**Ricardo Matos**

# Regulation of gene expression during Drosophila oocyte development.


**Department of Biomedical Sciences and Medicine**

**2020**

**Ricardo Matos**

# Regulation of gene expression during Drosophila oocyte development.

Master in Oncobiology – Molecular Mechanisms of Cancer

Work under the supervision of:

Rui Martinho, Ph. D

Rui Silva, Ph.D


**Department of Biomedical Sciences and Medicine**

**2020**

# Regulation of gene expression during Drosophila oocyte development.

Authorship Statement

I hereby declare to be the author of this work, which is original and unpublished. Authors and papers consulted are duly cited in the text and are listed in the included references.

# Abstract

Human oocytes can stay dormant for several decades in meiotic arrest, before reactivation and maturation. Similar, *Drosophila* oocyte is mostly transcriptionally quiescent during prophase I-arrest, transiently reactivating gene expression just before it progresses into metaphase I. Our aim is to better understand such reactivation and its role for oocyte maturation and female fertility.

This work is divided in 3 chapters. In chapter I we use an isolated loss of function allele of *dkdm5* to characterize the function of this enzyme during oogenesis. The results obtained with the loss of function allele of *dkdm5* are consistent, albeit stronger, with the phenotypes previously observed after RNAi depletion and hypomorphic allele of *dkdm5*. Yet, a qualitatively distinct phenotype has also been identified, suggesting a new function of *dkdm5* during oocyte maturation.

In chapter II we try to access chromatin quality and synaptonemal complex assembly throughout oocyte development. Proteins from chromatin remodelling complexes, as *dkdm5*, have been shown to disrupt the synaptonemal complex further leading to problems in meiotic progression. Here we observed that two core components of the *pho* repressive complex are required for female fertility. Further, we show that specific germline *dSfmbt* depletion leads to an abnormal increase of Corolla and defective synaptonemal complex morphology in the oocyte chromatin throughout development.

Chapter III is an independent chapter where we reveal Nine Teen Complex Protein Salsa as being particularly rate limiting for efficient splicing of short proximal introns and dorsoventral patterning of the *Drosophila* egg. We observed that, upon specific germline depletion of *Salsa*, *Gurken* transcript is poorly spliced leading to an abnormal localization, subsequentially leading to defects in the eggshell dorsoventral patterning and female fertility. Further we show that ectopic *Gurken* expression can suppress the dorsal ventral patterning defects after *Salsa* depletion. Our work aims to mechanistically understand our observations.



**Keywords:** Drosophila, oogenesis, dkdm5, meiosis, synaptonemal complex.

# Resumo

Os oócitos humanos têm a habilidade de para o seu ciclo celular na meiose durante diversas décadas, num estado de dormência transcricional, antes de reativarem a sua transcrição e maturarem. Semelhante ao que ocorre em humanos, os oócitos de *Drosófila* também estão, na grande parte do seu desenvolvimento, com o seu ciclo celular parado na prófase I da meiose num estado de quiescência transcricional. Antes da progressão meiótica para a metáfase I, existe uma reativação transiente da transcrição no nucleou do oócito.

Este trabalho está dividido em três capítulos. No primeiro capítulo, focamo-nos em utilizar um alelo isolado com a perda de função total de *dkdm5* para tentar caracterizar a função desta enzina durante a oogénese. Os resultados obtidos são semelhantes aos observados anteriormente em experiências feitas através de depleção por RNAi ou pelo uso de um alelo hipomórfico. Contudo os fenótipos observados na perda da função total de *dkdm5* são mais acentuados. Também, com este trabalho conseguimos desvendar alguns dos processos que são dependentes ou independentes da atividade de demetilase de *dkdm5.* Contudo, observamos um fenótipo qualitativamente distinto na cromatina do núcleo do oócito o que sugere uma nova função de *dkdm5* durante a maturação do oócito.

No segundo capítulo, avaliamos a qualidade da cromatina e a formação synaptonemal complex ao longo do desenvolvimento do oócito. Proteínas que integram complexos responsáveis pela remodelação da cromatina, como *dkdm5,* tem vindo a ser descritos como necessário para a correta formação do synaptonemal complex e subsequentemente uma correta progressão meiótica. Neste capítulo, mostramos que dois dos principais componentes de *pho* repressive complex, são necessários para a fertilidade feminina. Alem disso, mostramos também que após uma deleção especifica de *dSfmbt* na linha germinal, existe um aumento da expressão de Corolla e o aparecimento de morfologias defeituosas do SC durante o desenvolvimento do oócito.

O terceiro capítulo é um capítulo independente. Neste capítulo nós descobrimos que a proteína Salsa do Nine Teen Complex protein como sendo particularmente limitante

para a eficiência de splicing em intrões proximais curtos e padronização dorsoventral do ovo de *Drosófila.* Observamos que, após depleção especifica da linha germinal de *Salsa* que, o splincing do transcrito de Gurken não é feito corretamente levando a uma localização anormal desta proteína, subsequentemente levando a defeitos de padronização dorsoventral dos ovos e defeitos na fertilidade feminina. Contudo, mostramos que a expressão ectópica de *Gurken* pode suprimir os defeitos de padronização dorsoventral após a depleção de *Salsa.* O nosso objetivo reside em entender os mecanismos que estão por detrás das nossas observações.

**Palavras chave:** Drosófila, oogénese, dkdm5, meiose, synaptonemal complex.

## Table of contents

# List of Figures

# List of Tables

# List of Annexes

# Abbreviations

| | |
|---|---|
| **AQR** | Aquarius |
| **BAF** | Barrier to autointegration factor |
| **Brm** | Brahma |
| **$CO_2$** | Carbon dioxide |
| **CPED** | Cytoplasmic polyadenylation element binding protein |
| **DSB** | Double strand breaks |
| **Dv** | Dorsoventral/dorsal ventral |
| **ECL** | Enhanced chemiluminescence detection |
| **EJC** | Exon junction complex |
| **FGC** | Follicle somatic stem cells |
| **Grk** | Gurken |
| **GSC** | Germline stem cells |
| **IR** | Intro retention |
| **Jmjc** | Jumonji C |
| **LCMS** | Liquid chromatography–mass spectrometry |
| **LH** | Luteinizing hormone |
| **Lid** | Little imaginal discs |
| **MBT** | Mid blastula transition |
| **MeOH** | Methanol |
| **nhk-1** | Nucleosomal histone kinase 1 |
| **NTC** | Nine Teen Complex |
| **OR** | Oregon R |
| **Pcl** | Polycomblike |
| **Pho** | Pleiohomeotic |
| **Prp19** | Pre-mRNA processing factor 19 |
| **pSer5** | Phosphorylation of Serine 5 |
| **RNA pol II** | RNA polymerase II |
| **SC** | Synaptonemal complex |
| **Sfmbt** | Scm-related gene containing four mbt domains |

| | |
|---|---|
| **shRNAi** | Short hairpin RNAs |
| **spn-B** | Spindle B |
| **spn-D** | Spindle D |
| **TSS** | Transcription-start site |
| **WT** | Wild type |

# Chapter I

# 1 Introduction.

## 1.1 Oogenesis overview.

### *1.1.1 Homo sapiens oogenesis.*

Human oocyte development is mainly supported on the ability from the mother to provide nutrients required for oocyte maturation and embryonic development with the main role of reproductive success [1]. Oocyte development starts at around 6 weeks of gestation in female fetus. This event starts with the migration of the primordial germ cells to the genital ridges, generating millions of oogonia cells that will peak at over six million by the 26 gestation weeks [2–5].

As oogonia proliferate by asymmetric mitotic division, it gives rise to two daughter cells, specifically the future primary follicle and another oogonium. The future primary follicle ends proliferation and enters in prophase I of meiosis. During the fetal stages of female gametogenesis, oocytes advance through various stages of prophase I: leptotene, zygotene, pachytene, and transiently arresting their development at the diplotene stage of prophase I, acquiring a quiescent state [2,6]. Although the oogonia keeps proliferating throughout fetal stages of female gametogenesis, most of them die by apoptosis and only the survivors will continue differentiation [2].

In humans, the diplotene I arrest of the quiescent primary oocyte can last several years, or even decades, from fetal development until the onset of ovulatory cycles at puberty, being these dormant oocytes progressively reactivated during the entire women´s fertile period (Figure 1) [7].

Has oocyte differentiation proceeds, the layer of somatic cells that surrounds the oocyte and supports its development, further matures in a process called folliculogenesis [8]. This process begins in the fourth month of gestation and remains active until birth, when the containing primordial follicles arrested at the diplotene stage of prophase I are fully formed [8,9]. This layer of flatten granulosa cells, surrounding the arrested primordial follicles are also arrested at the G0 phase of the cell cycle, that will allow the production various growth factors

and cytokines that are crucial the viability, maturation and further oocyte development. The mechanisms underlying oocyte dormancy and follicular quiescence are still poorly understood [10,11].

At birth, the number of oocytes formed at the initial stages of development has already declined by 80% through a process of follicle degeneration, commonly known as atresia [12]. After puberty and until the onset of menopause, groups of primordial follicles are continuously reactivated for further oocyte maturation [13]. However, since the process of atresia continues after birth and throughout the reproductive life span of the individual, at puberty there are only around 400,000 primary oocytes from which only approximately 500 oocytes will fully mature and ovulate [5].

Follicle growth initiation and maturation result from crosstalk between the developing oocyte and the supporting somatic cells [14]. From the initial population of primordial follicles, there is an irreversible growth initiation of a pool of primordial follicles followed by the proliferation of cuboidal granulosa cells forming a structure that is the primary follicle (Figure 1). These morphological changes are associated with a significant change of the primary follicle transcriptome [15].

As granulosa cells continuously proliferate, theca cells are recruited to the growing secondary follicles. With the continuous proliferation of the granulosa cells, the follicle will reach the preantral stage were the capillary vessels form within the thecal layers, blood begins to circulate create the follicle and the oocyte becomes surrounded by a zona pellucida. Antral follicles (also known as tertiary follicles) are produced upon fluid accumulation in the antrum, dramatically expanding the oocyte volume (Figure 1). During the antral follicle development, the oocyte becomes competent and is finally released from the prophase I-arrest, progressing into meiosis II and transiently arresting once more in metaphase II until fertilization [16]. Meiotic resumption results from an endocrine cascade that induces a peak of luteinizing hormone (LH) secretion leading to ovulation.

After ovulation, the corpus luteum is formed from the terminal differentiation of remaining theca and granulosa cells which maintains the endometrium by secreting progesterone and estrogen [11,15,17].

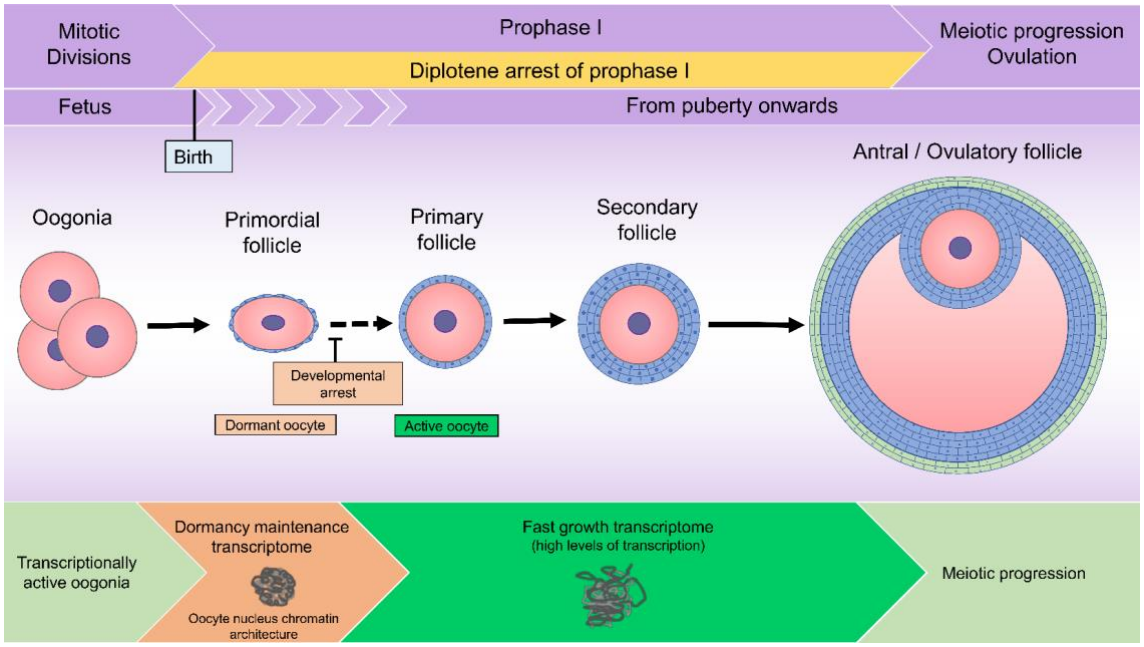



***Figure 1 - Homo sapiens oogenesis.*** *In Homo sapiens (human), the developing primary oocyte arrests for several decades in a dormant state, before reactivation after puberty. Reactivation of the dormant primordial follicle is associated with a dramatic change in the oocyte transcriptome and rapid growth*

### *1.1.2 Drosophila melanogaster oogenesis.*

*Drosophila melanogaster* has two ovaries, both are composed by 16 to 20 independent ovarioles, each with their own germ-line stem cell population [18,19]. Each ovariole is composed by a progression of distinct follicles, starting with the germarium at the anterior tip of the ovariole and progressing to a fully developed oocyte at the posterior end, all connected by somatic stalk cells [19]. Each *Drosophila* ovariole contains two main stem cell populations, the germ-line stem cells (GSC) and follicle somatic stem cells (FSC), which are harboured in an anterior structure named the germarium. Oogenesis takes approximately one week to be completed [20–22].

*Drosophila* female meiosis takes place at the anterior tip of the germarium where 2-3 GSCs divide asymmetrically to produce another stem cell and a daughter cell that will soon differentiate into a cystoblast. The newly formed cystoblast undergoes four mitotic divisions with incomplete cytokinesis forming a 16-cell cyst interconnected by the ring canals and a cytoskeleton structure known as fusome [20–22]. The synaptonemal complex (SC) assembly between the paired homologous chromosomes is the first cytological evidence of meiotic initiation occurring in region 2A of the germarium [18].

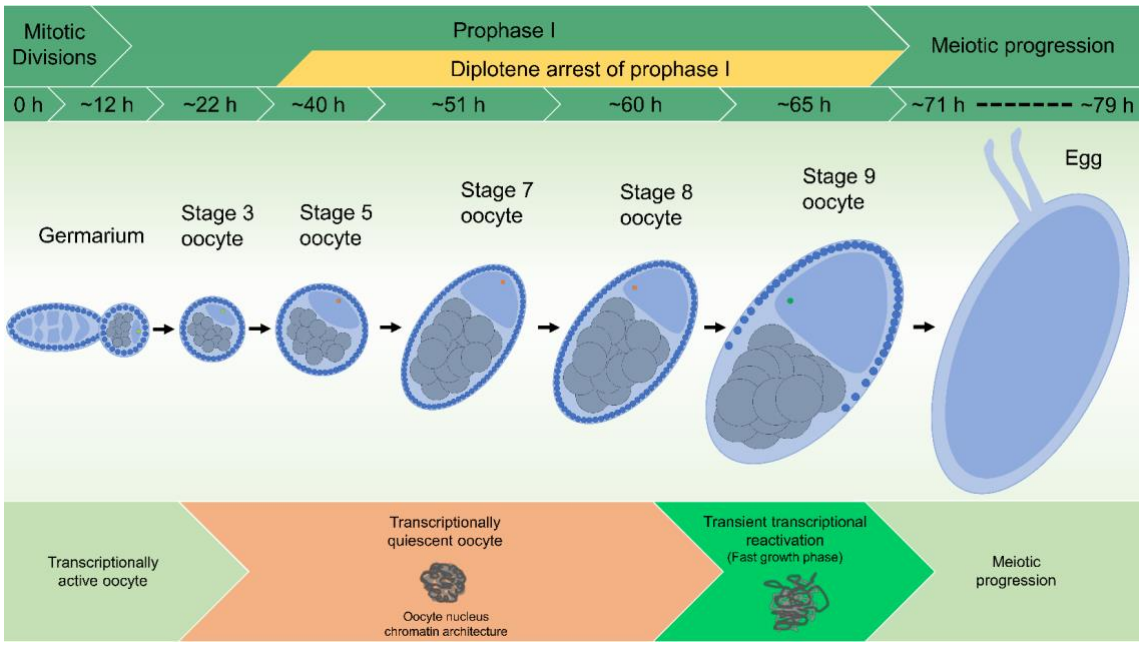



***Figure 2 - Drosophila melanogaster oogenesis.*** *In Drosophila melanogaster (fruit fly), the developing oocyte arrests at the diplotene stage of prophase I, with its nucleus forming a highly condensed structure, known as the karyosome, and becoming transcriptionally quiescent. Such quiescence does not impair oocyte growth as it mostly relies in the supporting nurse cells. The transcriptionally quiescent oocyte transiently reactivates gene expression before it progresses into metaphase suggesting a role in meiotic progression.*

The pairing and synapsis of homologous chromosomes by the SC is essential for subsequent meiotic processes including recombination and chromosome segregation [23]. After initiation of SC assembly between the paired homologous chromosomes attaching them together until anaphase, double-

stranded breaks (DSBs) are generated in order to create crossing over loci [24]. In region 3 only one cystoblast will remain in meiosis, whereas the other 15 cells of the 16-cell cyst will endoreplicate their own genome giving rise to the supporting polyploid nurse cells. This results in a single pro-oocyte by region 3 that maintains full-length SC along the arms of the chromosomes with most DSBs are already repaired [25], while the supporting nurse cells are required to transport and synthetize nutrients, cytoplasmic components, mRNAis and proteins that are important for correct oocyte growth and development. The oocyte is the only cell within the egg chamber that will progress through meiosis [21,26,27].

When the cystoblast leaves the germarium, an egg chamber is formed. The egg chamber is composed by the oocyte, the nurse cells and the somatic follicle cells that surround both the oocyte and the nurse cells. *Drosophila* oocyte Prophase I starts early in the region 2A of the germarium and arrests during the diplotene stage of prophase I at stage 5 during *Drosophila* oocyte development, with the oocyte chromatin compacted and condensed into a structure that is the karyosome [26–28]. In terms of meiotic progression, only a few events have been documented during stages 2–12 of oocyte development, namely karyosome formation and SC disassembly.

The arrangement of the karyosome relies, among other factors, on the ability of the oocyte nucleus chromatin to detached completely from the nuclear envelope allowing it's the full condensation and compaction. The phosphorylation of Barrier to autointegration factor (Baf) by Nucleosomal histone kinase-1 (*nhk-1*) is one of the elements that is required for the correct detachment of the oocyte chromatin from its nuclear envelope [29]. Loss of *nhk-1* is associated with a delay in the disassembly of euchromatic SC [30], creating aberrant chromosomal arrangements in the polar body, mitotic spindle defects, and metaphase I spindle defects [31]. Although, why the oocyte nucleus chromatin needs to detach from the nuclear envelope and form a highly condensed structure is still not well understood, *Drosophila* mutants that disrupt the karyosome architecture lead to female sterility, suggesting a correlation between the karyosome formation and female fertility [31], [32].

The oocyte nucleus arrest between stage 5 and 9 of *Drosophila* oogenesis, at the diplotene stage of prophase I, in a transcription quiescent stage with its chromatin condensed. Before the meiotic progression, there is however, a transient reactivation of gene expression accompanied by a remodelling of the oocyte chromatin architecture during late diplotene I-arrest (stage 9-10). During this period, the oocyte rapidly grows in size [32], [33].

The functional relevance of such transcriptional reactivation for meiotic progression and female fertility is still poorly understood, however Histone demethylase *dkdm5*/Lid not only avoids precocious reactivation of oocyte transcription [32], but is also important for the correct levels of Histone H3K4 trimethylation, karyosome architecture, SC maintenance, centromere pairing, meiotic progression and female fertility [32,34].

For *Drosophila* oocyte growth, the oocyte mostly relies in the ability of the supporting nurse cells, to synthetize and transport RNAs, proteins and other cytoplasmic components into the developing oocyte [27]. Such transport can be divided in two phases [35–37], a slow-initial phase that occurs between stage 2 until stage 10A and a rapid phase that occurs between stage 10B and 12 (nurse cell dumping), where an actin-myosin contraction of the nurse cells push their cytoplasm contents rapidly into the oocyte, through the connecting ring canals.

Once oocyte growth stops and the migrating follicle cells cap the oocyte, closing the connecting ring canals, meiosis progresses into metaphase I, where it is arrested again until egg activation [38]. *Drosophila* egg activation occurs independently of fertilization, being induced by the physically stimulated passing of the egg in the oviduct and rehydration [39]. During egg activation, there is considerable change of the oocyte transcriptome and proteome, which prepares the female gamete for fertilization and the oocyte-to-zygote transition [40,41].

## 1.2 The First meiotic arrest.

Prophase I during meiosis is divided in 5 stages. The first stage is designated has leptotene and during this stage of meiotic prophase I the chromosomes are unpaired and small segments of chromosome core structure appear in the nucleus. When the homologous chromosomes begin to pair and

synapse, the zygotene stage of prophase I take place. This stage is characterized in *Drosophila* by the observation of short patches of SC in early region 2A of the germarium. Pachytene is the third stage where the full-length SC is established allowing chromosomal crossover between paired chromosomes and recombination. The final stage of meiotic prophase I is the diplotene stage where meiotic recombination is already accomplished and in *Drosophila*, the SC disassembles from the chromosome arms, being still attached to the centromeric regions. During the diplotene stage, the oocyte enters a prolonged resting period [23], [42]. This meiotic arrest at the diplotene stage of prophase I is highly conserved across metazoans and it's important for oocyte differentiation and accumulation of maternal components [32].

Although the oocyte is arrested at the diplotene I stage, often the oocyte chromatin is associated with dramatic architecture rearrangements. Different species have different strategies to assure the correct maturation of oocyte at this stage. Mammal oocytes are smaller in size when compared to other vertebrates because they do not need to accumulate large stockpiles of storage proteins to sustain embryonic development after fertilization, has seen in *Xenopus laevis* and *Danio rerio* oocytes. Before birth, the primary oocytes enter meiotic diplotene I arrest, being surrounded by supporting granulosa cells and forming long-lived quiescent ovarian structures named primordial follicles [10]. After puberty initiation, there's a continuous activation of the primordial follicles, where the granulosa cells change from flat to a cuboid structure followed by a global alteration of the oocyte transcriptome. Subsequently, there is the resumption meiotic progression, oocyte growth and maturation [10].

*Drosophila* oogenesis has a prolonged meiotic arrest at the diplotene stage of prophase I. The oocyte arrests at diplotene stage of prophase I, being this arrest accompanied by condensation of the oocyte chromatin into the karyosome inducing a transcriptional quiescence of the oocyte between stage 5 and 9. Although the oocyte nucleus is in a quiescence state, this does not impair oocyte maturation as the nurse cells supports the oocyte with its requirements for growth and development [43,44]. Despite the supporting role of the nurse cells, prophase I arrested oocytes transiently reactivate gene expression just before

progress into metaphase I. This transient reactivation of transcription is associated with a global change in oocyte chromatin architecture. [32,33].

## 1.3 Oocyte Epigenetic Regulation.

Epigenetic modification play an important role guaranteeing a temporally and spatial regulation of gene expression during *Drosophila* oogenesis [45]. Although the prophase I-arrested oocyte is in a highly condensed and compact structure, its epigenome is quite dynamic, presenting an arrangement of euchromatic and heterochromatic markers that vary during oocyte quiescence and reactivation [32].

Although the biological significance of many of these epigenetic modifications is still poorly understood, the regulation of histone H3 lysine 4 trimethylation (H3K4me3) that is an well-known euchromatic marker and histone H3 lysine 27 trimethylation (H3K27me3) that is an well-known heterochromatic marker, have been reported to be rate-limiting for *Drosophila* oocyte chromatin architecture, meiotic completion and female fertility. During the transcriptional quiescence of *Drosophila* oocyte, H3K4me3 remain at low levels when compared to H3K27me3. Subsequentially, when the oocyte transiently reactivates its transcription before meiotic progression it has been reported a significant decrease in the H3K27me3 levels without any detectable change in the levels of H3K4me3, suggesting that the balance between these two epigenetic markers controls oocyte dormancy and reactivation [32].

Similar to *Drosophila*, Mammalian oocytes also present an assortment of euchromatic and heterochromatic markers. The correct regulation of H3K4me3 and H3K27m3 levels is important for meiotic progression and transcriptional reactivation, has these markers are dynamically regulated during oocyte development. The impaired regulation of these epigenetically markers can lead to an abnormal meiotic progression [10].

## 1.4 Histone Demethylases.

Mammalian cells encode four KDM5 paralogs, KDM5A, KDM5B, KDM5C, and KDM5D whereas organisms with smaller genomes, such as *Drosophila* that

have a single KDM5 ortholog. *Drosophila kdm5* histone demethylase (*dkdm5*; also known as Lid) is a strict H3k4m3 demethylase that localizes to the transcription-start site (TSS) of actively transcribed developmental genes and negatively regulates the levels of H3K4me3 by removing methyl marks from lysine 4 [46]. Not only *dkdm5* can repress gene expression through its demethylase activity, it can also act as a gene expression activator in a demethylase independent manner, by interacting with other proteins positively regulating gene expression [47]. Such dual function of *dkdm5* potentially explains why the expression of different subsets of genes are positively or negatively regulated.

Female germ-line specific depletion of *dkdm5*, leads to a significant increase in the levels of H3K4me3 and precocious transcriptional reactivation of the quiescent prophase I-arrested oocytes without disturbing other epigenetic markers. Although, upon female germ-line specific depletion of *dkdm5* the egg chamber growth and development are not impaired, the architecture of the oocyte chromatin is highly abnormal with a defective SC assembly leading to an abnormal meiotic progression and a significant reduction of female fertility [32], [34].

## 1.5 Aim of this work.

This work is divided in 3 chapters. In Chapter I we aim to define the roles of the histone demethylase *dkdm5* during oocyte reactivation, since previous experiments have been done in hypomorphic conditions. With the new isolated *dkdm5* loss of function we are able to completely abolish *dkdm5* [48] throughout oogenesis and not only we can distinguish the processes that are dependent or independent of *dkdm5* demethylase ability, but also hypothesize that *dkdm5* is required for other processes during oogenesis that yet have not been described.

In Chapter II we try to access chromatin quality and SC assembly throughout the prophase I arrested oocyte. The SC assembly is required for the correct meiotic progression. Chromatin remodelers or proteins that are associated to chromatin remodelling complexes such as *dkdm5* have been shown to disrupt the SC assembly further leading to problems in meiotic progression leading to female infertility [32,34]. Therefore, we hypothesize that there are other chromatin remodelers and polycomb subunits required for the correct SC assembly. Further in this chapter we characterize two proteins that form the pho repressive complex, *Pho* and *dSfmbt,* to be required for female fertility. Collectively, we hypothesise that, if *Pho* and *dSfbmt* are important for the correct architecture and remodelling of the oocyte chromatin during prophase I, then *Pho* and *dSfbmt* should regulate SC assembly and/or disassembly during meiosis.

The third chapter is an independent chapter based on the work of Om Rathore where I was integrated. In this chapter we undercover Nine Teen Complex (NTC) protein Salsa as being particularly rate limiting for efficient splicing of short proximal introns and dorsoventral patterning of the Drosophila egg. 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 dorsoventral patterning through its regulation of *Gurken* expression. Interestingly, we observed that splicing of the first intron of G*urken* transcript was significantly more affected than the second and third introns. Consistently, genome-wide analysis confirmed that splicing of short proximal

introns is particularly sensitive to depletion of *Salsa*. Our ongoing work aims to mechanistically understand our observations.

# 2 Methods.

## 2.1 Drosophila husbandry.

### *2.1.1 Stocks and Crosses.*

All *Drosophila* stocks were kept under standard procedures. Approximately, 15 virgin female flies were mated with 8 male flies for all crosses. After every forty-eight hours flies were transferred to fresh vials. Virgin female flies were collected during the day from vials kept at 25ºC. To maximize the number of virgins collected, vials were placed at 18ºC overnight and virgens collected in the morning. Female virginity was confirmed by absence of larvae after forty-eight hours at 25ºC. After this period, the vials were checked for larvae growth. If larvae were found, this indicates that at least one of the females was not virgin and the vial was discarded. Both females and males selected flies were between 3 to 6 days old to maximize the cross progeny. The cross date was also taken in count, in the view that after 18 days the second-generation progeny will start to hatch, whose genotype would be a complete mystery since it's almost impossible to deduce who the parents are.

### *2.1.2 Anesthetizing flies.*

*D. melanogaster* flies were anesthetized with carbon dioxide ($CO_2$). First, a hose is inserted to the vial or bottle and $CO_2$ is administered to the flies. The anesthetized flies are gently poured in a porous flat surface that delivers low levels of $CO_2$ gas making it possible to sort out the flies without them flying away. The administration of $CO_2$ to *D. melanogaster* flies will decrease the sensitivity of glutamate at the neural muscular junction, therefore, impairing motor ability [49]. There's no method of anesthetizing flies without risks. While the use of $CO_2$ isn't toxic, it may induce headaches or dizziness to the user while prolonged exposure and some studies [50] suggest that longevity and fecundity of the flies is affected as well as reproductive behaviour [51]. After the use of $CO_2$, flies were left to recover twenty-four hours before any experiment was conducted.

## 2.2 Fly work and genetics.

### *2.2.1 Additional Drosophila stocks.*

Genomic *dkdm5* transgenes were generated as previously described in [32]. Briefly, the demethylase-dead construct of *dkdm5* containing two-point mutations were generated with the following primers, JmjC_F: 50-CGCAGCCTTCTGCTGGGCCAACGCGGACCACTGGAGTA-30; JmjC_R: 50-TACTCCAGTGGTCCGCGTTGGCCCAGCAGAAGGCTGCG-30. For controls, an 11.4 kb wild type *dkdm5* genomic rescue transgene was generated. To generate this transgene, the following primers were used: pBGR_F1: 50-CTAAAGGGAACAAAAGCTGGCGCAGTGCGACGGCTCCAAATAC-30; pBGR_R1: 50-CGTGTTGTCCGCCTCAGTTTTGGCGGACATAGCTTTAAGA-30; pBGR_F2: 50-CATCTTAAAGCTATGTCCGCCAAAACTGAGGCGGACAACA-30; pGR_R2: 50-CCCCGGGCTGCAGGAATTAACTCGGACTTGCACAAGCAGAAC-30; pBGR_F3: 50-CTAAAGGGAACAAAAGCTGGCGGCGGAGATTGTG GCCAGCTT-30 and pBGR_R3: 50-CCCCGGGCTGCAGGAATTGCTTACGCCGATTGCATGTC-30.

Both transgenes contain a C-terminal 3xHA-tag and were subsequently analysed in a *dkdm5 null* mutant background showing equal expression (Figure 3B). The *dkdm5 null* allele we generated by imprecise excision of a P element inserted in the *kdm5* promoter region resulted in the deletion of three coding exons of the *kdm5* gene [52]. No full length or truncated KDM5 protein(s) were present in *kdm5 null* flies and homozygous *kdm5* adult mutants are 100% lethal [52].

### *2.2.2 Germline clones.*

Since homozygous and transheterozygous *kdm5* mutants is lethal, we performed germline clones using the FLP/FRTovoD system[53] in order to obtain heterozygous adult flies carrying homozygous mutant oocytes. For this purpose we crossed $w; \frac{KDM5^{140}FRT40A}{Cyo}$, $w; \frac{KDM5^{140}FRT40A}{Cyo}; \frac{gKDM5{:}HA^{WT}}{Tm6B}$ or $w; \frac{KDM5^{140}FRT40A}{Cyo}; \frac{gKDM5{:}HA^{Jmjc}}{Tm6B}$ virgin females with $w, hsFLP22; \frac{FRT\ ovo^{D}}{Cyo\ hsHid}$ males, and

heat shocking at 37ºC the progeny during second and third larval instar for one hour. For control $w; \frac{P(NeoFRT)40A,P(FRTw[hs])G13}{Cyo}$ females were crossed with $whsFLP22; \frac{FRT\ ovo^{D}}{Cyo\ hsHid}$ males followed by the heat shock.

### *2.2.3 Egg hatching and ventralized eggshell defects scoring.*

Cages containing 50 virgin females with control or *kdm5* germline clones and 20 wild type (Oregon R; OR) males, with ages between two and five days, were prepared. For wild-type control, 20 wild type females and 10 wild type males with ages between two and five days were also prepared. We reduced the number of flies in wild-type control in order to have equivalent number of eggs between wild-type control and germline clones. All cages were maintained at 25ºC. Eggs were collected in plaques containing agar, sugar and apple juice. Overnight egg collections were obtained since the germline clone females egg laying is extremely low. Upon collecting the agar plaques, the total number of eggs laid were scored and checked for dorsal-ventral patterning defects (with the criteria shown in figure 6). Subsequently, the plaques were incubated at 25ºC for forty-eight hours. After this period, the number of hatched eggs was counted.

## 2.3 Drosophila ovaries dissection, fixing and storage.

### *2.3.1 Dissection and fixing.*

Adult ovaries from females keep for 2 days in food vials supplemented with fresh yeast paste were dissected according to the standard procedures. To this purpose, flies were anesthetized by cold (5 minutes in -20ºc) and ovaries were dissected in 1x PBS pH=7,4. After dissection, ovaries were fixed for twenty minutes in fixing solution (PBS with 4% of formaldehyde (Thermo Fisher Scientific, cat. #28906), and 0.5% NP-40 (Sigma, cat. #I8896)) and mixed with 2 volumes of heptane (Fluka, cat. #34495) at 1:3 ratio. Subsequently, ovaries were washed in 0.2% PBST (PBS with 0.2% Tween-20 (Sigma, cat. #P5927)).

### *2.3.2 Storage.*

Fixed *Drosophila* ovaries were de-hydrated in Methanol (MeOH; Sigma, cat. 322415) by consecutive 10 minutes incubation in a solution made of 0.2% PBST and increased concentrations of methanol (30%MeOH, 50%MeOH, 70%MeOH, 100%MeOH). Afterwards, the ovary samples were stored at -20ºC.

## 2.4 Ovary immunofluorescence.

Adult ovaries (10 ovary pairs per sample per experiment) were dissected according to the standard procedures as described previously. For immunofluorescence, fresh ovaries were used without MeOH storage since, some antibody epitopes are particularly sensitive and MeOH may reduce the antibody - epitope affinity. After fixation, ovaries were gently detached into single ovarioles and incubated for 2h in a blocking and permeabilization solution (PBS with 0.2% Tween 20 (Sigma)), 1% Triton X-100 (Sigma, cat. #T8787), 1% (w/v) bovine serum albumin (BSA; Sigma, cat. #A7906) and 1% (w/v) donkey serum (Sigma, cat. #D9663)). Primary antibodies incubation was performed overnight at 4 ºC, with gentle rotation, in 0.2% PBST supplemented with 1% BSA and 1% donkey serum. Primary antibody was washed with 0.2% PBST and the secondary antibody incubation was performed at room temperature during 1h. Afterwards, secondary antibody was washed in 0.2% PBST and DNA was stained for 30 minutes at room temperature with DAPI (1: 10,000) (Fluka, cat. # 32670-5MG-F) diluted in 0.2% PBST. Subsequently, DAPI staining was washed with 0.2% PBST. Prior to mounting, ovaries were rinse in 1x PBS and mounted with mounting medium (DAKO; Lusopalex, cat. #411144).

## 2.5 Protein immunoblotting.

### *2.5.1 Protein extraction from ovaries.*

Ovary protein extraction was performed as described by P. Prudêncio and L. G. Guilgur (2015). Briefly, *Drosophila* ovaries were dissected as described above, and transferred to an ice-cold lysis buffer containing 150mM NaCl, 50mM Tris-HCL (pH 7.5), 2mM EDTA, 0.1% NP-40 2mM DTT, 10mM NaF and protease inhibitor (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail; Roche). Then,

samples were manually homogenized ensuring the complete breakdown of the tissue and centrifuged for twenty seconds at ~10,000 rcf (4ºC). Manual homogenization was repeated and then the samples were centrifuged for three minutes at ~20.000 rcf (4ºC). This step was repeated two times and between centrifugations the supernatant was transferred to new tubes, avoiding the lipid layer. The total protein extract was quantified in a spectrophotometer according to the Bradford protein assay. After quantification, all samples were diluted to a final concentration of 1 µg/µL and 2x Laemmli sample buffer was added (half of the final volume). After a quick spin down, samples were heated for five minutes at 100ºC and immediately frozen at -20ºC.

### *2.5.2 SDS-Page.*

Protein samples were run at 6% SDS-polyacrylamide electrophoresis gel. Samples were loaded into the gel and remaining empty wells was loaded with the same volume of sample buffer. The voltage for sample running started at 60v and after the stacking the voltage was increased to 90v. Afterwards, the protein was transferred to a nitrocellulose membrane at 100v for sixty minutes. After transfer, the membrane was briefly washed and blocked in a PBST 0.1% and 5% milk solution for one hour with orbital shaking. Following the blocking, the first antibody incubation was performed overnight at 4ºC in 50mL falcons with a PBST 0.1% and 5%milk, diluted according to the desired antibody. After, membranes were washed three times for fifteen minutes in PBST 0.1% and blocked for fifteen minutes in a PBST 0.1% and 5% milk solution. Following, the second antibody incubation was performed for two hours at room temperature on a roller in 50mL falcons with a PBST 0.1% and 5%milk, diluted according to the desired antibody. After the second antibody incubation, the membranes were washed three times for 10 minutes.

### *2.5.3 Enhanced chemiluminescence Detection (ECL).*

10mL of a homemade ECL solution (1M Tris-HCL pH 8.5, 250mM luminol, 90mM DMSO and water) supplemented with 6uL of 30% hydrogen peroxide was equally distributed this solution across the membranes. After one minute, the membranes were transferred to a cassette. In a dark room, the films were

exposed to the membrane with different time periods in order to have more or less exposure and were submerged in a developing solution following a fixing solution. Afterwards, the films were washed with water and dried.

## 2.6 Antibodies.

For immunofluorescence, the following antibodies were used: anti-phospho-Ser 5 CTD RNApol II rat (1:500 dilution, Merck Millipore 04-1572-I). For western blot, rat anti-HA (1:500 dilution, Roche 11867423001) and mouse anti-α-Tubulin (1:10,000 dilution, Sigma T6199). Secondary antibody for immunofluorescence detection was performed with Alexa 488 (Molecular Probes) at 1: 1,000 dilution and for western blot HRP-conjugated antibody at 1:4,000 (Jackson ImmunoResearch).

## 2.7 Protein immunoprecipitation.

For immunoprecipitation of the HA-tag in ovaries, ovary dissection and protein extraction protocol were performed as described previously. Per condition, 3mg of Pierce™ Anti-HA Magnetic Beads (Thermofisher, cat. #88836) were washed in protein lysis buffer. Subsequently, 1mg of protein extract diluted in protein lysis buffer was added to the beads following a one-hour incubation (4ºC). After incubation, the beads were washed, and all samples were diluted in 2x Laemmli sample buffer (half of the final volume). Input and supernatant samples were taken for western control and immunoprecipitation efficiency.

## 2.8 Quantitative imaging analysis.

Oocyte chromatin was assessed by using an automated global thresholding method (Huang's fuzzy thresholding; ImageJ).

Oogenesis stages were defined accordingly standard morphological components has described by Jia, Dongyu *et al.* 2016 [54]. For signal quantification, in each oocyte

## 2.9 Statistical analysis

Statistical analysis was performed using Prism six software. To analyse egg hatching and dorsal ventral patterning a statistical analysis of variance was used to evaluate the differences among group means in each sample. Signal quantification analysis scatter plots only present the mean of each sample and standard deviation.

# 3 Results

Previously our lab, with a work mainly performed by Paulo Navarro-Costa [32], showed that *dkdm5* histone H3K4me3 demethylase is a key regulator of *Drosophil*a oocyte reactivation. Germ line-specific depletion of histone demethylase *dkdm5* remodels *Drosophila* oocyte epigenome by drastically increasing H3K4me3 levels (euchromatic mark associated with the transcription start site of active genes) in prophase I oocytes, without affecting other epigenetic markers.

The dramatic increase of H3K4me3 levels in the oocyte nucleus, leads to premature transcription reactivation of the quiescent oocyte, abnormal remodelling of the oocyte chromatin during prophase I and deficient meiotic progression leading to a dramatic reduction in female fertility. Although, multiple functions have been described for *dkdm5* during *Drosophila* development, the phenotypes observed during oogenesis after *dkmd5* depletion are specific to the oocyte chromatin.

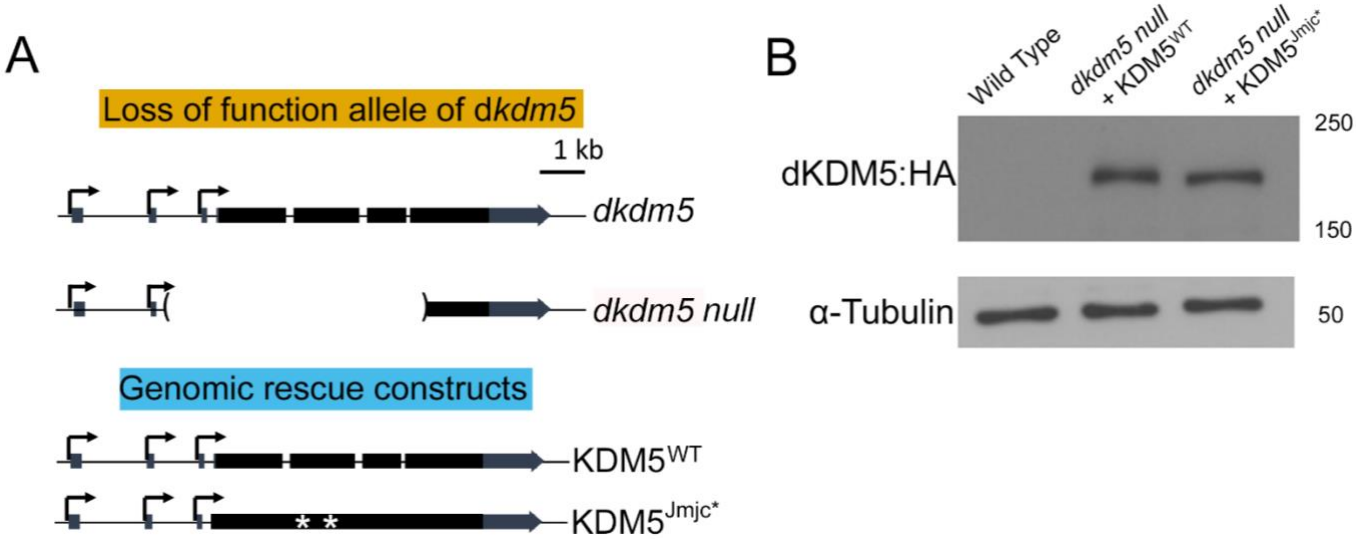



***Figure 3 - KDM5$^{WT}$ and KDM5$^{Jmjc*}$ genomic constructs are equally expressed in Drosophila oocytes.*** *(A) dkdm5 genomic sequence contains 3 alternative transcription start sites indicated by an arrow and depending on the TSS between 5 and 7 exons. dkdm5 null allele was generated by Julie et al (2018) through P element imprecise excision that resulted in the deletion of all isoforms*

*of kdm5 gene* [52]. *dkdm5 demethylase-dead (KDM5$^{Jmjc*}$ mutation* [55]*) enclose two point mutations that changed His637 and Glu639 to Alanine. For control, a wild type dkdm5 (KDM5$^{WT}$) genomic rescue transgene was generated. Both KDM5$^{Jmjc*}$ and KDM5$^{WT}$ transgenes contain a C-terminal 3xHA-tag (B) Western blot analysis confirming similar expression of KDM5$^{WT}$_HA and KDM5$^{Jmjc*}$_HA constructs in Drosophila ovaries. Ovaries dissected for Wild type (Oregon R/OR) were used as control. Three biological replicas were performed for this experiment.*

In contrast, other labs have not only reported that *dkdm5* regulates the H3K4me3 levels of the oocyte nucleus, but also that *dkdm5* depleted oocytes have meiotic recombination defects, chromosome alignment difficulties during Metaphase II and that *dkdm5* demethylase activity is not required for prophase I chromatin organisation [34].

All of the previous experiments done by both groups were performed in hypomorphs and under UAS-GAL4 RNAi system conditions. This system is known to lead to slightly different levels of protein depletion and may explain the discrepancies in the results obtained by both labs. Recently, a *dkdm5 null* allele was generated by a P element imprecise excision that removed 3 exons from *dkdm5* coding region abolishing all possible transcripts. (Figure 3A) [52]. In this work we aim to use this allele as a tool to clarify the divergent phenotypes previously described and further characterize what is the function of *dkdm5* in chromatin architecture.

## 3.1 *dkdm5* is required for female fertility.

To investigate if *dkdm5* compromises female fertility we utilize *dkdm5 null* flies (Figure 3A) provided by Julie Secombe [52] to generate germline clones of *dkdm5* using FLP/FRT system in order to obtain homozygous mutant ovaries in heterozygous flies. To test the impact of *dkdm5* demethylase activity, a *dkdm5* with impaired demethylase activity were generated (KDM5$^{Jmjc*}$) and expressed in flies with homozygous mutant ovaries. Conversely, as control wild type rescue transgene (KDM5$^{WT}$) was also expressed in this background (Figure 3A).

In *dkdm5 null* germline clones we observed that 99% of the maternal mutant eggs do not hatch (99 ±0.1% of eggs that didn't hatch, n= 1566). As germline clone control, almost all eggs were hatched (93 ±3.1% of eggs hatched, n= 596). Also, wild type flies (Oregon R; OR) were used as control showing the standard egg hatching (97 ±2.7%). *dkdm5 null* with the rescue transgene ($KDM5_{WT}$) were able to rescue the loss of *dkdm5* hatching (91 ±2.6% eggs hatched, n=830).

To test if female fertility is dependent of dKDM5 demethylase ability a demethylase-dead mutant allele of *dkdm5* ($KDM5_{JmjC}$*) was used in *dkdm5 null* background. We observed that in this condition egg hatching was significantly impair identical to *dkdm5 null* conditions (99 ±0.3% of eggs didn't hatch, n= 1800) (Figure 4 A). To be confident that the difference in egg hatching between rescue and demethylase-dead condition has not because of protein instability caused by the transgene insertion, a western blot analysis was performed (Figure 3B). Wild type flies that do not contain HA-tags were used as control. Western blot for the HA-tagged dKDM5 versions of rescue and demethylase-dead transgenes show that both are similarly expressed. As control, wild type flies without any construct were used. These results suggest that *dkdm5* is required for female fertility and this function is dependent of dKDM5 demethylase ability.

A

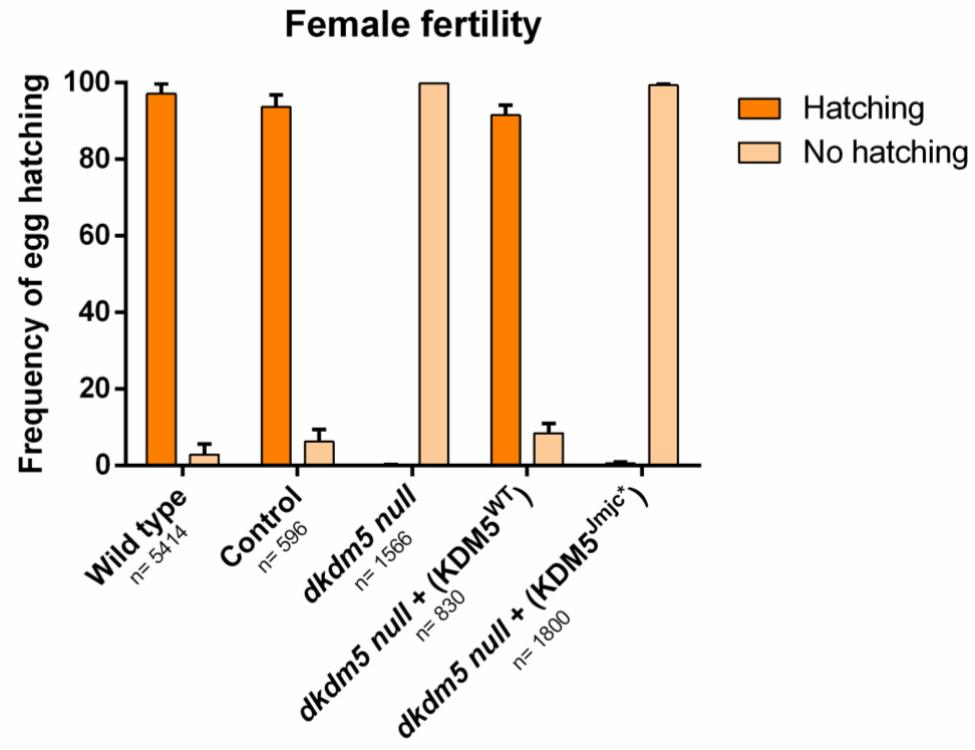


***Figure 4 dKDM5 demethylase activity is required for female fertility.*** *(A) Fertility was assessed after crossing virgin females with wild type males and egg hatching frequency was scored. KDM5$^{WT}$ but not KDM5$^{Jmjc*}$ genomic transgene was able to fully rescue dkdm5 null egg hatching. n represents the sum of eggs evaluated in at least four biological replicas.*

## 3.2 *dkdm5* regulates the levels of RNA polymerase II and chromatin architecture in prophase I arrested oocytes.

To confirm that as previously reported, *dkdm5 null* flies also have abnormal oocyte chromatin architecture and an early loading of RNA polymerase II during prophase I, we investigated *Drosophila* egg chambers [32]. First, with a DNA staining in *Drosophila* ovaries we investigated egg chamber morphology and development in *dkdm5 null* conditions and found it overall identical to the control and wild type (Figure 5 A). This indicates that *dkdm5* is not required for most developmental processes during oogenesis.

Then, we investigate if in *dkdm5 null* conditions the oocyte chromatin behaves as in RNAi depleted *dkdm5* conditions [32,34]. We found that in *dkdm5 null* conditions the oocyte chromatin also presents an abnormal morphology and a wider openness throughout oogenesis stages (Figure 6 A, B). In our control background, the oocyte chromatin has characteristic compact and round shape, until stage 8 of oogenesis were chromatin remodelling and transcription reactivation occurs as previously reported in the literature [33] . Rescue and demethylase dead construct (KDM5$^{WT}$ and KDM5$^{JmjC*}$ respectively) in *dkdm5 null* background, rescues the chromatin remodelling defects during prophase I observed in *dkdm5* homozygous mutant egg chambers.

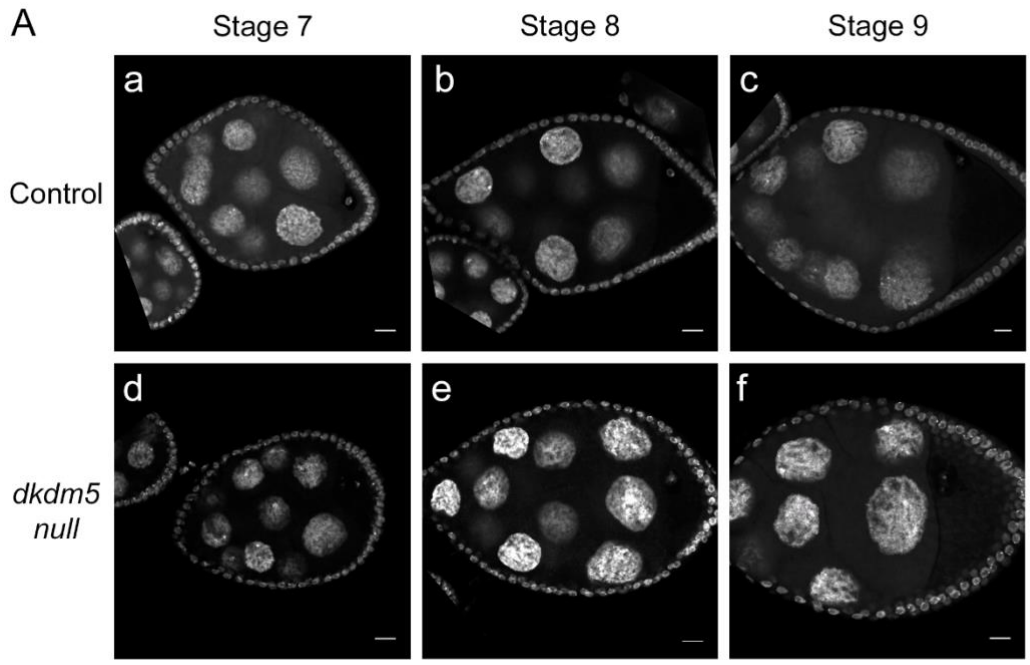



***Figure 5 – dkdm5 is not required for most developmental processes during oogenesis.*** *Morphology of the developing egg chamber's throughout oogenesis stages 7 (a, d), 8 (b, e) and 9 (c, f) in control (a, c) and dkdm5 null (d, f) backgrounds. Scale bar 10 µm.*

Also, we investigate if in *dkdm5 null* conditions oocyte transcription occurs earlier as in RNAi depleted *dkdm5* conditions [32]. Phosphorylation of the C-terminal domain of RNAPII regulates the polymerase activity and specifically phosphorylation of Serine 5 (pSer5) is suggestive of a promoter poised RNAP II. We observed by immunostaining, a significant increase of pSer5 RNA polymerase II levels in *dkdm5 null* oocytes and premature loading to the prophase I arrested nucleus (oogenesis stage 5 versus stage 8), as previously showed for RNAi depleted *dkdm5* conditions [32]. Under control conditions, pSer5 RNA polymerase II levels are lower than what is observed in *dkdm5 null* conditions, having an increase in signal by stage 8 of oogenesis (Figure 6 A, C). Rescue and demethylase dead construct ($KDM5^{WT}$ and $KDM5^{JmjC*}$ respectively) in *dkdm5 null* background, rescues transcription reactivation defects during prophase I observed in *dkdm5* homozygous mutant egg chambers.

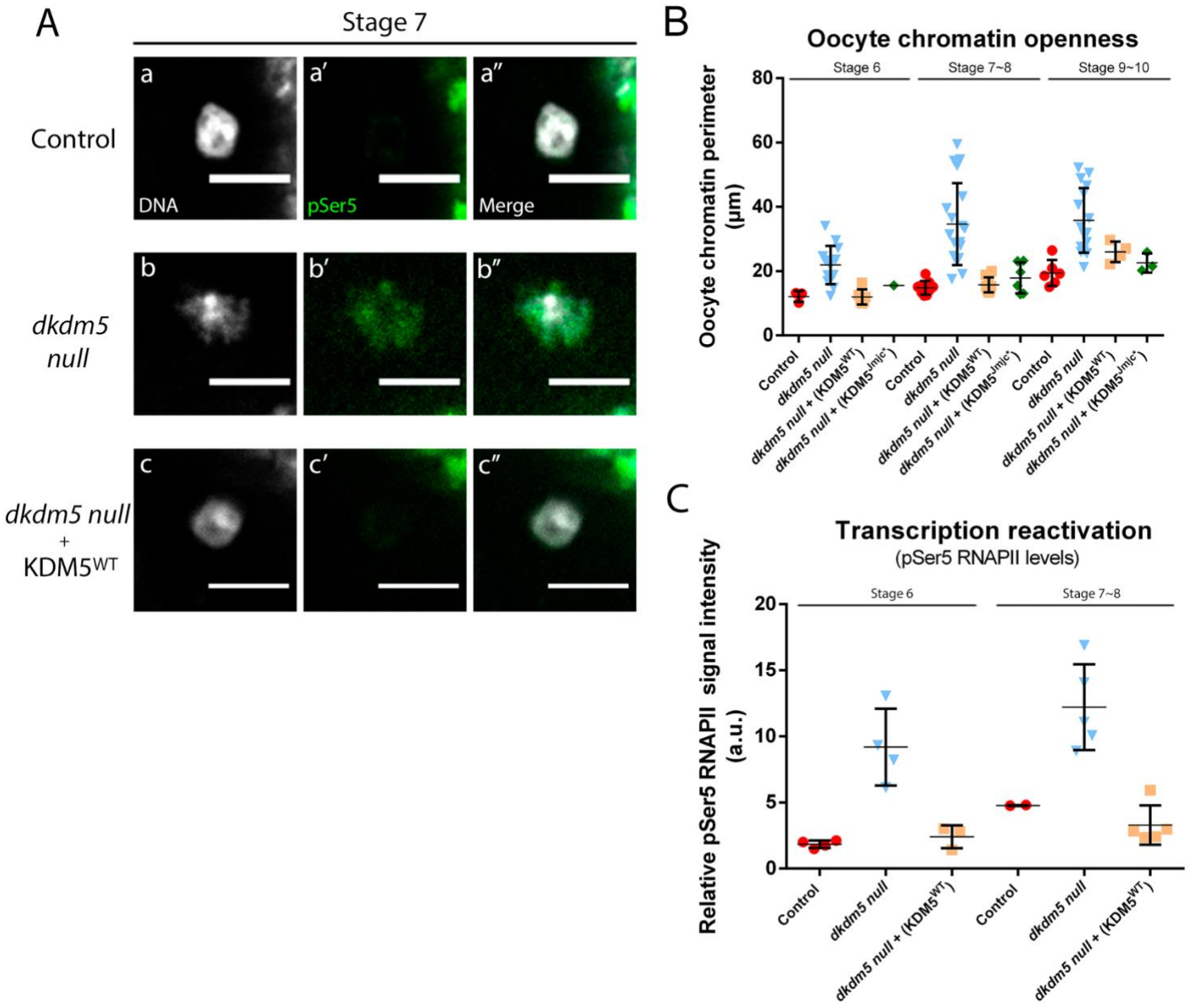



***Figure 6 – dkdm5 modulates transcription reactivation apparently independent of the demethylase ability.*** *(A-C) Loss of germline dkdm5 results in precocious transcription reactivation and abnormal chromatin architecture in prophase I arrested oocytes. Oocyte nucleus openness (karyosome openness) and polymerase loading into the oocyte chromatin is increased in dkdm5 null stage 7 egg chambers and this phenotype is rescued by the KDM5WT genomic construct. Chromatin openness was detected earlier in loss of dkdm5 (blue) when compared to the control (red), rescue construct (yellow) and the demethylase-dead construct (green). Karyosome morphology (DAPI) and detection of polymerase II (phospho-RNA polymerase II CTD (pSer5)) in control (a-a'), dkdm5*

*null (b-b') and dkdm5 null carrying a KDM5WT genomic construct (c-c') stage 7 egg chambers. (B) Chromatin openness quantification by measurement of the oocyte chromatin perimeter (µm) in control, dkdm5 null and dkdm5 null carrying a KDM5WT genomic construct stages 6-10 egg chambers. (C) Signal quantification of RNA polymerase II CTD (pSer5) throughout oogenesis. Oocyte RNA polymerase II loading was detected earlier in loss of dkdm5 (blue) when compared to the control (red) and the rescue construct (yellow).*

Together, these results suggest that *dkdm5* is required for chromatin architecture remodelling and loading of RNA polymerase II into prophase I arrested oocytes and that *dkdm5 null* phenocopies these RNAi depleted *dkdm5* phenotypes.

## 3.3 *dkdm5* is required for the correct dorsoventral patterning of *Drosophila* eggs.

During *dkdm5 null* egg chamber analysis, we observe that some prophase I arrested oocytes had a very distinct morphology of what has been described earlier, presenting a threadlike or fragmented DNA morphology versus a round and compact shape as seen in the control (Figure 7 A).

Similar morphology for prophase I arrested oocyte nucleus has been previously described by Trudi Schüpbach group in *spindle B* (*spn-B*) and *spindle D* (*spn-D*) mutant ovaries (Figure 7 B) [56]. *Spn-B* and *spn-D* (*Drosophila* orthologs from the human RAD51) are two proteins required for DNA repair in homologous recombination during meiosis by resolving double strand breaks (DSB) induced during recombination [57] leading us to hypothesise that *dkdm5* is required DNA repair in oocyte nucleus.

Since *dkdm5 null* oocyte nucleus presents morphology resemblances with the ones previously described in RAD51 mutants (without *spn-B* or *spn-D*).

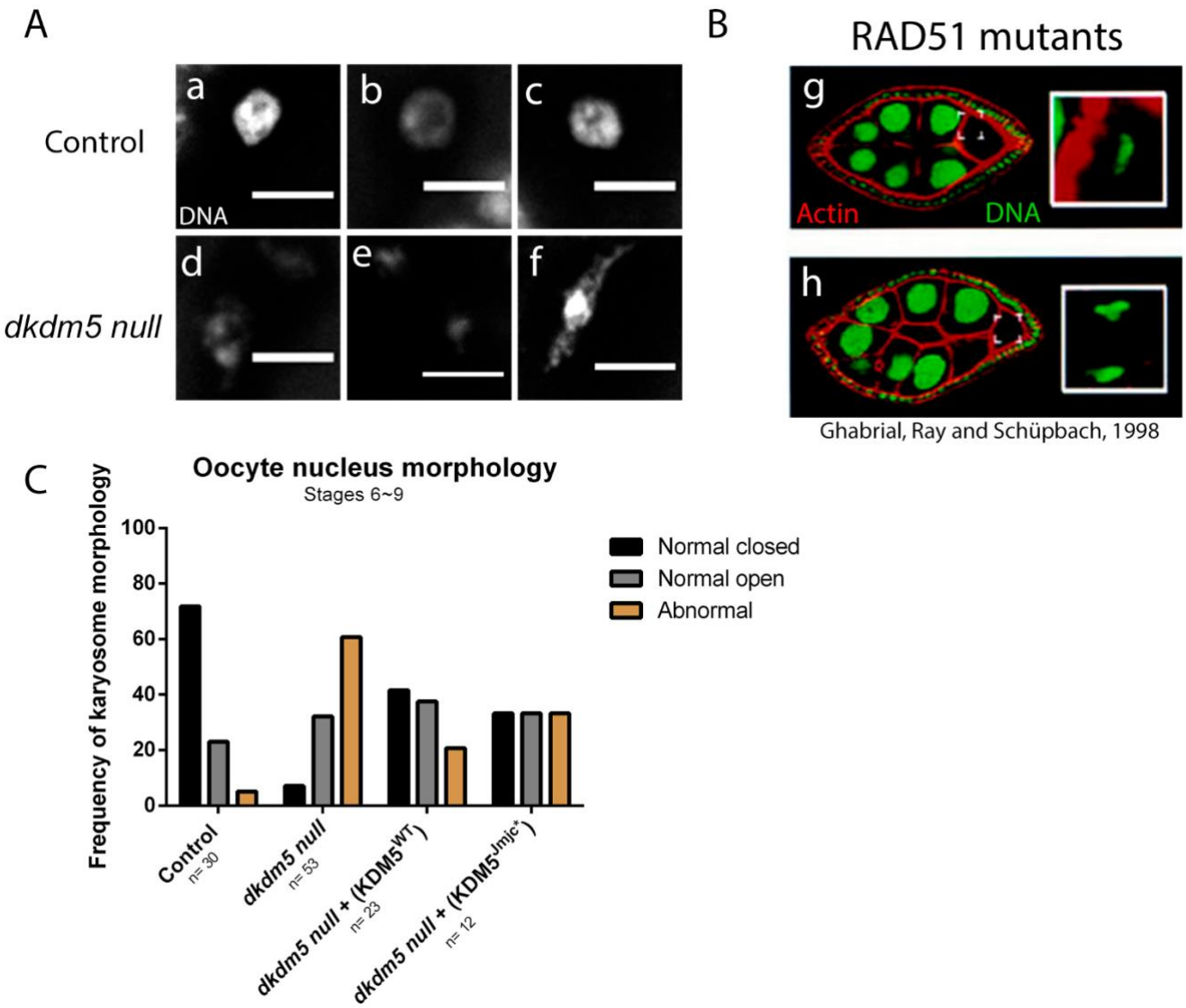


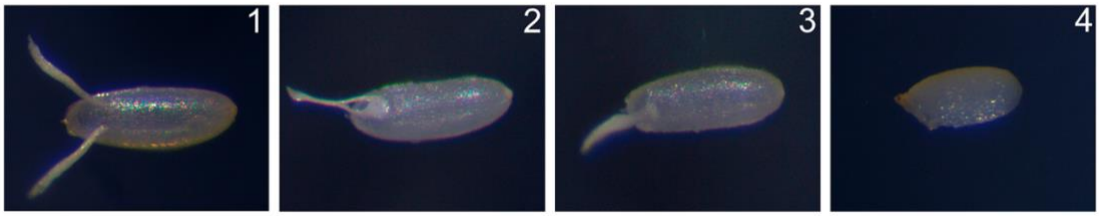

***Figure 7 - Highly abnormal oocyte nucleus morphology after dkdm5 depleted mutant.*** *(A) Karyosome morphology in ovaries from control and dkdm5 null. Ovaries from dkdm5 null flies present several oocyte nuclei with morphologies that resembled the DNA repair mutant phenotypes previously described in the literature (example shown in figure 5B). Scale bar 5 µm. (B) Typical karyosome morphology in the presence of extensive DNA damage induced by RAD51 mutant with chromatin extended (a) or shattered (b) across the nuclear envelope actin (red), chromatin (green) (adapted from* [56]*). (C) Quantification of oocyte nucleus morphology. n represents the number of oocytes nucleus summed from at least three independent biological replicas.*

To test this hypothesis, we performed germline clones to obtain *dkdm5* homozygous mutant ovaries in heterozygous flies and analysed the dorsoventral patterning of mutant eggs by scoring the position of the eggshell dorsal appendages. It is known that accumulation of DNA damage during *Drosophila* meiotic recombination will prolongate the meiotic checkpoint arrest, leading to an inefficient increase TGFα-like protein Gurken (Grk) affecting its localization and translation, creating an improper dorsoventral patternization of *Drosophila* eggs [58]. Therefore, analysis of dorsoventral patterning of mutant eggs can be used as a proxy for DNA damage in Drosophila oocyte nucleus.

For the analysis of dorsoventral patterning in *kdm5* null *Drosophila* eggs we scored the mutant eggs appendages in classes. We divided the observed phenotypes in four different classes: 1) wild type dorsal appendages, 2) partially merged dorsal appendages (merged at bottom), 3) complete merged dorsal appendages, and 4) short or extremely dorsal appendages (Figure 8 A).

In *dkdm5 null* mutant homozygous eggs we observed mild defects in the dorsal appendages of these eggs been 65 ±10% of the mutant eggs class 1, 20 ±5% class 2, 13 ±5% class 3 and 2 ±2% class 4 (Figure 8 B). As germline clone's control, almost all eggs showed wild type dorsal appendages, 98 ±2% class 1, 2 ±1% class 2, 0 ±0.3% class 3 and 1 ±0% class 4. Wild type flies (OR) were also used as control showing normal morphology of the dorsal appendages 99 ±0.1% class 1, 0 ±0.1% class 2 and no other classes recorded. The rescue transgene ($KDM5_{WT}$) in *dkdm5 null* background was able to rescue de mild phenotypes observed in the homozygous mutant eggs showing 93 ±4% of class 1 eggs, 4 ±2% class 2 eggs, 2 ±2% class 3 eggs and no class 4 eggs recorded. To test if the dorsoventral patterning is dependent of the demethylase ability of *dkdm5*, we used demethylase-dead mutant allele of *dkdm5* ($KDM5_{JmjC}^{*}$) in *dkdm5 null* background. In this condition we observed that the mild phenotypes present in *dkdm5 null* eggs were rescued presenting 90 ±5% of class 1 eggs, 7 ±4% of class 2 eggs, 3 ±0% of class 3 eggs and 1 ±0% of class 4 eggs. These results suggest that *dkdm5* is required for the dorsoventral patterning of *Drosophila* eggs. Also, we found that this function is independent from the demethylase ability of dKDM5.

A




Eggshell dorsal appendages scored classes

B

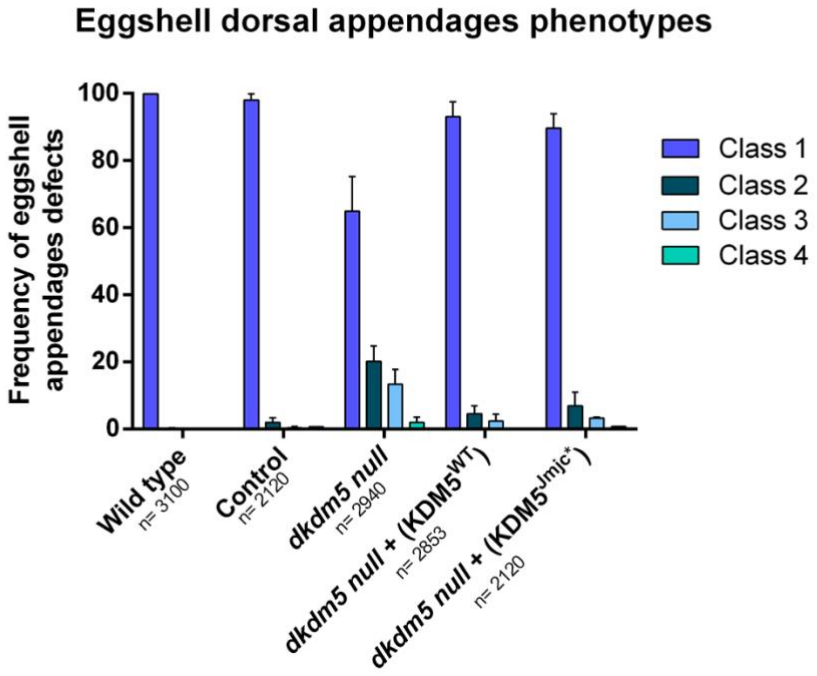
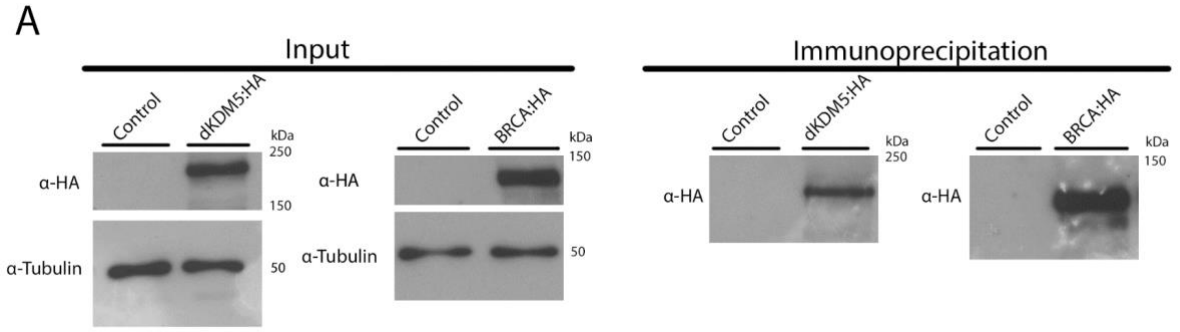



***Figure 8 – dKDM5 demethylase activity is not required for dorsoventral patterning of the eggs.*** *(A) Ventralization of Drosophila eggs was assessed by position of the dorsoventral appendages base. The observed phenotypes were categorized in four different phenotypic classes based on the eggshell dorsal appendages: class 1 (wild type appendages; two individualized dorsal appendages); class 2 (appendages fused at the bottom); class 3 (appendages totally fused; spindle phenotype); class 4 (short eggs without or with short dorsal appendages). (B) Quantification of ventralized eggshell phenotype. Both rescue construct and demethylase dead construct supresses the dorsoventral patterning defects. n represents the sum of the embryos evaluated in more than three biological replicas.*

## 3.4 *dkdm5* interacts with *Brca2* during recombination to avoid accumulation of DNA damage

*Drosophila Brca2* mutants also show a similar threadlike or fragmented DNA morphology of oocyte nucleus as seen in *dkdm5 null* egg chamber [59]. Further it was shown that *Drosophila Brca2* is required for the repair of meiotic double-stranded breaks and the efficient activation of the meiotic recombination checkpoint [59] leading to mild defects in the dorsoventral patterning of the eggs.

A




B

| Drosphila | | KDM5:HA | BRCA:HA | |
|---|---|---|---|---|
| CG | Gene | Rep1 | Rep1 | Rep2 |
| CG9088 | Lid/KDM5 | +++ | - | - |
| CG30169 | BRCA2 | - | ++++ | ++++ |
| CG10236 | Lamin A | +++ | - | - |
| CG7123 | LanB1 | +++ | - | - |
| CG3322 | LanB2 | ++ | - | - |

***Figure 9 – Absence of detectable protein interaction between dKDM5 and BRCA2 in protein immunoprecipitates from Drosophila ovaries.*** *(A) Western blot analysis of total extracts (left panel) and immunopurified extracts (right panel) obtained from ovaries of flies expressing dKDM5-HA and BRCA-HA. Ovaries from OR flies were used as control. (B) LC-MS analysis of immunopurified dKDM5-HA and BRCA-HA from ovaries. (−), (+), (++), (+++) correspond to 0, 1–9, 10–19, and >20 non-repeated peptides, respectively.*

In mouse, it has been shown that KDM5B is a bona fide DNA damage response protein and its recruitment to the double strand break sites is required for efficient BRCA1 binding, promoting the accumulation of BRCA1 during recombination [60]. These finding, together with our results led us to hypothesise

that *dkdm5* is important for *Brca2* recruitment to avoid the accumulation of DNA damage during meiotic recombination.

To test this hypothesis, we prepared protein extracts from ovaries of dKDM5WT endogenous flies that contain three HA-tags and a specific germline induced Brca2 with three HA-tags. In both extracts we immunoprecipitated dKDMA5-HA or Brca2-HA using anti-HA couple dynabeads (Figure 9). After Liquid chromatography–mass spectrometry (LC-MS) analysis of dKDM5 and Brca2 ovary protein immunoprecipitates, we failed to see an interaction between these two proteins in *Drosophila* ovaries.

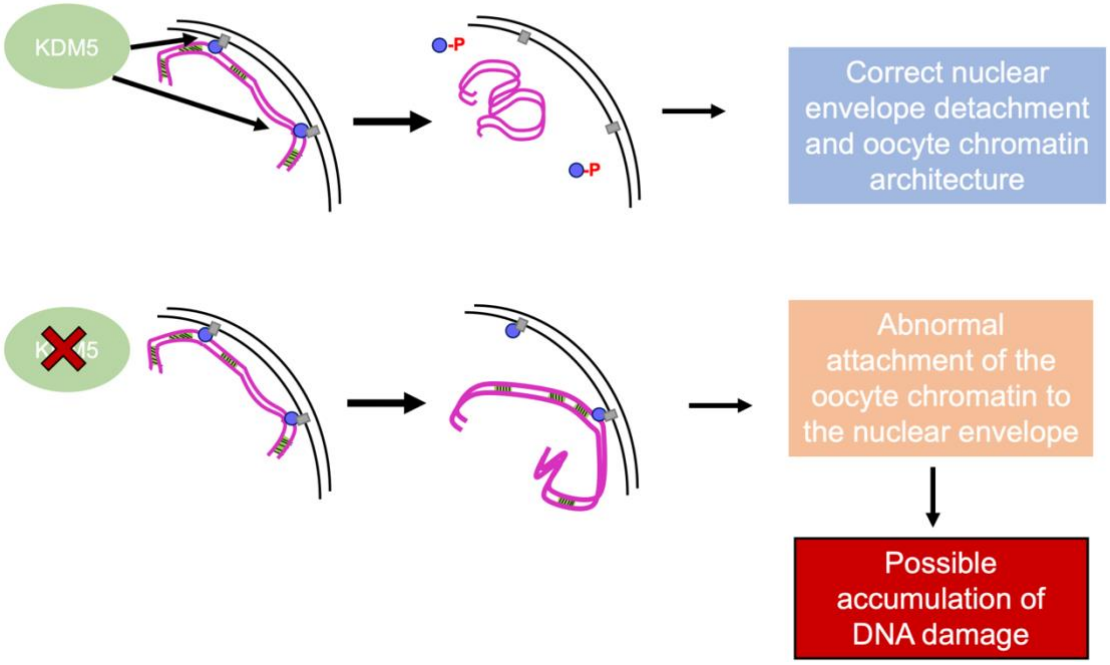



***Figure 10 – Proposed model of dKDM5 interaction with the nuclear envelope.*** *Under normal conditions, dKDM5 (green) interacts with components of the nuclear envelope (blue ball and grey square) allowing the correct detachment of the oocyte chromatin and correct establishment of the oocyte chromatin architecture (pink lines). Abolisment of dKMD5 leads to a atypical detachment of the oocyte chromatin from the nuclear envelope, leading to a possible accumulation of DNA damage.*

Although, the interaction between dKDM5 and Brca2 was not detected, LC-MS analysis of dKDM5 revealed a strong association with components of the nuclear envelope. Barrier to autointegration factor proteins are core components of the nuclear envelope that are essential for the crosstalk between chromatin and the nuclear envelope. During *Drosophila* development, for the correct formation of the oocyte nucleus, nucleosomal histone kinase-1 is required for BAF phosphorylation allowing the oocyte chromatin to detach from the nuclear envelope [29]. Prolonged activation of the meiotic checkpoint, provoked by defective DNA breaks repair, suppresses *nhk-1* activity preventing the detachment of the oocyte chromatin to the nuclear envelope, further on impairing the formation of the prophase I oocyte nucleus [30]. Therefore, our current working hypothesis is that *Drosophila dkdm5* is important for oocyte chromatin architecture because it regulates the crosstalk between oocyte chromatin and the nuclear envelope (Figure 10).

# 4 Discussion.

In this chapter we characterised *dkdm5* loss of function in *Drosophila melanogaster* oogenesis. *dkdm5 null* allele was generated by Julie Secombe from Albert's Einstein College of Medicine and the loss of function is characterized by a P element imprecise excision that abolishes all *dkdm5* transcripts [52].

Previously in our lab, we described histone H3K4me3 demethylase *dkdm5* to be required to set the correct H3K4me3 levels in the developing *Drosophila* oocytes nucleus [32]. Also, we have shown that specific germline depletion of *dkdm5* results in a premature reactivation of gene expression in otherwise transcriptionally quiescent oocytes anticipating transcriptional reactivation by approximately 14h. Jointly with the premature transcription reactivation there is also a precocious prophase I oocyte chromatin remodelling. Although we did not detect any metaphase I defects, we shown that upon *dkdm5* specific germline depletion, meiosis completion is impaired further leading to female fertility defects.

Other research groups have also shown that *dkdm5* is a major H3K4me3 regulator and is required for the correct *Drosophila* oocyte nucleus chromatin architecture. Conversely, it was also shown that *dkdm5* is required for the correct assembly of the SC complex during *Drosophila* meiotic recombination and correct centromere positioning throughout the oocyte development [34].

Regarding the relevance of *dkdm5* demethylase activity, in our lab, we observed that loss of *dkdm5* demethylase activity resulted in similar defects to those recorded after *dkdm5* germline depletion, showing premature oocyte transcription reactivation, increased oocyte pSer5 RNAPII levels, precocious oocyte chromatin remodelling, impaired entry into embryogenesis after fertilization and reduced fertility. On the other hand, it was also shown that *dkdm5* flies carrying the catalytically inactive transgene $KDM5_{JmjC}$* present a largely normal karyosome morphology, centromere clustering and SC morphology.

Having in account that most of the experimental procedure were performed under hypomorphic and under UAS-GAL4 RNAi system conditions,

where the target protein is not completely abolished but it is depleted within a certain spectrum, which can give raise to phenotypical divergences as referred previously. Additionally, although *dkdm5* is a strict H3K4me3 it has distinct functions during Drosophila development; such as been required to establish oxidative stress response [47], dMyc-induced cell growth [55], regulation of the circadian rhythms periods [61], female fertility [32] and being required for proliferation and differentiation of *Drosophila* larval hemocytes [62], it seems that during *Drosophila* oogenesis *dkdm5* has a specific role on the regulation of the oocyte chromatin without affect other cellular structures.

We hypothesise that with a loss of function of *dkdm5* we can undercover new mechanisms that *dkdm5* can regulate during oogenesis, regarding not only in terms of the oocyte chromatin architecture, but also unravelling the demethylase dependent and independent processes.

In this chapter we observe that in *dkdm5 null* oocytes, *dkdm5* seems to be dispensable for most developmental processes during *Drosophila* oogenesis, since no morphological structure, besides the karyosome, are affected throughout oogenesis. In addition, the oocyte has the correct development from the starting of germarium continuously progressing to a fully developed oocyte. Therefore, with this observation we assume that the role of *dkdm5* is not required for most developmental processes during oogenesis.

We observe that female fertility in *dkdm5 null* flies is severally impaired, albeit the phenotypes are more emphasised when compared to experiments done in hypomorphic or under UAS-GAL4 RNAi system conditions. Further, it is noteworthy to mention that *dkdm5* demethylase activity is required for female fertility showing comparable fertility defects has the ones observed in *dkdm5 null* conditions.

Concurrently with the results previously shown in our lab, *dkdm5* is required for the correct *Drosophila* oocyte transcription reactivation and normal oocyte chromatin architecture throughout oogenesis. *Drosophila* oocyte chromatin remodelling in normal condition takes place in oogenesis stage 9. Our observations regarding *Drosophila* oocyte chromatin architecture in *dkdm5 null* egg chambers suggest even more accentuated defects in oocyte chromatin when

compared with the same developmental stages both in normal conditions and in hypomorphic conditions. Importantly, it seems that oocyte chromatin architecture remodelling is mostly independent of *dkdm5* demethylase activity, since the demethylase dead construct in a homozygous mutant background egg chambers rescues the premature and abnormal chromatin architecture remodelling phenotypes. These results, confirm has suggested in UAS-GAL4 RNAi [34], that the demethylase activity of *dkdm5* may not be required for prophase I oocyte chromatin architecture.

We observed that, although in *dkdm5 null* flies the oocyte chromatin architecture is abnormal with a wider openness, some of them had a very specific structure that fell into two categories: threadlike or fragmented DNA morphology (Figure 7A). These structures have been previously described in RAD51 (*spn-B* or *spn-D*) mutant , where the oocyte nucleus chromatin is represented has an aberrant condensed nucleus [56]. During oogenesis, the activation of the meiotic DNA damage checkpoint after the induce DSB during meiotic recombination is required for the DNA damage repairing mechanism to take place. DSBs triggers the activation of a pathway that halts the cell cycle until the damaged DNA is repaired. The delay induced from the activation of the meiotic checkpoint is sufficient to allow a full repair of all DNA breaks. Under conditions in which DNA damage cannot be fully repaired, such as in RAD51 mutants, there is a persistence accumulation of unrepaired DSBs that are not resolved leading to a prolonged activation of the meiotic checkpoint and eventually oogenesis will proceed without DNA repairing. The prolonged meiotic checkpoint arrest results in Vasa phosphorylation that will further block the efficient translation of Grk. Has mentioned before Grk is required for the correct dorsoventral patterning of *Drosophila* eggs, therefore, if the efficient translation of Gkr is affected, this will lead to axial patterning defects in *Drosophila* eggs such as seen in RAD51 mutants and other well-known DNA damage repairing mutants such as *Brca2* [58,59].

Considering that in *dkdm5 null* flies the oocyte chromatin architecture defects resembles the ones reported in RAD51 mutants, we hypothesise that *dkdm5* is required for the correct dorsoventral patterning of *Drosophila* eggs.

Accordingly to our hypothesis, we show that *dkdm5* is required for the correct dorsoventral patterning of *Drosophila* eggs. The observation of a threadlike or fragmented DNA morphology in *Drosophila* oocyte chromatin and egg pattering defects haven't yet been described in previous works. This may possibly result due to the fact that the experimental procedures done in the past were performed in hypomorphic or RNAi conditions, were gene expression is reduced and not completely abolished, and low *dkdm5* levels can still accomplish their role during meiotic recombination. However, it is notable to mention that in this chapter we observed that *dkdm5* demethylase activity seems to be dispensable for *Drosophila* eggs dorsoventral patterning since, homozygous mutant eggs with *dkdm5* demethylase activity impaired show normal dorsoventral patterning, rescuing the effects observed in *dkdm5 null* eggs.

Even though, we fail to show a direct connection between *dkdm5* and DNA damage (data not shown), we can establish a correlation since the observed phenotypes in *dkdm5 null* oocytes and eggs are comparable the ones reported in DNA repair mutants. In addition to what was previously mentioned regarding RAD51 mutants, it has been shown in the literature that *Drosophila Brca2* is required for homologous DBS repair during meiotic recombination. *Drosophila Brca2* absence has been associated with defects in DNA damage repair with abnormal DNA morphology showing threadlike or fragmented structures and defects in the dorsoventral patterning of *Brca2* mutant eggs, similar to the ones we report in this work [59].

Further, it has been described that mouse KDM5B is a bona fide DNA damage response protein, being its recruitment to the DSB sites upon ionizing radiation required for efficient BRCA1 local accumulation and binding to DSBs during recombination [60]. Therefore, we hypothesise that *dkdm5* is important for *Brca2* recruitment to avoid the accumulation of DNA damage during meiotic recombination, being this function independent from *dkdm5* demethylase since, there is no correlation regarding oocyte chromatin morphology or egg dorsoventral defects in the results observed under demethylase impaired activity conditions.

To test this hypothesis, we tried to investigate if there is a protein interaction between *dkdm5* and *Brca2* in *Drosophila* ovaries by co-

immunoprecipitation and LC-MS analysis. Although we failed to see any interaction between these two proteins, it is noteworthy to notice that upon LC-MS analyses of *dkdm5* immunoprecipitated ovaries there is a strong association between *dkdm5* and components of the nuclear envelope more specifically nuclear lamins.

During *Drosophila* development, *nhk-1* is required to the correct oocyte nucleus formation. Mutants for *nhk-1* result in an anchorage of the meiotic chromosomes to the nuclear envelope forming an aberrant oocyte chromatin structure has the ones observed in *dkdm5 null* oocyte nucleus. The anchorage between the oocyte chromatin and the nuclear envelop have been described to be mediated by the phosphorylation of BAF proteins [29]. BAF proteins are a specific substrate of *nhk-1* and constitute core components of nuclear envelope, being essential for the crosstalk between the nuclear envelope and oocyte chromatin.

For the correct *Drosophila* oocyte development, it is required for the oocyte chromatin to completely detached from the nuclear envelope to form the karyosome. Therefore, has shown in the literature, *nhk-1* is required for BAF phosphorylation allowing the oocyte chromatin to detach from the nuclear envelope, has expression of non-phosphorylated BAF disrupts karyosome formation [29]. Subsequently, it has been shown that unrepaired DNA breaks suppress *nhk-1* kinase activity, being this suppression mediated by the meiotic checkpoint. Prolonged delay in the resumption of the meiotic checkpoint, induced by defects in DNA damage repair mechanisms, will suppresses *nhk-1* activity leading to an accumulation of non-phosphorylated BAF. This accumulation will prevent the correct detachment of the oocyte chromatin to the nuclear envelope, impairing the formation of the prophase I oocyte nucleus, has oogenesis will eventually proceed [30]. Our current working hypothesis is that *Drosophila dkdm5* is important for oocyte chromatin architecture because it regulates the crosstalk between oocyte chromatin and the nuclear envelope.

Collectively, our work regarding the loss of function of *Drosophila kdm5* revealed that, although *dkdm5* is involved in plenty of processes during *Drosophila* development it seems that its role is not required for most developmental processes during oogenesis. In addition, we show that *dkdm5* not only is required for female fertility, but also its demethylase activity is necessary. Regarding the oocyte chromatin architecture, we show that *dkdm5* is required for normal oocyte chromatin architecture although this process is mostly independent from its demethylase ability. Furthermore, we unravel a new undocumented phenotype regarding the oocyte chromatin morphology. Although we fail to prove a direct connection between *dkdm5* and DNA damage we show that *dkdm5* is required for correct dorsal ventral patterning, mostly independent of its demethylase activity. As far as we are aware, this is the first time that DNA damage like structures and defects in the dorsoventral patterning are seen in depleted *dkdm5* conditions.

# Chapter II

# 5 Introductory concept.

The synaptonemal complex is a protein complex that is assembled between homologous chromosomes during prophase I of meiosis being essential for recombination and chromosome segregation. Defects of SC assembly have been associated with meiotic recombination defects, further compromising female fertility [63]. *dkdm5* was also shown to be required for the correct formation and stability of the SC during meiotic recombination and further been crucial for female fertility, without depending on dKDM5 demethylase function [34].

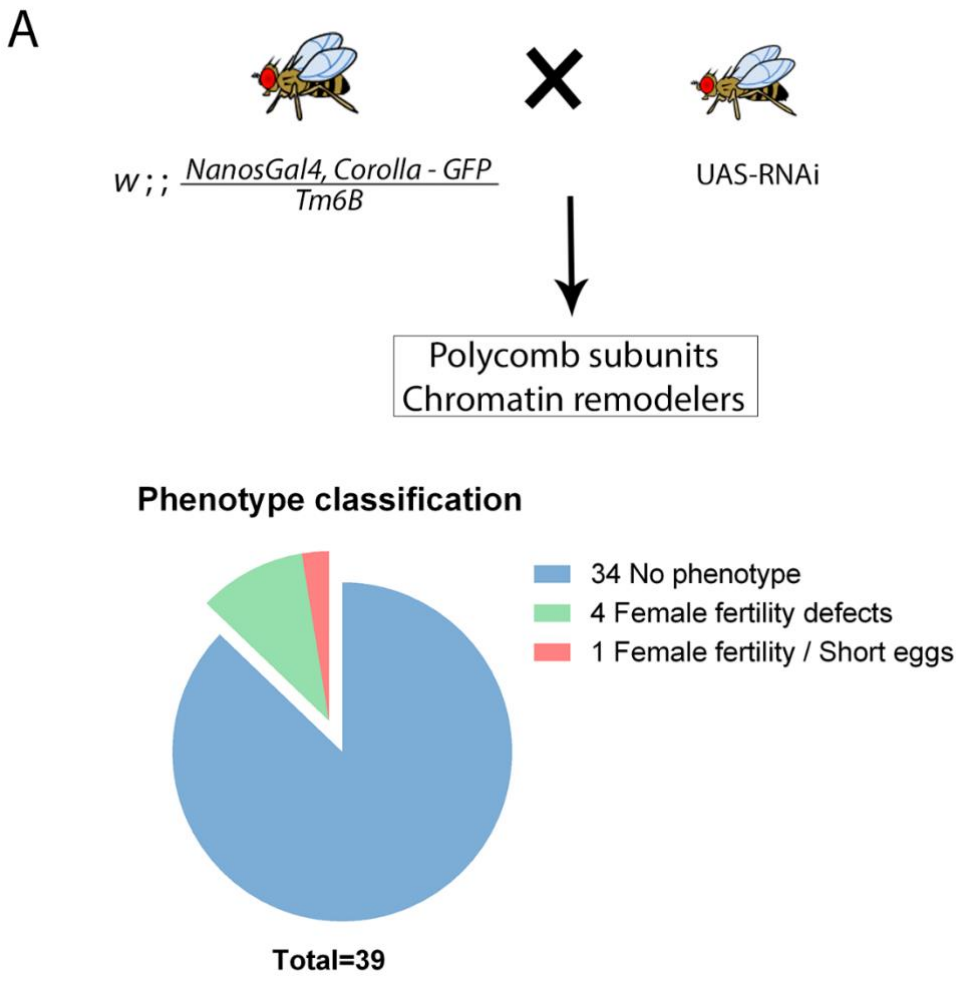



***Figure 11 – Chromatin remodelers RNAi screen.*** *(A) Schematic representation of the performed RNAi screen. We analysed a total of 39 polycomb subunits and chromatin remodelers related RNAi lines that are summarized in the pie chart. The primary screen was further characterized according to the egg shape, egg laying, and embryo eclosion rates. Total of 34 RNAi lines tested had no*

*phenotype (blue), 4 of the lines presented female fertility defects (green) and 1 presented not only fertility problems as well as dorsoventral patterning defects (orange).*

In our lab we also shown that *dkdm5* affects the assembly SC protein Corolla during *Drosophila* oocyte development (data not shown). We hypothesize that there are other proteins capable of regulating the SC assembly.

To test this hypothesis, we performed an RNAi screen for known chromatin remodelers. To investigate the SC assembly, we used flies that contain an endogenous construct of Corolla with a GFP-tag and a specific germline promoter (Nanos-Gal4) to achieve a germ line-specific expression of the RNAi (Gal4-UAS system [64]). In this screen we analysed 39 RNAi stocks for known chromatin remodelers (Table 1) and investigated the impact of RNAi-mediated knockdown by assessing defects in female fertility and abnormal dorsoventral patternization of *Drosophila* embryos (Figure 11).

## 5.1 *Drosophila dSfmbt* and *Pho* are required for female fertility.

From the 39 tested RNAi's, 34 RNAi lines do not present any detectable phenotype and female fertility was not comprised. 4 RNAi lines are associated to a reduction in female fertility and 1 RNAi line is associated to a reduction in female fertility and to a high percentage of short eggs laid by female flies (Figure 11).

We further characterized these 5 RNAis *(Pho* RNAi I, *Pho* RNAi II, *dSfmbt*, *PCL* and *BRM*) assessing female fertility and eggshell appendages position using apple agar plaques (Figure 12). Egg hatching in females with germline-specific depletion of *dSfmbt* is severely impaired when compared to the control (0% of the eggs hatch, n= 285). Germline-specific depletion with both *Pho* RNAi's leads an equivalent reduction in egg hatching, with 50 ±0.5% of egg hatching for *Pho* RNAi I (n=732) and 54 ±1% of egg hatching for *Pho* RNAi II (n= 811). Similar to *Pho*, germline-specific depletion *BRM* leads to a reduction of egg hatching (52 ±0.5%; n=1074). In this second analysis we were unable to confirm the previous results for germline specific depletion of PCL (93 ±1% eggs hatched, n=601) which indicates that this RNAi was a false positive in our first analysis.

***Table 1 – Representation of the RNAis that were tested for the preformed screen****. This table represented the complexes that each tested RNAi belongs to, the RNAi's tested and their Bloomington number. Also, in this table we have a preliminary analysis of egg eclosion and dorsoventral patterning defects (DV defects). The preliminary analysis was done by looking at the crosses food vial. RNAis were the flies had no egg laying could be discarded. Egg eclosion is shown in percentage and none of the test RNAis presented problems in dorsoventral patterning****.***

| Complex | Symbol | BL# | Egg eclosion | Dv defects | | Complex | Symbol | BL# | Egg eclosion | Dv defects |
|---|---|---|---|---|---|---|---|---|---|---|
| Prc1 | ph-p | 35207 | 95% | None | | PhoRC | Sfmbt | 32473 | 0% | None |
| | Psc | 35297 | 95% | None | | | pho | 35206 | 30% | None |
| | Psc | 38261 | 95% | None | | | pho | 42926 | 30% | None |
| | Su(z)2 | 57466 | 80% | None | | | phol | 64945 | 95% | None |
| | ph-d | 63018 | 90% | None | | | | | | |
| | Pc | 36070 | 95% | None | | Pr-DUB | calypso | 56888 | 95% | None |
| | ph-p | 33669 | 90% | None | | | Asx | 51677 | 95% | None |
| | | | | | | | Scm | 55278 | | |
| dRaf | Psc | 35297 | 95% | None | | | | | | |
| | Psc | 38261 | 95% | None | | Brm complex | brm | 35211 | 30% | None |
| | Su(z)2 | 57466 | 80% | None | | | osa | 35447 | 90% | None |
| | Su(z)2 | 33403 | 80% | None | | | mor | 35630 | No egg laying | |
| | Sce | 35446 | 95% | None | | | | | | |
| | Kdm2 | 33699 | 90% | None | | Chromatin Remodellers | BEAF-42 | 35642 | 80% | None |
| | | | | | | | ttk | 36748 | 90% | None |
| Prc2 | Jarid2 | 32891 | 95% | None | | | pgc | 33720 | 90% | None |
| | Su(z)12 | 33402 | 10% | None | | | pgc | 32386 | 95% | None |
| | Pcl | 33945 | 30% | None | | | Med | 43961 | 0% | None |
| | Pcl | 33946 | 10% | None | | | CG7839 | 63648 | 90% | None |
| | Caf1 | 34069 | 0% | None | | | | | | |
| | Jarid2 | 40855 | 90% | None | | | | | | |
| | E(z) | 36068 | No egg laying | | | | | | | |
| | E(z) | 33659 | No egg laying | | | | | | | |
| | jing | 55633 | No egg laying | None | | | | | | |
| | jing | 35750 | 90% | | | | | | | |

To confirm that the results obtained are not due to a genetic interaction with the Corolla GFP construct, we crossed the selected RNAi's with the specific germline promoter only. We observed similar results as seen previously for egg hatching and dorsoventral patterning. We observed 94 ±1% of eggs hatched in control conditions, 0% of eggs hatched in *dSfmbt* RNAi, 39 ±0.5% of eggs hatched in *Pho* RNAi I, 55 ±3% of eggs hatched in *Pho* RNAi II, 95 ±1% of eggs hatched in *PCL* and 55 ±0.5% eggs hatched in *BRM* RNAi (Figure 12A). The observed results suggest that *Drosophila dSfmbt* and *Pho* are required for female fertility.

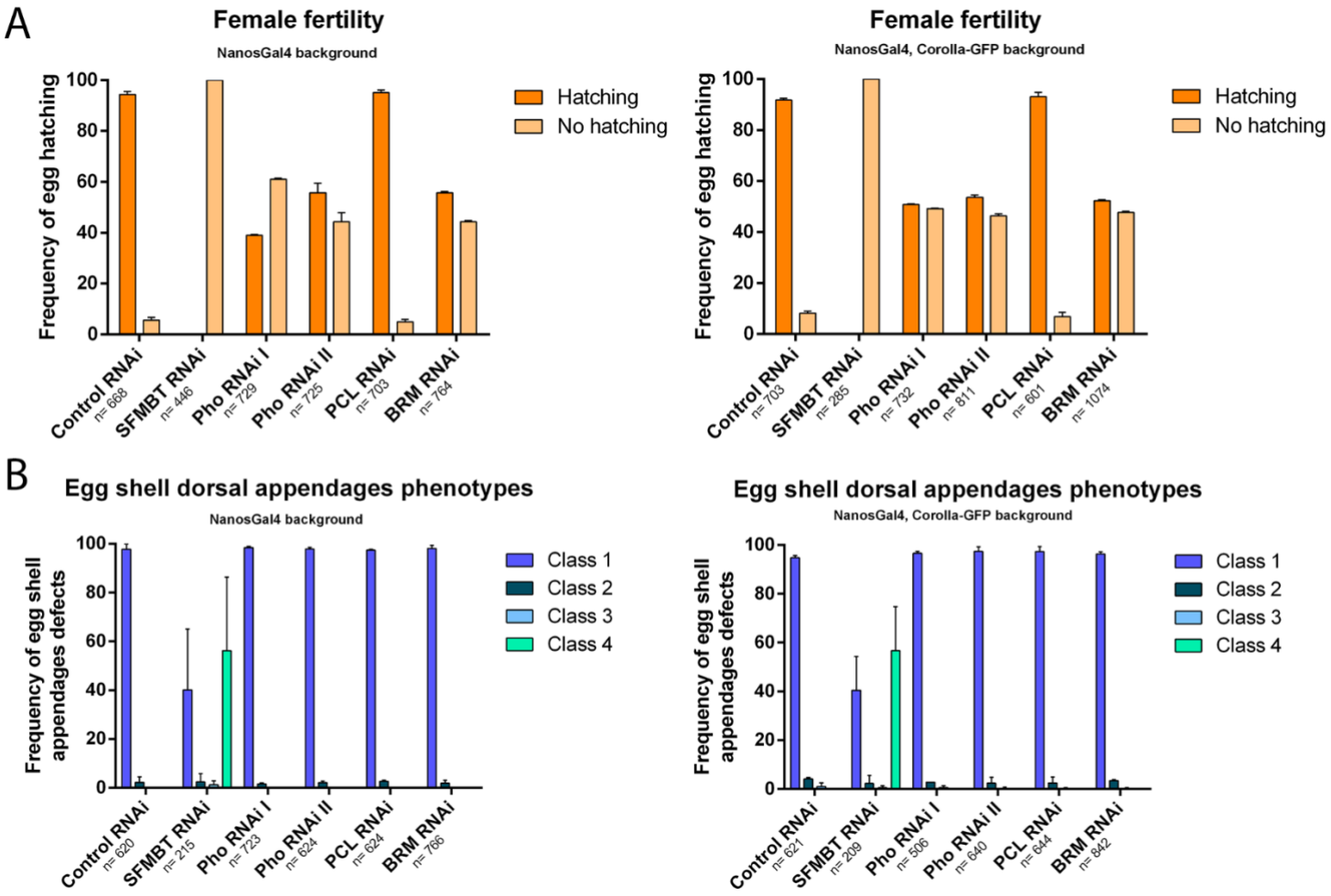



***Figure 12 - dSfmbt is required for female fertility.*** *(A) Frequency of egg hatching after specific germline depletion of Pho I, Pho II, dSfmbt, PCL and BRM by RNAi in females carrying a Corolla-GFP transgene and in a control background to prevent background or Corolla-GFP insertion specific phenotypes. Pho I and Pho II depletion results in a reduction of about 50% in egg hatching, compromising female fertility. PCL depletion does not affect the ration of egg*

*hatching. dSfmbt depletion dramatically decreases egg hatching compromising female fertility. n represents the sum of the embryos evaluated in two biological replicas. (B) Quantification of ventralized eggshell phenotype after specific germline depletion of Pho, dSfmbt, Pcl and BRM by RNAi in females carrying a Corolla-GFP transgene and in a control background. Only dSfmbt depletion induces ventralization phenotypes. Scored classes are the same and shown in figure 6B.*

Eggshell dorsal appendage position was normal for most of the analyzed genes when compared to the control (94 ±1% class 1 eggs, n= 621) displaying 96 ±1% class 1 eggs for *Pho* RNAi I (n= 506) , 97 ±2% class 1 eggs for *Pho* RNAi II (n= 640), 97 ±2% class 1 eggs for *Pcl* RNAi (n= 644) and 96 ±1% class 1 eggs for *BRM* RNAi (n= 842). Nonetheless, SFMBT RNAi presents a very specific phenotype with most eggs classified in class 1 (40 ±14%, n= 209) or class 4 (57 ±18%, n= 209) (Figure 12B).

Previously in our lab, we showed that homozygous mutant's eggs for *pkn* were significantly shorter, yet the position of dorsal appendages was normal in most cases, due to nurse cell dumping defects [36]. These shorter eggs resemble the class 4 eggs laid by females with germline specific depletion of *dSfmbt*, suggesting that *dSfmbt* also induces nurse cell dumping defects.

## 5.2 *dSfmbt* regulates synaptonemal complex components oocyte development.

To further examine the female fertility defects, we hypothesise that *Pho* and *dSfmbt* are important for the correct oocyte chromatin architecture and remodelling during prophase I. To test this hypothesis, we investigate SC integrity during oocyte development. In control conditions, Corolla-GFP co-localizes with the oocyte chromatin and its signal decreases throughout oocyte development. Beyond stage 8 Corolla-GFP signal is undetectable. Upon RNAi for *dSfmbt*, although the oocyte chromatin has a normal morphology, Corolla-GFP signal shows a very distinct scattered to filamentous like threads unseen in the control. Also, the signal intensity of Corolla-GFP upon *dSfmbt* RNAi is increased by two-

fold throughout oogenesis and its signal is still detectable at stage 10 whether in control conditions is not detectable. Both RNAi's for *Pho* showed a Corola-GFP behaviour similar to the control, with a decreasing expression along oogenesis and an undetectable signal beyond stage 8. With the observed results we conclude, *Pho* apparently is not essential for localization of Corolla to the oocyte nucleus and *dSfmbt* is required for the correct morphology and expression levels of Corolla in the oocyte nucleus. Currently, our on-going hypothesis is that *dSfmbt* is required for the correct SC assembly/disassembly during oocyte meiosis.

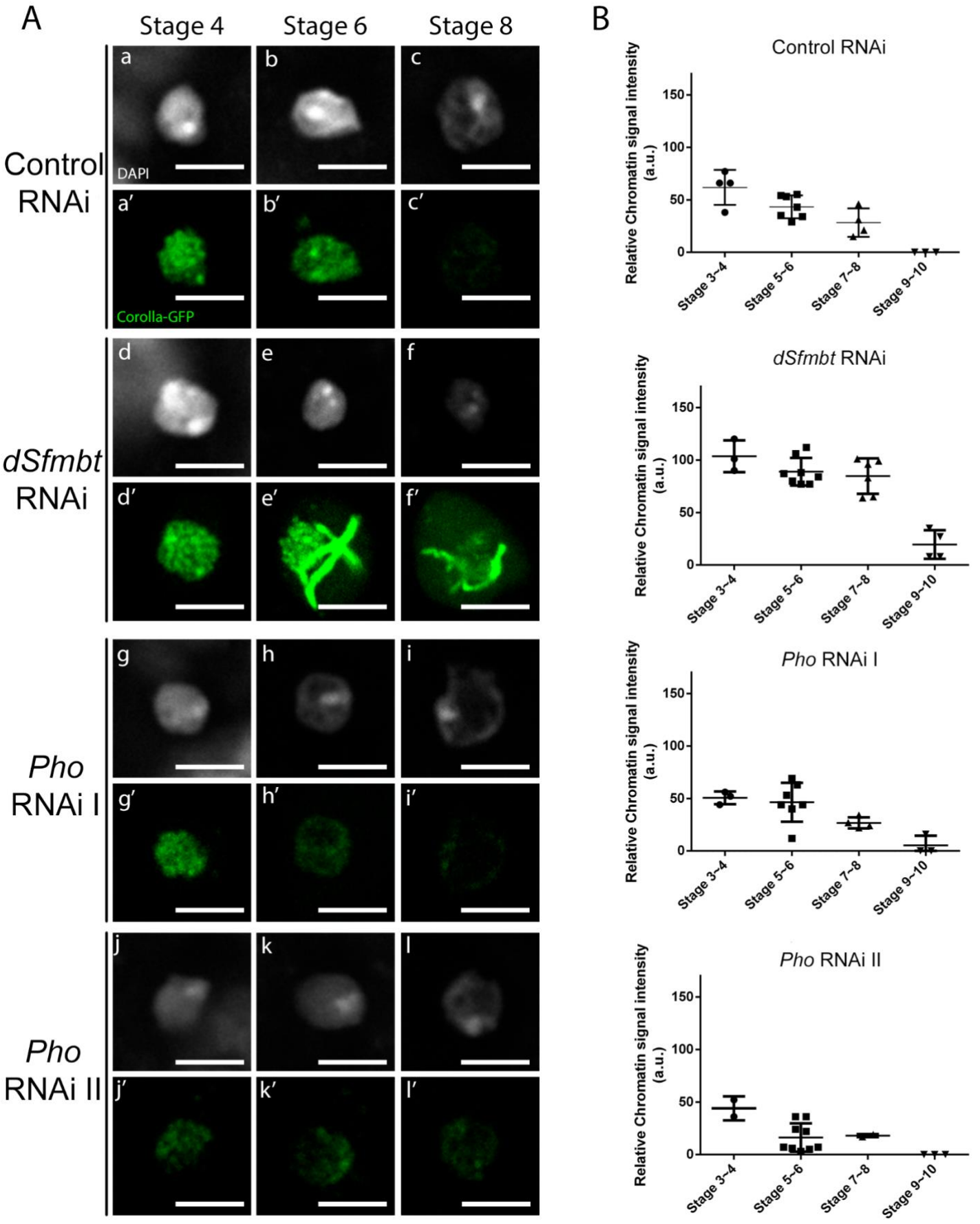

A
Stage 4
Stage 6
Stage 8
Control RNAi
dSfmbt RNAi
Pho RNAi I
Pho RNAi II
DAPI
Corolla-GFP
B
Control RNAi
dSfmbt RNAi
Pho RNAi I
Pho RNAi II
Relative Chromatin signal intensity (a.u.)
Stage 3~4
Stage 5~6
Stage 7~8
Stage 9~10

***Figure 13 – dSFMBT is required for the correct regulation of Corolla.*** *(A) Corolla-GFP localization and karyosome morphology (DAPI) in control (a-c') and specific germline depletion of SFMBT (d-f') and Pho (g-l'). Depletion Pho does not lead to detectable changes in SC assembly. SFMBT depletion leads to a strong accumulation of Corolla-GFP in the nucleus and appearance of abnormal structures. (B) Relative chromatin signal intensity quantification of Corolla-GFP throughout oogenesis represented per karyosome and is expressed in fluorescence arbitrary units (a.u.). For each graph it's represented together the total sample of three independent biological replicas. Scale bars 5 µm.*

# 6 Discussion.

The SC assembly between the paired homologous chromosomes is the first cytological evidence of meiosis initiation. The correct assembly of the SC is required the correct pairing and synapsis of the homologous chromosomes being also required for the subsequentially meiotic processes that occur during meiosis, such as recombination and chromosome segregation. Previous in the literature it has been shown that defects in SC assembly are associated with aberrant chromatin architecture leading to filamentous DNA structures [30] and defects in meiotic recombination that further will compromise female fertility [63].

*dkdm5* has been shown to also have a role during the SC assembly being required for the correct SC assembly along chromosome arms without affecting C(3)G at centromeres [34]. Previously in our lab, by the work of Ines Tinoco we have also reported defects in the SC assembly upon *dkdm5* specific germline depletion (data not shown). Due to the importance of the correct assembly of the SC during meiotic recombination we hypothesis that, such as *dkdm5*, there are other chromatin remodeler or polycomb subunits that are capable of regulating the SC assembly or disassembly.

To access our hypothesis, we performed a specific germline RNAi screen for known chromatin remodelers an DSBs triggers the activation of a pathway that halts the cell cycle until the damaged DNA is repaired. d polycomb subunits. In this screen we analyzed female fertility, eggshell dorsoventral patterning, oocyte chromatin morphology and the SC complex. In most of the tested RNAi's, we did not observe any defects in female fertility or eggshell dorsoventral patterning. Although many of the RNAi's that were tested are supposed to target genes that are required for the regulation of many processes during *Drosophila* development, we cannot take direct conclusions since negative results in a RNAi may be due to the fact that the levels of protein depletion that are being archived may not be sufficient to impair the processes that these proteins regulate.

From this screen, we were able to observe that *Drosophila dSfbmt* and *Pho*, two proteins that compose the core pho repressive complex, are required for female fertility. Additionally, we observed that specific germline depletion of *Brm* leads to defects in female fertility. Due to the extended literature that is

already done in this gene and due to the its pleiotropic effects, we choose not to pursue our analyses with this gene.

Regarding eggshell morphology, we observed that although depletion of *Pho* and *dSfbmt* was sufficient to impair female fertility, only *dSfbmt* is required for normal egg growth. Furthermore, we analyse the eggshell morphology of these specific germline depleted eggs by categorizing dorsoventral patterning defects from a phenotypical spectrum in four classes: 1) wild type dorsal appendages, 2) partially merged dorsal appendages (merged at bottom), 3) complete merged dorsal appendages, and 4) short eggs or extremely reduce dorsal appendages. When analysing specific germline depleted *dSfmbt* eggshells we noticed that we only had two classes, wild type dorsal appendages and eggs with short or extremely reduced appendages. Due to the polarizing phenotype observe and taking in count that the short eggs analysed had normal appendages morphology we conclude that the morphological defects that we are observing are not related to dorsoventral patterning defects.

Previously in our lab, by the work of Tânia Ferreira, we showed that homozygous mutant eggs for *pkn* are shorter when compared to normal conditions and, in most cases, the dorsal appendages morphology was normal. Also, we show that the defects observer in the egg's morphology were due to nurse cell dumping defects that affect the normal egg growth [36]. Consequentially, we assume that the defects observed in specific germline depletion of d*Sfmbt* eggs are a result of nurse cell dumping defects and not because dorsoventral patterning defects caused by Grk miss localization.

We hypothesize that *Pho* and *dSfbmt* are important for the correct architecture and remodelling of the oocyte chromatin during prophase I. If this is correct than, *Pho* and *dSfbmt* should regulate SC assembly and/or disassembly during meiosis. To access this hypothesis, we took advantage of flies that carry an endogenous construct of a core component of the SC with a GFP tag. Normal Corolla protein expression co-localizes with the oocyte chromatin throughout oogenesis and its expression progressively decreases during prophase I-arrest, being detectable until stage 8 of *Drosophila* oocyte development. Upon specific germline depletion of *Pho* there are no significant changes regarding Corolla protein expression during oocyte development when compared to control

condition. With this observation we conclude that *Pho* apparently is not essential for localization of Corolla to the oocyte nucleus. However, upon specific germline depletion of *dSfmbt* we observe that, although the oocyte chromatin remains normal, there is an overload of Corolla protein early on oogenesis and aberrant Corolla filamentous structures that are observed throughout the prophase I arrested oocyte until stage 10 of oocyte development. As far as we are aware, this is the first time that these abnormal loading and morphological SC structures were described. Therefore, we show that *dSfbmt* is required for the correct morphology and normal levels of Corolla expression in the oocyte nucleus.

Collectively, in this chapter we show that *dSfmbt*, a core component of the Pho repressive complex, is required for female fertility and normal egg growth. Regarding the SC, we show that d*Sfmbt* is required for setting the correct levels of Corolla protein in *Drosophila* oocytes and for the correct morphology of Corolla expression. Our future work will be focused on testing if the germline depleted *dSfmbt* short eggs are driven from nurse dumping defects and distinguish if the observed phenotypes regarding Corolla aberrant filamentous structures are due to a specific effect of *dSfmbt* to the Corolla protein assembly or, in fact, there is defective assembly or disassembly the entire SC during the prophase I arrest oocyte, since these results are only focused on one core protein of the SC.

# Chapter III

Chapter III is an independent chapter. Here I was integrated in the doctoral work of Om Rathore and performed the experimental procedures shown in figure 1C, figure 1D, figure 4A, figure 4B, figure 5, figure S3. All other figures were performed by Om Rathore and other collaborators. The text was written by Rui Martinho.

# 7 NineTeen Complex-subunit Salsa is required for dorsal-ventral patterning and splicing of small first introns.

Om Singh Rathore[1], Ricardo Matos[1], Mariana Ascensão-Ferreira[2], Margarida N. Tiago[1], Pedro Prudêncio[1,2], Rui D. Silva[1], Raquel P. Andrade[1], Jean-Yves Roignant[3], Nuno L. Barbosa-Morais[2], Rui Gonçalo Martinho[1,2,4,*]

[1] Center for Biomedical Research (CBMR), Universidade do Algarve, Faro, Portugal.

[2] Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

[3] Institute of Molecular Biology (IMB), 55128, Mainz, Germany

[4]Department of Medical Sciences and Institute for Biomedicine (iBiMED), Universidade de Aveiro, Aveiro, Portugal

* Corresponding author: Rui Gonçalo Martinho (rgmartinho@ua.pt)

# 8 Abstract

The NineTeen Complex (NTC), also known as Pre-mRNA-processing factor 19 (Prp19) complex, regulates distinct spliceosome conformational changes necessary for splicing. Splicing is particularly sensitive to mutations in NTC-subunit Fandango during *Drosophila* midblastula transition (MBT), suggesting differential requirements of NTC during development. Accordingly, we show that NTC-subunit Salsa, the *Drosophila* orthologue of human RNA helicase Aquarius, is particularly rate-limiting for splicing of a subset of small first introns during oogenesis, including the first introns of *gurken* and *transformer 2* transcripts. Germ line depletion of Salsa reduces female fertility, being oocyte dorsal-ventral patterning significantly impaired due to an abnormal expression of Gurken. Previously, it was shown that *Drosophila* first introns are more likely to be co-transcriptionally-retained than internal and terminal introns. We propose that Salsa RNA helicase is particularly rate-limiting for splicing of small first introns with weak 3´splice site, potentially minimizing the interference effect of transcriptional initiation factors binding to U1 snRNP.

Keywords: Drosophila / Splicing / NineTeen Complex / Dorsal-ventral patterning / Gurken / Tra2

# 9 Introduction

During *Drosophila* oogenesis, *gurken* mRNA localizes to the posterior cortex of the developing oocyte and Gurken signal is restricted to the underlying posterior follicle cells [65–68]. In response to a signal from the posterior follicle cells, there is a major reorganization of the cytoskeleton and a microtubule-dependent migration of the oocyte nucleus to the anterior cortex [65,68–70]. The anterior localized nucleus defines the dorsal-anterior region and provides the first detectable dorsal-ventral (D/V) asymmetry of the oocyte, being expression of both *gurken* mRNA and protein restricted to the cytoplasmic perinuclear region of the oocyte [66,68].

D/V patterning of the developing *Drosophila* egg is dependent on the dorsal-anterior localization of Gurken during mid-oogenesis [66,68,71]. Gurken is the ligand for the Torpedo/EGF receptor that locates to the apical surface of follicle cells that surround the developing oocyte [65–68]. Activation of Torpedo modifies the cell fate of the dorsal follicle cells and restricts the formation of Spätzle ligand to the ventral region of the [72,73], which is essential for normal morphogenesis of the eggshell dorsal appendages.

*gurken* mRNA is transcribed in the supporting nurse cells [74] and actively transported to the dorsal-anterior region of the oocyte by a dynein-mediated [75,76] dorsal-anterior localization of *gurken* mRNA relies on multiple elements localized to the transcript 5´UTR, 3´UTR and open-reading frame [77–79] being this localization crucial for its efficient [66,68,80] to the absence/low levels of oo18-RNA binding protein (Orb), the *Drosophila* homolog of cytoplasmic polyadenylation element binding protein (CPEB), *gurken* mRNA is not fully polyadenylated in the nurse cells and its translation is repressed [80,81] Orb is mostly absent from the nurse cells, it is nevertheless present at high levels in the oocyte [80,82,83] facilitating efficient translation of *gurken* transcript [80]. Casein kinase 2 (CK2) phosphorylation of Orb positively regulates its function [84], potentially facilitating posterior and, subsequently, dorsal-anterior translation of *gurken* mRNA [85].

The spliceosomal NineTeen Complex (NTC), also known as Pre-mRNA-processing factor 19 (Prp19) complex, regulates distinct spliceosome conformational changes necessary for efficient pre-mRNA splicing [86,87] composition is dynamic and comprises a subset of conserved core subunits and many transiently associated ones [86]. NTC also has a significant role in the crosstalk between transcription, co-transcriptional processing of the nascent RNA, and DNA repair, as distinct NTC subunits have been reported to be important for transcriptional elongation and genomic [87–89]. Human NTC-subunits PRP19, XAB2 and CDC5L are important for transcriptional elongation, transcription-coupled DNA repair, and activation of the ATM-related (ATR)-dependent DNA damage checkpoint [90–93] RNA Polymerase II (RNA Pol II) also promotes splicing activation through the recruitment of NTC [94].

Splicing during *Drosophila* early embryonic development is particularly sensitive to mutations in NTC-subunit Fandango [95], suggesting differential requirements of NTC during development [88]. Our working hypothesis is that NTC-dependent plasticity of the spliceosome is important for differential regulation of gene expression. To test this possibility, we decided to investigate the role of other NTC-subunits during *Drosophila* oogenesis and early embryonic development. We focused our initial work on gene *CG31368*, which encodes the ortholog of human Aquarius [96,97]. Since there is already a non-related *Drosophila* protease named Aquarius, we decided to rename CG31368 as Salsa.

Human Aquarius (AQR) (also known as intron-binding protein 160, IBP160) is an ATP-dependent RNA helicase that associates with NTC during spliceosome activation and formation of the activated B complex ($B_{ACT}$) [97,98] AQR binds to introns independently of sequence, but usually upstream of the branch-site (BS) and close to the associated U2 snRNP SF3a and SF3b proteins, being essential for intron-binding complex formation and efficient splicing [97,98]. Aquarius has also been suggested to be important for deposition of the exon junction complex (EJC) during the splicing reaction [99] and formation of intron-encoded snoRNAs [98], suggesting it regulates the cross-talk between splicing and other RNA processing events.

Here we found that female germ line depletion of Salsa, the *Drosophila* ortholog of AQR, is associated with a decrease in female fertility and significant D/V patterning defects of the eggshell. Salsa is required for efficient splicing of the first intron of *gurken* mRNA and normal dorsal-anterior expression of Gurken. Supporting the hypothesis that NTC-dependent plasticity of the spliceosome is likely to be important for differential regulation of gene expression, Salsa is particularly rate-limiting for splicing of a subset of small proximal introns, including the first intron of *sex-determining gene transformer 2* (*tra2*). We propose that Salsa potentially minimizes the interference effect of transcriptional initiation factors binding to U1 snRNP, by facilitating U2 snRNP spliceosome function.

# 10 Results

## 10.1 NTC-subunit Salsa is required for female fertility and eggshell dorsal-ventral patterning.

Our working hypothesis is that NineTeen Complex (NTC)-dependent plasticity of the spliceosome is important for differential regulation of gene expression. To test this possibility, we decided to investigate the role of distinct NTC-subunits during *Drosophila* oogenesis and early embryonic development. We focused our initial work on gene *CG31368*. Bidirectional blast analysis showed that *Drosophila* gene *CG31368* encodes the ortholog of human Aquarius (AQR) [96]. CG31368 physically interacts with NTC-subunits Fandango and Prp19 [95], and this interaction is independent of RNA (S1 Fig.). Since there is already a non-related *Drosophila* protease named Aquarius, and given its interaction with Fandango, we decided to rename *CG31368* as *salsa*.

To test the hypothesis that Salsa is rate-limiting during *Drosophila* oogenesis and early embryonic development, we depleted this RNA helicase specifically within the female germ line using *nanos*-Gal4 (*nos*-Gal4) germ line driver [100] and the upstream activating sequence (*UAS*)/Gal4 system [101–103] To confirm that the observed phenotypes were due to Salsa depletion, and not a consequence of an off-target effect of the used *salsa* RNAi construct, three non-overlapping short hairpin RNAs (shRNAs) were designed against *salsa* (*salsa* RNAi-1, *salsa* RNAi-2, and *salsa* RNAi-3). Depletion of Salsa within the female germ line resulted in a strong reduction of *salsa* mRNA levels in the ovaries and unfertilized eggs/embryos (Fig.1A, S2A Fig.). Depletion of *salsa* mRNA was stronger in unfertilized eggs/embryos (maternal contribution) than in the ovaries (S2A Fig.). This is most likely because Salsa is expressed in the soma (modEncode project; [104]) and the dissected ovaries contain supporting somatic cells not affected by the germ line-specific RNAi depletion. We next tested the effect of Salsa depletion in female fertility and observed that after being mated with wild-type males, there was a strong reduction in egg hatching (Fig.13B).

The reduced fertility of Salsa-depleted females was most likely related to the observed eggshell dorsal appendages defects (Fig.13C), as this is suggestive of dorsal-ventral (D/V) patterning defects [66,105]. For phenotypic quantification, we divided the observed eggshell dorsal appendages phenotypes in four different classes: class 0: wild-type dorsal appendages, class 1: dorsal appendages only fused at bottom, class 2: fused dorsal appendages (spindle phenotype), and class 3: short eggs without or with extremely short dorsal appendages (S3 Figure). Eggs laid by control females (mCherry RNAi) showed wild-type dorsal appendages, suggestive of normal D/V patterning of the oocyte, whereas a significant proportion of the eggs laid by females whose germ line was depleted for Salsa showed ventralized eggshells, with highly abnormal (fused or partially-fused) dorsal appendages (Fig.13C, D) (for more experimental detail see material and methods). Altogether these results show that Salsa is required within the germ line for female fertility, possibly because it is important for D/V patterning of the developing oocyte.

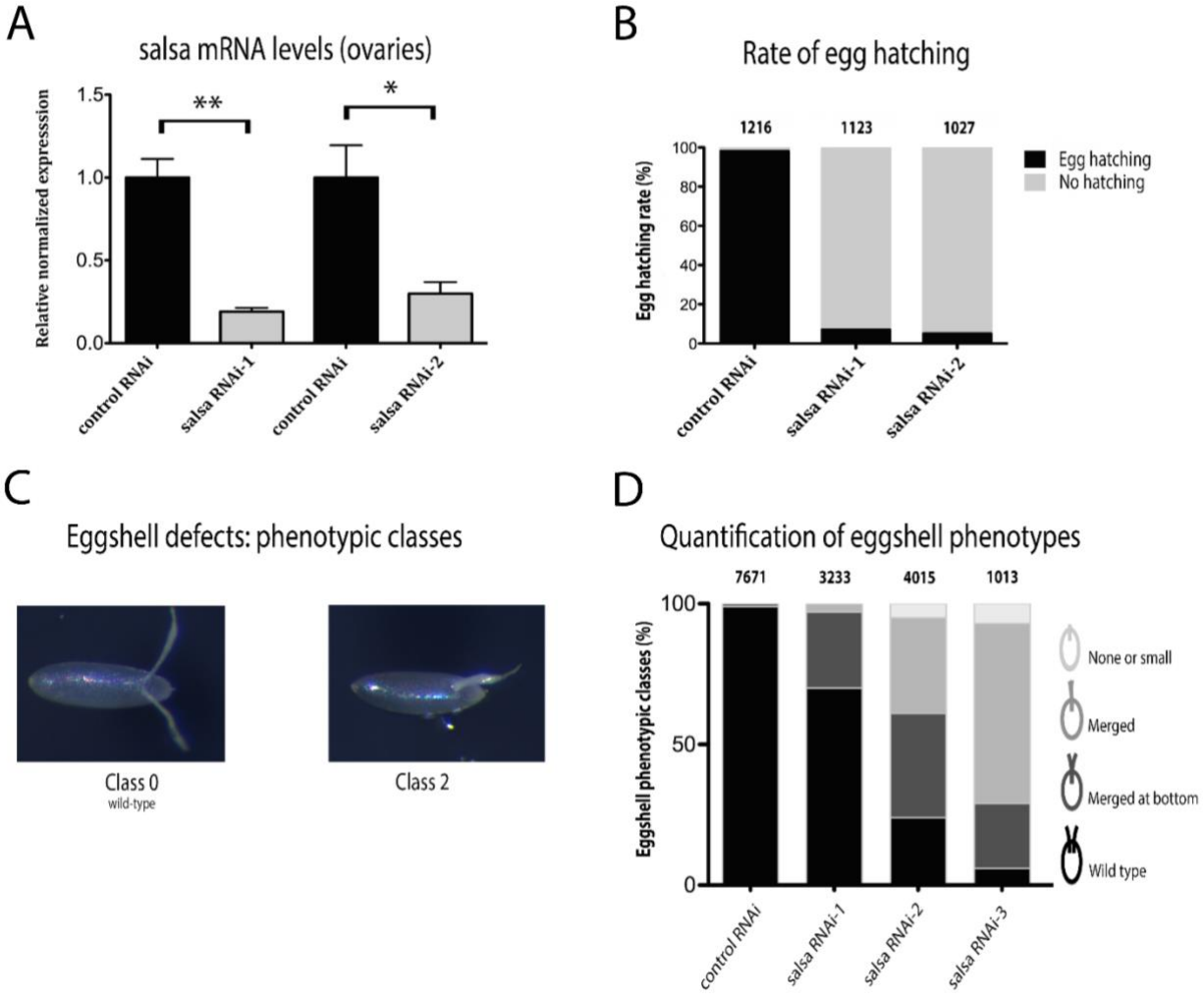


***Figure 14 - Drosophila Salsa is required for female fertility and eggshell dorsal-ventral patterning.*** *(A) Efficient depletion of salsa mRNA after germ line-specific RNAi (nanos-Gal4 and UAS-salsa RNAi). RT-qPCR analysis of salsa mRNA in control (mCherry RNAi) and Salsa-depleted ovaries (two non-overlapping shRNAs: 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 negative control (mCherry RNAi). At least three biological replicates were used for all shown results. Error bars indicate standard deviation. (B) Female fertility (egg hatching) was significantly reduced after germ line-specific depletion of Salsa. Virgin females were crossed with wild type males, and female fertility was calculated by the frequency of egg hatching 48 hours post oviposition. Number of eggs scored for each experiment are indicated above the bar plot. (C, D) Germ line-specific depletion of Salsa (three non-overlapping shRNAs: salsa RNAi-1, salsa RNAi-2, and salsa RNAi-3) impaired eggshell dorsal-ventral patterning. (C) A significant proportion of the eggs laid by females whose germ line was depleted for Salsa showed ventralized eggshells, with highly abnormal (fused or partially-fused) dorsal appendages (D) Quantification of eggshell dorsal appendages defects (eggshell ventralization). Observed phenotypes were categorized in four different phenotypic classes based on the eggshell dorsal appendages: class 0: wild type dorsal appendages - two individualized dorsal appendages); class 1: dorsal appendages only fused at bottom; class 2: fused dorsal appendages - spindle phenotype; and class 3: short eggs without or with extremely short dorsal appendages. Number of eggs scored for each experiment are indicated above the bar plot. Examples of scored phenotypes are shown in S3 Figure.*

## 10.2 Salsa is particularly rate limiting for splicing of the first intron of *gurken* mRNA.

The *gurken* gene encodes a single transcript with three introns [66], with the first intron localized to the gene 5´UTR, and the second and third introns localized to the gene open-reading frame (Fig.14A). Since the observed Salsa-depleted eggshell phenotypes were reminiscent of hypomorphic alleles of *gurken* [67] and Salsa is a highly conserved subunit of NTC [86,95,96] we hypothesized that Salsa was required for efficient splicing of *gurken* mRNA. To test this

hypothesis, we performed reverse transcription PCR (RT-PCR) using the exonic flanking primers for each intron of *gurken* transcript (open arrows shown in Fig.14A). Retention of the first, but not of the second and third introns, could be easily detected after germ line depletion of Salsa (Fig.14B). Subsequently, we performed quantitative RT-PCR (RT-qPCR) analysis of *gurken* transcript to quantify the relative levels of intron retention. 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 Fig.14A) (for used primer sequence see S1 Table). Consistent with our previous RT-PCR results, after depletion of Salsa there was a significant retention of the first intron, but not of the second and third introns, of *gurken* mRNA (Fig.14C). Altogether these results show that Salsa is particularly rate limiting for efficient splicing of the first intron of *gurken* mRNA.

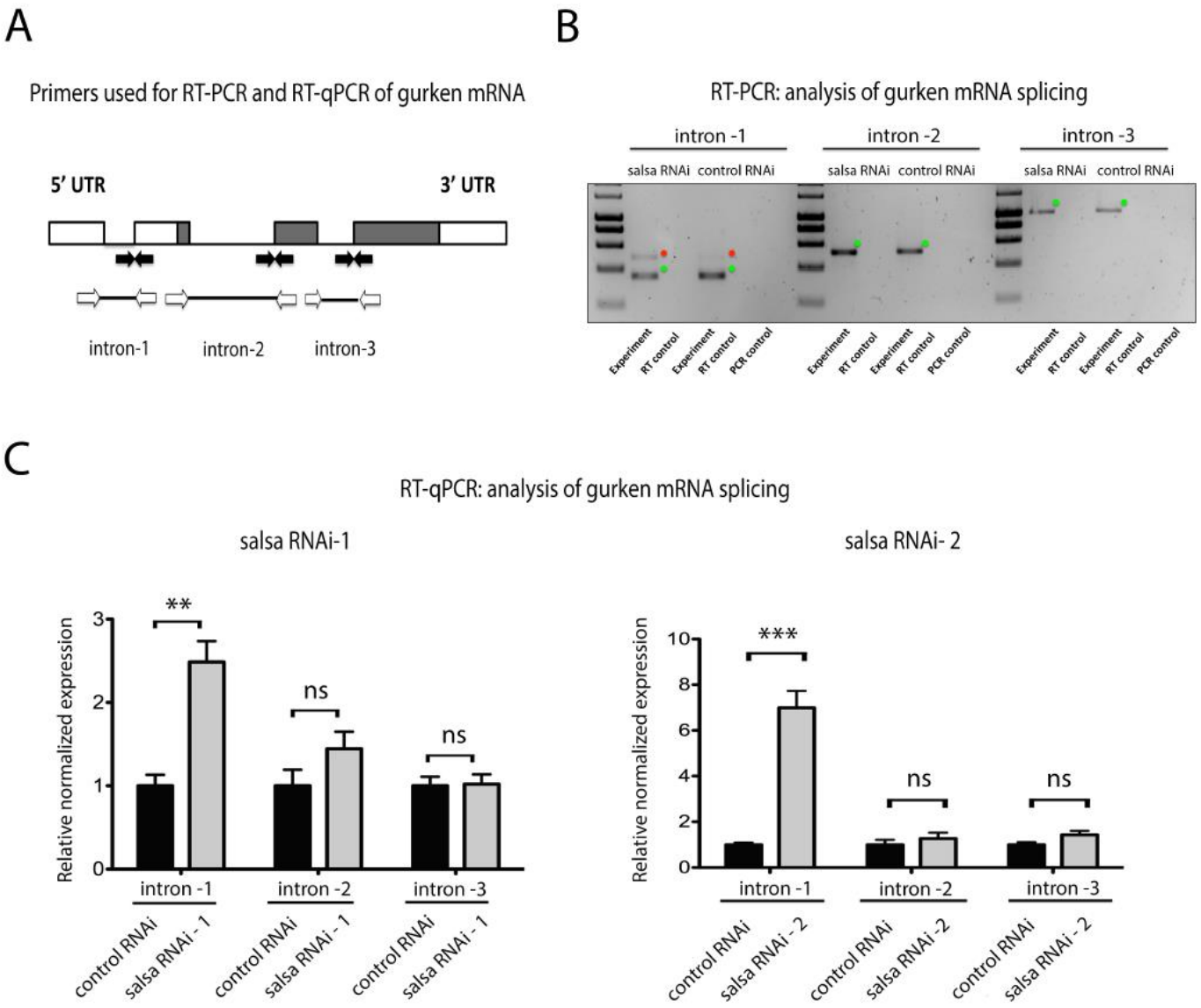


***Figure 15 - Salsa is particularly rate limiting for splicing of the first intron of gurken mRNA****. (A) The gurken gene encodes a single transcript with three introns. 5´ and 3´UTRs are shown in white, whereas the open-reading frame is shown in grey. Introns are represented as a thin black line. White and black arrows indicate, respectively, primers used for RT-PCR and RT-qPCR of gurken transcript. (B, C) Germ line-specific depletion of Salsa is associated with a significant increase in the retention of the first intron, but not of the second and third introns, of gurken transcript. (B) RT-PCR analysis of gurken mRNA using exonic flanking primers for each intron of gurken transcript and an iScript cDNA library (random hexamer and oligo (dT) reaction mix). Control ovaries yielded PCR products with the size predicted for the 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. Salsa-depleted ovaries showed an increase retention of the first intron of gurken mRNA (red dot, grk: 505bp), whereas the second and third introns were correctly spliced (green dots, grk2: 528 bp; grk3: 882 bp). PCR and RT controls 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) RT-qPCR analysis of gurken transcript using intron-exon primers and an iScript cDNA library (random hexamer and oligo (dT) reaction mix). Germ line-specific depletion of Salsa induced a significant retention of the first intron of gurken transcript, whereas the second and third introns were correctly spliced. Shown are the fold changes of intron retention. Relative normalized expression corresponds to values normalized with two distinct reference genes (β-actin and GAPDH) and relative to control conditions (mCherry RNAi). At least three biological replicates were used for all datasets. Error bars indicate standard deviation.*

## 10.3 Salsa is required for dorsal-anterior localization of *gurken* mRNA.

The correct dorsal-anterior localization of *gurken* mRNA relies on multiple elements localized to the transcript 5´UTR, 3´UTR and open-reading frame [77–79]. Since the 5´UTR of *gurken* mRNA is necessary for its correct dorsal-anterior [77,78] and Salsa depletion was associated with retention of the 5´UTR-localized first intron, we hypothesized that Salsa was required for dorsal-anterior localization of *gurken* mRNA.

To test this hypothesis, we performed fluorescence *in situ* hybridization with antisense RNA probes against *gurken* mRNA. Consistent with the observed eggshell dorsal appendages defects, depletion of Salsa was associated with a reduction of the dorsal-anterior localization of *gurken* mRNA (Fig.15A). For phenotypic quantification, we divided the observed *gurken* mRNA localization in three different classes: i) normal localization, ii) partial localization, and iii) no localization. Examples of scored phenotypes are shown in S4 Figure (for more experimental detail see material and methods). Egg chambers from control females (mCherry RNAi) showed a normal dorsal-anterior localization of *gurken* mRNA during mid-oogenesis, whereas a significant proportion of egg chambers from females whose germ line were depleted for Salsa showed a detectable reduction of dorsal-anterior localized *gurken* mRNA (Fig.15A, B). Interestingly, the reduction of localized *gurken* transcript was proportional to the severity of eggshell ventralization and *gurken* mRNA splicing defects (compare *salsa* RNAi-1 and RNAi-2 in Fig.13D, Fig.14C, and Fig.15A, B).

Since abnormally processed mRNAs are usually targeted for degradation [106], we next tested if the steady-state stability of *gurken* mRNA was reduced after salsa depletion, as this could easily explain the reduced levels of localized *gurken* transcript. We failed to detect any significant reduction in the levels of *gurken* mRNA in the ovaries (Fig.15C) and in unfertilized eggs/embryos (S2B Fig.) after depletion of Salsa. Furthermore, and consistent with the hypothesis that Salsa depletion did not elicit a significant degradation of *gurken* transcript and did not inhibit nuclear export from the nurse cells, significant levels

of intron retention could be detected within the maternally loaded embryonic *gurken* mRNAs (S2C Fig.).

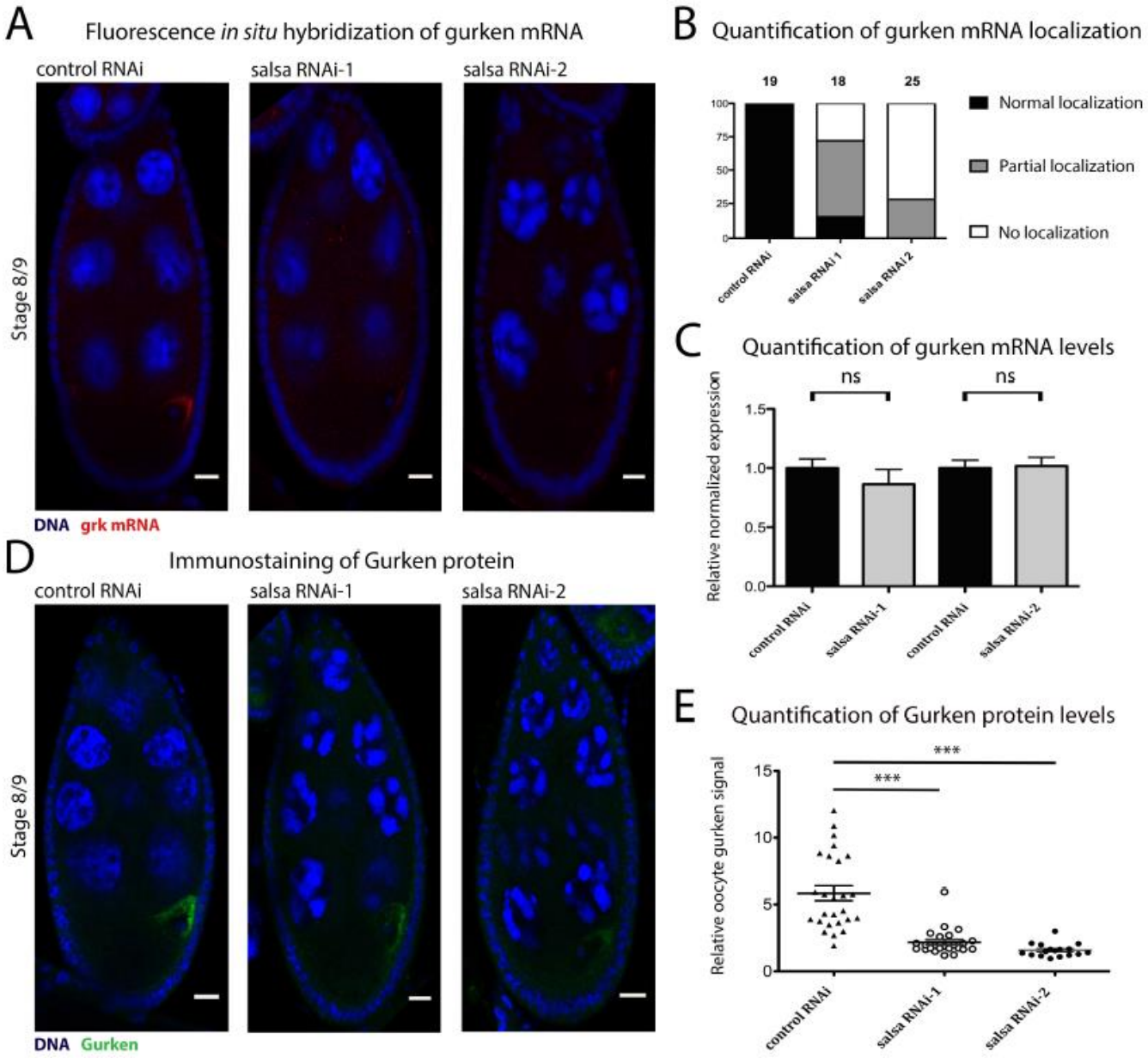



***Figure 16 - Salsa is required for dorsal-anterior expression of Gurken.*** *(A, B) Salsa is required for oocyte dorsal-anterior localization of gurken mRNA. (A) Germ-line specific of depletion of Salsa impaired dorsal-anterior localization of gurken mRNA during mid-oogenesis. Fragmented digoxigenin-labeled antisense RNA probes were used to detect gurken mRNA in situ. Visualization of probes was done using an anti-Digoxigenin Cyo3 secondary antibody. DNA (Blue) and gurken mRNA (red). Scale bar 10µm. (B) Quantification of anterior dorsal localization defects of gurken mRNA in stage 8/9 egg chambers using three phenotypic classes: "normal localization", "partial localization" and "no*

*localization". Additional examples of scored phenotypes are shown in S4 Figure. Negative control (mCherry RNAi): 100% "normal DV localization" (n=19); Salsa depletion (salsa RNAi-1): 16% "normal DV localization", 56% "partial DV localization", and 28% "no DV localization" (n=18); Salsa depletion (salsa RNAi-2): 0% "normal DV localization", 28% "partial DV localization", and 72% shows "no DV localization" (n=25). (C) Salsa is not required for stability of gurken mRNA. Real-time qPCR analysis detected no significant reduction of total levels gurken mRNA after depletion of Salsa. Relative normalized expression corresponds to values normalized with two distinct reference genes (β-actin and GAPDH) and relative to negative control (mCherry RNAi). At least three biological replicates were used for all datasets. Error bars indicate standard deviation. (D, E) Salsa is required for oocyte dorsal-anterior localization of Gurken. (D) Immunostaining of Gurken during mid-oogenesis (stage 8/9 egg chambers). Gurken was detected with an anti-Gurken monoclonal antibody and DNA was visualized with DAPI staining. DNA (Blue) and Gurken (green). Scale bar 10μm. (E) Relative oocyte Gurken dorsal anterior signal (arbitrary units (a.u.)) corresponds to Gurken signal pixel intensity (average of three different measurements taken from the oocyte dorsal-anterior region with the highest perinuclear signal) divided by the respective background signal (average of three different measurements from the respective nurse cells cytoplasm). 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).*

## 10.4 Salsa is required for dorsal-anterior localization of Gurken protein.

Since dorsal-anterior localization of *gurken* mRNA was reduced after depletion of Salsa (Fig.15A, B), we hypothesized that a similar reduction was likely to occur for Gurken protein. To test this hypothesis, we did immunofluorescence with an anti-Gurken antibody [107] in *Drosophila* egg chambers. Gurken accumulated at high levels in the cytoplasmic perinuclear

dorsal-anterior region of control oocytes (mCherry RNAi) (Fig. 15D), whereas such expression was significantly reduced after depletion of Salsa (Fig.15D, E) (see material and methods for Gurken levels quantification). We failed to detect the total levels of Gurken protein in the ovaries by western blot using currently available antibodies. Altogether, our results show that Salsa is required for normal dorsal-anterior expression of Gurken, possibly because it is rate limiting for efficient splicing of the first intron of *gurken* transcript.

## 10.5 The first intron of *gurken* is not required for Gurken expression and function.

The first intron of *oskar* is required for posterior localization of *oskar* mRNA, as its splicing recruits the exon junction complex (EJC) and generates a short RNA stem-loop structure both important for kinesin 1-dependent transport to the posterior [108,109]. In order to investigate if the first intron of *gurken* is similarly important for *gurken* mRNA localization, we decided to test if a genomic construct of *gurken* with its own endogenous promoter but without the first intron could complement a loss-of-function allele of *gurken* ($grk_{deltaFRT}$) [110]. Females homozygous for $grk_{deltaFRT}$ are viable but sterile, with eggs lacking detectable *gurken* mRNA [110]. Egg chambers and eggs produced by $grk_{deltaFRT}$ show severe anteroposterior and dorsoventral patterning defects and egg laying is severely compromised [110]. Genomic constructs of *gurken*, with its own endogenous *gurken* promoter and integrated in the same attP-site, but with or without the first intron (respectively, $grk_{WT}$ and $grk_{no\ 1st\ intron}$), showed identical expression levels (S5 Fig.). Since both constructs could similarly rescue the fertility (Fig.16A), egg laying, eggshell dorsal appendages (Fig.16B), and Gurken dorsal-anterior expression (Fig.16C) of *gurken* mutant females, we concluded that, contrary to *oskar*, the first intron of *gurken* is not required for expression of Gurken during oogenesis.

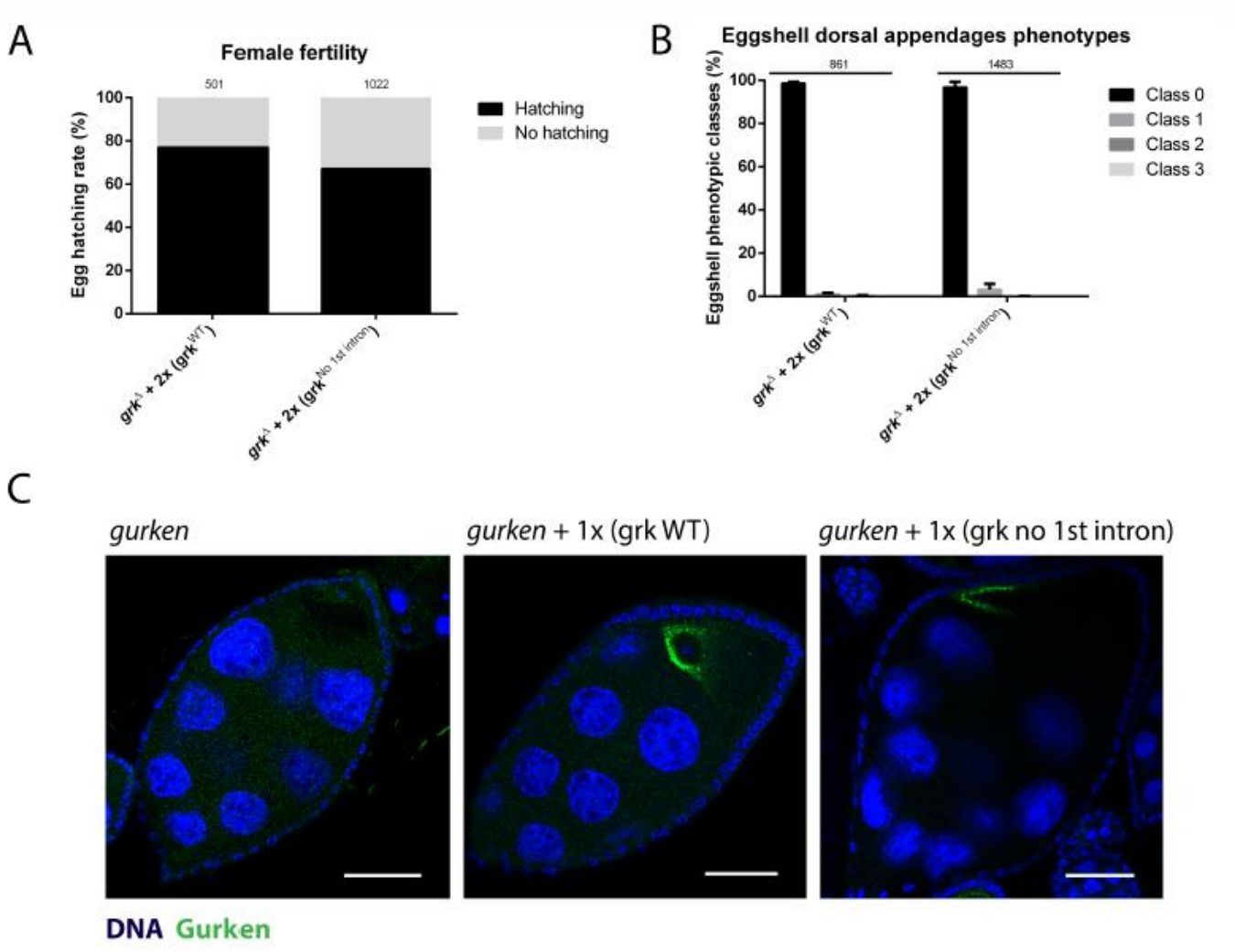



***Figure 17 - The first intron of gurken is not required for Gurken expression and function.*** *Females homozygous for grk$_{deltaFRT}$ are viable but sterile, being egg laying severely compromised* [110]*. (A) Expression of a genomic construct of gurken with its own endogenous promoter (two copies), with (grk$_{WT}$) or without the transcript first intron (grk$_{no\ 1st\ intron}$), rescued the egg laying and fertility (egg hatching) of females mutant for grk$_{deltaFRT}$ (grk$_{\Delta}$). Number of eggs scored for each experiment are indicated above the bar plot. (B) Expression of a genomic construct of gurken (two copies), with (grk$_{WT}$) or without the transcript first intron (grk$_{no\ 1st\ intron}$), rescued the eggshell dorsal appendages defects (eggshell ventralization) of eggs laid by females mutant for grk$_{deltaFRT}$ (grk$_{\Delta}$). For quantification of eggshell dorsal appendages defects (eggshell ventralization), observed phenotypes were categorized in four different phenotypic classes based on the eggshell dorsal appendages: class 0: wild type dorsal appendages - two individualized dorsal appendages); class 1: dorsal appendages only fused*

*at bottom; class 2: fused dorsal appendages - spindle phenotype; and class 3: short eggs without or with extremely short dorsal appendages. Number of eggs scored for each experiment are indicated above the bar plot. Examples of scored phenotypes are shown in S3 Figure. (C) Eggs from females mutant for gurken ($grk_{deltaFRT}$ / $grk_{HF}$) lack detectable Gurken. Expression of a genomic construct of gurken (one copy), with ($grk_{WT}$) or without the transcript first intron ($grk_{no\ 1st\ intron}$), rescued the dorsal anterior expression of Gurken. Gurken was detected with an anti-Gurken monoclonal antibody and DNA was visualized with DAPI staining. DNA (Blue) and Gurken (green). Scale bar 30μm.*

## 10.6 Increased levels of *gurken* expression rescue the D/V patterning defects observed after depletion of Salsa.

Our results show that Salsa is particularly rate-limiting for efficient splicing of the first intron of *gurken* mRNA and its depletion is associated with significant defects in Gurken dorsal-anterior expression. Since the 5´UTR of *gurken* mRNA is required for dorsal-anterior expression [77,78] and retention of the first intron of *gurken* is likely to severely affect the structure of its 5´UTR, we hypothesized that Salsa is required for D/V patterning because it regulates efficient splicing of the first intron of *gurken* transcript. Consistently, expression of a genomic construct of *gurken* lacking its first intron ($grk_{no\ 1st\ intron}$) efficiently suppressed the ventralized eggshell phenotypes observed after depletion of Salsa (compare *salsa* RNAi with *salsa* RNAi + 1x ($grk_{no\ 1st\ intron}$) in Fig.17A). Nonetheless, since expression of a wild-type genomic construct of *gurken* ($grk_{WT}$) also suppressed eggshell ventralization compare *salsa* RNAi with *salsa* RNAi + 1x ($grk_{WT}$) in Fig.17A) then it is still possible that Salsa regulates Gurken dorsal-anterior expression by additional mechanisms beyond splicing of *gurken* transcript.

## 10.7 Salsa is required for eggshell dorsalization after increased *gurken* dosage

Expression of a genomic construct containing wild-type *gurken* ($grk_{WT}$), in an otherwise wild-type genetic background, induced eggshell dorsalization

(Fig.17B; dorsalization corresponds do class -2 and -1). This is most likely due to an increase in *gurken* copy number (3 copies of *gurken*: two endogenous and one corresponding to the genomic transgene), as D/V patterning is sensitive to *gurken* dosage [105]. Interestingly, depletion of Salsa partially suppressed this phenotype (compare wild-type genomic construct of gurken (*grk*$_{WT}$) in Fig.17A and B), further confirming the requirement of Salsa for dorsal-anterior expression of Gurken during oogenesis.

Contrary to wild-type *gurken*, expression of *gurken* lacking its first intron did not induce eggshell dorsalization defects (compare *grk*$_{WT}$ with *grk*$_{no\ 1st\ intron}$ in Fig.17B). Since *gurken* mRNA expression levels were equivalent in both constructs (S5 Fig.), and regardless of the fact that *gurken* first intron is not required for *gurken* expression and function in an otherwise wild type background (Fig.16), this suggests that splicing *per se* might nevertheless facilitate, similarly to other 5´UTR localized introns [111], expression of Gurken protein.

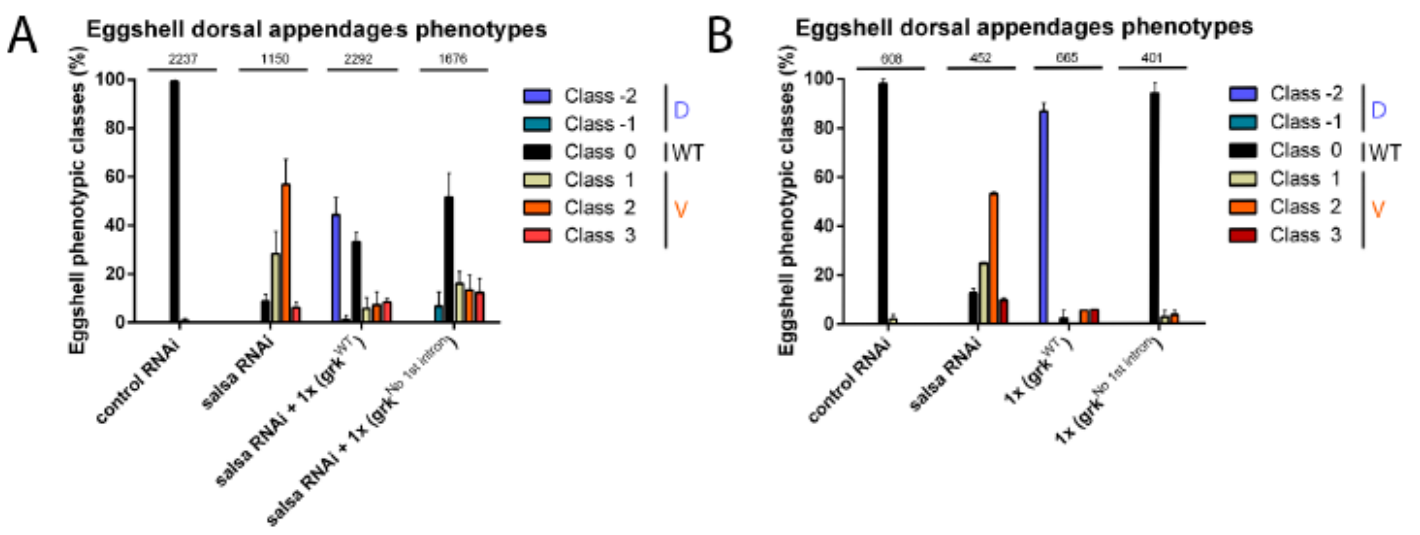



***Figure 18 - Increased levels of gurken rescues Salsa depletion eggshell D/V patterning defects.*** *(A) Eggs laid by control females (mCherry RNAi) showed wild type dorsal appendages (WT), whereas a significant proportion of the eggs laid by females whose germ line was specifically depleted for Salsa (nanos-Gal4 and salsa RNAi-2) showed eggshell ventralization (V), with highly abnormal (fused or partially-fused) dorsal appendages (classes 1 and 2). Expression of a genomic construct of gurken, with (grkWT) or without the first intron of gurken mRNA (grkno 1st intron), suppressed the eggshell ventralization defects of eggs*

*laid by females whose germ line depleted for Salsa. (salsa RNAi-2). (B) Expression of a genomic construct of gurken (grkWT) in an otherwise wild-type genetic background (1x grkWT; 3 copies of gurken: two endogenous and one corresponding to the genomic transgene) induce a significant dorsalization of the eggshells (D) (classes -2 and -1). Eggshell dorsalization was not observed after expression of a genomic construct of gurken without the first intron of gurken mRNA (1x grkno 1st intron). (A, B) Salsa is required for eggshell dorsalization (D) after increased gene copy number of gurken. For quantification of eggshell dorsal appendages defects, the observed phenotypes were categorized in six different phenotypic classes based on the eggshell dorsal appendages. Class 0 corresponds to wild type eggshell dorsal appendages (WT). Eggshell ventralization (V) of the eggshells corresponds to classes 1 to 3. Eggshell dorsalization (D) corresponds to classes -1 and -2. Class 0 (wild type dorsal appendages): two individualized dorsal appendages); class 1: dorsal appendages only fused at bottom; class 2: fused dorsal appendages - spindle phenotype; and class 3: short eggs without or with extremely short dorsal appendages. Class -1: corresponds to short eggs with dorsalized appendages; class -2: corresponds to a broad and thick crown of appendage material or a broad/fused appendage that covers the entire width of the eggshell. Examples of scored phenotypes are shown in S3 Figure*

## 10.8 Depletion of Salsa does not impair anterior cortical migration and normal morphology of the oocyte nucleus.

Caenorhabditis elegans EMB-4, which is the ortholog of human Aquarius and *Drosophila* Salsa, has recently been reported as playing an important role in the Piwi-interacting RNAs (piRNAs) pathway for efficient silencing of transposable elements [112]. *Drosophila* mutants for piRNAs biogenesis are similarly associated with reactivation of transposable elements and the accumulation of DNA [113,114] Accumulation of persistent DNA damage during oogenesis leads to the activation of the meiotic checkpoint that severely disrupts oocyte nucleus [114,115] and embryonic anterior-posterior and dorsal-ventral

axis specification by inhibiting the function of Vasa [114,115], which is a DEAD box RNA helicase required for efficient translation of *gurken* [116].

Although depletion of Salsa impaired the dorsal-anterior expression of *gurken* and D/V patterning of the eggshell was severely compromised, the microtubule-dependent migration of the posterior oocyte nucleus to the anterior cortex was mostly unaffected (Fig.15 and S4 Fig.). Furthermore, and contrary to the highly abnormal oocyte nucleus observed after the accumulation of persistent DNA damage [115], we failed to detect any oocyte nuclear abnormalities after Salsa depletion (S6 Fig.), clearly suggesting that the observed D/V patterning phenotypes were not mediated by persistent activation of meiotic checkpoint.

## 10.9 Salsa is required for nurse cells nuclear dispersal.

Heterogeneous nuclear RNA-binding protein (hnRNP) Squid (Sqd) interacts with Hrb27C, also an hnRNP, and Ovarian tumor (Otu) [117], and their respective mutants show mislocalized *gurken* mRNA, ectopic Gurken translation, and dorsalized eggshells [117,118] *half pint* (*hfp*) is the *Drosophila* ortholog of human PUF60, being its loss similarly associated with mislocalized *gurken* mRNA and dorsalized eggshell phenotypes [119].

During early oogenesis, the nurse cell chromosomes undergo several rounds of endoreplication that results in a characteristic polytene blob-like appearance [120]. This chromatin architecture disappears by stage 5/6 with the dispersal of the nurse cells polytene chromosomes. Besides the previously described D/V patterning defects, mutants for *sqd*, *hrb27C*, and *hfp* also show defects [117,119] suggesting that these proteins regulate multiple processes during oogenesis. Interestingly, whereas alternative splicing of *otu* was abnormal in *sqd* and *hfp* mutants, ectopic expression of the 104 kDa Otu isoform was sufficient to rescue the nurse cells nuclear dispersal phenotype of these mutants [117,119]

Although germ line depletion of salsa was associated to D/V patterning defects, there were however other phenotypes likely to contribute to the observed reduction of female fertility. For example, ovaries depleted for Salsa showed nurse cells nuclear dispersal defects (Fig.15D) identical to the ones previously

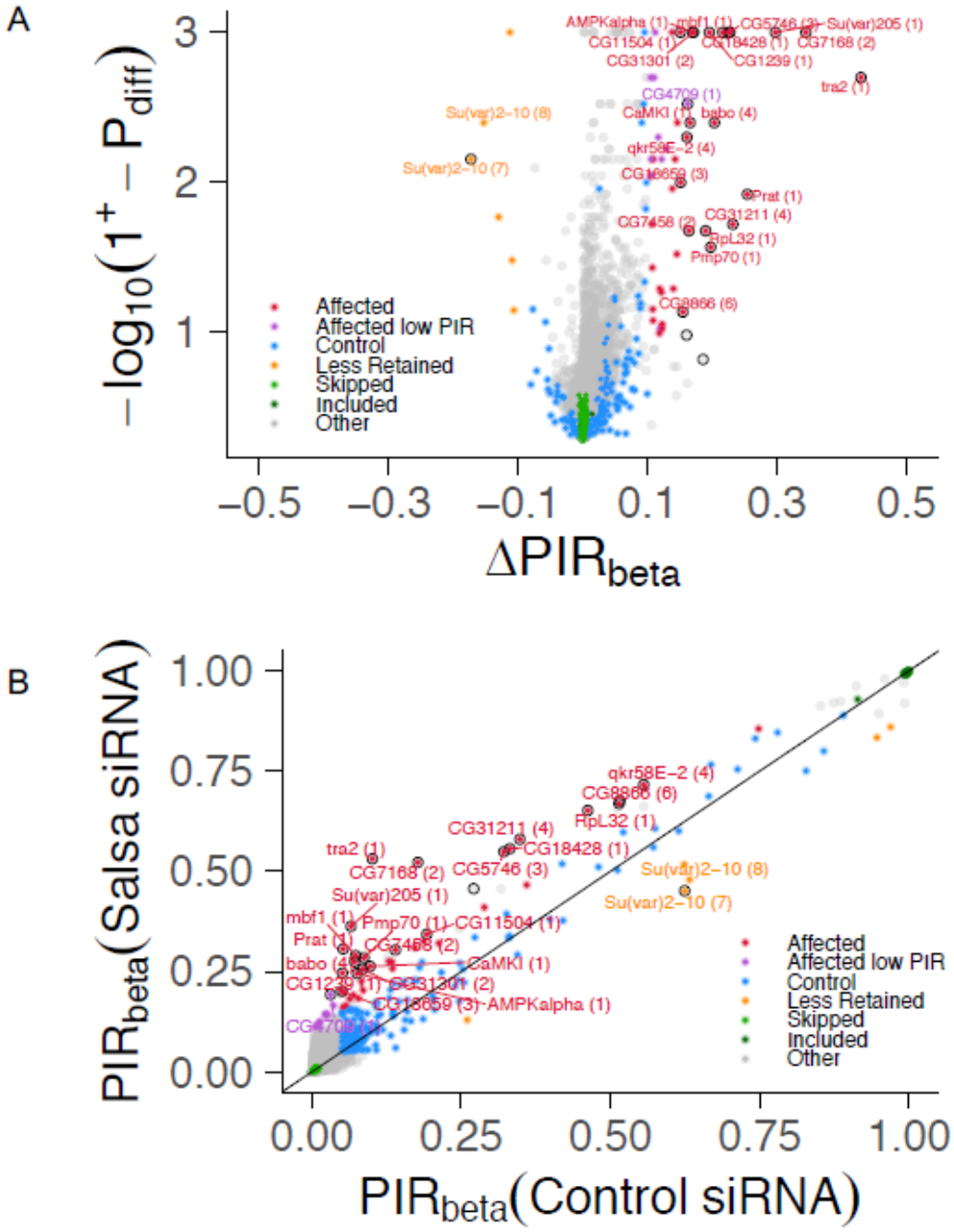
observed *in sqd* and *hfp* [117,119] Nonetheless, Salsa is most likely not functionally related with Sqd and Hfp, as Salsa depletion is associated with eggshell ventralization, and not dorsalization, there was no detectable physical interaction between Salsa and these two proteins, and Salsa depletion did not significantly impair alternative splicing of *otu* (S7 Fig.). Altogether our results suggest that during oogenesis Salsa regulates pre-mRNA splicing of other transcripts beyond *gurken*.

## 10.10 Salsa is particularly rate-limiting for splicing of a small subset of introns.

Our working hypothesis is that Salsa RNA helicase is particularly rate-limiting for efficient splicing of a small subset of introns. To identify the nature of such introns, and better understand the function that Salsa plays during development, the transcriptome from control and Salsa-depleted ovaries was sequenced. To minimize the identification of introns whose splicing was non-specifically affected by Salsa depletion, a *salsa* RNAi hairpin (*salsa* RNAi-1) with a mild D/V patterning defect (Fig. 13D) but without detectable egg chamber necrosis was used for this study. To better understand the function of Salsa in splicing plasticity, we focused our analysis on introns with good sequencing coverage and performed differential intron retention analysis between control and Salsa-depleted samples (for more experimental detail see material and methods).

Genome-wide analysis of the splicing defects observed after germ line depletion of Salsa showed that this RNA helicase is particularly rate-limiting for splicing of a small subset of introns, being the large majority of alternative splicing defects related to increased levels of intron retention (IR) upon Salsa depletion (S8 Fig.). Considering as biological relevant a difference in percent of intron retention (PIR) of at least 10% (see material and methods), we identified a significant enrichment for increased intron retention after Salsa depletion, with 47 introns increasingly retained and only 6 introns less retained in Salsa-depleted samples (red/purple and yellow points in Fig. 18A and 18B, respectively) (S3 Table). This proportion (47/53; 0.89) was significantly different from 0.5 (proportion's test p-value=1.78E-8). The fact that most identified cases were

related with increased levels of intron retention is consistent with the recently proposed function of human Aquarius before splicing catalysis for increased pre-mRNA splicing efficiency [97]. The few introns whose retention was significantly reduced after Salsa depletion (yellow points in Fig. 18A and 18B) (S3 Table) potentially resulted from an indirect effect on their pattern of alternative splicing.



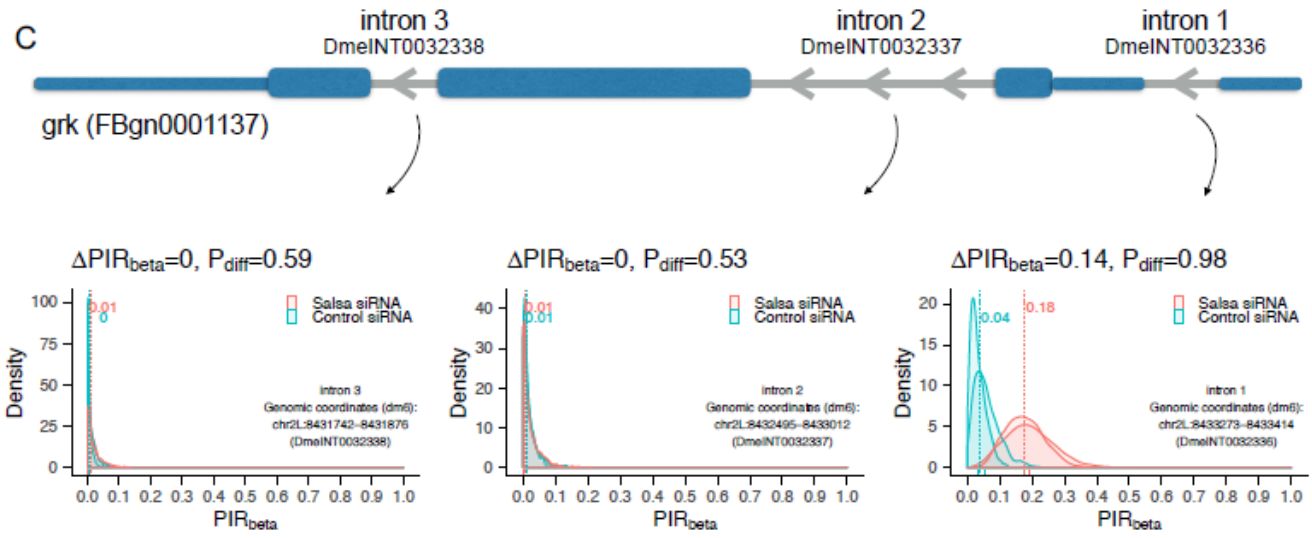



***Figure 19 - A small subset of introns showed increased levels of intron retention (IR) upon Salsa depletion.*** *(A) Volcano plot for beta distribution-based differential intron retention analysis, with differences in inclusion levels between Salsa-depleted (salsa RNAi-1) and control samples (mCherry RNAi) (**ΔPIRbeta** = PIRsalsa RNAi - PIRcontrol) in the x-axis and probability of differential retention (-log10(1+ - Pdiff)) in the y-axis, with 1+ = 1.001 used to avoid infinite values when Pdiff = 1. Differentially retained introns (Affected, Affected low PIR and Less Retained), identified by Pdiff ≥ 0.9, are highlighted in red, purple and yellow, respectively. Events with a **ΔPIRbeta** greater than 0.15 are labeled with gene name and the transcript's intron position. Control, Included and Skipped introns are also highlighted for comparison. (B) Scatterplot comparing PIRbeta values in Salsa-depleted and control samples, with intron groups highlighted/labeled as in (A). (C) Diagram of simplified gurken (grk) gene structure with exons (blue) and introns (gray) and density plots (smoothed histograms) reflecting points emitted from beta distributions used to model gurken intron 1 (right panel), 2 (middle) and 3 (left) retention for each sample and PIRbeta estimates for Salsa-depleted and control conditions (dashed vertical lines). Rug plots below the density curves reflect PIR values per sample. Event identifiers and genomic coordinates of the respective alternative sequences are part of the VAST-DB dm6 annotation [121].*

Fully supporting the *gurken* splicing defects shown in Fig.14, splicing of the first intron, but not of the second and third introns, of *gurken* was affected after Salsa depletion, with a significant PIR increase (Fig.18C). Analysis of the introns that were more retained after salsa depletion showed that this RNA helicase is particularly rate-limiting, among others, for efficient splicing of the first introns of the sex-determining gene *transformer 2* (*tra2*) (Fig.19), *suppressor of variegation 205* (*Su(var205)*), *phosphoribosylamidotransferase* (*Prat*), *CG18428* and *multiprotein bridging factor 1* (*mbf1*), the second 5´UTR-localized intron of *CG7168*, the fourth *CG31211* intron, the third *CG5746* intron and the fourth *baboon* (*babo*) intron (red points in Fig. 18A and 18B) (S3 Table).

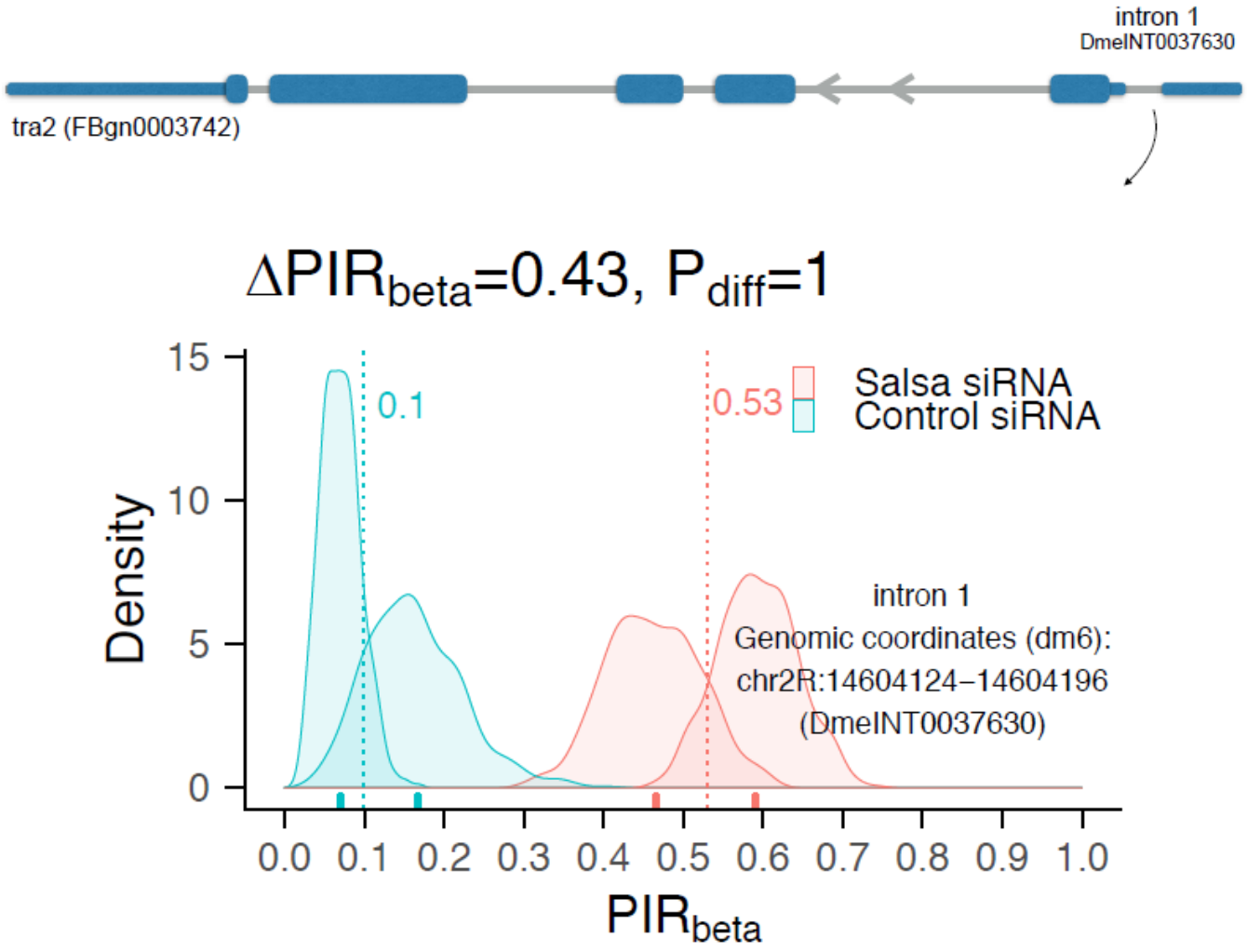



***Figure 20 - Salsa is particularly rate limiting for splicing of the first intron of tra2.*** *Increased retention of the first intron (DmelINT0037630) of sex-determining gene transformer 2 (tra2) upon Salsa depletion (salsa RNAi-1), when compared*

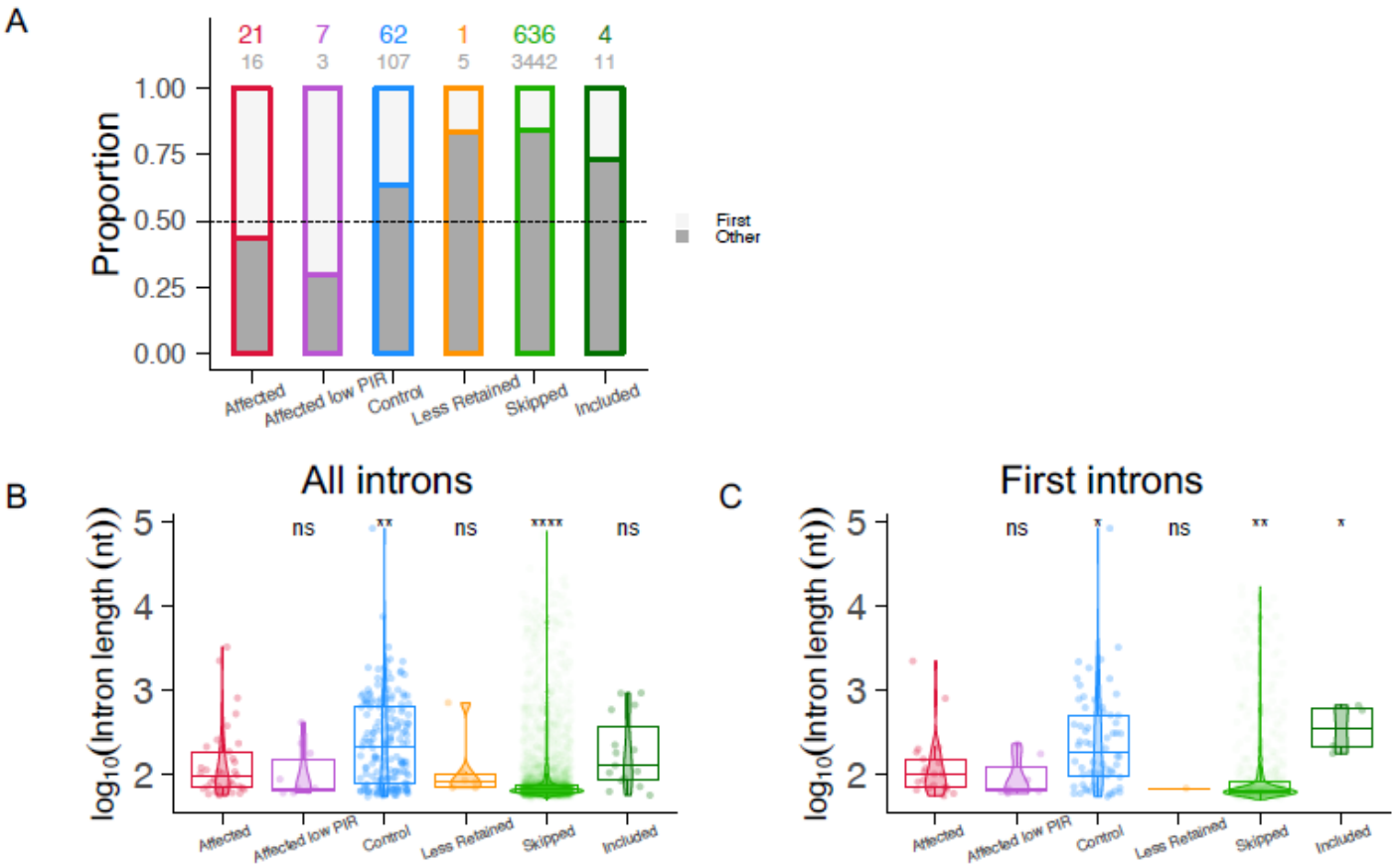
*to control conditions (mCherry RNAi). Diagram of simplified tra2 gene structure with exons (blue) and introns (gray) and density plots (smoothed histograms) reflecting points emitted from beta distributions used to model retention of the first intron of tra2 for each sample and PIRbeta estimates for Salsa-depleted and control conditions (dashed vertical lines). Rug plots below the density curves reflect PIR values per sample. Event identifier and genomic coordinates for the first intron of tra2 are part of the VAST-DB dm6 annotation [121].*

## 10.11 Salsa is particularly rate-limiting for splicing of small first introns

Since gene architecture influences splicing kinetics [122–124] we decided to investigate if there was any particular feature significantly associated to introns whose splicing is particularly sensitive to Salsa depletion. A comparison between differentially retained introns after Salsa depletion detected a significant bias for small proximal introns, as the affected introns had a higher probability of being first introns compared to control (proportion's test p-value=0.0011) (Fig.20A, S9 Fig.) or to commonly skipped introns (proportion's test p-value<2.2E-16) (Fig.8A, S9 Fig.), and were typically smaller than control (Wilcoxon's rank sum test with continuity correction p-value=0.0046) and skipped introns (Wilcoxon's p-value=1.85E-5) (Fig. 20B, C, S10A Fig.).

Recruitment of U1 to the proximal region of nascent pre-mRNAs is facilitated by interaction with the Cap-Binding Complex (CBC) [125–128] which protects against premature polyadenylation [129–131] and facilitates splicing of 5´end localized [132,133] The effect of CBC on splicing enhancement is distance dependent [134], with a significant positive correlation between transcription start site (TSS) distance and 5´splice site (5´ss) strength [135]. Consistent with the hypothesis that Salsa is particularly rate-limiting for CBC-dependent splicing, alternatively spliced introns with increased retention after Salsa depletion showed a bias for shorter distances between the transcription start site (TSS) and their 5´ss (average of 1060 base pairs) when compared to control introns (average of 3632 base pairs, Wilcoxon's p-value=0.0014) and typically skipped introns (6866

base pairs, Wilcoxon's p-value=3.42E-9) (Fig.20D, E). Nevertheless, since this distance bias was not significant amongst first introns themselves (Fig.20E, S10B Fig.), suggesting it is mostly related with a first intron position bias and not the distance *per se*, and since there was no significant bias for weaker 5´splice sites among the affected introns (S11A), it is unlikely that the observed biases are related with CBC-dependent splicing. Interestingly, affected introns showed a bias to weaker 3´splice sites when compared to both control (Wilcoxon's p-value=0.018) and typically skipped introns (Wilcoxon's p=3.82E-6) (Fig.20F, 20G, S10C Fig.), whereas there was no bias for higher GC content when compared to control introns (S11B, C). Altogether our data suggests that Salsa is particularly rate-limiting for efficient splicing of first introns independently of TSS distance, typically smaller and with weaker 3´slice sites.



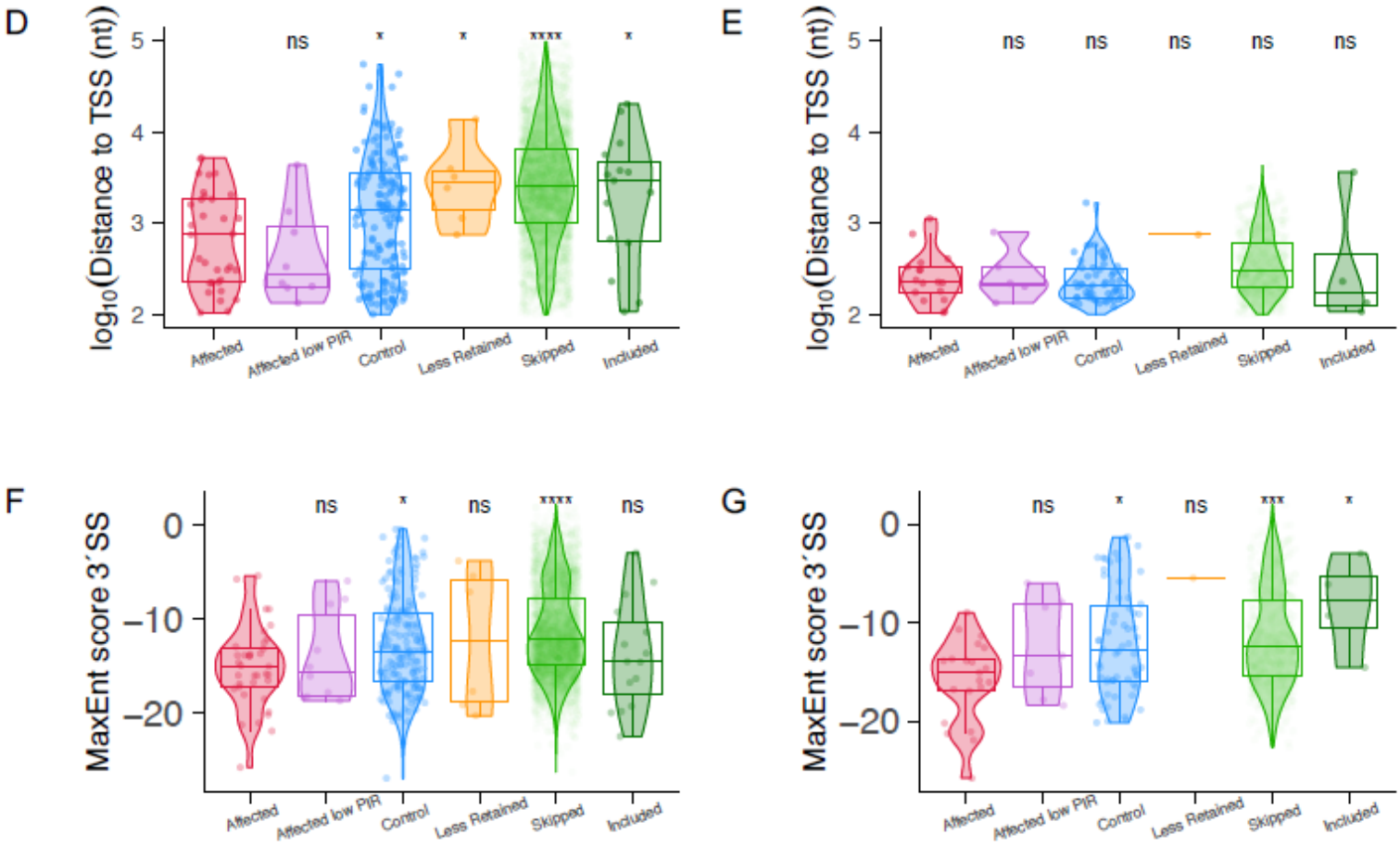



***Figure 21 - Salsa is particularly rate-limiting for splicing of small first introns.*** *(A) Proportion of first (white) and all other (gray) introns within each intron class, with the # first introns / # other introns indicated on top of each bar in color / gray, respectively. (B to G) Violin plots and boxplots (highlighting the median, 25th and 75th percentiles per intron class) summarizing distributions of observations (each point refers to one intron) considering all introns (left panels) and first introns only (right panels) of: (B-C) log10 of intron length (nt); (D-E) log10 of distance (nt) between intron start and the transcript transcription start site (TSS); (F-G) Maximum entropy (MaxEnt) scores* [136] *for estimation of 3´ splice site efficiency, obtained with MATT* [137]*. Comparison of metrics in (B to G) was performed between all intron classes and Affected: * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001 and ns non-significant (Wilcoxon rank-sum test with continuity correction).*

## 10.12 Salsa regulates the expression levels of a small subset of genes

Human NTC-subunit Xab2, the ortholog of *Drosophila* Fandango [95], has been reported to be rate-limiting, besides splicing, for transcription and transcription-coupled repair [92]. Since Salsa physically interacts with Fandango [95], it is possible that transcription was being similarly affected in Salsa-depleted ovaries. Differential expression analysis revealed that Salsa was particularly rate-limiting for the expression levels of only small subset of genes (Fig.21), as it positively regulated the expression levels of *CG4570* and *Activity-regulated cytoskeleton associated protein 1* (*Arc1*) (significantly downregulated in Salsa-depleted ovaries) (Fig.21, S4 Table), whereas negatively regulated the expression of *death resistor Adh domain containing target* (*Drat*), *Cyp4p2*, *sevenless* (*sev*) and *Insulin-like peptide 6* (*Ilp6*) (significantly upregulated upon Salsa depletion) (Fig.21, S4 Table). Since there were no significant alterations in alternative splicing of these genes (S12 Fig.), then these changes of gene expression are more likely the result of a minor role of Salsa in transcription or an indirect consequence of the splicing defects observed after Salsa depletion. Although the phenotypic relevance of these differences is unclear, it is nevertheless possible they give a contribution to the observed phenotypes. Expression of nonsense-mediated decay pathway genes showed no differences between conditions, suggesting the decay pathway is similarly efficient in control and Salsa-depleted samples (S13 Fig.). Altogether these results suggest that Salsa is not particularly rate-limiting for transcription and the observed splicing defects did not elicit a genome-wide change in the transcriptome steady-state stability. This is consistent with previously published results on *fandango* [95], which raises the possibility that splicing appears to be the main function of *Drosophila* NTC.

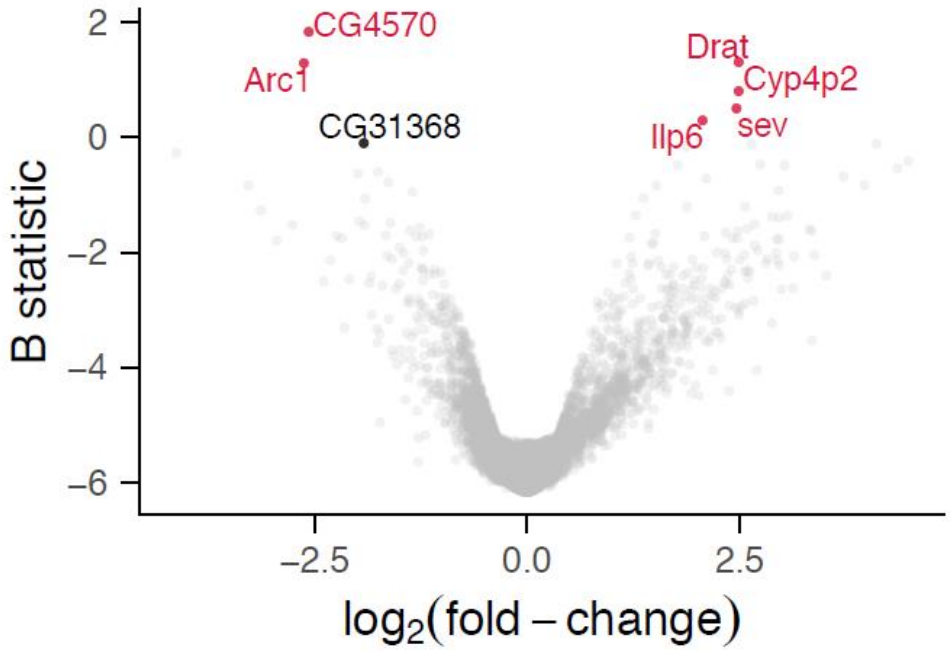



***Figure 22 - Salsa regulates the expression levels of a small subset of genes.*** *Volcano plot for differential gene expression analysis upon Salsa depletion (salsa RNAi-1), when compared to control conditions (mCherry RNAi). Each point indicates one gene considered in the analysis. Significantly differentially expressed genes (B statistic > 0) are highlighted in red, while salsa (CG31368) is highlighted in black as a positive control.*

# 11 Discussion

## 11.1 Salsa regulates expression and efficient splicing of *gurken*

Oocyte dorsal-anterior localization of *gurken* mRNA relies on multiple elements localized to the transcript 5´UTR, 3´UTR and open-reading [77–79] yet the relative importance of each element for mRNA localization is still unclear. The 5´ and 3´ UTRs of *gurken* were reported to be required for dorsal-anterior localization of *gurken* transcript [78]. Furthermore, and using a genomic *gurken* construct with a lacZ reporter inserted within the gene open-reading frame, it was shown that whereas *gurken* 5´UTR is required for transcript oocyte accumulation, its coding region and 3´UTR are necessary for its posterior and dorsal-anterior localization [77]. More recently it was nevertheless reported, using an oocyte injection assay, that a small stem loop located within the open-reading frame was necessary and sufficient for *gurken* transcript localization [79].

Our results suggest that efficient splicing of the first intron of *gurken* is required for mRNA dorsal-anterior localization and dorsal-ventral patterning. This is most likely because retention of the first intron impairs the secondary RNA structure of *gurken* 5´UTR, and the function of closely located RNA element important for its localization [78]. Furthermore, and although the first intron of *gurken* is not rate-limiting for transcript localization and function, its splicing is likely to facilitate Gurken protein expression as its presence enhanced the dorsalization phenotype associated with increased copy number of *gurken* gene [105]. Our results therefore fully support the role of *gurken* 5´UTR and strongly suggest that Salsa-dependent splicing is required for transcript localization and *gurken* function.

## 11.2 Salsa facilitates efficient splicing of small first introns

The precise function of human Aquarius in splicing is still poorly understood. This RNA helicase is recruited to the spliceosome as a pentameric complex known as intron-binding complex (IBC), which also contains hSyf1 (also known as Xab2), hIsy1, CypE and CCDC16 [97]. Co-immunoprecipitation experiments suggest a large interaction interface between IBC and U2 snRNP,

within the activated spliceosome ($B_{act}$ stage) and just before the first splicing reaction. Although Aquarius ability to bind and hydrolyze ATP is important for spliceosome activation and splicing efficiency, the role of its RNA unwinding activity is less clear.

Our work has identified a small subset of introns whose splicing is particularly sensitive to depletion of Salsa (the *Drosophila* ortholog of human Aquarius). The fact that splicing was only affected in a small number of introns is consistent with the observation that immunodepletion of human Aquarius from nuclear extracts only weakly impaired splicing *in vitro* [97]. This suggests that although this RNA helicase is apparently not critical for overall splicing, there is however a subset of introns whose efficient removal relies on the function of this enzyme.

Analysis of the introns whose splicing was particularly sensitive to Salsa depletion showed a clear bias for small first introns with weak 3´splice sites, independently of the distance between the transcription start site (TSS), 5´splice site strength and GC content. The bias for small introns suggests that Salsa is mostly rate-limiting when introns are recognized by intron definition [122], where the initial pairing between U1 and U2 snRNPs occur across the intron. Furthermore, the bias for introns with weak 3´splice sites is in accordance with the extensive interaction between IBC and U2 snRNP in the activated spliceosome and implies that depletion of Salsa is likely to impair, at least in a subset of introns, U2 snRNP function during splicing. The absence of any detectable bias for short distances between the TSS and 5´splice site, when evaluating affected and control first introns, or any bias for weak 5´splice site strength, suggests that Salsa is not likely to be particularly rate-limiting for Cap-Binding Complex-mediated splicing [134].

*Drosophila* first introns are more likely to be co-transcriptionally retained than internal and terminal introns [138]. This is not consistent with the kinetic competition model [139], where the fastest processes are the ones most likely to occur, suggesting additional constraints to first intron splicing. Although the precise nature of such constraints is still poorly understood, binding of transcriptional initiation factors to the 5´splice site associated [140,141] potentially restricts splicing efficiency, as it might impair the initial pairing between

U1 and U2 snRNPs. Our working hypothesis is that Salsa is particularly rate-limiting for splicing of small first introns with weak 3´splice site because this enzyme facilitates U2 snRNP function, minimizing the interference effect of transcriptional initiation factors on splicing. Since the number of affected introns was low it is nevertheless likely that there are other features among these introns, like secondary RNA structures, capable of making their splicing particularly sensitive to depletion of Salsa.

Expression of Salsa varies during development [104], being its expression particularly enriched in the central nervous system and in the female germ-line [142]. We propose that the differential expression of Salsa and other NTC-subunits gives an important contribution to spliceosome plasticity and differential gene expression during cell differentiation.

# 12 Material and Methods

## 12.1 Fly Husbandry

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

## 12.2 *Drosophila* RNAi stocks

Unless indicated, all stocks are available at the Bloomington Drosophila Stock Center. Depletion of Salsa was obtained using three different non-overlapping dsRNA hairpins. *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.

## 12.3 Generation of non-overlapping *salsa* RNAi (*salsa* RNAi-1 and RNAi-3)

We designed two non-overlapping RNAi hairpins against *salsa* (CG31368) using an algorithm that minimizes off target effects [143]. 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 top strand oligo (TS) of *salsa* 1 and *salsa* 3 are the following:

***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 bottom strand oligo (BS) of *salsa* 1 and *salsa* 3 are the following:

***salsa* RNAi -1 BS:**

aattcgcCGCTTGGATATGGACGATCTAtatgcttgaatataactaTAGATCGTCCATATCCAAGCGactg

***salsa* RNAi -3 BS:**

aattcgcCCACGATTATCTCCTACGCAAtatgcttgaatataactaTTGCGTAGGAGATAATCGTGGactg

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*. 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 digested with *NheI* and *EcoRI*) was added to the final reaction. The final volume of 20µl was made up with ddH20, mixed carefully and incubated overnight at 16°C. The reaction was stored at -20°C, until further use.

For transformation, 10µl of ligation reaction was transformed into 50µl TOP10 competent cells using electroporation method. Clones were selected through PCR using the standard primers for pVALIUM22 (S1 Table). The correct shRNA constructs were further confirmed by sequencing. The primer used for sequencing is shown in S1 Table. 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).

### *12.3.1 Germ line specific depletion of Salsa*

Salsa was specifically depleted within the female germ line by using the germ line-specific driver *nanos*-Gal4 (*nos*-Gal4) [100] and the upstream activating sequence (UAS)/Gal4 [101,103]

## 12.4 Ventralized eggshell phenotypes

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 to an egg collection cage supplemented with fresh yeast. Ventralized eggshell phenotypes were scored at least 2 days after assembly of collection cage (for optimal egg laying) and using fresh egg collections (4-5h) to facilitate unequivocal scoring.

## 12.5 Egg hatching

1-3 days old F1 females (n=20) were crossed with wild-type males (n=5) and added to an egg collection cage supplemented with fresh yeast. Eggs were collected using freshly prepared apple juice plates. Plates were incubated for 48h at 25°C and egg hatching was scored.

## 12.6 Ovaries immunostaining

Adult ovaries (20 ovary pairs per sample per experiment) were processed according to standard procedures [32]. All incubation steps involve a gentle mix with an orbital or nutator mixer. Briefly, ovaries were dissected from 2-4 days old females in ice cold Phosphate-buffered saline (PBS) and fixed for 20min with 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 three 10min washes in PBST (PBS +0.2% Tween 20), ovarioles were gently detached with a tungsten needle. Ovaries were subsequently blocked for 1h in PBS 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). After three 10min washes in PBST, ovaries were incubated for 1h at the room temperature with the appropriate secondary antibodies diluted in BBT solution. DNA was subsequently stained for 30 min at the room temperature using 1:10.000 DAPI in PBST. Prior to mounting, ovaries were washed three times, 10min each, with PBST. Ovaries were mounted in dako faramount aqueous mounting medium (Dako, California, USA) and were visualized using a Zeiss LSM710 Confocal microscope.

## 12.7 Quantification of ovaries Gurken immunostaining

For quantification of Gurken signal at the dorsal anterior region of the oocyte, maximum intensity projections of serial confocal optical sections of individual egg chambers stained with an anti-Gurken antibody [107] were obtained using Image J program (Grouped Zprojector, maximum pixel intensity). The average Gurken signal intensity within three oocyte dorsal anterior regions with highest levels of localized Gurken was obtained using ImageJ software. For each egg chamber, obtained values were divided by the respective average background signal intensity value obtained within three cytoplasmic regions of the nurse cells.

## 12.8 Preparation of a DIG-labelled *gurken* probe

A plasmid carrying a *gurken* cDNA sequence [66] 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). Synthesis of DIG-labeled antisense *gurken* RNA probes, was made according to the following reaction: 2ul of linearized plasmid (1ug) was mixed with the following reagents, in this order and at room temperature: 1µl of 10X transcription buffer (Roche), 1µl DIG-NTP mix (Roche), 0,3ul of RNAse inhibitor (Roche) and 1 µl T7 RNA polymerase (10 U/ µl; Roche). The final reaction volume was set up to 10µl

of DEPC water. The final reaction mixture was subsequently incubated at 37°C for 2h. Subsequently, 40µl DEPC water and 50µl 2X Carbonate Buffer (120mM-$Na_2CO_3$; 80mM $NaHcO_3$ at pH 10.2) were added to the probes and the mix was incubated at 65°C for 40min. Hydrolyzed probes were precipitated by adding 100µl 0.2M NaAc (pH 6), 10µl of 4M LiCl, 10ul of tRNA (Roche; stock solution 20mg/ul), and 600ul of absolute ethanol, on a -20ºC overnight incubation. Mix was centrifugated at 4°C for 15min. Pellet was washed twice with 70% EtOH, dried in air (10min), and dissolved in Hybridization MIX buffer (50% formamide, 5x SSC solution, 0.1% Tween-20, 100 µg/ml heparin, 100 µg/ml ssDNA). Probes were stored at -20ºC before use.

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

Adult ovaries from 2-4 days old females were dissected and fixed according to standard procedures [32]. All incubation steps involve a gentle mix with a nutator mixer. Ovaries were dissected from 2-4 days old females in ice cold Phosphate-buffered saline (PBS) and fixed for 20min with 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. After fix, ovarioles were gently teased apart with forceps and dehydrated in methanol by washes of 5 min with the following solutions: 30% methanol/PBT (PBS+0,2% Tween-20), 50% methanol/PBT and 70% of methanol/PBT. Ovaries were stored in 100% methanol at -20ºC before usage.

Prior usage, ovaries were rehydrated by washes with 70% Methanol/PBT; 50% Methanol/PBT and 30% Methanol/PBT. After rehydration, ovaries were washed with PBS+0,1% Tween-20, for 5 min, 3 times and transferred to a pre-hybridization buffer (40% formamide, 4x SSC (20xSSC (RNAse free), 0.1% Tween 20) and incubated for 1h at the room temperature. 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 55ºC. RNA probes were diluted at 5:100 in a pre-heated hybridization buffer. After removal of the hybridization solution, 50ul of pre-heated diluted probes were added to the ovaries for overnight incubation at 55°C. Following day, used probes were collected, and ovaries were transferred

to 1ml pre-hybridization solution for a 30min incubation at 55°C. They were subsequently washed 5 times with 1ml of PBS+0,1% Tween-20 for 30min each time, at the room temperature. After washes, ovaries were incubated in Roche Blocking Buffer (1:10 in PBS-0,1% Tween-20) for 1h at the room temperature. The mouse anti-Dig (1:400) was diluted in Roche Blocking Buffer (1:10 in PBS-0,1% Tween-20) and incubated overnight at 4°C in a rocking shaker. The ovaries were washed three times with 1ml PBS-0,1% Tween-20 for 10 minutes each at the room temperature and incubated with Roche Blocking Buffer as above for 30 minutes at the room temperature, following incubation (1h at RT) with secondary antibody anti-mouse-Cy3 (1:1000) diluted in Roche Blocking Buffer as above. Ovaries were washed three times with 1ml PBS-0,1% Tween-20, for 5 minutes each time, at room temperature. DNA was subsequently stained for 10 min at the room temperature using 1:10.000 DAPI in PBS-0,1% Tween-20. Prior to mounting, ovaries were washed in PBS-0,1% Tween-20 three times, 5min each, then rinsed with PBS. Ovaries were mounted in Vectashield mounting medium and they were visualized using a Zeiss LSM710 Confocal microscope.

## 12.10 Quantification of *gurken* mRNA localization

For semi-quantitative quantification of dorsal-anterior localization of *gurken* mRNA, stage 8/9 oocytes were divided it into three different classes based on *gurken* mRNA dorsal-anterior signal near to oocyte nucleus: a) normal dorsal-anterior localization of *gurken* mRNA, with a strong signal distributed in the periphery of the oocyte nucleus; b) partial dorsal-anterior localization, with weaker but clearly detectable *gurken* mRNA nearby the periphery of the oocyte nucleus; c) absence/almost absence of dorsal-anterior localization, with undetectable or hardly detectable *gurke*n mRNA signal nearby the periphery of the oocyte nucleus.

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

Total RNA was extracted from 0-1hr embryos (after egg laying) or from 2-4 days old adult female ovaries (after pupae eclosion), whose female germ line was specifically depleted for Salsa (*salsa* RNAi) or 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. 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. 200ng of RNA were used for cDNA synthesis, according to manufacturer's standard protocol. Primer combinations were designed with PrimerBLAST (NCBI, USA) and PCR was performed using NZYTech master mix (Lisbon, Portugal).

## 12.12 Real-Time quantitative PCR (Real-time qPCR)

### *12.12.1 Optimization of primers efficiency for target genes and reference genes*

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: $E=10^{(-1/slope)} -1$ [144]. Efficiency and regression curves for each primer sets are shown in S4 Table. 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 primer

sets with specificity and good efficiency were used for subsequent RT-qPCR analysis.

### *12.12.2 mRNA extraction and cDNA synthesis*

Total RNA was extracted from 0-1hr embryos (after egg laying) or from 2-4 days old adult female ovaries (after pupae eclosion), whose female germ line was specifically depleted for salsa (salsa RNAi) or 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. 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.

### *12.12.3 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. As positive control it was used genomic DNA. 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.

### *12.12.4 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 [145]. The relative expression ratio (RE) was calculated using the difference between ΔCt value of negative control (mCherry RNAi) and salsa RNAi, also known as the $2^{-\Delta\Delta ct}$ method.

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

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

$\Delta Ct_{\text{control}}$ 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 [146]. 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.

## 12.13 Protein extraction

0-1hr embryos (after egg laying) or 2-4 days old adult female ovaries (after pupae eclosion), whose female germ line was specifically depleted for salsa (salsa RNAi) or negative control (mCherry RNAi), were homogenized with a pestle in ice-cold NB-lysis buffer (50mM Tris (pH 7.5), 150mM NaCl, 2mM EDTA, 0.01% NP-40 (Igepal), 2mM DTT, 10 mM NaF, and protease inhibitors (Complete™ 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.

## 12.14 Western-blot analysis

For western blot analysis, total protein extracts were boiled for 5min at 95°C in 2xLaemmli buffer (Sigma). 10µg of protein were loaded per lane, and run in a SDS polyacrylamide gel electrophoresis gel (6% or 12%). Proteins were subsequently transferred onto Hybond-ECL nitrocellulose membrane (Amersham) using a Bio-Rad wet transfer apparatus (100V for 60 min). Western blotting was performed using standard procedures. Briefly, the Hybond-ECL membrane was blocked overnight with 5% non-fat milk in PBS-T (PBS +0.1% Tween-20) at 4°C with shaking. Primary antibodies were incubated overnight in 1% non-fat milk in PBS-T with gentle shaking at 4°C. Membranes were washed three times with PBS-T, for 15min at room temperature and with strong shaking. Secondary antibodies were incubated for 2h in 1% non-fat milk in PBS-T at the room temperature with gentle shaking. After three washes in PBS-T, for 5min at room temperature, and one quick wash with PBS, proteins of interest were detected with an ECL Plus western blotting detection system (GE Healthcare) and an ECL Hyperfilm (Amersham).

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

### *12.15.1 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) was used for site-directed mutagenesis, according to manufacturer's standard protocol. For primers sequence see S1 Table. For PCR amplification it was used the High Fidelity PCR system (Roche), with the following conditions: initial denaturation for

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

### *12.15.2 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.

### *12.15.3 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 100µL of Serum Free Medium (SFM), 400ng of DNA, and 4µL 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 4h of incubation at 25°C and cells were incubated for 48 hours at 25°C.

## 12.16 Co-immunoprecipitation

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

## 12.17 Differential intron retention analysis

Overall quality of raw RNA-seq data was analyzed using *fastqc* version 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) on the associated FASTQ files. *vast-tools* [121,147,148] version 2.0.0 was used for quantification of alternative sequence inclusion levels from FASTQ files using VAST-DB annotation for *Drosophila melanogaster* genome assembly dm6 [121,149] To ensure intron retention quantification with sufficient and reliable read coverage, we considered only *vast-tools*' events with a minimum mappability-corrected read coverage score of "OK" across all samples, corresponding to at least 20 mappability-corrected reads mapping either to the sum of skipping splice junctions or to one of the two inclusion exon-intron junctions and at least 15 of such reads mapping to the other inclusion splice junctions, as described in *vast-tools*'s GitHub description page (https://github.com/vastgroup/vast-tools). Also, only intron retention events without statistically significant evidence for a high imbalance in read counts between both upstream and downstream exon-intron junctions and the intron body [148] (binomial test's p-value $\geq 0.05$) for all samples were considered. Percent intron retention (PIR) [148] estimates are quantified as the ratio between the average of exon-intron junction reads that support intron inclusion (#inc) and the sum of these with exon-exon junction reads, supporting intron skipping (#exc).

Precision of PIR estimates from their supporting read coverage (#inc and #exc) was modelled using the beta distribution (conjugate prior probability distribution for the binomial), constrained to [0,1] and defined by shape parameters $\alpha$ and $\beta$ and a mean of $\alpha / (\alpha + \beta)$. Given that, for $\alpha, \beta \neq 0$ for the same mean, beta distributions get narrower with increasing shape parameters, emitting points from a beta distribution with $\alpha$ = #inc and $\beta$ = #exc allows the scattering of emitted values to serve as a surrogate for that PIR's dispersion (given the original read coverage supporting it), while the distribution's mean value, #inc / (#inc + #exc), is an approximation of the empirical PIR. For each considered intron retention event with nonzero #inc and #exc, the R function *rbeta* was used to emit, for each sample, 500 values from a beta distribution with shape parameters $\alpha$ = #inc and $\beta$ = #exc. However, while #inc or #exc can be equal to

zero, beta distributions with α or β = 0 are single-valued as 0 or 1, respectively, and are not appropriate to model inclusion levels. Therefore, an increment factor (incr) proportional to global coverage was added to zero #inc (PIR = 0) or #exc (PIR = 1) read counts, followed by emission, using *rbeta*, of 500 points from a beta distribution with α = #inc + incr and β = #exc + incr. Increment factors per coverage were set as the maximum value between 0.01 and 1 such that not more than 5% of simulated 500-point beta distributions with α = incr and β = coverage have a median that shows a deviation from theoretical PIR larger than 0.01.

For each considered event, *rbeta*-emitted values were grouped per condition (control and salsa RNAi), with their median values set as the respective summary PIRs ($PIR_{beta}$). The difference between these, $\Delta PIR_{beta} = PIR_{salsa\ RNAi} - PIR_{control}$, was used as the magnitude of differential intron inclusion. The significance of each $\Delta PIR_{beta}$ was set as the ratio between the number of differences between salsa-depleted and control emitted values that are greater than zero and the total number of differences, reflecting the probability of $PIR_{salsa\ RNAi}$ being greater than $PIR_{control}$ ($P_{diff}$).

Differentially retained introns were considered as those with $P_{diff} \geq 0.9$. From these, introns that were more retained upon salsa depletion than in Controls ($\Delta PIR_{beta} > 0.1$) were classified as *Affected,* while those showing decreased retention upon salsa depletion ($\Delta PIR_{beta} < -0.1$) were classified as *Less Retained*. Irrespectively of the significance of PIR differences, *Affected* introns showing very low inclusion levels in Control samples (mean PIR < 0.05) were classified into a separate group, *Affected low PIR*. *Included* introns were set as those showing consistently high inclusion, i.e. with a PIR ≥ 0.75 for all samples and an absolute $\Delta PIR_{beta} < 0.02$, while *Skipped* introns were classified as those always found spliced out, having PIR = 0 for all samples. Finally, *Control* introns were classified as those showing $PIR_{salsa\ RNAi}$ and $PIR_{control}$ between 0.05 and 0.9 (alternatively retained introns) but their retention does not seem to be affected by salsa depletion (absolute $\Delta PIR_{beta} < 0.02$).

*MATT* [137] was used to retrieve intronic features, such as intron, upstream or downstream exon length, splice site strength [136] and GC content across groups using function *get_ifeatures*. Ensembl Biomart [150] was used to extract

transcription start sites per transcript to calculate distance between these and the intron start (5'ss). Differential alternative splicing analysis considering all event types annotated in *vast-tools* (S8 Fig.) was performed using the same filtering criteria as described above for intron retention.

## 12.18 Differential gene expression analysis

Gene expression was quantified with *vast-tools*, that aligns RNA-seq reads against a reference transcriptome. Differential gene expression analysis was performed with the limma R package [151]. The magnitude of difference in expression was measured in log2 fold-change between control and salsa RNAi. Six genes were considered to be significantly differentially expressed (those with an associated positive empirical Bayes statistic giving the log-odds ratio of the gene being differentially expressed): CG4570 and Arc1 downregulated and Drat, Cyp4p2, sev and Ilp6 upregulated in salsa-depleted samples.

**ACKNOWLEDGEMENTS**

The authors wish to thank Julian König (IMB Mainz) for scientific discussions. We also thank Cláudia Florindo for assistance in fluorescence microscopy, the TRiP collection at Harvard Medical School (NIH/NIGMS R01-GM084947) for providing several of the transgenic RNAi fly stocks used in this study, and Manuel Irimia (CRG Barcelona) for advance access to the *Drosophila* VAST-DB annotation. The Microscopy Unit was partially supported by Portuguese national funding (FCT: PPBI-POCI-01-0145-FEDER-022122). This work was developed with the support of the research infrastructure Congento (project LISBOA-01-0145-FEDER-022170). Rui G. Martinho is supported by Portuguese national funding through Fundação para a Ciência e a Tecnologia [FCT grant refs. PTDC/BEX-BID/0395/2014, PTDC/BIA-BID/28441/2017, and UID/BIM/04773/2013 CBMR 1334]. Om Rathore was supported by Portuguese national funding through Fundação para a Ciência e a Tecnologia, ref. PD/BD/52428/2013 within the scope of the ProRegeM PhD program (Ref. PD/00117/2012, CRM:0027030). Om Rathore was also supported by a Federation of European Biochemical societies (FEBS) short-term fellowship. Nuno L. Barbosa-Morais is supported by EMBO

Installation Grant (3057), Investigador FCT Starting Grant (IF/00595/2014), UID/BIM/50005/2019, project funded by Fundação para a Ciência e a Tecnologia (FCT)/Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) through Fundos do Orçamento de Estado and iMM Lisboa start-up funds; Mariana Ascensão-Ferreira is supported by Fundação para a Ciência e Tecnologia (FCT) PhD fellowship (PD/BD/128283/2017) and Fundação AstraZeneca.

## AUTHORS CONTRIBUTIONS

O.S.R.: Investigation; R.M.: Investigation; M.A-F: Investigation, Writing (review & editing); M.N.T.: Investigation; P.P.: Investigation; R.D.S.: Investigation; R.P.A.: Conceptualization, J-Y.R.: Conceptualization, Writing (review & editing); N.L.B-M.: Conceptualization, Writing (review & editing); R.G.M.: Conceptualization, Writing (original draft + review & editing).

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

1. Rodrigues, P., Limback, D., Mcginnis, L.K., Plancha, C.E., and Albertini, D.F. (2008). Oogenesis: Prospects and challenges for the future. J. Cell. Physiol. *216*, 355–365.

2. Regan, S.L.P., Knight, P.G., Yovich, J.L., Leung, Y., Arfuso, F., and Dharmarajan, A. (2018). Involvement of Bone Morphogenetic Proteins (BMP) in the Regulation of Ovarian Function 1st ed. (Elsevier Inc.) Available at: http://dx.doi.org/10.1016/bs.vh.2018.01.015.

3. Marques-Mari, A.I., Lacham-Kaplan, O., Medrano, J. V., Pellicer, A., and Simón, C. (2009). Differentiation of germ cells and gametes from stem cells. Hum. Reprod. Update *15*, 379–390.

4. Johnson, J., Bagley, J., Skaznik-Wikiel, M., Lee, H.J., Adams, G.B., Niikura, Y., Tschudy, K.S., Tilly, J.C., Cortes, M.L., Forkert, R., *et al.* (2005). Oocyte generation in adult mammalian ovaries by putative germ cells in bone marrow and peripheral blood. Cell *122*, 303–315.

5. Feher, J. (2005). Female Reproductive Physiology, in Quantitative Human Physiology. Acad. Press Boston. p. 846-855.

6. Dutta, S., Burks, D.M., and Pepling, M.E. (2016). Arrest at the diplotene stage of meiotic prophase I is delayed by progesterone but is not required for primordial follicle formation in mice. Reprod. Biol. Endocrinol. *14*, 1–8. Available at: http://dx.doi.org/10.1186/s12958-016-0218-1.

7. Feher, J. (2012). Quantitative Human Physiology. Female Reprod. Physiol., pp 846-855.

8. Rimon-dahari, N., Yerushalmi-heinemann, L., and Alyagor, L. (2016). Molecular Mechanisms of Cell Differentiation in Gonad Development. *58*, 167–190. Available at: http://link.springer.com/10.1007/978-3-319-31973-5.

9. Hsueh, A.J.W., Kawamura, K., Cheng, Y., and Fauser, B.C.J.M. (2015). Intraovarian control of early folliculogenesis. Endocr. Rev. *36*, 1–24.

10. Xie, Z., Liu, H., Pan, Z., Shi, F., Li, Q., Li, Y., and Zhang, J. (2012). Current Advances in Epigenetic Modification and Alteration during Mammalian

Ovarian Folliculogenesis. J. Genet. Genomics *39*, 111–123. Available at: http://dx.doi.org/10.1016/j.jgg.2012.02.004.

11. Pépin, D., Vanderhyden, B.C., Picketts, D.J., and Murphy, B.D. (2007). ISWI chromatin remodeling in ovarian somatic and germ cells: revenge of the NURFs. Trends Endocrinol. Metab. *18*, 215–224.

12. Szmelskyj, I., Aquilina, L., and Szmelskyj, A.O. (2014). Anatomy and physiology of the reproductive system. Acupunct. IVF Assist. Reprod., 23–58.

13. Monniaux, D., Clément, F., Dalbiès-Tran, R., Estienne, A., Fabre, S., Mansanet, C., and Monget, P. (2014). The Ovarian Reserve of Primordial Follicles and the Dynamic Reserve of Antral Growing Follicles: What Is the Link?1. Biol. Reprod. *90*, 1–11.

14. Thomas, F.H., and Vanderhyden, B.C. (2006). Oocyte-granulosa cell interactions during mouse follicular development: Regulation of kit ligand expression and its role in oocyte growth. Reprod. Biol. Endocrinol. *4*, 1–8.

15. Ernst, E.H., Grøndahl, M.L., Grund, S., Hardy, K., Heuck, A., Sunde, L., Franks, S., Andersen, C.Y., Villesen, P., and Lykke-Hartmann, K. (2017). Dormancy and activation of human oocytes from primordial and primary follicles: Molecular clues to oocyte regulation. Hum. Reprod. *32*, 1684–1700.

16. Sagata, N. (1996). Meiotic metaphase arrest in animal oocytes: Its mechanisms and biological significance. Trends Cell Biol. *6*, 22–28.

17. Kim, J.Y. (2012). Control of ovarian primordial follicle activation. Clin. Exp. Reprod. Med. *39*, 10–14.

18. Hughes, S.E., Miller, D.E., Miller, A.L., and Hawley, R.S. (2018). Female meiosis: Synapsis, recombination, and segregation in Drosophila melanogaster.

19. Bratu, D., McNeil, G., and Walker, J.M. (2015). Drosophila Oogenesis: Methods and Protocols.

20. Kirilly, D., and Xie, T. (2007). The Drosophila ovary: An active stem cell

community. Cell Res. *17*, 15–25.

21. Hou, S.X., and Singh, S.R. (2008). Germline stem cells. Methods Mol. Biol. *450*, 1–20.

22. Voog, J., and Jones, D.L. (2010). Stem Cells and the Niche: A Dynamic Duo. Cell Stem Cell *6*, 103–115. Available at: http://dx.doi.org/10.1016/j.stem.2010.01.011.

23. Takeo, S., Lake, C.M., Morais-De-Sá, E., Sunkel, C.E., and Hawley, R.S. (2011). Synaptonemal complex-dependent centromeric clustering and the initiation of synapsis in Drosophila oocytes. Curr. Biol. *21*, 1845–1851.

24. Mehrotra, S., and McKim, K.S. (2006). Temporal analysis of meiotic DNA double-strand break formation and repair in Drosophila females. PLoS Genet. *2*, 1883–1897.

25. Joyce, E.F., Apostolopoulos, N., Beliveau, B.J., and Wu, C.T. (2013). Germline Progenitors Escape the Widespread Phenomenon of Homolog Pairing during Drosophila Development. PLoS Genet. *9*.

26. King, R.C., Aggarwal, S.K., and Aggarwal, U. (1968). The development of the female Drosophila reproductive system. J. Morphol. *124*, 143–165.

27. Roth, S., and Lynch, J.A. (2009). Symmetry Breaking During Drosophila Oogenesis. Cold Spring Harb. Perspect. Biol. *1*, a001891–a001891. Available at: http://cshperspectives.cshlp.org/lookup/doi/10.1101/cshperspect.a001891%5Cnpapers2://publication/doi/10.1101/cshperspect.a001891.

28. von Stetina, J.R., and Orr-Weaver, T.L. (2011). Developmental control of oocyte maturation and egg activation in metazoan models. Cold Spring Harb. Perspect. Biol. *3*, 1–19.

29. 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*, 817–824.

30. 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*, 1–10.

31. 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*, 2571–2582.

32. 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*.

33. Tiefert, M. (1970). Fine Structural Changes in the Drosophila Oocyte Nucleus during a Short Period of RNA Synthesis. *25*, 8–25.

34. Zhaunova, L., Ohkura, H., and Breuer, M. (2016). Kdm5/Lid Regulates Chromosome Architecture in Meiotic Prophase I Independently of Its Histone Demethylase Activity. PLoS Genet. *12*, 1–22.

35. Mahajan-Miklos S, C.L. (1994). Intercellular cytoplasm transport during Drosophila oogenesis. 10.1006/dbio.1994.1257.

36. 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. Available at: http://dx.doi.org/10.1016/j.ydbio.2014.08.008.

37. Bohrmann, J., and Biber, K. (1994). Cytoskeleton-dependent transport of cytoplasmic particles in previtellogenic to mid-vitellogenic ovarian follicles of Drosophila: time-lapse analysis using video-enhanced contrast microscopy. J. Cell Sci. *107 ( Pt 4*, 849–58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8056841.

38. Pritchett, T.L., Tanner, E.A., and McCall, K. (2009). Cracking open cell death in the Drosophila ovary. Apoptosis *14*, 969–979.

39. Mahowald, A.P., and Thomas, J. (1983). In Vitro Activation of Drosophila Eggs. *445*.

40. Bashirullah, A., Reed, B.H., Lipshitz, H.D., Semotok, J.L., Cooperstock, R.L., Houston, S.A., and Tadros, W. (2003). Regulation of maternal transcript destabilization during egg activation in Drosophila. Genetics *164*, 989–1001. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12871909.

41. Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG Is a Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation Is Activated by the PAN GU Kinase. Dev. Cell *12*, 143–155.

42. Snustad, D. P., & Simmons, M.J. (2015). Principles of genetics. New York John Wiley Sons.

43. Bastock, R., and St Johnston, D. (2008). Drosophila oogenesis. Curr. Biol. *18*, R1082–R1087.

44. Lake, C.M., and Hawley, R.S. The Molecular Control of Meiotic Chromosomal Behavior : Events in Early Meiotic Prophase in Drosophila Oocytes.

45. Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–19. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17320508.

46. Lloret-Llinares, M., Pérez-Lluch, S., Rossell, D., Morán, T., Ponsa-Cobas, J., Auer, H., Corominas, M., and Azorín, F. (2012). DKDM5/LID regulates H3K4me3 dynamics at the transcription-start site (TSS) of actively transcribed developmental genes. Nucleic Acids Res. *40*, 9493–9505.

47. Liu, X., Greer, C., and Secombe, J. (2014). KDM5 Interacts with Foxo to Modulate Cellular Levels of Oxidative Stress. PLoS Genet. *10*.

48. Drelon, C., Belalcazar, H.M., and Secombe, J. (2018). The histone demethylase KDM5 is essential for larval growth in Drosophila.

49. Nilson, T.L., Sinclair, B.J., and Roberts, S.P. (2006). The effects of carbon

dioxide anesthesia and anoxia on rapid cold-hardening and chill coma recovery in Drosophila melanogaster. J. Insect Physiol. *52*, 1027–1033. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16996534 [Accessed June 10, 2019].

50. Artiss, T., and Hughes, B. (2007). Taking the Headaches Out of Anesthetizing Drosophila. Am. Biol. Teach. *69*, e77–e80.

51. Champion De Crespigny, F.E., and Wedell, N. (2008). The impact of anaesthetic technique on survival and fertility in Drosophila. Physiol. Entomol. *33*, 310–315.

52. Drelon, C., Belalcazar, H.M., and Secombe, J. (2018). The histone demethylase KDM5 is essential for larval growth in Drosophila. Genetics *209*, 773–787.

53. Hsieh, C., and Thomson, R. (1973). Use of a Yeast Sitespecific Recombinase to Produce Female Germline Chimeras in Drosophila. J. Appl. Phys. *44*, 2051–2063.

54. Jia, D., and Xie, Q. (2016). Automatic Drosophila Egg Chamber Stages Identification From DAPI Images.

55. Secombe, J., Li, L., Carlos, L., and Eisenman, R.N. (2007). The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. Genes Dev. *21*, 537–551.

56. Ghabrial, A., Ray, R.P., and Schüpbach, 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.

57. Abdu, U., González-Reyes, A., Ghabrial, A., and Schüpbach, T. (2003). The Drosophila spn-D gene encodes a RAD51C-like protein that is required exclusively during meiosis. Genetics *165*, 197–204.

58. 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.

59. 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*.

60. Li, X., Liu, L., Yang, S., Song, N., Zhou, X., Gao, J., Yu, N., Shan, L., Wang, Q., Liang, J., *et al.* (2014). Histone demethylase KDM5B is a key regulator of genome stability. Proc. Natl. Acad. Sci. *111*, 7096–7101. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1324036111.

61. Valentin, V. V, Turken, U., Moore, S., Oaksford, M., Sephton, S.E., Hellhammer, D.H., Holt, C.S., Watson, D., Leeka, J., Mehl, M.R., *et al.* (2011). Histone Lysine Demethylase JARID1a Activates CLOCK-BMAL1 and Influences the Circadian Clock. 1881–1886.

62. Morán, T., Bernués, J., and Azorín, F. (2015). The Drosophila histone demethylase dKDM5/LID regulates hematopoietic development. Dev. Biol. *405*, 260–268. Available at: http://dx.doi.org/10.1016/j.ydbio.2015.07.011.

63. Collins, K.A., Unruh, J.R., Slaughter, B.D., Yu, Z., Lake, C.M., Nielsen, R.J., Box, K.S., Miller, D.E., Blumenstiel, J.P., Perera, A.G., *et al.* (2014). Corolla is a novel protein that contributes to the architecture of the synaptonemal complex of Drosophila. Genetics *198*, 219–228.

64. Rørth, P. (1998). Gal4 in the Drosophila female germline. Mech. Dev. *78*, 113–118.

65. Gonzâlez-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. Available at: https://doi.org/10.1038/375654a0.

66. Neuman-Silberberg, F.S., and Schüpbach, T. (1993). The drosophila dorsoventral patterning gene gurken produces a dorsally localized RNA and encodes a TGFα-like protein. Cell *75*, 165–174.

67. Schiipbach, T. (1987). Germ Line and Soma Cooperate during Oogenesis to Establish the Dorsoventral Pattern of Egg Shell and Embryo in Drosophila melanogaster. Cell *49*, 699–707.

68. 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.

69. González-Reyes, A., and St Johnston, D. (1994). Role of oocyte position in establishment of anterior-posterior polarity in Drosophila. Science (80-. ). *266*, 639–642. Available at: http://science.sciencemag.org/content/266/5185/639.abstract.

70. Zhao, T., Graham, O.S., Raposo, A., and St Johnston, D. (2012). Growing microtubules push the oocyte nucleus to polarize the Drosophila dorsal-ventral axis. Science (80-. ). *336*, 999–1003. Available at: http://science.sciencemag.org/content/336/6084/999.abstract.

71. Huynh, J.R., and St Johnston, D. (2004). The origin of asymmetry: Early polarisation of the Drosophila germline cyst and oocyte. Curr. Biol. *14*, R438–R449. Available at: http://www.sciencedirect.com/science/article/pii/S0960982204003677.

72. Morisato, D., and Anderson, K. V. (1994). The spätzle gene encodes a component of the extracellular signaling pathway establishing the dorsal-ventral pattern of the Drosophila embryo. Cell *76*, 677–688. Available at: http://www.sciencedirect.com/science/article/pii/009286749490507X.

73. Schneider, D.S., Jin, Y., Morisato, D., and Anderson, K. V (1994). A processed form of the Spätzle protein defines dorsal-ventral polarity in the Drosophila embryo. Development *120*, 1243–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8026333.

74. Caceres, L. (2005). Production of gurken in the nurse cells is sufficient for axis determination in the Drosophila oocyte. Development *132*, 2345–2353. Available at: http://dev.biologists.org/content/132/10/2345.abstract.

75. 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. Available at: http://www.sciencedirect.com/science/article/pii/S1534580703000583.

76. Delanoue, R., Herpers, B., Soetaert, J., Davis, I., and Rabouille, C. (2007). Drosophila Squid/hnRNP Helps Dynein Switch from a gurken mRNA

Transport Motor to an Ultrastructural Static Anchor in Sponge Bodies. Dev. Cell *13*, 523–538. Available at: http://www.sciencedirect.com/science/article/pii/S1534580707003449.

77. 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. Available at: http://www.sciencedirect.com/science/article/pii/S0012160600996908.

78. Saunders, C., and Cohen, R.S. (1999). 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. Available at: http://www.sciencedirect.com/science/article/pii/S1097276500801732.

79. 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. Available at: http://europepmc.org/abstract/MED/15992540.

80. Davidson, A., Parton, R.M., Rabouille, C., Weil, T.T., and Davis, I. (2016). Localized Translation of gurken/TGF-α mRNA during Axis Specification Is Controlled by Access to Orb/CPEB on Processing Bodies. Cell Rep. *14*, 2451–2462. Available at: http://europepmc.org/abstract/MED/26947065.

81. 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. Available at: https://doi.org/10.1002/dvdy.24311.

82. Tan, L., Chang, J.S., Costa, a, and Schedl, P. (2001). An autoregulatory feedback loop directs the localized expression of the Drosophila CPEB protein Orb in the developing oocyte. Development *128*, 1159–69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11245581.

83. Wong, L.C., and Schedl, P. (2011). Cup blocks the precocious activation of the Orb autoregulatory loop. PLoS One *6*, e28261. Available at: https://doi.org/10.1371/journal.pone.0028261.

84. Wong, L.C., Costa, A., McLeod, I., Sarkeshik, A., Yates, J., Kyin, S.,

Perlman, D., and Schedl, P. (2011). The functioning of the drosophila CPEB protein orb is regulated by phosphorylation and requires casein kinase 2 activity. PLoS One *6*, e24355. Available at: http://europepmc.org/abstract/MED/21949709.

85. Derrick, C.J., and Weil, T.T. (2017). Translational control of gurken mRNA in Drosophila development. Cell Cycle *16*, 23–32. Available at: https://doi.org/10.1080/15384101.2016.1250048.

86. 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. Available at: http://www.biochemsoctrans.org/content/38/4/1110.abstract.

87. Chanarat, S., and Sträßer, K. (2013). Splicing and beyond: The many faces of the Prp19 complex. Biochim. Biophys. Acta - Mol. Cell Res. *1833*, 2126–2134. Available at: http://www.sciencedirect.com/science/article/pii/S0167488913002139.

88. Martinho, R.G., Guilgur, L.G., and Prudêncio, P. (2015). How gene expression in fast-proliferating cells keeps pace. BioEssays *37*, 514–524. Available at: https://doi.org/10.1002/bies.201400195.

89. Mahajan, K. (2016). hPso4/hPrp19: A critical component of DNA repair and DNA damage checkpoint complexes. Oncogene *35*, 2279–2286. Available at: https://doi.org/10.1038/onc.2015.321.

90. Maréchal, A., Li, J.M., Ji, X.Y., Wu, C.S., Yazinski, S.A., Nguyen, H.D., Liu, S., Jiménez, A.E., Jin, J., and Zou, L. (2014). PRP19 Transforms into a Sensor of RPA-ssDNA after DNA Damage and Drives ATR Activation via a Ubiquitin-Mediated Circuitry. Mol. Cell *53*, 235–246. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24332808.

91. Alex, S., You, L., Daniel, L., Feng, J., Arriaga, E.A., Piskounova, E., Agathocleous, M., Murphy, M.M., Hu, Z., Huddlestun, S.E., *et al.* (2014). GraphPad Prism User Guide. Neoplasia *16*, 2591–2598. Available at: http://dx.doi.org/10.1016/j.neo.2014.03.006.

92. 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*, 940–950. Available at: http://europepmc.org/abstract/MED/17981804.

93. Wan, L., and Huang, J. (2014). The PSO4 protein complex associates with replication protein a (RPA) and modulates the activation of ataxia telangiectasia-mutated and RAD3-related (ATR). J. Biol. Chem. *289*, 6619–6626.

94. 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*, 972–982. Available at: https://www.ncbi.nlm.nih.gov/pubmed/21536736.

95. 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–e02181. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24755291.

96. 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. Available at: https://www.ncbi.nlm.nih.gov/pubmed/18981222.

97. 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. Available at: https://doi.org/10.1038/nsmb.2951.

98. Hirose, T., Ideue, T., Nagai, M., Hagiwara, M., Shu, M.-D., and Steitz, J.A. (2006). A Spliceosomal Intron Binding Protein, IBP160, Links Position-Dependent Assembly of Intron-Encoded Box C/D snoRNP to Pre-mRNA Splicing. Mol. Cell *23*, 673–684. Available at: http://www.sciencedirect.com/science/article/pii/S1097276506004916.

99. Ideue, T., Sasaki, Y.T.F., Hagiwara, M., and Hirose, T. (2007). Introns play an essential role in splicing-dependent formation of the exon junction complex. Genes Dev. *21*, 1993–1998. Available at: https://www.ncbi.nlm.nih.gov/pubmed/17675447.

100. Doren, M. Van, Williamson, A.L., and Lehmann, R. (1998). Regulation of zygotic gene expression in Drosophila primordial germ cells. Curr. Biol. *8*, 243–246. Available at: http://www.sciencedirect.com/science/article/pii/S0960982298700910.

101. Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401 LP – 415. Available at: http://dev.biologists.org/content/118/2/401.abstract.

102. Rørth, P. (1998). Gal4 in the Drosophila female germline. Mech. Dev. *78*, 113–118. Available at: http://www.sciencedirect.com/science/article/pii/S0925477398001579.

103. Ni, J.-Q., Zhou, R., Czech, B., Liu, L.-P., Holderbaum, L., Yang-Zhou, D., Shim, H.-S., Tao, R., Handler, D., Karpowicz, P., *et al.* (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nat. Methods *8*, 405. Available at: https://doi.org/10.1038/nmeth.1592.

104. Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., *et al.* (2010). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473. Available at: https://doi.org/10.1038/nature09715.

105. 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 LP – 2463. Available at: http://dev.biologists.org/content/120/9/2457.abstract.

106. Popp, M.W.-L., and Maquat, L.E. (2013). Organizing principles of mammalian nonsense-mediated mRNA decay. Annu. Rev. Genet. *47*, 139–165. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24274751.

107. Queenan, A.M., Barcelo, G., Van Buskirk, C., and Schüpbach, T. (1999).

The transmembrane region of Gurken is not required for biological activity, but is necessary for transport to the oocyte membrane in Drosophila. Mech. Dev. *89*, 35–42. Available at: http://europepmc.org/abstract/MED/10559478.

108. Ghosh, S., Marchand, V., Gáspár, I., and Ephrussi, A. (2012). Control of RNP motility and localization by a splicing-dependent structure in oskar mRNA. Nat. Struct. &Amp; Mol. Biol. *19*, 441. Available at: https://doi.org/10.1038/nsmb.2257.

109. Hachet, O., and Ephrussi, A. (2004). Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. Nature *428*, 959–963. Available at: http://europepmc.org/abstract/MED/15118729.

110. 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. Available at: http://europepmc.org/abstract/MED/21103393.

111. Nott, A., Meislin, S.H., and Moore, M.J. (2003). A quantitative analysis of intron effects on mammalian gene expression. RNA *9*, 607–617. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12702819.

112. 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. Available at: https://www.ncbi.nlm.nih.gov/pubmed/28787591.

113. Chen, Y., Pane, A., and Schüpbach, T. (2007). Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. Curr. Biol. *17*, 637–642. Available at: http://europepmc.org/abstract/MED/17363252.

114. Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H.A., and Theurkauf, W.E. (2007). Drosophila rasiRNA Pathway Mutations Disrupt Embryonic Axis Specification through Activation of an ATR/Chk2

DNA Damage Response. Dev. Cell *12*, 45–55. Available at: http://www.sciencedirect.com/science/article/pii/S1534580706005594.

115. 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. Available at: http://europepmc.org/abstract/MED/12361566.

116. 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 LP – 1578. Available at: http://dev.biologists.org/content/125/9/1569.abstract.

117. Goodrich, J.S., Clouse, K.N., and Schüpbach, T. (2004). Hrb27C, Sqd and Otu cooperatively regulate gurken RNA localization and mediate nurse cell chromosome dispersion in Drosophila oogenesis. Development *131*, 1949–1958. Available at: http://europepmc.org/abstract/MED/15056611.

118. 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*, 864–876. Available at: http://europepmc.org/abstract/MED/10197986.

119. Van Buskirk, C., and Schüpbach, T. (2002). Half pint regulates alternative splice site selection in Drosophila. Dev. Cell *2*, 343–353. Available at: http://europepmc.org/abstract/MED/11879639.

120. Dej, K.J., and Spradling, A.C. (1999). The endocycle controls nurse cell polytene chromosome structure during Drosophila oogenesis. Development *126*, 293 LP – 303. Available at: http://dev.biologists.org/content/126/2/293.abstract.

121. Tapial, J., Ha, K.C.H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallières, M., Permanyer, J., Sodaei, R., Marquez, Y., *et al.* (2017). An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. Genome Res. *27*, 1759–

1768. Available at: https://www.ncbi.nlm.nih.gov/pubmed/28855263.

122. 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*, e32537. Available at: https://www.ncbi.nlm.nih.gov/pubmed/29280736.

123. Sterner, D.A., Carlo, T., and Berget, S.M. (1996). Architectural limits on split genes. Proc. Natl. Acad. Sci. U. S. A. *93*, 15081–15085. Available at: https://www.ncbi.nlm.nih.gov/pubmed/8986767.

124. Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.-P., Baldi, P.F., and Hertel, K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc. Natl. Acad. Sci. U. S. A. *102*, 16176–16181. Available at: https://www.ncbi.nlm.nih.gov/pubmed/16260721.

125. 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*, 657–668. Available at: http://www.sciencedirect.com/science/article/pii/0092867494905304.

126. Izaurralde, E., McGuigan, C., Mattaj, I.W., Lewis, J.D., and Jarmolowski, A. (2007). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5’ splice site. Genes Dev. *10*, 1683–1698.

127. Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional Spliceosome Assembly Occurs in a Stepwise Fashion and Requires the Cap Binding Complex. Mol. Cell *19*, 53–63. Available at: http://www.sciencedirect.com/science/article/pii/S1097276505013134.

128. Pabis, M., Neufeld, N., Steiner, M.C., Bojic, T., Shav-Tal, Y., and Neugebauer, K.M. (2013). The nuclear cap-binding complex interacts with the U4/U6·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. RNA *19*, 1054–1063. Available at: https://www.ncbi.nlm.nih.gov/pubmed/23793891.

129. Ashe, M.P., Pearson, L.H., and Proudfoot, N.J. (1997). The HIV-1 5’ LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. EMBO J. *16*, 5752–5763. Available at:

https://www.ncbi.nlm.nih.gov/pubmed/9312033.

130. Gunderson, S.I., Polycarpou-Schwarz, M., and Mattaj, I.W. (1998). U1 snRNP Inhibits Pre-mRNA Polyadenylation through a Direct Interaction between U1 70K and Poly(A) Polymerase. Mol. Cell *1*, 255–264. Available at: http://www.sciencedirect.com/science/article/pii/S109727650080026X.

131. Guo, J., Garrett, M., Micklem, G., and Brogna, S. (2011). Poly(A) signals located near the 5' end of genes are silenced by a general mechanism that prevents premature 3'-end processing. Mol. Cell. Biol. *31*, 639–651. Available at: https://www.ncbi.nlm.nih.gov/pubmed/21135120.

132. Lin, K., and Zhang, D.-Y. (2005). The excess of 5' introns in eukaryotic genomes. Nucleic Acids Res. *33*, 6522–6527. Available at: https://www.ncbi.nlm.nih.gov/pubmed/16314314.

133. 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. Available at: http://europepmc.org/abstract/MED/12468090.

134. Qiu, J., Cheng, F., and Pintel, D. (2007). Distance-dependent processing of adeno-associated virus type 5 RNA is controlled by 5' exon definition. J. Virol. *81*, 7974–7984. Available at: http://europepmc.org/abstract/MED/17507471.

135. Lepennetier, G., and Catania, F. (2016). mRNA-Associated Processes and Their Influence on Exon-Intron Structure in Drosophila melanogaster. G3 (Bethesda). *6*, 1617–1626. Available at: https://www.ncbi.nlm.nih.gov/pubmed/27172210.

136. Yeo, G., and Burge, C.B. (2004). Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. J. Comput. Biol. *11*, 377–394. Available at: https://doi.org/10.1089/1066527041410418.

137. Gohr, A., and Irimia, M. (2019). Matt: Unix tools for alternative splicing analysis. Bioinformatics *35*, 130–132. Available at: http://europepmc.org/abstract/MED/30010778.

138. Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.-H.A., Marr 2nd, M.T., and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. *25*, 2502–2512. Available at: https://www.ncbi.nlm.nih.gov/pubmed/22156210.

139. Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24514444.

140. Kwek, K.Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N.J., and Akoulitchev, A. (2002). U1 snRNA associates with TFIIH and regulates transcriptional initiation. Nat. Struct. Biol. *9*, 800–805. Available at: https://doi.org/10.1038/nsb862.

141. Damgaard, C.K., Kahns, S., Lykke-Andersen, S., Nielsen, A.L., Jensen, T.H., and Kjems, J. (2008). A 5′ Splice Site Enhances the Recruitment of Basal Transcription Initiation Factors In Vivo. Mol. Cell *29*, 271–278. Available at: http://www.sciencedirect.com/science/article/pii/S1097276508000075.

142. Chintapalli, V.R., Wang, J., and Dow, J.A.T. (2007). Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat. Genet. *39*, 715. Available at: https://doi.org/10.1038/ng2049.

143. Vert, J.-P., Foveau, N., Lajaunie, C., and Vandenbrouck, Y. (2006). An accurate and interpretable model for siRNA efficacy prediction. BMC Bioinformatics *7*, 520. Available at: https://www.ncbi.nlm.nih.gov/pubmed/17137497.

144. Nolan, T., Hands, R.E., and Bustin, S.A. (2006). Quantification of mRNA using real-time RT-PCR. Nat. Protoc. *1*, 1559–1582. Available at: https://doi.org/10.1038/nprot.2006.236.

145. Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. *3*, 1101–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18546601.

146. 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–RESEARCH0034. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12184808.

147. 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*, 1511–1523. Available at: https://www.ncbi.nlm.nih.gov/pubmed/25525873.

148. 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*, 1774–1786. Available at: https://www.ncbi.nlm.nih.gov/pubmed/25258385.

149. Torres-Méndez, A., Bonnal, S., Marquez, Y., Roth, J., Iglesias, M., Permanyer, J., Almudí, I., O'Hanlon, D., Guitart, T., Soller, M., *et al.* (2019). A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. Nat. Ecol. Evol. *3*, 691–701. Available at: https://doi.org/10.1038/s41559-019-0813-6.

150. Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., *et al.* (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford). *2011*, bar030–bar030. Available at: https://www.ncbi.nlm.nih.gov/pubmed/21785142.

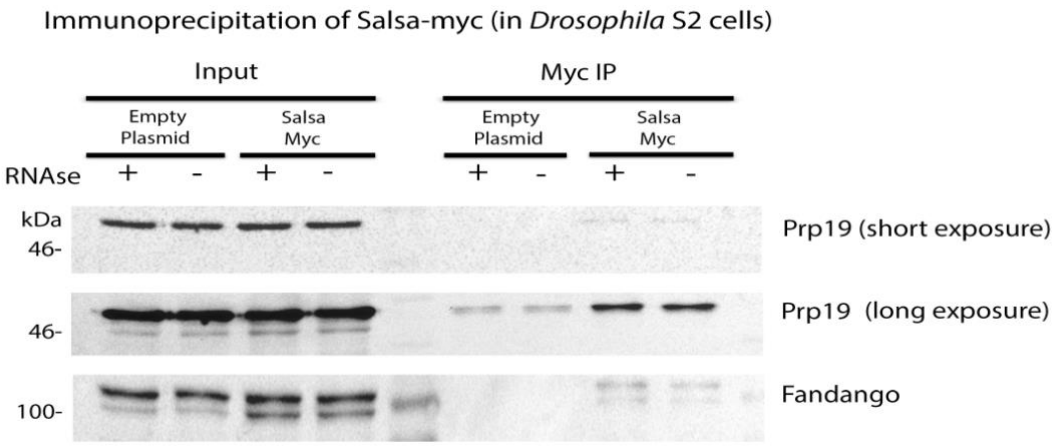
151. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47. Available at: https://www.ncbi.nlm.nih.gov/pubmed/25605792.

152. Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476. Available at: https://www.ncbi.nlm.nih.gov/pubmed/18978772.

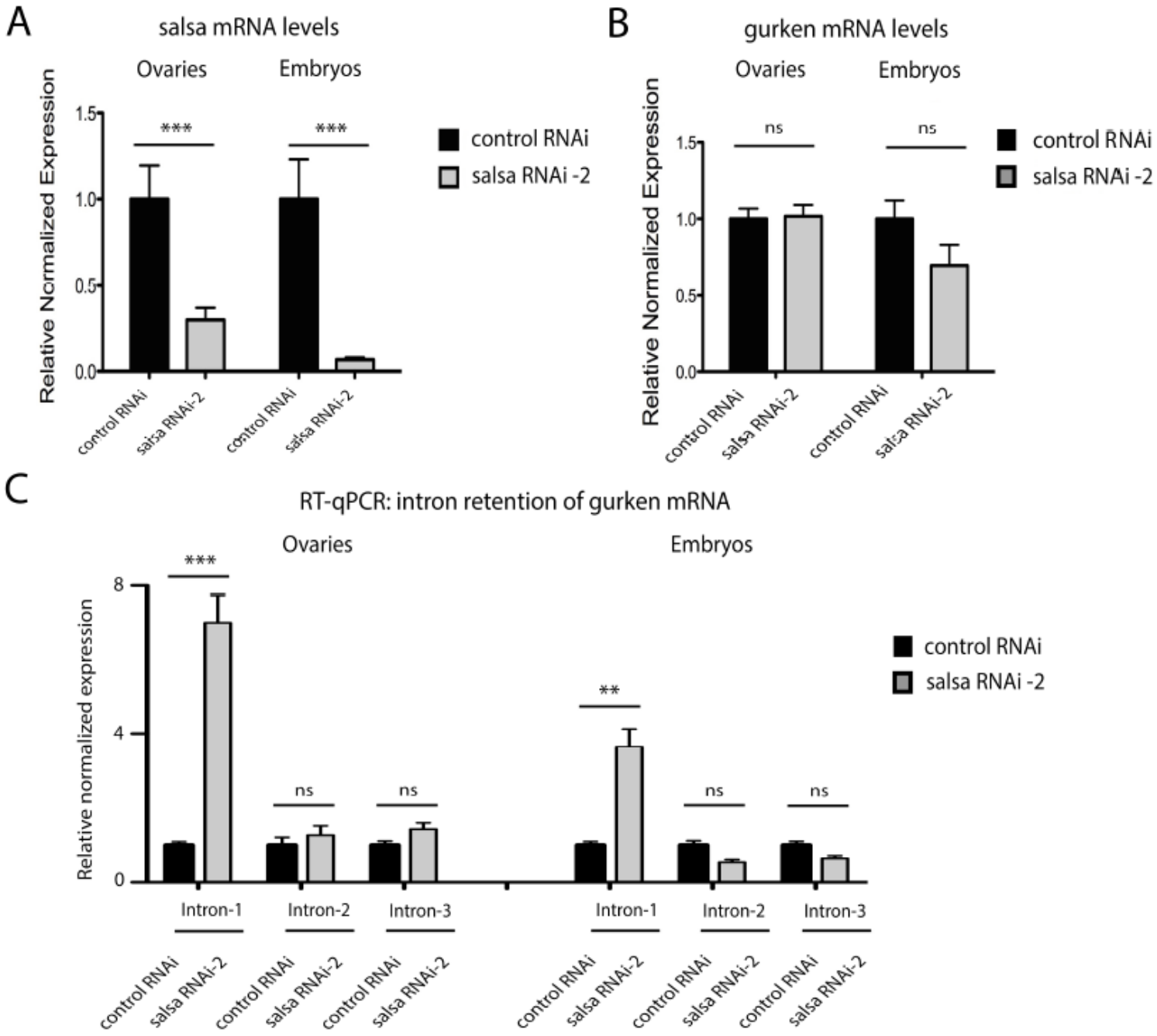
153. Labbé, R.M., Irimia, M., Currie, K.W., Lin, A., Zhu, S.J., Brown, D.D.R., Ross, E.J., Voisin, V., Bader, G.D., Blencowe, B.J., *et al.* (2012). A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. Stem Cells *30*, 1734–1745. Available at: https://www.ncbi.nlm.nih.gov/pubmed/22696458.

154. Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. EMBO J. *22*, 3960–3970. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12881430.

# 13 Annexes

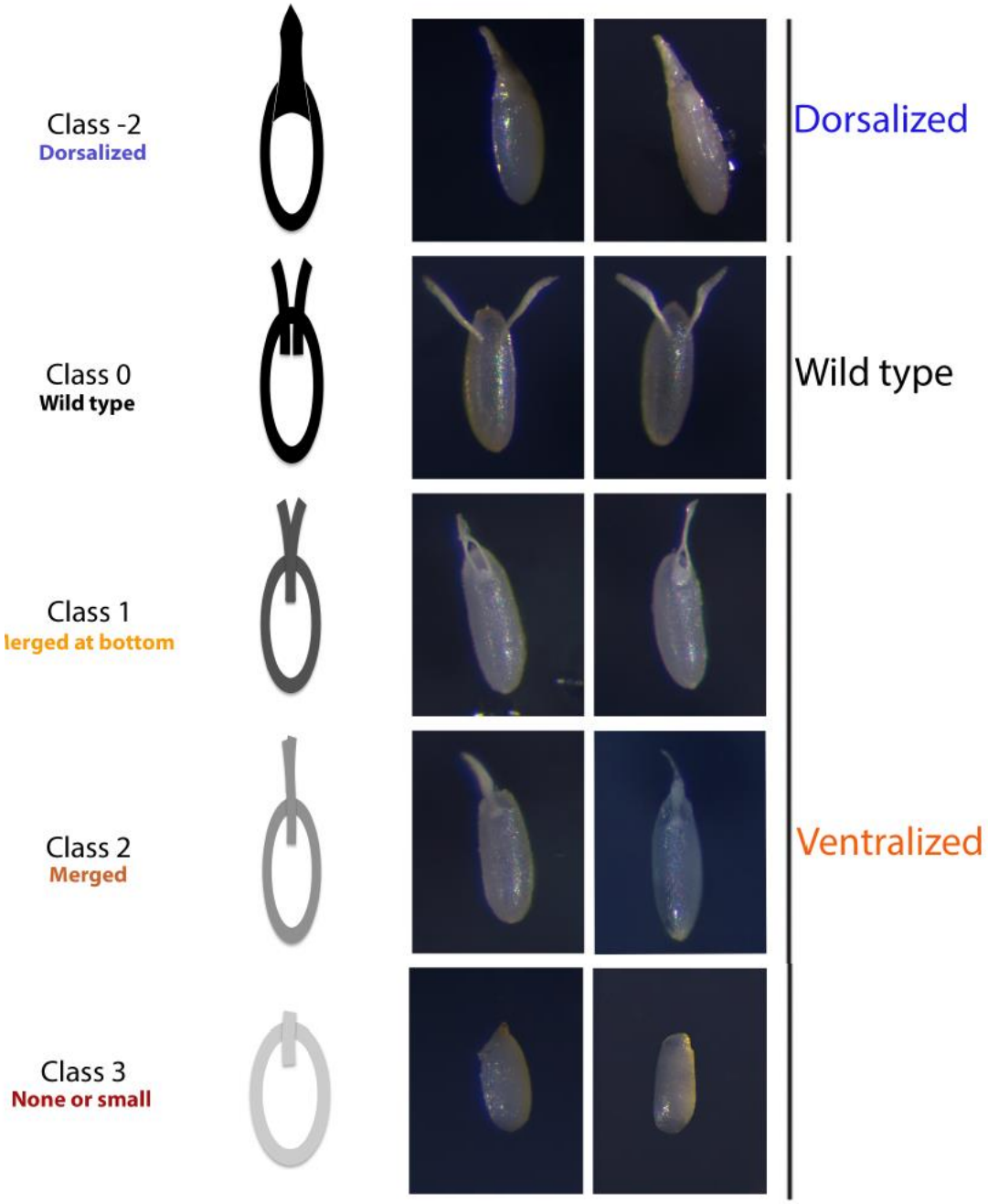



***Supplementary figure 1 - Drosophila Salsa physically interacts with spliceosome NineTeen Complex (NTC).*** *(A) Endogenous NTC proteins Fandango and Prp19 were efficiently co-immunoprecipitated by Salsa-Myc in an RNA-independent manner. Co-immunoprecipitation from total protein extracts from Drosophila Schneider 2 (S2) cells expressing a Myc-tagged Salsa (Salsa-Myc; N-terminal tag) and using dynabeads coated with an anti-Myc antibody. Protein extracts from Drosophila S2 cells transfected with an empty plasmid were used as negative control. Lysate were immunoprecipitated with (+) or without (-) RNAse treatment.*

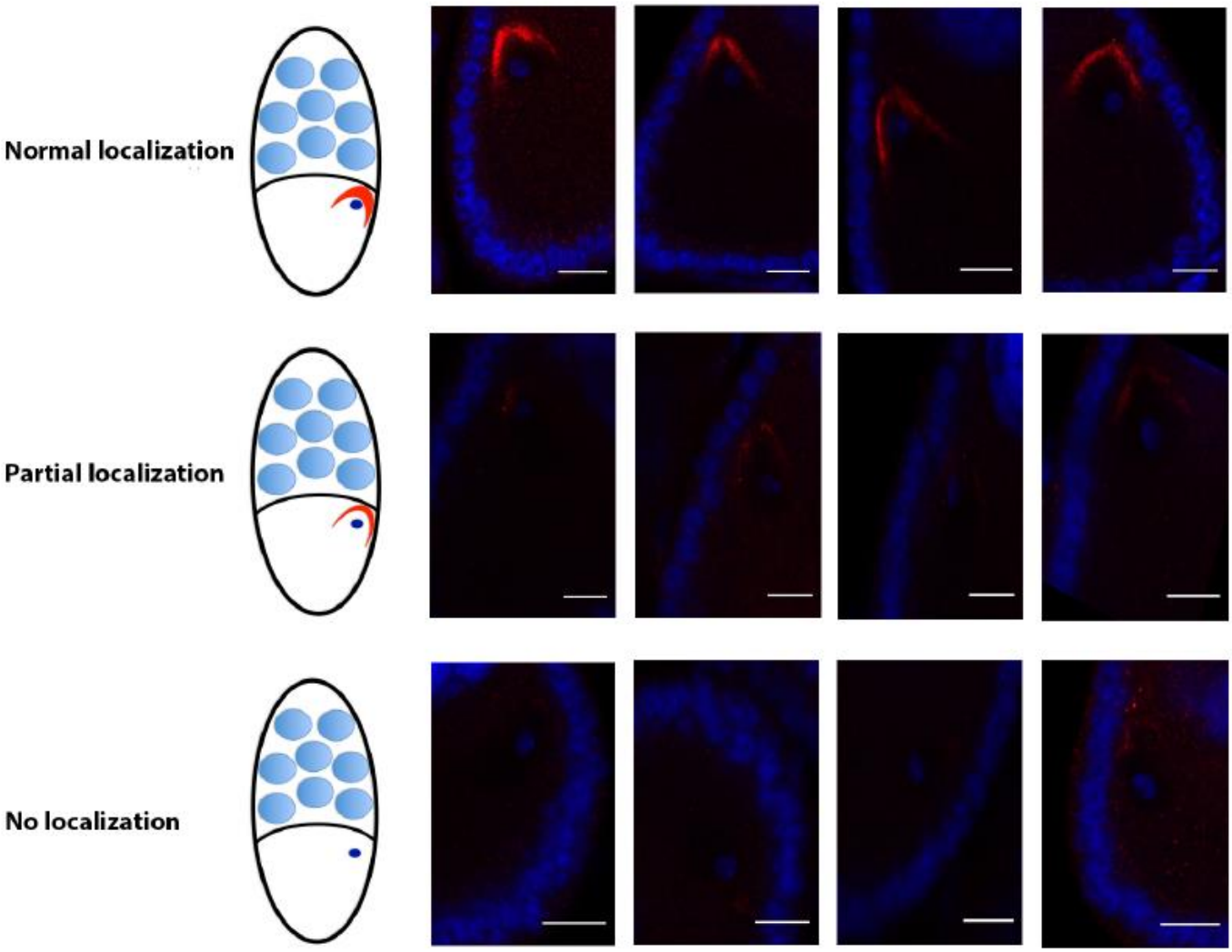



***Supplementary figure 2 - Depletion of Salsa impairs splicing of the first intron of gurken mRNA without eliciting a significant mRNA degradation.*** *(A) RT-qPCR analysis showed a significant reduction of salsa mRNA levels in ovaries and in embryos depleted for Salsa (salsa RNAi-2) when compared to control conditions (mCherry RNAi). (B) RT-qPCR analysis showed no significant reduction of gurken mRNA levels in ovaries and in embryos depleted for Salsa (salsa RNAi-2). Ovaries results are the same shown in Fig.14C. (C) RT-qPCR analysis using intron-exon primers showed a significant retention of the first intron of gurken mRNA, but not of the second and third introns, in ovaries and in*

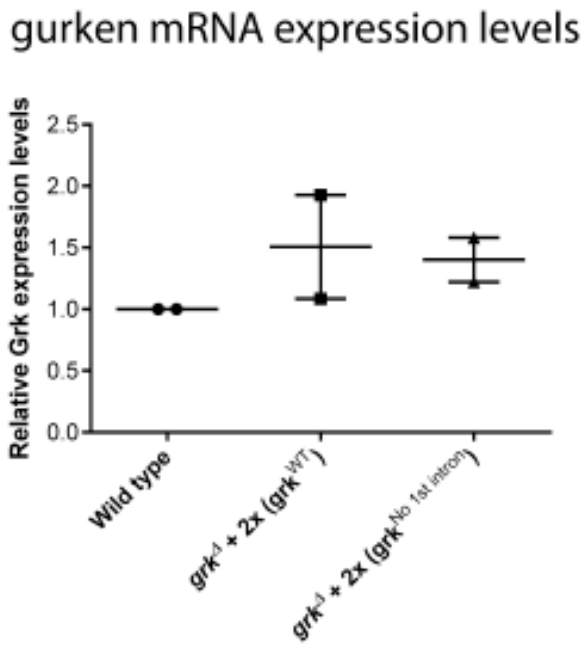
*unfertilized eggs/ embryos depleted for Salsa (salsa RNAi-2). For all panels, relative normalized expression corresponds to values normalized with two distinct reference genes (β-actin and GAPDH) and relative to negative control (mCherry RNAi). At least three biological replicates were used for ovaries, whereas two biological replicates were used for embryos. mCherry RNAi was used as a control RNAi from all panels. Error bars indicate standard deviation.*



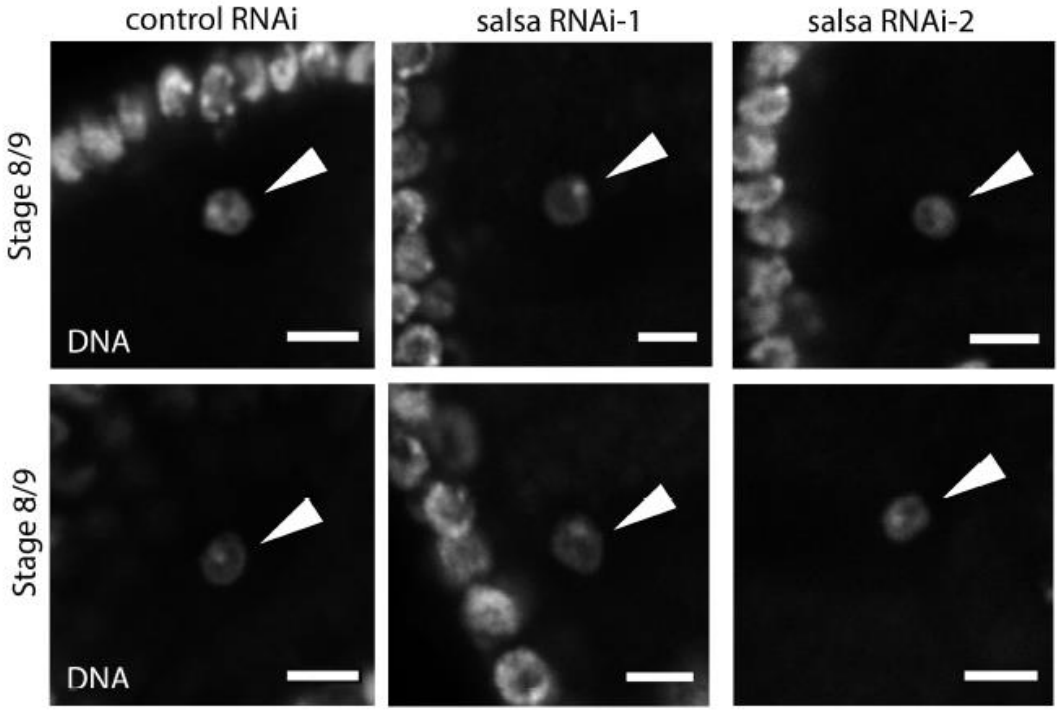
***Supplementary figure 3 - Classes of eggshell dorsal appendages defects.*** *Dorsal-ventral patterning defects of eggshells were categorized in four phenotypic classes: class 0: wild type appendages - two individualized dorsal appendages; class 1: appendages fused at their bottom; class 2: appendages totally fused - spindle phenotype); and class 3: short eggs without or with extremely short dorsal appendages.*



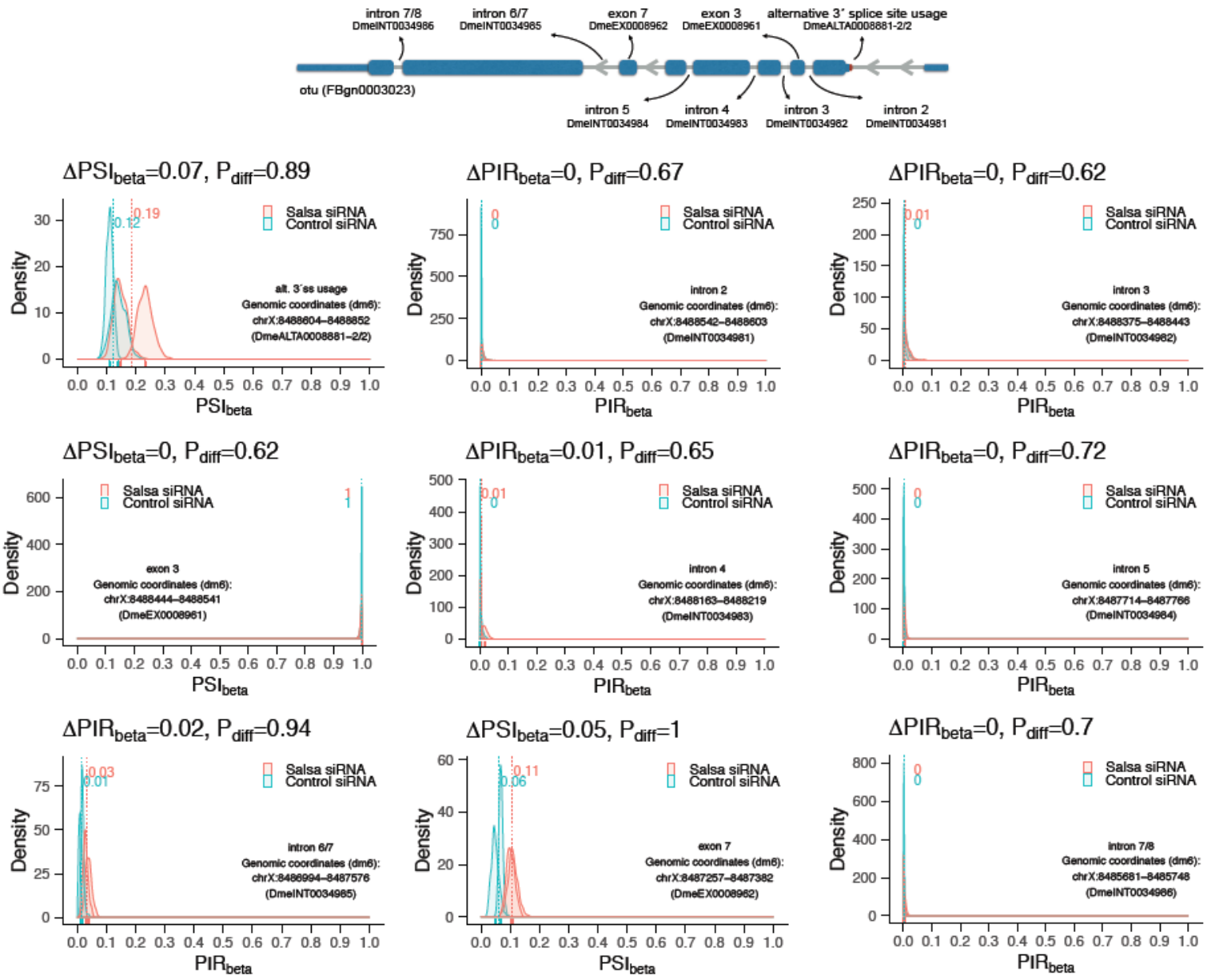

***Supplementary figure 4 - Classes of gurken mRNA dorsal-anterior localization defects.*** *Dorsal-anterior localization of gurken mRNA in stage 8/9 egg chambers was categorized in three phenotypic classes: "normal localization", "partial localization" and "no localization". DNA (Blue) and gurken mRNA (red). Scale bar 10µm.*




***Supplementary figure 5 - Deletion of the first intron does not impair total expression levels of gurken mRNA.*** *Genomic constructs of gurken, with its own endogenous gurken promoter and integrated in the same attP-site, but with (grk$_{WT}$) or without the transcript first intron (grk$_{no\ 1st\ intron}$), showed identical expression levels of gurken mRNA. Genomic constructs of gurken (two copies) were in grk$_{deltaFRT}$ mutant background. Real-time qPCR analysis of gurken mRNA levels was done using primers for the first exon of gurken transcript. Relative normalized expression corresponds to values normalized with two distinct reference genes (β-actin and GAPDH) and relative to wild type. Two biological replicates were used for all datasets. Error bars indicate standard deviation.*

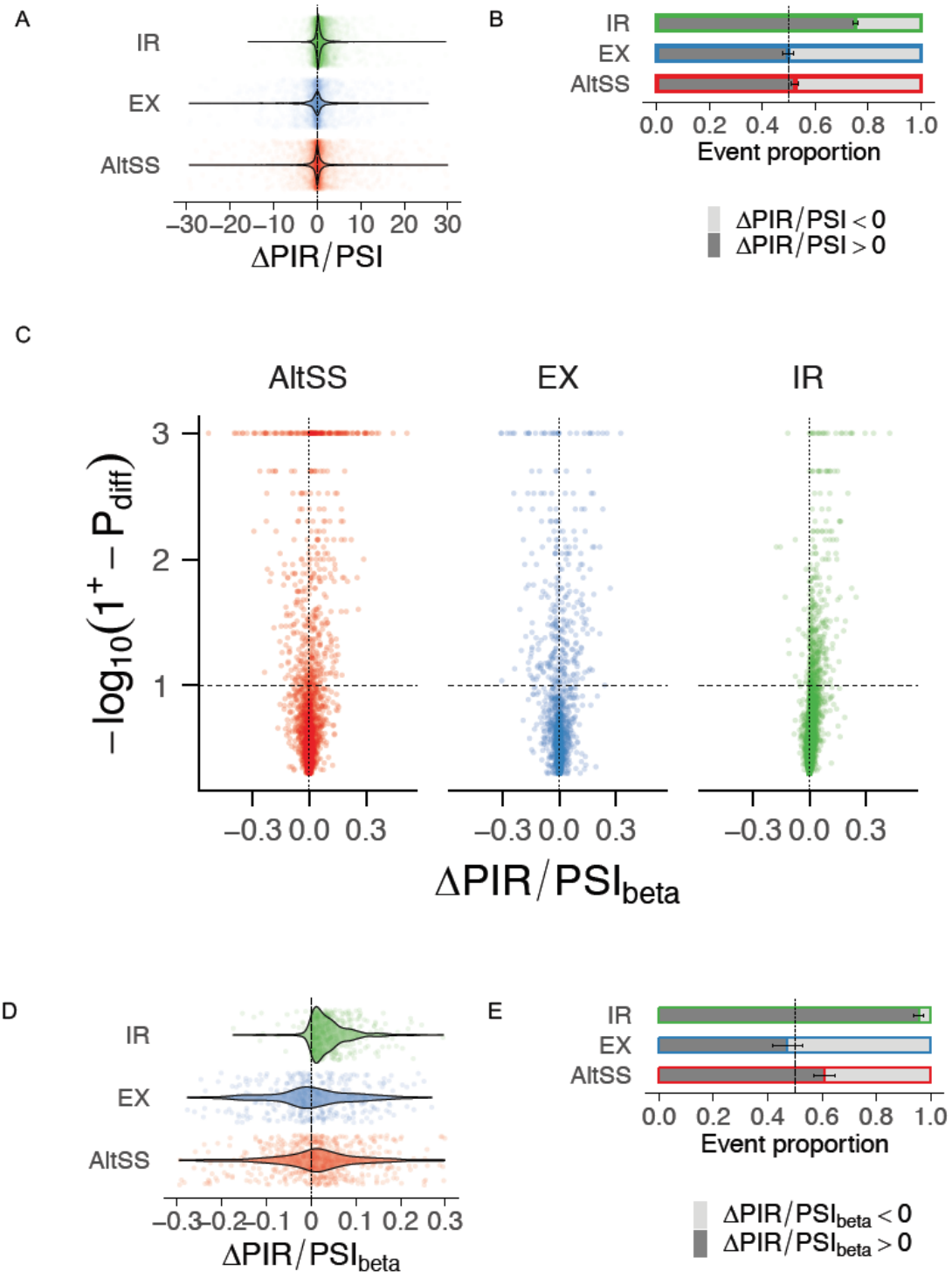



***Supplementary figure 6 - Depletion of Salsa does not impair the normal morphology of the oocyte karyosome (chromosomes).*** *During mid-oogenesis the oocyte chromatin is condensed into a structure called karyosome (arrowhead). Defects in karyosome morphology are indicative of the accumulation of DNA damage and abnormal activation of the meiotic checkpoint* [115]*. Germ line depletion of Salsa (salsa RNAi-2) did no impair the correct morphology of the oocyte karyosome. mCherry RNAi was used as a negative control. DNA was visualized with DAPI staining (grey). Arrowheads indicate the oocyte karyosome. Scale bars 5 µm.*



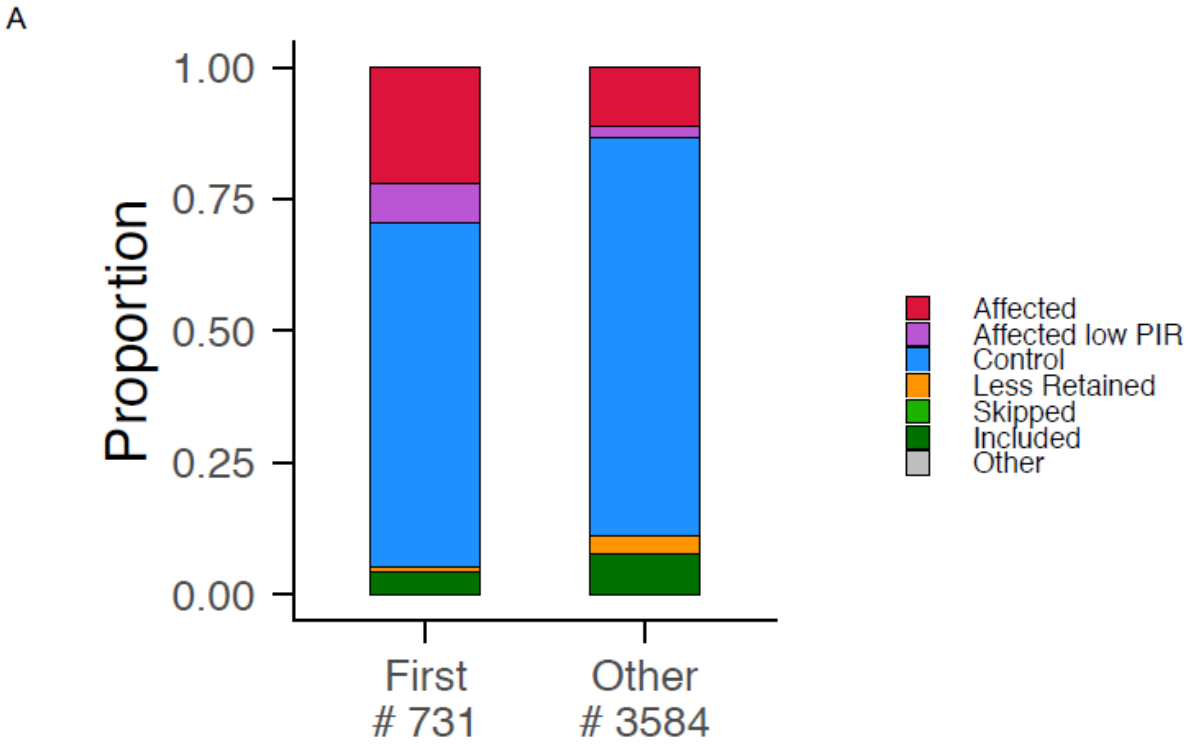

***Supplementary figure 7 - Depletion of Salsa does not impair alternative splicing of ovarian tumor (otu).*** *Diagram of simplified otu gene structure with exons (blue) and introns (gray) and density plots (smoothed histograms) reflecting points emitted from beta distributions used to model inclusion levels for alternative splicing events in otu gene. For each sample, density plots reflect PIRbeta estimates. Mild differences in inclusion levels could be detected in some cases between control (mCherry RNAi) and Salsa-depleted (salsa RNAi-1), yet none was significant under our criteria for differential splicing (Pdiff ≥ 0.9 and |ΔPIRbeta| > 0.10). Dashed vertical lines indicate PIRbeta estimates for Salsa-depleted and control conditions and rug plots below the density curves reflect PIR*

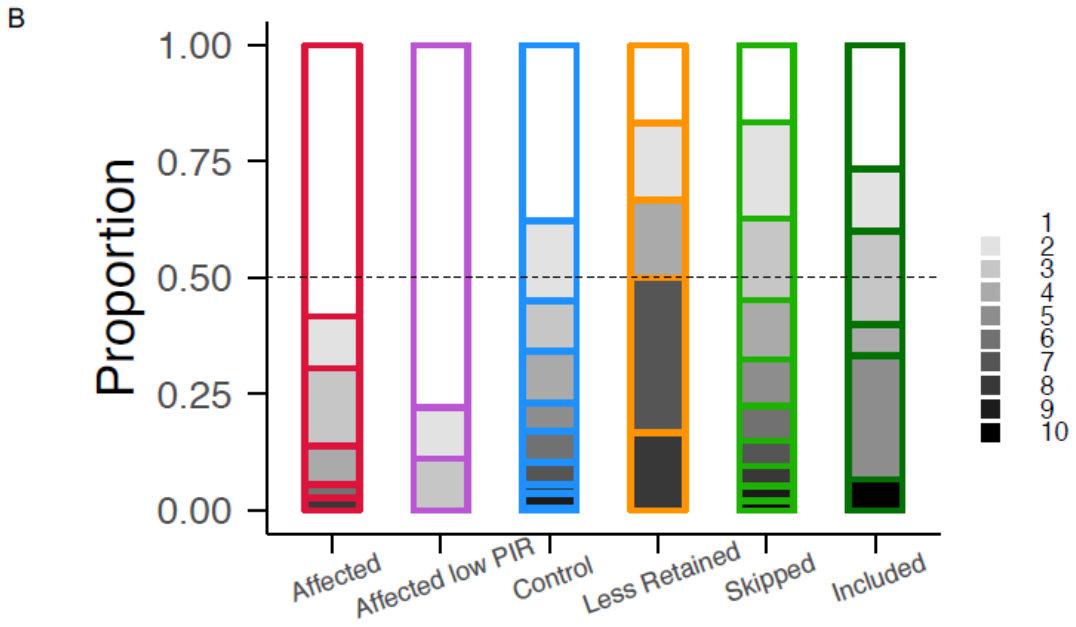
*values per sample. Event identifiers and genomic coordinates of the respective alternative sequences are part of the VAST-DB dm6 annotation* [121].



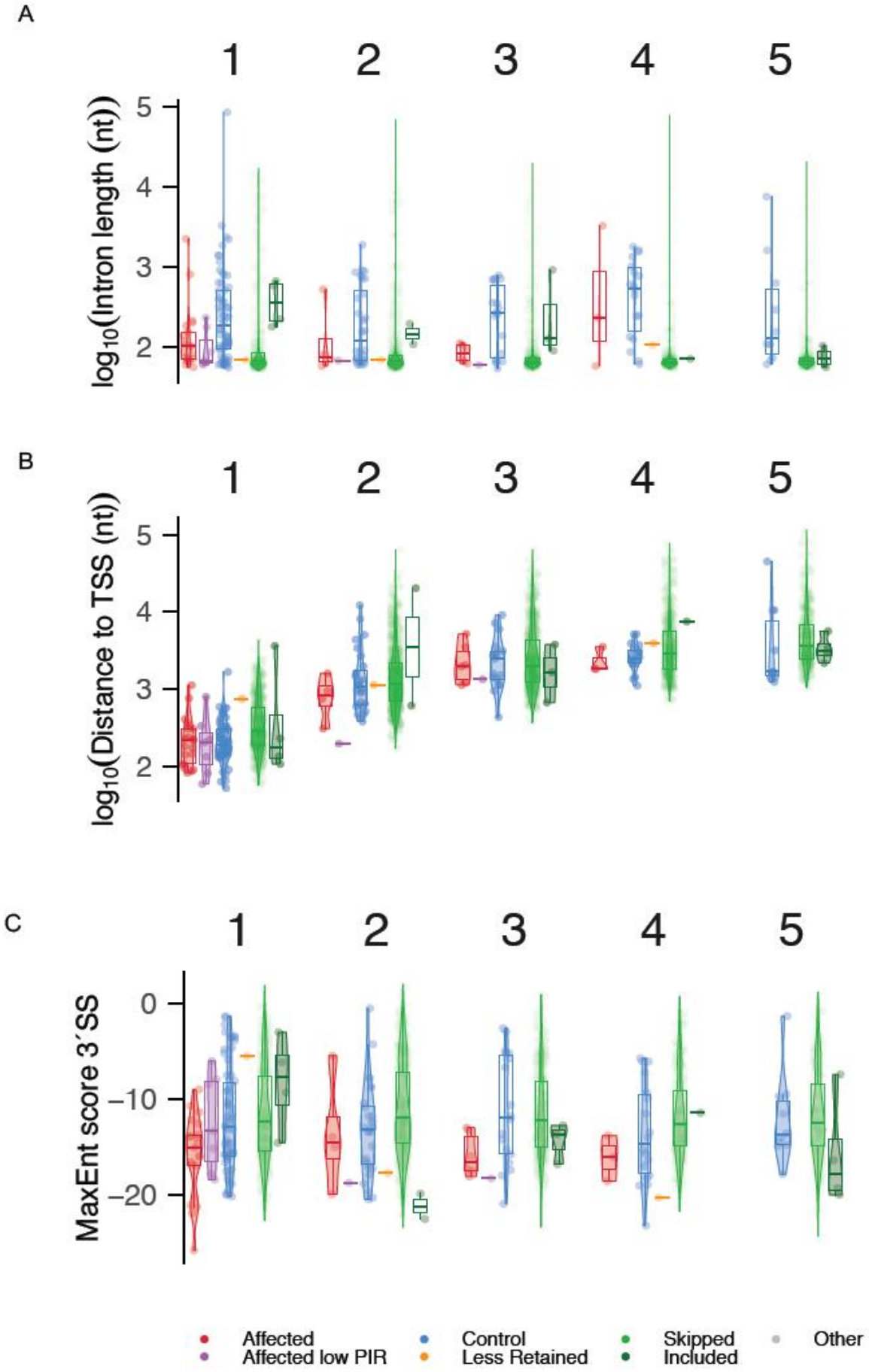

***Supplementary figure 8 - The most common form of alternative splicing defects after Salsa depletion is increased levels of intron retention (IR).*** *(A) Violin plot of distribution and (B) proportion of positive and negative differences*

*of mean PIR for intron retention (IR, 14504 events) and PSI (percent spliced-in* [152] *for exon skipping (EX, 13627 events) and alternative splice site usage (AltSS, 8385 events) values in Salsa-depleted (salsa RNAi-1) samples compared to control samples (mCherry RNAi) per type of alternative splicing. **Δ**PIR/PSI = PIR/PSIsalsa RNAi – PIR/PSIcontrol. (C) Volcano plot for beta distribution-based differential alternative splicing analysis per type of alternative splicing, with differences in inclusion levels between Salsa-depleted and control samples (**Δ**PIR/PSIbeta = PIR/PSIsalsa RNAi – PIR/PSIcontrol) in the x-axis and probability of differential splicing (-log10(1.001 - Pdiff)) in the y-axis, with 1+ = 1.001 used to avoid infinite values when Pdiff = 1. (D) Violin plot of distribution and (E) proportion of positive and negative differences of beta distribution-based PIR/PSIbeta values per type of alternative splicing. (B and E) Error bars in proportion bar plots reflect 95% confidence intervals of a proportion test against the null hypothesis that the proportion of positive differences is 0.5 (dashed vertical line).*

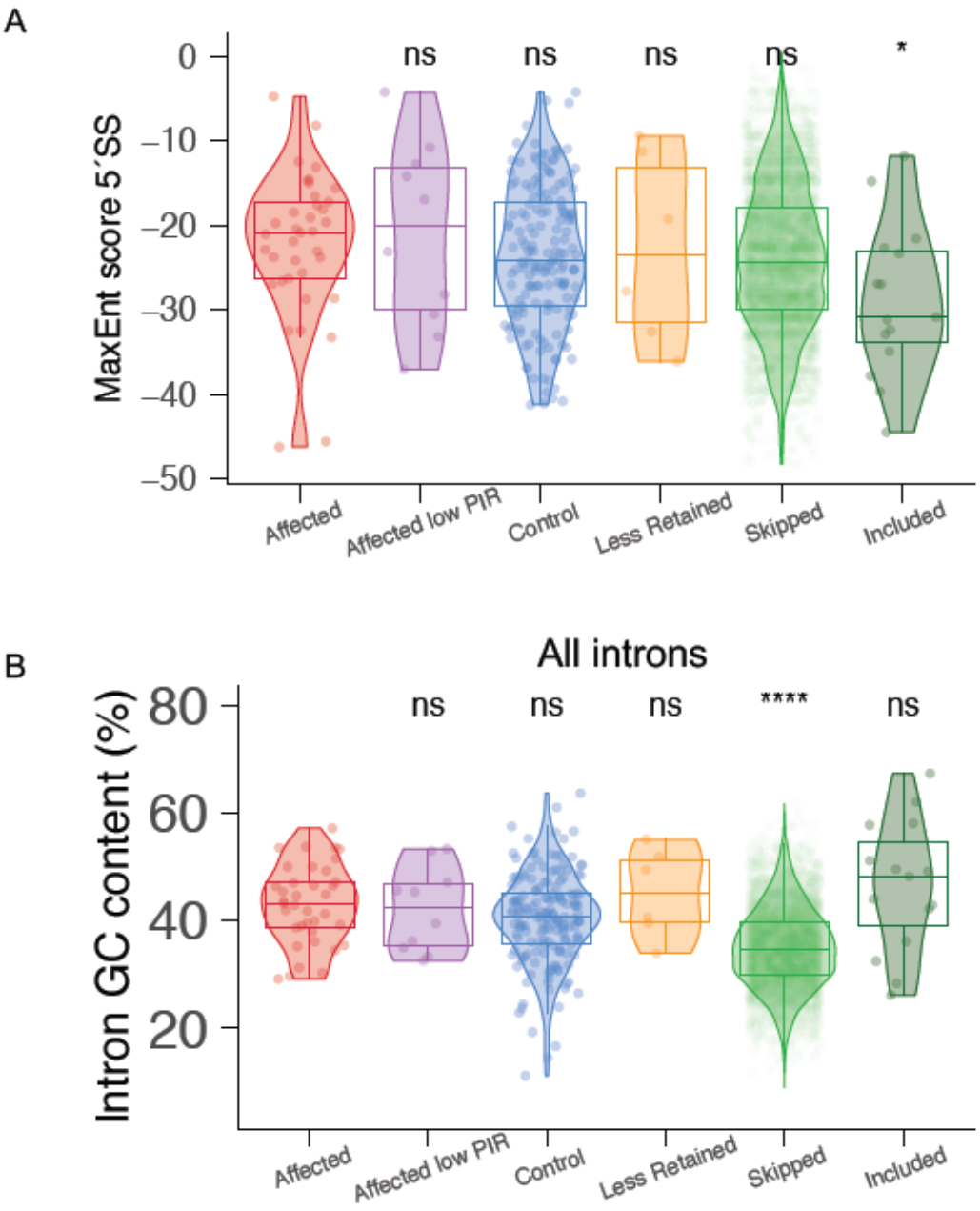


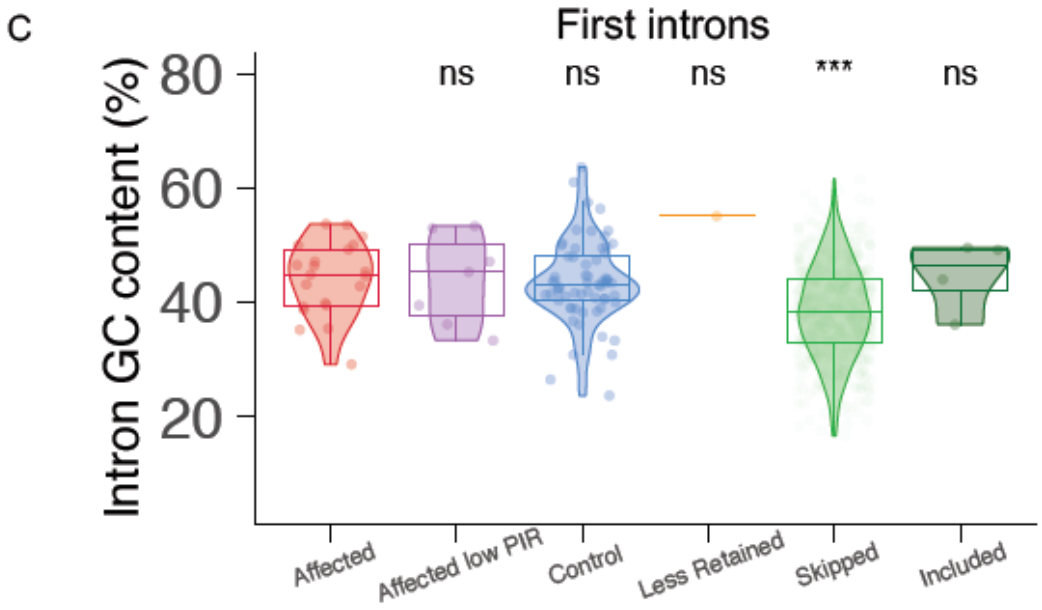



***Supplementary figure 9 - Salsa is particularly rate-limiting for splicing of small first introns.*** *(A) Proportion of first (left) and all other introns (right) that are part of each of the differential retention classes defined, showing increased proportion of Affected introns amongst first introns. (B) Proportion of first (white) and increasing rank (gradient of gray shades) of introns within each intron class, considering the subset on introns up to the 10th position in the transcript.*

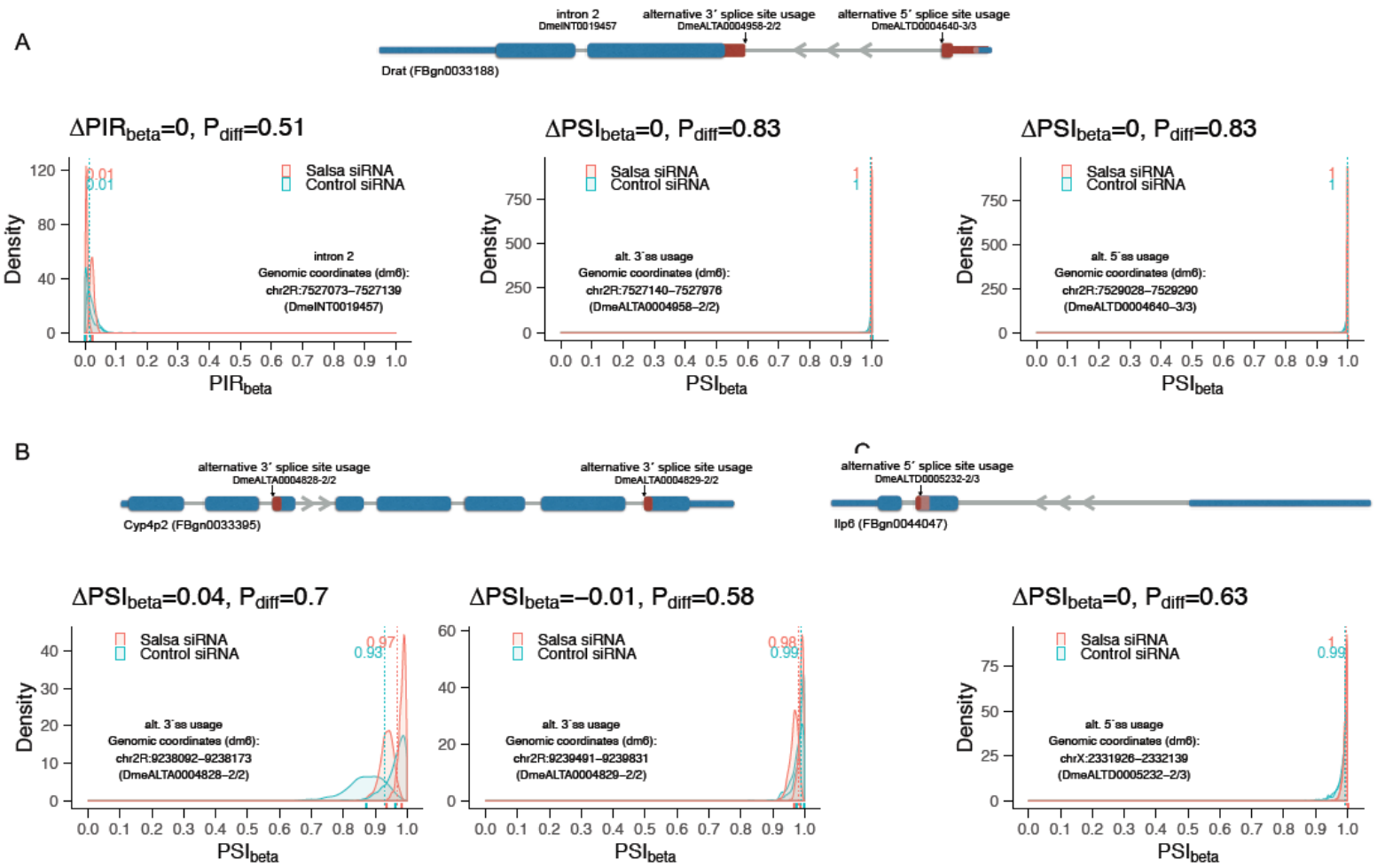

A
1
2
3
4
5
log10(Intron length (nt))
5
4
3
2
B
log10(Distance to TSS (nt))
C
MaxEnt score 3´SS
0
−10
−20
Affected
Affected low PIR
Control
Less Retained
Skipped
Included
Other

***Supplementary figure 10 - Intronic features per intron position when splicing is affected by Salsa depletion.*** *Violin plots and boxplots (highlighting the median, 25th and 75th percentiles per intron class) summarizing distributions of observations (each point refers to one intron) considering, from left to right, 1st to 5th introns of: (A) log10 of intron length (nt); (B) log10 of distance (nt) between intron start and the transcript transcription start site (TSS); (C) Maximum entropy (MaxEnt) scores* [136] *for estimation of 3´ splice site efficiency, obtained with MATT* [137].

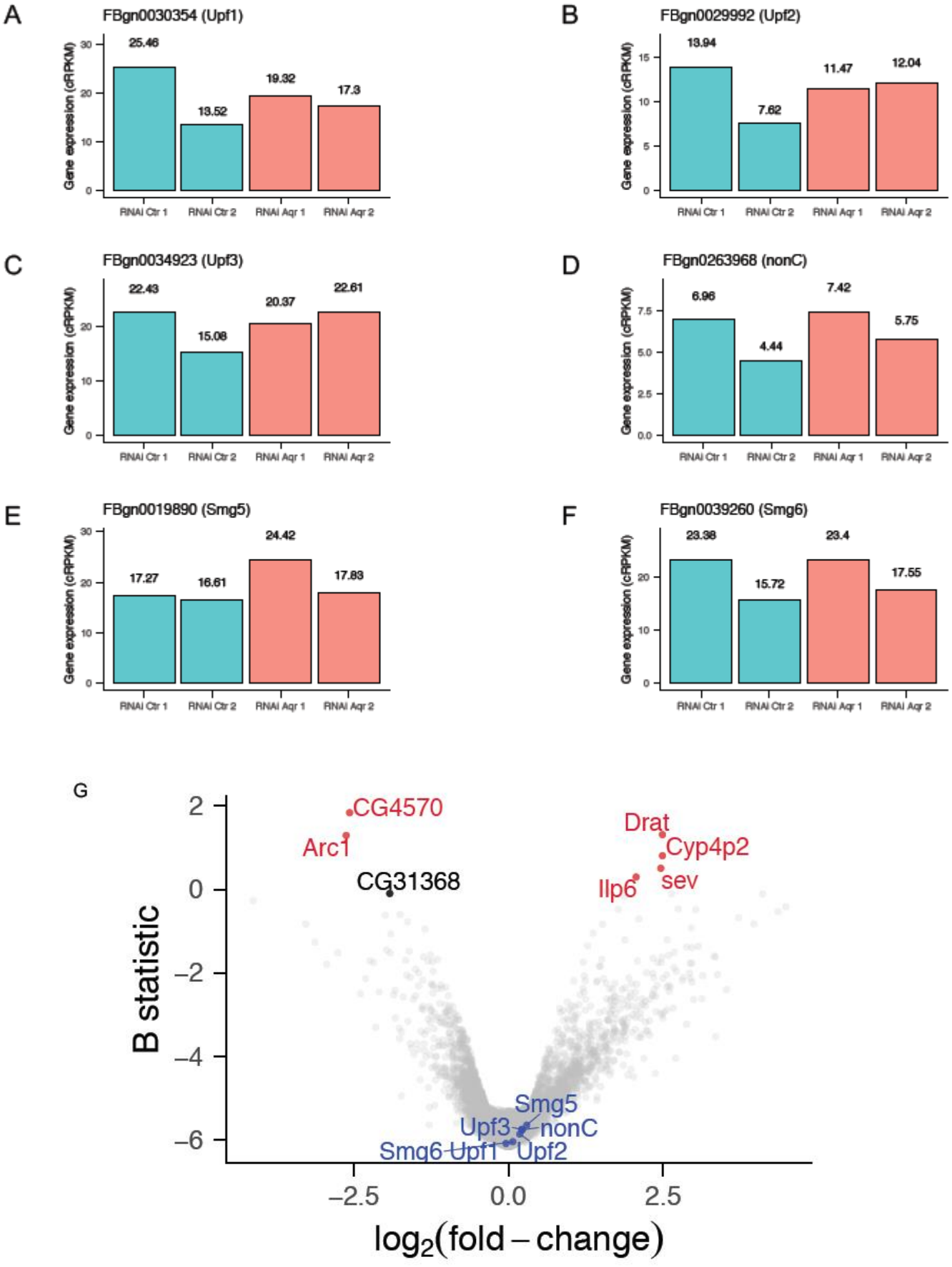





***Supplementary figure 11 - No obvious 5´splice site and GC content bias in introns affected by Salsa depletion.*** *(A to C) Violin plots and boxplots (highlighting the median, 25th and 75th percentiles per intron class) summarizing distributions of observations (each point refers to one intron) for (A) Maximum entropy (MaxEnt) scores for estimation of 5´ splice site efficiency, based on* [136] *and % GC content for (B) all and (C) first introns. Comparison of metrics was performed between all intron classes and Affected: * p-value < 0.05 and ns non-significant (Wilcoxon rank-sum test with continuity correction). Intronic features obtained with MATT* [137].




***Supplementary figure 12 - Genes whose expression levels were affected by Salsa depletion show no detectable changes in alternative splicing.*** *Diagram of simplified gene structures with exons (blue with alternative splice sites highlighted in red/pink), introns (gray) and density plots (smoothed histograms) reflecting points emitted from beta distributions used to model inclusion levels for each sample and PIRbeta estimates for Salsa-depleted (salsa RNAi-1) and control conditions (mCherry RNAi) for: (A) Drat intron 2 retention (left panel) and alternative usage of a 3´ (middle) and a 5´ splice site (right panel); (B) Cyp4p2 two alternative 3´ splice site usage events ; (C) Ilp6 alternative 5´ splice site usage Dashed vertical lines indicate PIRbeta estimates for salsa-depleted and control conditions and rug plots below the density curves reflect PIR values per sample. Event identifiers and genomic coordinates of the respective alternative sequences are part of the VAST-DB dm6 annotation* [121].


A
FBgn0030354 (Upf1)
25.46
13.52
19.32
17.3
B
FBgn0029992 (Upf2)
13.94
7.62
11.47
12.04
C
FBgn0034923 (Upf3)
22.43
15.08
20.37
22.61
D
FBgn0263968 (nonC)
6.96
4.44
7.42
5.75
E
FBgn0019890 (Smg5)
17.27
16.61
24.42
17.83
F
FBgn0039260 (Smg6)
23.38
15.72
23.4
17.55
Gene expression (cRPKM)
RNAi Ctr 1
RNAi Ctr 2
RNAi Aqr 1
RNAi Aqr 2
G
CG4570
Arc1
CG31368
Drat
Cyp4p2
sev
Ilp6
Smg5
Upf3
nonC
Smg6
Upf1
Upf2
B statistic
log2(fold – change)

***Supplementary figure 13 - Nonsense-mediated decay pathway genes show no change in expression after Salsa depletion.*** *(A to F) Barplots with expression levels of nonsense-mediated decay genes in corrected-for-mappability reads per kilobase and million mapped reads (cRPKM,* [153]*). (G) Volcano plot for differential gene expression analysis upon Salsa depletion (salsa RNAi-1) with nonsense-mediated decay genes Upf1 (FBgn0030354), Upf2 (FBgn0029992), Upf3 (FBgn0034923), Smg5 (FBgn0019890), Smg1/nonC (FBgn0263968) and Smg6 (FBgn0039260)* [154] *highlighted. Each point indicates one gene considered in the analysis. Significantly differentially expressed genes (B statistic > 0) are highlighted in red, while salsa (CG31368) is highlighted in black as a positive control.*

***Supplementary table 1 - 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 |

***Supplementary table 2 - Primer efficiency and regression curve (for 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-RB | 108.6 | 0.99 |
| 10 | eIF4E-RC | 109.7 | 0.99 |

***Supplementary table 3 - Differentially retained introns after Salsa depletion.*** *Results from beta distribution-based differential intron retention analysis. Intron description (FlyBase gene name, vast-tools' event identifier and Ensembl transcript ID, as well as intron's dm6 genomic coordinates, strand, length and order in transcript) followed by beta distribution-based inclusion levels for each sample, $PIR_{beta}$ for the two conditions (Salsa depletion (salsa RNAi-1) and control samples (mCherry RNAi)), the difference between these and the probability of differential intron retention. Final columns refer to the predicted impact of intron retention in the protein coding and the class the intron was assigned to following this analysis.*

***Supplementary table 4 - Differentially expressed genes after Salsa depletion.*** *Results from differential gene expression analysis performed with the*

*limma R package* [151]*. FlyBase gene identifiers (ID and name) are followed by the summary statistics of genes considered significantly differentially expressed: average expression across samples in log2-counts per million (logCPM), log2 fold-change in expression between Salsa depletion (salsa RNAi-1) and control conditions (mCherry RNAi), moderated t-statistic and respective nominal p-value and Benjamini-Hochberg-adjusted p-value, and empirical Bayes statistic (B statistic or log-odds ratio of the gene being differentially expressed).*