**OM SINGH RATHORE**

# Regulation of pre-mRNA splicing during *Drosophila* development

UAlg
UNIVERSIDADE DO ALGARVE
DEPARTAMENTO DE CIÊNCIAS BIOMÉDICAS E MEDICINA

**2018**

OM SINGH RATHORE

# Regulation of pre-mRNA splicing during *Drosophila* development

PhD Programme in Mechanisms of Disease and Regenerative Medicine

Supervisor:

Dr. Rui Goncalo Martinho

This dissertation was sponsored by Fundação para a Ciência e a Tecnologia, (PD/BD/52428/2013) within the scope of the ProRegeM PhD program (PD/00117/2012, CRM:0027030)

2018

# Regulation of pre-mRNA splicing during *Drosophila* development

**Declaração de autoria de trabalho**

Declaro ser o autor deste trabalho, que é inédito e original. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Om Singh Rathore

*To*

*My Grandparents*

(Nana-Nani)

*Almost all aspects of life are engineered at the molecular level, and without understanding molecules we can only have a very sketchy understanding of life itself.*

*-Francis Crick*

**Acknowledgments**

This thesis displays my name on the cover page, however, many have contributed to make it a success. I hereby dedicate this page with my great appreciation to everyone who had helped me during this incredible and really important journey that has been my PhD.

The first and for most I would like to express my sincere gratitude to my advisor, my guru, Dr. Rui Martinho for giving me the great opportunity to work with him and for the trust he granted me from the beginning. His knowledge of Biology and his enthusiasm for science is something that I will always admire and strive for in my own endeavors. I appreciate that he never giving up on me and pushing me through the hard times. I could not have imagined having a better supervisor and mentor for my Ph.D. studies, ascertaining him a commemorative plaque at deep of my heart. I greatly appreciate all his help.

I would like to thank to all the members of ProRegeM committee; Dr. Belo, Dr. Antonio, Dr. Bragança and my thesis advisory committee, Dr. Alisson and Dr. Alexandra for taking the time to follow this project and their helpful comments during annual meeting and evaluation.

I would like to thank the past and present members of my lab for their assistance and constant advices: Rui Silva, Ana (Tia Bakura), Margarida, Ricardo, Pedro, Paulo, Xana, Gaston and Bruno. You all had a tremendous impact on the success of my work. Rui Silva, thank you for all your support during my PhD and cheerful discussions.

I would like to thank Dr. Jean-Yves Roignant and the members of the Roignant research group for all the support, their kind assistance for experiments and scientific discussion, during my stay at IMB, Mainz, Germany.

I also appreciate all the help I've got from our collaborators Dr. Nuno Morais and Mariana Ferreira for their generous help in RNA-seq analysis, a great contribution for my thesis, thank you guys.

I would like to thank the entire CBMR members for being a supportive working environment. I would like to thank Dr. Raquel Andrade, and Dr. Álvaro Tavares particularly for their help, encouragement and conversations throughout the course. Dr. Raquel Andrade: Thanks for help and discussion about qPCR.

I am also grateful to the members of the core-facilities, in particular Claudia Florindo and Mauricia Vinhas for their kind help in microscopes and RNA analysis.

I'd like to thank all my friends in CBMR- Gil, Joao, Apolonia, Susana Rodrigues and Susana Machado, Goncalo, Tomas, Rui, Tinoni for their absolute moral support, I am very grateful to have them around me.

I have to thank ProRegeM family in particular the 1st edition crew with whom I shared some of the best moments of my life.

A minha familia: House of Horrors – Dino Mathias, Luis Casaca, Diogo Prata: You always there, to share good and bad time. Thank you brothers☺

Familia de BABAs (Vitor, Nitya, Honi, Amrit): Thank you very much for all your moral support that always motivates me for research.

A great thank to Vikram Sharma: My friend, my brother for being supportive, persistent and never losing the faith in my abilities.

I also thanks to Vaibhav, Rudra and Meenakshi for always being supportive. Himanshu and Shekhar (Nathu); guys you are awesome and thank you for all.

My Gwalior family: UKS uncle thank you very much to introduce me the "Research world " and your words of motivation, I always appreciate.

I would like to thank to Sushma and Lika for proofreading and useful comments; your help is highly appreciated.

I cannot finish my acknowledgements without expressing my gratitude to Prof. Dr. Hemant Malhotra and Dr. Bharti Malhotra for their supervision during my best time in Jaipur. During my three years with them, they taught me how to develop my own creativity, improve critical thinking and helped me to develop all the skills required for a diligent work.

Shipra madam: thank you very much for being always there; when I needed the most, I really appreciate.

Finally, I express a deep gratitude to my family for the immense support and cooperation. Thank you for always encouraging me to give my best in every situation.

A big thanks to all the above and all the others that I didn't mention and that somehow have been there in these last few years.

Bala-Bala: Thank you.

Muitos Obrigado ☺

**List of publications**

The work performed during my PhD contributed to following publications:

1. **Rathore OS**, Faustino A, Prudêncio P, Van Damme P, Cox CJ, Martinho RG. Absence of N-terminal acetyltransferase diversification during evolution of eukaryotic organisms. **Scientific Reports** 2016 Feb 10; 6:21304. Doi: 10.1038/srep21304. (2016)

2. Ribeiro AL, Silva RD, Foyn H, Tiago MN, **Rathore OS**, Arnesen T, Martinho RG. Naa50/San dependent N-terminal acetylation of Scc1 is potentially important for sister chromatid cohesion. **Scientific Reports**. 2016 Dec 20; 6: 39118. Doi: 10.1038/srep39118. (2016)

3. Silva RD, Mirkovic M, Guilgur LG, **Rathore OS**, Martinho RG, Oliveira RA. Absence of the Spindle Assembly Checkpoint restores mitotic fidelity upon loss of sister chromatid cohesion. **bioRxiv.** Doi**:** https://doi.org/10.1101/262204 (2018)

4. **Rathore OS**, Silva RD, Tiago MN, Prudêncio P, Roignant JY, Martinho RG. Salsa, a RNA helicase required for efficient splicing of weak splice transcripts during *Drosophila* development. (**In Preparation)**

This thesis is the result of the work I developed between January 2014 and December 2017 under the supervision of Dr. Rui Goncalo Martinho, Early Fly Development Laboratory at Center for Biomedical Research (CBMR), University of Algarve, Faro, Portugal.

The original manuscript 1, represent the side project developed during my first year of PhD and manuscript 2 and 3 are in collaboration with other lab members in different projects, supervised by Dr. Rui Gonaclo Martinho. The manuscript 4 (under preparation) corresponds to chapter II, where I performed most of the experiments described (exceptions are mentioned).

**Resumo**

O spliceosoma é uma máquina molecular extremamente dinâmica com variações constates na sua composição e conformação, sendo estas modificações cruciais para o splicing do pré-RNA mensageiro (pré-mRNA). A complexidade do spliceosoma contrasta com a simplicidade bioquímica da reação de splicing, que consiste essencialmente em dois ataques nucleofílicos. Esta complexidade deve-se ao facto de o spliceosoma ter de ser extremamente preciso no reconhecimento dos splice-sites e simultaneamente ter a plasticidade necessária para a realização de inúmeros eventos de splicing diferentes essenciais para a formação de diversidade transcricional. Vários estudos em levedura e células humanas demostraram que remoção de diferentes subunidades do spliceosoma afeta o splicing de diferentes subgrupos de intrões (Dix *et al*, 1999; Lygerou *et al*, 1999; Larson *et al*, 2016), sugerindo um elevado grau de plasticidade no splicing de diferentes tipos de intrões.

Para compreender melhor a contribuição desta plasticidade para a expressão génica durante o desenvolvimento de organismos multicelulares, realizámos um screen genético para identificar subunidades do spliceosoma cuja depleção por RNA de interferência (RNAi) está associada a fenótipos específicos durante a oogénese e/ou desenvolvimento embrionário. A nossa hipótese é que essas subunidades serão particularmente limitantes para o splicing de pequenos subgrupos de intrões. Com este screen nós identificamos Salsa, uma subunidade do complexo Nine Teen (NTC) do spliceosome, como sendo necessária para a formação de um gradiente dorso/ventral (D/V) no ovo de *Drosophila*.

A depleção de Salsa na linha germinal durante a oogénese originou um fenótipo de ventralização altamente penetrante (fenótipo de spindle) e reduziu drasticamente a fertilidade. Como a expressão de Gurken (um ligante da família TGFα) é crucial para a formação do gradiente D/V no oócito durante o desenvolvimento, decidimos investigar se Salsa era limitante para expressão de Gurken.

Observamos que o splicing do primeiro intrão do transcrito de gurken foi particularmente sensível à depleção de Salsa. Paralelamente verificamos que a localização dorso-anterior do mRNA de gurken e, consequentemente, da

proteína de Gurken, era altamente anormal após depleção de Salsa. Uma vez que foi previamente observado que o 5'UTR do mRNA de gurken é importante para sua localização dorso-anterior (Saunders & Cohen, 1999b), propomos que a retenção do primeiro intrão (dentro do 5'UTR do mRNA de gurken) interfere na sua conformação e consequentemente na formação de um motivo de reconhecimento de RNA importante para a sua correta localização.

Curiosamente, a análise do transcritoma de ovários de *Drosophila* após depleção de Salsa confirmou que esta helicase de RNA é particularmente limitante para o splicing de pequenos intrões próximos do início do transcrito; incluindo, por exemplo, o intrão localizado no 5'UTR do mRNA de tra2, um gene cujo splicing alternativo é crucial para a determinação do sexo.

**Summary**

The spliceosome is a very dynamic molecular machine, whose compositional and conformational remodeling is crucial for pre-mRNA splicing. Although splicing is biochemically simple, with essentially two nucleophilic attacks, the spliceosome is nevertheless remarkably complex, as it needs to be in one hand extremely precise in splice-site recognition, but in the other hand must accommodate an array of alternative splicing events capable of generating transcript diversity. Multiple studies in yeast and human cells have shown that loss of distinct subunits of the spliceosome impaired splicing of distinct subsets of introns (Dix *et al*, 1999; Lygerou *et al*, 1999; Larson *et al*, 2016), suggesting a significant degree of splicing plasticity among different classes of introns.

To better understand the contribution of such splicing plasticity to differential gene expression during development of multicellular organisms, we performed a screen to identify spliceosome subunits whose depletion by RNA interference (RNAi) was associated to specific phenotypes during oogenesis and/or early embryonic development. Our working hypothesis is that those subunits would be particularly rate limiting for splicing of small subsets of introns. We identified Salsa, a subunit of the spliceosome Nine Teen Complex (NTC), as being required for dorsal ventral (D/V) patterning of the Drosophila egg.

Germ-line specific depletion of Salsa during oogenesis produced a highly penetrant ventralization phenotype (spindle phenotype) and dramatically reduced fertility. Since Gurken expression (a TGFα family signaling ligand) is crucial for D/V patterning of the developing oocyte, we decided to investigate if Salsa was rate limiting for Gurken expression. We observed that splicing of the first intron of gurken transcript was particularly sensitive to Salsa depletion, whereas anterior dorsal localization of gurken mRNA, and consequently Gurken protein, was highly abnormal. Since it was previously suggested that the 5´UTR of gurken is important for its anterior dorsal localization (Saunders & Cohen, 1999a), our current working model is that retention of the proximal intron, within the 5´UTR of gurken transcript, disturbs the folding of an unknown RNA motif important for its correct localization.

Interestingly, analysis of the *Drosophila* ovaries transcriptome after depletion of Salsa further confirmed that this RNA helicase is particularly rate limiting for splicing of small proximal introns; including for example the proximal 5´UTR-localized intron of tra2, a gene whose alternative splicing is crucial for sex-determination.

**Abbreviations**

ATP: adenosine tri-phosphate

BCA: bicinchoninic acid

bp: base pair

BPS: branchpoint sequence

cDNA: complementary DNA

CTD: C-terminal Domain

DNA: Deoxyribonucleic acid

DTT: Dithiothreitol

ECL: enhanced chemiluminiscence

eIF4E; eukaryotic Initiation Factor 4E

EJC: Exon Junction Complex

EtOH: Ethanol

g : Gram/centrifugal force

h : hour/human

kb : kilo bases

kDa : Kilodalton

mRNA: messenger RNA

NCBI: National Center of Biotechnology Information

NTC: nineteen complexes

ORF: open reading frame

PBS Phosphate Buffered Saline

PCR:polymerase chain reaction

PFA: Paraformaldehyde

pH: preponderance of hydrogen

PIR: Percentage of intron retention

piRNAs: Piwi-interacting RNAs

Pre-mRNA: precursor-mRNA

RNA: ribonucleic acid

RNAi: RNA interference

RNAPII: RNA polymerase II

RNase: Ribonuclease

RNP: Ribonucleoprotein

RPM: revolutions per minute

RT:  room temperature

SDS: Sodium-Dodecyl-Sulfate

SS: Splice site

TE: Tris-EDTA

Tm: melting temperature

TREX: transcription and export

Tris: Tris-(hydroxymethyl)- aminomethane

UTR: untranslated region

UV: Ultraviolet

WT: wild-type

µ: micro

DAPI: 4',6-diamidino-2-phenylindole

°C : degree celcius

# Table of Contents

# 1

# General introduction

## Contents

## 1. Introductory note:

The identity of the different cells of an organism is given by the expression of distinctive sets genes. Therefore, differential regulation of gene expression is essential for establishing the fate of a cell and their function. Regulation of gene expression is composed by different layer of regulatory networks that ensure coordinated expression of required groups of genes. Defining gene regulatory networks and obtaining insights into their relationships with each other is essential for understanding any developmental program.

*Drosophila* oogenesis represents an invaluable experimental system amenable to investigate the complex process of gene regulation in a complex process of development. In next sections, I describe the process of *Drosophila* oogenesis and how different genes and regulatory networks contribute to patterning of the developing egg. Besides, I also provide a general introduction about the different layers of gene expression and regulation, where I mostly focus on splicing

# 1.1

# *Drosophila* oogenesis and axis specification

## Contents

### 1.1.1 *Drosophila* oogenesis

The adult female *Drosophila* contains a pair of ovaries and each ovary is composed of 16–20 developmentally ordered egg chambers called ovarioles (King, 1970; Spradling and Mahowald, 1979). Ovarioles are the functional units of egg production that support the development of a single oocyte/egg. The development of egg chambers can be divided into 14 developmental stages: early stage (1-6), mid stage (7-10), and late stage (11-14) of oogenesis. The anterior compartment of the ovariole is called germarium that contains several germline and somatic stem cells. The germarium is subdivided into four regions according to the developmental stage of the cyst- region1, 2a, 2b and 3 (Figure- 1.1.1).

Oogenesis begins in region 1, where the germline stem cell undergoes an asymmetric division; produces daughter stem cells to self- renew and a cystoblast. The cystoblast later undergoes four mitotic divisions, without completion of cytokinesis so that all-16 cells remain connected via cytoplasmic bridges, called ring canals (King, 1970; Mahowald and Strassheim, 1970). The newly formed 16-cell cysts are located in the region 2a of the germarium and all the cells cysts are similar in the size. By region 2b of the germanium, the cell cysts become lens shaped and one of the two cells with four ring canals becomes pro-oocyte. The remaining fifteen cells become the nurse cells and they do not divide again, but grow and undergo polytenization. The nurse cells transports- mRNAs, proteins, and endomembrane structures (e.g., ER and Golgi) via the ring canals, required for growth and maturation of the future oocyte. In region 3 (also called stage 1) of the germarium, the oocyte is already located at the posterior pole and somatic follicle cells migrate and surround the cyst to form an egg chamber. The developing egg chambers are connected by stalk cells, which are specialized inter-follicular cells. The stalk formation participates in establishing the position of the oocyte and axis determination of the egg chamber.

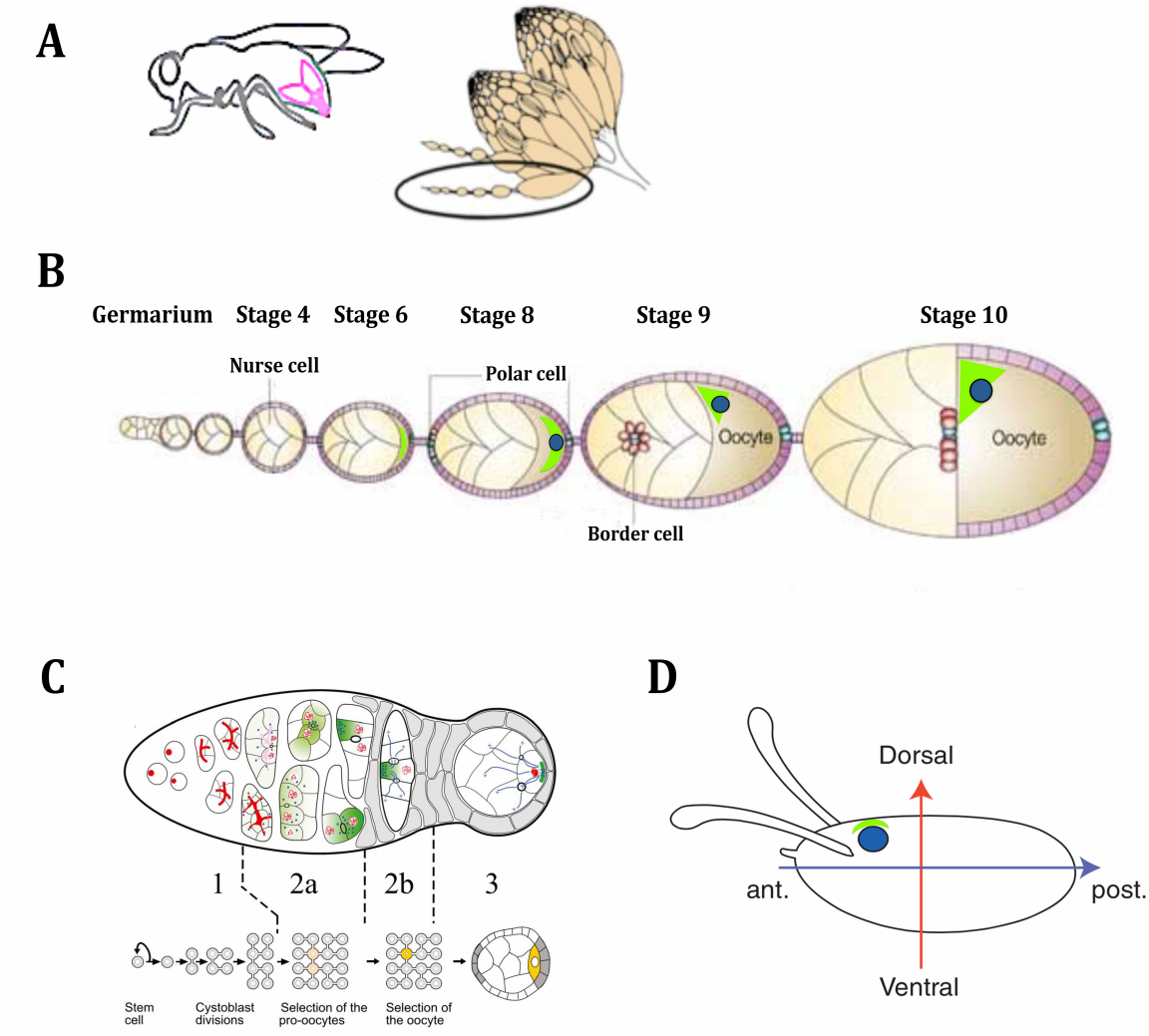



**Figure 1.1.1: *Drosophila* egg chamber development and axis determination of egg:** (A) A diagram of the *Drosophila* and pair of ovary. Each ovary contains multiple ovarioles that contains the developing egg chambers. (B) Structure and development of *Drosophila* ovarian follicles. A drawing of a *Drosophila* ovariole, showing the anterior germarium followed by a string of successively older ovarian follicles connected by inter-follicular stalks. The oocyte nucleus (blue) and *gurken* RNA (green) move from the posterior of the oocyte to an antero-dorsal position during mid-oogenesis. Gurken signals to the overlying follicle cells to determine the position of the future dorsal side of the egg. (C) Detailed scheme of the germarium structure: The germarium composed of three regions; in region 1, a single GSC division and cystoblast division that give rise to 16 interconnected cells. In region 2, several cells per cyst initiate the assembly of synaptic chromosomes and approximately two somatic stem cells (SSCs) become pro-oocytes with four ring canals. By region 2b, the oocyte has been selected and is the only cell to remain in meiosis. The somatic stem cells produce follicle cells and start to migrate and surround the germline cells. Region 3 contains

the stage 1 egg chamber, consisting of follicle cells, 15 germline nurse cells and an oocyte at the posterior end. (D) Dorso-ventral axis presentation of mature eggs; the gurken signaling and oocyte nucleus acquire an asymmetrical position, which determines the dorsal side of the egg and establishes orthogonality between the Anterior-posterior (AP) and the Dorso-ventral (DV) axes. Figure representation adapted from: (Jagut et al., 2013; Montell, 2003).

#### 1.1.1.1 Oocyte determination

The process of oocyte determination is very complex process, defined by the differential accumulation of specific RNAs and factors such as Oskar (osk) and Bicaudal-D (BicD), Egalitarian (Egl), Oo18 RNA-binding (Orb) and Cup proteins (Ephrussi et al., 1991; Keyes and Spradling, 1997; Kim-Ha et al., 1991; Lantz et al., 2010; Mach and Lehmann, 1997; Suter and Steward, 1991a; Wharton and Struhl, 1989). Bic-D and Egl play an important role in the oocyte-determining processes. The mutants for *BicD* and *egl*, fail to localize the RNAs and proteins crucial to define the oocyte (Mach and Lehmann, 1997; Mohler and Wieschaus, 1986; Schüpbach and Wieschaus, 1991; Suter and Steward, 1991b; Theurkauf et al., 1993). Consequently, *Bic-D* and *egl* mutants produce cysts lacking an oocyte and form egg chambers with sixteen polyploid nurse cells (Mohler and Wieschaus, 1986; Schüpbach and Wieschaus, 1991).

The microtubules (MTs) plays crucial role in the transport of different mRNAs and proteins to future oocyte during early oogenesis (Theurkauf, 1994; Theurkauf et al., 1993). MTs are the largest diameter cytoskeletal filaments and polarization emerges gradually after the last cystocyte division. The minus ends of MTs that grow slowly are initially enriched in both two pro-oocytes but eventually concentrate within one of them (Desai and Mitchison, 1997). The inherent polarity of the MT allows the selective transport to one end by unidirectional MT motor proteins. Dynein proteins plays crucial role for stable MT network, whereas BicD and Egl are cytoplasmic proteins that associate different cargo to the dynein/dynactin motor (Bullock and Ish-Horowicz, 2001; Mach and Lehmann, 1997). BicD and Egl have also been shown to recruit the dynein motor complex to mRNAs and initiate minus end-directed mRNA transport in other developmental contexts (Bullock and Ish-

Horowicz, 2001). Orb is a RNA-binding protein that facilitates translation of localized RNAs (Derrick and Weil, 2017). These observations provide evidence that a polarized microtubule network is necessary for the transport of various factors to the future oocyte and its eventual determination.

#### 1.1.1.2 Microtubule organization during *Drosophila* oogenesis

The large microtubule network of the *Drosophila* oocyte is highly dynamic throughout oogenesis and requires precise coordination for transport and localization of mRNA toward the oocyte (Mohler and Wieschaus, 1986; Schüpbach and Wieschaus, 1991; Suter et al., 1989; Theurkauf et al., 1993; Tomancak et al., 1998b). The oocyte nucleus changes its position several times during the oocyte development and this is accompanied by a reorganization of the MT cytoskeleton (Steinhauer and Kalderon, 2006; Theurkauf et al., 1992). Early in oogenesis, in region 2b in the germarium, a clearly defined anterior microtubule-organizing center (MTOC) is first observed in the oocyte and is maintained through stage 1. During stages 2-6 of oogenesis, microtubules extend from a MTOC positioned at the posterior pole of the oocyte and pass through the ring canals adjoining the oocyte and the neighboring nurse cells. (Theurkauf et al., 1993). Additionally, some microtubules are also observed in the nurse cells. These microtubules in the nurse cells do not appear to originate from the oocyte and do not nucleate from well-defined MTOCs. During stage 7 and early 8, the posterior MTOC degenerates as a result of a Gurken (Grk)-dependent signaling event and the majority of the microtubules reorganize at the anterior of the oocyte (Theurkauf et al., 1992). Microtubules are oriented in such a way that the minus-ends, the nucleation site of microtubules, are proximal to the oocyte cortex. The plus-ends of microtubules, the growing ends, are oriented towards the center of the oocyte. There is an evident gradient of microtubules along the cortex of the oocyte, with the highest abundance towards the anterior and with a decline in density towards the posterior pole. This anterior to posterior gradient in microtubule density is maintained throughout stages 8-10a. During stage 10b, the next major microtubule reorganization event takes place. The nurse cells dump their cytoplasmic content into the oocyte (fast nurse cells

dumping), which is concomitant with oocyte subcortical bundling of the microtubules and cytoplasmic streaming (Theurkauf et al., 1992) (Ferreira et al., 2014).

### 1.1.2 Axis specification

Axis specification of the *Drosophila melanogaster* egg is initially established during oogenesis by the asymmetric positioning of morphogenetic factors, mRNAs and proteins, within the oocyte (Lasko, 1999; St Johnston and Nüsslein-Volhard, 1992). Through much of oogenesis, the oocyte is mostly transcriptionally inactive, being the majority of the oocyte mRNAs and proteins, synthesized in the nurse cells and transported into the oocyte through the ring canals (Mahajan-Miklos and Cooley, 1994). Three major mRNAs oskar, bicoid, and gurken are localized to distinct compartments of the oocyte, before fertilization, and they determine the future axis of the egg (Riechmann and Ephrussi, 2001; van Eeden and St Johnston, 1999). Anterior-posterior (AP) patterning of the future egg is determined by the localization of the bicoid and oskar transcripts at the anterior and posterior poles of the developing oocyte, respectively. Whereas accumulation of gurken transcript at the anterior-dorsal region of the oocyte during mid-oogenesis is important for dorso-ventral (DV) patterning of the egg.

#### 1.1.2.1 Oskar

Oskar (osk) is one of the most important gene for germ-plasm assembly, which is required for both primordial germ cell formation and axis formation of *Drosophila* egg (Ephrussi et al., 1991; Ephrussi et al., 1991; Kim-Ha et al., 1991; Riechmann and Ephrussi, 2001). *osk* mRNA, like many other mRNAs, is transcribed in the nurse cells and transported to oocyte during early/mid oogenesis (Ephrussi et al., 1991; Kim-Ha et al., 1991; LASKO, 1999). A highly conserved pathway that involves two interacting proteins, Bicaudal-D (Bic-D) and Egalitarian (Egl), is required for *osk* mRNA transport and its localization to the posterior pole of oocyte (Clark et al., 1994; Riechmann and Ephrussi, 2001). The osk mRNA transport to the posterior cortical region is dependent on Gurken signaling and involves minus-end directed movement along the

microtubules (González-Reyes et al., 1995; Roth et al., 1995) and several rearrangements in microtubule cytoskeleton (Clark et al., 1994; Pokrywka and Stephenson, 1995).

The exon junction complex (EJC) components *mago nashi* (*mago*), *Y14* (also called *tsunagi*) and *eIF4AIII* have been implicated in posterior localization of *oskar* mRNA (Hachet and Ephrussi, 2001). The EJC complex is required for efficient splicing of the first intron of *osk* pre-mRNA, nuclear export, and mRNAs nonsense-mediated decay during oocyte development (St Johnston, 2005). The 3′ UTR of *osk* mRNA also contains sequence elements required for distinct steps in the *osk* mRNA localization process. The region between 532-791 nucleotides of the *osk* 3′ UTR is necessary for the transport of *osk* mRNA from nurse cells into oocyte, whereas the region between 1-242 nucleotides is required for its oocyte posterior localization (Kim-Ha et al., 1993). The other additional proteins that promote posterior localization of *osk* mRNA are putative RNA binding proteins, Ovarian tumor (Otu) and Staufen (Stau). Otu and Hrp48 (Heterogeneous nuclear ribo-nucleoprotein) associate with 3′ UTR of *osk* mRNA and appear to have a direct effect on *osk* mRNA localization, as the polarity of the microtubules cytoskeleton appears normal in these mutants (McBeath et al., 2004). Stau co-localizes with *osk* mRNA and regulates the microtubule-dependent localization of *osk* RNA to the posterior pole of developing oocyte (Ephrussi et al., 1991; Kim-Ha et al., 1991).

The exclusive localization of *osk* mRNA to the oocyte posterior pole happens by stage 9, when it is first translated (Ephrussi et al., 1991; Kim-Ha et al., 1991). The regulation of *osk* translation depends on the various regulatory elements present on both of the 3′ and 5′ UTRs. The 3′ UTR contains elements regulating the repression of translation (Kim-Ha et al., 1993; Kim-Ha et al., 1995). Bruno, encoded by the *arrest* gene, co-localizes with *osk* mRNA at the posterior pole and avoids premature translation of *osk* mRNA (Kim-Ha et al., 1995; Webster et al., 1997). Bicaudal-C (Bic-C) is an RNA binding protein that does not binds directly to *osk* mRNA but is implicated on its translational repression. In *Bic-C* mutants, *osk* mRNA is dispersed in the oocyte and is ectopically translated (Saffman et al., 1998). Various other proteins have also been implicated in *osk* mRNA translation, including Osk

itself (Markussen et al., 1995; Markussen et al., 1997), Stau (Kim-Ha et al., 1995; St Johnston and Nüsslein-Volhard, 1992; St Johnston et al., 1991), the DEAD-box RNA helicase Vasa, and Aubergine (Kim-Ha et al., 1995; Markussen et al., 1995; Rongo et al., 1995; Wilson et al., 1996). The localization of *osk* mRNA to the posterior pole and its translational regulation ensures the restriction of Osk protein to the oocyte posterior pole and the correct establishment of the anterior-posterior axis of developing oocyte.

#### 1.1.2.2 Bicoid

Bicoid (bcd) mRNA, the primary anterior defining morphogen, is synthesized in the nurse cells and transported through the ring canals to the oocyte anterior cortex (Berleth et al., 1988; Driever and Nüsslein-Volhard, 1988a; St Johnston et al., 1989). The transport of bcd mRNA into the oocyte and anchoring to the anterior cortex is dependent upon exuperantia (exu), swallow (sww), staufen (stau) and an intact microtubule network (Macdonald et al., 1991; Pokrywka and Stephenson, 1991; Pokrywka and Stephenson, 1995). Additionally, the 3'UTR of bcd mRNA has been identified as an essential cis-acting element for the proper localization of bcd mRNA (Macdonald and Struhl, 1988; MacDonald et al., 1993). Localization of bcd mRNA to the anterior region of the oocyte is comprised of three distinct steps. Initially, bcd mRNA associates with Exu protein in a microtubule dependent process, and the Exu-bcd mRNA complex is then transported by microtubules into the oocyte. Exu is not necessary for transport within the nurse cells, but is critical for anterior localization upon entering the oocyte (Cha et al., 2001; Macdonald and Struhl, 1988). A 53 base-pair element identified within the bcd 3'UTR, the bicoid localization element (BLE1), is required for the Exu-dependent anterior localization (Macdonald and Struhl, 1988; MacDonald et al., 1993). Once localized, Swallow is responsible for maintenance of bcd mRNA in the anterior region of the oocyte, by cortical anchoring of bcd mRNA at later stages of oogenesis (Stephenson et al., 1988). Lastly, Staufen protein has been shown to associate with the 3'UTR of bcd, including the BLE1, to prevent diffusion of the RNA once it is released into the egg cytoplasm during egg activation (Ferrandon et al., 1994). Following fertilization, bcd mRNA is translated to

produce a gradient of Bicoid (Bcd) protein extending over the anterior half of the embryo (Berleth et al., 1988; Driever and Nüsslein-Volhard, 1988b; St Johnston and Nüsslein-Volhard, 1992). Bcd protein guides the expression of zygotically expressed segmentation genes that establish anterior patterning in the developing egg (St Johnston and Nüsslein-Volhard, 1992).

#### 1.1.2.3 Gurken and dorso-ventral patterning

Gurken (grk), a member of the transforming growth factor α (TGFα) family, also transcribed in the nurse cells during oogenesis and plays a crucial role for the establishment of future APand DV axes of *Drosophila* egg (González-Reyes et al., 1995; Neuman-Silberberg and Schupbach, 1994; Ray and Schupbach, 1996; Roth et al., 1995). grk mRNA is transcribed and packaged into RNPs (ribonucleoproteins) in the nurse cells, support cells that are interconnected with each other and the developing oocyte. During early oogenesis, a Gurken-dependent signaling assigns posterior fate to a sub-population of follicle cells and inducing a major rearrangement of the oocyte microtubule cytoskeleton (González-Reyes et al., 1995; Roth et al., 1995). This reorganization of microtubules leads to the repositioning of the oocyte nucleus, which is essential for AP and DV polarity (Neuman-Silberberg and Schupbach, 1994; Price et al., 1989; Spradling, 1993). During mid-oogenesis, the oocyte nucleus, and concurrently grk mRNAs, are actively transported to the future anterior-dorsal region of the oocyte, where subsequent localized translation of Grk activates the *Drosophila* EGF receptor in the adjacent follicle cells, specifying the anterior-dorsal fate of these cells (González-Reyes et al., 1995; Januschke and Gonzalez, 2008; Neuman-Silberberg and Schupbach, 1994).

##### 1.1.2.3.1 Localization of gurken mRNA

grk mRNA, though transcribed in the nurse cells, localizes to the peri-nuclear region of the oocyte cytoplasm throughout of oogenesis. In early oogenesis, the oocyte nucleus is located at the posterior pole and grk mRNA also accumulates to posterior pole. Later, when the oocyte nucleus moves to anterior-dorsal position, grk mRNA forms a crescent between the apical

surface of the nucleus and the neighboring region of the cortex. Since grk mRNA is transcribed in nurse cells, it must be transported from the nurse cells into the oocyte and then distributed correctly. The initial transport of grk mRNA from the nurse cells to the oocyte requires, similarly to bcd mRNA, the Bic-D/Egl complex pathway. grk mRNA first accumulates along the anterior cortex, then it is transported laterally toward the oocyte nucleus (Jaramillo et al.; MacDougall et al., 2003). The second phase of grk mRNA localization within the oocyte also depends on dynein and the microtubule cytoskeleton, and the oocyte nucleus appears to nucleate a distinct population of microtubules, which are thought to mediate lateral displacement of oocyte nucleus (Januschke and Gonzalez, 2008; MacDougall et al., 2003). Various proteins and factors including Squid (Sqd), Hrp48, Otu, poly (A) binding protein (PABP) and IGF-II mRNA-binding protein (Imp) are also implicated in grk mRNA localization and translation. Sqd, a RNA binding protein with different protein isoforms, has multiple roles on grk mRNA localization and translational control (Johnstone and Lasko, 2001). Imp associates with Sqd and Hrp48, and interacts strongly with the grk mRNA 5'UTR (Saunders and Cohen, 1999a), suggesting a possible role of grk mRNA 5'UTR in localization and translation. Half pint (Hfp), an RNA binding protein, regulates pre-mRNA splicing of grk and is similarly required for localization during oogenesis (Van Buskirk and Schüpbach, 2002a). Studies of injected fluorescently tagged grk mRNAs implicated an element within the protein-coding region, termed the grk localization signal (GLS), as essential for both oocyte targeting and anterior-dorsal localization (Van De Bor et al., 2005). However, analysis of modified grk transgenes clearly indicated that the GLS is not sufficient for anterior-dorsal accumulation of grk mRNA, and another RNA element is likely to be involved during the process (Lan et al., 2010). Nuclear export of grk requires the RNA helicase UAP56, and hypomorphic mutation alleles of *uap56* impair efficient anterior-dorsal accumulation of grk mRNA. Fluorescently labeled grk mRNA injected into *uap56* mutant oocytes is localized only inefficiently, further supporting a role of UAP56 in cytoplasmic mRNA transport and grk localization (Cohen and Greenberg, 2008; Driever et al., 1990).

#### 1.1.2.3.*2 gurken* translation

Localized RNAs are usually translationally silent during transport, until they reach at their prescribed destination within the cell, or here in the developing oocyte. Translational repression during mRNA transport is imperative, as premature or ectopic translation of RNAs leads to developmental defects. Consistently, translation of grk mRNA only occurs once it is correctly localized at the posterior and anterior-dorsal region of the oocytes (Driever et al., 1990; Ephrussi and Lehmann, 1992; Gavis and Lehmann, 1992; Kim-Ha et al., 1995; Smith et al., 1992). Various hnRNP proteins have been implicated, not only in grk mRNA localization but also on its translational regulation. Depletion of *sqd* or *K10,* resulted in a mis-localized and precociously translated grk mRNAs. The strong dorso-ventralization phenotype of the oocyte results from the abnormal induction of dorsal follicle cell fates by the ectopically produced Grk (Johnstone and Lasko, 2001). During the transport of grk mRNA, Bruno (Bru) directly bind to the grk 3'UTR and significantly repress the grk translation. Furthermore, grk has an additional mode of translational repression through the action of Cup, the *Drosophila* homolog of the mammalian eukaryotic initiation factor eIF4E binding protein, 4E-transporter (4E-T) and functional homolog of *Xenopus* Maskin (Cao and Richter, 2002; Kamenska et al., 2014; Nakamura et al., 2004; Nelson et al., 2004; Richter and Sonenberg, 2005; Stebbins-Boaz et al., 1999). Grk expression in the oocyte is also dependent in the regulation of the polyadenylation status of grk mRNA (Norvell et al., 2015). In *orb* mutants, grk mRNA polyadenylation is significantly reduced, indicating that this RNA-binding protein enhances grk mRNA translation by increasing its polyadenylation status (Chang et al., 2001).

### 1.1.2.4 DNA damage and ventralization of egg

During *Drosophila* oocyte development, one of the earliest events is the formation of programmed double strand DNA breaks (DSBs) throughout the genome (Baudat et al., 2013). This process initiates meiotic recombination, facilitates homologous chromosome synapsis, and help to create new genetic combinations from parental genomes. During meiosis, programmed DNA

DSBs are generated by a highly evolutionarily conserved enzyme called sporulation-specific 11 (SPO11); which in *Drosophila* is encoded by the gene meiotic-W68 (mei-W68) (Baudat et al., 2000; Ghabrial et al., 1998; Keeney, 2008; Keyes and Spradling, 1997; McKim and Hayashi-Hagihara, 1998; Romanienko and Camerini-Otero, 2000). The response to DSBs prior to repair can be categorized into three events: first, sensing of the DSB, secondly, activation of Ataxia telangiectasia-mutated (ATM) and ataxia telangiectasia-related (ATR/MEI-41) kinases, and third is phosphorylation of several of their targets, including checkpoint-1 (Chk1/Grapes), checkpoint-2 (Chk2/MNK), and histone H2A variant (H2AX). Phosphorylation of H2AX (In *Drosophila* known as H2Av) at DSBs is an evolutionarily conserved response, which promotes the recruitment of repair factors and remodeling the local chromatin architecture (Madigan et al., 2002). Phosphorylated histone H2Av has been used to quantify the presence of DSBs and their resolution during meiotic progression (Joyce et al., 2011).

An accumulation of unrepaired DSBs during meiosis severely impairs oocyte development, and in *Drosophila* it is associated with an abnormal expression of Gurken (Grk) and DV patterning defects of the egg. Being Grk a dorsal-determinant, low level of its expression is associated to egg ventralization and fusion (or disappearance) of the egg dorsal appendages (known as the egg spindle phenotype). Defects in DSBs homologous recombination repair and defects in the Piwi-interacting RNA (piRNA) pathway are similarly associated with an abnormal expression of Gurken, implicating two seemingly unrelated pathways in the localization and translation of gurken mRNA. The spindle-class of genes (*spn-A*, *spn-B*, *spn-C*, *spn-D*, *spn-E* and *okra*) encodes homologues of yeast members of the Rad52 DSBs DNA repair and homologous recombination pathway, which are involved in the repair of DSBs (Abdu, 2006; Ghabrial et al., 1998; Joyce et al., 2011; McCaffrey et al., 2006). Mutation in any of these genes causes a failure to correctly repair the DSBs during recombination, which leads to an abnormal activation of the DNA damage meiotic checkpoint (meiotic checkpoint) and delay in cell cycle progression.

The prolonged activation of the meiotic checkpoint leads to reduction in the levels of Gurken protein, and consequently to defects in DV patterning of the egg (Abdu, 2006; Ghabrial et al., 1998; Joyce et al., 2011; McCaffrey et al., 2006). This phenotype is the result of meiotic checkpoint-dependent inhibition of Vasa phosphorylation, a RNA helicase with high homology to the general translational initiation factor eIF4A and crucial for gurken translation (Klovstad et al., 2008; Lasko and Ashburner, 1988; Styhler et al., 1998; Tinker et al., 1998; Tomancak et al., 1998a). Interestingly, mutations in mei-41 and mnk (which encode, respectively, *Drosophila* ATR and Chk2) suppress the DV patterning defects associated with the accumulation of DSBs, but not the DV defects seen in vasa mutant ovaries (Abdu et al., 2002; Ghabrial and Schüpbach, 1999; Ghabrial et al., 1998).

A prolonged activation of the meiotic checkpoint, due to the accumulation of unrepaired DSBs, is similarly associated with abnormal organization of the oocyte nucleus; which is usually organized during most diplotene I-arrest (prophase I but after the repair of the meiotic DSBs) in a highly compacted chromatin structure known as the karyosome (Ghabrial and Schüpbach, 1999; Lin, 2007; Mach and Lehmann, 1997). In *Drosophila*, highly conserved Histone 2A kinase, called NHK-1 (Nucleosomal Histone Kinase-1), plays a critical role in the formation and maintenance of the karyosome (Ivanovska et al., 2005). NHK-1 phosphorylates Barrier-to-Autointegration Factor (BAF) (Gorjánácz et al., 2007; Lancaster et al., 2007; Nikalayevich and Ohkura) and a mutation in BAF phosphorylation site leads to continuous association of DNA with the inner nuclear membrane and loss of karyosome formation (Lancaster et al., 2007). Activation the meiotic checkpoint inhibits NHK-1 activity, preventing the reorganization of the oocyte nucleus, including karyosome formation, until DSBs are correctly repaired (Lancaster et al., 2010).

In addition, several mutants for the Piwi-interacting RNA (piRNA) pathway also exhibit ventralized eggshell phenotypes, including armitage, maelstrom, aubergine, zucchini, squash, and cutoff (Barckmann et al., 2015; Cook, 2004; Findley, 2003; Handler et al., 2013; Pane et al., 2007; Teixeira et al., 2017). piRNA pathway represses the mobility of retro-transposons and the

accumulation of DNA damage within the *Drosophila* germ line (Handler et al., 2013; Mani and Juliano, 2013). Depletion of various spliceosome complex including EJC (Exon-Junction complex) also modulate piwi pathway and the transposon mobility thought splicing regulation during *drosophila* development (Malone et al., 2014a). Retro-transposons are a class of mobile elements that mediate transposition by reverse transcription of an RNA intermediate (O'Donnell and Boeke, 2007). Inactivation of the piRNA pathway leads to a failure in retro-transposon silencing, their retro-transposition within the germ line, and subsequent accumulation of DSBs.

# 1.2

# Gene expression and splicing

## Contents

### 1.2. 1 Introduction

Eukaryotic organisms contains their hereditary information encoded in molecules called deoxyribonucleic acid (D\NA), which are packed and organized in a structure called chromosomes into the nucleus of the cell. The DNA monomers are called nucleotides and are organized in a double-stranded helix (Watson and Crick, 1953; Wikins et al., 1953). Each nucleotide constituted by a phosphatase group, a deoxyribose, and a nitrogenous base called nucleobase. The genetic information in a DNA molecule is represented by the sequence of nucleotides containing one of four types of nucleobases: adenine (A), guanine (G), cytosine (C) and thymine (T). Following the Watson-Crick model, the two strands that constitute the DNA molecule are held together by hydrogen bonds that can only be established between specific pairs of nucleobases: A with T and G with C. Because of this restriction, both strains are complementary to one another and, therefore, contain the same genetic information.

Over the course of embryonic development, a fertilized egg cell gives rise to the entire organism with different cell types. However, each cell contains the genetic information that is almost an exact copy of the DNA that was in the fertilized egg cell. The distinct cell phenotypes are possible because different cell types make use of different stretches of the DNA molecule, called genes, to serve as templates to build functional cellular products in a process designated by gene expression

### 1.2. 2 Regulation of gene expression

Gene expression is a fundamental multi-step process where DNA is transcribed into RNA, which is then translated into protein (Jones, 2015; Moore, 2005). Gene expression is regulated at several levels from transcription to post-translation modifications to properly exert its biological functions. Dysregulation of gene expression at any level can lead to many diseases and syndromes including cancer, autoimmunity, neurological disorders, developmental syndromes, diabetes, cardiovascular diseases, and many more (Lee and Young, 2013). Transcription is one the first step of gene expression and is carried out in the nucleus (Nechaev and Adelman, 2011), under a highly structured organization (chromatin, distinct nuclear bodies,

etc.) and controlled by a dynamic group of proteins. RNA polymerase II has a significant role in mRNA processing and its largest subunit, the C-terminal domain (CTD), plays important roles at all the steps of transcription, including enhancing or modulating the efficiency of the RNA processing reactions (Armache et al., 2005). In Human, the CTD is composed of 52-conserved heptad repeats with the consensus sequence (Tyr-Ser-Pro-Thr-Ser-Pro-Ser) and provides multiple possible sites for phosphorylation. During the transcription cycle, the serine and tyrosine residues of each repeat undergo sequential phosphorylation (Egloff and Murphy, 2008), and these phosphorylation marks are critical for proper coordination of the transcription progression (Komarnitsky et al., 2000; Peterlin and Price, 2006). Through different patterns of phosphorylation and other modifications of CTD, the RNA polymerase II is processed to mature RNA through various co- and post-transcription modifications including the 5' end capping, 3' end processing and pre-mRNA splicing (Buratowski, 2009; Perales and Bentley, 2009; Phatnani and Greenleaf, 2006). The following section will describe these events in more detail.

#### 1.2.2.1 5' end capping of mRNA

5′ Capping is the first event of mRNA processing occurring during the transcription, and is required for efficient gene expression and cell viability. Formation of 5′ Capping occurs as soon as first 25-30 nucleotides of nascent pre-mRNA are transcribed by RNA polymerase II (Brody and Abelson, 1985; Coppola and Luse, 1984). Three sequential enzymatic steps are required for the 5′ Capping of pre-mRNA; 1) removal of the 5′ terminal γ-phosphate by RNA triphosphatase activity (TPase); 2) transfer of a GMP group from GTP to form diphosphate 5′ terminus by RNA guanylyltransferase activity (GTase) and 3) a methyltransferase modifies the N7 amine of guanosine by adding a methyl group by methyltransferase activity (MTase) (Ramanathan et al., 2016). Marking the 5' end of a pre-mRNA transcript with a 7-methyguanosine (m7G) cap helps to distinguish protein-coding mRNA from other types of RNA synthesized by RNA polymerase I or III (Houseley and Tollervey, 2009; Merrick et al., 2004; Meyer et al., 2004). The RNA cap structure at the 5′ end

of pre-mRNA is associated with the nuclear cap-binding complex (CBC), consisting of two proteins called CBP20 and CBP80 (Gonatopoulos-Pournatzis and Cowling, 2014; Izaurralde et al., 1994). These proteins rapidly recruited to this structure during pre-mRNA synthesis (Lahudkar et al., 2011; Wong et al., 2007). The CBC complex is the first protein complex that associates with nascent pre-mRNA and recruits several transcription factors directly and indirectly to promote the transcription elongation (Lahudkar et al., 2011; Proudfoot, 2011; Wong et al., 2007). During the first round of translation, the nuclear CBC is replaced by a different cap-binding protein complex, called eIF4F (Eukaryotic initiation factor 4F). eIF4F consists of the cap-binding protein eIF4E, the RNA helicase eIF4A, and the large scaffold protein eIF4G that facilitates circularization of pre-mRNAs via an interaction with polyA-binding protein 1 (PABP1) (Tarun and Sachs, 1996; Wakiyama et al., 2000). PABP1 serves as a central regulator of mRNA fate in the cytoplasm, coordinating the regulation of mRNA utilization and destruction (Brook et al., 2009; Kumar and Glaunsinger, 2010). This protein complex is an important regulator of bulk protein synthesis by recruiting the ribosomal pre-initiation complex to pre-mRNA and initiating ribosomal scanning (Sonenberg and Hinnebusch, 2009).

Several other 3' end-processing proteins have been identified including the poly (A) binding proteins (PABPs) and these are an important class of gene regulatory proteins that are required for the correct and efficient polyadenylation (Kühn and Wahle, 2004). PABPs affects both nuclear and cytoplasmic mRNA metabolism, where they control the length of the poly (A) tail. Cytoplasmic PABP are highly conserved throughout eukaryotes and plays a crucial role during translation initiation and mRNA surveillance by interacting with the m7G-cap and forming a ‘closed loop’ between the 5' end and the 3' end of the transcript (Kahvejian et al., 2001). Tight coupling of mRNA 3’ end processing, maturation and export ensures that only correctly terminated transcripts are transported to the cytoplasm (Hocine et al., 2010; Proudfoot, 2011).

#### 1.2.2.3 3’ end processing of mRNA

Similar to capping of the 5' end, formation of the 3′ end of the mature mRNA is a well-orchestrated process that involves components of the transcription, splicing and translation machinery (Drummond et al., 1985). With the exception of some histone mRNAs, all eukaryotic mRNAs possess poly(A) tails at their 3 end, which are produced by a two-step reaction involving endonucleolytic cleavage and subsequent addition of poly(A) tail (Proudfoot, 2004). The site of 3’ end cleavage requires multiple proteins and complex, which are highly conserved to regulate the 3’ end processing. These proteins include a poly (A) polymerase that is exposed on cleavage and 4 multi-subunit complexes: cleavage polyadenylation specificity factor (CPSF), cleavage stimulation factor (CSTF) and two cleavage factors (CF1A and CF1B) (Mandel et al., 2008; Proudfoot, 2004). Cleavage polyadenylation specificity factor (CPSF) is core component of the 3’ end-processing complex and recognizes the core sequence element AAUAAA in the polyadenylation site and catalyzes the cleavage (Chan et al., 2014; Schönemann et al., 2014). The cleavage stimulation factor (CSTF) specifically recognizes the U/GU-rich elements, and is required for cleavage but not polyadenylation (Mandel et al., 2008; Takagaki and Manley, 1997). The cleavage factors (CF) complex binding to the RNA element UGUA and facilitates assembly of the 3’ end-processing complex. The CF complex also enhances the efficiency of polyadenylation site cleavage (Chan et al., 2011; Tian and Graber, 2012). Polyadenylation is a dynamic process and up to 70% of human genes express mRNAs with different 3' ends, owing to alternative polyadenylation. Alternative polyadenylation produces mRNA with different 3'UTRs that influence the gene expression through mRNA stability, localization or transport (Liu et al., 2017). Lot of evidences suggests that transcripts with aberrant 3' end processing are retained in the nucleus and less stable (Burkard and Butler, 2000; Hilleren et al., 2001).

### 1.2.3 Pre-mRNA splicing

Pre-mRNA splicing is one of the most critical mRNA processing events, which consist the removal of noncoding intervening sequences, or called introns and

joining of exons of a pre-mRNA (Braunschweig et al., 2013a; Papasaikas and Valcárcel, 2016). In Eukaryotes the splicing reaction is catalyzed by a large multi-unit ribonucleoprotein (RNP) complex called the spliceosome (Bertram et al., 2017; Brody and Abelson, 1985). The introns vary considerably in length but flanked by conserved sequence elements consisting of 5′ and 3′ splice sites (SS) and contain a branch point sequence (BPS). Most of eukaryotic introns contain highly conserved GU dinucleotide at the 5' splice site and is preceded by a G in the exon. The intronic sequence normally ends with an AG dinucleotide, followed by a G as the first nucleotide of the 3' splice site (Rogozin et al., 2012). The branch point sequence are highly conserved adenosine and is located 18-40 nt upstream of the 3' splice site (Figure-1.2.1). Major-class of introns often contain a polypyrimidine tract consists of 10-15 pyrimidine nucleotides between the BPS and the 3' splice site, whereas the conservation of the described consensus sequences is less stringent in mammalian pre-mRNAs (Jurica and Moore, 2003).

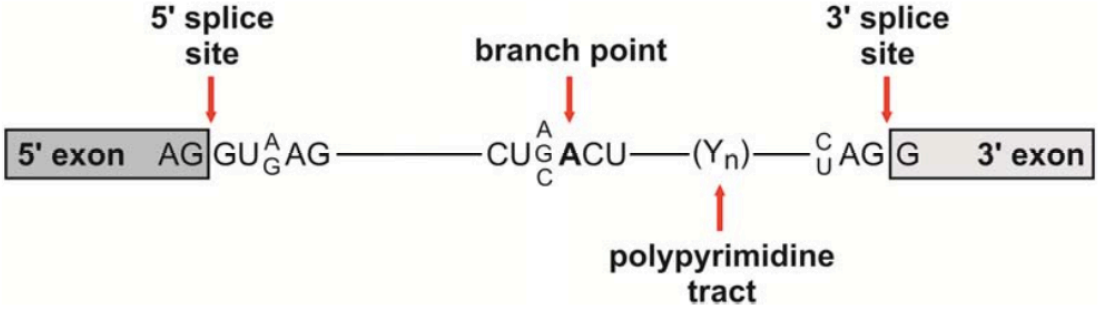



**Figure-1.2.1: Consensus sequences of pre-mRNAs.** The conserved consensus sequences of the 5' splice site (5' SS), 3' splice site (3' SS), and branch point sequence (BPS) (derived from diverse mammalian organisms) are shown. The branch point adenosine is indicated in bold lettering and the poly-pyrimidine tract by (Yn), where Y indicates a pyrimidine base. Two bases on top of each other illustrate an equal frequency of both bases in the consensus sequence. The exons are represented by shaded box and the intron by lettering or a solid line. Figure adapted from (Patel and Steitz, 2003).

Pre-mRNA splicing entails two consecutive, energy independent $S_N2$-type trans-esterification steps (Moore and Sharp, 1993). In the first, the 2'-hydroxyl group of an adenosine of the BPS in the intron carries a nucleophilic attack at the 5'SS and generating a 2-5'phosphodiester bond (see Figure-1.2.2). This

results in a free 5′-exon and a lariat-exon intermediate consisting of the branched intron and 3′-exon. In the second step, the free 3′ hydroxyl of the 5′ exon attacks the phosphate group between the intron and the 3′ exon, which joints two exons and release the lariat intron.

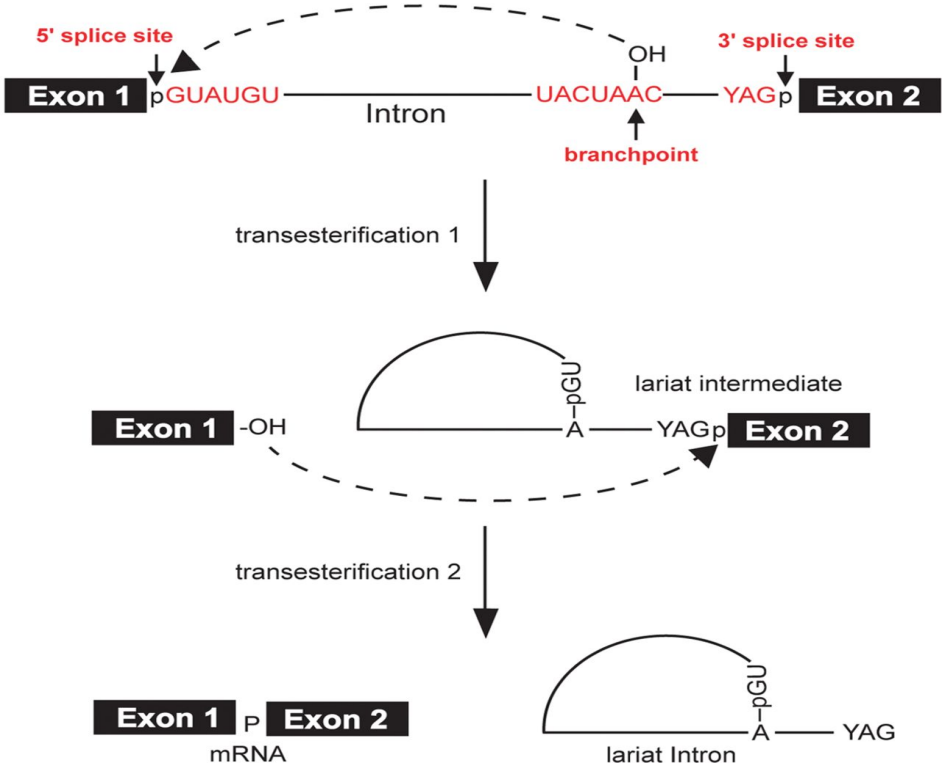



**Figure-1.2.2: Splicing takes place in two catalytic steps involving two consecutive trans-esterification reactions.** In the first step, the 2'-hydroxyl group of the A residue at the branch site attacks the phosphate at the GU 5'-splice site. This leads to cleavage of the 5' exon from the intron and the formation of a lariat intermediate. In the following step, a second trans esterification reaction, which involves the phosphate (p) at the 3' end of the intron and the 3'-hydroxyl group of the detached exon, ligates the two exons. This reaction releases the intron, still in the form of a lariat. In both steps, the nucleophilic attack is indicated by arrow. The consensus sequences (see Figure 1.2.1) and the branch point adenosine highlighted in red. The reactive groups of the pre-mRNA are indicated with the “p” for the phosphate groups, and “OH” for the hydroxyl group, together with their position on the ribose**.** (Adapted from (Chen and Cheng, 2012).

### 1.2.3.1 The dynamics of spliceosome complex

The splicing reaction is catalyzed by the spliceosome, a highly dynamic complex molecular machine composed by small nuclear RNPs (snRNPs). The major spliceosome is composed by five snRNPs the U1, U2, U4, U5, and U6 that sequentially associate with the pre-mRNA during the splicing reaction (Wahl et al., 2009). Interestingly, the prevailing view of the spliceosome is that it does not act as a preformed complex, but instead assembles stepwise on the pre-mRNA (Chen and Moore, 2014; Papasaikas and Valcárcel, 2016). In order to distinguish the individual snRNP complexes that form during splicing, the splicing intermediates have been designated as E, A, B, B* and C complex (Figure-1.2.3).

The first step of the splicing process is the assemble of E complex: the 5′ splice site (GU, 5′ SS) is bound by the U1 snRNP, and the splicing factors SF1/BBP and U2AF cooperatively recognize the branch point sequence (BPS), the polypyrimidine (Py) tract, and the 3′ splice site (AG, 3′ SS). The subsequent step is ATP-dependent and lead to the formation of pre-spliceosome or complex-A through base pairing of U2 snRNP with BPS of pre-mRNA. The next step leads to the binding of the U4/U5–U6 tri-snRNP to the 5′ splice site region and generate the B complex that is still catalytically inactive. The spliceosome undergoes a dramatic compositional and structural remodeling events including the loss of the U1 and U4 snRNPs and binding of the NineTeen Complex (NTC). The NTC complex is highly conserved protein complex and essential for pre-mRNA splicing activation and correct progression (next paragraph). These remodeling facilitate the first step of splicing and formation of the activated spliceosome (B* complex). Complex C is responsible for the two trans-esterification reactions at the splice sites. Additional rearrangements result in the excision of the intron, which is removed as a lariat RNA, and the ligation of exons. The U2, U5, and U6 snRNPs are then released from the complex and recycled for subsequent rounds of splicing process.

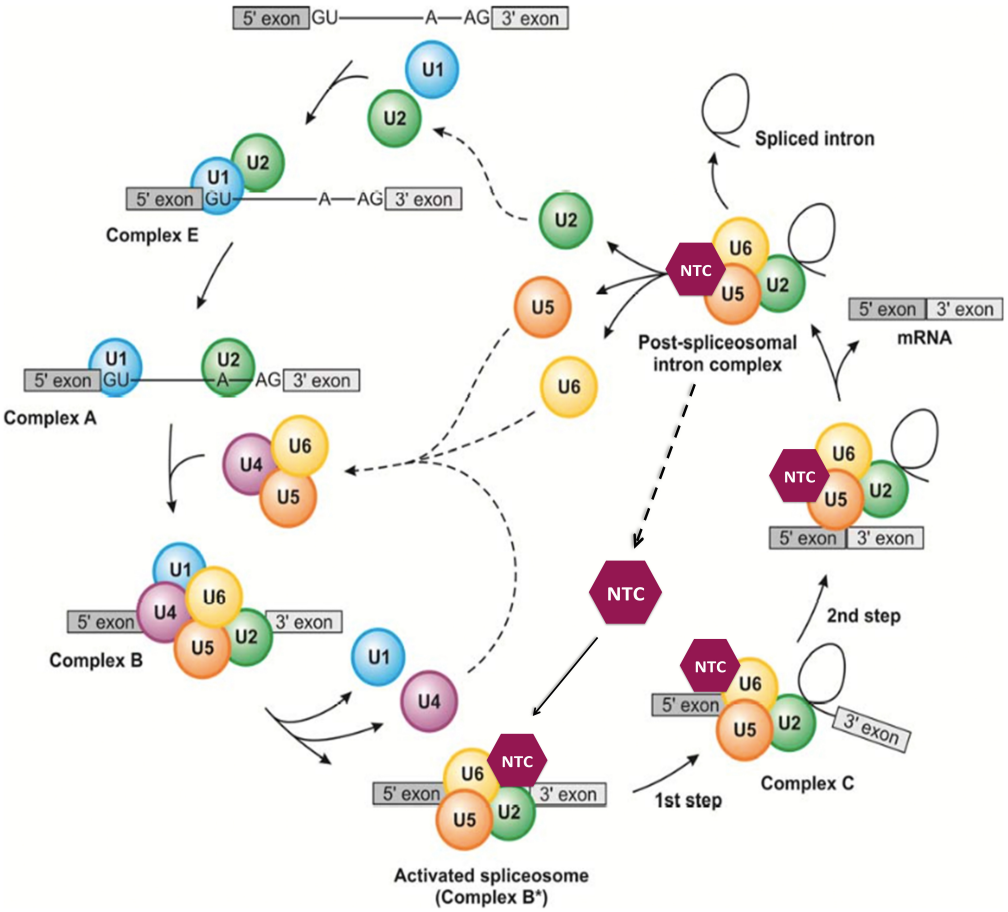



**Figure-1.2.3: Schematic representation of the spliceosome assembly**. The formation of spliceosome complex is a dynamic process, where spliceosomal snRNPs (indicated by colored circles) are assemble on the pre-mRNA. Spliceosome assembly is initiated by the binding of U1 and U2 snRNPs to the 5' SS, followed by the stable association of U2 snRNP with the branch site. After binding of the U4/U6.U5 tri-snRNP and NTC complex the spliceosome is catalytically activated, leading to the dissociation of U1 and U4 snRNPs. After splicing reaction, the spliceosome disassembles, and the snRNPs and NineTeen Complex (NTC) are recycled for subsequent rounds of splicing. The distinct spliceosomal complexes are indicated. (Adapted from (Chen and Moore, 2014).

### 1.2.3.2 Roles of different proteins in pre-mRNA splicing

The spliceosome is very complex and dynamic molecular machinery composed of more than 170 highly conserved proteins and 45 are snRNP-associated proteins (Wahl et al., 2009). Various other regulatory proteins and complexes also plays critical roles in splicing modulation and regulation. Since

most RNA-RNA interactions in the spliceosome are very short, many proteins required for the stability of spliceosome. U1 snRNP and SR proteins (serine-arginine-rich domains) stabilize the initial interactions of the U1 snRNA with the 5' SS (Bertram et al., 2017; Plaschka et al., 2017; Will and Lührmann, 1997). Prp8, Brr2 and Snu114 are integral components of the U5 snRNP. The highly conserved nuclear protein U5-specific protein Prp8 occupies the central position of spliceosome, interacting with many other spliceosomal proteins, snRNAs and both 5′ and 3′ splice sites (Boon et al., 2006; Grainer and Beggs, 2005). The DExD/H-box RNA helicase Brr2, is required for the release U4 from the spliceosome and provides a platform for interaction with other spliceosomal components (Bertram et al., 2017; Chang et al., 2009; Raghunathan and Guthrie, 1998). The Snu114 contain GTPase domain essential for the assembly of the U5 snRNP and it has been proposed to regulate spliceosome dynamics together with Brr2 (Absmeier et al., 2016; Small et al., 2006).

Several other DExD/H-box proteins are required for the splicing process including modulation and rearrangements of the spliceosome at different steps during splicing reaction (Awasthi et al., 2018; Staley and Guthrie, 1998; Wassarman and Steitz, 1991). For instance, Sub2 and Prp5 two ATP-dependent DExD/H-box proteins are involved in early steps of spliceosome assembly facilitating U2 binding to the spliceosome (Liang and Cheng, 2015; Libri et al., 2001). Other members of this family are also found to associate with the spliceosome during all steps of spliceosome assembly and maturation. Spliceosome rearrangements required for step I and II are catalyzed by Prp2 and Prp16, respectively (Egloff and Murphy, 2008; Lee and Young, 2013; Libri et al., 2001; Phatnani and Greenleaf, 2006) and U5-100K/Prp28, Prp22, and Prp43, catalyze the dynamic rearrangements of RNA-RNA and RNA-protein network (Papasaikas and Valcárcel, 2016; Schneider and Schwer, 2001; Staley and Guthrie, 1998).

#### 1.2.3.2.1 The NTC complex is a part of the heteromeric complex of spliceosome

The NTC complex (NineTeen Complex) plays an important role in regulating spliceosome conformations and fidelity. The molecular organization of NTC complex is considerably different but still evolutionary conserved from yeast to human. The NTC complex also known as Prp19 complex (Pre-mRNA-processing factor 19), composed in humans by eight highly conserved core proteins (Prp19, Cef1, Snt309, Xab2, Syf2, Clf1, Isy1, Ntc20) and thirteen associated proteins (Cdc5, Spf27, Prl1, AD-002, Skip, Syf3, ECM2, Aquarius, GCIPp29, MGC23918, G10, Cyp-E, PPIase-like1). NTC complex joins the spliceosome together with the U4/U6.U5 tri-snRNP, and form a stable heteromeric complex in human (the hPrp19/CDC5L complex; (de Almeida and O'Keefe, 2015; Grote et al., 2010; Proudfoot, 2011; Zhang et al., 2018) and yeast (Nature reviews. Genetics., 2000; Tarn et al., 1994; Zhang et al.; Zhang et al., 2018).

The NTC complex is recruited to spliceosome during the formation of the B complex and release of U1 and U4. (Tarn et al., 1994). The abundance of NTC proteins is greatly increased after activation of the B complex and plays important roles in spliceosome dynamics (De et al., 2015; Fabrizio et al., 2009). The NTC complex and NTC related proteins are stably associated with U5 snRNP during catalytic activation of the spliceosome and remain associated during both catalytic steps (Kühn and Wahle, 2004; Nancollis et al.; Nguyen et al., 2016). This complex specifies the dynamic interactions between both exon and U5, and between U6 and the 5'SS during spliceosome activation (Chan and Cheng, 2005; Chan et al., 2003). Therefore, it is suggested that NTC complex and related proteins play an important role for specifying snRNA interactions with the pre-mRNA during splicing process. The best-characterized function of NTC complex is in splicing, however NTC have also been implicated in many other non-splicing functions, including cellular senescence, DNA damage response, pi-RNA pathway, and miRNA biogenesis (Akay et al., 2017; Hogg et al., 2010; Jia et al., 2017; Kuraoka et al., 2008).

Prp19 (Pre-mRNA-processing factor 19), the core protein of the complex is required for the integrity of the NTC complex and facilitates spliceosomal rearrangements through ubiquitination activity (Song et al., 2010). The Human PRP19 protein was also found to have an important role in transcription-coupled RNA processing and genomic stability (David et al., 2011; Paulsen et al., 2009). CDC5L, one of the interacting partners of Prp19, is a known key regulator of mitotic progression and has been implicated in ATR (ATM (ataxia telangiectasia mutated) and Rad3-related signaling (Mu et al., 2014). Syf1/XAB2/Fandango, a tetratricopeptide-repeat (TPR) containing protein, interacts with most NTC/Prp19C components and play crucial role in splicing by mediating the assembly of early spliceosome (Guilgur et al., 2014; Kuraoka et al., 2008; Xu et al., 2002). Syf1 the yeast orthologous of Fandango/XAB2 is crucial for the recruitment of a coiled-coil protein called Yju2, an associated component of NTC complex and required for pre-mRNA splicing both *in vivo* and *in vitro*. Yju2 is associated with the spliceosome at nearly the same time as NTC but destabilized after the first catalytic reaction, and subsequently stabilize for further spliceosome activation (Liu et al., 2007).

### 1.2.4 Aim of the work

Spliceosome is an exquisitely dynamic macromolecular machine, and it is clear that the NTC is an integral component of the active spliceosome complex. NTC core and associated proteins are tightly coordinated in order to carry splicing activation and splicing process. Studies using budding yeast as model system have played an instrumental role in deciphering splicing mechanisms but the biggest challenge of the field is to understand the physiological relevance of splicing. Interestingly, Fandango, the *Drosophila* orthologue of human XAB2 (XPA binding protein 2), and core component of the Prp19 complex is particularly rate limiting for efficient splicing of early embryonic transcripts during *Drosophila* development. Since Fandango depletion causes a very specific phenotype, we hypothesize that other splicing factors are likely to be differentially rate limiting for splicing during development.

The aim of this wok was the identification of spliceosome subunits capable of modulating splicing efficiency during the development, and whose depletion produced specific phenotypes during oogenesis and early embryonic development. To identify such proteins, we performed a screen using the UAS/Gal4 system and a germ-line specific Gal4 driver (Nanos-Gal4), to specifically knockdown during oogenesis distinct spliceosome subunits. This screen identified the Nine Teen Complex (NTC) subunit Salsa, as being required during oogenesis, for dorsal ventral (DV) patterning of the *Drosophila* egg.

We showed that Salsa, the *Drosophila* orthologue of human Aquarius RNA helicase, is required for the correct splicing, localization, and expression of Gurken during oogenesis. Gurken is a TGF-alpha-like protein expressed in the female germ-line and whose function is crucial for DV patterning of the *Drosophila* egg. We showed that Salsa function is particularly rate limiting for splicing of the first intron of gurken mRNA, when compared to the second and third introns of this transcript. Interestingly, and further confirming such bias, transcriptome wide analysis during oogenesis suggested that splicing of a subset of small introns localized to the 5´-region of transcripts are particularly sensitive to Salsa depletion.

Currently, we are planning to do iCLIP (individual-nucleotide resolution UV crosslinking and immunoprecipitation), in collaboration with Dr. Jean Yves and Dr. Julian Konig (Institute of Molecular Biology, Mainz), to mechanistically understand such observation, as it is likely to give an important insight about spliceosome plasticity during development. Our aim is to identify genome wide RNA-binding sites of Salsa at individual-nucleotide resolution, providing an insight into a potential novel function of Salsa in splicing.

# 2

# Identification of splicing proteins rate-limiting for *Drosophila* oogenesis

## Contents

## 2.1 Introduction

*Drosophila* forward genetic screens are powerful tools to identify novel genes and pathways required for distinct biological process (St Johnston, 2002). Our working hypothesis is that splicing plasticity is important for differential gene expression during development. To test this hypothesis, we performed an RNAi screen for conserved spliceosome subunits, and other splicing factors, whose knockdown in the female germ line was associated with specific phenotypes during oogenesis and early embryonic development (Figure- 2.1). Our aim was to identity subsets of introns, whose splicing was particularly sensitive to depletion of distinct splicing factors, rationalize why splicing was affected, and contextualize some of the observed splicing defects with the observed developmental defects.

We performed an RNAi screen for distinct core-subunits of the spliceosome, and other highly conserved splicing factors, using female germ-line specific Gal4 drivers (e.g. Nanos-Gal4), and we identified 13 highly conserved splicing proteins, whose depletion lead to apparently specific developmental defects during oogenesis or early embryonic development. Based on the observed phenotypes, these hits were divided in three different classes: embryonic lethal, dorsal ventral patterning (DV) defects, and short eggs.

Since many of the depleted splicing proteins are essential for cell viability, the observed developmental phenotypes most likely resulted from the fact that RNAi-mediated knock-down of proteins was only partial, allowing for most splicing events to occur normally, whereas splicing was only affected in small subsets of introns. Better understanding the nature of such introns, and why is their splicing particularly sensitive to depletion of a given splicing factor, will give an important insight into spliceosome plasticity, and its role for differential gene expression during development.

## 2.2 Materials and methods

### 2.2.1 Fly husbandry

All flies were raised at 25°C unless otherwise indicated, under standard procedures.

### 2.2.2 *Drosophila* strains

Nanos-Gal4 maternal drivers were used for germ line specific depletion during *Drosophila* oogenesis (Nanos::Gal4-VP16) (Van Doren et al., 1998). The UAS-shRNA vector used in this screen is the VALIUM22, which carries the P-transposase core promoter, known to effectively expressing the RNAi sequences within the female germline. Details of the lines used in this study can be found in Supplementary Table 1.

### 2.2.3 Details of *Drosophila* screen

The crossing scheme and summary are illustrated in Figure 2.1. Approximately 10 virgin females of Nanos-GAL4 were crossed with 5 UAS-shRNA homozygous or heterozygous males at 25° C. F1 Females with Nanos-GAL4 and UAS-shRNA were collected from the previous cross and mated to wild type males. Egg and embryos were collected in apple juice plates supplemented with fresh yeast. The percentage of embryos hatching was determined by counting approximately two hundred embryos. The percentage of embryos that arrest at later stages of embryogenesis (embryos with a brownish color) was also evaluated. Whenever embryonic lethality was observed, embryos were examined morphologically and stained for early embryonic defects.

### 2.2.4 Fluorescence analysis

0-3h embryos (after egg-laying) were collected, dechorionated with 50% of bleach, fixed with 4% of formaldehyde, and devitellinized by vigorously shaking during 1 min in methanol. After rehydration, with a progressively increased percentage of PBST, DNA staining was performed with DAPI (1:10000) and nuclear envelopes were stained with Cy5-

conjugated WGA (1:1000). Embryos were mounted in dako faramount aqueous mounting medium (Dako, California, USA) and were visualized using a Zeiss LSM710 confocal microscope.

**2.2.5 Sequence alignment and functional domain arrangement.**

For identification of conserved helicase domains, orthologs of CG31368 from 6 distinct species representative of the higher eukaryotes (*Homo sapiens, Mus musculus, Caenorhabditis elegans, Danio rerio, Xenopus tropicalis, Drosophila melanogaster*) were retrieved using reciprocal bidirectional protein BLAST analysis. These retrieved sequences were multiple aligned using ClustalW84 program in the Geneious software (version 6.1.8) with default values.

## 2.3 Results and discussion

### 2.3.1 Overview of screen results

We performed an RNAi screen for conserved spliceosome subunits, and other splicing factors, whose depletion in the female germ line was associated with specific phenotypes during *Drosophila* oogenesis and early embryonic development. To achieve germ line-specific expression of the RNAi construct we used the UAS/Gal4 system (St Johnston, 2002) and the Nanos-Gal4 Driver (Van Doren et al., 1998).In this screen (Summarized in Figure- 2.1) we tested 140 RNAi stocks targeting highly conserved splicing factors (Herold et al., 2009). Initially, we investigate the impact of RNAi-mediated knock down in egg laying rate. Further we classified the laid eggs/embryos (hereafter referred to as knockdown embryos) in three different classes: embryonic lethal, dorsal ventral patterning (DV) and short eggs phenotype. Most of the tested RNAi (127 lines) either produced no detectable embryonic phenotype, with normal egg eclosion, or induced significant (apparently pleiotropic) oogenesis defects with a dramatic reduction of egg laying. In addition, 5 RNAi lines were associated to dorsal ventral patterning (DV) defects (eggshell defects; fused dorsal appendages), 6 RNAi lines were associated with morphologically normal eggs that failed to hatch, even when females were crossed with wild type males, and 2 RNAi lines were associated with abnormally short eggs (Summarized in Figure-2.2).

### 2.3.2 Preliminary analysis of isolated RNAi

#### 2.3.2.1 Phenotype: Embryonic lethality

To better characterize the RNAi lines whose embryos failed to hatch, we stained the laid embryos for DNA (DAPI) and nuclear envelope (WGA). Control embryos (mCherry RNAi) showed normal blastoderm cellularization, gastrulation, and germ-band extension, with correctly formed pole cells (Figure-2.3.) (Data not shown). The preliminary observation form embryonic phenotypes after RNAi-mediated knockdown (Acinus, Aubergine, NonA, Caper, exu and SMU1) are the following:

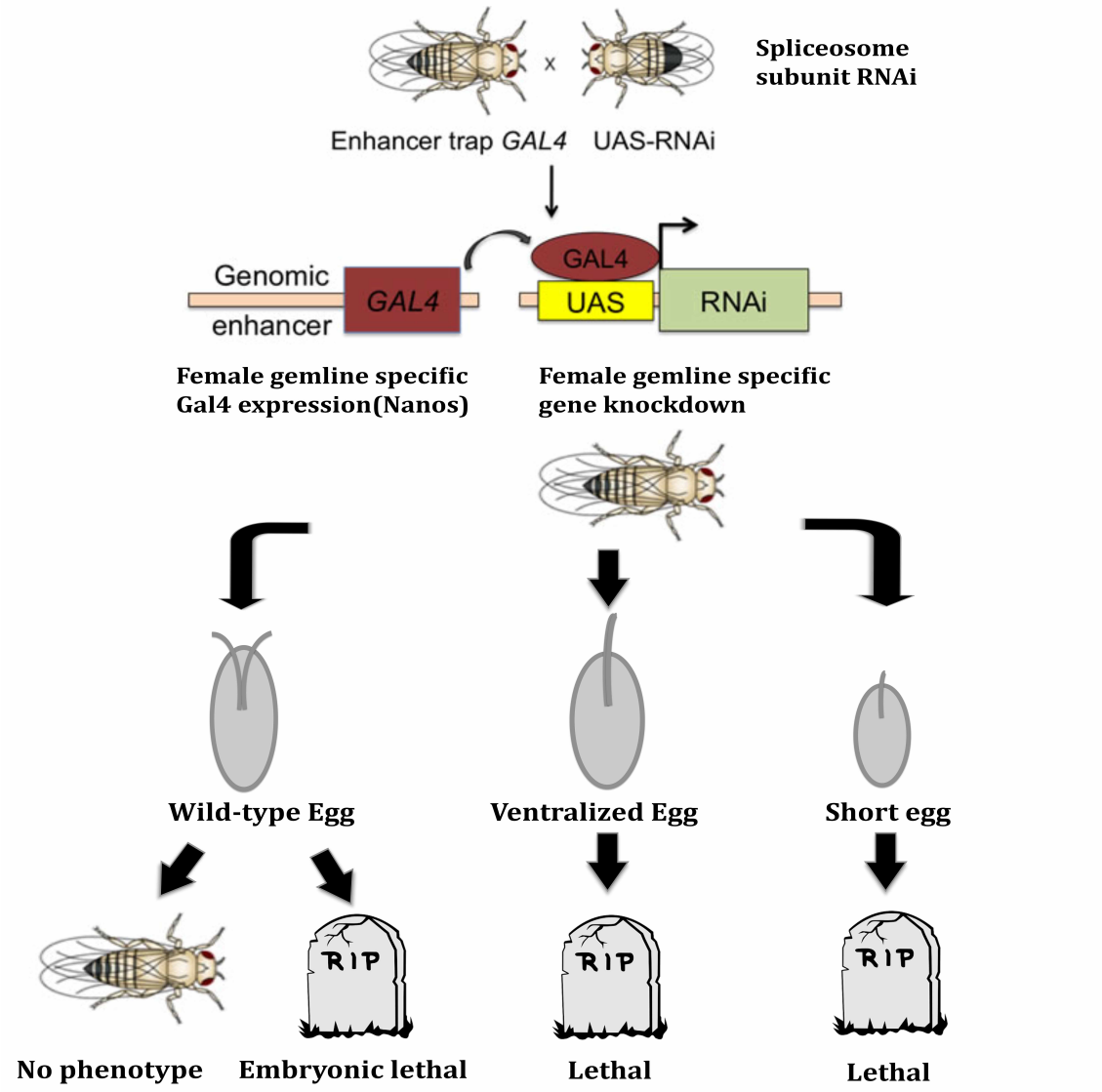



**Figure 2.1. Maternal screen for spliceosome subunits whole depletion is associated with specific phenotype during oogenesis and early embryonic development:** A *Drosophila* maternal screen was performed by crossing male carrying UAS responder (UAS-RNAi), with the females contains Nanos - GAL4 driver that specifically expressed in female germ cells. Total 140 spliceosome specific genes were screened as part of this project; those are highly conserved spliceosome protein from yeast to human. The F1 female progeny containing both elements (UAS RNAi and Nanos Gal4) of the system are produced were mated with wild type male and scored the egg/ embryonic phenotype derived from them. The phenotypes were classified based on morphological characterization and embryonic lethality. The phenotypes were divided into main three groups: embryonic phenotype, spindle phenotype and short eggs, depending upon the nature of the observed effect.

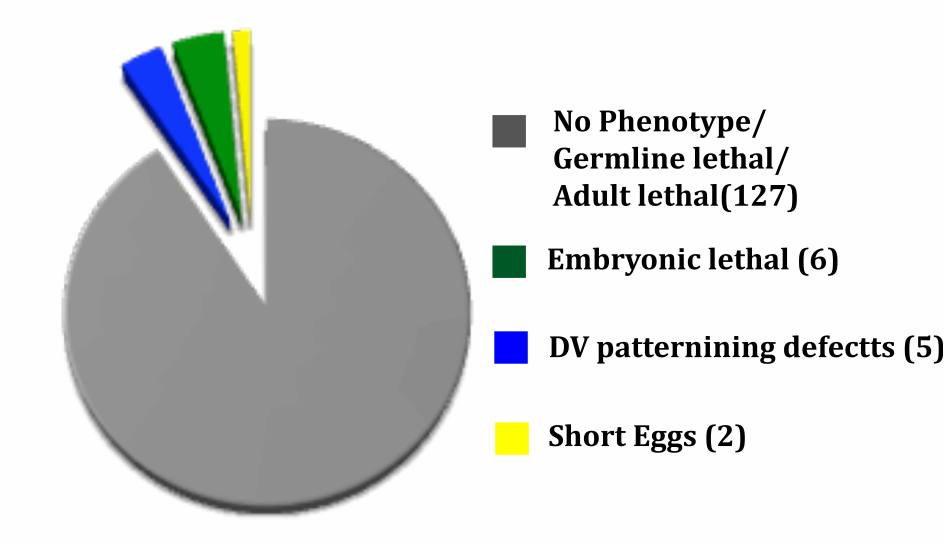



**Figure 2.2: Classification of the phenotype based on morphological features of eggs:** We analyzed total 140 spliceosome related RNAi lines that summarized in pie chart. The primary screen was further characterized according to their egg shape, egg laying, and embryo eclosion rates. Their embryos were stained with DAPI and analysed by confocal microscopy. Total five RNAi lines (in blue) gave rise to dorsal ventral (DV) patterning defects including positive control, squid. Six RNAi lines (in green) gave rise to distinct embryonic phenotypes, two with short egg phenotype (in yellow) and 127 RNAi lines (in gray) gave rise either no obvious embryonic phenotype or dramatically impaired oogenesis with a significant reduction in egg laying.

### Caper

Caper is a highly conserved RNA binding protein that regulates alternative splicing in breast adenocarcinoma cell lines (Brooks et al., 2015; Huang et al., 2012; Stepanyuk et al., 2016). In *Drosophila,* Caper is required for the development of distinct mechanosensory neuron subtypes at multiple stages of development (Olesnicky et al., 2017). In our screen, germ-line specific depletion of Caper impaired in embryonic development, but embryos only showed weak syncytial blastoderm nuclear division defects with normal blastoderm cellularization (Figure-2.3). Germ-band extension defects could be observed in some embryos but the overall observed phenotype was extremely heterogeneous.

**SMU1**

SMU1 (Small mushroom bodies) is highly conserved spliceosome protein known to regulate alternative splicing in *C. elegans* (Lundquist et al., 1996). In *A. thaliana*, smu1 is important during the first catalytic step of splicing and modulates alternative splicing (Kanno et al., 2017). In *Drosophila*, the functional role is not well characterized, but we observed that germ-line specific depletion of SMU1 did not impaired the syncytial blastoderm nuclear divisions and blastoderm cellularization was normal (Figure 2.3). We failed to identify the cause of the observed embryonic lethality, with no obvious defects during early embryonic development.

**Exuperantia**

Exuperantia (Exu) is a putative exonuclease and a highly conserved splicing factor. In *Drosophila*, exu is required for spermatogenesis, as well as for antero-posterior polarity of the developing oocyte; being required for anterior localization of bicoid mRNA (Lazzaretti et al., 2016). In our screen, germ line specific depletion for exu, as expected, to major morphogenetic defects of the developing early embryos (Figure- 2.3). Blastoderm cellularization was normal, with a mild bi-caudal phenotype.

**Aubergine**

Aubergine is a member of piwi family proteins and crucial for a wide variety of developmental processes, including retrotransposon silencing and maintenance of genome integrity during *Drosophila* oogenesis (Ku and Lin, 2014; Ma et al., 2017; Webster et al., 2015). Aubergine is known to destabilize mRNAs during embryonic maternal-to-zygotic transition and localization to the polar granules during development. In our screen, germ line specific depletion for aubergine impaired with egg fertilization. Some of the embryos were nevertheless fertilized, but they showed significant blastoderm syncytial nuclear divisions defects (Data not shown).

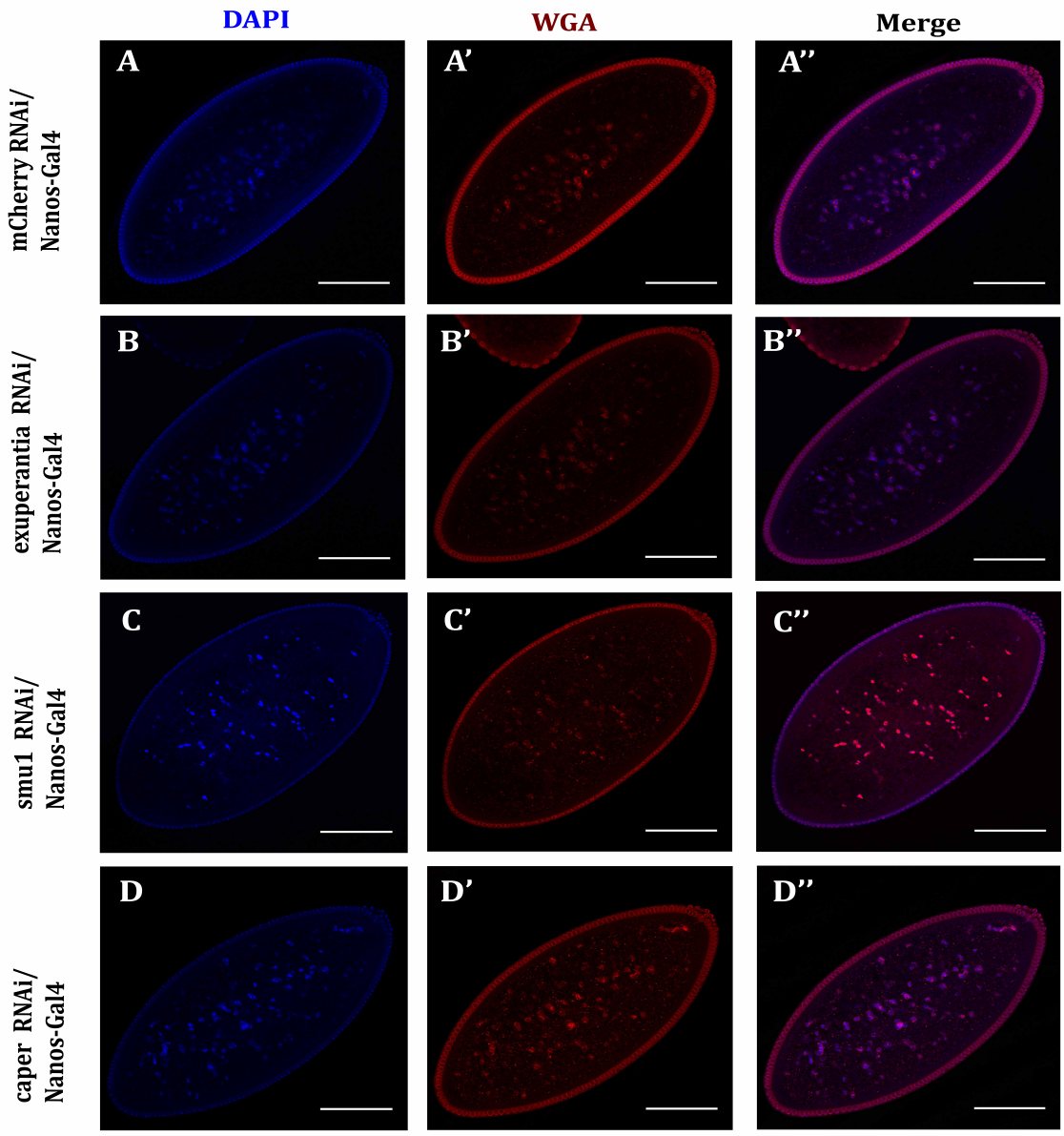



**Figure 2.3. Embryonic phenotype does not phenocopy the fandango mutant phenotype** (Guilgur et al., 2014) **(Blastoderm cellularlization defects) :** Embryos derived from RNAi knockdown mothers were stained for DNA (blue) and WGA (red) to visualize morphology of cells. 13-14 stages of embryos were used to analyze and comparison the phenotype with control RNAi. (A, A' and A'') Embryos derived from control RNAi (mCherry) used for the comparison and these embryos develop normally. (B, B' and B'') Embryos derived from exuperantia- RNAi develop norammy till stage 13-14 but later on shows diverse phenotype including, abnormal morphogenesis, mild or bi-caudal phenotype and many died after germ band extention. (C, C' and C'') Embryos derived from smu1- RNAi aslo develop normally till stage 13-14 and normal nuclear division but died after germ band extension. (D, D' and D'') Embryos derived from caper- RNAi show quite heterogeneous phenotype, where some embryos show abnormal nuclear division defects, some of them are unfertilized and strange germ band extension. Scale bar – 100um

**Acinus:**

Acinus (apoptotic chromatin condensation inducer in the nucleus) is a RNA-binding protein (RBP) and an auxiliary component of the exon junction complex (EJC), which is deposited at exon junctions as a consequence of pre-mRNA splicing. Depletion of acinus associated with abnormal splicing of *piwi*, an essential component of transposon pathway in *Drosophila* oogenesis (Malone et al., 2014b; Rodor et al., 2016). In our screen, germ-line specific depletion of acinus, shows very heterogeneous phenotype including, defects in blastoderm cellularization, some of them shows defects in nuclear division and very abnormal germ band extension (data not shown).

**NonA:**

The NonA protein of *Drosophila* is highly conserved splicing protein and binds with DNA/RNA in vitro. In *Drosophila* S2 cells, NonA binds with essential nuclear export factor NXF1 (Nuclear export factor 1) and shows function in mobility of mRNP particles (Kozlova et al., 2006). NonO, the human homologue of NonA believed to be involved in RNA splicing (Clark et al., 1997), but its precise function is still unknown. IN our screen, germ line specific knockdown of NonA shows mostly unfertilized eggs and very early mitotic phenotype (data not shown).

In conclusion, all identified splicing factors belonging to the "morphologically normal eggs that failed to hatch" class, were associated with previously reported phenotypes or with phenotypes apparently too pleiotropic to work with.

### 2.3.2.2 Phenotype: DV patterning defects

Next, we examined the RNAi lines whose laid eggs showed dorso-ventral patterning defects (fused dorsal appendages; spindle phenotype).

**Squid**

Squid (Sqd), also known as Hrp40 (heterogeneous nuclear RNA binding protein 40), is a highly conserved RNA binding protein that binds many different cellular RNAs and implicated in nascent mRNAs processing, and regulation of alternative splicing (Hartmann et al., 2011; Matunis et al., 1992). Squid is important for dorsal ventral patterning during oogenesis as it is important for anterior dorsal localization of gurken mRNA (Steinhauer and Kalderon, 2005). As expected, we observed that germ line specific depletion for Sqd was associated with fused dorsal appendages (spindle phenotype).

**CG7971**

The *Drosophila* CG7971 is a highly conserved serine/arginine (SR)-related nuclear matrix protein and its human orthologue known as SRm300. SRm300 is a well known splicing co-activator, promoting critical interactions between splicing factors and the associated pre-mRNAs (Grainger et al., 2009). In *Drosophila*, CG7971 is uncharacterized gene, and in our screen we observed that germ line specific depletion was associated with a weak eggshell spindle phenotype.

**PEP (protein on ecdysone puffs)**

The *Drosophila* protein PEP (protein on ecdysone puffs) is a component hnRNP complex that associates with distinct SR proteins and was reported to interact with Sqd (Grainger et al., 2009). Although conserved splicing protein, its precise function in splicing is not unknown. Germ line specific knockdown of PEP was associated with a weak eggshell spindle phenotype.

**CG5877**

CG5877 is uncharacterized gene in *Drosophila* but bidirectional BLAST suggest a conserved component of spliceosome that contain Half-A-Tetratricopeptide (HAT) motifs, known as NRDE-2 (Nuclear RNAi Defective conserved domain) (Aronica et al., 2016). Although, the role of NRDE2 in human cells is unknown, in *C. elegans* known as nuclear RNAi silencing protein (Ogami et al., 2017).

#### 2.3.2.3 Phenotype: Short eggs

Females whose germ line was knockdown for tsunagi and Zn72D laid abnormally short eggs with significant fused dorsal appendages, which was suggestive of significant defects in nurse cells dumping (Data not shown).

### 2.3.3 CG31368 / Salsa is a highly conserved RNA helicase

CG31368 is a highly conserved RNA helicase, whose human orthologue is known as Aquarius. In our lab and others previously observed that CG31368 / human Aquarius are subunits of the spliceosome NineTeen Complex (NTC) complex. We observed that germ line specific knock down of CG31368 was associated to a strong eggshell spindle phenotype.

Since CG31368 was the candidate gene whose germ line depletion gave rise to the most specific and still uncharacterized phenotype identified (fused dorsal appendages; which is suggestive of DV patterning defects), and given the fact that the encoded protein is a highly conserved splicing factor and a subunit of the spliceosome, we decided to pursue the characterization of this gene in order to better understand the role of splicing plasticity during development (see chapter 3 for more details).

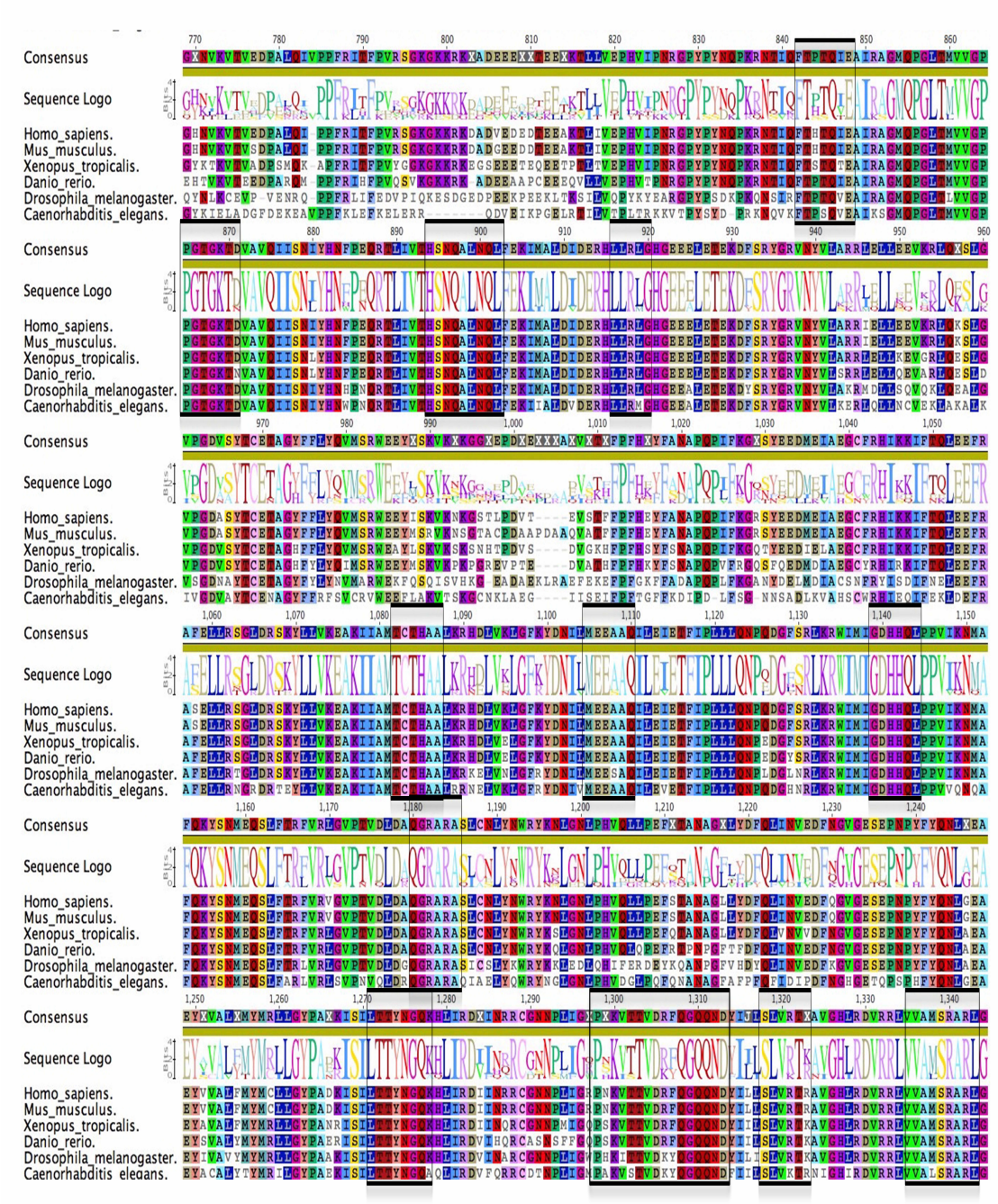



**Figure 2.4. CG31368 is highly conserved RNA helicase protein:** Identified *Drosophila* CG31368 is most likely an active RNA helicase, similarly to human Aquarius, as the helicase domains are highly conserved among all tested species. Full protein sequences were retrieved from the genomes of six distinct representatives of higher eukaryotes species, using reciprocal bidirectional protein BLAST analysis and multiple aligned using ClustalW program and the Geneious R7 software. Letters height represents the degree of conservation of each amino acid residue for that position. Conserved helicase sequence motifs are indicated by black lines on the both side of conserved sequences.

For *in silico* characterization of *D. melanogaster* CG31368 protein, the protein sequences of CG31368 were used in BLASTp search (in NCBI and Flybase) against the *Homo sapiens* proteome. We found that CG31368 was highly similar to a highly conserved spliceosome subunit named Aquarius; which is a subunit of the NTC complex. Since in *Drosophila* there is already a non-related gene named Aquarius, and since previously we showed that CG31368 physically interacts with another NTC-subunit named Fandango, we decided to name CG31368 as Salsa. To further test for functional homology of Salsa, we retrieved the orthologue protein sequences from five other eukaryotes on the basis of E-value of reciprocal best hit (NCBI) and performed multiple sequence alignment (Data not shown). The helicase domains of Salsa were highly conserved (Figure2.4), suggesting that Salsa, similarly to human Aquarius, is likely to be a catalytically active RNA helicase.

# 3

# Salsa is required for female fertility and dorsal-ventral patterning of *Drosophila* egg

## Contents

### 3.1 Introduction

One of the first crucial steps of animal development is to establishing the future embryonic axes, i.e., the Dorsal-Ventral (DV) axis. In the *Drosophila melanogaster*, the mother provides an mRNA called gurken, essential for establishing the future embryonic Dorsal-Ventral (DV) axis. In chapter 2, we identified a gene called Salsa, whose germ line specific knockdown produced a highly penetrant ventralization phenotype (spindle phenotype) and dramatically reduced egg eclosion.

In this chapter we investigate the role of Salsa in ventralization phenotype and Gurken expression. We observed that *gurken* transcript is poorly spliced and abnormally localized after depletion of Salsa. Gurken protein expression was similarly affected, as its localization was abnormal, leading to the working hypothesis that Salsa is required for DV patterning through its regulation of Gurken expression. Interestingly, we observed that after Salsa depletion splicing of the first intron of *gurken* transcript (that localizes to the transcript 5´UTR) was significantly more affected than splicing of the second and third introns.

In order to better understand the role of Salsa in pre-mRNA splicing during oogenesis, we analyzed the ovaries transcriptome after partial depletion of Salsa. Interestingly, we observed that only small subsets of introns particularly sensitive to Salsa depletion and these introns are significantly shorter and more proximal to the transcript start site.

### 3.1 1 Known functions of RNA helicase Aquarius/Salsa

Salsa is a highly conserved RNA helicase protein that is conserved in several eukaryotic model organisms, and has been studied to varying extents. Salsa is the *Drosophila* orthologue of human Aquarius, which also has ATPase and RNA helicase activities. Aquarius has been shown to be a core component of the human splicing machinery, as it directly interacts with known splicing factors such as Xab2, CCDC16, hIsy1 and CypE, which are also linked to the U2 snRNP (De et al., 2015; Guilgur et al., 2014). Aquarius was first identified as Intron Binding Protein 160 (IBP160), by direct binding to small nucleolar ribonucleoproteins (snoRNPs), which are distinct from the spliceosomal snRNPs and play an important role in nucleolytic processing and nucleotide

modification of precursor rRNAs (Kiss, 2006). Moreover, Aquarius was further characterized in the formation of EJC complex (Exon-Junction Complex) and its recruitment to spliced mRNA. EJC-binding to mRNA and its integrity are important factors that contribute to directing the mature transcripts into the nuclear export pathway rather than into the nonsense mediated decay (NMD) pathway (Ideue et al., 2007; Shiimori et al., 2013). The crystal structure of spliceosome and Salsa has also recently been solved, providing additional insights as to the activity of spliceosome activation and function of Salsa (Bertram et al., 2017; De et al., 2015; Egloff and Murphy, 2008).

Furthermore, human Aquarius is also involved in transcription-coupled repair (Sollier et al., 2014), whereas in *C. elegans* and *A. thaliana* it was suggested a role not only in pre-mRNA splicing but also small RNA pathways (Akay et al., 2017; Jia et al., 2017). Although it is still unclear whether the multiple functions of Aquarius are conserved among other eukaryotes, and whether they could implicate the NTC complex in interplay of multiple regulatory functions on gene expression during development, this is a likely possibility as other subunits of NTC (e.g. Xab2 and Prp19) are surprisingly multifunctional, being similarly involved in DNA repair, pre-mRNA splicing, and transcriptional elongation (Martinho et al., 2015).

## 3.2. Materials and methods

### 3.2.1 Fly husbandry

All flies were raised at 25°C, unless otherwise indicated, using standard techniques.

### 3.2.2 *Drosophila* RNAi stocks

All *Drosophila* UAS-RNAi stocks used in this work are listed in Supplementary Table 1. Unless indicated, all stocks are available at the Bloomington Drosophila Stock Center.

Depletion of Salsa was obtained using three different non-overlapping dsRNA hairpins (Supplementary Figure-1). Salsa RNAi-2 stock is available in BDSC (Bloomington number 55172; hairpin reference; P {TRiP.HMC03852}attP40;

Map: Chr 2,25C6, 2L:5108448..5108448.), whereas Salsa RNAi-1 and Salsa RNAi-3 were custom made.

### 3.2.3 Generation of non-overlapping Salsa RNAi (Salsa RNAi-1 and RNAi-3)

1. Oligos design: We designed two non-overlapping RNAi hairpins against Salsa (CG31368). The oligos were designed from 21-nucleotide sequence based on the algorithm (Vert et al., 2006). The oligo design minimizes off target effects starting at 16 nucleotides.

Based on a miR1 scaffold, for the top strand oligo (TS) of Salsa, ctagcagt nucleotides were added to the 5' end of the passenger strand DNA and tagttatattcaagcata nucleotides were added between passenger strand DNA and the guide strand DNA. In the end gcg nucleotides were added to the 3' end of guide strand DNA. The resulted top strand oligo (TS) of Salsa 1 and Salsa 3 will be:

**Salsa RNAi -1 TS:**

ctagcagtCGCTTGGATATGGACGATCTAtagttatattcaagcataTAGATCGTCCATATCCAAGCGgcg

**Salsa RNAi -3 TS:**

ctagcagtCCACGATTATCTCCTACGCAAtagttatattcaagcataTTGCGTAGGAGATAATCGTGGgcg

For the bottom strand oligo (BS) of Salsa, aattcgc nucleotides were added to the 5' end of the passenger strand DNA, and tatgcttgaatataacta nucleotides were added between passenger strand DNA and the guide strand DNA. In the end actg nucleotides were added to the 3' end of guide strand DNA. The resulted bottom strand oligo (BS) of Salsa 1 and Salsa 3 will be:

**Salsa RNAi -1 BS:**

aattcgcCGCTTGGATATGGACGATCTAtatgcttgaatataactaTAGATCGTCCATATCCAAGCGactg

**Salsa RNAi -3 BS:**

aattcgcCCACGATTATCTCCTACGCAAtatgcttgaatataactaTTGCGTAGGAGATAATCGTGGactg

2. Annealing the top and bottom strand oligos: 10µl top strand oligo (10µM) and 10µl bottom strand oligo (10µM) were added into 80µl annealing buffer (10mM Tris-HCl, pH 7.5, 0.1M NaCl, 1mM EDTA). The reaction mix was incubated at 95°C for 5 min, and then, slowly cooled down to room temperature. The resulting DNA fragment has overhangs for *NheI* and *EcoRI*.
3. Ligation: The resulting DNA fragments were directly cloned into a VALIUM22 vector, which has been linearized by *NheI* and *EcoRI* enzymes. For DNA ligation reaction, 6µl of the annealing product mixed with 2µl of 10X ligation buffer and 1µl of T4 DNA ligase (1U/µl). Finally 1µl of 40ng/µl backbone (gel purified VALIUM22 cut with *NheI* and *EcoRI*) was added in final reaction. Final volume 20µl made up with ddH20, mixed properly and incubated at 16°C for 1 hour. The reaction was stored at -20°C, till further use.
4. Transformation, colony selection and sequencing: 10µl of ligation reaction were transformed into 50µl TOP10 competent cells using electroporation method. Correct clones were selected through PCR using the standard primers for pVALIUM22 (Table 3). The correct shRNA constructs were further confirmed by sequencing. The primer used for sequencing is shown in Table 3. Plasmid amplification and isolation of the correct shRNA constructs was done using the standard protocol from NZY Miniprep isolation kit. (NZY Tech, Genes and Enzymes, Lisbon, Portugal). For production of transgenic *Drosophila* stocks, isolated plasmids were sent for injection (BestGene, Chino Hills, CA, USA).

### 3.2.4 Ventralized eggshell phenotypes

UAS-Salsa RNAi transgenic males were crossed with virgin females carrying a germ-line specific maternal Gal4 driver (Nanos: Gal4-VP16;) (Van Doren et al., 1998) Flies were cultured at 25°C and transferred to a new tube every 3 days. 0-3 days old F1 females (n=20) were crossed with wild-type males (n=5) and added into an egg collection cage supplemented with fresh yeast. Ventralized eggshell phenotypes were scored at least 2 days after assembly of collection cage (for productive egg laying) and using fresh egg collections (4-5hr) (to facilitate unequivocal scoring).

### 3.2.5 Egg hatching

F1 females (n=20) were crossed with wild-type males (n=5) and added into an egg collection cage supplemented with fresh yeast. Eggs were collected using freshly prepared apple juice plates. Plates were incubated for 48hr at 25°C and egg hatching was scored.

### 3.2.6 Ovaries immunostaining

Adult ovaries (20 ovary pairs per sample per experiment) were processed according to the standard procedures (Navarro-Costa et al., 2016). Briefly, ovaries were dissected from 2-3 days old females in ice cold Phosphate-buffered saline (PBS) and fixed for 20 min in a 3:1 mix of heptane (Fluka) and fixative aqueous mix. The fixative mix consisted of 4% formaldehyde (Polysciences) in PBS+0.5% NP-40 (Sigma) solution. Following gentle partial detachment of the ovarioles with a tungsten needle, ovaries were incubated with rotation for 2h in PBST (PBS +0.2% Tween 20). Ovaries were blocked by PBST supplemented with 1% Triton X-100 (Sigma), 1% (w/v) bovine serum albumin (BSA; Sigma) and 1% (w/v) donkey serum (Sigma).

Primary antibodies incubation was performed overnight at 4°C in PBST supplemented with 1% BSA and 1% donkey serum (BBT solution). Ovaries were subsequently incubated for 1h at the room temperature with the appropriate secondary antibodies diluted in BBT solution. Prior to mounting DNA was stained for 30 min at the room temperature using 1:10,000 DAPI in PBST. Ovaries were mounted in dako faramount aqueous mounting medium (Dako, California, USA) and were visualized using a Zeiss LSM710 Confocal microscope.

### 3.2.7 Quantification of ovaries Gurken immunostaining

For quantification of Gurken protein signal enrichment at the anterodorsal region of the oocyte it was used the following protocol: maximum intensity projections of serial confocal optical sections of Gurken immunostainings were obtained using Image J program (Grouped Zprojector, maximum pixel intensity). For each egg chamber, ImageJ software was subsequently used

to quantify the average signal intensity value of three oocyte regions with high levels of localized Gurken divided with the average signal intensity value of three regions in the nurse cells cytoplasm (background).

### 3.2.8 Preparation of DIG-labelled gurken probe

A plasmid carrying a gurken cDNA sequence (Gifted by Dr. Amanda Norvell, Department of Biology, The College of New Jersey, USA) was linearized with Sal I restriction enzyme and analyzed on agarose gel electrophoresis. The linearized plasmid was isolated and purified with a gel purification kit (Promega), using standard procedures. The purified linearized plasmid was subsequently quantified with spectrophotometer (Thermo Scientific NanoDrop 2000). For synthesis of DIG-labeled antisense grk RNA probes, linearized plasmid (2µg) was mixed with the following reagents in this order and at room temperature: 2µl of 10X transcription buffer (Roche), 2µl of 0.1M DTT (Roche; freshly diluted from 1M stock stored at -80°C), 2µl RNA DIG labelling mix (Roche), 1µl Placental ribonuclease inhibitor (100 U/ µl; Roche) and 1 µl T7 RNA polymerase (10 U/ µl; Roche). The final reaction volume was 20µl, with volume made up with sterile distilled water. The final reaction mixture was subsequently incubated at 37°C for 2h. To remove the DNA template, 2µl DNaseI (ribonuclease-free, Roche) was added to the reaction and it was subsequently incubated at 37°C for 15 min. The reaction suspend in 40µl DEPC water and add 50µl 2X Carbonate Buffer (120mM-$Na_2CO_3$; 80mM $NaHcO_3$ @ pH-10.2) and incubated at 65° C for 40 min. Add 100 µl 0.2M NaAc (pH-6). For probe purification, 100µl TE buffer (50 mM TrisHCl, 1 mM EDTA, pH 8.0), 10µl of 4M LiCl, and 300µl EtOH, were added to the mix, and probe was precipitated on dry ice, for 30 min. Probe was spin down in the microcentrifuge at 4°C for 30 min and the pellet was washed twice with 80% EtOH. Pellet was dried in air (10min) and dissolved in TE buffer at approximately 0.1µg/µl. The integrity and size of RNA transcripts was analyzed using gel electrophoresis on a 1% RNase free agarose gel with a 1kb RNA marker (New England BioLabs). The total RNA was further quantified by measuring absorbance at 260/280 using spectrophotometer (Thermo Scientific NanoDrop 2000).

### 3.2.9 Fluorescent *in-situ* hybridization of *Drosophila* ovaries

Adult ovaries from 2-3 days old females were dissected and fixed as described before. Ovarioles were gently teased apart with a tungsten needle and transferred to 500 µl of pre-hybridization buffer (50% formamide, 4x SSC (20xSSC (RNAse free), 0.1% Tween 20) and incubated for 1h at the room temperature. The ovaries were then transferred into 200µl hybridization buffer (50% formamide, 5x SSC (20xSSC (RNAse free), 0.1% Tween 20, 100 µg/ml heparin, 100 µg/ml ssDNA) and incubated for 1h at $55^0$C. The probes were diluted in hybridization buffer (1:100) and incubated overnight at 55°C.

Second day, the ovaries were washed with 500 µl pre- hybridization wash buffer for 30 minutes at 55 °C, with PBT (0.1% Tween 20) five times for 30 minutes at the room temperature and incubated in Roche Blocking Buffer (1:10 in PBT) for 30 minutes at the room temperature. The mouse anti-Dig (1:400) was diluted in Roche Blocking Buffer and incubated the ovaries overnight at 4°C.

Third day, the ovaries were washed with 500µl PBT (0.1% Tween 20) for 10 minutes at the room temperature, incubated in Roche Blocking Buffer (1:10 in PBT) for 30 minutes at the room temperature. The mouse anti-mouse-Cy3 (1:1000) and WGA-Cy5 (1:1000) were diluted in Roche Blocking Buffer and incubated for 1h at the room temperature. Ovaries were washed with 500µl PBT (0.1% Tween 20), for 5 minutes at the room temperature and DAPI staining for 30 minutes at the room temperature. The ovaries were then rinsed in PBS and mount in Vectashield.

### 3.2.10 Quantification of gurken mRNA localization

For semi-quantitative quantification of anterodorsal localization of gurken mRNA, stage 8/9 oocytes were divided it into three different classes based on the gurken mRNA signal near to oocyte nucleus: a) Normal anterodorsal localization of gurken mRNA, with a strong signal distributed nearby the periphery of the oocyte nucleus; b) Partial anterodorsal localization, with a weaker but clearly detectable gurken mRNA distributed nearby the periphery of the oocyte nucleus; c) absence/almost absence of anterodorsal

localization, with a an undetectable or hardly detectable gurken mRNA signal nearby the periphery of the oocyte nucleus.

### 3.2.11 RT-PCR (Reverse Transcription- PCR)

Total RNA was extracted from 0–1hr embryos (after egg laying), and from 2-3 days old adult female ovaries (after pupae eclosion), whose female germ line was specifically depleted for Salsa (salsa RNAi) or mCherry (negative control: mCherry RNAi). Extraction was performed following standard procedures (PureLink RNA Mini Kit, Ambion, NY, USA). Genomic DNA was removed from RNA samples using PureLink DNase (Invitrogen) on-column method. The concentration of RNA was determined using a spectrophotometer (Thermo Scientific NanoDrop 2000) (Thermo Scientific Massachusetts, USA). A260:A280 ratios were calculated for each sample and ratios smaller than 1.7 or greater than 1.9 were discarded. For reverse transcription, the iScript cDNA synthesis kit (Bio-Rad, California, USA) was used and 200ng of RNA were used for cDNA synthesis, according to manufacturer's instructions. Primer combinations were designed with PrimerBLAST (NCBI, USA) and PCR was performed using NZYTech master mix. (Lisbon, Portugal).

### 3.2.12 Real-time qPCR (Real –Time quantitative PCR)

#### 3.2.12.1 Optimizations of primers for Real-Time qPCR

Primers were designed using PrimerBLAST (NCBI, USA) and optimized to an equal annealing temperature 55°C. Primer efficiencies were initially tested with *Drosophil*a genomic DNA (gDNA). Five-fold serial dilutions of gDNA were made (40ng, 8ng, 1.6ng, 0.32ng and 0.064ng) and primer efficiencies for both target genes and reference genes were tested. Standard curves were constructed by the Cycle of threshold (Ct) (y-axis) versus log gDNA dilution (x-axis). The primer efficiency (E) of one cycle in the exponential phase was calculated according to the equation: $E=10^{(-1/slope)} -1$ (Nolan et al., 2006; Pfaffl, 2001). Efficiency and regression curves for each primer sets are shown in Table 4. Specificity of primer pair amplification was confirmed by melting curve analysis and the Ct values were obtained from dynamic range of the

standard curves. Only primers sets with specificity and good efficiency were used for subsequent RT-qPCR analysis.

#### 3.2.12.2 mRNA extraction and cDNA synthesis

Total RNA was extracted from 0–1hr embryos (after egg laying), and from 2-3 days old adult female ovaries (after pupae eclosion), whose female germ line was specifically depleted for Salsa (salsa RNAi) or mCherry (negative control: mCherry RNAi). Extraction was performed following standard procedures (PureLink RNA Mini Kit, Ambion, NY, USA). Genomic DNA was removed from RNA samples using PureLink DNase (Invitrogen). Experion™ Electrophoresis (Bio-Rad, California, USA) was used to assess the integrity and quantity of isolated RNA. Isolated RNA with less than 8 RQI was not considered for reverse transcription with iScript (as described above) and further Real-Time qPCR analysis.

#### 3.2.12.3 Quantitative PCR reaction (qPCR)

The equivalent to 4ng cDNA was used as template in each qPCR reaction. qPCR reactions were performed on a CFX96 Real-Time system (Bio-Rad, California, USA) using SsoFast EvaGreen Supermix (Bio-Rad, California, USA). Two reference genes (Actin and GAPDH) were used for normalization; "no template", "no RT" and “NEG” were used as negative controls for testing the qPCR reaction mixture contamination and and a positive control was performed using gDNA. For each reaction, the RT-qPCR master mix included 10µl of 2X EvaGreen mix, 1µl of forward primer (10µM), 1µl of reverse primer (10µM), and 4µl of nuclease-free water. The master mix (16µl) was added to a well of 96-multiwell plate (Bio-Rad, California, USA) followed by the addition (when appropriate) of 4µl cDNA (4ng). PCR protocol: denaturation programme (95°C for 30 seconds), amplification program repeated for 60 cycles (denaturation: 95°C for 5 seconds, annealing: 55°C for 5 seconds + plate read for fluorescence measurement), melting curve programme (65°C to 95°C with 0,5 ºC increments, for 5 seconds + plate read for fluorescence measurement), and cooling programme to 4°C.

### 3.2.12.4 Expression Analysis

The threshold cycle (Ct) of each gene transcript was determined by setting the threshold line above background levels in the linear region of the exponential curve before the gene expression levels were measured. The baseline and fluorescence signals are adjusted automatically by the CFX96 Real-Time machine software (default parameters) to calculate the ΔCt value (Schmittgen and Livak, 2008). The relative expression ratio (RE) was calculated using the difference between ΔCt value of control RNAi (mCherry RNAi) and salsa RNAi, also known as the $2^{-\Delta\Delta \mathbf{Ct}}$ method.

$RE=2 \wedge (-(\Delta Ct_{\ sample}{}^{(target\text{-}reference)} - \Delta Ct_{control}{}^{(target\text{-}reference)}))$

$\Delta Ct_{\ sample}$ is the Ct difference of target-reference genes in Salsa RNAi;

$\Delta Ct_{\ control}$ is the Ct difference of target-reference genes in control RNAi.

Total expression levels of the genes of interest were calculated by the mean of Ct values normalized to the expression levels of two reference genes (Actin and GAPDH) using the geNorm method (Vandesompele et al., 2002). Mean and standard deviation (SD) of RE values were calculated from three biological replicates. In addition, three technical replicates were performed for each sample. A technical replicate corresponds to three different cDNAs synthesized from the same isolated RNA. Biological replicates correspond to RNAs obtained from distinct biological samples.

### 3.2.13 RNA sequencing and analysis

Total RNA was extracted from 2-3 days old adult female ovaries (after pupae eclosion), whose female germ line was specifically depleted for Salsa (salsa RNAi-1) or mCherry (negative control: mCherry RNAi). Extraction was performed following standard procedures (PureLink RNA Mini Kit, Ambion, NY, USA). Genomic DNA was removed from RNA samples using PureLink DNase (Invitrogen). Experion™ Electrophoresis (Bio-Rad, California, USA) was used to assess the integrity and quantity of isolated RNA. cDNA libraries were generated using an Illumina® TruSeq® kit (polyA-tail enrichment using Oligo-dT beads for RNA capture) and sequenced on an Illumina® HiSeq®

2000 sequencing platform.

### 3.2.14 Detail protocol of RNAseq analysis of *Drosophila* ovaries

Transcriptome analysis was performed using four samples for Salsa RNAi or Control RNAi, as shown as below. RNA-sequencing (RNA-seq) data was analysed in order to profile alternative splicing (AS) – with major focus on intron retention – alterations that occur following depletion of Salsa in *Drosophila* ovaries. Analyses were performed using R software for statistical computing or specific tools mentioned along this document.

| Sample ID | Description | Condition | Replicate |
|---|---|---|---|
| A | RNAi Control replica 1 | Control | 1 |
| B | RNAi Salsa replica 1 | Salsa | 1 |
| C | RNAi Control replica 2 | Control | 2 |
| D | RNAi Salsa replica 2 | Salsa | 2 |

#### 3.2.14.1 RNA-seq reads pre-processing

Provided FASTQ files coming from RNA-seq were checked for the overall quality of sequencing reads using the quality control tool fastqc VAST-Tools (Vertebrate Alternative Splicing and Transcription Tools) (Irimia et al., 2014) was used for alignment and quantification of gene expression (GE) and alternative splicing (AS) for this dataset, considering the VASTDB annotation for *Drosophila melanogaster* (genome version dm6 - provided by collaborators). In the VAST-Tools pipeline GE quantification is performed by aligning RNA-seq reads against a reference transcriptome, while AS quantification requires a first step of read alignment against a reference genome, followed by alignment of unmapped reads against splice junction libraries.

#### 3.2.14.2 Gene expression quantification

Gene expression quantification is based on the counting of RNA-seq reads mapping to a given gene. However, for comparative analysis of gene expression, both between samples and within the same sample, normalization needs to be performed to reduce the intrinsic bias of read counts. Reads per

kilo-base of exon model per million mapped reads (RPKMs) are one widespread measure for normalized read counts. RPKMs are obtained, as explained in Equation 1, by taking the total number of reads assigned to a given gene and divide them by the length of the gene (in kilo-bases) times the total number of mapped reads (in millions). This normalized read count ensures faithful analysis of gene expression, without the bias for longer genes or samples sequenced at a deeper level (i.e. with more molecules being sequenced) (Conesa et al., 2016; Mortazavi et al., 2008).

**Equation1.**

$$\mathbf{RPKM}_{\text{gene A}} = \text{gene A reads} \times \frac{1\text{ kb}}{\text{gene A length}} \times \frac{1\text{ M}}{\text{total mapped reads}}$$

However, it is possible that some reads align to more than one region of the reference genome and therefore its mapping is ambiguous. RPKMs may be corrected for multiple mapping by considering only the portion of the gene sequence with uniquely mappable positions - the gene's "effective length" (Labbé et al., 2012). These normalized gene counts are then comparable between different genes within the same sample and between distinct samples, enabling a more accurate analysis of gene expression. VAST- Tools output tables containing raw RNA-seq read counts and cRPKMs for all genes were used in the following sections of this analysis. Salsa gene (also named CG31368 or FBgn0051368) relative decreased expression in Salsa RNAi samples was validated.

**Equation 2.**

$$\mathbf{cRPKM}_{\text{gene A}} = \text{gene A reads} \times \frac{1\text{ kb}}{\text{Uniquely mapable positions in gene A}} \times \frac{1\text{ M}}{\text{Total mapped reads}}$$

### 3.2.14.3 Splicing quantification

Splicing quantification depends on the profiling of RNA-Seq reads aligning to splice junctions (i.e. junction reads). Splicing was analysed based on the exon-centered approach, where each potential inclusion of an exon (or other alternative sequence) in a final transcript is called an event and the exon is treated as an alternative exon. Each event is defined by the genomic coordinates of its limiting alternative splice sites and its quantification is dependent on the number of junction reads that map to them. For each event, the percent-spliced-in (PSI) or percent-of-intron-retention (PIR) is obtained as the ratio between the reads supporting the inclusion of the exon and the total number of reads attributed to that event, that is the sum of reads supporting inclusion and reads supporting exclusion, as explained in Equation 3 (Irimia et al., 2014; Labbé et al., 2012). For intron retention events VAST-Tools checks the balance between the numbers of exon- intron junction reads and reads that map to the respective intron body sequence (that also support intron retention/inclusion but are not used for quantification). Events with a high imbalance between these two groups of reads may be reflecting alternative splice site events in the vicinity of exon-intron junctions or events in overlapping genes rather than intron retention (Braunschweig et al., 2014). Constitutive sequences tend to have many more reads supporting inclusion and consequently its PSI/PIR values will tend to 100, while alternative sequences will have more variable PSI/PIR values, near zero if seldom included.

VAST-Tools output tables used in the following steps contain PSI/PIR values for each alternative splicing event that is part of the fly VASTDB (the VAST-Tools database of alternative splicing events), as well as raw RNA-seq junction read counts and an associated label that classifies the number of reads in terms of their abundance as one of the following options: N, VLOW, LOW, OK, SOK. Moreover, the P-value of an exact binomial test for the null hypothesis that exon-intron junction reads and intron body reads are balanced is provided.

**Equation 3.**

$$\textbf{PSI or PIR} = \frac{\text{inclusion reads}}{\text{exclusion reads + inclusion reads}} \times 100$$

### 3.2.14.4 Differential gene expression

Differential gene expression analysis was performed by linear modeling using the limma R package (Ritchie et al., 2015). In this analysis, the expression of each gene is fitted considering as baseline the expression level of that gene across the Salsa samples and calculating the increment in expression on the samples where Salsa is present (Controls) with respect to the baseline. Global empirical Bayes statistics (B) are calculated for the model, allowing the identification of genes significantly differentially expressed between conditions.

### 3.2.14.5 Differential intron retention

Differential intron retention between conditions was profiled by calculating mean PIR per condition (Salsa RNAi samples and Control RNAi samples) and ΔPIR (Equation 4) for each intron retention event in *Drosophila melanogaster* VASTDB. Only events with reads supporting PIR at least classified as VLOW (10 actual reads mapping to the sum of exon-exon junctions or 10 actual reads mapping to one of the two exon-intron junctions and 5 to the other) were considered.

**Equation 4.** $\Delta\textbf{PIR} = \textbf{PIR(Salsa RNAi)} - \textbf{PIR(Control RNAi)}$

Direct comparison between mean PIR per condition allowed the definition of groups of intron retention events differentially affected by the knockdown of Salsa, highlighting different levels of susceptibility to the loss of Salsa. Groups of events were defined using get_vast function from Matt, a toolkit for downstream alternative splicing data preparation and analysis. Events with PIR levels within the alternative range (PIR values within 5 and 95, contrary to

constitutive events, with PIR levels closer to 0 or to 100) and showing increased PIR values in Salsa RNAi samples when compared to Control samples were labeled as *Salsa affected*. On the other hand, events with PIR values within the alternative range but that show differences between samples not greater than 5 are considered *Salsa controls*. Constitutive events were also selected as two other subsets of control events: *Skipped* events (with PIR values of 0 or 1) and *Included* events (with PIR values of 99 or 100), both with a ΔPIR of 0. Finally, events for which intron retention percentage is lower in Salsa RNAi samples compared to Control samples were defined simply as *Down*. Criteria used for definition of these 5 groups are compiled in points marked with crosses represent a previously considered subset of events with high coverage and a minimum ΔPIR of 10.

### 3.2.15 Protein extraction

0–1hr embryos (after egg laying) or 2-3 days old adult female ovaries (after pupae eclosion), whose female germ line was specifically depleted for Salsa (salsa RNAi) or mCherry (negative control: mCherry RNAi), were homogenized in a NB-lysis buffer containing 50mM Tris (pH 7.5), 150mM NaCl, 2mM EDTA, 0.01% NP-40 (Igepal) 2mM DTT, 10 mM NaF, and protease inhibitors (Complete$^{TM}$ protease inhibitor cocktail, Roche). The sample was centrifuged three times at 14000rpm, at 4°C, and for 5 minutes. Between each centrifugation, the supernatant was carefully collected into a new tube, avoiding the pellet and the upper lipid-rich layer. Bradford protein microassay (Bio-Rad) was performed to quantify the total amount of protein, and whenever needed, the total protein extract was frozen with liquid nitrogen and kept at -80°C until further usage.

### 3.2.16 Western-blot analysis

For western blot analysis, total protein extracts were boiled for 5 min at 95 $^{O}$ C in 2xLaemmli buffer (Sigma), loaded (10µg of protein/ per lane), and run onto a SDS polyacrylamide gel electrophoresis (6% or 12%). Proteins were subsequently transferred onto Hybond-ECL nitrocellulose membrane (Amersham) using a Bio-Rad wet transfer apparatus (100V for 60 min; with

agitation). Western blotting was performed using standard procedures. Briefly, the Hybond-ECL membrane was blocked overnight with 5% non-fat milk/PBT (0.1% Tween20, 1x PBS) at 4°C with agitation. Primary antibodies were incubated overnight in 1% non-fat milk/PBT with gentle agitation at 4 °C. Membranes were washed three times with PBT, for 15min at room temperature and with agitation. Secondary antibodies were incubated for 2 hours in 1% non-fat milk/PBT at the room temperature with gentle agitation. After extensive washes (3 times, 5min washes in PBT, and a quick wash in PBS) proteins of interest were detected with an ECL Plus western blotting detection system (GE Healthcare) and an ECL Hyperfilm (Amersham).

### 3.2.17 Co-immunoprecipitation

Co-immunoprecipitation was done using protein extracts from *Drosophila* S2 cells expressing C-terminal Myc-tagged Salsa or control (empty Myc-plasmid). Briefly, 1 mg of total protein extract was diluted in 1ml NB-lysis buffer and was incubated with anti-c-Myc Magnetic beads (Invitrogen, Grand Island, NY, USA) for 1 hr at 4 °C. Anti-c-Myc Magnetic beads were then washed three times with NB-lysis buffer, and boiled for 5 min at 95 $^{O}$ C with 2xLaemmli buffer (Sigma) for further western-blot analysis.

### 3.2.18 Expression of C-terminal Myc-tagged Salsa

#### 3.2.18.1 Cloning of Myc-tagged Salsa

For cloning of a C-terminal Myc-tagged Salsa, the open reading frame of CG31368 was amplified by Expand High Fidelity PCR System (Roche) from the cDNA clone RE 35509 (BDGP). Please note that a point mutation present in the original cDNA clone RE35509 was corrected by site-directed mutagenesis using the publically available Flybase sequence of CG31368 as reference. The QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Catalog no. – 200518) was used to correct point mutations, according to manufacturer's instructions. The primers used in correction listed in primer list (Supplementary Table- 3). Correction in Salsa CDNA was further confirmed by sequencing. To further amplify the corrected cDNA high fidelity pcr system (Roche) were used and the PCR condition: Initial denaturation for

1 cycle (95°C for 2 minutes), primary amplification programme repeated for 10 cycles (denaturation: 94°C for 15 seconds, annealing: 55°C for 30 seconds and elongation: 68 for 4 minutes), secondary amplification programme repeated for 10 cycles (denaturation: 94°C for 15 seconds, annealing: 55°C for 30 seconds and elongation: 72 for 4 minutes), final elongation for I cycle (72°C for 10 minutes) and cooling at 4°C for unlimited time. The PCR product was cloned into pAMW plasmid (Gateway) using the Gateway system (Invitrogen).

#### 3.2.18.2 Culture of *Drosophila* S2 cells

*Drosophila* Schneider 2 (S2) cells were cultured in Schneider's *Drosophila* complete medium: Schneider's insect medium (Sigma), supplemented with 1x L-glutamine, 1x PenStrep, and 10% Fetal bovine serum (Invitrogen) at 25°C.

#### 3.2.18.3 Transfection of Myc-tagged Salsa

Transient transfection in *Drosophila* S2 cells with a C-terminal Myc-tagged Salsa construct was done using FuGENE® HD Transfection Reagent and standard procedures (Promega). Briefly, a mix of 100 µL of Serum Free Medium (SFM), 400 ng of DNA, and 4 µL of Fugene HD were incubated at the room temperature for 15 min. Meanwhile, cells were plated into 6 well plates at the concentration of $2.5 \times 10^6$ cells/well in serum free medium. The mix (DNA and Fugene) was added drop by drop on the top of cells. To stop the transfection reaction, 2ml complete medium (with 10% Fetal bovine Serum) was added to the cells after 4 hours of incubation at 25°C and cells were incubated for 48 hours at 25°C.

## 3.3 Results and discussion

### 3.3.1 CG31368/ Salsa is a highly conserved RNA helicase that associates with the *Drosophila* NTC/Prp19 complex

As described in Chapter 2, CG31368 was identified in a *Drosophila* screen as gene whose germ line depletion during oogenesis resulted in abnormal dorsal appendages of the eggshell. In order to clarify the function of CG31368 in *D. melanogaster*, the protein sequence of CG31368 was used in a bi-directional BLASTp analysis (NCBI and Flybase) against the Homo sapiens proteome, showing that CG31368 encodes the *Drosophila* orthologue of a highly conserved human RNA helicase and spliceosome NTC (NineTeen Complex) subunit named Aquarius. In *Drosophila* the name Aquarius refers to an unrelated gene, and in order to avoid additional confusion, we decided to refer to CG31368 as salsa (as we previously observed that Salsa physically interacts with Fandango in the embryo (Guilgur et al., 2014) ; another highly conserved NTC subunit. To further validate this analysis, we identified the orthologs in five other eukaryotes on the basis of E-value of reciprocal BLASTp best-hit search and performed multiple sequence alignment (Results are not shown). As expected, and suggesting functional conservation as an active helicase, the RNA helicase domains are highly conserved (Figure-2.4)

Previous Liquid Chromatography Mass Spectrometry (LC/MS) analysis of immunoprecipitated fractions from Myc-tagged Fandango and Myc-tagged Prp19 showed a clear enrichment for Salsa, both in ovaries and embryonic protein extracts (Figure-3.1.1; adapted from (Guilgur et al., 2014). For further confirmation of these interactions, we immunoprecipiated the Myc-tagged Salsa from *Drosophila* S2 cells and analyzed the interaction of Salsa with endogenous Fandango and Prp19. Consistent with our previous result, we observed a binding of Fandango and Prp19 with Myc-tagged Salsa, and further showed that this interaction is independent of RNA, further confirming that Salsa is a *bona fide* subunit of *Drosophila* NTC/Prp19 complexes (Figure-3.1.2)

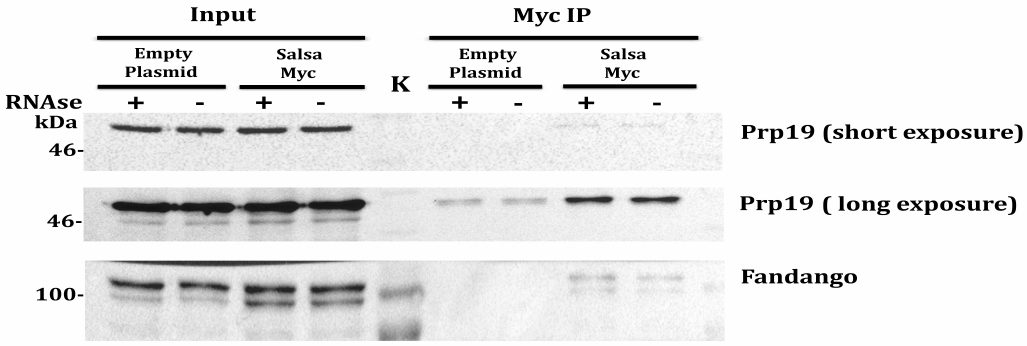
Previously we observed that depletion of NTC subunit Fandango was associated with a significant loss of NTC complex integrity, being the steady-

state stability of some of its subunits reduced (Guilgur et al., 2014). We therefore decided to investigate whether the integrity NTC complex was similarly affected after the depletion of Salsa. To do this, we analyzed the steady-state stability of several NTC subunits after depletion of Salsa, in the *Drosophila* ovaries and embryos. We failed to detect any obvious change to the total protein levels of the tested NTC subunits (Figure-3.1C). Altogether, these data suggest that Salsa is a highly conserved RNA helicase, which physically interacts with the *Drosophila* NTC complex but that is apparently not required for complex integrity.

**A.** **LC-MS analysis of co-immunoprecipitation assays from ovaries and embryos (Adapted from Guilgur et al. 2014 )**

| | | | Fandango-Myc | | | | Prp19-Myc | |
|---|---|---|---|---|---|---|---|---|
| | | | Ovaries | | Embryo | | Embryo | |
| Drosophila | | Human/Yeast | Rep1 | Rep2 | Rep1 | Rep2 | Rep1 | Rep2 |
| CG6197 | Fandango | Xab2/Syf1 | +++ | +++ | +++ | +++ | + | + |
| CG31368 | Salsa | AQR/- | +++ | +++ | +++ | +++ | + | + |
| CG5519 | Prp19 | PRP19/Prp19 | + | + | + | + | +++ | ++ |
| CG6905 | cdc5-like | CDC5L/Cef1 | + | + | + | + | +++ | ++ |

**B.** **Immunoprecipitation of Salsa-Myc (S2 cells)**




**C.** **Protein levels of Prp19 and Fandango after depletion of Salsa (ovaries and embryos)**

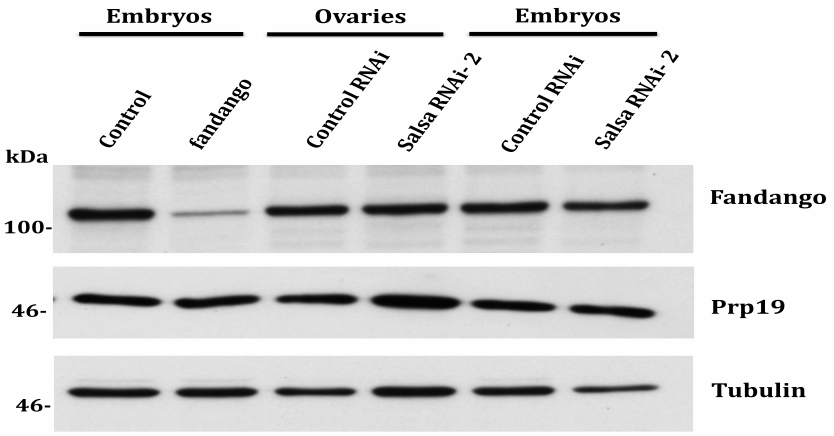



**Figure 3.1: *Drosophila* Salsa physically interacts with spliceosome NineTeen Complex (NTC) during oogenesis and embryogenesis. (A)** Salsa was efficiently immunoprecipitated with Myc-tagged Fandango and Myc-tagged Prp19 (both NTC subunits) expressed in *Drosophila* ovaries and embryos Adapted from (Guilgur et al., 2014). Immunoprecipitated complexes were analyzed by Liquid chromatography–mass spectrometry (LC-MS). Symbols (+), (++), (+++) corresponds to 1–9, 10–19, and >20 non-repeated peptides respectively. Rep1 and Rep2 correspond to biological replicas 1 and 2, respectively. None of the proteins shown were detected in the negative controls. Human and yeast homologues and the different sub-complexes are shown as described in (Herold et al., 2009). **(B)** Endogenous NTC proteins Fandango and Prp19 were efficiently co-immunoprecipitated by Salsa-Myc in

an RNA-independent manner. Co-immunoprecipitation from total protein extracts from *Drosophila* Schneider 2 (S2) cells expressing Myc-tagged Salsa (Salsa-Myc; N-terminal tag) and using anti-Myc coated dynabeads. Protein extracts from *Drosophila* S2 cells transfected with an empty plasmid were used as a negative control. Lysate were immunoprecipitated RNAse treatment (+) and without treatment (-). **(C)** Total protein levels of endogenous Fandango and Prp19 did not change after depletion of Salsa. Western-blot analysis of total protein extracts from ovaries (left; dissected from 3-5 days old females) and 0–3 hr embryos (right) from control and salsa RNAi (Salsa RNAi-2). A *fandango* mutant allele was used as a positive control for reduction of Fandango protein levels. α-Tubulin was used as a Western-blot loading control. Female germ-line specific expression of dsRNAs was obtained by using the UAS/Gal4 system and the germ-line-specific Nanos Gal4 driver.

### 3.3.2 Salsa required for female fertility and dorsal-ventral patterning of *Drosophila* egg

Germ line depletion of Salsa during oogenesis is apparently associated with abnormal eggshell dorsal appendages, which is suggestive of dorsal–ventral patterning defects. To confirm this phenotype, and better understand the function of Salsa during female oogenesis, we quantified the observed eggshell dorsal appendages defects. To do this, we divided the observed phenotypes in four different classes: i) wild type dorsal appendages, ii) partially merged dorsal appendages (merged at bottom) , iii) complete merged dorsal appendages, and iv) short or extremely dorsal appendages (Figure-3.2A) (Supplementary Figure-3.1). Almost all eggs laid by control females (mCherrry RNAi) showed wild type dorsal appendages, whereas a significant proportion of the eggs laid by females whose germ line were depleted for Salsa (Salsa RNAi; Salsa RNAi-2) showed highly abnormal (fused or partially-fused) dorsal appendages defects (Figure-3.2B) (for more experimental detail see methods).

To confirm that the observed eggshell dorsal appendages defects were due to Salsa depletion and not a consequence of an off-target effect of the used Salsa RNAi construct (Salsa RNAi-2), we designed two additional non-overlapping short hairpin RNA (shRNAs) constructs against salsa mRNA (Salsa RNAi-1 and Salsa RNAi-3). Both tested non-overlapping salsa shRNAs gave rise to qualitatively similar phenotypes, with significant defects in dorsal

appendages (Figure-3.2B). Squid RNAi was used as positive control (Figure-3.2B), as depletion of this protein during oogenesis was previously described as being associated to eggshell dorsal appendages and dorsal-ventral (DV) patterning defects (Norvell et al., 1999).

In the absence of an antibody capable of detecting endogenous Salsa, and in order to confirm Salsa RNAi depletion, the salsa mRNA levels in *Drosophila* ovaries and embryos/ unfertilized mature eggs were analyzed by RT-qPCR. As expected there was a significant reduction of salsa mRNA levels in both samples after Salsa RNAi (Figure-3.2D). salsa mRNA depletion in the embryos/ unfertilized mature eggs was significantly stronger than the one observed in the ovaries (Supplementary Figure-3.2A). This was most likely due to the fact that Salsa is ubiquitously expressed in the soma (modEncode project), we used a germ line specific Gal4 driver (Nanos-Gal4) to express the salsa RNAi construct, and the dissected ovaries necessarily contain supporting somatic cells. Although the observed phenotypes with the three non-overlapping salsa RNAi constructs were qualitatively identical, their phenotypic penetrance/ expressivity was nevertheless variable (Figure-3.2B). This was most likely due to distinct efficiencies of the used salsa shRNAs.

In order to investigate if Salsa was important for fertility, we counted egg hatching after germ line depletion of Salsa (Figure-3.2C). Females whose germ-line was depleted for salsa (Salsa RNAi) showed a dramatic reduction in fertility with egg hatching as low as 5%( n= 1123). Control females (mCherry RNAi) egg hatching was 98% (n = 1216).

Altogether these results suggest that Salsa is required within the germ line for female fertility and the correct development the eggshell dorsal appendages, possibly because it is important for DV patterning of the developing oocyte.

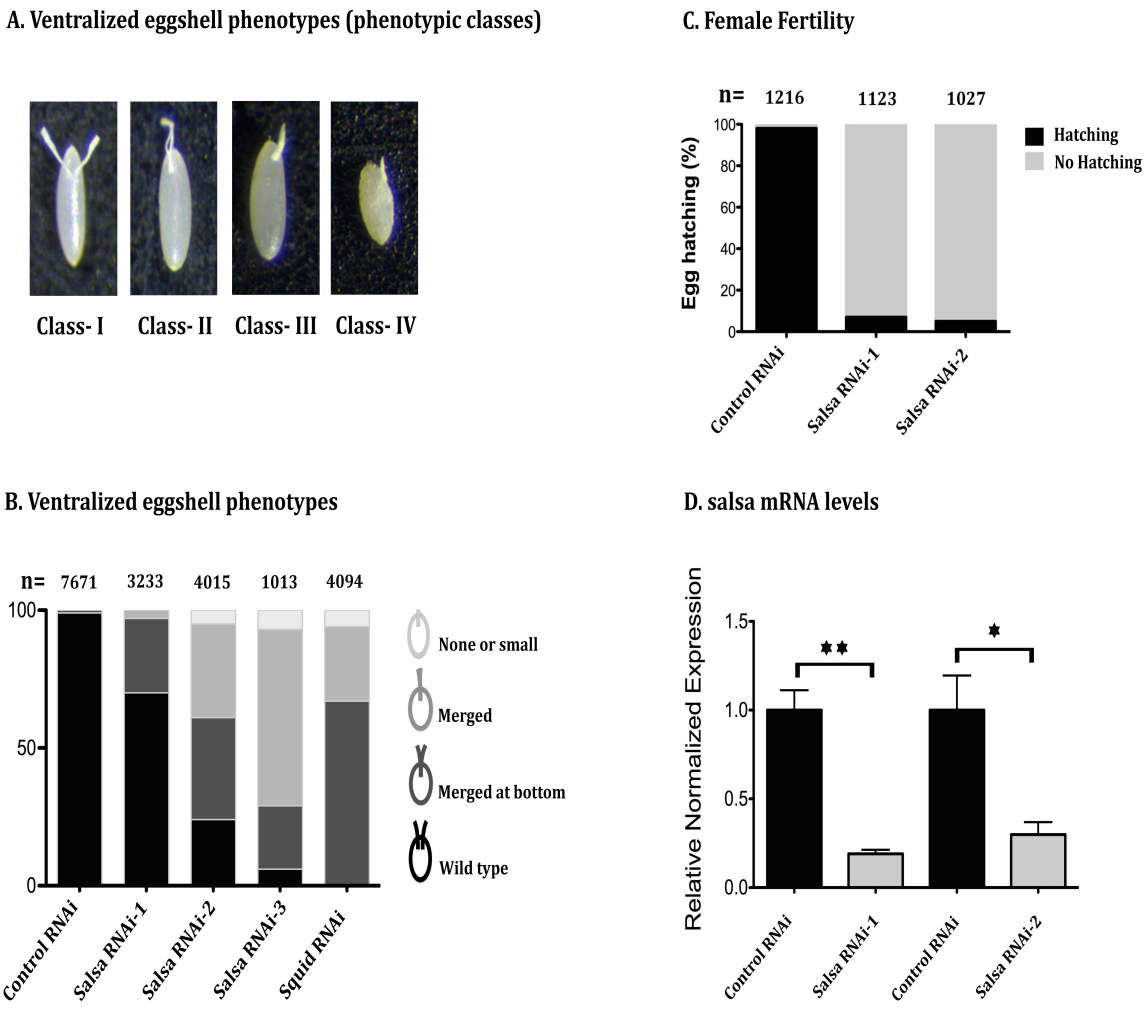
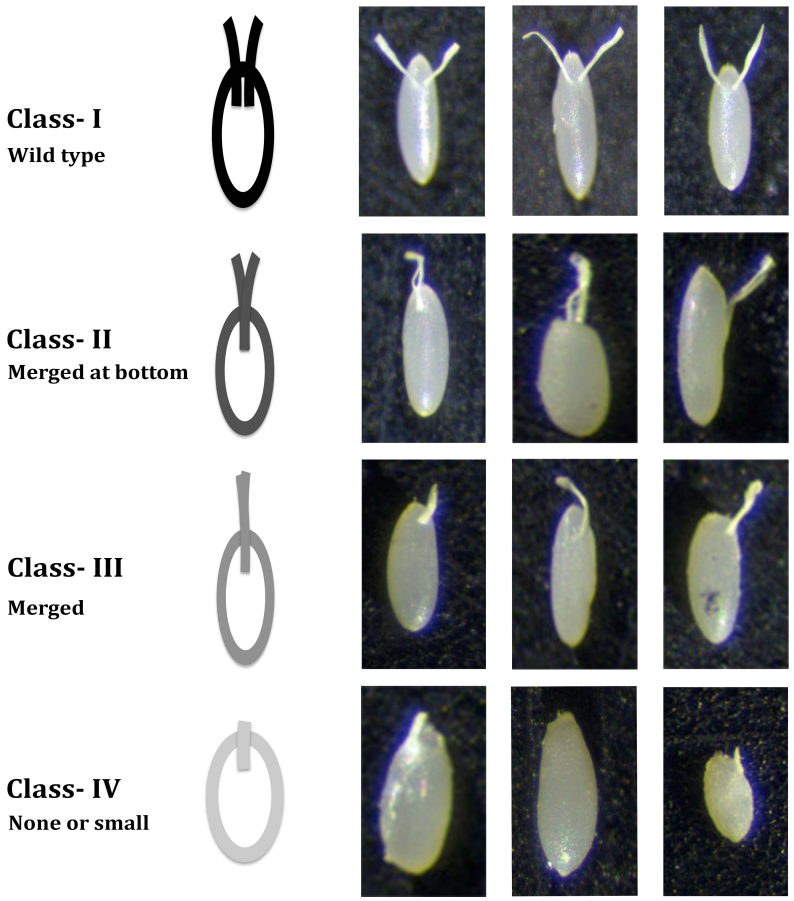



**Figure 3.2: *Drosophila* Salsa is required for eggshell dorso-ventral patterning and female fertility. (A)** Female germ line-specific depletion of Salsa significantly impaired dorsal-ventral patterning of the eggshell. Observed phenotypes were categorized in four different phenotypic classes based on the eggshell dorsal appendages: Class-I (wild type appendages; two individualized dorsal appendages); Class-II (appendages fused at their bottom); Class-III (appendages totally fused; spindle phenotype); and Class-IV (short eggs without or with short dorsal appendages). Additional examples for each phenotypic class are shown in Supplementary Figure-3.1. **(B)** Quantification of ventralized eggshell phenotype after germ line-specific expression of double-stranded RNAs (dsRNAs) hairpins against mCherry (negative control RNAi), salsa (three non-overlapping dsRNAs: Salsa RNAi-1, Salsa RNAi-2, and Salsa RNAi-3) and squid (positive control RNAi; Squid RNAi). Number of eggs analyzed for each experiment are indicated above the bar plot. **(C)** Quantification of female fertility (embryonic hatching) after female germ line-specific expression of dsRNAs against mCherry (negative control RNAi) and salsa (two non-overlapping dsRNAs: Salsa RNAi-1 and Salsa RNAi-2). Female fertility is defined by the frequency of egg hatching 48 hours post oviposition. Number of embryos analyzed for each experiment are indicated above the bar

plot. **(D)** Real-time qPCR analysis of the total steady-state levels of salsa mRNA in the ovaries after germ line-specific expression of dsRNAs against mCherry (negative control RNAi) and salsa (two non-overlapping dsRNAs: Salsa RNAi-1 and Salsa RNAi-2). Relative normalized expression corresponds to values normalized with two distinct reference genes (β-actin and GAPDH) and relative to control RNAi (mCherry RNAi). At least three biological replicates were used for all shown datasets. Error bars indicate standard deviation. Ovaries were dissected from 3-5 days old females (after pupae eclosion). Female germ-line specific expression of dsRNAs was obtained by using the UAS/Gal4 system and the germ-line-specific Nanos Gal4 driver.




**Supplementary Figure- 3.1: Classes of eggshell dorso-ventral patterning defects.** Dorsal–ventral patterning of the eggshell was categorized in four phenotypic classes: Class-I (wild type appendages; two individualized dorsal appendages); Class-II (appendages fused at their bottom); Class-III (appendages totally fused; spindle phenotype); and Class-IV (short eggs without or with short dorsal appendages).

### 3.3.3 Salsa is particularly rate limiting for splicing of the first intron of gurken mRNA

Dorsal-ventral (DV) patterning of the *Drosophila* egg is established during oogenesis and Gurken signaling plays an essential role in the establishment of axis formation of the developing embryo. Salsa is highly conserved subunit of the NineTeen complex, which is known to be required for efficient activation of the spliceosome. These observations lead us to raise the hypothesis that depletion of Salsa leads to an abnormal splicing of gurken transcript during oogenesis.

The gurken gene encodes a single transcript with three introns (Neuman-Silberberg and Schupbach, 1994). To test our splicing hypothesis, we performed RT-PCR using the exonic flanking primers for each intron of gurken transcript (open arrows shown in Figure-3.3A). Interestingly, we observed that the first intron (that localizes to the transcript 5´UTR) was particularly sensitive to Salsa depletion, with detectable levels of intron retention, whereas no detectable differences could be observed for the second and third introns (Figure-3.3B). We could also observe some degree of retention of the first intron in the control RNAi sample (Figure-3.3B), suggesting that it could be comparatively inefficiently spliced even in wild type embryos.

To quantitatively define the degree of gurken intron retention after depletion Salsa, we decided to perform RT-qPCR analysis of gurken transcript. Specific primers flanking the 3´splice site of all three introns of gurken were designed so that signal amplification would only occur in the presence of the intron (black arrows shown in Figure-3.3A). Sequence of all primers is listed in Table 3.

Consistent with our previous results, we observed a significant retention of the first intron of gurken after germ line depletion of Salsa (Salsa RNAi-1 and Salsa RNAi-2) when compared with the control RNAi (more than six fold intron retention for Salsa RNAi-2) (Figure- 3.3C', C''), whereas no significant increase could be detected for the second and third introns (Figure- 3.3C', C''). These data further confirmed by RNA-seq analysis from the ovaries depleted with Salsa (Supplementary Figure- 3.5). Interestingly, the maternally

loaded gurken transcripts similarly showed a significant first intron retention in the early embryo/ unfertilized mature eggs after Salsa depletion (Supplementary Figure- 3.2C). This further confirmed that Salsa is particularly rate limiting for splicing of the first intron of gurken transcript and suggested that such retention is not likely to elicit transcript degradation by nuclear surveillance machinery (Braunschweig et al., 2013b) or nonsense-mediated decay (NMD) (Wong et al., 2013). Consistently, total levels of gurken transcript levels in the ovaries and in the embryos / unfertilized mature eggs were not significantly affected after depletion of Salsa (Supplementary Figure- 3.2B).

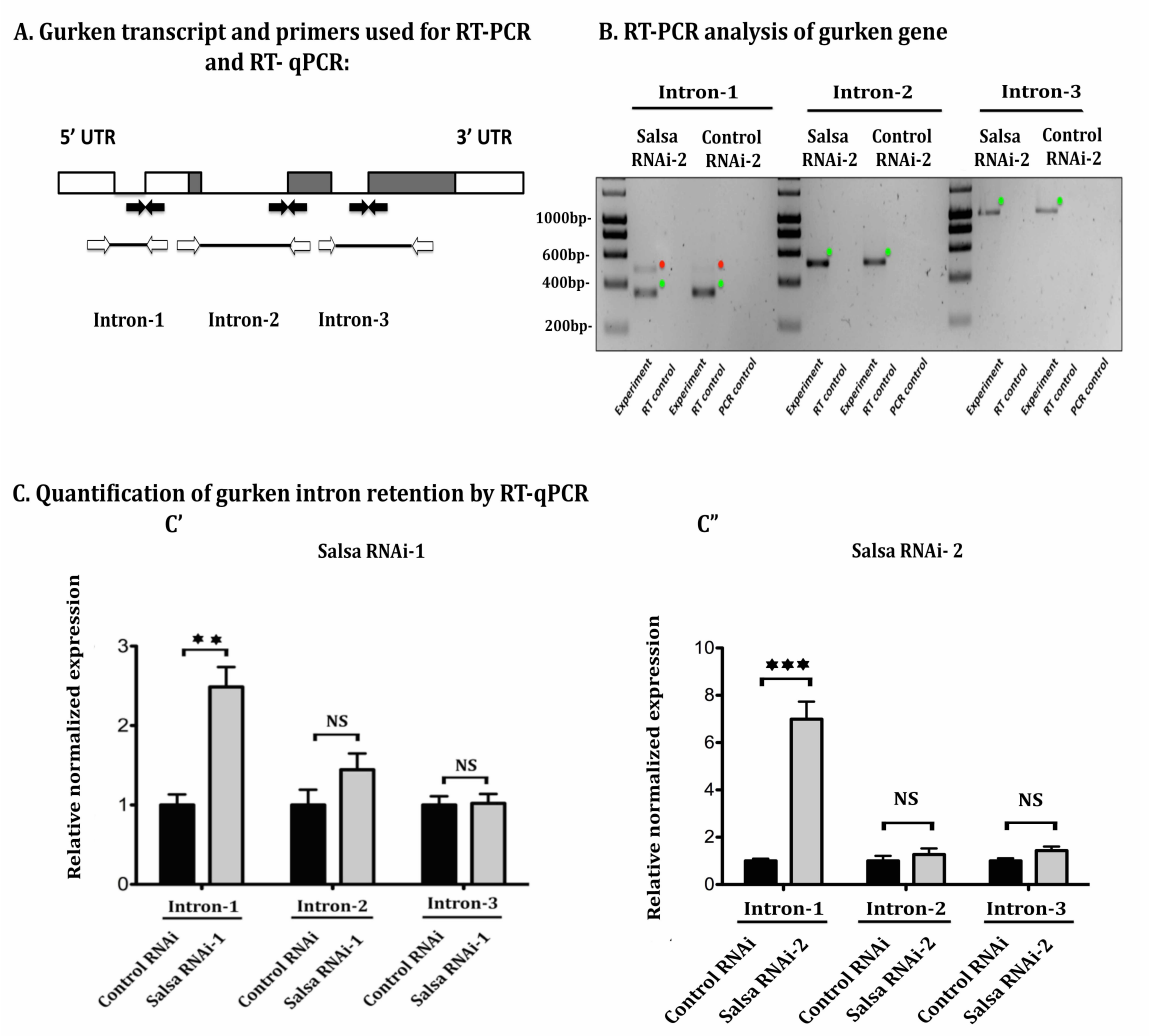
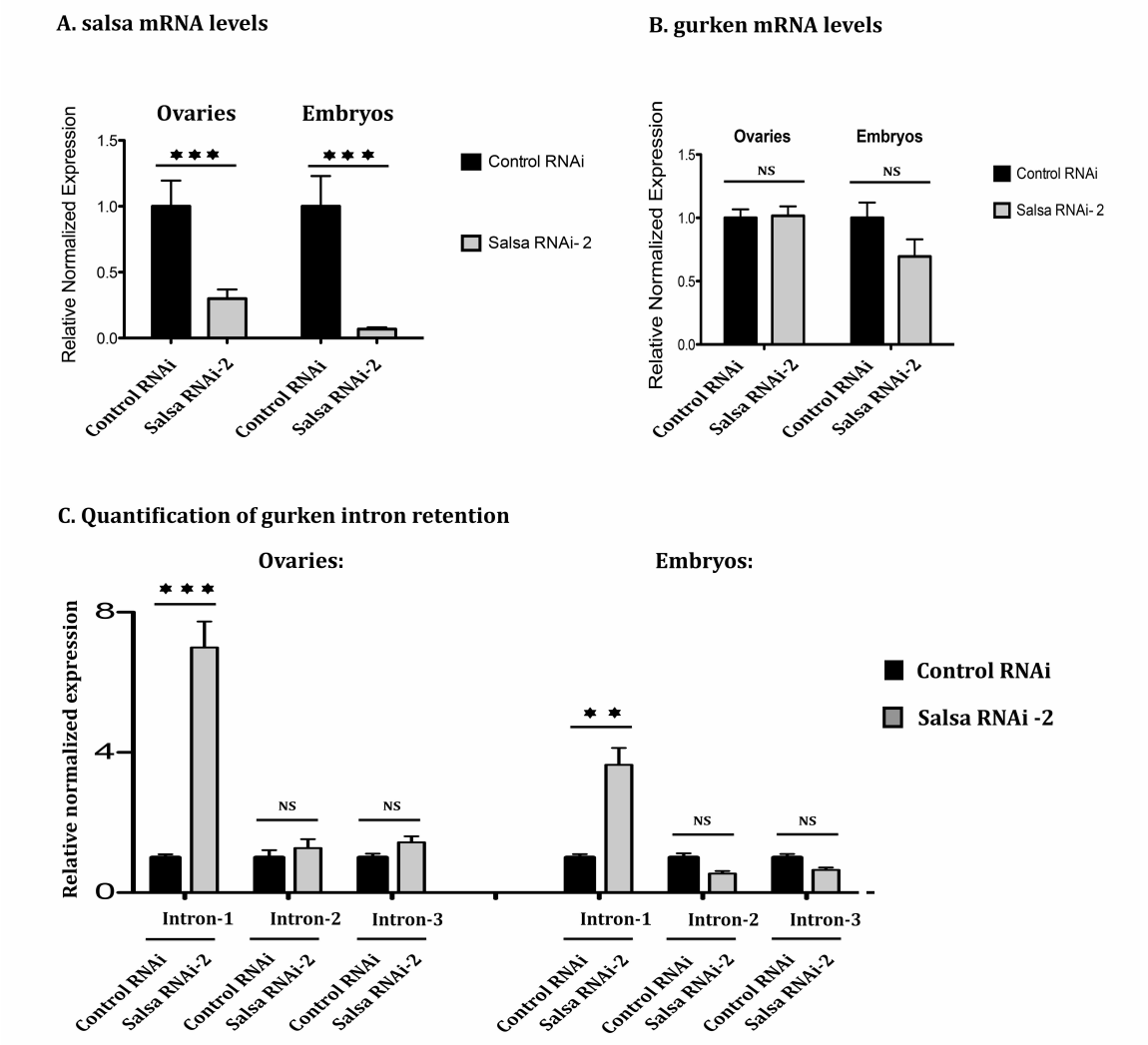



**Figure 3.3: Splicing of the first intron of gurken mRNA is particularly sensitive to depletion of Salsa. (A)** The gurken gene encodes a single transcript with three introns. 5′

and 3´UTRs are shown in white, whereas the open-reading frame is shown in grey. Black arrows indicate primers used for RT-qPCR of gurken, whereas white arrows (and the black line between them) indicate the primers used for RT-PCR of gurken**. (B)** RT-PCR analysis of gurken transcripts from iScipt cDNA library (random hexamer and oligo (dT) reaction mix) and using exon–exon (e–e) primers. Control ovaries yielded PCR products in the size predicted for correctly spliced gurken transcripts (green dots, *grk1*: 363 bp; *grk2*: 528 bp; grk3: 882 bp). DNA sequencing of the isolated bands confirmed correct splicing of gurken transcript. Ovaries whose germ-line was depleted for Salsa (salsa RNAi 1; Salsa RNAi-2) showed increased splicing defects (intron retention) of the first intron of gurken transcript (red dots, grk: 505bp), whereas the second and third introns were apparently correctly spliced without detectable intron retention (green dots, grk2: 528 bp; grk3: 882 bp). PCR and RT control used in all reactions. 'RT controls' (absence of reverse transcriptase (RT) reaction) yielded no amplification, meaning there was no contamination with genomic DNA in all samples tested. **(C)** Quantitative analysis of gurken transcript intron retention by RT–qPCR using intron-exon primers. Shown are the fold changes of unspliced gurken transcripts for each intron. Significant increase in gurken first intron retention (but not of the second and third introns) after depletion of Salsa. **C'** and **C´´**corresponds, respectively, to depletion of Salsa depletion using Salsa RNAi-1 and Salsa RNAi-2 hairpins. Relative normalized expression corresponds to values normalized with two distinct reference genes (β-actin and GAPDH) and relative to control RNAi (mCherry RNAi). At least three biological replicates were used for all datasets. Error bars indicate standard deviation. Ovaries were dissected from 3-5 days old females (after pupae eclosion). Female germ-line specific expression of dsRNAs was obtained by using the UAS/Gal4 system and the germ-line-specific Nanos Gal4 driver.




**Supplementary Figure- 3.2: Depletion of Salsa impairs splicing of the first intron of gurken mRNA without eliciting a significant mRNA degradation. (A)** Real-time qPCR analysis showed a significant reduction of salsa mRNA levels in ovaries and in embryos depleted for Salsa (Salsa RNAi-2 hairpin). **(B)** Real-time qPCR analysis showed no significant reduction of gurken mRNA levels in ovaries and in embryos depleted for Salsa (Salsa RNAi-2 hairpin). Results shown for gurken mRNA levels in the ovaries are the same shown in Figure 3.4C. **(C)** Real-time qPCR analysis using intron-exon primers showed a significant retention of the first intron of gurken transcript (but not the second and the third introns) in ovaries and in embryos depleted for Salsa (Salsa RNAi-2 hairpin). For all panels, relative normalized expression corresponds to values normalized with two distinct reference genes (β-actin and GAPDH) and relative to control RNAi (mCherry RNAi). At least three biological replicates were used for all ovaries datasets (two biological replicates for embryos). Error bars indicate standard deviation. Embryos were collected 0-1 hours after deposition and ovaries were dissected from 3-5 days old females (after pupae eclosion). Female germ-line specific expression of dsRNAs was obtained by using the UAS/Gal4 system and the germ-line-specific Nanos Gal4 driver.

### 3.3.4 Salsa is required for anterior dorsal localization of gurken mRNA

*half pint* (*hfp*), the *Drosophila* ortholog of human PUF60( poly(U) binding splicing factor 60), is required for splicing of distinct transcripts during oogenesis (Van Buskirk and Schüpbach, 2002b). Mutants for *hfp* show defects in the cystoblast mitotic divisions, the karyosome is morphologically abnormal, the nurse cells fail to undergo normal dispersal and remain polytene, and the eggs are short and show fused dorsal appendages (Van Buskirk and Schüpbach, 2002b). Oocytes mutant for *hfp* also show DV patterning defects with abnormal localization of gurken transcript. Interestingly, and although Hfp is specifically required for efficient splicing of the third intron of gurken transcript, this defect is not likely to cause its abnormal localization as the unspliced transcript is likely to have all the motifs necessary for its correct localization and previously it was shown that insertion of a lacZ fragment within the gurken cDNA in a position analogous to the third intron did not affect its localization (Thio et al., 2000). Hfp is also required for the correct splicing of *ovarian tumor* (*otu*). Interestingly, ectopic expression of an otu cDNA trangene, under the control of its endogenous promoter, rescued not only the nurse cells nuclear morphology phenotype but also the gurken transcript anterior dorsal localization defects (Van Buskirk and Schüpbach, 2002b).

The 5'UTR of gurken has been previously described as being important for the correct anterior dorsal localization of this transcript (Saunders and Cohen, 1999b). Since we observed that depletion of Salsa was specifically associated with splicing defects of the first intron of the gurken mRNA, which localizes to its 5´UTR, we hypothesized that the observed fused dorsal appendages were potentially due to defects in the anterior dorsal localization of gurken transcript during oogenesis.

To determine if depletion of Salsa was associated with an abnormal localization of gurken transcript, we carried out fluorescence *in situ* hybridization with antisense RNA probes against gurken mRNA. At mid-oogenesis (stage 6/7) gurken mRNA was mostly localized in the posterior region of the oocyte, but at later stages (stage 8/9), gurken mRNA expression

was restricted to cytoplasmic perinuclear anterior dorsal region of the oocyte (Klovstad et al., 2008; Neuman-Silberberg and Schupbach, 1994). Consistent with our hypothesis, depletion of Salsa (Salsa RNAi-1 and Salsa RNAi-2) was associated with defects in the anterior dorsal localization of gurken transcript (Figure-3.4A, B). Interestingly, the increased severity of the DV localization defects of gurken transcript in Salsa RNAi-2 ovaries (when compared to Salsa RNAi-1) correlates with the increased levels of first intron retention (Figure-3.3C', C''). Importantly, depletion of Salsa did not significantly impaired the total levels of gurken mRNA in the ovaries (Figure-3.4C).

Similar to *hfp* mutants, nurse cells depleted for Salsa (Salsa RNAi-2) also showed morphologically abnormalities, as they consistently failed to undergo normal dispersal and remain polytene (Figure-3.4A). Although it is still possible that the observed anterior posterior defects in the localization of gurken mRNA are similarly mediated by *otu* mRNA splicing defects this is nevertheless not likely as depletion of both splicing proteins (Otu and Salsa) specifically affected splicing of different introns within the gurken transcript (respectively, the third and the first), egg chambers depleted for Salsa showed a normal morphology of the oocyte karyosome (contrary to Otu mutants) (Supplementary Figure-3.4), and depletion mediated with Salsa RNAi-1 showed DV patterning defects without detectable (or at least with only mild) defects in nurse cells morphology (Figure-3.4A).

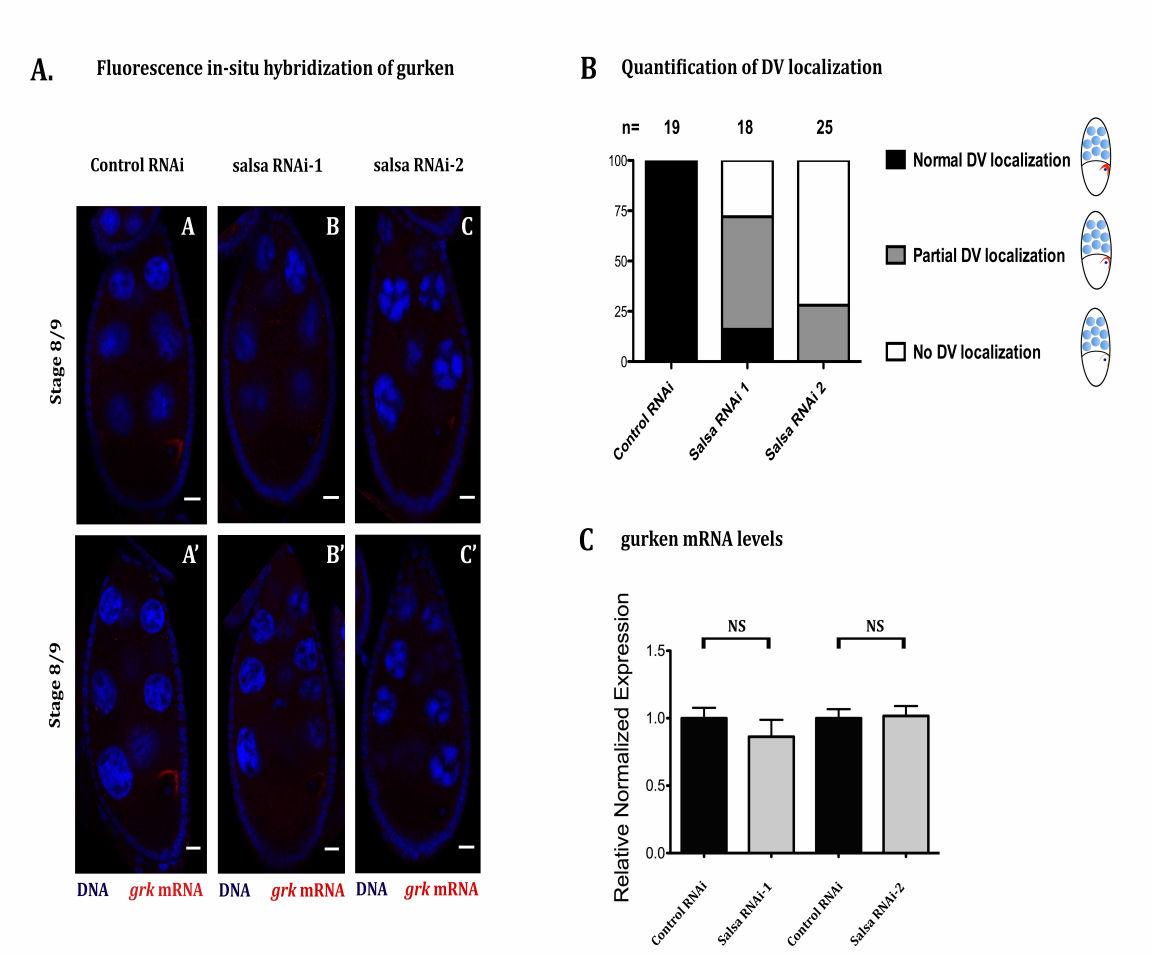
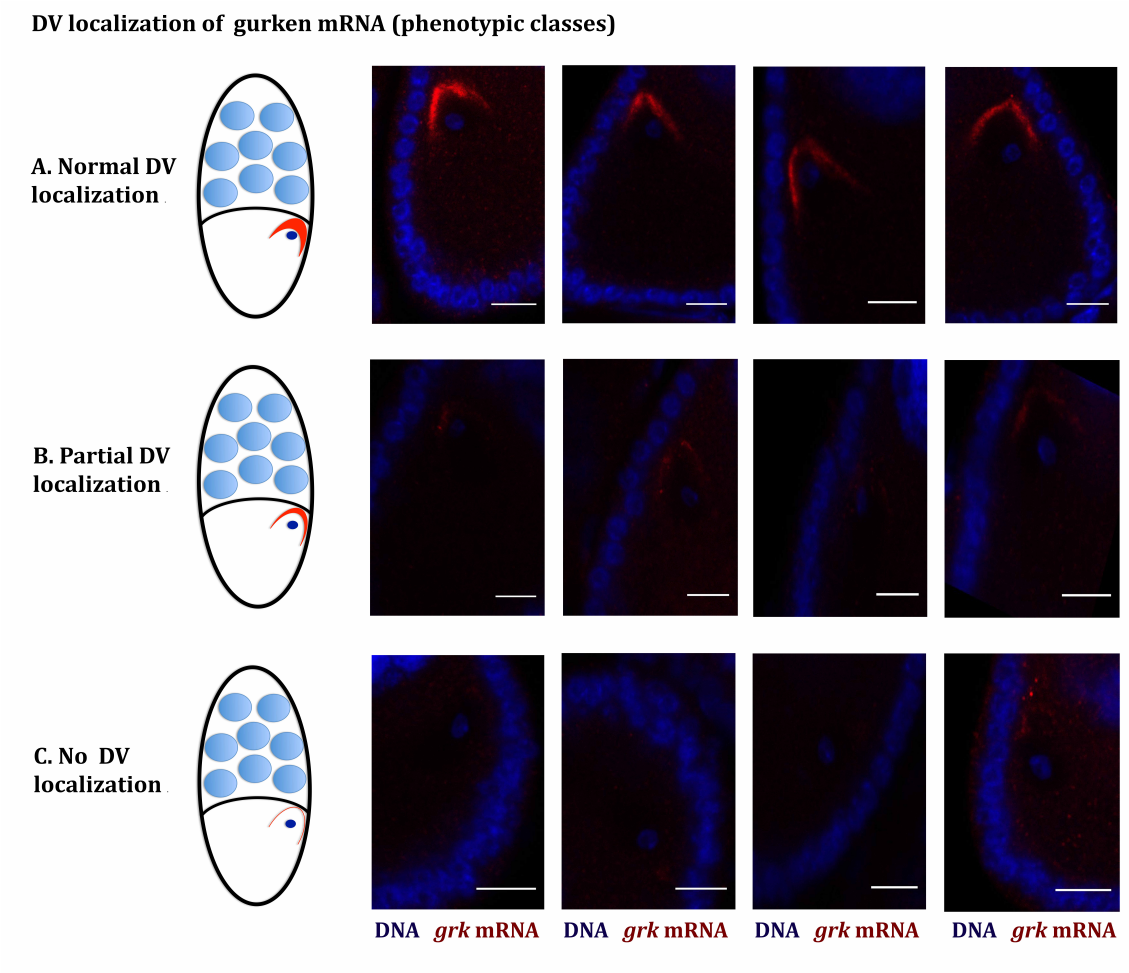


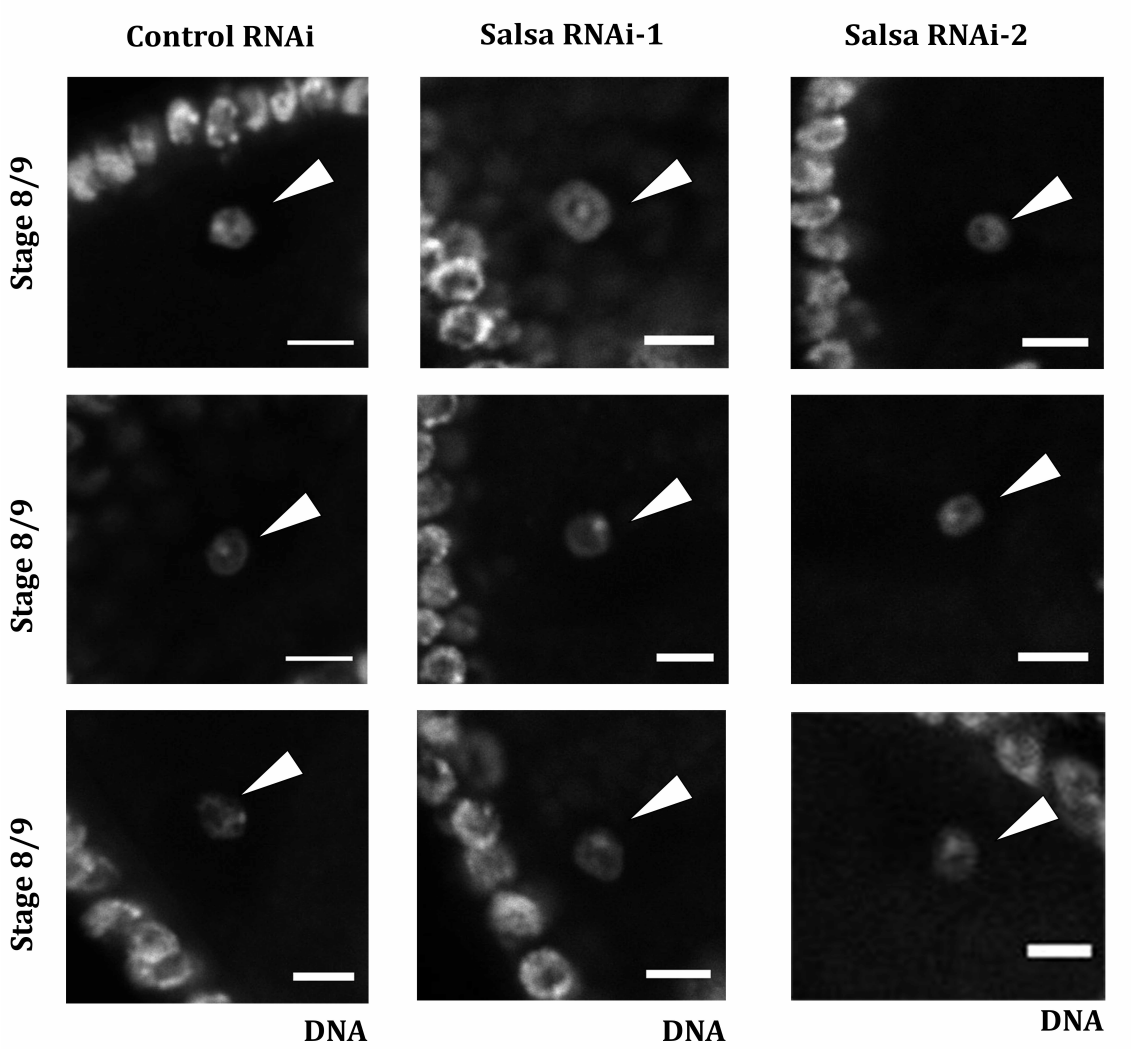

**Figure 3.4: Salsa is required for anterior dorsal localization of gurken mRNA. (A)** Fluorescent *in situ* hybridization showed that female germ-line specific of depletion of Salsa (Salsa RNAi-1 and Salsa RNAi-2) significantly impaired anterior dorsal localization of gurken mRNA during oogenesis. **(B)** Quantification of anterior dorsal localization defects of gurken mRNA in stage 8/9 egg chambers using three phenotypic classes: "normal DV localization", "partial DV localization" and "no DV localization". Additional examples for each phenotypic class are shown in Supplementary Figure-3.3. Control RNAi (mCherry RNAi) = 100% "normal DV localization" (n=19); Salsa RNAi-1 = 16% "normal DV localization", 56% "partial DV localization", and 28% "no DV localization" (n=18); Salsa RNAi-2 = 0% "normal DV localization", 28% "partial DV localization", and 72% shows "no DV localization". **(C)** Real-time qPCR analysis detected no significant reduction of total levels gurken mRNA after depletion of Salsa (Salsa RNAi-2 and Salsa RNAi-2). Results shown for gurken mRNA levels (control RNAi and Salsa RNAi-2) are the same results shown in Supplementary Figure 3.2B. Relative normalized expression corresponds to values normalized with two distinct reference genes (β-actin and GAPDH) and relative to control RNAi (mCherry RNAi). At least three biological replicates were used for all datasets. Error bars indicate standard deviation. Ovaries were dissected from 3-5 days old females (after pupae eclosion). Female germ-line specific expression of dsRNAs was obtained by using the UAS/Gal4 system and the germ-

line-specific Nanos Gal4 driver. Fragmented digoxigenin-labeled antisense RNA probes against full-length gurken transcript was used to detect gurken mRNA *in-situ*. Visualization of probes was done using an anti-Digoxigenin Cyo3 secondary antibody. DNA was visualized with DAPI staining. DNA (blue) and gurken mRNA (red). Scale bar 10um

.




**Supplementary Figure-3.3: Classes of gurken mRNA anterior dorsal localization defects.** Anterior dorsal localization of gurken mRNA in stage 8/9 egg chambers was categorized in three phenotypic classes: "normal DV localization", "partial DV localization" and "no DV localization". DNA (blue) and gurken mRNA (red). Scale bar 10um.




**Supplementary Figure-3.4: Depletion of Salsa does not affect the morphology of the oocyte karyosome (DNA).** During mid-oogenesis the oocyte chromatin is condensed into a sphere called karyosome (arrowhead). Defects in karyosome morphology are indicative of the accumulation of DNA damage and abnormal activation of the meiotic checkpoint. Germ line depletion of Salsa does no impair the correct morphology of the oocyte karyosome. DNA was visualized with DAPI staining (grey). Arrowheads point to the oocyte karyosome. Scale bars 5 μm.

### 3.3.5 Salsa is required for anterior dorsal localization of Gurken protein

Along with many other maternal mRNAs, gurken transcript is transcribed in the supporting nurse cells nuclei, and then transported and localized into the developing oocyte (González-Reyes et al., 1995; Neuman-Silberberg and Schupbach, 1994; Richter and Lasko, 2011). Translation of gurken mRNA is

repressed during transport and until it is correctly localized in the oocyte (Weil et al., 2012). Since depletion of Salsa impaired the correct anterior dorsal localization of gurken transcript, and since the correct RNA localization of gurken transcript is crucial for its translation, we hypothesized that Gurken protein expression would be similarly impaired after depletion of Salsa. To test this hypothesis, we did fluorescent immunostainings with an anti-Gurken antibody in *Drosophila* egg chambers. As expected, Gurken protein accumulated at high levels in the cytoplasmic perinuclear anterior dorsal region of control stage 8/9 oocytes (control RNAi; mCherry RNAi), whereas such expression was significantly dramatically reduced after depletion of Salsa (Figure-3.5A, B) (see material and methods for quantification details). Using the currently available antibodies, we failed to detect the total levels of Gurken protein in the ovaries by western blot.

Altogether, our results demonstrate that the spliceosome NTC subunit Salsa is important for DV patterning of the developing egg chamber because it is required for anterior dorsal localization of Gurken during oogenesis. This requirement is most likely because splicing of the first intron of gurken is particularly sensitive to depletion of Salsa.

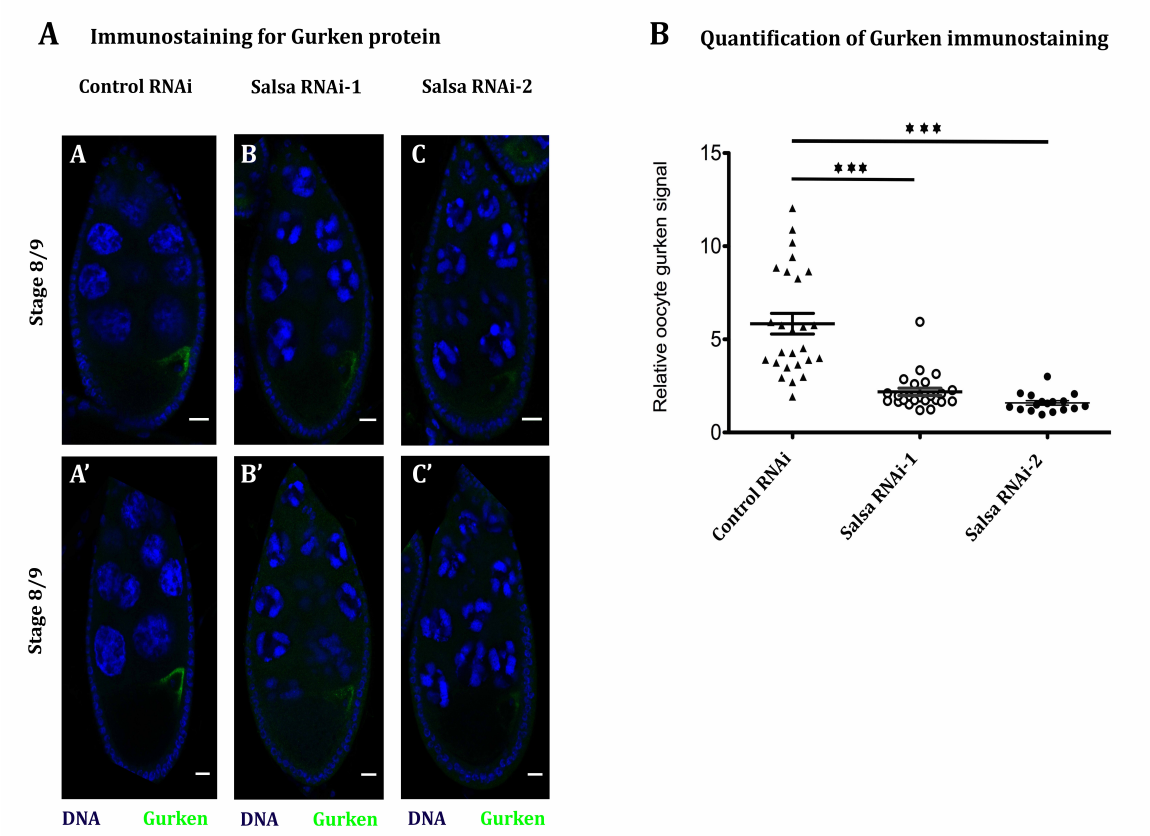



**Figure 3.5: Salsa is required for anterior dorsal localization of Gurken protein. (A)** Immunostaining for Gurken in stage 8/9 egg chambers showed that female germ-line specific of depletion of Salsa (Salsa RNAi-1 and Salsa RNAi-2) significantly impaired anterior dorsal localization of Gurken. **(B)** Quantification of Gurken anterior dorsal localization within the oocyte. Relative oocyte Gurken signal (arbitrary units (a.u.)) corresponds to the anterior dorsal Gurken signal intensity (three different measurements from the anterior dorsal region with the highest signal) normalized to signal detected in the respective nurse cells cytoplasm (three different measurements). To minimize sample variation all measurements were obtained from maximum intensity projections obtained from confocal Z stacks of stage 8/9 egg chambers. Each dot represents an individual stage 8/9-egg chamber. Horizontal lines specify mean values and asterisks indicate significant difference (two tailed unpaired t test; $P<0.001$). Gurken was detected with an anti-Gurken monoclonal antibody and DNA was visualized with DAPI staining. DNA (blue) and Gurken (green). Scale bar 10um.

### 3.3.6 Salsa is required for alternative splicing of eIF4E

The splicing machinery recognizes several motifs at pre-mRNA sequences for accurately recognizing the intended splice sites amidst a vast array of distinct exons and introns (Amit et al., 2012a). Based on the splicing strength, differential splice site recognition can allow the generation of a broad range of alternative splicing events. We observed that depletion of Salsa specifically impaired splicing of the first intron of gurken transcript. Bioinformatics analysis (http://spliceport.cbcb.umd.edu/) identified this intron as potentially having a weak 5´splice site, suggesting a role of Salsa for efficient splicing of weak splice site and regulation of alternative splicing.

To investigate if Salsa can potentially regulate alternative splicing, we analyzed the alternative splicing pattern of Eukaryotic Initiation Factor 4E (eIF4E). eIF4E protein binds to 5' end of mRNAs and it is important for translation initiation (Gingras et al., 1999). The *Drosophila eIF4E* gene encodes three alternatively spliced transcripts (Lavoie et al., 1996), giving rise to two different protein isoforms (eIF4E-I and eIF4E-II). We analyzed the alternative spliced transcripts of eIF4E-I using RT-PCR and previously published primers (black arrows shown in Figure-3.6A) (Van Buskirk and Schüpbach, 2002b), and observed an apparent change in the pattern of alternative splicing after Salsa depletion (Figure-3.6B'). To further confirm that Salsa is required for alternative splicing of *eIF4E*, we quantified its transcript variant levels using RT-qPCR and variant specific primers (open arrows shown in Figure-3.6A) (primers sequence showed in Table 3). Consistently, Salsa depletion was associated with a significant change in the levels of eIF4E-RB transcript variant after the depletion of Salsa (Figure-3.6B''). No detectable change could be observed for the total levels of eIF4E transcripts (Figure-3.4B´´).

The eIF4E-RB transcript variant encodes the protein isoform eIF4E-I, a subunit of eIF4F complex that plays a critical role in the regulation of translation initiation (Hiremath et al., 1989; Lavoie et al.). Since depletion of Salsa is associated with an abnormal expression of Gurken, it is possible that the observed DV patterning defects were being mediated (at least partially) by a loss of eIF4E. To test this hypothesis, we investigated if the protein levels of

eIF4E-I were impaired after Salsa depletion. Using western blot analysis from ovaries total protein extracts, we failed to detect any change in the total protein levels of eIF4E-I after depletion of Salsa (Figure-3.6C), suggesting that the observed Salsa phenotypes are not likely to be meditated by changes in eIF4E-I transcript.

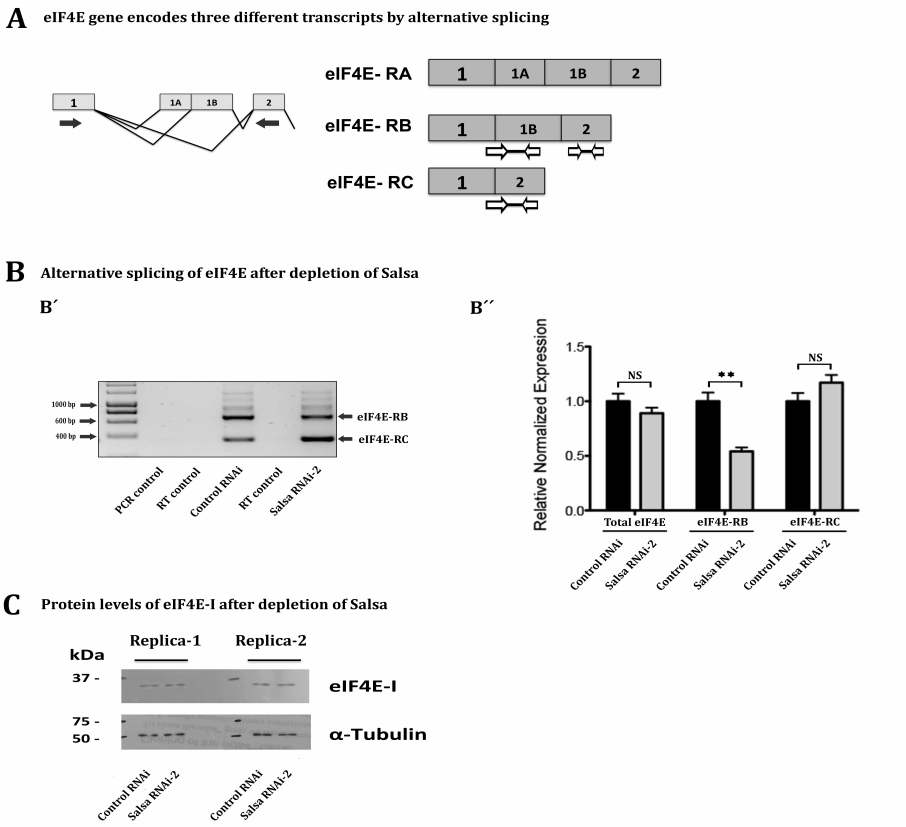



**Figure 3.6: Alternative splicing of eIF4E mRNA is sensitive to depletion of Salsa.(A)** The *Drosophila* eIF4E gene encodes three transcripts (RA, RB and RC) through alternative splicing. Black arrows indicate primers used for RT-PCR of gurken, whereas white arrows indicate the primers used for RT-qPCR of eIF4E transcripts**. (B-1)** RT-PCR analysis of eIF4E transcripts from iScipt cDNA library (random hexamer and oligo(dT) reaction mix) and using exon–exon primers that cover entire transcript. Control ovaries yielded PCR products in the size predicted for correct alternatively spliced eIF4E transcripts (eIF4E RB: 720 bp; eIF4E RC: 290 bp). DNA sequencing of the isolated bands confirmed correct splicing of eIF4E transcripts. Ovaries whose germ-line was depleted for Salsa (Salsa RNAi-2) showed increased eIF4E RC transcript (arrow), whereas decrease level of eIF4E RB transcript (arrow). PCR and RT control used in all reactions. 'RT controls' (absence of reverse transcriptase (RT) reaction) yielded no amplification, meaning there was no contamination with genomic DNA in all samples tested. **(B-2)** Quantitative analysis of eIF4E alternative

transcript by RT–qPCR using transcript specific primers. Shown are the fold changes of transcripts from alternative splicing of eIF4E. (Salsa RNAi-2) showed significant decrease in eIF4E RB transcript, whereas no significant changes in eIF4E RC transcript and total eIF4E level. Relative normalized expression corresponds to values normalized with two distinct reference genes (β-actin and GAPDH) and relative to control RNAi (mCherry RNAi). At least three biological replicates were used for all datasets. Error bars indicate standard deviation. Ovaries were dissected from 3-5 days old females (after pupae eclosion). Female germ-line specific expression of dsRNAs was obtained by using the UAS/Gal4 system and the germ-line-specific Nanos Gal4 driver. **(C)** Total protein levels of endogenous eIF4E-I did not change after depletion of Salsa. Western-blot analysis of total protein extracts from ovaries (dissected from 3-5 days old females) from control and salsa RNAi (Salsa RNAi-2). α-Tubulin was used as a Western-blot loading control. Female germ-line specific expression of dsRNAs was obtained by using the UAS/Gal4 system and the germ-line-specific Nanos Gal4 driver.

### 3.3.7 Salsa is required for splicing of small proximal introns

We found that Salsa is a *bona fide* subunit of the spliceosome NTC complex and its germ line specific depletion specifically impairs the efficient splicing of the first intron of gurken transcript. We also observed that the first intron of gurken potentially contains a 5´splice site significantly weaker than the ones present in the second and third introns of this transcript. We hypothesized that Salsa function is particularly rate limiting for efficient splicing of introns with weak splice site strength and/ or localized in the 5´region of transcripts. To test this hypothesis, and further examine the role of Salsa in pre-mRNA splicing, we decided to analyze the ovaries transcriptome after partial depletion of Salsa; using the weakness Salsa RNAi construct available (Salsa RNAi-1) in order to expose the introns whose splicing is particularly sensitive to Salsa depletion and minimize non-specific phenotypes due to egg chambers necrosis or other pleiotropic phenotypes. RNA sequencing (RNA-seq) of the ovaries transcriptome, after partial depletion of Salsa (Salsa RNAi-1) and poly(A) enrichment of the mRNAs, only revealed a minor effect on gene expression levels, with only 11 genes showing detectable changes in expression (Figure-3.7A).

In order to identify introns whose splicing is particularly sensitive to depletion of Salsa, we performed a differential intron retention analysis of the ovaries

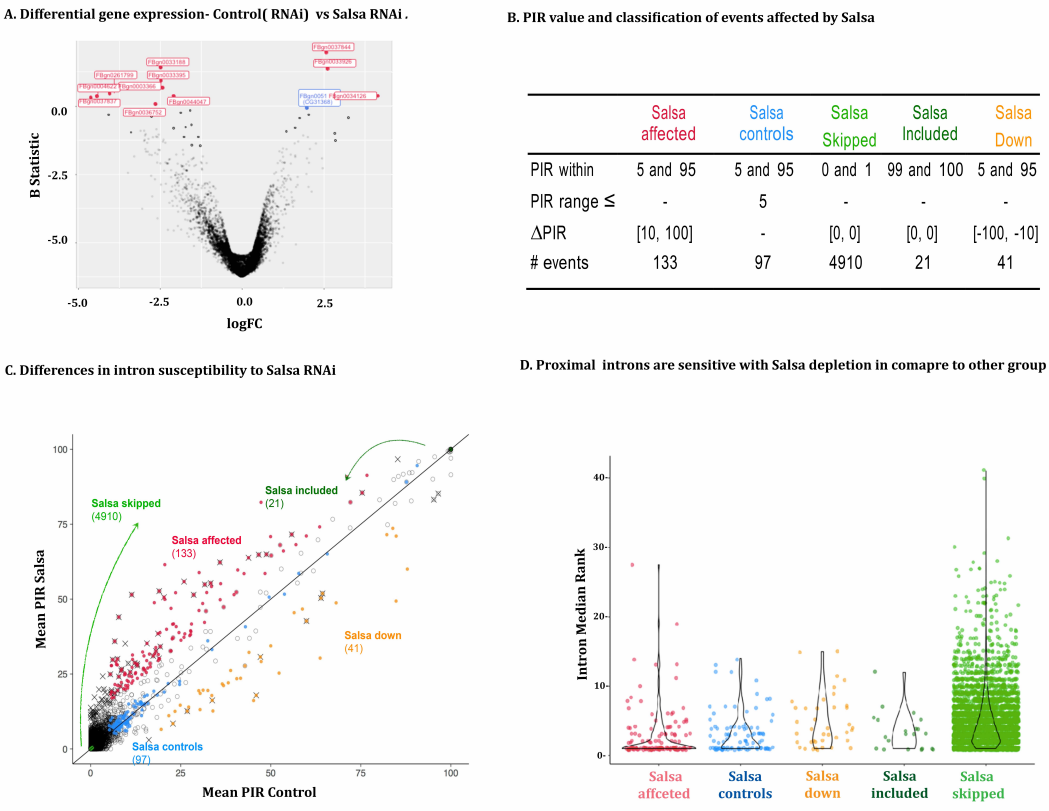
transcriptome by mean "Percent of intron retention" (PIR) per condition (control RNAi and Salsa RNAi; two biological replicas each) and calculated the ΔPIR for each intron retention event. Identified introns were divided into five classes; (a) alternative splicing introns whose splicing was affected by Salsa depletion (Salsa-dependent introns; marked in red); increased PIR values in Salsa RNAi samples compared to control RNAi; (b) alternative splicing introns whose splicing was not affected by Salsa depletion (Salsa control; marked in blue); PIR values within the alternative range but whose differences was not greater than 5. (c) Constitutive introns whose splicing is not affected by Salsa depletion (Salsa skipped; marked in light green); PIR values of 0 or 1. (d) Introns that are rarely spliced during oogenesis (Salsa included; marked in dark green); PIR values of 99 or 100. (e) Alternative introns whose splicing became more efficient after Salsa depletion (Salsa down; marked in yellow); decreased PIR values in Salsa RNAi samples compared to control RNAi. Criteria used for definition of these 5 groups are compiled in Figure-3.7B. This analysis revealed that partial depletion of Salsa during oogenesis only affected splicing (mostly, intron retention) of a small subset of alternatively spliced transcripts; only 2.25% (311 out of 5202), which suggested that Salsa is particularly rate-limiting for splicing of only a small subset of introns.

Since the RNAseq data was generated from mature RNA, we decide to analyze the splicing of gurken mRNA. As expected with our previous result, we observed that depletion of Salsa does not affect the total expression of gurken mRNA. Consistently, we observed depletion of salsa impairs the splicing of first intron, whereas the second and third intron spliced normally (Supplementary figure- 3.5). These observations again fit with our pervious data from Real-Time qPCR analysis and clearly suggest that depletion of Salsa impairs the splicing of particular first intron of gurken.

In order to better understand the physiological role of Salsa during oogenesis, we analyzed the different features of the introns whose splicing is particularly sensitive to Salsa depletion using the Kruskal-Wallis test (Ritchie et al., 2015).From the subset of intron features tested, introns particularly sensitive to Salsa depletion (Salsa RNAi-1) were significantly (P-value < 0.05) shorter

and more proximal to the transcript start(Figure- 3.7D), as well as enriched in upstream exons with lower GC content.

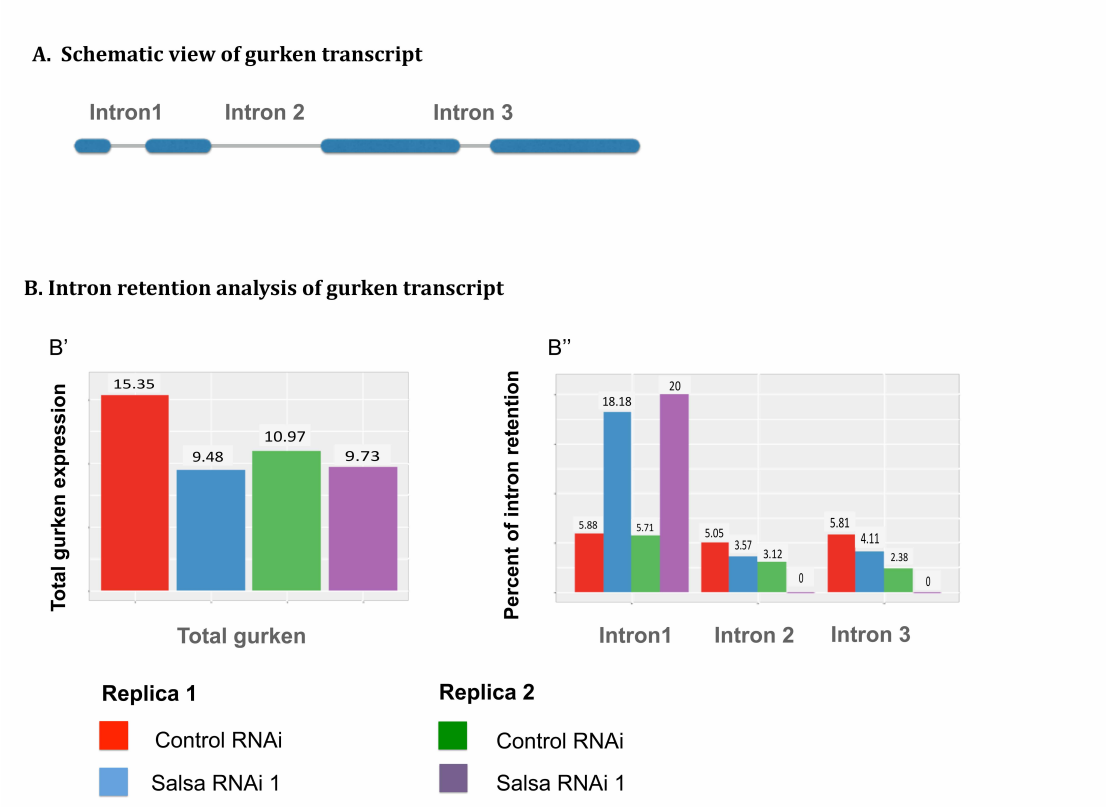
The spliceosome needs to accurately recognize the intended splice sites among a vast array of distinct exons and introns (Amit et al., 2012b). Lower GC content has been interpreted as implying less efficient transcription, splicing, or translation (Amit et al., 2012b; Ding et al., 2012). It has also been shown that the introns localized to the transcripts proximal end present a bias for weak 5´ splice sites, most likely because of the enrichment of U1 at this site to suppress premature cleavage and polyadenylation (Kaida et al., 2010). Altogether, and although the precise mechanism is still unknown, these results clearly suggest that introns with weak splice site and/ or located on the transcript proximal region are particularly sensitive to Salsa depletion.




| | Salsa affected | Salsa controls | Salsa Skipped | Salsa Included | Salsa Down |
|---|---|---|---|---|---|
| PIR within | 5 and 95 | 5 and 95 | 0 and 1 | 99 and 100 | 5 and 95 |
| PIR range ≤ | - | 5 | - | - | - |
| ΔPIR | [10, 100] | - | [0, 0] | [0, 0] | [-100, -10] |
| # events | 133 | 97 | 4910 | 21 | 41 |

**Figure 3.7: Splicing of small proximal introns is particularly sensitive to Salsa depletion.** Partial depletion of Salsa during oogenesis affects the expression of a small subset of transcripts. (A) Transcripts with significant differential gene expression (FC) after Salsa depletion (control RNAi *versus* Salsa RNAi) are shown in red. As expected salsa

(CG31368) expression is significantly reduced in Salsa RNAi (shown in blue). Differential gene expression (FC) is measured in log fold-change. (B) Percentage of Intron Retention (PIR) value and intron classification based on their responsiveness to Salsa depletion. Differential intron retention was profiled by calculating mean PIR per condition (Salsa RNAi samples *versus* control RNAi) and introns were classified in five major classes. Criteria used for such classification is shown in this panel. (C) Differential percentage of intron retention (PIR) after depletion of Salsa [control RNAi versus salsa RNAi]. Graph showing a direct comparison between mean PIR per condition allowed the definition of five major classes of introns: (a) alternative splicing introns whose splicing was affected by Salsa depletion (Salsa-dependent introns; marked in red); increased PIR values in Salsa RNAi samples compared to control RNAi; (b) alternative splicing introns whose splicing was not affected by Salsa depletion (Salsa control; marked in blue); PIR values within the alternative range but whose differences was not greater than 5. (c) Constitutive introns whose splicing is not affected by Salsa depletion (Salsa skipped; marked in light green); PIR values of 0 or 1. (d) Introns that are rarely spliced during oogenesis (Salsa included; marked in dark green); PIR values of 99 or 100. (e) Alternative introns whose splicing became more efficient after Salsa depletion (Salsa down; marked in yellow); decreased PIR values in Salsa RNAi samples compared to control RNAi. Criteria used for definition of these 5 groups are shown in panel B of this figure. (D) From the subset of intron features tested, introns particularly sensitive to Salsa depletion (Salsa RNAi-1) were significantly (P-value < 0.05) shorter and more proximal to the transcript start, as well as enriched in upstream exons with lower GC content.




**Supplementary Figure-3.5: The first intron retention of gurken is highly retained in RNA-seq data analysis:** (A) Schematic view of gurken transcript that contain three introns. (B') Germ line specific depletion of Salsa doesn't changes the total expression of gurken mRNA. The number on top is percentage of total expression in given RNAi. (B'') Germ line specific depletion of Salsa impairs the splicing of first intron, whereas the second and third intron spliced normally. In biological replica – 1, the control RNAi shows 5.88% of first intron retention and after Salsa depletion it goes to 18.18% of intron retention. Similarly, the biological replica-2 for control RNAi shows 5.71% of first intron retention and after Salsa depletion it goes to 20% of intron retention. These observations again fit with our pervious data from Real-Time qPCR analysis and clearly suggest that depletion of Salsa impairs the splicing of first intron and second and third intron splice normally. The color represents the different RNAi. Biological replica- 1: (Red- Control RNAi; Blue- Salsa RNAi1), Biological replica- 2: (Green- Control RNAi; Purple- Salsa RNAi1)

# 4

# General discussion and future directions

## Contents

## 4.1 General discussion

### 4.1.1 Salsa is particularly rate limiting for splicing of small proximal introns during *Drosophila* oogenesis.

In the last decade, the NineTeen Complex (NTC) (also known as Prp19 complex) has been extensively studied. It is a heteromeric complex and one of several non-snRNPs of the spliceosome, with a crucial role in splicing, as it is required for the activation of the spliceosome. This complex is nevertheless multifunctional, with several of its subunits being described as being important not only for splicing, but also for transcriptional elongation and DNA repair (Dujardin et al., 2013; Fededa and Kornblihtt, 2008; Fong and Zhou, 2001), raising extremely interesting questions regarding the role of this complex in functional crosstalk between gene expression and genomic stability.

The Prp19 complex plays an important role during development of multicellular organisms, being, for example, particularly rate-limiting for splicing of early zygotic transcripts during *Drosophila* development (Guilgur et al., 2014). Salsa is a highly conserved subunit of the Prp19 complex, being its human orthologue, a catalytically active RNA helicase named Aquarius (De et al., 2015) (Figure-2.4). Depletion of Salsa within the female germ line is associated with abnormal anterior dorsal localization of Gurken mRNA and dorso-ventral (DV) patterning defects. Interestingly, Salsa function is particularly rate limiting for the splicing of first intron of gurken mRNA, as its depletion specifically affects splicing of this proximal 5´UTR-localized intron; whereas splicing of the second and third introns is apparently unaffected.

Using a splice site analysis software (splice port), we observed that first intron contains weak 5'splice site score (-0.43) compared to second (1.4) and third (2.1) introns of gurken transcript, suggesting that Salsa is particularly rate limiting for splicing of introns with weak splice sites strength. To test this hypothesis, and better understand the general effect of Salsa in splicing, we decided to analyze whole transcriptome after the depletion of Salsa in *Drosophila* ovaries. This analysis was done in collaboration with Dr. Nuno Morais and Mariana Ferreira, at Institute of Molecular Medicine (IMM), Lisbon, Portugal.

The transcriptome-wide analysis of Salsa-depleted ovaries revealed that depletion of Salsa during oogenesis leads to intron retention in a small subset of transcripts; only 2.25% (311 out of 5202). Because only a small subset of introns was retained, we asked whether they had specific features capable of explaining why they were more sensitive to Salsa depletion. Our analysis showed that these introns had a significant bias (P-value < 0.05) for being proximal, slightly smaller in length, and with an upstream exon with a lower GC content. Interestingly, proximal introns were previously described as having a bias for comparatively weaker 5´splice sites, when compared with more distal ones (Lepennetier and Catania, 2016; Pai et al., 2017), suggesting an alternative (but potentially complementary) hypothesis that Salsa is instead particularly rate limiting for splicing not for introns with weak 5´ splice sites, but for proximal introns.

There are several hypotheses supporting a slower/ less efficient splicing of proximal introns. First, it has been demonstrated in *Drosophila* that various transcriptional enhancers are recruited to the first intron and making their splicing slower when compared to more distal introns (Arnold et al., 2013; Kharchenko et al., 2011). Second, high enrichment of U1 at the nascent gene transcript, and preferably at the first intron, it may slowdown the splicing at this position (Lin and Zhang, 2005; Ruvinsky and Ward, 2006; Sakurai et al., 2002). Thirdly, GC content influences splice site usage and it is associated with splicing kinetics, and it has been shown that low GC content negatively influences splicing efficiency. Interestingly, in *Drosophila* that the length of first intron is in average twice of size of the following introns (Horiuchi et al., 2003; Khodor et al., 2011; Lepennetier and Catania, 2016; Pai et al., 2017), suggesting that the observed bias for smaller introns is likely to be highly significant.

In our genome wide analysis of *Drosophila* ovaries depleted for Salsa, we identified transformer2 (tra2) transcript, as having high levels of intron retention (Supplementary Figure- 2). Tra2 is a splicing protein whose function is required for alternative splicing of doublesex, resulting in a female specific Doublesex mRNA. In the absence of TRA2, a male specific Doublesex mRNA is formed, inducing a male developmental fate. We confirmed the splicing

defects of Tra2 by RT-PCR (Supplementary Figure- 2) and we are currently quantifying the levels of intron retention with RT-qPCR. Consistently with our previous analysis, Salsa is particularly rate limiting for splicing of the proximal 5´UTR-localized intron of tra2, further confirming the importance of Salsa in the efficient splicing of this class of introns.

Studies in *Drosophila* and mammalian systems suggest that kinetics of alternative splicing is different and slower than constitutive splicing (Fong et al., 2014; Jonkers et al., 2014; Kwak et al., 2013; Pandya-Jones and Black, 2009). Since Salsa is potentially required for efficient splicing of weak splice sites, we hypothesized that Salsa depletion might also change the patterns of alternative splicing. To test this hypothesis, we analyzed the alternative splicing of the eukaryotic initiation factor 4E (eIF4E) (Gingras et al., 1999; Van Buskirk and Schüpbach, 2002b) by real time-qPCR, using transcript specific primers. We observed significant changes in the alternative splicing pattern of eIF4E transcripts after Salsa depletion.

### 4.1.2 Retention of the proximal intron of gurken is likely to impair its anterior dorsal localization.

Work performed in our laboratory showed that a wild type genomic construct of gurken fully complemented a loss-function allele of gurken with or without its first proximal intron (Rui Gonçalo Martinho, personal communication), suggesting that the anterior dorsal localization of gurken transcript does not require the first intron sequence and/or the splicing reaction per se. Nevertheless, it is still likely that retention of the proximal 5´UTR-localized intron of gurken mRNA impairs its anterior dorsal localization, as it was previously suggested that gurken 5´UTR is important for its localization and DV patterning of the egg (Saunders and Cohen, 1999b; Thio et al.).

Our working hypothesis is that the 5´UTR of the mature gurken mRNA contains a localization motif that is disturbed when the intron is not spliced correctly, impairing the recruitment of proteins (e.g. TREX or EJC) required for its nuclear export and localization. Consistently, and further supporting our hypothesis, a detailed bioinformatics analysis of the folding of the 5´UTR of mRNA gurken showed that its structure is dramatically changed when the intron is retained (Srivathsan Vranganathan and Prashanth Rangan, personal

communication). Alternatively, it is also possible that Salsa is required for the correct splicing of another gene (e.g. otu transcript), which is required for gurken mRNA localization.

### 4.1.3 Is Salsa required for the recruitment of TREX/EJC complex?

The TRanscription and EXport (TREX) is a highly conserved protein complex that links transcription, mRNA processing, and nuclear export (Katahira, 2012b; Lee and Tarn, 2013). TREX contains UAP56 and Aly proteins, as well as the multi-subunit THO complex (Gómez-González et al., 2011; Masuda et al., 2005a; Masuda et al., 2005b; Sträßer et al., 2002; Svejstrup, 2003). The THO complex is conserved throughout metazoan evolution (Katahira, 2012a), and in higher eukaryotes it contains six proteins; three proteins are homologous to yeast proteins (Hpr1/Thoc1, Thoc2, and Thoc3/Tex1), and three are unique (Thoc5/FMIP, Thoc6, and Thoc7) (Gómez-González et al., 2011; Sträßer et al., 2002). RNA interference and genetic analyses indicates that TREX is important for efficient nuclear mRNA export (Cheng et al., 2006; Gatfield et al., 2001; Hur et al., 2016; Le Hir et al., 2001; Masuda et al., 2005b).

Similarly to Salsa depletion, depletion of TREX complex cause defects in dorsal-vental patterning(Hur et al., 2016) and UAP56 (TREX/EJC) is also required for anterior dorsal localization of gurken mRNA localization (Meignin and Davis, 2008). Since it has been shown that the human orthologue of Salsa, Aquarius, is important for the loading of the exon junction complex (EJC) into the mature mRNAs, it is possible that Salsa may be similarly required for the recruitment of EJC complex into gurken mRNA. Supporting such hypothesis, and using tissue culture *Drosophila* S2 cells, we found that THO2, a major component of the *Drosophila* TREX complex, physically interacts with Salsa (Supplementary Figure- 3). Nevertheless, the relevance of Salsa-dependent recruitment of TREX/EJC for the correct localization of gurken mRNA is still unlikely. The fact that splicing of gurken proximal intron is *per se* not rate limiting for anterior dorsal expression of Gurken would necessarily imply an unlikely scenario of Salsa recruiting TREX/EJC to gurken mRNA independently of splicing. Furthermore, high levels of unspliced gurken

mRNAs, after depletion of Salsa, can be detect in the early embryo implying that proximal intron retention did not impair nurse cells nuclear export.

### 4.1.4 Is Salsa a RNA binding protein?

Aquarius, Salsa human ortholog, is a RNA helicase that does not have a clear RNA binding motif but its known to directly bind pre-mRNAs. Recent observations in *C. elegans* and *A. Thaliana*, suggested that Aquarius also binds to small RNAs (Akay et al., 2017; Jia et al., 2017). In *Drosophila,* Salsa is a member of the Prp19 complex and our results show that it is important for splicing. Therefore, we decided to investigate if Salsa also binds RNAs, and if it does so, if it binds with transcripts directly or in a complex dependent manner. In collaboration with Dr. Jean Yves Roignant (at Institute of Molecular Biology in Mainz, Germany), we immunoprecipitated a Myc-tagged version of Salsa from tissue culture *Drosophila* S2 and preliminary results suggested an interaction between Salsa and a small subset of RNAs (Supplementary Figure-4). Our preliminary data also indicated that the associated RNAs were small in size (bellow 50 nucleotides), but it is still unclear to what extend this small size results from being small RNAs or simply from RNA degradation.

## 4.2 Future directions

### 4.2.1 Show that proximal intron retention of gurken is rate-limiting for its anterior dorsal localization

If retention of the 5´UTR-localized intron is why Salsa depletion impairs anterior dorsal localization of gurken mRNA and is associated with DV patterning defects than expression of a genomic gurken construct without the first intron should be able to fully suppress those phenotypes.

If the reason why the first proximal intron is particularly sensitive to Salsa depletion is because it contains a weak 5´splice site than a genomic gurken construct with a strong 5´splice site in the first intron should be able to suppress the localization defects of gurken and DV patterning defects after Salsa depletion.

### 4.2.2 Show that Salsa binds gurken mRNA and define its *in vivo* RNA-binding sites

Salsa is a RNA binding protein (RBP) and most RBPs bind to mRNAs because they recognize short, degenerate RNA motifs. Genome wide accurate mapping of RBP-binding sites on mRNA is critical to elucidate the function of such RBPs and the way they regulate gene expression. To further confirm that Salsa directly regulates splicing of gurken mRNA, we are going to use RNA-Immunoprecipitation (RIP) followed by RT-qPCR to confirm that Salsa binds to gurken mRNA. Furthermore, and to better understand Salsa function *in vivo* we are going to define its in vivo RNA binding sites using individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP). This work will be done in collaboration with Dr. Jean Yves Roignant and Dr. Julian Konig (Institute of Molecular Biology in Mainz, Germany).

### 4.3 Concluding remarks

In this study I presented the identification and characterization of *Drosophila* Salsa as a subunit of the NTC complex, whose function was particularly rate limiting for splicing of small proximal introns. Consistently, depletion of Salsa impaired the efficient splicing of the first proximal intron of gurken transcript. We proposed that retention of the first intron of gurken impairs its anterior dorsal localization, possibly because it impairs the folding of a 5´UTR-localized RNA motif and the recruitment of proteins required for its localization. Defects in the anterior dorsal localization of gurken mRNA are most likely the reason why depletion of salsa is associated with striking DV patterning defects and a reduction in female fertility.

Nevertheless, more extensive research is required to fully unravel the molecular mechanism behind Salsa function during oogenesis. It remains to be elucidated how and when the rearrangements of the spliceosome activation through NTC assembly on pre-mRNA are influenced by Salsa, and why are small proximal introns particularly sensitive to Salsa depletion. Moreover, and also of great interest, is the identification of the 5´UTR RNA-binding motif responsible for the anterior dorsal localization of the gurken transcript, and the identification of its interacting proteins.

This study represents the first molecular characterization of Salsa in *Drosophila*, providing key insights into possible mechanisms regarding the way a highly conserved splicing factor can mRNA localization during development of multicellular organisms.

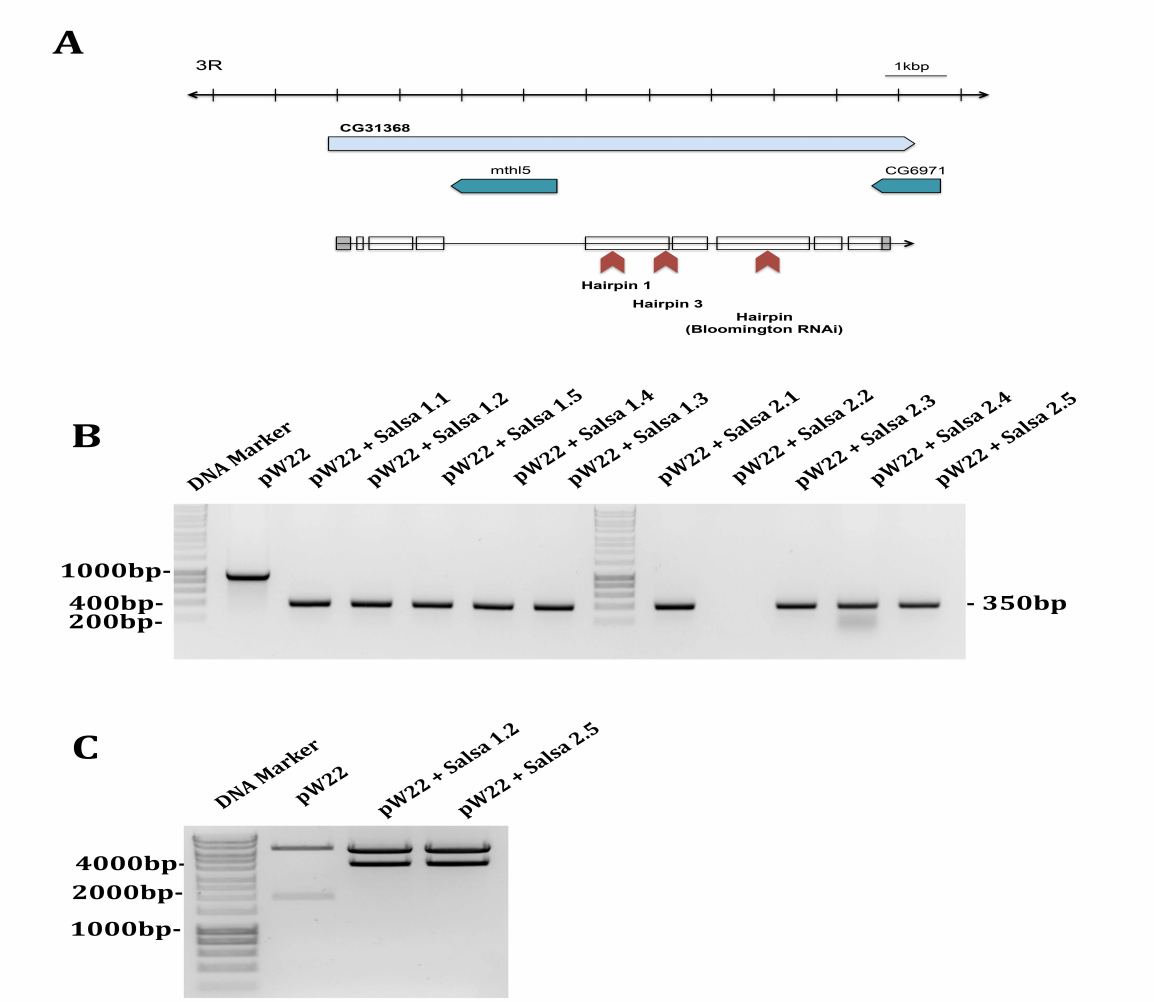
# 5

# Supplementary informations and references

## Contents

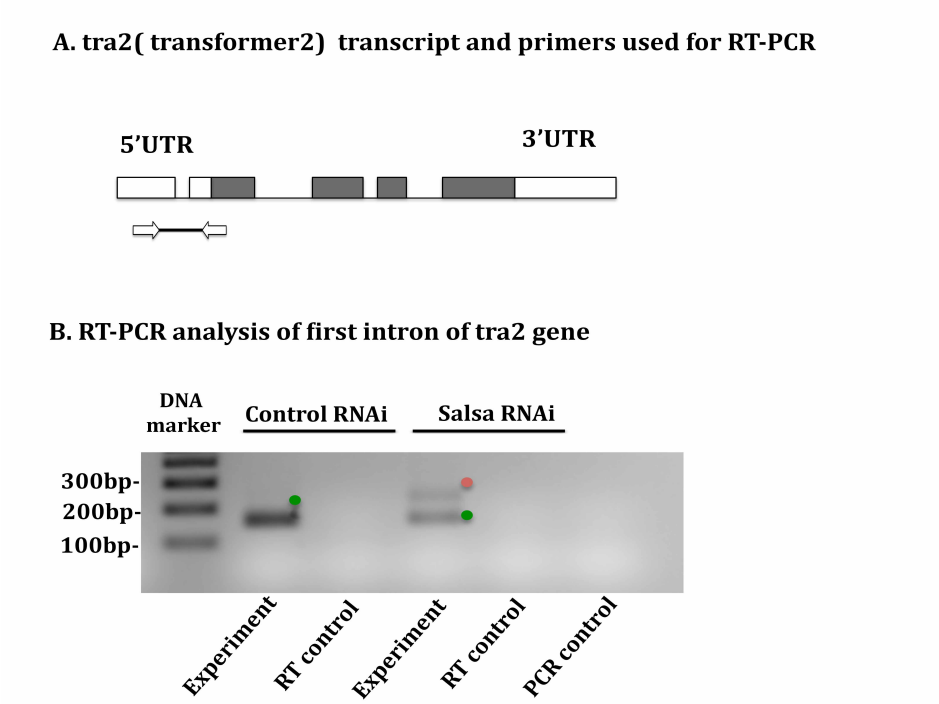
## 5.1 Supplementary figures

### 5.1.1 Supplementary figure 1. Generation of *Drosophila* Salsa (CG31368) RNAi lines



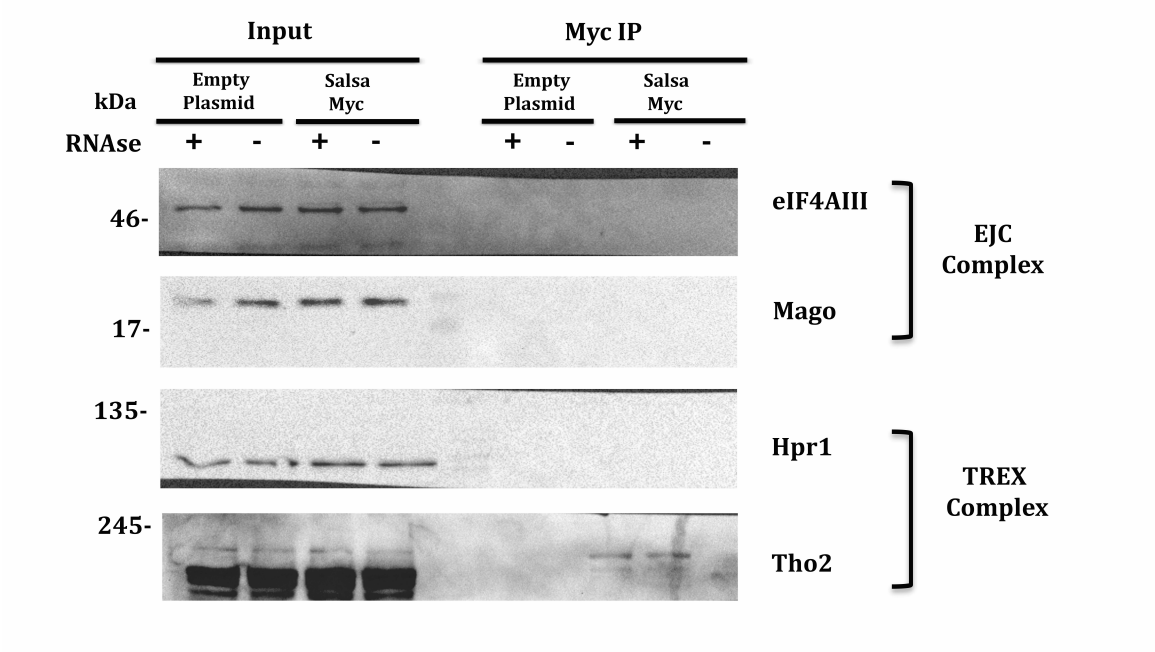

**Supplementary Figure 1: Generation of non-overlap Salsa (CG31368) RNAi:** (A) Genomic organization of CG31368 (light blue), and the orientation present the direction of transcript from 5'UTR to 3"UTR. The scheme represent the insertion of different hairpin in red arrow; Salsa RNAi-2 (BDSC RNAi), whereas Salsa RNAi-1 and Salsa RNAi-3 were custom made. (B) PCR amplification of top and bottom strand oligos specific for Salsa and the resulting fragment has overhangs for *NheI* and *EcoRI* and ligated in pW22 (350bp), insertion was further confirmed by sequencing. (C) The lgation reaction was also confirmed with with restriction digestion and sequencing.

### 5.1.2 Supplementary figure 2. Salsa is particularly rate limiting for the splicing of first intron of tra2 (transformer 2)



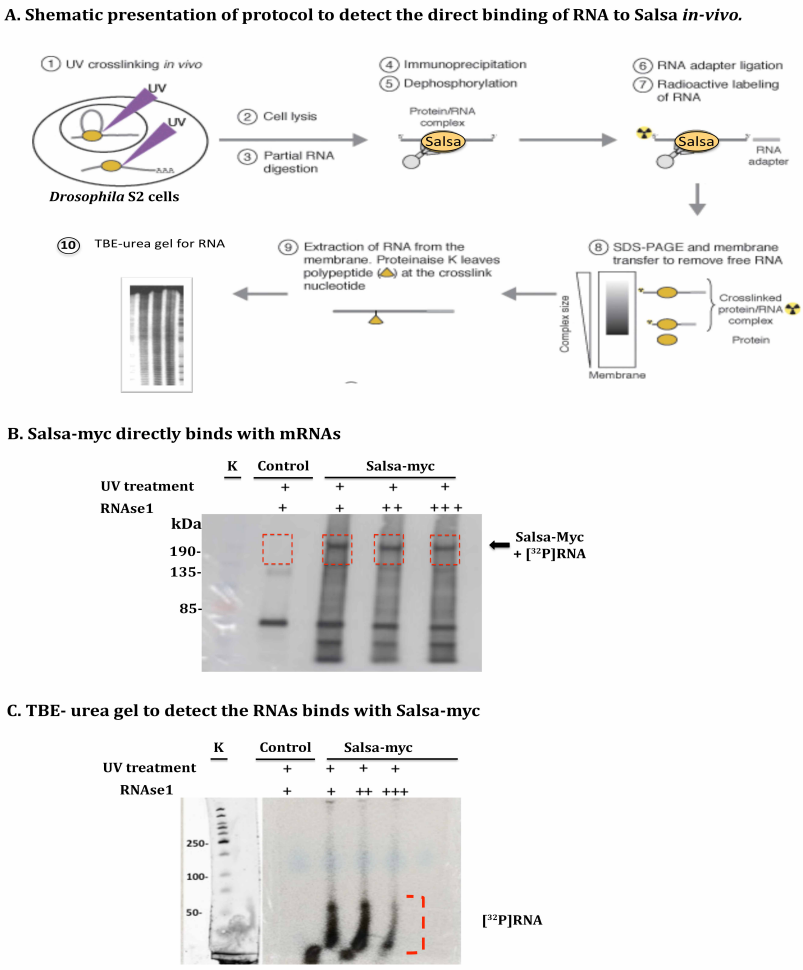

**Supplementary Figure 2: Salsa is particularly rate limiting for the splicing of first intron of tra2 (transformer 2).** (A) The Transformer2 (tra2) gene encodes transcript contain four introns, gray color shows CDS region and blank UTR region of Tra2 gene. Black arrow shows primer design specific for RT-PCR of gurken. (B) RT-PCR analysis of tra2 transcripts from poly (A) RNA and using exon–exon (e–e) primers. Control ovaries yielded PCR products in the size predicted for correctly spliced gurken transcripts (green dots, *tra2*: 170 bp). DNA sequencing of the isolated bands confirmed correct splicing of gurken transcript. Salsa depleted ovaries (Salsa RNAi) showed splicing defects (intron retention) of the first intron of gurken transcript (red dots, tra2: 230bp). PCR and RT control used in all reactions. 'RT controls' (absence of reverse transcriptase (RT) reaction) yielded no amplification, meaning there was no contamination with genomic DNA in the samples tested.

### 5.1.3 Supplementary figure 3. *Drosophila* Salsa physically interacts with spliceosome THO2, a conserved TREX subunit in *Drosophila* S2 cells




**Supplementary Figure 3: *Drosophila* Salsa physically interacts with spliceosome THO2, a conserved TREX subunit in *Drosophila* S2 cells:** Endogenous components of TREX complex (Hpr1 and Tho2) and EJC complex (eIF4AIII and Mago) were analyzed for interaction with *Drosophila* Salsa. TREX proteins THO2 was efficiently co-immunoprecipitated by Salsa-Myc in an RNA-independent manner, we failed to detect Hpr1 from the similar complex. Similarly, we failed to detect proteins from EJC complex in co-immunoprecipitated by Salsa-Myc. Co-immunoprecipitation from total protein extracts from *Drosophila* S2 cells expressing Myc-tagged Salsa (Salsa-Myc) and using anti-Myc coated dynabeads. Protein extracts from *Drosophila* S2 cells transfected with an empty plasmid were used as a negative control. Lysate were treated during immunoprecipiation with RNAse shown as (+) and without treatment (-).

## 5.1.4 Supplementary figure 4. Salsa-myc binds with mRNA in *Drosophila* S2 cells




**Supplementary Figure 4: Salsa-myc binds with mRNA in *Drosophila* S2 cells: (A)** Schematic illustration of CLIP (CrossLinked-ImmunoPrecipitation) procedure (Adapted from (Huppertz et al., 2014). **(b) Salsa directly binds with RNA;** Autoradiograph of immunopurified and 32P-labeled Salsa–RNA complexes transferred to a nitrocellulose membrane. The RNAse1 treatments are shown as +(1:50); ++ (1:5000) and +++(1:1,000). Control is the transfection with empty vector and cells were irradiated with UV-C light also indicated. The black arrow denotes the 32P-labeled Salsa–RNA adducts ([32$^{P}$]RNA-x-Salsa) on a nitrocellulose membrane. The indicated area (red square) was processed for RNA extraction through proteinase-K treatment. **(C) TBE-urea gels for RNA analysis.** The extracted RNA run on 6% TBE-urea precast gel. The RNAse1 treatments are shown as +(1:50); ++ (1:5000) and +++(1:1,000). Control is the transfection with empty vector and cells were irradiated with UV-C light also indicated. The indicated red area most likely RNA that bounds with Salsa, it can be either degraded during the handling of sample or only small

RNAs. We need to repeat and analyses these results with proper control in further experiments.

## 5.2 Supplementary tables

### 5.2.1 Supplementary table 1. List of *Drosophila* TRIP lines used in screen

| BL No. | TRiP No. | CG No. | Gene Symbol | Genotype |
|---|---|---|---|---|
| 44544 | HMC02939 | CG5352 | SmB | y1 v1; P{TRiP}attP40 |
| 34834 | HMS00150 | CG10753 | SmD1 | y1 sc v1; P{TRiP}attP2 |
| 55163 | HMC03839 | CG1249 | SmD2 | y1 sc v1; P{TRiP}attP40 |
| 42633 | HMS02469 | CG8427 | SmD3 | y1 sc v1; P{TRiP}attP40/CyO |
| 33664 | HMS00074 | CG18591 | SmE | y1 sc v1; P{TRiP}attP2 |
| 26734 | JF02276 | CG16792 | SmF | y1 v1; P{TRiP}attP2 |
| 33396 | HMS00274 | CG8749 | snRNP-U1-70K | y1 sc v1; P{TRiP}attP2 |
| 34593 | HMS01067 | CG4528 | snf | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 55668 | HMC03823 | CG4119 | CG4119 | y1 sc v1; P{TRiP}attP40 |
| 33711 | HMS00589 | CG3542 | CG3542 | y1 sc v1; P{TRiP}attP2 |
| 33671 | HMS00535 | CG1406 | U2A | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 34593 | HMS01067 | CG4528 | snf | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 34840 | HMS00157 | CG16941 | CG16941 | y1 sc v1; P{TRiP}attP2 |
| 55650 | HMC03799 | CG10754 | CG10754 | y1 sc v1; P{TRiP}attP40 |
| 34845 | HMS00163 | CG2925 | noi | y1 sc v1; P{TRiP}attP2 |
| 33650 | HMS00055 | CG2807 | CG2807 | y1 sc v1; P{TRiP}attP2 |
| 33651 | HMS00056 | CG3605 | CG3605 | y1 sc v1; P{TRiP}attP2 |
| 33731 | HMS00614 | CG13900 | CG13900 | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 36619 | GL00579 | CG13298 | CG13298 | y1 sc v1; P{TRiP}attP2 |
| 32344 | HMS00335 | CG9548 | CG9548 | y1 sc v1; P{TRiP}attP2 |
| 34788 | HMS00097 | CG11985 | CG11985 | y1 sc v1; P{TRiP}attP2 |
| 41654 | GL01236 | CG11107 | CG11107 | y1 v1; P{TRiP}attP2 |
| 41954 | HMS02351 | CG17540 | Spf45 | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 43199 | GL01543 | CG17454 | CG17454 | y1 sc v1; P{TRiP}attP40 |
| 50521 | GLC01639 | CG9998 | U2af50 | y1 sc v1; P{TRiP}attP2 |
| 50561 | GLC01681 | CG3582 | U2af38 | y1 v1; P{TRiP}attP2 |
| 34785 | HMS00094 | CG12085 | pUf68 | y1 sc v1; P{TRiP}attP2 |
| 52938 | HMC03680 | CG5836 | SF1 | y1 sc v1; P{TRiP}attP40 |
| 34864 | HMS00182 | CG11680 | mle | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 34651 | HMS01126 | CG16778 | CG16778 | y1 sc v1; P{TRiP}attP2 |
| 35204 | GL00079 | CG7843 | Ars2 | y1 sc v1; P{TRiP}attP2 |
| 34829 | HMS00144 | CG10279 | Rm62 | y1 sc v1; P{TRiP}attP2 |
| 42796 | GL01167 | CG3151 | Rbp9 | y1 v1; P{TRiP}attP2 |
| 44431 | GLC01382 | CG11266 | CG11266 | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 38904 | GL00675 | CG5654 | yps | y1 sc v1; P{TRiP}attP2 |

| | | | | |
|---|---|---|---|---|
| 32990 | HMS00790 | CG3606 | caz | y1 sc v1; P{TRiP}attP2 |
| 55173 | HMC03853 | CG2063 | CG2063 | y1 sc v1; P{TRiP}attP40/CyO |
| 55254 | HMC03941 | CG4602 | Srp54 | y1 sc v1; P{TRiP}attP40/CyO |
| 37519 | HMS01661 | CG10851 | B52 | y1 sc v1; P{TRiP}attP40 |
| 32367 | HMS00358 | CG6987 | SF2 | y1 sc v1; P{TRiP}attP2 |
| 55642 | HMC03790 | CG10203 | x16 | y1 sc v1; P{TRiP}attP40 |
| 50574 | GLC01696 | CG1987 | Rbp1-like | y1 v1/y1 sc v1; P{TRiP}attP2 |
| 52936 | HMC03678 | CG7971 | CG7971 | y1 sc v1; P{TRiP}attP40 |
| 36578 | GL00538 | CG11274 | SRm160 | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 33648 | HMS00052 | CG7035 | Cbp80 | y1 sc v1; P{TRiP}attP2 |
| 34797 | HMS00106 | CG12357 | Cbp20 | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 55209 | HMC03927 | CG30122 | CG30122 | y1 sc v1; P{TRiP}attP40/CyO |
| 35627 | GL00473 | CG16901 | sqd | y1 sc v1; P{TRiP}attP2 |
| 34622 | HMS01297 | CG8877 | Prp8 | y1 sc v1; P{TRiP}attP2 |
| 34024 | HMS00994 | CG5931 | l(3)72Ab | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 55701 | HMC03917 | CG4849 | CG4849 | y1 sc v1; P{TRiP}attP40 |
| 51909 | HMC03484 | CG6841 | CG6841 | y1 v1; P{TRiP}attP40 |
| 34857 | HMS00175 | CG10333 | CG10333 | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 38968 | HMS01882 | CG5198 | holn1 | y1 sc v1; P{TRiP}attP2 |
| 55207 | HMC03922 | CG3436 | CG3436 | y1 sc v1; P{TRiP}attP40/CyO |
| 55171 | HMC03851 | CG3058 | Dim1 | y1 sc v1; P{TRiP}attP40 |
| 44458 | GLC01610 | CG10418 | CG10418 | y1 sc v1; P{TRiP}attP2 |
| 55156 | HMC03829 | CG17764 | CG17764 | y1 sc v1; P{TRiP}attP40 |
| 54798 | HMJ21492 | CG9344 | CG9344 | y1 v1; P{TRiP}attP40/CyO |
| 52932 | HMC03674 | CG13277 | LSm7 | y1 sc v1; P{TRiP}attP40/CyO |
| 54007 | HMJ21430 | CG2021 | CG2021 | y1 v1; P{TRiP}attP40/CyO |
| 55256 | HMC03943 | CG7757 | CG7757 | y1 sc v1; P{TRiP}attP40 |
| 55210 | HMC03931 | CG6322 | U4-U6-60K | y1 sc v1; P{TRiP}attP40/CyO |
| 55364 | HMC04052 | CG6876 | Prp31 | y1 sc v1; P{TRiP}attP40 |
| 53267 | GLC01828 | CG6686 | CG6686 | y1 sc v1; P{TRiP}attP2 |
| 55632 | HMC03779 | CG7288 | CG7288 | y1 sc v1; P{TRiP}attP40/CyO |
| 43237 | GLC01424 | CG18426 | ytr | y1 sc v1; P{TRiP}attP2 |
| 44524 | HMC02918 | CG6905 | CG6905 | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 35684 | GLV21049 | CG4264 | Hsc70-4 | y1 sc v1; P{TRiP}attP2 |
| 32865 | HMS00652 | CG5519 | Prp19 | y1 sc v1; P{TRiP}attP2 |
| 34952 | HMS00382 | CG12135 | c12.1 | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 55251 | HMC03938 | CG2685 | CG2685 | y1 sc v1; P{TRiP}attP40 |
| 54852 | HMJ21589 | CG11820 | PQBP1 | y1 v1; P{TRiP}attP40 |
| 55172 | HMC03852 | CG31368 | CG31368 | y1 sc v1; P{TRiP}attP40 |
| 29535 | HM05215 | CG3193 | crn | y1 sc v1; P{TRiP}attP2 |
| 34777 | HMS00086 | CG8264 | Bx42 | y1 sc v1; P{TRiP}attP2 |
| 35611 | GL00451 | CG4886 | cyp33 | y1 sc v1; P{TRiP}attP2 |
| 31758 | HM04069 | CG9667 | CG9667 | y1 v1; P{TRiP}attP2 |
| 33380 | HMS00255 | CG1639 | l(1)10Bb | y1 sc v1; P{TRiP}attP2 |
| 55703 | HMC03926 | CG17168 | CG17168 | y1 sc v1; P{TRiP}attP40 |
| 55263 | HMC03950 | CG10466 | CG10466 | y1 sc v1; P{TRiP}attP40 |
| 35285 | GL00186 | CG7028 | CG7028 | y1 sc v1; P{TRiP}attP2 |

| | | | | |
|---|---|---|---|---|
| 55208 | HMC03925 | CG5808 | CG5808 | y1 sc v1; P{TRiP}attP40 |
| 55363 | HMC04051 | CG18005 | beag | y1 sc v1; P{TRiP}attP40 |
| 33739 | HMS01075 | CG5451 | Smu1 | y1 sc v1; P{TRiP}attP2 |
| 29536 | HM05216 | CG1017 | Mfap1 | y1 sc v1; P{TRiP}attP2 |
| 32838 | HMS00528 | CG8241 | pea | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 54824 | HMJ21543 | CG8833 | CG8833 | y1 v1; P{TRiP}attP40/CyO |
| 55365 | HMC04053 | CG3225 | CG3225 | y1 sc v1; P{TRiP}attP40 |
| 34565 | HMS01037 | CG6015 | CG6015 | y1 sc v1; P{TRiP}attP2 |
| 55697 | HMC03911 | CG12343 | CG12343 | y1 sc v1; P{TRiP}attP40 |
| 38202 | GL00641 | CG7483 | eIF4AIII | y1 sc v1; P{TRiP}attP40 |
| 36585 | GL00545 | CG8781 | tsu | y1 sc v1; P{TRiP}attP2 |
| 35453 | GL00378 | CG9401 | mago | y1 sc v1; P{TRiP}attP2 |
| 53676 | HMC03673 | CG10473 | Acn | y1 sc v1; P{TRiP}attP40 |
| 33666 | HMS00076 | CG7269 | Hel25E | y1 sc v1; P{TRiP}attP2 |
| 55743 | HMC03928 | CG8383 | Pnn | y1 sc v1; P{TRiP}attP40 |
| 34626 | HMS01301 | CG1101 | Ref1 | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 36781 | GL00127 | CG6046 | Bin1 | y1 sc v1; P{TRiP}attP2 |
| 55635 | HMC03782 | CG5215 | Zn72D | y1 sc v1; P{TRiP}attP40 |
| 34848 | HMS00166 | CG2910 | nito | y1 sc v1; P{TRiP}attP2 |
| 53247 | GLC01806 | CG5119 | pAbp | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 55692 | HMC03906 | CG5649 | kin17 | y1 sc v1; P{TRiP}attP40 |
| 38198 | GL00637 | CG7111 | Rack1 | y1 sc v1; P{TRiP}attP40 |
| 35572 | GL00088 | CG16910 | key | y1 sc v1; P{TRiP}attP2 |
| 34602 | HMS00553 | CG2163 | Pabp2 | y1 sc v1; P{TRiP}attP2 |
| 52933 | HMC03675 | CG4211 | nonA | y1 sc v1; P{TRiP}attP40 |
| 55204 | HMC03919 | CG10907 | CG10907 | y1 sc v1; P{TRiP}attP40 |
| 43285 | HMS02657 | CG1474 | Es2 | y1 sc v1; P{TRiP}attP40/CyO |
| 34025 | HMS00995 | CG6066 | CG6066 | y1 sc v1; P{TRiP}attP2 |
| 27703 | JF02783 | CG8597 | lark | y1 v1; P{TRiP}attP2 |
| 35325 | GL00232 | CG1362 | cdc2rk | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 55155 | HMC03814 | CG8610 | Cdc27 | y1 sc v1; P{TRiP}attP40/CyO |
| 35201 | GL00076 | CG6137 | aub | y1 sc v1; P{TRiP}attP2 |
| 31103 | JF01568 | CG5316 | CG5316 | y1 v1; P{TRiP}attP2 |
| 38931 | GL01090 | CG14648 | lost | y1 sc v1; P{TRiP}attP2 |
| 55252 | HMC03939 | CG7564 | CG7564 | y1 sc v1; P{TRiP}attP40 |
| 55159 | HMC03834 | CG3613 | qkr58E-1 | y1 sc v1; P{TRiP}attP40 |
| 55248 | HMC03935 | CG1591 | REG | y1 sc v1; P{TRiP}attP40/CyO |
| 32944 | HMS00738 | CG6143 | Pep | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 36880 | GL01044 | CG9684 | CG9684 | y1 sc v1; P{TRiP}attP2 |
| 41816 | GL01244 | CG8994 | exu | y1 v1; P{TRiP}attP2 |
| 55747 | HMC03934 | CG15747 | CG15747 | y1 sc v1; P{TRiP}attP40 |
| 35355 | GL00267 | CG1609 | Gcn2 | y1 sc v1; P{TRiP}attP2 |
| 33392 | HMS00269 | CG9412 | rin | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 41924 | HMS02321 | CG13597 | CG13597 | y1 sc v1; P{TRiP}attP2 |
| 55648 | HMC03797 | CG1622 | CG1622 | y1 sc v1; P{TRiP}attP40 |
| 55690 | HMC03904 | CG5877 | CG5877 | y1 sc v1; P{TRiP}attP40 |
| 55203 | HMC03916 | CG6209 | CG6209 | y1 sc v1; P{TRiP}attP40 |

| | | | | |
|---|---|---|---|---|
| 55356 | HMC04043 | CG12250 | ymp | y1 sc v1; P{TRiP}attP40 |
| 55654 | HMC03804 | CG8079 | CG8079 | y1 sc v1; P{TRiP}attP40 |
| 55960 | HMC04254 | CG8614 | Neos | y1 sc v1; P{TRiP}attP40/CyO |
| 55702 | HMC03923 | CG17187 | CG17187 | y1 sc v1; P{TRiP}attP40/CyO |
| 36640 | GL00600 | CG5343 | CG5343 | y1 sc v1; P{TRiP}attP2 |
| 33408 | HMS00287 | CG6539 | Gem3 | y1 sc v1; P{TRiP}attP2 |
| 29622 | JF03301 | CG3983 | ns1 | y1 v1; P{TRiP}attP2 |
| 50529 | GLC01647 | CG18497 | spen | y1 v1; P{TRiP}attP40 |
| 55741 | HMC03918 | CG33547 | Rim | y1 sc v1; P{TRiP}attP40 |
| 34809 | HMS00118 | CG7292 | Rrp6 | y1 sc v1; P{TRiP}attP2/TM3,Sb |
| 55369 | HMC04057 | CG17098 | CG17098 | y1 sc v1; P{TRiP}attP40/CyO |

**5.2.2 Supplementary table 2. List of preliminary splicing candidates form germ line specific screen**

| | Drosophila | | Human | Egg | | | | Larvae | | | Phenotype Descriptions |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Egg laying | | Egg Shape | | No Larvae | <10 Larvae | >10 Larvae | |
| | | | | Normal | Less | Normal | Abnormal | | | | |
| No. | CG | Gene | | | | | | | | | |
| 1 | Control | mCherry | | ✔ | | ✔ | | | | ✔ | Control |
| 2 | 16901 | Squid | | ✔ | | | ✔ | ✔ | | | Spindle Phenotype |
| 3 | 31368 | | Aquarius | ✔ | | | ✔ | | ✔ | | |
| 4 | 7971 | | SRm300 | | ✔ | | ✔ | | ✔ | | |
| 5 | 5877 | | | | ✔ | ✔ | | ✔ | | | |
| 6 | 6143 | Pep | | ✔ | | | ✔ | ✔ | | | |
| 7 | 10473 | Acinus | Acinus | ✔ | | ✔ | | | ✔ | | Embryonic Phenotype |
| 8 | 4211 | non-A | PSF | ✔ | | ✔ | | | ✔ | | |
| 9 | 11266 | Caper | RBM39/RNPC2 | ✔ | | ✔ | | ✔ | | | |
| 10 | 5451 | Smu1 | hSmu-1 | ✔ | | ✔ | | ✔ | | | |
| 11 | 6137 | aubergine | PIWIL1 | ✔ | | ✔ | | ✔ | | | |
| 12 | 8994 | exu | | ✔ | | ✔ | | | ✔ | | |
| 13 | 8781 | tsunagi | Y14 | ✔ | | | ✔ | ✔ | | | Short eggs phenotype |
| 14 | 5215 | Zn72D | ZFR2 | | ✔ | | ✔ | ✔ | | | |

### 5.2.3 Supplementary table 3. List of primers

| Complete list of primers | |
|---|---|
| **Primer name** | **Primer sequence** |
| Actin qRT Forward | TGGATACTCCTCCCGACACA |
| Actin qRT Reverse | AGTCTTTCGGTTTGGTGTCTCT |
| GAPDH qRT Forward | CGGCCATAGCGAAAATCGTG |
| GAPDH qRT Reverse | TTCTCGTGCGTCTCGTTGAT |
| Gurken qRT Forward | GCGCGCAACAAGACCTAAA |
| Gurken qRT Reverse | GTTAATCTAAAGAGCAGCAAGCG |
| Gurken First Intron qRT Forward | GCGTTCGTGCGACAGAAAATG |
| Gurken First Intron qRT Reverse | GGGGTCTAAACGATCGAGGG |
| Gurken Second Intron qRT Forward | AAGTTGCCGCACTAAAAACTGA |
| Gurken Second Intron qRT Reverse | TGTGCTGATGCTGCACAATTT |
| Gurken Third Intron qRT Forward | TGGAACGGATGGAACTAACGA |
| Gurken Third Intron qRT Reverse | CGCTGTTGGAGGCGAATAGA |
| Salsa qRT Forward | TATCGAAGATGCCGTCAGCC |
| Salsa qRT Reverse | CCACTTCGACAACGGCAAAG |
| Gurken First Intron RT Forward | GCGCGCAACAAGACCTAAAATC |
| Gurken First Intron RT Reverse | CATCATTGGAAAACGCTTGGG |
| Gurken Second Intron RT Forward | GCCACCCCAAGCGTTTTC |
| Gurken Second Intron RT Reverse | TGGGAGTCGTGGAGTCAGG |
| Gurken Third Intron RT Forward | GGTTGAGTTGCTCAATCGCC |
| Gurken Third Intron RT Reverse | CTTGTGCGTCTT TTGCAAC |
| eIF4E RB qRT Forward | GGAGACGGAGAAGTTTTTTGG |
| eIF4E RB qRT Reverse | GAGCGAAGCTTTTGATTTCTG |
| eIF4E RC qRT Forward | GTAACCTACGCAGCTTGAGTG |
| eIF4E RC qRT Reverse | CGCTGGTCTTCTCCGTCTC |
| eIF4E qRT Forward | GGACGCTGTGGTACCTTGAA |
| eIF4E qRT Reverse | GGGGGCTTGATGTGGTTGTA |
| eIF4E RT Forward | GCCCAGTAACCTACG CAGCTTGAG |

| | |
|---|---|
| eIF4E RT Reverse | CTCGTTTTGCATGTCCTCCCAGG |
| Salsa_attB1_Fwd | GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAT GAAGCGAAGAAGTCAAACTAG |
| Salsa_ attB2_Rev (stop) | GGGGACCACTTTGTACAAGAAAGCTGGGTCCTA CTAAGACTCTTCAGCTGGAGC |
| Salsa_ attB2_Rev (without Stop) | GGGGACCACTTTGTACAAGAAAGCTGGGTCAG ACTCTTCAGCTGGAGCC |
| Salsa_ RNAi -1- TS | ctagcagtCGCTTGGATATGGACGATCTAtagttatattc aagcataTAGATCGTCCATATCCAAGCGgcg |
| Salsa_ RNAi -1 - BS | aattcgcCGCTTGGATATGGACGATCTAtatgcttgaata taactaTAGATCGTCCATATCCAAGCGactg |
| Salsa_ RNAi -3- TS | ctagcagtCCACGATTATCTCCTACGCAAtagttatattca agcataTTGCGTAGGAGATAATCGTGGgcg |
| Salsa_ RNAi -3 - BS | aattcgcCCACGATTATCTCCTACGCAAtatgcttgaatat aactaTTGCGTAGGAGATAATCGTGGactg |
| pVALIUM22- Fw | GGTGATAGAGCCTGAACCAG |
| pVALIUM22- Rev | TAATCGTGTGTGATGCCTACC |
| Tra2 -Fw | GTTGAAGCGTGCGTCTTTCT |
| Tra2 -Rev | CGCCGATCGCTTGTGTTTAT |

### 5.2.4 Supplementary table 4. Primer efficiency and regression curve values of primers used in RT-qPCR

| No. | Name of primer | Efficiency | Regression curve |
|---|---|---|---|
| 1 | Actin | 108 | 0.994 |
| 2 | GAPDH | 108.9 | 0.992 |
| 3 | Gurken total | 107.1 | 0.981 |
| 4 | Gurken- First intron | 109.9 | 0.98 |
| 5 | Gurken- Second intron | 105.9 | 0.97 |
| 6 | Gurken- Third intron | 99 | 0.98 |
| 7 | Salsa | 108 | 0.98 |
| 8 | eIF4E-total | 108.1 | 0.99 |
| 9 | eIF4E-total | 108.6 | 0.99 |
| 10 | eIF4E-total | 109.7 | 0.99 |

### 5.3 References

**Abdu, U.** (2006). spn-F encodes a novel protein that affects oocyte patterning and bristle morphology in Drosophila. *Development* **133**, 1477–1484.

**Abdu, U., Brodsky, M. and Schüpbach, T.** (2002). Activation of a Meiotic Checkpoint during Drosophila Oogenesis Regulates the Translation of Gurken through Chk2/Mnk. *Curr. Biol.* **12**, 1645–1651.

**Absmeier, E., Santos, K. F. and Wahl, M. C.** (2016). Functions and regulation of the Brr2 RNA helicase during splicing. *Cell Cycle* **15**, 3362–3377.

**Akay, A., Di Domenico, T., Suen, K. M., Nabih, A., Parada, G. E., Larance, M., Medhi, R., Berkyurek, A. C., Zhang, X., Wedeles, C. J., et al.** (2017). The Helicase Aquarius/EMB-4 Is Required to Overcome Intronic Barriers to Allow Nuclear RNAi Pathways to Heritably Silence Transcription. *Dev. Cell* **42**, 241–255.e6.

**Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., Lev-Maor, G., Burstein, D., Schwartz, S., Postolsky, B., et al.** (2012a). Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition. *Cell Rep.* **1**, 543–556.

**Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., Lev-Maor, G., Burstein, D., Schwartz, S., Postolsky, B., et al.** (2012b). Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition. *Cell Rep.* **1**, 543–556.

**Armache, K.-J., Mitterweger, S., Meinhart, A. and Cramer, P.** (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem.* **280**, 7131–4.

**Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M. and Stark, A.** (2013). Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science (80-. ).* **339**, 1074–1077.

**Aronica, L., Kasparek, T., Ruchman, D., Marquez, Y., Cipak, L., Cipakova, I., Anrather, D., Mikolaskova, B., Radtke, M., Sarkar, S., et al.** (2016). The spliceosome-associated protein Nrl1 suppresses homologous recombination-dependent R-loop formation in fission yeast. *Nucleic Acids Res.* **44**, 1703–17.

**Awasthi, S., Verma, M., Mahesh, A., K Khan, M. I., Govindaraju, G., Rajavelu, A., Chavali, P. L., Chavali, S. and Dhayalan, A.** (2018). DDX49 is an RNA helicase that affects translation by regulating mRNA export and the levels of pre-ribosomal RNA. *Nucleic Acids Res.*

**Barckmann, B., Pierson, S., Dufourt, J., Papin, C., Armenise, C., Port, F., Grentzinger, T., Chambeyron, S., Baronian, G., Desvignes, J.-P., et al.** (2015). Aubergine iCLIP Reveals piRNA-Dependent Decay of mRNAs Involved in Germ Cell Development in the Early Embryo. *Cell Rep.* **12**, 1205–1216.

**Baudat, F., Manova, K., Yuen, J. P., Jasin, M. and Keeney, S.** (2000). Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Mol. Cell* **6**, 989–98.

**Baudat, F., Imai, Y. and de Massy, B.** (2013). Meiotic recombination in mammals: localization and regulation. *Nat. Rev. Genet.* **14**, 794–806.

**Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M.**

**and Nüsslein-Volhard, C.** (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. *EMBO J.* **7**, 1749–56.

**Bertram, K., Agafonov, D. E., Dybkov, O., Haselbach, D., Leelaram, M. N., Will, C. L., Urlaub, H., Kastner, B., Lührmann, R. and Stark, H.** (2017). Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* **170**, 701–713.e11.

**Boon, K.-L., Auchynnikava, T., Edwalds-Gilbert, G., Barrass, J. D., Droop, A. P., Dez, C. and Beggs, J. D.** (2006). Yeast ntr1/spp382 mediates prp43 function in postspliceosomes. *Mol. Cell. Biol.* **26**, 6016–23.

**Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. and Blencowe, B. J.** (2013a). Dynamic Integration of Splicing within Gene Regulatory Pathways. *Cell* **152**, 1252–1269.

**Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. and Blencowe, B. J.** (2013b). Dynamic Integration of Splicing within Gene Regulatory Pathways. *Cell* **152**, 1252–1269.

**Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M. and Blencowe, B. J.** (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24**, 1774–86.

**Brody, E. and Abelson, J.** (1985). The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228**, 963–7.

**Brook, M., Smith, J. W. S. and Gray, N. K.** (2009). The DAZL and PABP families: RNA-binding proteins with interrelated roles in translational control in oocytes. *Reproduction* **137**, 595–617.

**Brooks, A. N., Duff, M. O., May, G., Yang, L., Bolisetty, M., Landolin, J., Wan, K., Sandler, J., Booth, B. W., Celniker, S. E., et al.** (2015). Regulation of alternative splicing in Drosophila by 56 RNA binding proteins. *Genome Res.* **25**, 1771.

**Bullock, S. L. and Ish-Horowicz, D.** (2001). Conserved signals and machinery for RNA transport in Drosophila oogenesis and embryogenesis. *Nature* **414**, 611–616.

**Buratowski, S.** (2009). Progression through the RNA Polymerase II CTD Cycle. *Mol. Cell* **36**, 541–546.

**Burkard, K. T. and Butler, J. S.** (2000). A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p. *Mol. Cell. Biol.* **20**, 604–16.

**Cao, Q. and Richter, J. D.** (2002). Dissolution of the maskin-eIF4E complex by cytoplasmic polyadenylation and poly(A)-binding protein controls cyclin B1 mRNA translation and oocyte maturation. *EMBO J.* **21**, 3852–3862.

**Cha, S. H., Sekine, T., Fukushima, J. I., Kanai, Y., Kobayashi, Y., Goya, T. and Endou, H.** (2001). Identification and characterization of human organic anion transporter 3 expressing predominantly in the kidney. *Mol. Pharmacol.* **59**, 1277–86.

**Chan, S.-P. and Cheng, S.-C.** (2005). The Prp19-associated Complex Is Required for Specifying Interactions of U5 and U6 with Pre-mRNA during Spliceosome Activation. *J. Biol. Chem.* **280**, 31190–31199.

**Chan, S.-P., Kao, D.-I., Tsai, W.-Y. and Cheng, S.-C.** (2003). The Prp19p-

Associated Complex in Spliceosome Activation. *Science (80-. ).* **302**, 279–282.

**Chan, S., Choi, E.-A. and Shi, Y.** (2011). Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip. Rev. RNA* **2**, 321–35.

**Chan, S. L., Huppertz, I., Yao, C., Weng, L., Moresco, J. J., Yates, J. R., Ule, J., Manley, J. L. and Shi, Y.** (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. *Genes Dev.* **28**, 2370–2380.

**Chang, J. S., Tan, L., Wolf, M. R. and Schedl, P.** (2001). Functioning of the Drosophila orb gene in gurken mRNA localization and translation. *Development* **128**, 3169–77.

**Chang, Y., Zhang, X., Horton, J. R., Upadhyay, A. K., Spannhoff, A., Liu, J., Snyder, J. P., Bedford, M. T. and Cheng, X.** (2009). Structural basis for G9a-like protein lysine methyltransferase inhibition by BIX-01294. *Nat. Struct. Mol. Biol.* **16**, 312–317.

**Chen, H.-C. and Cheng, S.-C.** (2012). Functional roles of protein splicing factors. *Biosci. Rep.* **32**, 345–59.

**Chen, W. and Moore, M. J.** (2014). The spliceosome: disorder and dynamics defined. *Curr. Opin. Struct. Biol.* **24**, 141–9.

**Cheng, H., Dufu, K., Lee, C.-S., Hsu, J. L., Dias, A. and Reed, R.** (2006). Human mRNA Export Machinery Recruited to the 5′ End of mRNA. *Cell* **127**, 1389–1400.

**Clark, I., Giniger, E., Ruohola-Baker, H., Jan, L. Y. and Jan, Y. N.** (1994). Transient posterior localization of a kinesin fusion protein reflects anteroposterior polarity of the Drosophila oocyte. *Curr. Biol.* **4**, 289–300.

**Clark, J., Lu, Y.-J., Sidhar, S. K., Parker, C., Gill, S., Smedley, D., Hamoudi, R., Linehan, W. M., Shipley, J. and Cooper, C. S.** (1997). Fusion of splicing factor genes PSF and NonO (p54nrb) to the TFE3 gene in papillary renal cell carcinoma. *Oncogene* **15**, 2233–2239.

**Cohen, S. and Greenberg, M. E.** (2008). Communication Between the Synapse and the Nucleus in Neuronal Development, Plasticity, and Disease. *Annu. Rev. Cell Dev. Biol.* **24**, 183–209.

**Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., et al.** (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13.

**Cook, O.** (2004). brinker and optomotor-blind act coordinately to initiate development of the L5 wing vein primordium in Drosophila. *Development* **131**, 2113–2124.

**Coppola, J. A. and Luse, D. S.** (1984). Purification and characterization of ternary complexes containing accurately initiated RNA polymerase II and less than 20 nucleotides of RNA. *J. Mol. Biol.* **178**, 415–437.

**David, C. J., Boyne, A. R., Millhouse, S. R. and Manley, J. L.** (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev.* **25**, 972–983.

**De, I., Bessonov, S., Hofele, R., dos Santos, K., Will, C. L., Urlaub, H., Lührmann, R. and Pena, V.** (2015). The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat. Struct. Mol. Biol.* **22**, 138–144.

**de Almeida, R. A. and O'Keefe, R. T.** (2015). The NineTeen Complex (NTC) and

NTC-associated proteins as targets for spliceosomal ATPase action during pre-mRNA splicing. *RNA Biol.* **12**, 109–14.

**Derrick, C. J. and Weil, T. T.** (2017). Translational control of *gurken* mRNA in *Drosophila* development. *Cell Cycle* **16**, 23–32.

**Desai, A. and Mitchison, T. J.** (1997). MICROTUBULE POLYMERIZATION DYNAMICS. *Annu. Rev. Cell Dev. Biol.* **13**, 83–117.

**Ding, Y., Shah, P. and Plotkin, J. B.** (2012). Weak 5'-mRNA secondary structures in short eukaryotic genes. *Genome Biol. Evol.* **4**, 1046–53.

**Driever, W. and Nüsslein-Volhard, C.** (1988a). The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell* **54**, 95–104.

**Driever, W. and Nüsslein-Volhard, C.** (1988b). A gradient of bicoid protein in Drosophila embryos. *Cell* **54**, 83–93.

**Driever, W., Siegel, V. and Nüsslein-Volhard, C.** (1990). Autonomous determination of anterior structures in the early Drosophila embryo by the bicoid morphogen. *Development* **109**, 811–20.

**Drummond, D. R., Armstrong, J. and Colman, A.** (1985). The effect of capping and polyadenylation on the stability, movement and translation of synthetic messenger RNAs in Xenopus oocytes. *Nucleic Acids Res.* **13**, 7375–94.

**Dujardin, G., Lafaille, C., Petrillo, E., Buggiano, V., Gómez Acuña, L. I., Fiszbein, A., Godoy Herz, M. A., Nieto Moreno, N., Muñoz, M. J., Alló, M., et al.** (2013). Transcriptional elongation and alternative splicing. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829**, 134–140.

**Egloff, S. and Murphy, S.** (2008). Cracking the RNA polymerase II CTD code. *Trends Genet.* **24**, 280–288.

**Ephrussi, A. and Lehmann, R.** (1992). Induction of germ cell formation by oskar. *Nature* **358**, 387–392.

**Ephrussi, A., Dickinson, L. K. and Lehmann, R.** (1991). Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell* **66**, 37–50.

**Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H. and Lührmann, R.** (2009). The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome. *Mol. Cell* **36**, 593–608.

**Fededa, J. P. and Kornblihtt, A. R.** (2008). A splicing regulator promotes transcriptional elongation. *Nat. Struct. Mol. Biol.* **15**, 779–781.

**Ferrandon, D., Elphick, L., Nüsslein-Volhard, C. and St Johnston, D.** (1994). Staufen protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner. *Cell* **79**, 1221–32.

**Ferreira, T., Prudêncio, P. and Martinho, R. G.** (2014). Drosophila protein kinase N (Pkn) is a negative regulator of actin–myosin activity during oogenesis. *Dev. Biol.* **394**, 277–291.

**Findley, S. D.** (2003). Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. *Development* **130**, 859–871.

**Fong, Y. W. and Zhou, Q.** (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**, 929–933.

**Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.-D. and Bentley, D. L.** (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* **28**, 2663–76.

**Gatfield, D., Le Hir, H., Schmitt, C., Braun, I. C., Köcher, T., Wilm, M. and Izaurralde, E.** (2001). The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. *Curr. Biol.* **11**, 1716–1721.

**Gavis, E. R. and Lehmann, R.** (1992). Localization of nanos RNA controls embryonic polarity. *Cell* **71**, 301–313.

**Ghabrial, A. and Schüpbach, T.** (1999). Activation of a meiotic checkpoint regulates translation of Gurken during Drosophila oogenesis. *Nat. Cell Biol.* **1**, 354–357.

**Ghabrial, A., Ray, R. P. and Schupbach, T.** (1998). okra and spindle-B encode components of the RAD52 DNA repair pathway and affect meiosis and patterning in Drosophila oogenesis. *Genes Dev.* **12**, 2711–2723.

**Gingras, A.-C., Raught, B. and Sonenberg, N.** (1999). eIF4 Initiation Factors: Effectors of mRNA Recruitment to Ribosomes and Regulators of Translation. *Annu. Rev. Biochem.* **68**, 913–963.

**Gómez-González, B., García-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marín, A., Foiani, M. and Aguilera, A.** (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J.* **30**, 3106–19.

**Gonatopoulos-Pournatzis, T. and Cowling, V. H.** (2014). Cap-binding complex (CBC). *Biochem. J.* **457**, 231–42.

**González-Reyes, A., Elliott, H. and St Johnston, D.** (1995). Polarization of both major body axes in Drosophila by gurken-torpedo signalling. *Nature* **375**, 654–658.

**Gorjánácz, M., Klerkx, E. P. F., Galy, V., Santarella, R., López-Iglesias, C., Askjaer, P. and Mattaj, I. W.** (2007). Caenorhabditis elegans BAF-1 and its kinase VRK-1 participate directly in post-mitotic nuclear envelope assembly. *EMBO J.* **26**, 132–143.

**Grainger, R. J., Barrass, J. D., Jacquier, A., Rain, J.-C. and Beggs, J. D.** (2009). Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome. *RNA* **15**, 2161–73.

**GRAINGER, R. J. and Beggs, J. D.** (2005). Prp8 protein: At the heart of the spliceosome. *RNA* **11**, 533–557.

**Grote, M., Wolf, E., Will, C. L., Lemm, I., Agafonov, D. E., Schomburg, A., Fischle, W., Urlaub, H. and Lührmann, R.** (2010). Molecular architecture of the human Prp19/CDC5L complex. *Mol. Cell. Biol.* **30**, 2105–19.

**Guilgur, L. G., Prudêncio, P., Sobral, D., Liszekova, D., Rosa, A. and Martinho, R. G.** (2014). Requirement for highly efficient pre-mRNA splicing during Drosophila early embryonic development. *Elife* **3**, e02181.

**Hachet, O. and Ephrussi, A.** (2001). Drosophila Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. *Curr. Biol.* **11**, 1666–74.

**Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F. S. and Brennecke, J.** (2013). The genetic makeup of the Drosophila piRNA pathway. *Mol. Cell* **50**, 762–77.

**Hartmann, B., Castelo, R., Miñana, B., Peden, E., Blanchette, M., Rio, D. C., Singh, R. and Valcárcel, J.** (2011). Distinct regulatory programs establish widespread sex-specific alternative splicing in Drosophila melanogaster. *RNA* **17**, 453–68.

**Herold, N., Will, C. L., Wolf, E., Kastner, B., Urlaub, H. and Luhrmann, R.** (2009). Conservation of the Protein Composition and Electron Microscopy Structure of Drosophila melanogaster and Human Spliceosomal Complexes. *Mol. Cell. Biol.* **29**, 281–301.

**Hilleren, P., McCarthy, T., Rosbash, M., Parker, R. and Jensen, T. H.** (2001). Quality control of mRNA 3′-end processing is linked to the nuclear exosome. *Nature* **413**, 538–542.

**Hiremath, L. S., Hiremath, S. T., Rychlik, W., Joshi, S., Domier, L. L. and Rhoads$, R. E.** (1989). THE JOURNAL OF BIOLOGICAL CHEMISTRY In Vitro Synthesis, Phosphorylation, and Localization on 48 S Initiation Complexes of Human Protein Synthesis Initiation Factor 4E*. **264**, 1132–1138.

**Hocine, S., Singer, R. H. and Grünwald, D.** (2010). RNA processing and export. *Cold Spring Harb. Perspect. Biol.* **2**, a000752.

**Hogg, R., McGrail, J. C. and O'Keefe, R. T.** (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem. Soc. Trans.* **38**, 1110–1115.

**Horiuchi, T., Giniger, E. and Aigaki, T.** (2003). Alternative trans-splicing of constant and variable exons of a Drosophila axon guidance gene, lola. *Genes Dev.* **17**, 2496–501.

**Houseley, J. and Tollervey, D.** (2009). The Many Pathways of RNA Degradation. *Cell* **136**, 763–776.

**Huang, C., Zheng, X., Zhao, H., Li, M., Wang, P., Xie, Z., Wang, L. and Zhong, Y.** (2012). A Permissive Role of Mushroom Body α/β Core Neurons in Long-Term Memory Consolidation in Drosophila. *Curr. Biol.* **22**, 1981–1989.

**Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L. E., Sibley, C. R., Sugimoto, Y., Tajnik, M. and König, J.** (2014). iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods* **65**, 274–287.

**Hur, J. K., Luo, Y., Moon, S., Ninova, M., Marinov, G. K., Chung, Y. D. and Aravin, A. A.** (2016). Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in Drosophila. *Genes Dev.* **30**, 840–55.

**Ideue, T., Sasaki, Y. T. F., Hagiwara, M. and Hirose, T.** (2007). Introns play an essential role in splicing-dependent formation of the exon junction complex. *Genes Dev.* **21**, 1993–8.

**Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallières, M., Tapial, J., Raj, B., O'Hanlon, D., et al.** (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–23.

**Ivanovska, I., Khandan, T., Ito, T. and Orr-Weaver, T. L.** (2005). A histone code in meiosis: the histone kinase, NHK-1, is required for proper chromosomal architecture in Drosophila oocytes. *Genes Dev.* **19**, 2571–82.

**Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E. and Mattaj, I. W.** (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657–68.

**Jagut, M., Mihaila-Bodart, L., Molla-Herman, A., Alin, M.-F., Lepesant, J.-A. and Huynh, J.-R.** (2013). A mosaic genetic screen for genes involved in the early steps of Drosophila oogenesis. *G3 (Bethesda).* **3**, 409–25.

**Januschke, J. and Gonzalez, C.** (2008). Drosophila asymmetric division, polarity and cancer. *Oncogene* **27**, 6994–7002.

**Jaramillo, M. A., Callejas, R., Davidson, C., Smith, J. F., Stevens, A. C. and Tepe, E. J.** A Phylogeny of the Tropical Genus Piper Using ITS and the Chloroplast Intron psbJ–petA.

**Jia, T., Zhang, B., You, C., Zhang, Y., Zeng, L., Li, S., Johnson, K. C. M., Yu, B., Li, X. and Chen, X.** (2017). The Arabidopsis MOS4-Associated Complex Promotes MicroRNA Biogenesis and Precursor Messenger RNA Splicing. *Plant Cell* **29**, 2626–2643.

**Johnstone, O. and Lasko, P.** (2001). Translational Regulation and RNA Localization in *Drosophila* Oocytes and Embryos. *Annu. Rev. Genet.* **35**, 365–406.

**Jones, B.** (2015). Gene expression: Layers of gene regulation. *Nat. Rev. Genet. 2015 163*.

**Jonkers, I., Kwak, H. and Lis, J. T.** (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**, e02407.

**Joyce, E. F., Pedersen, M., Tiong, S., White-Brown, S. K., Paul, A., Campbell, S. D. and McKim, K. S.** (2011). *Drosophila* ATM and ATR have distinct activities in the regulation of meiotic DNA damage and repair. *J. Cell Biol.* **195**, 359–367.

**Jurica, M. S. and Moore, M. J.** (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**, 5–14.

**Kahvejian, A., Roy, G. and Sonenberg, N.** (2001). The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. *Cold Spring Harb. Symp. Quant. Biol.* **66**, 293–300.

**Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L. and Dreyfuss, G.** (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664–668.

**Kamenska, A., Simpson, C. and Standart, N.** (2014). eIF4E-binding proteins: new factors, new locations, new roles. *Biochem. Soc. Trans.* **42**, 1238–45.

**Kanno, T., Lin, W.-D., Fu, J. L., Matzke, A. J. M. and Matzke, M.** (2017). A genetic screen implicates a CWC16/Yju2/CCDC130 protein and SMU1 in alternative splicing inArabidopsis thaliana. *RNA* **23**, 1068–1079.

**Katahira, J.** (2012a). mRNA export and the TREX complex. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**, 507–513.

**Katahira, J.** (2012b). mRNA export and the TREX complex. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**, 507–513.

**Keeney, S.** (2008). Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. In *Recombination and Meiosis*, pp. 81–123. Berlin, Heidelberg: Springer Berlin Heidelberg.

**Keyes, L. N. and Spradling, A. C.** (1997). The Drosophila gene fs(2)cup interacts with otu to define a cytoplasmic pathway required for the structure and function of germ-line chromosomes. *Development* **124**, 1419–31.

**Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., Sabo, P. J., Larschan, E., Gorchakov, A. A., Gu, T., et al.** (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480–485.

**Khodor, Y. L., Rodriguez, J., Abruzzi, K. C., Tang, C.-H. A., Marr, M. T. and Rosbash, M.** (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.* **25**, 2502–2512.

**Kim-Ha, J., Smith, J. L. and Macdonald, P. M.** (1991). oskar mRNA is localized to the posterior pole of the Drosophila oocyte. *Cell* **66**, 23–35.

**Kim-Ha, J., Webster, P. J., Smith, J. L. and Macdonald, P. M.** (1993). Multiple RNA regulatory elements mediate distinct steps in localization of oskar mRNA. *Development* **119**, 169–78.

**Kim-Ha, J., Kerr, K. and Macdonald, P. M.** (1995). Translational regulation of oskar mRNA by Bruno, an ovarian RNA-binding protein, is essential. *Cell* **81**, 403–412.

**King, R. C.** (1970). The meiotic behavior of the Drosophila oocyte. *Int. Rev. Cytol.* **28**, 125–68.

**Kiss, T.** (2006). SnoRNP Biogenesis Meets Pre-mRNA Splicing. *Mol. Cell* **23**, 775–776.

**Klovstad, M., Abdu, U. and Schüpbach, T.** (2008). Drosophila brca2 Is Required for Mitotic and Meiotic DNA Repair and Efficient Activation of the Meiotic Recombination Checkpoint. *PLoS Genet.* **4**, e31.

**Komarnitsky, P., Cho, E. J. and Buratowski, S.** (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**, 2452–60.

**Kozlova, N., Braga, J., Lundgren, J., Rino, J., Young, P., Carmo-Fonseca, M. and Visa, N.** (2006). Studies on the role of NonA in mRNA biogenesis. *Exp. Cell Res.* **312**, 2619–2630.

**Ku, H.-Y. and Lin, H.** (2014). PIWI proteins and their interactors in piRNA biogenesis, germline development and gene expression. *Natl. Sci. Rev.* **1**, 205–218.

**Kühn, U. and Wahle, E.** (2004). Structure and function of poly(A) binding proteins. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1678**, 67–84.

**Kumar, G. R. and Glaunsinger, B. A.** (2010). Nuclear import of cytoplasmic poly(A) binding protein restricts gene expression via hyperadenylation and nuclear retention of mRNA. *Mol. Cell. Biol.* **30**, 4996–5008.

**Kuraoka, I., Ito, S., Wada, T., Hayashida, M., Lee, L., Saijo, M., Nakatsu, Y., Matsumoto, M., Matsunaga, T., Handa, H., et al.** (2008). Isolation of XAB2 complex involved in pre-mRNA splicing, transcription, and transcription-coupled repair. *J. Biol. Chem.* **283**, 940–50.

**Kwak, H., Fuda, N. J., Core, L. J. and Lis, J. T.** (2013). Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science (80-. ).* **339**, 950–953.

**Labbé, R. M., Irimia, M., Currie, K. W., Lin, A., Zhu, S. J., Brown, D. D. R., Ross, E. J., Voisin, V., Bader, G. D., Blencowe, B. J., et al.** (2012). A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30**, 1734–45.

**Lahudkar, S., Shukla, A., Bajwa, P., Durairaj, G., Stanojevic, N. and Bhaumik, S. R.** (2011). The mRNA cap-binding complex stimulates the formation of pre-initiation complex at the promoter via its interaction with Mot1p in vivo. *Nucleic Acids Res.* **39**, 2188–209.

**Lan, L., Lin, S., Zhang, S. and Cohen, R. S.** (2010). Evidence for a Transport-Trap Mode of Drosophila melanogaster gurken mRNA Localization. *PLoS One* **5**, e15448.

**Lancaster, O. M., Cullen, C. F. and Ohkura, H.** (2007). NHK-1 phosphorylates BAF to allow karyosome formation in the Drosophila oocyte nucleus. *J. Cell*

*Biol.* **179**, 817–24.

**Lancaster, O. M., Breuer, M., Cullen, C. F., Ito, T. and Ohkura, H.** (2010). The Meiotic Recombination Checkpoint Suppresses NHK-1 Kinase to Prevent Reorganisation of the Oocyte Nucleus in Drosophila. *PLoS Genet.* **6**, e1001179.

**Lantz, P. M., Golberstein, E., House, J. S. and Morenoff, J.** (2010). Socioeconomic and behavioral risk factors for mortality in a national 19-year prospective study of U.S. adults. *Soc. Sci. Med.* **70**, 1558–66.

**Lasko, P. F. and Ashburner, M.** (1988). The product of the Drosophila gene vasa is very similar to eukaryotic initiation factor-4A. *Nature* **335**, 611–617.

**LASKO, P.** (1999). RNA sorting in *Drosophila* oocytes and embryos. *FASEB J.* **13**, 421–433.

**Lavoie, C. A., Lachance, P. E. D., Sonenbergţ, N. and Lasko, P.** Alternatively Spliced Transcripts from the Drosophila eIF4E Gene Produce Two Different Cap-binding Proteins*.

**Lavoie, C. A., Lachance, P. E. D., Sonenberg, N. and Lasko, P.** (1996). Alternatively Spliced Transcripts from the ***Drosophila eIF4E*** Gene Produce Two Different Cap-binding Proteins. *J. Biol. Chem.* **271**, 16393–16398.

**Lazzaretti, D., Veith, K., Kramer, K., Basquin, C., Urlaub, H., Irion, U. and Bono, F.** (2016). The bicoid mRNA localization factor Exuperantia is an RNA-binding pseudonuclease. *Nat. Struct. Mol. Biol.* **23**, 705–713.

**Le Hir, H., Gatfield, D., Izaurralde, E. and Moore, M. J.** (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20**, 4987–97.

**Lee, K.-M. and Tarn, W.-Y.** (2013). Coupling pre-mRNA processing to transcription on the RNA factory assembly line. *RNA Biol.* **10**, 380–90.

**Lee, T. I. and Young, R. A.** (2013). Transcriptional Regulation and Its Misregulation in Disease. *Cell* **152**, 1237–1251.

**Lepennetier, G. and Catania, F.** (2016). mRNA-Associated Processes and Their Influence on Exon-Intron Structure in *Drosophila melanogaster*. *G3: Genes|Genomes|Genetics* **6**, 1617–1626.

**Liang, W.-W. and Cheng, S.-C.** (2015). A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. *Genes Dev.* **29**, 81–93.

**Libri, D., Graziani, N., Saguez, C. and Boulay, J.** (2001). Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes Dev.* **15**, 36–41.

**Lin, H.** (2007). piRNAs in the Germ Line. *Science (80-. ).* **316**, 397–397.

**Lin, K. and Zhang, D.-Y.** (2005). The excess of 5' introns in eukaryotic genomes. *Nucleic Acids Res.* **33**, 6522–6527.

**Liu, Y.-C., Chen, H.-C., Wu, N.-Y. and Cheng, S.-C.** (2007). A novel splicing factor, Yju2, is associated with NTC and acts after Prp2 in promoting the first catalytic reaction of pre-mRNA splicing. *Mol. Cell. Biol.* **27**, 5403–13.

**Liu, X., Freitas, J., Zheng, D., Oliveira, M. S., Hoque, M., Martins, T., Henriques, T., Tian, B. and Moreira, A.** (2017). Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in *Drosophila melanogaster*. *RNA* **23**, 1807–1816.

**Lundquist, E. A., Herman, R. K., Rogalski, T. M., Mullen, G. P., Moerman, D. G. and Shaw, J. E.** (1996). The mec-8 gene of C. elegans encodes a protein with two RNA recognition motifs and regulates alternative splicing of unc-52

transcripts. *Development* **122**, 1601–10.

**Ma, X., Zhu, X., Han, Y., Story, B., Do, T., Song, X., Wang, S., Zhang, Y., Blanchette, M., Gogol, M., et al.** (2017). Aubergine Controls Germline Stem Cell Self-Renewal and Progeny Differentiation via Distinct Mechanisms. *Dev. Cell* **41**, 157–169.e5.

**Macdonald, P. M. and Struhl, G.** (1988). Cis- acting sequences responsible for anterior localization of bicoid mRNA in Drosophila embryos. *Nature* **336**, 595–598.

**Macdonald, P. M., Luk, S. K. and Kilpatrick, M.** (1991). Protein encoded by the exuperantia gene is concentrated at sites of bicoid mRNA accumulation in Drosophila nurse cells but not in oocytes or embryos. *Genes Dev.* **5**, 2455–2466.

**MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S. A., James, M., Groot, N., et al.** (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **72**, 971–83.

**MacDougall, N., Clark, A., MacDougall, E. and Davis, I.** (2003). Drosophila gurken (TGFα) mRNA Localizes as Particles that Move within the Oocyte in Two Dynein-Dependent Steps. *Dev. Cell* **4**, 307–319.

**Mach, J. M. and Lehmann, R.** (1997). An Egalitarian-BicaudalD complex is essential for oocyte specification and axis determination in Drosophila. *Genes Dev.* **11**, 423–35.

**Madigan, J. P., Chotkowski, H. L. and Glaser, R. L.** (2002). DNA double-strand break-induced phosphorylation of Drosophila histone variant H2Av helps prevent radiation-induced apoptosis. *Nucleic Acids Res.* **30**, 3698–705.

**Mahajan-Miklos, S. and Cooley, L.** (1994). Intercellular Cytoplasm Transport during Drosophila Oogenesis. *Dev. Biol.* **165**, 336–351.

**Mahowald, A. P. and Strassheim, J. M.** (1970). Intercellular migration of centrioles in the germarium of Drosophila melanogaster. An electron microscopic study. *J. Cell Biol.* **45**, 306–20.

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

**Malone, C. D., Mestdagh, C., Akhtar, J., Kreim, N., Deinhard, P., Sachidanandam, R., Treisman, J. and Roignant, J.-Y.** (2014b). The exon junction complex controls transposable element activity by ensuring faithful splicing of the piwi transcript. *Genes Dev.* **28**, 1786–99.

**Mandel, C. R., Bai, Y. and Tong, L.** (2008). Protein factors in pre-mRNA 3'-end processing. *Cell. Mol. Life Sci.* **65**, 1099–122.

**Mani, S. R. and Juliano, C. E.** (2013). Untangling the web: the diverse functions of the PIWI/piRNA pathway. *Mol. Reprod. Dev.* **80**, 632–64.

**Markussen, F. H., Michon, A. M., Breitwieser, W. and Ephrussi, A.** (1995). Translational control of oskar generates short OSK, the isoform that induces pole plasma assembly. *Development* **121**, 3723–32.

**Markussen, F. H., Breitwieser, W. and Ephrussi, A.** (1997). Efficient translation and phosphorylation of Oskar require Oskar protein and the RNA helicase Vasa. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 13–7.

**Martinho, R. G., Guilgur, L. G. and Prudêncio, P.** (2015). How gene expression in fast-proliferating cells keeps pace. *BioEssays* **37**, 514–524.

**Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N. and Reed, R.** (2005a). Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev.* **19**, 1512–1517.

**Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N. and Reed, R.** (2005b). Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev.* **19**, 1512–7.

**Matunis, E. L., Matunis, M. J. and Dreyfuss, G.** (1992). Characterization of the major hnRNP proteins from Drosophila melanogaster. *J. Cell Biol.* **116**, 257–69.

**McBeath, R., Pirone, D. M., Nelson, C. M., Bhadriraju, K. and Chen, C. S.** (2004). Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment. *Dev. Cell* **6**, 483–95.

**McCaffrey, R., St Johnston, D. and González-Reyes, A.** (2006). Drosophila *mus301/spindle-C* Encodes a Helicase With an Essential Role in Double-Strand DNA Break Repair and Meiotic Progression. *Genetics* **174**, 1273–1285.

**McKim, K. S. and Hayashi-Hagihara, A.** (1998). mei-W68 in Drosophila melanogaster encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved. *Genes Dev.* **12**, 2932–42.

**Meignin, C. and Davis, I.** (2008). UAP56 RNA helicase is required for axis specification and cytoplasmic mRNA localization in Drosophila. *Dev. Biol.* **315**, 89–98.

**Merrick, C. J., Jackson, D. and Diffley, J. F. X.** (2004). Visualization of Altered Replication Dynamics after DNA Damage in Human Cells. *J. Biol. Chem.* **279**, 20067–20075.

**Meyer, P., Prodromou, C., Liao, C., Hu, B., Mark Roe, S., Vaughan, C. K., Vlasic, I., Panaretou, B., Piper, P. W. and Pearl, L. H.** (2004). Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. *EMBO J.* **23**, 511–9.

**Mohler, J. and Wieschaus, E. F.** (1986). DOMINANT MATERNAL-EFFECT MUTATIONS OF DROSOPHILA MELANOGASTER CAUSING THE PRODUCTION OF DOUBLE-ABDOMEN EMBRYOS. *Genetics* **112**,.

**Montell, D. J.** (2003). Border-cell migration: the race is on. *Nat. Rev. Mol. Cell Biol.* **4**, 13–24.

**Moore, M. J.** (2005). From Birth to Death: The Complex Lives of Eukaryotic mRNAs. *Science (80-. ).* **309**, 1514–1518.

**Moore, M. J. and Sharp, P. A.** (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**, 364–368.

**Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B.** (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628.

**Mu, R., Wang, Y.-B., Wu, M., Yang, Y., Song, W., Li, T., Zhang, W.-N., Tan, B., Li, A.-L., Wang, N., et al.** (2014). Depletion of pre-mRNA splicing factor Cdc5L inhibits mitotic progression and triggers mitotic catastrophe. *Cell Death Dis.* **5**, e1151.

**Nakamura, A., Sato, K. and Hanyu-Nakamura, K.** (2004). Drosophila Cup Is an

eIF4E Binding Protein that Associates with Bruno and Regulates oskar mRNA Translation in Oogenesis. *Dev. Cell* **6**, 69–78.

**Nancollis, V., Ruckshanthi, J. P., Novak Frazer, L. and O, R. T.** The U5 snRNA Internal Loop 1 Is a Platform for Brr2, Snu114 and Prp8 Protein Binding During U5 snRNP Assembly.

**Nature reviews. Genetics.** (2000). Nature Pub. Group.

**Navarro-Costa, P., McCarthy, A., Prudêncio, P., Greer, C., Guilgur, L. G., Becker, J. D., Secombe, J., Rangan, P. and Martinho, R. G.** (2016). Early programming of the oocyte epigenome temporally controls late prophase I transcription and chromatin remodelling. *Nat. Commun.* **7**, 12331.

**Nechaev, S. and Adelman, K.** (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1809**, 34–45.

**Nelson, M. R., Leidal, A. M. and Smibert, C. A.** (2004). Drosophila Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. *EMBO J.* **23**, 150–159.

**Neuman-Silberberg, F. S. and Schupbach, T.** (1994). Dorsoventral axis formation in Drosophila depends on the correct dosage of the gene gurken. *Development* **120**, 2457–63.

**Nguyen, T. H. D., Galej, W. P., Fica, S. M., Lin, P.-C., Newman, A. J. and Nagai, K.** (2016). CryoEM structures of two spliceosomal complexes: starter and dessert at the spliceosome feast. *Curr. Opin. Struct. Biol.* **36**, 48–57.

**Nikalayevich, E. and Ohkura, H.** The NuRD nucleosome remodelling complex and NHK-1 kinase are required for chromosome condensation in oocytes. *J. Cell Sci.*

**Nolan, T., Hands, R. E. and Bustin, S. A.** (2006). Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* **1**, 1559–1582.

**Norvell, A., Kelley, R. L., Wehr, K. and Schüpbach, T.** (1999). Specific isoforms of squid, a Drosophila hnRNP, perform distinct roles in Gurken localization during oogenesis. *Genes Dev.* **13**, 864–76.

**Norvell, A., Wong, J., Randolph, K. and Thompson, L.** (2015). Wispy and Orb cooperate in the cytoplasmic polyadenylation of localized *gurken* mRNA. *Dev. Dyn.* **244**, 1276–1285.

**O'Donnell, K. A. and Boeke, J. D.** (2007). Mighty Piwis Defend the Germline against Genome Intruders. *Cell* **129**, 37–44.

**Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J. J., Yates, J. R., Tian, B. and Manley, J. L.** (2017). An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev.* **31**, 1257–1271.

**Olesnicky, E. C., Bono, J. M., Bell, L., Schachtner, L. T. and Lybecker, M. C.** (2017). The RNA-binding protein caper is required for sensory neuron development in *Drosophila melanogaster*. *Dev. Dyn.* **246**, 610–624.

**Pai, A. A., Henriques, T., McCue, K., Burkholder, A., Adelman, K. and Burge, C. B.** (2017). The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. *Elife* **6**,.

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

**Pane, A., Wehr, K. and Schüpbach, T.** (2007). zucchini and squash Encode Two Putative Nucleases Required for rasiRNA Production in the Drosophila

Germline. *Dev. Cell* **12**, 851–862.

**Papasaikas, P. and Valcárcel, J.** (2016). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem. Sci.* **41**, 33–45.

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

**Paulsen, R. D., Soni, D. V., Wollman, R., Hahn, A. T., Yee, M.-C., Guan, A., Hesley, J. A., Miller, S. C., Cromwell, E. F., Solow-Cordero, D. E., et al.** (2009). A Genome-wide siRNA Screen Reveals Diverse Cellular Processes and Pathways that Mediate Genome Stability. *Mol. Cell* **35**, 228–239.

**Perales, R. and Bentley, D.** (2009). "Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions. *Mol. Cell* **36**, 178–191.

**Peterlin, B. M. and Price, D. H.** (2006). Controlling the Elongation Phase of Transcription with P-TEFb. *Mol. Cell* **23**, 297–305.

**Pfaffl, M. W.** (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45.

**Phatnani, H. P. and Greenleaf, A. L.** (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936.

**Plaschka, C., Lin, P.-C. and Nagai, K.** (2017). Structure of a pre-catalytic spliceosome. *Nature* **546**, 617.

**Pokrywka, N. J. and Stephenson, E. C.** (1991). Microtubules mediate the localization of bicoid RNA during Drosophila oogenesis. *Development* **113**, 55–66.

**Pokrywka, N. J. and Stephenson, E. C.** (1995). Microtubules Are a General Component of mRNA Localization Systems in Drosophila Oocytes. *Dev. Biol.* **167**, 363–370.

**Price, J. V., Clifford, R. J. and Schüpbach, T.** (1989). The maternal ventralizing locus torpedo is allelic to faint little ball, an embryonic lethal, and encodes the Drosophila EGF receptor homolog. *Cell* **56**, 1085–1092.

**Proudfoot, N.** (2004). New perspectives on connecting messenger RNA 3′ end formation to transcription. *Curr. Opin. Cell Biol.* **16**, 272–278.

**Proudfoot, N. J.** (2011). Ending the message: poly(A) signals then and now. *Genes Dev.* **25**, 1770–1782.

**Raghunathan, P. L. and Guthrie, C.** (1998). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8**, 847–855.

**Ramanathan, A., Robb, G. B. and Chan, S.-H.** (2016). mRNA capping: biological functions and applications. *Nucleic Acids Res.* **44**, 7511–26.

**Ray, R. P. and Schupbach, T.** (1996). Intercellular signaling and the polarization of body axes during Drosophila oogenesis. *Genes Dev.* **10**, 1711–1723.

**Richter, J. D. and Lasko, P.** (2011). Translational Control in Oocyte Development. *Cold Spring Harb. Perspect. Biol.* **3**, a002758–a002758.

**Richter, J. D. and Sonenberg, N.** (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* **433**, 477–480.

**Riechmann, V. and Ephrussi, A.** (2001). Axis formation during Drosophila oogenesis. *Curr. Opin. Genet. Dev.* **11**, 374–83.

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

**Rodor, J., Pan, Q., Blencowe, B. J., Eyras, E. and Cáceres, J. F.** (2016). The RNA-binding profile of Acinus, a peripheral component of the exon junction complex, reveals its role in splicing regulation. *RNA* **22**, 1411–26.

**Rogozin, I. B., Carmel, L., Csuros, M. and Koonin, E. V** (2012). Origin and evolution of spliceosomal introns. *Biol. Direct* **7**, 11.

**Romanienko, P. J. and Camerini-Otero, R. D.** (2000). The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol. Cell* **6**, 975–87.

**Rongo, C., Gavis, E. R. and Lehmann, R.** (1995). Localization of oskar RNA regulates oskar translation and requires Oskar protein. *Development* **121**, 2737–46.

**Roth, S., Shira Neuman-Silberberg, F., Barcelo, G. and Schüpbach, T.** (1995). cornichon and the EGF receptor signaling process are necessary for both anterior-posterior and dorsal-ventral pattern formation in Drosophila. *Cell* **81**, 967–978.

**Ruvinsky, A. and Ward, W.** (2006). A Gradient in the Distribution of Introns in Eukaryotic Genes. *J. Mol. Evol.* **63**, 136–141.

**Saffman, E. E., Styhler, S., Rother, K., Li, W., Richard, S. and Lasko, P.** (1998). Premature translation of oskar in oocytes lacking the RNA-binding protein bicaudal-C. *Mol. Cell. Biol.* **18**, 4855–62.

**Sakurai, A., Fujimori, S., Kochiwa, H., Kitamura-Abe, S., Washio, T., Saito, R., Carninci, P., Hayashizaki, Y. and Tomita, M.** (2002). On biased distribution of introns in various eukaryotes. *Gene* **300**, 89–95.

**Saunders, C. and Cohen, R. S.** (1999a). The role of oocyte transcription, the 5'UTR, and translation repression and derepression in Drosophila gurken mRNA and protein localization. *Mol. Cell* **3**, 43–54.

**Saunders, C. and Cohen, R. S.** (1999b). The Role of Oocyte Transcription, the 5′UTR, and Translation Repression and Derepression in Drosophila gurken mRNA and Protein Localization. *Mol. Cell* **3**, 43–54.

**Schmittgen, T. D. and Livak, K. J.** (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–8.

**Schneider, S. and Schwer, B.** (2001). Functional Domains of the Yeast Splicing Factor Prp22p. *J. Biol. Chem.* **276**, 21184–21191.

**Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A. R., Keller, W., Zavolan, M. and Wahle, E.** (2014). Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev.* **28**, 2381–2393.

**Schüpbach, T. and Wieschaus, E.** (1991). Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129**,.

**Shiimori, M., Inoue, K. and Sakamoto, H.** (2013). A specific set of exon junction complex subunits is required for the nuclear retention of unspliced RNAs in Caenorhabditis elegans. *Mol. Cell. Biol.* **33**, 444–56.

**Small, E. C., Leggett, S. R., Winans, A. A. and Staley, J. P.** (2006). The EF-G-like GTPase Snu114p Regulates Spliceosome Dynamics Mediated by Brr2p, a DExD/H Box ATPase. *Mol. Cell* **23**, 389–399.

**Smith, D. J., Proudfoot, A., Friedli, L., Klig, L. S., Paravicini, G. and Payton, M. A.** (1992). PMI40, an intron-containing gene required for early steps in yeast mannosylation. *Mol. Cell. Biol.* **12**, 2924–30.

**Sollier, J., Stork, C. T., García-Rubio, M. L., Paulsen, R. D., Aguilera, A. and**

**Cimprich, K. A.** (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **56**, 777–85.

**Sonenberg, N. and Hinnebusch, A. G.** (2009). Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**, 731–745.

**Song, E. J., Werner, S. L., Neubauer, J., Stegmeier, F., Aspden, J., Rio, D., Harper, J. W., Elledge, S. J., Kirschner, M. W. and Rape, M.** (2010). The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes Dev.* **24**, 1434–1447.

**Spradling, A. C.** (1993). Position Effect Variegation and Genomic Instability. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 585–596.

**Spradling, A. C. and Mahowald, A. P.** (1979). Identification and genetic localization of mRNAs from ovarian follicle cells of Drosophila melanogaster. *Cell* **16**, 589–98.

**St Johnston, D.** (2002). The art and design of genetic screens: Drosophila melanogaster. *Nat. Rev. Genet.* **3**, 176–188.

**St Johnston, D.** (2005). Developmental Cell Biology: Moving messages: the intracellular localization of mRNAs. *Nat. Rev. Mol. Cell Biol.* **6**, 363–375.

**St Johnston, D. and Nüsslein-Volhard, C.** (1992). The origin of pattern and polarity in the Drosophila embryo. *Cell* **68**, 201–19.

**St Johnston, D., Driever, W., Berleth, T., Richstein, S. and Nüsslein-Volhard, C.** (1989). Multiple steps in the localization of bicoid RNA to the anterior pole of the Drosophila oocyte. *Development* **107 Suppl**, 13–9.

**St Johnston, D., Beuchle, D. and Nüsslein-Volhard, C.** (1991). staufen, a gene required to localize maternal RNAs in the Drosophila egg. *Cell* **66**, 51–63.

**Staley, J. P. and Guthrie, C.** (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315–26.

**Stebbins-Boaz, B., Cao, Q., de Moor, C. H., Mendez, R. and Richter, J. D.** (1999). Maskin is a CPEB-associated factor that transiently interacts with elF-4E. *Mol. Cell* **4**, 1017–27.

**Steinhauer, J. and Kalderon, D.** (2005). The RNA-binding protein Squid is required for the establishment of anteroposterior polarity in the Drosophila oocyte. *Development* **132**, 5515–25.

**Steinhauer, J. and Kalderon, D.** (2006). Microtubule polarity and axis formation in theDrosophila oocyte. *Dev. Dyn.* **235**, 1455–1468.

**Stepanyuk, G. A., Serrano, P., Peralta, E., Farr, C. L., Axelrod, H. L., Geralt, M., Das, D., Chiu, H.-J., Jaroszewski, L., Deacon, A. M., et al.** (2016). UHM–ULM interactions in the RBM39–U2AF65 splicing-factor complex. *Acta Crystallogr. Sect. D, Struct. Biol.* **72**, 497.

**Stephenson, E. C., Chao, Y. C. and Fackenthal, J. D.** (1988). Molecular analysis of the swallow gene of Drosophila melanogaster. *Genes Dev.* **2**, 1655–1665.

**Sträßer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondón, A. G., Aguilera, A., Struhl, K., Reed, R., et al.** (2002). TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**, 304–308.

**Styhler, S., Nakamura, A., Swan, A., Suter, B. and Lasko, P.** (1998). vasa is required for GURKEN accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. *Development* **125**, 1569–78.

**Suter, B. and Steward, R.** (1991a). Requirement for phosphorylation and

localization of the Bicaudal-D protein in Drosophila oocyte differentiation. *Cell* **67**, 917–26.

**Suter, B. and Steward, R.** (1991b). Requirement for phosphorylation and localization of the Bicaudal-D protein in Drosophila oocyte differentiation. *Cell* **67**, 917–926.

**Suter, B., Romberg, L. M. and Steward, R.** (1989). Bicaudal-D, a Drosophila gene involved in developmental asymmetry: localized transcript accumulation in ovaries and sequence similarity to myosin heavy chain tail domains. *Genes Dev.* **3**, 1957–1968.

**Svejstrup, J.** (2003). Keeping RNA and DNA Apart during Transcription. *Mol. Cell* **12**, 538–539.

**Takagaki, Y. and Manley, J. L.** (1997). RNA recognition by the human polyadenylation factor CstF. *Mol. Cell. Biol.* **17**, 3907–14.

**Tarn, W. Y., Hsu, C. H., Huang, K. T., Chen, H. R., Kao, H. Y., Lee, K. R. and Cheng, S. C.** (1994). Functional association of essential splicing factor(s) with PRP19 in a protein complex. *EMBO J.* **13**, 2421–31.

**Tarun, S. Z. and Sachs, A. B.** (1996). Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J.* **15**, 7168–77.

**Teixeira, F. K., Okuniewska, M., Malone, C. D., Coux, R.-X., Rio, D. C. and Lehmann, R.** (2017). piRNA-mediated regulation of transposon alternative splicing in the soma and germ line. *Nature* **552**, 268–272.

**Theurkauf, W. E.** (1994). Chapter 25 Immunofluorescence Analysis of the Cytoskeleton during Oogenesis and Early Embryogenesis. *Methods Cell Biol.* **44**, 489–505.

**Theurkauf, W. E., Smiley, S., Wong, M. L. and Alberts, B. M.** (1992). Reorganization of the cytoskeleton during Drosophila oogenesis: implications for axis specification and intercellular transport. *Development* **115**, 923–36.

**Theurkauf, W. E., Alberts, B. M., Jan, Y. N. and Jongens, T. A.** (1993). A central role for microtubules in the differentiation of Drosophila oocytes. *Development* **118**, 1169–80.

**Thio, G. L., Ray, R. P., Barcelo, G. and Schü Pbach, T.** Localization of gurken RNA in Drosophila Oogenesis Requires Elements in the 5 and 3 Regions of the Transcript.

**Thio, G. L., Ray, R. P., Barcelo, G. and Schüpbach, T.** (2000). Localization of gurken RNA in Drosophila Oogenesis Requires Elements in the 5′ and 3′ Regions of the Transcript. *Dev. Biol.* **221**, 435–446.

**Tian, B. and Graber, J. H.** (2012). Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA* **3**, 385–96.

**Tinker, R., Silver, D. and Montell, D. J.** (1998). Requirement for the Vasa RNA Helicase ingurkenmRNA Localization. *Dev. Biol.* **199**, 1–10.

**Tomancak, P., Guichet, A., Zavorszky, P. and Ephrussi, A.** (1998a). Oocyte polarity depends on regulation of gurken by Vasa. *Development* **125**, 1723–32.

**Tomancak, P., Guichet, A., Zavorszky, P., Ephrussi, A. and Lasko, P.** (1998b). Oocyte polarity depends on regulation of gurken by Vasa. *Development* **125**, 1723–32.

**Van Buskirk, C. and Schüpbach, T.** (2002a). Half pint regulates alternative splice site selection in Drosophila. *Dev. Cell* **2**, 343–53.

**Van Buskirk, C. and Schüpbach, T.** (2002b). half pint Regulates Alternative Splice Site Selection in Drosophila. *Dev. Cell* **2**, 343–353.

**Van De Bor, V., Hartswood, E., Jones, C., Finnegan, D. and Davis, I.** (2005). gurken and the I Factor Retrotransposon RNAs Share Common Localization Signals and Machinery. *Dev. Cell* **9**, 51–62.

**Van Doren, M., Williamson, A. L. and Lehmann, R.** (1998). Regulation of zygotic gene expression in Drosophila primordial germ cells. *Curr. Biol.* **8**, 243–6.

**van Eeden, F. and St Johnston, D.** (1999). The polarisation of the anterior-posterior and dorsal-ventral axes during Drosophila oogenesis. *Curr. Opin. Genet. Dev.* **9**, 396–404.

**Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F.** (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, research0034.1.

**Vert, J.-P., Foveau, N., Lajaunie, C. and Vandenbrouck, Y.** (2006). An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics* **7**, 520.

**Wahl, M. C., Will, C. L. and Lührmann, R.** (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136**, 701–718.

**Wakiyama, M., Imataka, H. and Sonenberg, N.** (2000). Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for Xenopus oocyte maturation. *Curr. Biol.* **10**, 1147–50.

**Wassarman, D. A. and Steitz, J. A.** (1991). Alive with DEAD proteins. *Nature* **349**, 463–464.

**WATSON, J. D. and CRICK, F. H. C.** (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738.

**Webster, P. J., Liang, L., Berg, C. A., Lasko, P. and Macdonald, P. M.** (1997). Translational repressor bruno plays multiple roles in development and is widely conserved. *Genes Dev.* **11**, 2510–2521.

**Webster, A., Li, S., Hur, J. K., Wachsmuth, M., Bois, J. S., Perkins, E. M., Patel, D. J. and Aravin, A. A.** (2015). Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. *Mol. Cell* **59**, 564–575.

**Weil, T. T., Parton, R. M., Herpers, B., Soetaert, J., Veenendaal, T., Xanthakis, D., Dobbie, I. M., Halstead, J. M., Hayashi, R., Rabouille, C., et al.** (2012). Drosophila patterning is established by differential association of mRNAs with P bodies. *Nat. Cell Biol.* **14**, 1305–1313.

**Wharton, R. P. and Struhl, G.** (1989). Structure of the Drosophila BicaudalD protein and its role in localizing the the posterior determinant nanos. *Cell* **59**, 881–92.

**WILKINS, M. H. F., STOKES, A. R. and WILSON, H. R.** (1953). Molecular structure of deoxypentose nucleic acids. *Nature* **171**, 738–40.

**Will, C. L. and Lührmann, R.** (1997). Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **9**, 320–8.

**Wilson, J. E., Connell, J. E., Schlenker, J. D. and Macdonald, P. M.** (1996). Novel genetic screen for genes involved in posterior body patterning inDrosophila. *Dev. Genet.* **19**, 199–209.

**Wong, Y.-H., Lee, T.-Y., Liang, H.-K., Huang, C.-M., Wang, T.-Y., Yang, Y.-H.,**

**Chu, C.-H., Huang, H.-D., Ko, M.-T. and Hwang, J.-K.** (2007). KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res.* **35**, W588–W594.

**Wong, J. J.-L., Ritchie, W., Ebner, O. A., Selbach, M., Wong, J. W. H., Huang, Y., Gao, D., Pinello, N., Gonzalez, M., Baidya, K., et al.** (2013). Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation. *Cell* **154**, 583–595.

**Xu, Q., Modrek, B. and Lee, C.** (2002). Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* **30**, 3754–3766.

**Zhang, W., Murphy, C. and Sieburth, L. E.** Conserved RNaseII domain protein functions in cytoplasmic mRNA decay and suppresses Arabidopsis decapping mutant phenotypes.

**Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J. and Shi, Y.** (2018). Structure of the human activated spliceosome in three conformational states. *Cell Res.* **28**, 307–322.