A STUDY OF THE FUNCTION OF TWO NUCLEAR TRANSPORT RETROGENES (*DNTF-2R* AND *RAN-LIKE*) IN *DROSOPHILA MELANOGASTER*

by

SUSANA FILIPA SIMÕES LOPES DOMINGUES

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2014

ACKNOWLEDGEMENTS

I specially want to thank my committee members. I would like to start by expressing my special appreciation to my supervisor Dr. Esther Betrán. She is an amazing mentor and it has been an honor to work with her. Thank you for encouraging my research and for allowing me to grow as a scientist. The joy and enthusiasm you show for research are contagious and motivating, even when everything seems to fall apart. I also want to thank Dr. Cédric Feschotte, Dr. Jeff Demuth, Dr. Maeli Melotto, and Dr. Shawn Christensen for their guidance and help.

Thanks go also to everybody in the Betrán lab. Xavier for helping in the experiments and for keeping everything running. Anya, thank you for your help, ideas and support. Diwash, you are an amazing friend! Thank you for your support and company.

I also want to show gratitude to the staff and students in the Biology Department. Thank you Linda, Gloria, Paulette and Sherri for all the help you provided in these 6 years. A special thank you to Claudia and Catherine who have given me great friendship.

Mariana, Sofia, Tania, and Veronica thank you for aiding me even from far away.

Last but not least, I want to thank my parents for their love, support and encouragement. Finally thank you to my amazing wife for loving me everyday, and giving me the support I always need, encouraging me to continue even when I have no energy, and being patient mainly during this final period of my Ph.D. I know it wasn't always easy!

This dissertation work, and my Graduate Research Assistantship for 3 years were supported by two NIH grants to E.B. I got tuition funds from the Portuguese Science Foundation fellowship to S.F.S.L.D. and Phi Sigma Graduate Honors Society and the Biology Department provided travel funds. I also received funds from UT Arlington for the participation in the I ENGAGE program for undergraduate research instruction.

December 1, 2014

ABSTRACT

A STUDY OF THE FUNCTION OF TWO NUCLEAR TRANSPORT RETROGENES (*DNTF-2R* AND *RAN-LIKE*) IN *DROSOPHILA MELANOGASTER*

Susana Filipa Simões Lopes Domingues, Ph.D
The University of Texas at Arlington, 2014

Supervising Professor: Esther Betrán

Gene duplications are a class of mutations that contribute to the functional diversity within and between species. The analysis of gene duplications in fully sequenced genomes has revealed that they often evolve fast and acquire new functions. In Drosophila, it has been shown that many new duplicates have testes function. What the selective pressures are that ensure the preservation of these testes duplications and the exact function these genes are recruited to fulfill are still unknown. Here, I study two gene families of nuclear transport proteins in Drosophila, *Dntf-2* and *Ran* families, which have recurrently given rise to testes duplicates and are evolving under positive selection.

I start by providing an update on the number of duplications of *Dntf-2* and *Ran* genes in Drosophila using newly sequenced genomes. The analyses suggest a minimum of 6 duplications of *Dntf-2* and 10 duplications of *Ran* in Drosophila. Both, DNA- and RNA-mediated duplications were considered. Interestingly, more than 80% of the duplications occurred by means of an RNA intermediate despite the fact that RNA-mediated duplications make up a small fraction (~3%) of the duplicates in fly genomes.

RNA-mediated duplications do not include the regulatory regions of parental genes. In Chapter 3, I study the origin of the regulatory regions of two retrogenes in *D. melanogaster* (*Dntf-2r* and *Ran-like*). Interestingly, in the case of *Dntf-2r*, I demonstrate that the regulatory region was present in the genome before the retrogene insertion.

To elucidate the reasons for the recurrent duplication of *Dntf-2* and *Ran*, I have studied the function of *Dntf-2r* and *Ran-like* in *D. melanogaster*. I have gained a great understanding of the functions of these genes in spermatogenesis and revealed that they are also involved in somatic functions. Using fluorescent protein fusions, I show that both genes appear to have an array of functions from spermatocytes onwards in nuclear transport, nuclear membrane assembly, microtubule organization during chromosomal segregation, and during sperm head and tail elongation and individualization. The loss of *Dntf-2r* or *Ran-like* triggered by RNAi in male germline reveals strong effects in male fertility and in the soma leads to lethality. These new data reveal that, contrary to what I initially hypothesized, these retrogenes appear to have many functions in spermatogenesis and important functions in somatic cells. While it was initially hypothesized that *Dntf-2* and *Ran* genes duplicate recurrently to suppress selfish elements (i.e., transposable elements and meiotic drive systems) in male germline, the myriad of functions that these retrogenes currently have does not support this hypothesis.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

BACKGROUND

1.1. Genome novelty

Only a fraction of the DNA changes in the genomes contribute to changes in the individual's phenotype and adaptation (Biemont 2008; Hou and Li 2009). One kind of genomic change likely to have phenotypic consequences is the emergence of new gene. Several mechanisms can be responsible for the origin of new genes in the genomes. Gene duplication (Ohno 1970; Esnault, et al. 2000), protein domestication from transposable elements or viruses (Feschotte and Pritham 2007; Sinzelle, et al. 2009), *de novo* gene formation from non-coding regions in the genome (Levine, et al. 2006), horizontal gene transfer (Kurland 2000) and reorganization of existing genes (Hare, et al. 2008) are all different ways in which a new gene can emerge in a genome. Below, I will concentrate on gene duplication and its consequences because this is the mechanism that gave rise to the genes that are the focus of this study.

1.1.1. Gene duplication

Gene duplication is the process that results in the production of an additional copy of a gene. As reviewed and evaluated 44 years ago now, gene duplication is one of the major sources of genetic novelty (Ohno 1970). Gene duplication is a kind of mutation and, as such, it is a chance event and happens at random with respect to its fitness effects. Gene duplication can be fixed or removed from a population as a result of random genetic drift or the action of natural selection that selects against or in favor of a new mutation (Graur and Li 2000). The duplication of a gene can take place in several

ways: whole genome duplication, duplication of a single chromosome or part of a chromosome and duplication of a single gene or group of genes. In the evolutionary history of all organisms, gene duplicates have always been present and repeatedly given rise to new functions. They work as raw material for evolutionary forces to create evolutionary novelty (Ohno 1970). Gene duplication is, initially, often redundant and potentially expendable or even deleterious, a situation that frequently leads to the formation of a pseudogene and the eventual exclusion of this pseudogene from the genome (Lynch and Conery 2000). In the cases where the gene copy is fixed in the population, usually the assumption is that the duplicate fixes by drift (Ohno 1970), there is an additional copy that is free to evolve under different selective constraints than the parental gene. As mentioned above, new genes often acquire disablements becoming pseudogenes. However, under certain circumstances, new genes can be preserved (Wagner 2001). The preservation of a gene duplicate can occur by a wide range of mechanisms that I discuss below.

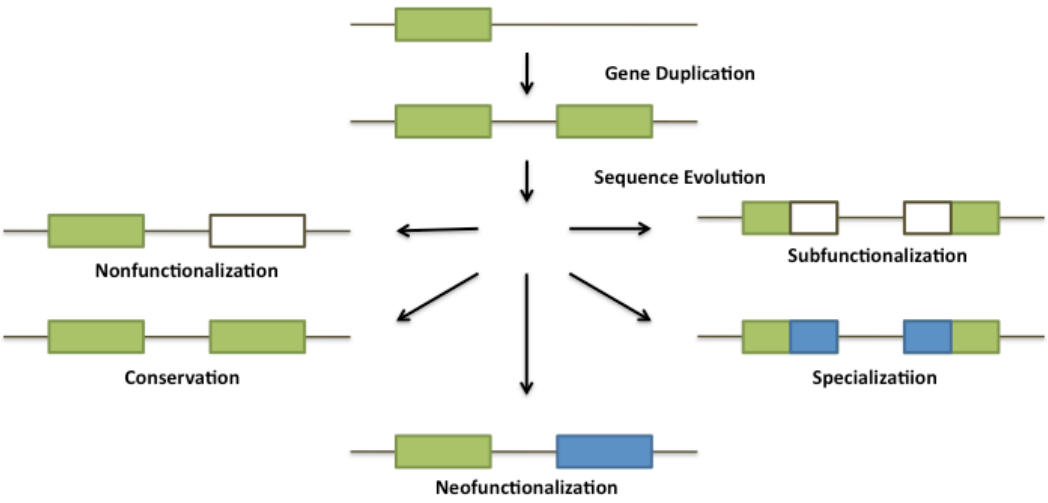


Figure 1.1 – Five potential fates of gene duplications.

In some instances, a new gene might stay functional in the genome, if it provides a beneficial extra dose of a gene product (Ohno 1970). Alternatively, it might undergo either a subfunctionalization or a neofunctionalization process wherein it can either retain some of the functions of the parental gene or evolve a new function altogether (Figure 1.1). During subfunctionalization there is a partition of the parental gene functions between the duplicated genes through complementary neutral disabling mutations in each gene copy. One copy becomes fixed for a neutral mutation that eliminates an essential subfunction, permanently preserving the second copy. The loss of alternate subfunction in the second copy reciprocally preserves the first copy. On the other hand, in the process of neofunctionalization, one gene copy retains the ancestral functions while the other acquires a novel function. While the figure above illustrates subfunctionalization and neofuntionalization for the protein-coding regions, both processes can also occur for regulatory regions (Ohno 1970; Force, et al. 1999; Lynch and Conery 2000; Prince and Pickett 2002; Chain and Evans 2006). It has often been proposed that for a new function to arise, the gene needs to go through a period of relaxation of selection, as the parental copy performs the function and the new copy is redundant. During this stage of relaxed selection, fixation of previously forbidden mutations can change the function of the new gene (Ohno 1970). During this period the DNA substitutions will modify the gene-coding region leading to a novel amino acid sequence that could confer increased fitness to the individuals. Currently, it is also known that after duplication both or one gene can change function by specialization without relaxation (Innan and Kondrashov 2010). During neofuntionalization or specialization beneficial mutations fix under positive selection and can create a new function or optimize a previously existing function after duplication if the gene is now released from pleiotropic effects (Assis and Bachtrog 2013). If a new gene undergoes any of the above

steps, it will be maintained in the genome by purifying selection (Long, et al. 2003A). Genes can, however, continue evolving under positive selection, if they are involved in pathways implicated in arms races. Examples of these are cell surface proteins in *E. coli* such as beta barrel porins (Petersen, et al. 2007), immune proteins (Bulmer 2001) and nucleoporins in Drosophila (Presgraves and Stephan 2007)

In some instances selection for a different allele might already occur in the parental gene and if that variant duplicates it will be fixed under positive selection. There are several models of how this can occur (Innan and Kondrashov 2010). Gallach and Betrán (2011) presented a new duplicative model of how intralocus sexual antagonism can be resolved through gene duplication. The existence of intralocus sexual antagonism means that there is polymorphism for a gene because males select for one allele and females select for the other allele. It is often the case that the new allele is the one beneficial to males because is good for testes. In this model, the end result will be the fixation of a sex- and tissue-specific new gene in the population and resolution of the intralocus sexual antagonism. The steps of the resolution are: the duplication of the antagonistic allele into another chromosomal location, evolution of sex- and tissue-specific expression of the duplicate and resolution of the sexual antagonism (males and females now select for the same allele at the parental gene). They propose that this might be the reason why there are so many new testis-specific genes (i.e., they might resolve intralocus sexual antagonism caused by the fast evolution in testes).

To gain information about all the details of the process and the roles that ensure the retention of new duplicate genes, it is useful to study young duplicated genes (Long, et al. 2003B). Using this approach it has become clear recently that newly duplicate genes can acquire essential functions despite being young (Chen, et al. 2010). It is still unclear how many of those genes have undergone subfunctionalization or

neofuntionalization but at least one has been studied in some detail and revealed to have acquired a new centromeric function that made it more essential than the parental gene (Ross, et al. 2013). In this work, I am also looking at young genes to understand the steps and evolutionary pressures that gave rise to them.

There are many molecular mechanisms that can give rise to gene duplications. There can be duplications of a few base pairs or of entire genomes. DNA-mediated gene duplications are the ones that have been traditionally studied. However, new gene copies can also originate through an RNA intermediate. This process is also known as retrotransposition (Brosius 1991; Long, et al. 2003A; Kaessmann, et al. 2009; Kaessmann 2010; Mendivil Ramos and Ferrier 2012).

1.1.1.1. DNA-mediated gene duplication

Tandem duplication by non-allelic homologous recombination (also known as unequal crossing over), fork stalling and template switching during replication, staggered break repair after the occurrence of rearrangements, interchromosomal recombination, transposition mediated by composite DNA transposable elements, whole genome duplication or polyploidization through chromosome non-disjunction in germline are mechanisms that produce DNA-mediated gene duplications (Ohno 1970; Samonte and Eichler 2002; Friedman and Hughes 2004; Conrad and Hurles 2007; Kaessmann 2010; Ranz, et al. 2007; Hastings, et al. 2009). Through these processes duplicates can occur in tandem, interspersed or as a polyploidization event. Recent segmental duplications or low copy number repeats interspersed in the DNA have high sequence identity and can be hot spots for non-allelic homologous recombination producing additional duplications (Bailey, et al. 2002; Bailey, et al. 2003; Marques-Bonet, et al. 2009). The immediate result from a DNA-mediated duplication of an ancestral gene is often considered to be

the generation of an identical copy of the gene in a new position. This is also what is assumed in most models presented above. Tandem duplications by non-allelic homologous recombination are the most abundant way of duplicating genes and represent a big fraction of the new young genes (Heger and Ponting 2007). In addition to DNA-mediated duplication we have RNA-mediated duplication that occurs much less frequently, but have been recognized to be more capable of generating genes with novel functions (Long, et al. 2003A; Kaessmann, et al. 2009).

1.1.1.2. RNA-mediated gene duplication

Retrotransposition is RNA-mediated duplication (Kaessmann, et al. 2009). An mRNA is mistakenly grabbed in the cytoplasm by the machinery of Non-LTR retrotransposable elements, transported to the nucleus and inserted in the genome by means of target-primed reverse transcription (Esnault, et al. 2000). This type of duplication results in a new gene copy that lacks introns. In some cases a poly-A tract at the end of the 3'-untranslated region and target site duplications that are other fingerprints of the event are still visible. The absence of the regulatory sequence of the parental gene is also expected. Just like for DNA-mediated copying, these copies are also subjected to different evolutionary forces, resulting eventually in their loss or fixation. Because of the lack of regulatory regions many more RNA-mediated duplicates than DNA-mediated duplicates might be rendered nonfunctional. Lacking these regulatory regions often leads to a retropseudogene formation (Emerson, et al. 2004). However, with time these copies can share or acquire regulatory sequences, resulting in functional proteins (Bai, et al. 2008). If the retrocopy is functional, it is called a retrogene. At this point we are still learning about the way retrogenes acquire regulatory regions. For the rest of this chapter I will focus on the mechanism and hallmarks of gene duplication

through retroposition, as this is the process that marks the formation of the two genes that are the main focus of my studies in Drosophila.

1.2. Retrogene formation

During the process of retroposition, an mRNA is carried from the cytoplasm back into the nucleus using the enzymatic machinery of Non-LTR retroelements (Esnault, et al. 2000). Essentially, the mRNA is captured by mistake by the reverse transcriptase/endonuclease (RT/EN) enzyme encoded by the Non-LTR retroelement and once in the nucleus the fragment is retrotranscribed and randomly inserted into the genome in a process named target-primed reverse transcription (Brosius 1991; Feng, et al. 1996). In the genome, the endonuclease domain of the RT/EN enzyme randomly nicks one strand of the genomic DNA leaving an exposed 3' hydroxyl residue. A DNA strand complementary to the mRNA is synthesized by the reverse transcriptase domain of the RT/EN enzyme beginning at the exposed 3' OH group (Feng, et al. 1996; Kazazian 2004) (Figure 1.2).

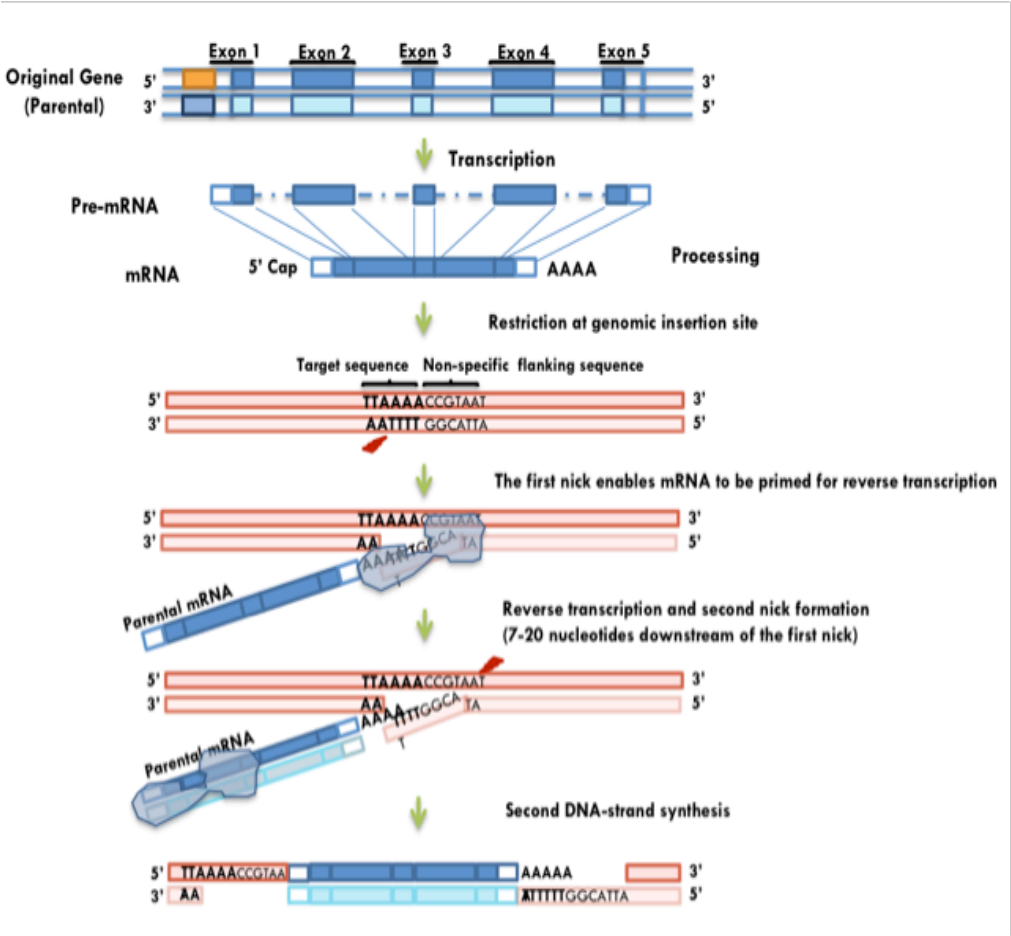


Figure 1.2 - Retrogene formation. Retroposition of a gene into new genomic region by target primed reverse transcription.

Several structural features characterize new retroposed copies of genes as mentioned above. They are intronless, flanked by direct repeats, have a remnant poly A tail and have a different location in the genome when compared with the parental gene (Brosius 1991). The poly A tail and direct flanking repeats are only intact and visible for very young retrocopies as there is no selective pressure for their maintenance. Those features have been used to annotate retrocopies in the genomes of *Drosophila melanogaster* (Betrán, et al. 2002; Emerson, et al. 2004; Bai, et al. 2007; Kabza, et al.

2014). Due to the fact that retrocopies do not carry their regulatory regions they were believed to become pseudogenes with time and were often annotated as pseudogenes (Jeffs and Ashburner 1991; Dunham, et al. 1999). Holver, work on several human functional retrocopies and the work done by Long and Langley shold that retrocopies can be expressed and evolve novel functions often under positive selection (McCarrey and Thomas 1987; Dahl, et al. 1990; Long and Langley 1993).

For a gene to be seen by natural selection it must be expressed at some stage or in some tissue in the organism where it resides. For retrogenes as they are initially devoid of regulatory regions, it is of interest to explore how these genes acquire regulatory regions (i.e., promoter and enhancers). Drosophila retrogenes show no evidence of carrying over upstream regulatory regions or regulatory regions from neighboring genes (Bai, et al. 2008). Transposable elements that could potentially be at the insertion site have also been studied, but there are not that many of them and they do not seem to donate regulatory regions to retrogenes (Bai, et al. 2008). In Drosophila, the kind of regulatory regions used in the parental genes (i.e., upstream promoter and enhancers over extended regions) makes them rarely, if ever, reverse transcribed (Bai, et al. 2007; Pei, et al. 2012). Studies have shown however that retrogenes are found in excess in testis neighborhoods (i.e., regions in the genome where many genes transcribed in testis are located) consequently close to genes expressed in testis (Vinckenbosch, et al. 2006; Bai, et al. 2008; Dorus, et al. 2008). An abundance of DRE and Inr motifs in promoter regions of retrogenes and of other motifs including the *β2 tubulin* 14bp testis motif (β2-UE1) have been described in Drosophila retrogenes (Bai, et al. 2008; Sorourian, et al. 2014). Details about how those regulatory regions come along are still scarce and Chapter 3 addresses this for the two retrogenes under study in this dissertation.

Retrogene formation has been studied using comparative genomics. In order to calculate the rate of retroposition in Drososphila Bai et al. identified functional retrogenes independent of the location of the parental gene in *D. melanogaster* using 50% protein identity level (Bai, et al. 2007). The rate of new gene formation through retroposition was estimated to be 0.5 genes per million years per genome (Bai, et al. 2007). Since the number of gene duplication in *Drosophila* has been estimated to be 17 genes per million years per genome (Hahn, et al. 2007), the rate of retrogene formation is just a fraction of this (i.e., 3%; Bai, et al. 2007). Although retrogenes are a small fraction of the duplicates in the genome, they have been of interest because in studies until now they have revealed trends in the duplication patterns in the genomes and trends in what gene functions are duplicated (See below).

1.2.1. Retrogene duplication patterns

Somatic tissues and gonads of males and females contain genes being differentially expressed. This has been shown in various studies in diverse species, including Drosophila (Meiklejohn, et al. 2003; Parisi, et al. 2003), worm (Reinke, et al. 2000), mouse (Khil, et al. 2004), and human (Wang, et al. 2005). In Drosophila, testes have a greater fraction of tissue-specific genes than other tissues (Ayroles, et al. 2009) and a fraction of these are retrogenes (Figure 1.3).

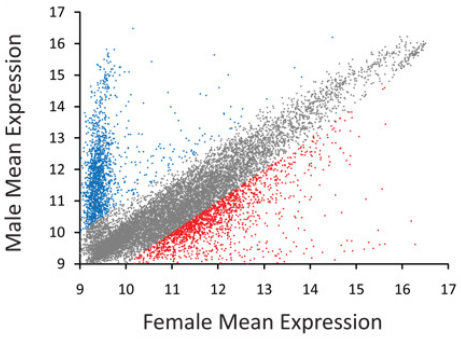


Figure 1.3 - Sex bias for gene expression among 40 wild-derived inbred lines. Blue and red dots represent genes showing a 2-fold difference in gene expression between males and females, respectively (This image is reprinted from Ayroles, et al. 2009 with permission from Nature Publishing Group).

Several genome-wide Drosophila analyses show an excess of retrogenes that have originated from genes on the X chromosome and have later inserted in an autosome and are transcribed in male gonads (Betrán, et al. 2002; Dai, et al. 2006; Bai, et al. 2007; Meisel, et al. 2009; Vibranovski, et al. 2009). Analogous patterns have been observed in mammals (Emerson, et al. 2004; Potrzebowski, et al. 2008). An excessive number, 299% in humans and 309% in mice, of functional retrogenes have been recruited to autosomes from the X chromosome compared to expectations. In addition, an excess of DNA mediated duplications and translocations have also been observed to autosomes from the X chromosome in Drosophila (Meisel, et al. 2009; Vibranovski, et al. 2009B). This pattern has been observed genome wide and has contributed to a demasculinization of the X chromosome (Parisi, et al. 2003; Sturgill, et al. 2007). Demasculinization of X chromosomes in the Drosophila genus has also been proposed to be in part due to the lack of dosage compensation in Drosophila male germline (Meiklejohn, et al. 2011). This "out-of-the-X" pattern does not appear to be explained by mutational biases. In Drosophila, the location of genes producing new polymorphic

retrogenes does not show this bias (Schrider, et al. 2011) and, in humans, retropseudogenes duplication patterns were compared to retrogene patterns and were found to be different and not to be biased to any particular chromosome (Emerson, et al. 2004). The conclusions of these works are that there are selective pressures involved in the recruitment of male-specific genes to the autosomes (Betrán, et al. 2002; Parisi, et al. 2003; Dorus, et al. 2006; Bai, et al. 2007; Dorus, et al. 2008; Vibranovski, Lopes, et al. 2009). With respect to gene expression, male-biased expression of retrogenes does not appear to results from insertion biases only. For example, although human retropseudogenes are often transcribed in testis, they are actually less often transcribed in testis than functional testis retrogenes (Vinckenbosch, et al. 2006). Thus, it appears that male-biased expression of retrogenes has been shaped by selection (either through preferential retention of retrogenes inserted in testes neighborhoods or through sequence changes in regulatory regions that result in testis-bias expression; Vinckenbosch, et al. 2006; Bai, et al. 2007; Dorus, et al. 2008; Gallach, et al. 2011; Sorourian, et al. 2014).

Interestingly, the out-of-the-X pattern of duplication has been also observed for DNA-mediated duplications in Drosophila and again attributed to selection (Meisel, et al. 2009; Vibranovski, et al. 2009B; Han and Hahn 2012). Some of the instances involve complete relocation of male-biased genes from X to autosomes (Meisel, et al. 2009; Han and Hahn 2012). Some gene ontologies are more affected by these patterns than others. Nuclear transport genes have been duplicated to autosomes and acquired testis expression (Bai, et al. 2007; Tracy, et al. 2010; Phadnis, et al. 2012). Proteins that constitute subunits of the proteasome have also been duplicated to autosomes (Belote and Zhong 2009). Many transcription factors have their testis-specific paralogs in Drosophila (Li, et al. 2009). An excess of X-chromosome to autosome duplications (RNA and DNA mediated) have been observed in nuclearly-encoded mitochondrial genes in

Drosophila (Gallach, et al. 2010). These duplicates seem to affect energy production pathways and give rise once again to testis-specific genes (Gallach, et al. 2010). Other gene relocations or duplications to autosomes show male-biased expression and are involved in chromosomal functions and meiosis (Han and Hahn 2012). Many of these genes have, in addition, been observed to evolve under positive selection (Li, et al. 2009; Gallach, et al. 2010; Quezada-Diaz, et al. 2010; Tracy, et al. 2010; Phadnis, et al. 2012).

There has also been an excess of retroduplicates entering the X in mammals (Emerson, et al. 2004; Potrzebowski, et al. 2010) and these genes often have female functions (Potrzebowski, et al. 2010). In Drosophila, some DNA-mediated duplications to the X with female functions also occur (Han and Hahn 2012). I, however, focus on male-specific genes and X to autosome duplications as they are most relevant to this work.

1.2.2. Hypotheses to explain the X-to-autosome duplication pattern of genes and their sex-biased expression

Four hypotheses have been proposed to explain male-specific expression and out-of-the-X pattern of duplicated genes: (1) Avoidance of male meiotic X-inactivation (MSCI); (2) Dosage compensation hypothesis; (3) Sexual antagonism; (4) Role of the new genes in genetic conflicts (meiotic drive and transposable elements suppression).

Avoidance of male meiotic X-inactivation (MSCI) is one of the hypotheses put forward to explain why so many retrogenes with male/testes-biased expression are found on autosomes. This hypothesis was first suggested by McCarrey and Thomas (1987) and states that it is advantageous for genes involved in male meiosis to be located on the autosomes where they can be expressed during meiotic cell divisions instead of on the X-chromosome which is subject to inactivation by *XIST* transcripts during male meiosis. Authors shold a complementary expression between the duplicated *Phosphoglycerate*

kinase 2 (*Pgk2)*, expressed during male meiosis only, and the X-linked *Pgk1*, expressed in spermatogenic cells before meiosis and somatic tissues (McCarrey and Thomas 1987). They concluded that the relocation of this retrogene to an autosome provides an active copy of the gene that would otherwise be transcriptionally silenced during spermatogenesis. That is, the production of an autosomal duplicate that acquires male germline expression is expected to be beneficial to the host (Lifschytz and Lindsley 1972; Hense, et al. 2007; Potrzebowski, et al. 2008; Vibranovski, et al. 2009B). A new mutant with a newly retroposed gene in an autosome will have advantage over an ancestral form since it is able, if the gene is functional, to carry out a role required by male germline cells after X chromosome inactivation (McCarrey 1994). However, male meiotic X-inactivation hypothesis does not fit all the collected data. Why not all the housekeeping genes of the X chromosome are moving out of the X? There must be other ways to cope with X inactivation. For example, it has been observed that X-linked genes are highly expressed before X inactivation in male germline (Wang, et al. 2001), suggesting that genes may be able to generate enough mRNA for proper cell functioning during inactivation. In addition, MSCI hypothesis does not explain the retroposition or DNA-mediated movements observed between the autosomes. Such duplications occur quite often and the new copy frequently acquires male-specific functions (Gallach, et al. 2010). The hypothesis also does not account for the recurrent positive selection that has been observed operating on many testis-specific duplicated genes and cannot explain why some of the duplicated genes are later lost (Tracy, et al. 2010). In addition, although the inactivation of sex chromosomes during meiosis is well established in mammals, its occurrence in Drosophila is still under debate (Meiklejohn, et al. 2011; Mikhaylova and Nurminsky 2012; Kemkemer, et al. 2014; Vibranovski 2014).

Another hypothesis proposed to explain the paucity of male-biased gene on the X chromosome is the dosage compensation hypothesis. Since dosage compensation in Drosophila occurs through hypertranscription of the X chromosome in males, new and old genes might not be able to attain as quickly or as high levels of expression in dosage compensated as in not-dosage compensated chromosomes (Swanson, et al. 2001; Vicoso and Charlesworth 2009). X-linked genes could be making copies to autosomes to further increase their level of expression. The X to autosome relocation could also occur in neo-X chromosomes before they evolve dosage compensation to attain dosage compensation (Meisel, et al. 2009) or to express in testis if there is interference between testis expression and the dosage compensation machinery (Bachtrog, et al. 2010). However, dosage compensation does not explain why RNA- or DNA-mediated movements are observed between the autosomes (Gallach, et al. 2010), and why relocated duplicates are only needed in spermatogenesis and again does not account for the observed positive selection and loss of duplicates (Tracy, et al. 2010).

The third hypothesis suggests sexual conflict as a major driver of the observed patterns. Early studies of sexual antagonism showed that recessive male-beneficial mutations and dominant female-beneficial mutations will increase in frequency more often if they occur on the X chromosome than if they occur on autosomes leading to enrichment of sexually antagonistic alleles on the X chromosome. Resolution of this intralocus sexual antagonism has been proposed to evolve by evolving sex-biased expression and would lead to an excess of sex-biased genes on X chromosome (Rice 1984). However, several studies in Drosophila have shown a paucity of male-specific genes on the X chromosome (Parisi, et al. 2003; Kalamegham, et al. 2007; Sturgill, et al. 2007) and out-of-the-X particular patterns of duplication giving raise to sex-biased genes. Recently, a new way for the resolution of intralocus sexual antagonism has been

proposed in our laboratory (Gallach and Betrán 2011; Gallach, et al. 2011) and suggests that gene duplication is a route for conflict resolution. Under this model, the allele beneficial for males might often be an allele that is actually good for testis (e.g., helping the production of more sperm or better sperm) but not for other male or female tissues. In this case duplication of the male-beneficial allele, accompanied by the establishment of testis-specific expression, is expected to resolve this conflict (Gallach and Betrán 2011; Gallach, et al. 2011). As discussed in Gallach and Betrán (2011) and further explored by Connallon and Clark (2011), all the steps in this model would occur under positive selection. Since duplications are dominant, they can be selected even in autosomes and autosomes might be a better location for male-biased genes because X-linked genes spend two thirds of the time on females (Rice 1987), might suffer male germline X inactivation (Lifschytz and Lindsley 1972; Hense, et al. 2007) or can attain higher level of expression in testis because it will have two copies (Swanson, et al. 2001; Vicoso and Charlesworth 2009). Therefore, any selective pressure that leads to intralocus sexual antagonism (i.e., male favoring a different allele for a gene) can fuel the process, including male-male competition, female-male antagonism, or adaptation to male germline selfish elements (transposable elements or meiotic drive systems; Gallach and Betrán 2011).

The last hypothesis that tries to explain the X to autosome duplication pattern of sex-biased genes proposes the involvement of the new duplicates in genetic conflict (i.e., meiotic drive) suppression. Meiotic drive systems are chromosomal selfish systems that increase the frequency at which they are transmitted to the next generation violating Mendel first law of equal segregation of homologous chromosomes. The ratio of two alleles in gametes is altered by the selfish element to increase its own representation in the next generation (Hurst 2001; Presgraves, et al. 2009). In Drosophila, there are

several examples of genes (often multi-locus complexes) that violate this law of inheritance and act selfishly to ensure their propagation (Meiklejohn and Tao 2010). Meiklejohn and Tao have proposed that new distorters are more likely to invade the X chromosome and suppressors are more efficient when they evolve on the autosomes or the Y chromosome and this could explain the gene duplication patterns. There is direct evidence that some retrogenes play a role in meiotic drive (Tao, et al. 2007) and that many translocated genes are involved in chromosomal processes and meiosis (Han and Hahn 2012). This role has been hypothesized for the nuclear transport retrogenes studied in this dissertation (Presgraves 2007; Tracy, et al. 2010) as discussed below. It is also relevant to mention that some selfish elements can create intralocus sexual antagonism (i.e., there will be different selective pressures on some genes in males and in females due to the presence of selfish elements in male germline) as mentioned above and there is the possibility that the model that involves resolution of intralocus sexual antagonism caused by selfish elements through gene duplication also applies to the evolution of suppressors of meiotic drive.

1.3. Retrogenes in Drosophila

The first retrogene in Drosophila to be discovered was *jingwei* (*jgw*). It was described to be a functional retrogene by Long and Langley (1993). With the discovery of *jgw* it was demonstrated that a gene retroposition event is able to create an entirely new gene. *Jingwei* was created when the *Alcohol dehydrogenase* (*Adh*) mRNA was reverse transcribed and then inserted into a duplication of the *yellow emperor* gene giving rise to a new chimeric gene (Long and Langley 1993; Wang, et al. 2000). Moreover, *jingwei* was

shown to have experienced accelerated evolution under positive selection (Long and Langley 1993).

A comprehensive genome-wide search for retrogenes in *Drosophila melanogaster* was performed by Bai, et al. (2007). These authors identified 94 functional retrogenes in *D. melanogaster*. They estimated that functional retrogenes originate and become fixed in the genome at a fairly constant rate of 0.5 genes per million years per lineage (Bai, et al. 2007). After a retrogene becomes fix in the genome it is shown that it can start evolving under positive selection to acquire a novel function (Betrán, et al. 2002; Betrán, et al. 2006; Quezada-Diaz, et al. 2010; Tracy, et al. 2010; Chen, et al. 2012A).

Insights about the selective pressures underlying the patterns of retrogene formation, their gain of male-germline expression and their rapid protein change can be gained if duplicates are studied in detail and their functions are scrutinized. For example, *Prosα6T*, an X to autosome retrogene studied using detailed pattern of expression and gene knockout, showed an important role in male fertility (Belote and Zhong 2009). Dorus, et al. (2008) reported four proteins of recent retrotransposition, two of them X to autosome, present in the sperm proteome. *Mojoless*, another retrogene with these characteristics, was demonstrated to be required for male germline survival (Kalamegham, et al. 2007). *K81*, a paternal effect gene, expressed in spermatogenesis that is transmitted in sperm to the embryo is also a retrogene (Loppin, et al. 2005). *Zeus* (also known as *Rcd1r*) encodes a transcription factor needed for male fertility (Chen, et al. 2012B).

In several instances, convergent retroduplications of the same genes in different Drosophila lineages have also occurred (Bai, et al. 2007; Belote and Zhong 2009; Tracy, et al. 2010; Han 2012; Han and Hahn 2012; Phadnis, et al. 2012). High levels of retrogene loss have also been proposed, revealing high level of turnover of some

retrogene families. These recurrent events are likely to be driven by strong selective pressures and often involve genes with chromosomal functions, nuclear transport and protein degradation. Two genes from two such gene families are studied in this Dissertation.

1.4. Recurrent retroduplication of *Dntf-2* and *Ran* in Drosophila

Dntf-2 and *Ran* are nuclear transport genes and their proteins interact with each other during nuclear transport (See more details in Chapter 2) and both genes have given rise to retrogenes expressed nearly exclusively in male germline in three instances in Drosophila (Tracy, et al. 2010). These retrogenes are evolving under positive selection as revealed by McDonald-Kreitman test (Betrán, et al. 2002; Tracy, et al. 2010) but sometimes are lost (i.e., a retroduplicate of *Ran* has been lost in *D. yakuba*; See more details in Chapter 2). Retroduplicates of *Dntf-2* and *Ran,* called *Dntf-2r* and *Ran-like,* exist in *D. melanogaster* and are amenable to functional analyses because *D. melanogaster* is a well-developed model organism for genetic and functional analyses. I *Dntf-2r* and *Ran-like* are an example of genes being recurrently duplicated from X to autosomes and one of the hypothesis is that they duplicated to suppress meiotic drive systems (Presgraves 2007; Tracy, et al. 2010). This could explain their expression pattern, mode of evolution, and their loss after the meiotic drive system is suppressed.

The reasons for this hypothesized role has its roots in a well-known male germline meiotic drive system called Segregation Distorter (SD). SD system was discovered close to 60 years ago in *D. melanogaster* (Reviewed in Larracuente and Presgraves 2012). SD system involves a duplication of a gene (*RanGAP*) known to be

involved in nuclear transport as well as other loci. To understand the SD system I will need to introduce nuclear transport first.

1.4.1. Ran and Ntf-2 function in nuclear transport

A double membranous barrier, the nuclear envelope, physically separates the nucleus from the cytoplasm. Nuclear pore complexes lined through the nuclear envelope control the exchange of all cellular materials between the inside and outside of the nucleus. It is known that *Ran* and *Ntf-2* proteins physically interact and play a central role in transport of proteins from the cytoplasm to the nucleus of the cells (Figure 1.4; Ribbeck, et al. 1998). These proteins are present virtually in every cell.

Ran is a member of Ras superfamily and exists in GDP-bound inactive form and GTP-bound active form in different species. RanGTP is present in the nucleus and RanGDP predominantly localizes in the cytoplasm. The import and export of cargo proteins across the nuclear membrane are resulting from the cytoplasm-nuclear gradient of RanGDP-RanGTP. The other nuclear transport protein Ntf-2 is a homodimer and interacts with RanGDP in the cytoplasm and carries it across the nuclear pore into the nucleus (Quimby, et al. 2000). Once in the nucleus, a catalytic enzyme RanGEF (Ran GTPase Exchange Factor), also known as RCC1, converts Ran from the GDP-bound to the GTP-bound form. RanGTP binds to importin β and induces conformational changes that lead to the dissociation of importin α/β heterodimer and release of the cargo protein. RanGTP ensures the release of cargo proteins in precise spatial and temporal pattern for the proper orchestration of downstream functions. RanGTP bound to importin β is then transported out of the nucleus (Isgro and Schulten 2007). RanGTP is also needed for assembly of export complexes and it is transported with these complexes to the cytoplasm (Kusano, et al. 2003; Matsuura and Stewart 2004). RanGAP (Ran GTPase

Activating Protein) helps the hydrolysis of RanGTP into RanGDP in the cytoplasm (Kusano, et al. 2002).

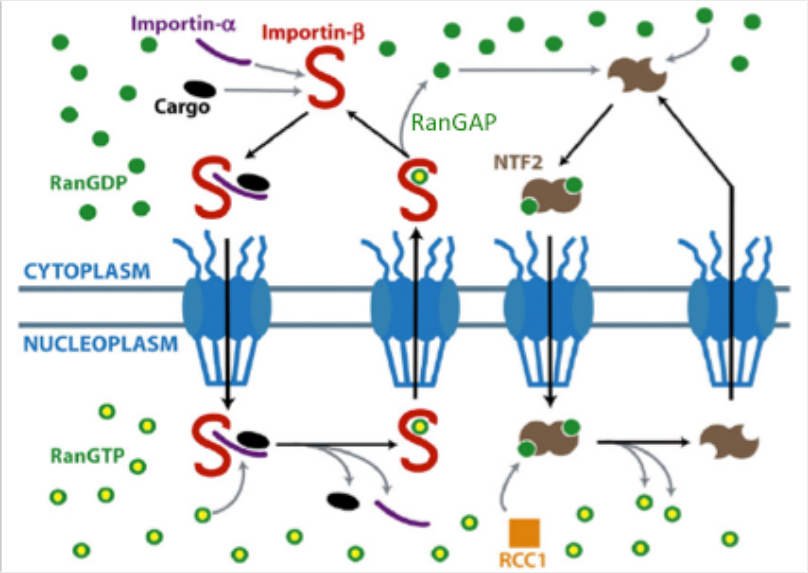


Figure 1.4 - Nuclear transport schema (This image is reprinted from Isgro and Schulten 2007 with permission from Elsevier).

Ntf-2-RanGDP complex is also actively involved in the import of filamentous actin capping protein CapG (Van Impe, et al. 2008). CapG belongs to the gelsolin superfamily, proteins that control actin organization by severing filaments, capping filament ends and nucleating actin assembly. RanGDP and Ntf-2 is essential for CapG transport (Van Impe, et al. 2008).

In addition to the transport functions, RanGTP concentration gradients are required during normal cellular divisions. RanGTP gradient is required for nuclear envelope assembly following cell division and also has a role in nuclear envelope assembly around sperm chromatin (Clarke and Zhang 2008). Ran is also responsible for synchronizing nuclear and chromosomal functions all through cell-division cycle. *Ran*

pathway has been shown to have a key role in spindle assembly. Furthermore, *Ran* is responsible for the formation and organization of the microtubule network being the signaling molecule that regulates microtubule nucleation during mitosis (Figure 1.5; Silverman-Gavrila and Wilde 2006; Clarke and Zhang 2008).

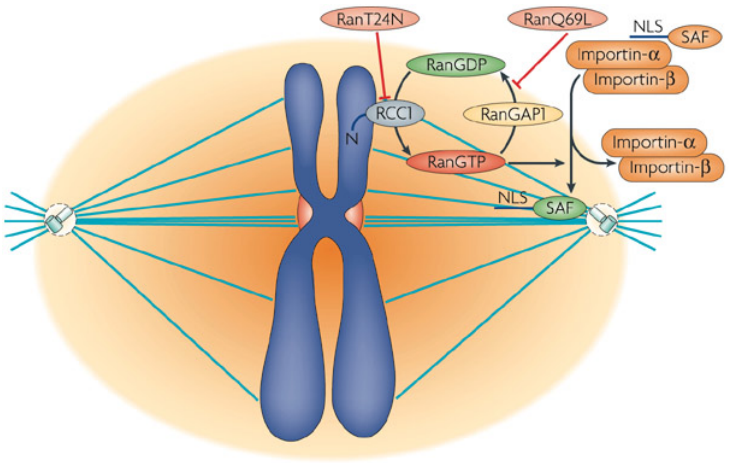


Figure 1.5 - RanGDP and GTP forms and RCC1 involvement in spindle assembly during mitosis (This image is reprinted from Clarke and Zhang 2008 with permission from Nature Publishing Group).

During mitosis there is a radical reorganization of the microtubule network of the cell in order to form the mitotic spindles. Microtubule changes and reorganization are achieved through the control of their nucleation and stability. RCC1 generates a high local concentration of RanGTP on the chromosomes around chromatin which, in turn, induces the local nucleation of the microtubules and assembly of all structures relative to chromosomes (Moore 1998; Clarke and Zhang 2008). The spindle formation progression is a two way process. In a first step RCC1 binds to the DNA and histone proteins 2A and 2B. RanGTP is then recruited to the location of spindle assembly on the chromosome. After this first phase is accomplished, high concentration of RanGTP is present around

the spindle assembly regions. The importins attract and release the spindle assembly factors (SAFs; Kalab, et al. 2006; Clarke and Zhang 2008; Fu, et al. 2010).

Loss of function of *Ran* produces lethality in Drosophila (Peter, et al. 2002). Work done by Cesario and McKim (2011) in Drosophila using Ran mutations in oocytes and embryos revealed a role of the Ran pathway on microtubule dynamics inside and outside the nucleus in the organization of acentrosomal spindle poles including chromosomal and chromosomal-independent microtubules and cytoplasmic microtubules, and the segregation of achiasmatic chromosomes like the 4th chromosome in females. Dominant-negative mutants in females carrying Ran that is locked into the GDP form showed fertility defects suggesting that RanGTP is necessary for normal fertility. In particular, the reductions in fertility occur because RanGTP is not participating in microtubule nucleation on the chromosomes as shown above.

In Drosophila, *Dntf-2* loss-of-function mutants are also lethal. Moreover, *Dntf-2* has also been implicated in normal eye development (Bhattacharya and Steward 2002). Even though *Dntf-2* loss-of-function mutants are lethal some hypomorphs are viable and fertile. These hypomorphs show an impaired eye phenotype where the number of omatidia in the eye is severely reduced. The third phenotype that has been observed for in *Dntf-2* mutants is that the larvae immune response gets compromised (Bhattacharya and Steward 2002).

1.4.2. SD system in D. melanogaster

As mentioned above genes involved in nuclear transport have been shown to be part of genomic conflicts (Presgraves 2007; Larracuente and Presgraves 2012). Segregation Distorter (SD) is a meiotic drive gene complex on chromosome 2 of *D. melanogaster*. This complex primary consists of a truncated duplicate of *RanGAP* that

lacks its nuclear export signal and accumulates in the nucleus (*Sd*; Kusano, et al. 2003; Presgraves 2007; Larracuente and Presgraves 2012), a short version of a satellite (insensitive responder; Rsp^{i}) and a set of inversions that link these loci. These three components are well characterized. Other components have been tentatively mapped (i.e., it is difficult to map effects in regions with inversions) and are modifiers and enhancers of the drive (Figure 1.5). The Segregation Distorter chromosome is known to cause meiotic drive where ~99% of the heterozygote males (SD/SD+) progeny receives the chromosome carrying the SD chromosome instead of the expected Mendelian ratio of 1:1 (Figure 1.7). This process happens in spermatogenesis because *SD* induces dysfunction of SD+ spermatids that do not transition between histones and protamines and do not mature into functional sperm (Presgraves 2007; Larracuente and Presgraves 2012; Gell and Reenan 2013). The retrogene *Dntf-2r* is located close to *Sd* (Figure 1.5) and this, in addition to its pattern of expression and mode of evolution, prompted the hypothesis of its involvement in this meiotic drive system, either as part of the drive or the suppression (Presgraves 2007; Larracuente and Presgraves 2012).

The SD system was initially hypothesized to have been originated in the Mediterranean area but it has now been found worldwide in natural populations of *D. melanogaster* including Africa (Presgraves, et al. 2009). *SD* was first revealed six decades ago and was found at a frequency of 3-5% in natural populations of *D. melanogaster* (Kusano, et al. 2002). This is around the same frequency at which it is found in Africa (Presgraves, et al. 2009). However, some SD chromosomes in Africa appear to have recently appeared and increase in frequency through the population regardless of not reaching high frequency (Presgraves, et al. 2009). This reveals a dynamic system that has chromosomes that are still changing and avoiding suppressors despite not reaching high frequency.

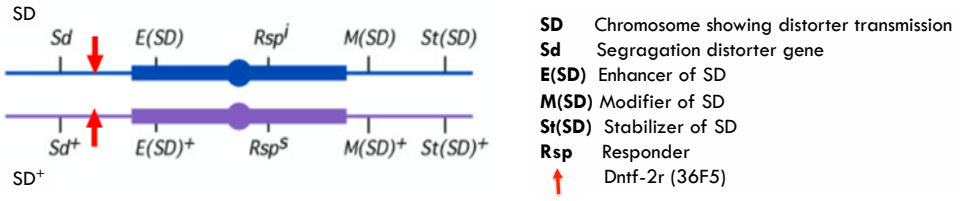


Figure 1.6 - The *SD* system in *D. melanogaster* (This image is reprinted from Presgraves 2007 with permission from John Wiley and Sons).

The most accepted model of how SD operates involves the perturbation of the RanGDP/RanGTP gradient by the misslocalization of RanGAP in the nucleus that in turn slows transport of small RNAs (i.e., piRNAs) that get amplified in the cytoplasm and transported back for the silencing of the responder satellite. When the transport slows the silencing/condensation of the long satellite (Rsp^s) that most chromosomes carry is prevented but not the condensation of the short satellite (Rsp^i) that SD chromosomes carry. This leads to the maturation of one kind of sperm but not the other. This model is supported by the fact that many nuclear transport genes affect SD (e.g., overexpression of *RanGAP* causes SD and overexpression of *Ran* or *RCC1* suppresses SD; Kusano, et al. 2002) and knocking-out a gene of the piRNA pathway (i.e., *Aubergine*) enhances segregation distortion (Gell and Reenan 2013).

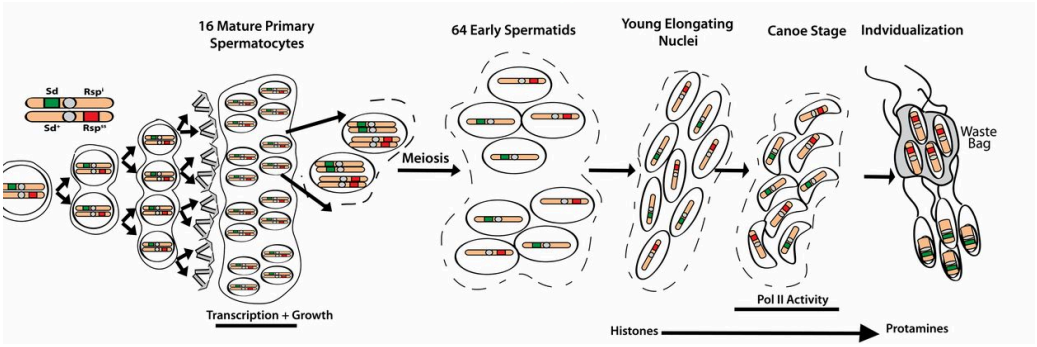


Figure 1.7 – Segregation Distorter (SD) in Drosophila spermatogenesis. Only Sd-bearing spermatids become mature sperm while the others are removed and end in the waste bag (This image is reprinted from Gell and Reenan 2013 with permission from Genetics Society of America.

After understanding nuclear transport and the SD system it becomes clear why it has been proposed that the recurrent duplication of *Dntf2* and *Ran* nuclear transport genes is part of a constant arms race between genes causing meiotic drive and duplicates compensating for it. In particular it has been proposed that many proteins are evolving fast under positive selection to suppress meiotic drive system including nucleoporins and RanGAP itself (Ganetzky 1977; Dermitzakis 2000; Presgraves 2007; Larracuente and Presgraves 2012). Fast evolving nucleoporins have also been shown to cause interspecies incompatibilities (Presgraves and Stephan 2007) underscoring the potential role of genomic conflicts in speciation. However, it is important to keep in mind that the SD system is relatively young as it is found only in populations of *D. melanogaster* and the *Dntf2* and *Ran* duplicates in *D. melanogaster* are much older (at least 5.4 My for the *Dntf2* duplicate and 12.8 My for the *Ran* duplicate but likely much older; Bai, et al. 2007). If these kinds of genomic conflicts are to explain the nuclear transport duplications, it needs to be assumed that systems similar to SD involving nuclear transport, satellites and rearrangements are a recurrent phenomenon in

Drosophila (Presgraves 2007). The mode of evolution of *Dntf2* and *Ran* duplicates (i.e., the positive selection observed acting on those proteins) could also be explained by these conflicts.

1.5. Dissertation objectives and outline

My dissertation has three main objectives that are a continuation of the work previously started in the laboratory on recurrent duplication of nuclear transport retrogenes for testes functions.

Objective 1: As more genomes are sequenced of Drosophila species, I would like to understand the amount of recurrent duplication of nuclear transport genes and the molecular mechanisms that give rise to them. For *Dntf-2* and *Ran*, the Betrán lab found recurrent retroduplication but, for importins, both DNA and RNA-mediated have been found (Phadnis, et al. 2012). What is the pattern when we consider all the data? Does the retroduplication bias still hold? Can we come up with hypotheses to explain it?

Objective 2: A long-standing question about retrogenes is how they acquire their often testis-specific regulatory regions. I study in detail the regulatory regions of *Dntf-2r* and *Ran-like* to provide two examples of how this occurs. One question we can answer is: Is retroduplication facilitating the acquisition of testes regulatory regions? How? How does the way the regulatory regions are acquired fit into/facilitate the proposed models of gene duplication?

Objective 3: I would like to understand the reasons why there is a recurrent duplication of nuclear transport retrogenes for testes functions. To answer this, I study the function of *Dntf-2r* and *Ran-like*. Does the function of these retrogenes explain the recurrent duplications of these genes and can I extrapolated from these data to other

duplicates? Does the current funciton of *Dntf-2r* and *Ran-like* reveals i fthey were duplicated to suppress ancient selfish systems similar to SD?

The presentation of the results is divided into 4 research chapters (Chapters 2 to 5). Chapter 2 includes an update about the number and nature of recurrent duplications of *Dntf2* and *Ran* in 22 available Drosophila genomes. It describes additional duplications of *Dntf2* and *Ran* including DNA-mediated and RNA-mediated duplications in different Drosophila species revealing even more gene turnover than previously described. In Chapter 3, I analyzed the regulatory regions of *Ran-like* and *Dntf-2r* and offer my contribution to the understanding of how the expression of testes genes is gained. Part of this work was recently published in MBE. In Chapter 4, my main focus is to understand *Dntf-2r* and *Ran-like* functions in testis in order to elucidate the selective pressures that lead to the duplication of these genes. In Chapter 5, I show data that reveals that *Dntf-2r* and *Ran-like* have actually important functions in somatic tissues as well. There is a final chapter (Chapter 6) where I summarize the results and come back to the questions I pose, the answers that I obtained and the questions that remain still open.

CHAPTER 2

RECURRENT DNA-MEDIATED AND RNA-MEDIATED DUPLICATIONS OF *DNTF-2* AND *RAN* IN DROSOPHILA

2.1. Introduction: Recurrent recruitment of *Dntf-2* and *Ran*

The study of retrogenes by Bai, et al. (2007) revealed recurrent and convergent duplications of two nuclear transport genes (*Dntf-2* and *Ran*) in different lineages of Drosophila. *Dntf-2* and *Ran* seem to have given rise to retroposed copies three independent times. *Dntf-2* transcript A was the transcript that gave rise to all *Dntf-2* retrocopies. A retrogene (*Dntf-2r*) is present in four species of the *D. melanogaster* complex: *D. melanogaster, D. simulans, D. mauritiana* and *D sechellia* (Betrán and Long 2003). Two other independent retroposition events occurred in the *D. ananassae* and *D. grimshawi* lineages. *Ran* seems to have given rise to retrogenes in the same lineages as *Dntf-2* as shown in Figure 2.1.

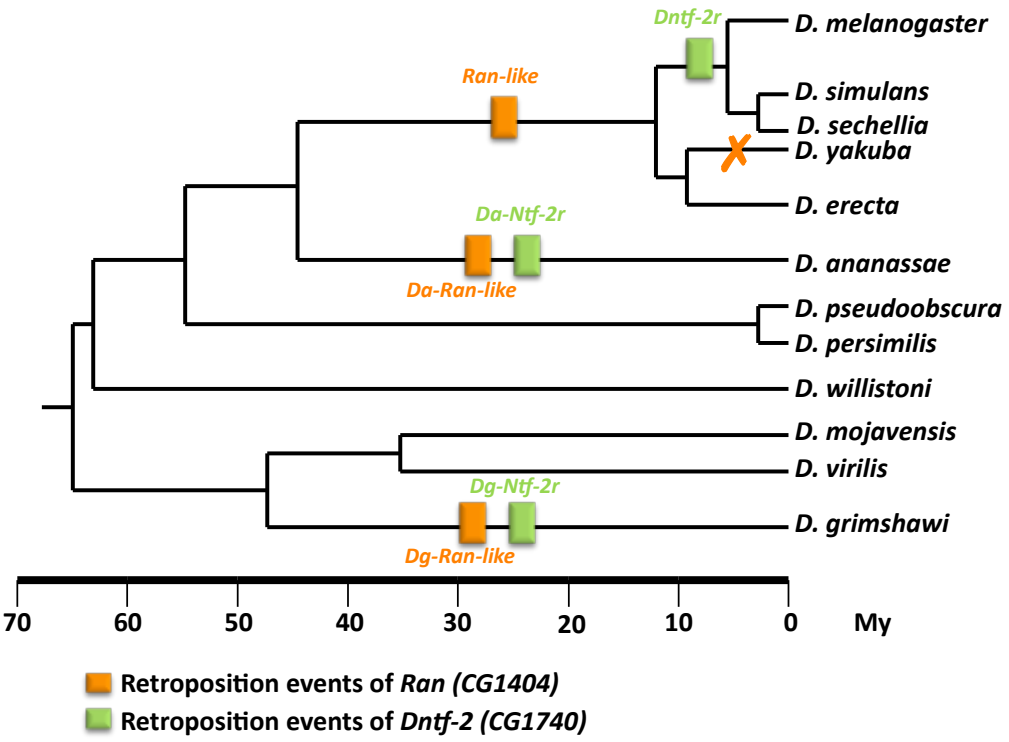


Figure 2.1 - Drosophila phylogeny indicating the retroposition events of *Dntf-2* and *Ran* in different lineages. "X" depicts the loss of *Ran-like* in the *D. yakuba* lineage.

The recurrent duplications of both genes (*Ran* and *Dntf-2*) are particularly interesting considering the fact that the two parental genes physically interact during nuclear transport (Chapter 1). It suggests that new duplicates may be recruited into a tissue-specific nuclear transport pathway as interacting members. Consistent with this possibility, the two new retrogenes have similar expression patterns. In *D. melanogaster,* both *Dntf-2r* and *Ran-like* are highly expressed in male germline (Long, et al. 2003B; Chintapalli, et al. 2007; Gelbart and Emmert 2013) while parental genes are present in every tissue. In *D. ananassae*, both retrogenes show high expression in male germline and lower expression in females whereas both parental genes are ubiquitously expressed. In *D. grimshawi,* both retrogenes and parental genes are expressed in males

and females. However, *Dntf-2 and Ran* retrogenes are expressed in male germline at higher levels than their parental genes (Tracy, et al. 2010; Gelbart and Emmert 2013).

Tracy, et al. (2010) showed that in all lineages both retrogene proteins are evolving at a faster rate than the parental gene proteins. This can be due to relaxation of selective constraint or positive selection acting on the duplicates. McDonald-Kreitman test using polymorphism data from *D. melanogaster* and *D. simulans* provided evidence that *Dntf-2r* and *Ran-like* are evolving under positive selection (Betrán and Long 2003; Tracy, et al. 2010). However, *Ran-like* in *D. yakuba* linage has accumulated many disablements and it is likely a pseudogene (Tracy, et al. 2010). This reveals a certain turnover of these genes as expected if these new genes are involved in arms races. McDonald-Kreitman test using polymorphism data from *D. ananassae* and *D. atripex* also provided evidence that the duplicates in these lineages are evolving under positive selection (Tracy, et al. 2010).

Interestingly, Phadnis et al. (Phadnis, et al. 2012) have recently described the gene duplications of another important component of nuclear transport, *importins* (Figure 1.4). Authors show that some members of *importin* gene family have been recurrently duplicated and acquired testis expression in different Drosophila lineages, while other members have been lost. Both DNA-mediated and RNA-mediated duplications contribute to this pattern. While these genes appear to have experienced positive selection in some instances, authors suggest that the additional dose of importin in testis might be the reason for the preservation of these new duplicates. Authors favor the hypothesis that these duplications occur to fight selfish segregation distortion systems (similar to the SD system described above) in male germline.

In this chapter, I address Objective 1. I would like to understand the amount of recurrent duplication of *Dntf-2* and *Ran* genes and the molecular mechanisms that give

rise to them. The Betrán lab found recurrent retroduplication of those genes, but, for importins, both DNA and RNA-mediated have been found (Phadnis, et al. 2012). What is the pattern of duplication of *Dntf-2* and *Ran* when we consider all the data? Does the retroduplication bias still hold? Can we come up with hypotheses to explain this bias, if it exists? So, using the available whole genome sequences for 22 Drosophila genomes from FlyBase, I provide an update on the number of times *Dntf-2* and *Ran* have produced duplicates in Drosophila. I search for RNA-mediated duplications as done before (Tracy, et al. 2010) and also for DNA-mediated duplications in these lineages. The analysis provides a broad overview of *Dntf-2* and *Ran* duplication patterns in the Drosophila phylogeny.

2.2. Material and Methods

2.2.1. Identification of Dntf-2 *and* Ran *duplicates*

In order to identify new DNA- and RNA-mediated duplications of *Dntf-2* and *Ran* across 22 Drosophila genomes, Blastp program was used to compare protein sequences from *D. melanogaster* with annotated proteins of other species (i.e., *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, and *D. grimshawi*). Since we blast to annotated proteins we can detect DNA- and RNA-mediated duplications as long as they are annotated. *Dntf-2* and *Ran* Blastp hits with identity >=50% and coverage of at least 100 amino acids were retained. For unannotated genomes tBlastn was used against the genome assembly of *D. takahashii*, *D. biarmipes*, *D. rhopaloa*, *D. elegans*, *D. eugracillis*, *D. ficusphila*, *D. kikkawai*, *D. bipectinata*, *D. miranda*, *D. virilis*, and *D. albomicans*. DNA-mediated duplications will appear as broken hits in the same scaffold. Those hits were pieced

together if they had identity >=50% and retained if the whole coding region was at least 100 amino acids. We recovered the parental genes in nearly all species in this way but few DNA-mediated duplications but we do not think our approach is biased (See Discussion). For hits that are more distantly related than was allowed by specified cutoff values, it was checked if these genes have hits above the cutoff values in other species, and in those instances the hits were retained. This data was used to map presence/absence of *Dntf-2* and *Ran* duplicates onto the Drosophila phylogeny.

2.2.2. Dntf-2r and Ran-like maximum-likelihood tree

All protein sequences for *Dntf-2* and *Ran* duplicates were aligned using accuracy-optimized L-INS-i option of MAFFT software (Katoh and Standley 2013). The maximum likelihood phylogenetic tree was constructed using PhyML 3.0 software (Guindon, et al. 2010). Branch support was assessed by approximate likelihood ratio test (aLRT SH-like support). This method for branch support largely agrees with branch support provided by bootstrap procedure, but is much faster (Anisimova and Gascuel 2006).

The trees illustrate the protein evolution. Long branches represent instances of acceleration in the rate of protein evolution as duplications have been inferred to have occurred in specific lineages from presence/absence and some synteny comparisons and do not appear to predate the Drosophila genus.

2.3. Results

The results from the identification all *Dntf-2* and *Ran* DNA- and RNA-mediated duplicates were compiled in a species cladogram (Figure 2.2). The protein sequences used are available in Appendix A.

In *D. melanogaster*, *Dntf-2* has 4 annotated transcripts. However, only the proteins produced from transcripts A (PA) and B (PB) were used in the blast searches as they represent all coding exons of *Dntf-2*. Previously described retrocopies are all derived from PA. *Dntf-2*, the parental gene, is present in all species but there are lots of missanotations for it. In many species PA and PB are annotated as a single transcript with extra exons. *D. melanogaster*, *D. simulans*, and *D. sechellia* all contain the parental and the known retrogene only. No *Dntf-2* DNA-mediated duplicates were found in any of these lineages. *D. yakuba* and *D. erecta* have no *Dntf-2* duplicates. *D. biarmipes* has a retro hit of 130 aa but no methionine and *D. rhopaloa* has a retro hit with 57.3% identity to PA and a length of 124 aa, but no clear start or stop codons. I put question marks on those because the genes are not completely annotated and could possibly be pseudogenes. Their presence likely indicates a shared old duplication event (See phylogenetic tree below). *D. ficusphila* has a very young retrogene 97.7% identity to PA. *D. ananassae* has an old retrogene 50% identity and a young retrogene 59.9% identity to PA. *D. bipectinata* is close related to *D. ananassae* and shares one retrogene with *D. ananassae* but lacks the other one. At this point it is difficult to say if the missing gene is actually present in *D. bipectinata,* or if the region has not been sequenced. I count it as a loss for now (Figure 2.2). Some losses in Figure 2.2 are confirmed but some are inferred from absence in the assembly and might represent unsequenced regions. *D. pseudoobscura*, *D. persimilis* and *D. miranda* have a DNA-mediated duplicate although it is difficult to annotate in *D. miranda* and no sequence was retrieved for this species. The

DNA duplicate in *D. pseudoobscura* (*GA25766*) is male-biased gene according to B. Oliver RNAseq data shown in FlyBase. *D. grimshawi* has a known RNA-mediated duplicate from PA. Therefore, we observe 7 duplications if we count the incomplete retrocopy of *Dntf-2* in *D. biarmipes* and *D. rhopaloa*.

For *Ran*, there is only one transcript in *D. melanogaster* and the protein encoded by that transcript was used in the blast searches. In the subgenus Sophophora, there is an old (i.e., very divergent, <50% in some lineages, and present in many species) retrogene that seems to be present in the *D. willistoni* and *D. obscura* groups, *D. rhopaloa, D. elegans* and *D. eugracilis* and the *D. melanogaster* subgroup of species except *D. melanogaster,* but not in the rest (inferred from the ML below; synteny has not been checked). This is a retrogene that was missed in Bai et al. (2007) work because identity to *D. melanogaster* Ran is <50% in some species. There is, in addition, a young retrogene in the *D. melanogaster* subgroup (*D. melanogaster, D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*) although *D. yakuba* has a disabled copy. *D. takahashii* and *D. rhopaloa* have an old Ran retrogene, but syntenic regions must be studied to see if this gene corresponds to *Ran-like* making *Ran-like* older and potentially adding some losses (See ML tree below). *D. biarmipes, D. bipectinata*, *D. grimshawi* and *D. rhopaloa* also have one young retrogene while *D. ficusphila* has two young retrogenes (i.e., a retrocopy that seems to have duplicated in tandem). *D. ananassae* has a old retrogene that has been lost in *D. bipectinata*. *D. mojavensis* and *D. virilis* share a DNA-mediated duplicate. This amounts to 10 duplications. It thus appears that *Ran* and *Dntf-2* experienced a lot more duplications and losses than was previously thought. It is also clear that RNA-mediated duplications contribute a lot more to Ran diversity than DNA-mediated duplications. The second retrogene in *D. ficusphila* should be counted as a DNA mediated duplication.

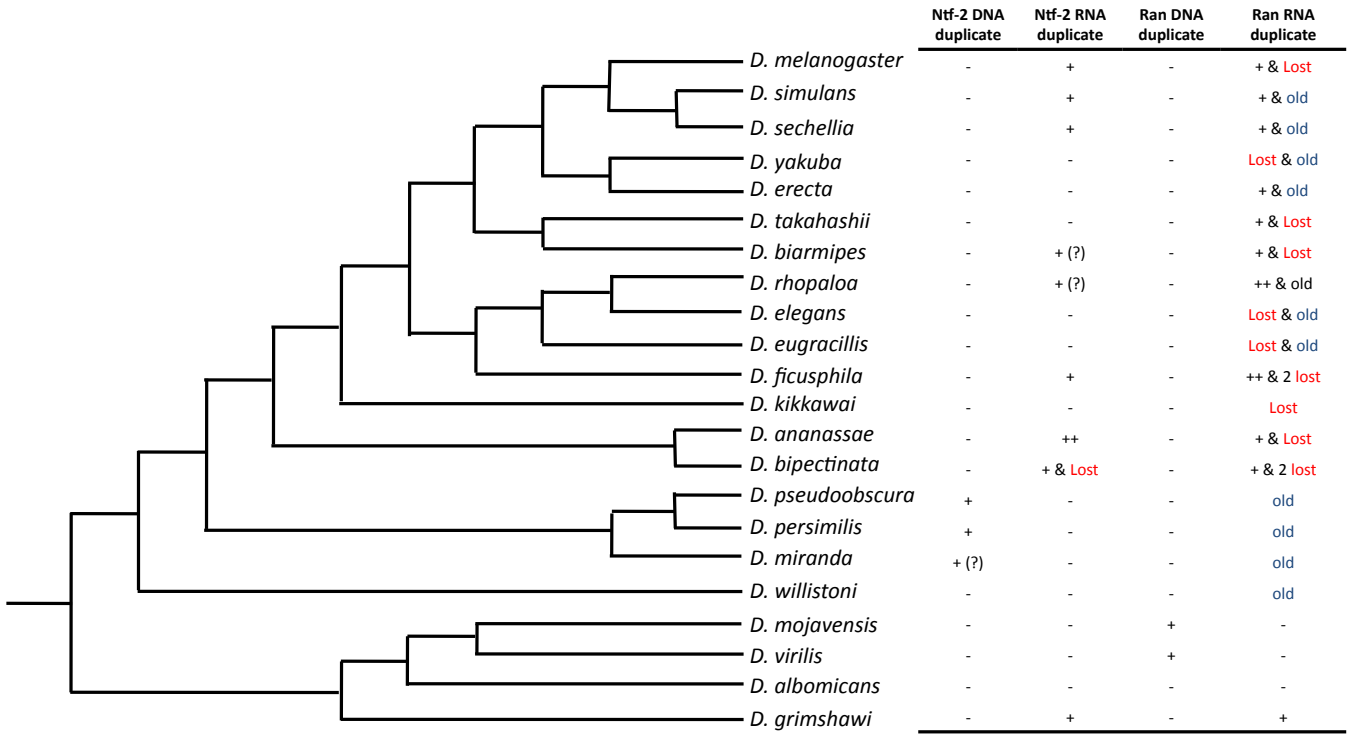


Figure 2.2 - Cladogram of DNA and RNA duplication events for *Ntf-2* and *Ran.* Plus sign indicates the presence and minus sign indicates the absence of the duplicate. Number of plus signs corresponds to number of detected duplicates. The word “lost” indicates the inferred loss of a duplicate. The question mark is present when the data doesn’t allow us to annotate the duplicate completely, but the annotated region does not have disablements. “Old” represents an old retrogene in that lineage (sequences are provided in Appendix

To describe the evolutionary relationships between the parental genes and their duplicates, I performed phylogenetic analysis using maximum likelihood approach. For *Dntf-2* (Figure 2.3), I see a distinct clade for *Dntf-2* PB with short length branches. This is indicative of a high degree of evolutionary constraint for this protein isoform. *Dntf-2* PA sequences are also conserved and form a distinct clade. *Dntf-2* PB and *Dntf-2* PA are only different in their last exon. A number of DNA duplicates and retroduplicates are nested within the *Dntf-2* PA clade indicating that this is the PA transcript that gives rise to new duplicates. This pattern is consistent with earlier findings in our laboratory. It was observed that all of the *Dntf-2* retroduplicates are retroduplicates of *Dntf-2* PA (Tracy, et al. 2010). Many new genes show longer branches than *Dntf-2* PA and *Dntf-2* PB revealing fast evolution for those proteins. Fast evolution can be a result of relaxation of selective constraint or positive selection acting on the duplicates. Since McDonald-Kreitman test using polymorphism data for the two of *Dntf-2* retrogenes studied revealed positive selection in the past (Betrán and Long 2003; Tracy, et al. 2010), I am inclined to suggest that this is also the case for the other lineages, but population data is needed to test this.

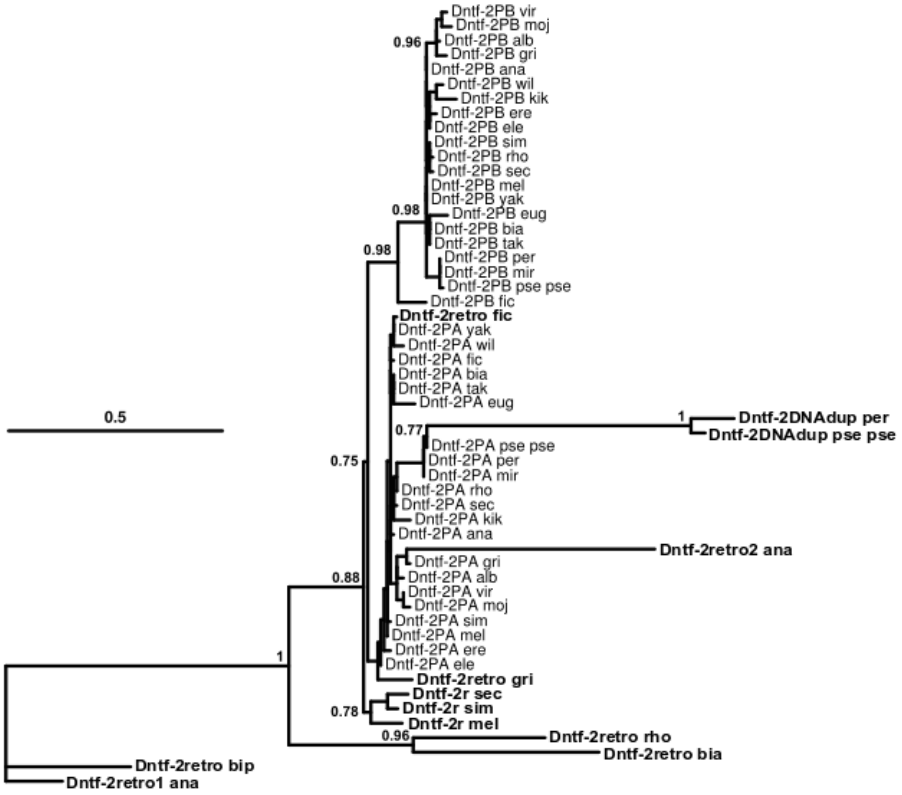


Figure 2.3 - Maximum-likelihood phylogenetic tree using protein sequences of all *Dntf-2* and *Dntf-2* duplicates (sequences are provided in Appendix A).

I used the same procedure for *Ran* duplicates and I can see that all parental orthologs are in the same clade with short branches. Ran protein is nearly identical in all species. Additional detailed synteny comparisons for some unannotated genomes should be performed to decide if some of the duplicates are actually orthologs. *Ran* DNA- and RNA-mediated duplications show longer branch lengths, suggesting again either less constrain for these genes or positive selection (Figure 2.4). Again, since McDonald-Kreitman test using polymorphism data for two of *Ran* retrogenes studied revealed positive selection in the past (Tracy, et al. 2010), I am inclined to believe that this is also

the case for the other lineages described here, but population data is needed to be sure of this.

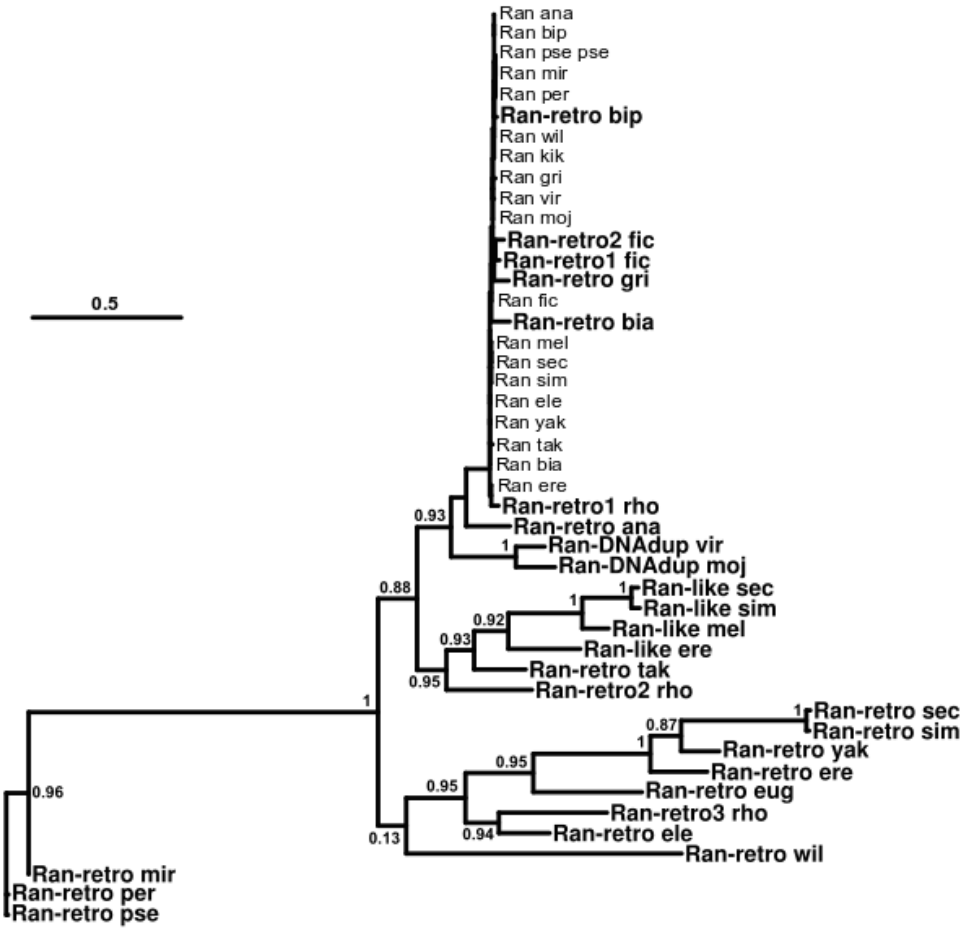


Figure 2.4 - Maximum-likelihood phylogenetic tree using protein sequences of *Ran* and *Ran* duplicates (sequences are provided in Appendix A).

2.4. Discussion

The divergence times between Drosophila species are still in dispute (Osbard et al. 2012. Estimating Divergence Dates and Substitution Rates in the Drosophila Phylogeny) and some divergence times between the species studied in this chapter have not been calculated. So, I did not calculate a rate of duplication for *Dntf-2* and *Ran* but I am able to observe old duplications that were missed in previous studies because were lost in some lineages and to describe quite a few new young duplications especially for *Ran*. So, from the analyses of duplications of *Dntf-2* and *Ran* in the 22 available Drosophila genomes, I reveal a higher rate of duplication of these genes and a higher level of turnover than was previously known. This suggests that the selective pressures leading to the duplication of these genes should be strong and ubiquitous. In this Chapter I showed that *Dntf-2* and *Ran* have been recurrently duplicated in more lineages than previously predicted. *Dntf-2* seems to have been duplicated 6 or 7 times being only one of these times a DNA duplication (Figure 2.2; Figure 2.3). For *Ran* I estimate that 10 duplications have taken place with two duplications being DNA duplications. Again, this amounts to more RNA-mediated duplications than DNA-mediated. In this case many losses have also occurred (Figure 2.2; Figure 2.4).

As explained in Chapter 1 many testis genes have been observed to evolve under positive selection. This could be in response to pressures like male–male competition, sexual antagonism, and/or genetic conflicts (segregation distortion or transposable elements) (Zhang, et al. 2004; Haerty, et al. 2007; Presgraves 2007; Presgraves and Stephan 2007; Tracy, et al. 2010; Phadnis, et al. 2012). These pressures can also explain why these gene duplicates are retained in the genomes. If this tissue is changing very fast, new functions are needed all the time.

In the particular case of *Dntf-2* and *Ran* duplicates, the new genes appear to evolve fast. They have been shown to evolve under positive selection in some lineages and we predict this is likely the case in other lineages, and they are later lost in some lineages revealing that they are likely involved in arms races. However, all of the above-mentioned pressures can lead to arms races. Male–male competition leads to an arms race between males. Sexual antagonism leads to an arms race between sexes and genetic conflicts lead to an arms race between the genome and selfish elements. It is impossible to differentiate between selection pressures by just looking at sequence evolution. So the fact that the additional gene duplicates are evolving fast and sometimes are lost does not provide evidence in favor of one or another selective pressure as the reason for retention of these gene duplications. Functional analyses are needed to elucidate this (Chapter 4).

The amount of recurrent duplications, however, reveals that the pressures leading to the retention of these duplicated genes should be strong and ubiquitous. It has been recently suggested that, if selection is strong, it might be affecting the parental genes prior to gene duplication leading to intralocus sexual antagonism that is resolved through subsequent gene duplication and acquisition of testis expression (Gallach and Betrán 2011). Acquisition of new expression patterns is facilitated by duplication through retrotransposition because this mechanism is accompanied by the relocation of the duplicate (Gallach and Betrán 2011). As described in Chapter 3 retrogenes can insert next to a testis-specific regulatory region (Sorourian, et al. 2014). I see that retrotransposition is the main mechanism underlying *Dntf-2* and *Ran* duplications, fitting this model really well. If this model is correct, it means that positive selection is acting at every step of the process. Additional information about the pattern of expression of the newly described duplicates is needed to further support this model.

I posed several questions at the beginning of this chapter: What is the pattern of duplication of *Dntf-2* and *Ran* when we consider all the data? Does the retroduplication bias still hold? Can we come up with hypotheses to explain this bias, if it exists? I have answered all of them.

CHAPTER 3

TESTES TRANSCRIPTION AND REGULATORY REGIONS OF *DNTF-2R* AND *RAN-LIKE* RETROGENES AND CHARACTERIZATION OF THE LACK OF DIRECTIONALITY OF THE TESTIS-SPECIFIC *β2-TUBULIN* GENE UPSTREAM ELEMENT 1 (β2-UE1)

3.1. Introduction

Several nuclear transport factors, including *Dntf-2, Ran* and several *importins*, have been recurrently duplicated in different Drosophila lineages. A majority of these duplications involves the mechanism of retroduplication (Bai, et al. 2007; Tracy, et al. 2010; Phadnis, et al. 2012; see also Chapter 2). The available expression data for some of the new nuclear transport genes indicates that they acquire testis-biased or testis-specific expression (Betrán and Long 2003; Tracy, et al. 2010; Phadnis, et al. 2012). The acquisition of testis expression by retrogenes occurs despite retroposition (i.e., duplication by means of an RNA intermediate without duplication of the parental regulatory regions). In addition, there is a high level of turnover for these genes; some are later lost and some go through additional duplications (Tracy, et al. 2010; Phadnis, et al. 2012, and data from Chapter 2). A high turnover of these nuclear transport genes as well as testis-biased expression and fast evolution (often driven by positive selection) of additional components of nuclear transport (nucleoporins and *RanGAP*) have been hypothesized to occur to defend the male germline against selfish segregation distortion systems and/or transposable elements (Presgraves 2007; Presgraves and Stephan 2007; Tracy, et al. 2010; Phadnis, et al. 2012). Whatever the function of the new genes might be, it is performed in testis and it is important to understand how new genes gain testis expression.

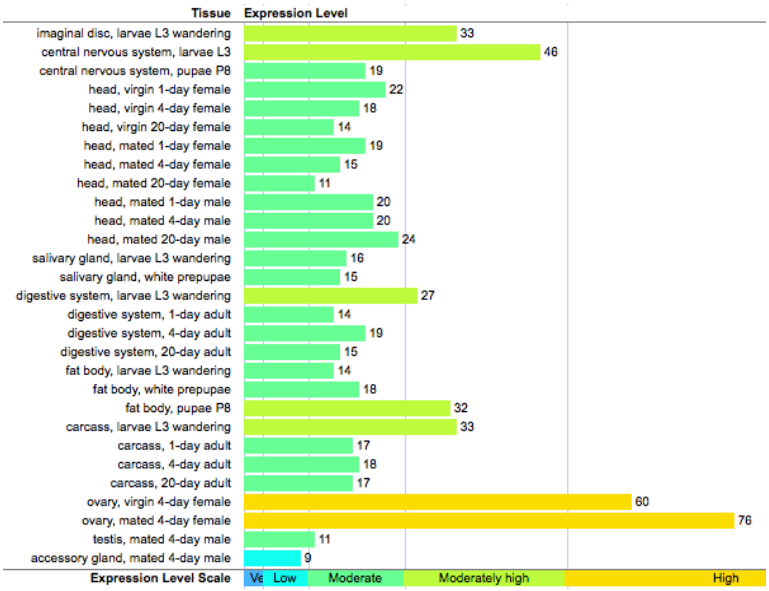
In this chapter I will address Objective 2 of my Dissertation. A long-standing question about retrogenes is how they acquire their often testis-specific regulatory regions. I study in detail the regulatory regions of *Dntf-2r* and *Ran-like* to provide two examples of how this occurs. One question we can answer is: Is retroduplication facilitating the acquisition of testes regulatory regions? How? How does the way the regulatory regions are acquired fit into/facilitate the proposed models of gene duplication?

I will first review what is known about expression patterns and regulatory regions of *Dntf-2r* and *Ran-like*. I will then describe the results of my work focusing on 1) detailed analyses of testes expression patterns for *Dntf-2r* and *Ran-like* using *in situ* hybridization and constructs with fused fluorescent proteins, 2) identification and comparison of regulatory regions that are required to drive testis expression of *Dntf-2r* and *Ran-like*, and 3) investigation of bidirectional transcriptional control by testis-specific cis-regulatory region of *β2 tubulin* gene. Part of this work is included in a recent publication in the journal Molecular Biology and Evolution (Sorourian, et al. 2014). The details obtained in these experiments will answer the questions above.

*3.1.1. Expression patterns of parental genes (*Dntf-2 *and* Ran*) and their retroduplicates (*Dntf-2r *and* Ran-like*): a review*

Both retrogenes, *Dntf-2r* and *Ran-like*, are present in *D. melanogaster* and this makes them amenable to the experimental characterization of their regulatory regions and to the detailed study of their expression patterns. *Dntf-2r* is derived from *Dntf-2* transcript A and *Ran-like* is derived from the only transcript of *Ran* gene (Bai, et al. 2007). *D. melanogaster Dntf-2r* and *Ran-like* exhibit a male germline-biased expression (Betrán and Long 2003; Chintapalli, et al. 2007; Tracy, et al. 2010). Recent detailed expression

profiles from modENCODE obtained using mRNA seq data from different fly tissues and stages (Graveley et al. 2011.4.13 Personal communication to FlyBase) confirm and complement these findings (Figure 3.3; Figure 3.4). *Dntf-2r* and *Ran-like* exhibit very high expression in male testis, high or moderate expression in imaginal discs, low or very low in fat body, accessory gland, salivary glands, digestive system and central nervous system. The expression of both genes is first detected in the third instar larvae (the assumption is that this expression is mostly due to male gonads) and is found in male adults but not females. Both parental genes are expressed in all tissues and at all developmental stages in agreement with the need for nuclear transport and *Dntf-2* and *Ran* functions in every cell (Chapter 1). Overall, *Ran* (Figure 3.2) shows higher levels of expression in all tissues and developmental stages compared to *Dntf-2*. *Dntf-2* is generally expressed at moderate levels, but with increased expression in embryos and female ovaries (Figure 3.1; Graveley, et al. 2011).



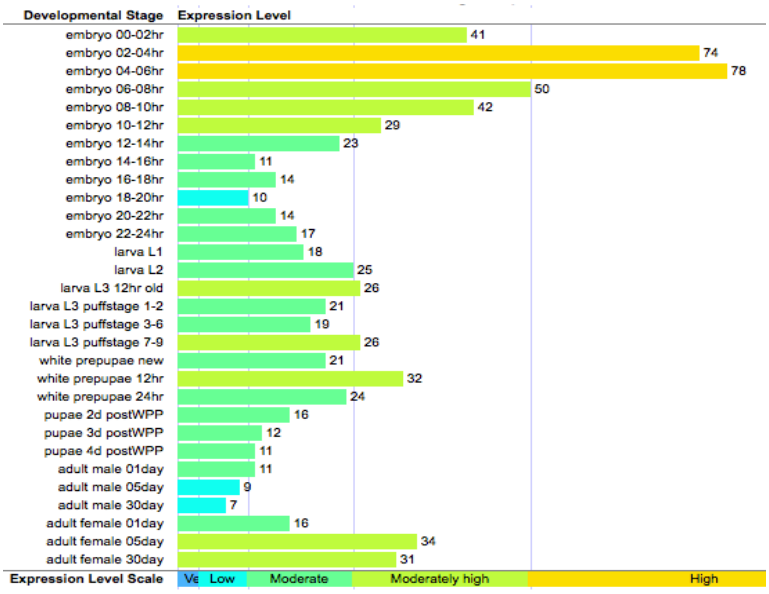


Figure 3.1 - Expression profiles from modENCODE obtained with mRNA seq data for Dntf-2 in different fly tissues (up) and during Drosophila development (bottom; FlyBase.org)

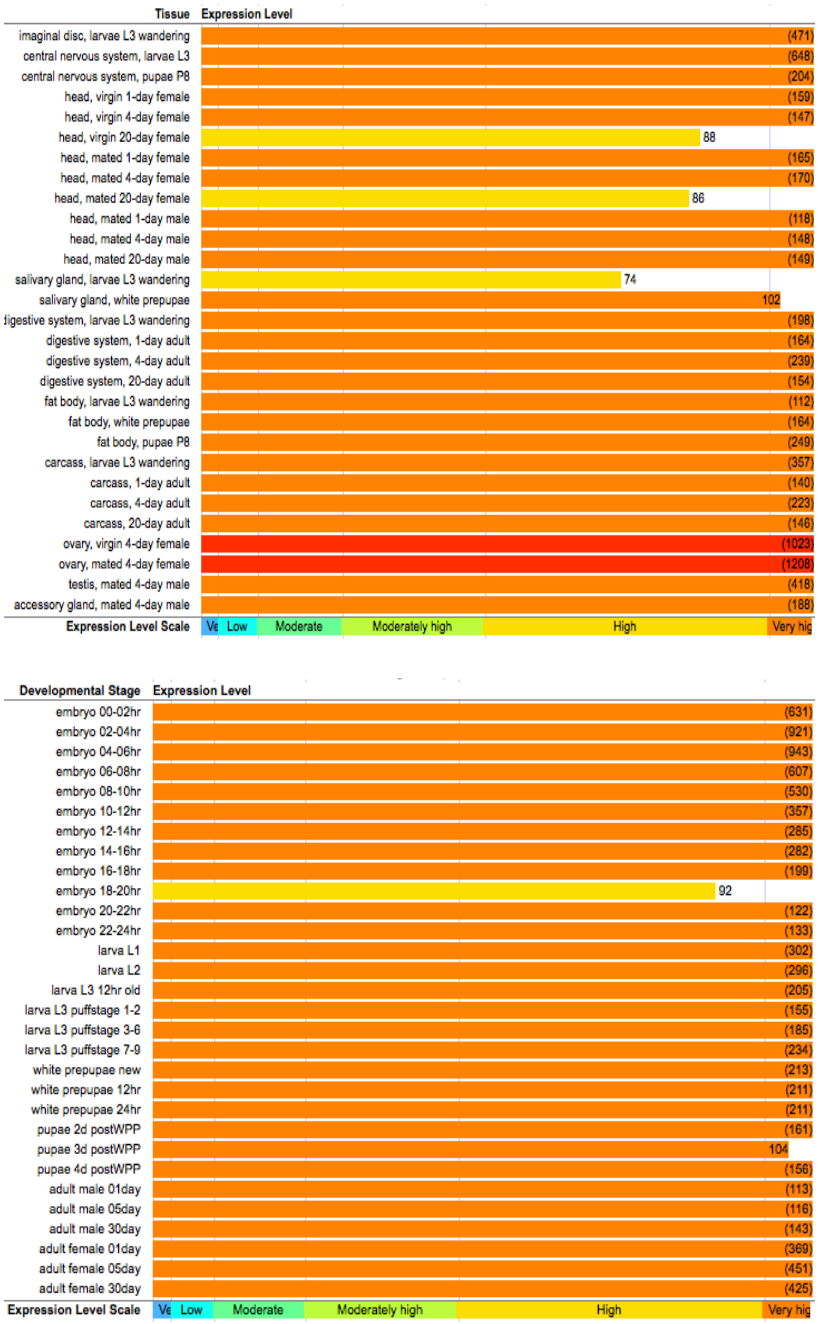


Figure 3.2 - Expression profiles from modENCODE obtained with mRNA seq data for Ran in different fly tissues (up) and during Drosophila development (bottom; FlyBase.org)

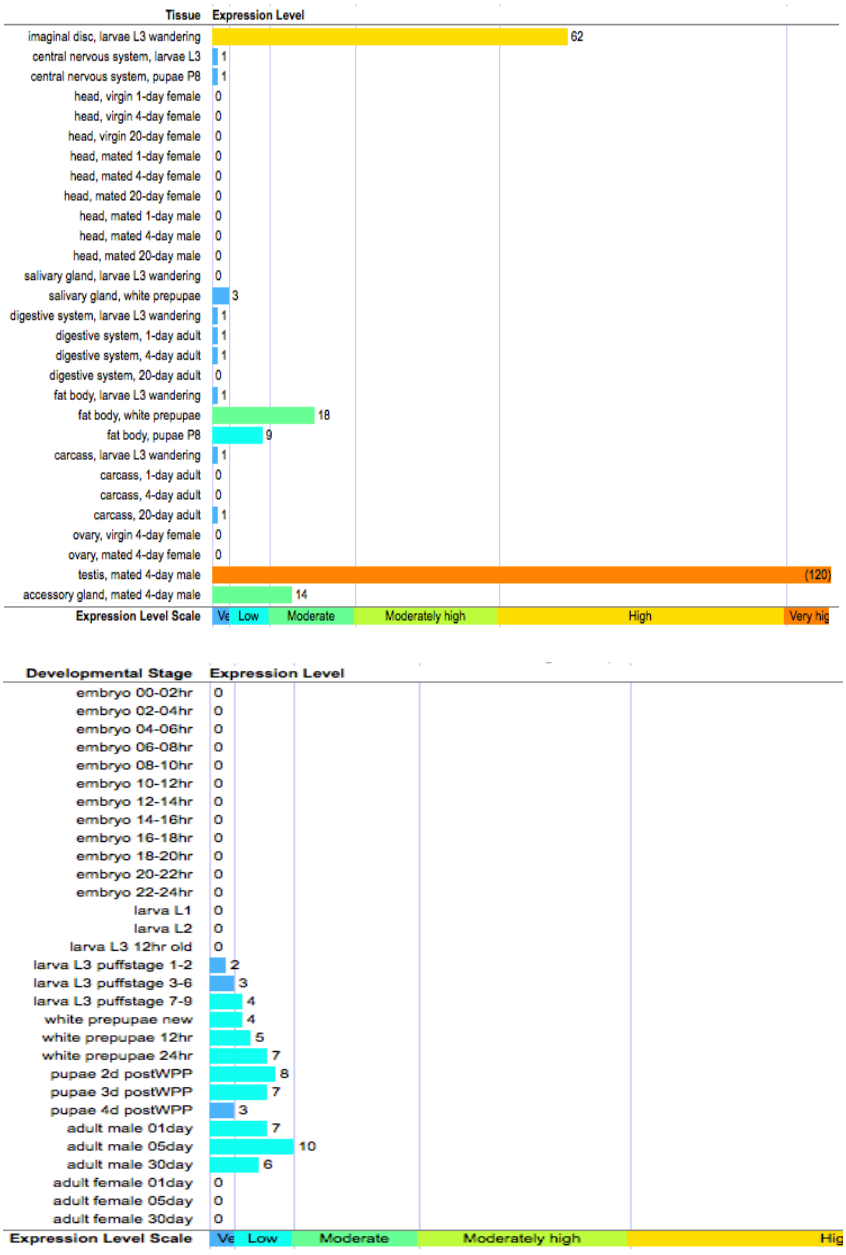


Figure 3.3 - Expression profiles from modENCODE obtained using mRNA seq data for Dntf-2r in different fly tissues (up) and at different Drosophila developmental stages (bottom; FlyBase.org).

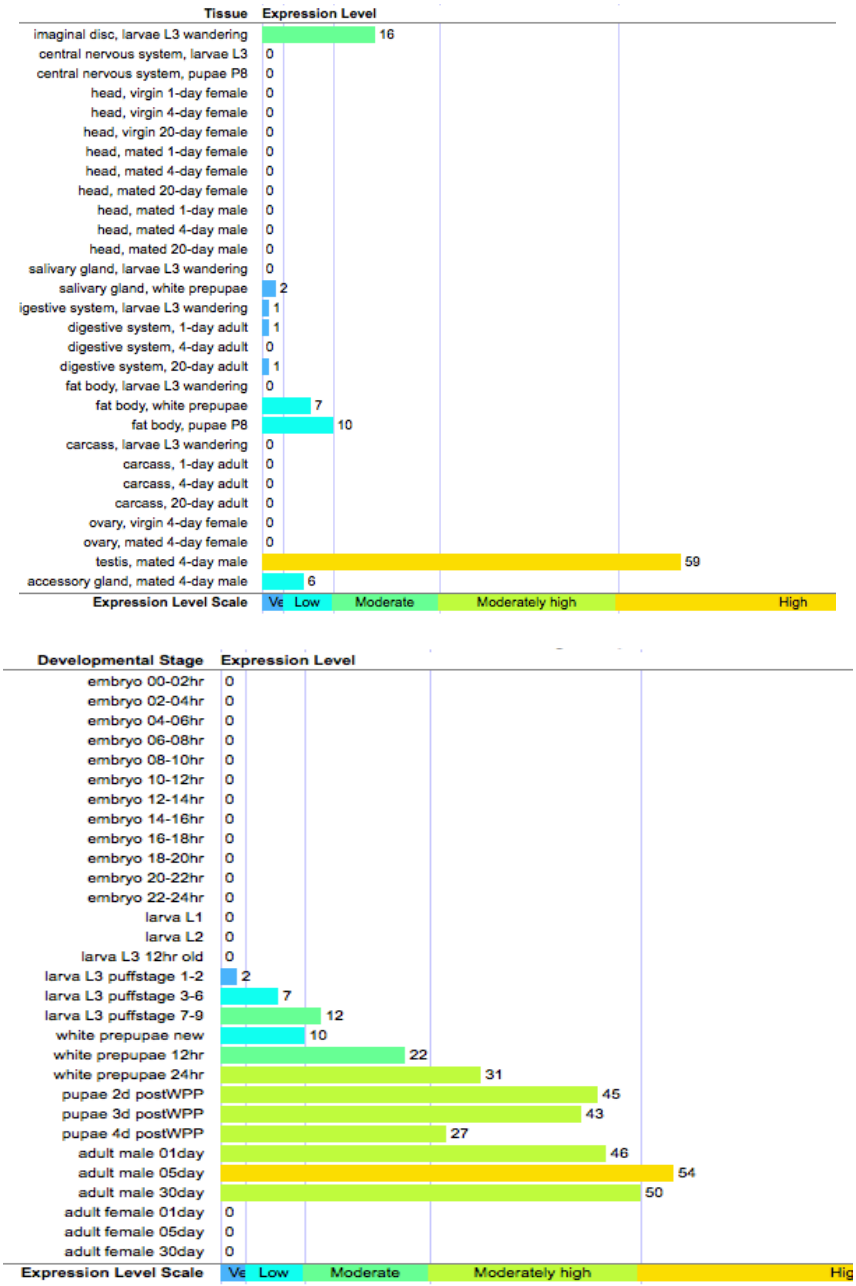


Figure 3.4 - Expression profiles from modENCODE obtained using mRNA seq data for Ran-like in different fly tissues (up) and at different Drosophila developmental stages (bottom; FlyBase.org).

3.1.2. Regulatory regions of Dntf-2r*: a review of previous work in the Betrán laboratory*

Sequence analysis of the region upstream of *Dntf-2r* transcription start site (TSS) from multiple species identified a short (i.e., 14 bp) cis-regulatory region. This sequence is similar (57% identity) to the β2 tubulin upstream element 1, (β2-UE1) needed for testis expression of the *β2-tubulin* gene (Michiels et al. 1989; Betrán et al. 2003). The region upstream of *Dntf-2r* was cloned together with the 5'UTR and coding region of the gene and fused to a reporter gene (i.e., EGFP) to study *Dntf-2r* regulatory region. The upstream region included 151 bp (i.e., all the region up to the next gene). To narrow down the testis-specific cis-regulatory region needed for *Dntf-2r* testis expression, shorter constructs were also made and transformed. The β2-UE1 promoter motif (Figure 3.2) was shown to be necessary and sufficient to drive testis expression of *Dntf-2r* in *D. melanogaster* (Sorourian, et al. 2014). In addition, the upstream region of *Dntf-2r* harbors a sequence that is identical to the 7 bp quantitative element of the *β2 tubulin* gene. The 14 cis-regulatory element has also been described in the gene *Sdic* (64% identity) and others (reviewed in Sorourian, et al. (2014) and Figure 3.5A). The low identity and length variation in this regulatory motif suggests a low specificity of the transcription factors that bind the motif. Moreover, the separation between the upstream element and the quantitative element in *Dntf-2r* is larger (8 bp more) than in the *β2 tubulin* gene. The same two motifs are conserved in *D. simulans*, *D. sechellia* and *D. mauritiana* (Sorourian, et al. 2014). Importantly, the examination of the *Dntf-2r* upstream region in species that do not have the retrogene (*D. yakuba, D. teissieri* and *D. erecta*) revealed that β2-UE1 is partially conserved and the quantitative element is completely conserved (Figure 3.5B). This finding suggests that a testis-specific regulatory region existed before the insertion of *Dntf-2r* and prompted my analyses of this region before *Dntf-2r* insertion (Figure 3.5B). I also studied the regulatory region of *Ran-like*.

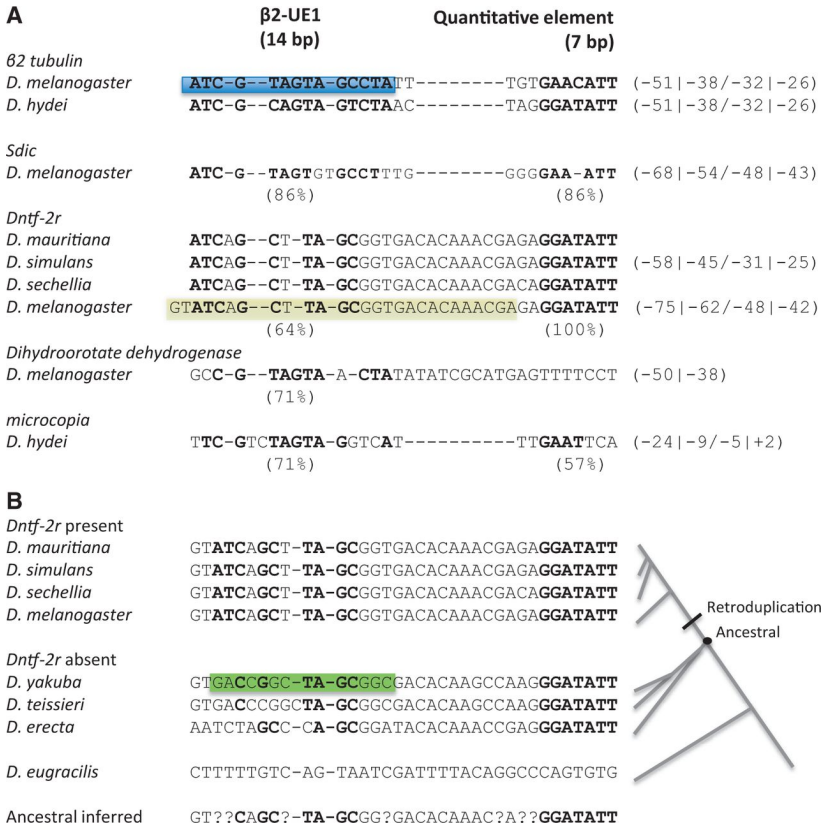


Figure 3.5 - (A) Percent identity of the upstream region of several testis-specific genes to the known 14-bp β2-UE1 motif and the 7-bp motif with quantitative effect of the *β2 tubulin* gene. Identical bases are shown in bold. Negative numbers show the location of both elements as distance to the TSS. For example, β2-UE1 motif of the *β2 tubulin* gene in *Drosophila melanogaster* spans the base pairs from −51 to −38 and the 7-bp motif with quantitative effects in the same gene and species spans the base pairs from −32 to −26. Yellow highlighted region shows the sequence required for testis-specific expression of *Dntf-2r* in *D. melanogaster*. (B) *Dntf-2r* upstream region in species where the gene is present and the same region in species where the gene is absent. The nucleotide state at the ancestral node was inferred when the same nucleotide was observed in species where *Dntf-2r* is present and in species where *Dntf-2r* is absent and not inferred (?) otherwise. I highlight in green the region from *D. yakuba* used to make a construct to show the current regulatory potential of the 14 bp in this species. This figure is from Sorourian et al. (2014) and consolidated from that study and work of others (Michiels et al. 1989; Lankenau et al. 1994; Yang et al. 1995; Nurminsky et al. 1998).

3.1.3. The study of directionality of β2-UE1 cis-regulatory motif

The β2-UE1 cis-regulatory motif is known to be necessary and sufficient to drive testis expression (Michiels *et al.* 1989) and was observed to have good hits in genomic analyses close to testis-specific genes, including several retrogenes in *D. melanogaster* (Sorourian, et al. 2014). Interestingly, these hits occur in either orientation suggesting that β2-UE1 motif can drive expression in both directions. To test this possibility, β2-UE1 cis-regulatory motif was flipped with respect to the transcription start site to see if the β2-UE1 cis-regulatory motif can still drive testis-specific expression This work was done in collaboration with a visiting schollar, Dr. Fulya Özdil from Kemal Üniversitesi, Agricultural Biotechnology, Turkey. Fulya made the β2-UE1 flipped clone for injection with my guidance. If the β2-UE1 cis-regulatory motif works bidirectionally, it could help the acquisition of testis expression of retrogenes and would be relevant to the questions posed above.

3.2. Materials and Methods

3.2.1. The study of transcriptional regulation of Dntf-2r *and* Ran-like: experimental design

Since a comparison of the regions between species that have and do not have *Dntf-2r* revealed partial and complete conservation of β2-UE1 and the quantitative elements in closely related species (*D. yakuba, D. teissieri* and *D. erecta*) that do not have the retrogene (Figure 3.5B), I made a construct for fly transformation including the 14 bp of the *D. yakuba* regions and a reporter EGFP protein to test if this region was able to drive testis-specific transcription in *D. melanogaster*.

I also made constructs to narrow down the testis-specific cis-regulatory region of *Ran-like*. I used genetic transformation and the red fluorescent protein reporter gene fused to *Ran-like* in *D. melanogaster* to be able to study protein localization as well (See Chapter 4). These constructs were examined to narrow down the regulatory regions that drive *Ran-like* expression in testes.

3.2.2. Drosophila stocks and fly handling

Several *D. melanogaster* strains were used in order to study the regulatory regions of *Dntf-2r*, *Ran-like* and *β2-tubulin* genes. A wild type strain of *D. melanogaster*, Besançon strain (Isofemale line captured in Besançon, France, and provided by Dr. P. Gilbert; Betrán and Long 2003) was used to amplify regions of DNA for cloning. Additionally, mutant strains of *D. melanogaster* were also utilized. The white mutant strain w^{1118} (Genetic Services, Cambridge, MA) and balancer stocks for second chromosome (w[*]; sna[Sco]/CyO, S[*] bw[1]) and for the third chromosome (w[*];2.3/Tm6b;sb). The balancer stocks (provided by Dr. A. Greenberg) were used to fix the P element insertions in the transformed flies. All stocks were maintained on corn medium and at room temperature.

3.2.3. Genomic DNA extraction

Genomic DNA was extracted from 30 young adult flies from the *Besançon* strain. The Wizard® Genomic DNA Purification Kit from Promega (Madison, WI) was used for the extractions.

3.2.4. Retrogene fragments amplification by PCR

To characterize *Ran-like* regulatory region, fragments including different lengths of the *Ran-like* upstream region (i.e., upstream of the transcription start size or TSS), the 5'UTR and the *Ran-like* coding region were amplified by PCR. PCR conditions were 1 cycle of 2 min at 95°C, followed by 35 cycles of 30 sec at 95°C, 30 sec at 52°C, and 1 min at 72°C, and finishing with 7 min incubation at 72°C. The longest fragment contained 500bp upstream of the TSS, the next fragment contained 100bp upstream of the TSS and the third fragment did not contain any upstream region (i.e., contained only the 5'UTR and the *Ran-like* coding region; See Figure 3.6).

3.2.5. Cloning and plasmid preparation for injection

Dntf-2r-EGFP constructs had been previously produced in the lab and were already available for analysis (Sorourian, et al. 2014). The complete Dntf-2r coding region and variable lengths of the upstream regions had been amplified from genomic DNA and cloned into the plasmid pEGFP1 (U55761; Clontech, Mountain View, CA) to put the Dntf-2r in frame with the EGFP gene and generate a fluorescent fusion protein. These regions containing different lengths of the putative regulatory region, 5'UTR of Dntf-2r, Dntf-2r-EGFP fused coding regions and the SV40 polyadenylation site were then further cloned into the P element Drosophila transformation vector – pCaSpeR 4 (X81645) and used for fly transformation. The flies carrying the longest upstream region of 151 bp (Sorourian, et al. 2014) were used in expression studies.

Similar to *Dntf-2r*-EGFP fusion and regulatory regions analyses (Sorourian, et al. 2014), the complete *Ran-like* coding region and variable lengths of the upstream region were amplified from genomic DNA and cloned into a destination plasmid. In this case, I

used pRed H-Pelican plasmid (provided by Dr. Barolo Lab, La Jolla, CA) to put *Ran-like* in frame with the Ds.RedT4 gene (red fluorescent protein; DsRed.T4) and generate a red fluorescent fusion protein. These clones containing different lengths of the putative regulatory region, 5'UTR of *Ran-like*, and *Ran-like*-DsRed.T4 fused coding regions. Before *Ran-like* region was inserted the TATA box was removed from the original pRed H-Pelican construct using AgeI and XhoI (Promega, Madison, WI) restriction enzymes. The three different inserts containing *Ran-like* and the diverse upstream regions were also digested using the same restriction enzymes and after cloned in frame with DsRed.T4 into the P element Drosophila transformation vector (pCaSpeR 4; X81645) and used for fly transformation (Figure 3.6). To clone the fragments into the plasmid

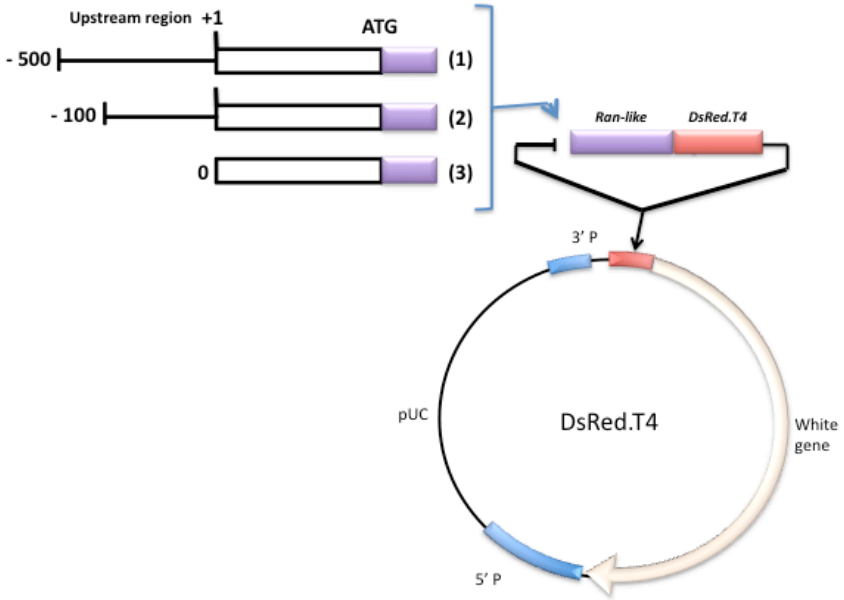


Figure 3.6 – Steps of the P-element construct production for fly transformation and the study of the *Ran-like* regulatory region. Representation of the three constructs sizes prepared for injection. (1) 500bp upstream region of *Ran-like*; (2) 100bp upstream region of *Ran-like*; (3) Construct with no upstream region.

Two other constructs were produced to study the regulatory regions of *Dntf-2r.* For these cloning, I set up the Gateway system in our laboratory. The 14bp β2-UE1-like sequence of *D. melanogaster* was replaced by the orthologous 14 bp from *D. yakuba* to understand if this sequence drives testis-specific expression (Figure 3.7). The region of Dntf-2r was amplified using the forward primer 5'CACC**GACCGGCTAGCGGC**GACACAAACGAGAGGATATTTG3' that added the *D. yakuba* sequence (bold) to the *D. melanogaster* region and the reverse primer 5´TTTAGTTCAAGTATATACGGGGTA3'. The PCR product was cloned with topoisomerase-catalyzed reaction using pENTRYTM Directional TOPO cloning kit (pENTR/D-TOPO Cloning Kit, Invitrogen catalog # K240020, Invitrogen, Carlsbad, CA). The forward primer added the necessary recombination site to the PCR product for directional cloning into the entry clone using the Gateway system. Colonies were screened by colony PCR using primers flanking the gene folloId by sequencing of the positive clones. Miniprep for a good clone was performed making this clone our master entry clone for recombination into the UASt-GFP P-element destination vector. To perform the recombination into the destination vector I used the Gateway LR Clonase Enzyme Mix (Invitrogen, Carlsbad, CA). Transformation was performed using Library Efficiency DH5alpha competent cells and the colonies where then sequenced by PCR to find positive clones. One good clone was chosen after sequencing and sent for fly transformation (Figure 3.9A).

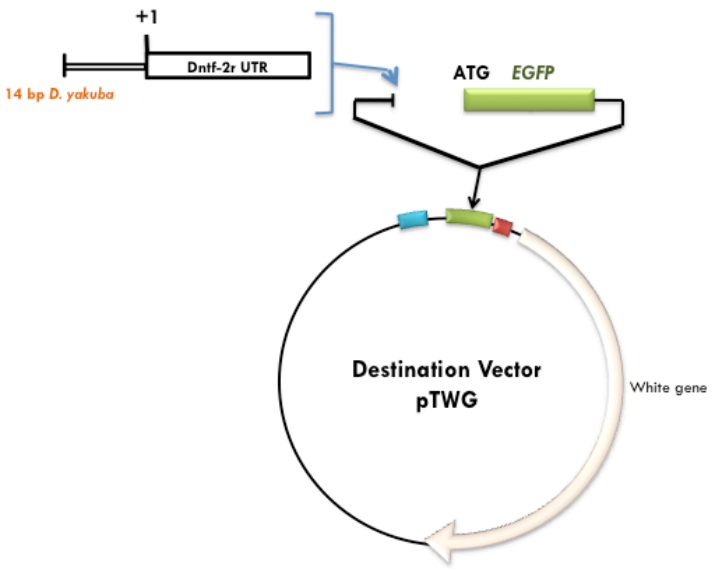


Figure 3.7 – A construct designed to test if the 14bp sequence present in the *D. yakuba* lineage can drive testis expression by itself.

Another construct was made that included the described regulatory regions of the β2-tubulin gene and its 5'UTR (Michiels, et al. 1989) but in which the β2-UE1 element was flipped. This region was introduced upstream of EGFP in the pCaSpeR 4 transformation vector and used for fly transformation (Figure 3.8; Figure 3.9B).

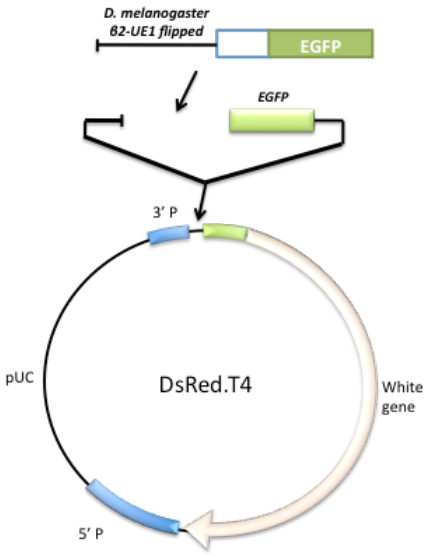


Figure 3.8 – A construct designed to test the lack of directionality of 14bp sequence of the flipped β2-UE1. β2-tubulin 5'UTR was inserted in front of EGFP.

More details of the two constructs introduced above are shown below (Figure 3.9).

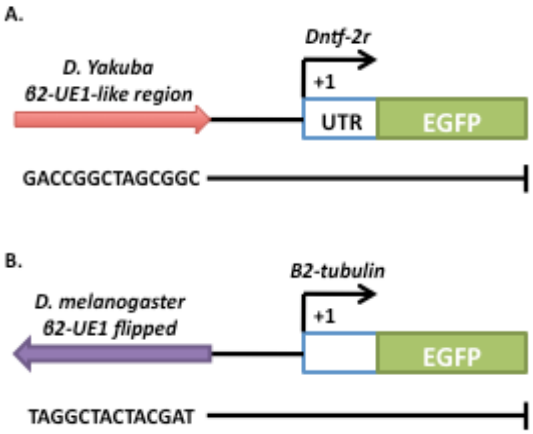


Figure 3.9 - (A) β2-UE1-like region in *D. yakuba* with the 5'UTR of *Dntf-2r* from *D. melanogaster* and EFGP. (B) β2-UE1 flipped sequence from *D. melanogaster* with the 5'UTR of β2-tubulin and EGFP.

3.2.6. Plasmid injections

P-element plasmids were sent for injection to Genetic Services, Inc. (Cambridge, MA). The white mutant stock w^{1118} was used for injection. A helper plasmid containing the P-element transposase gene was also injected along with the desired plasmid in order to provide the transposase function to excise the P-element region from the plasmid. After excision there is the insertion of the plasmid in the genome of the embryo's germline.

3.2.7. Fixation in the P-element transformants

Injected 1st instar larvae provided by Genetic Services Inc. (Cambridge, MA) were allowed to grow at room temperature until pupae eclosion. Single newly emerged virgin males and female flies were crossed with virgin w^{1118} flies. Any progeny with light or dark orange eye color indicates a successful insertion of the P-element plasmid. Each orange eye fly was crossed once again with w^{1118} flies, Curly wing flies or Stubble flies, balancer flies. This step was done in order to identify the chromosome of insertion of the plasmid and to fix the chromosome with the insertion. Each cross-contained one transformant male with 2 balancer virgin females or one transformant virgin female with 2 balancer males. The balancer phenotype for the 2nd chromosome is Curly wings and for the 3rd chromosome is Stubble. The crosses with transformant male with virgin females that gave rise to all white-eyed males and orange-eyed females in the next generation indicated an insertion in the X chromosome and were fixed using white mutant flies. After fixing, the lines were classified as independent insertions if they arose from different individuals or if they arose from the same individual but map to separate chromosomes.

3.2.8. Phase contrast microscopy of testes

To examine various stages of spermatogenesis, testes from male pupae and young males from w^{1118} and from the different transformant flies were dissected in 1% PBS (130 mM NaCl, 7 mM Na2HPO4, 3 mM KH2PO4, pH 7.4, Fisher Scientific, New Jersey). Testes were mounted in between two bridges to allow for the observation of the entire testis and others were squashed lightly under the light of a cover slip. The preparations were examined under phase contrast optics using (Olympus BX51TRF florescent microscope).

3.2.9. Fluorescence microscopy of testes

Testes were dissected from young transformant males carrying different TSS upstream regions and from w^{1118} males to allow a comparison of the fluorescence level. The tissues were dissected in 1% PBS. After dissection some testes were immediately mounted in 1% PBS. After mounting they were observed under the fluorescence microscope, Olympus BX51TRF florescent microscope, setting the UV exposure time manually at a level established by comparison to the control strain (w^{1118}). Sequentially, the remaining testes were fix in 4% paraformaldehyde for 30 minutes at room temperature. The fixed testes were washed twice with 1% PBS, mounted in FluorGlo® (Valley Scientific, Mayville, NY) and observed under the fluorescence microscope using again the mutant strain w^{1118} to control the exposure settings.

3.2.10. In situ *hybridization*

To compare the details of testis-specific expression profiles of parental *Dntf-2* and *Ran* genes with their respective retrogenes, *Dntft-2r* and *Ran-like*, I analyzed the transcription pattern of *Ran*, *Ran-like*, *Dntf-2* transcript A, *Dntf-2* transcript B, and *Dntf-2r*. Whole-mount *in situ* hybridizations using specific probes for *Ran*, *Ran-like*, *Dntf-2* transcript A, *Dntf-2* transcript B, *Dntf-2r* and *CG3927* as positive control were performed. I performed *in situs* using testes from young males and pupae gonads.

I set up the protocol for *in situ* hybridization of different tissues in our laboratory. This required multiple optimization steps. Below I described the protocols used for testes *in situs* with all incorporated changes. *In situ* hybridization of whole testes of young males of the *Besançon* strain and w^{1118} strain was performed to detect the transcript of *Dntf-2*, *Dntf-2r*, *Ran* and *Ran-like* following the protocol described by Morris et al. (2009). *In situ* hybridization in larvae was performed following the same protocol with a series of optimizations done in our lab. I used DNA probes instead of RNA probes. Nevertheless, RNA probes where tested to confirm that the observed expression did not depend on the type of probes used. To denature the probe, it was heated at 100ºC for 10 minutes. The prehybridization and hybridization steps were carried out at 45ºC. The overnight incubation at 4ºC was done with 0.1% BSA in PBST and the anti-digoxygenin antibody. After the color development, the reactions were stopped with PBST washes (four times for 10 minutes each). After the four washes, all the PBST was removed and 30% glycerol in PBST was added for 30 minutes, followed by 50% glycerol, and, finally, 70% glycerol.

For *Ran-like* and *Ran* probe production, specific primers were used to amplify each gene from *D. melanogaster Besançon* genomic DNA. DNA and RNA probes were

made according to protocol described by Morris et al. (2009). For *Dntf-2* and *Dntf-2r* only DNA probes were produced. A sense oligoprimer (5'CTTTTTTTCGGATCGGAAACTCAATCGTACCCCGTATATACTTGAACTAAAATGTC TCTGAATCTGCAGT3') allowed us to label the complementary DNA strand using the random priming procedure. This procedure was designed to detect *Dntf-2r* transcript specifically as the oligo corresponds to a 5'UTR region that is different from the four transcripts of the parental gene, *Dntf-2-RA*, *Dntf-2-RB*, *Dntf-2-RC* and *Dntf-2-RE*. For the parental *Dntf-2* gene, specific probes were design for transcript A and B (Table 3.1). *Ran* probes were designed in the coding region of the parental *Ran* and coding region with part of the 3' UTR of *Ran-like*. *CG3927* was chosen as control of testes hybridization due to its specific pattern in this tissue that alloId us to rule out background problems with our DNA probes. This probe was made using the primer set design by Helen White-Cooper's laboratory (Flyted.org).

Table 3.1 – Oligos used for PCR to produce probes. They were designed to amplify each specific gene or transcript. An oligo was designed to detect *Dntf-2r*.

Probe	**Sequence 5' to 3'**
Dntf-2A 3'	TGAGACCGCTGGCTGGCTTTG
Dntf-2A 5'	CTCAACATCCACAACTCTGC
Dntf-2B 3'	GCTCTCGTCTCGTGTCTG
Dntf-2B 5'	TATTCCGACTCTCGCTGC
Dntf-2r	CTTTTTTTCGGATCGGAAACTCAATCGTACCCCGTATATACTTGAACT AAAATGTCTCTGAATCTGCAGT
Ran 3'	GCTTACTTTGTGCCCATGG
Ran 5'	CCGAGCCTGCAATTTTACAC
Ran-like 3'	GGATTGGCAGGCGCAGATCGAGC
Ran-like 5'	CCGAGCCTGCAATTTTACACC

Probe labeling was performed using DIG-High Prime DNA Labeling and Detection Starter Kit I (Roche, Indianapolis, IN).

3.3. Results

3.3.1. Dntf-2r *and* Ran-like *expression pattern compared to the parental genes*

I performed *in situs* using testes from young males and pupae gonads. Figure 3.10 shows Drosophila testes and the stages of spermatogenesis to help with the interpretations of our results. The *in situ* results in adults Figure 3.7, show that *Ran* is strongly transcribed at the tip, hub and germline cells, and in early dividing cells, at 8-cell stage of mitotic cells. *Ran-like* is strongly transcribed later, at 16-cell stage spermatocytes and after, with no expression in the stem cells. Therefore, *Ran* and *Ran-like* show complementary transcription in testes.

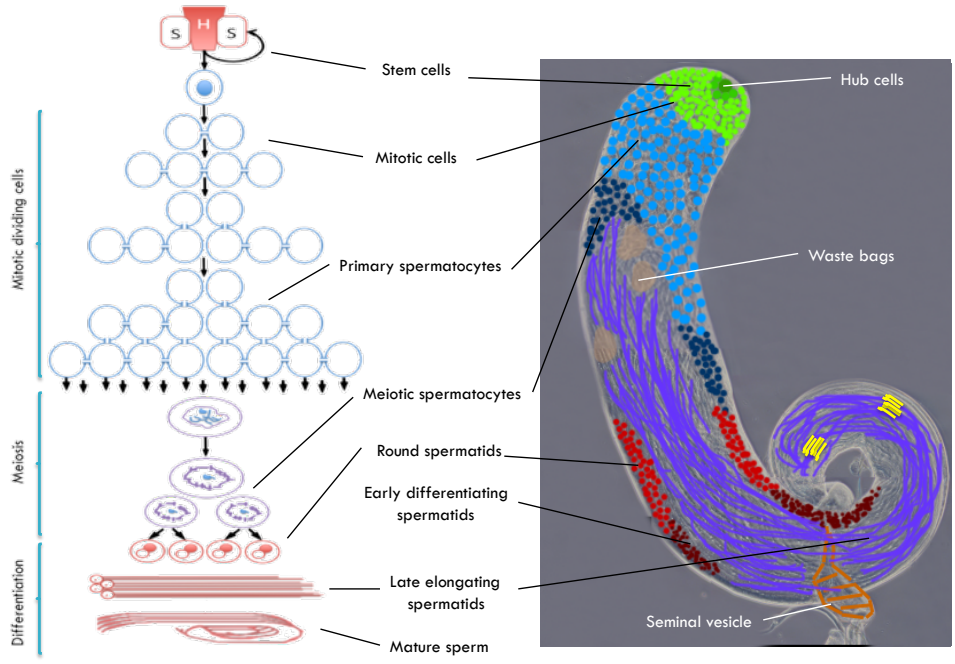


Figure 3.10 – Different stages of spermatogenesis starting from the tip of the testis with the germ cells that transition to mitotic dividing cells, then to meiotic dividing cells and finally to the sperm bundles elongating and the formation of mobile sperm.

Transcript A and B for *Dntf-2* show different expression patterns but they add up to transcription throughout the testis. *Dntf-2-RA* is present all over the testis and *Dntf-2-RB* is being transcribed only in spermatocytes. *Dntf-2-RB* is so slowly maturing that it appears mostly in the nucleus. *Dntf-2r* transcription pattern is quite similar to the pattern of *Ran-like* and exactly the one expected from a β2-UE1-like element (i.e., analogous to the pattern of *β2-tubulin* gene in testis (Michiels, et al. 1989; Santel, et al. 2000). *Dntf-2r* is not present at the tip of the testis and first appears in the spermatocytes. So, *Dntf-2* and *Dntf-2r* show overlapping transcription in testes with *Dntf-2* being more broadly transcribed. As a negative control I used fly gut tissue where *Ran-like* and *Dntf-2r* show no expression unlike *Ran* and *Dntf-2* transcripts. These results are consistent with modENCODE RNAseq data (FlyBase.org) shown above (Figures 3.1 and 3.2).

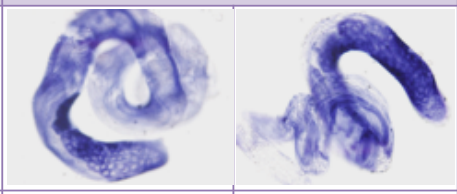
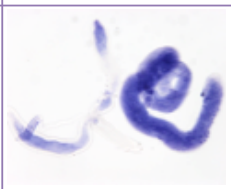
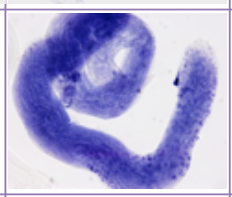
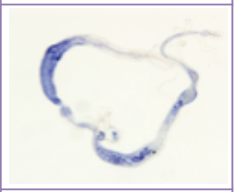
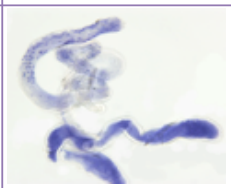
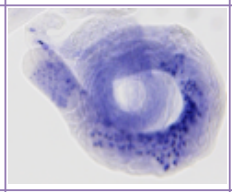
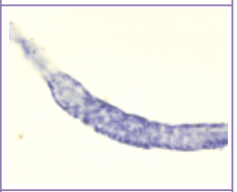
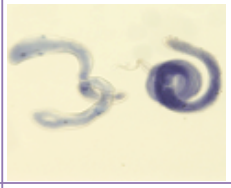
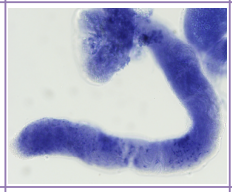
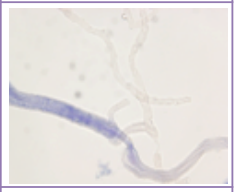
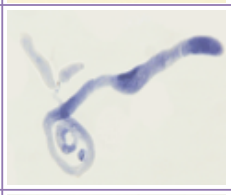
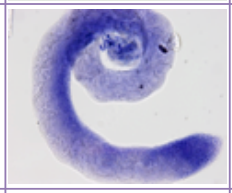
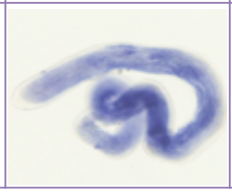
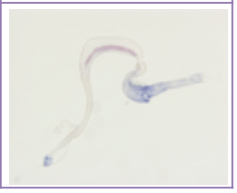
Genes		Testis		Gut
Control (CG3927)				
Dntf-2	***Dntf-2A***			
	Dntf-2B			
Dntf-2r				
Ran				
Ran-like				

Figure 3.11 – Results for the *in situ* hybridization for *Ran*, *Ran-like*, *Dntf-2* transcript A, *Dntf-2* transcript B, and *Dntf-2r* in young adult testis are shown. Gut is also shown for comparison. *CG38927* that has a known expression pattern in spermatogenesis was used as a positive control.

3.3.2. Dntf-2r-EGFP *expression and dissection of regulatory regions*

Sorourian et al. (2014) concluded that a 27 bp region including the β2-UE1-like element of *Dntf-2r* is needed to drive the expression of the *Dntf-2r*-EGFP fusion gene in testis. As explained above, they used different constructs containing the *Dntf-2r*-EGFP fusion and different lengths of the upstream region of the gene (Figure 3.12).

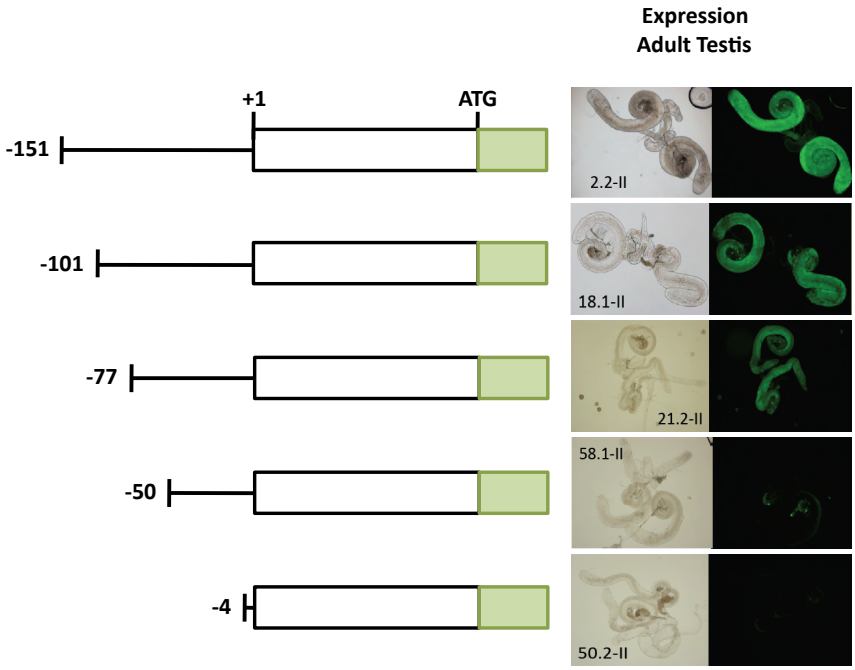


Figure 3.12 – The expression or lack of it of Dntf-2r-EGFP fluorescence tag protein in testis for different constructs is shown. A representative for every construct is shown (Sorourian et al. 2014). The longest construct goes up into the gene upstream of *Dntf-2r* (i.e., *bicoid stabilizing factor* or *bsf*; not shown). Clear field pictures of the testes are also shown.

Since a comparison of the upstream regions between species that have and do not have *Dntf-2r* revealed partial conservation of β2-UE1 and complete conservation of the quantitative element, the 14bp β2-UE1-like sequence of *D. melanogaster* was

replaced by the orthologous 14 bp of *D. yakuba* (without retrogene) to test if this sequence can drive testis-specific expression. From the fusion of the *D. yakuba* region with *D. melanogaster* upstream region, 5' UTR and EGFP, testis-specific fluorescence is observed in most of the independent transformants (Figure 3.13). The level of expression is lower than that for the *D. melanogaster* β2-UE1-like region, but consistent across constructs. The results suggest that this region drove testis expression before *Dntf-2r* insertion (Sorourian, et al. 2014).

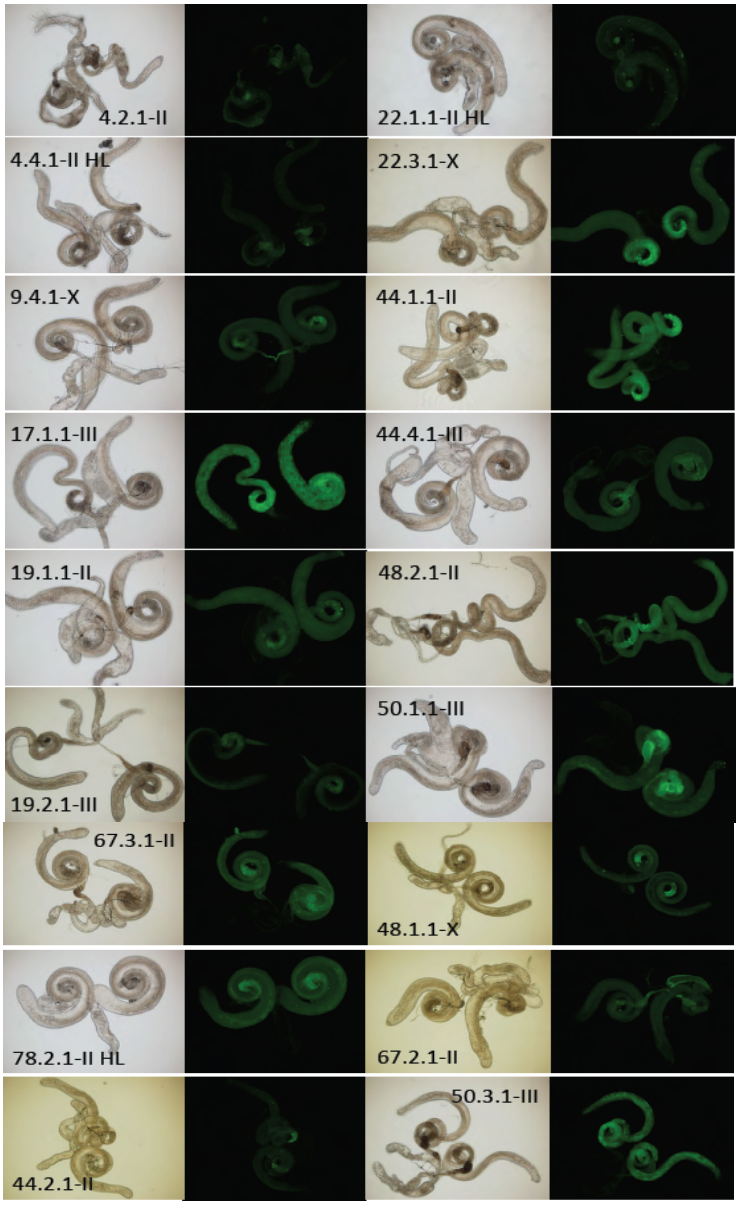


Figure 3.13 – Florescence in testes is shown for the different strains transformed with the *D. yakuba* region reporter construct. Pictures of the testes in clear field are shown as well. Low to high green fluorescence is observed in most of these transformed *D. melanogaster* lines. Codes correspond to the strain number, chromosome that has the insertion and lethality of the insertion, if lethality is observed (i.e., HL; homozygous lethal).

3.3.3. Dntf-2r-EGFP expression in larvae male gonads

The highest intensities of the green fluorescence were obtained using the longest construct containing 151bp upstream of the TSS (Figure 3.8 and Sorourian, et al. 2014). The longest construct and different individual transformant lines for it were used in this work to analyze the expression of *Dntf-2r* in larval testes and co-localization with other genes in Chapter 4.

As expected from *Dntf-2r* transcription in adult testis, *Dntf-2r*-EGFP fluorescence was observed in male gonads dissected from wondering larvae. Fluorescence could be seen in primary spermatocytes and 32 cell stages matching what was observed using *in situ* hybridization in larvae (Figure 3.14).

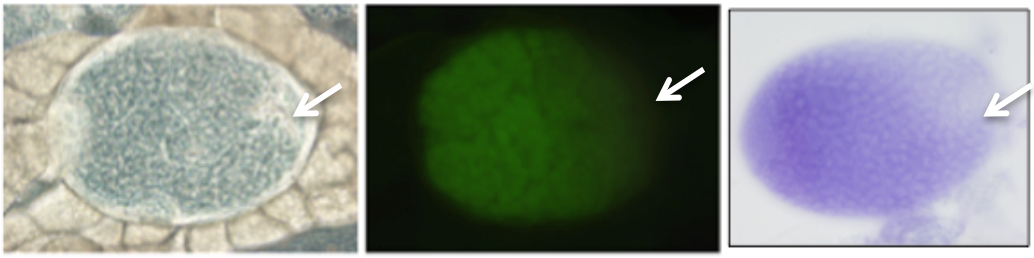


Figure 3.14 - A larval testis for one strain (1.53.2.2) transformed with the longest *Dntf-2r*-EGFP construct is shown. Phase contrast picture is shown. Expression of green fluorescent protein and *in situ* hybridization showing the same pattern are also shown. No expression in the hub cells is observed (white arrow).

3.3.4. Ran-like-DsRed.T4 *expression and regulatory region*

To identify the regulatory region that drives testis-specific expression of *Ran-like* in *D. melanogaster*, clones carrying the region upstream of the TSS, its 5'UTR and *Ran-like* coding region fused to Red Fluorescence Protein (DsRed.T4) as a reporter gene

were transformed in Drosophila using P element transformation technology. The longest transformed construct contained 500bp upstream of the TSS. *Ran-like* does not have any gene upstream, so I included 500 bp because testis regulatory regions are known to be close to genes (White-Cooper 2010). The second construct had 100bp upstream of the TSS and the last one started at the TSS (Figure 3.15). The transformed flies with 500bp and 100bp upstream regions respectively expressed the red fluorescent fusion protein in a pattern that mimics the *Ran-like in situ* profile in young adult and larva testis suggesting that the 100bp construct harbors the complete testis-specific regulatory region of *Ran-like*.

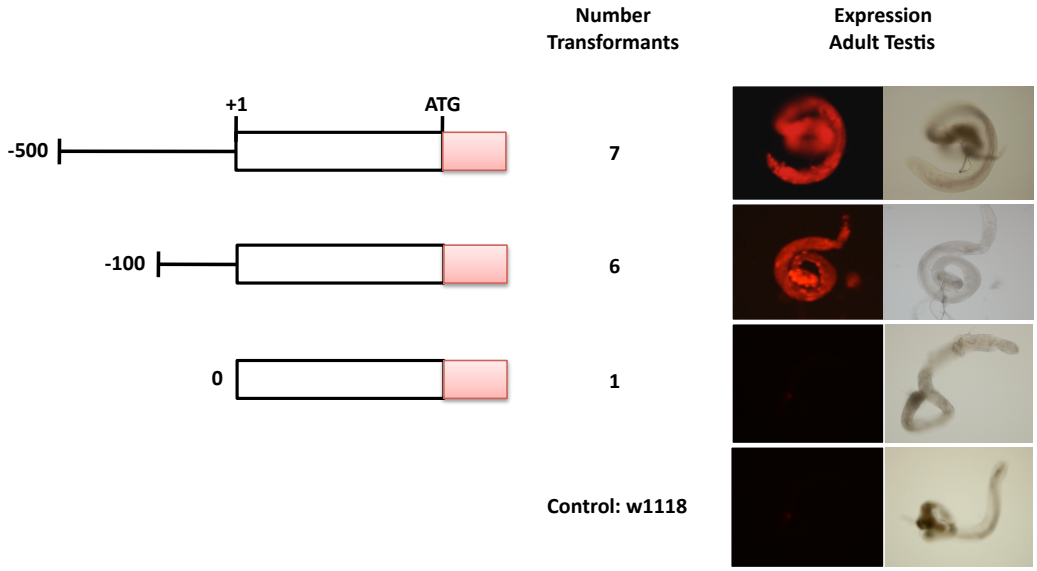


Figure 3.15 - The expression or lack of it of Ran-like-RFP fluorescence tag protein in testis for different constructs is shown. A strain representative of the construct is shown (7.24.2 for 500bp and 8.31.3 for 100bp). Number of strains examined is given. Clear field picture of the testes are shown. w^{1118} was used as control for auto fluorescence (more images in Appendix C).

The third construct didn't show fluorescence when compared to the same level of exposure that the white mutant control, w^{1118}. Independent insertion lines for the two constructs showed similar results indicating that the constructs themselves, and not the regions where the constructs were inserted, are responsible for the observations. Red fluorescence was also observed in the larval male gonads of the third instar larvae (Figure 3.16). As observed for adult testis the expression for the larval male gonads also mimics the results observed for the *in situ* hybridization.

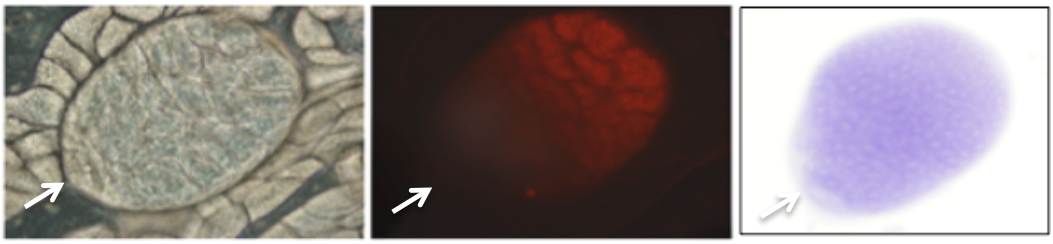


Figure 3.16 – A larval testis for one strain (7.24.2) transformed with the longest *Ran-like*-RFP construct (500 bp upstream) is shown. Phase contrast picture is shown. Expression of red fluorescent protein and in situ hybridization showing the same pattern for transcription are also shown. No expression in the hub cells is observed (white arrow).

3.3.3.1. *Ran-like-RFP in spermatogenesis*

I used fluorescent and confocal microscopy to study in detail the expression pattern of *Ran-like*-DsRed.T4 fusion protein in testis. Fluorescence was first observed at the 16-cell stage (primary spermatocytes). The expression continued in the 32-cell stage (meiotic spermatocytes) and in the 64-cell stage (round spermatids) and in elongating sperm heads and tails. No expression is observed in the somatic stem cells located at the tip of the testes as Ill as in mobile sperms (Figure 3.17).

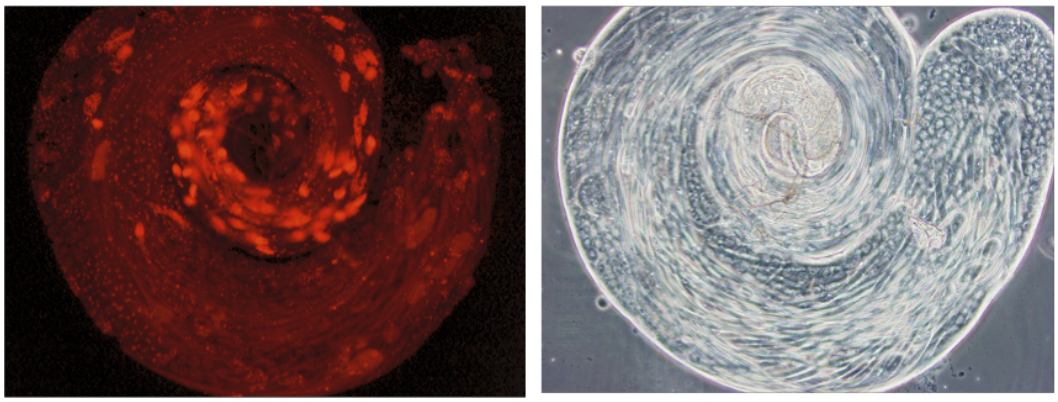


Figure 3.17 - *Ran-like*-DsRed.T4 expression along the testis in different stages of spermatogenesis for one-day-old males is shown. Phase contrast picture is also shown (7.12.1.1). transformed with the 500 bp upstream region construct was used (more lines in Appendix C).

3.3.4. Flipped β2-tubulin regulatory region

β2-UE1-like motifs were found in both strand orientations in testis-specific genes (Sorourian, et al. 2014), but have been initially described as directional testis-specific regulatory regions (Michiels, et al. 1989). The lack of directionality in β2-UE1-like motifs prompted us to generate a construct that carries the upstream region and 5'UTR of the β2-tubulin gene but where the 14-bp β2-UE1 motif has been flipped.

Ten independent transformants were obtained and all of them showed strong fluorescence in testis (Figure 3.18). Like for the β2-UE1, the expression was observed all over the testis with exception of the tip. I conclude that β2-UE1 can drive expression in testes in either orientation and would like to test if it can drive bidirectionally (i.e., two genes at once).

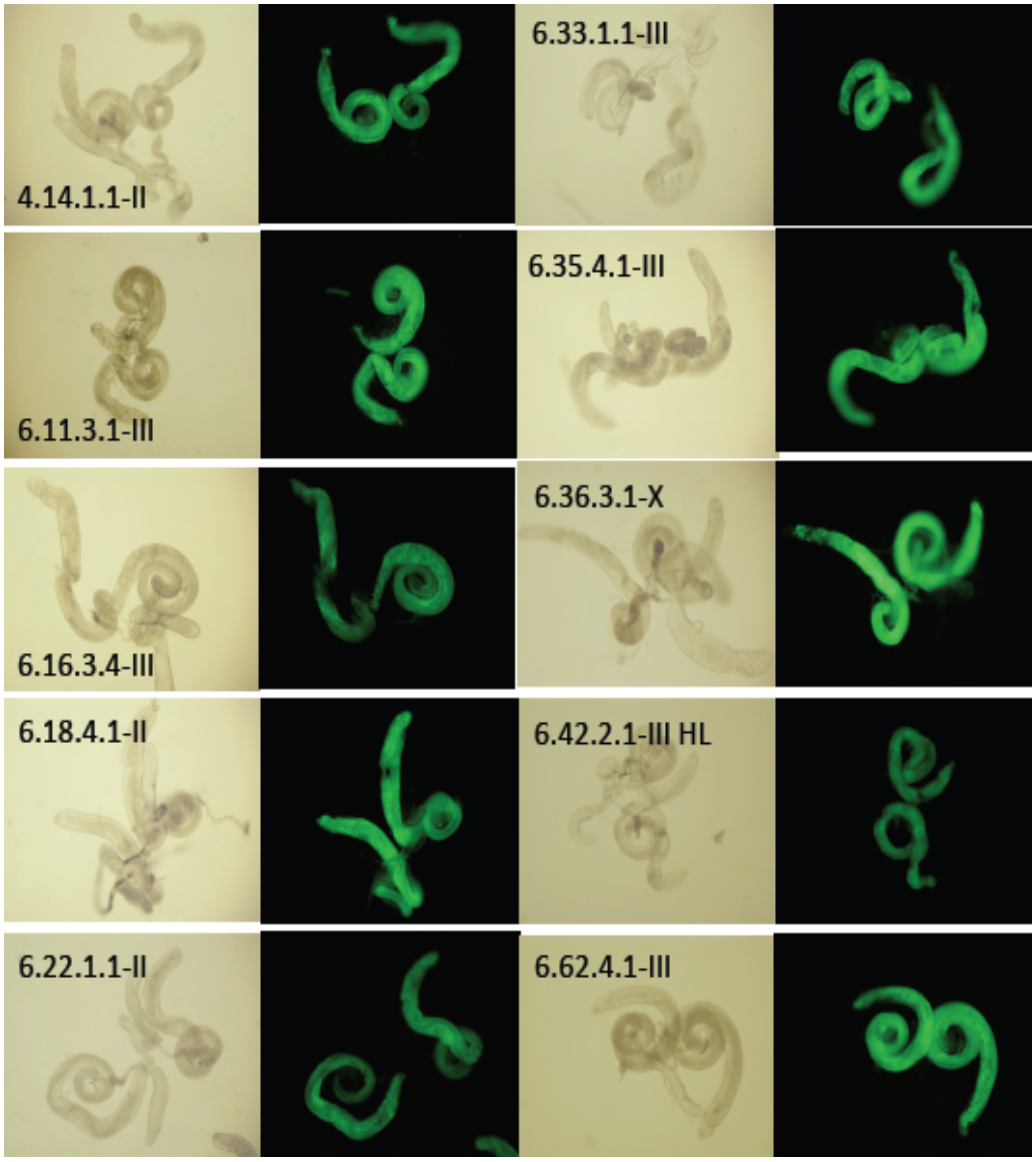


Figure 3.18 – Fluorescence in young male testes for strains transformed with the flipped 14-bp β2-UE1 construct is shown. Clear field picture of the testes is also shown. Codes correspond to the strain number, chromosome that has the insertion and lethality of the insertion, if lethality is observed (i.e., HL; homozygous lethal).

3.4. Discussion

3.4.1. Ran-like *and* Dntf-2r *expression in testes*

The results of *in situ* work demonstrate complementary transcription of *Ran* and *Ran-like* genes in testes, but an overlapping transcription of *Dntf-2* and *Dntf-2r*. I will provide more details about protein localization in Chapter 4, but this transcription data suggest that *Dntf-2r* does not replace *Dntf-2* during meiosis. The pattern of transcription of *Dntf-2-PA* and *Dntf-2-* during meiosis is different, with *Dntf-2-PB* being very slowly maturing and transcribed only in spermatocytes. Since the functional *Dntf-2* is a dimer (Figure 1.4). Dntf-2r, Dntf-2-PA and Dntf-2-PB could form homo and heterodimers in the tissues where their expression overlaps. The expression patterns of the *Ran* genes suggest that *Ran-like* might be replacing *Ran* in some cell types.

3.4.2. Dntf-2r regulatory region preceded retrogene insertion and likely facilitated its fixation

To address the evolutionary origin of the β2-UE1-like element of *Dntf-2r*, I compared the region before and after the insertion of the retrogene and attempted to reconstruct the ancestral state of this genomic location using maximum parsimony. The multiple alignment of the orthologous sequences revealed that the motifs upstream of *Dntf-2r* are highly conserved even in species that do not have the retrogene (Figure 3.5). I inferred with confidence that the motif with quantitative effects (GGATATT) was present at the time of insertion and that the β2-UE1-like region was at least partially present (Figure 3.5).

There is indirect evidence that support the possibility that this region was functional before the insertion of *Dntf-2r*. First, I find that the β2-UE1-like element and the motif with quantitative effects are partly and completely conserved, respectively, in *D. yakuba, D. teissieri* and *D. erecta*, where *Dntf-2r* is absent (Figure 3.5). In addition, I found two potential BEAF-32 insulator-binding sites at -135 –129 bp and -11 5-109 bp in *D. melanogaster*. These two BEAF-32 insulator sequences appear to be used at least during embryogenesis (Negre, et al. 2010) and, consistent with the previously described cases, might prevent co-expression of head-to-head genes (Yang, et al. 2012). Interestingly, these BEAF-32 binding sites are also conserved in *D. yakuba*, *D. teissieri* and *D. erecta* species that do not have *Dntf-2r* (Data not shown). This supports a head-to-head arrangement of *bsf* (a gene upstream of *Dntf-2r*) with another gene even before the insertion of *Dntf-2r*. Recently, a non-coding RNA gene (*let-7-C*; *CR43344*) has been annotated in *D. melanogaster* downstream of *Dntf-2r*. This non-coding RNA shows, according to modENCODE data, a male-biased pattern of expression similar to *Dntf-2r* (Marygold, et al. 2013). Additionally, its location is conserved in *D. yakuba*, *D. teissieri* and *D. erecta* (data not shown), indicating that *Dntf-2r* regulatory region likely predated the retrogene insertion.

To verify that this region is functional in species where *Dntf-2r* does not exist (i.e., species most likely carrying the ancestral configuration of this genomic region), I fused and transformed a construct containing the 14bp orthologous region from *D. yakuba* (see green highlight in Figure 3.9B) and the rest of the region from *D. melanogaster* in front of EGFP and checked if lines transformed with this contruct drive testis-specific expression. As summarized above, I observed testis-specific fluorescence in most of the independent transformants containing the construct and concluded that the region is capable of driving

testis-specific expression at a discernible level. All the evidence indicates that this region already contained a testis-specific regulatory element before the insertion of *Dntf-2r*.

3.4.3. Regulatory region of Ran-like retrogene

Ran-like regulatory region is within a 100 bp region upstream of TSS. This conforms to the observations made in other testis-specific genes (White-Cooper 2010). Testis-specific regulatory regions are, in general, short. However, longer construct (i.e., 500bp) expressed a bit higher than the shorter 100bp construct. This has also been observed before for *β2-tubulin* gene (Michiels, et al. 1989) and for *Dntf-2r* (Sorourian and Betrán 2010) and has been explained by the presence of quantitative or boundary elements (Michiels, et al. 1989; Sorourian and Betrán 2010). There is no motif similar to any known testis motifs in the 100 bp region that drives testis-specific expression of *Ran-like* and additional constructs would be required to dissect this region.

3.4.4. β2-UE1-flipped motif drives testis expression in either orientation

The analysis of the β2-UE1-flipped motif revealed that β2-UE1-flipped motif can drive testis expression in both orientations and has a potential for bidirectional regulation of transcription (potential to drive two genes; one in each orientation). There is an ongoing effort in the lab to study bidirectional activation of transcription by this element. If β2-UE1 is bidirectional it could facilitate the acquisition of testes expression by new genes.

As mentioned in the Introduction, long-standing question about retrogenes is how they acquire their often testis-specific regulatory regions. Here, I reveal that the regulatory regions of *Dntf-2r* were already in the region of insertion and that the *Ran-like*

regulatory region is short. So, it does appear from *Dntf-2r* that retroduplication facilitated the acquisition of testes regulatory regions by allowing the retrogene to insert downstream of a testis regulatory region. This is quite interesting because this case provides an excellent example of acquiring testis expression upon gene birth, leading to immediate preservation of a new gene duplicate and supports that duplication to a new location by retroposition facilitates the acquisition of testis-specific expression. This example alone answers all the questions we posed for this chapter. Is retroduplication facilitating the acquisition of testes regulatory regions? Yes. How? By mediating the insertion of a gene downstream of an existing testis regulatory region. How does the way the regulatory regions are acquired fit into/facilitate the proposed models of gene duplication? By retroduplication, genes can directly acquire expression in testis and this can guarantee their preservation. If the gene that is duplicated is under intralocus sexual antagonism, this duplication mechanism will immediately facilitate the acquisition of testis expression and every step of the duplication model will occur under positive selection (Gallach and Betrán 2011).

The regulatory region or *Ran-like* is short (<100bp) but we do not know what motif/s drive this expression, or how they were acquired. While retroposition might facilitate the acquisition of a new pattern of expression and given that studies have shown that retrogenes are found in excess in testis neighborhoods and consequently close to genes expressed in testis this can facilitate the evolution of testis expression (Vinckenbosch, et al. 2006; Bai, et al. 2008; Dorus, et al. 2008), Ran-like is far from any know gene. So, the only think we can say is that by being short it might be easy to evolve after retroposition.

CHAPTER 4

DNTF-2R AND *RAN-LIKE* RETROGENES FUNCTIONS IN SPERMATOGENESIS

4.1. Introduction

In the previous chapter I have confirmed testis expression of *Dntf-2r* and *Ran-like* retrogenes. The expression patterns of these genes within testes do not mirror the expression patterns of the parental genes completely, suggesting functional divergence between parental and daughter genes. In this chapter I will attempt to analyze the functions of the two retrogenes in spermatogenesis. This is part of Objective 3. In this objective, I would like to understand the reasons why there is a recurrent duplication of nuclear transport retrogenes. I will study the function of *Dntf-2r* and *Ran-like* in testes to answer this. I will use two approaches for this. I will examine the cellular localization and co-localization of both retrogenes with respect to each other and other known genes and structures involved in spermatogenesis, and I will inspect mutant phenotypes produced as a result of retrogene knockdowns using RNA interference (RNAi) and P-element insertion and excision lines for *Dntf-2r*. Since all of the work described in this chapter deals with the details of spermatogenesis, I will begin by providing an overview of this process as it occurs in Drosophila.

The testis of *D. melanogaster* is a well-structured organ. It contains germline and somatic stem cells whose cell divisions are tightly regulated by local signals and epigenetic mechanisms to produce cells that differentiate or help in the differentiation of cell into mobile sperm (de Cuevas and Matunis 2011). The germline stem cells (GSCs) are connected to somatic support cells at the tip of the testis. This part of the testis is called the hub. The GSCs asymmetrical divisions produce new GSCs that remain at the

hub as well as daughter cells (gonialblasts) that are displaced away from the hub and proceed to differentiate into sperm (Figure 4.1; Hardy, et al. 1979; Fuller 1993).

The newly formed daughter cells progress through spermatogenesis undergoing four mitotic cell divisions and forming groups of 16 spermatogonial cells. During these divisions cytokinesis is incomplete and the 16 cells stay connected by stable intercellular bridges called ring canals (White-Cooper 2010; de Cuevas and Matunis 2011). With the exception of cell size, germline stem cells, gonialblasts and spermatogonia are morphologically identical. Cyst cells derived from somatic stem cells surround the germline cells (Figure 4.1).

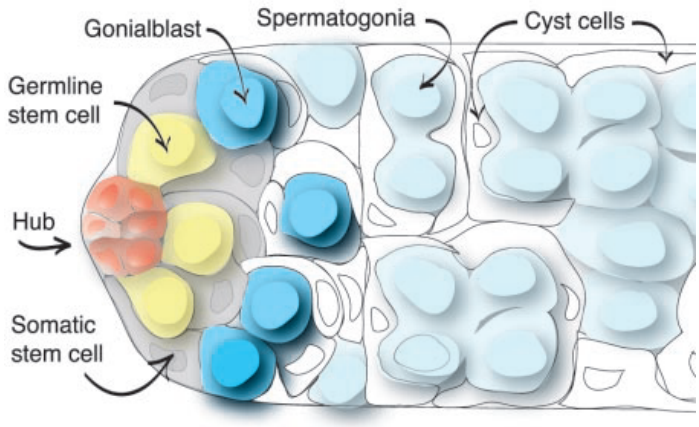


Figure 4.1 – The tip of Drosophila testis. Germline stem cells (yellow) attach to a cluster of about 12 somatic hub cells (red). The germline cells differentiate into gonialblasts (blue). Gonialblasts undergo four incomplete mitotic divisions, forming 16 interconnected spermatogonia (light blue). Each gonialblast is surrounded by two somatic support cells (cyst cells, colorless) (This image is reprinted from Tulina and Matunis 2001 with permission from The American Association for the Advancement of Science).

The 16-cell spermatogonia stage is followed by meiosis ending with the formation of 64 spermatids. During spermiogenesis (the process of spermatid transformation into functional sperm), spermatids undergo extensive structural changes to assume the final needle shape and individualize to form the mobile sperm.

In the first step of spermatid transformation, a substantial remodeling of mitochondria and nuclei is taking place. All the mitochondria fuse into two giant aggregates that wrap around one another producing a giant spherical and dense structure known as Nebenkern. This stage with round spermatids containing Nebenkern mitochondria is also called the onion stage due to onion-like layers of mitochondria (Figure 4.2).

The next step is the elongation stage that initiates with the formation of the flagella and acrosomes. The Nebenkern then starts to elongate, and the polarization of elongating cysts occurs in order to form the sperm tails. A microtubule-rich structure, the dense complex, helps positioning the basal body and keeps the strength of the nuclei as it undergoes condensation. During this process extra cytoplasmic material is removed, the cell nucleus is condensed and the histones are substituted by protamines (Figure 4.2; Texada, et al. 2008; Fabian and Brill 2012; Chen and Megraw 2014). The axonema, elongating from the basal body, and the mitochondria also elongate with the help of microtubules to produce the sperm tail (Figure 4.3).

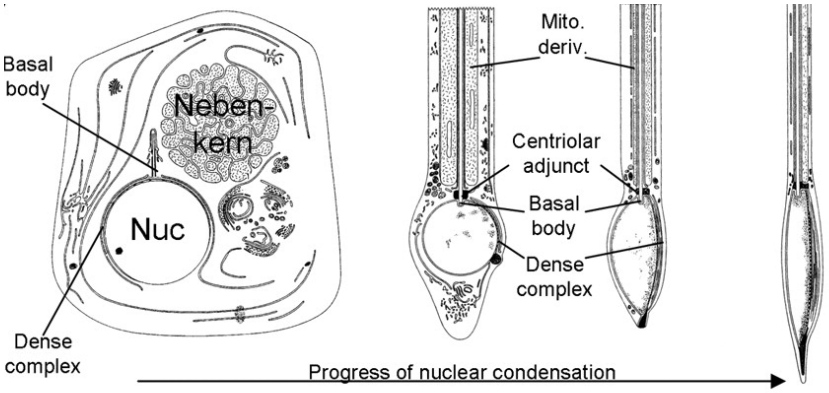


Figure 4.2 - Positioning of the dense complex, basal body and centriolar adjunct during spermatid nuclear condensation and sperm tail elongation (This image is reprinted from Texada et al. 2008 with permission from Company of Biologists Ltd).

The next step is the individualization. During the individualization stage, a motile filamentous actin system (known as actin cones) goes through the entire length of the sperm tail removing excess of cytoplasm and undesired proteins, and covering each sperm in a plasma membrane and removing the ring canals. The undesired proteins end in waste bags. At this point the sperm heads are located at the base of the testis and the waste bags are apical. Newly formed mobile sperm coils up and passes to the seminal vesicle to be stored (Tokuyasu, et al. 1972; Fuller 1993; de Cuevas and Matunis 2011).

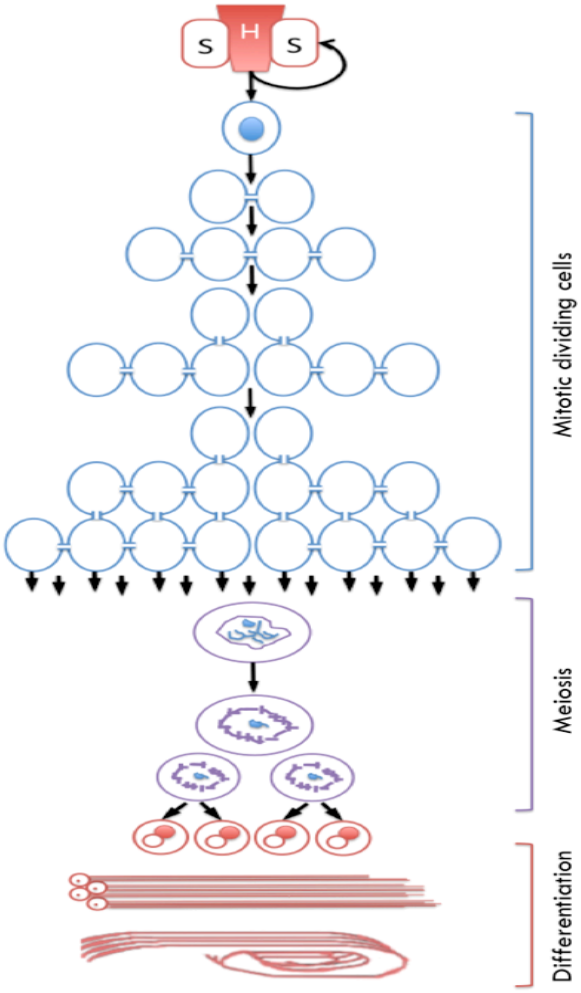


Figure 4.3 - Spermatogenesis stages in Drosophila. See text for details.

4.2. Materials and Methods

4.2.1. Functional studies: approach

To understand the function of *Dntf-2r* and *Ran-like* retrogenes in spermatogenesis, I studied their cellular localization. I also studied the co-localization of the retrogenes and other known genes and structures within testes by antibody staining

or other stainings. For the cellular localization of *Dntf-2r* and *Ran-like* I used the protein fusions expressed under the native testes regulatory regions described in Chapter 3 (i.e., Dntf-2r-EGFP and Ran-like-RFP).

I also performed a detailed study of the effects of knocking down *Dntf-2r* and *Ran-like* retrogenes in male germline using double-stranded RNA-mediated interference (RNAi) technique (Dietzl, et al. 2007). RNAi is a broadly used technique for generating knockdown phenotypes for genes in Drosophila. RNAi can be activated through the expression of a double-stranded 'hairpin' RNA from a transgene. This transgene must contain the gene fragment of interest cloned as an inverted repeat. When the two strands of complementary RNA are combined, a double stranded RNA (dsRNA) is produce and is subsequently chopped by Dicer resulting in small fragments of double stranded RNA molecules. The RNA separates into single stranded RNA and pairs up with complementary sequence of the gene's transcript, leading to its degradation. For such transcriptional knockdown mechanism to work, the GAL4/UAS expression system is used. By using the GAL4 driver, the RNAi transgenes can be used to target a gene of interest in almost any desired cell type at any stage of the Drosophila life cycle (Brand and Perrimon 1993; Dietzl, et al. 2007). To examine mutant phenotypes of *Dntf-2r*, I also used a line with a P-element insertion in the coding region of *Dntf-2r* and lines that carry perfect and imperfect excisions of that P-element.

4.2.2. Drosophila stocks and fly handling

In order to study the localization and co-localization of the retrogenes, both *Dntf-2r*-EGFP and *Ran-like*-DsRed.T4 lines were used and compared with the white mutant strain w^{1118} (Genetic Services, Cambridge, MA) that is the strain that was injected to produce the gene fusion lines.

For RNAi studies, UASt lines for *Ran* (108549), *Ran-like* (12293), *Dntf-2* (17755), and *Dntf-2r* (109227) were obtained from the stock center in Vienna (VDRC). All UASt stocks used were produced by Dietzl, et al. (2007). These stocks contain a hairpin sequence that can induce RNA interference (RNAi) upon production of small RNAs. *Dntf-2* stock is described as potentially having off target effects on its retrogene *Dntf-2r*. However, I did not observe these effects in our crosses (See Results). All UASt stocks were kept as homozygous since no deleterious effects were observed of the transgene insertions. Every UASt stock was crossed with *nanos*- (*nos*-) and *bam*-GAL4 drivers obtained from Bloomington stock center (stock number 4937) and send by Dr. Michael Buszczak, respectively. *nos*-GAL4 and *bam*-GAL4 are both expressed in germline, but have different temporal expression profiles. *nos* is expressed early during spermatogenesis while *bam* is expressed later (White-Cooper 2012). w^{1118} was used as a control for all the experiments because this was the line that was used to generate the transgenic lines.

Additional *Dntf-2r* functional studies were performed using a line ($Ntf\text{-}2r^{EY05573}$) that contains a P-element insertion in the coding region of *Dntf-2r* (Bellen, et al. 2004). The stock was ordered from Bloomington stock center. The line ordered (16658) had to be out-crossed and fixed again as it became clear that it came from Bloomington containing two P-element insertions instead of a single independent insertion. I called this newly fixed line 5.1.1. Two other *Dntf-2r* lines were produced from this line. One of these lines was produced by excising the P-element (line 2.1) and the second line, a control line (7.2), resulted from excising the P-element followed by a perfect repair of this region. Additionally, knockout lines for *Ran* (11800; Cesario and McKim 2011) and *Dntf-2* (109227; Bhattacharya and Steward 2002) were also ordered from Bloomington and

mutant rescue studies were performed using both parental and retrogene constructs to study if the genes have changed functions from the parental genes.

The rescue of the parental knockout lines was done using constructs made in the lab with UASt upstream of the gene of interest fused to EGFP. This process allows for the overexpression of a gene fusion and the attempt to rescue the other gene mutant. Details of the particular crosses and genotypes will be given below.

4.2.3. Knockdown crosses

Knockdown crosses were performed using RNAi driven by the GAL4/UAS system in particular cells. Five young (1-2 days old) UASt-RNAi males and five GAL4 virgin females (1-2 days old) driving GAL4 in a particular pattern were crossed. Three replicates of every cross were performed. In a first round of crosses, each cross was incubated at 25ºC and 29ºC (Duffy 2002). Twenty-nine degrees Celsius has been described as the best temperature for the GAL4/UAS system (Duffy 2002). So, these two temperatures were used initially to find out the best way to drive RNAi without producing side effects. In this step the progeny produced with each temperature was compared side by side for dead larvae and pupae in the vials, and the emerging flies were carefully observed under the dissecting scope. The results indicated that the best temperature to drive RNAi is 29ºC. However, this temperature can generate negative viability and fertility effects (Duffy 2002). To make sure that the results obtained were not due to temperature, additional crosses were performed using crosses between drivers and w^{1118} and between UASt lines and w^{1118}. Knockdown offspring were also used in an assay to test fertility. For the fertility assay, 1 male of one line and 3 females of the other line were crossed. Reciprocal crosses were also performed. Testes from pupae and young males from these crosses were dissected and stained with DAPI (DNA), mitotracker (mitochondria)

and phalloidin (F-actin cones) in order to observe the cytological effects of knocking down our genes. See details for the stainings below. Additionally, fly testes were dissected, fixed and frozen to perform *in situ* hybridization analysis (Figure 4.4).

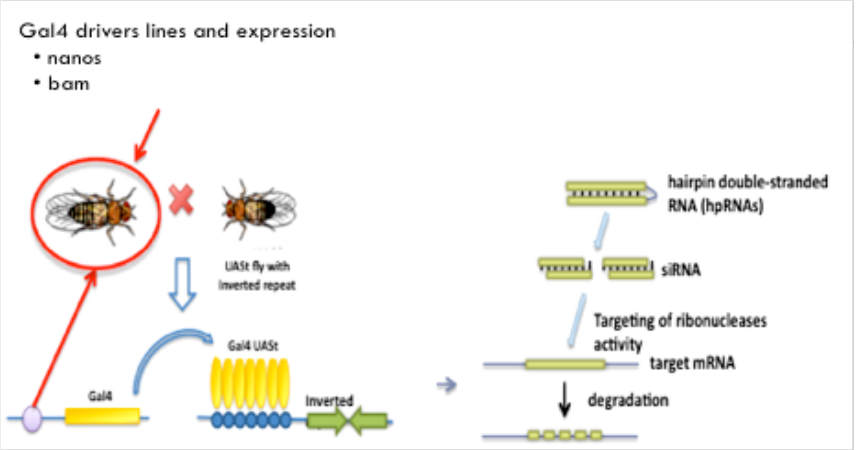


Figure 4.4 – GAL4 lines used to drive RNAi expression and knockdown of the genes of interest in germline. Schematic representation of the RNAi driven by the GAL4/UAS system.

4.2.4. Fertility assays for RNAi lines

Fertility assays were carried out with the progeny from RNAi crosses and also with the P-element insertion line and excision lines for *Dntf-2r*. For the RNAi assay, three virgin females (driven line or w^{1118}) were crossed with one young (1-2 days old) male (GAL4 driver or w^{1118}) in order to test the fertility of each male. Ten replicates were performed for each cross. After 7 days the progeny was counted for 10 days. In the case of the P-element line and excision lines, both w^{1118} and the P-element perfect excision line were used as controls. All progeny was counted for every cross and a t-test was used to test for differences between the average numbers of offspring produced by the RNAi crosses and the controls.

4.2.5. Fertility assay for the P-element line and excision lines for Dntf-2r

Males from each stock were tested for their fertility levels. Two methods were employed to check for infertility of *Dntf-2r* P-element insertion and excisions. The first method used was a simple male fertility assay based on assays done by Dyer et al. (2011). Five males from the P-element line were crossed with 5 females w^{1118} and allowed to mate for 10 days at room temperature. For each cross, 10 replicates were produced. After 10 days the progeny was counted.

The second method is a male sperm exhaustion assay. It was performed following Sun, et al. (2004). This is a more sensitive assay that allows the detection of small differences in sperm production through time. Individual 1-day-old males were crossed with three 1-day-old virgin females. Crosses were made over a period of 10 days and each day a male was allowed to mate with three new virgin females. Each female was placed in an independent vial and allowed to lay eggs. Once pupae were observed the females were removed from the vials. Progenies were counted for 10 days (Figure 4.5).

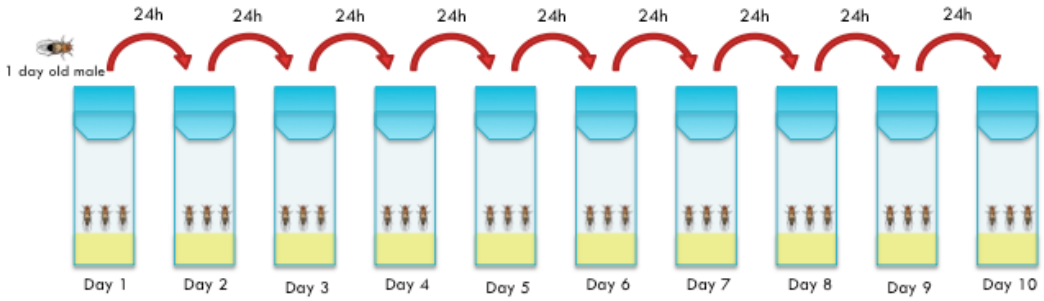


Figure 4.5 – Schematic representation of the sperm exhaustion assay. Ten replicates were generated for each line.

4.2.6. UASt transformants for mutant rescue

Using the gateway system, *Dntf-2-PA*, *Ran* and *Ran-like* constructs were produced containing the UASt upstream of the coding region of the gene and in-frame EFGP downstream of the gene. I had difficulties with the UASt-*Dntf-2r*-EGFP construct and I do not have it yet, but I still intend to produce it. The crosses were carried out according to Bhattacharya and Steward (2002). Each UASt fusion gene was expressed under the control of the *arm*–GAL4, *Act5C*-GAL4 and *tubP*-GAL4 drivers (stock number 4414, 1560, and 5138) and checked for their ability to rescue *Ran* knockout line or *Dntf-2* knockout line. These are all ubiquitous drivers. The first overexpression-driving cross was done using five 2-days-old males and five 2-days-old virgin females and kept initially at 29ºC. However, a new cross was done at 25ºC due to the lethality observed when overexpressing *Ran-like* with any of the drivers. The progeny of the second cross was counted and checked for non-bar males that identified rescued individuals (Figure 4.6). As I explained in Chapter 1, the parental genes (*Dntf-2* and *Ran*) are located on the X chromosome and are recessive lethal, requiring the use of balancers to keep the stocks. When checking the progeny I expect 25% of the non-bar males to be rescued as only this percentage inherits both the Gal4 driver and the UASt transgenes and should potentially to rescue the X mutant.

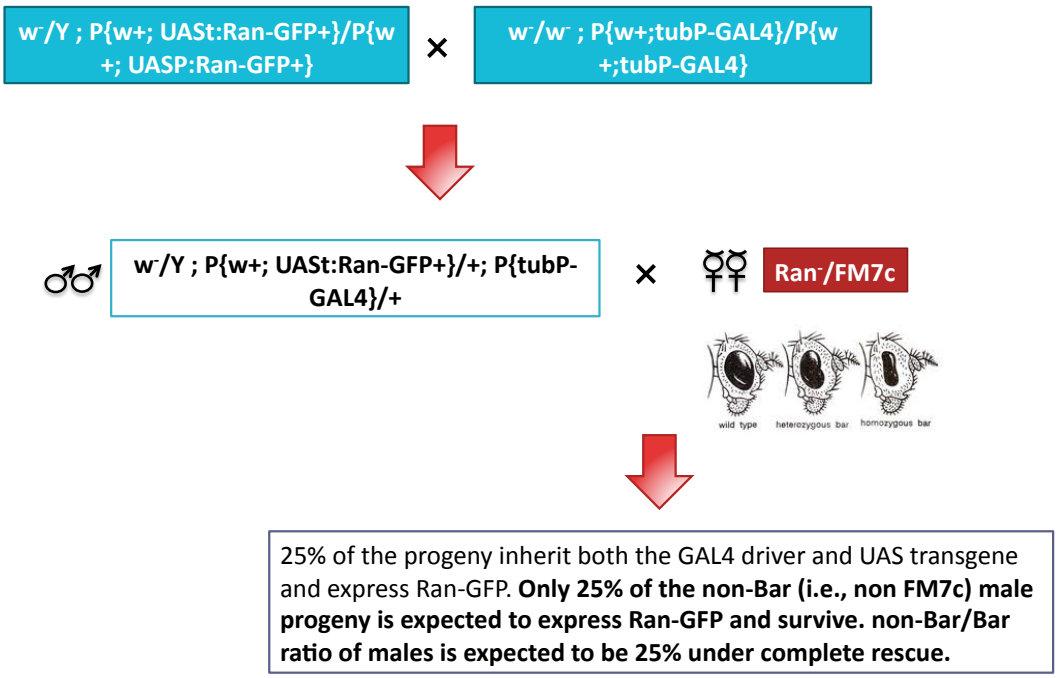


Figure 4.6 - Diagram of the first and second crosses done for mutant rescue of *Ran*. Using the driver *tubP*-GAL4 to overexpress the parental gene. The same was done using *Ran-like* to rescue *Ran*.

4.2.7. Detection of mobile sperm

I dissected testes under a dissecting scope and examined them under the fluorescence scope. To help with the observation of the mobile sperm morphology, DNA and mitochondria stainings were also carried out (See details of the stainings below). Knockdowns F1 progeny and 7.2, 2.1, and 5.1.1 mutants 8-days-old virgin males were examined. Wild type 8-day-old virgin males are expected to have large amounts of mobile sperm inside the seminal vesicles. By making squashes of seminal vesicles I was able to look for the presence or absence of sperm.

4.2.8. RNA extraction

The Qiagen RNeasy mini kit was used for RNA extraction from 30 adult flies from each line. *D. melanogaster* Besançon was used as control. This procedure allows us to check for the presence of the transcript in the knockdown individuals.

4.2.9. Retrogenes and parental transcription detection by RT-PCR

RT-PCRs were used to check for the presence of transcripts in the *Dntf-2r* mutant lines. RNA was first digested with the DNase I enzyme to digest any contaminant genomic DNA. Reverse transcription was performed using oligo (dT) primers (Promega, Madison, WI) and Superscript II reverse transcriptase (Promega, Madison, WI). From the cDNA obtained, PCR was performed using *Dntf2r*-specific primers (Table 4.1).

Table 4.1 – Primers used to study the presence of *Dntf-2r* transcript in the P-element stock and in the excision stocks.

Primers	**Sequence 5' to 3'**
3' Race 1	TTGTCCAGCAGTACGCC
GSP 1	AGCCACGAAGAGGGATCCTC
Dntf-2r_For	GGGGATCGTCATCGCATTT
Dntf-2r_Rev	TTGTCCAGCAGTACTACGCC

4.2.10. Antibody staining

In order to study possible co-localization of both retrogene fusions with other known genes expressed during spermatogenesis, antibody stainings were performed using different antibodies of genes with known localization in spermatogenesis. Anti-lamin, anti-centrosomin (cnn), anti-alpha-tubulin (tub), and anti-dynein were order from the Developmental Studies Hybridoma Bank, University of Iowa. Lamin antibody was

used to label the internal part of the nucleus in mitotic cells before meiosis starts (Chen, et al.2013). To label the centrosomes during cell division I used the centrosomin (Cnn) antibody (Megraw et al 1999). Alpha tubulin antibody was used with to label microtubules during cell division and also to check for the presence or absence of Dntf-2r and Ran-like in the waste bags since this gene is not present in this structure (Bo et al 1989). Finally, dynein antibody was used to mark the dense body in the spermatids (Texada, et al. 2008).

Testes were dissected from pupae and 0-days-old males from different transformed strains and from w^{1118}. Some testes were squashed using the cover slip and frozen for 1 min at -80C. The cover slip was then removed and the tissues were dipped in methanol for 10 min, followed by 30 seconds in acetone and then 5 min in PTW. In the last step the samples were washed in PBS 2 times for 10 minutes and allowed to pre-hybridize for 2 hours in 1% PBS with BSA. Different antibodies were used one at a time in a concentration of 1:50 overnight in a wet chamber. The next day, samples were again washed in 1% PBS and secondary antibody was applied. For secondary antibodies I used Alexa Fluor® 488 Goat Anti-Rabbit and Alexa Fluor® 488 Goat Anti-mouse (Molecular Probes®, Life Technologies, Carlsbad, CA). Before observations, samples were washed one more time for 30 minutes. Whole testes were also dissected and fixed with PFA and washed 2 times with 1% PBS for 10 minutes folloId by staining with primary antibody at 1:100 overnight at 4ºC degrees. After overnight hybridization the samples were washed with 1% PBS and followed by staining with secondary antibody at 1:500. Finally the testes were washed and mounted in FluorGlo® mounting solution (Valley Scientific, Mayville, NY) for examination by fluorescence microscopy (Olympus BX51TRF florescent microscope) and/or confocal microscopy (Zeiss LSM 150)

4.2.11. DAPI staining

Testes were dissected from pupae or freshly emerged males of the desired transformed line and w^{1118} and fixed in 4% paraformaldehyde in 1× PBS for 30 minutes at room temperature. Tissues were then washed with 1% PBS for 10 minutes and in order to stain DNA, incubated with DAPI (1 µg/ml; Sigma, USA) for 1 hour in the dark at room temperature. Samples were mounted in FluorGlo® mounting solution (Valley Scientific, Mayville, NY) for examination by fluorescence microscopy.

4.2.12. Mitotracker and phalloidin stainings

To recognize mitochondria and F-actin cones in testes, mitotracker and phalloidin stainings were used. Testes were dissected from 0-days-old males from different transformed strains and w^{1118} and fixed following the above protocol. Tissues were then washed with 1% PBS for 10 minutes and 0.3µl of 1mM mitotracker (Cell Signaling Technology®, Danvers, MA) and/or 0.5µl of 10µM phalloidin (Biotium Inc., Hayward, CA) were added to 750µl of 1% PBS and left 1 hour in the dark at room temperature. After 1 hour the testes were washed twice with 1% PBS and mounted in FluorGlo® mounting solution (Valley Scientific, Mayville, NY) for examination by fluorescence microscopy (Olympus BX51TRF fluorescence microscope) and/or confocal microscopy (Zeiss LSM 150)

4.2.13. Phase contrast microscopy of testes

To examine various stages of spermatogenesis, testes from pupae males and young males from w^{1118} and from different transformant flies were dissected in 1% PBS (130 mM NaCl, 7 mM Na2HPO4, 3 mM KH2PO4, pH 7.4; Fisher Scientific, New Jersey). Testes were mounted in between two bridges to allow for the observation of the entire

testis and others were squashed lightly under the weight of a cover slip. The preparations were examined under phase contrast optics using (Olympus BX51TRF fluorescence microscope).

4.2.14. Fluorescence microscopy of testes

Testes were dissected from young transformant males with constructs containing different lengths of upstream regions and w^{1118} males to allow a comparison of the fluorescence level. The tissues were dissected in 1% PBS. After dissection some testes were immediately mounted in 1% PBS. After mounting they were observed under the fluorescence microscope, Olympus BX51 TRF fluorescence microscope, setting the UV exposure time manually at a level established by the comparison to the control strain (w^{1118}). Sequentially, the remaining testes were fixed in 4% paraformaldehyde for 30 minutes at room temperature. The fixed testes were washed twice with 1% PBS, mounted in FluorGlo® (Valley Scientific, Mayville, NY) and observed under the fluorescence microscope using the mutant strain w^{1118} to control exposure settings.

4.3. Results

4.3.1. Dntf-2r-EFGP and Ran-like-DsRed.T4 localization and co-localization studies

Using the longest gene fusion constructs for Dntf-2r-EGFP and Ran-like-DsRed.T4 (i.e., the ones containing the most of the upstream regulatory region) described in Chapter 3, I explored the detailed expression pattern and cellular localization of both proteins during spermatogenesis. Moreover, antibodies against proteins known to be express in spermatogenesis were also used to help determine the exact localization of the retrogenes and make inferences about their function.

As expected from the pattern of transcription of the genes (i.e., *Dntf-2r* and *Ran-like* are not transcribed at the tip or end of the testis, the hub and mobile sperm respectively), I observe fluorescence starting at the 16-cell stage (i.e., primary spermatocytes that will start meiosis; Figure 4.7).

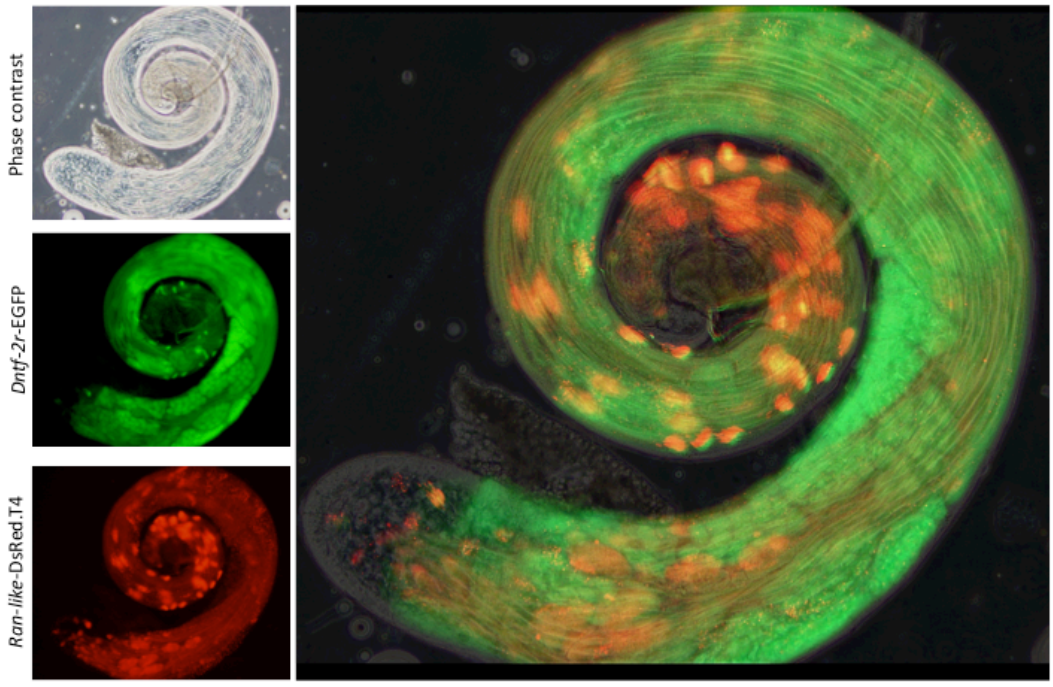


Figure 4.7 – Whole testis showing Dntf-2r-EGFP and Ran-like-RFP localization in the different cell stages. Dntf-2r in green and Ran-like in red. The total magnification is 100x.

All along the testis, Ran-like-DsRed.T4 (Ran-like in this Dissertation section) has a sharper localization than Dntf-2r-EGFP (Dntf-2r in this Dissertation section). The earliest presence of both retrogenes is observed in the primary spermatocyte cells that will enter meiosis. Dntf-2r is present in the cytoplasm, nuclear membrane and inside the nucleus. Ran-like co-localizes with Dntf-2r in the nuclear membrane and nucleus, but is not present in the other cellular compartments (Figure 4.8). This is the expected localization for nuclear transport genes that translocate from cytoplasm to nucleus, as

they are involved in nuclear transport and nuclear functions during cell division (i.e., envelope disintegration and restructure of the envelope after).

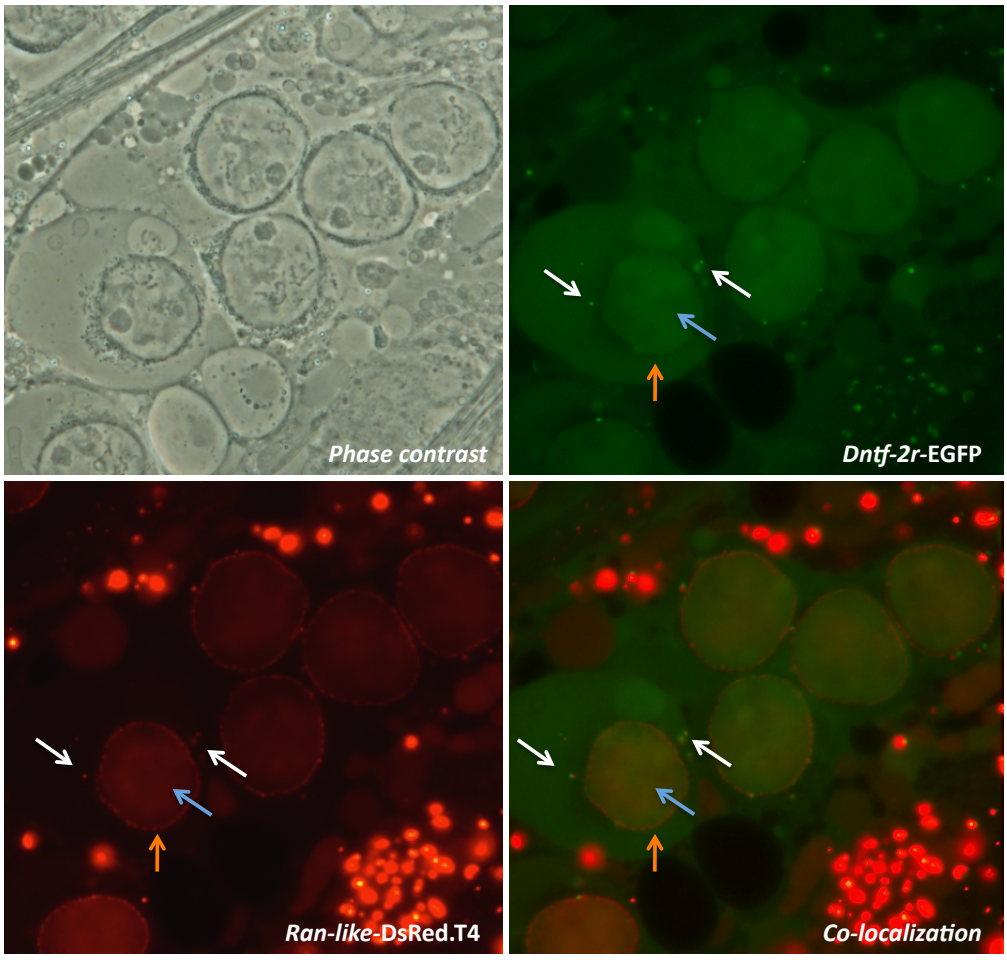


Figure 4.8 - Dntf-2r and Ran-like fusion proteins and their localization at the 16-cell stage spermatocytes in different cellular structures: (1) Nuclear membrane – orange arrow; (2) Nucleus – blue arrow; (3) Granules – white arrow. *Dntf-2r* in green and Ran-like and red. The total magnification is 400x.

Dntf-2 localization has been studied in yeast and it has been shown to have a similar cellular localization as the one observed for Dntf-2r here. Quimby et al. (2000) showed that Dntf-2 is always present in the cytoplasm and nucleus of the different cell

types. The cellular localization of Ran has been observed in Drosophila embryos (Trieselmann and Wilde 2002). Ran localization is similar to the one observed for Ran-like in spermatocytes. It is around the nuclear membrane and nucleus and during cell division in the cell spindles. From these results and the comparison between retrogenes and parental genes, I infer that our fusions are producing functional proteins. Additional data on the localization of the retrogenes shown below additionally support this.

Dntf-2r and Ran-like are also observed in cytoplasmic granules (Figure 4.8). These granules that might be piRNA granules appear in 16-cell spermatocytes and seem to ensure the posttranscriptional regulation of germline repeat transcripts and silencing in the nucleus (Nosov, et al. 2014). Aubergine and RanGAP have been previously observed to co-localize in these granules in spermatocytes (Gell and Reenan 2013). It is known that to export small RNAs precursor from the nucleus, ribonucleoproteins containing exportin and RanGTP need to be assembled (Kohler and Hurt 2007), exit the nucleus and deliver the transcripts to the granules/piNG-bodies where the ping-pong small RNA amplification occurs (Kohler and Hurt 2007; Gell and Reenan 2013). This localization of Dntf-2r and Ran-like makes us think that both retrogenes may be involved in piRNA pathways during spermatogenesis. To confirm this inference, I will need to study the co-localization localization of Dntf-2r and Ran-like with Aubergine and RanGAP.

Throughout meiosis Ran-like appears in the dividing cells localizing with the cell spindles and at the cell poles. Dntf-2r localization is more disperse; Dntf-2r is present in the same structures as Ran-like, but also in the cell cytoplasm (Figure 4.9). This is exactly how parental genes have been observed to localize (Quimby, et al. 2000; Trieselmann and Wilde 2002) and confirms our inference that the fusions are functional.

If Ran-like has retained Ran functions, Ran-like is expected to be transported to the nucleus if bound to GDP, or induce microtubule nucleation next to chromosomes if bound to GTP (see Chapter 1 for a description of all known functions of Ran). From our *in situ* data (Chapter 3), the parental Ran is likely not present at these spermatogenesis stages, and our interpretation is that all the orchestration of cell division is mediated by Ran-like including membrane reassembly (Clarke and Zhang 2008). The parental Dntf-2 is likely present, as it is transcribed throughout spermatogenesis, and might produce homodimers and/or heterodimers with Dntf-2r of both PA and PB proteins. The function of these alternatively spliced forms is unknown.

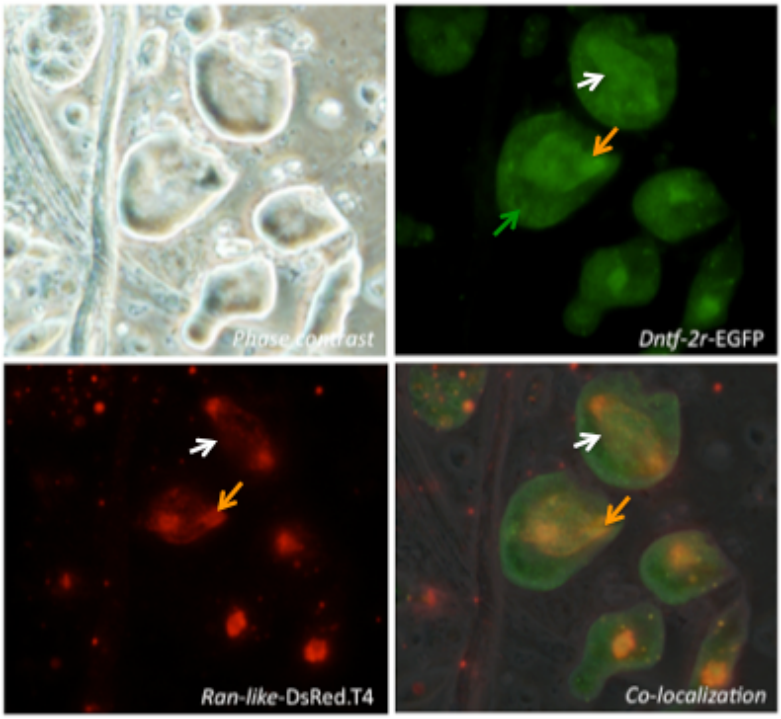


Figure 4.9 – Dntf-2r and Ran-like fusion tag protein and their localization in meiotic dividing cells. Cell poles – orange arrow; Meiotic spindles – white arrow; Cytoplasm – green arrow. Dntf-2r in green and Ran-like in red. The total magnification is 400x.

At the onion stage of spermatogenesis (i.e., when the mitochondria have fused into a round multilayered dense structure), bundles of 64 cells can be found along the testis (Figure 4.10) just before the mitochondria start elongating. Both retrogenes are in a structure named dense body that is surrounding the nucleus where the nuclear pores are accumulating and microtubules are assembling (Fabian and Brill 2012). They are also present at the beginning and at the end of the mitochondria where microtubules are organizing for tail axoneme and mitochondria elongation (See next step below). Dntf-2r can also be observed overlaying with the cell nucleus (Figure 4.10).

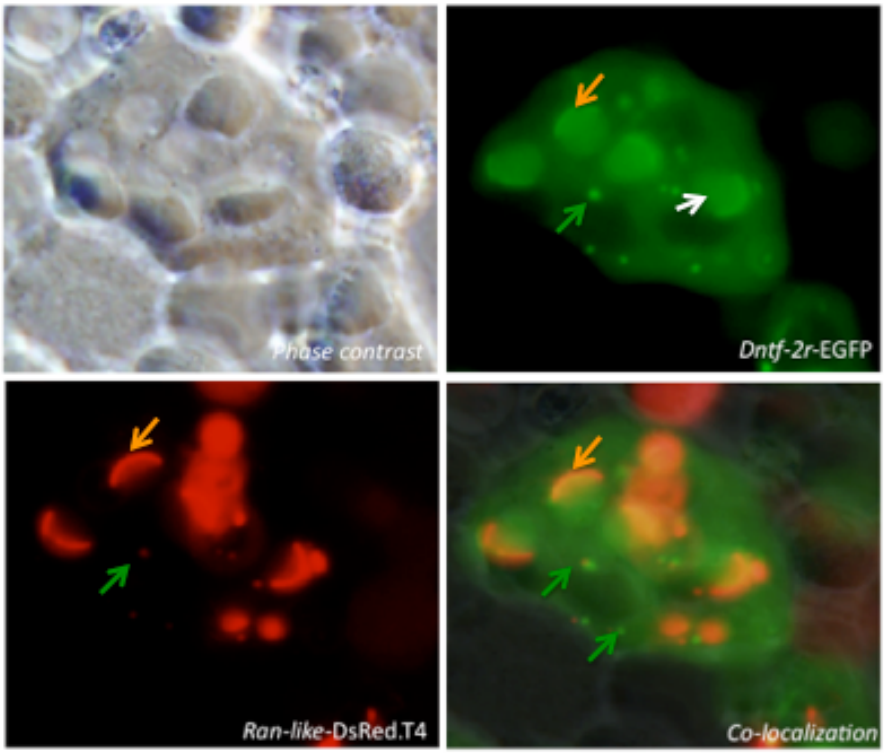


Figure 4.10 - Dntf-2r and Ran-like fusion proteins and their localization in onion stage cells (See Figure 4.11A for details on the structures). Dense body – orange arrow; Nucleus – white arrow; Mitochondria poles – green arrows. *Dntf-2r* in green and Ran-like in red. The total magnification is 400x.

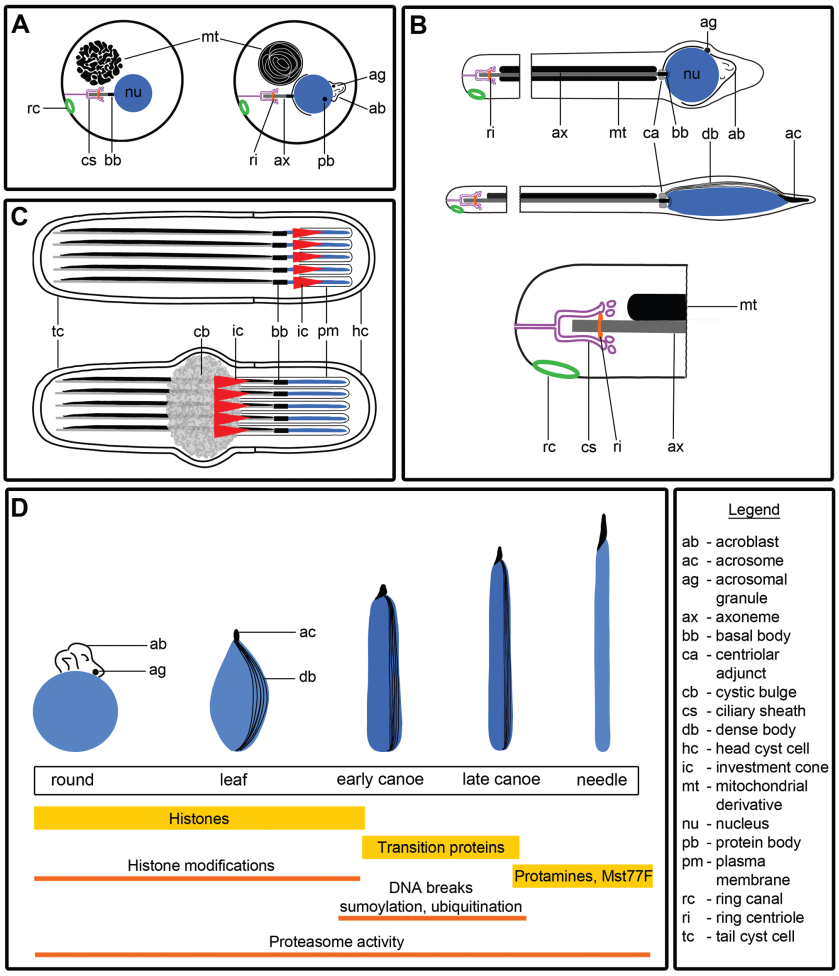


Figure 4.11 – Spermiogenesis stages and chromatin condensation details. Starting from the 64-cell stage (early round spermatids) and before mitochondria elongation (A). Spermatid elongation with representation of the axoneme and the mitochondrial derivatives (B). The 64 spermatids are surrounded by two somatic cyst cells: a head cyst cell and a tail cyst cell and undergo individualization with help of the F-actin cones (C). As the nuclei elongate, they go through leaf, early canoe, late canoe and needle-shaped stages and they replace histones with protamines (D). (This image is reprinted from Fabian and Brill 2012 with permission from Landes Bioscience).

The different cells and the retrogenes localization can be observed in detail in round spermatids under the oil objective. Dntf-2r and Ran-like co-localize in the dense body, centriolar adjunt, and around the mitochondria to nucleate microtubules in preparation for its elongation (Fabian and Brill 2012; Figure 4.13). Dntf-2r is located in the acroblast but Ran-like fusion does not appear to locate there.

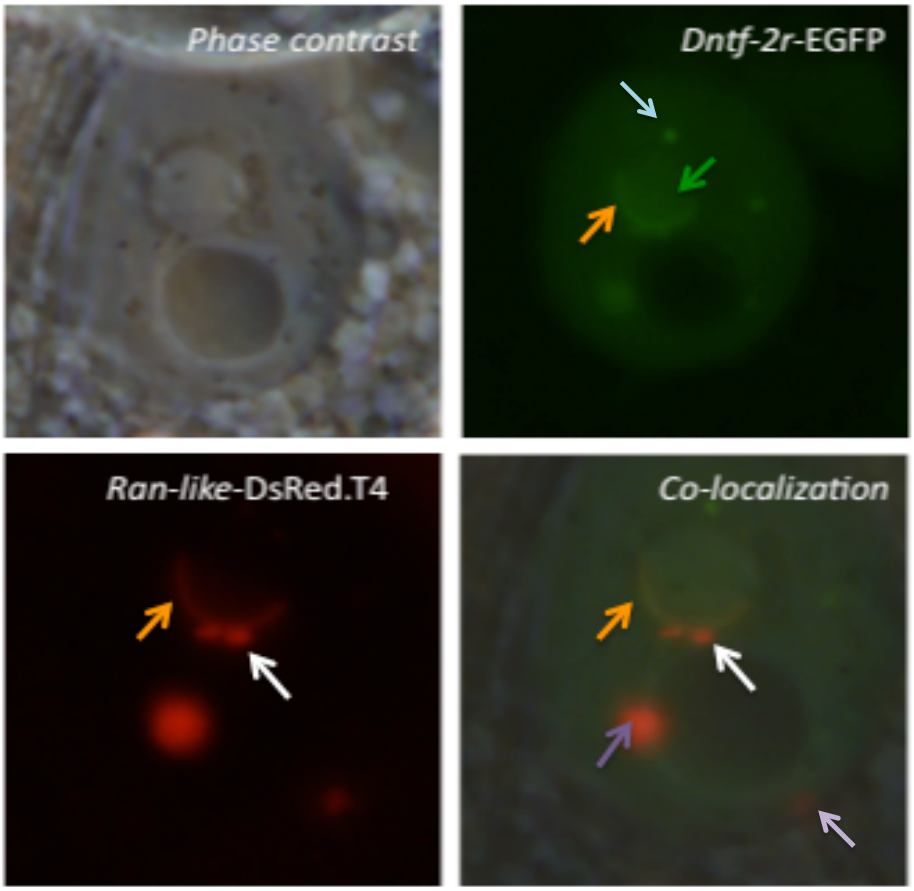


Figure 4.12 - Dntf-2r and Ran-like detailed localization in round spermatids. Dense body – orange arrow; Centriolar adjunt – white arrow; Nucleus – green arrow; Mitochondria poles – purple arrow; Acroblast - blue arrow. Dntf-2r in green and Ran-like in red. The total magnification is 1000x.

After the formation of the 64 cells bundles the mitochondria and the nucleus start elongation (Figure 4.14) giving rise to the sperm bundles (Figure 4.16) that will go through differentiation and form the mobile sperm.

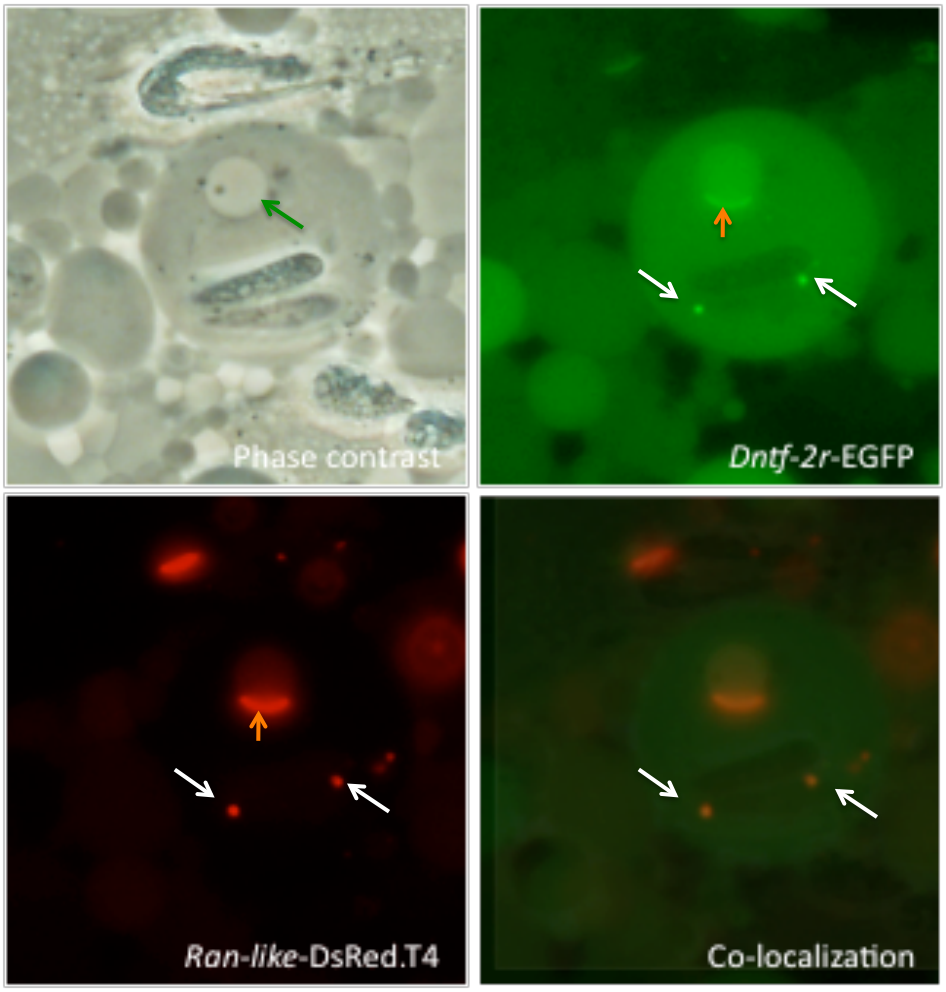


Figure 4.13 - Dntf-2r and Ran-like detailed localization during the start of mitochondria elongation. Dense body – orange arrow; Nucleus – green arrow; Mitochondria poles – white arrows. Dntf-2r in green and Ran-like in red. The total magnification is 1000x.

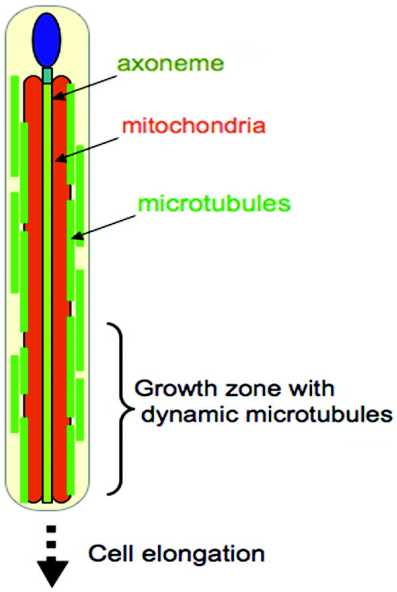


Figure 4.14 - Model for the dynamics of mitochondria-dependent elongation of sperm tail proposed by Noguchi, et al. (2012). The two giant mitochondria elongate simultaneously with microtubules and push cell membrane of elongating sperm tail. (This image is reprinted from Noguchi et al. 2012 with permission from Landes Bioscience).

Dntf-2r is present all along the sperm bundles, however Ran-like seems to have a very precise localization along the sperm bundle tails. At this point there are dynamic microtubules helping the cell and mitochondria to elongate (Noguchi, et al. 2012; Figure 4.15), and Ran-like appears to be helping microtubules nucleation as it does during cell division. Dr. Jieyen Chen (Florida State University) presented in the 2014 Drosophila Meeting about a new splice variant of *centrosomin* that converts spots on the mitochondria into MTOCs (microtubule organizing centers) to facilitate sperm elongation. We think that Ran-like likely localizes closer to those centers (Figure 4.16).

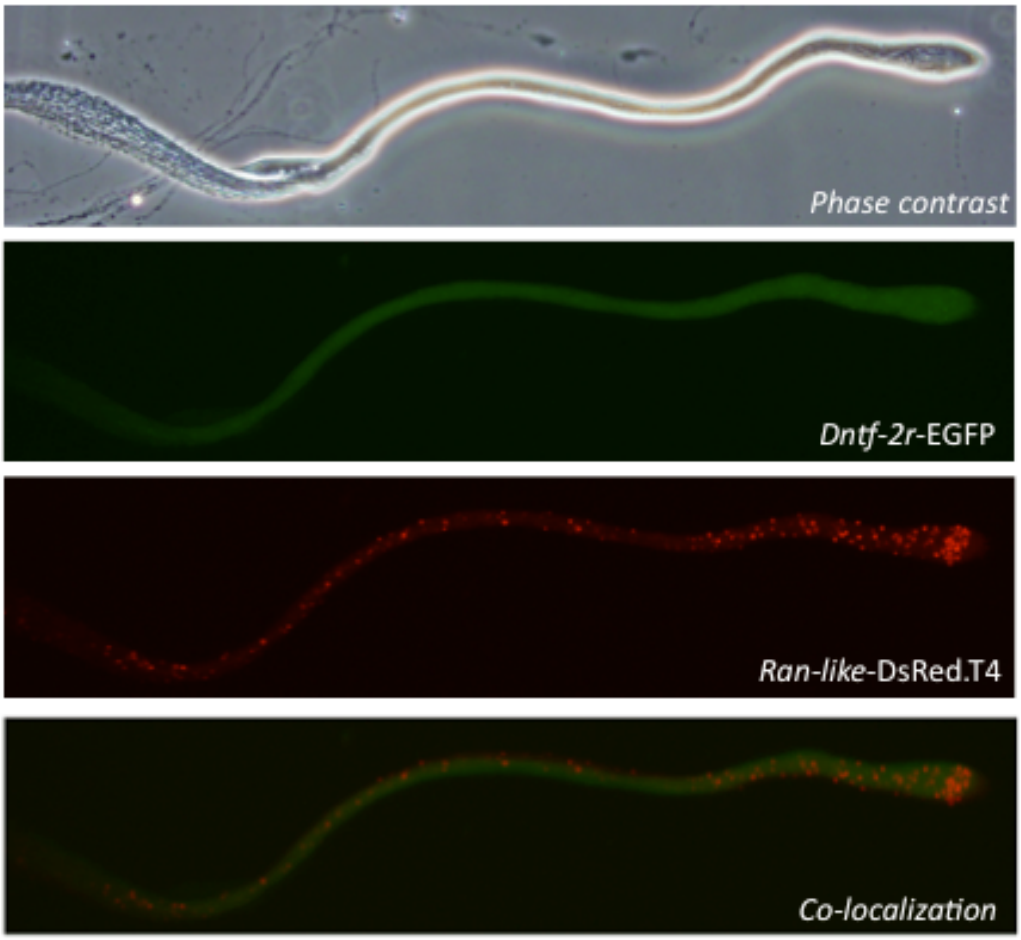


Figure 4.15 – Dntf-2r and Ran-like localization in sperm bundles. Dntf-2r in green and Ran-like in red. The total magnification is 50x.

Getting a close up view on the bundle heads (Figure 4.17) during elongation, I see that both retrogenes co-localize next to the nucleus where the elongating dense body is. This is a structure involved in nuclear transport and nuclear reshaping and it is rich in nuclear pores and microtubules (Fabian and Brill 2012; Figure 4.14).

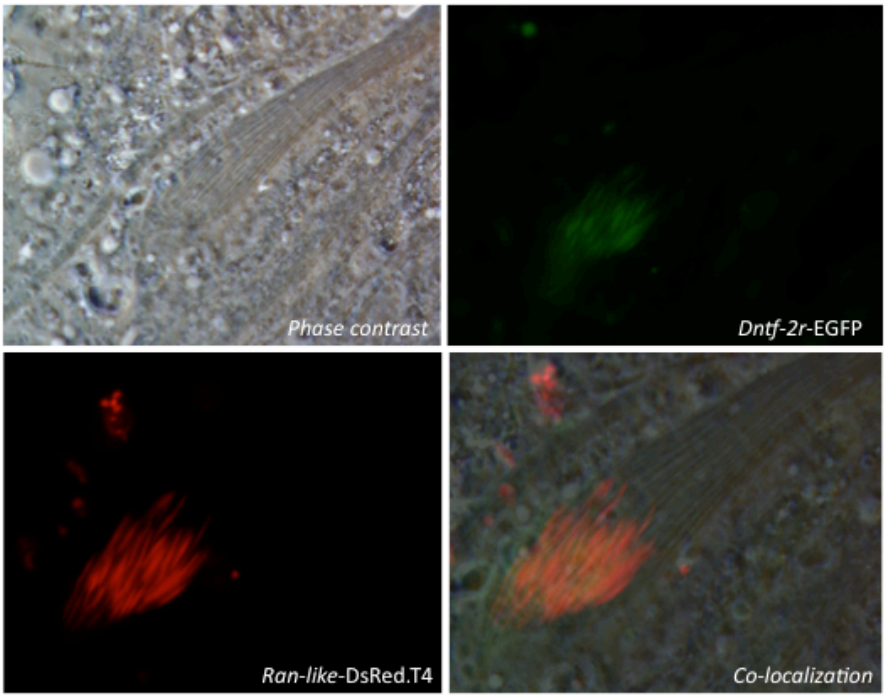


Figure 4.16 – Dntf-2r and Ran-like localization in the nuclei of the bundle of elongating spermatids. Dntf-2r in green and Ran-like in red. The total magnification is 400x.

I also used confocal microscopy to see more details and confirm the expression of Dntf-2r and Ran-like in some of these cell types (Figure 4.18).

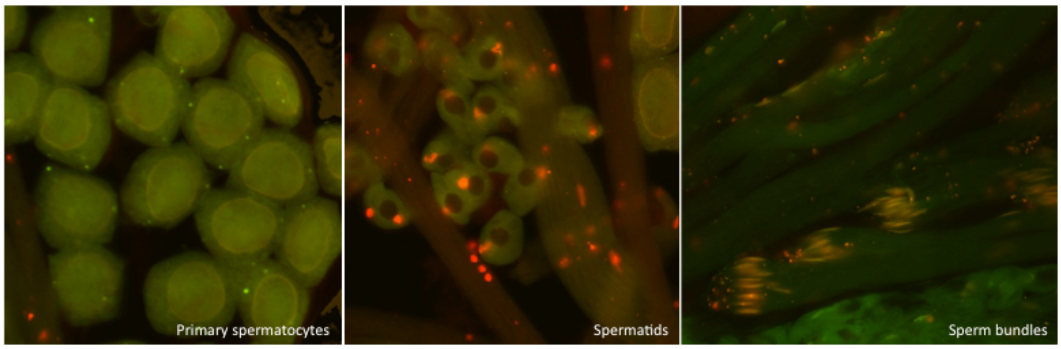


Figure 4.17 – Dntf-2r and Ran-like localization in different cell types. *Dntf-2r* in green and Ran-like in red. The total magnification is 400x.

4.3.2. Retrogene fusion proteins co-localization with known spermatogenesis proteins

Above I checked for the localization of Dntf-2r and Ran-like proteins during the different stages of spermatogenesis. In this part of Chapter 4, I will look at the co-localization between Ran-like that has a sharper localization than Dntf-2r and different proteins known to be expressed during spermatogenesis.

To label the first dividing cells in spermatogenesis I used an antibody against *lamin* protein (Shevelyov, et al. 2009). This antibody stains the nuclear membrane of 2, 4, 8 and 16 cells stages. Co-localization between Ran-like and lamin was observed only in 16-cell stage (i.e., primary spermatocytes; Figure 4.19).

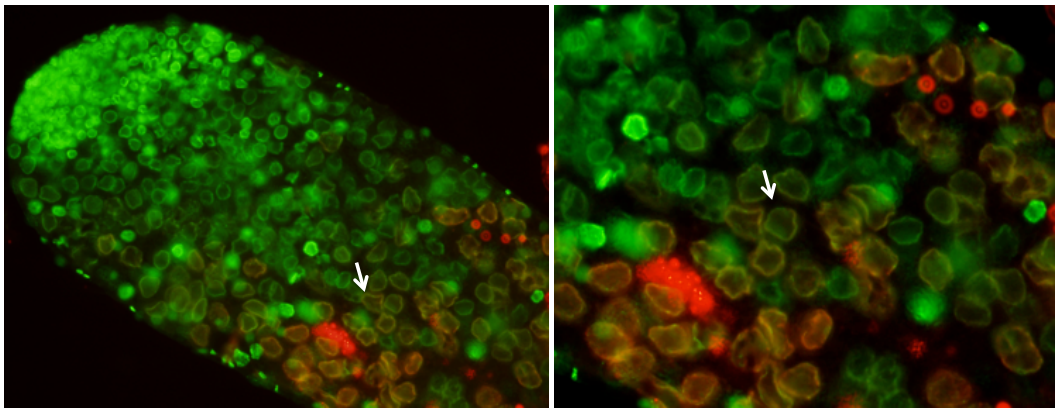


Figure 4.18 – Ran-like fusion protein (red) and lamin (green) co-localizing in primary spermatocytes (overlay between red and green; white arrows).

Anti-Cnn is an antibody that detects centrosomin protein (Cnn) located in the centrosomes, a microtubule organizing centers of the cell and also in primary spermatocytes and in cells during meiosis. Anti-Cnn does not co-localize with Ran-like, but close to it as microtubules are nucleating at centrosomes (Figure 4.20A-C; Anderson, et al. 2009). It seems that there is a gradient of Ran-like strong next to centrosomes and

farther and close to the chromosomes and farther. Similar to what has been proposed for Ran during mitosis. The antibody used for Cnn does not recognize the alternatively spliced variant that is located in the mitochondria during elongation (Dr. Jieyen Chen presentation). So I cannot check co-localization or closeness between these proteins at that stage. Intriguingly, just like Ran and Dntf-2r, Cnn has recently been described to evolve very fast by means of amino acid changes, indels and changes in the reading frame (Eisman and Kaufman 2013). Authors discuss how this result is completely unexpected as this protein is central to cell division.

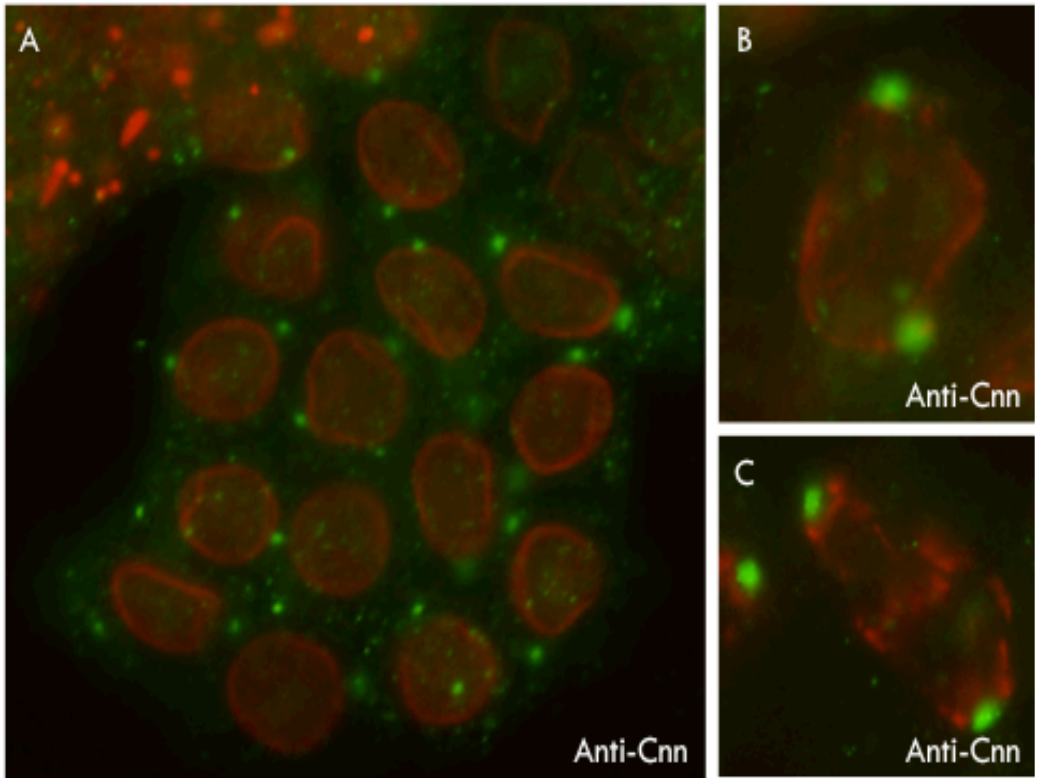


Figure 4.19 – Ran-like (red) and centrosomin (green) in primary spermatocytes (A) and during cell division (B-C).

Since Ran-like appears to help in microtubule nucleation, I also used an anti-alpha-tubulin antibody (Dorogova, et al. 2008). Ran-like co-localizes with alpha-tubulin

(one the two components that make up microtubules) in the cell spindles during meiosis close to the DNA (Figure 4.21A-B) as well as at the microtubules. Using the same antibody, I can also observe the presence of Ran-like during sperm head elongation (Figure 4.21D) in the dense body. It also appears to nucleate the microtubules during elongation and overlaps with the DNA in the sperm bundles heads (Figure 4.21C; Fabian and Brill 2012).

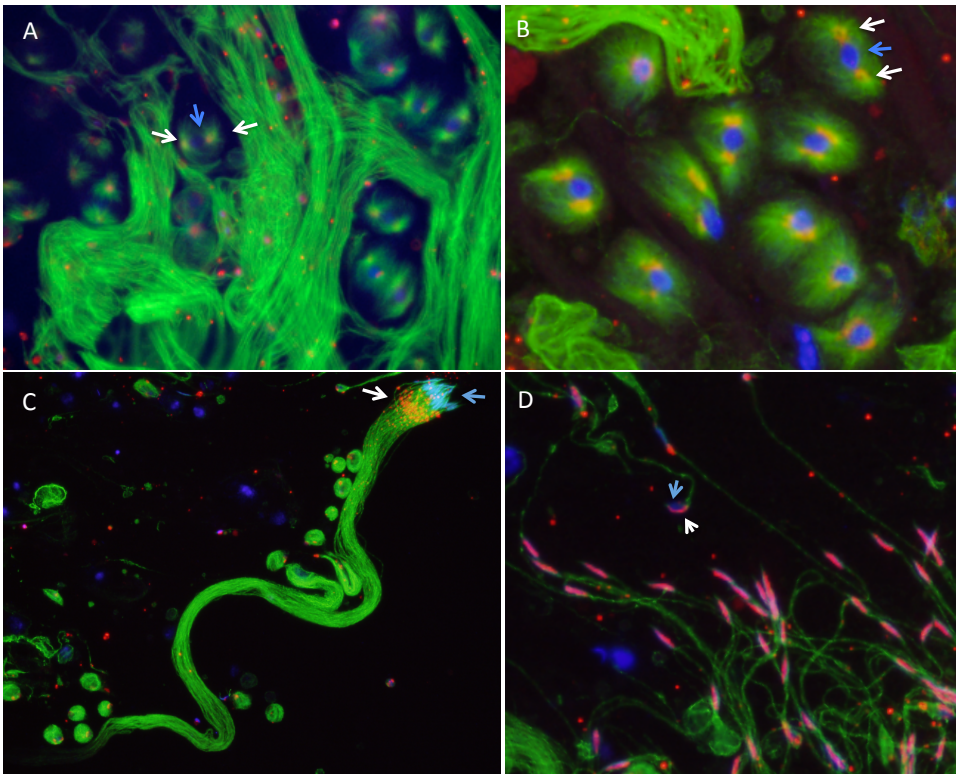


Figure 4.20 – Alpha-tubulin in green, Ran-like in red (white arrows) and DNA staining in blue (blue arrows) during cell division (A-B); in the sperm bundles (C) and during sperm head elongation (D).

To confirm the localization of the retrogenes in the dense body during the onion stage, an antibody against dynein was used. Anti-dynein is an antibody that binds to

many different places in the cell making its visualization difficult compared to the other antibodies (Sitaram, et al. 2012). The results (Figure 4.21) show that in round spermatids Ran-like overlaps with dynein. Dynein is expressed in the dense body helping with sperm head elongation (Anderson, et al. 2009; Sitaram, et al. 2012).

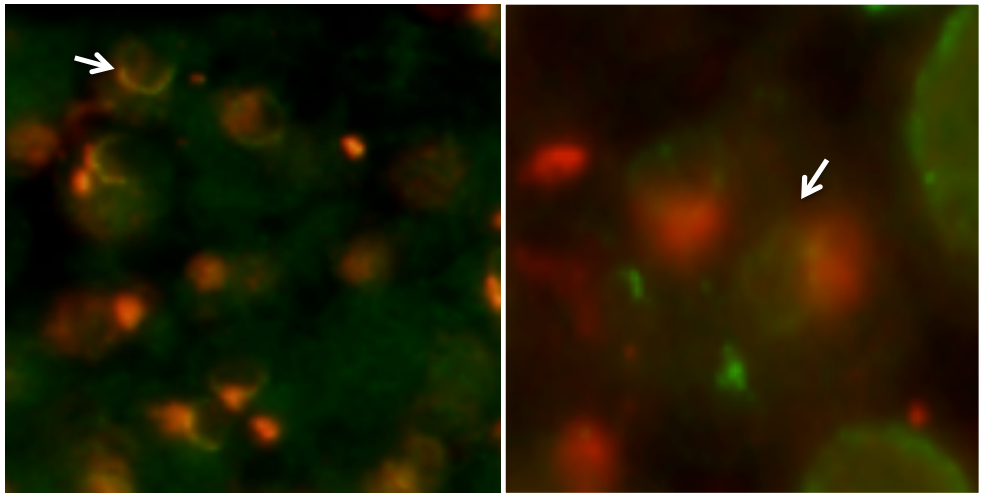


Figure 4.21 – Anti-dynein in green and Ran-like in red in spermatids. Both proteins co-localize in the dense body (white arrows).

DNA and the F-actin cones were also visualized using DAPI and phalloidin during sperm individualization. At this point these cones are responsible for the removal of undesired proteins from the sperm bundles, breaking the bridges between the haploid cells and leading to individual mobile sperm. *Ran-like* is proximal to the F-actin cones (Figure 4.23) and again it seems to be organizing microtubules as everything is being pushed to the waste bags (Figure 4.24).

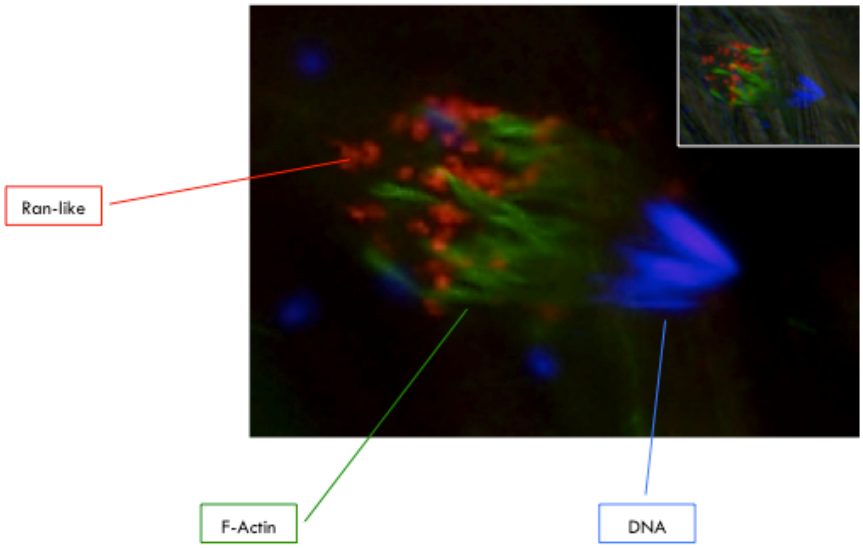


Figure 4.22 – Ran-like (red) and its localization during sperm individualization with the F-actin cones (green) and DNA (blue). The total magnification is 1000x.

Ran-like stays close to F-actin cones during the entire period of individualization of the sperm bundles (Figure 4.24; Noguchi, et al. 2012). Ran-like localization is very similar to the localization of the proteasomes, protein complexes that are degrading proteins as individualization proceeds (Zhong and Belote 2007). However, proteasomes are not directly involved in Ran-like degradation as I observe that Ran-like remains undigested in the waste bag (See below). This likely indicates a more proximal localization of Ran-like in comparison to the proteasomes.

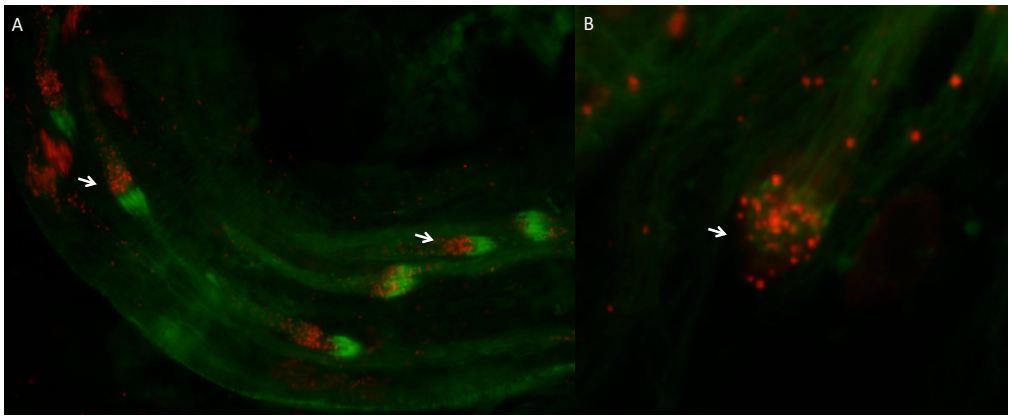


Figure 4.23 – Ran-like (red) and F-actin cones (green) during spermatid individualization. Ran-like is being pushed down as the F-acting cones move through the sperm tail (A) until it reaches the end of the bundle (B).

Proteins that are no longer required and have not been degraded by proteasomes, accumulate in the waste bags (Ghosh-Roy, et al. 2005). Using the anti-alpha-tubulin antibody, alpha-tubulin protein that does not end in the waste bags, because it is required after individualization helps me see the different fate for Ran-like (Figure 4.25). Ran-like is visible in the waste bags.

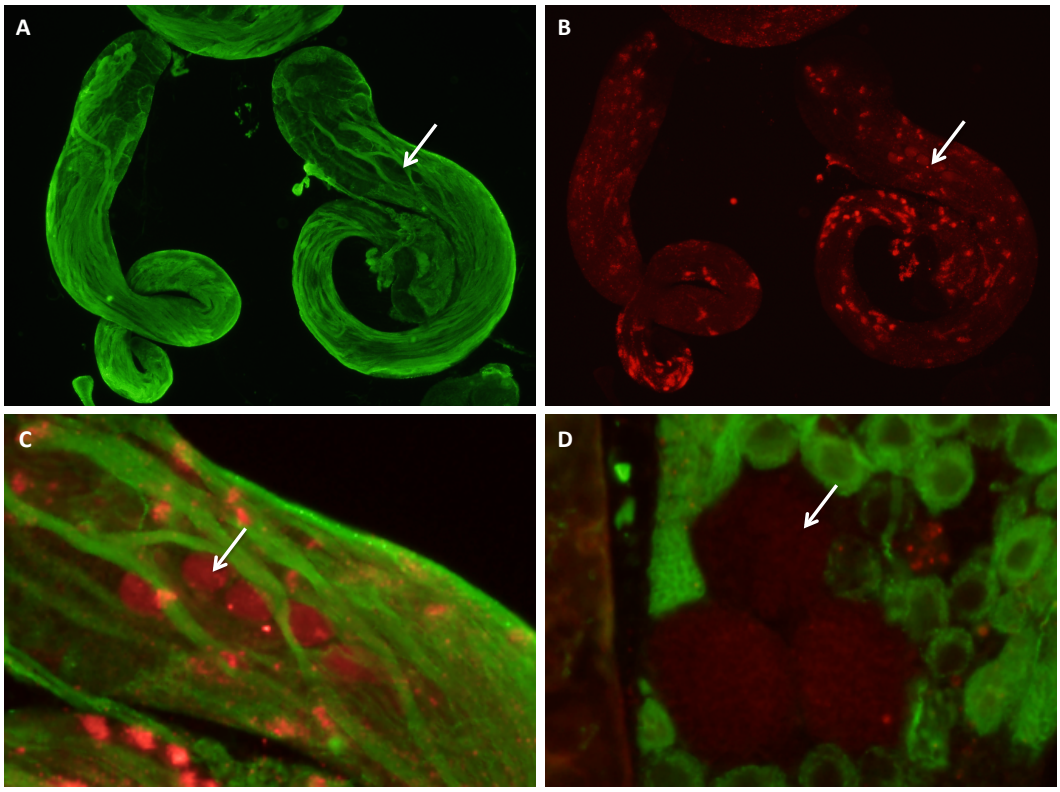


Figure 4.24 – (A) Anti-tubulin staining in whole testis showing its absence in the waste bags. (B-D) Ran-like (white arrow) in the waste bags.

I also wanted to study the localization of the parental proteins during spermatogenesis. Contrary to Dntf-2, anti-Ran antibody is commercially available. I made testes and gut mounts and compared the expression of the parental protein and the retrogene protein fusion. Just like in *in situ* hybridization experiments, Ran is expressed at the tip of the testis where the retrogene is not. However, I observe that both genes are present in the sperm bundles heads. I also observe expression of both genes in primary spermatocytes. Ran-like is more concentrated around the nuclear membrane and Ran is in the nuclear membrane, but also inside the cell concentrated in a particular point. I have not yet identified what this point corresponds to (Figure 4.26E,F).

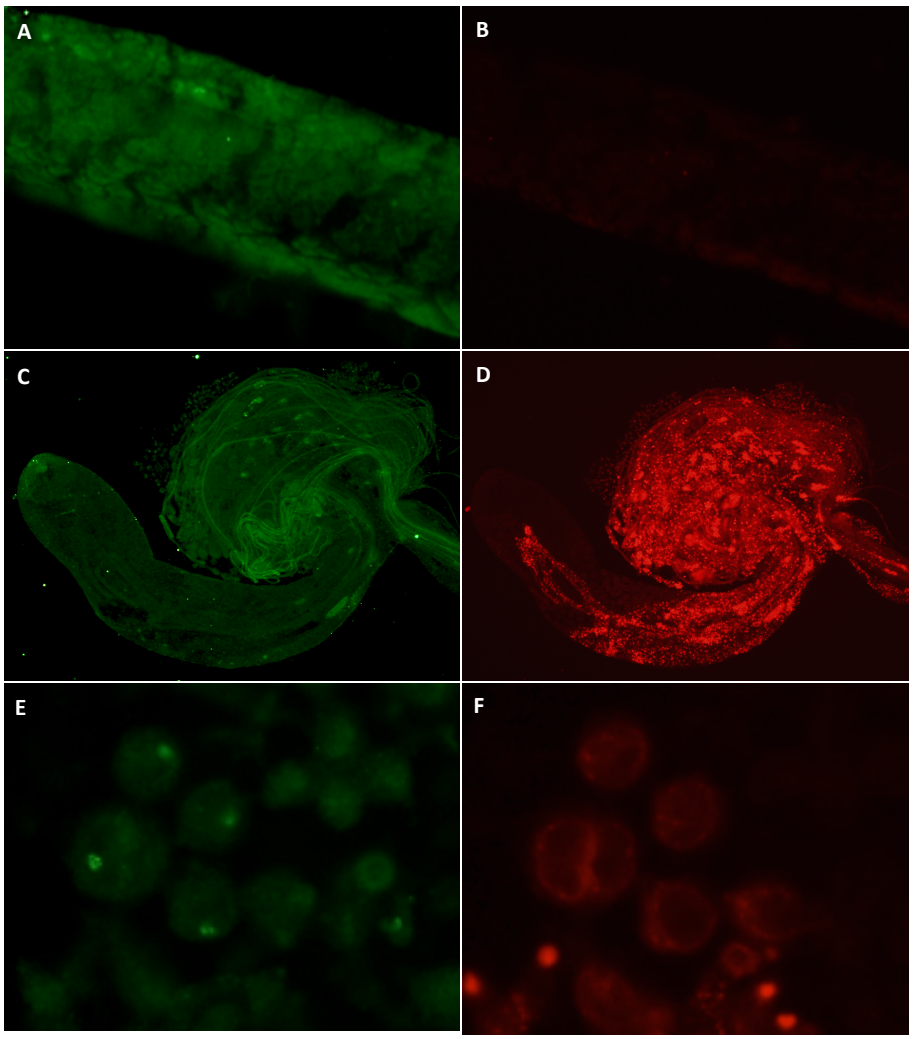


Figure 4.25 – Ran (green) expression in gut (A), whole testis (C) and primary spermatocytes (E). Ran-like (red) in the same organs and cell types: gut (B), whole testis (D) and primary spermatocytes (F).

These results show that Ran and Ran-like proteins co-localize during spermatogenesis in some cell types and cell structures but not always. These results suggest that Ran and Ran-like have different functions during spermatogenesis.

4.3.3. Dntf-2r *and* Ran-like *RNAi knockdown during spermatogenesis*

The knockdowns of *Dntf-2r* and *Ran-like* were initially performed in somatic and germline tissues using UASt males and GAL4 females, and vice versa. In this chapter, I focus on the crosses done for germline and, in the next Chapter (Chapter 5), I discuss the effect of knocking down the retrogenes in somatic tissues.

GAL4/ UAS RNAi crosses were carried out at 29ºC. The F1 progeny from these crosses was checked for viability first and knockdown males and females were selected and crossed with w^{1118} flies. Both germline drivers (*nos*-GAL4 and *bam*-GAL4) had viable progeny. The reduction in transcript levels in testes was checked using *in situ* hybridization performed with retrogene and parental probes to make sure there were no off target effects affecting the parental genes. The fertility crosses were performed as described in section 4.2.3.

In situ hybridizations were carried out in the testes from F1 young males (0 to 1 days old) and from dark pupae just before eclosing. The identification of the cell types in which each probe is detected is difficult due to changes in morphology of some knockdown testes. However, this allowed us to check for the present or absence of each gene transcript and the parental transcript in each cross (Figure 4.27; Figure 4.28; Figure 4.29; Figure 4.30; Figure 4.31). Wild type flies raised at 29ºC were also used as control to ensure that the temperature did not affect testes morphology and probe localization (Figure 4.27).

The results showed that the knocking down of the parental genes using *nos*- and *bam*-GAL4 in testes does not seem to affect the transcription of the retrogenes. Similarly, the transcription levels of parental genes are not affected by the knockdown of the retrogenes. (Figure 4.28; Figure 4.29; Figure 4.30; Figure 4.31). These results indicate

that the knockdowns are specific to either parental or daughter gene without noticeable off target effects.

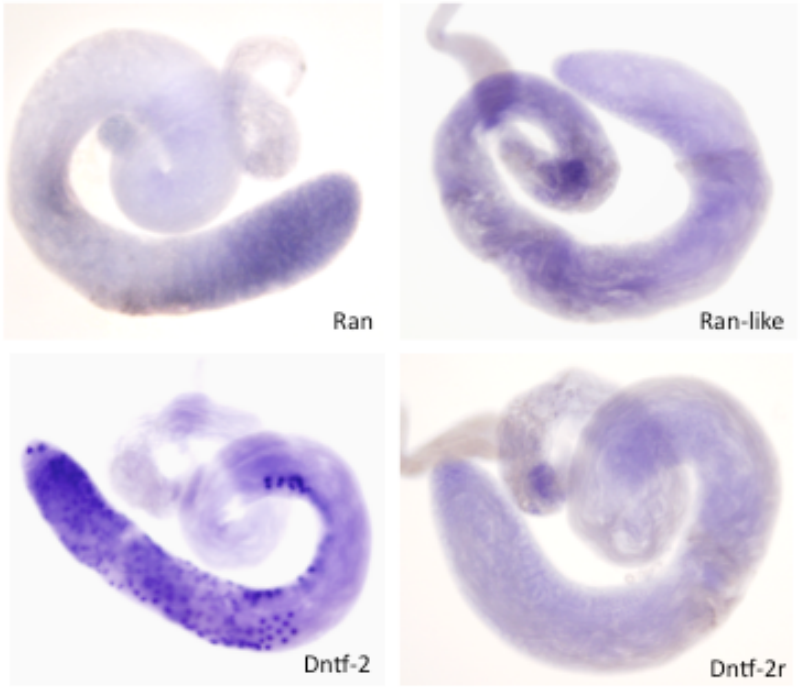


Figure 4.26 – *In situ* hybridization using *Ran*, *Ran-like*, *Dntf-2* and *Dntf-2r* probes. *In situ* performed in wild type pupae testes from flies kept at 29ºC.

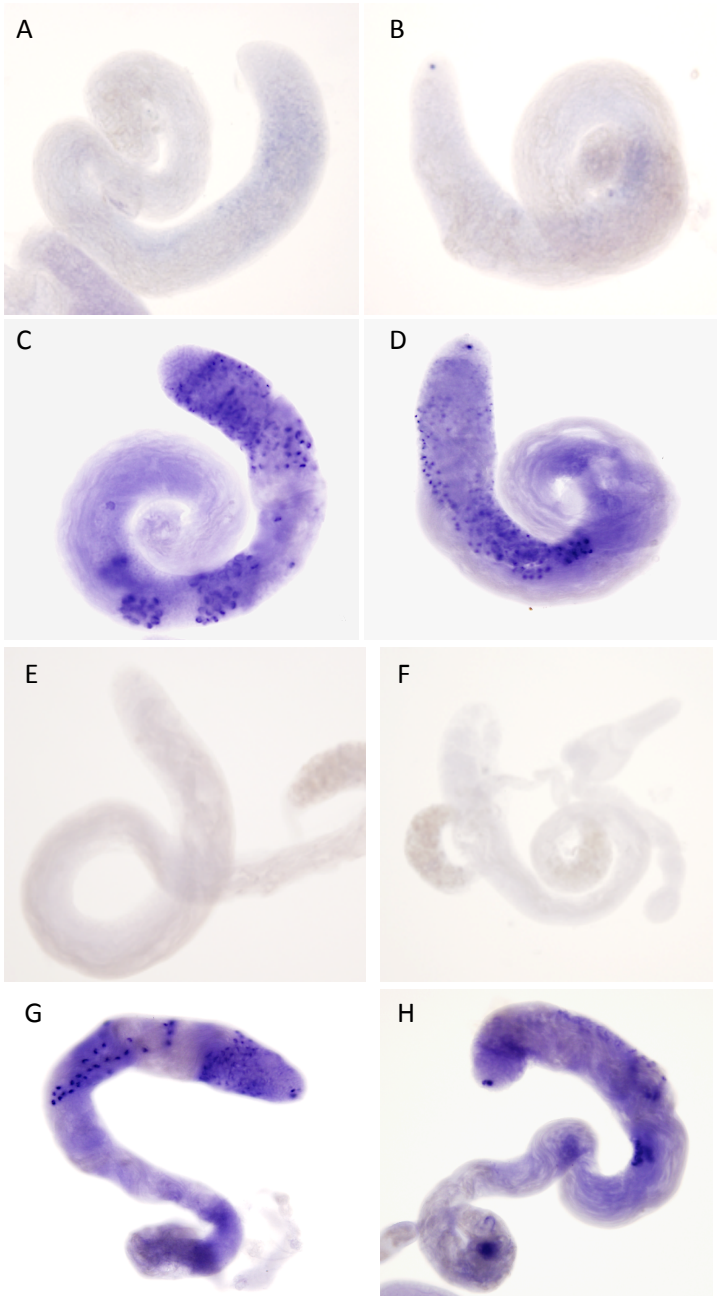


Figure 4.27 – *In situ* hybridization in *Dntf-2r* knockdown testis using *nos*-GAL4 driver (A-D) and *bam*-GLA4 driver (E-H). Hybridization using *Dntf-2r* probe (A-B, E-F) and *Dntf-2* probe (C-D, G-H).

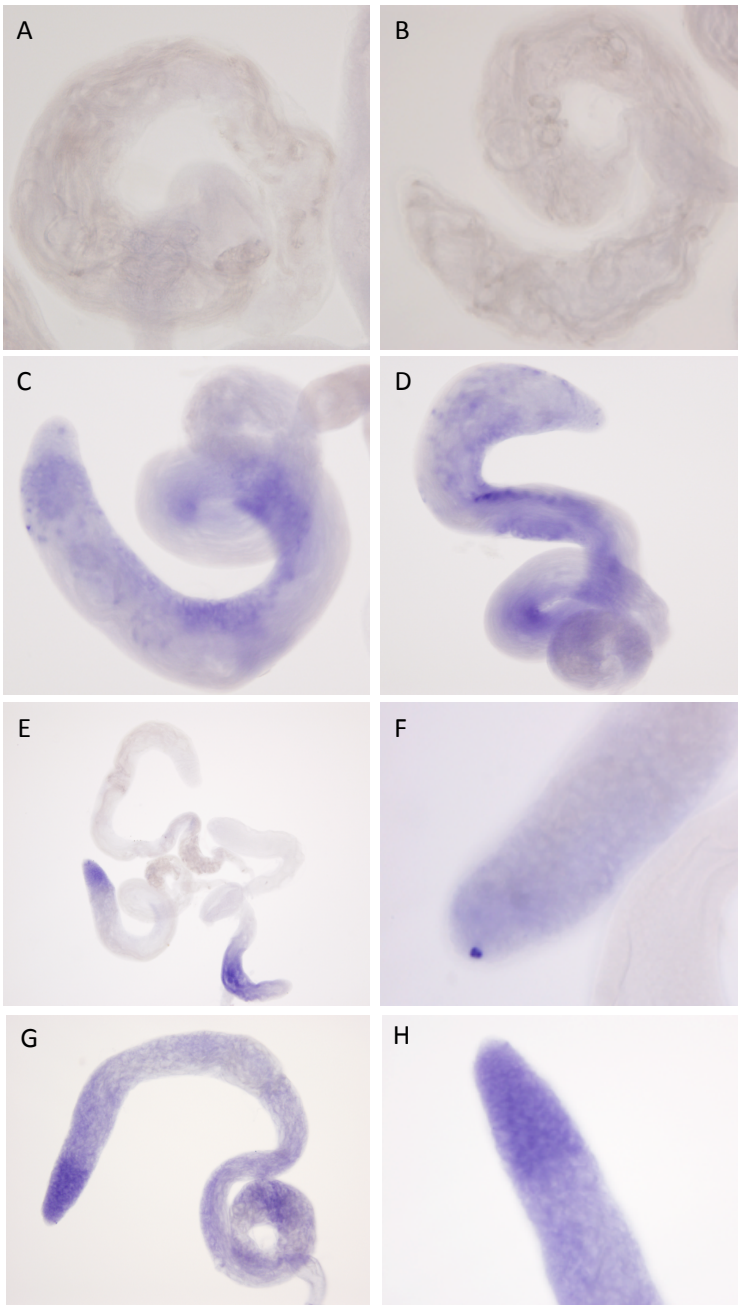


Figure 4.28 – *In situ* hybridization in *Dntf-2* knockdown testis using *nos*-GAL4 driver (A-D) and *bam*-GAL4 driver (E-H). Hybridization using *Dntf-2r* probe (A-B, E-F) and *Dntf-2* probe (C-D, G-H).

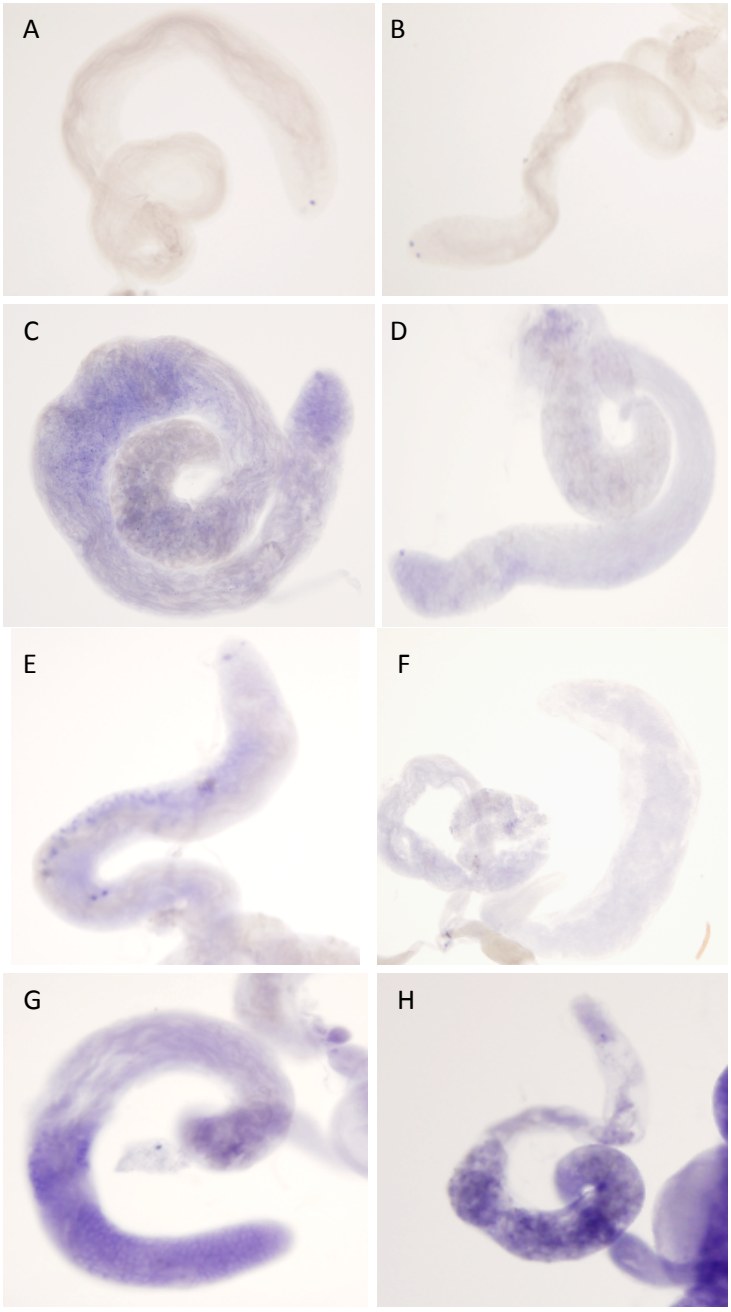


Figure 4.29 – *In situ* hybridization in *Ran* knockdown testis using nos-GAL4 driver (A-D) and bam-GAL4 driver (E-H). Hybridization using *Ran* probe (A-B, E-F) and *Ran-like* probe (C-D, G-H).

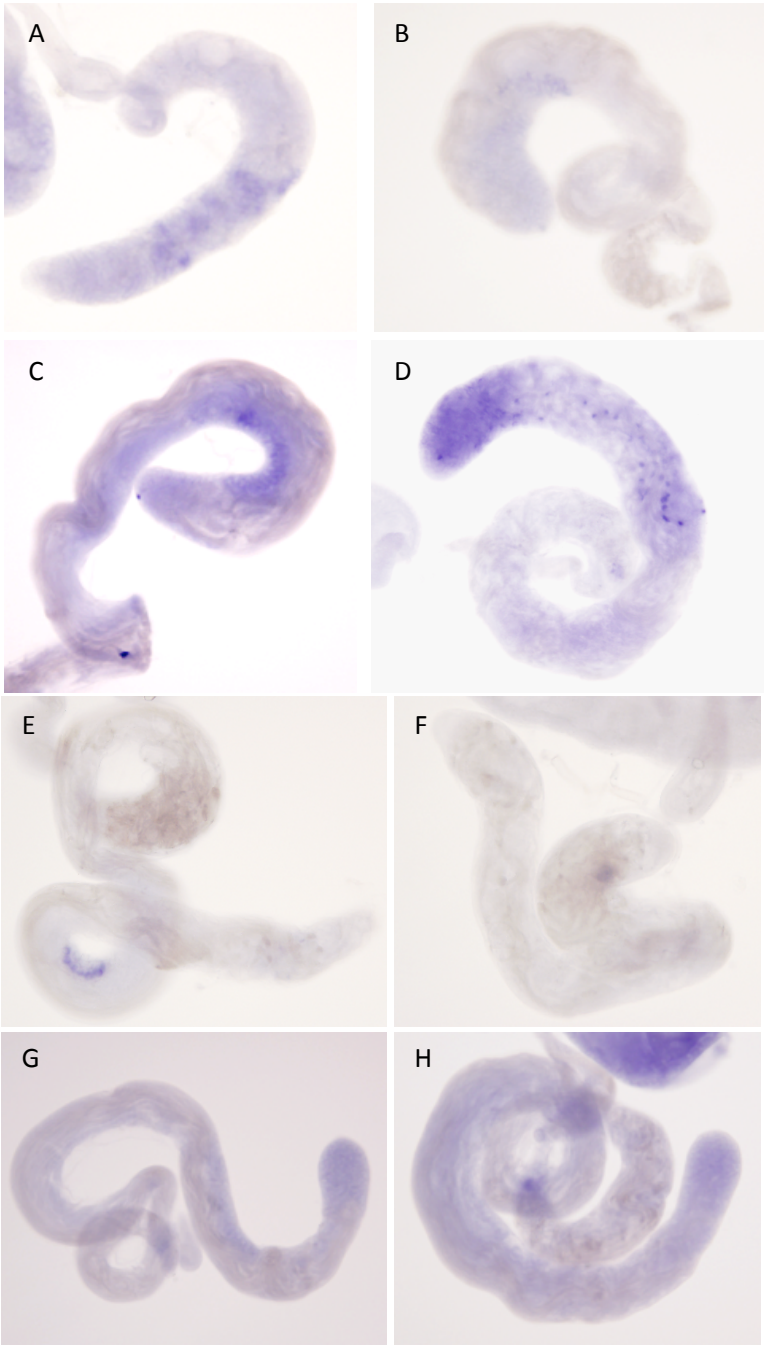


Figure 4.30 – *In situ* hybridization in *Ran-like* knockdown testis using nos-GAL4 driver (A-D) and bam-GAL4 driver (E-H). Hybridization using *Ran-like* probe (A-B, E-F) and *Ran* probe (C-D, G-H).

In the crosses using the F1 with *Dntf-2r* and *Ran-like* knockdown driven by an early germline driver (*nos*-GAL4 driver; White-Cooper 2010), I can see lower fertility in males compared to the control (w^{1118}) and compared to the females (Figure 4.31). *nos* is expressed in male and female germline (Figure 4.31) but RNAi should have no effect in female germline because the retrogenes are not transcribed in ovaries and I am using UASt that does not work well in ovaries (White-Cooper 2012). The same results were observed for the F1 from the *bam*-GAL4 cross (Figure 4.31). Knockdowns with the *bam*-GAL4 driver, a later germline driver (White-Cooper 2010), show less progeny than the crosses with the *nos*-GAL4 driver when males are used. I conclude that the loss of either of the retrogenes has fertility effects. This is consistent with the cytological effects detailed below for those crosses.

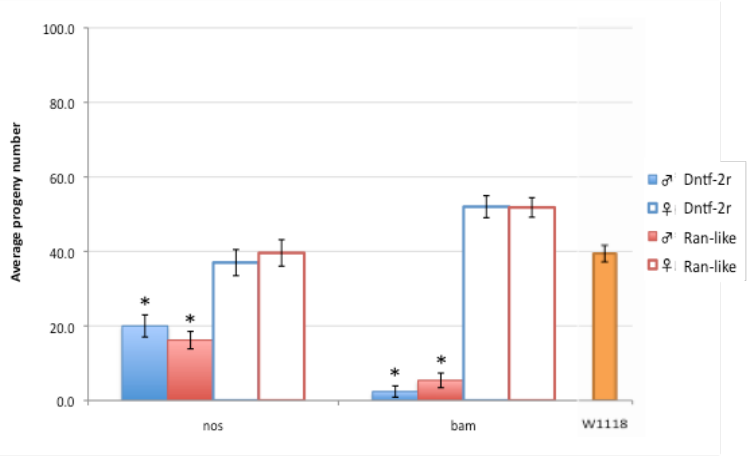


Figure 4.31 – Average progeny number using nos-GAL4 and bam-GAL4 drivers knocking down each retrogene independently in the germline. w^{1118} was used as control line as all transgenic lines were geenrated in this background. Error bars indicate the confidence intervals. (*) Significantly different average values between males and females ($p<0.05$, $n=10$).

4.3.4. Cytological changes in testes of Dntf-2r *and* Ran-like *RNAi males*

Testes from knockdown lines were dissected and different stages of spermatogenesis were observed under the microscope. For each cross DNA staining with DAPI was performed to check the presence of all the spermatogenesis stages.

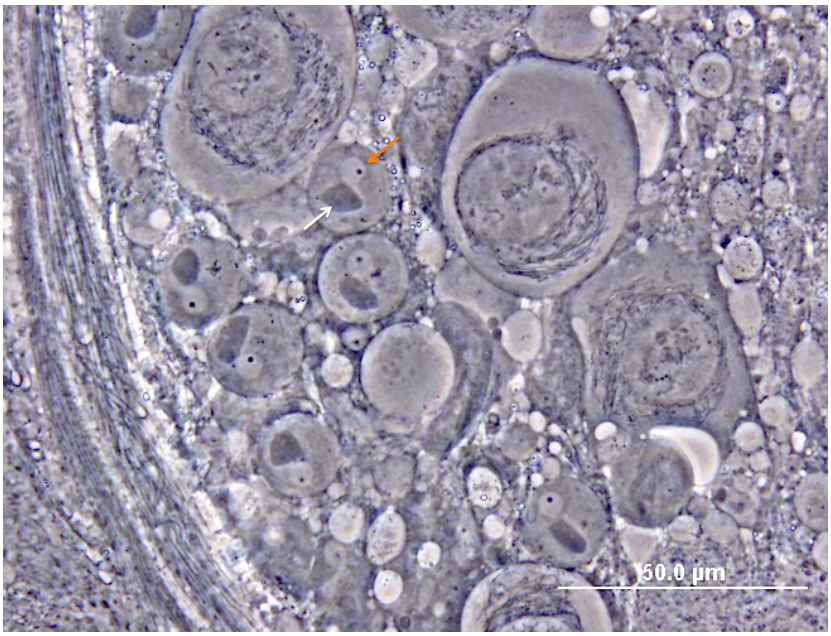


Figure 4.32 – Spermatids under the microscope in phase contrast. Each individual cell during this stage has the nucleus and the mitochondria close to each other as the mitochondria starts to elongate. Nucleus (orange arrow) and mitochondria (white arrow) before elongation in wild type testes.

In the spermatids of the wild type Drosophila testes, the nucleus is located next to the large mitochondrial aggregate, known as Nebenkern (Chen and Megraw 2014). In the spermatids with *Dntf-2r* knockdown, some nuclei appear to be missing (Figure 4.33).

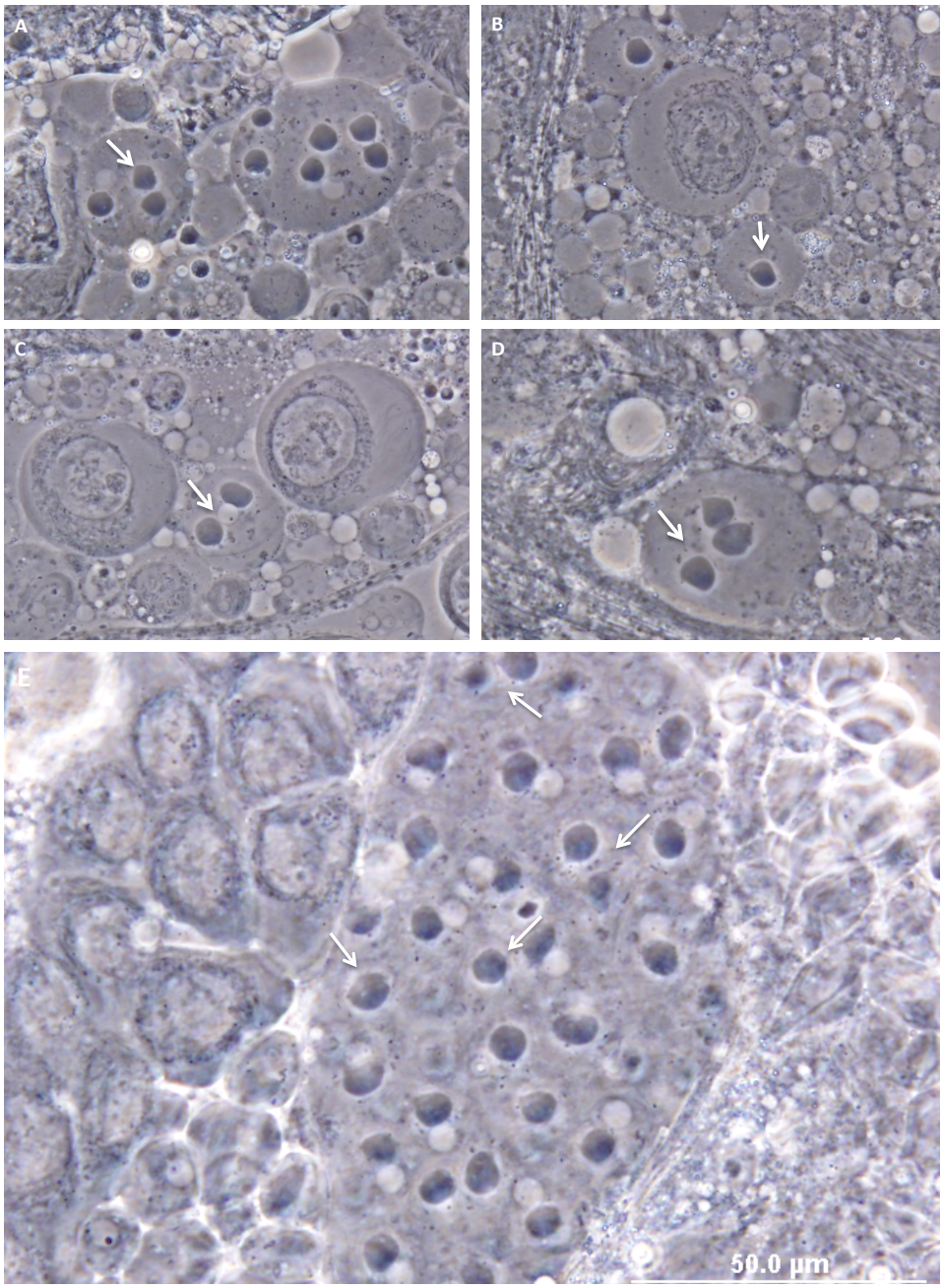


Figure 4.33 – Testis from F1 progeny of *Dntf-2r* knockdown males using *nos*-GAL4 driver. Spermatids (A-D) and a bundle with 64 spermatids (E) under the microscope in phase contrast. Missing nuclei and smaller nuclei are observed (white arrow).

The same results are obtained when knocking down *Dntf-2r* in the germline with *bam*-GAL4 (Figure 4.34). In this knockdown, I observed some spermatids with smaller nucleus compared to the size of the mitochondria (red arrows; Figure 4.34). In the onion stage (spermatids), nuclei and mitochondria have similar size in wild type Drosophila (Casal, et al. 1990). The smaller size of the nucleus might reveal non-disjuction, and the absence of nuclei is likely due to disorganization and loss during cytokinesis or problems during nuclear membrane reassembly. Inspection of *Ran-like* knockdown testes with phase contrast microscopy did not show obvious differences from the controls (Data not shown).

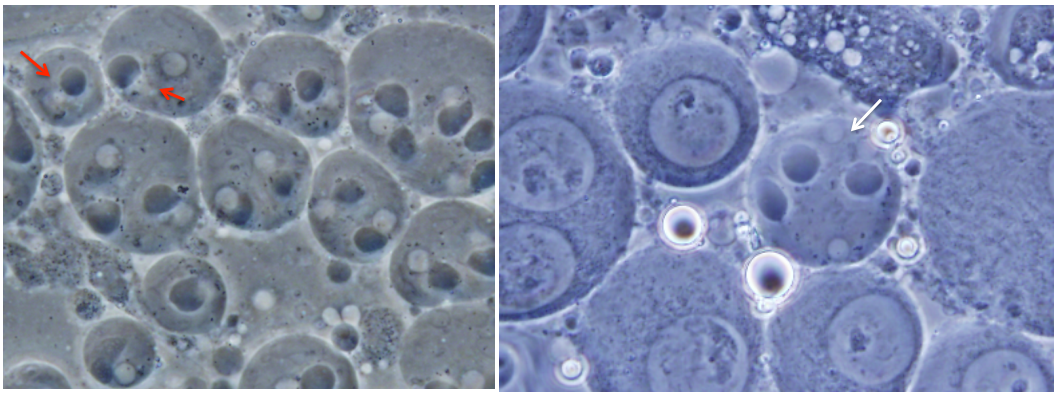


Figure 4.34 – Testis from F1 progeny of the *Dntf-2r* knockdown males using bam-GAL4 driver. Spermatids without nuclei (white arrow) and spermatids with small nuclei (red arrows) under phase contrast.

I also stained knockdown testes with DAPI and mitotracker for the visualization of DNA and mitochondria, respectively (Timakov and Zhang 2001). In the control testis DNA is present in the heads of the sperm bundles after the elongation stage (Figure 4.35). The bundles of sperm heads are close to the basal part of the testis and the bundles of

sperm tails are closer to the apical area, as expected. The elongating mitochondria occupy a larger portion of the whole testis as males get older (Tokuyasu, et al. 1972).

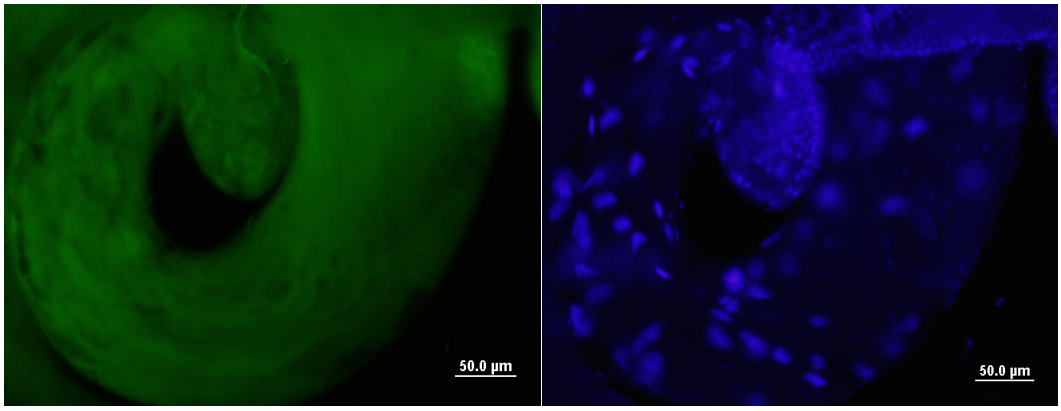


Figure 4.35 – Control testis showing align and organized sperm bundles DNA (right image blue) and the sperm bundle tails stained with mitotracker (left green).

Dntf-2 knockdown in testes, using both nos-GAL4 and bam-GAL4, results in a "baggy" testes phenotype: there is an enlargement of the tip and the midsection of the testes. Most importantly, no mobile sperm is observed in the seminal vesicle (Figure 4.36). These defects are more pronounced when the bam-GAL4 driver drives the *Dntf-2* knockdown.

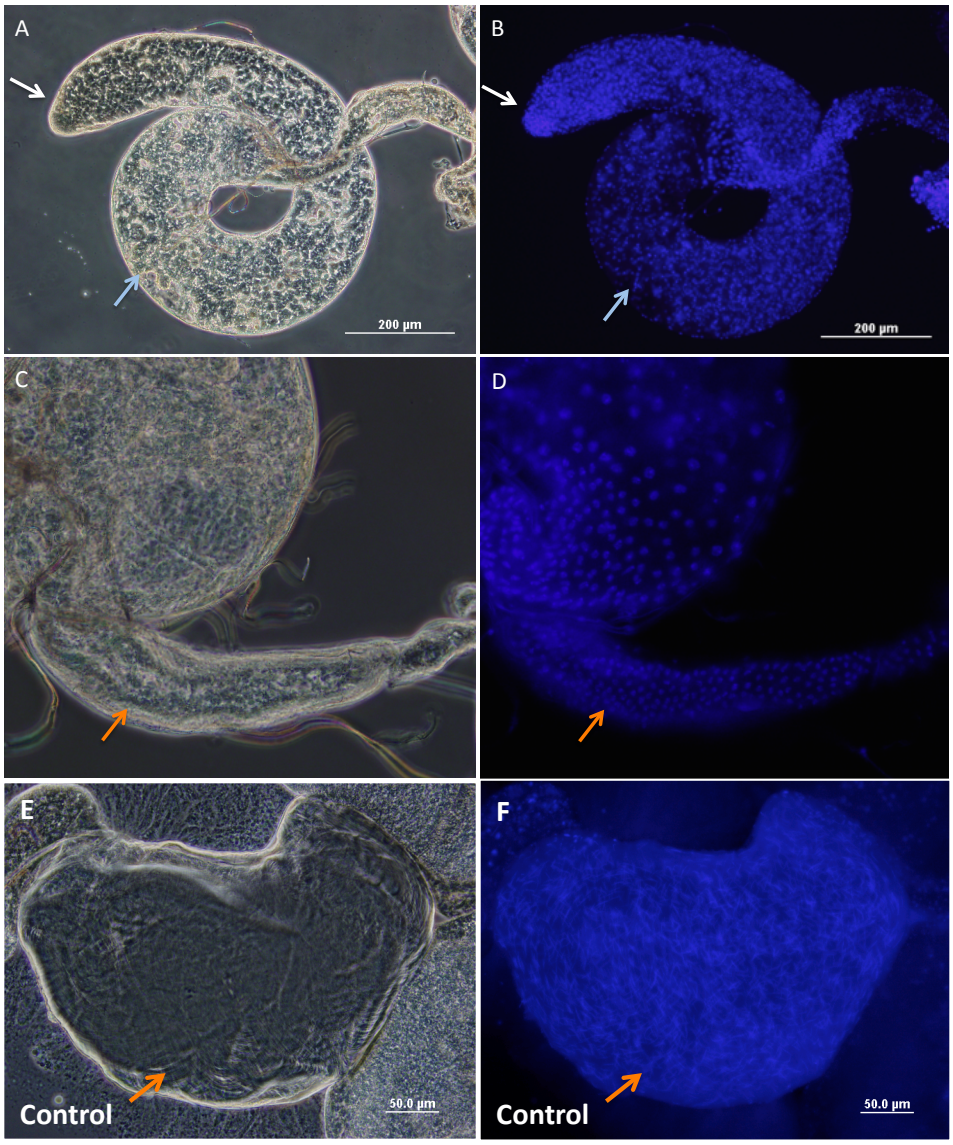


Figure 4.36 – *Dntf-2* knockdown with *bam*-GAL4 in 1-2 days old males showing an enlargement of the middle area of the testes where the sperm bundles can be found before moving to the seminal vesicle (A-B; blue arrow). Extra germ cells are present at the tip of the testis (A-B; white arrow) and no sperm is seen in seminal vesicles (C-D; orange arrows). Control lines seminal vesicle (E-F; orange arrows) full of mobile sperm.

Knockdown of *Dntf-2r* with nos-GAL4 driver results in cytological differences compared to the control testis, with some disorganization in the sperm bundles heads that was not observed in every sperm bundle (Figure 4.37).

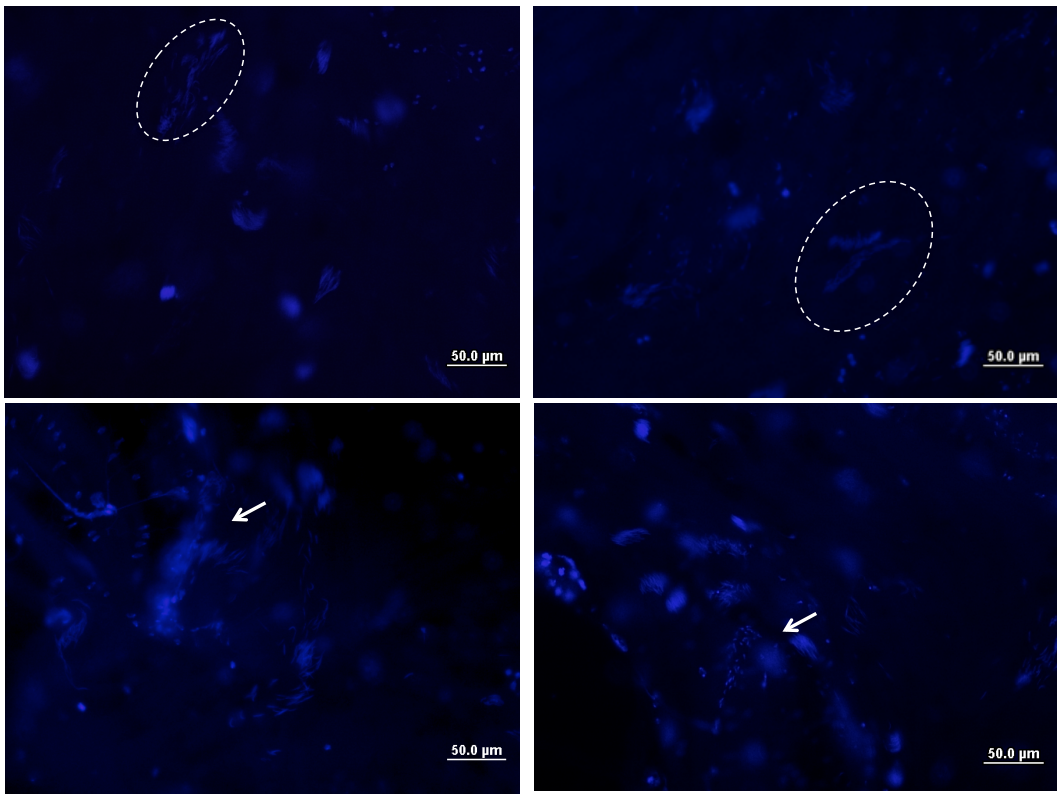


Figure 4.37 – Testes of the progeny of the *Dntf-2r* knockdown males with nos-GAL4. Sperm bundle heads with disorganized DNA (white arrows and circles).

The cytological differences were more pronounced in the F1 progeny of the cross between *Dntf-2r*-UASt-RNAi and bam-GAL4. In this case, I observed the disorganization of the DNA in the bundles and the lack of elongation of the sperm head at the tips of the testis suggesting that this mechanism is compromised in this crosses (Figure 4.38). As explained before *Dntf-2r* and *Ran-like* both co-localize in the dense body in spermatids during the process of formation of the needle-shaped sperm heads. The proteins in this

area are important and help in this spermatogenesis step (Kracklauer, et al. 2010; Fabian and Brill 2012).

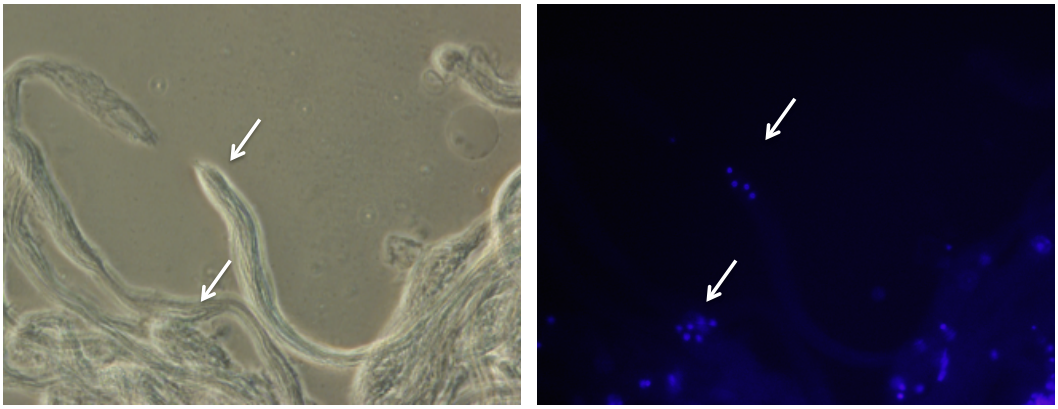


Figure 4.38 – *Dntf-2r* knockdown testes using the bam-GAL4 driver. Few heads of sperm bundles are observed in some bundles and do not elongate (white arrows).

In the case of *Ran* knockdown in germline, using both nos-GAL4 and bam-GAL4 drivers, results are similar to those previously observed for *Dntf-2r*: a loss of the testes morphology and cell organization. Neither young nor older males show mobile sperm accumulation in the seminal vesicle. This defect is manifested in the enlargement of the basal portion of the testis, mainly in the knockdown with nos driver (Figure 4.38). In the bam knockdown, it seems that there is an earlier arrest of the spermatogenesis and consequently, slower production of sperm bundles. Similar phenotypes have been previously described. For example, neurotransmitter transporter-like (*Ntl*), a Drosophila SLC6 gene, is expressed only in the male germline and its mutant shows this phenotype. In *Ntl* mutants sperm morphogenesis appears normal, however there is an accumulation of bundles at the base of the testes just like observed for both parental genes and *Ran*-

like (Figure 4.39; Chatterjee, et al. 2011). Potential genetic interactions with this mutant could be studied.

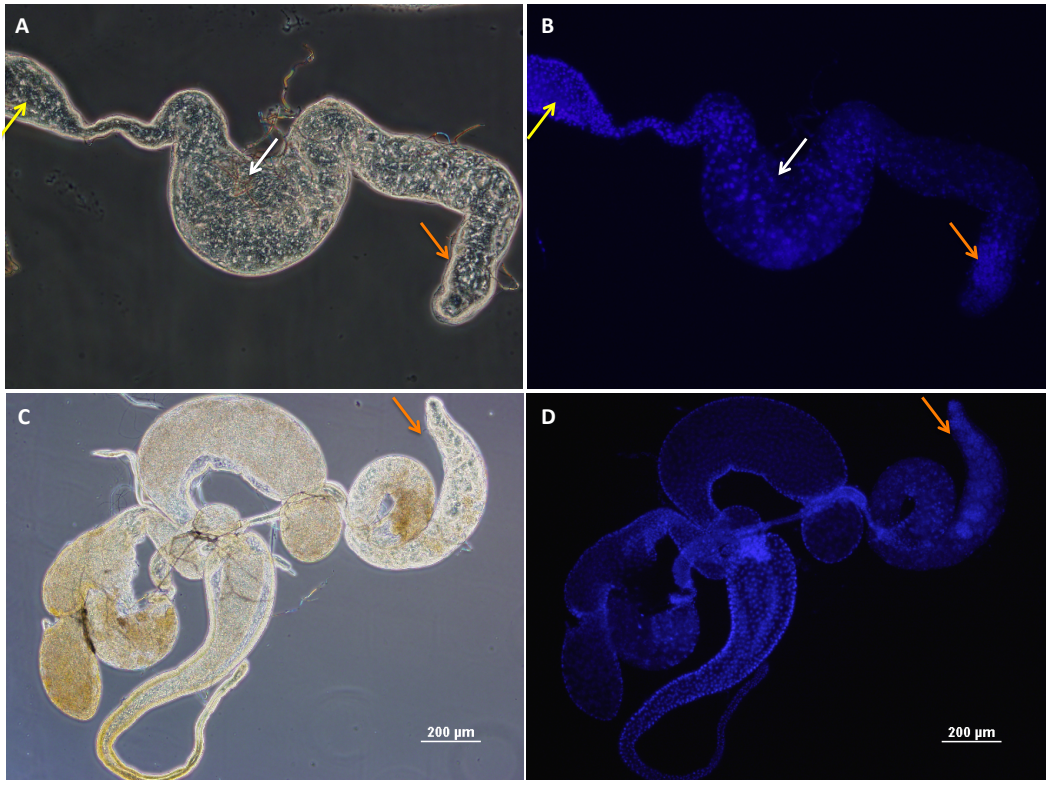


Figure 4.39 – Phase contrast image of a 5-days-old (A, C) *Ran* knockdown testis using nos- (A-B) and bam-GAL4 (C-D). DNA (blue) staining showing cell distribution along the testis. Almost empty seminal vesicles (yellow arrow), basal end of the testis full of mature sperm (white arrows) and tip of the testes (orange arrow).

Ran-like knockdown results in an increase of sperm bundles at the base of the testes and no passage of sperm to the seminal vesicle. Additionally, I observe an expansion of the germ cells at the tip of the testis and the loss of the testis morphology (Figure 4.40).

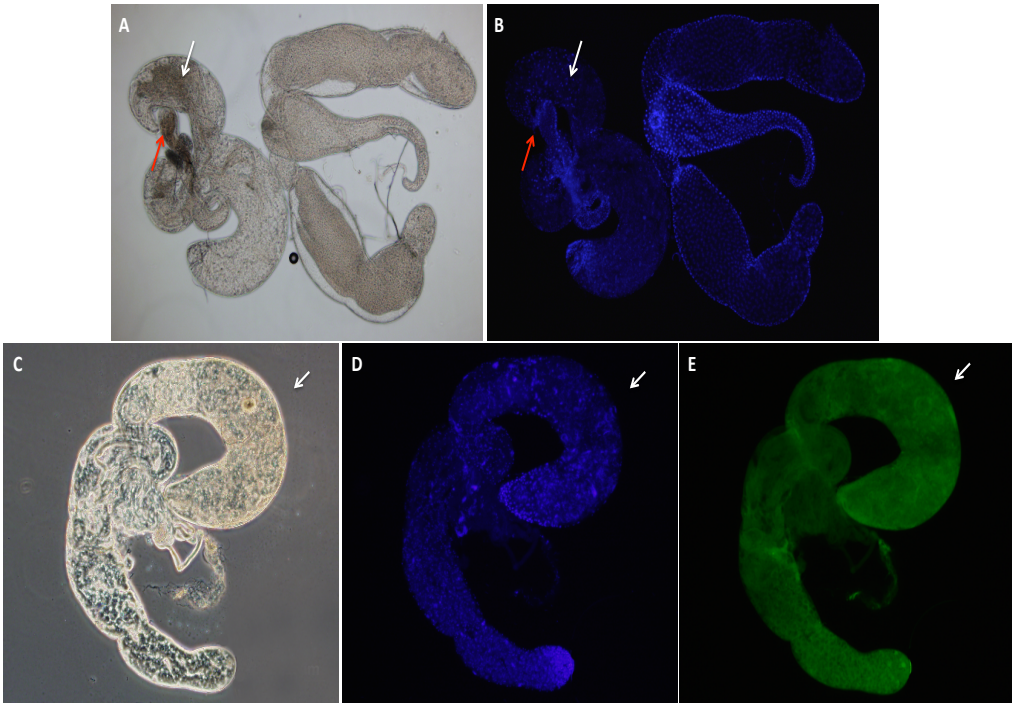


Figure 4.40 – Phase contrast image of a 5- (A) and 10- (C) days-old *Ran-like* knockdown testis using bam-GAL4. DNA (blue; B and D) and mitochondria (green; E) staining showing cell distribution along the testis. White arrow showing the basal part of the testis and red arrow point at seminal vesicle. Seminal vesicle depleted of mobile sperm.

I also observe the disorganization of the DNA sperm bundles when I knockdown *Ran-like* using either nos-GAL4 (Data not shown) or bam-GAL4 (Figure 4.41).

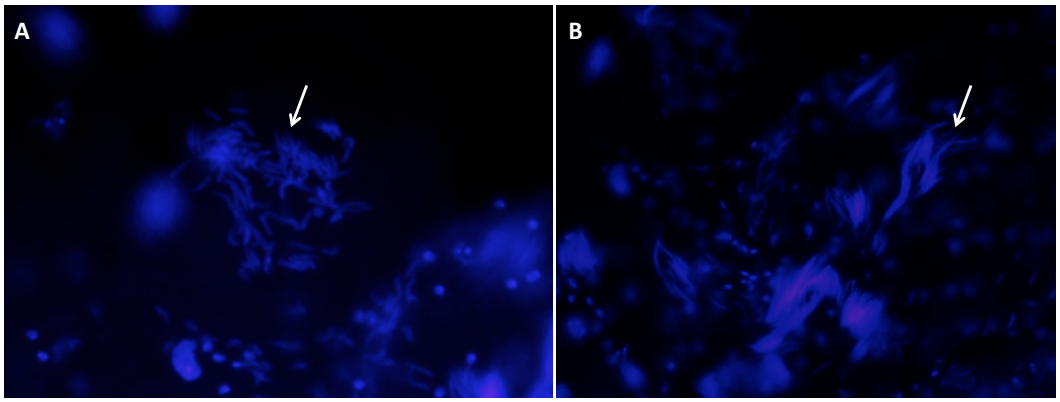


Figure 4.41 – 5-days-old *Ran-like* knockdown testis using bam-GAL4. Disorganization of the DNA distribution (blue) in the sperm bundles heads is shown in A and B.

4.3.5. Ran *mutant rescue with* Ran-like-*EGFP*

Croses were performed to drive *Ran-like*-EGFP in the soma and rescue a *Ran* mutant. Young males from the driving cross between *arm*-GAL4 and UASt-*Ran-like*-EGFP and between *arm*-GAL4 and UASt-*Ran*-EGFP were selected and crossed with female Ran^{-}/FM7c flies. From the progeny of this cross, the number of males with no Bar phenotype and the number of Bar males were recorded and compared to the 1:4 expected ratio between those kinds of males under full rescue.

Table 4.2 – Expected values for Ran mutant rescues and observed averages for three preliminary tests crosses to check which driver would be best (bottom blue numbers) and for second round of mutant rescue crosses using *arm*-GAL4 (top black numbers).

	Expected	Observed		
Ran	25%	**24.8%** 23.6%	**24.4%** 24.1%	**23.9%**
Ran-like	25%	**18.2%** 17.8%	**-** 23.8%	**-**

The *Ran* mutant viability was always fully rescued when the parental gene was overexpressed. When *Ran-like* was used to rescue the parental gene the values were lower than the expected however a partial rescue (i.e., 18.2%, 17.8, and 23.8%) were observed for the crosses performed (Table 4.2). Neither *Act5C*-GAL4 nor *tubP*-GAL4 could be used as drivers because no flies survived past pupae in crosses between these strong drivers and *Ran-like*. I show results about this below. I do not have a good explanation for these results. It could be that Ran-like at a low level can rescue Ran mutant but at high level is lethal, revealing potentially that the genes have same or at least similar functions but different "activity" levels. It could also be that Ran and Ran-like have a negative interaction when driven in the same cells, although this is not evident in spermatocytes.

4.3.6. Ran *and* Ran-like *overexpression*

As described in section 4.3.5. in order to perform the rescue of *Ran* knockout using UASt-*Ran*-EGFP or/and UASt-*Ran-like*-EGFP line, I had to drive the UAS lines by crossing them to the GAL4 drivers first. Driving *Ran*-EGFP independently of the GAL4

driver line used to drive it generated healthy progeny without viability problems. However, driving *Ran-like*-EGFP in somatic tissues resulted in developmental arrest. In the crosses using UASt-*Ran-like* and *Act5C*-GAL4 the arrest happened during the embryo development (Figure 4.42). These embryos appeared rotten and larvae never hatched from them. These crosses were performed at 25ºC and 29ºC with the same result The same was observed in the crosses using *tub*-GAL4 to overexpress *Ran-like*.

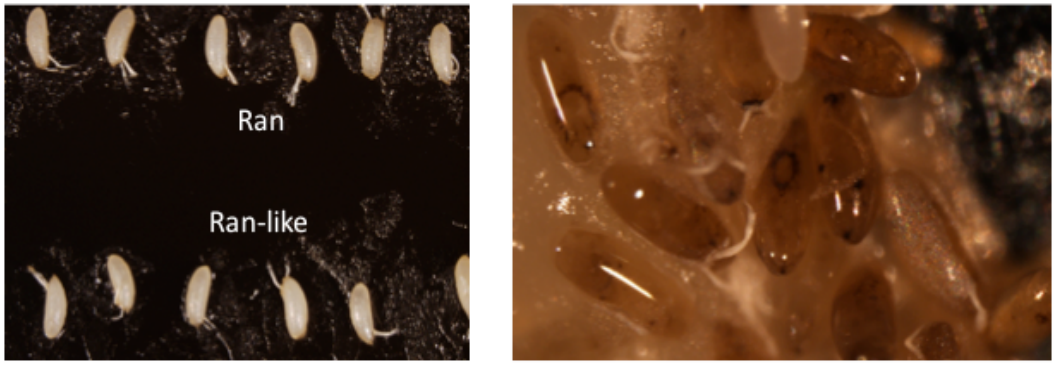


Figure 4.42 – Embryos resulting from driving *Ran* and *Ran-like* with *Act5C*-GAL4. On the left 1-day-old embryos are shown for *Ran* and *Ran-like*. On the right 5-days-old embryos are shown for *Ran-like*. These embryos didn't hatch and look rotten.

Arm-GAL4 is a weak somatic driver and due to it the last overexpression crosses were done by crossing it with UASt-*Ran*-EGFP and UASt-*Ran-like*-EGFP. After the appearance of the first black pupae in the vials all crosses were checked and pupae were observed under the scope. Normal development was observed in the *Ran* crosses, but driving *Ran-like* resulted in pupae arrest in most of the progeny (Figure 4.43). This result either demonstrates the functional divergence between Ran and Ran-like proteins or the

potentially negative interactions between these proteins in the soma, as mentioned above.

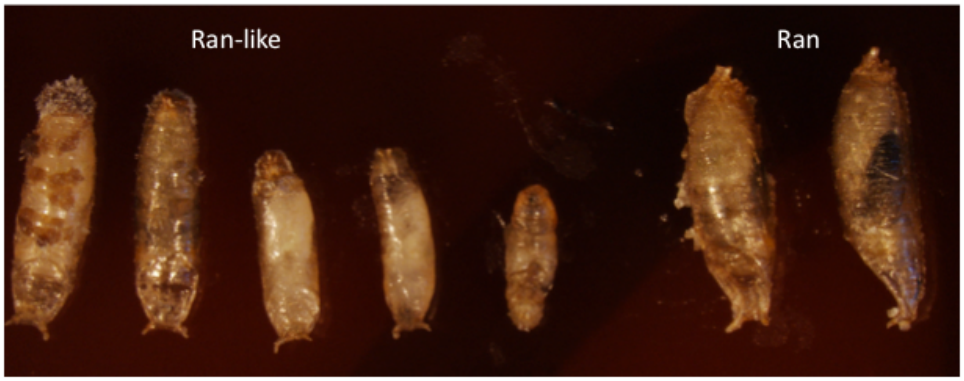


Figure 4.43 - Embryos resulting from driving Ran and Ran-like with *arm*-GAL4. On the left pupae arrest in the *Ran-like* lines and on the right normal pupae observed in the *Ran* crosses.

5.3.6. Dntf-2r *P-element insertion and excision lines phenotypes*

I intended to have knockout lines for both retrogenes to check the loss-of-function phenotype for both genes. However, at this point I have only two mutants available for *Dntf-2r* and none for *Ran-like*. The first *Dntf-2r* mutant is a mutant with a P-element insertion (line 5.1.1) and the second mutant is a P-element excision mutant (line 2.1) that contains a premature stop codon introduced after part of the P-element and target site duplications were left behind during excision and repair (Figure 4.44).

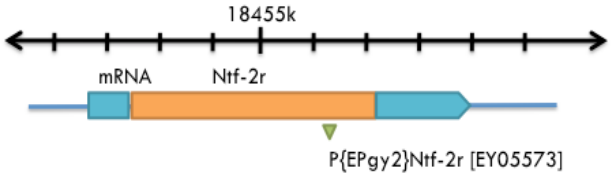


Figure 4.44 – *Dntf-2r* CDS showing location of the P-element insertion in line 5.1.1. In line 2.1 the P-element was removed and a premature stop codon was introduced (See text for details).

The exact location of the P-element insertion is shown below in a protein alingment that also shows the protein domains in Dntf-2r (Figure 4.45). It can be seen that in both mutants all interactions with nucleoporins and RanGDP are retained with the exception of the last one.

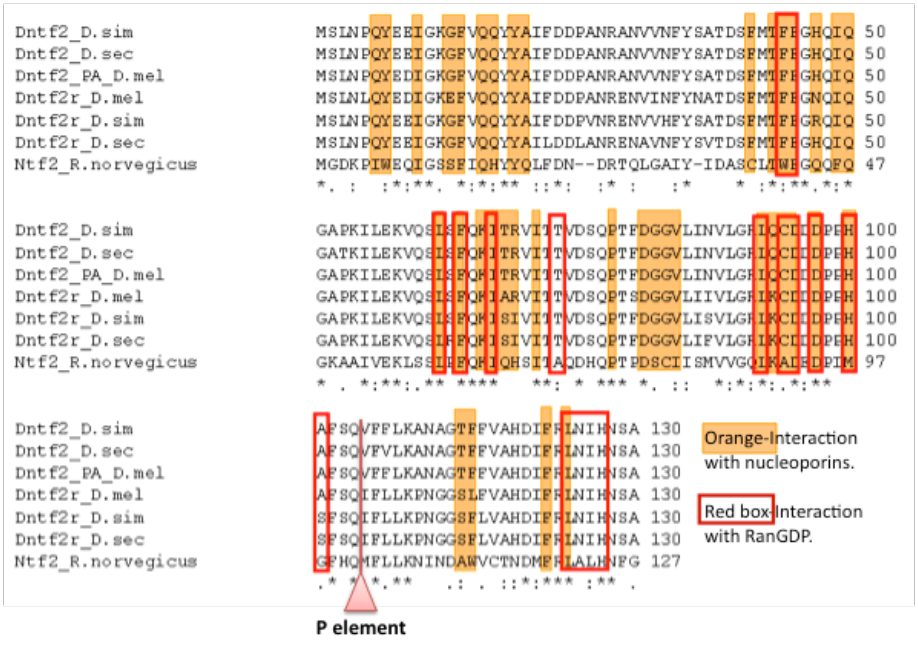


Figure 4.45 – Dntf-2r protein alignment between Drosophila species, predicted interactions and exact position of the P-element insertion in line 5.1.1. In line 2.1 the P-element was removed and a stop codon was introduced.

These two mutant lines were subjected to fertility and sperm exhaustion assays and compared to the control line (line 7.2) where the P-element was excised and the repair of the excision was perfect. No differences were found for the fertility assays between mutants and the control (Data not shown). However, when these lines where subject to a 10 days sperm exhaustion assay, there was a decrease in progeny number after day 7 compared to the control line (Figure 4.46).

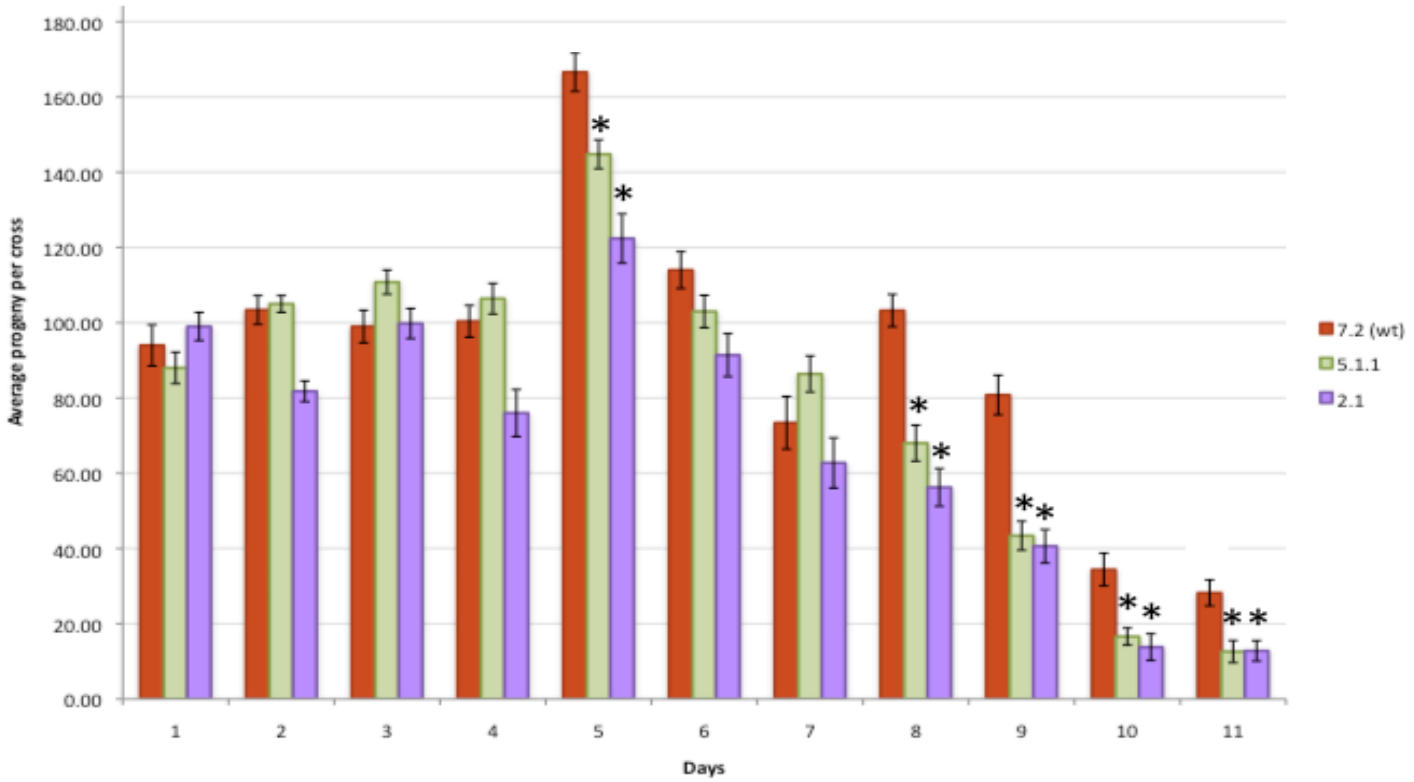


Figure 4.48 - Sperm exhaustion assay for *Dntf-2r* mutants, P-element insertion line (5.1.1) and P-element excision line (2.1), compared to the perfectly repaired excision line (7.2.) * $p < 0.05$. Error bars indicate confidence intervals.

RNA was extracted from males of the 2.1 line and of a wild type isogenic line (i.e., *D. melanogaster* from Besançon) and reverse transcribed. RT-PCRs were performed using two primer sets (Figure 4.47). A longer PCR product as expected from the sequence details described above was observed for the 2.1 line. The translation of this transcript should replace the last 26 amino acids of Dntf-2r with 21 different amino acids and a new stop codon.

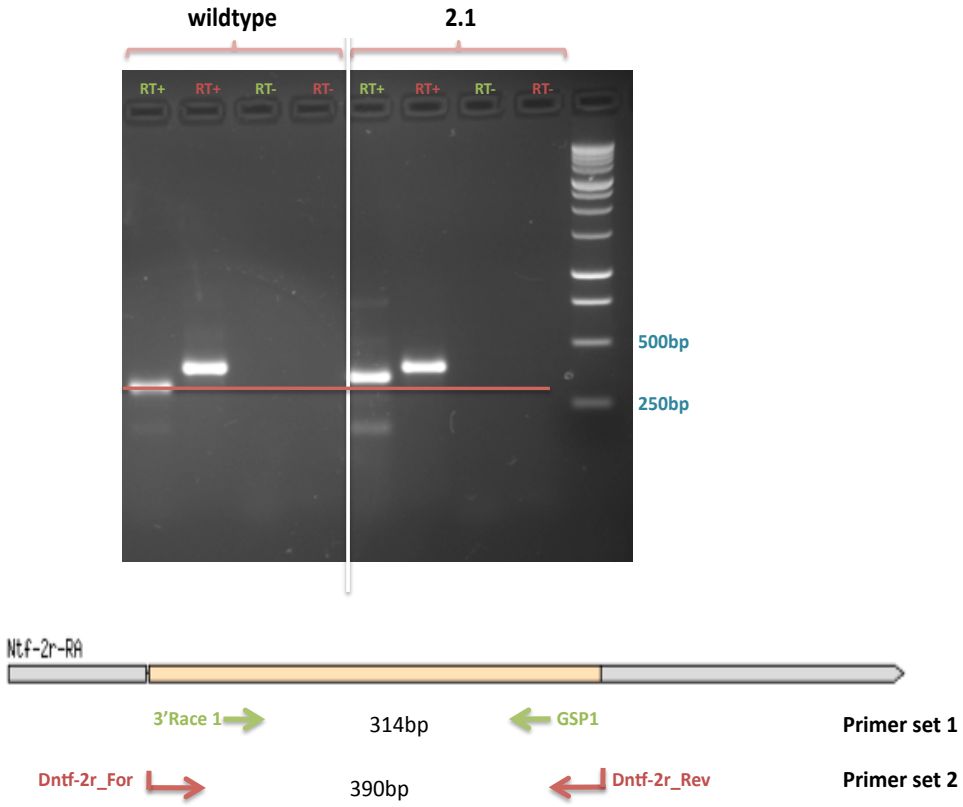


Figure 4.47 – RT-PCR results for wild type (*D. melanogaster* from Besançon) flies and line 2.1. RT+ and RT- for each primer set can be seen in green and red at the top. Primer sets for the RT-PCR and the lengths of expected PCR products are shown at the bottom. A line is drawn to visualize the difference in the PCR product sizes between wild type flies and line 2.1.

To look for possible morphological defects in the mutant, virgin males from the mutant strains were separated from the females and allowed to age for 10 days. The most striking phenotype was observed in the line 2.1, where the testis tip after 7 days becomes a bulb-looking structure (Figure 4.48) filled with primary spermatocytes. These results were not observed for line 5.1.1. The testes tip in line 5.1.1.is similar to wild type. Additionally in line 2.1, a disorganization of the sperm bundles heads was observed (Figure 4.48E,F – red arrows). This disorganization is also present in some of the testis from line 5.1.1. but the phenotype is not so drastic (Image not shown).

This bulb-looking testis tip has been observed when spermatid differentiation is suppressed or slowed down (Grice and Liu 2011). Since this is not a phenotype that is observed in the RNAi crosses, it is possible that the observed effects are not due to the loss-of-function of Dntf-2r, but due to the different protein (truncated Dntft-2r fused to 21 new amino acids before the new stop codon) produced by the 2.1 mutant. I would need to study if this phenotype is rescued by a Dntf-2r wildtype construct once I produce it.

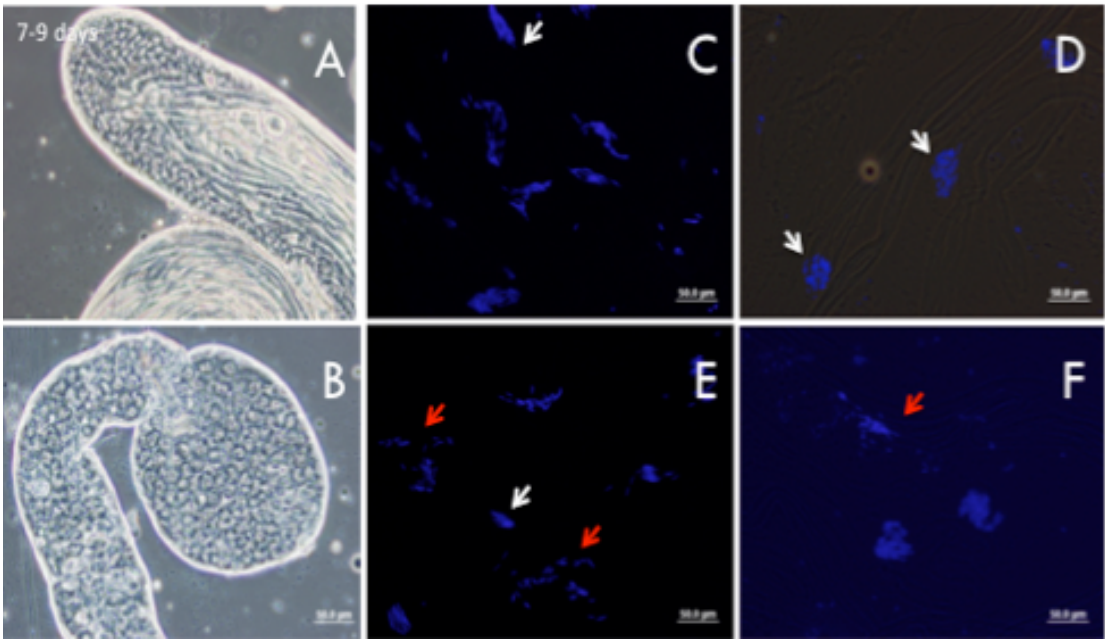


Figure 4.48 – Phenotype observed for a control line 7.2 (top) and for P-element excision line 2.1 (bottom) in 7-9-days-old males. Sperm bundle DNA stained with DAPI (blue). Visualization of disorganized sperm heads (red arrows) and normal sperm bundles (white).

However, *Dntf-2r* P-element excision line 2.1 and *Dntfr-2r* RNAi knockdown show similar spermatid phenotype using phase contrast microscopy. In the P-element line, I was able to observe, as seen for the *Dntf-2r* knockdowns above, the loss of nuclei at the onion stage of spermatogenesis (Figure 4.51).

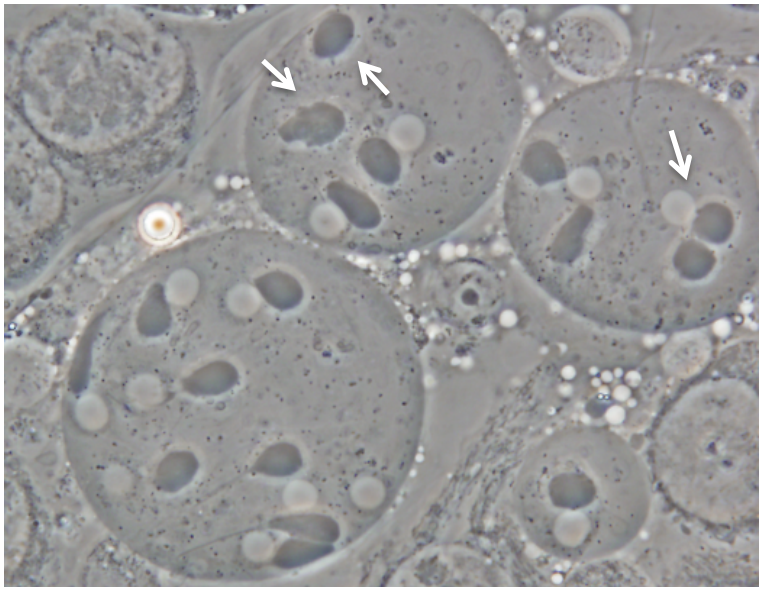


Figure 4.49 – Phase contrast image showing the loss of nuclei compared to the number of mitochondria in spermatids (white arrows) as the Nebenkern starts its elongation in the excision line 2.1.

In the mutant line 2.1, I also observe in some cases what seems to be the lack of aggregation of the mitochondria derivatives. Mitochondria do not wrap around each other to form the spherical Nebenkern (Figure 4.52). Similar pattern has been seen for the *fuzzy onions* (*fzo*) mutant (Hales and Fuller 1997). During spermatogensis the flagellar axoneme grows and the mitochondria in early postmeiotic spermatids aggregate, fuse, and elongate. In *fzo* mutants, males are sterile because they lack proper mitochondrial fusion and elongation (Hales and Fuller 1997). Given the inferred role of Dntf-2r and Ran-like in mitochondria structure and elongation that is likely mediated by microtubule organization (See figures 4.10, 4.12 and 4.13 and text), this phenotype is not unexpected. I, however, do not think that the P-element mutants are complete loss-of-

function mutants because they do not show viability effects. I will talk about the viability effects cause by knocking down *Dntf-2r* in soma in Chapter 5.

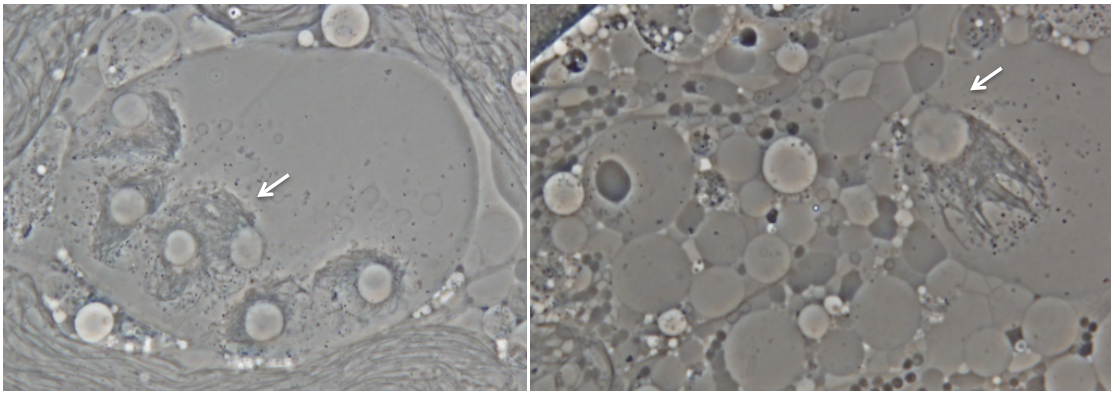


Figure 4.50 – Phase contrast image of the onion stage cells without fusion of the mitochondria derivatives close to the nuclei (white arrow) in the excision line 2.1.

4.4. Discussion

Dntf-2r and *Ran-like* are transcribed highly in testes and appear to require only a small cis-regulatory region to achieve this expression pattern (Chapter 3). Despite this short regulatory region, many cell types from primary spermatocytes onwards show the presence of their transcripts. This is not unexpected as the same regulatory region is found in the *β2-tubulin* gene that is transcribed from primary spermatocytes onwards in male germline (Michiels, et al. 1989). Consistent with their pattern of transcription Dntf-2r-EGFP and Ran-like-DsRed.T4 under their native regulatory region are also express from primary spermatocytes onwards in male germline allowing for the study of the function of these genes in this chapter.

Initially, I intended to produce knockout mutants for *Dntf-2r* and *Ran* and rescue them with our Dntf-2r-EGFP and Ran-like-DsRed.T4 fusions to show the functionality of these fusions. A company (Genetic Services) has injected twice a clone to transcribe RNAs to guide Cas9 mediated mutations in *Dntf-2r* and *Ran-like* using the CRISP/Cas9 system but they have not yet been successful at producing the mutants. Although I have not been able to obtain knockout mutants for *Dntf-2r* and *Ran* and rescue them with our fusions Dntf-2r-EGFP and Ran-like-DsRed.T4 three features of the behavior of our fusions reassure us that they are functional and that their location recapitulates the wildtype location and function of these proteins providing a lot of functional information. First, their localization is the one expected for nuclear transport genes and additional cellular functions known for those genes. This is not trivial. Dntf-2r-EGFP and Ran-like-DsRed.T4 appear inside the nucleus and interacting with nuclear structures and that should not be the case unless they act as transport proteins because they do not carry nuclear localization signal. Second, as expected because Dntf-2 interacts with Ran-GDP, I observe the colocalization of both genes at many stages. Finally, their localization respect to cellular structures and other known genes that I manage to detect using antibodies also makes a lot of sense.

I have learn much about the function of Dntf-2r and Ran-like from the localization of Dntf-2r-EGFP and Ran-like-DsRed.T4 during spermatogenesis. Interestingly, in primary spermatocytes both genes are present in the nuclear membrane with *Ran-like* simply surrounding the membrane and *Dntf-2r* being observed also in the inside the nucleus and in the cytoplasm. When the nuclear envelop breaks down for meiosis I Ran-like colocalizes with microtubules and the spindle. After meiosis localizes with the centriolar adjunt, axoneme and basal body hemisphere. During nuclear elongation *Dntf-2r* and *Ran-like* are present in the dense body. Kracklauer, et al. (2010) showed that

spag4 (*sperm-associated antigen 4*) protein (a SUN protein) and the *yuri* protein have this same localization and are two essential proteins for male fertility (Kracklauer, et al. 2010). SUN proteins maintain the interaction of the nucleus and microtubule organizers to position and reshape the nucleus. *spag4* is a testis specific SUN protein. Spag4 and Yuri are present in the dense body of the spermatids as well. The nucleus of *spag4*$^{-}$ individuals show disorganized nucleus in the cyst and distal curlying similar to the observations in our RNAi crosses with *Ran-like* and *Dntf-2r*. This *spag4*$^{-}$ phenotype was attributed to the lack of centriole and nucleus association during elongation and Ran-like and Dntf-2r appear to participate in this association as well.

Spag4 and Yuri role during nuclear elongation is clear but it is unknown what they do during meiosis close to the nuclear membrane and in the spindles colocalizing with Ran-like and Dntf-2r (Kracklauer, et al. 2010). These proteins might colocalize with Ran-like and Dntf-2r in preparation for the next step.

Ran-like and Dntf-2r, however, appear to have also a function in mitochondria elongation by nucleating microtubules and during individualization retracting the microtubules and proteasomes as F-actin cones advance and ending together with F-actin cones in the waste bag and undigested after individualization.

RNAi knockdown experiments were performed to understand the loss of *Dntf-2r* and *Ran-like* function during sperm production. I wanted to observe if having only the parental genes would have no effect on fertility or this isn't the case and retrogenes are also required in spermatogenesis. *nos*- and *bam*-GAL4 drivers, 4 cells and 8-16 cells stages drivers, respectively, were selected to knockdown the genes. I wanted to carry out these crosses with drivers expressed earlier in the testes however; no testes driver is available to knockdown genes in the testes germ cell (White-Cooper 2012) other than tubulin. In our case tubulin was not an option because it provoked lethality of the F1

progeny. I will enter in more detail on the effects of knockdown *Dntf-2r* and *Ran-like* in somatic cells in Chapter 5. The knockdown results showed that when I knockdown the parental and retrogenes individually in testes male fertility is compromised. *bam*-GAL4 was the driver that showed the highest effect in fertility. These results agreed with the pattern of expression of *Dntf-2r* and *Ran-like* that starts in the 16 cell stage time when nos-GAL4 driver is only weakly present (White-Cooper 2012). The observed phenotypes include nuclear maturation and elongation and sperm tail maturation including mitochondria and nucleus orientation and localization for elongation at several stages leading to the lack of mobile sperm in the seminal vesicles.

Fertility assays and sperm exhaustion assays were also completed using the available mutant lines for *Dntf-2r*. The results obtained were not completely consistent with the results observed for the *Dntf-2r* knockdowns fertility assay. For example the simple fertility test was not observed to reveal differences to the control in these mutants. However, sperm exhaustion and the dissection of the phenotype, DNA staining and observation under the microscope of young and older males, revealed similar phenotype. The truncated *Dntf-2r* protein of 5.1.1 and 2.1 might be hypomorphs (i.e., partial loss of function). An excess of mitochondrial miss structuring has been seen before for other genes involved in microtubule organization and cell division (Casal, et al. 1990). Suggesting that *Dntf-2r* may be indispensable during cell division in male germline.

From the results of trying to rescue *Ran* with its retrogene. I observed that *Ran-like* is lethal when overexpressed somatically using strong drivers Act5C-, tubP- or arm-GAL4 drivers. This makes us think that *Ran-like* function is specialized for spermatogenesis however, since some males survived and were able to partially rescue the Ran^{-} mutant, I infer that some parental functions are still kept by the retrogene. This is also supported by the behaviour of the fusion protein.

I could not perform similar rescue experiment for *Dntf-2* and its retrogene because I am still making the UASt- *Dntf-2r*-EGFP construct.

All data collected in this chapter points to *Ran-like* having a complementary pattern of expression to *Ran* and both genes being indispensable and having individual essential functions during spermatogenesis. *Ran* after duplication may have had its function partitioned and *Ran-like* might have suffered specialization under positive selection but still keeps some parental functions. In the case of *Dntf-2r*, the parental gene is still expressed in the same tissues and *Dntf-2r* could have acquired a new function in the testis interacting with *Ran-like*. The duplication of *Ran-like* predates *Dntf-2r* duplication and this would make sense.

Our study of *Dntf-2r* and *Ran-like* functions does not, however, answer the initial objective of this chpater. I wanted to study the function of *Dntf-2r* and *Ran-like* in testes to understand the reasons why there is a recurrent duplication of nuclear transport retrogenes and why *Dntf-2* and *Ran* gave rise to *Dntf-2r* and *Ran-like* that were retained in the genome. *Dntf-2r* and *Ran-like* are quite old (>6 My old) and currently appear to have many functions in nuclear and chromosome organization during the meiotic cell divisions, during nuclear elongation and during mitochondria and axoneme elongation in a big part by participating in microtubule nucleation. Any of these roles that might enhance spermatogenesis could explain their duplication. It is also unclear what of all the molecular interactions due to all these functions leads to the fast evolution of these proteins. Therefore, it does not appear that *Dntf-2r* and *Ran-like* function is currently suppressing meiotic drive systems or the entry of transposable elements into the nucleus but these could be the aspect of their function that drives their fast evolution. If *Dntf-2r* and *Ran-like* originated to suppress selfish elements they currently have many additional functions. Younger duplications like some of the newly duplicated nuclear transport

genes described in Chapter 2 should be studied, if we are to find a direct link between the event of duplication and the suppression of selfish elements. I have not studied if loosing *Dntf-2r* and *Ran-like* (i.e., in RNAi crosses) enhances the drive of the *D. melanogaster* SD system. It is not trivial to put all the transgenes in a single cross, make sure the second chromosome has a sensitive responder and have chromosomes marked to observe the drive and I hope to do it with the knockouts and will do it when those are produced. However, for what is worth, the P-element insertion in *Dntf-2r* has been assayed and did not modify the drive (McElroy, et al. 2008).

CHAPTER 5

DNTF-2R AND *RAN-LIKE* FUNCTIONS IN SOMATIC TISSUES

5.1. Introduction

Dntf-2r and *Ran-like* are not only expressed in testes, but are also transcribed at moderately high levels in imaginal discs, fat body and accessory glands (Figure 3.1; Figure 3.3). Imaginal discs are tissue-specific progenitors that remain dormant during embryonic and larval life (Harbecke, et al. 1996). Larval tissues include imaginal discs and differentiated larval cells. The 17th century Dutch Biologist, Jan Swammerdam, was the first person to describe the imaginal discs of insects. Always present in pairs in the body of the insect larva, they develop into the external parts of the adult during metamorphosis (e.g., larvae wing disc will become the wings of the adult fly; Figure 5.1). During the pupa stage, each disc everts, elongates and it is broken down. All structures in larva that give rise to adult structures undergo rapid development. Even though during larval stages disc cells appear undifferentiated, their developmental fate is already determined (Harbecke, et al. 1996; Stieper, et al. 2008; Weaver and Krasnow 2008). Most larval cells die during metamorphosis with the exception of some neuron and muscle cells that are retained in the adult tissues (Harbecke, et al. 1996).

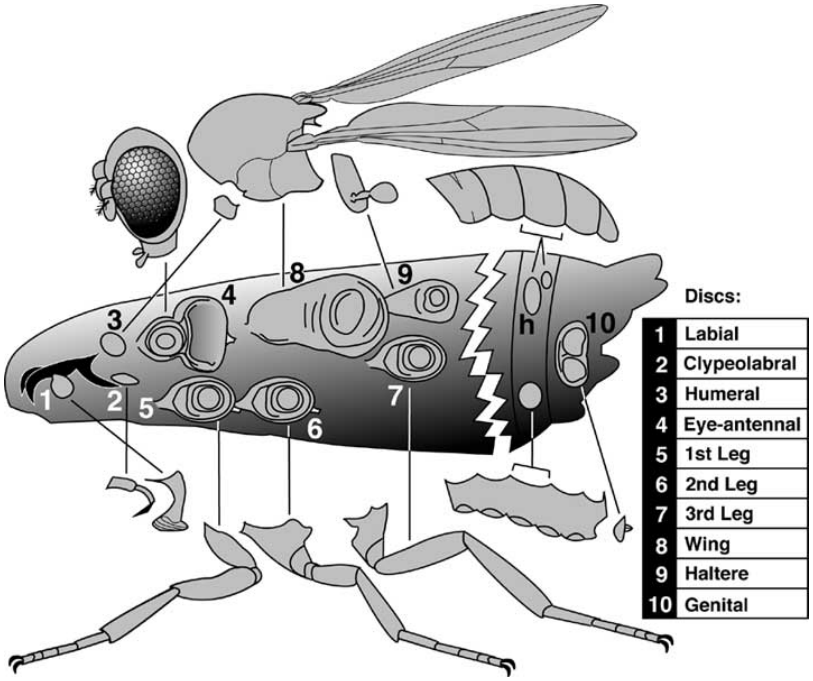


Figure 5.1 – Schematic representation of the different imaginal discs in Drosophila and the corresponding parts in adult after pupal metamorphosis (Image courtesy of Lewis I. Held, Jr,. FlyBase.org).

Exposure to heat damages larval imaginal discs, and a number of studies have shown that when Drosophila larval imaginal discs do not develop correctly, it results in a delay of pupation or even an arrest of metamorphosis (Russell 1974; Simpson, et al. 1980; Stieper, et al. 2008). If the damage occurs after the 3rd instar larval stage, the effect is reduced or eliminated suggesting that at this point the critical size for development of the larva and adult tissues has been achieved (Simpson and Schneiderman 1975; Stieper, et al. 2008).

There are known examples of retrogenes with functions in imaginal discs. Septins are cytoskeletal protein complexes involved in many biological processes. *Sep5* is a recent retroposed copy of *Sep2* in *D. melanogaster*. A knockout of *Sep2* has no

phenotypic effects, but double mutants for *Sep2* and *Sep5* have an early pupal lethal phenotype and lack imaginal discs. Being both genes working still interacting and also both are required in development (O'Neill and Clark 2013).

In my third objective, I wanted to understand the reasons why there is a recurrent duplication of nuclear transport retrogenes and, to answer this, I decided study the function of *Dntf-2r* and *Ran-like* in testis. It is now known that *Dntf-2r* and *Ran-like* transcribe in somatic tissues as well and I need to understand their somatic function to have a complete picture of the function of those genes. In this chapter, I will demonstrate that *Dntf-2r* and *Ran-like* retrogenes have an indispensable function in somatic tissues. I will first confirm and extend the modENCODE expression data with *in situ* hybridization and fluorescent protein tag studies. I will then describe in detail the effects of knocking down *Dntf-2r* and *Ran-like* in the whole body and in specific tissues including imaginal discs, brain and fat body.

5.2. Material and Methods

5.2.1. Drosophila stocks and fly handling

The same UASt-RNAi stocks introduced above were used here: *Ran* (108549), *Ran-like* (12293), *Dntf-2* (17755), and *Dntf-2r* (109227). Every UASt stock was crossed with the somatic GAL4 drivers *Actin 5C*, *armadillo*, and α*tubulin84B* (Table 5.1). They were also crossed with the imaginal discs (indicate discs drivers here), brain and fat body drivers Ap, Lov, N, 1782, Lz and Lsp2 obtained from Bloomington stock center (Table 5.1). *Act5C*- and *arm*-GAL4 are expressed somatically, *tubP*-GAL4 is expressed in soma and germline, *Ap*-GAL4 is expressed in the apterous gene pattern from embryo to adult fly and is involved in an anterior-posterior pattern formation during morphogenesis, *Lov*-

GAL4 is expressed in the brain and CNS during embryogenesis, N-GAL4 drives in the wing discs, *1782*-GAL4 drives in all imaginal discs of the 3rd instar larvae, Lz-GAL4 drives in the eye disc and Lsp2-GAL4 drives in larval fat body. Flies from the GAL4 lines, UASt lines and white mutant flies w^{1118} were used as negative controls in all comparisons.

Table 5.1 - Driver names, stock numbers, expression patterns and genotypes of the imaginal discs, brain and fat body drivers used in RNAi experiments.

Driver (Stock number)	Driver Expression	Genotype
Actin 5C 4414	Soma (strong driver)	y^1 w*; P{Act5C-GAL4}25FO1/CyO, y^+
Armadillo 1560	Soma (weak driver)	w*; P{GAL4-arm.S}11
αTub84B 5138	Ubiquitous	y^1 w*; P{tubP-GAL4}LL7/TM3, Sb^1 Ser^1
Notch (N) 36554	All or most cells of wing disc	y[1] w[*], P{w[+mW.hs]=GawB}N[MD776]/FM6
Lozenge (Lz) 6313	Eye disc	P{w[+mW.hs]=GawB}lz[gal4]; P{w[+mC]=UAS-GFP.S65T}Myo31DF[T2]
1782	Imaginal discs	w[*]; P{w[+mW.hs]=GawB}32B
Apterous (ap) 3041	Moderate to high expressed in imaginal discs (*ap* pattern)	y[1] w[1118]; P{w[+mW.hs]=GawB}ap[md544]/CyO
Lov 3737	Expresses GAL4 in third instar larva: brain - throughout CNS, strong in central brain, not in discs	w[*]; P{w[+mW.hs]=GawB}lov[91Y]
Larval serum protein 2 (Lsp2) 6357	GAL4 expressed in third instar fat body	y[1] w[1118]; P{w[+mC]=Lsp2-GAL4.H}3

5.2.2. In situ *hybridization*

I set up the protocol for *in situ* hybridization in different tissues in the laboratory. I modified the *in situ* hybridization protocol for whole testes described by Morris et al.

(2009) to work in imaginal discs and other larval tissues. To denature the probe, it was heated at 100ºC for 10 minutes. The prehybridization and hybridization steps were carried out at 45ºC. The overnight incubation at 4ºC was done with 0.1% BSA in PBST and the anti-digoxygenin antibody. After the color development, the reactions were stopped with PBST washes (four times for 10 minutes each). After the four washes, all the PBST was removed and 30% glycerol in PBST was added for 30 minutes, followed by 50% glycerol, and, finally, by 70% glycerol.

For *Ran-like* and *Ran* probe production, specific primers were used to amplify each gene from *D. melanogaster Besançon* wildtype strain genomic DNA. DNA and RNA probes were made according to protocol described by Morris et al. (2009) and used to test if using DNA or RNA probes would make a difference in the results obtained for the *in situs* in larvae. I again used both RNA and DNA probes to verify that the results are independent of the type of probe used. For the *Dntf-2* parental gene and *Dntf-2r* retrogene only DNA probes were produced.

Probe labeling was performed using DIG-High Prime DNA Labeling and Detection Starter Kit I (Roche, Indianapolis, IN) following the protocol described in Chapter 3.

5.2.3. RNAi knockdown crosses

Knockdown crosses were performed as described in Chapter 4 by crossing five young (1-2 days old) UASt males with five virgin (1-2 days old) GAL4 females and vice versa. For each cross, 3 replicates were made at different times. These crosses were incubated at 29ºC for 10 days. The knockdown offspring were checked under the dissecting microscope for phenotypic effects. Flies from the GAL4 lines, UASt lines and just white mutant flies w^{1118} were also kept at 29ºC and used as negative controls for the

comparisons (Figure 5.2). In these knockdown crosses 9, different drivers were chosen based on the mRNA seq and *in situ* data collected in this chapter.

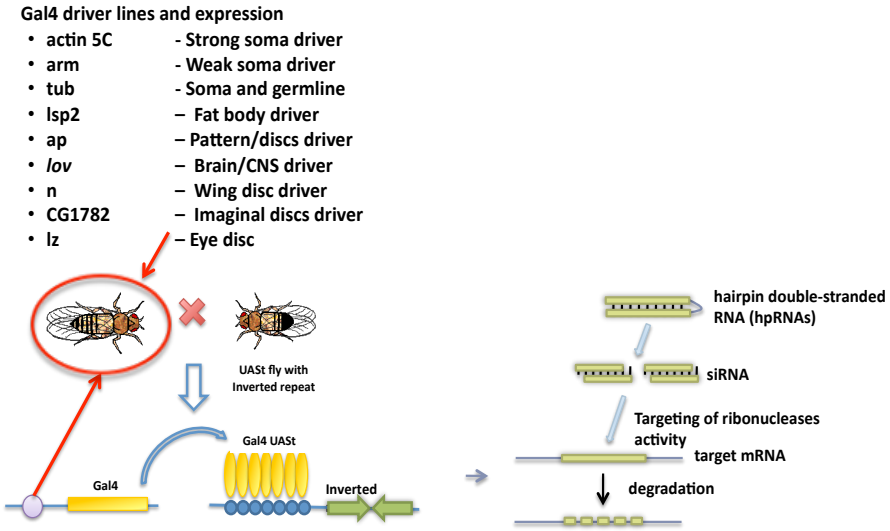


Figure 5.2 – GAL4 lines used to drive RNAi and knockdown of *Dntf-2r*, *Ran-like* and their parental genes in somatic cells. Schematic representation of UASt-GAL4 system and the RNAi mechanism.

5.2.4. RNA extraction

The Qiagen RNeasy mini kit was used for RNA extraction from the top halves of 60 larvae and from 30 adult flies. The RNA extraction from the top halves of the larvae that lack gonads was done to ensure that genes under study are expressed in tissues other than testes.

5.2.5. Retrogenes and parental genes transcription detection by RT-PCR

RNA was first digested with DNase I enzyme to digest any contaminant genomic DNA. Reverse transcription was performed using oligo (dT) primers (Promega INFO) and

Superscript II reverse transcriptase (Promega, Madison, WI). From the cDNA obtained, PCR was performed using specific primers *Dntf2r* and *Ran-like* (Table 4.1).

5.2.6. Phenotypes studied

All crosses were first checked for progeny after 6 days. The vials with no larvae or pupae were observed under the dissecting scope for the presence of eggs or small larvae and then returned to the incubator. From the 6th day on, all crosses were checked daily for emerging flies. The progeny that emerged from each cross was observed under the scope and compared to control flies also raised at 29ºC. Wings, eyes, thorax, abdomen, bristles and hair development were carefully checked. Pictures of eggs, larvae, pupae and adult flies were taken using the dissecting scope in the Demuth laboratory (11389-218 VistaVision, VWR).

5.3. Results

5.3.1. Transcription patterns of parental genes and their retrogenes in somatic tissues

A set of whole-mount *in situ* hybridizations were performed in 3rd instar larvae from the w^{1118} and *D. melanogaster* Besançon strains to obtain details of the transcription of *Ran-like* and *Dntf-2r* in different tissues during larval development. Brain, prothoracic gland, salivary glands, gut, imaginal discs, and testes (See Chapter 4) were dissected from the larvae after staining (Figure 5.3).

As expected, both parental genes are moderately to highly transcribed in all tissues analyzed and nearly in all cell types (Kumar, et al. 2001; Figure 5.3). *Dntf-2* appears to be absent from the ventral nerve cord. *Dntf-2r* is weakly expressed in the

brain including the ventral nerve cord and *Ran-like* is not present in the brain in agreement with the modENCODE mRNA seq data (Figure 3.1; Figure 3.3). In the prothoracic gland, *Ran-like* shows weak expression while *Dntf-2r* is expressed at high levels. *Dntf-2r* and *Dntf-2* are present in what seems to be the apical membrane of the salivary glands and *Ran* and *Ran-like* is in the cell cytoplasm basolateral membrane. The expression of these genes is also detected in the imaginal discs of wings, eyes, antennae and legs (Figure 5.3; Figure 5.4).

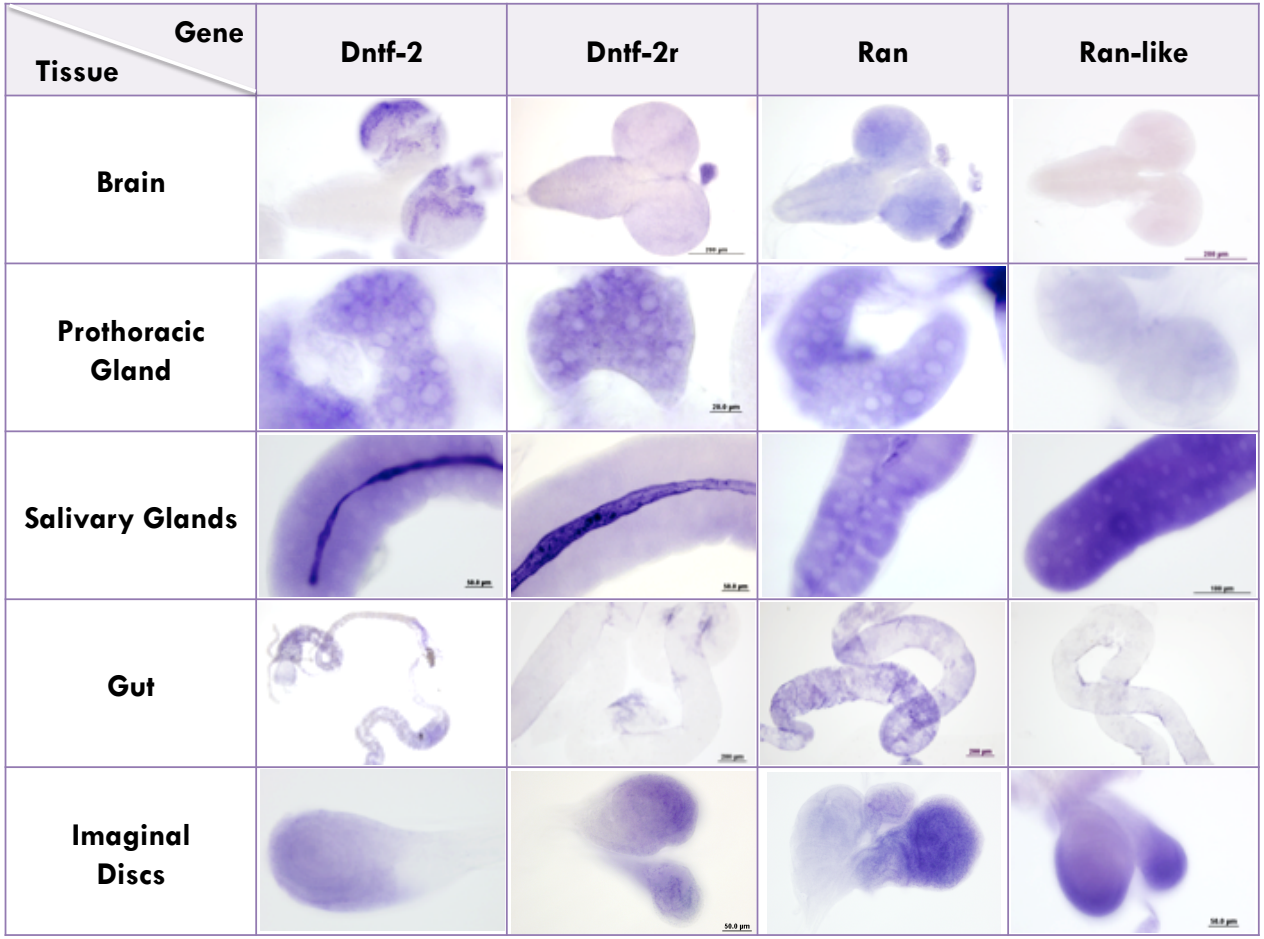


Figure 5.3 - *In situ* hybridization for *Ran*, *Ran-like*, *Dntf-2* transcript A, *Dntf-2* transcript B, and *Dntf-2r* in larval wildtype tissues.

It takes about 10 hours of development for the leg and wing discs to become visible which corresponds to the mid embryonic stage. The development of the imaginal discs continues during larvae growth giving rise to the adult fly legs and wings respectively. Each specialized cluster of cells from the imaginal discs invaginates, grows inside the developing larva and differentiates into adult structure during metamorphosis (French and Daniels 1994).

Observation of the different imaginal discs under the microscope allowed me to detect the presence of *Dntf-2* transcripts in the wing, leg and eye discs. Similarly, *Dntf-2r* is expressed in the leg discs and also in the haltere disc, wing and genital discs. In the wing disc it is possible to observe a stronger expression in its ventral and posterior part (Figure 5.4).

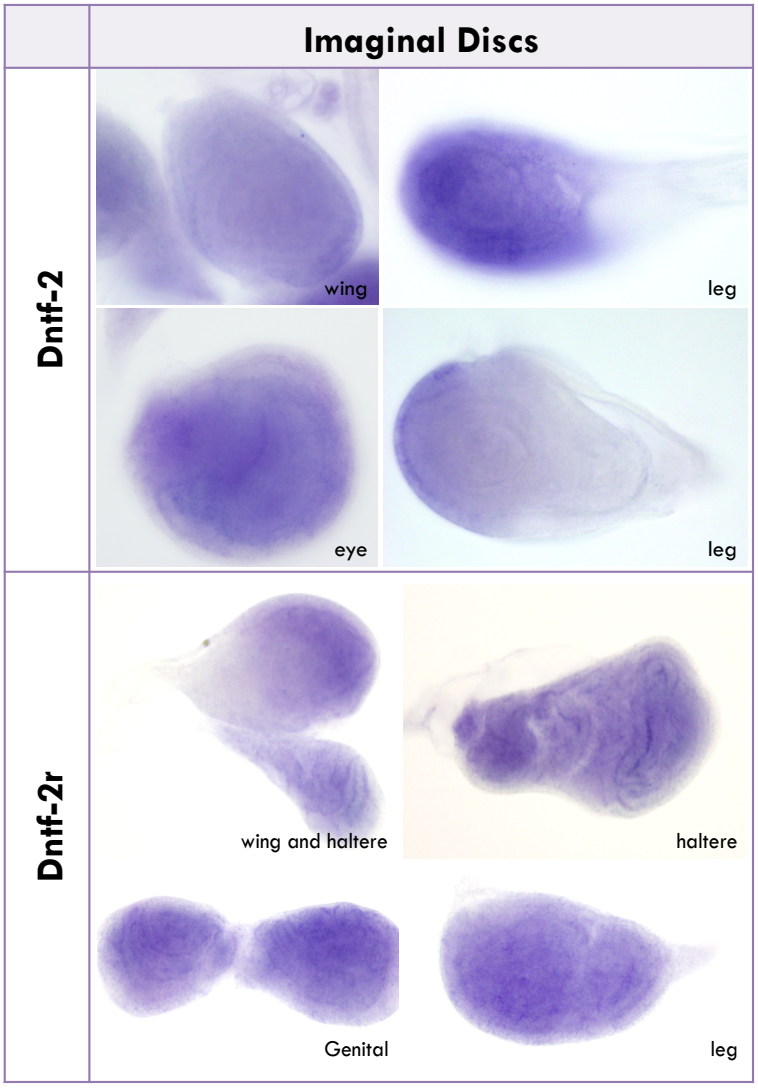


Figure 5.4 – *Dntf-2* and *Dntf-2r* transcription patterns detected with *in situ* hybridization in wildtype imaginal dics.

Ran and *Ran-like* transcription pattern in imaginal discs was also observed. *Ran* was detected in the eye and in the different leg discs. Ran shows almost no expression in the wing disc (Figure 5.3), *Ran-like* shows a strong expression in the ventral part of the wing disc. *Ran-like* is also expressed in the leg discs and in the haltere disc (Figure 5.5).

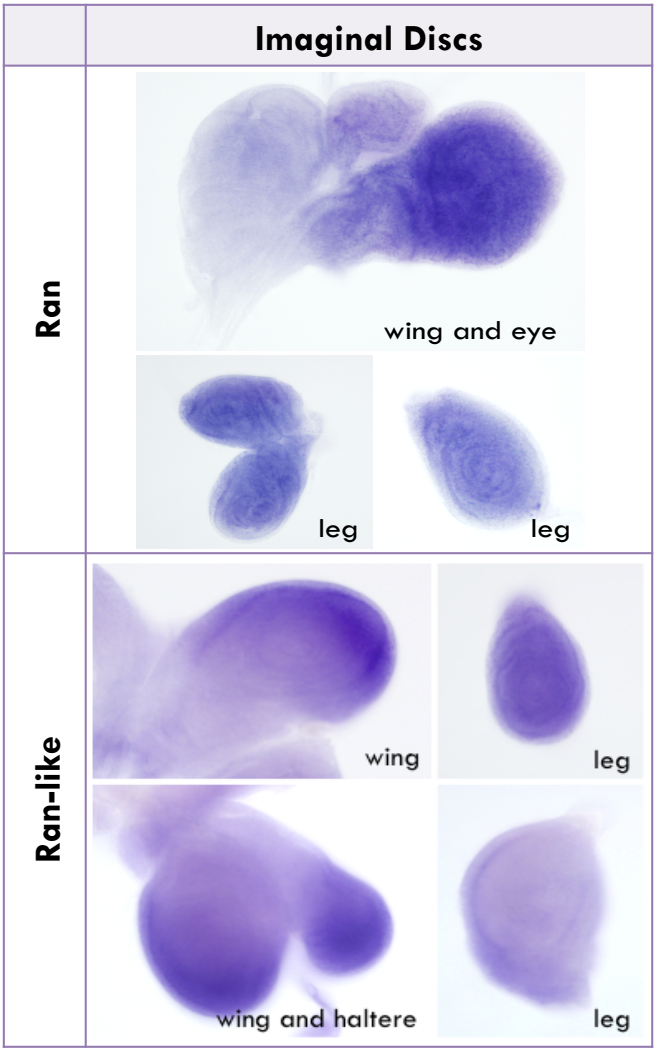


Figure 5.5 – *Ran* and *Ran-like* transcription patterns detected with *in situ* hybridization in wildtype imaginal dics.

Below I describe expression patters of *Dntf-2r* and *Ran-like* in the soma using the fusion/tagged proteins (see Chapter 3) and show the consequences of knocking down *Dntf-2r* and *Ran-like* in imaginal discs. Results for those experiments are consistent with the *in situs* presented above.

5.3.2. Dntf-2r-EGFP *expression and dissection of* Dntf-2r *somatic regulatory regions*

As shown in Chapter 3, Sorourian et al. (2014) concluded that a 27bp region including the β2-UE1-like element of *Dntf-2r* is needed to drive the expression of the *Dntf-2r*-EGFP fusion gene in testis. Here I show that the 151bp upstream region of *Dntf-2r* drives expression in the imaginal discs and the brain (Figure 5.8). The fluorescence in these tissues disappears in transformed flies that carry only the longest region (i.e., up to the next gene 5' of *Dntf-2r*) but not in the flies that carry only 101bp upstream.

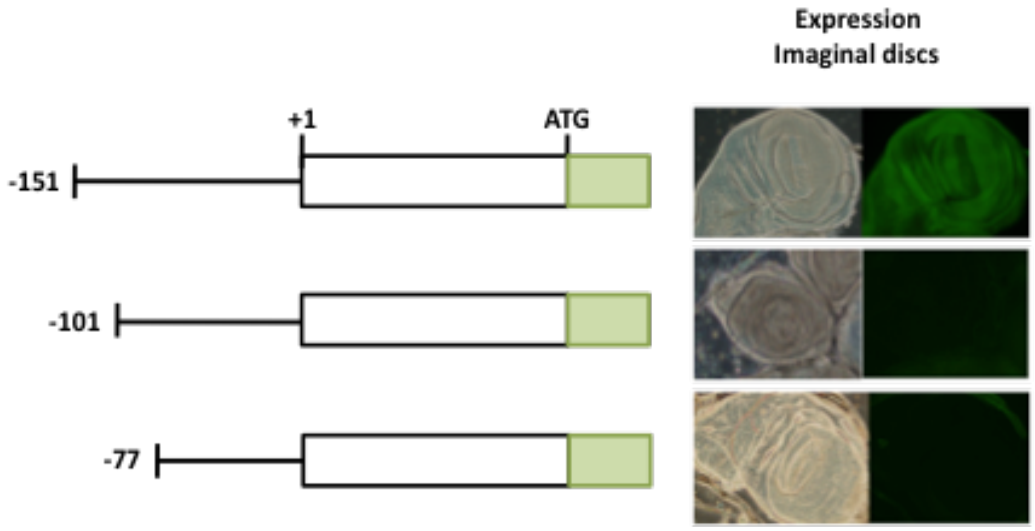


Figure 5.6 – Dntf-2r-EGFP fluorescence tag protein showing expression in imaginal discs (wing disc showed in the figure) in flies transformed with a construct containing different lengths of the upstream region. Image for 151bp from line 1.53.2.2.

To confirm Dntf-2r expression in imaginal discs, I dissected imaginal discs from different individual insertion lines transformed with the long construct containing *Dntf-2r*-EGFP fusion and observed fluorescence under the microscope (Table 5.2). A total of five constructs out of 8 constructs showed expression in larval imaginal discs.

Table 5.2 - Presence or absence of fluorescence in imaginal discs or larval testis dissected from flies transformed with 151bp construct. Images for these results are provided in Appendix B.

Dntf-2r Construct (line number)	Chromosome insertion	Fluorescence Imaginal discs	Fluorescence Larva Testis
1.53.2.2	II	Yes	Yes
1.53.5.1	III	No	Yes
1.53.6.2	X	Yes (Weak)	Yes
1.53.8.1	IV	No	No
1.53.8.3	II	No	Yes
1.53.44.3	X	Yes (weak)	Yes
1.53.45.1	II	No	Yes
1.53.76.1	II	Yes	Yes
1.53.76.8	III HL	Yes (Weak)	Yes
Total Constructs with fluorescence		5	8

Figure 5.8 shows that the fluorescence of Dntf-2r-*EGFP* in the wing disc follows anterior/posterior pattern. This result supports that Dntf-2r has a function during metamorphosis and the formation of the anterior posterior part of the adult fly. Fluorescence is also observed in leg discs, haltere and genitalia discs.

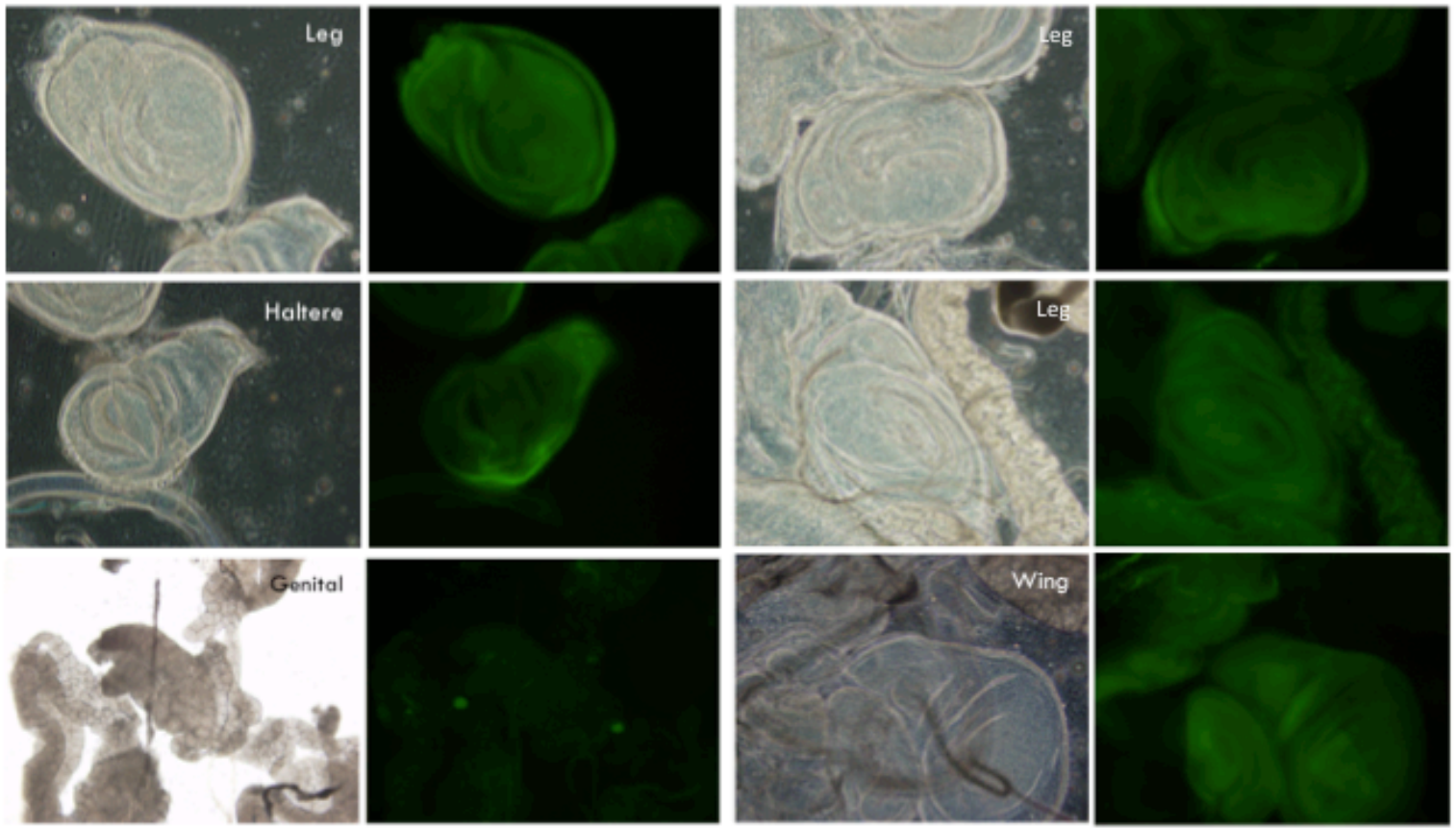


Figure 5.8 –Leg, haltere, genitalia and wing discs showing green fluorescence of Dntf-2r-*EGFP*. Line 1.53.76.1 containing the longest upstream region (151bp) was used. See Appendix B for imaginal discs in other lines.

The last larval tissue to observe under the florescence microscope was the brain where *Dntf-2r*-EGFP shows strong green fluorescence for lines containing 151bp upstream region (Figure 5.9). These results mimic results obtained with the *in situ* hybridization studies.

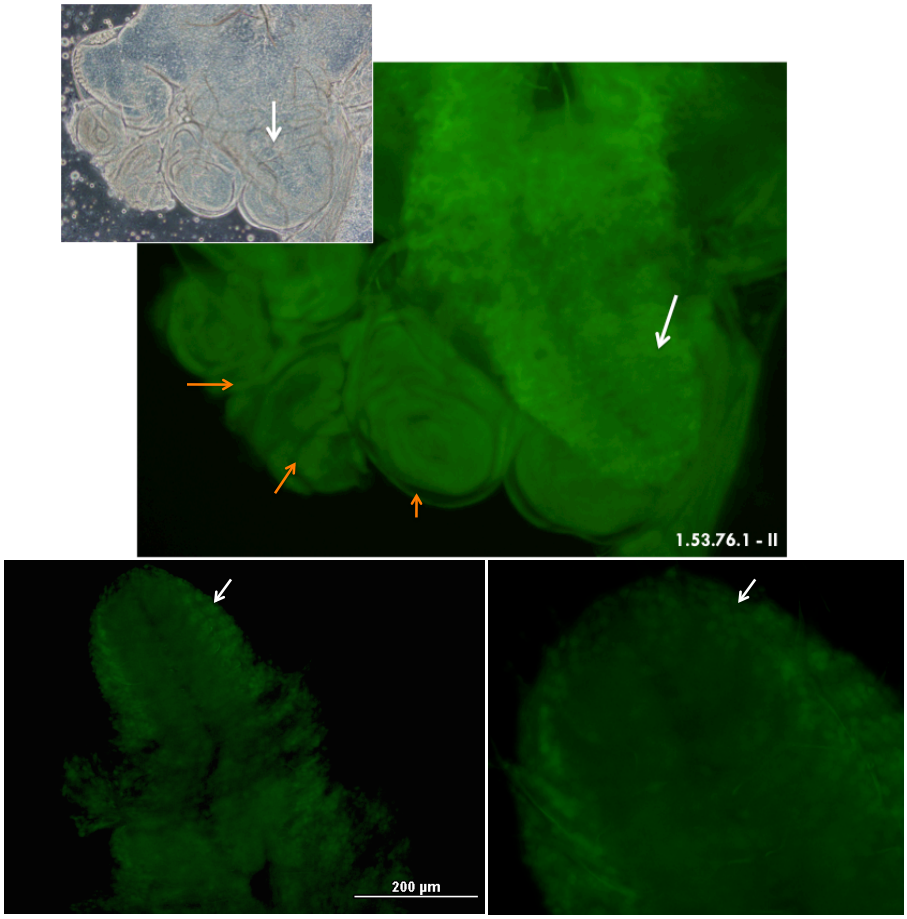


Figure 5.9 – Green fluorescence in the brain cells in the middle showing higher expression than in imaginal disc of Dntf-2r-*EGFP*. Line 1.53.76.1 containing the longest upstream region (151bp) was used. Phase contrast image in the left corner. White arrow indicates the brain area. Fluorescence appears to be highest in the nucleus of the cells but more detail is required to be certain of cellular localizations. Orange arrows indicate the different imaginal discs.

From the analysis of all the transformed lines available for *Dntf-2r*-EGFP fusion constructs containing different sizes of the upstream regions I verified that only the longest constructs show fluorescence in the imaginal disc and brain (Figure 5.6) and conclude that the region between 151bp and 101bp upstream of *Dntf-2r* must contain the regulatory region needed for somatic expression of *Dntf-2r*.

I also tried to observe the cellular localization of the retrogene in detail in some of the somatic tissues. Just like before I can see in salivary glands that Dntf-2r is inside the nucleus and in the cytoplasm (Figure 5.9). For Ran-like-DsRed.T4 (Images not shown) I see it in the nuclear membrane. Since I observe the same expression pattern in most *Dntf-2r*-EGFP lines I ruled out the possibility of expression of the fusion due only to the insertion site of the construct.

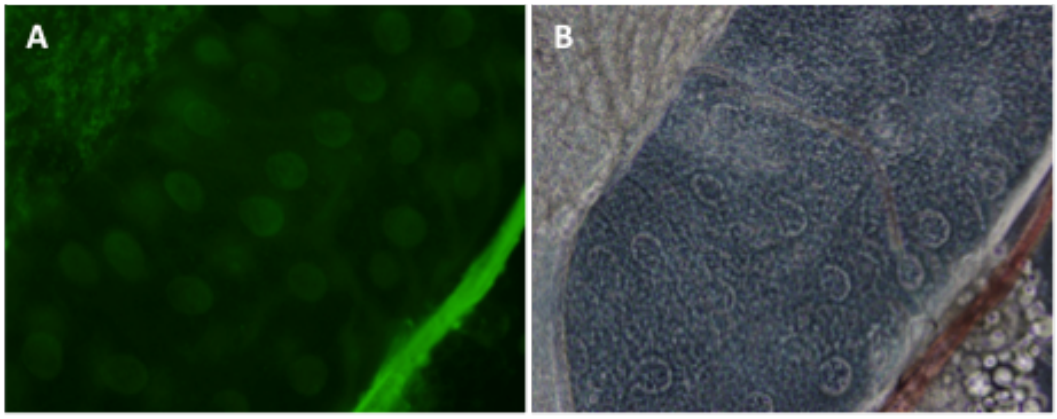


Figure 5.7 – Accessory gland cells showing Dntf-2r-EGFP inside de nucleus and weak expression in the cytoplasm (A) and phase contrast image (B).

5.3.3. Ran-like-RFP *expression and regulatory region in somatic tissues*

Tissues were dissected from larva and from adults, but no consistent fluorescence was observed in gut, ovaries and fat body for the Ran-like-RFP longest

insertion (i.e., 500bp) when compared to the control (w^{1118}). However, fluorescence was observed for the accessory glands. In the accessory glands red fluorescence can be observed in the nuclear membrane and in nucleus of the cell as well as in the membrane. Imaginal discs and brain were also dissected from wondering larva and no expression was detected in these tissues (Figure 5.10) in any of the lines transformed with the 500bp or 100bp regions. This reveals that only the accessory gland regulatory region in addition to the testes regulatory region is contained in our longest construct.

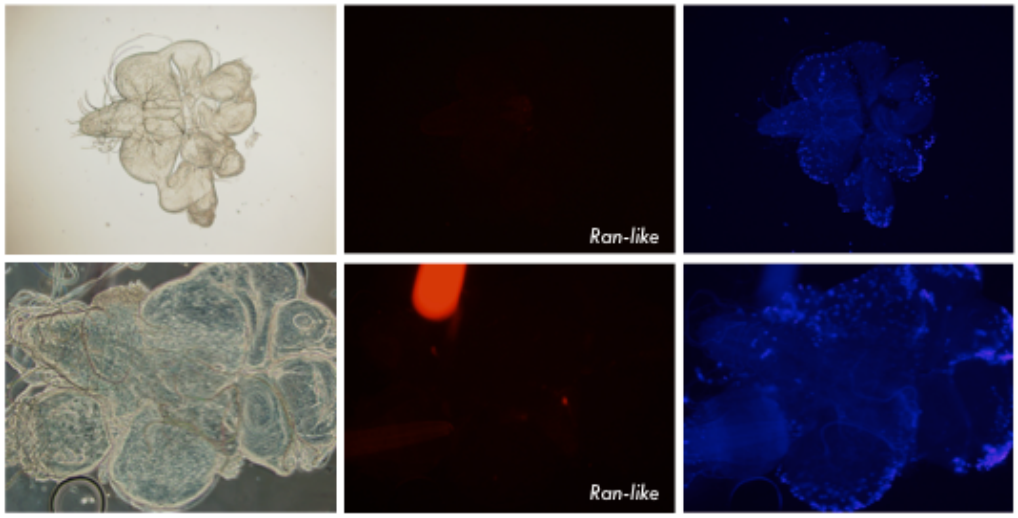


Figure 5.8 – Brain (top) and brain with imaginal discs (bottom) for line 7.24.2 with 500bp upstream region showing no red fluorescence. DNA is stained using DAPI (blue) and allows for visualization of cell nuclei.

5.3.3. RT-PCR confirming somatic transcription of Ran-like *and* Dntf-2r

Although I was able to detect the expression of both genes in imaginal disc with *in situ* hybridization and fluorescence protein tags, I also performed RT-PCR in wild type larvae for additional confirmation of my results. To make sure we separated the gonads, the larvae were observed under the microscope and cut in half. Cutting the larvae in half

using the dissecting scope insured us that no RNA from male gonads is present in the final RNA extraction. After the RNA was extracted and reverse transcribed, RT-PCRs were performed with two primer sets (primer set 1 and 2; Figure 5.11).

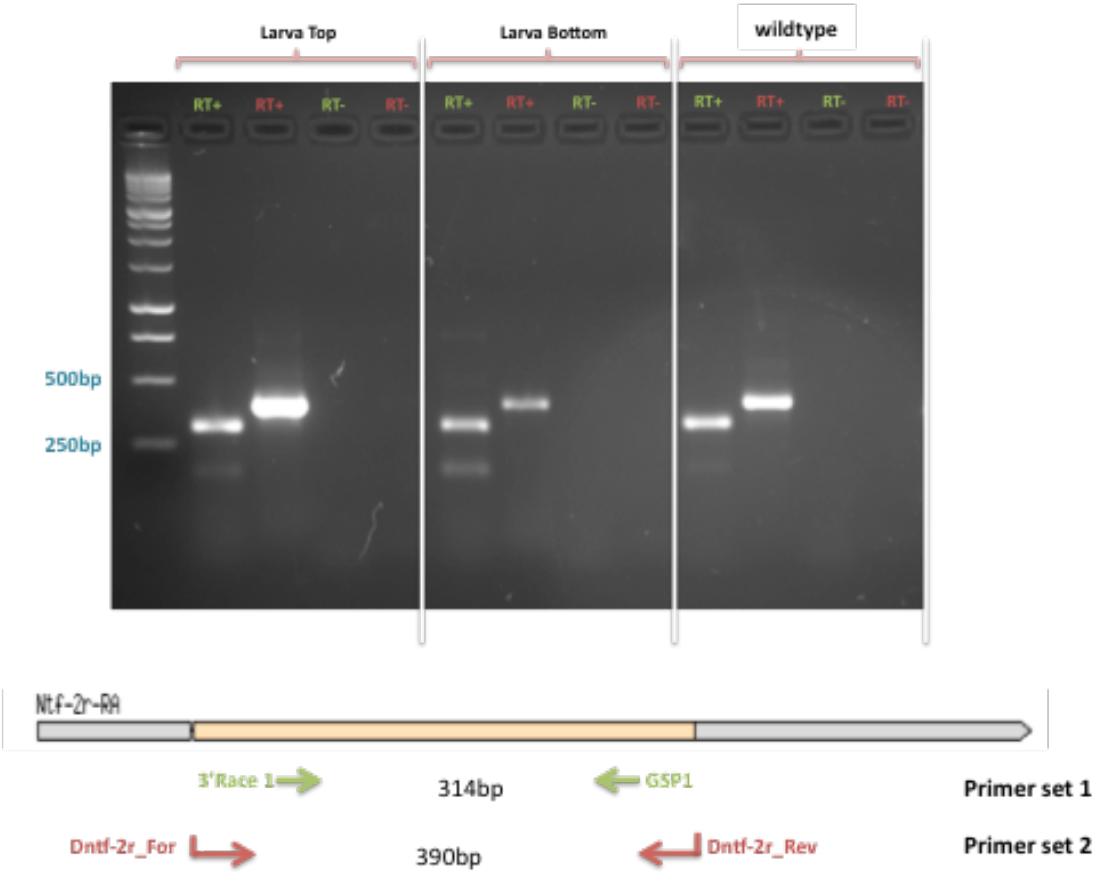


Figure 5.9 – RT-PCR for Dntf-2r in wild type larvae and in whole fly (*D. melanogaster* Besançon). The "larva top" includes larval head, imaginal discs, mouth and the top part of the gut. The "larva bottom" includes the gonads, genital disc and some fat body. Primer sets and expected PCR products lengths are also shown.

RT-PCR for *Dntf-2r* were positive for all larval tissues (Figure 5.12). The results presented so far demonstrate that the two retrogenes are expressed in tissues other than

testes, *Ran-like* image not shown. In order to decide if this expression is biologically relevant and of functional importance, I performed RNAi experiments to determine if knocking down retrogene expression in those tissues would result in mutant phenotypes.

5.3.4. Somatic tissues knockdowns

I carried out ubiquitous somatic and tissue-specific (brain, CNS and imaginal discs; i.e., wing discs and eye discs) knockdown crosses at 29ºC for *Dntf-2*, *Dntf-2r*, *Ran* and *Ran-like*. The progeny of all the crosses and replicates were checked for viability and possible mutant phenotypes under the dissecting scope. Crosses were performed with five young (1-2 days old) UASt males and five virgin females GAL4 (1-2 days old). For each cross 5 replicates were set up. The flies were kept at 29ºC (Duffy 2002). Additionally, UASt, GAL4 and w^{1118} flies were grown at 29ºC and used as controls. The summary of all results from the knockdown experiments is given in Table 5.3.

Table 5.3 – Table of the results of the knockdown crosses in embryos, larvae, pupae and adults. (+) Normal healthy progeny; (-) No progeny observed; (Few) reduced number of progeny; (Few death) Reduce number of this stage and deaths; (Eggs) Fertilized eggs but no pupae after; (Arrest) Developmental arrest; (Small) Small size compared to control.

UAS line	Gal4 line (Driver)	Tissue	Stages			
			Embryos	Larvae	Pupae	Adults
Dntf-2	Act	Soma (strong)	Eggs	-	-	-
	Arm	Soma (weak)	+	Few	Few death	-
	Tub	Soma/germline	Eggs	-	-	-
	Ap	*Apterous* pattern	+	+	+	-
	Lov	Brain/CNS	+	+	+	+
	N	Wing disc	+	+	+	Few
	1782	Imaginal discs	-	-	-	+
	Lz	Eye disc	+	+	+	+
	Lsp2	Fat Body	+	+	+	+
Dntf-2r	Act	Soma (strong)	+	Arrest	-	-
	Arm	Soma (weak)	+	+	+	+
	Tub	Soma/germline	+	Arrest	-	-
	Ap	*Apterous* pattern	+	+	+	Few
	Lov	Brain/CNS	Eggs	-	-	-
	N	Wing disc	+	+	+	-
	1782	Imaginal discs	+	+	-	-
	Lz	Eye disc	+	+	+	+
	Lsp2	Fat Body	+	+	+	+
Ran	Act	Soma (strong)	+	Small	-	-
	Arm	Soma (weak)	+	+	+	+
	Tub	Soma/germline	+	-	-	-
	Ap	*Apterous* pattern	Eggs	-	-	-
	Lov	Brain/CNS	+	+	Few	Few
	N	Wing disc	+	+	+	+
	1782	Imaginal discs	+	+	+	+
	Lz	Eye disc	+	+	+	+
	Lsp2	Fat Body	+	+	+	+
Ran-like	Act	Soma (strong)	+	+	Few	Few
	Arm	Soma (weak)	+	+	+	+
	Tub	Soma/germline	+	+	Few	Few
	Ap	*Apterous* pattern	+	+	-	-
	Lov	Brain/CNS	+	+	+	+
	N	Wing disc	+	+	+	+
	1782	Imaginal discs	+	+	+	+
	Lz	Eye disc	+	+	+	+
	Lsp2	Fat Body	+	+	+	+

Dntf-2 knockdowns with *Act5C*-, *tubP*- and *arm*-GAL4 drivers that are expressed ubiquitously resulted in developmental arrest during embryo and larvae stages. *Dntf-2r* knockdowns also show developmental arrest with *Act5C*- and *tubP*-GAL4 drivers. In this case, the arrest occurs later during larval development and in some cases in pre-pupa stage (Figure 5.12). In the case of *Ran*, I observed some flies emerging when using *arm*-GAL4, which is a weak somatic driver. However, no progeny was seen when the *Ran* was crossed with strong somatic drivers, *Act5C*- and *tubP*-GAL4. Ubiquitous knockdown of *Ran-like* results in reduced number of progeny. I expected 50% of the progeny from the crosses between *Act5C*-GAL4, if there is complete lethality, because this stock carries a balancer chromosome. However, 6% of the individuals had no balancer meaning that some driven individuals survived.

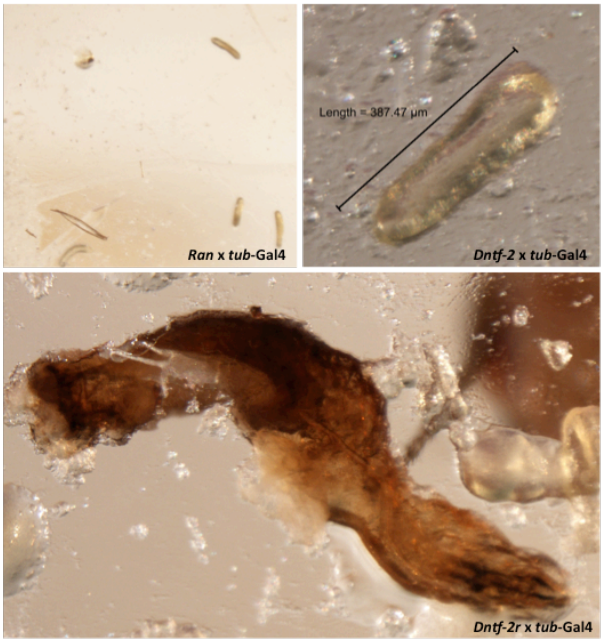


Figure 5.10 – Lethality in the tub-GAL4 driver crosses occurs early in larval development in both parental genes and later on for *Dntf-2r*.

The observed lethality and the reduction in the number of progeny in the knockdown crosses of *Dntf-2r* and *Ran-like*, respectively, is an unexpected result. This is because these retrogenes were known to have strong testis-biased pattern of expression (Betrán and Long 2003; Tracy, et al. 2010) and were expected to function in male germline only. This is, *Act5C*- and *tubP*-GAL4 drivers were not expected to alter levels of retrogene expression in somatic tissues because they only drive in somatic tissues. The observed phenotypes are only expected after the confirmation presented above that reveals that these retrogenes are expressed in somatic tissues quite specifically. It should be noted that UASt, GAL4 and w^{1118} lines were also kept at 29ºC as experimental control to guarantee that the phenotypes observed were not due to food problems or due to the increased temperature. In all controls, as expected, I was able to observe all stages of normal development from an egg to an adult fly (Figure 5.13).

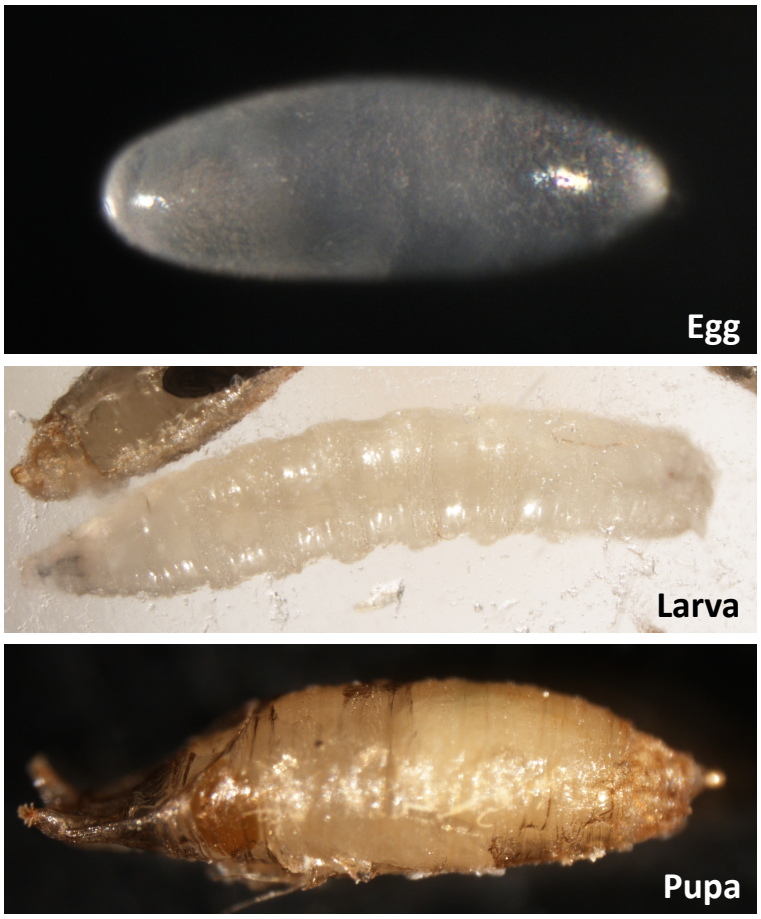


Figure 5.11 – w^{1118} control flies grown at 29ºC and not showing any problems at the egg, larva and pupa stages.

5.3.4.1. Tissue specific knockdowns of Dntf-2

As explained above and can been seen in Figure 5.14 no adult progeny was observed for the *Dntf-2* using any of the ubiquitous somatic drivers. When using *ap*-GAL4 driver, driver for *apterous* gene pattern in embryo and adults, I observed the lack of the abdominal area of the pupae where the mid-gut should be found. Additionally, I observed absence of spiracles during molt and morphogenesis. This phenotype has been previously observed when knocking down genes involved in control of steroid hormones

(Rewitz, et al. 2010; Ou, et al. 2011). Given that *Dntf-2* is expressed in every cell and from early development to the adult stages (see Chapter 3) these results are consistent with that.

Figure 5.12 – *Dntf-2* knockdown flies showing larval arrest (tub-GAL4) and pupal arrest with Ap-GAL4 driver. Missing abdominal area (white arrows).

5.3.4.2. Tissue-specific knockdowns of Dntf-2r

When using the *arm*-GAL4, the weak ubiquitous somatic driver, to knockdown *Dntf-2r*, arrest was observed in larvae and no pupae or adult suggesting that viability is

compromised. *Dntf-2r* also differed from all the other lines by not showing progeny when this gene is knockdown in brain and CNS (Table 5.3). When using *ap*-GAL4 driver, a strong driver that starts in embryo in *apterous* gene pattern, I observed the lack of the abdominal area of the pupae where the mid-gut should be found. Additionally, the anterior spiracles are also absent (Figure 5.15). The same phenotype observed for the parental gene. The wing disc driver (*N-GAL4*) crossed with UAS-*Dntf-2r* showed that suppressing *Dntf-2r* at this step results in pupae arrest with the half of the pupae body being absent (Table 5.3; Figure 5.15).

The results obtained for *Dntf-2r* using individual tissues knockdowns are not consistent with the results observed for the *arm*-GAL4. They phenotypes are more extreme than for the ubiquitous driver. However, it is known that when driving RNAi in whole fly, the effect of the driver might be not as strong as when driving RNAi in a specific tissue (White-Cooper 2012). Even among somatic drivers some have been shown to drive stronger than others (Ni, et al. 2009; Perrimon, et al. 2010; White-Cooper 2012). Given that *Dntf-2r* is expressed almost in all discs as well as the brain, the tissue-specific knockdown results described here are consitent with the gene expression.

Figure 5.15 – *Dntf-2r* knockdown flies showing larvae arrest (*act*- and *tub*-GAL4) and pupae arrest with N-GAL4 driver. Missing bottom area (white arrows) and anterior spiracles (orange arrows). In the control (green arrows) I can see the anterior spiracles and the pupae full body. Ap-GAL4 knockdowns resulted sometimes in pupae arrest and in other cases adult flies showed asymmetric hair and smaller bristles comparing to control line (blue arrows).

5.3.4.3. Tissue-specific knockdowns of Ran

Similarly to the results with *Dntf-2*, no adult progeny was obtained for *Ran* knockdown with *Ap*-GAL4 (Figure 5.16). However, the arrest in this case occurred earlier and many embryos did not develop. As was mentioned above knockdown of *Ran* with ubiquitous arm-GAL4 show no lethality effects. This is surprising as *Ran* knockout is lethal (Cesario and McKim 2011). *Ap*-GAL4 seems to be more efficient than arm-GAL4 supporting the idea that tissue-speficic knockdown is often more efficient than the ubiquitous knockdown. *Ran* is highly expressed along the entire life of the fly and virtually in every cell. I expected that even small reductions of the *Ran* transcript in any tissue would cause a phenotype, but this was not the case.

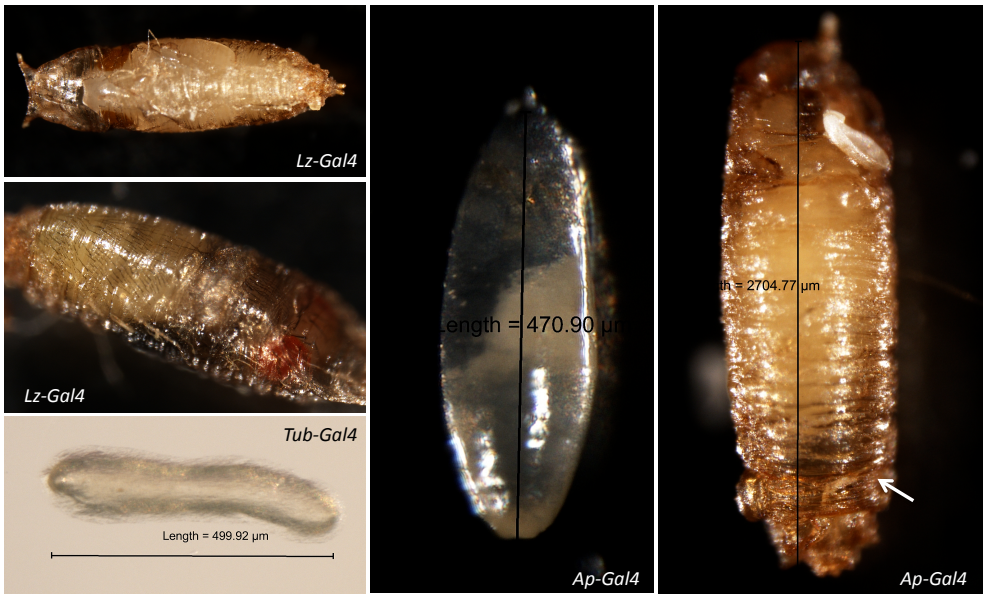


Figure 5.14 – Ap-GAL4 crosses with *Ran* result in pupae and egg arrest. Early larvae arrest was observed in *tub*-GAL4. Lz-GAL4 driven flies did not show any mutant phenotype.

5.3.4.4. Tissue-specific knockdowns of Ran-like

Ran-like and *Ran* knockdowns presented quite different outcomes showing that the results cannot be attributed to off target effects. In the *Ran-like* knockdown crosses using ubiquitous drivers I observed a drastic lowering in progeny number for *Act5C*- and *tubP*-GAL4 drivers. When using *arm*-GAL4 the number of F1 flies is higher than for the strong drivers, but still less than the progeny of the controls (Table 5.3). Similar to the results of the cross between *Dntf-2r* and *Ap*-GAL4 (Figure 5.17) adult flies in the cross between *Ran-like* and *Ap*-GAL4 showed asymmetric hair and smaller bristles (Figure 5.17). The imaginal discs are necessary to generate different structures in the adult fly after metamorphosis and since *Ran-like* is expressed in the imaginal discs I expected some phenotype related to these structures.

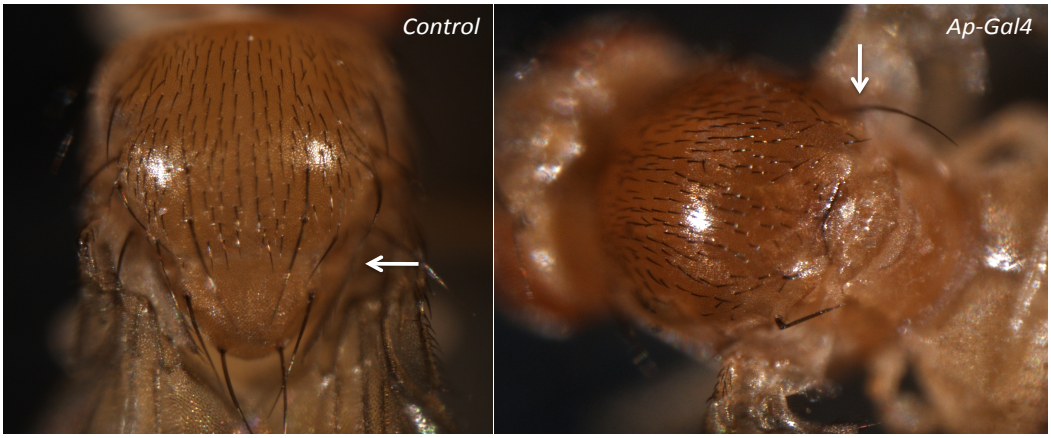


Figure 5.15 –*Ran-like* knockdown using Ap-GAL4 showing smaller bristles (white arrows) compared to the control line.

5.3.5. Somatic overexpression of Ran *and* Ran-like

In Chapter 4 I showed that overexpression of *Ran-like* with ubiquitous, somatic *Act5C*-, *arm*- and *tubP*-GAL4 drivers in order to rescue the parental knockout mutant results in the lethality of F1 progeny. Ectopic expression of a gene has been widely used in Drosophila to study the functions of a given gene in a specific tissue. Being one of the best know examples the overexpression of the *eyeless* gene (Halder, et al. 1995; Phelps and Brand 1998). I decided to investigate the consequences of overexpression of *Ran* and *Ran-like* in a particular tissue by using tissue specific drivers. *Ran* and *Ran-like* were both overexpressed and the F1 of the two overexpression crosses compared with each other and against the w^{1118} control line. Overexpressing *Ran* in different tissues had no effects in viability. In contrast to these results, overexpressing *Ran-like* causes progeny lethally in all crosses with the exception of the cross between *Ran-like* and 1782-GAL4, a driver expressed in imaginal disc. I also overexpressed *Ran-like* with *bam*- and *nos*-GAL4 (two-germline drivers), and obtained viable and fertile male progeny. This results show that overexpressing *Ran-like* in tissues where it is already expressed has no phenotypic effects while overexpression of *Ran-like* appears to be toxic when it is present in ectopic places. This is due to functional differences between *Ran* and *Ran-like* as mentioned in Chapter 4.

Figure 5.19 – Overexpressing *Ran* and *Ran-like* using Lz-GAL4 (antenna and eye disc driver) abolishes the development of the head in *Ran-like* pupae (white arrow) and of the anterior spiracles (orange arrow).

When I overexpress *Ran-like* using *Lz*-GAL4 (eye and antenna disc driver) I observe that the head regions as well as the anterior spiracles in the pupae are missing (Figure 5.19). The spiracles are a tree of tracheae that form the respiratory system of the larva. They are sense organs derived from the antenna progenitor cells in the eye-antennal disc. Initial disc development begins during embryogenesis when a set of cells acquires antennal identity (Sen, et al. 2010). *Ran-like* is not expressed in the eye-antennal disc according to *in situ* hybridization results, leading us to think that overexpressing this gene in this disc has a deleterious effect.

5.4. Discussion

Using *in situ* hybridization I have confirmed that *Dntf-2r* and *Ran-like*, the two young retrogenes, are not only expressed in male testes, but are also expressed in imaginal discs, salivary glands and in the prothoracic gland. Furthermore, *Dntf-2r* is also expressed in the brain. The fluorescence-tagged proteins also confirm these results. For *Dntf-2r*, I used the construct with the longest regulatory region available to us, and narrowed down the region required for imaginal disc epxression but since the fluorescence was not observed in all the tissues identified by in situ hybridizations except for accessory glands and testes, I think that I did not capture all the regulatory regions of the gene in the longest construct. In the case of Ran-like, the 500bp upstream region showed no fluorescence except for accessory glands and testes making us believe that a longer region may be required for expression of the gene in the rest of tissues.

The effects of the knockdowns were initially surprising because the genes were believed to be male specific but I show that the observed phenotypes are consistent with the now confirmed patterns of expression of these genes in somatic tissues. There seem to be no off target effects. First, I do not expect off target effect for these lines except for *Dntf-2* that might target *Dntf-2r* given the descriptions provided by Vienna Stock center. Second, I can see that the results in many cases are different for the parental gene and the retrogene. For example, when I use the ap-GAL4 driver for *Dntf-2* and *Dntf-2r*, I see that F1 progeny with a retrogene knockdown have short and disorganized bristles while there is no progeny at all in case of a parental gene knockdown. Third, as I mentioned above, the effects are consistent with the patterns of expression of the genes. Fourth, in the tissues where I have checked the knockdown of the particular gene using *in situ* (i.e., testes; Chapter 4) I did not reveal any off target effects.

Interestingly, knocking down *Dntf-2r* and *Ran-like* in different tissues using tissue-specific drivers has uncovered the multi-functional nature of these genes. It is now clear that these genes have functions in biological processes other than spermatogenesis. This is not completely unexpected if we think about how old these genes are. *Dntf-2r* is at least 6 My old (Figure 2.1) and *Ran-like* could be between 10 and 45 My old (Figure 2.1). If we consider Figure 2.4, it suggests that *Ran-like* could be present in *D. takahashii*, and *D. rhopaloa* and much older than 10 My old. Even if these genes entered the genome with the initial function in testes (Chapter 3), they could have gained additional functions through the acquisition of new expression patterns (i.e., new cis-regulatory regions). I could examine this possibility further by studying the evolution of imaginal disc expression of *Dntf-2r* given that I have narrowed down the region that drives expression in imaginal discs and I can compare this regions to the regions in close related species.

Traditionally, gene functions in Drosophila and other organisms have been studied through the analysis of knockdown effects and loss-of-function mutant phenotypes. However, these techniques have limitations since knockdowns are not always effective and not all genes have a loss-of-function phenotype (Brand and Perrimon 1993; White-Cooper 2012; Jonchere and Bennett 2013). I am now waiting to get the loss-of-function mutants for these genes. However, I now suspect that these mutants will not be viable and this means that the RNAi observations that I have at this point are the best way to help us dissect the functions of these genes. Once I get these mutants lines I will need to rescue them using the UAS-GAL4 system driving the UASt-gene fusions in tissues that cause lethality as revealed by our RNAi tissue knockdowns. This will further confirm that our in situ data is correct.

Chen, et al. (2010) studied 195 protein-coding genes that originated between 3 and 35 million years ago in Drosophila and investigated how quickly they become essential. Using RNAi techniques they show that 30% of these new genes are essential for viability (Chen, et al. 2010). They carried out their experiments at room temperature, but otherwise followed approaches analogous to the ones I followed here. However, they did not detect viability effects of *Dntf-2r* and *Ran-like* and I think this is because the efficiency of RNAi is lower at room temperature (Duffy 2002). I hope to get loss-of-function mutants for these genes to confirm the viability effects observed here.

By ectopically expressing *Ran-like* with different GAL4 somatic drivers, *Act5C-*, *arm-* and *tubP*-GAL4, I observed the lethal effect of overexpressing this gene in somatic tissues. This result led us to conclude that *Ran-like* is only required in certain tissues and for a specific function, for example in spermatogenesis and imaginal discs where the overexpression of these genes does not seem to produce adverse effects. Little information about gene functions can be gained from overexpression experiments. In this case I can only conclude that *Ran-like* does not have the exact same function as the parental *Ran* gene. When I overexpressed this retrogene with the eye and antenna driver in tissues where the gene is not normally expressed, I got lethality at the pupae stage, with flies missing the heads. Dissection of these flies showed that the organs above the head area were present and legs were also correctly formed. Halder, et al. (1995) were able to induce ectopic eyes in wings, legs, antennae, and halteres in Drosophila by targeted gene expression of the eyeless gene. However, these results were obtained using E132-GAL4 being expressed only in small patches. Whenever they tried to express the gene at high levels, the expression resulted in early death during pupae (Halder, et al. 1995) revealing that some overexpression experiments can be of little relevance to the understanding of the function of the ectopically expressed gene.

In my third objective, I wanted to understand the reasons why there is a recurrent duplication of nuclear transport retrogenes and, to answer this, I decided study the function of *Dntf-2r* and *Ran-like*. In this chapter, I revealed that *Dntf-2r* and *Ran-like* have important functions in somatic tissues. We think that those functions likely appear much after the duplication and were not the reason why those genes were retained. Younger genes, i.e., as we argued in the Discussion in Chapter 4, need to be studied to understand nuclear transport factor new gene retaintion.

CHAPTER 6

GENERAL DISCUSSION, CONCLUSIONS, AND FUTURE WORK

It has been observed that many new duplicated genes acquire testis functions not only in Drosophila but in other organisms including humans (Emerson, et al. 2004; Potrzebowski, et al. 2008). However, the specific functions of the new genes that ensured their preservation in the genome are not yet fully understood. In this dissertation, I described the study, in Drososphila, of genes in two gene families (*Dntf-2* and *Ran* gene families) that include recent nuclear transport duplicates highly expressed in testis and address three important aspects of their evolution:

1) The extent of recurrent duplications of nuclear transport genes in the Drosophila genus

2) The origin and molecular features of the regulatory regions that helped their retaintion

3) Their function and the degree of functional divergence from the parental genes of two of the duplicates (*Dntf-2r* and *Ran-like*) present in *D. melanogaster* and amenable to functional analyses.

Nuclear transport duplicates including *Dntf-2r* and *Ran-like* are duplicates of housekeeping genes (e.g., *Dntf-2* and *Ran*), and they have been often seen to evolve under positive selection in Drosophila. Why would housekeeping genes that are needed in every cell type duplicate and change function? Scientist have come to understand that actually, the recruitment of duplicated housekeeping genes to perform testis-specific functions is the norm rather than the exception, and we need to understand at least some

of the examples in detail in order to comprehend the selective pressures that lead to these duplications. In the case of *Dntf-2* and *Ran*, I also know that there have been recurrent duplications and losses (i.e., gene turnover) and would like to understand why.

The analyses I performed of the duplications of *Dntf-2* and *Ran* in 22 Drosophila genomes (Chapter 2) suggest that *Dntf-2* and *Ran* have been recurrently duplicated in more lineages than previously thought. I estimate that 6 or 7 duplications occurred for *Dntf-2* and only one of these is a DNA-mediated duplication (14-17%). If I count absences as losses (at least for now), many losses of these duplicates have occurred (Figure 2.2; Figure 2.3). In case of *Ran*, I estimate 10 duplication events and again, only two of these are DNA-mediated duplication (20%). Many losses of *Ran* duplicates have occurred as well (Figure 2.2; Figure 2.4). These results point to the importance of retroposition (as a mechanism of duplication) for the preservation of duplicates. Since retroposition produces a duplicate in a new genomic location, there is little opportunity for gene conversion with the parental gene. At the same time, relocation opens a possibility to evolve or recruit a regulatory region that will most likely differ from that of the parental gene (Gallach and Betrán 2011). Together, these consequences of retroposition make the evolution of new functions more likely for RNA-mediated than DNA-mediated duplicates. I suspect that many of the recurrent duplicates are either transcribed in testis or are male-biased. This is the case for the *Dntf-2* DNA duplicates in *D. pseudoobscura* and for previously described RNA-mediated duplicates in *D. ananasae* and *D. grimshawi* (Tracy et al. 2010). I see some old and some young duplications for both genes. For any old duplication described in this work, I see that the proteins are changing very fast. These are the results that reveal that the selective pressures leading to the duplication of these genes are strong and ubiquitous. It is also clear that although there is some gene turnover, some duplicates are maintained in the genomes for a long time.

It is relevant at this point to discuss the previously proposed out-of-the-testes hypothesis for the origin of new genes. The out-of-the-testes hypothesis states that new retrogenes emerge in testes and evolve other functions as they age. This is based on the observation in mammals and Drosophila that a higher proportion of young retrogenes is expressed in testis while older retrogenes are expressed in more tissues (Vinckenbosch, et al. 2006; Assis and Bachtrog 2013). Our observation is that there are many losses of fast-evolving retrogenes and this would point to a different explanation for the observed pattern. Many retrogenes emerge for testes functions, but the function they fulfill is short-lived because testis is a tissue that changes very fast in response to various selection pressures. That is, the new duplicates are functional for a short period of time and are lost when no longer required. This scenario would explain why previous work has seen less testis-specific duplicates in the old cohorts of retrogenes. Unlike genes that perform testes functions, retrogenes that emerge for other more stable functions might be longer lived explaining the observed pattern. Gene turnover and changes in expression patterns in many lineages must be studied in detail to distinguish these hypotheses or distinguish how many cases fit one or the opther of these hypotheses.

Dntf-2r and *Ran-like* are two duplicates that originated through retroposition, a mechanism that should produce dead-on-arrival gene copies because they lose all parental regulatory regions. Instead of this, it has been observed that these retrogenes and many others as described above often acquired testis expression. Exactly how this happens is a question that deserves careful consideration. Very interesting results have come out of our analyses of the regulatory regions of *Dntf-2r* and *Ran-like*. The testis regulatory regions are short and potentially easy to acquire after retroposition. Even more interestingly, in the case of *Dntf-2r*, the regulatory region was present in the genome before the insertion of retrogene, likely driving the testis expression of another gene, a

non-coding RNA (Sorourian, et al. 2014). So, I propose that retrotransposition in the male germline to regions of open chromatin helps retrogenes acquired testis expression directly. Previous work in our lab pointed to this, as an excess of retrogenes in testis neighborhoods was found (Bai, et al. 2007). Work in humans also suggests this (Vinckenbosch, et al. 2006). *Dntf-2r* analysis provides a detailed example of that. The fact that the *β2-tubulin* testis-specific upstream regulatory region can be flipped and still drive testis expression points to an additional way in which retroposition can help acquire testis expression directly: if an existing regulatory region can drive the expression in both directions. This mechanism was proposed for a paternal effect retrogene because the retrogene (ms(3)K81) was inserted in head-to-head orientation with a testis-specific gene (Loppin, et al. 2005). Both the upstream region of ms(3)K81 and the β2-UE1 motif should be tested to study if they drive testis expression bidirectionally.

Interestingly, as mentioned in Chapter 1, in instances of intralocus sexual antagonism driven by testis, a duplication of the male-beneficial allele, accompanied by the establishment of testis-specific expression, is expected to resolve this conflict (Gallach and Betrán 2011; Gallach, et al. 2011). If testis-specific expression already occurs upon insertion of the retrogene as inferred for Dntf-2r, rapid fixation of the duplicate is expected under positive selection suporting the model in that all the steps of this model of intralocus sexual conflict resolution by gene duplication might occur under srtong positive selection and there will be no waiting time to establish the testis-specific expression.

It has recently been shown that intronless genes in general are enriched for testis transcription and their transcription and export migth be tightly coupled and specialized in testis (Caporilli, et al. 2013). An mRNA export protein, Ntx1, is needed to export genes

transcribed by the testis specific trancription complex and mutations of this gene have particular effects on intronless genes in male germline (Caporilli, et al. 2013). So this can be an additional gene regulatory feature of testis that favors retrogenes and migth contribute to explain the excess of RNA-mediated duplicates retained for testis function.

To elucidate the reason for the recurrent duplication of *Dntf-2* and *Ran*, I have taken a functional approach that is described in Chapters 4 and 5. I have gained remarkable understanding of the function of *Dntf-2r* and *Ran-like* in spermatogenesis. Localization studies show that in spermatogenesis, both genes appear to have many functions from spermatocytes onwards related to nuclear transport, nuclear membrane assembly, microtubule organization during chromosomal segregation, and sperm head and tail elongation. As a consequence, the deficit of *Dntf-2r* or *Ran-like* strongly affects male fertility and is manifested cytologically as abnormalities at different stages of spermatogenesis. I observe loss of nucleus and smaller nucleus expected if non-disjunction and/or failure to assemble the nuclear membrane are occurring and smaller number of sperm heads in the bundles as a consequence. I observe dot-shaped sperm heads and disorganized sperm heads as a consequence of the faulty sperm head elongation. This included a lack of the reshaping of the mitochondria and asynchronous elongation, culminating in the lack of sperm in seminal vesicles. Both genes appear to have kept most functions of the parental genes (nuclear transport, nuclear membrane assembly, and microtubule organization) but appeared to have undergone specialization. This is clear in the case of Ran-like. Ran-like can partially rescue Ran loss-of-function mutant but causes lethality when it is overexpressed in the soma.

Even though the initial aims were to understand the function of these genes in spermatogenesis, a whole new array of possibilities opened with the confirmation of *Dntf-*

2r and *Ran-like* expression in other tissues (Chapter 5). I have now gained some understanding about the function of *Dntf-2r* and *Ran-like* in the soma. Both genes are expressed in imaginal discs, salivary glands, accessory glands and prothoracic gland. *Dntf-2r* is expressed much higher than *Ran* in these tissues and this might be the reason why RNAi effects for *Dntf-2r* are stronger. In addition, *Dntf-2r* is expressed in the brain and much higher than its parental gene in imaginal discs, unlike *Ran-like*. So, these migth be the reasons why knockdown of *Dntf-2r* in the soma causes complete lethality, while *Ran-like* knockdown results in viability problems. These phenotypes were not observed by Chen, et al. (2010) that performed a series of knockdown experiments for young duplicates in Drosophila. Since I drove RNAi at 29ºC (optimal temperature for efficient RNAi), I think that I am being able to knockdown these genes while they were not because the RNAi was driven at 25 ºC. I also infer that the P-element insertion line and the excision line that introduced and alternative protein end are hypomorphs because they do not produce the lethality that RNAi of *Dntf-2r* produces in soma.

Our functional analyses point to a currently multifaceted function of *Dntf-2r* and *Ran-like*. Since the genes are expressed very highly during spermatogenesis, I still think that the duplication occurred under the testes selective pressures. However, even in testes, *Dntf-2r* and *Ran-like* have multiple functions and any of these could explain why these genes duplicated and in every instance the selective pressure could be different. For example, function in sperm tail elongation could have led to gene duplications and could have evolved as a response to male-male competition (Immler, et al. 2011). Initially, as I discussed in Chapter 1 and Chapter 4, it was proposed that nuclear transport genes could be fixed in the populations if they acted as suppressors of meiotic drive systems. It is believed that the *D. melanogaster* SD prevents the maturation of sperm that carries long DNA satellites likely by slowing nuclear transport (import and export of small

RNA precursors and small RNAs; Gell and Reenan 2013). If the function of *Dntf-2r* and *Ran-like* was to suppress meiotic drive systems similar to SD, I would expect these genes to be involved in transport during meiosis but not in that many other functions. So, if *Dntf-2r* and *Ran-like* were duplicated to suppress one of these conflicts, I infer that they have gained additional functions later. To show if *Ran* and *Dntf-2* duplicates originate as suppressors of selfish meiotic drive systems younger duplicates should be studied to demonstrate it.

There are some experiments that I need to undertake immediately to prepare this work for publication. I need to generate *Dntf-2r* and *Ran-like* mutants. I need them to confirm that the loss-of-function of *Dntf-2r* or *Ran-like* affects viability. These mutants are being produced by Genetic Services, Inc. However, they have been trying now for almost a year, but have not been successful so far. I proposed a region to be targeted in both genes using the CRISPR/Cas9 system (Bassett, et al. 2003). The company has been injecting a plasmid that should express the guide RNA that Cas9 uses to digest the DNA creating a deletion in germline. They are now in the process of trying another approach. The RNA will be directly injected to guide Cas9 DNA digestion in germline. The fusion proteins to make sure these proteins have the wild type function should then rescue the mutants. I think they do but this will be the definitive prove. If this doesn't work in the near future I will go forward and make RNAi resistant constructs to inject and try to rescue the RNAi phenotype observed for the different genes (Jonchere and Bennett 2013). These are constructs that can drive a modified coding region that is different enough not to be targeted by the RNAi that targets the wildtype transcript. For the UASt-Dntf-2r-EGFP construct I am missing I will use a different technique to try to produce this construct by using a normal P-element construct instead of the gateway system.

It has been known for a long time that testes and spermatogenesis differ among Drosophila species in major ways. As an example, a study of 1990 compared the spermatogenesis of *D. melanogaster* and *D. virilis* (Klasterska and Ramel 1990). The two species differ in the morphology and coloration of the testes as well as in the details of meiosis and sperm differentiation. During meiosis a large portion of the DNA appears as a dense structure during prophase in *D. virilis*, but this does not happen in *D. melanogaster*. The nucleolus also behaves differently during meiosis. It appears to be more active in *D. virilis*. In addition, centrioles are not clearly discernable in *D. virilis*. Visible differences in spermiogenesis include the presence of a dot where chromatin is condensed in sperm heads during elongation, different length and thickness of sperm head and tail, and different levels of organization of sperm bundles. I am likely observing two new genes that contribute to the molecular bases of some of these differences. These differences are the product of strong and diverse selective pressures and illustrate rapid evolution of this tissue that is achieved to a large extent through gene duplication.

In summary, I have made major contributions to the field that I hope will materialize in two additional publications:

1. I have shown that *Dntf-2* and *Ran* have duplicated many more times and with a huge bias towards retroduplication that needs to be explained and I discuss several reasons for this. I provide very young events that can be the ones that need to be studied in the future. I think this chapter can be published as a Brief Communication to MBE.

2. I have also studied the regulatory regions of *Dntf-2r* and *Ran-like*. In particular, my work revealed that the ancestral region where *Dntf-2r* already contained a testis-specific regulatory region making this the first example of this. This has been published in MBE.

3. I have studied in detail the expression and thge many functions of *Dntf-2r* and *Ran-like* in spermatogenesis and revealed that they also have important functions in soma. This reveals that understanding why those genes duplicated cannot be answered by analysing their current function and younger genes need to be studied but enourmously contributes to understanding the current function of these genes. The functional work will be published in a journal like Genes and Development.

APPENDIX A

Parental and duplicates protein sequences

>Dntf-2r_mel
MSLNLQYEDIGKEFVQQYYAIFDDPANRENVINFYNATDSFMTFEGNQIQGAPKILEKVQSLSF
QKIARVITTVDSQPTSDGGVLIIVLGRLKCDDDPPHAFSQIFLLKPNGGSLFVAHDIFRLNIHNSA

>Dntf-2r_sim
MSLNPQYEEIGKGFVQQYYAIFDDPVNRENVVHFYSATDSFMTFEGRQIQGAPKILEKVQSLSF
QKISIVITTVDSQPTFDGGVLISVLGRLKCDDDPPHSFSQIFLLKPNGGSFLVAHDIFRLNIHNSA

>Dntf-2r_sec
MSLNPQYEEIGKGFVQQYYAILDDLANRENAVNFYSVTDSFMTFEGHQIQGAPKILEKVQSLRF
QKISIVITTVDSQPTFDGGVLIFVLGRLKCDDDPPHSFSQIFLLKPNGGSFLVAHDIFRLNIHNSA

>Dntf-2retro_fic
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAYSQIFVLKANAGTFFVAHDIFRLNIHNSA

>Dntf-2retro_bia
VKMNPQFEAIAKSFVEEYYSSFDDQTNHKNLSKFYSETDSLMTLDGYQLKGTKKILEKLKSVDF
KKIERQISSVDSQPTPDGGVLIQVIGSFKWEGTLPLNFSQVFLLKTSGNKFYVGNDIFRIILPNKA

>Dntf-2retro_rho
IKMNQNFESIGKTFVGQYYTRFDDPAKQNTLEELYSDTESLMTFDGHQILGGKKILEKIKSLQFK
KINREISSVDSQPSLDGAVMVHVVGRLQCDKNPSIAFSQVFLLKPSKNTFYVVHDIFRIV

>Dntf-2retro1_ana
MELNRDFEEITSLFVDQYYTLFDDPEKREELCNCYNSSSSLLSFQGEQIRGPKISEKLKNLPVQ
KINRIIRSVDSQPTCDGGVLIYVHGSLQCEEEVPVNFSQIILLHNGEQGIFIAHDIFRTEVI

>Dntf-2retro_bip
DFEEISSSFIDQYYGLFDDPDKREELSNFYNSSSSLLSFQGEQIPGPRISDKFYSLPNQRIKRIVR
SMDAQPVGDDRILIHVHGVLQCEEDLPQNFSQVFLLQSGEHGYFIAHDIFRTQVI

>Dntf-2retro2_ana
MPLNPHYEPMGQEFVKQYYVIFDNPATRALTATFFSHNDSFMTFEGEQVLGYYKIFEKVKSLPF
QKVNRTLTNVDCQPTGDGGILMSVLGRLQCDDDPSLSFSQIFVLKPDTSPNAYYLSHDIFRLNI
HDTE

>Dntf-2DNAdup_pse_pse
MSLNLQFENIANSFVQEYYTLLDSPENRTRVAHFYKAKESLMTVEGLRLEGASQILETIQNLSFK
KIHHMITVVDAQPTIDGGVLICVMGRLKIDDGSPFSFSQVFVLKAVGNSFFVENEIFRLSELKSP

>Dntf-2DNAdup_per
MDPFLNQPDTDYEDVEEWGEFVEVLPHNLLMPLKTGVLIDFVQEYYTLLDNPENRTRVAHFYK
AKESLMTVEGLRLEGASQILETIQNLSFKKIHHIITVVDAQPTIDGGVLICVMGRLKIDDGPPFAFS
QVFVLKAVGNSFFVENEIFRLSELKSP

>Dntf-2retro_gri
MAINPQYEAVGKGFVQQYYAIFDDPANRANVVNFYSTTDSFMTFEGHQIQGAPKILEKVQSLSF
QKINRIITTIDSQPTFDGGVLINVLGRLQCDEDPPHSYSQVFLLKANAGSFFVAHDIFRLNIHNSA

>Dntf-2PA_mel
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFFLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_sim

MSLNPQYEEIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFFLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_sec
MSLNPQYEEIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGATKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFVLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_yak
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAYSQVFVLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_ere
MSLNPQYEDIGKGFVQQFYGIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKINRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFFLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_fic
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLTF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFVLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_eug
MSLNPQYEDIGKGFVQQYYAIFDDPANRANKFLIFQATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFLLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_bia
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYQATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFVLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_tak
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYQATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFVLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_ele
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKINRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFLLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_rho
MSLNPQYEEIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFMLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_kik
MSLNLQYEEIGKGFVQQYYSIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKINRVITTVDSQPTFDGGVIINVLGRLQCDDDPPHAFSQLFVLKANAGTFFVAHDIFRLNIHNSA

>Dntf-2PA_ana
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDEDPPHAFSQVFVLKANAGTFFVAHDIFRLNIHNS
A

>Dntf-2PA_pse_pse
MSLNPQYEDIGKGFVQQYYALFDDPANRASVVNFYSATESFMTFEGHQIQGAPKILEKIQSLSF
QKITRVITAVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFMLKANANSFFVAHDIFRLNIHNT
A

>Dntf-2PA_per
MSLNPQYEDIGKGFVQQYYALFDDPANRASVVNFYSATESFMTFEGHQIQGAPKILEKVQSLS
FQKITRVITAVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFMLKANANSFFVAHDIFRLNIHN
TA

>Dntf-2PA_mir
MSLNPQYEDIGKGFVQQYYALFDDPANRASVVNFYSATESFMTFEGHQIQGAPKILEKVQSLS
FQKITRVITAVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFMLKANANSFFVAHDIFRLNIHN
TA

>Dntf-2PA_wil
MSLNPQYEDIGKGFVQQYYSIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKISRVITTVDSQPTFDGGVLINVLGRLQCDEDPPHAYSQVFVLKANAGTFFVAHDIFRLNIHNS

>Dntf-2PA_moj
MALNPQYEDIGKGFVQQYYGIFDDPANRANVVNFYSATESFMTFEGHQIQGAPKILEKVQSLTF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFVLKANAGTYFVAHDIFRLNIHNS
A

>Dntf-2PA_vir
MALNPQYEDIGKGFVQQYYSIFDDPANRANVVNFYSATESFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFVLKANAGTYFVAHDIFRLNIHNS
A

>Dntf-2PA_alb
MALNPQYEDIGKGFVQQYYAIFDDPENRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQCDDDPPHAFSQVFVLKTNAGTYFVAHDIFRLNIHNS
A

>Dntf-2PA_gri
MALNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKISRVITTVDSQPTFDGGVLINVLGRVQCDDDPPLAFSQVFVLKANAGTYYVAHDIFRLNIHNS
A

>Dntf-2PB_mel
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_sim
MSLNPQYEEIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_sec
MSLNPQYEEIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGATKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_yak
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_ere
MSLNPQYEDIGKGFVQQFYGIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKINRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_fic
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLTF
QKITRVITTVDSQPTFDGGVLINVLGRLQADDDPPLAFNQTFLLKPGGCGFYVERDIFRLSLHNV

>Dntf-2PB_eug
MSLNPQYEDIGKGFVQQYYAIFDDPANRANKFLIFQATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_bia
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYQATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_tak
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYQATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_ele
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKINRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_rho
MSLNPQYEEIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSYFVQHDIFRLSLHDV

>Dntf-2PB_kik
MSLNLQYEEIGKGFVQQYYSIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKINRVITTVDSQPTFDGGVIINVLGRLQTDDDQPHAYIQTFVLKSVSGTFFVQHDIFRLSLHDV

>Dntf-2PB_ana
MSLNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_pse_pse
MSLNPQYEDIGKGFVQQYYALFDDPANRASVVNFYSATESFMTFEGHQIQGAPKILEKIQSLSF
QKITRVITAVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHDV

>Dntf-2PB_per
MSLNPQYEDIGKGFVQQYYALFDDPANRASVVNFYSATESFMTFEGHQIQGAPKILEKVQSLS
FQKITRVITAVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHD
V

>Dntf-2PB_mir
MSLNPQYEDIGKGFVQQYYALFDDPANRASVVNFYSATESFMTFEGHQIQGAPKILEKVQSLS
FQKITRVITAVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGGSFFVQHDIFRLSLHD
V
>Dntf-2PB_wil
MSLNPQYEDIGKGFVQQYYSIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKISRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVSGSFFVEHDIFRLSLHDV

>Dntf-2PB_moj
MALNPQYEDIGKGFVQQYYGIFDDPANRANVVNFYSATESFMTFEGHQIQGAPKILEKVQSLTF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGVSFFVQHDIFRLALHNV

>Dntf-2PB_vir
MALNPQYEDIGKGFVQQYYSIFDDPANRANVVNFYSATESFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDDDQPHAYIQTFVLKPVGVSFFVQHDIFRLALHDV

>Dntf-2PB_alb

MALNPQYEDIGKGFVQQYYAIFDDPENRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKITRVITTVDSQPTFDGGVLINVLGRLQTDEDQPHAYIQTFVLKPVGVSFFVQHDIFRLALHDV

>Dntf-2PB_gri
MALNPQYEDIGKGFVQQYYAIFDDPANRANVVNFYSATDSFMTFEGHQIQGAPKILEKVQSLSF
QKISRVITTVDSQPTFDGGVLINVLGRVQTDEDQPHAYIQTFVLKPVGISFFVQHDIFRLALHDV

>Ran-like_mel
MQPQEEVKAIFKLILIGDGGTGKTTLVKRHLTGEFKMQYNATLGVEVEQLLFNTNRGVFRIDVW
DTAGQERYGGLRDGYFVQSQCAIIMFDVASSNTYNNVKRWHRDLVRVCGNIPIVICGNKVDIM
HKKTWKKGVDFDRKTNIYLIEMSAKSNYNVEKPFVYLLRKLVGDPSLQLVQSPAIQPPKVVFTD
EMSRQVESLFNEAKSKPLPTIYDIDL

>Ran-like_sim
MQSQEEVKATFKLILIGDGESGKTTFVKRHLTGEFNVQHNATLGVEVNHLLFHTNRGVFRFDV
WDIAGHEMFDGLRDGYFIRSQCAIIMFDTAKANTYNNVNRWHRDLVGVCGDIPIVICGNKVDIM
PKKSWKTCINFDRKSILYHIEMSARTNYNIEKPFVYLLKKLVGDPSLKLVQSPAIQPPKVVFTDE
MSCQVERFLDEATSYTLPPTYDFDL

>Ran-like_sec
MQSQEEVKATFKLILIGDGESGKTTFVKRHLTGEFNVQHNATLGVEVNHLLFHTNRGVFRFDV
WDTAGHEMFDGLRDGYFIRSQCAIIMFDTSKANIYNNVNRWHRDLVGVCGDIPIVICGNKLDIM
PKKSWKTCINFDHKSILYHIEMSAKTNYNVEKPFVYLLKKLVGDPSLKLVQSPAIQPPKVVFTDE
MSRQVERLLDEATSYRLPPIYDFDL

>Ran-like_ere
MQPQEKEVASFKLLLLGDGATGKTTFVKRHLTGEFEKRYIASPGAMPHSLLFHTSRGCYRFIV
WDIAGQETLDGLREGYYTEGQCAIIMFDVNSRTTYKNVTSWHRDLVRVCGNIPIVICGNKVDIE
HRKDWPKWFDFHSKNYIEMSAKSNYNFERPFVYLLRKLVDDPNLKLVQSPALKPPQVVFTDEI
RRQVESDLMEASSYPLPTSVDDDS

>Ran_Retro_sim
MIEYKLVILGDAQVGKTSLMNRVLYEMFEEEYKPTIGVDVNSFTIQTNRGFIRFIVWDTAGQEEL
CGLRDGYYINAQCAIIMVDEASHRTYRAWHLDLVRVCKNIPVVICGNKSENQKERMSALQVTF
GRMNLVYYGISVKFKMNLEMPFMYLARKLTGDRSLTLVRPPEPNFLEELKETPPEEELEREDL

>Ran_Retro_sec
MIEYKLVILGDAQVGKTSLMNRVLYEMFEEEYKPTIGVDVNSFTIQTNRGFIRFIVWDTAGQEEL
CGLRDGYYINAQCAIIMVDEASHRTYRAWHLDLVRVCKNIPVVICGNKSENQKERTSALQVTFG
RMNLVYYGISVKFKMNLEMPFMYLARKLTGDRSLTLVRPPEHNFLEVLEETPPEEELEREDL

>Ran_Retro_yak
MNRYKLVLLGDAQVGKTSLMSRLLYGMYQERYEPTDGVQLTSFAIQTDHGNIRFKVWDTAGQ
AELCGLRDGYYIGAQCAIIMVDVTSPGSYQSWHRDLVRVCNKIPVVICGNKNEPMAERTERNL
QASFGRMNLAYYGISVKCKTNLEKPFLYLARRLTGDPKLELLKPPKPPSLITLSKETPSEEEHEG

>Ran_Retro_ere
MKGYKLVLLGDAQVGKTALINRLLHGRRLEKYEPTEGVDLTPVAIQTNHGSIRFIVWDTAGQQE
LGGLRDGYYIDAQCAIIMIDVTLPYSSYMNWHRDLVRVCSKIPVVICGNKNELRTERTERTGCTL
QAAFGRMNLVYYDISLKFKNNLEKPFLYLARQLTGNQKLKLAMLPKAPSLIRLTEEENEC

>Ran_Retro_1_fic

MAQEGQDMPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAMRFN
VWDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTHKNVPNWHRDLVRVCDNIPIVLCGNKV
DIKDRKVKAKNIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVK
MDKDWQVQIERDLQEAQATALPDEDEDL

>Ran_Retro_2_fic
MAQEGQDMPTFKCVLIGDGGTGKTTFVKRHKTGEFEKRYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCDYVPIVLCGNKVD
IKDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKM
DKDWQVQIERDLHEAQATALPDEDEDL

>Ran_Retro_eug
MKVFKLVLLGDAEVGKSSFIERLMLGTYQESYEATNGVELARLTFRTNRGNVGFIVWDTAGRE
ELGGLRDGYYIKADCAIIMFDVTSMKSNINLEKWHRDIVRACSQIPVVFCGNKVEMTRNRMVND
NSATFGRKNNLVYYDISAKLNYNMEKPFLYLSRQLLKDPKLEFVTNFEPPKL

>Ran_Retro_1_bia
MAQEAQDMPTFKCVLVGDGGTGKTTFVKRHMTGEFEKRYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGECALILFDVTSRVTYKNVPNWHRDLVRICETIPIVLCGNKVDVK
DRKVKAKNVVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLAFVAMPALLPPEVKMD
REWQAQIERDLQEAQATALPDEDEDL

>Ran_Retro_tak
MNSQNEVVAAFKLVLLGDGGTGKTTFVKRHLTGEYQKRYLATMGVEVHPLLFSTSRGMFRFD
VWDTAGQEKLGGLRDGYYVQGQCAIIMFDLASKITYKNVGNWYRDLLRVCGNIPIVLCGNKADI
RYRKIKAKQVTFHRKNNLHYIEMSAKSNYNYERPFVYLLRQLVGDPQLELTVSPALKPPEVSFT
EEMRLEAERNLLEASACPLPPNDDD

>Ran_Retro_ele
FKLVLLGEAEVGKSTFIERHLSGKFQWKYAPTVGVSLNRLTFRTDRGGIPFIVWDTAGQEKLGG
MRDGYYIQAHCAIIMFDVTSRKSYDKLEGWHRDLVRVCNNIPIVICGNKVDLKEKRQVDDTQVK
FHRKKNLMYFDISAKSGFNLDKPFIYLARQLLNDRKLRLVEQPAIPKMITSPELWSQSQDELME
AKA

>Ran_Retro_1_rho
MTQEGPDMPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDVYYIQGQCSVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIIFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQAQIERDLQEAQATALP

>Ran_Retro_2_rho
EVAAFKLILLGDGGTGKTTFVKRHLTGEFEKKYEATLGVEVHPIFFHTTRGLFRINIWDTAGQEK
FGGLSEGYCVKGQCAILMFDVTSRTTYNNVPFWYQDVERVCGNIPMVLCGNKVDVRDRKVRA
KQVTFHLKKNLPFIEISAKSNYNYEKPFVYLLRELTGDPQLELVRQPALKPPEVILTDEQRIQME
QELEAAKYLQFPSEKDYEL

>Ran_Retro_3_rho
FKLVLIGEAKVGKSTLMGCHLFGKFQKNYEPTLGVDVNPLTFKTDHGEIRFMVWDTAGQERLG
GMRDGYYIQAHCAIMMFDVTSRTSYDRLEKWHRDLVRVCGNIPIVICGNKADSLFEAETFVKFY
RQKQLMFFKISAKSNYNLDKPFLYLARQLLGDRKQEVAKHPSFPKLIMTAEQLSQLQEDLKAAR
ACDLPPESDD

>Ran_Retro_ana
MSGSGDGIPSFKCVLIGDGGTGKTTFVKRHSTGEFEKRYVATLGVEVRPLLFNTTRGSIRFNV
WDTAGQEKFGGLRDGYYIEGQCAIIMFDVTSRITYKNVPNWYRDLVRVCDSVPIVLCGNKVDIK

DRKVKAKAITFHRQKNLQYYDISAKSNYNFEKPFLWLARKLVGDPLLEFVVMPALLPPEIHMGP
ELLAQVERDLVEAKATALPDDDDDDL

>Ran_Retro_bip
MAQEDQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALMPPEVKM
DKDWQLQIERDLQEAQATALPDEDEDL

>Ran_Retro_pse
MFKDTFKLVLVGDGQTGKTSLIRRHLKGTFDEFYRPTVGHEVKDLEFQTKRGPVCFSVWDTAG
QDKYAGLRDIYYTGSHCAIIMFDVTSPSSYSRVAYWHREILATCGRIPIVLCGNKAEIIKDRKVFK
GHKVHHWEFQYFDISVKNNYCCEEPFLSLFYQLIGERFKYCAIVRPHLKMIQEVENEDKEDLDM
EIEHKDDLEMEIEEEAQEALPLVDF

>Ran_Retro_per
MFKDTFKLVLVGDGQTGKTSLIRRHLKGTFDEFYRPTVGHEVKDLEFQTKRGPVCFSVWDTAG
QDKYAGLRDIYYTGSHCAIIMFDVTSPSSYSRVAYWHREILATCGRIPIVLCGNKAEIIKDRKVFK
GHKVHHFQYFDISVKNNYCCEEPFLSLFYQLIGERFKYCAIVRPHLTMIQEVENEDKEDLEMEIE
HKDDLEMEIEEEAQQALPLVDF

>Ran_Retro_mir
MLTDTFKLVLVGDGQTGKTSLIKRHLKGTFDEFYRPTVGHDVHNLEFQTKRGPVCFSVWDTAG
QNKYAGLRDIYYTGSHCAIIMFDVTSPSSYSRVAYWHREILATCGRIPIVLCGNKVEIIKDRKVFK
SHKVRQWEFQYFDISVKNNYCCEEPFLSLYYQLIGERFKYCAIMRPHLKMIQEVENEDKEDLEI
EHKDDLEMEIEEEAQHSLPLVDF

>Ran_Retro_wil
MSDYLDMKCVLLGDGKTGKSTFIKRQLTGQYQRDYLPTDNVDLSTLKFHTNQGTISFKVWDTA
GQEELRGLRDIYYMQARCAIIMFDVTRPKTFLNVTEWRRELIRICPDIPIVLCGNKADFMLPLSIE
STFHLHNLPYYPMSVKTGYNYDKPFLWLARRVTDNPKLRFVKQPVNFSPQ

>Ran_DNAdup_moj
MAQGDAIPTFKLVLVGDGGTGKSTFVKRHMTGEFEKRYLATMGIEVHPLKFHTSRGPLCFNIW
DTAGQEKFGGLREGYYIQAQCALIFFDVTSRTTYKNVPHWYRDLIRICGHIPIVLCGNKIDIKNCK
IRPIKPRTSDFYSKKNMQYFPISAKSNRNFEKPFRWLARVLVGDPHLEFISMPALQPPEVHMSK
SCQLKMELELEQATDLPLDDDDDDL

>Ran_DNAdup_vir
MAYDCGDAIPSFKLVLVGDGGTGKTTFVKRHMTGEYQKRYLATMGVEVHPLKFHTSRGPVQF
HVWDTAGQEKFGGLREGYYIQAQCAIIFFDVTSRTTYKNVPNWYRDLMRTCGHIPIVLCGNKIDI
KDCKVKPKGLDFHRKKNIQYFPISAKSNYNFEKPFRWLARVLVGDPRLEFVEMPALQPPDVRM
AKSWQLKMEHELAQAGLLLEDDDENL

>Ran_Retro_gri
MAQDMPTFKCILVGDGGTGKTTFVKRHLSGEFEKKYVATLGVEVHPLVFHTNRGAIRFNVWDT
AGQEKFGGLRDGYYIQGQSAVIMFDVTSRVTYKNVPNWHRDLVRVCDNIPIVLCGNKVDIKDR
KVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDANLEFVAMPALLPPDVKMDTD
WQLQIERDLQEAQATALPDDDEDL

>Ran_mel
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQAQIERDLQEAQATALPDEDEEL

>Ran_sim

MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQAQIERDLQEAQATALPDEDEEL

>Ran_sec
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQAQIERDLQEAQATALPDEDEEL

>Ran_yak
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQAQIERDLQEAQATALPDEDEDL

>Ran_ere
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCSVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQAQIERDLQEAQATALPDEDEDL

>Ran_fic
MAQEGQDMPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQAQIERDLQEAQATALPDEDEDL

>Ran_bia
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQAQIERDLQEAQATALPDEDEDL

>Ran_tak
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQAQIERELQVAQATALPDEDEDL

>Ran_ele
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQAQIERDLQEAQATALPDEDEDL

>Ran_kik
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQLQIERDLQEAQATALPDEDEDL

>Ran_ana
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI

KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQLQIERDLQEAQATALPDEDEDL

>Ran_bip
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQLQIERDLQEAQATALPDEDEDL

>Ran_pse_pse
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVVMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVD
IKDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKM
DKDWQLQIERDLQEAQATALPDEDEDL

>Ran_per
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVVMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVD
IKDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKM
DKDWQLQIERDLQEAQATALPDEDEDL

>Ran_mir
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVVMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVD
IKDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKM
DKDWQLQIERDLQEAQATALPDEDEDL

>Ran_wil
MAQEGQDIPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQLQIERDLQEAQATALPDEDEDL

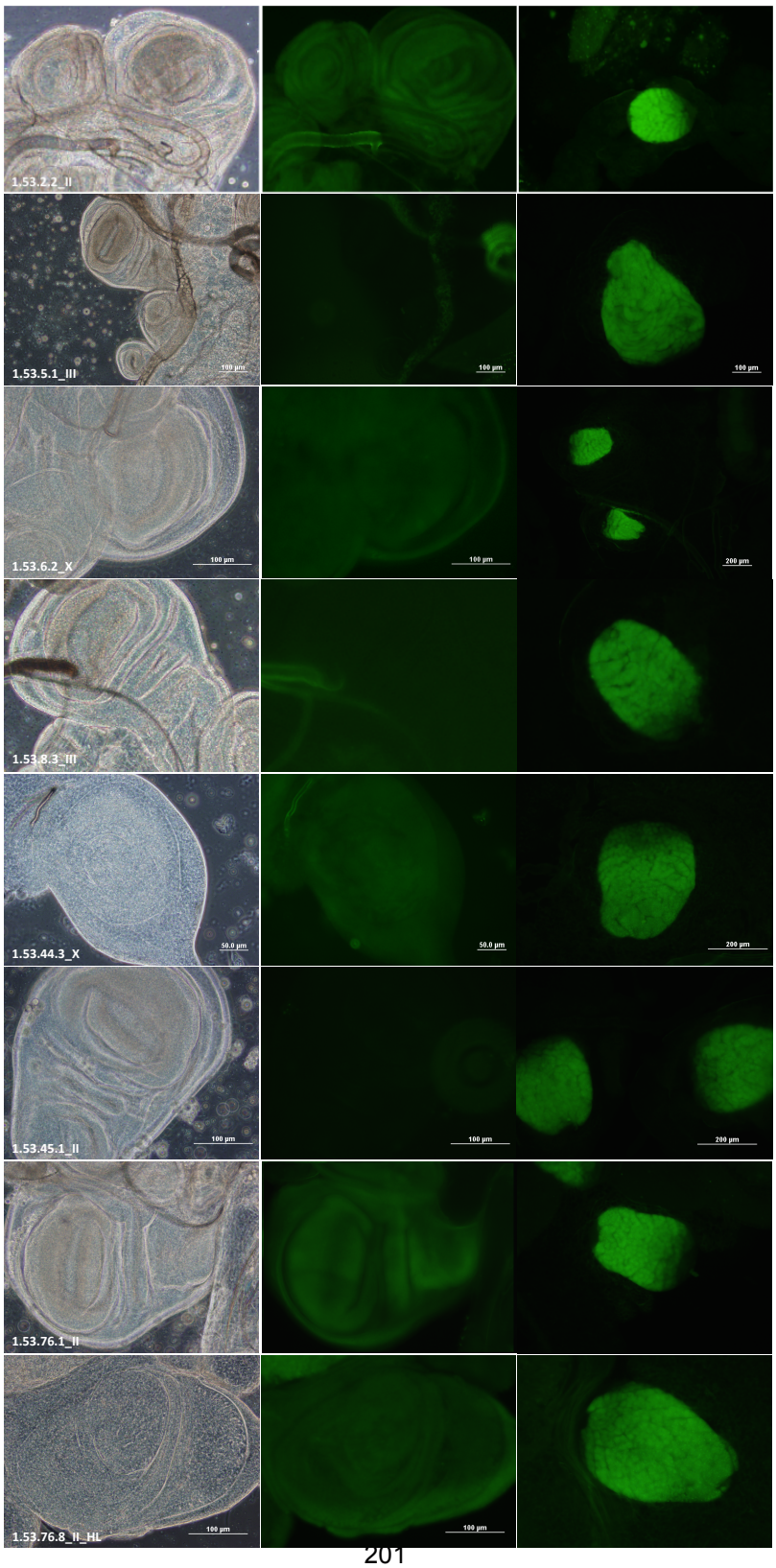
>Ran_moj
MAQEGQDMPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQLQIERDLQEAQATALPDEDEDL

>Ran_vir
MAQEVQDMPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKSIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
KDWQLQIERDLQEAQATALPDEDEDL

>Ran_gri
MAQEGQDMPTFKCVLVGDGGTGKTTFVKRHMTGEFEKKYVATLGVEVHPLIFHTNRGAIRFNV
WDTAGQEKFGGLRDGYYIQGQCAVIMFDVTSRVTYKNVPNWHRDLVRVCENIPIVLCGNKVDI
KDRKVKAKTIVFHRKKNLQYYDISAKSNYNFEKPFLWLARKLVGDPNLEFVAMPALLPPEVKMD
RDWQLQIERDLQEAQATALPDEDEDL

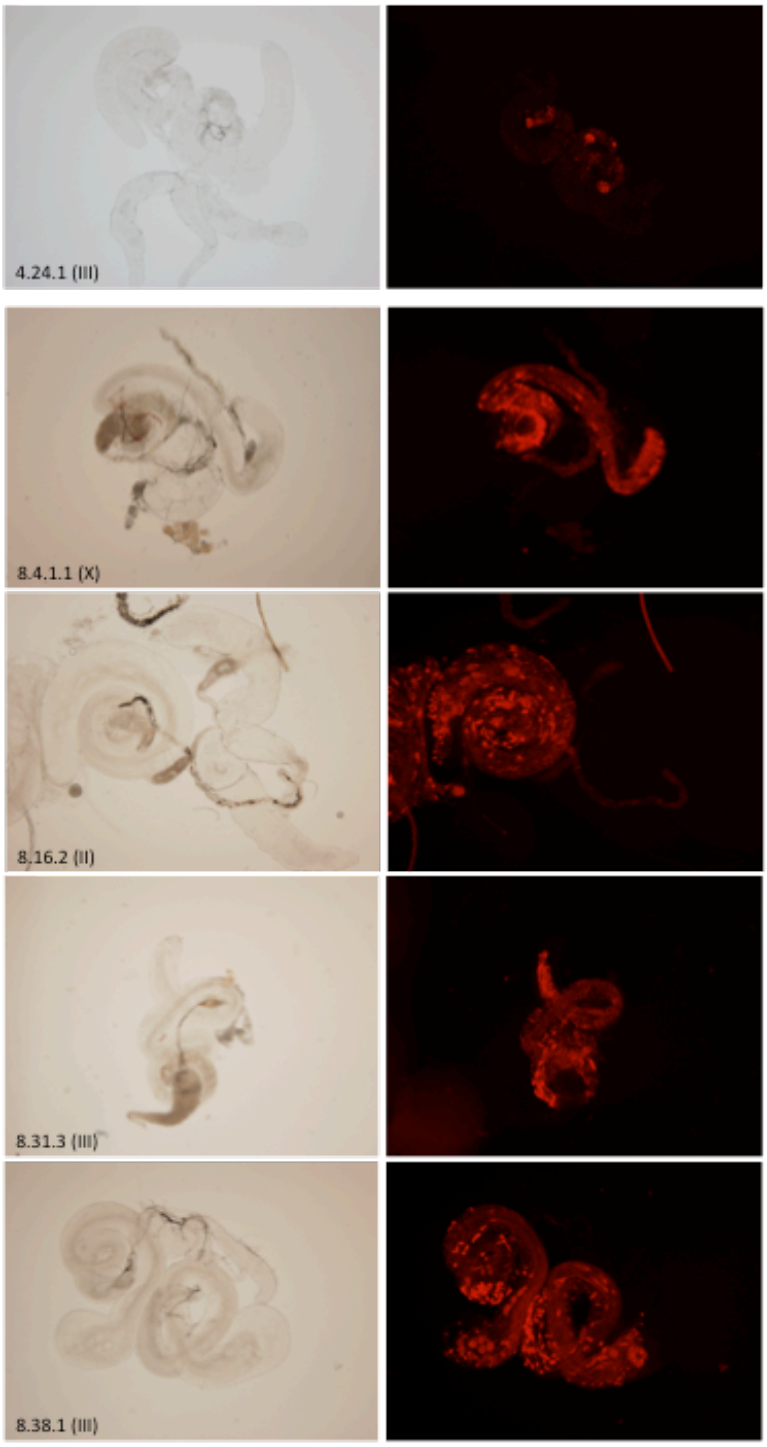
APPENDIX B

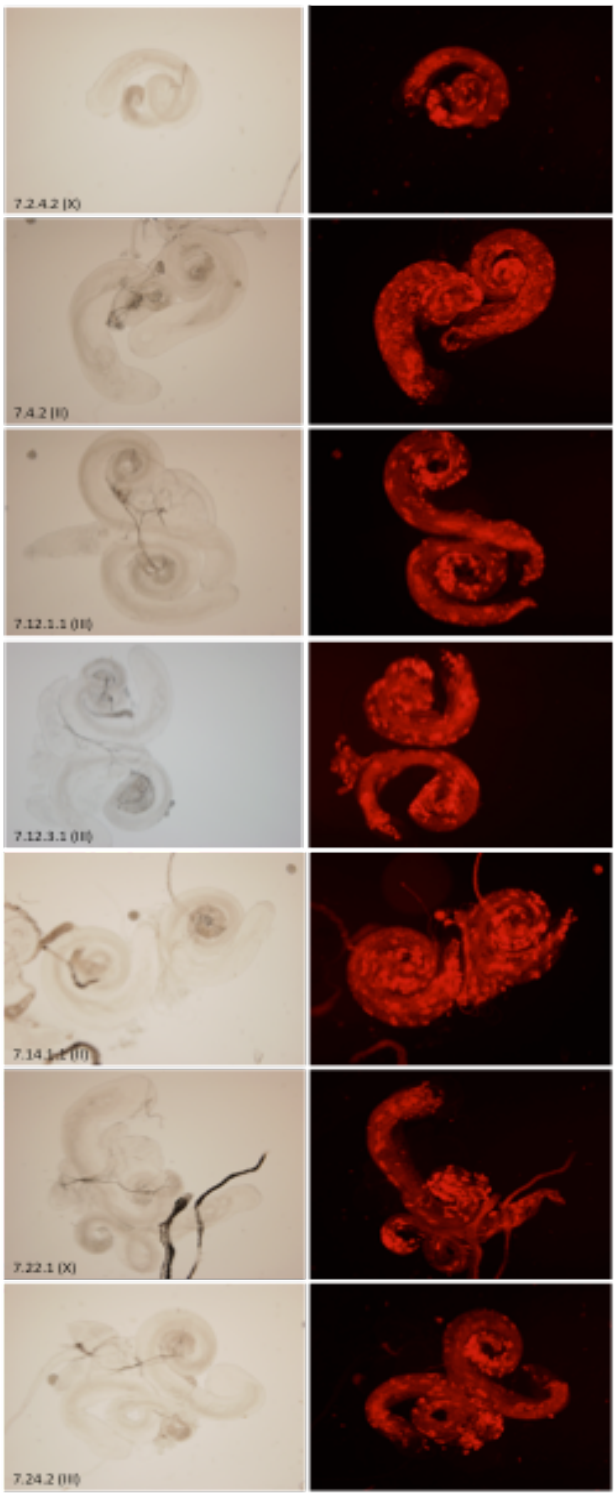
Additional lines for presence and/or absence of Dntf-2r-EGFP in imaginal discs for the construct containing the longest upstream region from the gene


1.53.2.2_II
1.53.5.1_III
100 µm
1.53.6.2_X
100 µm
200 µm
1.53.8.3_III
1.53.44.3_X
50.0 µm
200 µm
1.53.45.1_II
100 µm
200 µm
1.53.76.1_II
1.53.76.8_II_HL
100 µm

APPENDIX C

Additional lines for presence of Ran-like in testis for the three constructs containing the different upstream regions. Construct 4.24.1 as 0bp upstream region. Constructs 8.4.1.1, 8.16.2, 8.31.3, and 8.38.1 have 100bp upstream of the TSS. Constructs 7.2.4.2, 7.4.2, 7.12.1.1, 7.12.3.1, 7.14.1.1, 7.22.1, and 7.24.2 all have the longest upstream region (500bp). The chromosome location for each line is in parenthesis.


4.24.1 (III)
8.4.1.1 (X)
8.16.2 (II)
8.31.3 (III)
8.38.1 (III)


7.2.4.2 (X)
7.4.2 (II)
7.12.1.1 (II)
7.12.3.1 (II)
7.22.1 (X)
7.24.2 (II)

REFERENCES

Anderson MA, Jodoin JN, Lee E, Hales KG, Hays TS, Lee LA. 2009. Asunder is a critical regulator of dynein-dynactin localization during Drosophila spermatogenesis. MBoC. 20:2709-2721.

Anisimova M, Gascuel O. 2006. Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. Syst Biol 55:539-552.

Assis R, Bachtrog D. 2013. Neofunctionalization of young duplicate genes in Drosophila. Proc Natl Acad Sci U S A 110:17409-17414.

Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, et al. 2009. Systems genetics of complex traits in *Drosophila melanogaster*. Nat Genet 41:299-307.

Bachtrog D, Toda NR, Lockton S. 2010. Dosage compensation and demasculinization of X chromosomes in Drosophila. Curr Biol 20:1476-1481.

Bai Y, Casola C, Betrán E. 2008. Evolutionary origin of regulatory regions of retrogenes in Drosophila. BMC Genomics 9:241.

Bai Y, Casola C, Feschotte C, Betrán E. 2007. Comparative Genomics Reveals a Constant Rate of Origination and Convergent Acquisition of Functional Retrogenes in Drosophila. Genome Biol 8:R11.

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. 2002. Recent segmental duplications in the human genome. Science 297:1003-1007.

Bailey JA, Liu G, Eichler EE. 2003. An Alu Transposition Model for the Origin and Expansion of Human Segmental Duplications. The American Journal of Human Genetics 73:823-834.

Bassett Andrew R, Tibbit C, Ponting Chris P, Liu J-L. 2003. Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. Cell Rep 4:220-228.

Bellen H, Levis R, Liao G, He Y, Carlson J, Tsang G, Evans-Holm M, Hiesinger P, Schulze K, Rubin G, et al. 2004. The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. Genetics 167:761-781.

Belote JM, Zhong D. 2009. Duplicated proteasome subunit genes in Drosophila and their roles in spermatogenesis. Heredity 103:23-31.

Betrán E, Bai Y, Motiwale M. 2006. Fast protein evolution and germ line expression of a Drosophila parental gene and its young retroposed paralog. Mol Biol Evol 23:2191-2202.

Betrán E, Long M. 2003. Dntf-2r, a young Drosophila retroposed gene with specific male expression under positive darwinian selection. Genetics 164:977-988.

Betrán E, Thornton K, Long M. 2002. Retroposed New Genes Out of the X in Drosophila. Genome Res. 12:1854-1859.

Bhattacharya A, Steward R. 2002. The Drosophila homolog of NTF-2, the nuclear transport factor-2, is essential for immune response. EMBO Rep 3:378-383.

Biemont C. 2008. Genome size evolution: within-species variation in genome size. Heredity 101:297-298.

Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118:401-415.

Brosius J.. 1991. Retroposons-seeds of evolution. Science 251:753.

Bulmer MS. 2001. Evolution of Immune Proteins in Insects. In. eLS: John Wiley & Sons, Ltd.

Caporilli S, Yu Y, Jiang J, White-Cooper H. 2013. The RNA Export Factor, Nxt1, Is Required for Tissue Specific Transcriptional Regulation. PLoS Genet 9:e1003526.

Casal J, Gonzalez C, Wandosell F, Avila J, Ripoll P. 1990. Abnormal meiotic spindles cause a cascade of defects during spermatogenesis in asp males of Drosophila. Development 108:251-260.

Cesario J, McKim KS. 2011. RanGTP is required for meiotic spindle organization and the initiation of embryonic development in Drosophila. J Cell Sci 124:3797-3810.

Chain FJ, Evans BJ. 2006. Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog Xenopus laevis. PLoS Genet 2:e56.

Chatterjee N, Rollins J, Mahowald AP, Bazinet C. 2011. *Neurotransmitter Transporter-Like*: A Male Germline-specific SLC6 Transporter Required for *Drosophila* Spermiogenesis. PLoS ONE 6:e16275.

Chen JV, Megraw TL. 2014. Spermitin: a novel mitochondrial protein in Drosophila spermatids. PLoS ONE 9:e108802.

Chen S, Ni X, Krinsky BH, Zhang YE, Vibranovski MD, White KP, Long M. 2012A. Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene. EMBO J 31:2798-2809.

Chen S, Spletter M, Ni X, White KP, Luo L, Long M. 2012B. Frequent recent origination of brain genes shaped the evolution of foraging behavior in Drosophila. Cell Rep 1:118-132.

Chen S, Zhang YE, Long M. 2010. New Genes in Drosophila Quickly Become Essential. Science 330:1682-1685.

Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39:715-720.

Clarke PR, Zhang C. 2008. Spatial and temporal coordination of mitosis by Ran GTPase. Nature Reviews Genetics 9:464-477.

Connallon T, Clark AG. 2011. The Resolution of Sexual Antagonism by Gene Duplication. Genetics 187:919-937.

Conrad DF, Hurles ME. 2007. The population genetics of structural variation. Nat Genet 39:S30-36.

Dahl HH, Brown RM, Hutchison WM, Maragos C, Brown GK. 1990. A testis-specific form of the human pyruvate dehydrogenase E1 alpha subunit is coded for by an intronless gene on chromosome 4. Genomics 8:225-232.

Dai H, Yoshimatsu TF, Long M. 2006. Retrogene movement within- and betIen-chromosomes in the evolution of Drosophila genomes. Gene 385:96-102.

de Cuevas M, Matunis EL. 2011. The stem cell niche: lessons from the Drosophila testis. Development 138:2861-2869.

Dermitzakis ET, Masly, J.P., Waldrip, H. M. and Clark A.G. 2000. Non-mendelian segregation of sex chromosomes in heterospecific drosophila males. Genetics 154:687-694.

Dietzl G, Chen D, Schnorrer F, Su K, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, et al. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448:151-156.

Dorogova N, Akhmametyeva E, Kopyl S, Gubanova N, Yudina O, Omelyanchuk L, Chang L-S. 2008. The role of Drosophila Merlin in spermatogenesis. BMC Cell Biology 9:1.

Dorus S, Busby S, Gerike U, Shabanowitz J, Hunt D, Karr T. 2006. Genomic and functional evolution of the Drosophila melanogaster sperm proteome. Nat Genet 38:1440-1445.

Dorus S, Freeman ZN, Parker ER, Heath BD, Karr TL. 2008. Recent origins of sperm genes in Drosophila. Mol Biol Evol 25:2157-2166.

Duffy JB. 2002. GAL4 system in drosophila: A fly geneticist's swiss army knife. Genesis 34:1-15.

Dunham I, Shimizu N, Roe BA, Chissoe S, Hunt AR, Collins JE, Bruskiewich R, Beare DM, Clamp M, Smink LJ, et al.. 1999. The DNA sequence of human chromosome 22 [see comments] [published erratum appears in Nature 2000 Apr 20;404(6780):904]. Nature 402:489-495.

Dyer KA, White BE, Bray MJ, Piqué DG, Betancourt AJ. 2011. Molecular Evolution of a Y Chromosome to Autosome Gene Duplication in Drosophila. Mol Biol Evol 28:1293-1306.

Eisman RC, Kaufman TC. 2013. Probing the boundaries of orthology: the unanticipated rapid evolution of Drosophila centrosomin. Genetics 194:903-926.

Emerson JJ, Kaessmann H, Betrán E, Long M. 2004. Extensive Gene Traffic on the Mammalian X Chromosome. Science 303:537-540.

Esnault C, Maestre J, Heidmann T.. 2000. Human LINE retrotransposons generate processed pseudogenes. Nat Genet 24:363-367.

Fabian L, Brill JA. 2012. Drosophila spermiogenesis. Spermatogenesis 2:197-212.

Feng Q, Moran JV, Kazazian HH, Jr., Boeke JD. 1996. Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. Cell 87:905-916.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331-368.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J.. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531-1545.

French V, Daniels G. 1994. Pattern Formation: The beginning and the end of insect limbs. Curr Biol 4:34-37.

Friedman R, Hughes AL. 2004. Two Patterns of Genome Organization in Mammals: the Chromosomal Distribution of Duplicate Genes in Human and Mouse. Mol Biol Evol 21:1008-1013.

Fu J, Jiang Q, Zhang C. 2010. Coordination of Cell Cycle Events by Ran GTPase. Nature Education 3:32.

Fuller MT. 1993. Spermatogenesis. Cold Spring Harbor, N.Y.: Cold Spring Harbor Press.

Gallach M, Betrán E. 2011. Intralocus sexual conflict resolved through gene duplication. Trends Ecol Evol 26:222-228.

Gallach M, Chandrasekaran C, Betrán E. 2010. Analyses of nuclearly-encoded mitochondrial genes suggest gene duplication as a mechanism for resolving intralocus sexual antagonistic conflict in Drosophila. Genome Biol Evol 2:835-850.

Gallach M, Domingues S, Betrán E. 2011. Gene duplication and the genome distribution of sex-biased genes. Int J Evol Biol 2011:989438.

Ganetzky B. 1977. On the components of segregation distortion in Drosophila melanogaster. Genetics 86:321-355.

Gelbart WM, Emmert DB. 2013. FlyBase High Throughput Expression Pattern Data. FlyBase analysis.

Gell SL, Reenan RA. 2013. Mutations to the piRNA Pathway Component Aubergine Enhance Meiotic Drive of Segregation Distorter in *Drosophila melanogaster*. Genetics 193:771-784.

Ghosh-Roy A, Desai BS, Ray K. 2005. Dynein Light Chain 1 Regulates Dynamin-mediated F-Actin Assembly during Sperm Individualization in Drosophila. Mol Biol Cell 16:3107-3116.

Graur D, Li W-H. 2000. Fundamentals of Molecular Evolution. Sunderland, MA: Sinauer Associates Inc.

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al. 2011. The developmental transcriptome of Drosophila melanogaster. Nature 471:473-479.

Grice SJ, Liu J-L. 2011. Survival Motor Neuron Protein Regulates Stem Cell Division, Proliferation, and Differentiation in *Drosophila*. PLoS Genet 7.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol 59:307-321.

Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, et al. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. Genetics 177:1321-1335.

Hahn MW, Han MV, Han S-G. 2007. Gene Family Evolution across 12 Drosophila Genomes PLoS Genet 3:e197.

Halder G, Callaerts P, Gehring WJ. 1995. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science 267:1788-1792.

Hales K, Fuller M. 1997. Developmentally regulated mitochondrial fusion mediated by a conserved, novel, predicted GTPase. Cell 90:121-129.

Han MV. 2012. Characterizing gene movements between chromosomes in Drosophila. Fly 6:121-125.

Han MV, Hahn MW. 2012. Inferring the History of Interchromosomal Gene Transposition in Drosophila Using n-Dimensional Parsimony. Genetics 190:813-825.

Harbecke R, Meise M, Holz A, Klapper R, Naffin E, Nordhoff V, Janning W. 1996. Larval and imaginal pathways in early development of Drosophila. Int J Dev Biol 40:197-204.

Hardy RW, Tokuyasu KT, Lindsley DL, Garavito M. 1979. The germinal proliferation center in the testis of Drosophila melanogaster. J Ultrastruct Res 69:180-190.

Hare EE, Peterson BK, Eisen MB. 2008. A careful look at binding site reorganization in the even-skipped enhancers of Drosophila and sepsids. PLoS Genet 4:e1000268.

Hastings PJ, Lupski JR, Rosenberg SM, Ira G. 2009. Mechanisms of change in gene copy number. Nat Rev Genet 10:551-564.

Heger A, Ponting CP. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 Drosophila genomes. Genome Res 17:1837-1849.

Hense W, Baines JF, Parsch J. 2007. X chromosome inactivation during Drosophila spermatogenesis. PLoS Biol 5:e273.

Hou Y, Lin S. 2009. Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: gene content estimation for dinoflagellate genomes. PLoS ONE 4:e6978.

Hurst LD. 2001. Evolutionary genomics. Sex and the X. Nature 411:149-150.

Immler S, Pitnick S, Parker GA, Durrant KL, Lüpold S, Calhim S, Birkhead TR. 2011. Resolving variation in the reproductive tradeoff between sperm size and number. Proc Natl Acad Sci U S A 108:5325-5330.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11:97-108.

Isgro TA, Schulten K. 2007. Association of nuclear pore FG-repeat domains to NTF2 import and export complexes. J Mol Biol 366:330-345.

Jeffs P, Ashburner M. 1991. Processed pseudogenes in Drosophila. Pro. R. Soc. Lond. B 244:151-159.

Jonchere V, Bennett D. 2013. Validating RNAi Phenotypes in Drosophila Using a Synthetic RNAi-Resistant Transgene. PLoS ONE 8:e70489.

Kabza M, Ciomborowska J, Makałowska I. 2014. RetrogeneDB - A Database of Animal Retrogenes. Mol Biol Evol 31(7):1646-1648.

Kaessmann H. 2010. Origins, evolution and phenotypic impact of new genes. Genome Res 20:1313-1326.

Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet 10:19-31.

Kalab P, Pralle A, Isacoff EY, Heald R, Weis K. 2006. Analysis of a RanGTP-regulated gradient in mitotic somatic cells. Nature 440:697-701.

Kalamegham R, Sturgill D, Siegfried E, Oliver B. 2007. *Drosophila mojoless*, a retroposed GSK-3, has functionally diverged to acquired an essential role in male fertility. Mol Biol Evol 24:732-742.

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30:772-780.

Kazazian HH, Jr. 2004. Mobile Elements: Drivers of Genome Evolution. Science 303:1626-1632.

Kemkemer C, Catalan A, Parsch J. 2014. 'Escaping' the X chromosome leads to increased gene expression in the male germline of Drosophila melanogaster. Heredity 112:149-155.

Khil PP, Smirnova NA, Romanienko PJ, Camerini-Otero RD. 2004. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. Nat Genet 36:642-646.

Klasterska I, Ramel C. 1990. A comparative study of male meiosis in Drosophila melanogaster and D. virilis. Genetica 81:89-98.

Kohler A, Hurt E. 2007. Exporting RNA from the nucleus to the cytoplasm. Nature reviews Mol Cell Biol 8:761-773.

Kracklauer MP, Wiora HM, Deery WJ, Chen X, Bolival B, Romanowicz D, Simonette RA, Fuller MT, Fischer JA, Beckingham KM. 2010. The Drosophila SUN protein Spag4 cooperates with the coiled-coil protein Yuri Gagarin to maintain association of the basal body and spermatid nucleus. J Cell Sci 123:2763-2772.

Kumar JP, Wilkie GS, Tekotte H, Moses K, Davis I. 2001. Perturbing Nuclear Transport in Drosophila Eye Imaginal Discs Causes Specific Cell Adhesion and Axon Guidance Defects. Dev Biol 240:315-325.

Kurland CG. 2000. Something for everyone. Horizontal gene transfer in evolution. EMBO Rep 1:92-95.

Kusano A, Staber C, Chan HY, Ganetzky B. 2003. Closing the (Ran)GAP on segregation distortion in Drosophila. Bioessays 25:108-115.

Kusano A, Staber C, Ganetzky B. 2002. Segregation distortion induced by wild-type RanGAP in Drosophila. Proc Natl Acad Sci U S A 99:6866-6870.

Larracuente AM, Presgraves DC. 2012. The selfish Segregation Distorter gene complex of Drosophila melanogaster. Genetics 192:33-53.

Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. Proc Natl Acad Sci U S A 103: 9935–9939.

Li VC, Davis JC, Lenkov K, Bolival B, Fuller MT, Petrov DA. 2009. Molecular evolution of the testis TAFs of Drosophila. Mol Biol Evol 26:1103-1116.

Lifschytz E, Lindsley DL. 1972. The role of X-chromosome inactivation during spermatogenesis. Proc Natl Acad Sci U S A 69:182-186.

Long M, Betrán E, Thornton K, Wang W. 2003A. The origin of new genes: glimpses from the young and old. Nat Rev Genet 4:865-875.

Long M, Deutsch M, Wang W, Betrán E, Brunet FG, Zhang J. 2003B. Origin of new genes: evidence from experimental and computational analyses. Genetica 118:171-182.

Long M, Langley CH.. 1993. Natural selection and the origin of jingli, a chimeric processed functional gene in Drosophila. Science 260:91-95.

Loppin B, Lepetit D, Dorus S, Couble P, Karr TL. 2005. Origin and neofunctionalization of a Drosophila paternal effect gene essential for zygote viability. Curr Biol 15:87-93.

Lynch M, Conery JS. (20525849 co-authors). 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155.

Marques-Bonet T, Girirajan S, Eichler EE. 2009. The origins and impact of primate segmental duplications. Trends Genet 25:443-454.

Marygold SJ, Leyland PC, Seal RL, Goodman JL, Thurmond J, Strelets VB, Wilson RJ. 2013. FlyBase: improvements to the bibliography. Nucleic Acids Res 41:D751-757.

Matsuura Y, Stewart M. 2004. Structural basis for the assembly of a nuclear export complex. Nature 432:872-877.

McCarrey JR. 1994. Evolution of tissue-specific gene expression in mammals: How a new phosphoglycerate kinase was formed and refined. BioScience 44:20-27.

McCarrey JR, Thomas K. 1987. Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. Nature 326:501-505.

McElroy JM, McLean RA, McLean JR. 2008. Defects in nuclear transport enhance segregation distortion. Fly 2:280-290.

Meiklejohn CD, Landeen EL, Cook JM, Kingan SB, Presgraves DC. 2011. Sex Chromosome-Specific Regulation in the *Drosophila* Male Germline But Little Evidence for Chromosomal Dosage Compensation or Meiotic Inactivation. PLoS Biol 9:e1001126.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in Drosophila. Proc Natl Acad Sci U S A 100:9894-9899.

Meiklejohn CD, Tao Y. 2010. Genetic conflict and sex chromosome evolution. Trends Ecol Evol 25:215-223.

Meisel RP, Han MV, Hahn MW. 2009. A Complex Suite of Forces Drives Gene Traffic from Drosophila X Chromosomes. Genome Biol Evol 1:176-188.

Mendivil Ramos O, Ferrier DEK. 2012. Mechanisms of Gene Duplication and Translocation and Progress towards Understanding Their Relative Contributions to Animal Genome Evolution. Int J Evol Biol 2012:10.

Michiels F, Gasch A, Kaltschmidt B, Renkawitz-Pohl R.. 1989. A 14 bp promoter element directs the testis specificity of the Drosophila beta 2 tubulin gene. EMBO J 8:1559-1565.

Mikhaylova L, Nurminsky D. 2012. No severe and global X chromosome inactivation in meiotic male germline of Drosophila. BMC Biology 10:50.

Moore MS. 1998. Ran and Nuclear Transport. J Biol Chem 273:22857-22860.

Morris CA, Benson E, White-Cooper H. 2009. Determination of gene expression patterns using in situ hybridization to Drosophila testes. Nat Protoc 4:1807-1819.

Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M. 2006. The 160-Kilobase Genome of the Bacterial Endosymbiont Carsonella. Science 314:267.

Negre N, Brown CD, Shah PK, Kheradpour P, Morrison CA, Henikoff JG, Feng X, Ahmad K, Russell S, White RA, et al. 2010. A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet 6:e1000814.

Ni J-Q, Liu L-P, Binari R, Hardy R, Shim H-S, Cavallaro A, Booker M, Pfeiffer BD, Markstein M, Wang H, et al. 2009. A Drosophila Resource of Transgenic RNAi Lines for Neurogenetics. Genetics 182:1089-1100.

Noguchi T, Koizumi M, Hayashi S. 2012. Mitochondria-driven cell elongation mechanism for competing sperms. Fly 6:113-116.

Nosov GA, Kibanov MV, Olenina LV. 2014. Dynamic properties of a germinal granule piNG-body in the testes of *Drosophila melanogaster*. Molecular Biology 48:701-708.

O'Neill RS, Clark DV. 2013. The Drosophila melanogaster septin gene Sep2 has a redundant function with the retrogene Sep5 in imaginal cell proliferation but is essential for oogenesis. Genome 56:753-758.

Ohno S. 1970. Evolution by gene duplication. New York: Springer.

Ou Q, Magico A, King-Jones K. 2011. Nuclear Receptor DHR4 Controls the Timing of Steroid Hormone Pulses During *Drosophila* Development. PLoS Biol 9:e1001160.

Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Oliver B. 2003. Paucity of genes on the Drosophila X chromosome showing male-biased expression. Science 299:697-700.

Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu XJ, Harte R, Balasubramanian S, Tanzer A, Diekhans M, et al. 2012. The GENCODE pseudogene resource. Genome Biol 13:R51.

Pellicer J, Fay MF, Leitch IJ. 2010. The largest eukaryotic genome of them all? Botanical Journal of the Linnean Society 164:10-15.

Perrimon N, Ni J-Q, Perkins L. 2010. In vivo RNAi: Today and Tomorrow. Cold Spring Harb Perspect Biol 2(8):a003640.

Peter A, Schottler P, Irner M, Beinert N, Dol G, Burkert P, Mourkioti F, Dentzer L, He Y, Deak P, et al. 2002. Mapping and identification of essential gene functions on the X chromosome of Drosophila. EMBO Rep 3:34-38.

Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R. 2007. Genes under positive selection in Escherichia coli. Genome Res 17:1336-1343.

Phadnis N, Hsieh E, Malik HS. 2012. Birth, death, and replacement of karyopherins in Drosophila. Mol Biol Evol 29:1429-1440.

Phelps CB, Brand AH. 1998. Ectopic Gene Expression in Drosophila Using GAL4 System. Methods 14:367-379.

Potrzebowski L, Vinckenbosch N, Kaessmann H. 2010. The emergence of new genes on the young therian X. Trends Genet 26:1-4.

Potrzebowski L, Vinckenbosch N, Marques A, Chalmel F, Jegou B, Kaessmann H. 2008. Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. PLoS Biol. 6:e80.

Presgraves DC. 2007. Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila? Bioessays 29:386-391.

Presgraves DC, Gerard PR, Cherukuri A, Lyttle TW. 2009. Large-scale selective sIep among Segregation Distorter chromosomes in African populations of Drosophila melanogaster. PLoS Genet 5:e1000463.

Presgraves DC, Stephan W. 2007. Pervasive adaptive evolution among interactors of the Drosophila hybrid inviability gene, Nup96. Mol Biol Evol 24:306-314.

Prince VE, Pickett FB. 2002. Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3:827-837.

Quezada-Diaz JE, Muliyil T, Río J, Betrán E. 2010. Drcd-1 related: a positively selected spermatogenesis retrogene in Drosophila. Genetica 138:925-937.

Quimby BB, Lamitina T, L'Hernault SW, Corbett AH.. 2000. The mechanism of ran import into the nucleus by nuclear transport factor 2. J Biol Chem 275:28575-28582.

Ranz J, Maurin D, Chan Y, von Grotthuss M, Hillier L, Roote J, Ashburner M, Bergman C. 2007. Principles of genome evolution in the Drosophila melanogaster species group. PLoS Biol 5:e152.

Reinke V, Smith HE, Nance J, Wang J, Van Doren C, Begley R, Jones SJ, Davis EB, Scherer S, Ward S, et al. 2000. A global profile of germline gene expression in C. elegans. Mol Cell 6:605-616.

Rewitz KF, Yamanaka N, O'Connor MB. 2010. Steroid Hormone Inactivation Is Required during the Juvenile-Adult Transition in Drosophila. Dev Cell 19:895-902.

Ribbeck K, Lipowsky G, Kent HM, Stewart M, Gorlich D. 1998. NTF2 mediates nuclear import of Ran. EMBO J 17:6587-6598.

Rice WR. 1987. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination betIen primitive sex chromosomes. Evolution 41:911-914.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. Evolution 38:735-742.

Ross BD, Rosin L, Thomae AW, Hiatt MA, Vermaak D, de la Cruz AFA, Imhof A, Mellone BG, Malik HS. 2013. Stepwise Evolution of Essential Centromere Function in a Drosophila Neogene. Science 340:1211-1214.

Russell MA. 1974. Pattern formation in the imaginal discs of a temperature-sensitive cell-lethal mutant of Drosophila melanogaster. Dev Biol 40:24-39.

Samonte RV, Eichler EE. 2002. Segmental duplications and the evolution of the primate genome. Nat Rev Genet 3:65-72.

Santel A, Kaufmann J, Hyland R, Renkawitz-Pohl R. 2000. The initiator element of the Drosophila beta2 tubulin gene core promoter contributes to gene expression in vivo but is not required for male germ-cell specific expression. Nucleic Acids Res 28:1439-1446.

Schrider DR, Stevens K, Cardeño CM, Langley CH, Hahn MW. 2011. Genome-wide analysis of retrogene polymorphisms in Drosophila melanogaster. Genome Res 21:2087-2095.

Sen S, Hartmann B, Reichert H, Rodrigues V. 2010. Expression and function of the empty spiracles gene in olfactory sense organ development of *Drosophila melanogaster*. Development 137:3687-3695.

Shevelyov YY, Lavrov SA, Mikhaylova LM, Nurminsky ID, Kulathinal RJ, Egorova KS, Rozovsky YM, Nurminsky DI. 2009. The B-type lamin is required for somatic repression of testis-specific gene clusters. Proc Natl Acad Sci U S A 106:3282-3287.

Silverman-Gavrila RV, Wilde A. 2006. Ran is required before metaphase for spindle assembly and chromosome alignment and after metaphase for chromosome segregation and spindle midbody organization. MBoC 17:2069-2080.

Simpson P, Berreur P, Berreur-Bonnenfant J. 1980. The initiation of pupariation in Drosophila: dependence on growth of the imaginal discs. J Embryol Exp Morphol 57:155-165.

Simpson P, Schneiderman HA. 1975. Isolation of temperature sensitive mutations blocking clone development in Drosophila melanogaster, and the effects of a temperature sensitive cell lethal mutation on pattern formation in imaginal discs. Development Genes and Evolution 178(3):247-275.

Sinzelle L, Izsvak Z, Ivics Z. 2009. Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell Mol Life Sci 66:1073-1093.

Sitaram P, Anderson MA, Jodoin JN, Lee E, Lee LA. 2012. Regulation of dynein localization and centrosome positioning by Lis-1 and asunder during Drosophila spermatogenesis. Development 139:2945-2954.

Sorourian M, Betrán E. 2010. Turnover and lineage-specific broadening of the transcription start site in a testis-specific retrogene. Fly 4:3-11.

Sorourian M, Kunte MM, Domingues S, Gallach M, Özdil F, Río J, Betrán E. 2014. Relocation Facilitates the Acquisition of Short Cis-Regulatory Regions that Drive the Expression of Retrogenes during Spermatogenesis in Drosophila. Mol Biol Evol 31:2170-2180.

Stieper BC, Kupershtok M, Driscoll MV, Shingleton AW. 2008. Imaginal discs regulate developmental timing in Drosophila melanogaster. Dev Biol 321:18-26.

Sturgill D, Zhang Y, Parisi M, Oliver B. 2007. Demasculinization of X chromosomes in the Drosophila genus. Nature 450:238-241.

Sun S, Ting CT, Wu CI. 2004. The normal function of a speciation gene, Odysseus, and its hybrid sterility effect. Science 305:81-83.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proc Natl Acad Sci U S A 98:7375-7379.

Tao Y, Araripe L, Kingan SB, Ke Y, Xiao H, Hartl DL. 2007. A sex-ratio meiotic drive system in Drosophila simulans. II: an X-linked distorter. PLoS Biol 5:e293.

Texada MJ, Simonette RA, Johnson CB, Deery WJ, Beckingham KM. 2008. Yuri gagarin is required for actin, tubulin and basal body functions in Drosophila spermatogenesis. J Cell Sci 121:1926-1936.

Timakov B, Zhang P. 2001. The hsp60B gene of Drosophila melanogaster is essential for the spermatid individualization process. Cell Stress Chaperones 6:71-77.

Tokuyasu KT, Peacock WJ, Hardy RW. 1972. Dynamics of spermiogenesis in Drosophila melanogaster. Z Zellforsch Mikrosk Anat 127:492-525.

Tracy C, Rio J, Motiwale M, Christensen SM, Betrán E. 2010. Convergently recruited nuclear transport retrogenes are male biased in expression and evolving under positive selection in Drosophila. Genetics 184:1067-1076.

Trieselmann N, Wilde A. 2002. Ran localizes around the microtubule spindle in vivo during mitosis in Drosophila embryos. Curr Biol 12:1124-1129.

Tulina N, Matunis E. 2001. Control of stem cell self-renewal in Drosophila spermatogenesis by JAK-STAT signaling. Science 294:2546-2549.

Van Impe K, Hubert T, De Corte V, Vanloo B, Boucherie C, Vandekerckhove J, Gettemans J. 2008. A new role for nuclear transport factor 2 and Ran: nuclear import of CapG. Traffic 9:695-707.

Vibranovski M. 2014. Meiotic sex chromosome inactivation in Drosophila. J Genomics 2:104-117.

Vibranovski MD, Lopes HF, Karr TL, Long M. 2009A. Stage-Specific Expression Profiling of Drosophila Spermatogenesis Suggests that Meiotic Sex Chromosome Inactivation Drives Genomic Relocation of Testis-Expressed Genes. PLoS Genet 5:e1000731.

Vibranovski MD, Zhang Y, Long M. 2009B. General gene movement off the X chromosome in the Drosophila genus. Genome Res. 19:897-903.

Vicoso B, Charlesworth B. 2009. The deficit of male-biased genes on the D. melanogaster X chromosome is expression-dependent: a consequence of dosage compensation? J Mol Evol 68:576-583.

Vinckenbosch N, Dupanloup I, Kaessmann H. 2006. Evolutionary fate of retroposed gene copies in the human genome. Proc Natl Acad Sci U S A 103:3220-3225.

Wagner A. (21233560 co-authors). 2001. Birth and death of duplicated genes in completely sequenced eukaryotes. Trends Genet 17:237-239.

Wang PJ, McCarrey JR, Yang F, Page DC. 2001. An abundance of X-linked genes expressed in spermatogonia. Nat Genet. 27:422-426.

Wang W, Zhang J, Alvarez C, Llopart A, Long M. 2000. The origin of the Jingli gene and the complex modular structure of its parental gene, yellow emperor, in Drosophila melanogaster. Mol Biol Evol 17:1294-1301.

Wang W, Zheng H, Yang S, Yu H, Li J, Jiang H, Su J, Yang L, Zhang J, McDermott J, et al. 2005. Origin and evolution of new exons in rodents. Genome Res. 15:1258-1264.

Iaver M, Krasnow MA. 2008. Dual origin of tissue-specific progenitor cells in Drosophila tracheal remodeling. Science 321:1496-1499.

White-Cooper H. 2010. Molecular mechanisms of gene regulation during Drosophila spermatogenesis. Reproduction 139:11-21.

White-Cooper H. 2012. Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the Drosophila testis. Spermatogenesis 2:11-22.

Yang J, Ramos E, Corces VG. 2012. The BEAF-32 insulator coordinates genome organization and function during the evolution of Drosophila species. Genome Res 22:2199-2207.

Zhang Z, Hambuch TM, Parsch J. 2004. Molecular Evolution of Sex-Biased Genes in Drosophila. Mol Biol Evol 21:2130-2139.

Zhong L, Belote JM. 2007. The testis-specific proteasome subunit Prosalpha6T of D. melanogaster is required for individualization and nuclear maturation during spermatogenesis. Development 134:3517-3525.

BIOGRAPHICAL INFORMATION

Susana Domingues was born in Lisbon, Portugal in 1981. She received her B.S., a 5 years degree in Portugal, from the University of Algarve in 2005. Her B.S. is in Marine Biology with major in Molecular and Developmental Biology. During her B.S. she worked under the supervision Dr. Leonor Cancela in a Biochemistry and Molecular Biology Laboratory. After obtaining her degree she moved to Germany where she performed research on the effects of vitamin A in zebrafish development. Since arriving at the University of Texas at Arlington her research has focused on the study of the regulatory regions and the function of Drosophila retrogenes. To her credit, to do this work she obtained a partial fellowship from the Portuguese Science Foundation and that has paid for part of her tuition in the last few years. She also participated in the I ENGAGE program in 2014 that is designed for senior graduate students to mentor an undergraduate student over summer. From 2009 to 2014, she has been multiple times a laboratory instructor in the Biology Department at the University of Texas at Arlington for three years. Susana has also been a Graduate Research Assistant for 3 years supported by the NIH to E.B. Susana plans to be involved in clinical research, including product research and development, in the future. She is a member of Phi Sigma Honors Society for biology graduate students and has volunteered at Cook Children's Hospital in the past.