of a control fly and flies with indicated KD




observation in human cells showing that the usage of general transcription inhibitors improve splicing efficiency on two EJC-mediated exon skipping events[47]. Of note, two recent studies suggested a third mechanism for splicing regulation by the EJC that involves the repression of cryptic splice sites (PMID: 30388410, 30388411).

The notion that splicing takes place co-transcriptionally is now a general consensus and two non-exclusive models regarding the impact of transcription on splicing have been proposed. Through the ability of the C-terminal repeat domain (CTD) of its large subunit, RNA Pol II can recruit a wide range of proteins to nascent transcripts[74–77], thereby influencing intron removal. Pol II can influence splicing via a second mechanism referred to as kinetic coupling. According to the model, changes of elongation rates can alter the recognition of exons containing weak splice sites[78–80]. In regards to pre-EJC's activity we favor the first model.

**Fig. 7** The role of Mago on promoter proximal pausing is conserved in human. **a** Metagene profiles based on averaged enrichment over input from two independent biological replicates of total Pol II occupancies in control and Magoh-depleted HeLa cells with standard error of the mean for all the Pol II bound genes. Log2 fold changes against input control are shown on *Y*-axis while *X*-axis depicts scaled genomic coordinates. **b** Metagene profiles of averaged enrichment from two independent biological replicates of total Pol II occupancies in control and Magoh-depleted HeLa cells with standard error of mean for all the Pol II bound genes, centered at the TSS in a ±1 Kb window. Log2 fold changes against input control are shown on *Y*-axis while *X*-axis depicts genomic coordinates. **c** The ECDF plot of PRR in HeLa cells treated with either control or *Magoh* siRNA. *p*-value is derived from two-sample Kolmogorov–Smirnov test. **d** Track examples of total Pol II ChIP-Seq from HeLa cells extracts, after either control or Magoh knockdown. The tracks are the average of two independent biological replicates after input normalization. **e** Western blots performed using antibodies against total Pol II, Ser2P, Ser5P, and Magoh, in HeLa cells treated with either control or *Magoh* siRNA. Similar to *Drosophila*, the loss of Magoh leads to elevated level of Ser2P without affecting Ser5P. **f**–**h** Co-immunoprecipitation of Magoh from HeLa cells extracts, using antibody directed against Magoh. Western blots using Pol II, Ser5P and Ser2P antibody reveal RNA-dependent interaction of Magoh with Pol II and Ser5P. There was no detectable interaction of Magoh with Ser2P. The specific bands for Pol II are highlighted by arrowheads in **f**. The lane labeled with "Lad" indicates ladder. **i** Co-immunoprecipitation of Cdk9 using anti-Cdk9 antibody from HeLa cells extracts treated with either control or *Magoh* siRNA. Western blot was performed with anti-Ser5P antibody. The quantification of the intensity was performed with ImageJ. **j** Model: The pre-EJC stabilizes Pol II pausing by restricting P-TEFb binding at promoter, and possibly by sequestering Cdk12. This activity is required for proper recognition of exons

First, our genome wide studies demonstrate a global impact of the pre-EJC on promoter proximal pausing. Second, we did not observe substantial alteration of the average rate of transcription elongation upon Mago depletion. We did find however some gene-to-gene differences but they poorly correlate with the degree of exon inclusion. Still, this effect might be a secondary consequence of splicing defects, as a previous study suggested the existence of a splicing-dependent elongation checkpoint[81]. Third, reducing the speed of Pol II or depleting the function of transcription elongation factors failed to rescue Mago-splicing defects, arguing that the positive impact of reducing P-TEFb levels on exon definition is dependent on its function in Pol II release rather than in regulating the elongation stage. Thus, in the light of previous model regarding the interplay between pre-mRNA capping and transcription, we propose that by stabilizing Pol II pausing the EJC provides enough time for the recruitment of additional splicing factors that play a critical role in exon definition.

Pol II pausing is found more prominently at developmentally regulated genes, which tend to be long and frequently regulated by alternative splicing. We found a size dependency for Mago-bound genes as well as for Mago-regulated gene expression, suggesting that pre-EJC function is adapted to regulate exon definition of large genes by enhancing their promoter proximal pausing. Interestingly, a recent study shows that genes with long introns tend to be spliced faster and more accurately[82]. Whether this function depends on EJC binding to nascent RNA constitutes an interesting possibility. The next important challenge will be to address the precise mechanisms by which promoter proximal pausing influences pre-mRNA splicing at these developmental genes.

## Methods

**Cloning**. The plasmids used for chromatin immunoprecipitation (ChIP) and co-IP assays in *Drosophila* S2R+ cells were constructed by cloning the corresponding cDNA in the pPAC vector either with N-terminal Flag—3× HA tag or with HA-SBP tag. The CDS were cloned in pPAC vector with N-terminal tags between EcoRV and NotI. The lambdaN and Box-B constructs are derived from the plasmids described earlier[83]. The lambdaN constructs were made by cloning different CDS in frame at the C-terminal, between EcoRV and NotI sites. The boxB constructs were made by cutting out the 3′ boxB sites, and cloning it upstream of luciferase gene with KpnI site. For endogenous genes, the luciferase CDS was first removed using SpeI, followed by blunting the ends. BoxB sites were reintroduced using NotI and StuI sites, and these sites were used to clone endogenous genes.

**RNA isolation and RT-PCR**. RNA was extracted from cells using Trizol reagent, following the manufacturer's protocol. For reverse transcription, cDNA was synthesized using MMLV reverse transcriptase (Promega, Cat No-M1701). For semi-quantitative RT-PCR 2 µg of RNA was reverse transcribed. Five microliter of the cDNA was amplified using the respective primers in 50 µl PCR reaction, using One Taq polymerase (NEB, Cat No-M0480). After 40 cycles of amplification half of the PCR product was loaded on 1% agarose gel to qualitatively analyze the splicing products. For real time PCR, RpL15 was used as an internal control. Relative abundance of transcripts was calculated by the $2^{\Delta}$ Ct method. PCR primers used for semi-quantitative and real time PCR are listed in Supplementary Table 1.

**Cell culture, RNAi, and transfection**. *Drosophila* S2R+ cells were cultured in Schneider Cell's Medium (GIBCO, Cat No-21720) supplemented with 10% FBS and 2% Penicillin/Streptomycin. The plasmids expressing various transgenes were transfected with Effectene transfection reagent (Qiagen, Cat No-301425), following manufacturer's protocol. For knock down experiments, dsRNA was synthesized overnight at 37 °C using Hi-Scribe T7 transcription kit (NEB, Cat No-E2040). dsRNA was transfected in S2R+ cells by serum starvation for 6 h. The treatment was repeated three times and cells were harvested 7 days after the first treatment for Mago. For knockdown of other pre-EJC components and Btz, the treatment was repeated two times and cells were harvested 5 days after the first treatment. The primers used for generating dsRNA are listed on Supplementary Table 1. S2R+ cells were treated with 50 µg/ml of α-amanitin for 7 h to block transcription. Triptolide treatment in S2R+ cells was performed at 10 µM for 6 h.

HeLa cells were cultured in standard RPMI medium supplemented with 10% FBS and 2% Penicillin/Streptomycin. For siRNA knockdown, cells were transfected with 10 nM of siRNA using RNAiMax (Invitrogen) according to manufacturer's protocol. Cells were harvested 48 h after transfection. A mixture of three siRNA was used to deplete Magoh, two against MagohA isoform (siRNA sequence; 1-CGGGAAGTTAAGATATGCCAA; 2-CAGGCTGTTTGTATATTTAAT) and one targeting MagohB isoform (siRNA sequence; GATATGCCAACAACAGCAA).

**Antibodies**. The following antibodies were used in this study. For total Pol II ChIP, RBP1 (Diagenode Cat No-15200004) was used. Ser2P ChIP was performed using ab5095 (Abcam); 3E10 (Chromotek) was used for western blotting. Anti-Ser5P Pol II (Chromotek, Cat No-3E8) and ARNA3 (Pol II) antibodies (Progen, Cat No-65123) were used for western blot assays. For immunoprecipitation experiments, anti-Flag M2 (Sigma, Cat No-F3165) and M-280 streptavidin beads (Thermo-Fisher, Cat No-11205) were used. A polyclonal rabbit anti-Mago antibody was generated from Metabion (Germany). Anti-hCdk9 (Santa Cruz, Cat No-8338) was used for immunoprecipitation from HeLa cells extracts. The Cdk9 western blot from S2R+ cell extract was performed by anti-dCdk9 antibody, a kind gift from Akira Nakamura.

**Immunostaining**. The primary antibodies used were rat anti-Elav (1:5; Developmental Studies Hybridoma Bank) and guinea pig anti-Senseless (1:1000)[84]. Eye imaginal discs were dissected in 0.1 M sodium phosphate buffer (pH 7.2) and then fixed in PEM (0.1 M PIPES at pH 7.2 mM $MgSO_4$, 1 mM EGTA) containing 4% formaldehyde. Washes were done in 0.1 M phosphate buffer with 0.2% Triton X-100. Appropriate fluorescent-conjugated secondary antibodies were used (1:1000; Jackson Immunoresearch Laboratories). Images were collected on Zeiss TCS SP5 confocal microscope.

**Co-IP assay and western blot analysis**. For co-IP assay in S2R+ cells, cells were plated in 10 cm cell culture dish, and respective transgenes were transfected using Effectene transfection reagent, according to manufacturer's protocol. Forty-eight hours post transfection cells were collected, washed once with PBS and re-suspended in swelling buffer (10 mM Tris pH 7.5, 2 mM $MgCl_2$, 5 mM $MgCl_2$, 3 mM $CaCl_2$, and protease inhibitors). After incubating 10 min on ice, the suspension was spun at 600 g for 10 min at 4 °C. After discarding the supernatant the pellet was resuspended in lysis buffer (10 mM Tris pH 7.5, 2 mM $MgCl_2$, 5 mM $MgCl_2$, 3 mM $CaCl_2$, 0.5% NP-40, 10% glycerol and protease inhibitors) and centrifuged for 5 min at 600 × *g*. Nuclei were resuspended in lysis buffer (40 mM

HEPES pH7.4, 140 mM NaCl, 10 mM $MgCl_2$, 0.5% Triton X-100 and protease inhibitors) and sonicated in the bioruptor plus (Diagenode) for 6 cycles with 30 s "ON/OFF" at low settings. Protein concentrations were determined using Bradford reagent (BioRad, Cat No-5000006). For IP 2 mg of proteins were incubated with respective antibody in lysis buffer and rotated head-over-tail O/N at 4 °C. The beads were washed 3× for 10 min with lysis buffer and IP proteins were eluted by incubation in 1× SDS buffer at 85 °C for 10 min. Immunoprecipitated and input proteins were analyzed by western blot, after separating them on 4–15% gradient SDS-PAGE gel (BioRad, Cat No-4561083) and transferred to PVDF membrane (Millipore, Cat No-IPVH00010). After blocking with 5% milk in TBST (0,05% Tween in 10 mM Tris pH 7.4 and 140 mM NaCl) for O/N at 4 °C, the membrane was incubated with respective primary antibody in blocking solution O/N at 4 °C. The antibodies were used at following dilution: Ser2P: 1:500; Ser5P: 1:500; ARNA-3: 1:1000; HA: 1:2500; Mago: 1:2000; and Magoh: 1:1000. Membrane was washed 4× in TBST for 15 min and incubated 1 h at RT with secondary antibody in blocking solution. Blots were developed using SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Scientific, Cat No-34080) and visualized using BioRad Gel documentation system. Full-length blots with molecular weight standards are provided in the Supplementary Figure 11.

For co-IP assay in HeLa, cells were plated in 10 cm cell culture dish. Afterwards, cells were collected washed once with PBS and re-suspended in swelling buffer (10 mM Tris pH 7.5, 2 mM $MgCl_2$, 5 mM $MgCl_2$, 3 mM $CaCl_2$, and protease inhibitors), identical to the approach as in S2R+ cells. After incubating 10 min on ice, the suspension was spun at 600 *g* for 10 min at 4 °C. After discarding the supernatant the pellet was resuspended in lysis buffer (10 mM Tris pH 7.5, 2 mM $MgCl_2$, 5 mM $MgCl_2$, 3 mM $CaCl_2$, 0.5% NP-40, 10% glycerol and protease inhibitors) and centrifuged for 5 min at 600 × *g*. Nuclei were resuspended in lysis buffer (40 mM HEPES pH7.4, 140 mM NaCl, 10 mM $MgCl_2$, 0.5% Triton X-100 and protease inhibitors) and sonicated in the bioruptor plus (Diagenode) for 6 cycles with 30 s "ON/OFF" at low settings. Protein concentrations were determined using Bradford reagent (BioRad, Cat No-5000006). For IP 2 mg of proteins were incubated with anti-Magoh antibody in lysis buffer and rotated head-over-tail O/N at 4 °C. The beads were washed 3× for 10 min with lysis buffer and IP proteins were eluted by incubation in 1× SDS buffer at 85 °C for 10 min, immunoprecipitated and input proteins were analyzed by western blot, as described above.

**RNA extraction and RNA-seq**. RNA was extracted from cells using Trizol reagent, following the manufacturer's protocol. RNA was further cleaned for organic contaminants by RNeasy MinElute Spin columns (Qiagen, Cat No-74204). The purified RNA was subjected to oligodT (NEB, Cat No-S1419S) selection to isolate mRNA. The resulting mRNA was fragmented and converted into libraries using illumina TruSeq Stranded mRNA Library Prep kit (illumina, Cat No- 20020594) following manufacturer's protocol.

**ChIP-qPCR and ChIP-Seq**. S2R+ cells and HeLa cells were fixed with 1% formaldehyde for 10 min at room temperature, and harvested in SDS buffer resuspended in RIPA buffer (140 mM NaCl, 10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% DOC), and lysed by sonication. The lysate was cleared by centrifugation, and incubated with respective antibodies overnight at 4 °C. Antibody complexes bound to protein G beads were washed once with 140 mM RIPA, four times with 500 mM RIPA, once with LiCl buffer and twice with TE buffer for 10 min each at 4 °C. DNA was recovered after reverse crosslinking and phenol chloroform extraction. After precipitating and pelleting, DNA was dissolved in 30 µl of TE. Control immunoprecipitations were done in parallel with either tag alone or knock down controls, and processed identically. Five microliters of immunoprecipitated DNA were used for checking enrichment with various primer pairs (listed in Supplementary Table 1) on Applied Biosystem ViiA™ 7 real time machine using SYBR green reagent (Life technologies, Cat No-4367659). To examine whether these changes in Pol II distribution were widespread, we performed ChIP-Seq experiments in control and Mago KD conditions. To exclude the possibility of changes in Pol II occupancy driven by differences in immunoprecipitation efficiency and technical variance during library preparation in different knock down conditions, we used yeast chromatin as "spike-in" control (Orlando et al.[85]). With this approach, we confirmed the decrease in Ser2P levels and Pol II at the promoter region and an increase within the gene body of MAPK. After validating enrichment, the recovered DNA was converted into libraries using NebNext Ultra DNA library preparation kit, following manufacturer's protocol. DNA libraries were multiplexed, pooled and sequenced on Illumina HiSeq 2000 platform.

**DRB-4sU-Seq**. S2R+ cells were grown in Schneider's Cell Medium with 10% bovine serum supplemented with antibiotics and maintained at 25 °C. 5,6-dichlorobenzimidazole 1-β-d-ribofuranoside (DRB) from Sigma (D1916) was used at a final concentration of 300 µM, dissolved in water, for 5 h. 4-thiouridine (4sU) was purchased from Sigma (Cat No-T4509) and used at a final concentration of 100 µM. Control and Mago KD was performed as described before. All the samples were labeled for 6 min with 4-thiouridine, and transcription was allowed to proceed after DRB removal for 0, 2, 8, and 16 min along with one non-DRB treated control.

A total of 100–130 µg RNA was used for the biotinylation reaction. 4sU-labeled RNA was biotinylated with EZ-Link Biotin-HPDP (Thermo Scientific, Cat No-21341), dissolved in dimethylformamide (DMF, Sigma Cat No-D4551) at a concentration of 1 mg/mL. Biotinylation was done in labeling buffer (10 mM Tris pH 7.4, 1 mM EDTA) and 0.2 mg/ml Biotin-HPDP for 2 h with rotation at room temperature. Two rounds of chloroform extractions removed unbound Biotin-HPDP. RNA was precipitated at 20,000×*g* for 20 min at 4 °C with a 1:10 volume of 5 M NaCl and an equal volume of isopropanol. The pellet was washed with 75% ethanol and precipitated again at 20,000×*g* for 10 min at 4 °C. The pellet was left to dry, followed by resuspension in 100 µl RNase-free water. Biotinylated RNA was captured using Dynabeads MyOne Streptavidin T1 beads (Invitrogen, Cat No-65601). Biotinylated RNA was incubated with 50 µl Dynabeads with rotation for 15 min at 25 °C. Beads were magnetically fixed and washed with 3× Dynabeads washing buffer. RNA-4sU was eluted with 100 µl of freshly prepared 100 mM dithiothreitol (DTT), and cleaned on RNeasy MinElute Spin columns (Qiagen, Cat No-74204). For the untreated 4sU-Seq version used for calculating polymerase release ratio (PRR), an identical approach was used with following modifications. During the period when biotinylated RNA was incubated with 50 µl Dynabeads with rotation for 15 min at 25 °C, RNAse T1 was added in order to fragment RNA to 100 bp. Beads were magnetically fixed and washed with 3× Dynabeads washing buffer, as described before. RNA-4sU was eluted with 100 µl of freshly prepared 100 mM DTT, and cleaned on RNeasy MinElute Spin columns (Qiagen, Cat No-74204). Enriched nascent RNAs were converted to cDNA libraries with *Drosophila* Ovation Kit (Nugen- Cat No-7102–32) with integrated ribosomal depletion workflow. Amplified cDNA libraries were pooled, multiplexed, and sequenced on two lanes of Illumina HiSeq 2000.

**MNase-Seq**. S2R+ cells were fixed with 1% formaldehyde for 10 min at RT. Cells were harvested and 20 million nuclei were spun at 3500 g at 4 °C for 10 min. Nuclear pellets were resuspended in 300 µl of MNase digestion buffer (0.5 mM spermidine, 0.075% NP40, 50 mM NaCl, 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM β-mercaptoethanol and complete protease inhibitors). Reaction was spun at 3200 g, 4 °C for 10 min and resuspended in 50 µl of MNase digestion buffer and digested with 30U of MNase at 37 °C for 10 min at 300 rpm in mixing block. The MNase digestion reaction was quenched with EDTA at 10 mM final concentration. After 10 min on ice, the nuclei were washed once with 1 ml of RIPA buffer (140 mM NaCl and complete protease inhibitors). Pellets were resuspended in 300 µl of RIPA buffer (140 mM) and sonicated (3 cycles, medium intensity, 30 s on/off intervals) and centrifuged at 18,000 × *g*, 4 °C for 10 min. DNA was recovered after reverse crosslinking and phenol chloroform extraction. After precipitating and pelleting, DNA was dissolved in 30 µl of TE and resolved on agarose gel. The ~147 bp fragments corresponding to the mono nucleosomal fragments were gel extracted and after library prepararion, were subjected to 50 bp paired end sequencing on Illumina HiSeq 2500 platform.

**DamID-Seq**. pUAST- LT3- ORF1 vector (kind gift from A. Brand) was used to clone Cdk9 as a C-terminal Dam-fusion protein. The Dam-Cdk9 itself was cloned downstream of mcherry (as a primary ORF) separated by stop codon. This ensured low level expression of the Dam-Cdk9 fusion protein. S2R+ cells were plated in 10 cm dish and subjected to control and Mago knockdowns using dsRNA, as described earlier. On the sixth day of knockdown pUAST-LT3-Dam-Cdk9 was co-transfected with pActin-Gal4 vector to induce Dam-Cdk9 expression, using effectene transfection reagent according to manufacturer's protocol. The Dam alone control was similarly transfected in control and Mago depleted S2R+ cells. DNA was isolated from cells after 16 h of transfection and subsequent treatments were performed as described[86]. Purified and processed genomic DNA of two biological duplicates was subjected to library preparation using the NebNext DNA Ultra II library kit (New England Biolabs) and sequenced on a NextSeq500.

**Computational analysis**. RNA-Seq: The libraries were sequenced with a read length of 71 bp in paired end mode. Mapping was performed using STAR[87] (v. 2.5.1b) against ENSEMBL release 84. Counts per gene were derived using htseq count (v.0.6.1p1). Differential expression analysis was done using DESeq2[88] (v.1.10.1), differential expressed genes were filtered for an FDR of 1% and a fold change of 1.3. Splicing analysis was done using DEXSeq[89] (v. 1.16.10), and rMATS[90] (v. 3.2.1b) with 10% FDR filtering. Genes were defined as expressed if they had coverage above 1 rpkm in the averaged control samples.

ChIP-Seq: The libraries were sequenced on a HiSeq2500/NextSeq500 in either paired end or single end mode. De-multiplexing and fastq file conversion was performed using blc2fastq (v.1.8.4/v2.19.1) for all libraries save the Ser2P ChIPs. Ser2P ChIP libraries were de-multiplexed using 6 bp front tags. After sorting, the tags and the A-overhang base were trimmed (7 bp in total). Reads from ChIP-Seq libraries were mapped using bowtie2[91] (v. 2.2.8), and filtered for uniquely mapped reads. The genome build and annotation used for all *Drosophila* samples was BDGP6 (ensemble release 84). The genome build and annotation used for the HeLa samples was hg38 (ENSEMBL release 84). Peak calling was performed using MACS2[92] (v 2.1.1–20160309). Further processing was done using R and Bioconductor packages. Input normalized bigwig tracks were produced using Deeptools[93]. The spike-in normalization was done according to the Orlando et al.[85]. All the libraries (including input) where calibrated using a normalization

factor; defined as the number of reads mapped to the yeast genome (used as a reference/non-test genome) per million of reads mapped to the *Drosophila* genome. Once the libraries were calibrated according to the respective normalization factor, the enrichment was computed for every condition against their respective input.

To assign the target genes bound by pre-EJC components, peaks were called using MACS2 with 2.0-fold enrichment as cut-off. The resulting peaks were annotated with the ChIPseeker package on Bioconductor, using nearest gene to the peak summit as assignment criteria. The intersection of genes bound by all pre-EJC components, i.e., Mago-HA, Y14-HA, and eIF4A3-HA, was defined as pre-EJC bound.

4sU-Seq: The libraries were sequenced with a read length of 50 bp in single end mode. Mapping was performed using STAR (v. 2.5.1b) and ENSEMBL release 84 for *Drosophila*. Multimapped reads were filtered out, and uniquely mapped reads to the transcript were considered.

Calculation of polymerase release ratio (PRR): PRRs were calculated as follows: for each gene, the TSS region was defined as 250 bp upstream to 250 bp downstream of the TSS. The gene body was defined as 500 bp downstream of the TSS to 500 bp upstream of the TES. The PRR ratio was calculated as the log2 ratio between the enrichment in the downstream region towards the enrichment at the TSS. For each gene, the TSS with the highest average signal around the TSS in the Control condition was selected. Enrichment calculations were based on the enrichment over the input (ChIP-Seq) or t0 (4sU-Seq). The TSS reference was taken from ENSEMBL release 84. Genes with a length smaller than the required length for the calculation were excluded. All libraries containing spike-in controls were normalized to spike-ins following the method described before.

Calculation of elongation rate: For elongation rate calculation, all the genes longer than 10 kb were divided into 100 bp bins (to a total of 20 kb) and the transcriptional wave front was identified in the bin with lowest local minimum signal. The distance in base pairs covered by the wave front between 2 min after DRB removal and 8 min is then divided by the corresponding time interval of 6 min to calculate elongation rates (bp/min).

MNase-Seq: The libraries were sequenced with a read length of 50 bp in paired end mode. De-multiplexing and fastq file conversion was performed using blc2fastq (v.1.8.4). Libraries were de-multiplexed using 6 bp front tags. After sorting, the tags and the A-overhang base were trimmed (7 bp in total). Reads from MNase-Seq libraries were mapped using bowtie2 (v. 2.2.8), and filtered for uniquely mapped reads. The genome build and annotation used for all *Drosophila* samples was BDGP6 (ensemble release 84). Further processing was done using R and Bioconductor packages. Heatmaps and input normalized tracks were produced using Deeptools (v. 2.2.3). Metagene profiles were produced using NGS.plot (v. 2.61).

Targeted DamID-Seq: The libraries were sequenced with a read length of 50 bp in paired end mode. The first read was mapped to *Drosophila melanogaster* genome (BDGP6) using bowtie (v.2.2.9), binned to GATC fragments, and normalized against the Dam-only control[94] using the available damidseq_pipeline on GitHub. The resulting bedgraph files were averaged and smoothened using BEDOPS[95] (v. 2.4.30). The smoothened bedgraph files were converted to bigwig file using SeqPlot, and processed through Deeptools (v. 2.2.3) to generate heatmaps. To quantify the changes at the promoter, the signals in the bedgraph were mapped to the promoters using bedmap tool available in BEDOPS software. The further quantification and plots were generated using R (v 3.4.2), and ggplot2 package available on Bioconductor.

**NELF microarray analysis**. To compare our data with NELF knockdown we used data from this study[60]. The NELF data was retrieved from GEO (GSE20471) and processed in R. Preprocessing was done as described in the original paper. The TSS used for pausing index calculations were determined as the ones with the highest average enrichment + −250 bp around the TSS in the control condition. Calculation of pausing index was done as described in the study.

NELF bound genes were determined as the genes which have an average enrichment >1.3 in the + −250 bp around the TSS (based on the TSS used for pausing index calculations). Any gene ids not matching our reference were converted using the Flybase ID converter tool[96].

**Quantification and statistical analysis**. Statistical parameters and significance are reported in the Figures and the Figure legends. For comparisons of the distribution of different classes we used ANOVA. *t*-Test, two-sample Kolmogorov-Smirnov test and Fisher's test were used for testing the statistical significance. Number of genes used in the box plots is indicated in Supplementary Table 2.

**Reporting Summary**. Further information on experimental design is available in the Nature Research Reporting Summary linked to this Article.

## Data availability

Datasets from RNA-Seq, ChIP-Seq, 4sU-Seq, MNase-Seq, and DamID-Seq have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE92389. A Reporting Summary for this Article is available as a Supplementary Information file. All other data supporting the findings of this study are available from the corresponding author upon request.

Received: 11 May 2018 Accepted: 4 January 2019
Published online: 31 January 2019

## References

1. Brugiolo, M., Herzel, L. & Neugebauer, K. M. Counting on co-transcriptional splicing. *F1000Prime Rep.* **5**, 9 (2013).
2. Girard, C. et al. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat. Commun.* **3**, 994 (2012).
3. Khodor, Y. L. et al. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev.* **25**, 2502–2512 (2011).
4. Tilgner, H. et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).
5. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163–175 (2014).
6. Custodio, N. & Carmo-Fonseca, M. Co-transcriptional splicing and the CTD code. *Crit. Rev. Biochem. Mol. Biol.* **51**, 395–411 (2016).
7. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* **18**, 637–650 (2017).
8. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–177 (2015).
9. Saldi, T., Cortazar, M. A., Sheridan, R. M. & Bentley, D. L. Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *J. Mol. Biol.* **428**, 2623–2635 (2016).
10. Gilmour, D. S. & Lis, J. T. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Mol. Cell. Biol.* **6**, 3984–3989 (1986).
11. Rougvie, A. E. & Lis, J. T. The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**, 795–804 (1988).
12. Levine, M. Paused RNA polymerase II as a developmental checkpoint. *Cell* **145**, 502–511 (2011).
13. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
14. Gaertner, B. et al. Poised RNA polymerase II changes over developmental time and prepares genes for future expression. *Cell Rep.* **2**, 1670–1683 (2012).
15. Kwak, H. & Lis, J. T. Control of transcriptional elongation. *Annu. Rev. Genet.* **47**, 483–508 (2013).
16. Smith, E. & Shilatifard, A. Transcriptional elongation checkpoint control in development and disease. *Genes Dev.* **27**, 1079–1088 (2013).
17. Liu, X., Kraus, W. L. & Bai, X. Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends Biochem. Sci.* **40**, 516–525 (2015).
18. Gressel, S. et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *elife* **6**, e29736 (2017).
19. Luo, Z., Lin, C. & Shilatifard, A. The super elongation complex (SEC) family in transcriptional control. *Nat. Rev. Mol. Cell Biol.* **13**, 543–547 (2012).
20. Marshall, N. F. & Price, D. H. Purification of P-Tefb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* **270**, 12335–12338 (1995).
21. Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
22. Marshall, N. F., Peng, J., Xie, Z. & Price, D. H. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* **271**, 27176–27183 (1996).
23. Fujinaga, K. et al. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol. Cell. Biol.* **24**, 787–795 (2004).
24. Ni, Z. et al. P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo. *Mol. Cell. Biol.* **28**, 1161–1170 (2008).
25. Chen, F. X. et al. PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. *Cell* **162**, 1003–1015 (2015).
26. Yu, M. et al. RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350**, 1383–1386 (2015).
27. 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).
28. 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).
29. 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. USA* **109**, 21313–21318 (2012).
30. 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).
31. 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).
32. Steckelberg, A. L., Altmueller, J., Dieterich, C. & Gehring, N. H. CWC22-dependent pre-mRNA splicing and eIF4A3 binding enables global deposition of exon junction complexes. *Nucleic Acids Res.* **43**, 4687–4700 (2015).
33. Shibuya, T., Tange, T. O., 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).
34. Kataoka, N., Diem, M. D., Kim, V. N., Yong, J. & Dreyfuss, G. Magoh, a human homolog of *Drosophila* mago nashi protein, is a component of the splicing-dependent exon–exon junction complex. *EMBO J.* **20**, 6424–6433 (2001).
35. Hachet, O. & Ephrussi, A. *Drosophila* Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. *Curr. Biol.* **11**, 1666–1674 (2001).
36. Kim, V. N. et al. The Y14 protein communicates to the cytoplasm the position of exon-exon junctions. *EMBO J.* **20**, 2062–2068 (2001).
37. 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).
38. Choudhury, S. R. et al. Exon junction complex proteins bind nascent transcripts independently of pre-mRNA splicing in *Drosophila melanogaster*. *eLife* **5**, e19881 (2016).
39. Boehm, V. & Gehring, N. H. Exon junction complexes: supervising the gene expression assembly line. *Trends Genet.* **32**, 724–735 (2016).
40. Gerbracht, J. V. & Gehring, N. H. The exon junction complex: structural insights into a faithful companion of mammalian mRNPs. *Biochem. Soc. Trans.* **46**, 153–161 (2018).
41. Le Hir, H., Sauliere, J. & Wang, Z. The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.* **17**, 41–54 (2016).
42. Tange, T. O., Shibuya, T., Jurica, M. S. & Moore, M. J. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* **11**, 1869–1883 (2005).
43. 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).
44. 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).
45. 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).
46. 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).
47. Wang, Z., Murigneux, V. & Le Hir, H. Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biol.* **15**, 551 (2014).
48. Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. & Blencowe, B. J. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**, 1252–1269 (2013).
49. Moehle, E. A., Braberg, H., Krogan, N. J. & Guthrie, C. Adventures in time and space: splicing efficiency and RNA polymerase II elongation rate. *RNA Biol.* **11**, 313–319 (2014).
50. Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* **84**, 165–198 (2015).
51. Schwalb, B. et al. TT-seq maps the human transient transcriptome. *Science* **352**, 1225–1228 (2016).
52. Singh, J. & Padgett, R. A. Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* **16**, 1128–1133 (2009).
53. Fuchs, G. et al. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol.* **15**, R69 (2014).
54. Thummel, C. S., Burtis, K. C. & Hogness, D. S. Spatial and temporal patterns of E74 transcription during *Drosophila* development. *Cell* **61**, 101–111 (1990).
55. O'Brien, T. & Lis, J. T. Rapid changes in *Drosophila* transcription after an instantaneous heat shock. *Mol. Cell. Biol.* **13**, 3456–3463 (1993).
56. Yao, J., Ardehali, M. B., Fecko, C. J., Webb, W. W. & Lis, J. T. Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation. *Mol. Cell* **28**, 978–990 (2007).
57. Ardehali, M. B. & Lis, J. T. Tracking rates of transcription and splicing in vivo. *Nat. Struct. Mol. Biol.* **16**, 1123–1124 (2009).
58. Lis, J. Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 347–356 (1998).
59. Baron-Benhamou, J., Gehring, N. H., Kulozik, A. E. & Hentze, M. W. Using the lambdaN peptide to tether proteins to RNAs. *Methods Mol. Biol.* **257**, 135–154 (2004).
60. Gilchrist, D. A. et al. Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**, 540–551 (2010).
61. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424–428 (2000).
62. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat. Protoc.* **2**, 1467–1478 (2007).
63. Battaglia, S. et al. RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases. *Elife* **6**, ee25637 (2017).
64. Blythe, A. J. et al. The yeast transcription elongation factor Spt4/5 is a sequence-specific RNA binding protein. *Protein Sci.* **25**, 1710–1721 (2016).
65. Cheng, B. & Price, D. H. Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic Acids Res.* **36**, e135 (2008).
66. Crickard, J. B., Fu, J. & Reese, J. C. Biochemical analysis of yeast suppressor of Ty 4/5 (Spt4/5) reveals the importance of nucleic acid interactions in the prevention of RNA polymerase II arrest. *J. Biol. Chem.* **291**, 9853–9870 (2016).
67. Lee, C. et al. NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. *Mol. Cell. Biol.* **28**, 3290–3300 (2008).
68. Narita, T. et al. Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Mol. Cell. Biol.* **23**, 1863–1873 (2003).
69. Choe, J. et al. eIF4AIII enhances translation of nuclear cap-binding complex-bound mRNAs by promoting disruption of secondary structures in 5′ UTR. *Proc. Natl Acad. Sci. USA* **111**, E4577–E4586 (2014).
70. Chuang, T. W., Chang, W. L., Lee, K. M. & Tarn, W. Y. The RNA-binding protein Y14 inhibits mRNA decapping and modulates processing body formation. *Mol. Biol. Cell* **24**, 1–13 (2013).
71. Chuang, T. W., Lee, K. M., Lou, Y. C., Lu, C. C. & Tarn, W. Y. A point mutation in the exon junction complex factor Y14 disrupts its function in mRNA cap binding and translation enhancement. *J. Biol. Chem.* **291**, 8565–8574 (2016).
72. Ji, X. et al. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* **153**, 855–868 (2013).
73. Singh, G. et al. The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* **151**, 750–764 (2012).
74. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
75. de Almeida, S. F. & Carmo-Fonseca, M. The CTD role in cotranscriptional RNA processing and surveillance. *FEBS Lett.* **582**, 1971–1976 (2008).
76. Misteli, T. & Spector, D. L. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol. Cell* **3**, 697–705 (1999).
77. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936 (2006).
78. Kornblihtt, A. R. Chromatin, transcript elongation and alternative splicing. *Nat. Struct. Mol. Biol.* **13**, 5–7 (2006).
79. Kornblihtt, A. R. Coupling transcription and alternative splicing. *Adv. Exp. Med. Biol.* **623**, 175–189 (2007).
80. Fong, N. et al. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* **28**, 2663–2676 (2014).
81. Chathoth, K. T., Barrass, J. D., Webb, S. & Beggs, J. D. A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. *Mol. Cell* **53**, 779–790 (2014).
82. Pai, A. A. et al. The kinetics of pre-mRNA splicing in the *Drosophila* genome and the influence of gene architecture. *eLife* **6**, e32537 (2017).
83. Hilgers, V., Lemke, S. B. & Levine, M. ELAV mediates 3' UTR extension in the *Drosophila* nervous system. *Genes Dev.* **26**, 2259–2264 (2012).
84. Frankfort, B. J., Nolo, R., Zhang, Z., Bellen, H. & Mardon, G. Senseless repression of rough is required for R8 photoreceptor differentiation in the developing *Drosophila* eye. *Neuron* **32**, 403–414 (2001).
85. Orlando, D. A. et al. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170 (2014).
86. Marshall, O. J. et al. Cell-type-specific profiling of protein-DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nat. Protoc.* **11**, 1586–1598 (2016).
87. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
88. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
89. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017 (2012).

90. Shen, S. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl Acad. Sci. USA* **111**, E5593–E5601 (2014).
91. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with bowtie 2. *Nature Methods* **9**, 357–359 (2012).
92. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
93. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
94. Marshall, O. J. & Brand, A. H. damidseq_pipeline: an automated pipeline for processing DamID sequencing datasets. *Bioinformatics* **31**, 3371–3373 (2015).
95. Neph, S. et al. damidseq_pipeline: an automated pipeline for processing DamID sequencing datasets. *Bioinformatics* **28**, 1919–1920 (2012).
96. Gramates, L. S. et al. FlyBase at 25: looking to the future. *Nucleic Acids Res.* **45**, D663–D671 (2017).

## Acknowledgements

We thank the Bloomington *Drosophila* Stock Center, the Transgenic RNAi Project in Harvard and the Vienna *Drosophila* Resource Center for fly reagents. We also thank Akira Nakamura for the Cdk9 antibody. Support by the IMB Genomics Core Facility and the use of its NextSeq500 (INST 247/870–1 FUGG) is gratefully acknowledged. We also thank the IMB Bioinformatics Core Facility for tremendous support; members of Ulrich lab, especially Lilliana Batista for help with yeast chromatin for "spike-in" control; members of the Roignant lab for fruitful discussion; and Enrico Cannavo, Yad Ghavi-Helm, Guillaume Junion, Jessica Treisman for critical reading of the manuscript. This work was supported by the Marie Curie CIG 334288.

## Author contributions

J.A. and J.Y.R. designed the experiments. J.A performed the experiments and analyzed the data, D.B. helped with RIP experiments. N.K., F.M., J.A., and H.B performed the bioinformatics and statistical analyses. G.K.M. carried out the in vivo and cell culture rescue experiments. J.A. and J.Y.R. wrote the paper.

## Additional information

**Supplementary Information** accompanies this paper at https://doi.org/10.1038/s41467-019-08381-0.

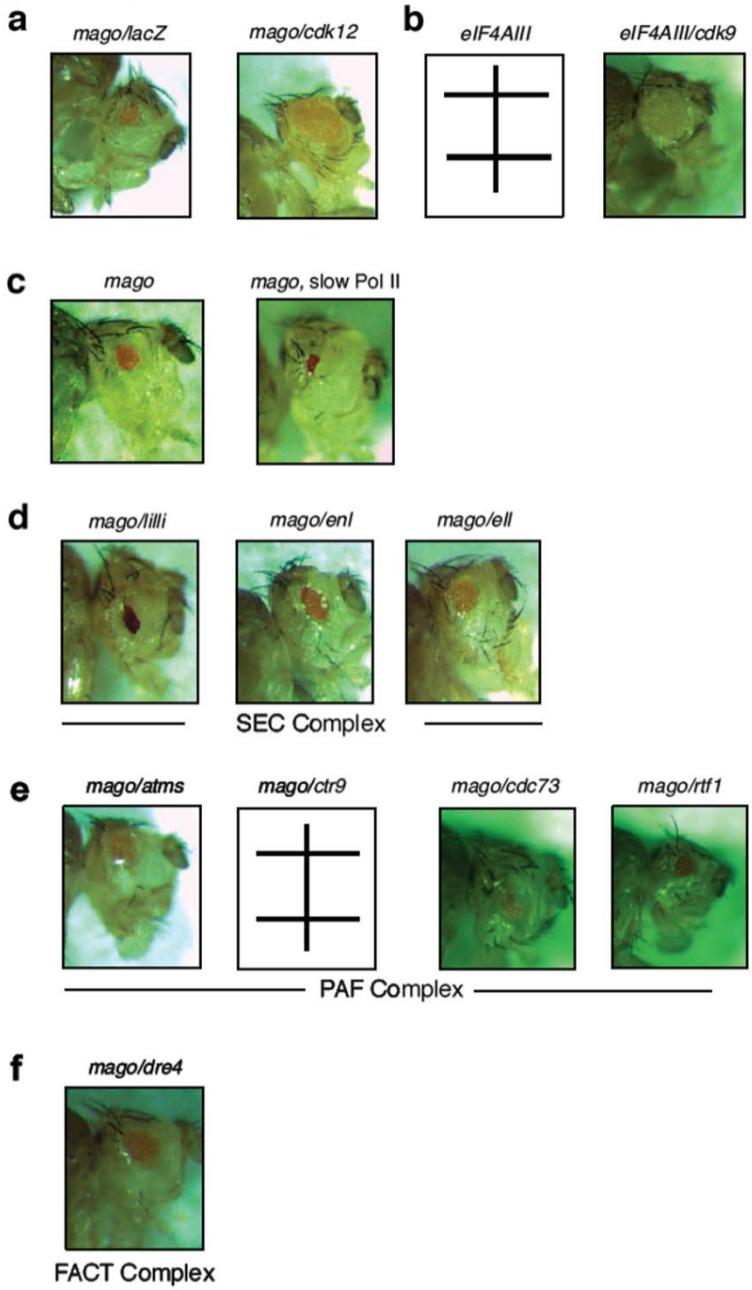
**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**Journal peer review information**: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Supplementary Figure 10**




**Depletion of elongation factors or slowing down RNA Pol II does not rescue Mago photoreceptor differentiation defects (a-f)**
*Drosophila* adult eyes in different conditions. **(a)** Loss of Mago in the eye results in impairment of eye development due to lack of photoreceptor differentiation. Simultaneous depletion of Cdk12 restores photoreceptor differentiation. **(b)** Loss of eIF4AIII in the eye results in lethality. Simultaneous depletion of Cdk9 restores viability and photoreceptor differentiation. **(c)** Slowing down the kinetics of Pol II using a slow *Pol II* mutant fails to rescue Mago's effect on photoreceptor differentiation. **(d-f)** Knockdowns of SEC complex components **(d)**, PAF complex components **(e)** or the Dre4 FACT complex subunit **(f)**, do not substantially rescue the eye phenotype resulting from Mago depletion.

# 10. List of abbreviations

| | |
|---|---|
| ALS | Amyotrophic lateral sclerosis |
| AS | Alazami Syndrome |
| BDSC | Bloomington Drosophila Stock Center |
| BR | Basic rich region |
| CNS | Central nervous system |
| CTD | Carboxy-terminal domain |
| DPE | Downstream promoter element |
| DRB | D-ribofuranosyl benzimidazole |
| DSE | Distal Sequence Element |
| DSIF | DRB sensitivity-inducing factor |
| EJC | Exon-junction complex |
| ENL | Eleven-nineteen leukemia |
| GTF | General Transcription factors |
| IITM | Indian Institute of Technology Madras |
| IMB | Institute of Molecular Biology |
| LARP | La and La-related proteins |
| LEC | Little Elongation Complex |
| NELF | Negative elongation factor |
| PB | Pause Button |
| PCR | Polymerase chain reaction |
| PD | Primordial dwarfism |
| PSE | Proximal Sequence Element |
| SEC | Super elongation complex |
| TBP | TATA-binding protein |
| TEVC | Two-Electrode Voltage Clamp |
| TF | Transcription factors |
| TSS | Transcription start site |
| VNC | Ventral nerve cord |
| ATP | Adenosine Triphosphate |

| | |
|---|---|
| CRISPR | clustered regularly interspaced short palindromic repeats |
| DAPI | 4′,6-diamidino-2-phenylindole |
| DDX | DEAD box |
| DRE | Damage Responsive Elements |
| DSHB | Developmental Studies Hybridoma Bank |
| FACS | Fluorescence-activated cell sorting |
| GFP | Green Fluorescent Protein |
| eGFP | enhanced Green Fluorescent Protein |
| NMJ | Neuromuscular Junction |
| PBS | Phosphate Buffer Saline |
| PBT | Phosphate Buffer Saline TritonX |
| PFA | Paraformaldehyde |
| PVDF | Polyvinylidene difluorideRD |
| RNA | Ribonucleic acid |
| UAS | Upstream Activation sequence |
| UTR | Untranslated region |

## 11. Table 2: List of flies used in this study

| Name on tube | Genotype Description | Chromosome modified | Health of Fly | External Link | Creator | Remark | Remark2 |
|---|---|---|---|---|---|---|---|
| **shGFP 41553** | short hair pin against GFP | III | mild lethality in crosses | #41553 | Bloomington | | |
| **shGFP 41552** | short hair pin against GFP | II | | #41552 | Bloomington | | |
| **dsGFP 9331** | double stranded RNA against GFP | II | | #9331 | Bloomington | | |
| **KG00819-7SK PE** | P element | III | | | Bloomington | | |
| **aCC-RP2 Gal4 (RRK)/Tm3** | RP2 and some subtypes moto neurons | III | | #42739 | Bloomington | | |
| **CQ-Gal4** | CQ neurons | II | | #7468 | Bloomington | | |
| **CQ-Gal4::CD8-GFP** | CQ neurons | II | | #7465 | Bloomington | | |
| **RN2-Gal4** | RP2 and some subtypes f moto neurons | | | #7470 | Bloomington | | |
| **mCherry shRNA** | | | | #BL35785 | Bloomington | | |
| **Per-Gal4** | period | | | #7127 | Bloomington | | |
| **dsCDK9 CG5179** | RNAi against CDK9 | | | #34982 | Bloomington | Strong | |
| **dsGFP 9330** | double stranded RNA against GFP | III | | #9330 | Bloomington | | |
| **Dam PolII** | Original | III | | | Brand | | |
| **Dam Control** | Original | III | | | Brand | | |
| **PolII Dam::7SK** | DAM fused polII expression in 7SK KO | III | Weak!! | | Brand and Mohana | Recombination Validated by PCR | |
| **Dam::7SK KO** | Dam control | III | | | Brand and Mohana | Validated | |
| **Dam::7SK KO/Tm6c** | Dam control | III | | | Brand and Mohana | Validated | |
| **Larp7 KO;;Dam** | Dam control | X, III | | | Brand and Mohana | Validated | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Dam-CDK9/Cyo;7SK KO** | DAM fused CDK9 expression in 7SK KO | II,III | | | Mohana | Validated | |
| **Dam CDK9/Cyo** | UAS Dam CDK9 - Plasmid from Brand lab, CDK9 cloned | II | | | Mohana | Injected by best gene | |
| **eGFP Larp7;UAS - tdTomato/Cyo** | | X,III | | | Mohana | Recombined - validated | |
| **Larp7 KO;;UAS Stinger Red** | | X,III | | | Mohana | Recombined - used for single cell RNA seq | |
| **UAS Stinger Red;7SK KO/Tm6b** | | II,III | | | Mohana | Recombined - used for single cell RNA seq | |
| **UAS-Dicer::eGFP-Larp7; Tub Gal4/Cyo** | Dicer recombined to enhance the KD | X, II | | | Mohana | for RNAi of Larp7 | |
| **UAS-Dicer::eGFP-Larp7; sh-eGFP 41552** | Dicer recombined to enhance the KD | X, II | | | Mohana | for RNAi of Larp7 | |
| **UAS-Dicer::eGFP-Larp7; sh-eGFP 41552/Cyo** | Dicer recombined to enhance the KD | X, II | | | Mohana | for RNAi of Larp7 | |
| **UAS-Dicer::eGFP-Larp7; engrail/Cyo** | Dicer recombined to enhance the KD | X, II | | | Mohana | for RNAi of Larp7 | |
| **UAS-Dicer::eGFP-Larp7; ;24B10/Tm6c** | Dicer recombined to enhance the KD | X, III | | | Mohana | for RNAi of Larp7 | |
| **UAS-Dicer::eGFP-Larp7; Tdc/Cyo** | Dicer recombined to enhance the KD | X, II | | | Mohana | for RNAi of Larp7 | |
| **eGFP Larp7/Fm6;UAS - Stinger DsRed/Cyo** | | X,III | | | Mohana | Recombined - validated | |
| **UAS CD8-RFP** | | II | | | Olaf | | |
| **UAS CD8-RFP** | | III | | | Olaf | | |
| **Insc-Gal4 (Mz1407)** | Neuroblast expression | | | | Olaf | | |
| **D42-Gal4** | All motoneurons | II | | | Olaf | | |
| **OK6-Gal4** | motoneuron expression | | | #64199 | Olaf | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Larp7 KO;;Dam PolII** | DAM fused polII expression in Larp7 KO | X, III | | | Brand and Mohana | Validated | |
| **Larp7 KO;;Dam PolII** | DAM fused polII expression in Larp7 KO | X, III | | | Brand and Mohana | Validated | |
| **Larp7 KO;Dam Cdk9** | DAM fused CDK9 expression in Larp7 KO | X, II | | | Brand and Mohana | Validated | |
| **D42-Gal4,Cha-Gal80** | motoneuron expression | | | | Carsten lab | | |
| **UAS-nCD8::GFP;D42-Gal4,Cha-Gal80** | motoneuron expression | | | | Carsten lab | | |
| **OK371 ; CD8 GFP** | motoneuron expression | II | | | | | |
| **dMP2-Gal4/Tm3,Sb** | MP2 neurons | III | | | | | |
| **dMP2-Gal4:UAS-myc.eGFP/Tm3,Sb** | MP2 neurons | III | | | | | |
| **20XUAS IVS-GCamp6** | expression of Gcamp6 | II | | #42746 | From Carsten Duch | | |
| **Sp/Cyo;Prnt/Tm6b** | Balancer line | II;III | | | Lab | used for II and III chromosome crosses | |
| **engrail-Gal4** | | | | | Lab | | |
| **Ddc-Gal4/Tm6c** | | | | | Lab | | |
| **Repo Gal4** | Glia | | | | Lab | | |
| **7SK KO #1** | CRISPR KO-7SK | III | Weak!! | | Mohana | Homozygous | |
| **7SK KO #2** | CRISPR KO-7SK | III | Weak!! | | Mohana | Homozygous | |
| **7SK gRNA** | CRISPR-gRNA | II | NA | | Mohana | Homozygous | U6.2B |
| **Larp7 KO;;7Sk KO** | CRISPR KO - larp7 and 7SK | X and III | Weak!! | | Mohana | Homozygous | Recombined |
| **Larp7 KO #2** | CRISPR KO-Larp7 | X | Weak!! | | Mohana | Homozygous | |
| **Larp7 KO #1** | CRISPR KO-Larp8 | X | Weak!! | | Mohana | Homozygous | |
| **Larp7 gRNA** | CRISPR-gRNA-Larp7 | II | NA | | Mohana | Homozygous | U6.2B |
| **7SK- 17313/Cyo; +/Tm6C** | pUAST 7SK (Promoter+FL) | II | NA | | Mohana | Homozygous | Rescue line injected by Best Gene |
| **TUbGal4(UAS eGFP-Larp7) /Cyo** | pUASTeGFP Larp7 CDS | II | NA | | Mohana | Heterozygous | Rescue line injected by Best Gene |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Control - No deletion CG9650/Fm6,W** | Experiment control No KO | NA | NA | | Mohana | Homozygous | |
| **Larp7 KO/Fm6,w; UAS-eGFP Larp7** | pUASTeGFP Larp7 CDS | II | NA | | Mohana | Heterozygous | Rescue line injected by Best Gene |
| **Larp7 KO/Fm6,w; TubGal4/Cyo** | | II | | | Mohana | Heterozygous | |
| **7SKO 17313/Cyo; 7SK KO/Tm6C** | pUAST 7SK (Promoter+FL) | II | NA | | Mohana | Homozygous | Rescue line injected by Best Gene |
| **7SKO 17313; 7SK KO** | pUAST 7SK (Promoter+FL) | II | NA | | Mohana | Homozygous | Rescue line injected by Best Gene |
| **UAS-eGFP Larp7/Cyo** | pUASTeGFP Larp7 CDS | II | NA | | Mohana | Heterozygous | Rescue line injected by Best Gene |
| **TubGal4/Cyo** | | II | | | Mohana | Heterozygous | |
| **Larp7 gRNA (endo)** | CRISPR gRNA | II | NA | | Mohana | Homozygous | U6.2B - gRNA target the ATG of Larp7 -Used to insert eGFP |
| **ElavC155 Gal4 (Larp7 KO)/Fm6,w** | Larp7 KO recombined with Elav | X | NA | | Mohana | Homozygous | |
| **eGFP Larp7::(Act>Cas9)** | CRISPR - endo tagged Larp7 | X | | | Mohana | Homozygous | Expression of cas9, but the flies are completely normal. |
| **RRK::CDK9(ds)/Tm6b #5** | RRK drive expression of dsRNA against CDK9 | III | | | Mohana | Recombined - validated | |
| **eGFP Larp7/Fm6,w;;7SK KO/Tm6c** | | X,III | | | Mohana | Recombined - validated | |
| **UAS - Stinger red/Cyo; 7SK KO** | | II,III | | | Mohana | Recombined - validated | |
| **Larp7 KO; GCamp6/Cyo** | | X,II | | | Mohana | Recombined - validated | |

## 12.Table 3: Primers used in this study

| | |
|---|---|
| **7SK - tRNA -HindIII FP** | cccAAGCTTgggACTTTAAAACGCCTCGCAAA |
| **7SK - tRNA RP - EcoRI - RP** | cgGAATTCcgTCATATTGCAATGCTCATGGA |
| **7SK dsRNA FP** | taatacgactcactatagggAGTAATTCTGCCTGGCGTTG |
| **7SK dsRNA RP** | taatacgactcactatagggCGCGTACATGGAAGTGTTCT |
| **7SK FL1 CC9 FP** | CTTC GTAATTCCCAAGTGCTTATT |
| **7SK FL1 CC9 RP** | AAAC AATAAGCACTTGGGAATTAC |
| **7SK FL2 CC9 FP** | CTTC GTTCGCTGCAGCAAAAGAAT |
| **7SK FL3 CC9 RP** | AAAC ATTCTTTTGCTGCAGCGAAC |
| **7SK gSQ 2-FP** | ACTTTAAAACGCCTCGCAAA |
| **7SK gSQ 2-RP** | TTGCAATGCTCATGGAAAAG |
| **7SK Pol III- HindIII FP** | cccAAGCTTgggTCCCAATTTGTGTCTCAACTG |
| **7SK qPCR 2 FP** | TGTGATTGTTCTGTACATTGATCG |
| **7SK qPCR 2 RP** | GGAAACACAAAGGTGTGACG |
| **7SK qPCR FP** | CGTCACACCTTTGTGTTTCC |
| **7SK qPCR RP** | GATAACCCGTCGTCATCCAG |
| **7SK SEQ3 FP** | ATCTGCTTCCCGTTACTTCG |
| **7SK SEQ3 RP** | TTGCAATGCTCATGGAAAAG |
| **7SK SP6 promo IS FP** | CATACGATTTAGGTGACACTATAGAAT TTTGAAAGTGGCTGACTCG |
| **7SK t7 promo IS FP** | GGATCCTAATACGACTCACTATAGGGA GTAATTCTGCCTGGCGTTG |
| **7SK-CC9-FP 1** | CTTCGCTCGGCAATGCCGCGTACA |
| **7SK-CC9-FP 2** | CTTCGCAACGCCAGGCAGAATTAC |
| **7SK-CC9-RP 1** | AAACTGTACGCGGCATTGCCGAGC |
| **7SK-CC9-RP 2** | AAACGTAATTCTGCCTGGCGTTGC |
| **GFP_ex1_F2** | tatatcatggccgacaagca |
| **GFP_ex2_R2** | gaactccagcaggaccatgt |
| **HuLa- qPCR1-FP** | AAAACAGCCAACAGGGAAGA |
| **HuLa- qPCR1-RP** | TGTTTCCTGCCAGGTAGAGG |
| **HuLa- qPCR2-FP** | ATTTGGGAAATGTGGCAATG |
| **HuLa- qPCR2-RP** | TCAATTGCTTTTGCTGCTTG |
| **Hu-LARP7- KpnI RP** | ctgtacaagggtaccTCAATCATATTCAGAAAATCTTATATG |
| **Hu-LARP7- NotI FP** | ctgtacaaggcggccgcaATGGAAACTGAAAGTGGAAATC |
| **Insert Validation DRSC FP** | CGCAGCTGAACAAGCTAAAC |
| **Insert Validation DRSC RP** | TAATCGTGTGTGATGCCTACC |
| **LARP7 3UTR T7 FP** | taatacgactcactatagggAGTTGAGGGGTGAGTTTGGA |
| **LARP7 3UTR T7 RP** | taatacgactcactatagggTTTCAATTTCAC TAGGAAGTTTCTTTT |
| **LARP7 5UTR qPCR FP** | AAAAACCAAGTGCAATATTTTATTC |
| **LARP7 5UTR qPCR RP** | CTGACCCTCGCCCTTCTC |
| **Larp7 gfptag gRNA1 FP** | CTTCGACTTAGATTACTGGTTAAC |

| | |
|---|---|
| **Larp7 gfptag gRNA1 RP** | AAACGTTAACCAGTAATCTAAGTC |
| **Larp7 gfptag gRNA2 FP** | CTTCGGCTGCCAAGAGTATCGAGA |
| **Larp7 gfptag gRNA2 RP** | AAACTCTCGATACTCTTGGCAGCC |
| **LARP7 GR SEQ FP** | GGAGCGTGGGAAATAATGGT |
| **LARP7 GR SEQ RP** | GCTCGATTTTTCGCCAATAG |
| **LARP7 orf KpnI FP** | CGGGGTACCGAATGGCTGCCAAGAGTATCGA |
| **LARP7 orf KpnI RP** | CGGGGTACCTCAGTCCATGGCGGCCGC |
| **Larp7 Promoter BglII FP** | AGATCTCTTGCGATGCGACTGTGATA |
| **Larp7 Promoter BglII RP** | AGATCTAACCAGTAATCTAAGTCACTCG |
| **Larp7 Rev KpnI HR RP** | cggggtaccGATGCGACGCTGTAGATTCA |
| **LARP7-CC9-FP 1** | CTTCGGCTGCCAAGAGTATCGAGA |
| **LARP7-CC9-FP 2** | CTTCGCCAGCCAAGCGTCGCAAGG |
| **LARP7-CC9-RP 1** | AAACTCTCGATACTCTTGGCAGCC |
| **LARP7-CC9-RP 2** | AAACCCTTGCGACGCTTGGCTGGC |
| **Lig4[169] FP** | ccttgttagccacctttgc |
| **Lig4[169] RP** | cggtcacattgcttgaacag |
| **Mu -LARP7- KpnI RP** | ctgtacaagggtaccTCAGTCATACTCAGAGA AACGGATGTGCTG |
| **Mu -LARP7- NotI FP** | tacaaggcggccgcaATGGAAACTGAAAACC AAAAAACTATGG |
| **U6.2/B seq Rev** | TACACTTTATGCTTCCGGCTCGTATGTTGT |
| **RPL15 FP** | AGGATGCACTTATGGCAAGC |
| **RPL15 RP** | GCGCAATCCAATACGAGTTC |
| **Xe -LARP7- KpnI RP** | ctgtacaagggtaccTCAGAAGCCATCACTGAATC |
| **Xe -LARP7- NotI FP** | tacaaggcggccgcaATGACTGCGATTGAAACTGACACCC |
| **Zf -LARP7- KpnI RP** | ctgtacaagggtaccTTAATCATCAAAGCGGATGTG |
| **Zf -LARP7- NotI FP** | tacaaggcggccgcaATGAAAGTGTGTCAACGC |

# 13.Bibliography


Abasi, M. *et al.* (2016) '7SK small nuclear RNA transcription level down-regulates in human tumors and stem cells', *Medical Oncology*. Springer US, 33(11), pp. 1–5. doi: 10.1007/s12032-016-0841-x.

Adelman, K. *et al.* (2005) 'Drosophila Paf1 Modulates Chromatin Structure at Actively Transcribed Genes', *Molecular and Cellular Biology*. doi: 10.1128/mcb.26.1.250-260.2006.

Adelman, K. and Henriques, T. (2018) 'Transcriptional speed bumps revealed', *Nature*, 560, pp. 560–561.

Aebi, M. and Weissman, C. (1987) 'Precision and orderliness in splicing', *Trends in Genetics*, 3(C), pp. 102–107. doi: 10.1016/0168-9525(87)90193-4.

Aikawa, M. *et al.* (2014) 'Cloning of Hexamethylene-bis-acetamide-inducible Transcript, HEXIM1, in Human Vascular Smooth Muscle Cells', *Biomedical Research*, 20(5), pp. 273–279. doi: 10.2220/biomedres.20.273.

Akhtar, J. *et al.* (2019) 'Promoter-proximal pausing mediated by the exon junction complex regulates splicing', *Nature Communications*. Nature Publishing Group, 10(1), p. 521. doi: 10.1038/s41467-019-08381-0.

Alazami, A. M. *et al.* (2012) 'Loss of function mutation in LARP7, chaperone of 7SK ncRNA, causes a syndrome of facial dysmorphism, intellectual disability, and primordial dwarfism', *Human Mutation*, 33(10), pp. 1429–1434. doi: 10.1002/humu.22175.

Ammosova, T. *et al.* (2005) 'Dephosphorylation of CDK9 by protein phosphatase 2A and protein phosphatase-I in Tat-activated HIV-I transcription', *Retrovirology*, 2, pp. 1–15. doi: 10.1186/1742-4690-2-47.

Ammosova, T. *et al.* (2011) 'Protein phosphatase-1 activates CDK9 by dephosphorylating ser175', *PLoS ONE*, 6(4). doi: 10.1371/journal.pone.0018985.

Ashton-Beaucage, D. *et al.* (2010) 'The Exon Junction Complex Controls the Splicing of mapk and Other Long Intron-Containing Transcripts in Drosophila', *Cell*, 143(2), pp. 251–262. doi: 10.1016/j.cell.2010.09.014.

Barboric, M. *et al.* (2005) 'Interplay between 7SK snRNA and oppositely charged regions in HEXIM1 direct the inhibition of P-TEFb', *EMBO Journal*, 24(24), pp. 4291–4303. doi: 10.1038/sj.emboj.7600883.

Barboric, M. *et al.* (2009) '7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0903188106.

Bayfield, M. A., Yang, R. and Maraia, R. J. (2010) 'Conserved and divergent features of the structure and function of La and La-related proteins (LARPs)', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1799(5–6), pp. 365–378. doi: 10.1016/j.bbagrm.2010.01.011.

Bensaude, O. (2011) 'Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity?', *Transcription*, 2(3), pp. 103–108. doi: 10.4161/trns.2.3.16172.

Bousquet-Antonelli, C. and Deragon, J. M. (2009) 'A comprehensive analysis of the La-motif protein superfamily', *Rna*, 15(5), pp. 750–764. doi: 10.1261/rna.1478709.

Boyd, D. C. *et al.* (1995) 'Functional redundancy of promoter elements ensures efficient transcription of the human 7SK gene in vivo', *Journal of Molecular Biology*, 253(5), pp. 677–690. doi: 10.1006/jmbi.1995.0582.

Braberg, H. *et al.* (2013) 'From structure to systems: High-resolution, quantitative genetic analysis of RNA polymerase II', *Cell*. Elsevier Inc., 154(4), pp. 775–788. doi: 10.1016/j.cell.2013.07.033.

Brand, A. H. and Perrimon, N. (1993) 'Targeted gene expression as a means of altering cell fates and generating dominant phenotypes', *Development*. doi: 10.1101/lm.1331809.

Briese, M. *et al.* (2018) 'hnRNP R and its main interactor, the noncoding RNA 7SK, coregulate the axonal transcriptome of motoneurons', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1721670115.

Cai, L. *et al.* (2002) 'Identification of a Novel Cardiac Lineage-Associated Protein(cCLP-1): A Candidate Regulator of Cardiogenesis', *Developmental Biology*, 208(1), pp. 210–221. doi: 10.1006/dbio.1998.9180.

Cardona, A., Larsen, C. and Hartenstein, V. (2009) 'Neuronal fiber tracts connecting the brain and ventral nerve cord of the early Drosophila larva', *Journal of Comparative Neurology*, 515(4), pp. 427–440. doi: 10.1002/cne.22086.

Chance, M. R. *et al.* (2015) 'Phosphorylation of HEXIM1 at Tyr271 and Tyr274 Promotes Release of P-TEFb from the 7SK snRNP Complex and Enhances Proviral HIV Gene Expression', *Proteomics*, 15(12), pp. 2078–2086. doi: 10.1002/pmic.201500038.

Chen, F. X. *et al.* (2015) 'PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II', *Cell*. Elsevier Inc., 162(5), pp. 1003–1015. doi: 10.1016/j.cell.2015.07.042.

Cheng, B. and Price, D. H. (2007) 'Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M702936200.

Cheng, Y. *et al.* (2012) 'LARP7 is a potential tumor suppressor gene in gastric cancer', *Laboratory Investigation*. doi: 10.1038/labinvest.2012.59.

Cho, S. *et al.* (2011) 'Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly', *Proceedings of the National Academy of Sciences*, 108(20), pp. 8233–8238. doi: 10.1073/pnas.1017700108.

Christian, H. *et al.* (2013) 'Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system', *Genes and Development*, 23, pp. 413–428. doi: 10.1101/gad.207779.112.Spliceosome.

Connelly, S. and Manley, J. (1988) 'A functional .mRNA polyadeny.lation signal is required for transcnpnon termination by RNA polymerase II', *Genes & Development*, pp. 440–452.

Conte, M. R. *et al.* (2012) 'The association of a La module with the PABP-interacting motif PAM2 is a recurrent evolutionary process that led to the neofunctionalization of La-related proteins', *Rna*, 19(1), pp. 36–50. doi: 10.1261/rna.035469.112.

Contreras, X. *et al.* (2007) 'HMBA releases P-TEFb from HEXIM1 and 7SK snRNA via PI3K/Akt and activates HIV transcription', *PLoS Pathogens*, 3(10), pp. 1459–1469. doi: 10.1371/journal.ppat.0030146.

Cramer, P. *et al.* (1997) 'Functional Association between Promoter Structure and Transcript Alternative Splicing Author ( s ): Paula Cramer , C . Gustavo Pesce , Francisco E . Baralle and Alberto R . Kornblihtt Source : Proceedings of the National Academy of Sciences of the United', 94(21), pp. 11456–11460.

Dahm, R. (2005) 'Friedrich Miescher and the discovery of DNA', *Developmental Biology*. Academic Press, 278(2), pp. 274–288. doi: 10.1016/J.YDBIO.2004.11.028.

Deleuze, J.-F. *et al.* (2018) 'LARP7 variants and further delineation of the Alazami syndrome phenotypic spectrum among primordial dwarfisms: 2 sisters', *European Journal of Medical Genetics*. Elsevier, 62(3), pp. 161–166. doi: 10.1016/j.ejmg.2018.07.003.

Deragon, J. M. and Bousquet-Antonelli, C. (2015) 'The role of LARP1 in translation and beyond', *Wiley Interdisciplinary Reviews: RNA*, 6(4), pp. 399–417. doi: 10.1002/wrna.1282.

Diribarne, G. and Bensaude, O. (2009) '7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor', *RNA Biology*, 6(2), pp. 122–128. doi: 10.4161/rna.6.2.8115.

Dujardin, G. *et al.* (2014) 'How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping', *Molecular Cell*, 54(4), pp. 683–690. doi: 10.1016/j.molcel.2014.03.044.

Eichhorn, C. D., Chug, R. and Feigon, J. (2016) 'hLARP7 C-terminal domain contains an xRRM that binds the 3′ hairpin of 7SK RNA', *Nucleic Acids Research*, 44(20), pp. 9977–9989. doi: 10.1093/nar/gkw833.

Erdmann, I. *et al.* (2015) 'Cell-selective labelling of proteomes in Drosophila melanogaster', *Nature Communications*. doi: 10.1038/ncomms8521.

Frank, D. and Guthrie, C. (1992) 'An essential splicing factor, SLU7, mediates 3' splice site choice in yeast', *Genes and Development*, 6(11), pp. 2112–2124. doi: 10.1101/gad.6.11.2112.

Fujinaga, K. *et al.* (2004) 'P-TEFb Phosphorylates RD

and Dissociates Negative Effectors from the Transactivation Response Element', *Molecular and Cellular Biology*, 24(2), pp. 787–795. doi: 10.1128/MCB.24.2.787.

Gilmour, D. S. and Lis, J. T. (1986a) 'RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells.', *Molecular and Cellular Biology*. doi: 10.1128/MCB.6.11.3984.

Gilmour, D. S. and Lis, J. T. (1986b) 'RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells.', *Molecular and cellular biology*. American Society for Microbiology Journals, 6(11), pp. 3984–9. doi: 10.1128/MCB.6.11.3984.

Gomez-Marin, A. and Louis, M. (2012) 'Active sensation during orientation behavior in the Drosophila larva: More sense than luck', *Current Opinion in Neurobiology*. Elsevier Ltd, 22(2), pp. 208–215. doi: 10.1016/j.conb.2011.11.008.

Grainger, R. J. and Beggs, J. D. (2003) 'Prp8 protein: At the heart of the spliceosome', *Stroitel'nye Materialy*, (4), pp. 31–34. doi: 10.1261/rna.2220705.RNA.

Graña, X. *et al.* (1994) 'PITALRE, a nuclear CDC2-related protein kinase that phosphorylates the retinoblastoma protein in vitro', *Proc. Natl. Acad. Sci. USA*, 91(9), pp. 3834–3838. doi: 10.1016/0272-7358(92)90024-3.

Green, C. H., Burnet, B. and Connolly, K. J. (1983) 'Organization and patterns of inter- and intraspecific variation in the behaviour of Drosophila larvae', *Animal Behaviour*, 31(1), pp. 282–291. doi: 10.1016/S0003-3472(83)80198-5.

Gruber, A. R. *et al.* (2008) 'Arthropod 7SK RNA', *Molecular Biology and Evolution*, 25(9), pp. 1923–1930. doi: 10.1093/molbev/msn140.

Gunning, P. W. *et al.* (1981) 'Differential and coordinate regulation of the eukaryotic small molecular weight RNAs.', *Journal of Biological Chemistry*, 256(13), pp. 6663–6669.

Guo, S. *et al.* (2000) 'A regulator of transcriptional elongation controls vertebrate neuronal development', *Nature*. doi: 10.1038/35042590.

Hamm, J. *et al.* (1990) 'The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal', *Cell*, 62(3), pp. 569–577. doi: 10.1016/0092-8674(90)90021-6.

He, N. *et al.* (2008) 'A La-Related Protein Modulates 7SK snRNP Integrity to Suppress P-TEFb-Dependent Transcriptional Elongation and Tumorigenesis', *Molecular Cell*, 29(5), pp. 588–599. doi: 10.1016/j.molcel.2008.01.003.

Hendrix, D. A. *et al.* (2008) 'Promoter elements associated with RNA Pol II stalling in the Drosophila embryo', *Proceedings of the National Academy of Sciences*, 105(22), pp. 7762–7767. doi: 10.1073/pnas.0802406105.

Hirose, Y., Tacke, R. and Manley, J. L. (1999) 'Phosphorylated RNA polymerase II stimulates pre-mRNA splicing', *Genes and Development*, 13(10), pp. 1234–1239. doi: 10.1101/gad.13.10.1234.

Ho, C. K. and Shuman, S. (1999) 'Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme', *Molecular Cell*, 3(3), pp. 405–411. doi: 10.1016/S1097-2765(00)80468-2.

Hollink, I. H. I. M. *et al.* (2016) 'Broadening the phenotypic spectrum of pathogenic LARP7 variants: Two cases with intellectual disability, variable growth retardation and distinct facial features', *Journal of Human Genetics*. doi: 10.1038/jhg.2015.134.

Holohan, B. *et al.* (2016) 'Impaired telomere maintenance in Alazami syndrome patients with LARP7 deficiency', *BMC Genomics*. BMC Genomics, 17(Suppl 9), pp. 1–9. doi: 10.1186/s12864-016-3093-4.

Howe, K. J., Kane, C. M. and Ares, M. (2003) 'Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae', *Rna*, 9(8), pp. 993–1006. doi: 10.1261/rna.5390803.

Huang, Y. *et al.* (2012) 'Mediator Complex Regulates Alternative mRNA Processing via the MED23 Subunit', *Molecular Cell*. Elsevier Inc., 45(4), pp. 459–469. doi: 10.1016/j.molcel.2011.12.022.

Ilagan, J. O. *et al.* (2013) 'Rearrangements within human spliceosomes captured after exon ligation', *Rna*, 19(3),

pp. 400–412. doi: 10.1261/rna.034223.112.

J.C., S. *et al.* (2012) 'FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2', *Genes and Development*, 26(24), pp. 2690–2695. doi: 10.1101/gad.204602.112 LK - http://sfxprd.tau.ac.iI:3410/sfxIc141?sid=EMBASE&issn=08909369&id=doi:10.1101%2Fgad.204602.112&atitle=FUS+binds+the+CTD+of+RNA+polymerase+II+and+regulates+its+phosphorylation+at+Ser2&stitle=Genes+Dev.&title=Genes+and+Development&volume=26&issue=24&spage=2690&epage=2695&aulast=Schwartz&aufirst=Jacob+C.&auinit=J.C.&aufull=Schwartz+J.C.&coden=GEDEE&isbn=&pages=2690-2695&date=2012&auinit1=J&auinitm=C.

James, C. (2005) 'Preparation and Immunostaining of Polytene Chromosome Squashes (PROT01)', *The EPIGENOME Network of Excellence*, 3, pp. 4–6.

Jennings, B. H. (2013) 'Pausing for thought: Disrupting the early transcription elongation checkpoint leads to developmental defects and tumourigenesis', *BioEssays*, 35(6), pp. 553–560. doi: 10.1002/bies.201200179.

Jeronimo, C. *et al.* (2007) 'Systematic Analysis of the Protein Interaction Network for the Human Transcription Machinery Reveals the Identity of the 7SK Capping Enzyme', *Molecular Cell*, 27(2), pp. 262–274. doi: 10.1016/j.molcel.2007.06.027.

Ji, X. *et al.* (2014) 'LARP7 suppresses P-TEFb activity to inhibit breast cancer progression and metastasis', *eLife*. doi: 10.7554/eLife.02907.

Jong-bok, Y. *et al.* (1995) 'Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes.', *Molecular and Cellular Biology*, 15(4), pp. 2019–2027. doi: 10.1128/mcb.15.4.2019.

Juven-Gershon, T. *et al.* (2008) 'The RNA polymerase II core promoter — the gateway to transcription', *Current Opinion in Cell Biology*. Elsevier Current Trends, 20(3), pp. 253–259. doi: 10.1016/J.CEB.2008.03.003.

Kedinger, C. *et al.* (1970) 'α-Amanitin: A specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus', *Biochemical and Biophysical Research Communications*. Academic Press, 38(1), pp. 165–171. doi: 10.1016/0006-291X(70)91099-5.

Keene, A. C. and Sprecher, S. G. (2012) 'Seeing the light: Photobehavior in fruit fly larvae', *Trends in Neurosciences*. Elsevier Ltd, 35(2), pp. 104–110. doi: 10.1016/j.tins.2011.11.003.

Keramati, F. *et al.* (2015) '7SK small nuclear RNA inhibits cancer cell proliferation through apoptosis induction', *Tumor Biology*, 36(4), pp. 2809–2814. doi: 10.1007/s13277-014-2907-8.

Keshishian, H. (2002) 'The Drosophila Neuromuscular Junction: A Model System for Studying Synaptic Development and Function', *Annual Review of Neuroscience*, 19(1), pp. 545–575. doi: 10.1146/annurev.neuro.19.1.545.

Khodor, Y. L. *et al.* (2011) 'Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila.', *Genes & development*. Cold Spring Harbor Laboratory Press, 25(23), pp. 2502–12. doi: 10.1101/gad.178962.111.

Kim, J. B. and Sharp, P. A. (2001) 'Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent Kinase-activating Kinase', *Journal of Biological Chemistry*, 276(15), pp. 12317–12323. doi: 10.1074/jbc.M010908200.

Kim, M. D., Wen, Y. and Jan, Y. N. (2009) 'Patterning and organization of motor neuron dendrites in the Drosophila larva', *Developmental Biology*. Elsevier Inc., 336(2), pp. 213–221. doi: 10.1016/j.ydbio.2009.09.041.

Kitajima, T. *et al.* (2018) 'Novel compound heterozygous variants in the LARP7 gene in a patient with Alazami syndrome', *Human Genome Variation*. The Author(s), 5(December 2017), p. 18014. doi: 10.1038/hgv.2018.14.

Kleinert, H., Bredow, S. and Benecke, B. J. (2018) 'Expression of a human 7S K RNA gene in vivo requires a novel pol III upstream element.', *The EMBO Journal*, 9(3), pp. 711–718. doi: 10.1002/j.1460-2075.1990.tb08164.x.

Kohsaka, H. *et al.* (2012) 'Development of larval motor circuits in Drosophila', *Development Growth and Differentiation*, 54(3), pp. 408–419. doi: 10.1111/j.1440-

169X.2012.01347.x.

Kolesnikova, O. *et al.* (2016) 'An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb', *Proceedings of the National Academy of Sciences*, 113(45), pp. 12721–12726. doi: 10.1073/pnas.1612331113.

Kramer, A. (1996) 'THE STRUCTURE AND FUNCTION OF PROTEINS PRE-mRNA SPLICING', *Annual review of biochemistry*, 65, pp. 367–409. doi: 10.1146/annurev.biochem.65.1.367.

Krueger, B. J. *et al.* (2010) 'The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK', *PLoS ONE*, 5(8). doi: 10.1371/journal.pone.0012335.

Kumar, J. P. (2016) 'Single-Fly Genomic DNA Extraction', *Lab Protocol*, (http://www.indiana.edu/~kumarlab/7_labresources/protocols/016%20Single%20Fly%20Genomic%20DNA%20Extraction.pdf), p. 200. Available at: http://www.indiana.edu/~kumarlab/7_labresources/protocols/016 Single Fly Genomic DNA Extraction.pdf.

De La Flor, M. *et al.* (2017) 'Drosophila increase exploration after visually detecting predators', *PLoS ONE*, 12(7), pp. 1–17. doi: 10.1371/journal.pone.0180749.

de la Mata, M. *et al.* (2003) 'A Slow RNA Polymerase II Affects Alternative Splicing In Vivo', *Molecular Cell*. Cell Press, 12(2), pp. 525–532. doi: 10.1016/J.MOLCEL.2003.08.001.

Landgraf, M. *et al.* (2018) ' The Origin, Location, and Projections of the Embryonic Abdominal Motorneurons of Drosophila ', *The Journal of Neuroscience*, 17(24), pp. 9642–9655. doi: 10.1523/jneurosci.17-24-09642.1997.

Landgraf, M. and Thor, S. (2006a) 'Development and Structure of Motoneurons', *International Review of Neurobiology*, 75(06), pp. 33–53. doi: 10.1016/S0074-7742(06)75002-4.

Landgraf, M. and Thor, S. (2006b) 'Development of Drosophila motoneurons: Specification and morphology', *Seminars in Cell and Developmental Biology*, 17(1), pp. 3–11. doi: 10.1016/j.semcdb.2005.11.007.

Lence, T. *et al.* (2016) 'M6A modulates neuronal functions and sex determination in Drosophila', *Nature*. Nature Publishing Group, 540(7632), pp. 242–247. doi: 10.1038/nature20568.

Levene, P A La Froge, F. B. (1912) 'Über die Hefe-Nucleinsäurw. V. Die Struktur der Pyrimidin-Nucleoside', *European Journal of Inorganic Chemistry*, (March).

Levene, P. A. (1917) 'Yeast Nucleic Acid . II', *Journal of Biological Chemistry*.

Levene, P. A. and Jacobe, W. A. (1911) 'Uber die Hefe-nucleinsäure', *European Journal of Inorganic Chemistry*.

Lew, Q. J. *et al.* (2013) 'Hexim1, a new player in the p53 pathway', *Cancers*, 5(3), pp. 838–856. doi: 10.3390/cancers5030838.

Lin, C. *et al.* (2010) 'AFF4, a Component of the ELL/P-TEFb Elongation Complex and a Shared Subunit of MLL Chimeras, Can Link Transcription Elongation to Leukemia', *Molecular Cell*. Elsevier Ltd, 37(3), pp. 429–437. doi: 10.1016/j.molcel.2010.01.026.

Ling, T. T. and Sorrentino, S. (2016) 'Compound heterozygous variants in the LARP7 gene as a cause of Alazami syndrome in a Caucasian female with significant failure to thrive, short stature, and developmental disability', *American Journal of Medical Genetics, Part A*, 170(1), pp. 217–219. doi: 10.1002/ajmg.a.37396.

Luo, Z., Lin, C. and Shilatifard, A. (2012) 'The super elongation complex (SEC) family in transcriptional control', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 13(9), pp. 543–547. doi: 10.1038/nrm3417.

Maraia, R. J. and Intine, R. V. (2002) 'La protein and its associated small nuclear and nucleolar precursor RNAs', *Gene Expression*, 10(1–2), pp. 41–57.

Marais, G. *et al.* (2005) 'Intron size and exon evolution in Drosophila', *Genetics*, 170(1), pp. 481–485. doi: 10.1534/genetics.104.037333.

Marz, M. *et al.* (2008) 'Invertebrate 7SK snRNAs', *Journal of Molecular Evolution*, 66(2), pp. 107–115. doi: 10.1007/s00239-007-9052-6.

Marz, M. *et al.* (2009) 'Evolution of 7SK RNA and its protein partners in metazoa', *Molecular Biology and Evolution*, 26(12), pp. 2821–2830. doi: 10.1093/molbev/msp198.

Matrone, G. *et al.* (2015) 'CDK9 and its repressor LARP7 modulate cardiomyocyte proliferation and response to injury in the zebrafish heart', *Journal of Cell Science*, 128(24), pp. 4560–4571. doi: 10.1242/jcs.175018.

Matsuis, T. *et al.* (1980) 'Multiple Factors Required for Accurate Initiation of Transcription by', *The Journal of Biological Chemistry*, 255(24), pp. 11992–11996.

Mbonye, U. R. *et al.* (2013) 'Phosphorylation of CDK9 at Ser175 Enhances HIV Transcription and Is a Marker of Activated P-TEFb in CD4+T Lymphocytes', *PLoS Pathogens*, 9(5). doi: 10.1371/journal.ppat.1003338.

McCracken, S. *et al.* (1996) 'The C-terminal domain of RNA polymerase II couples mRNA processing to transcription', *Tomarev, S. I. & Piatigorsky, ). Eur. ]. Biochem*, 10(9), pp. 807–809.

McCracken, S. *et al.* (1997) '5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II', *Genes and Development*, 11(24), pp. 3306–3318. doi: 10.1101/gad.11.24.3306.

McNamara, R. P. *et al.* (2016) 'KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II', *Molecular Cell*, 61(1), pp. 39–53. doi: 10.1016/j.molcel.2015.11.004.

Mennie, A. K., Moser, B. A. and Nakamura, T. M. (2018) 'LARP7-like protein Pof8 regulates telomerase assembly and poly(A)+TERRA expression in fission yeast', *Nature Communications*. Springer US, 9(1). doi: 10.1038/s41467-018-02874-0.

Menon, K. P., Carrillo, R. A. and Zinn, K. (2013) ' Development and plasticity of the Drosophila larval neuromuscular junction ', *Wiley Interdisciplinary Reviews: Developmental Biology*, 2(5), pp. 647–670. doi: 10.1002/wdev.108.

Michels, A. *et al.* (2003) 'MAQ1 and 7SK RNA Interact with CDK9 / Cyclin T Complexes in a Transcription-Dependent Manner', *Molecular and cellular biology*, 23(14), pp. 4859–4869. doi: 10.1128/MCB.23.14.4859-4869.2003.

Michels, A. A. *et al.* (2004) 'Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor', *EMBO Journal*, 23(13), pp. 2608–2619. doi: 10.1038/sj.emboj.7600275.

Michels, A. A. and Bensaude, O. (2018) 'Hexim1, an RNA-controlled protein hub', *Transcription*. Taylor & Francis, 9(4), pp. 262–271. doi: 10.1080/21541264.2018.1429836.

Moehle, E. A. *et al.* (2014) 'Adventures in time and space', *RNA Biology*. Taylor & Francis, 11(4), pp. 313–319. doi: 10.4161/rna.28646.

Morchikh, M. *et al.* (2017) 'HEXIM1 and NEAT1 Long Non-coding RNA Form a Multi-subunit Complex that Regulates DNA-Mediated Innate Immune Response', *Molecular Cell*. doi: 10.1016/j.molcel.2017.06.020.

Morris, D. P. and Greenleaf, A. L. (2000) 'The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA Polymerase II', *Journal of Biological Chemistry*, 275(51), pp. 39935–39943. doi: 10.1074/jbc.M004118200.

Mugat, B. *et al.* (2003) 'Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in Drosophila', *Rna*, 9(3), pp. 299–308. doi: 10.1261/rna.2154103.

Muniz, L., Egloff, S. and Kiss, T. (2013) 'RNA elements directing in vivo assembly of the 7SK/MePCE/Larp7 transcriptional regulatory snRNP', *Nucleic Acids Research*, 41(8), pp. 4686–4698. doi: 10.1093/nar/gkt159.

Murphy, S. *et al.* (1989) 'Purified octamer binding transcription factors stimulate RNA polymerase III-mediated transcription of the 7SK RNA gene', *Cell*, 59(6), pp. 1071–1080. doi: 10.1016/0092-8674(89)90763-0.

Murphy, S. and Liegro, C. Di (1987) 'he In Vitro Transcription of the 7SK RNA Gene by RNA Polymerase Ill Is Dependent Only on the Presence of an Upstream Promoter trscrs-CCCCACCCATCT6CAA66CATTCT66ATA6T6TcAAAAcA6', *Cell*, 51, pp. 81–87.

Murphy, S., Tripodi, M. and Melli, M. (1986) 'A

sequence upstream from the coding region is required for the transcription of the 7SK RNA genes', *Nucleic Acids Research*, 14(23), pp. 9243–9260. doi: 10.1093/nar/14.23.9243.

Napolitano, G. *et al.* (2005) 'Increased HEXIM1 expression during erythroleukemia and neuroblastoma cell differentiation', *Journal of Cellular Physiology*, 206(3), pp. 603–610. doi: 10.1002/jcp.20502.

Narasimha, S. *et al.* (2019) 'Drosophila melanogaster cloak their eggs with pheromones, which prevents cannibalism', *PLoS biology*, 17(1), p. e2006012. doi: 10.1371/journal.pbio.2006012.

Nechaev, S., Fargo, David C, *et al.* (2010) 'Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 327(5963), pp. 335–8. doi: 10.1126/science.1181421.

Nechaev, S., Fargo, David C., *et al.* (2010) 'Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila', *Science*. doi: 10.1126/science.1181421.

Nguyen, D. *et al.* (2012a) 'The Drosophila 7SK snRNP and the essential role of dHEXIM in development', *Nucleic Acids Research*, 40(12), pp. 5283–5297. doi: 10.1093/nar/gks191.

Nguyen, D. *et al.* (2012b) 'The Drosophila 7SK snRNP and the essential role of dHEXIM in development', *Nucleic Acids Research*, 40(12), pp. 5283–5297. doi: 10.1093/nar/gks191.

Nguyen, V. T. *et al.* (2001) '7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes', *Nature*, 414(6861), pp. 322–325. doi: 10.1038/35104581.

O'Keeffe, B. *et al.* (2000) 'Requirement for a kinase-specific chaperone pathway in the production of a Cdk9/cyclin T1 heterodimer responsible for P-TEFb-mediated Tat stimulation of HIV-1 transcription', *Journal of Biological Chemistry*, 275(1), pp. 279–287. doi: 10.1074/jbc.275.1.279.

Okamura, D. *et al.* (2012a) 'Cell cycle gene-specific control of transcription has a critical role in proliferation of primordial germ cells', *Genes and Development*. doi: 10.1101/gad.202242.112.

Okamura, D. *et al.* (2012b) 'Cell cycle gene-specific control of transcription has a critical role in proliferation of primordial germ cells', *Genes and Development*, 26(22), pp. 2477–2482. doi: 10.1101/gad.202242.112.

Orphanides, G., Lagrange, T. and Reinberg, D. (1996) 'The general transcription machinery of RNA polymerase II.', *Genes & Dev.*, 10(7), pp. 2657–2683.

Osheim, Y. N., O.L. Miller, J. and Beyer, A. L. (1985) 'RNP particles at splice junction sequences on Drosophila chorion transcripts', *Cell*, 43(1), pp. 143–151. doi: 10.1016/0092-8674(85)90019-4.

Van Oss, S. B. *et al.* (2016) 'The Histone Modification Domain of Paf1 Complex Subunit Rtf1 Directly Stimulates H2B Ubiquitylation through an Interaction with Rad6', *Molecular Cell*. doi: 10.1016/j.molcel.2016.10.008.

Pabis, M. *et al.* (2013) 'The nuclear cap-binding complex interacts with the U4/U6·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells', *Rna*, 19(8), pp. 1054–1063. doi: 10.1261/rna.037069.112.

Páez-Moscoso, D. J. *et al.* (2018) 'Pof8 is a La-related protein and a constitutive component of telomerase in fission yeast', *Nature Communications*. Springer US, 9(1), pp. 1–11. doi: 10.1038/s41467-017-02284-8.

Parchem, R. J. *et al.* (2015) 'miR-302 Is Required for Timing of Neural Differentiation, Neural Tube Closure, and Embryonic Viability', *Cell Reports*. doi: 10.1016/j.celrep.2015.06.074.

Pastor-Pareja, J. C. and Xu, T. (2011) 'Shaping Cells and Organs in Drosophila by Opposing Roles of Fat Body-Secreted Collagen IV and Perlecan', *Developmental Cell*, 21(2), pp. 245–256. doi: 10.1016/j.devcel.2011.06.026.

Perales, R. and Bentley, D. (2009) '"Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions', *Molecular Cell*. Elsevier, 36(2), pp. 178–191. doi: 10.22266/ijies2017.0630.45.

Perrimon, N., Zirin, J. and Bai, J. (2011) 'Primary Cell Cultures from Drosophila Gastrula Embryos', *Journal of Visualized Experiments*, (48), pp.

40–43. doi: 10.3791/2215.

Peterlin, B. M., Brogie, J. E. and Price, D. H. (2012) '7SK snRNA: A noncoding RNA that plays a major role in regulating eukaryotic transcription', *Wiley Interdisciplinary Reviews: RNA*, 3(1), pp. 92–103. doi: 10.1002/wrna.106.

Peterlin, B. M. and Price, D. H. (2006) 'Controlling the Elongation Phase of Transcription with P-TEFb', *Molecular Cell*. doi: 10.1016/j.molcel.2006.06.014.

Porrua, O. and Libri, D. (2015) 'Transcription termination and the control of the transcriptome: why, where and how to stop', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(3), pp. 190–202. doi: 10.1038/nrm3943.

Prasanth, K. V. *et al.* (2010) 'Nuclear Organization and Dynamics of 7SK RNA in Regulating Gene Expression', *Molecular Biology of the Cell*, 21, pp. 4184–4196. doi: 10.1091/mbc.e10-02-0105.

Proudfoot, N. J. (1989) 'How RNA polymerase II terminates transcription in higher eukaryotes', *Trends in Biochemical Sciences*, 14(3), pp. 105–110. doi: 10.1016/0968-0004(89)90132-1.

Pugh, B. F. (1996) 'Mechanisms of transcription complex assembly', *Current Opinion in Cell Biology*. Elsevier Current Trends, 8(3), pp. 303–311. doi: 10.1016/S0955-0674(96)80002-0.

Radhakrishnan, S. K. and Gartel, A. L. (2006) 'CDK9 phosphorylates p53 on serine residues 33, 315 and 392', *Cell Cycle*, 5(5), pp. 519–521. doi: 10.4161/cc.5.5.2514.

Raghunathan, P. L. and Guthrie, C. (1998) 'RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2', *Current Biology*, 8(15), pp. 847–855. doi: 10.1016/S0960-9822(07)00345-4.

Ramanathan, Y. *et al.* (2001) 'Three RNA Polymerase II Carboxyl-terminal Domain Kinases Display Distinct Substrate Preferences', *Journal of Biological Chemistry*, 276(14), pp. 10913–10920. doi: 10.1074/jbc.M010975200.

Reddy, R. *et al.* (1984) 'Primary and secondary structure of 7-3 (K) RNA of Novikoff hepatoma', *Journal of Biological Chemistry*, 259(19), pp. 12265–12270.

Richard, P. and Manley, J. L. (2009) 'Transcription termination by nuclear RNA polymerases', *Genes and Development*, 23(11), pp. 1247–1269. doi: 10.1101/gad.1792809.

Roeder, R. G. and Rutter, W. J. (1969) 'Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms', *Nature*. doi: 10.1038/224234a0.

Rogers, S. L. and Rogers, G. C. (2008) 'Culture of Drosophila S2 cells and their use for RNAi-mediated loss-of-function studies and immunofluorescence microscopy', *Nature Protocols*, 3(4), pp. 606–611. doi: 10.1038/nprot.2008.18.

Roignant, J. Y. and Treisman, J. E. (2010) 'Exon Junction Complex Subunits Are Required to Splice Drosophila MAP Kinase, a Large Heterochromatic Gene', *Cell*. Elsevier Inc., 143(2), pp. 238–250. doi: 10.1016/j.cell.2010.09.036.

Rougvie, A. E. and Lis, J. T. (1988) 'The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged.', *Cell*, 54(6), pp. 795–804. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3136931 (Accessed: 13 February 2019).

Roy, A. L. and Singer, D. S. (2015) 'Core promoters in transcription: Old problem, new insights', *Trends in Biochemical Sciences*. doi: 10.1016/j.tibs.2015.01.007.

Sabo, A. *et al.* (2008) 'Acetylation of Conserved Lysines in the Catalytic Core of Cyclin-Dependent Kinase 9 Inhibits Kinase Activity and Regulates Transcription', *Molecular and Cellular Biology*, 28(7), pp. 2201–2212. doi: 10.1128/MCB.01557-07.

Sainsbury, S., Bernecky, C. and Cramer, P. (2015) 'Structural basis of transcription initiation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(3), pp. 129–143. doi: 10.1038/nrm3952.

Schaub, M. *et al.* (1997) 'Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III', *EMBO Journal*, 16(1), pp. 173–181. doi: 10.1093/emboj/16.1.173.

Schindelin, J. *et al.* (2012) 'Fiji: An open-source platform for biological-image analysis', *Nature Methods*, 9(7), pp. 676–682. doi: 10.1038/nmeth.2019.

Schor, I. E. *et al.* (2009) 'Neuronal cell depolarization

induces intragenic chromatin modifications affecting NCAM alternative splicing', *Proceedings of the National Academy of Sciences*, 106(11), pp. 4325–4330. doi: 10.1073/pnas.0810666106.

Schroeder, S. C. *et al.* (2004) 'A Function of Yeast mRNA Cap Methyltransferase, Abd1, in Transcription by RNA Polymerase II', *Molecular Cell*, 13(3), pp. 377–387. doi: 10.1016/S1097-2765(04)00007-3.

Schuster, C. M. *et al.* (1996) 'Genetic dissection of structural and functional components of synaptic plasticity. II. Fasciclin II controls presynaptic structural plasticity', *Neuron*, 17(4), pp. 655–667. doi: 10.1016/S0896-6273(00)80198-1.

Schwer, B. and Gross, C. H. (1998) 'Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing', *EMBO Journal*, 17(7), pp. 2086–2094. doi: 10.1093/emboj/17.7.2086.

Schwer, B. and Guthrie, C. (1991) 'PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome', *Nature*, 349(6309), pp. 494–499. doi: 10.1038/349494a0.

Seifart, K. H. and Sekeris, C. E. (1969) 'α Amanitin, a specific inhibitor of transcription by mammalian RNA-polymerase', *Zeitschrift für Naturforschung B*. Verlag der Zeitschrift für Naturforschung, 24(12), pp. 1538–1544. doi: 10.1515/znb-1969-1211.

Shi, X. *et al.* (2015) 'Paf1p, an RNA polymerase II-associated factor in Saccharomyces cerevisiae, may have both positive and negative roles in transcription.', *Molecular and Cellular Biology*. doi: 10.1128/mcb.16.2.669.

Shilatifard, A. *et al.* (1996) 'An RNA polymerase II elongation factor encoded by the human ELL gene', *Science*, 271(5257), pp. 1873–1876. doi: 10.1126/science.271.5257.1873.

Shimba, S. and Reddy, R. (1994) 'Purification of human U6 small nuclear RNA capping enzyme. Evidence for a common capping enzyme for ??-monomethyl-capped small RNAs', *Journal of Biological Chemistry*, 269(17), pp. 12419–12423.

Singh, N., Morlock, H. and Hanes, S. D. (2011) 'The Bin3 RNA methyltransferase is required for repression of caudal translation in the Drosophila embryo', *Developmental Biology*. Elsevier Inc., 352(1), pp. 104–115. doi: 10.1016/j.ydbio.2011.01.017.

Sinha, K. M. *et al.* (2002) 'Adenylation of Small RNAs in Human Cells', *Journal of Biological Chemistry*, 273(12), pp. 6853–6859. doi: 10.1074/jbc.273.12.6853.

Skene, P. J. and Henikoff, S. (2017) 'An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites', *eLife*, 6, pp. 1–35. doi: 10.7554/elife.21856.

Slomnicki, L. P. *et al.* (2016) 'Nucleolar Enrichment of Brain Proteins with Critical Roles in Human Neurodevelopment', *Molecular & Cellular Proteomics*, 15(6), pp. 2055–2075. doi: 10.1074/mcp.m115.051920.

Spector, D. L. and Misteli, T. (1999) 'RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo', *Molecular cell*, 3(6), pp. 697–705.

Staknis, D. and Reed, R. (1994) 'SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex.', *Molecular and Cellular Biology*, 14(11), pp. 7670–7682. doi: 10.1128/MCB.14.11.7670.

Sulkowski, M. J. *et al.* (2016) 'A Novel, Noncanonical BMP Pathway Modulates Synapse Maturation at the Drosophila Neuromuscular Junction', *PLoS Genetics*, 12(1), pp. 1–31. doi: 10.1371/journal.pgen.1005810.

Sun, J. S. and Manley, J. L. (1995) 'A novel U2-U6 snRNA structure is necessary for mallalian messenger-RNA splicing', *Genes Dev.*, 9, pp. 843–854.

Suzuki, R. and Arita, T. (2012) 'Coevolution of game strategies, game structures and network structures', *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, 87 LNICST, pp. 143–154. doi: 10.1007/978-3-642-32615-8_17.

Takahashi, H. *et al.* (2011) 'Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors', *Cell*. Elsevier Inc., 146(1), pp. 92–104. doi: 10.1016/j.cell.2011.06.005.

Uchikawa, E. *et al.* (2015a) 'Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7', *Nucleic Acids Research*, 43(6), pp. 3373–3388. doi: 10.1093/nar/gkv173.

Uchikawa, E. *et al.* (2015b) 'Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7', *Nucleic Acids Research*. doi: 10.1093/nar/gkv173.

Ullu, E. and Melli, M. (1982) 'Cloning and characterization of cDNA copies of the 7S RNAs of HeLa cells', *Nucleic Acids Research*, 10(7), pp. 2209–2223. doi: 10.1093/nar/10.7.2209.

Vijendravarma, R. K., Narasimha, S. and Kawecki, T. J. (2013) 'Predatory cannibalism in Drosophila melanogaster larvae', *Nature Communications*. Nature Publishing Group, 4, pp. 1788–1789. doi: 10.1038/ncomms2744.

Vos, S. M., Farnung, L., Boehning, M., *et al.* (2018) 'Structure of activated transcription complex Pol II–DSIF–PAF–SPT6', *Nature*. doi: 10.1038/s41586-018-0440-4.

Vos, S. M., Farnung, L., Urlaub, H., *et al.* (2018) 'Structure of paused transcription complex Pol II–DSIF–NELF', *Nature*. Nature Publishing Group, 560(7720), pp. 601–606. doi: 10.1038/s41586-018-0442-2.

Wada, T. *et al.* (1996) 'DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs', pp. 343–356.

Wang, Y. *et al.* (2008) 'Phosphatase PPM1A regulates phosphorylation of Thr-186 in the Cdk9 T-loop', *Journal of Biological Chemistry*, 283(48), pp. 33578–33584. doi: 10.1074/jbc.M807495200.

Watanabe, Y. *et al.* (2000) 'Modulation of TFIIH-associated kinase activity by complex formation and its relationship with CTD phosphorylation of RNA polymerase II', *Genes to Cells*, 5(5), pp. 407–423. doi: 10.1046/j.1365-2443.2000.00336.x.

Weinmann, R. and Roeder, R. G. (1974) 'Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 71(5), pp. 1790–4. doi: 10.1073/PNAS.71.5.1790.

Wieland, T. *et al.* (1968) 'Poisonous principles of mushrooms of the genus Amanita. Four-carbon amines acting on the central nervous system and cell-destroying cyclic peptides are produced.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 159(3818), pp. 946–52. doi: 10.1126/science.159.3818.946.

Wiesner, S. *et al.* (2002) 'Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40', *Journal of Molecular Biology*, 324(4), pp. 807–822. doi: 10.1016/S0022-2836(02)01145-2.

Wu, H., Xiong, W. C. and Mei, L. (2010) 'To build a synapse: signaling pathways in neuromuscular junction assembly', *Development*, 137(7), pp. 1017–1033. doi: 10.1242/dev.038711.

Wu, Z. *et al.* (1991) 'Small nuclear ribonucleoproteins and heterogeneous nuclear ribonucleoproteins in the amphibian germinal vesicle: Loops spheres, and snurposomes', *Journal of Cell Biology*, 113(3), pp. 465–483. doi: 10.1083/jcb.113.3.465.

Xue, Y. *et al.* (2009) 'A capping-independent function of MePCE in stabilizing 7SK snRNA and facilitating the assembly of 7SK snRNP', *Nucleic Acids Research*, 38(2), pp. 360–369. doi: 10.1093/nar/gkp977.

Yamada, T. *et al.* (2006) 'P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation', *Molecular Cell*, 21(2), pp. 227–237. doi: 10.1016/j.molcel.2005.11.024.

Yamaguchi, Y. *et al.* (1999) 'NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation Cooperates with DSIF to Repress RNA Polymerase II Elongation', 97, pp. 41–51.

Yang, J. S. *et al.* (2011) 'OASIS: Online application for the survival analysis of lifespan assays performed in aging research', *PLoS ONE*. doi: 10.1371/journal.pone.0023525.

Yang, Z. *et al.* (2001) 'The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control tra...: EBSCOhost', 414(November). Available at: http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=8b344c04-04be-4b8c-bb7d-3bee57731211%40sessionmgr111&vid=2&hid=127.

Yang, Z. *et al.* (2005) 'Recruitment of P-TEFb for stimulation of transcriptional elongation by the

bromodomain protein Brd4', *Molecular Cell*, 19(4), pp. 535–545. doi: 10.1016/j.molcel.2005.06.029.

Yik, J. H. N. *et al.* (2003) 'Inhibition of P-TEFb (CDK9/cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA', *Molecular Cell*, 12(4), pp. 971–982. doi: 10.1016/S1097-2765(03)00388-5.

Yik, J. H. N. *et al.* (2005) 'Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive positive transcription elongation factor b complexes for control of transcription', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M500912200.

Yu, Y. and Reed, R. (2015) 'FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP', *Proceedings of the National Academy of Sciences*, 112(28), pp. 8608–8613. doi: 10.1111/tri.12405.

Zhang, Z. *et al.* (2016) 'Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis.', *Genes & development*. Cold Spring Harbor Laboratory Press, 30(18), pp. 2106–2118. doi: 10.1101/gad.285395.116.

Zhou, M. *et al.* (2001) 'TFIIH Inhibits CDK9 Phosphorylation during Human Immunodeficiency Virus Type I Transcription', *Journal of Biological Chemistry*, 276(48), pp. 44633–44640. doi: 10.1074/jbc.M107466200.

Zhou, M. *et al.* (2009) 'Bromodomain Protein Brd4 Regulates Human Immunodeficiency Virus Transcription through Phosphorylation of CDK9 at Threonine 29', *Journal of Virology*, 83(2), pp. 1036–1044. doi: 10.1128/JVI.01316-08.

Zhou, Q. *et al.* (2004) 'A Human Immunodeficiency Virus Type 1 Tat-Like Arginine-Rich RNA-Binding Domain Is Essential for HEXIM1 To Inhibit RNA Polymerase II Transcription through 7SK snRNA-Mediated Inactivation of P-TEFb', *Molecular and Cellular Biology*, 24(12), pp. 5094–5105. doi: 10.1128/mcb.24.12.5094-5105.2004.

Zhu, Y. *et al.* (1997) 'Transcription elongation factor P-TEFb is required for HIV-1 Tat transactivation in vitro', *Genes and Development*, 11(20), pp. 2622–2632. doi: 10.1101/gad.11.20.2622.

Zieve, G. and Penman, S. (1976) 'Small RNA species of the HeLa cell: metabolism and subcellular localization.', *Cell*, 8(1), pp. 19–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/954090 (Accessed: 19 February 2019).

Zylber, E. A. and Penman, S. (1971) 'Products of RNA polymerases in HeLa cell nuclei.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 68(11), pp. 2861–5. doi: 10.1073/PNAS.68.11.2861.

Kent, ENCODE DCC., Jim. n.d. "KentUtils." https://github.com/ucscGenomeBrowser/kent.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R

Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. http:

//ggplot2.org.

# 14. Acknowledgements

# 15.Curriculum Vitae