ANALYSIS OF NOVEL REGULATORY REGION AND FUNCTION OF A YOUNG *DROSOPHILA* RETROGENE: *Dntf-2r*

by

MANSI MOTIWALE KUNTE

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

August 2009

ACKNOWLEDGEMENTS

First and foremost I offer my sincerest gratitude to my supervisor, Dr Esther Betrán, who has supported me throughout my thesis with her patience and knowledge whilst allowing me the room to work in my own way. It has been my honor to work with her. I would also like to thank Dr. Cedric Feschotte, Dr. Jeffery Demuth, Dr. Andre Pires Da Silva, and Dr. Pawel Michalak for their support and guidance as my committee members. Special thanks are due to Dr. Shawn Christensen for his valuable comments in my research. It was my delight to interact with these and other faculty members from whom I have learnt a lot and they have contributed to some or the other aspect of my research.

In my daily work I have been blessed with friendly and cheerful fellow students. I want to thank everybody form Betrán lab especially Xavier, who has helped me in every stage of my research and given friendly advise from time to time. I am grateful to all the past and present members of the lab, especially Charles, Susana, Jorge and Miguel who have been a part of healthy discussions about work and more importantly life.

Also I want to thank other students of the department who have given me great friendship and provided me a lot of emotional as well as scientific support. I am especially grateful to Claudia for performing all the radioactive labeling for me.

Finally I am indebted to my family for always being there for me, especially my husband and daughter who have been very patient and supportive of me.

July 14,2009

ABSTRACT

ANALYSIS OF NOVEL REGULATORY REGION AND FUNCTION

OF A YOUNG *DROSOPHILA* RETROGENE:

Dntf-2r

Mansi Motiwale Kunte, PhD.

The University of Texas at Arlington, 2009

Supervising Professor: Esther Betrán

Dntf-2r is a young retroposed gene that originated from the nuclear transport gene *Dntf-2*. Here the function, quality and evolutionary origin of the regulation of this retrogene are explored. In Chapter 1, I introduce the background of what is known of *Dntf-2r* as well as the function of its parental gene. This includes their evolutionary histories, and known functional and regulatory aspects. Chapter 2 describes my work on the novel regulatory region recruited by the retrogene and I explore the tissue specific expression pattern (i.e. testis expression) driven by this region. In Chapter 3 I set myself up to review what is known about retrogene regulation. I do this to be able to understand the mechanisms by which young retrogenes can be transcribed to become functional and provide some general understanding of this important feature of retrogenes. In Chapters 4 and 5 I address the function of *Dntf-2r*. In Chapter 4 I look in detail at the protein compared to the parental protein. I also study its cellular localization and protein interactions using a fusion to EGFP and look at the spermatogenesis and fertility of a *Dntf-2r* knockout mutant. Finally, as many nuclear transport genes have been implicated or suggested to be involved in causing/suppressing meiotic drive, in Chapter 5, I explore the role of *Dntf-2r* in meiotic drive using the known Segregation Distortion (*SD*) system in *D. melanogaster*.

Results reveal a short 14 bp motif in the 5' flanking region is identified as regulatory sequence driving testis specific expression of *Dntf-2r*. This is a motif is conserved in all species where the gene is present and is similar to the known β2 tubulin promoter motif. Few changes in the region of insertion were required to evolve this element.

Our review of how retrogenes are regulated reveals that promoter recruitment and transcription appears to be in many cases highly dependent on the region of insertion. In other instances the initial survival depends on the quality of the transcript: aberrant transcript carrying regulatory regions from parental gene or downstream regulatory regions carried in the normal transcript. In addition, it is evident in some cases of male germline expression, that selective pressures created and or improved the pattern of expression.

Functional analysis revealed that *Dntf-2r* likely retains some interactions of the parental gene; however, it seems to be essentially dispensable for the organism since all fertility assays performed fail to detect significant decrease in fertility. This supports the emerging genetic conflict hypothesis. Additionally, partial knockout of *Dntf-2r* lowers the affect of otherwise strong *SD*-5 chromosome suggesting that it may be acting as an enhancer of this meiotic drive system, supporting a meiotic drive involvement.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Gene duplicates serve as a raw material for evolutionary forces to create evolutionary novelty (Ohno 1970). Gene duplication has given rise to many new functions, one of the most remarkable being the evolution of developmental genes in various organisms (Prince *et al.* 2002). Duplicates provide functional redundancy in the genome where selection can play upon one copy to mould it anyways. However, not all duplicates are, from the beginning, functional or stay functional. The most common fate of such duplicates is to accumulate deleterious mutations and become pseudogenes. It was observed by Lynch *et al.* (Lynch *et al.* 2000) that right after duplication the genes undergo relaxed selection. After a few million years the duplicate may either become a pseudogene or experience a strong purifying selection suggesting a functional role. According to their study the average rate of duplication of eukaryotic gene is 0.01/gene/MY (Lynch *et al.* 2000). If a duplicate does not degenerate after a few million years, it is preserved in the genome through either subfunctionalization, where the two copies undergo complimentary substitutions such that together they can partition the ancestral function (Force *et al.* 1999) or neofunctionalization, which is usually considered a rarer event where one copy can undergo adaptive evolution and give rise to a novel function. In mammals, a study of young duplicates reveals about 10% of lineage specific genes to be evolving under positive selection (Han *et al.* 2009).

One mechanism of gene duplication is through retroposition. Retroposed genes are processed copies of genes, where the mRNA is believed to be carried back into the nucleus using the enzymatic machinery of LINE elements (Esnault *et al.* 2000) and undergo reverse transcription and insertion into the genome. Retrogenes are intronless and may be flanked by

direct repeats and have a remnant poly A tail. Retrogenes constitute a fraction of genes originating from duplication. In *Drosophila* the rate of gene duplication is estimated to be 0.0023 duplication/gene /million years (Hahn, 2007). Bai *et al.* estimated the rate of origin of new genes via retroposition to be 0.51genes/MY/lineage (Bai *et al.* 2007). They identified functional retrogenes independent of the location of the parent gene in *D. melanogaster* using 50% protein identity level. They also studied other 11 sequenced genomes for presence and absence of retrogenes to infer the age of each retrogene independent of sequence divergence analysis and to calculate the minimum rate of generation of functional retrogenes. Ninety-four candidate functional retroposition events were identified by this approach. Thirty-two of those occurred in the last 63 My (Bai *et al.* 2007).

Retrogenes can lead to the formation of new genes via various pathways. When a retrogene forms a chimera with other duplicates a novel gene is created that carries the functional domains that were previously not together (Long *et al.* 1993, Jones *et al.* 2005, Arguello *et al.* 2007). Besides forming a new chimera, retrogenes can donate a novel functional domain to an existing gene through alternative splicing (Guldner *et al.* 1999). Novel functions can also arise when a retrogene undergoes adaptive evolution or positive selection to make changes at the protein level thus giving rise to new protein different from the parent. In various lineages, retrogenes have been under positive selection (Long *et al.* 1993; Kalamegham *et al.* 2007; Betran *et al.* 2003; Cusack *et al.* 2007; Tracy *et al.* submitted) and some have known functional value (Sayah *et al.* 2004; Zhang *et al.* 2004; Timakov *et al.* 2001; Bradley *et al.* 2004; Donaldson *et al.* 2004; Rohozinski *et al.* 2006; Kalamegham *et al.* 2007; Zhong *et al.* 2007).

Respect to retrogenes, it is important to note that as retrogenes arise from mRNA they essentially lack any regulatory region (Reviewed by (Vinckenbosch *et al.* 2006; Bai *et al.* 2008). Exceptions to this fact are the retrogenes that arise from aberrant transcript of the parent (McCarrey 1990) which might carry the regulatory region of the parent. All the other transcribed retroposed sequences have to either land downstream of an existing promoter or recruit de novo regulatory region/s. Neighboring genes at the site of insertion have also been suggested

or implicated as source of context (Kalmykova *et al.* 2005) or regulatory regions (Hofmann *et al.* 1991). A study in *Drosophila* shows that most retrogenes do not carry regulatory region form the parent gene, and usually neighboring genes and transposable element also do not seem to donate regulatory regions to retrogenes (Bai *et al.* 2008). Retrogenes that are in excess in testis neighborhoods or close to genes (Bai *et al.* 2008; Vinckenbosch *et al.* 2006; Dorus *et al.* 2008) and particular insertion sites having sequences that can transform easily into regulatory regions via few substitutions appear to be the most common routes to recruit a regulatory region (Bai *et al.*. 2008). Chapter 3 of this thesis provides a review of the available data on recruitment of cis-regulatory region by retrogenes from various lineages.

Interestingly, many retrogenes show testis-specific or –biased expression patterns (Betran *et al.* 2002; Emerson *et al.* 2004; Bai *et al.* 2007) and an excess of testis specific or –biased retrogenes represent an X chromosome to autosome duplication (Betran *et al.* 2002; Emerson *et al.* 2004; Bai *et al.* 2007). In mammals, these biases are not explained by the patterns of retropseudogenes (Emerson *et al.* 2004) revealing that they do not represent mutational biases and deeming natural selection as the most reasonable explanation. In Drosophila, there is no data on retropseudogenes that could be used to infer if there are any mutational biases (i.e. there are very few retropseudogenes to study; Bai *et al.* 2007). Therefore other evidence has to be gathered to uncover any mutational biases. In this direction, it was found that this excess is not explained by the insertional biases recently described for retrotransposable elements (Fontanillas *et al.* 2007), allegedly encoding the machinery used for retrotransposition of retrogenes (Esnault *et al.* 2000). Since the X to autosome biases of retrogenes remains unexplained by the insertional biases of TEs, selection might be a driving force.

In addition, some of these retrogenes are known to be needed for spermatogenesis. For example, *Utp14a* is an X linked gene that has given rise to autosomal retrogenes in both mice (*Utp14b*) and humans (*Utp14c* (Rohozinski *et al.* 2004; Rohozinski *et al.* 2006). Mutations in these retrogenes affect spermatogenesis and fertility. Interestingly, *Utp14b* exists as

pseudogene in human. So, recurrent events of retroposition have been observed in mammals and flies (see below). This suggests that a strong selective force preserves the autosomal duplicates of X linked genes. Similarly in *Drosophila*, *mojoless* is a retroposed gene essential for male fertility. It is derived from X linked *shaggy* gene and inserted into third chromosome (Kalamegham *et al.* 2007). For long X chromosome was believed to be the most advantageous place for male specific genes because of sexual antagonism (Rice 1984). However, studies in *Drosophila* showed scarcity of male specific genes on the X chromosome (Parisi *et al.* 2003; Sturgill *et al.* 2007). The demasculinisation of the X chromosome along with a high rate of male specific retrogenes originating from X chromosome implies a strong selective force favoring such movement. Namely two evolutionary forces have been proposed to cause such a trend. First, sexual antagonism suggests that as X chromosome spend two thirds of its life in the female, any mutation that is beneficial to the males but deleterious for the female will be removed (Wu *et al.* 2003; Ranz *et al.* 2003). Secondly, during meiotic phase of spermatogenesis, the X and Y chromosome form a highly condensed XY body and are silenced (Turner 2007). Therefore, it might be advantageous for the genes required during spermatogenesis to have a copy located on the autosome (McCarrey 1990). The general biases on the location of male genes in *Drosophila* have also been argued to be partially explained by the level of expression that genes can attain in dosage and not-dosage compensated chromosomes (Yi *et al.* 2000; Swanson *et al* .2001; Vicoso *et al.* 2009). X-linked genes could be making copies to autosomes to increase level of expression. In this thesis, we propose meiotic drive as a third hypothesis for the X to autosome duplication of a young retrogene – *Dntf-2r*. Meiotic drive has been proposed as a strong selective force involved in shaping of genomes as it causes genetic conflict (Burt *et al.* 2006), and genes involved in such drive systems evolve under positive selection (Presgraves, 2007). However, even if the genes were recruited for a new function, one of the hypotheses or several might explain why the genes originate in autosomes.

1.1 *Dntf-2r*: Background

Drosophila nuclear transport factor-2 related (*Dntf-2r*) is a young retroposed gene present on the left arm of chromosome two. It has originated from *Dntf-2* (*Drosophila NuclearTransport Factor-2*) which is located on the X chromosome. This gene was first described by Betrán *et al.* (Betrán *et al.* 2003). It is present in only four species of Drosophila – *D. melanogaster*, *D. simulans*, *D. sechellia* and *D. mauritiana*. This suggests that the gene is ~ 3-12 My old as that is the time between the common ancestor of all four species and the split of these species with the closest related species (*D. yakuba* and *D. erecta*). In *D. melanogaster*, *Dntf-2r* shows male germline specific expression where as the parent is ubiquitously expressed (Betrán *et al.* 2003). This is a typical example of X to autosome retroduplication where the autosomal copy evolves male biased expression. Interestingly, in all four species, the 5' upstream region of retrogene insertion site harbors a region of sequence similarity with the promoter element of a known testis specific gene *β 2-tubulin* (Michiels 1989). This promoter element comprises of a 14 bp upstream element (β2UE1) and a 7bp quantitative element (Michiels 1989). Both these regions are conserved in *D. melanogaster* and were predicted to drive male germline expression of *Dntf-2r* (Betrán *et al.* 2003). In this work we identify the cis-regulatory region of *Dntf-2r* using P element transformation technology (see goals and chapter 3). The retrogene is inserted near the *Bicoid stability factor* gene (*Bsf*) in a head to head orientation. *Bsf* produces an RNA binding protein that binds to the 3'UTR of *Bicoid* (*Bcd*) mRNA and stabilizes it during oogenesis (Mancebo *et al.* 2001). *Bsf* is highly conserved among all eukaryotes, which makes it easy to find the syntenic region of *Dntf-2r* in related *Drosophila* species and compare it to see the lack of the element (Betran *et al.* 2003). In addition, the expression pattern of *Bsf* is different from that of *Dntf-2r* suggesting that *Dntf-2r* does not borrow promoter from neighbouring genes.

Betrán *et al.* demonstrated that *Dntf-2r* is evolving at a faster rate than the parent (Betrán *et al.* 2003). They calculated Ka/Ks ratios for both genes using PAML software. For *Dntf-2* Ka/Ks is 0.0499 whereas for *Dntf-2r* Ka/Ks is 0.5405 and significantly smaller than one

suggesting, in addition, that both genes are under purifying selection. Using the McDonald-Kreitman test a significant excess of amino acid substitution was observed in *Dntf-2r* suggesting positive selection for this gene.

Many possible scenarios for *Dntf-2r* evolution have been put forward (Betrán *et al.* 2003). One, which was proposed for X to autosome retrogenes, is to compensate for the parental gene which may be silenced due to X inactivation during spermatogenesis. *Dntf-2* would be silenced in certain stages of spermatogenesis, where the nuclear transport function would be carried out by the retrogene. The prediction is that if a gene replaces the other, it should conserve its function (i.e. should be similar and under purifying selection). While there could be some specialization, it is not expected that a gene will keep evolving under positive selection under this hypothesis. A well-known example of a retrogene that substitutes for the function of the parent is *Pgk2* (McCarrey 1990). However, the positive selection acting on *Dntf-2r* is puzzling. It would be interesting to know the transcription and availability of *Dntf-2* protein during spermatogenesis and to see if *Dntf-2* is available all throughout spermatogenesis and *Dntf-2r* has just a new function. In this case, the positive selection acting on *Dntf-2r* may indicate that has evolved a new male germline specific function that is under recurrent selective pressure.

Recently, genes involved in nuclear transport have been suggested as playing a role in genomic conflict: meiotic drive and viral and transposable element defenses have been suggested as pressures that might explain fast evolution (Presgraves 2007, Presgraves *et al.* 2007). A constant arms race between genes causing meiotic drive and those compensating for it results in their fast evolution. Such genes show signs of positive selection that may be due to selective sweep rather than adaptive evolution (Presgraves 2007). Interestingly, *Dntf-2r* is located close to *Segregation distorter* (*Sd*) gene - a truncated duplicate of *RanGap* which is also involved in nuclear transport. *Sd* has been known to cause meiotic drive where ~99% of the progeny receive the chromosome carrying *Sd*. The close proximity of *Dntf-2r* to *Sd* further strengthens the hypothesis of its involvement in meiotic drive (Presgraves 2007). We study the

function of *Dntf-2r* from two view points in chapters 4 and 5: 1) function during spermatogenesis and 2) possible role in meiotic drive. We explore its possible role in meiotic drive using the *SD* system. Genes involved in nuclear transport may also be involved in either facilitating or counteraction transport of certain viral genomes and TE's (Presgraves *et al.* 2007). Involvement of *Dntf-2r* is such pathogenic attacks against the genome will not be directly addressed in this study, but it remains an additional possibility.

1.2 *Dntf-2*: Background

Dntf-2 is the drosophila homolog for *Nuclear transport factor 2*. *NTF-2* (*Nuclear Transport Factor 2*) plays an indirect role in the transport of proteins containing nuclear localization signal (NLS) across the nuclear membrane by setting up the RanGDP/Ran GTP gradient. NTF-2 is a dimer that interacts with two and possibly three molecules of Ran bound to GDP (Quimby 2000). It also interacts with the FxFG repeats of nuclear pore complex (Isgro 2007). *Ran* is a member of Ras superfamily and exists in GDP bound inactive state in the cytoplasm, and GTP bound active state in the nucleus. The RanGDP predominantly localizes in the cytoplasm and RanGTP in the nucleus. The RanGDP-RanGTP gradient is important for the directionality of import and export of cargo proteins across the nuclear membrane. In the cytoplasm, NTF-2 interacts with Ran-GDP molecules and carries them into the nucleus via the nuclear pore where NTF-2 loosely binds to the FxFG repeats of the nuclear pore complex. Once into the nucleus, a chromatin associated protein RCC1 converts RanGDP to RanGTP. RanGTP binds to importinβ and induces conformational changes that lead to the dissociation of importinα/β heterodimer and release of the cargo protein. These cargo proteins mostly comprise of proteins that are involved in the spindle assembly during mitosis (Ciciarello *et al.* 2007). RanGTP ensures the release of cargo proteins in precise special and temporal pattern for the proper orchestration of downstream functions. Ran-GTP, bound to importin β is transported out of the nucleus. In the cytoplasm, another catalytic enzyme - RanGAP hydrolyses Ran-GTP to Ran-GDP (figure 1.1). The crystal structure of rat NTF2 and canine Ran and their interaction with each other as well as other proteins is available in the Protein database website

(http://www.rcsb.org/pdb/home/home.do)(Stewart *et al.* 1998, Bullock *et al.* 1996, Renault *et al.* 2001, Vetter *et al.* 1999; Seewald *et al.* 2002). In chapter 4, we compare these structures with proteins sequence from the *Drosophila* retrogenes and predict their potential function.

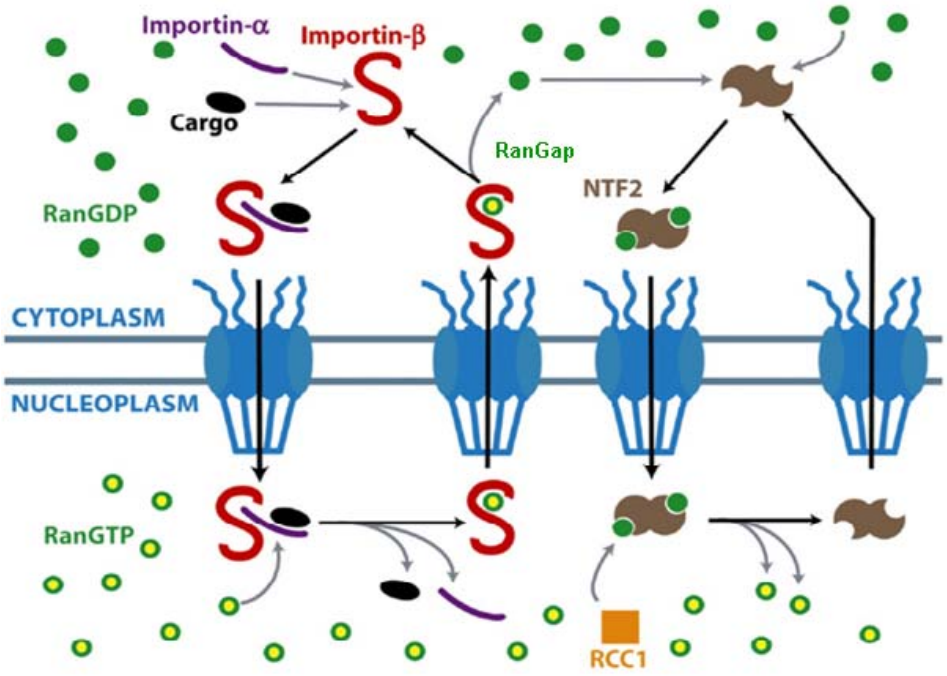


Figure 1.1. Nuclear Transport. Representation of protein transport across the nuclear membrane. Modified from Isgro *et al*. (Isgro and Schulten 2007).

Interestingly, it has recently been revealed that NTF-2-Ran complex is actively involved in the import of filamentous actin capping protein CapG (Katrien Van Impe *et al.* 2008). CapG belongs to the gelsolin superfamily which constitutes of proteins that control actin organization by severing filaments, capping filament ends and nucleating actin assembly (Silacci *et al.* 2004). CapG binds directly to NTF-2, however the interaction between Ran and NTF-2 is essential for its transport. Mutants of NTF-2 and Ran that inhibit interaction with each other also inhibit the transport of CapG (Katrien Van Impe *et al.* 2008). It is unknown if they also participate in the transport of other proteins or ribonucleoproteins.

Recently, an overexpression of NTF-2 has been implicated in protection against diabetic retinopathy (DR; (Li *et al.* 2009). DR and consequent vision loss or impairment is a common complication in patients with Type 2 Diabetes mellitus. Diabetes mellitus patients with DR had lower levels of NTF-2 than the ones without DR. From mice studies it was observed that dose of Ntf-2 inversely affected the expression of vascular endothelial growth factor (VEGF). It is suggested that NTF-2 may provide resistance against diabetic retinopathy by lowering the expression of VEGF (Li *et al.* 2009) however; the exact mechanism is still unknown.

Dntf-2 has also been implicated in normal eye development (Bhattacharya *et al.* 2002) in drosophila. In *Drosophila*, *Dntf-2* mutants are lethal. However, some hypomorphs are viable and fertile and show impaired eye phenotype where the number of omatidia in the eye is severely reduced. This mutant phenotype is rescued by mutation in chickadee that encodes Drosophila Profilin (Minakhina *et al.* 2005). Profilin is an actin-binding protein involved in the dynamic turnover and restructuring of the actin cytoskeleton. Interestingly, similar suppression of the mutant eye phenotype is caused by *Sd* mutants (Minakhina *et al.* 2005). *Sd* is a truncated duplicate of RanGAP that functions in maintaining the Ran-GDP Ran-GTP gradient (see above). Minakhina *et al.* suggested that as the mutant phenotype can be suppressed by over expression of RanGAP, some abnormalities in nuclear trafficking could be the apparent cause of the rescue (Minakhina *et al.* 2005).

The immune response in the larvae is also affected by the partial loss of *Dntf-2*. Normally, upon infection three NF-κB/Rel proteins (Dorsal, Dif and Relish) target the nucleus of the larval fat bodies. In the hypomorphs, the NF-κB/Rel proteins do not enter the nucleus and this in turn impairs the expression of anti microbial peptide genes (Bhattacharya *et al.* 2002). Some NF-κB/Rel proteins are believed to be transported in a complex and may interact with NTF-2-RanGDP complex at the nuclear pore but others may be transported via the importin complex (Bhattacharya *et al.* 2002).

1.3 Recurrent recruitment of *Dntf-2* and *Ran*

Besides *D. melanogaster* complex, *Dntf-2* has given rise to retrogenes independently in the *D. ananassae* and *D. grimshawi* lineages (Bai *et al.* 2007). Interestingly, *Ran* has also given rise to retrogenes three times in overlapping lineages. Once in the *Drosophila* subgroup (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. mauritiana*, *D. yakuba* and *D. erecta*) where the retrogene is named *Ran-like* (Tracy *et al.* - submitted) and additionally, it has independently given rise to retrogenes in *D. grimshawi* and *D. ananassae* lineage (Bai *et al.* 2007). All the retroposition events involve X to autosome movement of retrogenes. As previously stated, *Ran* and *Dntf-2* encode proteins that interact with each other during nuclear transport. Their evolving a duplicate copy in overlapping lineages could suggest an adaptive role especially if they have overlapping expression. In *D. melanogaster* both *Dntf-2r* and *Ran-like* are highly expressed in male germline (Betrán *et al.* 2003; Chintapalli *et al.* 2007). In *D. ananassae*, both retrogenes show high expression in male germline and lower expression in females where as both parents are ubiquitously expressed. In *D. grimshawi* both retrogenes and parent genes are expressed in both males and females (Tracy *et al.* - submitted). However, *Dntf-2* retrogene is expressed much higher in males and *Ran* retrogene is expressed higher in females (Tracy *et al.* - submitted).

Tracy *et al.* also observed that in all the lineages both the retrogenes are evolving at a faster rate than the parent genes. This can be due to relaxation of constrain or positive selection acting on the duplicates. As commented above, McDonald-Kreitman test using polymorphism data from *D. melanogaster* and *D. simulans* provided evidence that *Dntf-2r* is evolving under positive selection (Betrán *et al.* 2003). Tracy *et al.* (Tracy *et al.*- submitted) revealed that *Ran-like* is also evolving under positive selection and often changing so much that is likely to have lost some parental functions (some of them trough deletions). In addition in *D. yakuba*, *Ran-like* has accumulated many out of frame deletions in the coding region and it is likely a pseudogene (Tracy *et al.* - submitted). It has been suggested that convergent recruitment of X to autosome duplicates by two interacting genes could be due to a strong adaptive advantage (Bai *et al.*

2007). However, this scenario does not support the loss of function of *Ran-like* in one of the lineage where, interestingly, *Dntf-2r* is not present. A hypothesis that we would explore here for *Dntf-2r* is the involvement of both retrogenes in meiotic drive. This was introduced above and may lead to a constant arms race that could explain the recurrent positive selection acting on these retrogenes. It can also explain the loss of functions and genes when meiotic drive role is not needed or the meiotic drive system has been fixed or lost.

1.4 Goals

This study aims to identify the cis-regulatory region as well as the function of the young retrogene *Dntf-2r*. A concise list of all the issues I wish to address follows:

1. The expression pattern of *Dntf-2r* in various strains of *D. melanogaster*, as well as in *D. simulans*, *D. sechellia* and *D. mauritiana*. This is important, as it will indicate if the same pattern of expression exists for the gene arguing for a high likelihood that the same regulatory region is utilized by the retrogene in various species.
2. Identify the transcription start site (TSS) in *D. simulans* to position where we expect the regulatory regions to be and compare with the known TSS of *D. melanogaster*.
3. Experimentally identify the cis-regulatory region of *Dntf-2r* in *D. melanogaster* required for testis specific expression and make inferences about its nature and origin. This will be done by making P element constructs that include the putative regulatory regions and *Dntf-2r* fused to EGFP and studying the fluorescence.
4. Study the transcription pattern of wild type *Dntf-2r* using in situ hybridization technique. A comparison between wild type expression and expression of EGFP tagged fusion protein will indicate if the identified regulatory region can recapitulate the original regulatory region.
5. Analyze the affect on fertility of P element knock off of *Dntf-2r* using a very sensitive assay (i.e. male sperm exhaustion assay).
6. Compare the amino acid sequence of *Dntf-2r* and *Ran-like* to the crystal structures of their mammalian homolog to analyze if the retrogene encoded proteins have lost or maintain any functional domain and to predict the functional capabilities of these proteins.

7. Test if, like its parent, *Dntf-2r* protein can form a dimer and interact with Ran using western blot analysis.

8. Explore the potential effects of *Dntf-2r* in meiotic drive using the *SD* system currently present in *D. melanogaster*.

CHAPTER 2

IDENTIFICATION AND CHARACTERIZATION OF THE NOVEL CIS-REGULATORY REGION OF *Dntf-2r*

As introduced in Chapter 1, retrogenes are processed copies of genes that originate through reverse transcription of a parental gene and reinsertion into the genome. Most retrogenes can be identified in a genome as intronless genes and may be flanked by direct repeats and/or have a remnant of poly A tail in the case of recent events of retropositions. While retrogenes have a parental homolog that may or may not have introns, most identified retrogenes are copies of intron containing genes. This is because it is more difficult to identify retrogenes that originated from intronless genes, as they will have to contain additional features (direct repeats or poly-A tail). Retrogenes can also acquire introns trough gene fusion or recruitment of flanking regions into genes (Long *et al.* 1993; Vinckenbosch, Dupanloup *et al.* 2006; Wang *et al.* 2006; Bai *et al.* 2007). In Drosophila, there has been an excess of retrogenes that have originated from genes on the X chromosome and are inserted in an autosome (Betran *et al.* 2002; Dai *et al.* 2006; Bai *et al.* 2007; Vibranovski *et al.*2009; Meisel *et al.* unpublished results). In addition, many retrogenes have evolved testis biased expression (Betran *et al.* 2002; Bai *et al.* 2007). Similar patterns have been observed in human and mouse genome (Betrán *et al.* 2004; Emerson *et al.* 2004; Potrzebowski *et al.* 2008).

At the time of retroposition, all retroposed copies essentially lack a promoter, except for the copies that are formed from an aberrant transcript of the parent which may harbor the promoter (McCarrey 1994; Zhang *et al.* 1998; Kleene *et al.* 1999). The window of opportunity for a retrogene to recruit a promoter might be small as an untranscribed region will accumulate deleterious mutations which can disrupt the intact coding region or be lost by deletion. Thus, it is important to either land downstream of an existing promoter, or evolve a simple promoter

requiring a few substitutions in order to be transcribed. It is also possible that recruitment of transcription could be facilitated by inserting close to transcribed genes or in the case of testis expressed retrogenes in testis neighborhoods (Kalmykova *et al.* 2005) and an excess of these type of retrogenes has been found (Loppin *et al.* 2005; Vinckenbosch *et al.* 2006; Bai *et al.* 2008; Dorus *et al.* 2008). Although the mechanism/s that would facilitate transcription in those cases is/are unknown, transcriptional coregulation has been proposed as a mechanism (Loppin *et al.* 2005; Vinckenbosch *et al.* 2006) and has been described in *Drosophila*. However, we argue that just being in an open chromatin domain or close to a gene does not ensure the initiation of transcription of a gene. It would still require a cis-regulatory region for the binding of the polymerase and likely tissue specific transcription factors. As most of the X to autosome retrogenes have evolved testis specific expression (Betran *et al.* 2002; Emerson *et al.* 2004; Vinckenbosch *et al.* 2006; Bai *et al.* 2007), we postulate that initially the tissue specific promoter should be, in addition, simple and easy to recruit. With time, elaborate regulatory mechanisms like enhancer elements and epigenetic regulation can evolve as in case of *Pgk2* (Zhang *et al.* 1999; McCarrey *et al.* 2005). Thus, it is important to study the regulatory region of young retrogenes which will lead us to understand the origin of cis-regulatory regions that can drive testis-specific expression.

Interestingly, there are different lines of evidence that directly or indirectly support small/easy to evolve cis-regulatory regions in male germline genes. Spermatogenesis is a complex process that gives rise to mature sperms for fertilization and it is known that such rigorous cell differentiation requires a major reprogramming at the epigenetic as well as genetic level (Sassone-Corsi 2002). The gene regulatory mechanisms are different from the somatic cells, where the general transcription factors are either differentially regulated or have a testis-specific isoform (Sassone-Corsi 2002; Hiller *et al.* 2004). In spite of such drastic changes in regulatory mechanism, many testis specific genes have a relatively simple and short cis-regulatory regions close to the transcription start site (TSS; (Kuhn *et al.* 1988; Michiels *et al.* 1989; Robinson *et al.* 1989; Schulz *et al.* 1990; Yanicostas *et al.* 1990; Kuhn *et al.* 1991;

Gigliotti *et al.* 1997; Nurminsky *et al.* 1998a; Blumer *et al.* 2002; Reddi *et al.* 2002; Han *et al.* 2004; Somboonthum *et al.* 2005; Di Cara *et al.* 2006; Hempel *et al.* 2006; Hense *et al.* 2007). So the first line of evidence is the data on the known testis-specific regulatory regions. An additional and indirect line of evidence comes from the fact that testis expression pattern seems to have very high turnover and that might be a consequence of the requirement of simple sequences for that type of expression but see additional discussion by Zhang *et al.*

Here, we describe the testis-specific expression pattern of *Dntf-2r* in most species where it is present and in a variety of *D. melanogaster* strains, describe its TSS in the species where it is present and experimentally identify the short (i.e. 14 bp) cis-regulatory region needed for testis expression of this gene. Given the relatively small similarity of this sequence and the β2-tubulin testis promoter we infer a low specificity of the transcription factors that bind to this sequence. Through comparative genomics analyses we confirm that few changes to the original sequence were needed to produce the regulatory region of *Dntf-2r* after its insertion as previously suggested.

2.1 Material and Methods

2.1.1 Drosophila stocks and fly handling

The Besançon wild type strain was used for Drosophila melanogaster. Other strains used for *D. melanogaster* were Cantonese (CS2), rinanga, strains from Seattle and two strains from Ecuadore EC151 and EC152. Other species used were *D. simulans*, *D. sechellia* and *D. mauritiana*. White mutant strains w118, and w1118 and balancer stocks CyO (for second chromosome (w;Sco/Cyo,s)) and Tm6b (for third chromosome (w;2.3/Tm6b;sb)) were used to fix the P element insertions in transformed flies. All stocks were maintained on corn media at 25^{o}C with 12hr day and night periods.

2.1.2 Genomic DNA extraction

DNA was extracted from 15-20 adult flies using the Puregene DNA purification system – cell and tissue kit from Gentra systems.

2.1.3 RNA extraction

The Qiagen RNeasy mini kit was used for RNA extraction from about 15 adult males and virgin females. When extracting from fly halves ~ 35 flies were used. RNA was extracted from fly halves of *D. melanogaster* CS2 and Besançon and *D. simulans*. For higher concentration of RNA (as needed for 5' RACE) ~ 45 whole flies were used. For analyzing testis specific expression RNA was extracted from testis, accessory glands gonadectomised bodies dissected from 100 males from *D. melanogaster* and *D. simulans*. The protocol provided by the manufacturer was followed.

2.1.4 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) and Ssuperscript II reverse transcriptase (Invitrogen). From the cDNA obtained, PCR was performed using the primers Dntf2r_3'RACE1 – 5' 'TTTGTCCAGCAGTACTACGC' and Dntf2r_nestedGSP – 5' 'GGGGGATCGTCATCGCATTT'. These primers worked in all the four species. For *D. melanogaster* another primer Dntf2r_GSP1 – 5'AGCCACGAAGAGGGATCCTC 3' was used along with Dntf2r_3'RACE1.

2.1.5 5'RACE

RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE) was performed to identify the transcription start site for *Dntf-2r* in *D. simulans* using First choice RLM-RACE kit from Ambion, Inc. One µg of RNA extracted from adult males and virgin females was used. Gene specific outer primer CG10174 5'RACE2 - 5' CCGTTGGGCTTCAGCAAAAAGAT 3' and the inner primer 5'RACE1 - 5'CATCGCATTTTAGTCTTCCAAGGACG 3' were used an long with the 5'RACE outer and inner primers provided by the manufacturer respectively. The transcription start site in *D. melanogaster* was known (Betran and Long 2003).

2.1.6 Preparation of the construct

Dntf-2r along with 157 bp upstream region was amplified from genomic DNA using the primers Dntf2r_Xho15' – 'CTCGAGTCTCTTCGCGCCTATCGATG' and Dntf2r_sacII3' -

'CCGCGGAGAGTTGTGGATGTTCAGAC' and TA cloned using TOPO TA cloning kit (Invitrogen). These primers added XhoI and SacII restriction sites to the PCR product for directional cloning. Transformed cells were plated on LB agar plate with ampicillin and grown overnight. Next day colonies were PCR screened and cells with the expected size of PCR product were grown over night. Plasmid was extracted (QIAprep spin miniprep kit) from the overnight grown cultures. A good clone was identified by sequencing and XhoI and SacII cassette was extracted and inserted in XhoI and SacII site of plasmid pEGFP1 (U55761; Clontech) which put the *Dntf-2r* in frame with the EGFP gene. Again a good clone was identified by PCR screening and sequencing. The clone now contained the putative promoter region, *Dntf-2r* - EGFP fusion gene and SV40 poly A site. For transformation we cloned this region in the P element transformation vector - pCaSpeR 4 (Thummel *et al.* 1991). As no common restriction site was found between the plasmid EGFP clone and pCaSpeR 4, they were first digested with AflII and KpnI respectively and then blunt ended using mung bean exonuclease. Both plasmids were then digested with XhoI and ligated overnight. The next day the ligation mix was used to transform XL-blue super competent cells (Stratagene) and plated on ampicillin plates. Colonies were screened using PCR and plasmid was extracted from overnight grown cultures. A clone was considered good when it produced expected band length after digestion and had the correct boundaries. Such a clone was used for transformation and sent for transformation to Genetic Services Inc. (Figure 2.1). Several independent transformant were obtained.

Once EGFP (i.e. fluorescence) was detected in testis of the transformants from the first clone subsequent clones were made using DNA extracted from transformed flies. Each clone used the same primer at the 3' end - Kpn1-EGFP-3' 'GGTACCAACCACAACTAGAATGC'. For the 5' end of the fusion gene different primers were used, each making the putative promoter region shorter and shorter. These were XhoI5'-2 CTCGAGGGGACCAAAACGTTCAAATC, XhoI5'-A' CTCGAGTATCAGCTTAGCGGTGA, Xho15'-3 CTCGAGGATATTTGAACAGTGCTAAG and XhoI5'-4 CTCGAGTTTGCTTTTTTTCG GATCGG.

These primers introduced XhoI and KpnI restriction sites to the PCR product that comprise of putative promoter region, *Dntf-2r* fused to EGFP gene and SV 40 poly A site. The PCR product was cloned using the TA TOPO clonig vector 2.1 (Invitrogen) to increase the copy number. The TA clone was then digested using XhoI and KpnI restriction enzymes. The band of interest was gel extracted and ligated into XhoI – KpnI site of pCaSpeR 4 plasmid vector. One good clone was chosen by sequencing and sent for transformation to Genetic Services Inc.

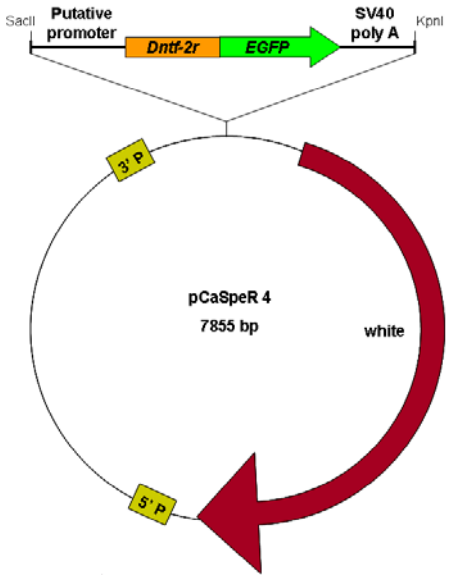


Figure 2.1. The pCaSpeR 4 fusion construct with the insert containing the putative promoter, *Dntf-2r* coding region fused in frame to EGFP coding region, and SV40 poly A termination signal.

2.1.7 Injection of the plasmid

Plasmids were sent for injection to Genetic Services Inc. The stock used for injection is white mutants (w118). A helper plasmid containing transposase gene is injected along with the desired plasmid.

2.1.8 P element fixation in the transformants

Injected larvae were obtained from the company. These were allowed to grow at room temperature. After individuals eclosed they were crossed with virgin flies from white mutant w118 provided by the Genetic Services Inc. Any progeny with orange eye color indicates a

successful insertion of the plasmid. These flies were crossed once again with w118. To identify the chromosome of insertion, the progeny were crossed with balancers. Each cross contained one transformant male with 2-3 balancer virgin females or one transformant virgin female with 2 balancer males. The balancer phenotype for 2nd chromosome is curly wings and for 3rd chromosome is short chaeta. 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 arise from different individuals or if they arise from the same individual but map to separate chromosomes.

2.2 Results

2.2.1 Dntf-2r pattern of expression

Dntf-2r is a young retrogene (i.e. less than 12 My old) present only in *D. melanogaster*, *D. simulans*, *D. sechellia* and *D. mauritiana* (Betran *et al.* 2003). Testis biased expression of *Dntf-2r* is known in *D. melanogaster* (Betran *et al.* 2003; Chintapalli *et al.* 2007). We studied the expression pattern in *D. simulans*, *D. sechellia* and *D. mauritiana* using RT-PCR. Figure 2.2 – 2.4 shows the results. In all the species *Dntf-2r* has male biased expression (Figure 2.2). When we performed RT-PCR using RNA from *D. simulans* testis, we observed testis biased expression (Figure 2.3). We also studied expression in various strains of *D. melanogaster* (Figure 2.4). In all the strains and species, a male biased expression is observed whereas the parent *Dntf-2* is ubiquitously expressed. Sometimes lower expression is also observed in females. The females used for RNA extraction for those experiments were old and may show leaky expression. No such expression was observed in younger females (this study and Betrán *et al.* 2003).

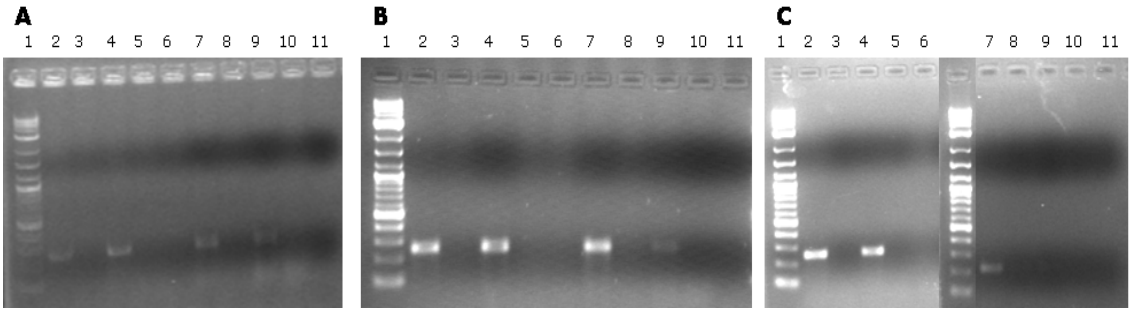


Figure 2.2 *Dntf-2r* expression in various *Drosophila species.* A) *D. simulans*; B) *D. sechellia*; C) *D. mauritiana*. In some cases a leaky expression in older females is observed. Lane 1 – ladder; Lane 2 – *Dntf-2* male RT+; Lane 3 – *Dntf-2* male RT-; Lane 4 – *Dntf-2* female RT+; Lane 5 – Dnt-f2 female RT-; Lane 6 – *Dntf-2* PCR-; Lane 7 – *Dntf-2r* male RT+ ; Lane 8 – *Dntf-2r* male RT-; Lane 9 – *Dntf-2r* female RT+; Lane 10 – *Dntf-2r* female RT-; Lane 11 – *Dntf-2r* PCR-

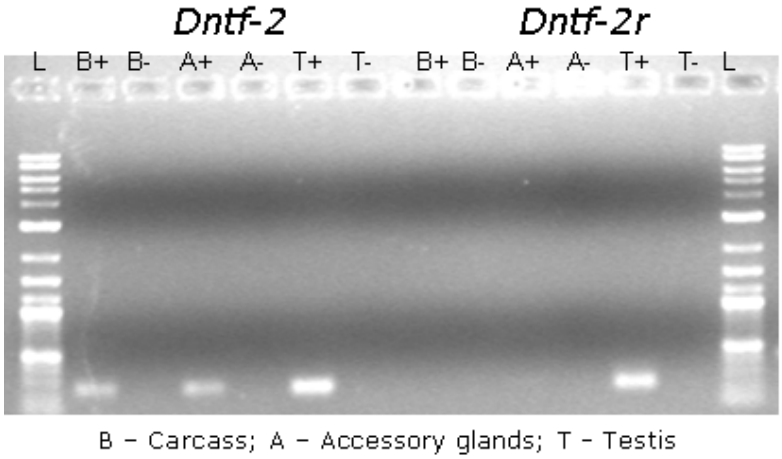


Figure 2.3. Tissue specific expression in *D. simulans.* Expression of *Dntf-2* and *Dntf-2r* in gonadectomised males, accessory glands and testis. (+ denotes RT positive and – denotes RT negative)

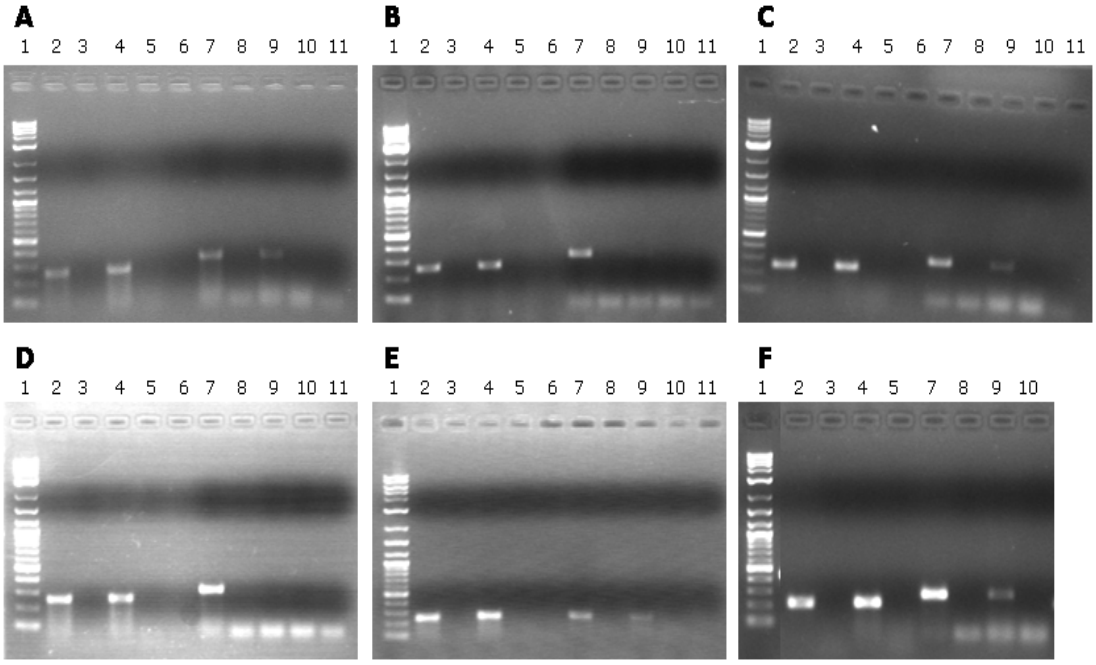


Figure 2.4. Expression pattern of *Dntf-2r* in whole males and females in various strains of *D. melanogaster*. A) CS2; B) EC151; C) Seattle; D) Besançon; E) EC152; F) Rinanga. Lane 1 – ladder; Lane 2 – *Dntf-2* male RT+; Lane 3 – *Dntf-2* male RT-; Lane 4 – *Dntf-2* female RT+; Lane 5 – Dnt-f2 female RT-; Lane 6 – *Dntf-2* PCR-; Lane 7 – *Dntf-2r* male RT+ ; Lane 8 – *Dntf-2r* male RT-; Lane 9 – *Dntf-2r* female RT+; Lane 10 – *Dntf-2r* female RT-; Lane 11 – *Dntf-2r* PCR-

As observed from figure 2.2 – 2.4, *Dntf-2r* is male biased in all the species and strains suggesting that a single promoter is involved. On comparing the 5' flanking region of all species where *Dntf-2r* is present and many strains of *D. melanogaster*, a β2UE like sequence is observed which is highly conserved (see below). We show that this conserved region drives the testis specific expression of fusion protein in *D. melanogaster* and, based on expression analysis, also in other species.

2.2.2 Transcription start sites

A description of the TSS of a gene is needed to adequately predict the putative location of promoter and other cis-regulatory elements. TSS for *Dntf-2r* in *D. melanogaster* is known (Betrán *et al.* 2003). We performed 5'RACE to identify the transcription start site in *D. simulans*. The TSS in both species differs by a five base pairs (Figure 2.5). The fact that the retrogene in

these two species show similar TSS suggests again that a common cis-regulatory region may be driving the expression in all the species.

```
D.simulans          CGAATGTATCAGCTTAGCGGTGACACAAACGAGAGGATATTTGA----------GTCTG 49
D.melanogaster      CAAATGTATCAGCTTAGCGGTGACACAAACGAGAGGATATTTGAACAGTGCCTAAGCCTG 60
                    * ****************************************          * ***
D.simulans          TCAGCT-AACCGATTTCATTTGCATTTGTTTTCGTATTGGGATTCCCCAAAAATCTAAGC 108
D.melanogaster      TCAGCTTAACCGATTTAATTTGCTTTT--TTTCGGATCGG---------AAACTCTAATC 109
                    ****** ********* ****** ***  ***** ** **         *** ***** *
D.simulans          GTACC--GCATATACTTCAACTGAAATG 134
D.melanogaster      GTACCCCGTATATACTTGAACTAAAATG 137
                    *****  * ******** **** *****
```

Figure 2.5. Alignment of 5' flanking region of *Dntf-2r* in *D. simulans* and *D. melanogaster*. Transcription start site in *D. melanogaster* is from Betrán *et al.* (Betran *et al.* 2003). Blue letters in italics show the mRNA and boldface letters show the translational start site.

2.2.3 Identification the regulatory region of Dntf-2r

To identify the regulatory region of *Dntf-2r* in *D. melanogaster*, clones carrying the region upstream of TSS and *Dntf-2r* fused to EGFP as reporter gene were transformed in *Drosophila* using P element technology (see Materials and Methods). The first transformed construct was the longest and consisted of 157bp upstream region from the TSS (i.e the region up till the adjacent gene *Bicoid stability factor* (*Bsf*)). The transformed flies expressed fluorescence in testis suggesting that this region harbored the complete testis specific regulatory region of *Dntf-2r* (Figure 2.6). Consequently, we generated constructs carrying -107 bp, -78 bp, -54 bp, -9bp upstream of the TSS (Figure 2.6). As seen in the figure (Figure 2.6) the first three clones show testis specific fluorescence. The latter two did not show any fluorescence and were comparable to the auto fluorescence of the white mutant control. All independent insertion lines showed similar result indicating that the clone itself, and not the region of insertion of the clone, was responsible for the signal. As is clear from figures 2.6, a 27 bp region is responsible for driving the expression of the *Dntf-2r* - EGFP fusion gene. The testis specific nature of this promoter was confirmed by the lack of fluorescence in various other tissues and developmental stages.

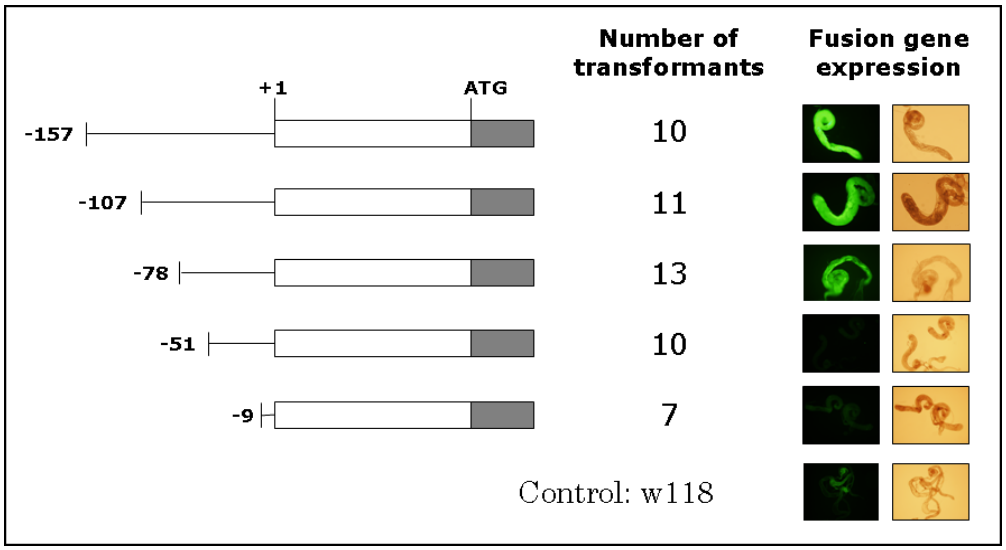


Figure 2.6. Constructs used for transformation. Fluorescence observed from the testis specific expression of *Dntf-2r*-EGFP fusion gene. All construct except the last two show strong fluorescence.

This is the regions that had been predicted to drive testis expression in this gene (Betran *et al.* 2003) because it harbors a sequence similar (57% identity) to the β2 tubulin upstream element (UE) promoter motif (Figure 2.7) that was shown to be necessary and sufficient to drive testis expression (Michiels *et al.* 1989). Our analyses therefore experimentally confirm this prediction. The upstream region of *Dntf-2r* harbors in addition a sequence that is identical (Figure 2.7) to the 7 bp quantitative element of the *β2 tubulin* gene. This promoter element has also been described in the gene *Sdic* (64% identity). The length variation in this regulatory motif and low identity reveals a low specificity of the transcription factor/s that bind the motif. Additionally, the separation between the upstream element and the quantitative element in *Dntf-2r* is larger (8 bp more) than in the *β2 tubulin* gene. A great separation (42bp) between these two elements has been shown to prevent expression likely due to the importance of the position of the element within the promoter (Michiels *et al.* 1989). β2UE1 was described to be a directional and position-dependent promoter (located at -38) that had the ability to recruit polymerase II and also drive testis specific expression. A longer distance from the TSS abolishes the expression of the reporter gene (Michiels *et al.* 1989). This element is in

the same orientation and similar distance (-41) in *Dntf-2r*. The same two motifs are conserved in *D. simulans*, *D. sechellia* and *D. mauritiana* (Figure 2.7). However, the position of the motif in *D. simulans* (where we have data of the TSS) is closer to the TSS (i.e. -24). Despite this length difference we still hypothesize that this element will be used in *D. simulans* and in *D. sechellia* and *D. mauritiana* as these are similar to that for β2 tubulin gene.

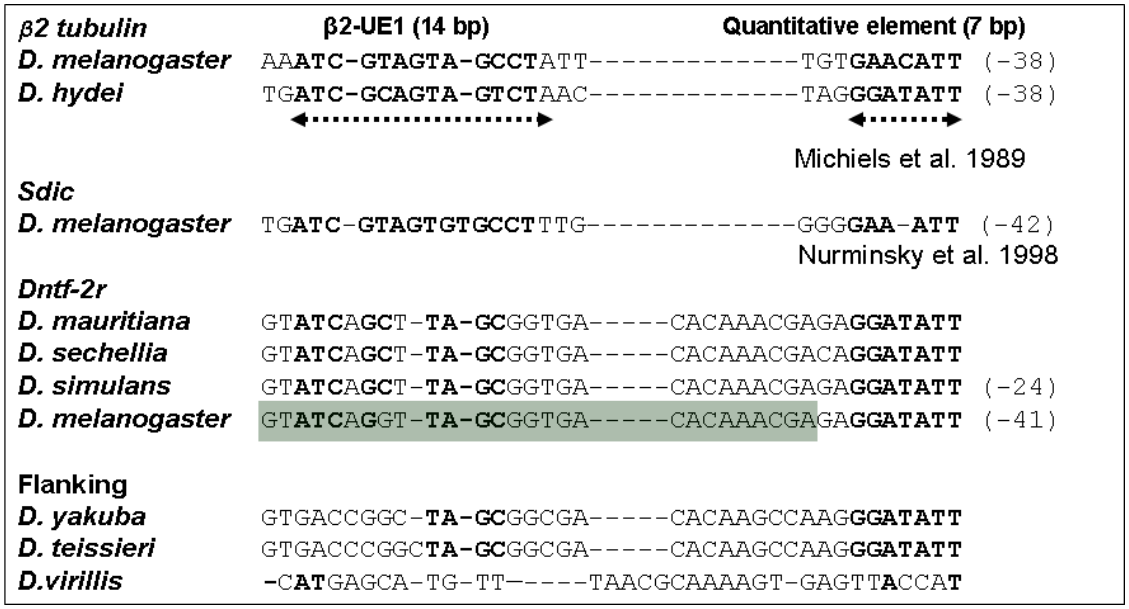


Figure 2.7. Similarity of the upstream region of *Dntf-2r* to the known β2-UE1 promoter and quantitative element. Related sequence were observed in syntenic region in distantly related species suggesting that only a few substitutions were required to evolve the tissue specific promoter. Highlighted region shows the sequence required for testis specific expression.

2.2.4 The regulatory motif is conserved in D. melanogaster and related species

Unpublished data of sequences from 5' flanking region of *Dntf-2r* from 24 *D. melanogaster* alleles, one *D. simulans* and *D. sechellia* allele and two *D. mauritiana* alleles from previous study from Dr. Betrán was used to study the conservation of the regulatory region (Figure 2.8). The β2UE1 promoter element and the quantitative element are conserved among all the alleles of the various species studied. A gap of 11 bp is observed in *D. simulans* as well as *D. sechellia*. Also a gap of 6 bp is observed in 9 alleles of *D. melanogaster*. These gaps

bring the regulatory region closer to the TSS. As discussed earlier the β2UE1 promoter functions in a position dependent manner. A longer than 59 bp region between the promoter and the TSS has been known to prevent transcription (Michiels *et al.* 1989). However, a variation in the position of the promoter element is observed and less than -42 bp from the TSS does not seem to hamper expression (Figure 2.7).

The expression analysis of *Dntf-2r* in various strains of *D. melanogaster*, *D. simulans*, *D. sechellia* and *D. mauritiana* demonstrate that possibly the same promoter is used for the male biased expression in all these species.

```
mau_1       CGAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAGCAGTGACTA-GCCTGTCAGCT-AACCGAT
mau_2       CGAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAGCAGTGACTA-GCCTGTCAGCT-AACCGAT
sech        CGAATGT**ATCAGCTTAGCGGT**GACACAAACGACA**GGATATTT**GAG----------TCTGTCAGCT-AAC-GAT
sim         CGAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAG----------TCTGTCAGCT-AACCGAT
yep18       CAAATGT**ATCAGGTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCCTAAGCCTGTCAGCTTAACCGAT
yep25_2     CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCCTAAGCCTGTCAGCTTAACCGAT
2s56_1      CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCCTAAGCCTGTCAGCTTAACCGAT
2s56_2      CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCCTAAGCCTGTCAGCTTAACCGAT
yep25_1     CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCCTAAGCCTGTCAGCTTAACCGAT
Rio_1       CAAATGT**ATCAGGTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCC------TGTCAGCTTAACCGAT
yep3_1      CAAATGT**ATCAGGTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCC------TGTCAGCTTAACCGAT
Cof3        CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCC------TGTCAGCTTAACCGAT
OK17        CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACATTGCC-AAGCCTGTCAGCTTAACCGAT
Rio_2       CAAATGT**ATCAGGTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCC------TGTCAGCTTAACCGAT
yep3_2      CAAATGT**ATCAGGTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCC------TGTCAGCTTAACCGAT
253.30      CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCC------TGTCAGCTTAACCGAT
Closs3      CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCC------TGTCAGCTTAACCGAT
y10         CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCC------TGTCAGCTTAACCGAT
HG84        CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCCTAAGCCTGTCAGCTTAACCGAT
Closs10     CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACATTGCC-AAGCCTGTCAGCTTAACCGAT
253.27      CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCC------TGTCAGCTTAACCGAT
Closs16     CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACATTGCC-AAGCCTGTCAGCTTAACCGAT
BLI5        CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACATTGCC-AAGCCTGTCAGCTTAACCGAT
Closs19     CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACATTGCC-AAGCCTGTCAGCTTAACCGAT
y2          CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCCTAAGCCTGTCAGCTTAACCGAT
Besancon_1  CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCCTAAGCCTGTCAGCTTAACCGAT
Besancon_2  CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCCTAAGCCTGTCAGCTTAACCGAT
Seattle     CAAATGT**ATCAGCTTAGCGGT**GACACAAACGAGA**GGATATTT**GAACAGTGCCTAAGCCTGTCAGCTTAACCGAT


mau_1       TTAATTTGCATTTGTTTTCGTATTGGGATTCCCCAAAAATCTAAGCGTACC--GCATATACTTCAACGGAAATG
mau_2       TTAATTTGCATTTGTTTTCGTATTGGGATTCCCCAAAAATCTAAGCGTACC--GCATATACTTCAACGGAAATG
sech        TTCATTTGCATTTGTTTTCGTACTGGGATTGACCAAAAATCTAAGCGCACA--GCATATGCTTCAATTGAAATG
sim         TTCATTTGCATTTGTTTTCGTATTGGGATTCCCCAAAAATCTAAGCGTACC--GCATATACTTCAACTGAAATG
yep18       TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
yep25_2     TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
2s56_1      TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
2s56_2      TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
yep25_1     TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
Rio_1       TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
yep3_1      TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
Cof3        TTAATTTGT--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
OK17        TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
Rio_2       TTAATTTGT--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
yep3_2      TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
253.30      TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
Closs3      TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
y10         TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
HG84        TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
Closs10     TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
253.27      TTAATTTGT--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
Closs16     TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
BLI5        TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
Closs19     TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
y2          TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
Besancon_1  TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
Besancon_2  TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
Seattle     TTAATTTGC--TTTTTTTCGGATCGG---------AAACTCTAATCGTACCCCGTATATACTTGAACTAAAATG
```

Figure 2.8 Alignment of the 5' flanking region of *Dntf-2r* in 24 *D. melanogaster* alleles, in one allele of *D. simulans*, in one allele of *D. sechellia* and two alleles of *D. mauritiana*. The first codon (ATG) of *Dntf-2r* is shown at the end of the alignment. The putative conserved nucleotides of the promoter and quantitative element are shown in bold.

2.2. 5 Expression of fusion protein during spermatogenesis

We used confocal microscopy to study the detailed expression pattern of the fusion protein during spermatogenesis. Fluorescence was observed in the 16 celled stage (primary spermatocytes), 32 celled stage (meiotic spermatocytes), in the 64 celled stage (round spermatids) and sperm heads of the elongating sperms (Figure 2.9). No expression is observed in the somatic stem cells (i.e. spermatogonia) located at the tip of the testes as well as in mature sperms. Such expression pattern is analogous to the expression pattern of β2tubulin gene and is the characteristic of the 14 bp β2UE1 promoter.

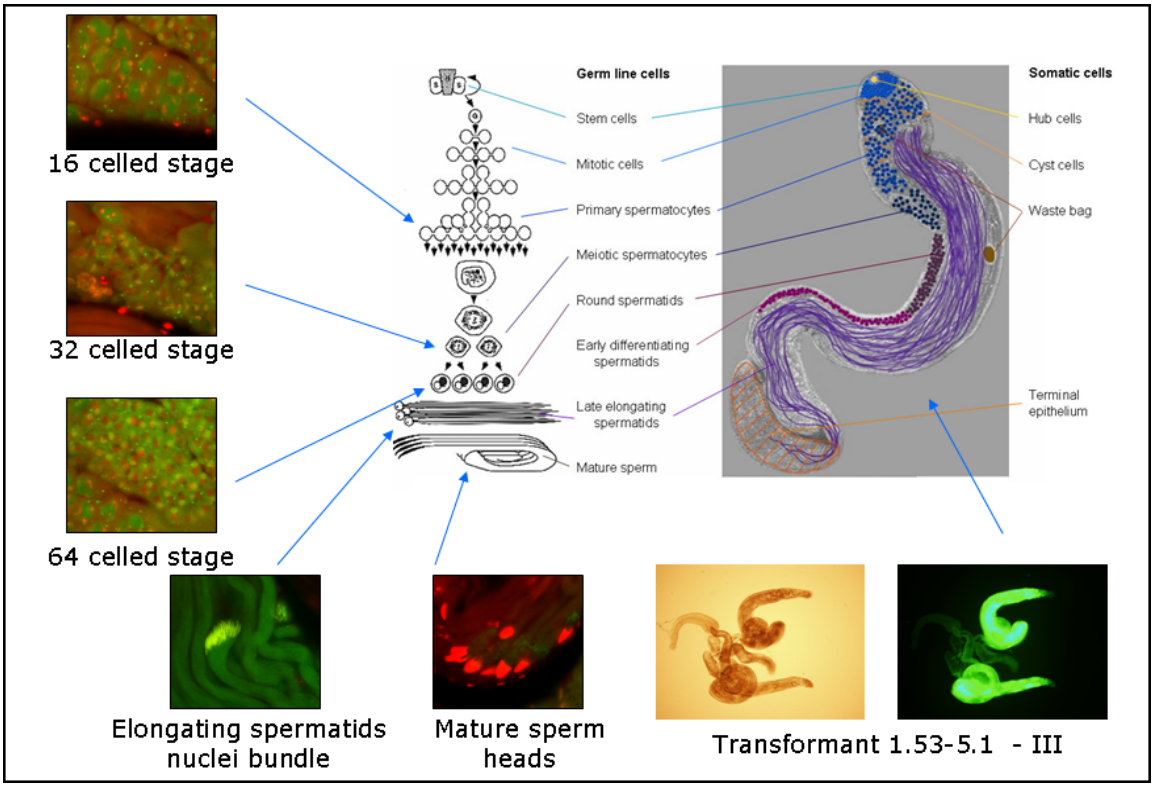


Figure 2.9. Expression of fusion protein in different stages of spermatogenesis. The DNA is stained with Draq5 and pseudocolored as red. Green color depicts the fluorescent protein.

2.3 Discussion

2.3.1 Quality of regulatory region of Dntf-2r

Consistent with the results of various transgenic experiments for many testis-specific genes (see chapter 3), we find that a very simple and short basal cis regulatory region is required for the expression of *Dntf-2r* in male germline. The element has 57% similarity with the β2UE1 promoter of the *β2 tubulin* gene in *D. melanogaster*. The 7 bp quantitative element of the *β2 tubulin* gene is also present which enhances the expression (Michiels *et al.* 1989) and deletion of this element does not prevent transcription. Besides the 14 bp core promoter and 7 bp quantitative element, an initiator element (Inr) was found in the upstream region of this promoter (Santel *et al.* 2000). Inr elements act as cis-regulatory element in vivo and have been shown to quantitatively regulate the tissue specific expression of reporter genes (Santel *et al.* 2000). However, deletion of this element does not affect the Inr mediated tissue specificity. We have no data on the presence of a initiator like element as found in the *β2 tubulin* gene.

2.3.2 Evolutionary origin

Dntf-2r originated from the processed mRNA of *Dntf-2* and thus did not contain promoter sequence from the parent (Betran *et al.* 2003). It instead recruited a novel 5' regulatory sequence from the insertion site requiring only a few substitutions. On examining the orthologus region of the insertion site from *D. yakuba*, *D. teissieri* and *D. virilis* (Figure 2.7), a similar element is detected in the close related species. Farther related *D. virilis* does not show much identity with the regulatory sequence. *Dntf-2r* inserted very close to another gene *Bicoid stabilizing factor* (*bsf*) in the head to head orientation. However, *bsf* does not share the expression pattern of *Dntf-2r* (Mancebo *et al.* 2001) (see introduction) and thus it is not likely that it may be using the same promoter as *Dntf-2r*.

Thus, it seems that *Dntf-2r* recruited a novel promoter sequence from a random genomic sequence which happened to be similar to the β2UE1 regulatory sequence. It fortuitously landed downstream of a sequence similar to the β2tubulin and a few substitution

lead to the formation of a novel testis specific regulatory sequence. In later chapters we discuss the evolutionary forces selecting for such a tissue specific expression.

2.3.3 Specificity and position of transcription factors binding β2UE1

The β2UE promoter has been observed to drive the expression of many testis-specific genes and on comparing the sequence of individual promoter we find that the 14bp promoter sequence and the quantitative element is not highly conserved (Michiels *et al.* 1989; Nurminsky *et al.* 1998). As discussed earlier it is clear that this promoter has position effect and a longer distance from the TSS will abolish expression of downstream genes. However, the position effect is not stringent and anywhere between -24 to -42 bp upstream location seems to be enough for its proper functioning.

A transcription factor, Modulo, has been identified as the protein binding the promoter region of *Sdic* (Mikhaylova *et al.* 2006). It binds to the promoter element of both *Sdic* and *β2Tubulin* genes. From figure 2.7 we observe that the promoter elements of *Sdic* shares 85% similarity with *β2Tubulin* promoter from *D. melanogaster* and 64% similarity with that from *D. hydei*. Elements from both *D. melanogaster* and *D. hydei* drive testis specific expression (Michiels *et al.* 1989). Thus, it seems that there is room for some mismatch in the identification of modulo target site. *Dntf-2r* promoter element is also very similar to that of *β2tubulin* and *Sdic*, and shares the same expression. It may be possible that modulo is the transcription factor that binds to this novel promoter and directs downstream expression. Studies analyzing the affect of modulo knockout on *Dntf-2r* expression are underway.

CHAPTER 3

EVOLUTIONARY ORIGIN OF CIS-REGULATORY REGIONS OF RETROGENES: COMPILATION OF STUDIES

In Chapter 1 I described that retrogenes are processed copies of genes reverse transcribed and reinserted randomly into the genome. The retrogenes contain intact coding region and thus are good candidates to form functional new genes. However, they usually lack a regulatory region necessary for transcription and the most common fate of such genes is to accumulate deleterious mutations and become functionless pseudogenes (Emerson *et al.* 2004). A process termed as nonfunctionalization (Vanin 1985). If by chance, a retrosequence inserts in a region downstream of an already existing promoter or a new promoter evolves in a short period of time; such genes can be transcribed and can become functional. Transcribed retropseudogenes (Zheng *et al.* 2005) reveal that transcription is a condition that is necessary but not sufficient to stay functional. The window of opportunity for a retrogene to recruit a promoter is very small as it is very easy to accumulate deleterious mutations which can disrupt the intact coding region. Despite having a very low chance of being transcribed, many genes have successfully recruited promoters and some have even evolved elaborate regulatory mechanism after time. Such genes are good candidates to study the effect of various evolutionary forces. Many have been studied and their mode of expression analyzed.

It has been observed that a big fraction of retrogenes show testis-specific or -biased expression (Betran *et al.* 2002; Marques *et al.* 2005; Vinckenbosch *et al.* 2006; Bai *et al.* 2007). The percentage is significantly higher than expected from the fraction of genes expressed specifically in testis for the whole genome. In addition, an excess of retrogenes are located on the autosomes and have the parental copy on the X chromosome and are testis specific.

There are therefore two main questions to be answered about retrogene regulation. How retrogenes are initially transcribed? What is the cause of the male germline bias in expression?

In this article we have tried to compile examples of retrogene regulatory region analyses to give an overview of the possibilities. We compile examples that have been studied in some detail or studies that address the particular questions of how retrogenes acquire regulatory regions and the evolutionary origins of the observed biases of expression in male germline to answer the two questions above. In most of the examples below, research is being performed in relatively young retrogenes to have enough sequence information about the particular DNA changes that have occurred since retrogene inserted. In some other work positional biases of retrogenes are studied, transcription of old and young retrogenes and/or transcription of retrogenes and pseudogenes are compared to gain some insides. This compilation reveals some trends in the events responsible for transcription and consequently endurance of the retrogenes found in the genomes.

3.1 First a surprise: a retrogene has replaced a parental that is now testis specific

Unlike many retrogenes (Betran *et al.* 2002; Marques *et al.* 2005; Vinckenbosch *et al.* 2006; Bai *et al.* 2007), Drosophila *e(y)2b* (enhancers of yellow 2b) poses a unique example where the retrogene is inserted on the X chromosome and has functionally replaced the parental gene. The parental gene is on chromosome 3R and, interestingly, has evolved a new male germline specific expression. *e(y)2b* gave rise to a retrogene *e(y)2* before the divergence of all the *Drosophila* species (Krasnov *et al.* 2005). The parent *e(y)2b* has male germline specific expression and the retrogene is ubiquitously expressed. However, the homologs of the *e(y)2b* in human as well as yeast are ubiquitously expressed, indicating that this was the ancestral expression (Krasnov *et al.* 2005). After the creation of the retrogene, the ubiquitous expression pattern was taken over by the duplicate copy and the parental gene developed male germline restricted expression. This example further strengthens the argument made in Chapter 1 that X chromosome is not a favorable location for male biased genes. Generally, out of the X

retrogenes evolve male specific expression. Here, X linked retrogene evolved a ubiquitous expression (including ubiquitous expression in testis) and the autosomal parent evolved male biased expression further supporting a strong selective force favoring the autosomal location of male biased genes. Genes having ubiquitous expression sometimes use a specific promoter for expression in spermatogenesis. In case of *e(y)2b* authors postulate that when the selective constraint was relaxed due to presence of another copy, it accumulated substitution that demolished the ubiquitous expression (Krasnov *et al.*. 2005) and maintained the testis specific expression. This is possible only if the retrogene landed downstream of a strong promoter that could induce high levels of expression (Krasnov *et al.* 2005). A program design to detect promoter that have a combination of known features (i.e. TATA, DPE, Inr and others) found strong signal of a promoter at -30 nt of the initiation of transcription of the retrogene (Krasnov *et al.* 2005). However, it is not know when and how this regulatory region originated. Additionally, the parental protein could not restore the phenotypic changes in a *e(y)2* mutant strain (Krasnov *et al.* 2005) indicating that the parent may have evolved to better adapt to male germline functions.

3.2 Chimeric retrogenes

Chimeras are formed when a retrogene is inserted in the non coding region of a functional gene or a duplicate copy of a gene and a fusion gene is created. In general, chimeras can be formed by illegitimate recombination or transposition (Arguello *et al.* 2006) tandem duplication and deletion (Ponce *et al.* 2006; Begun 1997) or via retroposition (Begun 1997; Long *et al.* 1999; Jones *et al.*. 2005). Here we will focus on formation of chimeras involving retrogenes. A high rate of the functional retrogenes has recruited coding and non-coding exons from the surrounding region to form a chimeric transcript in plant genomes (Wang *et al.* 2006), as well as in animal genomes (Buzdin *et al.* 2003). As discussed earlier, retrogenes can give rise to novel functions by instantly donating a novel protein domain to an existing or a duplicate gene. In both scenarios a preexisting regulatory region is used for retrogene transcription.

3.2.1 Insertion in a duplicate gene

Gene duplicates can create genes with novel function by the formation of chimeric genes, thus bringing two earlier independent functional domains together in one protein. A duplication event causes copying or movement to another location in the genome of a gene along with its regulatory region. Such duplicates might be dispensable in the genome and an insertion of a retrogene might not affect the organism as it has another copy to perform the ancestral function. Retrogenes inserted in such duplicates, generally form the 3' end of the chimera as they carry the polyadenylation signal. The non-retroposed part of the fusion, which may still have the regulatory sequence of its parent gene, drives the initial expression of the newly formed chimeric gene.

Jingwei (*Jgw*) is one such example of a chimeric gene formed by the fusion of a retroposed copy of *Alcohol dehydogenase* (*Adh*) with a duplicate copy of *Yellow emperor* (*Ymp*; (Long *et al.* 1999). *Adh* is a gene that encodes a protein required for the oxidation of ethanol. In Drosophila *Adh* is required for the digestion of ethanol coming from decaying fruit and thus is a highly expressed in the mid gut, fat body and female ovaries (Ashburner 1998; Betran *et al.* 2000). A retroposed copy of *Adh* landed in the duplicate copy of *Ymp* and captured three 5' exons to form the chimeric gene *Jgw*. It is ~2 Myr old and is present in *D. yakuba* and D. teisseiri (Long *et al.* 1993). In *D. melanogaster Ymp* is composed of twelve exons and produces two transcripts and has not known function. Both transcripts share the first three exons and have testis specific expression pattern (Wang *et al.* 2000). *Jgw* retains the expression pattern of *Ymp* in *D. teisseri* however shows ubiquitous expression in *D. yakuba*. The similar expression pattern in *D. teisseri* suggests that the chimeric gene uses the regulatory region of the duplicate copy of *Ymp*. In *D. yakuba* it is ubiquitously expressed with higher expression in males (Long *et al.* 1993) but the nucleotide changes that produce the change in expression in this lineage are not known.

Another chimeric gene formed by the retroposition of *Adh* is *Adh-twain* (*AdhT*; (Jones *et al.* 2005). It is a young gene that originated about 3 Mya in the common ancestor of *D.*

subobscura, *D. guanche* and *D. madeirensis*. The *Adh* retrosequence inserted within another gene *CG9010* which has *Gapdh* like function. *CG9010* is present in single copy in *D. melanogaster* and *D. pseudoobscura*, where as in *D. subobscura*, *D. guanche* and *D. madeirensis* it has a duplicate copy (Jones *et al.* 2005). It is not clear if the original gene or the duplicate copy forms a part of the fusion gene (Jones *et al.* 2005). The 5' UTR and upstream region of chimeric *AdhT* is very similar to that of *CG9010* and the retrogene contributes to the 3' end proving the 3' UTR and the polyadenylation signal. The expression pattern of the fusion gene is same as that of *CG9010* where both are expressed in male, female and larvae. A TATA-less promoter like region is highly conserved between *AdhT* and *CG9010* which may serve as a putative promoter (Jones *et al.* 2005).

3.2.2. Insertion in single copy genes

When inserted in a duplicate copy, a retrogene does not disrupt any existing function and are likely tolerated better by the organism, however, in case of single copy genes, retrogenes are generally inserted in the non coding region (introns, 3' UTR or 5' UTR) and often form alternative spliced transcripts. They can form alternatively spliced transcripts (Luo *et al.* 2006), or donate an extra exon to the host gene (exon shuffling; (Guldner *et al.* 1999).

The insertion of the retrogene can be on the sense strand or on the complimentary strand, in the same or opposite orientation of the host gene. If inserted on the sense strand and in the same orientation as the host gene, these genes can recruit the upstream exons of the host gene along with the host regulatory machinery to form a chimeric gene. See below.

3.2.2.1. Insertion in an intron

YY2 is a retrogenes on the X chromosome derived from the parent gene *YY1* present on chromosome 14. *YY2* is conserved among all placental mammals and both *YY1* and *YY2* are functional transcription factors and are involved in the expression and repression of many other genes (Nguyen *et al.* 2004). *YY2* is inserted in the fifth intron of *membrane-bound transcription factor protease site 2* (*Mbtps2*; (Luo *et al.* 2006). *Mbtps2* has 11 exons in all and three transcripts are formed involving *Mbtps2* and *YY2*. First is the transcript from the *YY2*

retrogenes. This transcript can potentially encode for 371 amino acid long zinc finger protein and thus have a function similar to that of the parent *YY1* (Nguyen *et al.* 2004). The retrogene does not use the same transcription start site (TSS) as the parent but has recruited a start codon from the intron of *Mbtps2.* Another transcript produced is a fusion transcript formed by first 5 exons of *Mbtps2* and the retrogene - *YY2*. It is interesting to note that *Mtbs2* is in frame with the zinc finger exon of the retrogene (Nguyen *et al.* 2004). The third transcript produced is that from the entire *Mbtps2* where the retrogene along with the fifth intron is spliced out (Nguyen *et al.* 2004). The expression pattern of all the three genes varies with various tissues. *YY1* is ubiquitously expressed. Expression analysis of *YY2* reveals that it shares the expression pattern of *Mbtps2*. In some tissues *YY2* has a expression pattern different from the rest. In the neuronal and glial cells of cerebral cortex, *YY2* has low expression where as *Mbtps2* and *YY1* are highly expressed. All three are highly expressed in adult testis but *YY2* is not expressed in sperm cell where the rest have high expression. Also, in ovary follicles *YY1* is highly expressed but both *YY2* and *Mbtps2* are not expressed. It seems that the fusion transcript uses the promoter of the *Mbtps2* but besides that *YY2* retrogene has recruited an independent promoter which accounts for the difference in expression in the neuronal and glial cells of cerebral cortex and sperm cells (Nguyen *et al.* 2004).

As described in Chapter 1, *Utp14a* (*U3 small nucleolar ribonucleoprotein homolog A*) has given rise to retrogene independently in human (*Utp14c*) and mouse (*Utp14b*) lineage. Homolog of *Utp14a* in yeast encodes a component of the small subunit processome which is a 28 protein complex required for the pre-18S rRNA processing (Dragon *et al.* 2002). In human *Utp14b* homolog has degenerated into a pseudogene suggesting that *Utp14a* has recruited retrogenes recurrently for male biased functions. *Utp14a* is X linked and is ubiquitously expressed, however in mouse the expression is greatly reduced in pachytene spermatocytes (Zhao *et al.* 2007). Both the retrogenes have developed testis specific expression pattern and presumably have an important function in spermatogenesis as mutations rendering them non functional cause male sterility (Bradley *et al.* 2004; Rohozinski *et al.* 2004; Rohozinski *et al.*

2006). It is postulated that selection may favor *Utp14b* and *Utp14c* to develop male specific function in order to compensate for X chromosome inactivation during spermatogenesis (Rohozinski *et al.* 2004; Rohozinski *et al.* 2006).

The *Utp14b* it is located within the 2nd intron of *acyl-CoA synthatase long-chain family member 3* (*Acsl3*) on mouse chromosome 1. *Ascl3* has fifteen exons and the coding region begins in the third exon. It forms two transcripts, one involves both non coding exons 1 and 2 and the second transcript includes only exon 1 and excludes the second exon. Both transcripts have the same transcription start site (TSS) (Zhao *et al.* 2007). The retrogene *Utp14b* is inserted in its second intron and has 5 different transcripts. The variant 1 includes the two upstream non coding exons of *Ascl3* and has the same transcription start site as the host gene. The expression pattern of this variant is same as the host gene. *Ascl3* is ubiquitously expressed with high levels of expression in brain and testis. The variant 1 of *Utp14b* shows ubiquitous expression with higher level of expression in testis than the host gene (Zhao *et al.* 2007). Other four transcripts of *Utp14b* include exon 2 of the host gene and additional non coding exons from the first intron of the host. For these variants the transcription starts within the first intron of *Ascl3*, 517 bp downstream of the common TSS of the host and variant 1. All the four variants have spermatogenesis specific expression and may share a common testis specific promoter. The parent *Utp14a* is ubiquitously expressed and shows very low expression in pachytene spermatocytes (Rohozinski *et al.* 2004). In these cells the variant 1 has the highest expression of all. This indicates that the initial expression was derived from the host promoter and later other testis specific variants evolved (Zhao *et al.* 2007). Higher level of variant 1 in testis than the *Ascl3* maybe because either the *Utp14b* is more stable than the host or the splicing machinery in testis favors the splice acceptor site in the retrogene over that in the third exon of the host gene (Zhao *et al.* 2007).

3.2.2.2. Insertion in 3'UTR of host gene

As described above, *Utp14a* has given rise to retrogenes in two lineages via independent retroposition events. In human, the retrogene *Utp14c* inserted in the 3' UTR of a

novel putative glucosyl transferase-containing gene *GT8* on the chromosome 13 and in the opposite orientation. Syntenic region in mouse does not show any signs of a retroposed gene (Rohozinski *et al.* 2006). *GT8* has four exons and the last exon overlaps the second exon of *Utp14c*. *Utp14c* has recruited an upstream non coding exon and the transcription is initiated within the third exon of *GT8* (Rohozinski *et al.* 2006). The coding region of the retrogene starts downstream of the termination codon of the host gene. Both the parent – *Utp14a* and the host gene – *GT8* are ubiquitously expressed, whereas the expression of *Utp14c* is restricted to the ovaries and testis (Rohozinski *et al.* 2006). Mutations in *Utp14c* have been related to infertility in males (Rohozinski *et al.* 2006). While this retrogene forms a chimeric structure with GT8 it is inserted in the opposite orientation and it is unclear how it may have later acquired the male and female germline specific promoters likely from the third exon of the host gene (Rohozinski *et al.* 2006).

Exon shuffling plays a big role in new gene formation and thus in genome evolution (Long *et al.* 2003). Retrogenes can be a source of new exons that can donate a novel function to already existing genes. *Sp100-HMG* is one such example. The host gene *Sp100* encodes for a nuclear protein - *Sp100*. Many splice variants of the *Sp100* gene at the 3' terminal have been reported (Guldner *et al.* 1999). One of the splice variants, *Sp100-HMG*, has an additional DNA binding domain. This is because of the incorporation of the last exon which is derived from a processed copy of *High mobility group 1* (*HMG1*) gene (Rogalla *et al.* 2000). *HMG1* belongs to a family of high mobility group proteins, which contain HMG-Box domain which has DNA binding and bending activity, thus affecting transcription (Strichman-Almashanu *et al.* 2003). The retrogene *HMG1L3* inserted in the 3'UTR of *SP100* before the divergence of Old World and New World monkeys about 35 – 40 million years ago (Rogalla *et al.*. 2000; Devor 2001). A part of *HMG1L3* becomes the last exon of the fusion gene. It starts 72 bp downstream of the 5' border of the original open reading frame of the *HMG1*. This position has gained 3' splice acceptor sequence for intron-exon boundary after a G to A transition as compared to the parent gene (Rogalla *et al.* 2000). Also a poly (T) stretch follows this site. Here is it interesting to

notice that the insertion was such that the DNA binding motif of the parental *HMG1* was in frame with the *SP100* and thus donated an active DNA binding domain. Thus, a fusion protein with a novel function is created. Regulation was again attained by this retrogene by means of producing a chimera with an upstream gene. It uses the regulatory mechanism of an existing gene and is transcribed as an integral part of an existing protein.

3.2.3. Retrogenes as a source of promoter region

There are examples where retrogenes have provided promoter region to other genes. In *Drosophila bipectinata* complex a chimeric gene – *siren* has derived its regulatory region from a retroposed gene (Nozawa *et al.* 2005). *Siren* is a chimera between a tandem duplicate of *Adh* and *CG11779*. The insertional sequence of *siren* is formed by the insertion of retroposed copy of *nanos* which contains a part of the overlapping gene *CG11779* and its regulatory region in the complimentary strand (Nozawa *et al.* 2005). The second and third exons of *siren* resemble that of *Adh* and the first exon and the regulatory sequence comprises of a duplicate copy of the first exon of *CG11779* and its regulatory region. This duplicate of *CG11779* and its regulatory region was carried by a retroposed copy of *nanos* which inserted on the complimentary strand. This is the only described case where a retroposed gene is a source of regulatory region for another chimeric gene. We have seen above that retroposition plays an important role in exon shuffling and origin of new functions, but this example opens doors to a whole new world where the role of retroposed genes has been much enhanced.

3.3 Carryover from the parent gene

During the earlier studies of retrogene, it was believed that the retroposition event will always lead to formation of a pseudogene and only when originated from an aberrant transcript (i.e. longer at its 5' end and containing the regulatory region of the parental) would it have a chance to survive in the genome. Under this assumption, retrocopies in the human genome were annotated always initially as pseudogenes (Dunham *et al.* 1999). Now we know that there are various other ways a retrogene can acquire transcription (see this work). However, carry over from the parent gene may ensure expression of the retrogene at the time of its origin. A

retrogene can carry a promoter from the parent in two situations. First, when the parental gene forms an aberrant transcript which starts upstream of the original transcription start site. Such an aberrant mRNA will have the regulatory region of the parental gene and that might drive its initial pattern of expression. Secondly, when there are regulatory sequences downstream of the transcription start site i.e. the 5'UTR or the 5' exons of the parental might harbor downstream elements (Lim *et al.* 2004; Burke 1997). In this case the regulatory sequences will be carried with the transcript. There are instances of both types of events. However, see below that with time, the retrogenes may be selected to evolve a new pattern of expression.

Phosphoglycerate kinase-2 (*Pgk-2*) gene is one of the first studied retrogenes. It is another example of X to autosome duplication. The parental gene *Pgk-1* is X linked and ubiquitously expressed and its copy *Pgk-2* is autosomal and expressed only during spermatogenesis. The duplication event is about 125 million years old, before the divergence of eutherians and metatherians (McCarrey 1990; Potrzebowski *et al.* 2008). The ubiquitous expression of *Pgk-1* gene is regulated by CpG island. Presence of CpG island is a common feature of many ubiquitously expressed genes. This is the only highly conserved region between the *Pgk-1* promoters in human and mouse and thus seems to be sufficient for the ubiquitous expression of both genes. An aberrant mRNA of this gene, the transcription of which initiated at an upstream start point, was responsible for the formation of the retrogene. Thus, initially the *Pgk-2* gene was regulated by the same regulatory region as its parental gene and thus, was ubiquitously expressed. During the course of evolution the CpG in the promoter region was substituted by a TA rich region (McCarrey 1990) and gradually due to selection, the *Pgk-2* gene lost its ubiquitous expression in mouse and became testis specific and developed an elaborate regulation.

Studies from transgenic mice showed that the core promoter of *Pgk-2* is contained in a 188bp region upstream from the transcription start site and consists of two transcription factor binding sites. It has a GC box – binding site for SP1 transcription factor and the CAAT box – binding site for the nuclear factor 1. These are essential elements for promoter activity, however

a 323bp upstream region is required for transcription initiation (Robinson *et al.* 1989). CAAT box is known to bind to general transcription factors and enhance the expression of down stream genes. The Sp1 also has a wide range of target genes, however recently some variants have been known to drive male germ cell specific expression. An enhancer element (E1/E4) present 40bp immediately upstream of the core promoters directs the tissue specific expression (Gebara *et al.* 1992). Besides these, the *Pgk-2* gene has evolved additional upstream enhancer sequences, which direct the lower level expression of the gene in absence of the E1/E4 element or the Sp1 binding site (Zhang *et al.* 1999). This shows that besides a basal promoter sequence, the tissue specific *Pgk-2* gene has developed an elaborate regulatory mechanism. The above-mentioned element drives the tissue specific expression, which has to be suppressed in somatic tissues. Such suppression is mediated by CpG methylation (Zhang *et al.* 1998). *Pgk-2* derived its initial expression from its parental gene, however, it has evolved elaborate regulation with time.

Poly (A) Binding Protein 2 (*Pabp2*) is a retrocopy of the *Pabp1* gene that originated ~80 Mya in mouse. *Pabp2* has testis specific expression pattern where as the parent is ubiquitously expressed (Kleene *et al.* 1998). The 5' upstream region of the retrogene shows 72.5% homology to the human *Pabp1* gene which shows that it carries the complete 5'UTR of the parental gene (Kleene *et al.* 1999). The 5' region of the retrogene harbors AP2 and MSY-2 transcription factor binding sites. Both have been involved in germ cell specific expression of various genes (U. Kevin Ewulonu 1996; Yiu *et al.* 1997). The 5' UTR of the parent harbors the AP2 and Sp1 binding sites. This indicates that the retrogene uses the binding sites present in 5'UTR of the parental gene (Kleene *et al.* 1999).

Another example of promoter recruitment from 5'UTR is the woodchuck *N-myc2* gene. This retrogene *N-myc2* found in woodchuck and squirrels recruited its promoter from the non coding part of exon 2 of the parental gene *N-myc1* (Fourel *et al.* 1992). Both *N-myc1* and *N-myc2* belong to the myc family of proto-oncogenes. The retrogene is expressed in the brain of a healthy woodchuck and in the liver in a tumor condition. The parent however is expressed in all

tissues except liver, and high level of expression is observed in brain and testis. In this case, the transcription of the retrogene *N-myc2* starts ~280 bp downstream of the 5' boundary of the retrosequence. The 5' flanking region of his gene has Sp1 binding site, CAAT box, two purine boxes and a TATA-like sequence. The TSS of the retrogene is located ~37 bp downstream of the TATA box. By deletion mapping it was observed that removal of Sp1 and CAAT binding sites and purine boxes reduced the level of expression; however removal of TATA box resulted in a complete loss of expression (Fourel *et al.* 1992). Besides CAAT, all the other binding sites are conserved between *N-myc1* and *N-myc2*. The CAAT box is specific for the retrogene and may have evolved after the insertion (Fourel *et al.* 1992). The TATA box corresponds to a region in the exon 2 of the parent gene and is highly conserved among woodchuck, mouse, rat and human *N-myc1* genes. It is assumed that this region could be involved in the expression of *N-myc1*, however, no experimental evidence supports this hypothesis (Fourel *et al.* 1992). Here the retrogene has utilized a putative promoter located in the 5'UTR of parent and has also evolved other elements for higher expression.

3.4 Some Retrogenes Are Transcribed from Bidirectional Cis-Regulatory Regions

Another interesting example of retrogene inserted in the intron of another gene is *Germ cell specific gene 2* (*Gsg2*) which encodes the Haploid Germ Cell-specific Nuclear Protein Kinase (Haspin). It was created before the divergence of human and mice. The progenitor of *Gsg2* is not known, however it has all the hallmarks of a retrogene (Tanaka *et al.* 2001). It is intronless and flanking target site duplications are observed in both human and mouse. The remnants of a poly (A) tract in the 3' end are only observed in mouse. These may have been erased from the human lineage (Tanaka *et al.* 2001). It is male germ cell specific with high level of expression in haploid spermatids (Tanaka *et al.* 1999). Some expression is also seen in adult thymus and bone marrow and various adult and fetal tissues (Higgins 2001). It is inserted in the 26th intron of the *integrin αE* gene (Schon *et al.* 1999) which comprises of 31 exons. The retrogene is located close to the exon 27 on the sense strand but in opposite orientation. Another transcript of the αE gene in humans (*hAED*) is known to start from the 27th exon and

includes exons 29, 30 and 31 (Higgins 2001). The *hAED* transcript and *Gsg2* are in a head to head orientation, about 70 bp apart from each other. This 70bp region harbors a CpG island along with three transcription factor binding sites, namely AP2, E2F and Sp1 (Higgins 2001). Presence of a CpG island is a common feature of genes present in head to head orientation (Adachi *et al.* 2002). The transcription factor AP2 has a wide range of target genes, and has been involved in testis specific expression of genes through interaction with germ cell specific transcription factors (U. Kevin Ewulonu 1996). E2F targets genes having diverse function. These include apoptosis, development and differentiation (Bracken Ciro *et al.* 2004). Sp1 sites are associated with CpG promoters and are bidirectional. These have been known to target housekeeping gene, however recently some variants of the Sp1 transcription factor have been associated with regulation of germ cell specific genes (Thomas *et al.* 2007).

The transcription start site for *hAED* transcript lies 26bp downstream of the transcription start site of *Gsg2* (Tokuhiro *et al.* 2007). A GC rich 193bp region 5' of the *Gsg2* gene has been detected as the region responsible for the bidirectional and the testis specific expression of the *Gsg2* and the hEAD transcript in mice (Tokuhiro *et al.* 2007). This region predominantly drives the testis specific expression of *Gsg2*. Although GC rich, this region is not methylated for silencing of these genes in somatic tissues, and thus may use transcriptional factors specific for testicular germ cell for the tissue specific expression. Results from gel retardation assay suggest that specific factors from testicular germ cell lysate bind to the *Gsg2* promoter (Tokuhiro *et al.* 2007). The specific factors responsible for the expression of *Gsg2* have not been identified; however, a relatively short sequence is required for its testis specific expression.

While a bidirectional promoter seems to drive the expression of *Gsg2*, comparative genomic analyses remain to be done to see if the described region was there before the retrogene insertion or evolved afterwards.

3.5 Regulatory region from transposable elements

Transposable elements (TE) were first discovered by Barbara McClinctoc in 1952. Since then, for long, they were considered as junk and parasites (Doolittle *et al.* 1980). However, after the release of genome sequence of many organisms, they seem to be one of the major factors influencing genome dynamics (Makalowski 2003). TE's have a tendency to insert in the 5' upstream region of many genes (Thornburg *et al.* 2006). Once inserted the TE's can evolve de novo regulatory sequence through point mutations or promoter elements pre-existing in them can be co-opted to act as regulators of the downstream genes (Mariño-Ramírez *et al.* 2005; Feschotte 2008). Thus, they affect regulation, which if not deleterious may be selected for. This nature of TE's makes them an important source of regulatory region for many genes (Jordan *et al.* 2003). Retrogenes are no exception. Theoretically retrogenes use the LINE machinery for insertion in the genome (Esnault *et al.* 2000) and it is possible that both insert in close proximity to each other. Below we review the actual evidence.

The LINE1 elements are a class of transposable elements that move in the genome via an RNA intermediate formed by the reverse transcriptase encoded within the element. The elements have a 5' untranslated region that harbors an internal promoter, two ORFs of which one has RNA binding activity and the other has reverse transcriptase and endonuclease activity, and a 3' UTR region ending in a poly A tail (see (Moran 1999) for review). A functional retrogene in mouse - *PSME2b* is an example of retrogene deriving its promoter region from a TE. The parental gene *PSME2* is located on the chromosome 14 and encodes for a protein PA28β, which is a subunit of the proteasome activator complex PA28 (Kohda *et al.* 1998). The retrogene *PSME2b* inserted in a transcriptionally active LINE1 element on chromosome 11, 470bp downstream of the F-type murine LINE1 promoter (Zaiss *et al.* 1999). It is flanked by target site duplications and a poly A tail is also observed (Zaiss *et al.* 1999). The F-type LINE 1 promoter upstream of this retrogene is an active promoter as it can drive the expression of luciferase reporter gene (Zaiss *et al.* 1999). The in vitro translation of the retrogene resulted in

the PA28β active protein, suggesting that *PSME2b* is an expressed retrogene deriving its expression from F-type LINE1 promoter (Zaiss *et al.* 1999).

Two retrogenes *CP-2* and *CP-5* (*1-CysPrx -2* and *5*) derived from *CP-3* (*1-Cys peroxiredoxin*) are inserted in mouse L1 repetitive elements (Lee *et al.* 1999). These are intronless genes flanked by 11 bp direct repeats and a Poly A tail and show 85% sequence similarity with the parent gene. These are identical to each other and can putatively encode 224 amino acid residues. They are inserted in the reverse transcriptase domain of mouse L1 element (Lee *et al.* 1999) and might have inherited expression from the TE.

3.6 De novo recruitment

Another mechanism by which a retrogene can recruit a promoter is when it fortuitously lands downstream of a region that resembles a promoter and by only a few substitutions can become functional. We call this de novo recruitment, where a regulatory region did not exist before the insertion but after insertion mutations leading to functional promoter are selected for. Such events can only be evaluated when regions can be compared in closely related species that do not have the insertion to trace back the steps of evolution.

We describe the regulatory region of *Dntf-2r* at length in Chapter 2.

Another example of such recruitment is *Pdha2*. It is one of the highly studied retrogenes to understand gene regulation during spermatogenesis. Its origin involved X chromosome to autosome retroposition of *Pdha1* which encodes for the somatic E1α subunit of Pyruvate Dehydrogenase (PDH) complex. The retrogene *Pdha2* is located on chromosome 4 in humans and chromosome 19 in mouse (Dahl, Brown *et al.* 1990; Fitzgerald, Hutchison *et al.* 1992). It arose after the divergence of eutherians and marsupials. It shares 86% protein identity with the *Pdha1*. This gene is expressed specifically during the spermatogenesis and high level of transcript has been observed at the pachytene spermatocyte stage (see (Julia C. Young 1998) for review). Such a specific expression needs to be highly regulated. In mice the core promoter was identified within the nucleotide positions -187 and +22 of the gene. In transgenic mice, this region induced 6-12 fold higher expression of CAT reporter gene in testis than in

somatic tissues. This region has been shown to direct the testis and time specificity of *Pdha2* (Iannello *et al.* 1997). The ~187bp core promoter contains binding sites for four transcription factors namely - Sp1, ATF/CRE, YY1 and MEP-2. The MEP-2 binding site drives testis specific expression of many other genes too (Iannello *et al.* 1997). Sp1, ATF/CRE and YY1 are transcription factors that generally target house keeping genes however, some variants of Sp1 can induce male germ cell specific expression. For a gene to be tissue specific, it is important that its expression in the rest of the tissues is suppressed. This may be achieved by CpG methylation. In vitro, methylation of CpG dinucleotide located within the ATF/CRE binding site has been shown to be responsible for reduced activity of the reporter gene. In vivo too, this site is critical for the functional activation of the promoter (Iannello *et al.* 2000). We consider this example as a de novo recruitment of promoter as the source of the promoter element is not known, on studying the genomic region of the gene on vista genome browser we observed that it is not close to any other gene or transposable element to be able to 'hitch hike' its promoter in any of the species (probably revealing the ancestral situation) and the parental gene does not contain these motifs.

Another example of de novo recruitment is the *ARF-lik4* (*Arl4)*. The ARF (ADP-ribosylation factors) family is a subfamily of Ras related GTPases and consists of six ARF's and six ARF-like (ARL's) proteins (Moss *et al.* 1998). One of these - *Arl4* has originated as a result of retroposition of *Arf* or an ARF like isoform. This retroposed gene inserted in the mouse chromosome 12 and a homolog is found on human chromosome 7 (Jacobs *et al.* 1998). The CDS, the 3'UTR and a considerable portion of the 5'UTR is present in a single exon. Additionally *Arl4* has recruited two alternatively spliced exons (1A and 1B) from the flanking region (Jacobs *et al.* 1998). These alternative exons are included in the transcript in a tissue specific manner. The upstream exon (1A) is included in the transcript produced in the testis and the downstream exon (1B) is used in the transcript from 3T3-L1 cells, brain, spleen, kidney and muscle. The tissue specific expression of this gene is regulated by two separate promoters. The

luciferase fusion construct of the 5'flanking region of both the exons 1A and 1B showed promoter activity when expressed in the COS-7 cells (Jacobs *et al.* 1998).

Both the promoters A and B have GC rich regions. The 419 bp promoter B causing the ubiquitous expression of *Arl4* has two CAAT boxes and four Sp1 binding sites. These are well known transcription factors which drive ubiquitous expression of many genes. The 403bp promoter A which shows testis specific expression has six Sp1 binding sites, a cAMP response element (CRE) like motif, two motifs matching presumed testis specific expression of proacrosin gene and a palindromic motif flanked by CAAT boxes (Jacobs *et al.* 1998). The CRE-like motif has been known to enhance testis specific expression in various genes and certain Sp1 variants can also drive germ cell specificity.

Another interesting example of de novo recruitment could be the case of *tact1* and *tact2*. These are claimed to be a retrogenes (Mizue Hisano *et al.* 2003). However, they represent a very old retroposition event and have lost the flanking duplicates and the poly(A) tail, and have greatly diverged from the parental gene. We have included it in this review as it is a unique case where a retrogene has duplicated and both copies are functional. The *tact1* and *tact2* genes are located on the 4^{th} chromosome of mouse genome. These two are very similar to each other and are located about 2 kb apart in head to head orientation. Orthologs of *tact1* and *tact2* in humans are named *ACTL7B* and *ACTL7A* respectively showing that these were created before the divergence of human and mouse. These are intronless genes that have lost rest of the hallmarks of retroposition. These have amino acid similarity to proteins involved in ATPase activity and have 42% identity with conventional actins. However, a single gene of origin is not known (Mizue Hisano *et al.* 2003). As these are present in a head to head orientation, they share the 5' upstream region. The 2 kb region between the two genes has been shown to direct the haploid spermatids specific expression of both the genes (Mizue Hisano *et al.* 2003). This region lacks any conventional promoter like the TATA box, but harbors CRE-like motifs. These motifs are conserved between the two genes and the human homologs. The *tact1* has 5'-TGACATCA-3' and the *tact2* gene has 5'-TGATGTCA-3' which are similar to

the consensus sequence for cAMP response element 5'-TGACGTCA-3' (Mizue Hisano *et al.* 2003). From the transient transfection analysis it was observed that CREMτ transcription factor induced eight fold and twelve fold increase in the activation of *tact1* and *tact2* genes respectively. This indicates that retrogenes have upstream regulatory elements which may in turn be regulated by the CREMτ. The CRE-like motifs have been known to drive the expression of many testis specific genes. *Pdha2* is one of the testis specific retrogene regulated by such motif (Iannello *et al.* 2000). The origin of this regulatory sequence is not clear as the parent gene for *tact1* and *tact2* genes is not known. The origin of these retrogenes is interesting though. It may have been created either by insertion of two independent DNA copies of the parent mRNA or by insertion of one copy followed by duplication. In the first case the highly conserved 5' upstream region may be derived from the parental transcript. In the second case the high conservation of the 5' upstream region can be attributed to the duplication of the gene along with its 5' regulatory sequence. Both the scenarios make this a very interesting retroposition event.

3.7 Epigenetic regulation of retrogenes

Gene imprinting is a kind of epigenetic regulation where only one allele of an imprinted gene, either maternal or paternal, is functional. The other allele is silenced. This is one type of monoallelic expression. Genomic imprinting has been observed in various plants (Scott 2006), and animal species (Paulsen *et al.* 1998; Peters *et al.* 1999).

Imprinted genes are in association with differentially methylated regions (DMR) embedded within CpG islands which are methylated in sex specific manner. This differential methylation is considered to be involved in the sex specific expression of linked genes. The imprinted genes are generally found in clusters. Large imprinted chromosomal domains are coregulated by a genomic region termed as IC (imprinting center) which has imprinted status like the DMR. In human and mouse many imprinted clusters are known (Paulsen *et al.* 1998; Peters *et al.* 1999). Insertion of a retrogene in such imprinted clusters can influence its pattern of expression. In a recent study, out of 11 retroposed imprinted genes in mouse, 8 were

inserted in imprinted clusters and only 3 were imprinted in spite of the surrounding region having biallelic expression (Wood *et al.* 2007) showing that a majority of time when a retrogene inserts in close proximity of imprinted genes, it take up the status of the surrounding region. However, it is essential for such retrogenes to fortuitously land downstream of a CpG island and a promoter sequence that confers the imprinted status (see below).

The chromosome 7C in mouse harbors many imprinted clusters. One such cluster is the prader-willi syndrome (PWS) region which shows paternal allele dominant expression. The loss of paternal allele expression of the orthologus region in human (chromosome 15q11-q13) causes the prader-willi syndrome (reviewed in (Horsthemke *et al.* 2006)). Six intronless genes map to this region in mouse (Mkrn3 (Jong, Gray *et al.* 1999), Ndn, Magel2, *Frat3* (Chai *et al.* 2001), C150rf2 and Atp5l-ps1 (Chai *et al.* 2001)) out of which three (Mkrn3, *Frat3* and Atp5l-ps1) have originated by retroposition. The origin of the other three intronless genes is unknown.

The *Frat3* (*Frequently rearranged in advanced T-cell lymphomas 3*) is a recently evolved retroposed paralog of *Frat1* (18-26 mya) (Chai *et al.* 2001). *Frat1* has bialleleic expression as both the alleles are not methylated. The retrogene – *Frat3*, however, shows differential methylation, and is only paternally expressed in mouse brain consistent with paternal expression of PWS region. As commonly found in imprinted genes, a CpG island overlaps 163nt 5' region of *Frat3*, a region that includes 5'UTR and part of the coding region. The progenitor *Frat1* lacks the CpG island. The imprinting status of the PWS region is attributed to imprinting center (IC) overlapping the snrpn (small nuclear ribonucleoproteins) promoter region, deletion in which result in the condition of PWS (Bielinska *et al.* 2000). This IC has a high number of matrix attachment regions (MAR's) which have cis-acting epigenetic regulatory function (Greally *et al.*. 1999). Before the insertion of the retrogene, the ancestral state had CpG island and promoter region which confers the surrounding paternal allele dominant expression to the new gene (Chai *et al.* 2001). Here the IC co regulates the expression of the paternal allele of the surrounding genes including the retrogenes. Interestingly, the retrogene landed downstream of an existing CpG island which conferred its imprinting status.

There are also examples where the surrounding region of a mono allelic imprinted retrogene has biallelic expression (Nabetani *et al.* 1997; Smith *et al.* 2003; Choi *et al.* 2005). *U2af1-rs1* is derived from a non-imprinted gene *U2af1-rs2* located on the X chromosome. It is in the first intron of *Murr1* gene on the mouse chromosome 11 (Nabetani *et al.* 1997) and is transcribed in the opposite direction. This is a result of recent retroposition that occurred after the split of human and mouse. The retrogene *U2af1-rs1* is an imprinted gene expressing exclusively from the paternal allele where as the host gene *Murr1* has biallelic expression in all tissues (Nabetani *et al.* 1997) except in adult mouse brain where it has maternal allele predominant expression (Wang *et al.* 2004). The orthologus gene *MURR1* in humans also does not show any monoallelic expression (Zhang *et al.* 2006). Imprinted genes are generally present in clusters and are controlled by a distant ICR (see above). When the surrounding region of *Murr1* in mouse (Nabetani *et al.* 1997) and its syntenic region in humans (Zhang *et al.* 2006) was analyzed for imprinted expression, no differential methylation pattern was observed. Thus, indicating that in the *Murr1* region, *U2af1-rs1* is the only gene having monoallelic expression, and it may interfere with the expression of *Murr1* in adult mouse brain causing it to have a maternal allele predominant expression. (Wang *et al.* 2004)

The promoter region for *U2af1-rs1* is not known. Out of eight CpG islands found in the *Murr1* region only one that is linked to the *U2af1-rs1* retrogene showed methylation. In humans, where the retrogene is absent, no such methylation was observed. The CpG island linked with *U2af1-rs1* overlaps its 5' region (Zhang *et al.* 2006) which includes the putative promoter region, the complete 5' UTR and part of the open reading frame. Within the CpG island, a specific region in the 5' end shows oocyte-specific methylation and thus may act as the imprint control region for this gene. (Zhang *et al.* 2006).

A review by Wood *et al.* (Wood *et al.* 2007) compares three retrogenes (*U2af1-rs1*, *Inpp5f_v2* and *Nap1l5*) which have monoallelic expression despite the surrounding region being biallelic. All three are inserted in introns of somatic genes and are derived from progenitors on the X chromosome and all show paternal allele dominant expression. Paternal conflict has been

known to be the selective force behind the evolution of imprinting (Hore *et al.* 2007). A gene that favors paternal interest at the expense of mother may be methylated in maternal genome and another gene that may be counteracting the paternal interest may be silenced in the father's genome. However, retrogenes inserted in imprinted locus may be `innocent bystanders' and may not signify evolution due to such conflict (Hore *et al.* 2007). Other evolutionary forces that could shape the imprinted status of a gene may be dose compensation after duplication, or intra locus genetic conflict (Hore *et al.* 2007).

Thus, to study how a retrogene recruited a promoter it is not only important to study its immediate 5' flanking region, but the surrounding region where a retrogene lands and the expression pattern of neighboring genes should also be considered.

3.8 Functional retrogenes are often close to genes

In humans retropseudogene location and level of expression were compared to functional retrogenes. Functional retrogenes are higher expressed and tend also to be closer to genes (Vinckenbosch *et al.* 2006). It would seem that inserting closer to genes leads to higher expression and more chances to remain functional. It was hypothesized that chimeric retrogenes, and chromosomal domains that favor transcription would explain why being closer to genes leads to higher expression. This particular comparison is not possible in other genomes like Drosophila where the number of pseudogenes is very small.

3.9 Retrogenes in excess in testis neighborhoods

Recently, biases in the location of some retrogenes were revealed (Bai *et al.* 2008; Dorus *et al.* 2008). Looking at previously reported testis domains (Parisi *et al.* 2004) an excess of retrogenes was observed mapping in these neighborhoods (Bai *et al.* 2008). Sperm proteome also contains retrogenes that map to testis domains (Dorus *et al.* 2008). A comparison with retropseudogenes would tell us if this is a selective pattern but we do not dispose of this data. For now the only comparison has been performed with TE insertions revealing that retrogenes do not follow TE distribution in Drosophila. TEs are in excess in the X chromosome and inserted close to female and male germline genes unlike retrogenes.

3.10 Out of testis?

It was recently observed for humans retrogenes that young functional retrogenes are initially expressed in testes, which may contribute to their immediate preservation, but older acquire a higher and broader tissue expression, which may eventually lead to the evolution of other new functions. This was called the 'out of the testes' hypothesis (Vinckenbosch *et al.* 2006).

Similar analyses were performed in Drosophila, but no change in level of expression or male germline bias was observed between young and old retrogenes. Many Drosophila retrogenes are expressed highly and primarily in testes regardless of their age (Bai *et al.* 2007).

3.11 Conclusion

Here we review the literature trying to compile the various modes by which retrogenes are expressed to get some insides on how they might recruit a promoter region. We also include work that tries to explain retrogene expression and pattern of expression by studying biases in their location or comparing to retropseudogenes.

It is clear from the above data that there are diverse ways in which retrogenes are currently being expressed. Some of the studies directly address the initial expression manner: chimeric genes, parental aberrant transcripts, inserted in TEs, and our de novo study reveals a lot of details but other just describe current regulation (e.g. bidirectional motifs used in *Gsg2*). In the case of the bidirectional motifs used in *Gsg2* we wonder if the motifs were already present at the time of insertion or evolved afterwards. More detail comparative genomic analyses should answer those questions. In some instances the signatures of the events might be gone.

It is interesting to note that promoter recruitment and transcription appears to be in many cases highly dependent on the region of insertion: chimeras, bias towards insertion close to genes, inserted in testis neighborhoods, inserted in imprinted cluster, insertions in TEs, or regulatory potential of the 5' region. In other instances the initial survival depends on the quality of the transcript: aberrant transcript carrying regulatory regions from parental gene or downstream regulatory regions carried in the normal transcript.

In humans, it has been observed that retropseudogenes express in male germline but never at the level or percentage of functional retrogenes. This could be explained by testis expressed genes that are preserved more often or retrogenes additionally evolving testis expression (Vinckenbosch *et al.* 2006). From the data above, it is evident in some cases of male germline expression that selective pressures created and or improved the pattern of expression. *Pgk2* originated from an aberrant transcript of the parental gene and ulterior changes in the sequence lead to the likely needed testis-specific expression. Even in the case of *Dntf-2r*, it is likely that the improvements in the testis-driving motif are fixed by positive selection. Although genes might insert in testis neighborhoods, we postulate that it is likely that de novo regulatory regions emerge through positive selection in many cases. Context is only a contributing factor in some instances. Interestingly, most of the retrogenes listed above have currently short regulatory sequence from a few base pairs to couple of hundreds. Some older retrogenes like *Pgk2* and P*adh*a2 have evolved elaborate regulation to express in male germline, however a short region is sufficient to drive their tissue specific expression. This is consistent with the time frame available for new retrogenes.

CHAPTER 4

ANALYSES OF *Dntf-2r* FUNCTION

Dntf-2r originated from a nuclear transport gene *Dntf-2*. It has testis specific expression in all species where it is present as described in Chapter 2. The NTF2 protein is functional as a dimer and it is involved in the transport of GDP bound Ran across the nuclear membrane into the nucleus. It does this interacting with nucleoporins (nuclear pore proteins). It has also been observed to transport other proteins to the nucleus when bound to Ran (see Chapter 1). Initially (i.e. after duplication), *Dntf-2r* could essentially perform the same function as the parental gene. With time and after positive selection has been acting on the duplicate (Betran *et al.* 2003) its function may have changed. It may or may not have lost some functions of the parental gene or may have gained additional functions. Additionally, *Ran* has also a young retroduplicate that is testis specific called *Ran-like*.

Here, we begin exploring the function of *Dntf-2r*. We compare the amino acid sequence of *Dntf-2r* to that of the parental *Dntf-2*. The functional interactions of the parental gene have been extensively studied, the dimer has been crystallized interacting with Ran and most of the functional domains have been identified. On comparing the amino acid sequences of these two proteins it can be identified if any functional domains have been lost or maintained in the retrogene encoded proteins. Additionally, taking advantage of the fusion protein from previous chapter, the expression of fluorescent protein during spermatogenesis is analyzed using confocal microscopy. The parent is nuclear transport protein and functional data suggest that it localizes near the nuclear membrane. The localization of the fusion protein (*Dntf-2r*-EGFP) will indicate the cell types where the gene is present and its cellular localization will be related to its function. In vivo interactions of this fusion are also being studied by protein co-

immunoprecipitation. A P element insertion line that knocks out *Dntf-2r* is used to identify any mutant phenotype. A fertility assay and a more sensitive male sperm exhaustion assay are used as no other easily identifiable phenotype is observed. Chapter 5 addresses a potential role of *Dntf-2r* in meiotic drive using this knockout line.

4.1 Material and Methods

4.1.1 Stocks used

Besançon strain of *D. melanogaster* is used as wild type. Three stocks having P element insertion close to *Dntf-2r* were ordered from Bloomington stock center. Line EY05573 had P element insertion in the coding region and line KG00588 had P element inserted in 5' upstream region (Figure 4.1). Another line (EY12961) having P element in the second chromosome but was not close to any known functional region was ordered . Same strains, P element and procedures were used to generate strains EY05573 and EY12961 and hence come from an identical genetic background. EY12961 strain was used for comparison with EY05573 because the P element of EY05573 is defective and cannot be excised.

The genotype of each line is as follows:

EY05573: y[1] w[67c23]; P{w[+mC] y[+mDint2]=EPgy2}Ntf-2r[EY05573]

KG00588: y[1] w[67c23]; P{y[+mDint2] w[BR.E.BR]=SUPor-P}KG00588

EY12961: y[1] w[67c23]; P{w[+mC] y[+mDint2]=EPgy2}EY12961

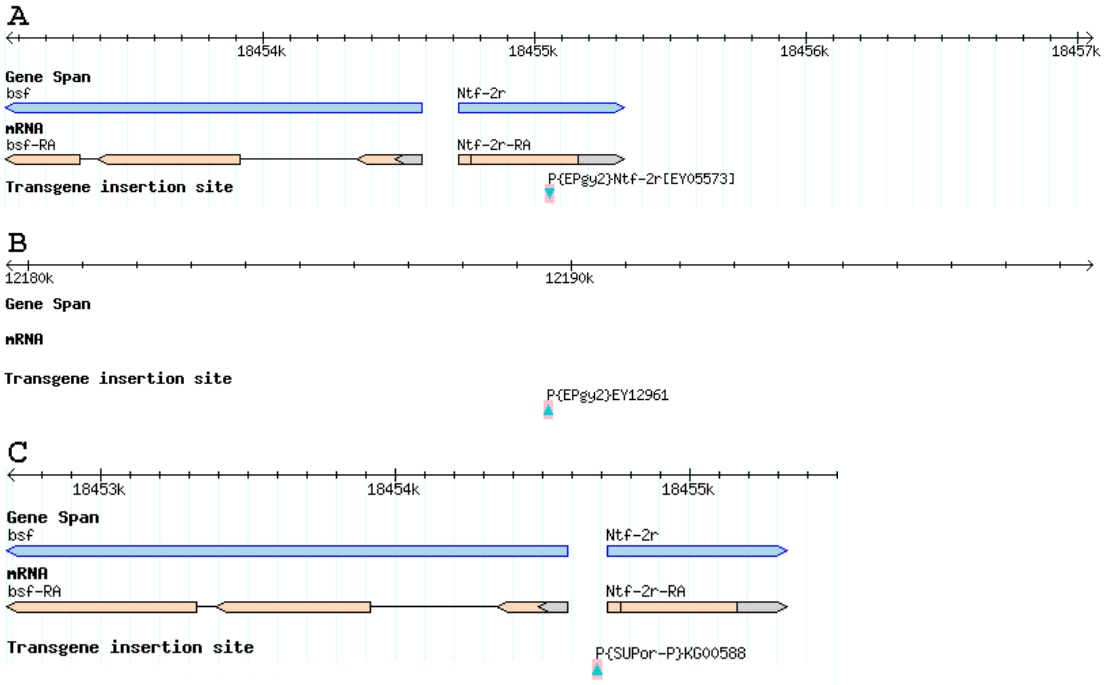


Figure 4.1. A schematic representation of P element insertion. A. EY05573; B. EY12961; C. KG00588 Blue arrow indicates the site of insertion.

A stock (1.53 45.1-II) carrying *Dntf-2r*-EGFP fusion P element insertion (from Chapter 2) was used to detect localization of fusion protein and for co-immunoprecipitation and western blotting.

4.1.2 DNA extraction

DNA was extracted from 15-20 adult flies using the Puregene DNA purification system – cell and tissue kit from Gentra systems. Flies used were from *D. melanogaster* wild type, EY05573, EY12961 and 1.53-45.1-II (line with P element insertion carrying the fusion gene). The 1.53 line carries the longest (-157bp) promoter element (see chapter 2).

4.1.3 RNA extraction

The Qiagen RNeasy mini kit was used for RNA extraction from about 15 adult males from lines fixed for EY05573 and KG00588 P element insertion.

4.1.4 PCR and RT-PCR

PCR was performed from DNA isolated from KG00588 and EY05573. Primers used for EY05573 are Dntf2r_3'RACE1 – 5' 'TTTGTCCAGC AGTACTACGC' and Dntf2r_GSP1 – 5'AGCCACGAAGAGGGATCCTC 3'. These primers were used in combination with primers specific for P element 3.SUP.seq1 TATCGCTGTCTCACTGCTCCCACATT and 5.SUP.seq1 TCCAGTCACAGCTTTGCAGC (Bellen *et al.* 2004). The expected product length for each combination was calculated using P element sequence from gene disruption project database (http://flypush.imgen.bcm.tmc.edu/pscreen; (Bellen *et al.* 2004).

Retrogenes are intronless. This makes it important to remove any genomic DNA contamination as a RT-PCR product cannot be differentiated from that from a genomic DNA contamination. For this purpose, all RNA samples are digested with DNAse I enzyme before further processing. Reverse transcription was performed using oligo (dT) primers (Promega) and Superscript II reverse transcriptase (Invitrogen). From the cDNA obtained, PCR was performed using the primers Dntf2r_3'RACE1 – 5' 'TTTGTCCAGCAGTACTACGC' and Dntf2r_GSP1 – 5'AGCCACGAAGAGGGATCC TC 3'. These primers flank the P element insertion site in EY05573.

4.1.5 Inverse PCR

Genomic DNA from line KG00588 was digested with HpaII at 37^{o}C for 4 hrs. The enzyme was inactivated by heating at 65^{o}C for 20 minutes. To test for digestion 5µl of digested DNA was run on 1% agarose gel. Once complete digestion was obtained, it was precipitated following the protocol from Bellen et. al (Bellen, Levis *et al.* 2004)

Ligation was performed overnight at 4^{o}C. A large volume (400µl) of ligation reaction was used to facilitate generation of circular products. PCR was performed from the ligated mix using primers Pry1 (CCTTAGCATGTCCGTGGGGTTTGAAT) and Pry4 (CAATCATATCGCTGTCTCACTCA) and primers Plac1 (CACCCAAGGCTCTGC TCCCACATT) and Pwht1 (GTAACGCTAATCACTC CGAACAGGTCACA).

4.1.6 5'RACE

RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE) was performed to identify the transcription start site for *Dntf-2r* in *D. simulans* using First choice RLM-RACE kit from Ambion, Inc. One µg of RNA extracted from adult males and virgin females was used. Gene specific outer primer CG10174 5'RACE2 - 5' CCGTTGGGCTTCAGCAAAAAGAT 3' and the inner primer 5'RACE1 - 5'CATCGCATTTTAGTCTTCCAAGGACG 3' were used an long with the 5'RACE outer and inner primers provided by the manufacturer respectively.

4.1.7 Protein function identification

The parent protein has been crystallized as a dimer (from rat) and in association with RanGDP (Ntf2 from rat and RanGDP from dog). Structures were downloaded from Protein database (http://www.rcsb.org/pdb) and analyzed using PyMOL software (http://www.pymol.org). As the crystal structures were not from drosophila, required changes were made in the amino acid sequence to make them identical for Dntf-2 and *Dntf-2r*. The functional domains were identified from literature and conserved domain database (CDD) (Marchler-Bauer, Anderson *et al.* 2009).

4.1.8 Co-immunoprecipitation/ Western blot

Testes from about 40 males carrying the *Dntf-2r*-EGFP fusion P element insertion, were dissected in ice cold ringer's solution and homogenized in cold extraction buffer containing 50mM Tris/HCl (pH 8), 150mM NaCl, 0.05% Triton X-100 and protease inhibitor cocktail. The lysate was then immunoprecipitated using anti GFP antibodies (Abcam). A immunoprecipitation kit containing magnetic beads coated with protein G was used (Invitrogen cat# 100-07D). Beads were conjugated to anti GFP antibodies for 20 mins at room temperature and these conjugated beads were exposed to tissue lysate for 20 mins at room temperature. A control, where unconjugated beads were exposed to the lysate was also performed. The lysates were resolved on a criterion 10% Tris/HCl gel (BioRad). The gel was allowed to run with buffer containing SDS for a few minutes before loading the samples. After the sample was resolved the gel was stained with Corsi tangerine protein stain to check for enrichment of proteins after co-

immunoprecipitation. The proteins were then transferred to PVDF membrane and blotted with anti Ran antibody (Abcam). Membrane was then blotted with a secondary anti rabbit IgG antibody and signal was detected using Amersham ECL western blotting detection kit (GE Healthcare). The next day, the membrane was stripped and again blotted with anti GFP antibody and secondary antibody and signal was detected.

4.1.9 DNA staining of fixed testes

Testes were dissected from young males fixed for the P element insertion of the *Dntf-2r* EGFP fusion gene (see previous chapter). Testes were collected in 4% Paraformaldehye and incubated overnight at 4^{o}C. Next day they were washed two times with 1X PBS for 10 minutes with rotation. For staining, testes were incubated with staining solution (355µl 1X PBS, 40µl 10mg/ml BSA, 4µl 10 % saponin, 1µl Draq5 and 0.5µl RNase) for 45 minutes at room temperature. They were again washed with 1X PBS for 10 min with rotation three times. Stained testes were mounted on slides using 70% glycerol and observed under the confocal microscope.

4.1.10 Fertility assay

Males from stock EY05573 were tested for fertility defects and males from EY12961 were used as control. Two methods were employed to check the fertility effects of *Dntf-2r* knockout. One is a simple test of male fertility where three males from line EY05573 and line EY12961 were used. Three to four days old virgin males reared in controlled density (50 eggs/ 10-12 ml corn media) were individually allowed to mate with 10 white mutant virgin females (w1118). Females were then kept in different vials and allowed to lay eggs for 4 days. Progeny from each female were counted from day 10- 14 from the first day of laying eggs. Total progeny from each male was calculated and the average for each stock was compared using a t test.

A second more sensitive male sperm exhaustion assay was also performed. Individual males 3 days old reared in controlled density (50 eggs in 10-12 ml of corn media) were crossed with 2-3 days old virgin females. Crosses were made over a period of five days and each day a male was allowed to mate with two virgin females for about 24 hrs. Each female was placed in

independent vial and allowed to lay eggs. Once pupae were observed the females were removed from the vials. Progeny were counted from the first day they eclose till the 15th day from the day cross was made. Average number of progeny for day 1, 3 and 5 were calculated.

4.2 Results

4.2.1 Has Dntf-2r lost any Dntf-2 fucntions?

The parental gene for *Dntf-2r* and its role in nuclear transport has been characterized in quite a lot of detail (see Chapter 1). The heretrospecific complex between rat NTF-2 and canine RanGDP has been crystallized (Stewart *et al.*. 1998) and the interacting amino acids are known. Ntf-2 works in a dimer form as well as interacts with FxFG repeats of nucleoporins. These interactions have also been crystallized and the amino acids forming the interface between two molecules are also known (Bullock *et al.* 1996; Isgro *et al.* 2007). Dntf-2 and the retrogenes were aligned with the rat NTF-2 to explore any possible changes in functional domains (Figure 4.2). Based on the amino acid sequence conservation it is observed that most of the amino acids involved in interaction with RanGDP are conserved between Dntf-2 and Dntf-2r in *D. melanogaster* lineage. In other lineages they represent conservative changes. The amino acids involved in interaction with FxFG repeats of nucleoporins show many conservative changes. Nucleoporins are known to be under positive selection (Presgraves *et al.* 2007), and changes in interacting amino acid of Dntf-2r may be to keep up with the evolving nucleoporins.

Dntf-2r has been evolving under positive selection in *D. melanogaster* subgroup (Betran *et al.* 2003). Other retroduplicates of *Dntf-2* in D. ananassae and D. grimshawi are also shown. The retroduplicate has been evolving under positive selection in D. ananassae and D. atripex (close related to D. ananassae) lineages revealed by a McDonald-Kreitman test (Tracy et. al - submitted). However, from the amino acid sequence alignment it is observed that Dntf-2r has conserved interaction with RanGDP as well as nucleoporins, two most important associations of the parental Dntf-2. Recently, Ntf2 was found to have additional interaction with CapG proteins which are filamentous actin capping proteins and participate directly in the transport of this protein to the nucleus (Katrien Van Impe *et al.* 2008). The interacting residues

in this instance are not known. It is possible that *Dntf-2r* may be evolving under positive selection to attain new interactions and transport additional proteins to the nucleus. It might also have changed to modulate the RanGDP/RanGTP gradient. It is also possible that Dntf-2r is currently interacting strongly with Ran-like but less strongly with Ran. Work by Tracy *et al.* predicted that Ran-like has likely lost some nuclear functions and the transport of Ran-like efficiently in male germline might modulate this other functions of Ran. Related to this it was postulated that this genes might be involved in male germline conflicts (either meiotic drive or TEs and viral resistance). Current meiotic drive role of *Dntf-2r* is explored in Chapter 5.

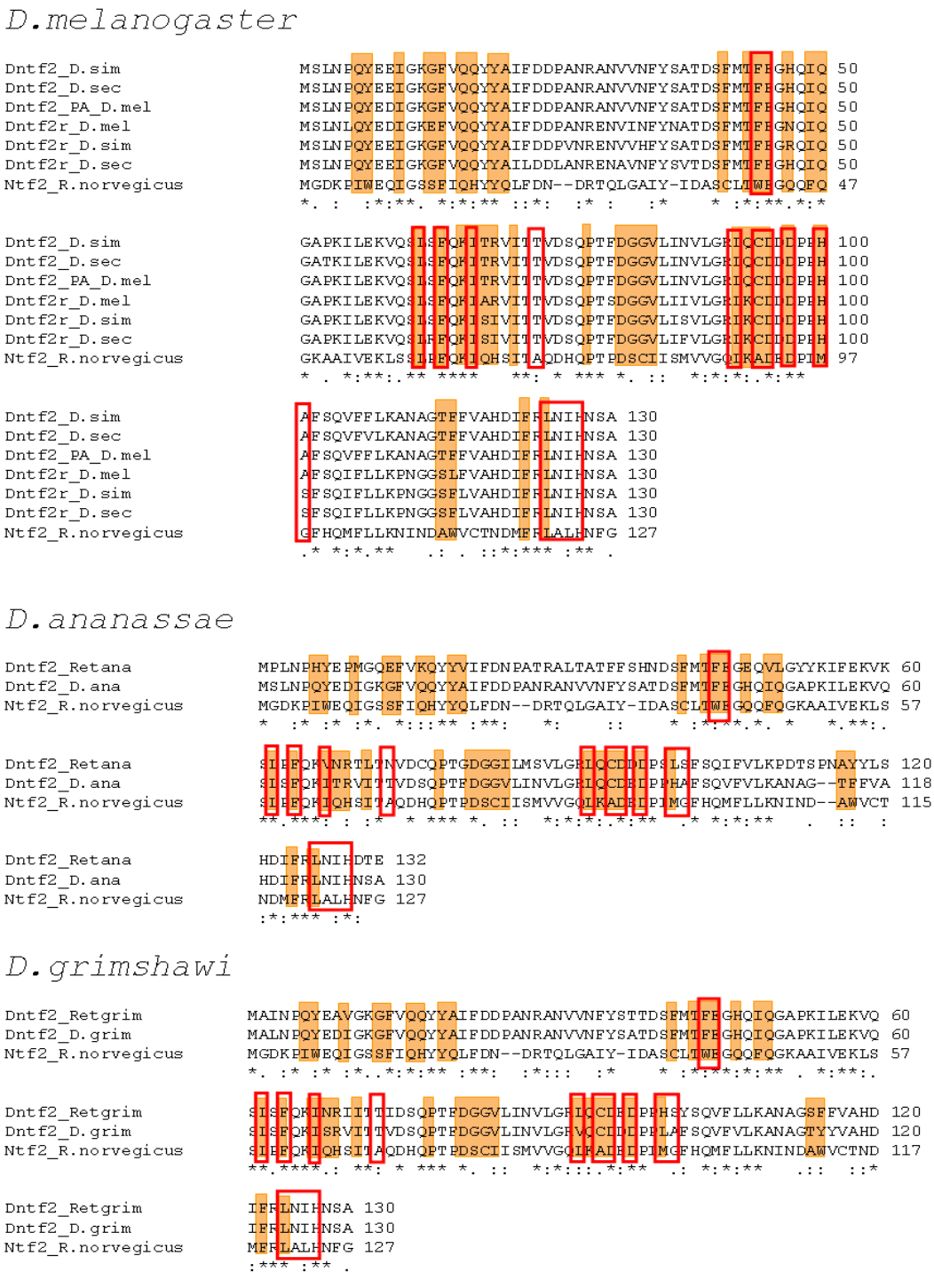


Figure 4.2. Protein alignment of *Dntf-2* retrogenes with the parental gene and with Rat Ntf-2 in *D. melanogaster* and related species, D. ananassae and D. grimshawi. Red box - Interacting interface with RanGDP. Orange – Amino acids interacting with FxFG repeats of nucleoporins. Amino acids interacting with RanGDP are highly conserved between the parent and retrogenes.

4.2.2 Protein expression pattern and cellular localization

Taking advantage of *Dntf-2r*-EGFP fusion protein produced to study the regulatory regions (Chapter 2) we explore the details of its expression pattern and cellular localization to gather additional functional information. In Chapter 2 the testis-specific regulatory region of *Dntf-2r* is described. This regulatory region drives the expression of the fusion gene (green fluorescence in Figure 4.3) exclusively during meiosis (16, 32 and 64 cell stage), spermatid elongation and early mature sperm. Completely mature sperm lacks the green fluorescence. DNA is labeled using Draq5 (see Materials and Methods) and it is shown in red. This allows us to state that the fusion protein localizes mostly in the cytoplasm and also in the nucleus in the stages where nuclear membrane is present. In dividing cells, it is located at the periphery of the cell. Fluorescence is observed in elongating sperm heads but is lost in completely mature sperms (Figure 4.3).

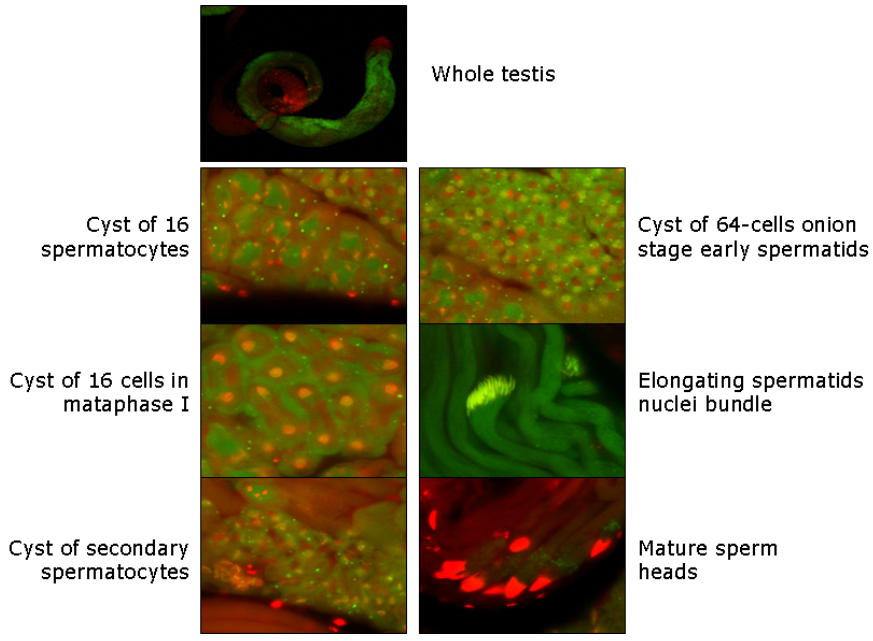


Figure 4.3. The localization of *Dntf-2r*-EGFP fusion protein during spermatogenesis is shown in green. DNA is stained in red using Draq5 (see above). From the staining of whole testis no expression of fluorescence is observed in the tip of the testis.

The observed pattern is consistent with the expectations. As it is a transport protein, we expect it to have both cytoplasmic and nuclear localization. Similar study is under way for the parental *Dntf-2* to study its detail expression pattern and cellular localization during spermatogenesis for comparison.

The above protein fusion localization would recapitulate the wild type cellular localization only if *Dntf-2r* maintains function in the fused protein and it is able to dimerize and interact with the proteins that its is usually interacting in the different cell types. It is known that monomers do not interact with Ran (Bullock *et al.* 1996). To confirm we are looking at a functional fusion we performed co-immunoprecipitation of the fusion protein and try to detect Ran/Ran-like interaction. The results confirm that we are looking at a functional fused protein (see below).

4.2.3 Co-immunoprecipitation/Western blotting

To verify if the prediction that Dntf-2r retains its interaction with Ran/Ran-like, co-immunoprecipitation with immobilized anti GFP antibody was performed. Testes lysate, in non denaturing conditions, from males carrying the fusion *Dntf-2r*-EGFP protein was used for co-immunoprecipitation. An enrichment of proteins binding to *Dntf-2r*-EGFP was observed (Figure 4.4). Beads unconjugated to the anti GFP antibody and incubated with the crude extract did not show any bands indicating that proteins did not bind directly to the protein G coated magnetic beads (negative control). As the immunopercipitate was diluted most of the lighter bands which may indicate non specific binding are lost.

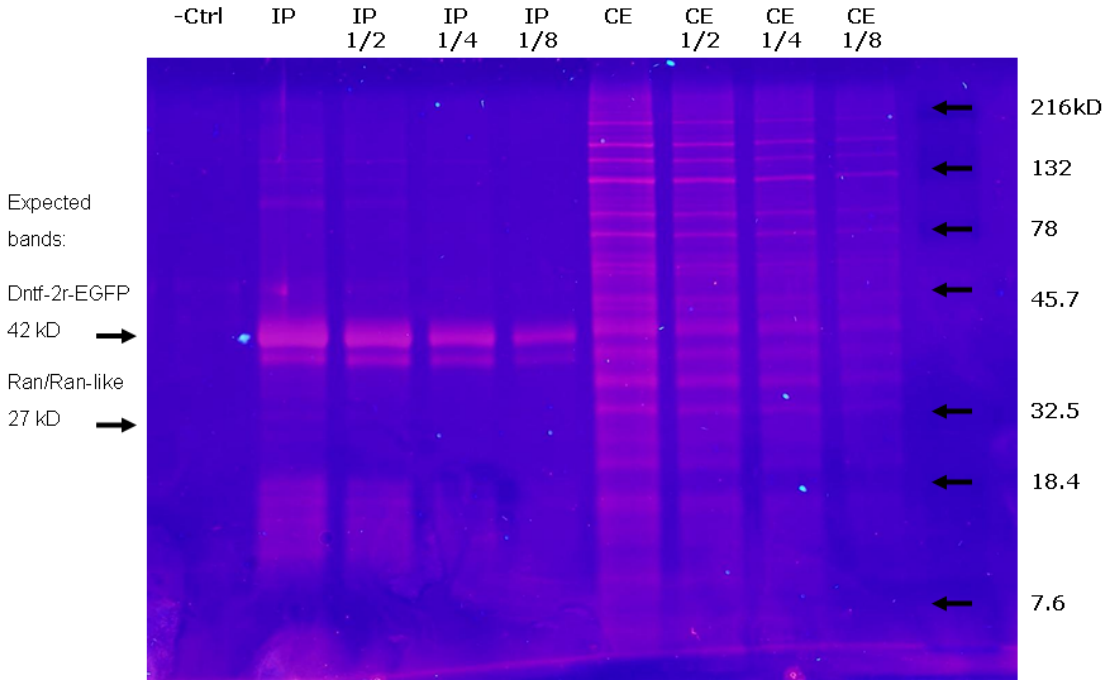


Figure 4.4. SDS Page gel of immunoprecipitates of the protein bound to *Dntf-2r*-EGFP protein. Ctrl – Control. IP – immunoprecipitates with various dilutions CE – crude extract with various dilutions.

The proteins from the gel were transferred to a PVDF membrane and blotted with anti Ran antibody (Figure 4.5). Anti Ran antibody is a rabbit polyclonal antibody and the anti rabbit IgG antibody is used as a secondary antibody. From the figure 4.5, we observe that the anti Ran antibody binds to many proteins at higher concentration. However when diluted only one band is observed indicating that is the band of interest. Ran is a 29 kDa protein, however, the band observed corresponds to higher molecular mass. We know from previous studies that Ntf2 binds to RanGDP, and the GDP can cause the protein to run slower and thus showing a higher molecular mass. However, the observed 42 kDa band corresponds to the expected size of the Dntf-2r-EGFP protein.

To test if the band corresponds to the fusion protein, the membrane was stripped and re-blotted with anti GFP antibody (Figure 4.5). Again the 42 kDa band was tagged. Two scenarios can lead to this result. First, both Ran/Ran-line and fusion protein co-localizes at the

same spot on the SDS page gel. This however, seems less likely as only one band is observed and there does not seem to be any overlap. Secondly, it may be possible that as anti Ran is a polyclonal antibody its may have some non specific binding thus binding strongly to the most abundant protein (i.e. Dntf-2r fusion protein).

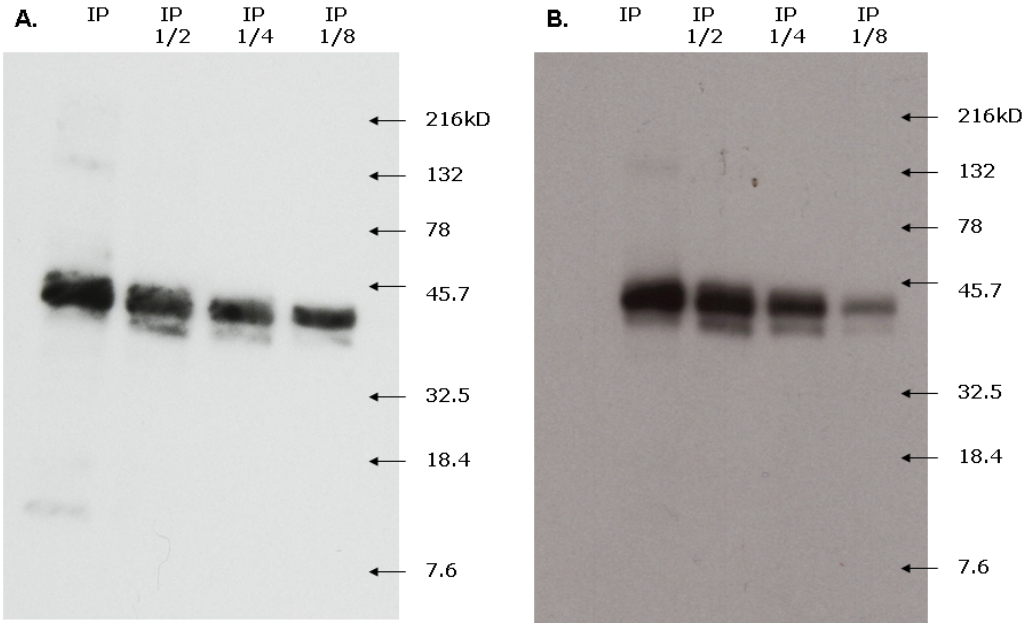


Figure 4.5. Western blot of immunoprecipitates from figure 4.4. A) using anti Ran antibody B) using anti GFP antibody. Higher molecular mass suggest that Ran may be phosphorylated. Ntf2 is known to interact with RanGDP.

In a recent study Ntf2 was tagged with EGFP in the C terminal and was observed to interact with RanGDP (Katrien Van Impe *et al.* 2008). If we assume that anti Ran antibody is specific and RanGDP moves slower, our results also shows that EGFP tag does not interfere with Ntf2 binding with Ran/Ran-like. Also, dimerisation of NTF2 is essential for binding with Ran (Bullock *et al.* 1996). Any association between *Dntf-2r* –EGFP and Ran also signifies that *Dntf-2r*-EGFP exist in dimer form. It can be inferred from this study and previous study (Katrien Van Impe *et al.* 2008) that *Dntf-2r*-EGFP preserves all interactions of wild type Dntf-2r, and any analysis of the fusion protein function and localization will be a good representation of that of

Dntf-2r. We currently do not know if the fusion protein is interacting with Ran or Ran-like because the polyclonal antibodies should recognize both. Additional fusion construct of Ran and Ran-like fused to Red Fluorescent Protein (RFP) are being generated to study this. A fusion construct of *Dntf-2* and EGFP is also being generated. Additional antibodies will be ordered against other nuclear transport proteins to further assess the interactions of all these fusions.

4.2.4 Mutant strains

The most effective study to understand the function of a gene is to create mutant lines and observe any mutant phenotype. Fortunately, two presumably P element knockout lines were created in the BDGP gene disruption project (Bellen *et al.* 2004). One has a P element inserted in the coding region of *Dntf-2r* (EY05573) and the other has P element inserted in the 5' upstream region of *Dntf-2r* (Figure 4.1). The size of the P elements is about 9kb and thus these insertions will disrupt either the coding region or block transcription of *Dntf-2r*. When the stocks arrived they were first checked for homozygosity. All flies with red eyes were homozygous for the P element insertion. Flies with orange eyes indicated heterozygosity. The stock KB00588 was homozygous as all flies had red eyes. EY05573 stock had some flies with orange and white eyes and thus was heterozygous. Both stocks were again fixed for the P element insertion using CyO balancers. Once the lines seem to be fixed for the P element insertion, PCR from DNA sample using primers flanking the insertion site was performed to confirm the homozygosity. The PCR conditions used could amplify short sequences but could not amplify the 9kb P element insert. The exact locations of the P element insertion were identified by using primers in the P element and in the flanking region (see Materials and Methods). Only one combination produced bands. One band was of expected length; however, one was much shorter than expected. On sequencing the bands it was observed that the P element was missing its 5' end as Figure 4.6 indicates.

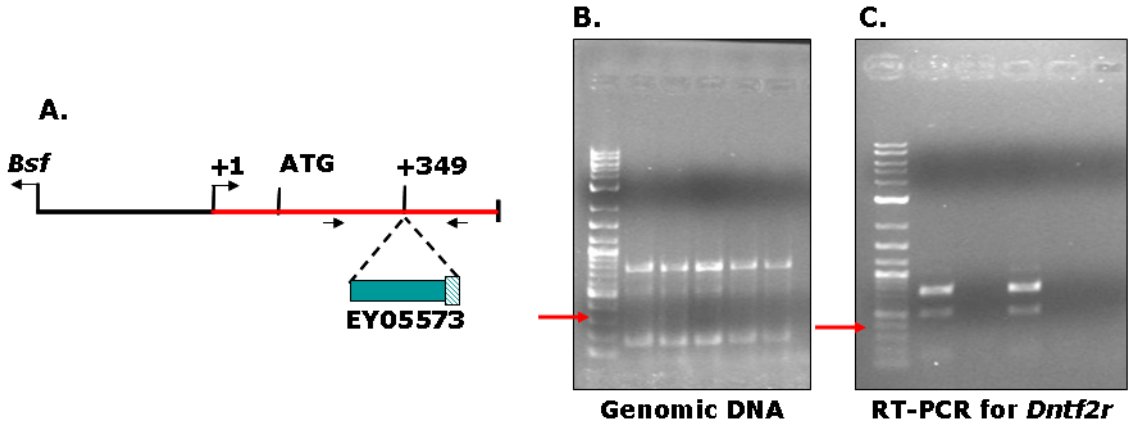


Figure 4.6 EY05573 P element insertion analysis. A) Diagram showing the insertion site of P element (EY05573) and primer positions (black arrows). B. 5 lines were fixed for P element insertion EY05573. PCR from genomic DNA using flanking primers did not show any band of expected length (Red arrows indicate the expected band length). C. RT-PCR results from two of the fixed lines did not show any product of expected length. Some spurious bands were observed. Sequences from these bands indicated non specific primer binding.

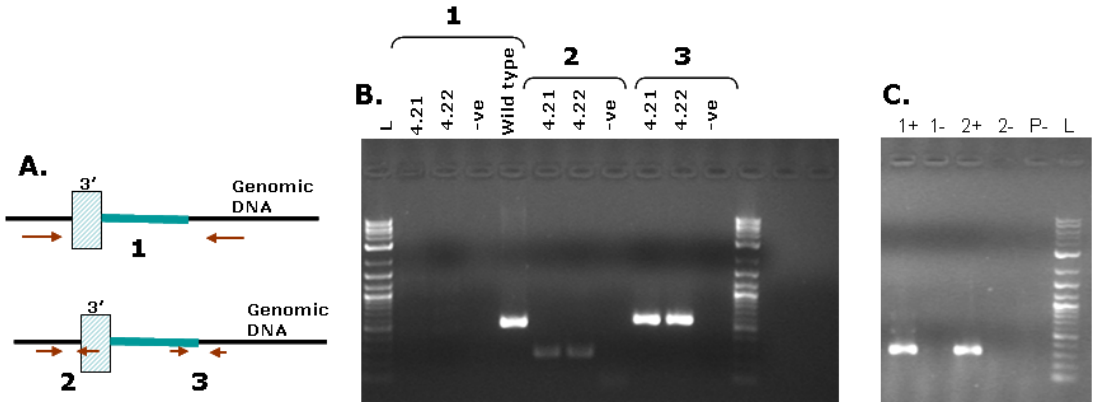


Figure 4.7 KG00588 P element insertion analysis. A. Diagram showing primer combinations used. B. PCR from genomic DNA from KG00588 using the primer combinations indicated in A. Two lines were fixed for the P element insertion (4.21 and 4.22) –ve band indicates the PCR negative. C. RT-PCR result for the lines fixed. (1+) 4.21 RT positive, (1-) 4.21 RT negative, (2+) 4.22 RT positive (2-) 4.22 RT negative, (P-) PCR negative. Both lines show expression of *Dntf-2r*.

RNA was extracted from the lines fixed and RT-PCR was performed to check the expression of *Dntf-2r*. Line EY05573 did not give any band of expected length thus indicating that the intact coding region was disrupted. Another longer band was observed and sequenced

indicating non specific primer binding in absence of *Dntf-2r*. RT-PCR from line KB00588 shows expression of *Dntf-2r* (Figure 4.7). In this line, P element is inserted between the promoter region and transcription start site however still the mRNA is produced. *Dntf-2r* is suspected to use another transcription start site (TSS) in the P element. 5'RACE experiment was performed to identify the TSS. As expected the TSS was found to be within the P element suggesting that some random promoter sequence embedded in the P element was used (Figure 4.8).

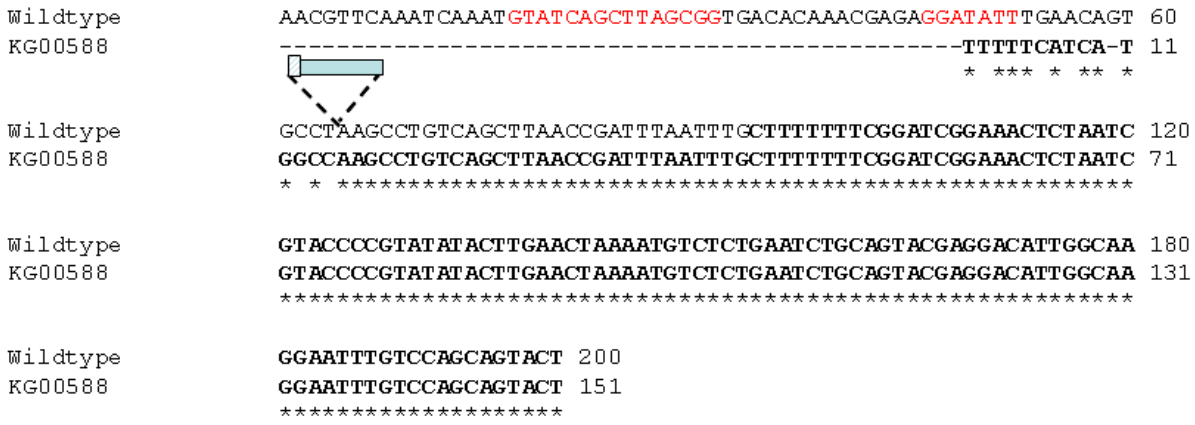
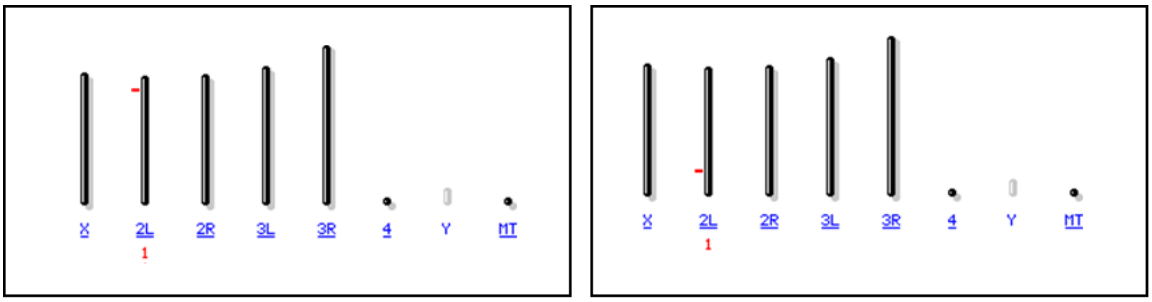


Figure 4.8. Results from 5'RACE for *Dntf-2r* in P element insertion line KG00588. Bold letters indicate the mRNA. *D. melanogaster* wildtype was known (Betran and Long 2003). mRNA from line KG00588 begins in the P element.

While working with the flies of this line, it was observed that they had dark red eyes which were red even when crossed with a white mutant (heterozygotes). Another P element insertion on the second chromosome was suspected. To confirm this observation an inverse PCR was performed (see material and methods). As expected another P element insertion on the left arm of chromosome 2 was observed (Figure 4.9). Because of the second P element insertion it was too complicated to work with this line so only EY05573 was used for further studies.



Insert II ***Dntf-2r***

Figure 4.9. Diagram showing the two insertion sites of P element in line KB00588. Red mark on the chromosome 2L indicates the insertion site.

4.2.5 Knockout Phenotype

Dntf-2r knockout line – EY05573 was used to detect any phenotypic changes from the wild type. As P element insertion and background by themselves may have an affect on phenotype, EY12961 was used as a control. All flies were fertile and did not show any drastic spermatogenesis phenotype when looking at the different cell types under phase contrast (Figure 4.10).

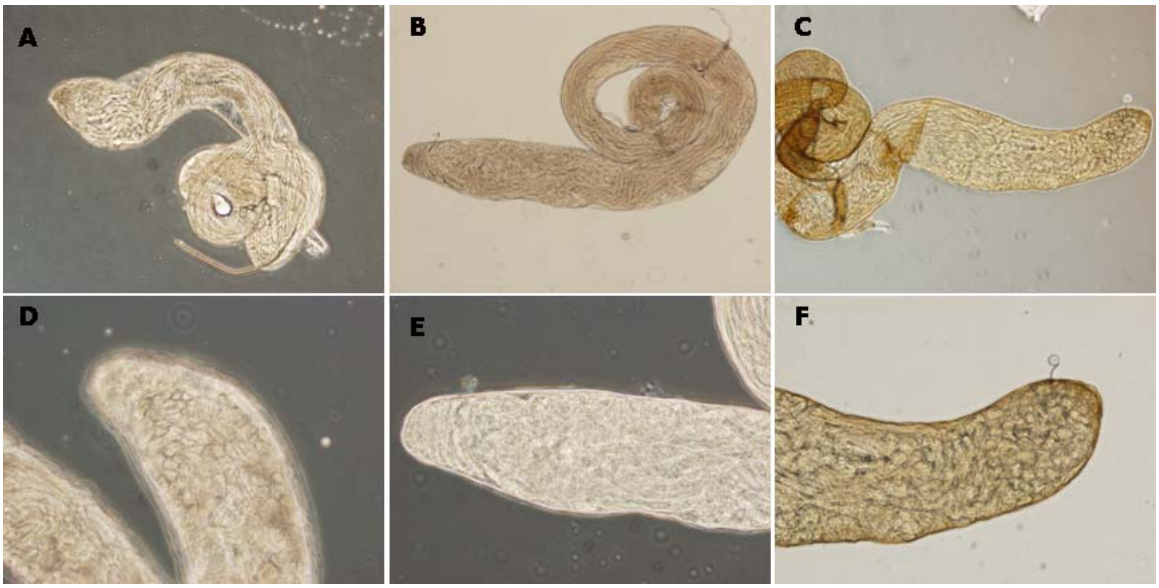


Figure 4.10. Phase contrast pictures of testis. 2-4 days old males of EY05573, EY12961 and wildtype. A, B, C - EY05573, EY12961 and Wild type respectively. D,E,F – Testis tip of EY05573, EY12961 and wildtype respectively.

Figure 4.10 shows the whole testis and the tip of the testis producing the expected cell types. Actively diving spermatocytes are located close to the tip and we see that in all three strains. Apart from the fact that wild type *D. melanogaster* testis are yellow in color whereas the testis of the mutants are white because the are w-, no obvious phenotypic defect is observed in the knockout line during spermatogenesis. This suggests that knocking out *Dntf-2r* does not have any big effect. It is possible, however, that a knocking out *Dntf-2r* has more subtle effect on fertility. A fertility assay was performed to assess that.

4.2.6 Fertility assay

Two fertility assays were performed for *Dntf-2r* knockout line – EY05573. As a control another line having P element insertion in the second chromosome was used (EY12961). A genomic view of this insertion shows that it is not close to any known functional region (Figure 4.1). EY12961 was obtained by the same procedure as EY05573 and came from the same genetic background. Results of fertility assay are shown in figure 4.11. Average number of progeny is calculated from crosses between three males and ten females (see Materials and Methods). A t test is performed to compare the two averages. A P value of 0.59 is obtained indicating that the difference is not significant, revealing that fertility is not affected in the mutant strain compared to another strain of similar background.

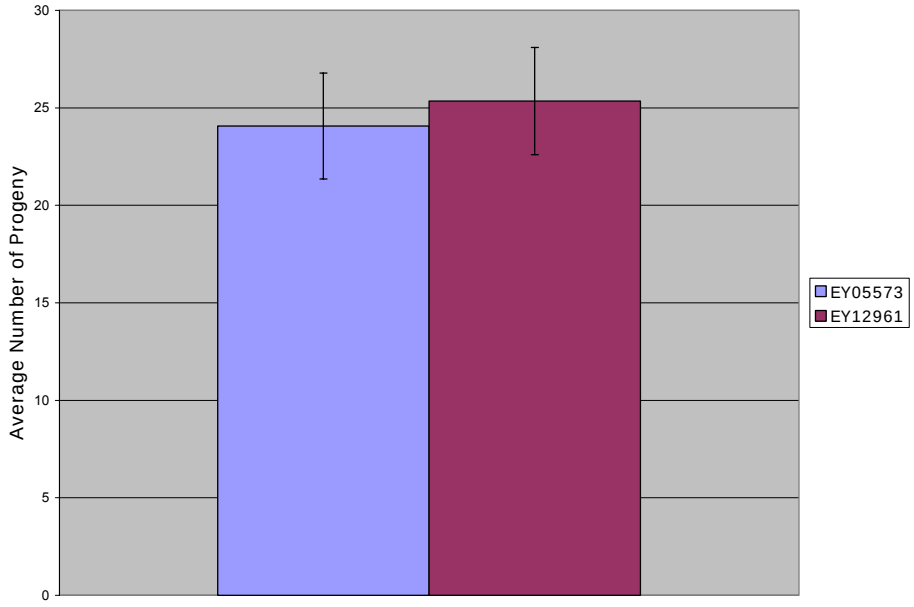


Figure 4.11 Result of male fertility assay. The difference in fertility of the males of EY05573 and EY12961 was not significant (P = 0.59)

As no significant difference was obtained from the fertility assay, a more sensitive male sperm exhaustion assay was performed. This is an assay that measures how males recover from mating and reveals if sperm is exhausted and not replenish after mating (Sun *et al.* 2004). Three males from each line were individually mated with two virgin females everyday over a period of five days (see Materials and Methods). This experimental design assumes that the males will probably mate and remate with the different females using a lot of sperm every day and with time the males might show depletion in number of sperms. The results are shown in figure 12. Lowering of average number of progeny with time is not observed in either strain and the two strains behave equally at every point in time. The effects we see (i.e. increase of progeny with time) are likely due to the genetic background and not to the P-element insertions. There is a trend that EY05573 is always lower than EY12961 and more males could be counted to see if that becomes significant.

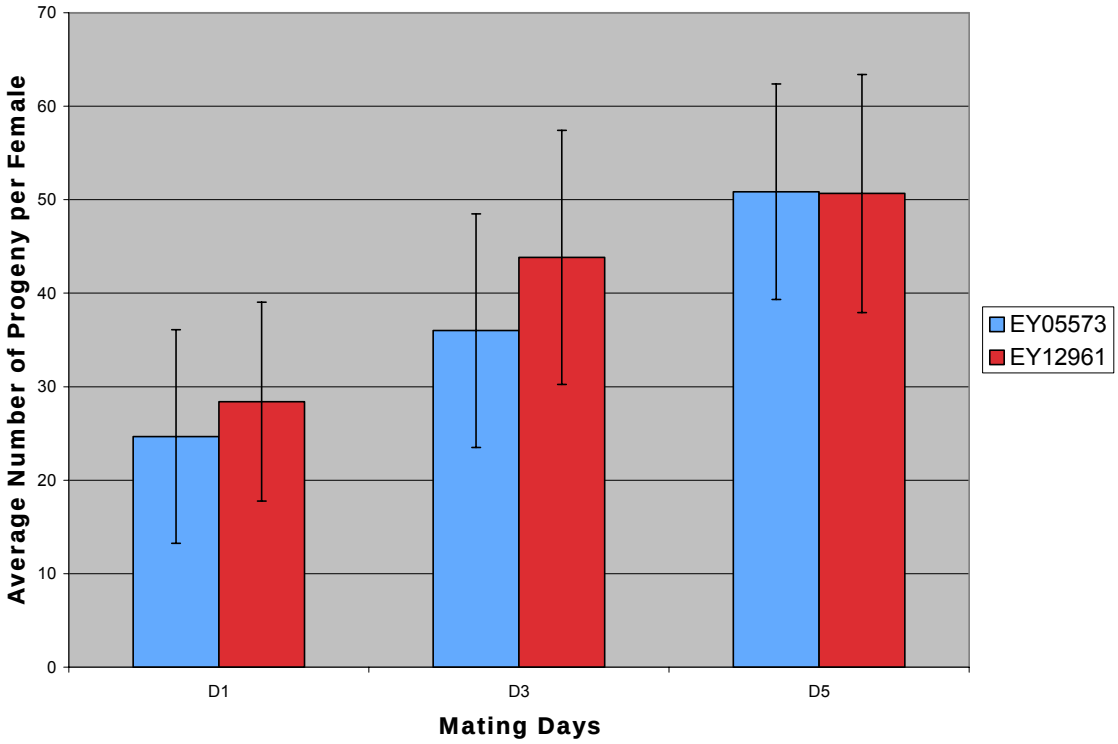


Figure 4.12 Results from male sperm exhaustion assay. Average number of progeny per female was not found to be significantly different in the two species. P value obtained for each day is 0.56, 0.28 and 0.98 respectively.

4.3 Discussion

Dntf-2r originated from a gene with an essential function *Dntf-2*. A complete knockout of the parental gene causes lethality (Bhattacharya *et al.*2002). A partial knockout has very few surviving progeny which shows defects in eye development and immune system, however no effect on fertility was observed ((Bhattacharya *et al.* 2002) and see Chapter 1). As described earlier *Dntf-2* is located on the X chromosome and has given rise to autosomal retrogene *Dntf-2r*, and it was proposed that this new gene might be replacing the parental gene during X inactivation in spermatogenesis (Bertrand *et al.* 2002) or be a better copy for male germline due to sexual antagonism (Wu *et al.* 2003). However, in both hypotheses the retrogene will have an essential function and replace the parental gene during spermatogenesis.

Here we infer that *Dntf-2r* likely retains at least some functional abilities of the parent and interaction with Ran/Ran-like during meiosis and spermatid individualization, however it does not seem to be essential during spermatogenesis. Knocking out of *Dntf-2r* does not show any defect in phenotype or in fertility. It seems essentially dispensable for the organism. Thus, replacing the parental during X inactivation or having a copy to perform an important function better because of sexual antagonism does not seem to be the factors driving the evolution of this gene.

The role of nuclear transport in genetic conflicts (meiotic drive or related to TE and viral expansion) is another selective force that has been suggested to explain the origin and evolution of this gene (Presgraves 2007). As described earlier it seems that Ran - the protein that interacts with both *Dntf-2* and *Dntf-2r* has also given rise to a retrogene *Ran-like*. Both *Dntf-2r* and *Ran-like* are evolving under positive selection. We know now that *Dntf-2r* is expendable for *D. melanogaster*. Similarly, *Ran-like* also seem to be expendable as it has lost many functional domains of the parent and is a pseudogene in *D. yakuba* (Tracy et. al. - submitted). This further strengthens the argument of genetic conflict being the driving force behind their evolution. In the next chapter we address the possible current meiotic drive role of *Dntf-2r* using segregation distorter system in *D. melanogaster*.

In any case more detailed fertility analyses should be carried out possibly backcrossing to a white mutant strain that shows exhaustion to see the effects of this mutant in other backgrounds. We should also observe the fusion protein in heterozygotes to see if there is postmeiotic expression. In addition, if *Dntf-2r* is dispensable, is the parental gene carrying nuclear transport functions during spermatogenesis? The *Dntf-2*-EGFP fusion should answer this question. Detail interactions should also be tested between all these genes and other nuclear transport genes using additional fusion proteins and additional antibodies. The mutant could also be tested in a strain with active TEs for more or less transposition.

CHAPTER 5

POTENTIAL EFFECTS OF *DNTF-2R* IN MEIOTIC DRIVE

According to Mendel's first law homologous chromosomes have equal probability of forming a gamete and hence transmitting to the offspring. This law holds in males and females for different reasons. In females, the law holds if the chromosomes segregate randomly in meiosis I and II and end with equal probability in the oocyte. In males, this law holds if all meiotic and postmeiotic cells mature independently of the chromosomes they carry.

Meiotic drive is an exception to this law where one chromosome or one gametic type is over or under represented in the gametes formed during meiosis and thus in the next generation. Again, this phenomenon basically arises from different reasons in males and females. In females, the different probability arises from centromeric drive (i.e., the best centromere ends in the cell that will give rise to the egg). In males, it arises from gamete "competition" where the gametes carrying the "driver" allele is healthy, and the one carrying the alternative allele is subjected to dysfunction (Lyttle 1991). Examples of such meiotic drive systems have been described in natural populations of a variety of organisms such as plants, fungi, insects and mammals (Lyttle 1991).

In Drosophila melanogaster, there is a naturally occurring meiotic drive system called Segregation Distorter (*SD*). It was first discovered about five decades ago (Sandler *et al.* 1959), and since then has been a subject of interest for many researchers. It is found in 3-5% of natural populations of *D. melanogaster* (Reviewed in Kusano *et al.* 2003). In this system, males heterozygous for an *SD* bearing chromosome (*SD*/*SD*$^{+}$) preferentially transmit it to almost 100% of the offspring (Ganetzky 1977). *SD* is believed be a selfish locus that successfully transmits itself to the next generation by disrupting the sperms carrying the wild type chromosome (*SD*$^{+}$).

SD^{+} bearing sperms show defective lysine-rich to arginine-rich (protamine or protamine-like) histone transition (Hauschteck-Jungen *et al.*1982) which causes abnormalities in chromatin condensation and compaction (Tokuyasu *et al.* 1977). Thus, *SD*+ carrying sperms do not progress to elongation and maturation.

The *SD* locus comprises of *Sd* gene, which is the main driver, along with other components like the *Enhancer of SD* (*E(SD)*), *Modifier of SD* (*M*(*SD*)) and *Stabilizer of SD* (*St*(*SD*)) that altogether enhance the driving *capacity* of *Sd*. It is important to note that *SD* refers to the whole chromosome and *Sd* refers to the particular gene on the *SD* chromosome. The genetic linkage of all the components is crucial for *SD* drive (Charlesworth *et al.* 1978). The *SD* locus is clustered around the centromere of chromosome 2, which is an area of low recombination. The chances of crossing over are further lowered as *SD* chromosomes have incurred pericentric or paracentric inversions. The centromeric location coupled with the inversions allows the various components of *SD* to be in linkage disequilibrium and makes up a successful chromosome showing a strong drive (Figure 5.1).

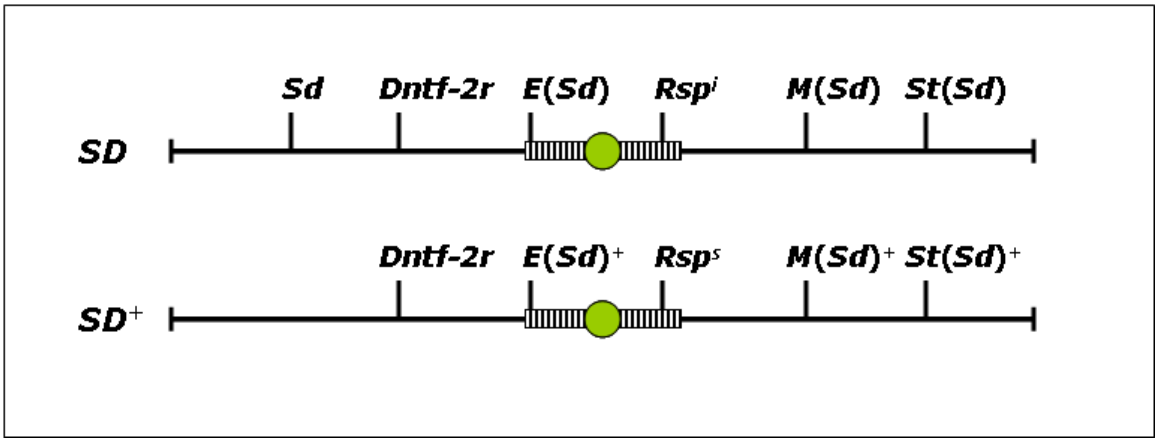


Figure 5.1. Chromosomal organization of *SD* chromosome. All components involved are clustered around the centromere. *SD*$^{+}$ - Chromosome showing distorted transmission, *SD* – segregation distorter gene, *E(SD)* – Enhancer of SD, *M*(*SD*) – Modifier of SD, *St*(*SD*) – Stabilizer of SD, Rsp – Responder locus.

The target of *Sd* is the responder locus (Rsp) that maps to heterochromatic region on chromosomal arm 2R (h39 locus) close to the centromere (Pimpinelli *et al.* 1989). It comprises

of XbaI repeats (i.e., 240bp AT rich sequence repeats). The repeats are termed as XbaI repeats because each 240bp unit is flanked by XbaI restriction site and after digestion with XbaI restriction enzyme a ladder of 240bp is observed in southern blots (Wu *et al.* 1988). Each 240bp XbaI repeat is a dimer of 120bp repeats in head to head or head to tail orientation (Houtchens *et al.* 2003). A chromosome can be classified as responder insensitive (Rsp^{i}), responder intermediate sensitive (Rsp^{in}), responder sensitive (Rsp^{s}) or responder super sensitive (Rsp^{ss}) based on the number of 240bp repeats it harbors. Sensitivity of a chromosome is directly proportional to the number of repeats. Insensitive responder harbors very small number (25-30) of repeats. 200-500 repeats make a chromosome sensitive and more than 1000 repeats make a super sensitive chromosome. As is obvious, the responder carried on the *SD* chromosome is insensitive. Experimentally recombining a sensitive responder into *SD* chromosome forms a suicidal combination (Hartl 1974).

Sd is a truncated duplicate of RanGAP (Ran-GTPase Activating Proteins) on the left arm of chromosome 2 (2L). RanGAP is a major component nuclear transport (see Chapter 1). RanGAP is predominantly localized in the cytoplasm where it hydrolyses RanGTP into RanGDP. Transcription of *Sd* forms an aberrant mRNA of *RanGAP* which forms a mutated protein (Sd-RanGAP) that lacks 234 amino acids from the C terminus (Merrill *et al.* 1999). Sd-RanGAP retains the enzymatic activity of RanGAP, however it is mislocalized in the nucleus (Kusano *et al.* 2001). Two features responsible for the cytoplasmic localization of RanGAP are deleted from Sd-RanGAP. First, the lack of a conserved K residue in the C terminal end that is a site for a ubiquitin-like molecule (SUMO) posttranslational modification. This modification is essential for the localization of RanGAP to the nucleus in association with the nuclear pore (Mahajan *et al.* 1998). Secondly, Sd-RanGAP lacks the nuclear export signals (NES) carried by RanGAP. Both these deletions cause the Sd-RanGAP to mislocalize to the nucleus. The presence of enzymatically active Sd-RanGAP in the nucleus causes the hydrolysis of RanGTP which in turn creates an imbalance in the RanGTP – RanGDP gradient across the nuclear membrane. Although the exact mechanism is not known, this altered gradient is believed to be

responsible for the drive. Interestingly, in the absence *Sd*, an over-expression of wild type *RanGAP* in the germline using the β2tubulin promoter, shows the same distorting ability as the *Sd-RanGAP*. A germ-line specific over expression of *Ran* or *RanGEF* (Ran GTpase Exchange Factor also called RCC1), increases the RanGTP concentration in the nucleus and restores the gradient, suppressing the distortion caused by *SD* (Kusano *et al.* 2002). Distorting ability can also be obtained by doubling the dose of Enhancer of segregation (*E(SD)*) in the absence of *Sd* (Temin 1991). It is believed that *E(SD)* causes distortion by facilitating the nuclear import of wild type RanGAP and hindering its export. Thus, *E(SD)* enhances the drive caused by *Sd*. Besides *E(SD)*, there can be other factors associated with the *SD* locus, which may enhance the driving ability of *Sd*. These may not have a dramatic effect on the drive and thus have not been identified. As previously stated (Chapter 1), *Dntf-2r* is located between *Sd* and *E(SD)* (Figure 5.1) and thus it is close to the *SD* locus. We would like to see if the *Dntf-2r* allele present on the *SD* chromosome is different from the rest of the population (See below). If it is different, this allele may specifically function as a helper when associated with *SD*. We want to characterize the sequence of the particular allele and its level of expression.

Basically, any change that causes the imbalance of RanGTP – RanGDP gradient across the nuclear membrane, may cause segregation distortion. In a study by Presgraves, wild type RanGAP was found to be under positive selection and was proposed that the selective sweeps could be due to the potential role as suppressor or driver of RanGAP in similar meiotic drive systems (Presgraves 2007). He suggested that other genes that could alter the RanGTP concentration gradient (e.g., RanGAP) may be caught up in constantly appearing genetic conflicts similar to *SD*. Six nucleoporin genes were also found to be evolving under positive selection and selective sweep could be one of factors influencing these genes as well. Direct involvement of various nuclear transport genes and chromatin condensation genes in segregation distortion has been observed (McElroy *et al.* 2008), suggesting again that any changes that influence transport across nuclear membrane and in turn chromatin condensation, could result in segregation distortion. At its origin, *Dntf-2r* could essentially act as an extra dose

of *Dntf-2* that could influence nuclear transport. In *D. melanogaster*, *Dntf-2r* has male germline specific expression using a regulatory element similar to the *β2 tubulin* upstream element (β2UE1) promoter (see Chapter 2). This is interesting as over expression of *E(SD)* using the same promoter causes distortion even in absence of *Sd* (Kusano *et al.* 2002) revealing that *Dntf-2r* is expressed in the cells where it can affect *SD*.

Interestingly, as introduced in Chapter 1, there has been recurrent recruitment of *Dntf-2* retrogenes in two other lineages and also show mostly male biased expression (Tracy *et al.* - submitted). In all cases, these retrogenes are X to autosome and several selective forces are discussed in Chapter 1. Interestingly, in all lineages where it was possible to perform McDonald-Kreitman test (excluding *D. grimshawi* where there are no strains available to gather polymorphism) *Dntf-2* retrogenes seem to be under positive selection. This is in agreement with the mode of evolution proposed for genes involved in meiotic drive functions as proposed by Presgraves (Presgraves 2007). Although other conflicts or new functions could also explain this mode of evolution, in this chapter, we try to address if *Dntf-2r* has currently any effect in meiotic drive using *Dntf-2r* knockout chromosome in the *SD* background by observing the consequences on the drive of loosing one copy of *Dntf-2r*.

5.1 Materials and methods

5.1.1 Drosophila strains used

Besançon strain of *D. melanogaster* is used as wild type. P element insertion line EY05573 (Chapter 4) is used as a knockout line for *Dntf-2r*. EY12961 line (Chapter 4) is used for comparison with EY05573 because it carries a P element insertion in another region but was generated simultaneously (Bellen *et al.* 2004). *SD* chromosome comes from the stock SD-5 which was isolated from natural populations and has been described to be a strong driving chromosome (Wu *et al.* 1988). *Cn bw* is used as a standard sensitive chromosome. All stocks were obtained from Bloomington stock center.

5.1.2 Chromosomes used

The following chromosomes were used in this study:

SD chromosome: In(2R)SD5, In(2R)NS, Dp(2;2)RanGap[SD], RanGap[SD]/SM1 SD-5 recovered from natural population in Madison, Wisconsin (Sandler *et al.* 1959). It carried two non overlapping inversions in 2R.

Test chromosomes: a) *cn bw*: (cn[1] bw[1]) standard chromosome to test for drive . It carried a sensitive responder. b) *Dntf-2r* knockout chromosome (EY05573): y[1] w[67c23]; P{w[+mC] y[+mDint2]=EPgy2}Ntf-2r[EY05573] This chromosome carries a P element insertion in the *Dntf-2r* CDS [23]. c) Control chromosome (EY12961): y[1] w[67c23]; P{w[+mC] y[+mDint2]=EPgy2}EY12961 This was generated using the same P element (EPgy2) and the same starter stock (w[67c23]) as EY05573 by Bellen's lab (Bellen *et al.* 2004). It carries a P element insertion on the second chromosome but does not disrupt any known functional region (figure 5.2). We initially assumed both the chromosomes may have the same responder allele and any difference in distortion will indicate the affect of *Dntf-2r* knockout on distorting ability of SD-5 but see our estimates of the number of XbaI repeats below. The best would have been to compare the *Dntf-2r* knockout chromosome with the same chromosome after the excision of the P element but the P element lost one of the ends and can not be excised. So we used EY12961 for comparison but see below for results and discussion.

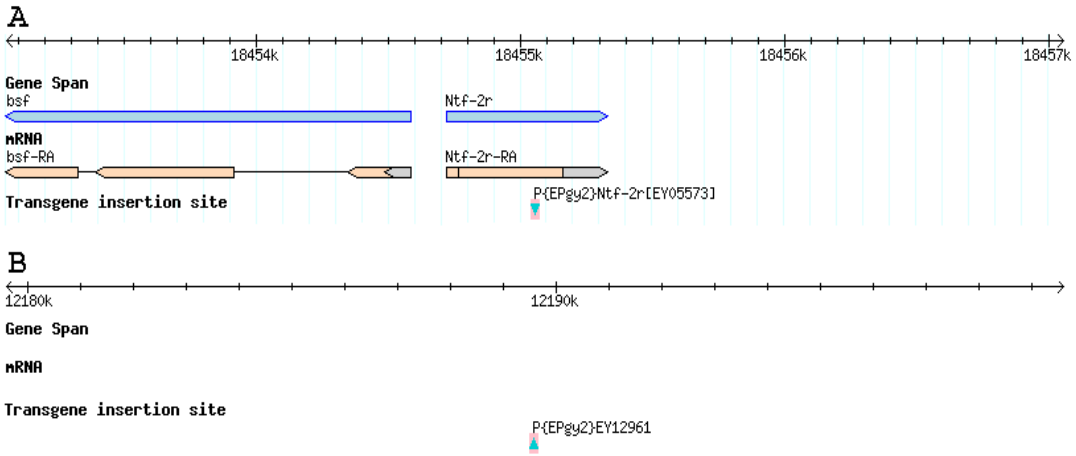


Figure 5.2. Schematic showing insertion site of the two P elements EY05573 and EY12961 (Blue arrow). EY05573 disrupts the *Dntf-2r* coding region and EY12961 is not in close proximity to any known functional region.

5.1.3 Tests for distortion

SD-5 males were crossed with virgin females from line EY05573 and line EY12961. The male progeny heterozygous for *SD* and test chromosome (In(2R)SD5, In(2R)NS, Dp(2;2)RanGap[SD], RanGap[SD]/ P{w[+mC] y[+mDint2]=EPgy2}Ntf-2r[EY05573] or In(2R)SD5, In(2R)NS, Dp(2;2)RanGap[SD], RanGap[SD]/ P{w[+mC] y[+mDint2]=EPgy2}EY12961) were recognized as males with orange eyes and straight wings (Figure 3). It should be noted that the heterozygotes for *Dntf-2r* knockout still have one intact copy of the gene on the *SD* chromosome. Six to eight of these heterozygous males were crossed with 9 to 12 white mutant w1118 virgin females and the progeny was scored for white (i.e. carrying *SD* chromosome) and orange eyes (i.e. carrying EY05573 or EY12961 second chromosome) (Figure 3). The distorting ability of a chromosome is calculated as K value (i.e. the proportion of progeny carrying the *SD* chromosome). To check if the test chromosomes in absence of *SD*-5 show any deviation from the expected K value of 0.5 (i.e., existence of any inviability effects), males heterozygous for test chromosome and chromosome from w1118 background were crossed with w1118 virgin females and progeny were scored for white and orange eyes.

To check the distorting ability of *SD*-5 against a sensitive chromosome, 6-8 males heterozygous for *SD* chromosome and *cn bw* were crossed with 9-12 *cn bw* virgin females. Progeny carrying *SD* chromosome show red eye phenotype and those homozygous for *cn bw* have white eyes. The viability for the *SD*-5/*cn bw* heterozygote flies is tested using reciprocal cross (Figure 5.3)

Figure 5.3. Crosses made to test for the distorting ability of *SD*-5 against EY05573 and EY12961.

5.1.4 DNA isolation

DNA was extracted from 15-20 adult males heterozygous for *SD* and *Dntf-2r* knock out chromosome using the Puregene DNA purification system – cell and tissue kit from Gentra systems. This DNA was used to PCR amplify regulatory region and coding region of *Dntf-2r* allele on the *SD* chromosome. DNA was also extracted from 20-30 flies from *SD*-5, insertion line EY05573, EY12961 and *cn bw*. This DNA was used to characterize the type of repeat and repeat number using slot blot and Southern blot hybridization analyses.

5.1.5 PCR

To amplify the *Dntf-2r* allele on the *SD* chromosome DNA from heterozygous males for *SD*-5/*Dntf-2r* knockout was used. DNA was amplified from its flanking sequence using primers 5'CGCGCCTATCGATGTTTACCT3' and 5'GCTGGGCGTCTTTGGATGTC3'. We observe from chapter 4 that primers flanking the P element insertion do not give the expected band size after PCR amplification. If a band if expected length is obtained, we assume that it represents the allele on the *SD* chromosome (Figure 5.4). The band of expected length was sequenced.

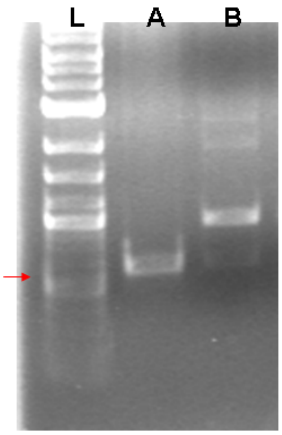


Figure 5.4. PCR amplification of *Dntf-2r* flanking region. L – Ladder; A – Heterozygote for *SD* and EY05573; B – Homozygote for EY0557. Homozygote for *Dntf-2r* knockout gives a spurious band due to non specific primer binding. Arrow indicates the expected length.

5.1.6 Probe preparation

Probe was designed based on the sequence of a well studied clone H_0 (Wu, Lyttle *et al.* 1988). A 353 bp sequence from the known responder region (accession number: M21213) was used for making a synthetic clone (Figure 5.5). The company (Gene script) delivered the sequence in a clone named as pUC57 which is used in the slot blot analysis. This clone was sequenced to check the probe sequence. The 353bp region was extracted from the clone by restriction digestion using EcoRI and BamHI restriction enzyme. This region was used to make the probe. The probe was labeled with (α32P) dCTP and cleaned using Quiagen PCR purification kit.

```
M21213.1        TCAACTGGTACGCAAAAACAGTAAATTGCCTAAGTTTTACATTTTAAGCGGTCAAAATGG 600
probe           ------------------------------------------------------------

M21213.1        GTAATTTTCCGATTTCAAGTACCAGACAAACAGAAGATACCTTCTACAGATTATTTAAAC 660
probe           ----------------------------------------TCTACAGATTATTTAAAC 18
                                                        ******************

M21213.1        CTAGTACACAAAAACAATAAATTGACTAAGTTATGTCATTTTAACGGTCAAAATGGGTGA 720
probe           CTAGTACACAAAAACAATAAATTGACTAAGTTATGTCATTTTAACGGTCAAAATGGGTGA 78
                ************************************************************

M21213.1        TTTTTCGATTTCAAGTACCAGGCGAAAAGAAGACACCTTCTAGAGATTCTGTTCAACTGG 780
probe           TTTTTCGATTTCAAGTACCAGGCGAAAAGAAGACACCTTCTAGAGATTCTGTTCAACTGG 138
                ************************************************************

M21213.1        TAAGGAAAAACAGTAAATTGCCTAAGTTTTACATTTTAAGCGGTCAAAATGGGTGATTTT 840
probe           TAAGGAAAAACAGTAAATTGCCTAAGTTTTACATTTTAAGCGGTCAAAATGGGTGATTTT 198
                ************************************************************

M21213.1        CCGATTTCAAGTACCAGACAAACAGAAGACTATAACTATAAATTGACTAATATAAATAAA 900
probe           CCGATTTCAAGTACCAGACAAACAGAAGACTATAACTATAAATTGACTAATATAAATAAA 258
                ************************************************************

M21213.1        TTGACTAAGTTGACTAAATTGACTAAGTTATGTCATTTTAACGGTCAAAATGGGTGATTT 960
probe           TTGACTAAGTTGACTAAATTGACTAAGTTATGTCATTTTAACGGTCAAAATGGGTGATTT 318
                ************************************************************

M21213.1        TTCGATTTCAAGTACCAGGCGAACAGAAGACACCTTCTAGAGATTCTGTTCACACTGGTA 1020
probe           TTCGATTTCAAGTACCAGGCGAACAGAAGACACCT------------------------ 353
                **********************************
```

Figure 5.5. Alignment showing the sequence similarity between the probe sequence and the responder sequence.

5.1.7 Southern Blot

Six µg of DNA from stocks EY05573, EY12961 and *cn bw* were digested overnight using EcoRI restriction enzyme. Once completely digested, the DNA was run in 0.8% agarose gel overnight. After transferring to nylon membrane, the DNA was fixed to the membrane using UV crosslinker at 1245 J energy for 45 seconds. It was allowed to hybridize to the probe overnight followed by stringent washing conditions. First wash with 2X SSC with 0.1% SDS for 10 minutes at room temperature. Two times second wash with 0.5X SSC with 0.1% SDS for 20 minutes at 65^{o}C.

5.1.8 Slot Blot

The concentration of genomic DNA from line EY05573, EY12961, *cn bw* and *SD*-5 along with the pUC57 clone was determined using nano drop and DNA was diluted to appropriate concentration in water. All the dilutions were brought up to 220µl with water. One

hundred and sixty µl of 1M NaOH and 20 µl of 200mM EDTA were added to make a final concentration of 0.4M NaOH and 10mM EDTA. Samples were denatured for 10 min in a water bath at 100^{o}C before vacuum blotting to nylon membrane. After transfer, DNA was fixed to the membrane by UV crosslinker at 1250J energy for 45 seconds. Hybridization and washing steps were same as during Southern blot.

5.2 Results and Discussion

5.2.1 Meiotic drive crosses

First, we determined the distorting ability of *SD*-5 by crossing *SD*-5/*cn bw* males with *cn bw* virgin females. The resulting K value was 0.998 which suggest that *SD*-5 is a strong driver (Table 5.1). This value is corrected for viability as the reciprocal cross results show deviation from the expected 0.5.

Crosses between *SD*-5/EY05573 males and w1118 virgin females show a K value of 0.58 (Table 5.1). The control cross between males heterozygous for *SD*-5/EY12961 and w1118 virgin females gave a K value of 0.86 (Table 5.1). A reciprocal cross was made to check for the viability of mutant chromosome in heterozygous state. A relative viability effect was estimated by w = number of white eye flies/number of orange eye flies from the reciprocal cross. The corrected k value k_c = Number of progeny with *SD* chromosome/ [number of progeny with *SD* + (number of progeny with mutant chromosome/w]). The k value from the test and control cross (Table 5.1) were not corrected for viability as no deviation from the expected 0.5 was observed in the reciprocal cross.

This difference in K value could suggest that knocking down *Dntf-2r* reduces the distortion efficiency of *SD*-5. A lowering of K value could also be observed if both the chromosomes carry different responder allele since this locus has been described to evolve very rapidly (Wu *et al.* 1988), where EY12961 could be a sensitive responder and EY05573 could be an insensitive responder.

Table 5.1. Crosses used to check the drive and results. a. *SD*-5/*cn bw* males X *cn bw* females. b. Reciprocal cross *SD*-5/*cn bw* females X *cn bw* males. k value for this cross was corrected for viability. A. Test cross and control cross. B. Reciprocal cross to check for viability. K value for these was not corrected for viability as results from viability cross are not significantly different from expected 0.5. All three K values are significantly different from expected 0.5 ($p < 0.5$). 0.87 is significantly different from 0.99 and 0.58 ($p < 0.0001$).

Genotype of parent	Type of cross	Phenotype of progeny		*k* value
		SD	Mutant	
SD-5/cn bw	a	1057	1	0.998
	b	389	251	
SD-5/EY05573	A	968	685	0.58
	B	117	128	
SD-5/EY21961	A	1267	182	0.87
	B	194	233	

5.2.2 Repeat number of the responder locus

To determine if the difference in K values from the two test crosses is due to difference in the responder locus, we perform slot blot analysis. A synthetic probe is used to hybridize the membrane. We tested the probe by performing Southern blot hybridization to reveal if the pattern observed is the one expected for the repeats previously described on the second chromosome (Wu *et al.* 1988). The banding pattern on the Southern blot shows a strong band of a long length (>8kb) in both *Dntf-2r* knockout (EY05573) and control (EY12961) chromosomes (Figure 6). Such a pattern is similar to those observed in previous studies (Wu *et al.* 1988) suggesting that the probe hybridizes to the Rsp locus in the lines tested. The band intensity of EY12961 is lower than that of EY05573 suggesting that it may have a lower number of repeats as confirmed below by slot blot. In addition, it seems that the responder locus on the *cn bw* has acquired an EcoRI restriction site. The smear shows that it has a high copy number of the repeats as confirmed by slot blot. EcoRI restriction site was not observed in previous

studies (Wu *et al.*1988). This may be because the *cn bw* line used in this study was ordered directly from Bloomington stock center and may not be the same one used in previous studies.

The same probe was used to hybridize the slot blot membrane to estimate Rsp repeat number. Appropriate dilutions of plasmid carrying the 353 bp probe (Puc57p1) are used to calculate copy number of the XbaI repeats. The results were surprising (Figure 5.6). The *Dntf-2r* knockout chromosome (EY05573) carries less than 335 XbaI repeats which correspond to a semi sensitive responder. EY12961 chromosome, on the other hand, carries less than 100 repeats however it gave k value corresponding to a sensitive responder (0.86). As there is no cut off between semi sensitive responder and insensitive responder we cannot designate the EY12961 responder as semi sensitive or insensitive. Generally less than 200 repeats is considered as semi insensitive or insensitive responders (Wu *et al.* 1988). Intuitively a lower number of repeats should give a lower k value which is not what we observe here. It has been previously observed (Wu *et al.* 1988) that when naturally occurring insensitive chromosomes are subjected to strong drivers like *SD*-5, some sensitivity is observed. Our study supports that observation and reinforces the fact that because *SD*-5 is a strong driver, a chromosome with lower number of repeats also shows a comparatively large drive. In addition, other loci we are not controlling for, that are unexpectedly different between strains might affect the drive.

However, low drive is observed in the case of EY05573 (*Dntf-2r* knockout) chromosome (k value = 0.58) while the Rsp locus is sensitive and should show a stronger drive. This suggests that *Dntf-2r* could play an enhancer role for *Sd* and knocking down *Dntf-2r* lowers the distortion ability of *SD*-5 chromosome which is otherwise a strong driver.

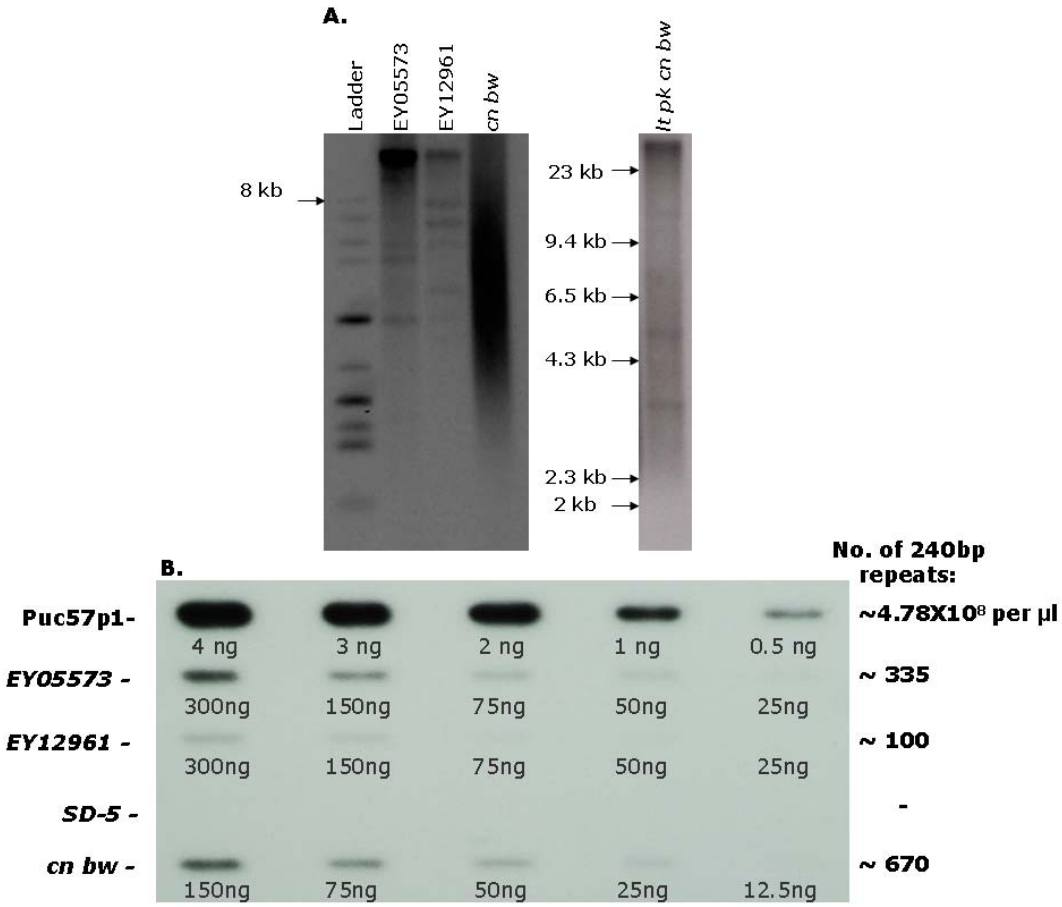


Figure 5.6. Responder locus analysis. A. Southern blot hybridization results. 6µg of DNA is digested using EcoRI restriction enzyme (3 µg for *lt pk cn bw*). DNA from EY05573, EY12961 and *lt pk cn bw* show similar banding pattern as observed in previous studies (Wu, Lyttle *et al.* 1988). *Cn bw* is digested by EcoRI. B. Slot blot analysis – Puc57p1 is the plasmid carrying probe sequence and is used as positive control to count the number of repeats. From band intensities it is observed that control EY12961 carries one-third of EY05573 repeats.

As described above, *Dntf-2r* has been suggested to have possibly played a role in meiotic drive as it is a member of nuclear transport machinery and also shows signs of positive selection that can be explained by the gene being involved in a genetic conflict (Presgraves 2007). Recently, in a study by McElroy *et al.* no distortion by *SD*-5 chromosome (i.e., k= 0.52) was observed in the same *Dntf-2r* knockdown line used in this work (McElroy *et al.* 2008). Authors argued that as no drive was observe by the knockout chromosome most likely it carries

an insensitive responder. No quantification of the responder locus was performed. Our data supports a semi sensitive responder and possibly some effect of loosing a *Dntf-2r*.

Recently, it was observed that *SD* spread in the African populations very recently i.e. within a few thousands years ago (Presgraves *et al.* 2009). Clearly, *Dntf-2r* arose via retroduplication much earlier than the *SD* system and it maybe initially maintained in the genome because of some functional advantage or some ancestral drive system (Betran *et al.*2003; Presgraves *et al.* 2009). When the *Sd* gene was formed, it quickly spread in the population dragging along the whole chromosome with it (Presgraves *et al.* 2009). *Dntf-2r* is in close proximity to the *Sd* (Figure 5.1) and might contain a different allele that contributes to the *SD* system. We sequenced the coding region and the regulatory region of *Dntf-2r* in *SD*-5 chromosome. From the coding region no signs of fast evolution were observed. The sequence is similar to one of the most common haplotypes of *Dntf-2r* (Figure 5.7) (Betran *et al.* 2003).

```
Dntf-2r_SD      GATGTTTACCTGTGACAATCGATTGTTTTCGGTGGGACCAAAACGTTCAAATCAAATGTA 60
Dntf-2r_WT      GATGTTTACCTGCGACAATCGATTGTTTTCGGTGGGACCAAAACGTTCAAATCAAATGTA 60
                ************ ********************************************

Dntf-2r_SD      TCAGCTTAGCGGTGACACAAACGAGAGGATATTTGAACAGTGCCT------GTCAGCTTA 114
Dntf-2r_WT      TCAGCTTAGCGGTGACACAAACGAGAGGATATTTGAACAGTGCCTAAGCCTGTCAGCTTA 120
                ******************************************      *********

Dntf-2r_SD      ACCGATTTAATTTGTTTTTTTTCGGATCGGAAACTCTAATCGTACCCCGTATATACTTGA 174
Dntf-2r_WT      ACCGATTTAATTTGCTTTTTTTCGGATCGGAAACTCTAATCGTACCCCGTATATACTTGA 180
                ************** ********************************************

Dntf-2r_SD      ACTAAAATGTCTCTGAATCTGCAGTACGAGGACATTGGCAAGGAATTTGTCCAGCAGTAC 234
Dntf-2r_WT      ACTAAAATGTCTCTGAATCTGCAGTACGAGGACATTGGCAAGGAATTTGTCCAGCAGTAC 240
                ************************************************************

Dntf-2r_SD      TACGCCATATTCGATGACCCGGCGAATCGGGAGAACGTGATTAATTTCTATAACGCTACC 294
Dntf-2r_WT      TACGCCATATTCGATGACCCGGCGAATCGGGAGAACGTGATTAATTTCTATAACGCTACC 300
                ************************************************************

Dntf-2r_SD      GATTCTTTCATGACCTTTGAAGGCAACCAAATACAGGGAGCACCCAAGATTCTGGAAAAA 354
Dntf-2r_WT      GACTCTTTCATGACCTTTGAAGGCAACCAAATACAGGGAGCACCCAAGATTCTGGAAAAA 360
                ** ********************************************************

Dntf-2r_SD      GTTCAGAGTCTGAGCTTTCAGAAGATTGCCAGAGTGATAACCACAGTGGATTCGCAGCCA 414
Dntf-2r_WT      GTTCAGAGTCTGAGCTTTCAGAAGATTGCCAGAGTGATAACCACAGTGGATTCGCAGCCA 420
                ************************************************************

Dntf-2r_SD      ACTTCCGATGGCGGAGTTCTGATCATCGTCCTTGGAAGACTAAAATGCGATGACGATCCC 474
Dntf-2r_WT      ACTTCCGATGGCGGAGTTCTGATCATCGTCCTTGGAAGACTAAAATGCGATGACGATCCC 480
                ************************************************************

Dntf-2r_SD      CCACATGCATTCTCGCAGATCTTTTTGCTGAAGCCCAACGGAGGATCCCTCTTCGTGGCT 534
Dntf-2r_WT      CCACATGCATTCTCGCAGATCTTTTTGCTGAAGCCCAACGGAGGATCCCTCTTCGTGGCT 540
                ************************************************************

Dntf-2r_SD      CACGACATCTTCCGTCTGAACATCCACAACTCTGCCTAGGAGCACTCCAGACCCATATGT 594
Dntf-2r_WT      CACGACATCTTCCGTCTGAACATCCACAACTCTGCCTAGGAGCACTCCAGACCCATATGT 600
                ************************************************************

Dntf-2r_SD      ACACCACACATAATCGACATCC 616
Dntf-2r_WT      ACACCACACATAATCGACATCC 616
                *********************
```

Figure 5.7. Alignment showing the sequence of *Dntf-2r* allele on the *SD* chromosome as compared to the wild type *D. melanogaster* sequence. Green highlight show the regulatory region and yellow highlight show *Dntf-2r* coding region.

It has been previously observed that the dose of certain proteins involved in nuclear transport could affect meiotic drive of *SD* locus (see above). Over expression of wild type RanGAP, Ran and RanGEF either enhance or compensate for the drive (Kusano *et al.* 2002). It is possible that although the coding region of *Dntf-2r* allele has not diverged, the regulation could be altered to affect the dose of the protein formed. As we have identified the regulatory region of *Dntf-2r* (see Chapter 2), we compare it to the one on the *SD* chromosome and observe no difference in the regulatory sequence (Figure 7). A deletion of 6bp is observed

which brings the regulatory region closer to the transcription start site. However, this deletion is observed in many strains of *D. melanogaster* and does not seem to affect the transcription and tissue specificity of *Dntf-2r* as the promoter region is not affected by its distance from the transcription start site (see Chapter 2).

Dntf-2r is a product of X to autosome retroduplication event. As discussed earlier, there is an abundance of such duplication events in Drosophila and mammalian genomes (Betran *et al.* 2002; Emerson *et al.* 2004; Bai *et al.* 2007). Just like other retroduplicates, male germline X inactivation, level of expression and sexual antagonism may be the selective forces responsible for the recurrent recruitment and fixation of *Dntf-2r* in three lineages (Tracy et. al - submitted). In all cases, the retrogene should have an important functional role as may be adapted for specific male germline functions. However, the fact that this retrogene does not have a clear effect in fertility (i.e., might be viewed as dispensable) does not support the above hypothesis. While X inactivation, level of expression and sexual antagonism could be responsible for its initial fixation, its current role and evolutionary pattern fit better its involvement in a conflict. Ran that interacts with *Dntf-2* has also given rise to several X to autosome retrogenes that evolve under positive selection, are lost or lose functions in some lineages. These two duplicates might both play a role in these genomic conflicts (segregation distortion or other related to viruses' and transposable elements). The effect of *Dntf-2r* if any on these conflicts could be related to its role in the RanGTP/Ran GDP gradient or related to its recently described direct role in transport. For long *Dntf-2* protein was known only as a transporter of RanGDP into the nucleus. A latest study shows that CapG binds directly to Ntf-2 and Ran-Ntf2 complex is responsible for its transport across the nuclear membrane (see chapter 1)(Katrien Van Impe *et al.* 2008). This finding opens new doors to the function of *Ntf-2* as well as *Dntf-2r*. They may be responsible for direct transport of some other proteins. The fast evolution of *Dntf-2r* under positive selection could suggest that the protein may be acquiring new transport functions. It has been suggested that genomic conflict may arise due to the nuclear transport of various transposable elements and virus genomes (Presgraves *et al.* 2007). Genes involved in nuclear transport may be evolving to

counteract such pathogenic attacks. However, a dispensable duplicate like *Dntf-2r* can be having either function. From the current study it seems that *Dntf-2r* might play a role in *SD* meiotic drive system but it might be an indirect effect unrelated to its current function. As no divergence was observed in the coding sequence as well as regulatory region of the allele associated with *SD*, it is difficult to postulate a direct role.

From this study, we observe that the chromosome containing the knockout of *Dntf-2r* lowers the drive caused by *SD* system. As shown above, our data supports a semi sensitive responder and possibly some effect of loosing a *Dntf-2r*. We can however not rule out that the effect we see of lower drive than expected by the type of responder that the chromosome is carrying might be due to other loci in that chromosome. Recombination of *Dntf-2r* knockdown allele into a known super sensitive responder chromosome would allow us to separate the effects. However this could be an indirect effect as discussed above. Therefore, further characterization of the interactors of *Dntf-2r* and of this and other knockout/knockdown lines (i.e., RNAi of *Dntf-2r* would decrease further the level of the gene) are needed. Functional domains and identification of other proteins that could potentially bind *Dntf-2r* would increase our understanding of the gene function.

REFERENCES

Adachi, N. and M. R. Lieber (2002). Bidirectional Gene Organization: A Common Architectural Feature of the Human Genome, Cell 109(7): 807-809.

Arguello, J. R., Y. Chen. *et al.* (2006). Origination of an X-Linked Testes Chimeric Gene by Illegitimate Recombination in Drosophila. PLoS Genet 2(5): e77.

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

Bai, Y., C. Casola, *et al.* (2008). Evolutionary origin of regulatory regions of retrogenes in Drosophila, BMC Genomics 9(1): 241.

Begun, D. J. (1997). Origin and Evolution of a New Gene Descended From alcohol dehydrogenase in Drosophila, Genetics 145(2): 375-382.

Bellen, H. J., R. W. Levis, *et al.* (2004). The BDGP Gene Disruption Project: Single Transposon Insertions Associated With 40% of Drosophila Genes, Genetics 167(2): 761-781.

Betran, E. and M. Ashburner. (2000). Duplication, dicistronic transcription, and subsequent evolution of the Alcohol dehydrogenase and Alcohol dehydrogenase-related genes in Drosophila, Mol Biol Evol 17(9): 1344-52.

Betran, E. and M. Long (2003) *Dntf-2r*, a young Drosophila retroposed gene with specific male expression under positive Darwinian selection, Genetics 164(3): 977-88.

Betran, E., K. Thornton, *et al.* (2002). Retroposed new genes out of the X in Drosophila, Genome Res 12(12): 1854-9.

Bhattacharya, A. and R. Steward (2002) The Drosophila homolog of NTF-2, the nuclear transport factor-2, is essential for immune response, EMBO reports 3(4): 378-83.

Bielinska, B., S. M. Blaydes, *et al.* (2000). De novo deletions of SNRPN exon 1 in early human and mouse embryos result in a paternal to maternal imprint switch, 25(1): 74-78.

Bracken, A. P., M. Ciro, *et al.* (2004). E2F target genes: unraveling the biology. Trends in Biochemical Sciences, 29(8): 409-417.

Bradley, J., A. Baltus, H. Skaletsky, M. Royce-Tolland, K. Dewar, & D. C. Page. (2004). An X-to-autosome retrogene is required for spermatogenesis in mice, Nat Genet 36:872-876.

Bullock, T. L., D. W. Clarkson, *et al.* (1996). The 1.6 Å Resolution Crystal Structure of Nuclear Transport Factor 2 (NTF2), Journal of Molecular Biology 260(3): 422-431.

Burke, T. W. a. J. T. K. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila, Genes Devel. 11: 3020-3031.

Burt, A., and R. Trivers. (2006). Genes in conflict. The biology of selfish genetic elements, Harvard University Press, Cambridge, Massachusetts.

Buzdin, A., E. Gogvadze, *et al.* (2003). The human genome contains many types of chimeric retrogenes generated through in vivo RNA recombination. Nucl. Acids Res, 31(15): 4385-4390.

Chai, J.-H., D. P. Locke, *et al.* (2001). Retrotransposed genes such as *Frat3* in the mouse Chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site, Mammalian Genome V12(11): 813-821.

Charlesworth, B. and D. L. Hartl (1978). Population Dynamics of the Segregation Distorter Polymorphism of Drosophila Melanogaster, Genetics 89(1): 171-192.

Chintapalli, V. R., J. Wang, *et al.* (2007). Using FlyAtlas to identify better Drosophila melanogaster models of human disease, 39(6): 715-720.

Choi, J. D., L. A. Underkoffler, *et al.* (2005). A Novel Variant of Inpp5f Is Imprinted in Brain, and Its Expression Is Correlated with Differential Methylation of an Internal CpG Island, Mol. Cell. Biol. 25(13): 5514-5522.

Ciciarello, M., R. Mangiacasale, *et al.* (2007). Spatial control of mitosis by the GTPase Ran, Cellular and Molecular Life Sciences (CMLS) 64(15): 1891-1914.

Cusack, B. P. and K. H. Wolfe (2007). Not Born Equal: Increased Rate Asymmetry in Relocated and Retrotransposed Rodent Gene Duplicates, Mol Biol Evol 24(3): 679-686.

Dahl, H.-H. M., R. M. Brown, *et al.* (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(2): 225-232.

Daven C. Presgraves. (2007). Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila, BioEssays 29(4): 386-391.

Devor, E. (2001). Molecular archeology of an *SP100* splice variant revisited: dating the retrotranscription and Alu insertion events, Genome Biology 2(9): research0040.1 - research0040.6.

Donaldson, T. D., M. A. Noureddine, P. J. Reynolds, W. Bradford, and R. J. Duronio. (2004). Targeted disruption of Drosophila Roc1b reveals functional differences in the Roc subunit of Cullin-dependent E3 ubiquitin ligases, Mol Biol Cell 15:4892-4903.

Doolittle, W. F., and C. Sapienza. (1980). Selfish genes, the phenotype paradigm and genome evolution. 284(5757): 601-603.

Dorus, S., Z. N. Freeman, E. R. Parker, B. D. Heath, and T. L. Karr. (2008). Recent origins of sperm genes in Drosophila, Mol Biol Evol 25:2157-2166.

Dorus, S., Z. N. Freeman, *et al.* (2008). Recent origins of sperm genes in Drosophila, Mol Biol Evol 25(10): 2157-66.

Dragon, F., J. E. G. Gallagher, *et al.* (2002). A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis, 417(6892): 967-970.

Dunham, I., N. Shimizu, *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(6761): 489-95

Emerson, J. J., H. Kaessmann, E. Betran, and M. Long. (2004). Extensive Gene Traffic on the Mammalian X Chromosome, Science 303:537-540.

Esnault, C. c., J. l. Maestre, *et al.* (2000). Human LINE retrotransposons generate processed pseudogenes, 24(4): 363-367.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks, 9(5): 397-405.

Fitzgerald, J., W. M. Hutchison, *et al.* (1992). Isolation and characterisation of the mouse pyruvate dehydrogenase E1[alpha] genes, Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1131(1): 83-90.

Fontanillas, P., D. L. Hartl, and M. Reuter. (2007). Genome organization and gene expression shape the transposable element distribution in the Drosophila melanogaster euchromatin, PLOS Genet 3:2256-2267.

Force, A., M. Lynch, *et al.* (1999). Preservation of Duplicate Genes by Complementary, Degenerative Mutations, Genetics 151(4): 1531-1545.

Fourel, G., C. Transy, *et al.* (1992). Expression of the woodchuck *N-myc2* retroposon in brain and in liver tumors is driven by a cryptic N-myc promoter, Mol. Cell. Biol. 12(12): 5336-5344.

Ganetzky, B. (1977). On the components of segregation distortion in drosophila melanogaster, Genetics 86(2): 321-355.

Gebara, M. M. and J. R. McCarrey (1992). Protein-DNA interactions associated with the onset of testis-specific expression of the mammalian *Pgk-2* gene, Mol. Cell. Biol. 12(4): 1422-1431.

Genome and RNA: Expression and Functions - Papers presented at a S*ymp*osium in Puntarenas, Costa Rica, 26 February-2 March 2005 365: 104-110

Greally, J. M., T. A. Gray, *et al.* (1999). Conserved characteristics of heterochromatin-forming DNA at the 15q11-q13 imprinting center, PNAS 96(25): 14430-14435.

Gu, Z., A. Cavalcanti, *et al.* (2002). Extent of Gene Duplication in the Genomes of Drosophila, Nematode, and Yeast, Mol Biol Evol 19(3): 256-262.

Guldner, H., C. Szostecki, *et al.* (1999). Splice variants of the nuclear dot-associated *Sp100* protein contain homologies to HMG-1 and a human nuclear phosphoprotein-box motif, J Cell Sci 112(5): 733-747.

Hahn, M. W., M. V. Han, and S.G. Han. (2007). Gene Family Evolution across 12 Drosophila Genomes PloS, Genetics 3:e197.

Han, M. V., J. P. Demuth, *et al.* (2009). Adaptive evolution of young gene duplicates in mammals, Genome Res 19: 859-867.

Hartl, D. L. (1974). Genetic Dissection of Segregation Distortion. I. Suicide Combinations of Sd Genes, Genetics 76(3): 477-486.

Hauschteck-Jungen, E. and D. L. Hartl. (1982). Defective histone transition during spermiogenesis in heterozygous segregation distorter males of drosophila melanogaster, Genetics 101(1): 57-69.

Higgins, J. M. G. (2001). The Haspin gene: location in an intron of the Integrin [alpha]E gene, associated transcription of an Integrin [alpha]E-derived RNA and expression in diploid as well as haploid cells, Gene 267(1): 55-69.

Hofmann, A., M. D. Garfinkel, and E. M. Meyerowitz. (1991). cis-acting sequences required for expression of the divergently transcribed Drosophila melanogaster Sgs-7 and Sgs-8 glue protein genes, Mol Cell Biol 11:2971-2979.

Hore, T. A., R. W. Rapkins, *et al.* (2007). Construction and evolution of imprinted loci in mammals, Trends in Genetics 23(9): 440-448.

Horsthemke, B. and K. Buiting (2006). Imprinting defects on human chromosome 15, Cytogenet Genome Res 113: 292-299.

Houtchens, K. and T. W. Lyttle (2003). Responder (Rsp) Alleles in the Segregation Distorter (*SD*) System of Meiotic Drive in Drosophila may Represent a Complex Family of Satellite Repeat Sequences, Genetica 117(2): 291-302.

Iannello, R. C., J. A. Gould, *et al.* (2000). Methylation-dependent Silencing of the Testis-specific Pdha-2 Basal Promoter Occurs through Selective Targeting of an Activating Transcription Factor/cAMP-responsive Element-binding Site, J. Biol. Chem. 275(26): 19603-19608.

Iannello, R., J. Young, *et al.* (1997). Regulation of Pdha-2 expression is mediated by proximal promoter sequences and CpG methylation, Mol. Cell. Biol. 17(2): 612-619.

Isgro, T. A. and K. Schulten (2007). Association of Nuclear Pore FG-repeat Domains to NTF2 Import and Export Complexes, Journal of Molecular Biology 366(1): 330-345.

Jacobs, S., A. SchÃ¼rmann, *et al.* (1998). The mouse ADP-ribosylation factor-like 4 gene: two separate promoters direct specific transcription in tissues and testicular germ cell, Biochem. J. 335(2): 259-265.

Jones, C. D. and D. J. Begun (2005). Parallel evolution of chimeric fusion genes, PNAS 102(32): 11373-11378.

Jones, C. D., A. W. Custer, *et al.* (2005). Origin and Evolution of a Chimeric Fusion Gene in Drosophila subobscura, D. madeirensis and D. guanche, Genetics 170(1): 207-219.

Jong, M., T. Gray, *et al.* (1999). A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the Prader-Willi syndrome critical region, Hum. Mol. Genet. 8(5): 783-793.

Jordan, I. K., I. B. Rogozin, *et al.* (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements, Trends in Genetics 19(2): 68-72.

Julia C. Young, J. A. G., Ismail Kola, Rocco C. Iannello (1998). Review: Pdha-2, past and present, The Journal of Experimental Zoology 282(1-2): 231-238.

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

Kalmykova, A. I., D. I. Nurminsky, *et al.* (2005). Regulated chromatin domain comprising cluster of co-expressed genes in Drosophila melanogaster, Nucl. Acids Res. 33(5): 1435-1444.

Katrien Van Impe, Thomas Hubert, *et al.* (2008). A New Role for Nuclear Transport Factor 2 and Ran: Nuclear Import of CapG, Traffic 9(5): 695-707.

Kleene, K. C. and M.-A. Mastrangelo (1999). The Promoter of the Poly(A) Binding Protein 2 (*Pabp2*) Retroposon Is Derived from the 5'-Untranslated Region of the *Pabp1* Progenitor Gene, Genomics 61(2): 194-200.

Kleene, K. C., E. Mulligan, *et al.* (1998). The Mouse Gene Encoding the Testis-Specific Isoform of Poly(A) Binding Protein (*Pabp2*) Is an Expressed Retroposon: Intimations That Gene Expression in Spermatogenic Cells Facilitates the Creation of New Genes, Journal of Molecular Evolution 47(3): 275-281.

Kohda, K., T. Ishibashi, *et al.* (1998). Characterization of the Mouse PA28 Activator Complex Gene Family: Complete Organizations of the Three Member Genes and a Physical Map of the ~150-kb Region Containing the {alpha}- and {beta}-Subunit Genes, J Immunol 160(10): 4923-4935.

Krasnov, A. N., M. M. Kurshakova, *et al.* (2005). A retrocopy of a gene can functionally displace the source gene in evolution, Nucl. Acids Res. 33(20): 6654-6661.

Kusano, A., C. Staber, *et al.* (2002). Segregation distortion induced by wild-type RanGAP in Drosophila, PNAS 99(10): 6866-6870.

Kusano, A., S. Cynthia, *et al.* (2001). Nuclear Mislocalization of Enzymatically Active RanGAP Causes Segregation Distortion in Drosophila, Developmental Cell 1(3): 351-361.

Lee, T.-H., S.-L. Yu, *et al.* (1999). Characterization of the murine gene encoding 1-Cys peroxiredoxin and identification of highly homologous genes, Gene 234(2): 337-344.

Li, B., H.-Q. Zhang, *et al.* (2009). Overexpression of nuclear transport factor 2 may protect against diabetic retinopathy, Molecular Vision 15: 861-869.

Lim, C. Y., B. Santoso, *et al.* (2004). The MTE, a new core promoter element for transcription by RNA polymerase II, Genes Dev. 18(13): 1606-1617.

Long, M. and C. Langley (1993). Natural selection and the origin of *Jingwei*, a chimeric processed functional gene in Drosophila, Science 260: 91 - 95.

Long, M., M. Deutsch, *et al.* (2003). Origin of new genes: evidence from experimental and computational analyses, Genetica 118(2-3): 171-82.

Long, M., W. Wang, *et al.* (1999). Origin of new genes and source for N-terminal domain of the chimerical gene, *Jingwei*, in Drosophila, Gene 238(1): 135-141.

Luo, C., X. Lu, *et al.* (2006). Rapid evolution of a recently retroposed transcription factor YY2 in mammalian genomes, Genomics 87(3): 348-355.

Lynch, M. and J. S. Conery (2000). The Evolutionary Fate and Consequences of Duplicate Genes, Science 290(5494): 1151-1155.

Lyttle, T. W. (1991). Segregation Distorters, Annual Review of Genetics 25(1): 511-581.

Mahajan, R., L. Gerace, *et al.* (1998). Molecular Characterization of the SUMO-1 Modification of RanGAP1 and Its Role in Nuclear Envelope Association, J. Cell Biol. 140(2): 259-270.

Makalowski, W. (2003). GENOMICS: Not Junk After All, Science 300(5623): 1246-1247.

Mancebo, R., X. Zhou, *et al.* (2001). BSF Binds Specifically to the bicoid mRNA 3' Untranslated Region and Contributes to Stabilization of bicoid mRNA, Mol. Cell. Biol. 21(10): 3462-3471.

Marchler-Bauer, A., J. B. Anderson, *et al.* (2009). CDD: specific functional annotation with the Conserved Domain Database, Nucl. Acids Res. 37(suppl_1): D205-210.

Mariño-Ramírez, L., K. C. Lewis, *et al.* (2005). Transposable elements donate lineage-specific regulatory sequences to host genomes, Cytogenet Genome Res 110: 333-341.

Marques, A. C., I. Dupanloup, *et al.* (2005). Emergence of Young Human Genes after a Burst of Retroposition in Primates, PLoS Biology 3(11): e357.

McCarrey, J. R. (1990). Molecular evolution of the human *Pgk-2* retroposon, Nucl. Acids Res. 18(4): 949-955.

McElroy, J., R. McLean, *et al.* (2008). Defects in nuclear transport enhance segregation distortion, Fly 2(6): 280-290.

Merrill, C., L. Bayraktaroglu, *et al.* (1999). Truncated RanGAP Encoded by the Segregation Distorter Locus of Drosophila, Science 283(5408): 1742-1745.

Michael Ashburner (1998). Speculations on the subject of alcohol dehydrogenase and its properties in Drosophila and other flies, BioEssays 20(11): 949-954.

Michiels, F., A. Gasch, *et al.* (1989). A 14 bp promoter element directs the testis specificity of the Drosophila beta 2 tubulin gene, EMBO Journal 8(5): 1559-65.

Mikhaylova, L. M., A. M. Boutanaev, *et al.* (2006). Transcriptional regulation by Modulo integrates meiosis and spermatid differentiation in male germ line, PNAS 103(32): 11975-11980

Minakhina, S., R. Myers, *et al.* (2005). Crosstalk between the actin cytoskeleton and Ran-mediated nuclear transport, BMC Cell Biology 6(1): 32.

Mizue Hisano, S. Y., Hiromitsu Tanaka, Yoshitake Nishimune, Masami Nozaki, (2003). Genomic structure and promoter activity of the testis haploid germ cell-specific intronless genes, *Tact1*and *Tact2*, Molecular Reproduction and Development 65(2): 148-156.

Moran, J. (1999). Human L1 retrotransposition: insights and peculiarities learned from a cultured cell retrotransposition assay, Genetica 107(1): 39-51.

Moss, J. and M. Vaughan (1998). Molecules in the ARF Orbit, J. Biol. Chem. 273(34): 21431-21434.

Nabetani, A., I. Hatada, *et al.* (1997). Mouse *U2af1-rs1* is a neomorphic imprinted gene, Mol. Cell. Biol. 17(2): 789-798.

Nguyen, N., X. Zhang, *et al.* (2004). Molecular Cloning and Functional Characterization of the Transcription Factor YY2, J. Biol. Chem. 279(24): 25927-25934.

Nozawa, M., T. Aotsuka, *et al.* (2005). A novel chimeric gene, *siren*, with retroposed promoter sequence in the Drosophila bipectinata complex, Genetics: genetics.105.041699.

Nurminsky, D. I., M. V. Nurminskaya, *et al.* (1998). Selective sweep of a newly evolved sperm-specific gene in Drosophila, 396(6711): 572-575.

Parisi, M., R. Nuttall, *et al.* (2003). Paucity of Genes on the Drosophila X Chromosome Showing Male-Biased Expression, Science 299(5607): 697-700.

Parisi, M., R. Nuttall, *et al.* (2004). A survey of ovary-, testis-, and soma-biased gene expression in Drosophila melanogaster adults, Genome Biol 5(6): R40.

Paulsen, M., K. Davies, *et al.* (1998). Syntenic organization of the mouse distal chromosome 7 imprinting cluster and the Beckwith-Wiedemann syndrome region in chromosome 11p15.5, Hum. Mol. Genet. 7(7): 1149-1159.

Peters, J., S. F. Wroe, *et al.* (1999). A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2, PNAS 96(7): 3830-3835.

Pimpinelli, S. and P. Dimitri (1989). Cytogenetic Analysis of Segregation Distortion in Drosophila melanogaster: The Cytological Organization of the Responder (Rsp) Locus, Genetics 121(4): 765-772.

Ponce, R. and D. L. Hartl (2006). The evolution of the novel *Sdic* gene cluster in Drosophila melanogaster, Gene 376(2): 174-183.

Potrzebowski, L., N. Vinckenbosch, *et al.* (2008). Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes, PLoS Biol. 6(4): e80.

Presgraves, D. C. and W. Stephan (2007). Pervasive Adaptive Evolution among Interactors of the Drosophila Hybrid Inviability Gene, Nup96, Mol Biol Evol 24(1): 306-314.

Presgraves, D. C., P. R. GÃ©rard, *et al.* (2009). Large-Scale Selective Sweep among Segregation Distorter Chromosomes in African Populations of Drosophila melanogaster, PLoS Genet 5(5): e1000463.

Prince, V. E. and F. B. Pickett (2002). Splitting pairs: the diverging fates of duplicated genes, 3(11): 827-837.

Ranz, J. M., C. I. Castillo-Davis, C. D. Meiklejohn, and D. L. Hartl. 2003. Sex-dependent gene expression and evolution of the Drosophila transcriptome, Science 300:1742-1745.

Rice, W. (1992). Sexually antagonistic genes: experimental evidence, Science 256(5062): 1436-1439.

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

Robinson, M. O., J. R. McCarrey, *et al.* (1989). Transcriptional Regulatory Regions of Testis-Specific PGK2 Defined in Transgenic Mice, PNAS 86(21): 8437-8441.

Rod J. Scott, M. S. (2006). Deeper into the maize: new insights into genomic imprinting in plants, BioEssays 28(12): 1167-1171.

Rogalla, P., B. Kazmierczak, *et al.* (2000). Back to the Roots of a New Exon--The Molecular Archaeology of a *SP100* Splice Variant, Genomics 63(1): 117-122.

Rohozinski, J. and C. E. Bishop (2004). The mouse juvenile spermatogonial depletion (jsd) phenotype is due to a mutation in the X-derived retrogene, m*Utp14b*, PNAS 101(32): 11695-11700.

Rohozinski, J., D. J. Lamb, *et al.* (2006). *Utp14c* Is a Recently Acquired Retrogene Associated with Spermatogenesis and Fertility in Man, Biol Reprod 74(4): 644-651.

Sandler, L., Y. Hiraizumi, *et al.* (1959). Meiotic drive in natural populations of drosophila melanogaster. I. The cytogenetic basis of segregation-distortion, Genetics 44(2): 233-250.

Sassone-Corsi, P. (2002). Unique Chromatin Remodeling and Transcriptional Regulation in Spermatogenesis, Science 296(5576): 2176-2178.

Sayah, D. M., E. Sokolskaja, *et al.* (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1, 430(6999): 569-573.

Schon, M. P., A. Arya, *et al.* (1999). Mucosal T L*ymp*hocyte Numbers Are Selectively Reduced in Integrin {alpha}E (CD103)-Deficient Mice, J Immunol 162(11): 6641-6649.

Silacci, P., L. Mazzolai, *et al.* (2004). Gelsolin superfamily proteins: key regulators of cellular functions, Cellular and Molecular Life Sciences (CMLS) 61(19): 2614-2623.

Smith, R. J., W. Dean, *et al.* (2003). Identification of Novel Imprinted Genes in a Genome-Wide Screen for Maternal Methylation, Genome Res. 13(4): 558-569.

Stewart, M., H. M. Kent, *et al.* (1998). Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the ras-family GTPase ran, Journal of Molecular Biology 277(3): 635-646.

Strichman-Almashanu, L. Z., M. Bustin, *et al.* (2003). Retroposed Copies of the HMG Genes: A Window to Genome Dynamics, Genome Res. 13(5): 800-812.

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

Sun, S., C.T. Ting, *et al.* (2004). The Normal Function of a Speciation Gene, Odysseus, and Its Hybrid Sterility Effect, Science 305(5680): 81-83.

Susumu Ohno (1970). Evolution by gene duplication

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner, and C. F. Aquadro. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila, Proc Natl Acad Sci U S A 98:7375-7379.

Tanaka, H., N. Iguchi, *et al.* (2001). Cloning and characterization of human haspin gene encoding haploid germ cell-specific nuclear protein kinase, Mol. Hum. Reprod. 7(3): 211-218.

Tanaka, H., Y. Yoshimura, *et al.* (1999). Identification and Characterization of a Haploid Germ Cell-specific Nuclear Protein Kinase (Haspin) in Spermatid Nuclei and Its Effects on Somatic Cells, J. Biol. Chem. 274(24): 17049-17057.

Temin, R. G. (1991). The Independent Distorting Ability of the Enhancer of Segregation Distortion, *E(SD)*, in Drosophila melanogaster. Genetics 128(2): 339-356.

Thomas, K., J. Wu, *et al.* (2007). SP1 transcription factors in male germ cell development and differentiation, Molecular and Cellular Endocrinology 270(1-2): 1-7.

Thornburg, B. G., V. Gotea, *et al.* (2006). Transposable elements as a significant source of transcription regulating signals, Gene

Thummel, C. S. and V. Pirrotta (1991). Technical Notes: New pCaSpeR P-element vectors, 2.

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

Tokuhiro, K., Y. Miyagawa, *et al.* (2007). The 193-Base Pair *Gsg2* (Haspin) Promoter Region Regulates Germ Cell-Specific Expression Bidirectionally and Synchronously, Biol Reprod 76(3): 407-414.

Tokuyasu, K., W. Peackock, *et al.* (1977). Dynamics of spermiogenesis in Drosophila melanogaster VII. Effects of Segregation Distorter (SD) chromosomes, J Ultrastruct Res 58: 96-101.

Turner, J. M. A. (2007). Meiotic sex chromosome inactivation, Development 134(10): 1823-1831.

U. Kevin Ewulonu, L. S., Lee M. Silver, John C. Schimenti, (1996). Promoter mapping of the mouse Tcp-10bt gene in transgenic mice identifies essential male germ cell regulatory sequences, Molecular Reproduction and Development 43(3): 290-297.

Vanin, E. F. (1985). Processed Pseudogenes: Characteristics and Evolution, Annual Review of Genetics 19(1): 253-272.

Vicoso, B., and B. Charlesworth. (2009). The Deficit of Male-Biased Genes on the *D. melanogaster* X Chromosome Is Expression-Dependent: A Consequence of Dosage Compensation, J Mol Evol.

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

Wang, W., H. Zheng, *et al.* (2006). High Rate of Chimeric Gene Origination by Retroposition in Plant Genomes, Plant Cell 18(8): 1791-1802.

Wang, W., J. Zhang, *et al.* (2000). The Origin of the *Jingwei* Gene and the Complex Modular Structure of Its Parental Gene, *Yellow emperor*, in Drosophila melanogaster, Mol Biol Evol 17(9): 1294-1301.

Wang, Y., K. Joh, *et al.* (2004). The Mouse *Murr1* Gene Is Imprinted in the Adult Brain, Presumably Due to Transcriptional Interference by the Antisense-Oriented *U2af1-rs1* Gene, Mol. Cell. Biol. 24(1): 270-279.

Wood, A. J., R. G. Roberts, *et al.* (2007). A Screen for Retrotransposed Imprinted Genes Reveals an Association between X Chromosome Homology and Maternal Germ-Line Methylation, PLoS Genetics 3(2): e20.

Wu, C.-I., T. W. Lyttle, *et al.* (1988). Association between a satellite DNA sequence and the responder of segregation distorter in *D. melanogaster*, Cell 54(2): 179-189.

Yi, S., and B. Charlesworth. (2000). A selective sweep associated with a recent gene transposition in Drosophila Miranda, Genetics 156:1753-1763.

Yiu, G., M. Murray, *et al.* (1997). Deoxyribonucleic acid-protein interactions associated with transcriptional initiation of the mouse testis-specific cytochrome c gene, Biol Reprod 56(6): 1439-1449.

Yuan, X., M. Miller, *et al.* (1996). Duplicated Proteasome Subunit Genes in Drosophila melanogaster Encoding Testes-Specific Isoforms, Genetics 144(1): 147-157.

Zaiss, D. M. W. and P.-M. Kloetzel (1999). A second gene encoding the mouse proteasome activator PA28[beta] subunit is part of a LINE1 element and is driven by a LINE1 promoter, Journal of Molecular Biology 287(5): 829-835.

Zhang, J., A. M. Dean, *et al.* (2004). Evolving protein functional diversity in new genes of Drosophila, PNAS 101(46): 16246-16250.

Zhang, L. P., J. C. Stroud, *et al.* (1998). A Gene-Specific Promoter in Transgenic Mice Directs Testis-Specific Demethylation Prior to Transcriptional Activation In Vivo, Biol Reprod 59(2): 284-292.

Zhang, L. P., J. Stroud, *et al.* (1999). Multiple Elements Influence Transcriptional Regulation from the Human Testis-Specific PGK2 Promoter in Transgenic Mice, Biol Reprod 60(6): 1329-1337.

Zhang, Z., K. Joh, *et al.* (2006). Comparative analyses of genomic imprinting and CpG island-methylation in mouse *Murr1* and human *MURR1* loci revealed a putative imprinting control region in mice, Gene 366(1): 77-86.

Zhao, M., J. Rohozinski, *et al.* (2007). *Utp14b*: A unique retrogene within a gene that has acquired multiple promoters and a specific function in spermatogenesis, Developmental Biology 304(2): 848-859.

Zheng, D., Z. Zhang, *et al.* (2005). Integrated pseudogene annotation for human chromosome 22: evidence for transcription, J Mol Biol 349(1): 27-45.

Zhong, L., and J. M. Belote. 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

Mansi Motiwale Kunte was born in Indore, India in 1977. She received her B.S. in microbiology from PMB Gujrati College in Indore in 1999 and M.S. in biochemistry from School of Biochemistry, DAVV, Indore in 2001. She received her Ph.D. from University of Texas – Arlington. Her research has focused on evolution of new gene functions, specifically via retrogenes. In her research she has gained proficiency in various genetic tools like molecular genetics, comparative genomics, cell biology and biochemistry.

In future, she wants to change direction and learn about human genetics. She would do a post doc in clinical genetics to apply her current knowledge to various human genetic diseases.