Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
Ibrahim Avsar Ilik, B.Sc.
Born in Kayseri, Turkey

Date of the oral examination:

A biochemical study of roX2 interacting proteins

Referees:

Prof. Dr. Matthias W. Hentze

Prof. Dr. Ed Hurt

Dünya bir düştür. Evet, dünya..Ah! Evet, dünya bir masaldır.

Uzun İhsan Efendi

Hiç bilenlerle bilmeyenler bir olur mu?

Zümer/9

Table of Contents

Summary

Dosage compensation is a biological phenomenon where a sexually dimorphic organism balances the inequality in gene expression that results from unequal distribution of sex chromosomes. Different organisms have invented different ways to carry out dosage compensation. For instance, in mammals females transcriptionally down-regulate one of the two X-chromosomes they possess to match the male gene expression. Flies, on the other hand, hyper-transcribe the single male X-chromosome to reach the transcriptional output generated by two X-chromosomes in females. A ribonucleoprotein complex, called the Male Specific Lethal (MSL) complex is essential for fly dosage compensation. It is composed of five proteins and two non-coding RNAs called the roX RNAs. The complex contains at least two enzymes: MOF, an acetyltransferase that specifically acetylates Histone 4 Lysine 16; and MLE, a DNA/RNA helicase. Both enzymatic activities are indispensible for dosage compensation.

Evidence coming from genetic studies have shown that male flies lacking both roX RNAs die due to a failure in dosage compensation, although these ncRNAs are redundant in function and only one of the two is enough to rescue male lethality. MLE was shown to be required for the incorporation of these RNAs into the MSL complex although it is not clear if this is the only function of MLE in dosage compensation

During the first part of my PhD I carried out a tandem affinity purification to reveal proteins that interact with MLE in an effort to understand its role in dosage compensation. In accord with previous observations, I have not been able to detect any protein that stably interacts with MLE under various purification conditions. This work thus supports the view that MLE is a lone RNA-helicase and is recruited to the X chromosome by its interaction with the roX RNAs.

In the second part of my thesis I describe the biochemical purification and analysis of roX2 interacting proteins *in vitro*. With this approach I have identified MLE and two novel proteins; CG5787 and CG3613 that interact with roX2 RNA specifically. Interestingly, CG5787 and CG3613 were found to co-localize with each other and

with MLE on chromatin. CG3613 was also shown to interact with roX2 RNA *in vivo* validating the initial *in vitro* approach. By using ChIP analysis I was able to detect CG3613 on high-affinity sites on the X-chromosome, which might indicate that it is recruited to the X-chromosome via roX RNA similar to MLE. CG3613 was further characterized and found to be a phosphoprotein *in vivo*. CG3613 also co-localizes with RNA polymerase II and is recruited to heat-shock loci after a brief heat-shock, indicating a strong relationship with transcription. Knocking-down CG3613 in flies was lethal, but in S2 cells reduction of its levels lead to the stabilization of intron-containing pre-mRNA suggesting a role in pre-mRNA processing.

This work was the first attempt to biochemically define roX2 interacting proteins in flies and shows that the MSL proteins are not the only proteins that roX RNAs interact with. In fact, this study suggests that roX RNAs' role may be to act as a platform that brings together various proteins in addition to the MSL complex to hyper-transcribe the male X chromosome.

Zusammenfassung

Die Dosis-Kompensation ist ein biologisches Phänomen, in welchem ein geschlechtlich dimorpher Organismus das Ungleichgewicht in der Genexpression, das aus der ungleichmäßigen Verteilung der Sex-Chromosomen resultiert, ausgleicht. Verschiedene Organismen haben verschiedene Wege entwickelt, wie die Dosis-Kompensation erreicht wird. Bei Säugetieren z.B. unterdrückt das Weibchen transkriptional eines der beiden X-Chromosomen, die sie besitzt, um mit der männlichen Genexpression gleichzuziehen. Auf der anderen Seite benutzt die Fruchtfliege Drosophila Melanogaster die Hypertranskription des einzigen männlichen X-Chromosoms, um die transkriptionelle Leistung , die durch die zwei X-Chromosome bei Weibchen generiert wird, zu erreichen. Ein Ribonucleoprotein-Komplex, den man MSL (Male Specific Lethal) Komplex nennt, ist in der Dosis-Kompensation von Fruchtfliegen dabei unverzichtbar. Er besteht aus fünf Poteinen und zwei nicht kodierenden RNA's, den sogenannten roX RNA's. Dieser Komplex beinhaltet mindestens zwei Enzyme: MOF, eine Azetyltransferase, die spezifisch das Histone 4 an der Position Lysine 16 azetyliert, und MLE, eine DNA/RNA -Helicase. Beide enzymatische Aktivitäten sind essentiell für die Dosis-Kompensation.

Genetische Studien belegen, dass männliche Fruchtfliegen, denen beide roX RNA's fehlen, an einem Fehler in der Dosis-Kompensation sterben, obwohl diese ncRNA's in der Funktion überflüssig sind und nur eine der Beiden ausreicht, um die männliche Lethalität zu retten. Es wurde gezeigt, dass MLE für die Inkorporation dieser RNA's in den MSL Komplex notwendig ist, obwohl nicht klar ist, ob dies die einzige Funktion von MLE in der Dosis-Kompensation ist.

Um die Rolle von MLE in der Dosis-Kompensation zu verstehen, führte ich während des ersten Teils meiner Doktorarbeit eine Tandem-Affinitäts-Reinigung durch, um jene Proteine erkennbar zu machen, die mit MLE zusammenwirken.
In Übereinstimmung mit früheren Beobachtungen konnte ich kein neues Protein entdecken, das stabil mit MLE unter verschiedenen Aufreinigungsbedingungen zusammenwirkt. Diese Arbeit untermauert deshalb die Ansicht, dass MLE eine

einzelne RNA-Helicase ist und durch das Zusammenwirken mit den roX RNA's zu den X-Chromosomen rekrutiert wird.

Im zweiten Teil meiner Thesis beschreibe ich die biochemische Aufreinigung und Analyse von mit roX2 zusammenwirkenden Proteinen *in vitro*. Mit diesem Ansatz habe ich MLE und zwei neuartige Proteine identifiziert; CG5787 und CG3613, die mit roX2 -RNA spezifisch zusammenwirken. Interessanterweise hat sich gezeigt, dass CG5787 und CG3613 miteinander und mit MLE auf Chromatin kolokalisieren. CG3613 bindet roX2 RNA auch *in vivo*, was den vorherigen *in vitro* Ansatz bestätigte. Durch eine ChIP Analyse war es mir möglich, CG3613 auf hochaffinen Stellen auf dem X-Chromosom zu detektieren; dies könnte ein Hinweis dafür sein, dass die Rekrutierung zum X-Chromosom, ähnlich wie bei MLE, mit Hilfe der roX RNA gewährleistet wird.
CG3613 wurde weiter charakterisiert und dabei festgestellt, dass es sich *in vivo* um ein Phosphoprotein handelt. CG3613 tritt auch gemeinsam mit der RNA Polymerase II auf und wird nach einem kurzen Hitzeschock zu Hitzschockloci rekrutiert, was ein Hinweis auf eine starke Verbindung zur Transkription ist. Ein Knock-down von CG3613 in Fruchtfliegen war lethal, jedoch führte ihre Reduktion in S2 Zellen zu einer Stabilisierung von intronhaltigen pre-mRNA's. Dies gibt zu der Vermutung Anlass, dass es eine Rolle in der pre-mRNA Verarbeitung spielt.

In dieser Arbeit wurden zum ersten Mal die mit roX2 interagierenden Proteine in der Fruchtfliege biochemisch definiert und dabei gezeigt, dass die MSL Proteine nicht die einzigen Proteine sind, mit denen roX RNAs zusammenwirken. Tatsächlich lässt diese Untersuchung darauf schließen, dass roX RNA's als Plattform dienen, um diese verschiedenen Proteine zusätzlich zu dem MSL-Komplex, zusammenzubringen und damit die Hypertranskription des männlichen X-Chromosoms zu gewäherleisten.

1. Introduction

The cell is the most basic form of life that we know of. It is a machine that replicates itself, responds and adapts to the environment. Almost all cells (with the exception of highly-specialized cells such as mature erythrocytes in mammals) contain DNA as their hereditary material that is replicated and passed onto the next generation during cell division. Another very important macromolecule and a major constituent of a cell is proteins. The "muscles" of the cell, proteins give the cell its shape as cytoskeleton, they form motor molecules such as myosin that gives mobility to the cells, and they make very efficient catalysts, called enzymes, that facilitate some of the most fundamental chemical reactions necessary for life. All proteins within a cell are encoded within its DNA, which is kept inside the nucleus in eukaryotic cells in a compacted form that is also known as the chromatin. The basic unit of chromatin is a nucleosome, which is made up of a tetramer of Histones H3 and H4 and two dimers of Histones H2A and H2B wrapped around by ~147bp of DNA [5]. Another core histone, H1, is known as the linker histone and usually found in regions of chromatin where a high level of compaction is observed. Nucleosomes are not only instrumental in blindly compacting DNA, they also provide a complex regulation platform that is acted upon by numerous chromatin modifying enzymes, such as methytransferases, acetyltransferases, demethylases, deacetylases, ubiquitin ligases, various ATP-dependent chromatin remodelers and many more (see [6], [7] and [8]).

In a process called transcription, message encoded in DNA is converted into a chemically similar molecule, RNA, that can be moved out of the nucleus where it is decoded into protein by a large ribonucleoprotein complex, the ribosome.

There are different types of RNA molecules produced in a cell. There used to be three categories of RNA that were functionally defined: ribosomal RNAs (rRNAs) that form the ribosome together with ribosomal proteins, transfer RNAs (tRNAs) that are carriers of activated amino acids which are used during protein synthesis and finally messenger RNAs (mRNAs) that carry the protein code from the DNA to the ribosomes to be deciphered. Today, however, we know that there are many more types of RNA, large and small, that participate in almost every aspect of cellular life.

In the next part I will mostly focus on the biogenesis of mRNAs and how its production is regulated.

1.1 The biogenesis and degradation of mRNA

From transcription to degradation, the life of an mRNA molecule involves many stages which are initiation of transcription, modification of the 5'end, splicing, termination of transcription and modification of the 3' end, export from the nucleus to the cytoplasm, translation and finally degradation. Almost all of these stages are highly regulated and many of them are interconnected: the outcome of one process affects the efficiency of the other and some of them occur simultaneously as the RNA is transcribed by the polymerase.

1.1.1 Initiation of Transcription

Arguably the most important point of regulation is the initiation of transcription. Although the DNA includes all the information necessary for the production all proteins in the cell, when a particular gene should be expressed, where and how much should it be expressed must be determined with accuracy and precision. The final molecule that integrates all the information and initiates transcription is the RNA polymerase (RNAP). A polymerase molecule interacts with a plethora of factors (protein and RNA) to initiate transcription at certain stretches on the DNA (transcription start sites, TSSs) under the direction of certain DNA elements (promoters, enhancer cis-regulatory modules etc.).

How does the polymerase know where to initiate? The human genome, for instance is made up of 3 billion nucleotides in 23 chromosomes. A very small fraction of this DNA actually codes for proteins (~1.5%, [9]), hence it is very important for the polymerase to find these genes and initiate transcription at those sites and not randomly. Promoters are special DNA sequences that are near or at the transcription start sites of genes. Eukaryotes have three types of polymerases (with the exception of flowering plants [10]) and different polymerases require

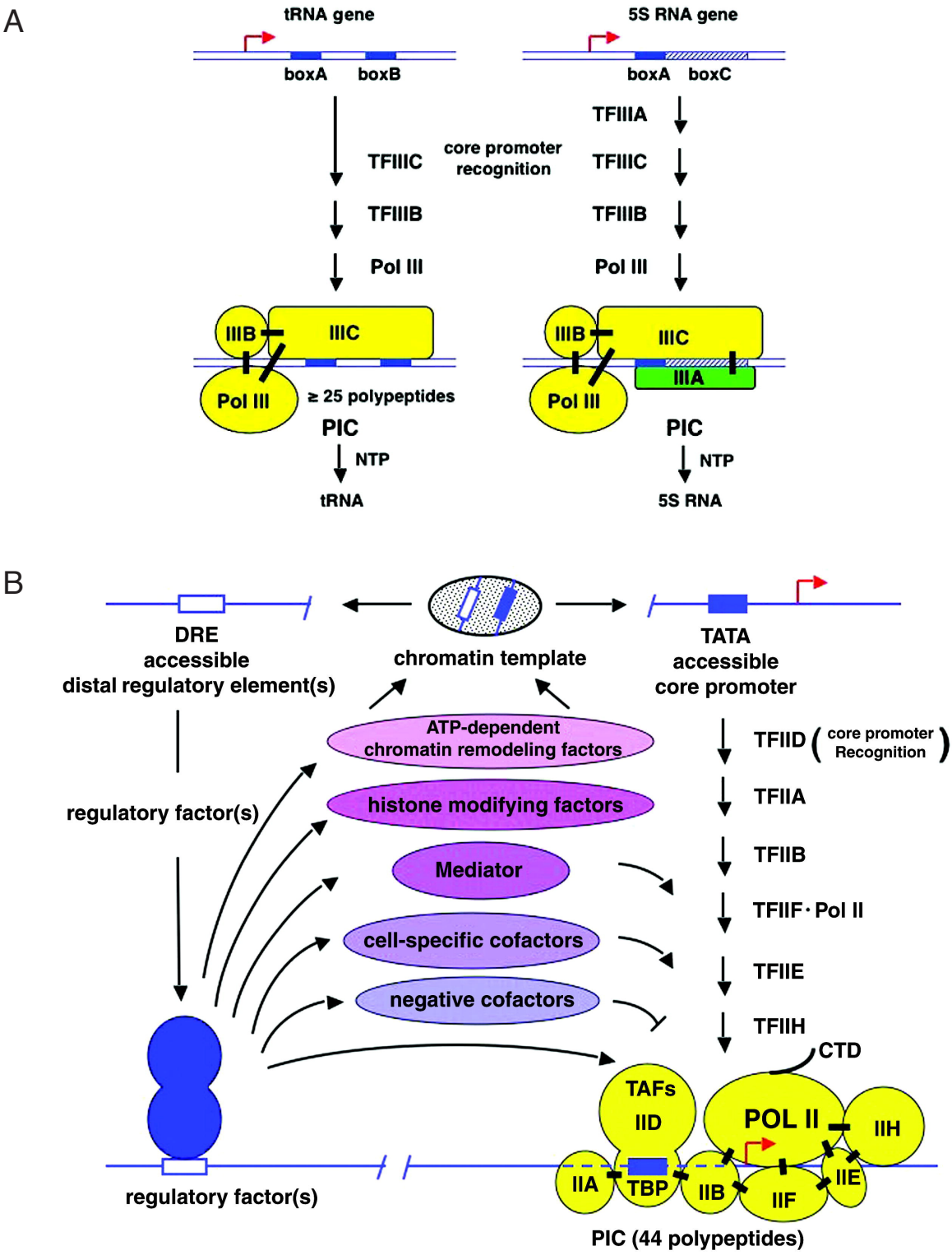

Figure 1 Initiation of transcription involves many factors In (A) two types of RNAPIII genes are depicted with the set of accessory factor required for the recognition of their promoter. In (B) the vast number of protein factors necessary for correct and timely initiation at an RNAPII promoter is shown. Adapted from Roeder, 2005 [3]

different types of promoters to initiate transcription. RNA polymerase I (RNAPI) is responsible for the transcription of a single gene which gives rise to 18S, 5.8S and 28S ribosomal RNA after several processing steps [11]. RNAPI also produces another transcript that is transcribed from the spacer region between rDNA genes, whose role was only recently determined by the Grummt laboratory as an epigenetic silencer of rRNA transcription [12],[13]. As would be expected, RNAPI does not require so many accessory factors for promoter recognition and initiation of transcription [UBF, TIF-IB (a stable complex of TBP, TAF_I48, TAF_I68 and $TAF_I95/110$), TIF-IA and TIFI-C] as it has a single promoter element in the genome. RNA Polymerase III, on the other hand, has several different genes that it transcribes which include the 5S rRNA gene, all tRNAs, U6 snRNA, RNaseP and other small RNAs [14]. There are two factors required for promoter recognition and correct initiation of RNAPIII genes, TFIIIC and TFIIIB [15]; however the 5S RNA gene requires an additional factor, TFIIIA that has to interact with the internal promoter elements (a peculiarity of RNAPIII genes) prior to TFIIIC and TFIIIB (Figure 1A).

Unlike RNAPI, which produces a single transcript and RNAPIII that produces several, RNAPII transcripts vary widely in terms of their sequence, structure and length. The promoter elements and other proximal or distal regulatory sequences that determine when and where a productive initiation will happen are very diverse and thus require a vast number of polypeptides to decipher them (Figure 1B).

The expression of a gene by RNAPII requires the formation of the pre-initation complex (PIC) on the core-promoter of that gene. PIC consists of RNAPII itself and general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. Distinct co-activators and ATP-dependent chromatin remodelers are required to expose the promoter by covalently modifying amino terminal tails of histones [16] and changing their positions along DNA. Sequence specific transcription factor binding to upstream sequences called enhancers can result in the recruitment of co-activators or co-repressors to a gene of interest. These factors usually relay their positive or negative affect on the rate of transcription via a highly conserved multi-protein complex called Mediator [17]. Finally, the DNA template is locally melted and RNAPII catalyzes the formation of the first few phosphodiester bonds of the mRNA in a template dependent manner. Productive elongation requires the phosphorylation of RNAPII C-terminal

domain (CTD) first by Cdk7 subunit of TFIIH and later by the Cdk9 subunit of PTEFb which is required to release the "promoter-proximal paused" RNAPII into the gene [18]. As the RNA emerges from the polymerase, it starts to interact with a number of factors that help the RNA turn into a mature mRNA that can be translated into protein in the cytoplasm by the ribosomes [19] ,[20].

1.1.2 Capping the 5' end

When the nascent transcripts is ~30 nucleotides long, all mRNAs are co-transcriptionally modified at their 5'-end by the addition of a guanosine moiety with an unusual 5'-5' triphosphate linkage which is immediately methylated at least once at position N7. This structure is known as the m^7GpppN cap [21]. Productive capping requires promoter proximal pausing and phosphorylation of ser5 residues of RNAPII CTD by TFIIH [22]. The 5' cap structure, and the cap binding complex (CBC) that binds to it in the nucleus confers stability to the RNA by protecting it from nuclear exonucleases [23], and is required in almost all the processes that follow transcription, i.e. splicing [24], export [25], 3'end formation [26] and translation initiation (where the CBC is replaced by eIF4E [27]) and failure in capping results in decreased mRNA stability [28]. It is generally accepted that the nuclear 5'-3' exonucleases such as Rat1 eliminate un-capped mRNA in the nucleus thus providing a checkpoint for proper mRNA transcription (Figure 3, [29]).

1.1.3 Splicing

In higher eukaryotes most mRNAs are interrupted by non-coding regions called introns. These sequences are removed from the transcribed RNA by the concerted action of hundreds of proteins and small RNAs, collectively called as the Spliceosome, and the phenomenon itself is called splicing of mRNA. Splicing is essentially two *trans*-esterification reactions where the 3'end of an exon is joined to the 5' end of the following exon, whereas the intron is released as a "lariat" (Figure 2 and see [4] for the molecular details of this process).

Many eukaryotic mRNAs contain short RNA sequences that regulate splicing by stimulating (enhancers) or inhibiting (silencers) the use of proximal splice sites. The decision to splice or retain an intron is influenced by the concerted action of many RNA binding proteins such as serine/arginine (SR) proteins, hnRNPs and other RNA-binding proteins (i.e. Nova, PTB) that interact with these RNA elements [30].

Splicing is also used as a means of regulation: the pattern of splicing can be altered by extracellular stimuli, which in the end may lead to different isoforms of a protein that may have different physical properties that result in different biological activities [31].

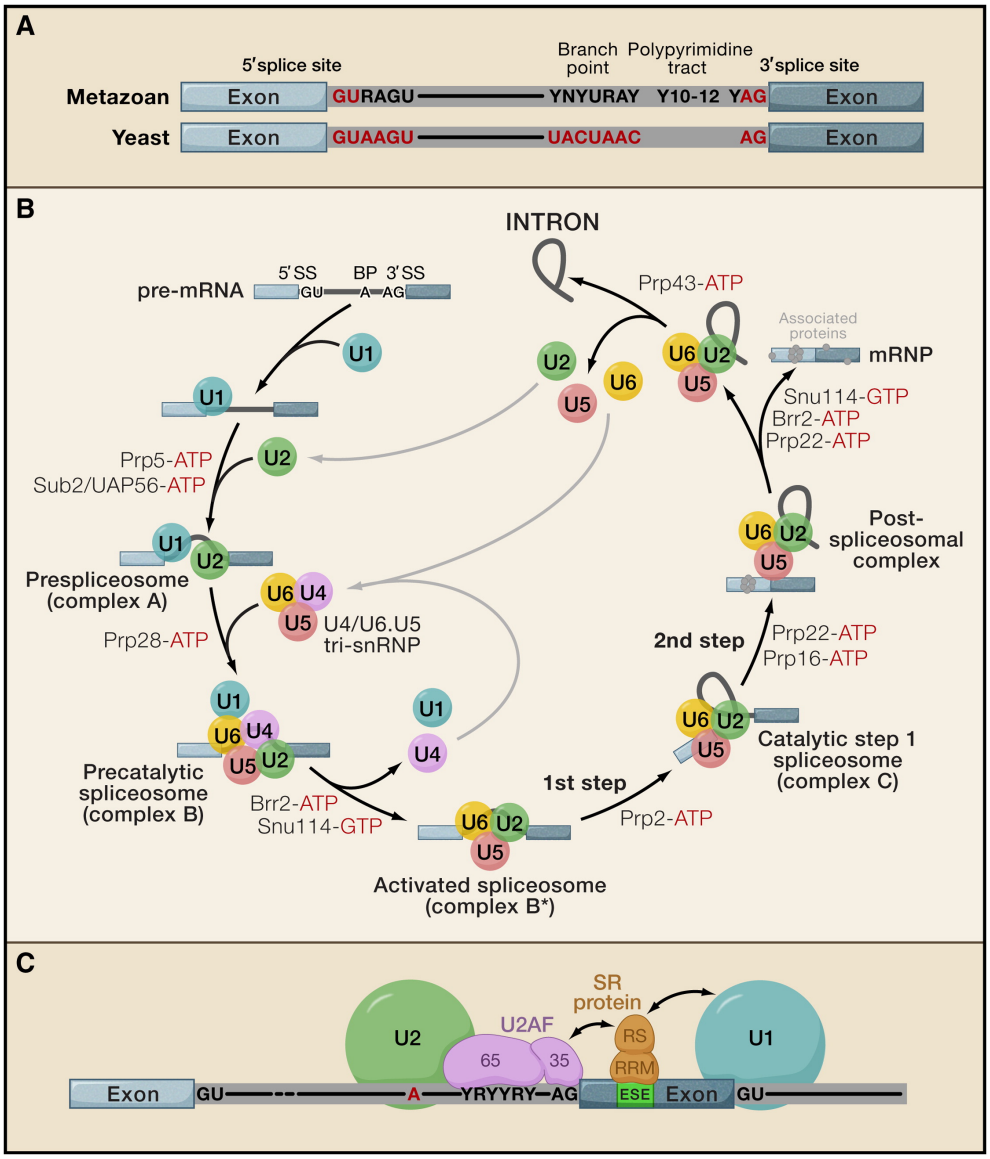

Figure 2 pre-mRNA splicing is a highly regulated multi-step process (A) Conserved motifs that define an intron in metazoa and yeast. (B) The splicing cycle involving the sequential action of snRNP (circles) and other proteins on an intron is depicted. (C) SR proteins can influence splice site selection by interacting with RNA elements within exons or introns and proteins that bind to 5'-or-3'-splice sites like U1 and U2AF. This can lead to splicing out an intron that would otherwise be retained in the final mRNA or retention of an intron that would otherwise be spliced depending on the nature of the SR protein. Adapted from Wahl, Will and Luehrmann, 2009 [4] .

1.1.4 Termination of transcription and polyadenylation of the 3' end of mRNA

When the polymerase reaches the end of a gene, transcription has to be terminated. In eukaryotes, polymerase does not fall off from the template as in bacteria [32], instead a cleavage/polyadenylation protein complex (polyA complex) by virtue of directly interacting with termination signals within the nascent pre-mRNA and with the Ser2 phosphorylated RNAPII CTD creates mRNA ends that are competent for export and translation [33].

Formation of a proper 3' end is thought to stimulate pre-mRNA splicing and ultimately protein expression [34]. Interestingly, components of polyA complex interact with general transcription factors such as TFIIB [35] and TFIID [36] that are situated at the core promoter of a gene. Such interactions are thought to be the result of gene looping, where the terminator of a gene interacts with the promoter of the same gene [37]. Failure to make a polyadenylated 3'end results in the retention of the pre-mRNA at the nucleus and finally to its degradation by the exosome [38], [39].

There is at least one more termination pathway in yeast that is independent of the polyA complex mentioned above. This pathway requires three polypeptides, Nrd1, Nab3 and Sen1 and is mainly used to form snRNA and snoRNA ends [40]. Nrd1 interacts with the CTD of RNAPII and recognizes the sequence GUAA/G on the RNA whereas Nab3 recognizes the sequence UCUU and maturation of the RNA ensues by the action of the nuclear exosome assisted by the TRAMP complex. It was recently shown that cryptic unstable transcripts use this pathway of termination rather than the canonical polyA complex mediated termination in yeast [41].

1.1.5 Export of mRNA to the cytoplasm

In eukaryotes, all mRNA have to be transported from the nucleus to the cytoplasm where they serve as templates for translation carried out by ribosomes. A general receptor composed of TAP-p15 heterodimer (also known as Mex67-Mtr2 or NXF1-NXT1) facilitates the transport of almost all mRNAs from the nucleus to the cytoplasm. The physical interaction of TAP/p15 with the Phe-Gly domains of FG-nucleoporins enables them to overcome the permeability barrier imparted by the

nuclear pore complex and enables the transport of mRNA from nucleus to cytoplasm [42].

Although TAP/p15 can interact with RNA, adaptor proteins usually mediate its interaction with diverse mRNAs. These adaptor proteins can be loaded onto the pre-mRNA in a way that is coupled to transcription and/or splicing. The mRNA packaging complex THO (absence of which results in RNA:DNA hybrids that inhibit elongation [43]) together with UAP56/Sub2 and Aly/Yra1 form the TREX (Transcription/Export) complex that facilitates the export of mRNAs into the cytoplasm. In yeast, THO complex co-transcriptionally loads the generic adaptor Yra1/Aly/REF together with Sub2 (UAP56) onto the emerging pre-mRNA throughout transcription [44]. In humans, on the other hand, loading TREX onto the pre-mRNA depends on splicing and on the cap binding protein CBP80 [25]. Aly/REF interacts directly with TAP/p15 and functions as a general adaptor molecule for many mRNAs. THO complex is certainly not the only way to provide adaptors to the mRNAs [45], SR (Ser/Arg-rich) proteins, although well known for their roles in splicing, can also interact with and recruit TAP/p15 onto mRNAs in a way regulated by phosphorylation cycles [46].

1.1.6 RNA surveillance

The cell is a dynamic bio-factory that has to recycle its products and as a major constituent of the cell RNA is no exception. There are broadly two reasons to return molecular RNA to its monomers: Either the RNA of interest is no longer needed, or that it has not completed its processing steps properly and/or has a pre-mature stop codon.

An RNA molecule can be degraded in three ways: it can be churned from the 5'- or 3'end by enzymes called exonucleases (5'-3' and 3'-5' exonucleases, respectively) or it can be digested in the middle by enzymes called endonucleases. Exosome is one of the most important protein complexes in eukaryotes and is the principle 3'-5' exonuclease both in the nucleus and in the cytoplasm [47]. In the nucleus, it is necessary for tRNA, snRNA, snoRNA and rRNA maturation and surveillance, whereas in the cytoplasm it is mainly utilized in the degradation of mRNA.

Almost all stages of mRNA production is subject to a checkpoint that determines if the mRNA will make it to the cytoplasm and translated. For instance, due to the presence of 5'-3' exonucleases such as Rat1, all mRNAs must acquire the m^7GpppN cap structure to escape degradation in the nucleus [29]. Promoter-proximal pausing of RNAPII is a way to ensure that all mRNAs are capped before productive elongation ensues. The cap structure is then needed for loading of TREX to the mRNA that facilitates its export to the nucleus [25]. Removal of the cap structure by Dcp1-Dcp2 protein complex in the cytoplasm makes the RNA a substrate for the abundant cytoplamsic 5'-3' exonuclease Xrn1 and leads to the degradation of that mRNA [48] .

Failure to splice out an intron can lead to the introduction of pre-mature stop codons in the final mRNA. The cell has various strategies to eliminate such transcripts from being translated. The first response is to get rid of these intron-containing pre-mRNAs within the nucleus, prior to their export into the cytoplasm [49]. Although the details are not clear, the nuclear exosome is reported to be involved in this process as loss of exosome function lead to a 20-50 fold increase in the abundance of unspliced pre-mRNAs in the nucleus. In yeast, nuclear pore complex (NPC) component Mlp1 is involved in the retention of unspliced mRNA providing a final block in the way of these aberrant mRNAs reaching to the nucleus [50]. If these intron-containing mRNAs that now have premature terrmination signal reach the cytoplasm, there is another way to dispose of these transcripts that depend on the deposition of a group of proteins, called the Exon Junction Complex (EJC), on the mRNA in a sequence-independent manner during splicing in mammals. The EJC is deposited ~20 nucleotides upstream of exon-exon junctions following splicing [51]. The EJC core is composed of four proteins (eIF4AIII, MAGOH, MLN51 and Y14) and it travels with the mRNA into the cytoplasm. Once the mRNA engages with the ribosome, a "pioneer" round of translation takes place [52]. In a normal mRNA, all EJCs should be deposited downstream of a stop codon, in an mRNA with a premature termination signal, however, this is not the case and recognition of the premature stop codon by the release factors (eRF1 and eRF2) recruit a group of conserved proteins (UPF and SMG proteins) that, together with the EJC and CBP80 trigger events that lead to the rapid degradation of mRNA. This process is called non-sense mediated decay (NMD) and it highly conserved among eukaryotes [53].

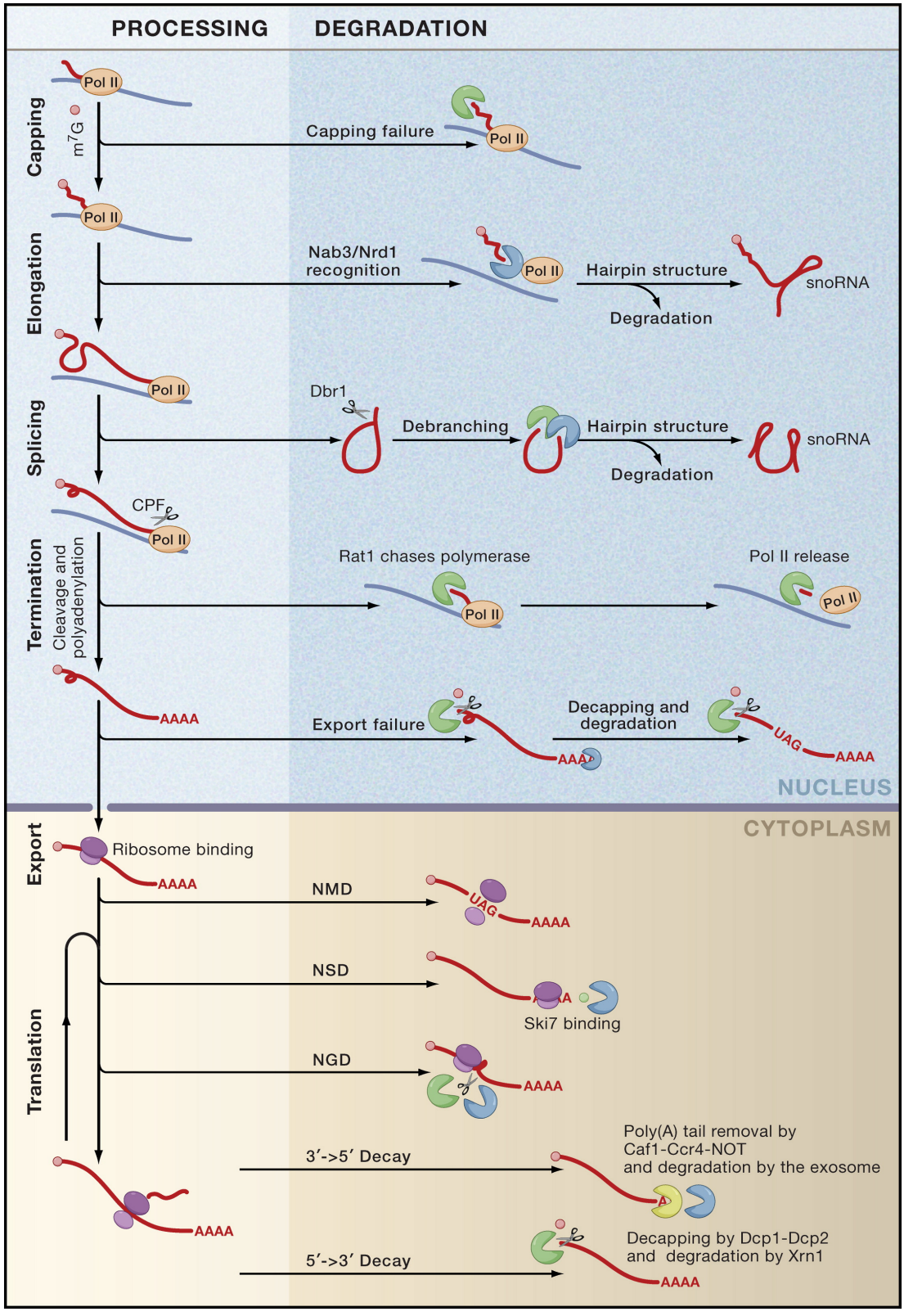

Figure 3 Many pathways of RNA degradation The cell has multiple layers of control over the production mRNAs both in the nucleus and in the cytoplasm. Adapted from Houseley and Tollervey, 2009 [1].

As mentioned before, proper termination of transcription and 3'-end formation is a very important step in the life of an mRNA and pre-mRNAs that fail to get a polyA tail at the end of the transcription are retained at their sites of synthesis. This retention is dependent on Rrp6 and the nuclear exosome [54]. These RNAs that would become poor templates for protein synthesis in the cytoplasm are sequestered in the nucleus by the complex that will degrade them [55].

Regulated degradation of RNA is an important tool used by the cell for rapid response to extracellular stimuli. For instance, some mRNAs contain AU-rich elements (ARE) at their 3'UTR that make them unstable in the cytoplasm. These mRNAs usually code for proteins that are only transiently activated upon extracellular stimuli (i.e. proto-oncogenes, cytokines) hence the mRNAs that code for these proteins must have short half-lifes. Cytoplasmic exosome is known to be mainly responsible for the degradation of these mRNAs, but the exosome requires adaptor proteins, collectively called as ARE-binding proteins, that recognize and bind to these AREs in the 3'UTR of target genes [56].

Cryptic initiation of transcription in the genome is probably unavoidable due to the stochastic nature of molecular events. The cell eliminates most of these transcripts rapidly and their detection usually requires tempering with Exosome function [57]. All the details of this process is not known, but a pathway involving Nrd1-Nab3 (which, as discussed before, is used for the formation of snRNA and snoRNA ends) is well studied in yeast. [41]. The sequence that Nrd1 recognizes is very short (GUAA/G) and although many pre-mRNAs would be predicted to have it, this pathway does not eliminate most of these pre-mRNAs. It was proposed that such a system would make any mRNA that has trouble with one of the processing steps a target for degradation by Nrd1-Exosome pathway, but a correctly processed pre-mRNA would escape this fate and become a mature mRNA that is transported into the cytoplasm [47].

1.2 Large non-coding RNAs

A very large proportion of a cell's recourses go into the production of mRNAs and to the machinery that decodes its message into proteins, which are by and large the effector molecules of a cell. Yet, RNA is not just the medium between DNA and protein; catalytic RNAs, known as ribozymes, are known to exist for a long time [58]. Our knowledge about the repertoire of RNAs produced in the nucleus has increased significantly over the last two decades. It now seems like almost all of the genome of eukaryotes is transcribed [59] and this lead to the term "eukaryotic genomes as RNA machines" [60]. The debate about the functional relevance of these transcripts is ongoing (are these transcripts functional or are they just genomic noise, (see [61]) however the list of ncRNAs with an assigned function is growing rapidly ([62], [63]) and a description of all of them is far beyond the scope of this text. I will, however, discuss some of the large non-coding RNAs that take part in gene regulation at the chromatin level and especially dosage compensation in mammals and in *Drosophila* (see Table 1 for some of these RNAs).

Long ncRNAs identified so far range from a few hundred base pairs to kilobases (Table 1). Classical examples include Xist and Tsix RNAs in dosage compensation in mammals [64] Air RNA [65], H19 RNA involved in genomic imprinting in mammals [88] and roX RNAs involved in dosage compensation in Drosophila. However, it is clear that the repertoire of these long RNAs and the roles they assume in the cell is ever increasing ([89]; [90] and [91]). For example, recently discovered, HOTAIR (~2kb) is situated in the HOXC loci in mammals, and when expressed leads to the repression of genes in the HOXD loci *in trans* [70]. Its depletion leads to loss of H3K27Me3 and PRC2 protein Suz12 from the HOXD cluster. HOTAIR was further shown to directly interact with Suz12. Detailed work on PRC2 showed that there are many other ncRNAs it interacts with ([92],[93]) and these interactions may be modulated by its phosphorylation [94]. Although the abovementioned examples may suggest so, ncRNA in mammals are not only used for down-regulating transcription. For example, heat-shock-RNA-1 (HSR-1) was shown to interact with HSF and activate heat-shock responsive genes [83] Shiekhattar laboratory has recently shown that some long ncRNAs can positively

regulate neighboring protein coding genes, acting similar to enhancer elements [87]. These examples show that mammals can regulate gene expression by altering chromatin via ncRNAs both *in cis* and *in trans* and both negatively and positively.

Eukaryotic long ncRNAs that play a role in gene regulation

	Name	Size	Organism	Function	References
Repressive RNAs	*Xist*	14-18 kb	Human, Mouse	X chromosome inactivation	[64]
	Tsix	40kb	Mouse	Down regulates Xist expression	[64]
	7SK	331nt	Vertebrates	Inactivation of P-TEFb, repression of RNAPII transcription	[66]; [67]
	Airn	108kb	Mouse	Monoallelic silencing of *Igf2r in cis*; *Slc22a2* and *Slc22a3 in trans*. Interacts with G9a H3K9 HMT.	[68],[69]
	HOTAIR	2.2kb	Human	Interacts with PRC2, represses target genes *in trans*	[70]
	B2	178nt	Mouse	Binds RNAPII, represses transcription	[71],[72]
	$ncRNA_{CCND1}$	200-330nt	Human	Represses transcription of *CCND1 in cis* by interacting with TLS	[73]
	NRON	2-4kb	Human	Repression of NFAT	[74]
	Kcnq1ot1	91kb	Mouse	Interact with G9a and PRC2, represses transcription from *Kcnq1* locus *in cis*	[75]
	ANRIL	3.8kb	Human	Interacts with PRC2 and PRC1 to repress INK4b/ARF/INK4a locus	[76],[77]
Activating RNAs	roX1, roX2	600nt – 3.7kb	Drosophila	X chromosome hypertranscription	[78]
	SRA	883nt	Mammals	Co-activator of nuclear receptors	[79]
	tre-1, tre-2, tre-3	950, 1109, 351nt	Drosophila	Activates transcription of target genes by Ash1 recruitment	[80]
	U1	165nt	Human	Interacts with TFIIH, activates transcription	[81]
	Evf-2	3.8kb	Human, Mouse	Interacts with Dlx-2 and increases its enhancer activity	[82]
	HSR1	600nt	Mammals	Interacts with eEF1A and activates HSF1 upon heat-shock	[83]
	LINoCR	1.9kb	Chicken	Activates lysozyme expression upon LPS exposure	[84]
	Jpx	3.8kb	Mouse	Induction of Xist expression	[85]
	HOTTIP	3.7kb	Chicken	Regulates HoxA locus	[86]
	ncRNA-a1-7	~400nt	Human	Activation of specific genes within ~300kb, enhancer-like RNAs (eRNAs)	[87]

Table 1 A selection of long ncRNAs that are involved in regulation of gene expression

1.2.1 Dosage compensation in mammals

A recurring theme in evolution, dosage compensation is a system that evolves when the sex in a species is determined genetically by sex chromosomes, where one copy of the chromosome has degenerated in time, which causes an imbalance in gene dosage since one sex has one copy of the sex chromosome whereas the other one has two (discussed in detail [95]). Dosage compensation in mammals, as mentioned above, is a classical example where a long ncRNA, Xist, is paramount in the silencing of one X chromosome in females ([64],[96]). Xist is a ~17kb transcript in humans (~14kb in mice) that is exclusively transcribed from the inactive X chromosome. It is known that Xist only functions *in cis*, when moved to an autosome it leads to the silencing of flanking loci, and can not be detected on the X chromosome. One apparent advantage of such a mechanism is that the ncRNA can immediately mark the region to be silenced and bypasses the need to be transported to the cytoplasm and imported back to the nucleus to find its target [64]. Exact mechanism by which Xist functions is a matter of intense investigation. However, accumulating evidence suggests that Xist is unlikely to be the only player regulating this complicated process and epigenetic factors also pay at an important contribution. For instance, it has been shown that coating of the X chromosome with Xist RNA is followed by the accumulation of heterochromatin marks such as trimethylation of histone H3 lysine 27 [97]. This histone mark is thought to be catalyzed by Polycomb Repressive Complex 2 (PRC2) that contains the H3K27 specific histone methyl transferase (HMTase) EZH2, and is maintained by Polycomb Repressive Complex 1 (PRC1), which is thought to contribute towards silencing of the X chromosome [98].

1.2.2 Dosage compensation in Flies

Dosage compensation in flies, in principle, does the same job as in mammals that is to equalize the gene output from the unequal number of X chromosomes between the two sexes. The flies however, carry out dosage compensation by hypertranscription of the single X chromosome in males rather than silencing one of the two X chromosomes in females. This process is carried out by a ribonucleoprotein complex named the Dosage Compensation Complex (DCC) or the MSL complex. It is comprised of Male Specific Lethal (MSL) proteins; MSL1, MSL2, MSL3, a histone

H4 lysine 16 specific acetytransferase MOF (*m*ales absent *o*n the *f*irst), and RNA/DNA helicase MLE (maleless). The complex also harbors two large ncRNAs, *roX1* and *roX2* (RNA on X) (Figure 4). The utilization of large ncRNAs and their indispensability in dosage compensation parallels between mammals and flies, although the mechanism by which they achieve the final effect seems to be entirely

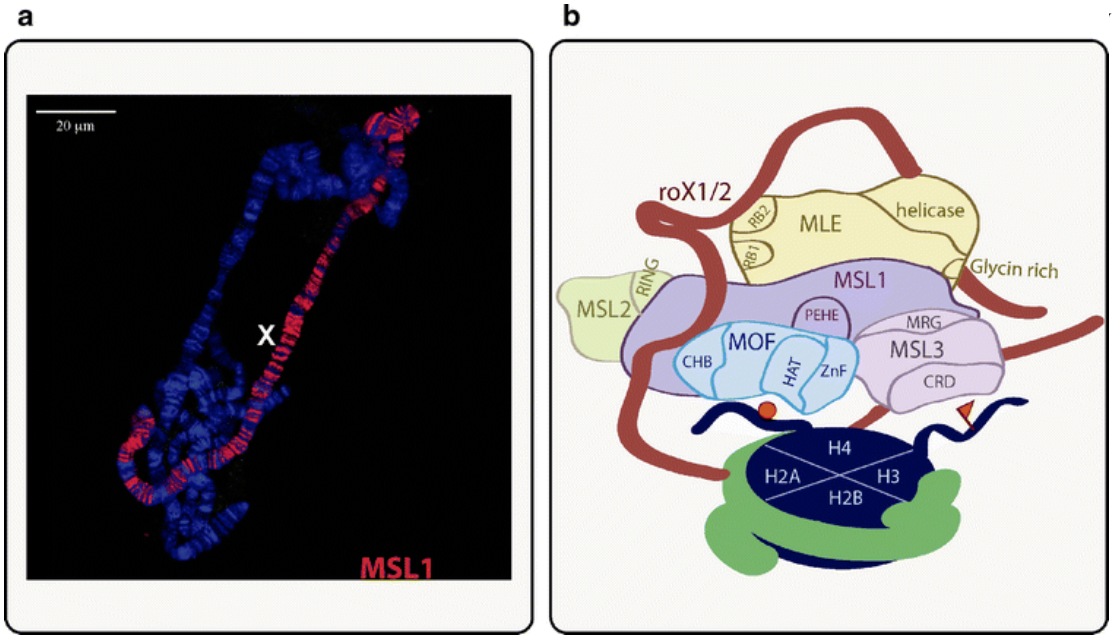

Figure 4 The MSL complex is required for dosage compensation (A) A polytene staining showing the classical X-localization of MSL1. All of the other four proteins of the MSL complex give similar staining patterns. (B) The MSL complex is a ribonucleoprotein complex. Various domains that facilitate protein-protein, protein-RNA and possibly preotin-DNA in these proteins are depicted here. Adapted from Hallacli and Akhtar, 2010 [2].

1.2.3 Discovery of roX RNAs

The *roX* RNAs were initially discovered in a screen looking for RNAs that are exclusively expressed in one sex but not in the other [99]. It was immediately understood that the regulation of these RNAs were under the control of the dosage compensation system as lack of *msl-1, msl-3* or *mle* lead to the disappearance of the *roX* RNAs [99]. In an independent study, an enhancer trap strategy was used to look for *lacZ* expression in the mushroom bodies and one line showed expression of the reporter only in females [100]. The characterization of this line revealed that the transgene had landed on the *roX1* gene. This study also looked in detail into the nuclear localization of the *roX1* RNA using *in situ* hybridization on salivary glands which not only showed that MSL-1 and *roX1* co-localized on the male X

chromosome, but *roX1* was still "painting" the male X chromosome when its locus was translocated onto an autosome. So whatever this RNA might have been doing, it was clear that it could function *in trans* and this was in marked contrast to the biology of the *Xist* ncRNA which only works in *cis* [101], [64].

1.2.4 Similar yet different: roX1 and roX2

Both *roX* RNAs seemed to be transcribed by the RNAPII, since they could be isolated using oligo dT probes and contained putative introns which were spliced out in the isolated cDNAs ([100], [99]). In addition, recent studies showed strong RNAPII binding on both genes using ChIP-chip analysis, further indicating that RNAPII is the polymerase responsible for *roX* transcription [102].

It is important to note that, these two RNAs looked very different in both size and sequence; *roX1* RNA being 3.7 kb *roX2* being 0.6kb in length ([99], [100] and [103]). Yet, *roX1* and *roX2* RNAs seemed to be similar in many ways. They are both stably expressed in males and both genes map to the X chromosome **(**roX1: Location 3F, roX2: Location 10C) and they are stable only in the presence of the known MSL complex members [101],[64]. These RNAs could, at this point, be thought of as by-products of a chromosome-wide dosage compensation system, since they were obviously expressed only from the hypertranscribed male X chromosome; but it was also possible that they were actually not the result of this hypertranscription but were taking part in its realization and hence were only detectable in males where this system is up and running. Initial experiments showed that *roX1* was completely dispensable, its deletion had no apparent phenotype [100], which would argue that it has nothing to do with dosage compensation. Yet, the RNA coated the X chromosome in male nuclei, co-localizing with MSL-1 protein. One possibility would be that these two RNAs were redundant in function, although bearing almost no sequence similarity, and hence deletion of one would not result in any apparent phenotype, whereas a double mutant would reveal if they are both involved in dosage compensation. This hypothesis was tested with a deficiency that removed the *roX2* gene, which was combined with the $roX1^{MB710}$ used as a null allele. The results were striking: The MSL proteins were no longer localized to the X chromosome as judged from immunostaining of male embryos with antibodies against MSL-2. The transcriptional effect of this loss of localization on dosage compensation could not be

determined as the deficiency that removed *roX2* also removed essential genes like the large subunit of RNA polymerase II [101]. In the same study, a 30nt similarity between the otherwise completely different *roX1* and *roX2* RNAs was noticed. Later studies showed that deletion of this region as well as another ~110bp stretch of similar sequence in the two RNAs (the *D*NAseI *H*ypersensitive *S*ite, DHS), the latter being able to recruit the MSL complex from *roX1* or *roX2*, resulted in no obvious phenotype when these transgenes were the only source of *roX* RNA ([104], [105] and [106]). These results indicated that *roX* RNAs carry out their functions by virtue of degenerate primary, secondary or even tertiary structures that are very difficult to identify computationally, or that there might be further redundant elements in these RNAs so that deletion of one functional region would not yield a phenotype since it will be complemented by another element in these large RNAs. Such a redundancy was systematically tested for *roX1* [106] and it was found that apart from a putative stem loop region at the 3' end of the RNA, deletions of approximately 400 bases did not affect the rescue of male lethality by these truncated RNAs in $roX1^{-}$ $roX2^{-}$ mutant background. An analogous study on the *roX2* RNA gave similar results: deletion of most of the conserved residues in small chunks (17bp) did not change the rescue of male lethality when the mutated RNA was the only source of *roX* RNA in these flies [107]. The redundant nature of these RNAs might be explained by their lengths. If they interact with their cognate proteins via an induced fit mechanism where the RNA rather than the protein assumes a stable 3D conformation upon binding, bulk of the primary sequences might be dispensable giving the RNA molecule much more flexibility [108]. These now dispensable sequences might evolve into novel protein interaction modules in time.

Interestingly, the embryonic expression pattern of *roX1* and *roX2* is not identical. *roX1* transcripts, for instance, are detected very early (2hrs AEL) and are detected in both sexes; *roX2* on the other hand is not detectable before 6hrs AEL and thereafter is detected only in males [109]. Ectopic expression of MSL-2 in females, which leads to the formation of DCC in female flies and results in female lethality [110] has revealed another striking difference between the two RNAs: When females express MSL-2 ectopically but are also mutant for *mle*, both *roX* RNAs were detected only at their sites of transcription, but not on other chromosomal sites when polytene squashes from these flies were investigated by *in situ* hybridizations carried out with probes

against *roX1* and *roX2*. In *msl-3* mutants, however, *roX1* was still only detected at its site of synthesis, whereas *roX2* could be detected in other regions, mostly on the X chromosome, including the 3F band which is the site of the *roX1* gene [111].

1.2.5 Why males can't do without -at least one of- them?

The data from $MSL2^{+}$ females (females that over-express MSL2 in *msl3* mutant background) discussed above had multiple implications: First of all, in the absence of MLE, it seems that *roX* RNAs are unable to co-localize with the rest of the complex at the remaining "high-affinity" sites on the X chromosome. Which would mean that whatever is left from the MSL complex is devoid of *roX* RNAs in this situation. Nevertheless, when MSL3 is absent, *roX2* is able to co-localize with the partial MSL complexes on the X chromosome. This lead to the idea that MLE was indispensable in the integration of *roX* RNA into the complex and that these *roX*-free partial complexes were able to bind some of the X chromosomal sites. Co-localization, detected by microscopy, however, is no proof that the detected molecules do interact *in vivo*. However, immunoprecipitation (IP) analysis carried out with antibodies against MSL1, MSL3 and MOF and MLE showed that *roX2* RNA is physically associated with the MSL complex ([112], [113] and [103]).

The RNase sensitive chromatin association of MLE and the following discovery of the *roX* RNAs had begged the question whether MLE, as an ATP dependent RNA/DNA helicase associated with the *roX* RNAs and together with the other MSL proteins regulates dosage compensation ([114], [115] and [116]). Further evidence in support of this view emerged when the chromodomains (now known as the chromo-barrel domain) of MOF and MSL-3 were shown to interact with RNA [112]. In the same study, association of MOF with the X chromosomal territory in SL2 cells was shown to be RNase sensitive. In addition, an IP with antibodies against MOF under conditions where MLE protein was no longer co-precipitated with the rest of the complex was carried out and yet *roX2* RNA could still be detected with levels close to a less stringent IP where MLE protein is robustly detected. This interaction was dependent on the presence of the wild-type chromodomain of MOF. The presence of *roX* RNAs, and the interaction of MOF with these RNAs seems to be important for proper association of MOF with the X chromosome. Notably, a recent structural study

on the MOF homologue in yeast, Esa1 has revealed that the chromodomain together with an N-terminal extension is competent in binding RNA *in vitro* and interestingly, mutations that disrupt this structure were lethal. This might indicate that interaction with RNA might be a common feature that is necessary for MOF function conserved from yeast to flies [117].

The results mentioned above clearly showed that the *roX* RNAs are stable members of the MSL complex; they actually seem to be even more stably associated with the rest of the complex than MLE, the protein they are required for stability and association with other MSL proteins *in vivo* [112], [118]. Their stability, however, is compromised in mutants with a single residue change that renders MLE incapable of hydrolyzing NTPs ([119],[115] and [120]). It is interesting to note that, in these mutants the MSL proteins are detected on the "chromatin entry sites" where they co-localize with each other in total absence of both *roX* RNAs. Similarly in *roX1*$^{-}$ roX2^{-} mutants, MSL proteins are mistargeted to autosomes and chromocenter where they co-localize with H4K16 acetylation [121]. Another line of evidence that might indicate that a *roX*-free MSL complex exist comes from the observation that in wild-type females where MSL-2 is not present, the seemingly non-specific chromatin binding of MSL-1, MSL-3 and MLE is completely lost when either one of these proteins is not expressed [122]. MSL1 is also shown to directly interact with MSL-3 and MOF in addition to its better known interaction with MSL-2 without which it is less stable [123], [124]. In a very intriguing experiment, MSL1 along with MSL2 is over-expressed in flies in the absence of *roX* RNAs. The results showed that this co-over-expression of the core MSL proteins lead to a clear increase in the number of male progeny in the absence of *roX* RNAs (~6 to 70-fold) and improved the X chromosomal staining of the MSLs, with the notable exception of MLE. However, MSL proteins were still strongly localized to ectopic autosomal sites and the chromocenter [125].

These studies raise the following question: If the MSL complex is able to form, land on the X chromosome and acetylate H4K16 in the absence of detectable *roX* RNA, then why do these RNAs exist and why does their absence result in male lethality? First of all, although all three claims are correct, they are at least inaccurate. There are indications that the complex does form without *roX* RNA –as mentioned above–; but the amount, stability and/or the stoichiometry of the individual components in these partial complexes are not known; and all of these parameters could be vital to the

ultimate function of the MSL complex. These partial complexes also seem to bind to the X chromosome, but almost only to the "high-affinity sites" reported for *msl3* mutant males as well as to the chromocenter and the autosomes and these complexes are engaged in acetyltransferase activity [126], [121]. In females that express MSL2 ectopically and have only catalytically inactive MOF (mof^1), the MSL complex localizes to a reduced number of sites on the X chromosome as judged from polytene squashes. Yet, when salivary gland nuclei are investigated, MSL proteins look like dispersed in the nucleoplasm but still co-localize. *roX1* also follows this pattern by giving a diffuse staining in the nucleus [119].

Thus it seems likely that the function of *roX* RNAs is to make the MSL complex more competent in targeting the X chromosome rather than the autosomes and facilitate the binding of the MSL complex to more genes on the X chromosomes perhaps by increasing the local concentration of the DCC.

1.2.6 RNA independent role of roX?

The *roX1* and *roX2* loci coincide with the 35-40 "high-affinity sites", named so because MSL1 and MSL2 was able to bind these sites in the absence of MSL3, MOF or MLE. In early experiments where genomic clones of *roX1* or *roX2* were integrated onto an autosome, the whole set of MSL complex proteins could be detected on the transgene; moreover, rather unexpectedly, "spreading" of the complex to the flanking chromatin was observed [127]. These experiments showed that the MSL complex could be recruited onto an autosomal position when a "high-affinity site" was transferred there; but more importantly they showed that the complex is recruited to chromatin that normally cannot recruit the complex but is now able to do so because it is now adjacent to a "high-affinity site". A model for the X chromosome targeting of the MSL complex was proposed based on these observations: MSL1/MSL2 can recognize a small number of sites on the X chromosome in a sequence specific manner (the "high-affinity" sites) and then spreading of the complex occurs *in cis* to recognize all dosage compensated genes on the X chromosome [127]. This model was later challenged by experiments that looked into MSL recruitment in X to autosome or autosome to X transpositions. It was shown that independent of the presence of a previously defined "high-affinity site", any piece of X chromosome larger that ~40kb was able to recruit MLS complex when inserted onto an autosome and "spreading" of

MSL proteins into flanking chromatin was not observed. Likewise, an autosomal region that normally does not recruit MSL proteins will still not do so when inserted onto the X [128]. Together with the findings of [129], a new hypothesis emerged that explains MSL recruitment, the "affinities model" [130], [131]. According to this model, the "high-affinity" sites are simply more efficient in recruiting the MSL complex and are not qualitatively different from "low-affinity" sites. When the MSL complex is limiting, or its composition is compromised, only "high-affinity sites" seem to recruit the MSL complex; but what happens is that they simply recruit the MSL complex much better than the "low-affinity sites" and hence are the only sites visible when analyzed by immunofluorescence. This model, however does not explain why MSL binding and up-regulation of transcription is observed in the loci that flank *roX* transgenes that are moved to autosomal positions. Furthermore, the recent findings show that MSL3, like its homologue in yeast Eaf3, can bind to trimethylated histone H3 lysine 36 (H3K36Me3), a histone modification that is found on the body of actively transcribed genes, and this modification seems to stabilize the MSL complex on target genes [132], [133] and [134]. These results implicate transcription, rather than specific sequence elements as cues that lead to MSL recruitment at least to some genes. As specific examples, the MOF and CG3016 genes, have been shown to be sitting in between these two models for MSL targeting. MOF, as an X-linked gene, is targeted by the MSL complex, and this does not change when it is inserted to an autosome. The gene retains MSL binding in the absence of *msl3*, making it technically a "high-affinity site". Interestingly, however the binding of the MSL complex is dependent on the transcription of the gene and more interestingly the direction of the transcription is not important as long as polymerase runs through the gene. Finally, targeting of the complex is diminished when the 3' end of the gene is deleted [135]. This lead the authors to propose a model where targeting requires redundant DNA elements at the 3' end, but these elements have to be made accessible via transcription through the gene. It is interesting to note here that Kuroda and colleagues mapped the "high-affinity sites" by ChIP-on-chip and ChIP-seq in *msl3* mutants and identified enriched motifs on the X and named these motifs *M*SL *r*ecognition *e*lements (MREs). The reported X chromosomal versus autosomes enrichment of these motifs is around 2-fold, but it increases to 4-fold when active and especially 3'end of the genes are taken into account [136].

1.2.7 Relationship between roX and H4K16 acetylation

Recent reports have suggested that evolutionarily conserved stretches present at the 3'end of *roX1* and *roX2* RNAs are functionally important [107], [137]. A putative stem-loop was previously suggested to play an important role in *roX1* function since its deletion resulted in the reduction of male viability [106]. *roX1* contains three of these elements (roX1-box1-3) whereas roX2 has two. The 3' stem of the stem-loop of roX1 entails roX1-box2. It was noticed that rox2-box1 could also reside in a stem-loop at the end of the *roX2* RNA. This stem-loop is reported to be able to recruit the MSL complex onto the X chromosome when concatemerized into a 6-mer and expressed in the absence of endogenous roX RNA [138]. The surprise was that, although the staining patterns of MSL proteins on polytene chromosomes were seemingly indistinguishable from wild-type (including H4K16Ac, but except occasionally for MLE), the rescue of male lethality was very low, ~17% but was significantly higher than a mutant that cannot form this stem-loop or a 6mer with the anti-sense sequence (~1%). A follow up study further established the importance of this putative secondary structure. In this study, the authors showed that deleting or mutating this secondary structure in *roX2* leads to male lethality [107]. When polytene squashes were examined, it was seen that targeting of the MSL complex to the X chromosome was still normal, however H4K16Ac was not detectable. Since these conserved sequences were also present in *roX1*, their functionality were also tested in that context. As mentioned above, Stuckenholz et al. had previously reported that a stem-loop at the 3' end of *roX1* was important for the function of this RNA. Since this stem loop contained rox1-box2 in its 3' stem, the authors tested MSL localization and H4K16Ac in these mutant flies where the stem-loop is deleted and found that all the stainings were normal. In accordance, they report a 55% male rescue as opposed to the 22% reported by [106]. When a mutation of rox1-box3, that did not alter MSL localization or histone acetylation by itself, was combined with the stem-loop deletion, severe reduction of H4K16Ac from the X chromosome was observed while all other MSL complex members seemed to still co-localize on the X chromosome with a male rescue frequency of 9% [107].

2. Aims and Objectives

In flies, males hypertranscribe their single X-chromosome in order to equalize their gene expression to that of females which have two X-chromosomes. This process, called dosage compensation, is dependent on the presence of a ribonucleoprotein complex, the MSL complex, which coats the male X-chromosome in its entirety and acetylates H4K16. roX RNAs are an integral part of this complex, and although their exact role is unknown they are necessary for dosage compensation to take place.

Genetic studies aimed at understanding the role of roX RNAs in dosage compensation have been very fruitful; nevertheless, biochemical approaches that complement these studies have been lagging behind. It is therefore essential to look into roX RNAs using biochemical tools in order to understand their function in the context and perhaps beyond dosage compensation.

In my PhD work, I wanted to understand how roX RNAs take part in dosage compensation. We tackled this problem from two angles. The first approach was to affinity purify MLE, which is an RNA/DNA helicase that is part of the MSL complex and was shown to be required for the incorporation of roX RNAs into the MSL complex. This work is summarized in the first part of the Results section.

In the second approach I purified roX2 interacting proteins *in vitro* using GRNA chromatography and then analyzed them using biochemical, cytological and immunological methods. This part of the work is summarized in the second part of the Results section.

3. Results

3.1 Tandem Affinity Purification of Maleless

Affinity purification of protein complexes can give invaluable information about the function of a protein if the co-purifying components are well-studied proteins, or if the components of the complex have protein domains that suggest a function altogether.

Maleless, especially together with its mammalian orthologue RNA Helicase A, can be considered as a well-studied protein ([115, 116, 139 and 141]). Its activity as an RNA/DNA helicase is studied in detail, mostly *in vitro* and *in vivo* ([115], [139, 142 and 143]. And yet, activity does not necessarily determine function: although MLE is known to be required for dosage compensation, it is not known how its helicase/ATPase activity is utilized for its function as a male specific lethal protein.

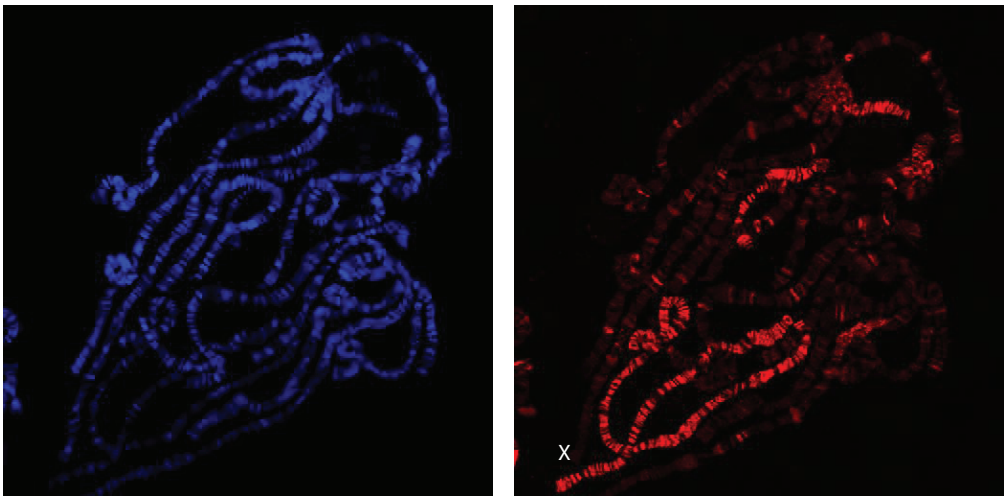

Figure 5 Maleless stains autosomal loci in addition to the male X-chromosome Salivary glands from a 3rd instar male are squashed and stained with antibodies against MLE (red). DNA is counter-stained with DAPI (blue). X-chromosome is indicated with an "X".

MLE protein seems to have additional functions beside dosage compensation. A simple immunological staining of polytene squashes of male third instar larvae shows strong inter-band staining in some autosomal loci in addition to the classical X-chromosomal staining (Figure 5). Encouraged by these observations, we wanted to carry out an affinity purification of MLE and identify proteins that co-purify with it.

There are many different affinity tags available to a biochemist, all with their advantages and disadvantages. Perhaps the most commonly used affinity purification approach utilizes the interaction between *S. aureus* protein A and IgG as the first purification step, followed by the Calmodulin binding protein (CBP) calmodulin interaction in the presence of calcium as the second purification step [144]. TAP-tag forms a relatively bulky domain (~21kDa) and when used as an affinity tag it may interfere with the function of the protein of interest and/or alter its affinity towards some of its interaction partners. And although the elution from IgG beads is carried out under native conditions, a protease is used for this purpose. The elution from calmodulin beads is carried out by chelating calcium out of the solution using EGTA; but although EGTA is a strong chelator of Ca^{2+} it also chelates other divalent metal ions such as Mg^{2+} and Zn^{2+}, which may destabilize the structure of certain proteins (i.e. Zinc-finger containing proteins) that require these ions.

I have used FLAG [145] and hemagglutinin (HA) [146] tags in tandem for the affinity purification of MLE. HA tag corresponds to the amino acids 98-106 of human influenza hemagglutinin glycoprotein and has been used extensively for isolation, detection and purification of many proteins. The tag is very small (9 amino acids) and highly specific monoclonal antibodies as well as polyclonal antibodies are available that bind to it. Elution occurs under perfectly native conditions by competing the bound protein by high concentrations of HA-peptide. FLAG tag is a designer peptide selected for its hyrophilicity, which makes it readily accessible for detection/capture and is only 8 amino acids long. Elution, similar to the HA tag, is carried out by peptide competition.

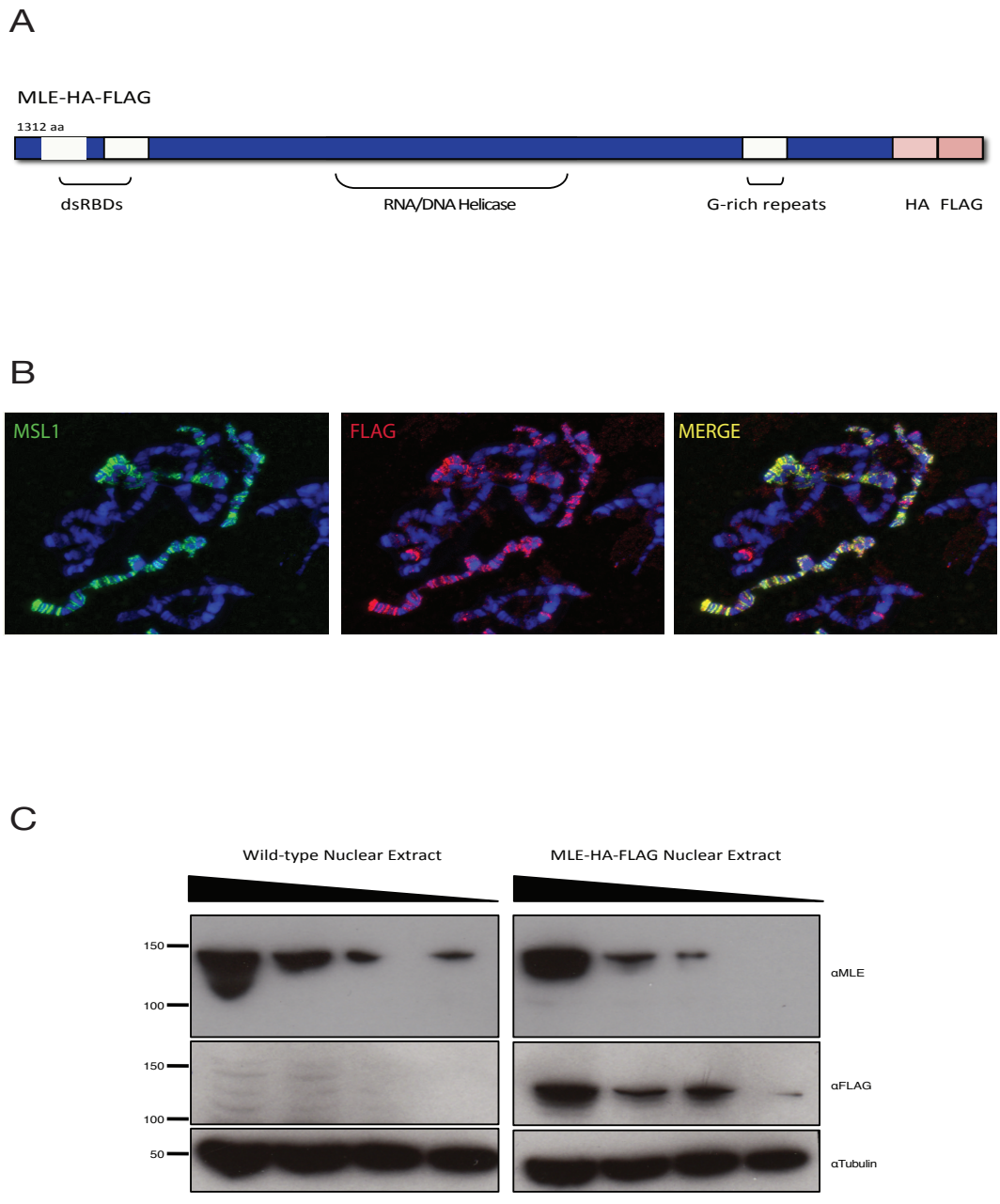

Figure 6 MLE tagged with HA-FLAG (A) MLE is tagged with an HA and a FLAG tag at its C-terminus. (B) Polytene chromosome spreads from 3rd instar male larvae carrying the HA-FLAG tagged MLE transgene is stained with antibodies against MSL1 (green) and FLAG (red). An extensive co-localization of the two proteins shows that the transgene is targeted to the X-chromosome like the wild-type protein. Also note the autosomal puffs stained with FLAG antibody. (C) 10, 20, 30 and 40 μg of nuclear extract prepared from wild-type or MLE-HA-FLAG expressing embryos is analyzed by western blotting. The transgene is expressed only in the transgenic flies, and is not over-expressed with respect to the endogenous MLE protein as judged by the MLE immunoblot.

I have tagged full length MLE with an HA (YPYDVPDYA) and a FLAG (DYKDDDDK) tag in tandem at its C-terminus (Figure 6A). This construct was inserted into the Drosophila genome by P-element mediated transformation. Several insertions were recovered. Before carrying out a purification, I wanted to make sure that the transgene I have introduced into the fly was functional similar to the wild-type protein. There were two ways to check this: First, the transgene should rescue the male specific lethality phenotype observed when the male fly lacks endogenous MLE protein. Second, the transgene should stain the X chromosome, co-localizing with other MSL proteins. Two of the lines I recovered had an insertion on the 3rd chromosome which makes it possible to complement mutant *mle* since the endogenous gene resides on the 2nd chromosome. These two lines were able to rescue the male specific lethality phenotype observed in *mle*^{1} mutants (45-50% of the flies were males in both lines when the only source of MLE protein was the transgene I introduced; there are normally no escaper males in *mle* mutants), so the first and the most important criterion was met; the transgene carries out its function as if it was the wild-type protein. I decided to continue with one of the lines and when polytene squashes from this line is immunostained with antibodies against the FLAG peptide and MSL1, an extensive co-staining could be observed on the male X chromosome, as can be seen in Figure 6B. Knowing that the protein is functional, and is targeted to the X chromosome like the wild-type protein, I expanded this line into population cages in order to get enough material for a tandem affinity purification. Figure 6C shows a representative nuclear extract prepared from wild-type embryos and from the embryos carrying the MLE-HA-FLAG allele on their 3rd chromosome. It is not possible to differentiate wild-type MLE with the affinity tag containing transgene due to the very subtle size difference imparted by the tag, but it is clear that the transgene is not over-expressed with respect to the wild-type MLE.

MLE-HA-FLAG and putative associated polypeptides were then immunoprecipitated out of the prepared nuclear extracts, first using FLAG M2 mAb, immobilized on agarose beads, followed by elution by FLAG peptide. This eluate was then loaded onto agarose beads coated with anti-HA mAb, and after several washes the immunoprecipitated material was eluted with 2xLaeammli buffer. The eluate was separated by PAGE, and the gel is silver stained to reveal the proteins (Figure 7). The background introduced by proteins that bind non-specifically to the beads was

controlled by a mock purification, carried out by using a nuclear extract prepared from wild-type embryos. Three most prominent bands were cut from the lane that contained the eluate coming from the purification carried out with the transgenic embryos. Surprisingly, all of these bands turned out to be MLE and its degradation products. Considering that MLE's association with the MSL complex is sensitive to high concentrations of salt, I have repeated the purification at a lower salt concentration (60mM NaCl, instead of 110mM). Apart from the increasing background the result was the same: there were no other proteins in the eluates, but MLE.

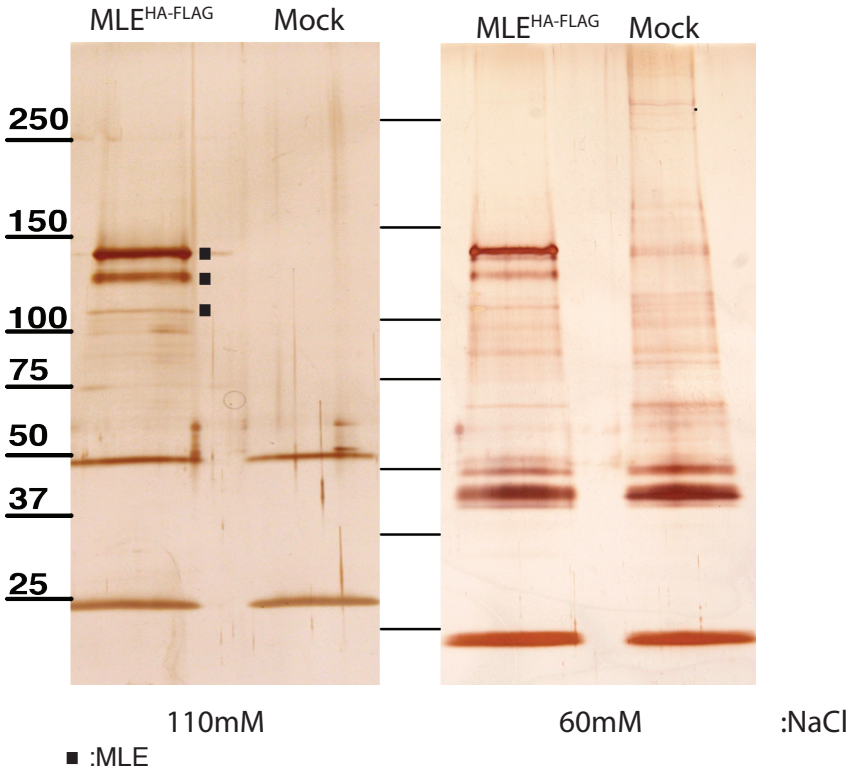

Figure 7 Tandem affinity purification of MLE Nuclear extracts prepared from 0-12 hrs embryos (extraction with 420mM NH_4SO_4, extract finally equilibriated to 100mM KCl) are used for the affinity purification of MLE. FLAG purification using anti-FLAG M2 sepharose beads (Sigma) is followed by HA purification using anti-HA agarose beads (Sigma). Protein bound to the anti-HA agarose beads is eluted with 2XLaemmli buffer, separated by PAGE and revealed by silver staining. Two separate purifications are carried, using two different salt concentrations that are indicated at the bottom of the gel.
Protein marker is indicated to the left. $MLE^{HA-FLAG}$: Nuclear extract made from transgenic embryos; Mock: Nuclear extract made from wild-type embryos

Substoichiometric interactions are difficult to catch during a tedious tandem affinity purification, yet these interactions can be important to understand the function of a protein. Since the MLE transgene could rescue male specific lethality, and thus we could essentially replace all cellular MLE by the MLE-HA-FLAG that we have introduced, I decided to expand this fly line, make nuclear extracts from its embryos and carry out a purification using that extract. In addition, the classical way to prepare a nuclear extract involves lysing the nuclei with 420mM ammonium sulfate and extracting nuclear proteins at this salt concentration, and the precipitating the extracted proteins by increasing the salt concentration up to 4M, after which the precipitated proteins are re-solubilized and dialyzed down to 100mM salt. It might be possible that, in such an extraction procedure, all MLE associated proteins fall off, and hence cannot be co-precipitated with MLE, because the salt concentration is elevated to 420mM, kept there for a while, and further increased to precipitate almost all the proteins in the solution. It was therefore essential to test different ways of extraction to see if we could purify possible MLE associated proteins.

In order to increase my chances to isolate polypeptides that interact with MLE, I have expanded the transgenic flies in which the only source of the transgene was the affinity tagged MLE that I had cloned. I have used two “milder” extraction methods to generate the material that was used for the affinity purification. Figure 8A shows two different extracts, and the efficiency of the FLAG pull-down in depleting MLE out of the extract, and the efficiency of elution by FLAG peptide competition as judged by the leftover MLE on the beads following elution. I have then proceeded with a small-scale purification using these extracts (Figure 8B). Although the background was high, due to the little amount of material I have used, it looks clear that there no major proteins bands in the lanes where I have immunoprecipitated MLE with respect to the mock purifications.

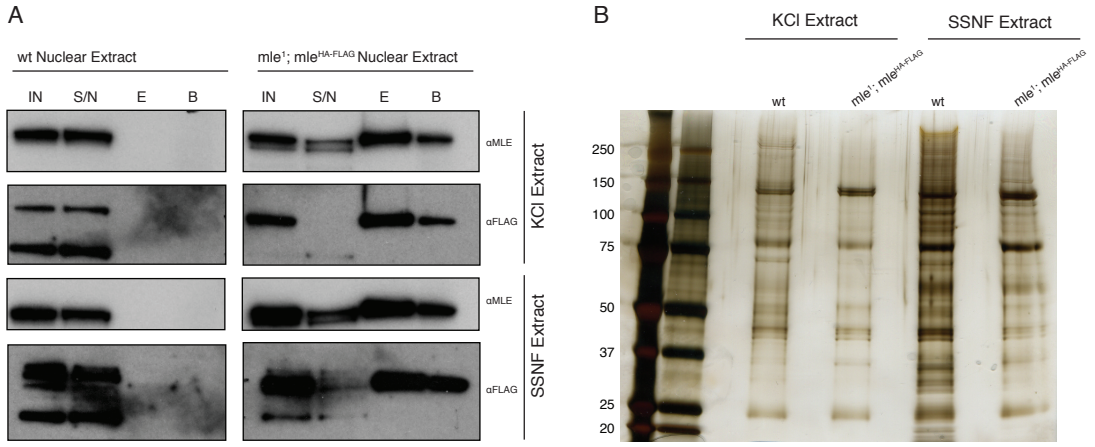

Figure 8 MLE purification in an *mle* mutant background (A) Nuclear extracts prepared from wild-type embryos and transgenic embyos which lacked endogenous MLE protein which was replaced by the MLE-HA-FLAG we introduced were immunoprecipitated with anti-FLAG beads. Two types of extracts were prepared: KCl and SSNF (procedure is explained in detail in Material and Methods section). (B) A small scale purification using KCl and SSNF extracts does not show proteins enriched in the experimental lanes (mle^{1}; $mle^{HA-FLAG}$) compared to mock lanes (wt). IN: Input, S/N: Supernatant, E: Elution, B: Beads boiled after FLAG elution

I have concluded, then, that if there are proteins that do interact with MLE, these interactions are very likely to be substoichiometric, possibly mediated by RNA (see Discussion) and very difficult to identify using a tandem affinity purification approach.

These results in themselves are quite intriguing as MLE is a classical dosage compensation complex member that co-localizes with the rest of the MSL complex on the X-chromosome. Yet this protein does not co-purify when other members of the MSL complex are tagged, neither does it co-purify MSL proteins when when they are epitope-tagged. This suggest that MLE's interaction with the X-chromosome is perhaps not mediated by interactions with other proteins but via, RNA with the most likely candidates being the roX RNAs.

3.2 Identification of novel roX interacting proteins

In the second part of my thesis, I changed gears and concentrated on studying the roX RNAs. The aim was to establish an *in vitro* system where we could detect the interaction between roX RNAs and members of the MSL complex and also to find additional polypeptides that specifically interact with roX RNAs. In our case, we know that the roX RNAs interact with the MSL complex ([103],[113] and [112]), and in the absence of both roX RNAs, the MSL complex cannot be targeted to the male X chromosome, dosage compensation fails and male flies die [101], [121]. However, exactly how roX RNAs are required for dosage compensation is currently not known. We reasoned that, if we can identify proteins that interact with roX RNAs, besides the MSL complex, we might get a better understanding of its function, and thus, we have utilized GRNA chromatography (a scheme of which is given in Figure 9) to identify novel proteins that interact with roX RNA *in vitro.* GRNA chromatography, developed in the Hentze [147] and Mattaj [148] laboratories have been used successfully to identify novel proteins associated with an RNA of interest. Since then, similar methods have also been used by other laboratories to the same purpose [149, 150]. For technical reasons , I have used roX2 RNA instead of roX1 in this study: long RNAs are more difficult to immobilize on beads, the probability of getting a correctly folded RNA *in vitro* gets lower and the background binding of non-specific proteins increases. Since roX RNAs are redundant in function, their RNP composition should also be similar. This assumption will be put to test in future studies.

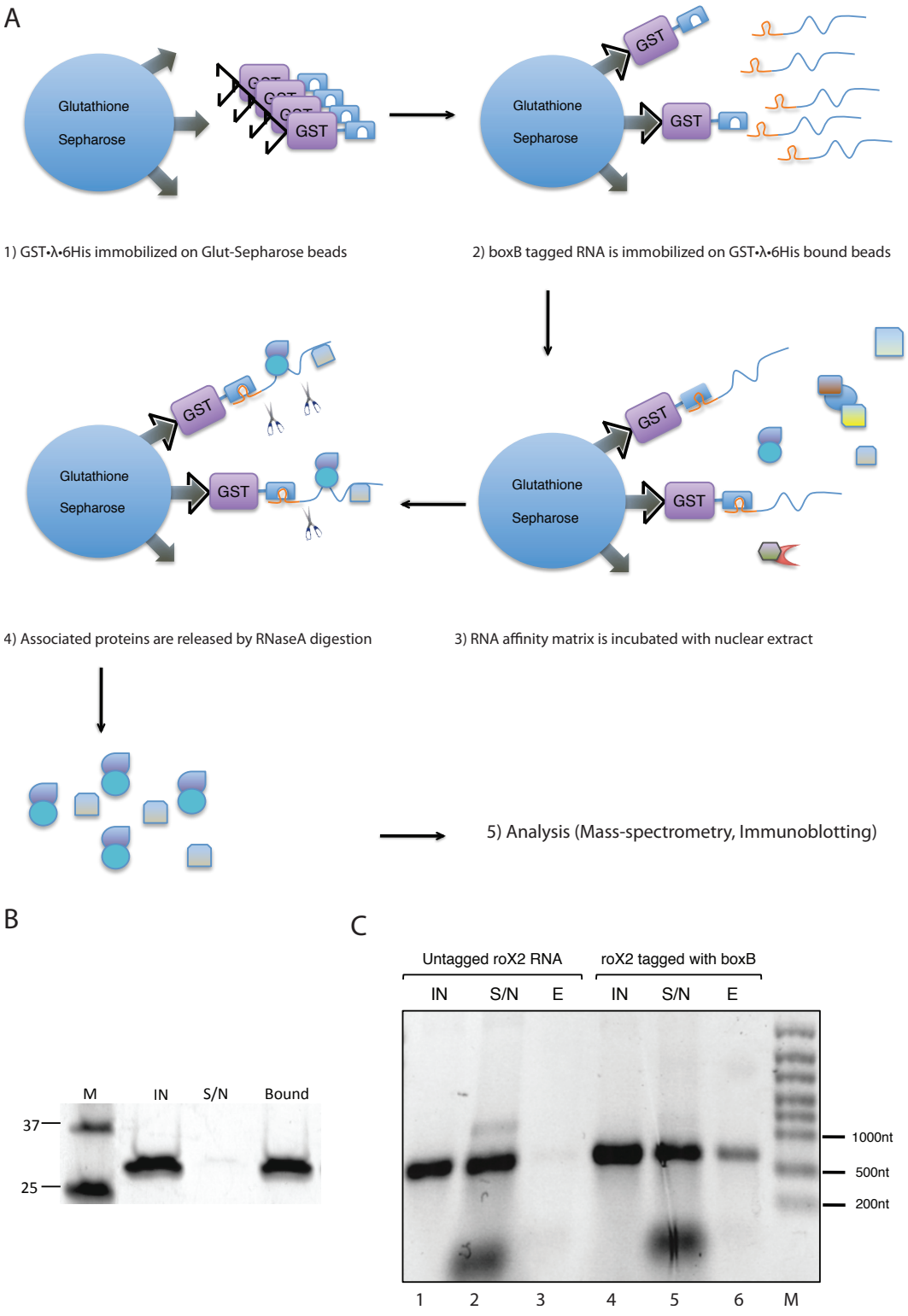

Figure 9 An overview of the GRNA purification.

(A) GST•λ•6His, a fusion protein that contains a GST moiety at its N-terminus is immobilized on sepharose beads coated with glutathione. An RNA of interest, tagged at its 5'-end with two boxB hairpins is incubated with beads that have GST•λ•6His bound to them. Due to the high affinity interaction between the λN22 peptide and boxB hairpin (Kd~ 20nM), the tagged RNA in immobilized on the beads in a directional manner. This affinity matrix is then incubated with a Drosophila embryonic

extract, after which the proteins that interact with the matrix is releases with RNase A digestion. The eluate is then analyzed either silver-staining or immunoblotting followed by PAGE. (B) Immobilization of GST•λ•6His on Glutathione-Sepharose beads is specific and efficient. M: Bio-Rad Precision Plus All Blue protein marker; IN: input, S/N: supernatant, Bound: Protein immobilized on the beads. (C) The interaction between λ_{N22} and boxB is highly specific. When untagged roX2 is incubated with beads that have GST•λ•6His immobilized on them, the RNA is not depleted from the supernatant (lane 2) and cannot be immobilized on beads (lane 3). When however, the same RNA is tagged with boxB, it is depleted from the supernatant (lane 5) and can be immobilized on GST•λ•6His bound beads (lane 6). M: Fermentas High-range RNA marker; IN: input, S/N: supernatant, E: Eluted RNA from the beads. The RNA is electrophoresed on a 1.2% agarose, 0.8M formaldehyde gel, after which it is revealed by SYBR Gold staining.

3.2.1 GRNA Chromatography

GRNA chromatography relies on the high affinity interaction between the λ_{N22} peptide and a 19-nt RNA hairpin called boxB. This peptide is tagged with a GST moiety at its N-terminus and six histidines at its C-terminus [148]. I have expressed GST•λ•6His fusion protein in *E. coli* BL21-CodonPlus cells and purified it using IMAC, yielding a highly pure protein as can be seen in Figure 10. This fusion protein is immobilized on Sepharose beads coated with glutathione. These beads are the incubated with *in vitro* transcribed RNAs tagged with boxB. Since the λ_{N22} binds specifically to the boxB hairpin, the RNA with the tag is immobilized on the beads in a directional, non-random manner. This interaction is highly specific, as you can see from Figure 9C that an RNA without the boxB hairpin does not bind to the column. This affinity matrix is then incubated with a nuclear extract prepared from 0-12hrs wild-type *Drosophila* embryos. Bound proteins are eluted with RNaseA, thus making sure that the eluate contains proteins that interact with the RNA not with the tube, Sepharose or GST•λ•6His non-specifically. A representative purification shows that the background coming from the beads bound to GST•λ•6His is very low and that almost all the bands that are visible comes from an interaction with the RNA on the column (Figure 11, lane 1).

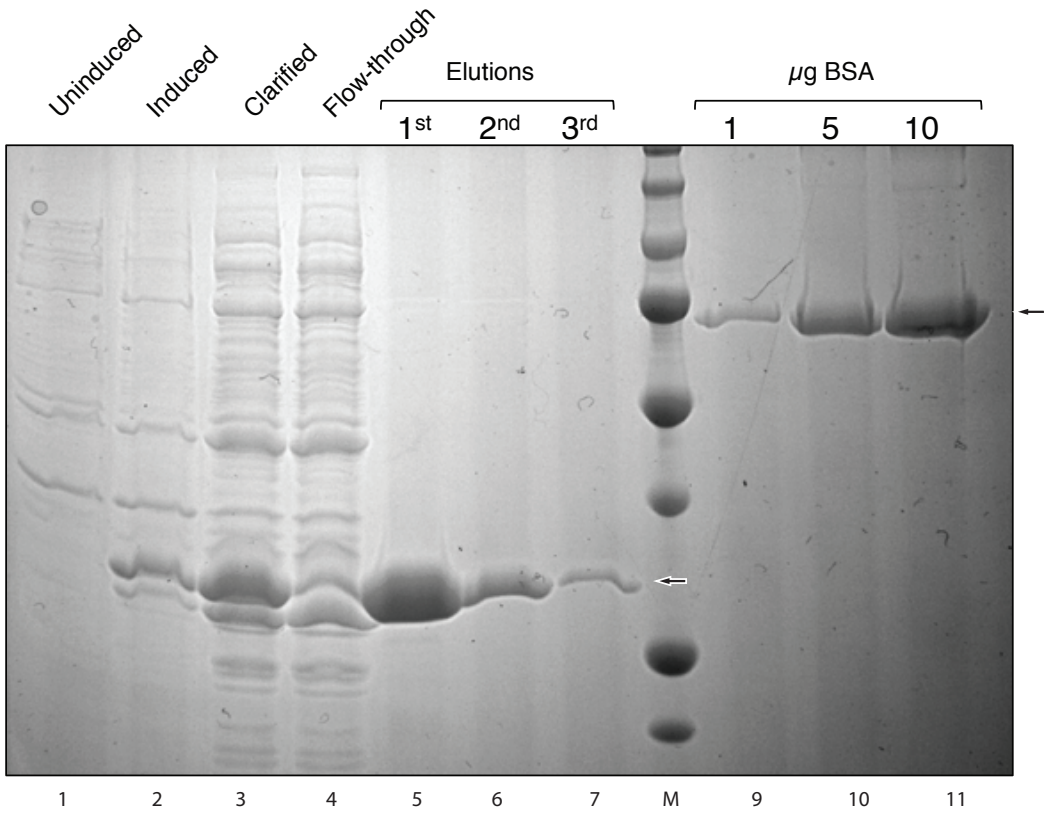

Figure 10 Expression and purification if GST•λ•6His GST•λ•6His was expressed in BL21 (DE3) CodonPlus (Novagen) cells and purified using Ni-NTA chromatography (Qiagen). The samples are loaded to a 12% polyacrylamide gel and stained with GelCode Blue. Arrowheads show GST•λ•6His and BSA.

Upon visual inspection of the silver-stained polyacrylamide gels, I have cut out and sent bands for mass-spectrometric analysis to reveal the proteins that were only present or highly enriched in the roX purifications with respect to GFP purifications. There were four proteins that consistently interacted with roX2 but not with GFP: MLE, CG5792, CG5787 and CG3613.

MLE was being pulled-down by roX RNA with very high efficiency thus showing that at least some, if not all aspects of roX RNA is present in the *in vitro* transcribed RNA. The interaction between roX2 and MLE seems to be fairly strong, as MLE was often depleted from the extracts that I have used during GRNA purifications. The efficiency of the MLE pull-down is very interesting indeed, since this protein is hardly ever detected when other MSL proteins are immunoprecipitated from cellular

extracts and is certainly not detected in previous purifications of the MSL complex in our laboratory [118]. This observation, coupled with the MLE purification that I have carried out (see the first part of the Results) suggests that MLE interacts mainly with RNA rather than other proteins and binds to roX2 with high affinity. The fact that I could not detect any other MSL proteins even after immunoblotting and that MLE interacts strongly with roX2 might be an indication of how the MSL complex assembles *in vivo*.

CG5787 is a 100kDa protein with a weak Nop17p homology domain close to its C-terminus and has a glycine/asparagine rich N-terminal domain. It was recently shown to co-purify with Zn72D, a zinc-finger protein that is required for the productive splicing of *mle* pre-mRNA [151].

CG3613 is a 45 kDa protein with a single KH domain in the middle and has a serine/tyrosine rich C-terminus. It is a member of RNA binding proteins that are known as STAR (Signal Transducer and activator of RNA) proteins. Prototypical STAR protein Sam68 is known to bind RNA when it is non-phosphorylated, and this RNA interaction does not take place when Sam68 is phosphorylated [152]. Sam68 also interacts with other nuclear factors, such as CBP and represses transcription; or with AR (Androgen Receptor) and activates transcription. These latter functions are independent of Sam68's ability to bind RNA and mediated by direct protein-protein interactions.

CG5792 has a complex gene structure where two almost complete separate polypeptides, and another polypeptide encompassing the two can be expressed. The small polypeptide (37kDa) that can be synthesized from this gene is the *Drosophila* homologue of yeast PIH1 protein [153]. In yeast cells, PIH1 was shown to be part of a complex (together with Hsp90, Tah1, Rvb1 and Rvb2) that is required for assembly and/or maintenance of box C/D snoRNP but does not interact with RNA directly [154].

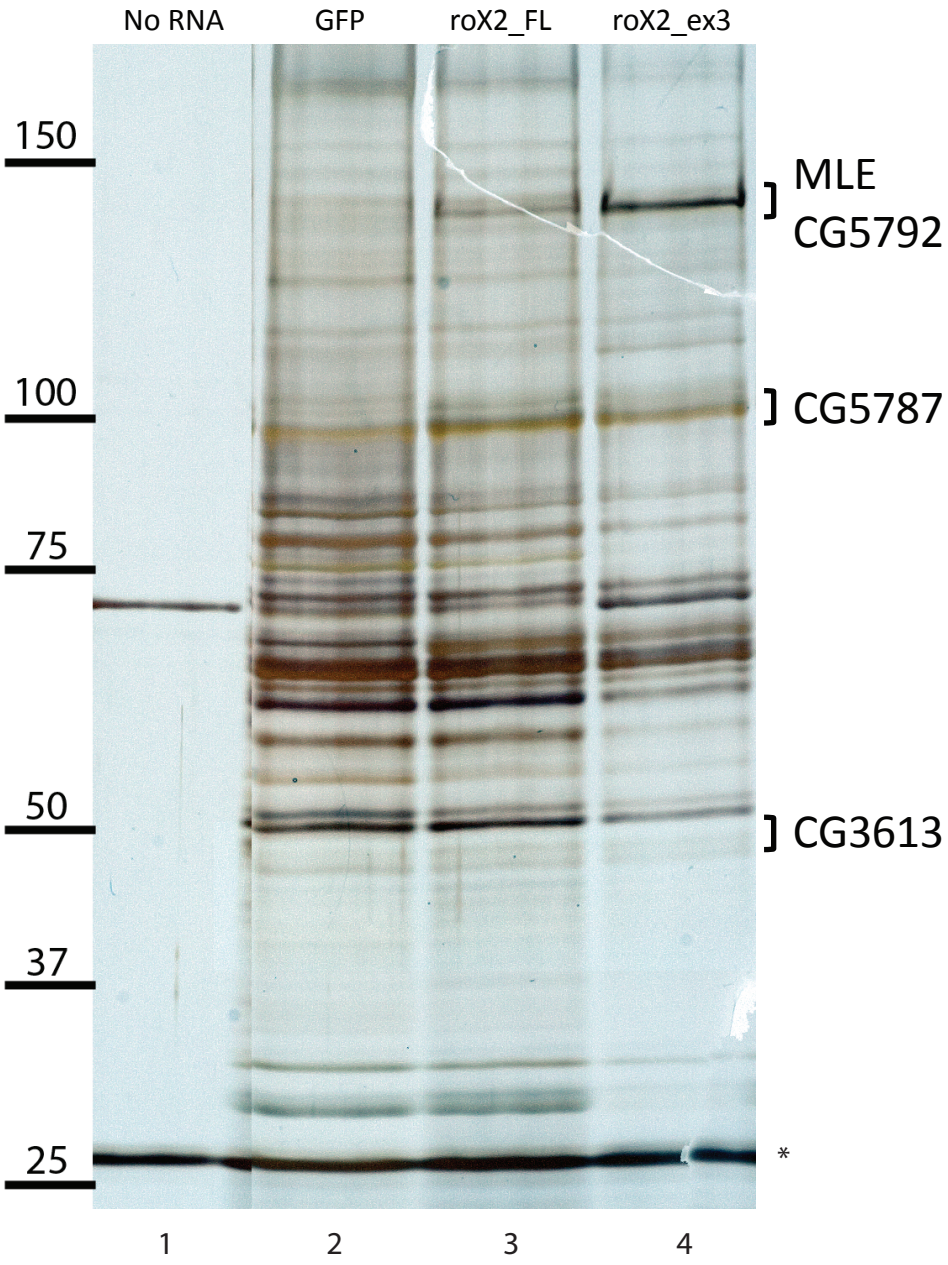

Figure 11 GRNA chromatography of roX2. At the end of GRNA chromatography, the proteins associated with the beads (no RNA, lane 1), GFP RNA (lane 2), full-length roX2 RNA (lane 3), and exon3 of roX2 RNA (lane 4) are eluted with RNaseA, run on a 4-20% polyacrylamide gel and silver-stained. After visual inspection of the gel, proteins that are enriched on roX2 lanes are cut and analyzed by mass spectrometry along with the corresponding bands from the GFP lane.

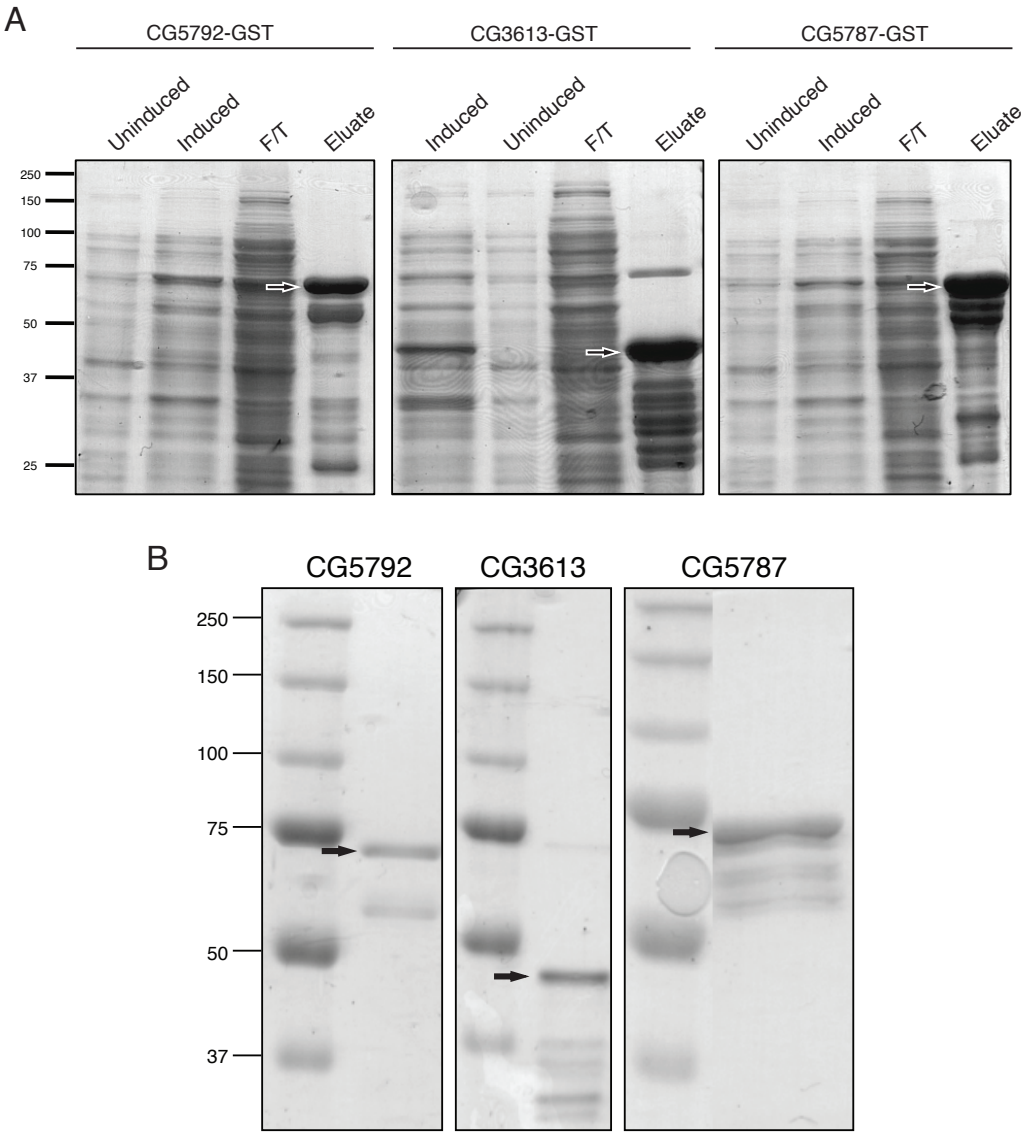

Figure 12 Expression and purification of GST fusion proteins for antibody production (A) Fragments of CG5792 (2-334 aa), CG5787 (168-571 aa) and CG3613 (220-395 aa) were cloned into pET41b vector and expressed in BL21 (DE3) cells. "Uninduced" lanes show total *E. coli* protein before the induction of the cells with 0.5mM IPTG, and "induced" lanes show afterwards. "F/T" lanes show the proteins that did not bind to the Glutathione Sepharose resin. The bound protein is eluted with 2X Laemmli buffer and can be seen in lanes "eluate". (B) The proteins that are injected; after purification over Glutathione Sepharose resin, dialysis and concentration. The expected sizes are 70 kDa for CG5792-GST fusion, 53 kDa for CG3613-GST fusion and 73 kDa for CG5787-GST fusion. CG579 was injected to 2 rabbits, CG3613 into 2 rabbits and CG5787 into 3 rats. Arrowheads indicate the respective protein bands.

3.2.1.1 Raising antibodies against CG5792, CG5787 and CG3613

I wanted to study the abovementioned proteins in more detail therefore I raised antibodies against all three novel proteins that were identified following GRNA chromatography. For this purpose, fragments of CG5792 (amino acids 2-334), CG5787 (amino acids 168-571) and CG3613 (amino acids 220-395) were cloned into pET41b, creating GST fusions. All three proteins were expressed in *E. coli* and purified over Glutathione Sepharose beads (Figure 12). As predicted *in silico,* anti-sera against CG5787 recognized a band at 100kDa, anti-sera against CG3613 recognized a band at 45kDa and anti-sera against CG5792 recognized a band at 37kDa. These bands were absent when the membranes were blotted with the corresponding pre-immune sera. To further validate the specificity of the antibodies, I knocked-down these proteins in S2 cells, and checked if the bands revealed by these anti-sera could be diminished upon reducing the protein from the cells. The result, shown in Figure 13, proves that all the anti-sera specifically recognize the proteins that they were raised against.

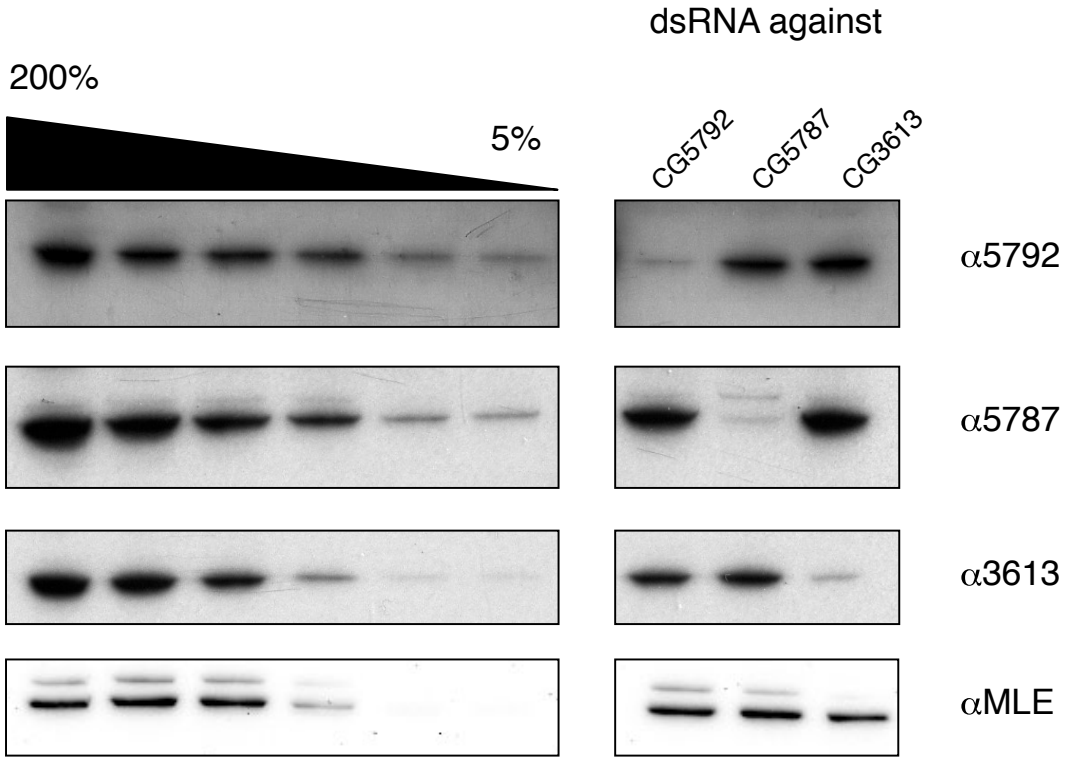

Figure 13 Antibodies against CG5792, CG5787 and CG3613 are specific Drosophila S2 cells are treated with dsRNA against CG5792, CG5787 and CG3613. Three days after incubation with dsRNA, the cells are lysed with 2XLaemmli buffer and analyzed by immunoblotting following PAGE.

3.2.2 Confirming the GRNA results

3.2.2.1 Antibodies

I have repeated the GRNA chromatography, on a smaller scale, to see if MLE, CG3613 and CG5787 interacted with roX2 RNA specifically. This time I have separated the eluates by PAGE and transferred the proteins to a PVDF membrane after which I revealed proteins by immunoblotting using the antibodies that I have generated (Figure 14). In this experiment antisense roX2 RNA joined GFP RNA as a negative control. The result was clear; roX2 interacts with MLE, CG5787 and CG3613, much more efficiently than anti-sense roX2 RNA, or GFP RNA (Figure 14), hence confirming earlier GRNA experiments. Since western-blotting is a semi-quantitative method, I wanted to know if the RNAs bound to the beads at the end of the experiment were comparable to each other. One could get a higher amount of binding of a certain protein to certain RNA, not because the affinity of that protein is appreciably higher towards the RNA of interest rather than the negative control, but because the RNA of interest is simply bound more to the beads. To rule this out, I split the sample into two before RNase A elution, and eluted the RNA from the beads using RNeasy columns (Qiagen) and visualized it by Sybr gold staining after Formaldehye-agarose gel electrophoresis.

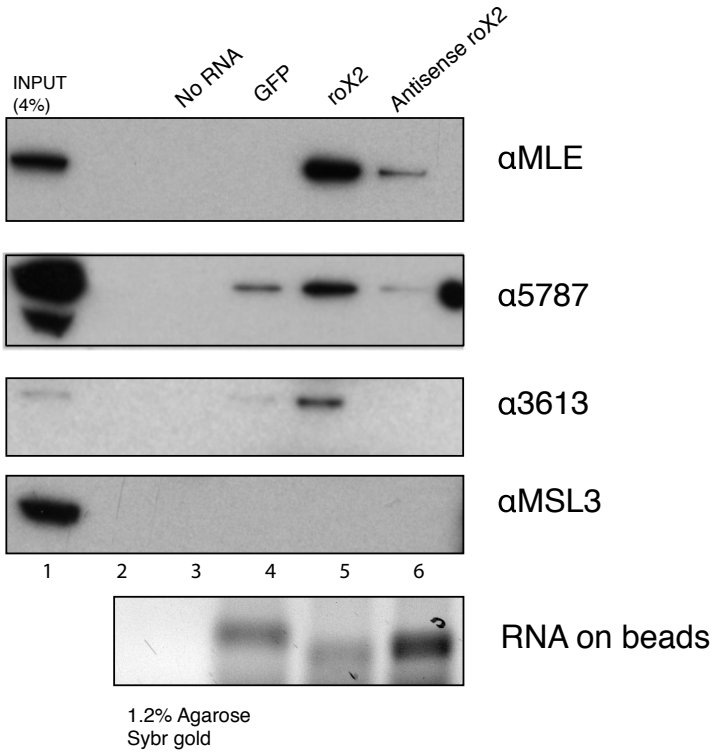

Figure 14 GRNA chromatography on a smaller scale. GRNA chromatography was carried out as in Figure B3. Before the elution, beads were divided into two and RNA was extracted from one half which separated by denaturing agarose electrophoresis and stained with Sybr gold. RNase A eluates were separated by PAGE and immunoblotted.

3.2.2.2 PP7 pull-down

Although the results from the GRNA chromatography were convincing, I have decided to change the affinity tag on the RNA to exclude any effect the affinity tag might have on the result of the experiment. To do that, I have used another viral protein: PP7 coat protein (PP7Cp) and its cognate RNA hairpin ([67]).

I tagged GFP and roX2 RNA with a single PP7 hairpin at their 5' end. The only difference with the GRNA chromatography was the use of Sepharose beads coated with rabbit IgG instead of Glutathione-Sepharose beads to immobilize the viral peptide since this protein is a fusion protein that contains two Protein A moieties at its N-terminus attached to the PP7Cp. Similar to GST•λ•6His purification I have utilized IMAC (Ni-NTA, Qiagen) for the purification of the PP7Cp (Figure 15A, the IN lane shows the purity of the protein).

The result was the same; MLE, CG3613 and CG5787 interact preferentially with roX2 RNA as compared to GFP RNA (Figure 15B); hence these interactions are independent of the affinity pairs used to immobilize the RNA of interest on beads.

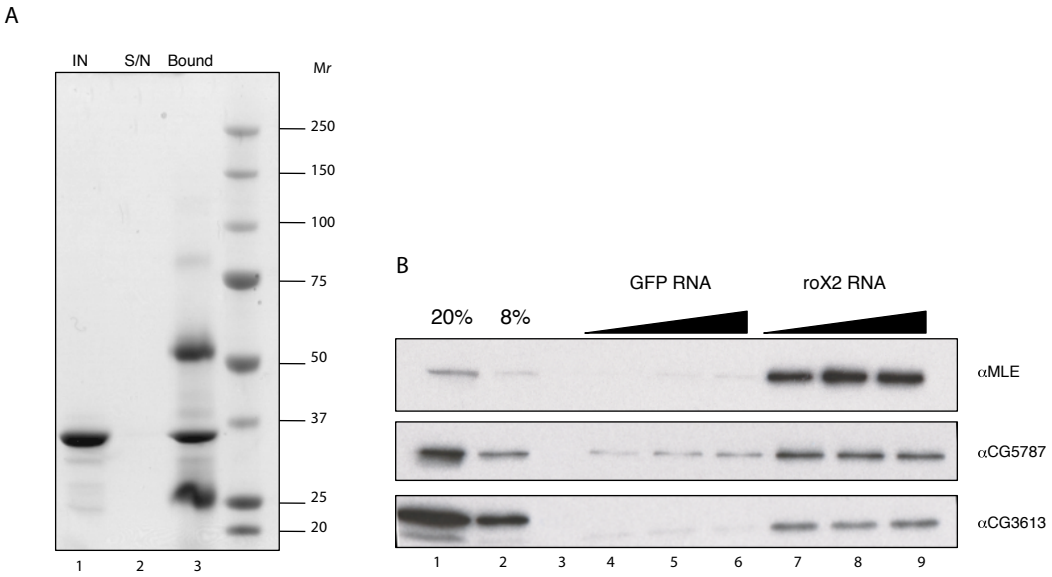

Figure 15 PP7 pull-down gives identical results with GRNA chromatography (A) PP7Cp purified using IMAC (Ni-NTA, Qiagen) is free of contaminants as seen in lane IN: Input. This protein is immobilized on IgG conjugated sepharose beads (see text for details) which is very efficient as judged by the depletion of supernatant (lane S/N) and elution by boiling in 2XLaemmli (lane Bound). (B) GFP RNA and roX2 RNA are used in this pull-down, in increasing amount to see if the system is saturated for RNA (GFP: 2.7*μ*g, 5.4*μ*g, 10.8*μ*g; roX2: 3.1*μ*g, 6.2*μ*g, 12.4*μ*g; these amounts correspond to equimolar loading). As in GRNA chromatography, roX2 associates with MLE, CG5787 and CG3613 more than GFP does.

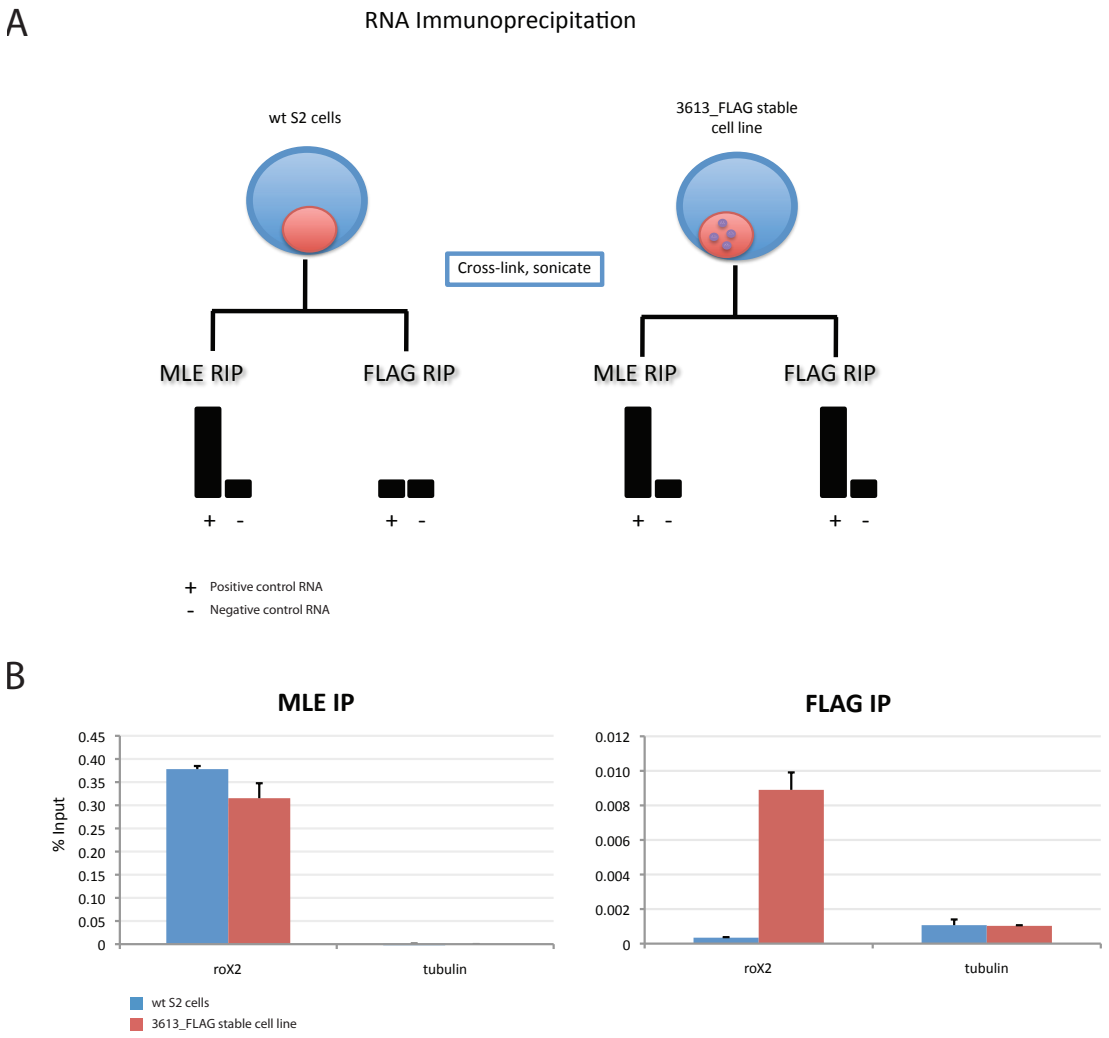

Figure 16 CG3613 interacts with the roX2 RNA *in vivo*
(A) The scheme of RNA Immunoprecipitation (RIP) is presented. A stable cell line expressing 3XFLAG tagged CG3613 is used as a positive control, whereas wild-type cells are used to determine the background of RNA binding to the FLAG antibody and to the beads. (B) The actual experiment, RNA from the RIP experiment is reverse-transcribed and analyzed by quantitative PCR.

3.2.2.3 CG3613 interacts with roX2 RNA *in vivo*

I have isolated CG3613 as a roX2 interacting protein using an *in vitro* experiment. Interactions such as these might be spurious and may occur only under the special conditions set-up for the *in vitro* RNA pull-down experiment. In order to test this interaction whether this interaction also occurs *in vivo*, I have created a stable cell line, which expresses full-length CG3613 tagged with the 3xFLAG tag at its C-terminus. I have carried out an RNA Immunoprecipitation experiment with the FLAG

M2 mAb to pull-down CG3613 and checked if roX2 RNA co-precipitates with it (an outline of the RNA Immunoprecipitation experiment is presented in Figure 16A).

In this experiment, wild-type S2 cells were used as a negative control, since S2 cells without any transgene would lack an antigen that would be recognized by the FLAG antibody, thus, one would expect not to pull-down any roX2 RNA. MLE anti-sera, on the other hand, was used as a positive control, with the expectation that it would pull-down the same amount of roX2, but not an unrelated RNA such as tubulin, from the transgenic cell line containing tagged CG3613 and from wild-type S2 cells without any transgene. In MLE immunoprecipitations, roX2 RNA was pulled down robustly in both cell lines, whereas tubulin RNA was not. FLAG immunoprecipitations did not show any enrichment neither for roX2 nor tubulin RNA when the extract was made from wild-type S2 cells, interestingly, significant amounts of roX2 was pulled-down when the extract was from the stable cell line expressing FLAG tagged 3613, whereas co-precipitated tubulin RNA stayed at background levels (Figure 16B). This experiment shows that CG3613 protein interacts with roX2 RNA *in vivo* as well as *in vitro* and complements the results of the GRNA chromatography.

3.2.3 CG3613, CG5787 and MLE co-localize on polytene chromosomes

Both CG3613 and CG5787 were purified using roX2 as a bait. MLE, a well-known MSL protein was also purified along with these novel factors in these purifications. I have next used polytene squashing of salivary glands of third-instar larvae, followed by fluorescent immunostaining of the squashed chromosomes to see if there is any co-localization between these factor, given that they were all identified as roX2 interacting proteins.

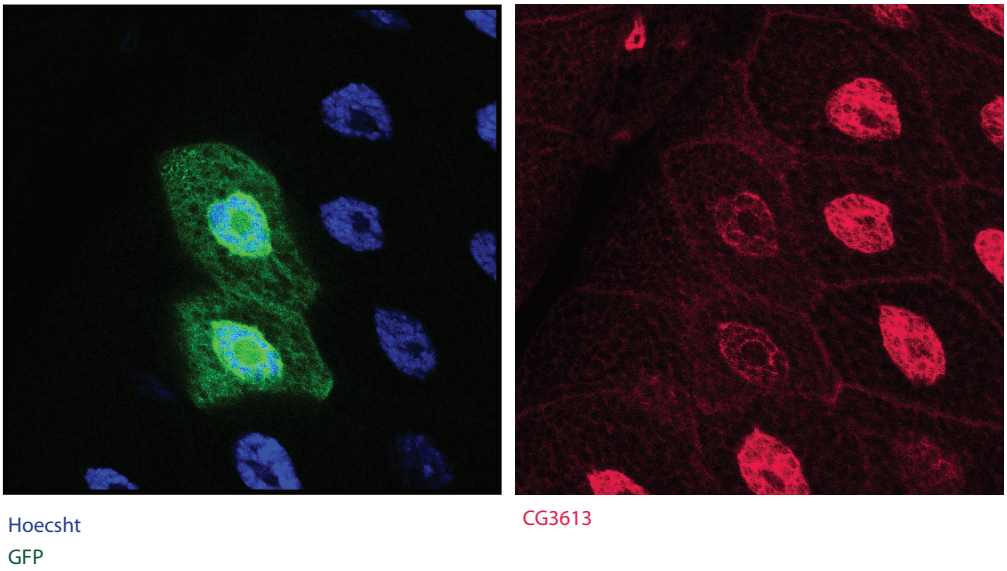

Figure 17 Whole-mount salivary gland staining of CG3613 8-10 hour embryos with the genotype *P*{hsFLP}; *P*{Act<FRT> y^+ <FRT>GAL4} *P*{UAS.GFP} / UAS.dsRNACG3613 are heat shocked at 37°C for 1hr. This creates patches of cells within a tissue that knock-down CG3613 and express GFP concomitantly hence marking these cells. The depletion of signal coming from CG3613 staining (red) in green cells show that the antibody specifically recognizes CG3613 and corroborated biochemical data that shows that CG3613 is nuclear.

In order to test whether CG3613 antibody recognizes the correct protein in fixed tissue, I have generated a fly which expresses patches of cells within a tissue, where dsRNA against CG3613 is expressed along with GFP to mark these cells. In this way, one can create cells that have reduced amounts of a protein, CG3613 in this case, surrounded by wild-type cells. Such a salivary gland is shown in Figure Figure 17. As can be seen, the CG3613 anti-sera recognize CG3613 specifically, as the nuclear staining is diminished only in the cells that are knocked-down for CG3613 ("green cells"). Thus, this anti-sera appears to be suitable for immunostaining of salivary glands fixed with formaldehyde.

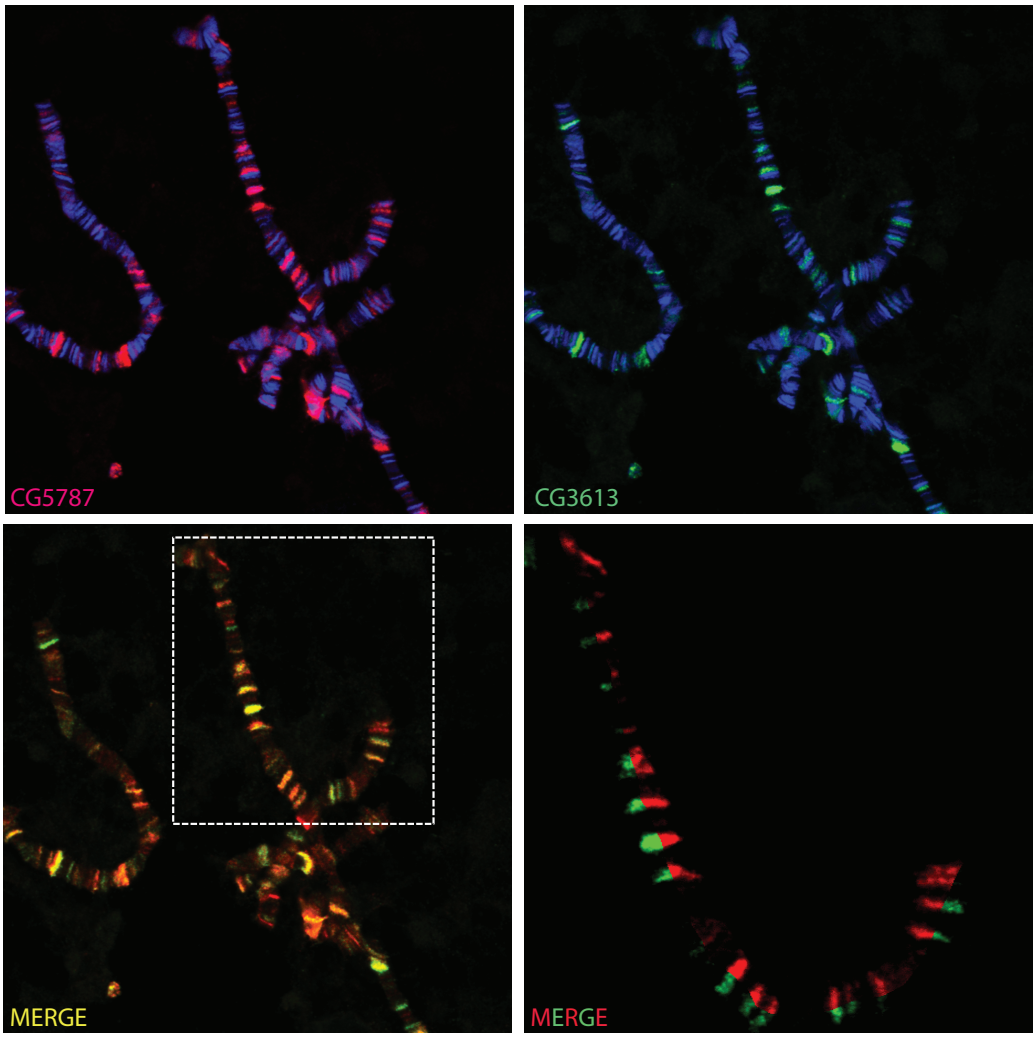

Figure 18 CG3613 and CG5787 co-localize on inter-bands Polytene squashes made from 3rd instar larvae are stained with antibodies against CG3613 and CG5787 and fluorescently labeled secondary antibodies against CG3613 and CG5787 that do not cross-react. DNA is stained with Hoechst 33258.

Intriguingly, a co-staining of a polytene squash with anti-sera against CG3613 and CG5787 reveals that these proteins co-localize with each other on many inter-bands and puffs (Figure 18). Inter-bands are considered to be sites of active transcription [155], and this type of staining suggests that these two proteins function on a similar set of genes and are involved in a process related to transcription.

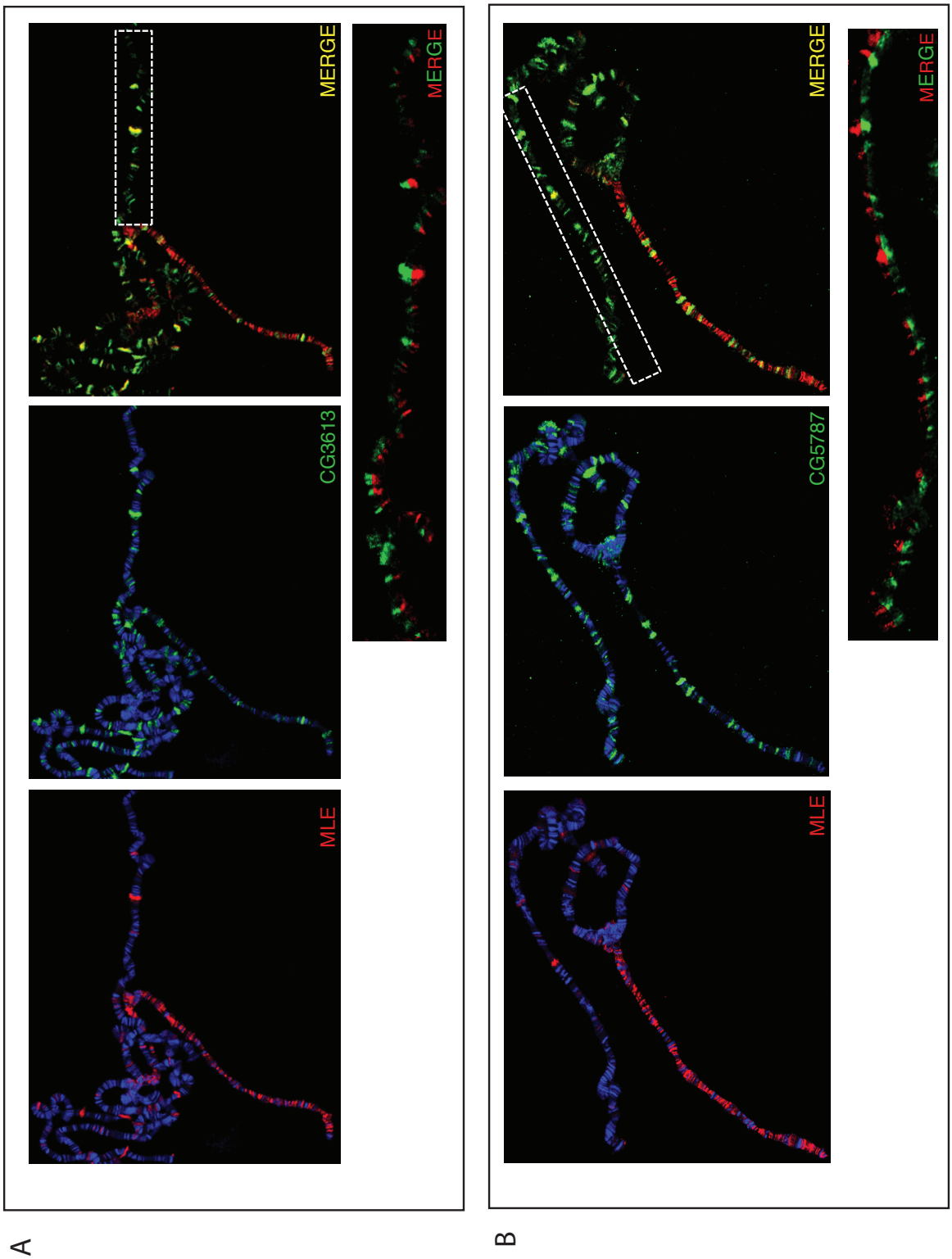

Figure 19 CG3613 and CG5787 co-localize with MLE on autosomal inter-bands. Polytene squashes made from male 3rd instar larvae are stained with antibodies against CG3613 and MLE or with CG5787 and MLE and fluorescently labeled secondary antibodies against do not cross-react. MLE localizes to the X chromosome in both cases, as expected, and the well-known autosomal sites of MLE binding seem to contain CG3613 (A) and CG5787 (B). DNA is stained with Hoechst 33258

All the MSL proteins stain the X chromosome in a very particular way in polytene squashes. MLE, in addition to its X-chromosome staining, has additional autosomal loci to which it binds to (Figure 5), and it was very interesting to see whether CG3613 and CG5787 co-localize with MLE either on the X chromosome or on the autosomes. Co-staining of polytene squashes of male third instar larvae with antibodies against CG5787 and MLE or against CG3613 and MLE was carried out (Figure 19). MLE, as expected, decorated the X-chromosome and some other loci on the autosomes. Interestingly, some of the densely stained autosomal sites were also occupied by both CG5787 and CG3613. When the intensity of autosomal staining was increased for MLE (since the X-staining is very bright, it may obscure less intense autosomal bands), it seems like many inter-bands that are stained with MLE are also co-stained with CG5787 and CG3613 (see inlets in Figure 19).

3.2.4 Cellular localization

GRNA chromatography was carried out using a nuclear extract. Polytene squash analysis showed that CG5787, CG3613 and MLE co-localize on chromatin. I wanted to see if these proteins were indeed nuclear, or if they also have a cytoplasmic pool. A fractionation experiment using S2 cells revealed that CG3613 and CG5787 were predominantly nuclear, whereas, to our surprise, CG5792 was almost completely cytoplasmic (Figure 20). The gene structure of CG5792 can shed light this unexpected outcome. There are, as it were, two different polypeptides produced from the CG5792 locus, and these two polypeptides are totally different from each other except for a very short common N-terminus. One of these polypeptides is 37kDa, and the other is ~100kDa. In GRNA purification, CG5792 was detected at ~150kDa. This form of the protein is no longer annotated in FlyBase, but it can be transcribed from the CG5792 locus and it is more or less the fusion of the two polypeptides

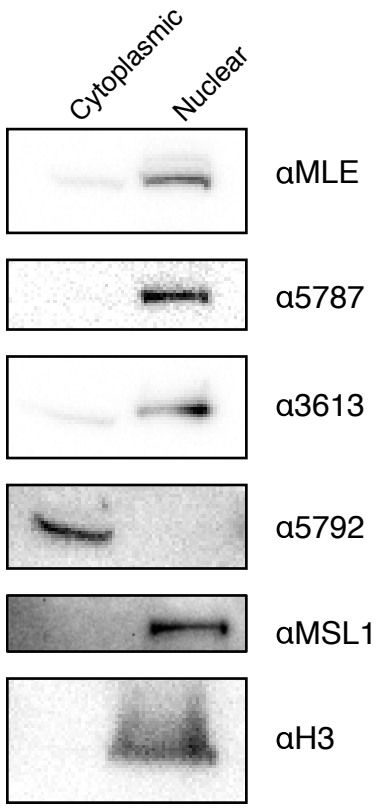

Figure 20 Cellular localization of CG5792, CG5787 and CG3613 Drosophila S2 cells are lysed in a hypotonic lysis buffer to separate nuclei from the cytoplasm (see Methods for details). The nuclear fraction and the cytoplasmic fraction are then analyzed by immunoblotting following PAGE

mentioned above. This large protein would contain the 37kDa fragment against which I raised the anti-sera, it should have been possible to detect it. However, I cannot distinguish this polypeptide in western blots from the background, perhaps due to its scantiness of this form of the protein in the cell. For this reason I have not pursued the characterization of this protein any further.

3.2.5 Nature of the chromatin association

A simple fractionation experiment that separated nucleus from the cytoplasm shows that CG3613 and CG5787 are nuclear proteins (Figure 20). A protein can be nuclear, yet it could spend its time in the nucleoplasm, or on the chromatin or both. In order to reveal where CG3613 and CG5787 reside in the nucleus, I have isolated nuclei from S2 cells, and extracted the soluble proteins at physiological salt concentration of 140mM, and at higher salt concentrations (240mM, 340mM and 420mM). I call the soluble protein as the "nucleoplasmic fraction" (np), and the protein in the pellet as "chromatin bound" (ch). At physiological salt, CG3613 and CG5787 seem to be mostly chromatin bound (lane 2), with very little protein in the nucleoplasm (lane 1). As the amount of salt is increased, both CG3613 and CG5787 are rapidly extracted from the chromatin; and almost all is extracted into the nucleoplasm when the salt concentration reaches 340mM (lanes 5-6). In contrast, members of the MSL complex (MSL1, MSL3, MOF and MLE) are extracted gradually, and a rather significant amount is left at the chromatin even at 420mM sodium chloride.

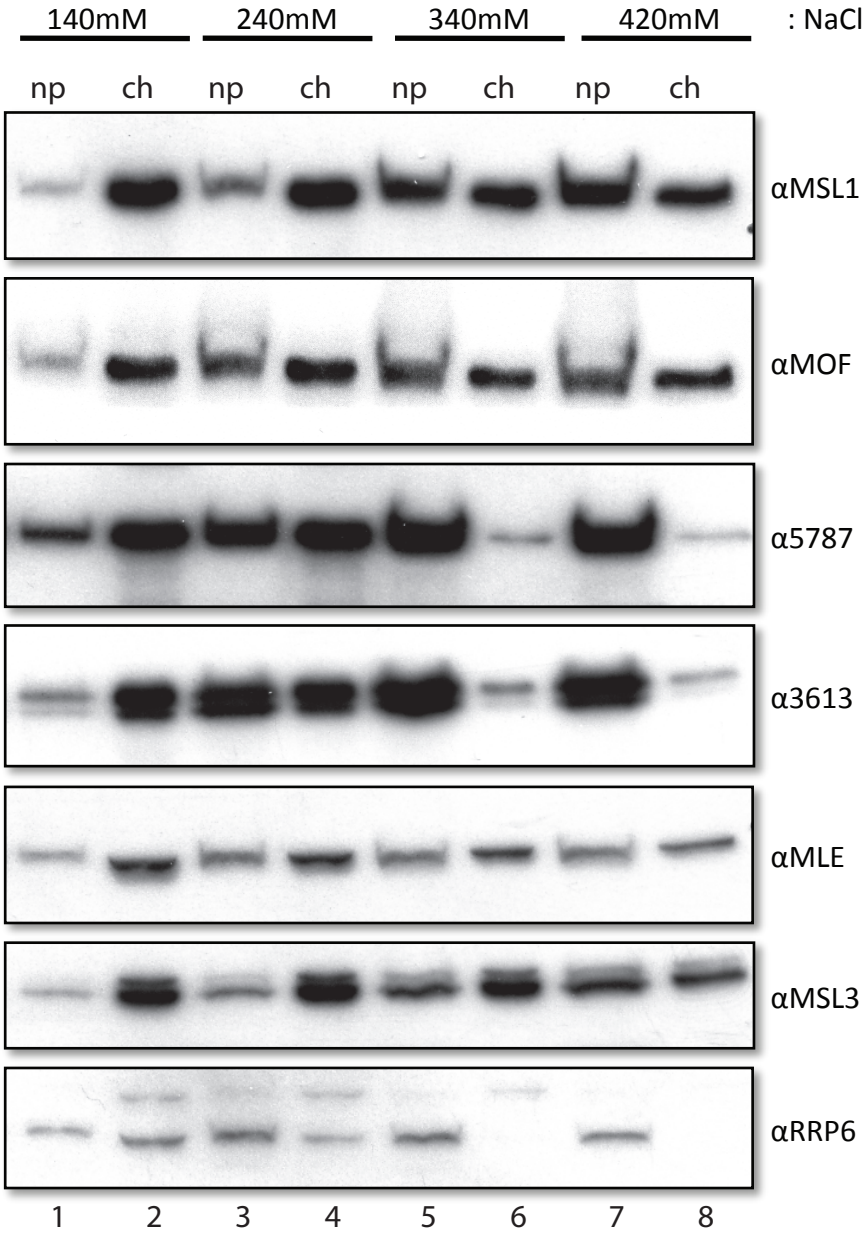

Figure 20 CG3613 and CG5787 are chromatin bound but relatively sensitive to high salt concentrations In order to assess the localization of CG5788 and CG3613 in the nucleus, and to check their reaction to increasing salt, nuclei were isolated from Drosophila S2 cells by hypotonic lysis with 1% NP-40. After washing with lysis buffer without detergent, nuclei are re-suspended in buffers with different amounts of NaCl in them (140mM, 240mM, 340mM and 420mM). After 20 minutes, the nuclei are pelleted and supernatant is collected as "nucleoplasmic fraction" (np) whereas the insoluble proteins in the nuclei (which are solubilized in 2XLaemmli buffer). as "chromatin bound" (ch). All the fractions are then run on a polyacrylamide gel, transferred to a PVDF membrane and analyzed by immunoblotting.

3.2.5.1 Sensitivity to RNase treatment

Some proteins associate with the chromatin via an RNA bridge, one the most notable examples being the MLE protein [116]. MSL3 and MOF also show some RNAse sensitivity when it comes to their chromatin binding, indicating that at least some of their interactions with the chromatin may be bridged by RNA [112].

Since I have isolated CG5787 and CG3613 via roX2 RNA, and I have shown that they are chromatin bound proteins, the next logical step was to check whether this association with the chromatin is mediated by RNA. Nuclei isolated from S2 cells, then, are treated with RNAse A at physiological salt, or not; and as before, soluble protein is called as the "nucleoplasmic fraction" and what is left behind in the nucleus as "chromatin bound" (Figure 21). As expected, MLE is extracted from the chromatin almost quantitatively with RNase A treatment (compare lanes 1-2 and 3-4). This result, together with the result of my affinity purification can be consistent with an MLE protein that works on RNA in the nucleus by itself with very little or no direct interaction with other polypeptides. The MSL proteins, apart from MLE, seem relatively unperturbed with the addition of RNase A. Most importantly the bulk of CG5787 and CG3613 interact with the chromatin in an RNA-independent manner. Of particular interest is the doublet of CG3613 clearly visible in the nucleoplasm. Upon

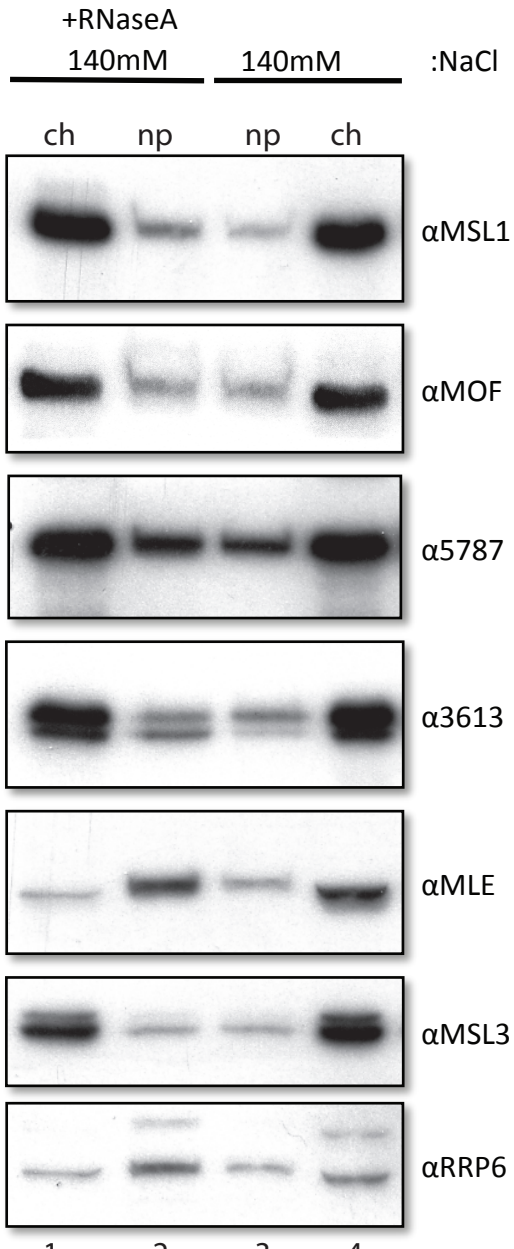

Figure 21 CG3613 and CG5787 interact with the chromatin in an RNase-insensitive manner As proteins that are potentially RNA-binding, it was necessary to check if the chromatin association of CG3613 and CG5787 were bridged by RNA molecules. As in Figure 20, isolated nuclei is re-suspended in a buffer with 140mM NaCl, but treated with 1μg/μL RNaseA in one sample. The resulting fractions are analyzed by immunoblotting. MLE, as expected, is almost completely detached from the chromatin upon RNase treatment, whereas CG3613 and CG5787 were mostly unaffected. np: "**n**ucleo**p**lasmic", ch: "**ch**romatin bound".

careful examination one can realize that the fast migrating CG3613 protein seems to be released into the nucleoplasm upon RNase A treatment (lane 2 vs lane 3), whereas the slow migrating CG3613 protein looks undisturbed by the presence of RNase A.

3.2.6 CG3613 is a phosphoprotein

As discussed before, CG3613 belongs to a family of proteins called STAR proteins that are involved in relaying extracellular signals to the nucleus and modulate transcription and related processes such as splicing and RNA export. Sam68, a hallmark STAR protein, binds to RNA when non-phosphorylated, and this interaction is inhibited when it is phosphorylated [156]. CG3613 contains many serine and tyrosine residues at its C-terminus that may be phosphorylated *in vivo*. Prompted by the doublet seen on Figure 21 we wanted to see if the slower-migrating form of CG3613 represents a phosphorylated form of the protein. In order to look into that, I have made a fractionation experiment, and treated each fraction (cytoplasmic, nucleoplasmic and nuclear) with λ-phosphatase that removes all phosphate groups from serine, tyrosine and threonine residues [157]. The experiment showed clearly that CG3613 is phosphorylated; very likely on multiple residues judging by the mobility shift upon dephosphorylation and that the antibody that I have generated recognizes both forms (Figure 22). Interestingly, this phosphorylated form seems to be mostly on the chromatin associated fraction.

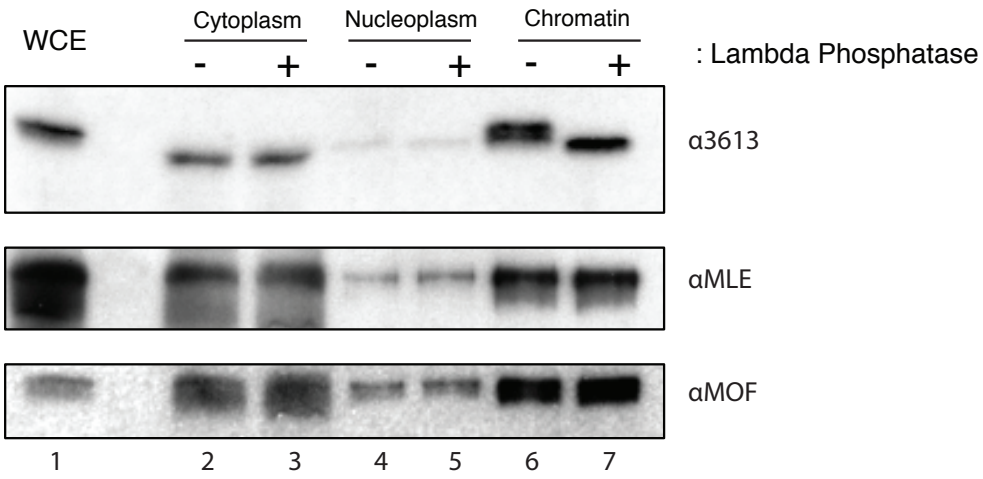

Figure 22 CG3613 is a phosphoprotein As a STAR protein, and with a Y/S rich C-terminus, CG3613 is likely to be a phosphoprotein. Cytoplasmic, nucleoplasmic and nuclear fractions from S2 cells are treated with λ-Phosphatese (4u/μL, NEB) and then analyzed by immunoblotting. Chromatin bound CG3613 seems to have a phosphorylated fraction, that can be converted into the non-phosphorylated form by λ-Phosphatese treatment. WCE: Whole cell extract.

Knowing that CG3613 can be phosphorylated *in vivo*, and taking the data from Sam68's RNA binding properties into account, one might argue that the RNase-sensitive CG3613 seen on Figure 21 is probably a pool of CG3613 that is hypophosphorylated or non-phosphorylated and that interacts with the chromatin via RNA. How phosphorylation modulates CG3613 function, and how is the phosphorylation itself is regulated remains to be investigated.

3.2.7 CG3613 and CG5787 associate with transcriptionally active loci

Previous polytene squashes showed that CG3613 and CG5787 co-localize on puffs, which are usually loci of active transcription. In order to check if CG5787 and CG3613 are associated with transcriptionally active loci, I co-stained squashed polytene chromosomes with CG5787 or CG3613 and a monoclonal antibody against

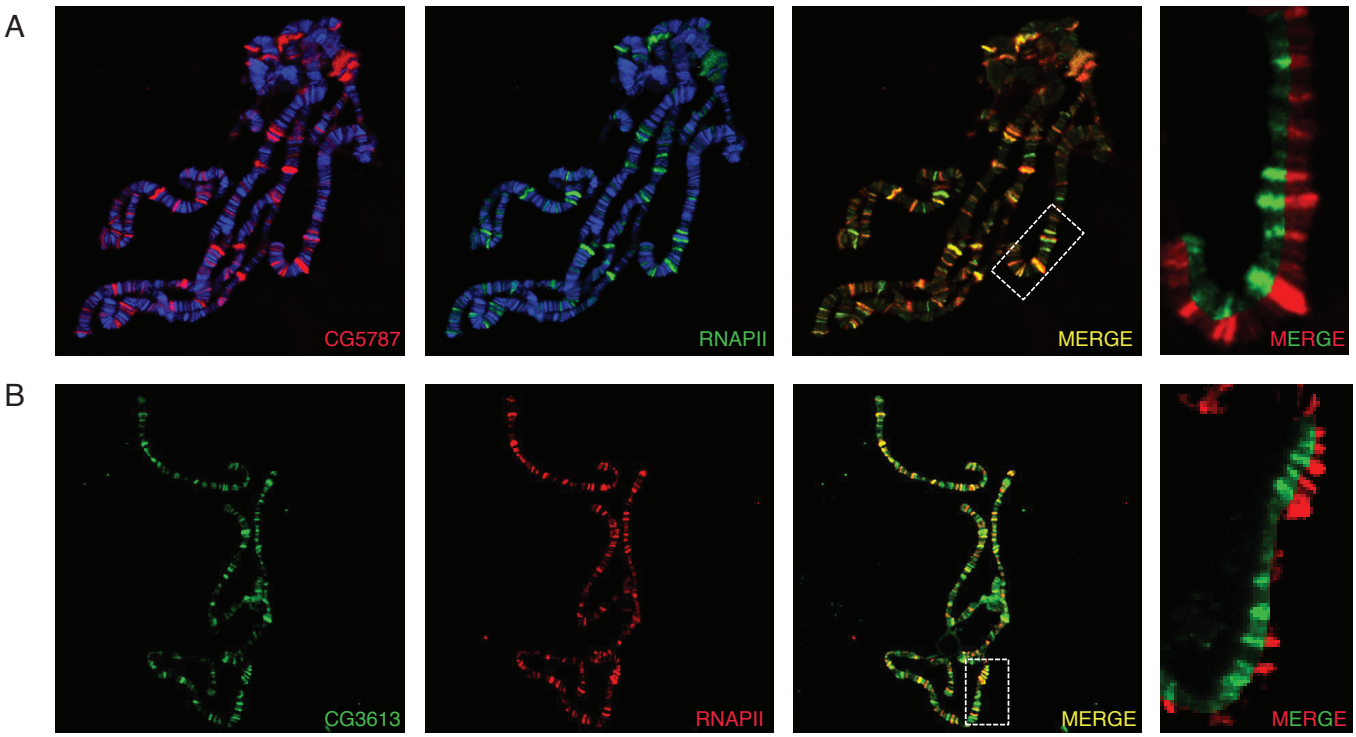

Figure 23 CG3613 and CG5787 co-localize with the active form of RNA polymerase II To verify that the inter-bands bound by CG3613 and CG5787 represent transcriptionally active loci, co-staining of polytene squashes are carried out with the mAb against Serine-5 phosphorylated RNAPII CTD (H14, Covance) and with antibodies against CG5787 (A) or CG3613 (B). DNA is stained with Hoechst 33258 in (A).

serine-5 phosphorylated form of RNA polymerase II CTD. The staining shows that almost at each locus with which CG5787 or CG3613 is detected, active form of RNA polymerase II can also be detected, indicating that CG3613 and CG5787 associate with actively transcribed regions (Figure 23).

3.2.8 CG3613 cross-links to MSL rich regions on the X chromosome

Polytene squash analysis revealed that CG3613 and CG5787 associated with active loci, and frequently co-localized on what looks like transcriptional puffs with each other and with MLE. There was, however, not a particular bias for the staining of the X chromosome. The resolution and the dynamic range of polytene squash analysis is low, thus another technique was needed to see the binding pattern of these proteins in more detail and as a result Chromatin Immunoprecipitation followed by qPCR (ChIP-qPCR) approach is used.

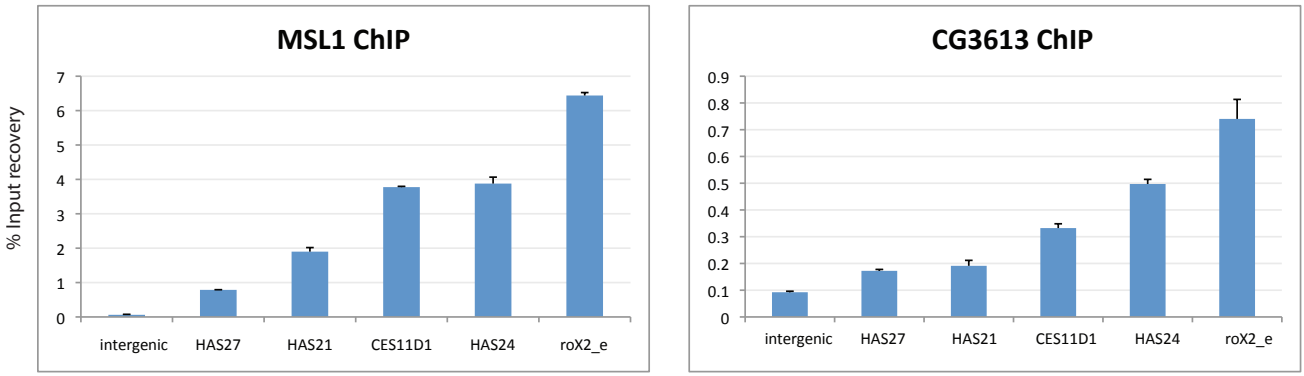

Figure 24 CG3613 can be detected at loci enriched for the MSL complex Wild-type S2 cells are fixed with formaldehyde, sonicated and chromatin prepared from these cells is used for ChIP-qPCR analysis. HAS: High-Affinity Site, CES: Chromatin Entry Site. roX2_e: roX2 high-affinity site.

Anti-sera against CG5787 did not give us any enrichment in any of the target regions that I have checked, suggesting that this antibody does not ChIP efficiently and was excluded from further analysis. Anti-sera against CG3613, on the other hand, revealed a very interesting binding pattern (Figure 24). There was a relatively low, but consistent enrichment of CG3613 on the High Affinity Sites (HAS, [136], [158]), loci that are defined by their ability to recruit partial complexes in the absence of MSL3 but are also known to be sites of high MSL complex occupancy. Although this pattern was very consistent, the percentage recovery was low. A better-controlled experiment

was called for, so I have used the stable cell line, which expresses full-length CG3613 tagged with the 3xFLAG tag at its C-terminus that I had used for the RIP experiment once again. The lack of the FLAG epitope in the wild-type S2 cells was once again the rational behind this experiment: the recovery coming from these cells served as the absolute background; enrichment over this level can only be explained by the presence of FLAG tagged CG3613.

This experiment confirmed the previous ChIP experiment carried out with CG3613 anti-sera (Figure 25): FLAG tagged CG3613 also cross-links to genomic regions of high MSL complex occupancy. MSL1 was used as a positive control, and as seen in Figure 25 it cross-links the same amount of DNA in both wild-type S2 cells and the stable cell line that contains FLAG tagged CG3613.

Putting together the results of the GRNA chromatography, RNA Immunoprecipitation and Chromatin Immunoprecipitation, one can argue that CG3613 interacts with roX2 RNA *in vivo*, and probably through its interaction with roX2, travels along with the MSL complex. The physiological role of this interaction in the context of dosage compensation remains to be investigated.

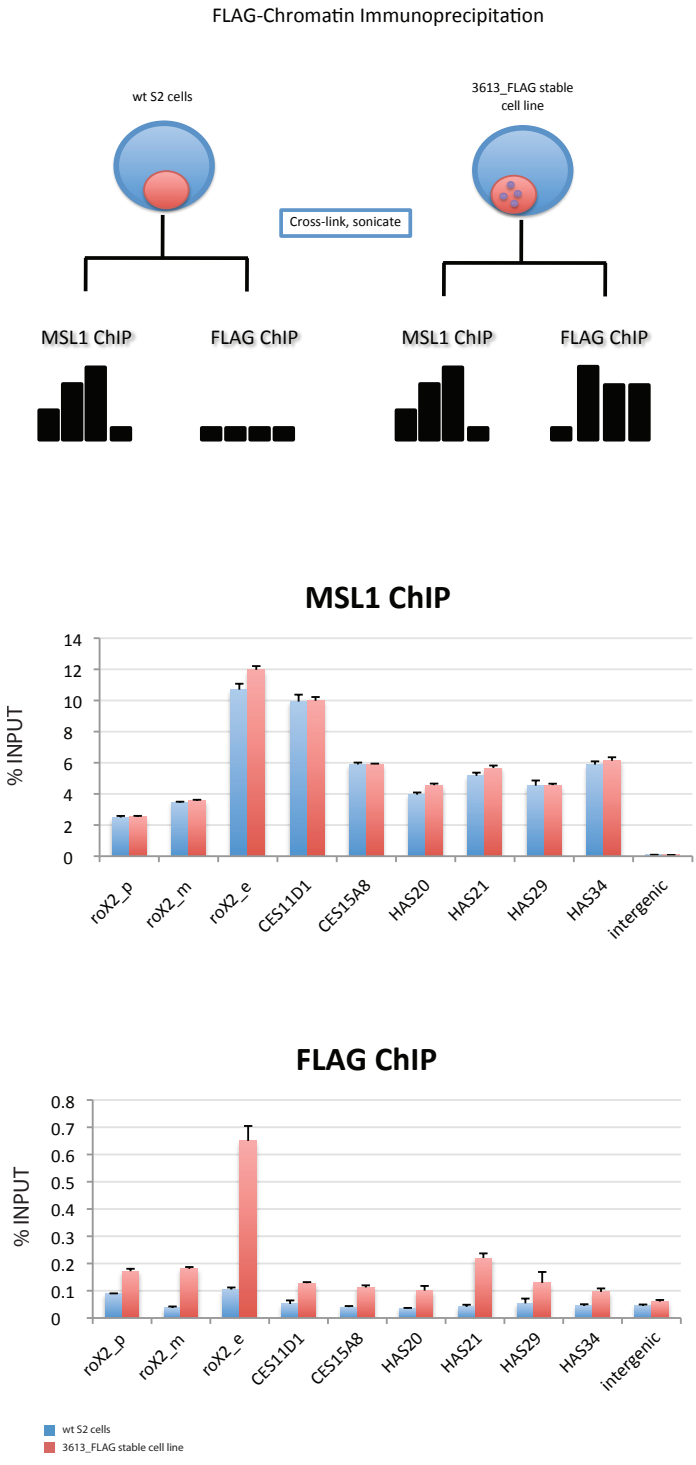

Figure 25 FLAG-tagged CG3613 is also localized to MSL-rich loci (A) The schematic representation of FLAG-Chromatin Immunoprecipitation (FLAG-ChIP) protocol. A stable cell line expressing 3XFLAG tagged CG3613 is used as a positive control, whereas wild-type cells are used to determine the background of chromatin binding to the FLAG antibody and to the beads. (B) The DNA from the ChIP experiment is reverse-cross-linked and analyzed by quantitative PCR. MSL1 is enriched similarly in both cell lines (wild-type or CG3613-3xFLAG cell line) on High-Affinity Sites (or Chromatin Entry Sites (CES)) whereas FLAG antibody shows enrichment only in the CG3613-3xFLAG cell line since FLAG epitope is only present there and not in wild-type cells.

3.2.9 CG3613 is re-distributed to some heat-shock loci upon heat-shock

Transcriptional response to heat-shock is an extensively studied phenomenon in flies. When flies are exposed to elevated temperatures, a rapid response takes place where many transcription-related proteins (i.e. RNAPII, Spt6, TopoI, PTEFb [159]) are recruited to a couple of genomic loci that code for heat-shock proteins (Hsps). Since polytene squashes of third instar larvae showed an extensive overlap of CG3613 and RNAPII (Figure 23) we wanted to know if CG3613 would also be recruited to heat-shock genes following heat-shock.

As mentioned above, CG3613 co-localizes with Serine-5 phosphorylated form of RNA polymerase II, which suggests that it is associated with active transcription. In addition, CG3613 can be cross-linked to MSL-rich regions of the genome, yet polytene squash analysis shows that there are many more loci that CG3613 interacts with. Some of these regions are the promoter of an autosomal gene, *sda* and two promoter elements of the autosomal gene *shn* at room temperature of 25˚C. However, when the temperature is increased to 37˚C, CG3613 seems to leave these promoters and localize to some heat-shock genes such as *hsp22* and *hsp26*, where they localize to the body of the gene instead of the promoter (Figure 26). While the amount of CG3613 at the promoters of *sda* and *shn* decreases, CG3613 at MSL-regions is not affected; meaning that CG3613 interacting with roX2 escapes this re-distribution following heat-shock. This re-distribution of CG3613 protein after heat-shock to actively transcribed heat-shock genes indicates that it is involved in a process that involves transcription by RNAPII.

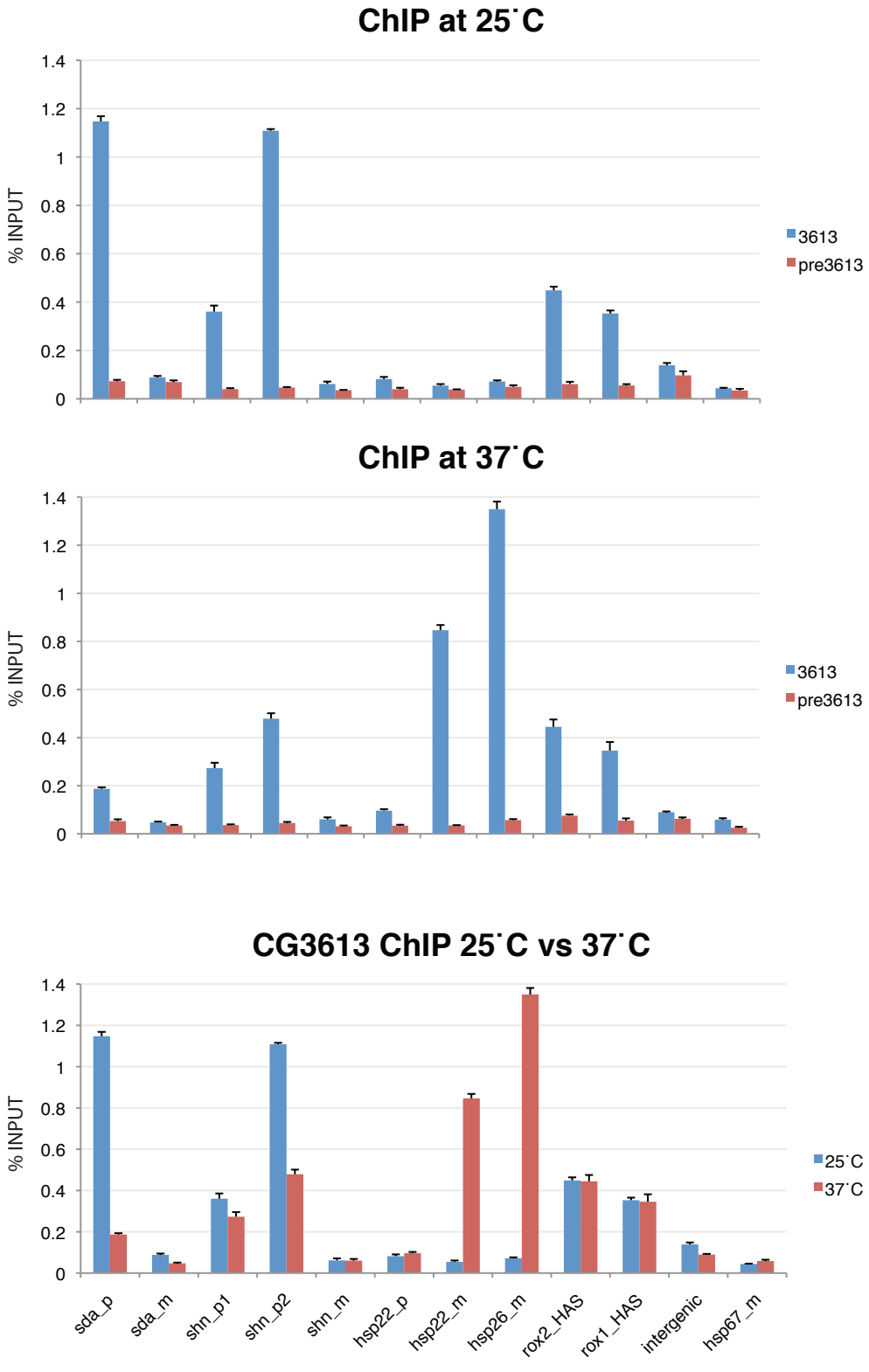

Figure 26 CG3613 is depleted from promoters and recruited to heat-shock genes following heat shock Chromatin prepared from S2 cells kept at 25˚C and cells switched to 37˚C for 30 minutes is immunoprecipitated with antibodies against CG3613 and the corresponding pre-immune sera. Upon heat-shock, promoter bound CG3613 decreases whereas heat-shock gene bound CG3613 increases. Interestingly, CG3613 cross-linked to MSL-rich sites do not change with heat-shock.

3.2.10 CG3613 is required for the viability of adult flies

In order to see if the newly identified factors interacting with roX2 affect the viability of flies, I have ordered RNAi lines from Vienna and crossed them with a fly that carries GAL4 transgene driven by a tubulin promoter. A parallel cross between wild-type flies and the tubulinGAL4 driver line was used to rule out any viability effects resulting from GAL4 over-expression. In this experiment, knocking-down CG5792 did not show a major viability defect. Interestingly, CG5787 knock-down resulted in a partial-male-specific lethality, and knocking-down CG3613 resulted in complete lethality in both sexes (Figure 27).

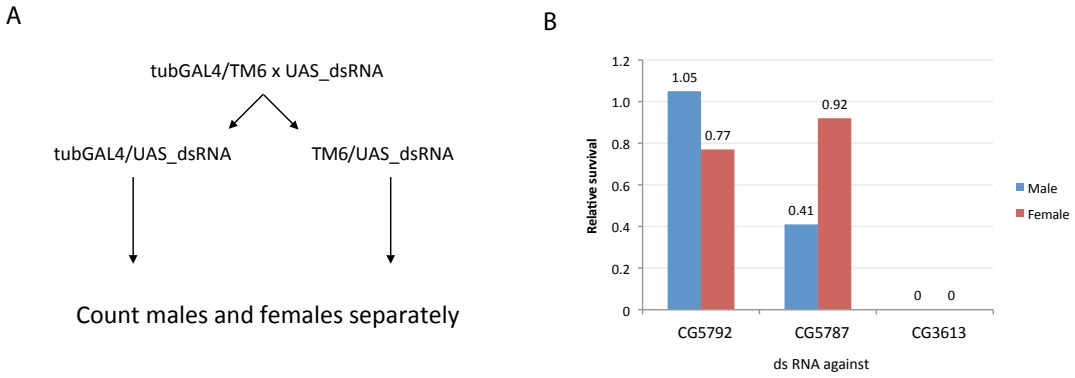

Figure 27 Knock-down of CG3613 kills the flies whereas CG5787 reduces the viability of males (A) RNAi lines from Vienna Biocenter, that contain dsRNA against the gene of interest under a UAS promoter is crossed with a driver fly line that contains GAL4 under the control of a tubulin promoter. A parallel cross between the GAL4 driver line and w^{1118} flies is carried out for normalization and correction of discrepancies from Mendelian segregation due to the presence of the TM6 balancer and GAL4 over-expression (*n*= 678). (B) Knocking-down CG3613 leads to complete lethality in both sexes (*n*= 136) whereas knocking-down CG5787 lead to a partial lethality in males (*n*= 473). Knock-down of CG5792 lead to a slight decrease in female viability (*n*= 558).

3.2.11 Depletion of CG3613 results in pre-mRNA stabilization

Since we do not have CG3613 mutants available to us, we decided to knock it down in S2 cells by RNAi in order to see if it affects gene expression and/or mRNA stability. Interestingly, there was a general, but slight increase in the RNA levels of some genes that we checked (Figure 27). This, however, is probably not because of elevated transcription levels because when we checked the expression of a subset of genes with primer pairs that span intronic regions, we have observed a striking up-regulation, a 2-3 fold difference as compared to control cells. This pre-mRNA stabilization is not specific to the X chromosome as genes like pyruvake kinase (PyK, 3R) and Thor (2L) are also affected similarly. It is important to note that the up-regulation observed in X-chromosomal genes (Suv4-20, armadillo and pcx, which are also high-affinity sites) is more pronounced when the primer pairs amplify exonic regions. Whether this is a general phenomenon, or restricted to a subset of genes remains to be seen.

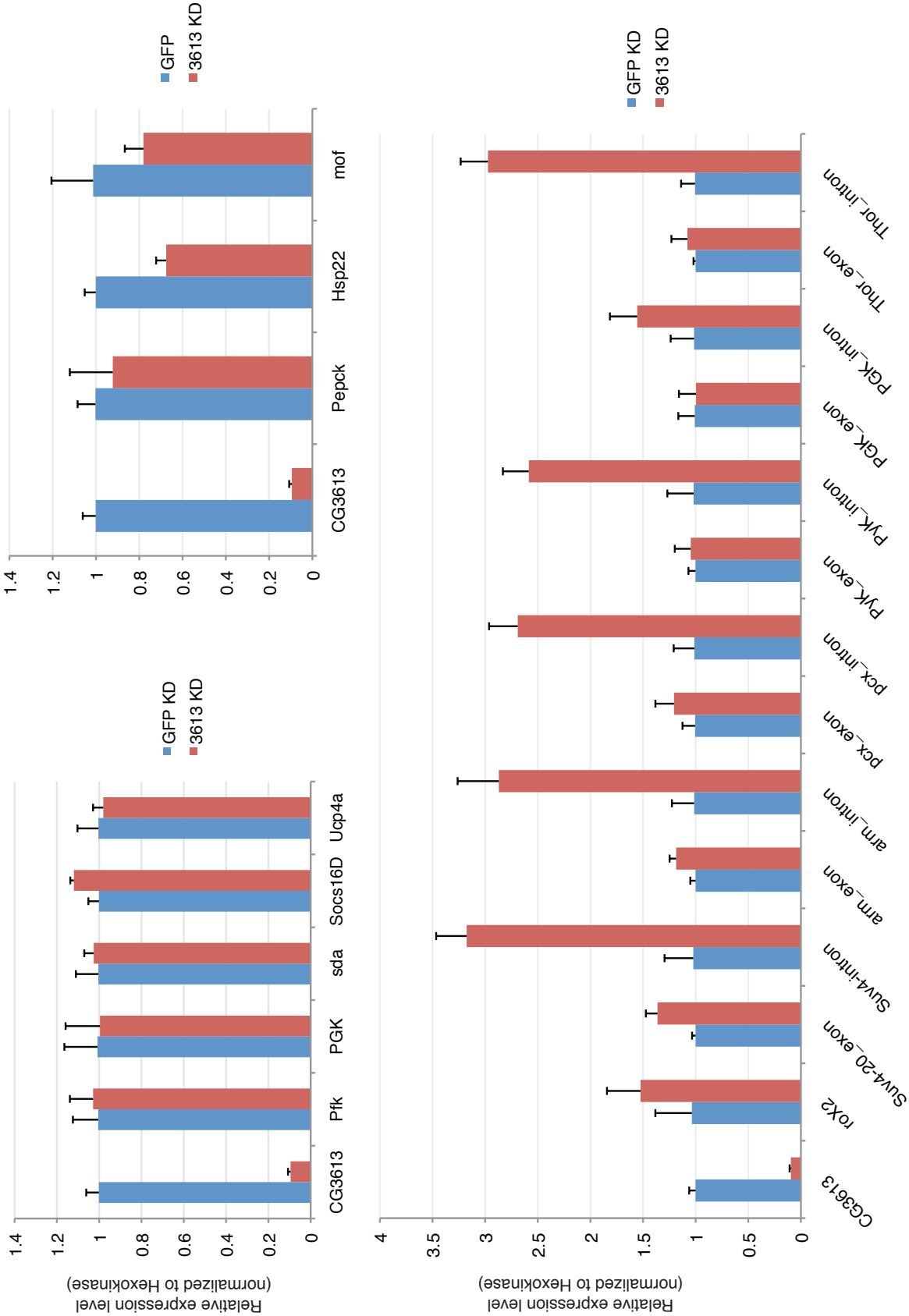

Figure 28 CG3613 knock-down results in intron containing pre-mRNA stabilization CG3613 was knocked-down in S2 cells for 3 days using dsRNA against CG3613 (two independent knock-downs with two different dsRNA, one against the 3'UTR one against an exon of CG3613, not shown). Total RNA is reverse transcribed and analyzed by RT-qPCR. All primers amplify exonic regions except for the ones designated as "intron" which have at least one primer in an intron of the indicated gene.

4. Discussion

Sex determination and dosage compensation in flies have been excellent model systems to study many cellular processes and how these processes exert genome-wide, far reaching consequences. For instance, initiation of dosage compensation is a prime example of concentration dependent modulation of gene expression by transcription factors: The so-called X-linked numerator genes lead to the expression of the master sex regulator protein, *sxl*, in females when there are two copies of them, but not in males when there is a single copy of each. Alternative splicing is also central to dosage compensation, as SXL, in turn, regulates its own splicing and leads to an mRNA that codes for full-length SXL, forming an auto-regulatory loop in females, whereas in males an exon containing a stop codon is retained in the mRNA, leading to truncated, non-functional protein. Stable expression of SXL results in two events: repression of *msl-2*, the key protein that is required for the formation of the dosage compensation complex in males, and alternative splicing of *dsx*, which yields two sex-specific proteins DSXM and DSXF that in turn lead to sexual differentiation in both sexes. Upon expression of *msl-2* the dosage compensation complex is formed (MSL1, MSL2, MSL3, MOF and MLE), which is targeted to the male X-chromosome, where it leads to approximately 2-fold up-regulation of transcription together with the acetylation of histone 4 lysine 16 throughout the male X-chromosome, thus making dosage compensation a suitable system for studying epigenetic control of gene expression.

The aspect that was the subject of my thesis, however, was the role of ncRNAs in the transcriptional regulation of the male X chromosome, another phenomenon that one can study by using dosage compensation as a model system. The dosage compensation complex, in addition to the aforementioned polypeptides, contains two redundant ncRNAs that are stably associated with the complex, roX1 and roX2. Previous work on the role of these ncRNAs suggests that these RNAs are required for targeting the complex to the X-chromosome [101], and perhaps are required for the acetyltransferase activity of MOF on target genes [107]. We sought to investigate the role of these RNAs further, and as a strategy we decided to identify the polypeptides that roX RNAs interact with and investigate their biology with respect to roX RNAs.

4.1 MLE: a male specific lethal protein

Among the known components of the DCC, MLE was the first one found to be localized to the male X-chromosome [114] and also the first one that was shown to be associated with the chromatin mostly via RNA [116]. MLE, as an RNA/DNA helicase with domains that are shown to interact with RNA, was the primary suspect for the interaction with roX RNAs, and indeed was documented to be required for the incorporation of the roX RNAs into the DCC [113]. There is one caveat though, it is notoriously difficult to co-immunoprecipitate MLE together with other members of the DCC, and biochemical purifications of MOF and MSL-3 are essentially MLE-free [118]. This contrasts with the immunological stainings that show clear enrichment of MLE on the X chromosome in males and of course the genetic studies that lead to its discovery showing MLE's indispensability for dosage compensation.

We decided to purify MLE from *Drosophila* embryos using a tandem affinity purification approach. The reason was to identify additional polypeptides that interact with MLE which could explain its role in the incorporation of roX RNAs into the DCC and perhaps also give us clues about other functions of MLE within the cell beyond dosage compensation. The result of the purification was elusive: there were no proteins we could detect aside from MLE itself. There are two possible explanations for these observations: either MLE does not interact with any other protein in the nucleus, or it does interact with other proteins but these interactions are too weak to be detected by a biochemical purification. Either way, based on this purification, it is safe to say that there is no stable MLE-complex in the nucleus that can be purified by tandem affinity purification and analyzed.

As mentioned above, MLE was the first MSL protein shown to localize to the X chromosome, and strangely this interaction was shown to be mostly RNase-sensitive, arguing against the hypothesis that it stably interacts with the DCC. I have also shown that MLE's interaction with chromatin is very sensitive to RNase treatment and I could not detect any protein interacting with MLE after an affinity purification. And yet, when roX2 RNA is used to pull-down proteins from a nuclear extract using GRNA chromatography, MLE is robustly purified under conditions where no other

MSL protein can be detected. This data is consistent with a model where MLE acts as a general "RNA chaperone" ([160],[161] and [162]) which interacts directly with many cellular RNAs that are kinetically trapped in an unfavorable conformation, including but not restricted to roX RNAs (perhaps by recognizing some RNA sequences, or secondary structures with its many RNA binding domains [139]), re-models their overall folding by melting secondary structures that form after transcription and/or dissociate certain RNA binding proteins and allow others to bind to the re-modeled RNA. An excellent support for this model comes form *para* slicing. MLE is involved in the splicing of a Na^+ channel, *para*, however in a very peculiar way. In the mutant background of mle^{napts} the splicing of *para* disturbed and only 20% of the cDNAs can be identified as full-lentgh transcripts. Upon detailed analysis of this phenotype Reenan and colleagues [140] found that MLE resolves a stable secondary structure between an exon and the following intron, a structure which is evolutionarily conserved and required for the editing of the *para* transcript by ADAR, but when left unresolved leads to various exon-skipping events and the splicing catastrophe is the end result. Interestingly, MLE is not absolutely required for this re-modeling, in the background of a null allele of MLE, both *para* editing and splicing seems to be unperturbed, indicating that there are other RNA helicases that can compensate for the lack of MLE, which is in accord with the observation that females lacking MLE protein look the same as wild-type females in terms of viability and fertility.

In a way similar to the resolution of the stable secondary structure in the *para* pre-mRNA which is initially required for ADAR to edit the RNA but is detrimental when left un-resolved, MLE may be responsible for re-modeling roX RNAs, without which the RNAs cannot interact with the MSL complex and targeted for degradation in the cytoplasm if they are exported, or in the nucleus if they are retained there. MLE's robust interaction with roX2 can enable it to interact with the RNA emerging from RNAPII and rapidly make it available for MSL interaction. In this scenario, MLE does not have to interact with any other DCC component but the roX RNAs and yet it would be required for the stabilization of roX RNAs and for a functional DCC (Figure 29).

Designing an RNA that does not require MLE's re-modeling activities for DCC interaction is the most straightforward way to test this hypothesis. Such an RNA

would be able to function as the wild-type roX RNAs in the absence of MLE protein assuming that MLE is not required anymore once DCC incorporates the roX RNAs (although it was shown that roX2 indeed keeps on interacting with the DCC in the absence of MLE [112], it is not clear if such a complex is fully functional *in vivo*). Notably, Park and colleagues [138] have shown that a 6-mer of an evolutionarily conserved stem-loop structure in the roX2 RNA can target MSL proteins to the X-chromosome. Interestingly, MLE is reported to show a heterogeneous staining unlike other MSL proteins, suggesting that it can interact with other RNA species than roX RNAs or that roX RNAs have additional uncharacterized functions in the autosomes.

4.2 CG3613: a roX2 interacting protein

An RNA that is transcribed in the nucleus associates with many proteins as soon as it is transcribed. Some of these proteins are necessary for the correct splicing of the RNA, some are necessary for proper 3' end formation, some are necessary for its export, some are necessary to degrade the RNA if one of the RNA processing steps fail or if the RNA is a cryptic transcript that is not meant to freely diffuse in the nucleoplasm. All of these decisions depend on the nature of the RNP, and the proteins an RNA interacts with ultimately determine its fate. Until this study, the only proteins that were known to interact with roX RNAs were MSL proteins. MLE was proposed to be the primary point of contact, but immunoprecipitation experiments carried out under conditions where MLE protein can no longer be co-immunoprecipitated, one can still detect roX2 co-immunoprecipitating with the rest of the complex. However, all the MSL proteins, with the exception of MSL1 has some RNA binding potential ([112], [115] and [163]), making it difficult to pinpoint exactly which member or members of the complex roX RNAs interact with.

Finding out additional proteins that interact with roX RNAs can lead a way to shed light to their function in the context of dosage compensation or it can help us understand what makes roX RNAs what they are: male specific lethal RNAs. I have used GRNA chromatography to this end and identified four polypeptides that were consistently purified with roX RNA but not with control RNA. These were MLE,

CG5792, CG5787 and CG3613. MLE was the only MSL protein that I could detect in these purifications.

By generating antibodies against these polypeptides, I was able to study them in more detail. MLE, as a reasonably well-studied MSL protein, is mostly nuclear with some cytoplasmic fraction. The form of CG5792 that I could pick up in western blots turned out to be completely cytoplasmic, and although it is very well possible that it might have a role in the biogenesis of the MSL complex in the cytoplasm by regulating perhaps a structural aspect of roX RNAs I have decided not to pursue it and restrict my studies on the other proteins.

By using the antibodies I have generated during the course of this study, I have shown that both CG3613 and CG5787 are chromatin bound proteins with a detectable cytoplasmic fraction. Unlike MLE, these proteins do not interact with the chromatin in an RNase-sensitive manner. CG3613, however, seems to have a minor, fast-migrating fraction that is indeed released from the chromatin by RNase treatment. In addition, chromatin bound CG3613 seems to be mostly phosphorylated, as treating extracts with λ-phosphatase converts all slow-migrating species to fast-migrating species. It remains to be seen if the fast-migrating RNase-sensitive fraction of CG3613 represents roX-interacting fraction of the protein on chromatin.

CG3613 and CG5787 stain large and small puffs and other inter-bands on polytene spreads. Their staining pattern is very similar and a generally overlap. Interestingly, the autosomal loci stained by MLE are also mostly occupied by CG3613 and CG5787, suggesting that these proteins might work on a similar subset of RNAs.

I have confirmed the *in vitro* interaction between roX2 and CG3613 *in vivo* by RNA immunoprecipitation. If CG3613 does interact with roX2, similar to MLE, one would predict to immunoprecipitate it on MSL target sites. As expected, I was able to show, both with antibodies against CG3613 generated in this study, and by using a FLAG-tagged transgene and antibodies against the FLAG antigen that CG3613 is present at loci where we can detect MSL binding by chromatin immunoprecipitation.

In order to elucidate the role CG3613 plays in the nucleus, I have knocked it down in S2 cells and checked changes in gene expression. There is a reproducible, ~3-fold increase in many intron-containing mRNAs (pre-mRNAs) in cells that are knocked-down for CG3613. The spliced mRNAs also increase to approximately 1.1-to 1.3-fold of wild-type levels. This result is confirmed by using two different dsRNAs targeting two different portions of CG3613 mRNA (one targets an exon and the other targets the 3'UTR) showing that the effects that we observe do not result from the reduction of an off-target gene. At steady-state, the rate of transcription, the rate of degradation of mis-spliced, mis-processed transcripts in the nucleus, and the rate of degradation of cytoplasmic mRNA should be at equilibrium. Under these circumstances, barring the mRNAs with extremely short half-lives, the amount of mRNA should greatly exceed the amount of pre-mRNA that are just being transcribed by RNAPII and processed co-transcriptionally. It is, then, reasonable to assume that a 3-fold increase in the pre-mRNA levels can account for the 10-20% increase in the mRNA levels detected upon CG3613 knock-down. Such a result can be obtained when the splicing is machinery is disturbed, leading to mRNA with introns in them. However, there are various factors in the nucleus and in the cytoplasm that deal with such transcripts, destroying them as fast as possible [38], [164]. Similar pre-mRNA stabilization has been shown to occur in yeast cells depleted of a component of the exosome, Rrp41p by the Tollervey laboratory [49]. Like our observations in this study, Tollervey and colleagues report a 2- to 8-fold increase in the unspliced pre-mRNA levels for some genes where the increase in the mRNAs levels for the same genes are around 1.2- to 1.5-fold. They also show that yeast cells harboring a *prp2-1* TS mutant, an RNA-helicase that is required for the first transesterification step during splicing, but not required for the assembly of the Spliceosome on the pre-mRNA, do not show this pre-mRNA stabilization phenotype. Yeast cells lacking *prp2* and *rrp41* together, however, show a much higher level of pre-mRNA stabilization, 20- to 50-fold.

A direct link between CG3613 and the exosome has not yet been established, however, changes in gene expression suggests that such a link exists. If that is the case, roX2 might be targeted for degradation by CG3613 and that would be the role of this protein. Indeed, an increase in roX2 levels is observed upon CG3613 knock-down, but only about 1.5-fold. Alternatively, roX2 might be acting as a platform to recruit CG3613 and the complex associated with it to the hypertranscribed X-

chromosome as a measure against increased cryptic transcripts and/or mis-processed mRNAs. Further experiments are required to elucidate the exact role of CG3613 in general and within the context of dosage compensation.

In summary, the work detailed in this thesis describes the identification of a novel roX2 interacting protein, CG3613, which is required for the viability of *Drosophila* adults. A member of the STAR family of proteins, CG3613 is phosphorylated *in vivo* and is very likely to be regulated by phosphorylation. The protein localizes to MSL-rich regions, also to other transcriptionally active loci, and is recruited to heat-shock genes upon heat shock. Knocking it down in S2 cells results in the stabilization of intron containing pre-mRNAs transcribed from the X-chromosome and the autosomes. Its role in dosage compensation has not been clarified, but it is possible that roX2 acts as a platform for recruiting CG3613 and associated proteins to cope with hypertranscription from the X chromosome in males.

Due to time constraints I have not been able to complete the analysis of CG3613 protein. In the near future, we are planning to utilize mutants of MSL proteins to see if the MSL complex affects localization and function of CG3613 protein. Likewise, we are generating mutants of CG3613 and we will use these flies to see if CG3613 modulates MSL function. There is a possibility that binding of CG3613 to RNA depends on its phosphorylation levels. We will investigate this possibility and see if this “phospho-switch” and CG3613’s interaction with roX2 is important for dosage compensation.

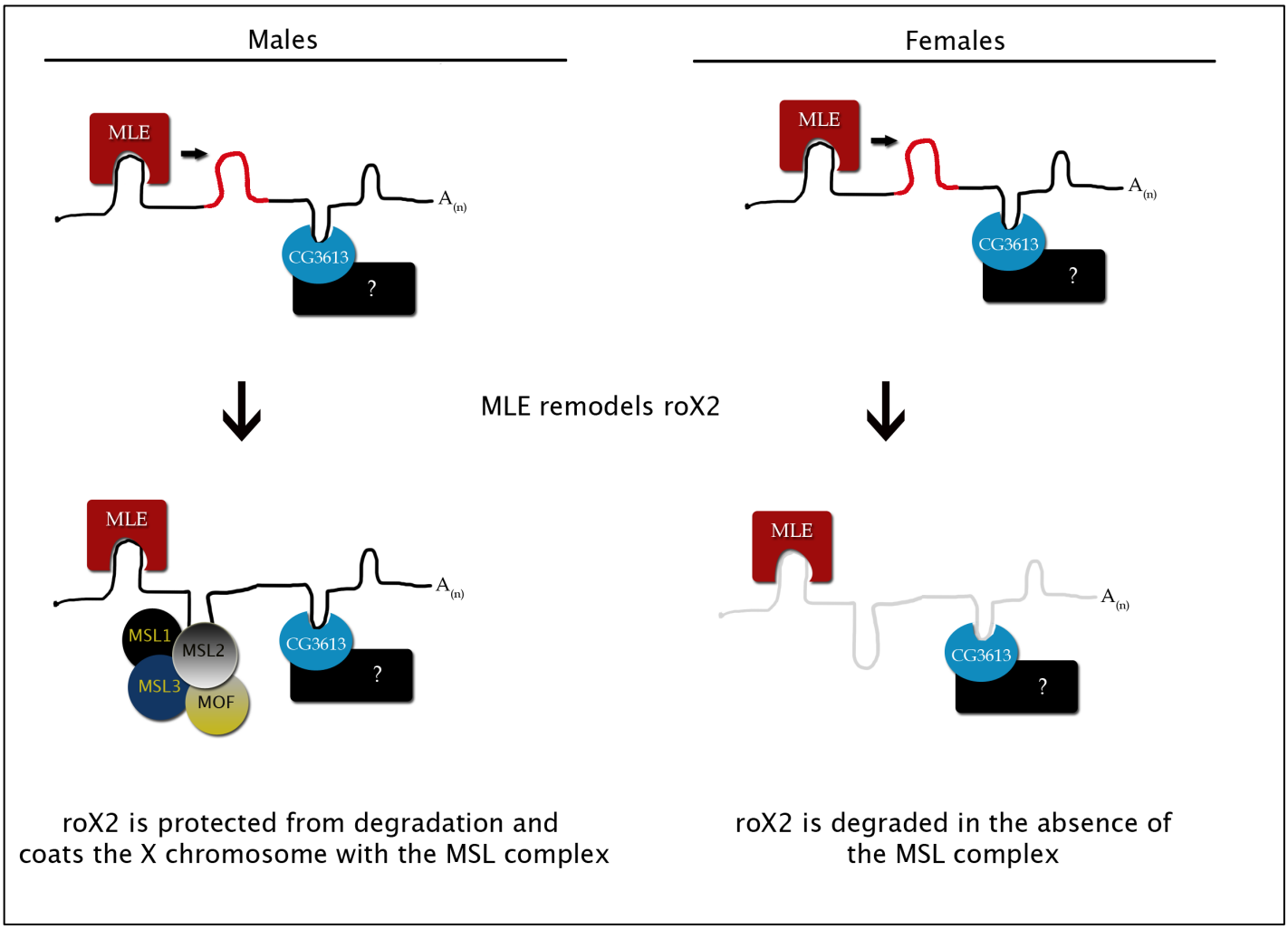

Figure 29 The interdependency of roX RNAs and the MSL proteins roX RNAs interacts with many polypeptides, and these interactions determine their role and fate in the cell. In this model, MLE interacts only with roX RNA and re-models its structure. This re-modeling makes the RNA competent to bind the MSL complex (MSL1, 2 and 3 and MOF). This interaction protects roX from degradation and the complete-MSL complex is targeted to hundreds of sites on the X chromosome. CG3613 also travels with the MSL complex via its interaction with roX2, but is not required for roX2 stability. In females, although MLE is present and very likely interacts with scarcely expressed roX, lack of the MSL complex leads to the degradation of roX RNA. This explains the robust accumulation of roX RNAs and proper targeting of the MSL complex to the X chromosome in females ectopically expressing MSL2. According to this model, in males lacking MLE, roX RNA would not be able to interact with the MSL complex, leading to its degradation and failure of dosage compensation.

5. Materials and Methods

Cloning MLE-HA-FLAG

cDNA of full-length MLE, missing the last 6 amino acids was digested from pBBHisMLE vector with NcoI (blunt) and NotI. An oligo that has the C-terminal 6 aa that is lost after NotI digestion, an HA tag and a FLAG tag was assembled in a way that creates a NotI site at its 5' end and an XbaI site at its 3' end. A three way ligation using MLE (NcoI[blunt]-NotI), the oligo (NotI-XbaI) and pCasper4 (with a tubulin promoter) digested with Asp718(blunt)-XbaI created the vector pMLE-HA-FLAG. Flies containing this vector are generated by P-element mediated transformation

Oligo:
gcGGCCGCTGGGGTAACTTTTACCCATACGATGTTCCAGATTACGCTCCCG
ACTACAAGGACGACGATGACAAGTAATctaga

Last 6 amino acids of MLE, HA-tag. FLAG-tag.
NotI and XbaI restriction sites are underlined.

Nuclear Extract Preparation from Embryos

Three types of embryonic nuclear extracts were prepared and used in this thesis. All three protocols start with the same procedure for isolation of the nuclei from embryos but differ in the way protein was extracted from these nuclei.

Nuclei isolation
12-16hrs embryos laid on large apple-juice agar plates placed in population cages were collected by the help of a series of sieves and dechorionated. Dechorionated embryos are immediately rinsed with 120mM NaCl, 0.04% Triton X, and washed extensively with water. The embryos are then dried over a vacuum trap (they are left humid, not dried completely) and weighed. 3mL of NU-I (15mM HEPES-KOH 7.6, 10mM KCl, 5mM $MgCl_2$, 0.1mM EDTA, 0.5mM EGTA, 350mM sucrose, 2mM DTT, 1x Protease Inhibitors) per gram of embryos was added and homogenization

was carried out using a Potter-Elvehjem type Teflon pestle mounted on a rotor (1000 rpm, 30-40 strokes). The homogenate was filtered through Miracloth twice and centrifuged at 10,000g for 15 minutes using a GSA rotor. The nuclear pellet is very loose after this step.

Conventional nuclear extraction

This type of extract was used for the initial MLE purifications (Figure 7). It involves protein extraction from the nuclei using ~400mM ammonium sulfate, and later precipitation of all proteins by ~3M ammonium sulfate in order to obtain a concentrated protein extract in the end.

The nuclei are resuspended in 1mL NU-II (15mM HEPES-KOH 7.6, 110mM KCl, 5mM $MgCl_2$, 0.1mM EDTA, 1x Protease Inhibitors) per gram of embryo. The suspension is distributed into SW40 ultracentrifucation tubes and 1/10 volume of 4M ammonium sulfate is added on top. The tubes are sealed and rotated for 20 minutes in the cold room. Then, ultracentrifugation is carried out in an SW40 rotor at 38 krpm for 1hr at 4˚C. The lipid layer on top is removed and the nuclear extract is collected into a cold beaker. While stirring and over a period of 5 minutes, 0.3g of ammonium sulfate per milliliter extract is added. The mixture is stirred for another 10 minutes and then transferred into SS34 bottles and spun at 15 krpm for 30 minutes at 4˚C. The supernatant is discarded. Per gram of embryo 20µL of HEMG40 (25mM HEPES-KOH pH7.6, 40mM KCl, 12,5mM $MgCl_2$, 0.1mM EDTA, 10% Glycerol, 1mM DTT, 1x Protease Inhibitors) is added and the nuclei is re-suspended with the help of a glass rod. Then 180µL of HEMG40 per gram of embryo is added and the mixture is fully re-suspended with the help of a 5mL pipette. Dialysis was against HEMG40 and allowed to continue until the conductivity of the extract was equal to HEMG100 (25mM HEPES-KOH pH7.6, 100mM KCl, 12,5mM $MgCl_2$, 0.1mM EDTA, 10% Glycerol, 1mM DTT, 1x Protease Inhibitors). The precipitated material is removed by spinning the extract in an SS34 rotor at 15 krpm for 5 minutes at 4˚C. The clear supernatant is snap-frozen in liquid nitrogen and kept at -80˚C until use.

KCl and Sonicated Soluble Nuclear Fraction (SSNF) extraction

"KCl" type of extract was used for GRNA purifications (Figures 11, 14, 15) and also for small-scale MLE purifications (Figure 8). It involves using only KCl as the salt and skips total protein precipitation by ammonium sulfate. SSNF type extract was only used for small-scale MLE purifications (Figure 8). This protocol mainly relies on squeezing out nuclear proteins by ultracentrifugation, but a brief sonication is introduced before centrifucation to increase the protein yield. Since there is no salt-extraction in this protocol dialysis is not required. Theses two protocols differ after the re-suspension of the nuclei in HEMG20.

For both protocols, the nuclei are re-suspended in 1mL buffer AB (15mM HEPES-KOH 7.6, 110mM KCl, 5mM $MgCl_2$, 0.1mM EDTA, 2mM DTT, 1xProtease Inhibitor Coctail) per gram of embryos. Yellow yolk is left behind, and the nuclei are dispersed with loose pestle (A, if Kontes, B if Wheeton) by three strokes. The suspension is centrifuged at 10,000g in the GSA rotor for 10 minutes at 4˚C. Supernatant is discarded and the pellet is dissolved in 100µl HEMG20 (25mM HEPES-KOH 7.6, 100mM KCl, 12.5mM $MgCl_2$, 0.1mM EDTA, 20% Glycerol, 1 x Protease Inhibitors, 2mM DTT) for every gram of embryo.

For KCl extration, the nuclei suspension in HEMG20 is distributed to ultracentrifugation tubes (SW40, 14mL capacity), KCl is added to 420mM from 3M stock, tubes are sealed with parafilm and incubated in the cold room for 30 minutes and then centrifuged in an SW40 rotor at 38 krpm for 1hr at 4˚C. The lipid layer on top is sucked away and the supernatants from different tubes are pooled for dialysis. Nuclear debris at the bottom of the tube is avoided. The dialysis was against HEMG20, and was allowed to continue until the salt concentration of the extract was equal to that of HEMG20 (i.e. 100mM KCl). The precipitated material is removed by centrifugation in an SS34 rotor at 15 krpm for 10 minutes at 4˚C and the clear supernatant is snap-frozen in liquid nitrogen and kept at -80˚C until use.

For SSNF extraction, the nuclei suspension in HEMG20 is mildly sonicated using a Branson sonicator (Output: 2, Duty cycle 40, 3 times 5 pulses) and then distributed into ultracentrifugation tubes (SW40, 14mL capacity). Centrifugation was in an SW40

rotor for 1hr at 38 krpm at 4˚C. The lipid layer on top is sucked away and the brownish supernatant is transferred into fresh tubes. Dialysis is not required thus the extract is snap-frozen in liquid nitrogen and kept at -80˚C until use.

Tandem Affinity Purification

Nuclear extract prepared from embryos (wild-type and transgenic in parallel) are diluted in Binding buffer (BB, 25mM HEPES-KOH 7.6, X mM NaCl, 5mM $MgCl_2$, 0.05% Igepal CA 630, 2mM DTT, 5% Glycerol, 40u/mL RNasin, 1xProtease Inhibitor Coctail. Concentration of NaCl differed between purifications, In Figure 7 it was 110mM or 60mM, in Figure 8 it was 125mM). For 10mL of solution, 100µL of 50% anti-FLAG(M2) agarose beads (Sigma) was used to collect FLAG-tagged MLE, for 2hrs at 4˚C. The beads were washed 4 times with 5mL of BB. The elution was with 500µL of BB + 250µg/µL FLAG peptide for 1hr at 4˚C, twice. The eluate is incubated with 50µL of 50% slurry of anti-HA agarose (Sigma) for 90 minutes at 4˚C. The beads are washed 3 times with BB. The bound proteins were eluted with 50µL of non-reducing 1XLaemmli sample buffer at 70˚C for 30 minutes. The eluate is examined by immunoblotting or silver staining after PAGE.

Silver staining

Mass-spectrometry friendly silver-staining protocol is adapted from Schevchenko, 1997 ([165]). The gel is fixed in 50% MeOH, 12% AcOH, 0.05% formalin for 2 hrs. It is washed three times with 35% EtOH for 20 mins each, and twice with water for 10 minutes each. The gel is then sensitized with 0.02% $Na_2S_2O_3$ for 2 mins and washed with water three times, 5 minutes each. Gel is Stained with 0.2% AgNO3, 0.076% formalin for 30 mins in the cold room after which it is rinsed with water twice, 1 minute each. The gel is developed in 6% Na_2CO_3, 0.05% formalin, 0.0004% $Na_2S_2O_3$ until the bands were sufficiently visible. The gel is rinsed with water and staining is stopped with 50% MeOH, 12% acetic acid. The gel is rinsed with water and kept in 1% acetic acetic in the cold room.

Coomassie staining

A commercial preparation of colloidal Coomassie stain G-250 was used (GelCode Blue, Thermo). After PAGE, the gel is washed three times with water, 5 minutes each. Then GelCode Blue solution is added onto the gel and incubation at room temperature is carried on until the bands were sufficiently visible. De-staining is not necessary, but gels are kept in water to get rid of the background staining when desired.

Mass spectrometry

All mass-spectrometry related data presented in this work was carried out by Innova Proteomics (Rennes, France). Briefly, silver stained 1-D gel pieces were cut and sent to Innova Proteomics where they are reduced and alkylated (DTT/iodoacetamide), digested with 6.7ng/µL trypsin overnight at 37˚C followed by salt removal with POROS 20 R2 (Perseptive Biosystems) and deposition on MTP AnchorChip 600/384 S/N 11169 with HCCA matrix (a-Cyano-4-hydroxycinnamic acid). Mass spectra was acquired using a MALDI-TOF/TOF system (Ultraflex, Bruker Daltonics). Peptide fingerprint acquisition was carried out by FlexControl, FlexAnalysis and Proteinscape software from Bruker Daltonics. NCBInr was used for database searches.

GRNA Chromatography and PP7 pull-downs

GRNA protocol is adapted from Czaplinski [148] and Duncan [147]

Cloning and synthesis of boxB and PP7 tagged RNAs

All RNAs used in GRNA chromatography and PP7 pull-downs were made by *in vitro* transcription using Ribomax T7 kit from Promega. The run-off transcripts were purified using Megaclear columns (Ambion). The templates were all cloned into pBlueScript KS and linearized with Asp718 or XhoI after which the 3'-ends were filled in with Klenow (NEB) and purified by phenol/chloroform extraction.

The sequence of the boxB tag used in this study is as follows:

GGGCCCTGAAGAAGGGCCC

The sequence of the PP7 hairpin used in this study is as follows:
AAGGAGTTTATATGGAAACCCTT

All the tags were cloned at the 5'-end of the RNA of interest.

GFP RNA is the NheI/XhoI digest from pEGGP-C2 (Clontech).

roX2_FL is the NotI/XhoI digest from the vector roX2 78.2.2 (a gift from Mitzi Kuroda) and contains the fist and third exons of roX2 and a partially spliced second exon.

roX2_ex3 is the amplicon generated by amplifying roX2 78.2.2 with these primers:
L: ATACTGCAGTAGCTCGGATGGCCATCG
R: ACTCGAGTATTATTTGGCAATTGTTAAG
PstI and XhoI sites are underlined.

anti-sense_roX2_ex3 is created by switching the restriction sites in the primers given above and cloning into the same boxB containing vector.

To generate templates that were used to synthesize RNAs for use in PP7 pull-downs, the constructs above were sub-cloned into a vector that contained the PP7 hairpin at the 5'-end.

Expression and purification of GST•λ•6His and ZZ•tev•PP7cp•6His

Plasmid expressing GST•λ•6His was a gift from Matthias Hentze, plasmid encoding ZZ•tev•PP7cp•6His was a gift from Kathleen Collins ([67]). GST•λ•6His requires co-expression of rare codon expressing bacteria and BL21(DE3) CodonPlus was used for this purpose. ZZ•tev•PP7cp•6His can be expressed readily using BL21 (DE3) bacteria. Both proteins were purified to near homogeneity using single-step IMAC (Ni-NTA, Qiagen).

GRNA Protocol

30µg GST•λ•6His was incubated with 25µL of Glutathione-Sepharose beads (GE Healthcare) for 1hr at 4˚C in binding buffer (BB, 50mM Tris•Cl pH 7.6, 100mM KCl, 2mM $MgSO_4$, 0.1% Igepal CA-630 (Sigma), 0.1mg/mL tRNA (Roche), 0.01mg/mL Heparin, 1x Protease inhibitor coctail (Roche)) and washed several times with BB. Then 100pmol of RNA (tagged with boxB if to be immobilized on beads) was incubated with the GST•λ•6His bound beads in BB+40u/mL RNasin (Promega) for 12-16hrs. The beads are washed twice with BB and are then incubated with 250-500µg of embryonic nuclear extract (KCl method, see above) in 300µL BB+40u/mL Rnasin for 45 minutes at 18˚C. The beads are washed with BB extensively. The elution was with 30µL BB with 100mM NaCl instead of KCl and without tRNA and with 0.5µg Protease-free RNaseA (Chembiochem) at 30˚C for 30 minutes. The eluate was analyzed by silver staining or by immunoblotting.

PP7 pull-downs

GRNA protocol was followed exactly, only that GST•λ•6His is replaced by ZZ•tev•PP7cp•6His and Glutathione-sepharose beads were replaced by IgG beads (GE healthcare). The RNA to be immobilized contained a PP7 hairpin instead of boxB.

Chromatin Immunoprecipitation from S2 cells

Wild-type S2 cells maintained in Drosophila S2 medium (Invitrogen) + 10% FCS were fixed with 1.8% Formaldehyde for 15 minutes at room temperature. Formaldehyde is quenched with 125mM Glycine for 5 minutes at room temperature. All the steps that follow were carried out either on ice or in the cold room (~4˚C). The cells were pelleted by centrifugation at 2000 rcf for 2 minutes. The cells were washed (i.e. resuspended in the given buffer and rotated in the cold room for 5 minutes each step) two times with Paro 1 (10mM Tris pH 8.0, 10mM EDTA, 0.5mM EGTA, 0.25% TritonX-100), two times with Paro 2 (10mM Tris pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA) and twice with RIPA (140mM NaCl, 25mM HEPES pH 7.5, 1mM EDTA, 1% TritonX-100, 0.1% SDS, 0.1% DOC). Cell pellets not exceeding

100µL were resuspended with 500µL RIPA and sonicated (Branson sonifier. Power output: 3, Duty cycle: 40, 30 cycles. Each cycle: 20 seconds on, 50 seconds off). After sonication the sample is centrifuged at 14 krpm for 30 min and the pellet is discarded. The supernatant is incubated with 50µL of Protein A Sepharose beads for pre-clearing. After a brief centrifugation the supernatant is aliquoted and snap-frozen in liquid nitrogen if not used immediately.

For each IP, chromatin coming from ~$1x10^6$ cells is diluted in RIPA and antibodies are allowed to form immune-complexes for 12-16 hrs, rotating in the cold room (CG3613, 4RAP: 5µl, MLE, rat1 or rat2: 2µl, MSL1, 3BVM: 3µl, anti-FLAG(M2): 3µl). All the IPs were centrifuged for 10 min at maximum speed, and supernatants were transferred into new tubes that contain 40µl of 50% protein A sepharose (GE healthcare) if the antibody comes from a rabbit, or protein G agarose (Roche) if it comes from a rat (blocked with 1µg/µl BSA (NEB), 1µg/µl salmon sperm DNA (Invitrogen) in RIPA, for 1hr). After collecting the immune-complexes with protein A or G beads, the IPs are centrifuged for 2min at 2000rpm. The beads are washed 4 times with RIPA and once with LiCl buffer (10mM Tris-Cl 8.0, 250mM LiCl, 1mM EDTA, 0.5% Igepal CA 630, 0.5% DOC) and then once again with RIPA. The supernatant is removed as much as possible and 100µl of 10% Chelex beads (Bio-rad, in water) is added to each IP. De-crosslinking was carried out by incubating the beads for 15 minutes at 95˚C. The tubes are then spun at 10 krpm for 1 minute and allowed to cool down to room temperature. 1µl of 10mg/mL Proteinase K (Roche) is added to each tube, and incubated at 56˚C for 30 minutes. Proteinase K is then de-activated by boiling the beads at 95˚C for 15 minutes. The supernatant is then transferred to a fresh tube, diluted with water and used in qPCR.

Primer pairs used in qPCR:

roX2_p	AGCTAGATGTTGCGGCATTC	**Shn_p1**	TCACTTTGGGTGTGCTCTTG
	CGCTTACCTAAACGCTCGAC		AAACAGAGCCAGCGAGAATC
roX2_m	TCCAACCAGTGTAGCATCCA	**Shn_p2**	TCTCTTTACGGTTGATCGGC
	AGGATTGTCATAGGCGCAAC		TGGTGCCCACTCTCTCTTTC
roX2_e (HAS)	ACGGTGCTGGCTTAGAGAGA	**Shn_m**	ACCAAGAAGCGGACATATCG
	GGCGGAAATGTATTTGCAGT		AGATGCTGATGCTGATGCTG
CES11D1	GTGGAAACGGACAGCGTAA	**Hsp22_p**	CTCGCACTCAGAAAGCTGTG
	CACATCAGCGACAAGAGGC		AGCGAAGTTCTCTCTCGCTG
HAS20	ATCTCGCTGTTATCGGTTGC	**Hsp22_m**	CGAGCTAAAGGTCAAGGTGC
	TCAAGTTGATCGTGGAGCAG		TCTGATTTTCCCTCCACCAG
HAS21	GTCTGCAAGGATGAACAGCA	**Hsp26_m**	AAGGATGGATTCCAGGTGTG
	GATTGGGTTTGGTGGAATTG		ACCAAGATGGAGTCGTCCAC
HAS29	GCGACTATATAGGACCGCCA	**roX1_HAS**	ACTGCAAGTCCCGAAAGAGA
	TCCAGAAGTCCGTGTTCCAT		ACTTTTTCCTTTCCGAGGGA
HAS34	GCTAAGTTTGGATTGCCCTG	**Hsp67_m**	CATCGAAGAGGAGCAAAAGG
	CTGCATATAGGGTTTGCCGT		TGGAGATTTCCGGTTTCTTG
Sda_p	CCTTAACTCGCATCGCTTTC	**Intergenic**	ACTCGGTTCAGATCCTGTGG
	CGAAAAAGTAAAACCGTCGC		GGCCAGTGGGCTTGTAATAA
Sda_m	TTTGGAGAGCATTGCTGTTG		
	GAAATCTGTGTTGCAATGCG		

Antibody production

Plasmids expressing GST-enterokinase cleavage site-Antigen-6His fusion proteins were cloned with the following primers and cDNA clones.

	Clone name	Covered antigen	Forward primer	Reverse primer
CG5792	LD15349	38kDa/38kDa	taatgacaagTGTCGgCGTCGTTCCAAT	aatactcgagACTGAACACTGGAATGCGAA
CG5787	LD23647	40kDa/100kDa	tatagacaaaggTCACTATCGAGGTCCCCAG	tatactcgagCTGTGGTCCTGTCCTGTTT
CG3613	GH05812	20kDa/45kDa	tatagacaaagGtCTTCGGTTGGAGCAACT	tatactcgagTATTTTCGGATATGGAGCCG

The primers introduce a PshAI at the 5'-end and a XhoI site at the 3'-end. These restriction sites were used to clone all amplicons into pET41b. All proteins were expressed in BL21(DE3) cells. Briefly, BL21(DE3) bacteria were transformed with the expression vectors and single colonies were picked into LB media containing 50µg/mL kanamycin and grown at 37˚C overnight. The next day 1L culture is inoculated with 10mL of overnight grown culture and incubated at 30˚C until the OD reached 0.4-0.6 at which point protein expression was induced with 0.5mM IPTG. The bacteria was let to express the protein for 4hrs at 25˚C after IPTG induction and collected in a GSA rotor by centrifugation at 5000rpm for 15 minutes. The pellet was then re-suspended in Lysis buffer (50mM Tris 8.0, 500mM NaCl, 5% Glycerol, 1xProtease Inhibitors) and the suspension is then made to 0.1% TritonX and 1mM DTT. Sonication was used to break the bacteria, and the lysate is cleared by centrifugation using an SS34 rotor, 20,000g for 20 minutes at 4˚C. 0.5mL of Glutahione coated Sepharose (GE Healthcare) resin is used to pull-down GST fusion proteins. After several washed with the Lysis buffer, bound proteins were eluted using the lysis buffer supplemented with 20mM Glutathione for 1 hour in the cold room. The eluates were then dialyzed against PBS + 0.1% TritonX, concentrated with Centricon spin-columns (Millipore) when necessary, snap-frozen in liquid nitrogen and stored at -80˚C. Titre-Max (Sigma) is used as an adjuvant when immunizing animals with these protein preps.

For CG5792 two rabbits (4QYT and 4QZF) were injected for 5 times, for CG5787 three rats for 6 times and later 2 rabbits for 5 times (ATOM and ATDE) and for CG3613 two rabbits for 5 times (4RAP and 4QRM).

RNA Immunoprecipitation from S2 cells

Adapted from Selth, Gilbert and Svejtrup [166]

S2 cells, wild-type or the CG3613_3xFLAG6His line, are grown in S2 media + 10% FCS. 12-16hrs before the experiment expression of CG3613 was induced with 0.25mM $CuSO_4$. The cells were fixed with 1% formaldehyde at room temperature for 15 minutes. Formaldehyde is quenched with 200mM Glycine for 5 minutes at room temperature. The cells are pelleted by centrifugation at 1000g for 5 minutes and washed twice with ice-cold PBS. Then the cells are re-suspended in 500µL of FA

lysis buffer (50mM HEPES-KOH 7.6140 mM NaCl, 1 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) sodium deoxycholate, 40u/mL RNasin, 1xProtease inhibitors) and spun at 3000g for 30 seconds. Supernatant is discarded and the pellet is re-suspended in 600µL of FA lysis buffer and sonicated using a Branson sonifier (Power output: 3, Duty cycle: 40, 10 cycles. Each cycle: 20 seconds on, 50 seconds off). Sonicated extract is transferred into a new tube and centrifuged at maximum speed for 10 minutes at 4˚C. Supernatant is transferred to a new tube and the centrifugation is repeated once. The extract is then pre-cleared over 50µL of Protein A sepharose beads for 1 hr at 4˚C. Beads are removed by centrifuging at 1000g for 2 minutes at 4˚C. Supernatant is transferred into a new tube and made to 25mM $MgCl_2$ and 5mM $CaCl_2$. 3µL RNasin and 6µL RQ1 RNase-free DNase (Promega) is added and the extract is incubated at 37˚C for 30 minutes. The reaction is stopped by adding EDTA to 20mM. Insoluble material is disposed of by centrifuging at maximum speed for 10 min at 4˚C. Supernatant is either used immediately or snap-frozen in liquid nitrogen and stored at -80˚C until use.

Extract coming from about a million cells is diluted to 500µL with FA lysis buffer for each IP. 50µL of this solution is saved as INPUT. 3µL of anti-FLAG(M2) or 3µL of anti-MLE antibody is added and the tubes are incubated in the cold room overnight with end-over-end rotation. The RIPs are centrifuged at maximum speed for 10 minutes at 4˚C and the supernatants are transferred into a new tube that contains 40µL of Protein A or G slurry. Incubation was for 90 minutes in the cold room. Then, the beads are pelleted by centrifugation at 1000g for 2 minutes and washed once with FA lysis buffer, once with FA500 buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) sodium deoxycholate, 1xProtease Inhibitors, 40u/mL RNasin), once with LiCl buffer (10 mM Tris·Cl, pH 8, 250 mM LiCl, 0.5% (v/v) Igepal CA 630, 0.1% (w/v) sodium deoxycholate 1 mM EDTA, 1xProtease Inhibitors, 40u/mL RNasin) and once with TE buffer (10 mM Tris·Cl, pH 8, 1 mM EDTA, 100 mM NaCl) for 5 minutes each in the cold room with end-over-end rotation. TE buffer is removed as much as possible and the immune complexes are eluted with 75µL of RIP elution buffer (100 mM Tris·Cl, pH 8, 10 mM EDTA, 1% (w/v) SDS, 40U/mL RNasin) by incubation at 37˚C for 10 minutes. The beads are pelleted by centrifugation at 1000g for 2 minutes at room temperature, the supernatant is transferred into a new tube and the elution is repeated with another 75µL of RIP elution buffer.

Eluates are pooled, made to 200mM NaCl and 0.13µg/µL Proteinase K. Protease digestion was carried out by incubating the eluates at 42˚C for 1 hour, followed by reversal of formaldehyde cross-links by incubating at 65˚C for 1 hour. Input samples saved before are processed in parallel. Onto each RIP, 100µL nuclease-free water and 250µL acid-phenol:chloroform is added. Phase separation is carried out in MaxTrak tubes (Qiagen) by centrifugation at 10000g for 3 minutes at room temperature. The aqueous phase is transferred into a new tube and the RNA is precipitated by adding 25µL of 3M sodium acetate, 20µg glycogen and 625µL ice-cold absolute ethanol and incubating the mixture at -20˚C overnight. RIPs are centrifuged at maximum speed for 30 minutes are 4˚C. Supernatant is discarded and the pellet is washed with ice-cold 70% ethanol. After a final centrifugation at maximum speed for 5 minutes at 4˚C, supernatant is discarded and the pellet is air-dried for 5-10 minutes. 90µL nuclease-free water is used for re-solubulizing the RNA pellet. 10µL TURBO DNase buffer and 1µL TURBO DNase (Ambion) is added and the mixture is incubated at 37˚C for 30 minutes. 10µL DNase inactivation reagent is pipetted to each tube and the tubes are incubated at room temperature for 2 minutes with occasional shaking. The tubes are centrifuged at 10,000g for 2 minutes and the supernatants are transferred into fresh tubes. These RNA samples are either used immediately for reverse transcription-qPCR or stored at -80˚C until use. Reverse transcription and PCR were carried out in the same tube by using reverse transcriptase and RNasin in the SYBR mix, and adding a step of 50˚C for 30 minutes before the start of PCR amplification.

Primer pairs used in RT-qPCR:

roX2	L: TCGCAATGCAAACTGAAGTC
	R: AGGCGCGTAAAACGTTACC
Tubulin	L: AACCTGAACCGTCTGATTGG
	R: GTCAGATCCACGTTAAGGGC

Immunostaining of polytene chromosomes

Adapted from Johansen et al. [167]

Wandering 3rd instar larvae were dissected in PBS and the salivary glands are separated from larval imaginary discs and other tissue as much as possible. A pair of glands is then transferred into Fix1 (1x PBS, 3.7% formaldehyde, 1% TritonX) and

kept there for 2 minutes and then transfer to Fix2 (3.7% formaldeyde, 50% acetic acid) and kept in this solution for another 2 minutes. The gland is then transferred onto a clean Sigmacoated coverslip that has a drop of Lactoacetic acid solution (50% acetic acid, 17% lactic acid) on it. Then a poly-L-lysine coated microscope slide is lowered on top of the coverslip. Cell lysis is achieved by moving the coverlip back and forth 3-4 times with a pencil without applying any vertical pressure on it. Tapping the coverslip a few times with a pencil can be helpful in spreading the chromosomes at this stage. To flatten the chromosomes, using the inside of a thumb, pressure is applied on top of the coverslip, absolutely avoiding and lateral movement at this stage, which breaks the chromosomes. Putting the slide on top of a stack of Kim Wipes, and putting another one on top of the coverslip can be helpful. The slide is then dipped into liquid nitrogen and the coverslip is removed with the help of a razor and placed into a Coplin jar filled with PBS and kept at 4˚C until all slides are ready. The slides are then washed with PBST (1xPBS, 0.4% TritonX) for 30 minutes. The slides are taken out of the jar, touched on the corner with a Kim-Wipe to get rid of excess PBST. 50µL of blocking buffer (PBST, 0.2% BSA, 5% normal goat serum) is applied on top of the slide and the slide is covered with a piece of parafilm. The slide are kept in a humid chamber for 1 hour. The primary antibody is diluted with blocking buffer (CG3613, 4RAP: 1:100; MSL1: 1:200; MLE: 1:200; H14: 1:100) and 50µL of this dilution is applied on top of the slides after the blocking buffer is drained by touching the corner of the slide with a Kim-Wipe. The slide is covered with a 20x44mm coverslip, placed in a humid chamber and kept in the cold room overnight. The slides are then washed three times with PBST for 10 minutes each time. The fluorescently labeled secondary antibody is diluted with blocking buffer (1:200) and 50µL of it is applied to the slides similar to the primary antibody. Incubation was for 2 hours in the dark. The slides are then washed 3 times with PBST, 10 minutes each and twice with PBS for 5 minutes each, in dark. Then, the DNA is stained with Hoechst solution (1xPBS, 0.2µg/mL Hoechst 33258) for 30 minutes. Finally the slides are washed twice with PBS for 5 minutes each. Two-three drops of Fluromount G is added on top of the tissue, and a coverslip was placed upon the slide. After 2-3 hours, the coverslip is sealed with nail polish. Images were captured with an AxioCamHR CCD camera on a Leica SP5 (Leica Microsystems) using an Apochromat NA 1.32 oil immersion objective. Images were arranged with Adobe Illustrator.

Fractionation of S2 cells

Wild-type S2 cells grown in S2 cell medium + 10% FCS were pelleted by spinning them at 500g for 5 minutes. The pellet is washed once with PBS and pelleted again. All the steps that follow should performed in cold. The cell pellet is re-suspended with 10 pellet volume of HLB (10mM HEPES-KOH, 10mM KCl, 1,5 mM $MgCl_2$, 1X Protease Inhibitors) and kept on ice for 15 minutes. Then, 1/10 volume of 10% Igepal CA 630 is added to this suspension and immediately vortexed at full-speed for 30 seconds. Nuclei is pelleted by centrifugation at 2000g for 2 minutes at 4˚C. Supernatant can be saved as cytoplasmic extract, but it is necessary to increase the salt concentration to 140mM NaCl and to centrifuge the extract at maximum speed for 10 minutes to remove insoluble material before storage. The nuclear pellet is quickly re-suspended in 5 cell volume of HLB (half of the initial amount) to remove the detergent and spun again at 2000g for 2 minutes at 4˚C. The supernatant is discarded and the nuclear pellet is re-suspended in extraction buffer at 150mM monovalent salt (25mM HEPES-KOH, 140mM NaCl, 10mM KCl, 1,5mM $MgCl_2$, 0.1% TritonX, 0.2mM EDTA, 25% Glycerol, 1XProtease Inhibitors). Once the nuclei are properly re-suspended with the help of a micropipette, the salt concentration is raised using a stock of 5M NaCl (up to 240-420mM) and extraction is carried out by rotating the tubes in the cold room for 30 minutes. When desired, RNase A is added to the solution at this stage at a final concentration of 1µg/µL. Following the extraction the samples are centrifuged at maximum speed for 15 minutes to remove the nuclear debris and insoluble material. Supernatant is saved as the nuclear extract (or nucleoplasm when the salt is kept at 150mM). The pellet is re-suspended in 1XLaemmli sample buffer, sonicated briefly to breake the genomic DNA and stored as the "chromatin fraction".

Whole gland immunostaining

Flies with the genotype *P*{hsFLP}1, y^1 w^{1118}; *P*{AyGAL4}25 *P*{UAS_GFP.S65T}T2 were crossed to RNAi flies from Vienna that have inverted repeats of RNA constructs targeting CG3613 under the control of a UAS promoter. 12 hour old embryos are heat

shocked at 37˚C for 1hr in an air incubator. This heat shock leads to the expression of FLP recombinase, which removes the yellow gene that seperates a GAL4 gene and an actin promoter, leading to GAL4 expression randomly in some cells. These cells start to express GFP and also knock-down CG3613 as both are under the control of UAS promoters.

Wandering 3rd instar larvae were dissected in PBS by cutting the animal in half and inverting the anterior half to expose salivary glands. These inverted larvae are fixed with 4% formaldehyde, freshly prepared from paraformaldehyde by dissolving the powder in PBS. After 15 minutes of fixation at room temperature the larvae are washed extensively with PBS+0.1% TritonX. The glands were permeabilized with PBS + 0.4% Triton X for 30 min, changing the buffer every 10 minutes. Blocking was with PBS + 3% BSA + 0.4% TritonX for 1hr at room temperature. Primary antibody (CG3613, 4RAP) is added to the blocking solution at 1:100 dilution, and the glands are kept at 4˚C overnight. Excess antibody is washed with PBS+0.4% Triton X, for 3 times 10 minutes each. The TRITC conjugated goat anti-rabbit secondary antibody (Molecular Probes) was used at 1:200 dilution in the blocking buffer mentioned above for 2 hours at room temperature. Larvae are washed three times with PBS+0.4% TritonX and once with PBS for 10 minutes each and then incubated with PBS+10µg/mL Hoechst 33258 (Invitrogen) for 10 minutes. After two more washes with PBS, the larvae are allowed to settle in Fluromount G (Southern Biotech) overnight in the cold room or several hours at room temperature. Finally salivary glands are dissected on a glass slide and the coverslip is sealed with nail polish before microscopy. All the steps involving fluorescently labeled secondary antibody were carried out in dark.

Quantitative Real-Time PCR (qPCR)

Immunoprecipitated material (DNA from ChIP samples or RNA from RIP samples) and total RNA samples reverse-transcribed by random hexamers were analyzed by quantitative PCR using SYBR Green Master Mix and Applied Biosystems 7500 Fast Real-Time PCR system. For ChIP samples, 10% and 1% INPUT material was used to

calculate PCR efficiency and cycle differences are normalized using this information. $2^{\Delta\Delta Ct}$ method was used to calculate recoveries and fold-changes.

RNAi in S2 cells, total RNA isolation, reverse transcription

Double stranded RNA was generated using PCR products with T7 promoter sequences attached to both forward and reverse primers as templates and Ribomax Large Scale T7 *in vitro* transcription system (Promega) according to manufacturer's recommendations. RNA was purified using Megaclear columns (Ambion). To generate double stranded RNA, purified RNA is made to 250mM NaCl and incubated at 70 ˚C for 10 minutes, then slowly cooled down to room temperature.

10-20µg of RNA was transfected to 1-2 million cells in 6-well dishes using RNAiMAX transfection reagent (Invitrogen). 3 days later, media is removed from the plates, cells are washed with PBS and lysed with Buffer RLT on the plate. The lysate is homogenized with Qiashredder columns (Qiagen) and total RNA is purified using RNeasy kit (with on-column DNAse digestion, Qiagen). Total RNA is reverse transcribed using Superscript III (Invitrogen) and random hexamers as primers (Invitrogen). The cDNA is diluted with water and used for qPCR.

Primer pairs used to amplify IVT templates:

CG5792	L: GCATCCTTCTCGTCGATCTC
	R: CTCCTGGCTCCTATCGAGTG
CG5787	L: CTCCTGGCTCCTATCGAGTG
	R: AACAGGAAACGCATACGGTC
CG3613 (exon)	L: GTGCTTCTTCTCGAGCGTCT
	R: AATCGGTGATGCCGATTTAC
CG3613 (3'UTR)	L: TTTTCCAAAACCCCAAGAAA
	R: TCGGACAACAGTTGCAATAAA
GFP	L: TGAAGTTCATCTGCACCACC
	R: AGTTCACCTTGATGCCGTTC

The T7 promoter sequence (ATATAATACGACTCACTATAGGG) precedes each primer.

Primer pairs used in RT-qPCR:

CG3613	ATCCAGCAAAACCAAGCAAC	**Suv4-20 (int)**	GTCTGCAAGGATGAACAGCA
	AGATGAATGTGATCCCGAGC		GATTGGGTTTGGTGGAATTG
Pfk	CTGAGGGCAAGTTCAAGGAG	**Arm (exon)**	TCAGGTGGTAGTTTCCCAGG
	AAGCCACCAATGATCAGGAG		TTGCTGTTAGAGATGGCACG
Sda	TTTGGAGAGCATTGCTGTTG	**Arm (intron)**	AGAGTGGGGTGAGGTTGTTG
	GAAATCTGTGTTGCAATGCG		CAACAACAACCGAAATGACG
Socs16D	AGCTTTAGCATCATGCCACC	**Pcx (exon)**	AACAAATCGTCTCTGGTGGG
	TGGAACTCCACTATCATCGC		CTCTTAATGCGCCTCACTCC
Ucp4a	CTCCATTTGGATTTGCACCT	**Pcx (intron)**	GCTAAGTTTGGATTGCCCTG
	GGCAAGGAGTTCACACAGAA		CTGCATATAGGGTTTGCCGT
Pepck	CATTGCGTGGATGAAGTTTG	**Pyk (exon)**	AATGGTGAAGAAGCCACGTC
	TTGGTCTCCATTGAGGTTCC		CCAGACAACATGACGCAATC
Hsp22	CGAGCTAAAGGTCAAGGTGC	**PyK (intron)**	ACCATCTACGATGAGGCACC
	TCTGATTTTCCCTCCACCAG		TTTCCTCCCCAATGACTCTG
Hsp26	AAGGATGGATTCCAGGTGTG	**Pgk (exon)**	AGCTGGGCGATGTCTATGTC
	ACCAAGATGGAGTCGTCCAC		TTGTTCAGCAACAGACCAGC
Hsp67	CATCGAAGAGGAGCAAAAGG	**Pgk (intron)**	ATAAACATTGCCCGTGCTTC
	TGGAGATTTCCGGTTTCTTG		TCTGGTTGCTGGTGATCTTG
Mof	AGGAGGGCGTAATCGGTAGT	**Thor (exon)**	TGGAAGATAATCCCGAGCAC
	CCCAATAGCTGCGATAGCTC		GAGTAAACGCTTCTTTGCCG
roX2	TCGCAATGCAAACTGAAGTC	**Thor (intron)**	GCTAAGATGTCCGCTTCACC
	AGGCGCGTAAAACGTTACC		AACCTTCCTGGTGATCATGG
Suv4-20 (ex)	TGGGGAACAAGACTTTCTGG	**Hexokinase**	AGTGTGTACCGCTTCCATCC
	CTAAACCTAACCGGCAGCAG		ATCAGATCGAAGGTGATGCC

Generation of the CG3613 stable cell line

A cDNA clone containing full-length CG3613 was obtained from GeneCore, EMBL (clone id: GH05812). This clone is digested with EcoRI-HpaI and cloned into EcoRI-XhoI (blunt) cut pIBU1.C2 (a custom made plasmid that contains MtnA promoter to drive gene expression, an SV40 terminator and a neomycine resistance cassette) in frame with 3XFLAG and 6His tags at the C-terminus. The cloning was verified by sequencing.

This plasmid is transfected into S2 cells by Effectene reagent (Qiagen). 24hrs after the transfection, the cells were split into three (1:2, 1:5 and 1:10 dilutions) and let to recover for 24 hours. The cells were put under selective pressure by adding 1µg/µl geneticin (Gibco) in fresh media. The media is replaced every two days until there were no living cells left in mock transfected dishes. The cells that contained CG3613 stably integrated into their genome were expanded and kept under 0.5µg/µl geneticin.

Phosphatase treatment of cell extracts

Wild-type S2 cells were fractionated as described above with the exception that after incubating the nuclei in extraction buffer at 150mM salt for 20 minutes in cold room, the nuclei are pelleted by centrifugation at 10,000g for 2 minutes and the pellet is re-suspended in extraction buffer at 420mM salt and incubated in the cold room for another 20 minutes. All three fractions were made to 150mM NaCl, 2mM $MnCl_2$ and 1X Phosphatase Buffer (NEB). λ-Phosphatase (NEB) was added to 4U/µL for "+ Phosphatase samples" and omitted for "- Phosphatase samples". Samples are incubated at 30˚C for 15 minutes. Reactions were stopped by addition of 4XLaemmli buffer to 1X and boiling at 90˚C for 5 minutes. Samples are separated by PAGE and analyzed by immunoblotting.

RNAi in flies

The fly lines containing an inducible UAS-RNAi construct targeting CG5792 (Stock#: 34143), CG5787 (Stock#: 35043) and CG3613 (Stock#: 26332) were obtained from Vienna Drosophila RNAi Center. RNAi was induced by crossing these flies with a tubGAL4/TM6 line (a gift from Anne Ephrussi). As a control, w^{1118} flies were crossed to tubGAL4/TM6 line. Adult flies were counted as they eclosed from pupae. The flies were kept at 25˚C and 70% humidity.

6. References

1. Houseley, J. and D. Tollervey, *The many pathways of RNA degradation.* Cell, 2009. **136**(4): p. 763-76.
2. Hallacli, E. and A. Akhtar, *X chromosomal regulation in flies: when less is more.* Chromosome Res, 2009. **17**(5): p. 603-19.
3. Roeder, R.G., *Transcriptional regulation and the role of diverse coactivators in animal cells.* FEBS Lett, 2005. **579**(4): p. 909-15.
4. Wahl, M.C., C.L. Will, and R. Luhrmann, *The spliceosome: design principles of a dynamic RNP machine.* Cell, 2009. **136**(4): p. 701-18.
5. Davey, C.A., et al., *Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution.* J Mol Biol, 2002. **319**(5): p. 1097-113.
6. Smith, E. and A. Shilatifard, *The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes.* Mol Cell, 2010. **40**(5): p. 689-701.
7. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
8. Varga-Weisz, P.D. and P.B. Becker, *Regulation of higher-order chromatin structures by nucleosome-remodelling factors.* Curr Opin Genet Dev, 2006. **16**(2): p. 151-6.
9. Consortium, I.H.G.S., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.
10. Wierzbicki, A.T., J.R. Haag, and C.S. Pikaard, *Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes.* Cell, 2008. **135**(4): p. 635-48.
11. Grummt, I., *Wisely chosen paths--regulation of rRNA synthesis: delivered on 30 June 2010 at the 35th FEBS Congress in Gothenburg, Sweden.* FEBS J, 2010. **277**(22): p. 4626-39.
12. Mayer, C., et al., *Intergenic transcripts regulate the epigenetic state of rRNA genes.* Mol Cell, 2006. **22**(3): p. 351-61.

13. Santoro, R., et al., *Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans.* EMBO Rep, 2010. **11**(1): p. 52-8.
14. Dieci, G., et al., *The expanding RNA polymerase III transcriptome.* Trends Genet, 2007. **23**(12): p. 614-22.
15. Geiduschek, E.P. and G.A. Kassavetis, *The RNA polymerase III transcription apparatus.* J Mol Biol, 2001. **310**(1): p. 1-26.
16. Suganuma, T. and J.L. Workman, *Crosstalk among Histone Modifications.* Cell, 2008. **135**(4): p. 604-7.
17. Malik, S. and R.G. Roeder, *The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation.* Nat Rev Genet, 2010. **11**(11): p. 761-72.
18. Nechaev, S. and K. Adelman, *Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation.* Biochim Biophys Acta, 2011. **1809**(1): p. 34-45.
19. Dreyfuss, G., et al., *hnRNP proteins and the biogenesis of mRNA.* Annu Rev Biochem, 1993. **62**: p. 289-321.
20. Hocine, S., R.H. Singer, and D. Grunwald, *RNA processing and export.* Cold Spring Harb Perspect Biol, 2010. **2**(12): p. a000752.
21. Moteki, S. and D. Price, *Functional coupling of capping and transcription of mRNA.* Mol Cell, 2002. **10**(3): p. 599-609.
22. Sims, R.J., 3rd, R. Belotserkovskaya, and D. Reinberg, *Elongation by RNA polymerase II: the short and long of it.* Genes Dev, 2004. **18**(20): p. 2437-68.
23. Beelman, C.A. and R. Parker, *Degradation of mRNA in eukaryotes.* Cell, 1995. **81**(2): p. 179-83.
24. Izaurralde, E., et al., *A nuclear cap binding protein complex involved in pre-mRNA splicing.* Cell, 1994. **78**(4): p. 657-68.
25. Cheng, H., et al., *Human mRNA export machinery recruited to the 5' end of mRNA.* Cell, 2006. **127**(7): p. 1389-400.
26. Proudfoot, N.J., A. Furger, and M.J. Dye, *Integrating mRNA processing with transcription.* Cell, 2002. **108**(4): p. 501-12.
27. Shatkin, A.J. and J.L. Manley, *The ends of the affair: capping and polyadenylation.* Nat Struct Biol, 2000. **7**(10): p. 838-42.

28. Schwer, B., X. Mao, and S. Shuman, *Accelerated mRNA decay in conditional mutants of yeast mRNA capping enzyme.* Nucleic Acids Res, 1998. **26**(9): p. 2050-7.
29. Kim, M., et al., *The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II.* Nature, 2004. **432**(7016): p. 517-22.
30. Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing.* Annu Rev Biochem, 2003. **72**: p. 291-336.
31. Grabowski, P.J. and D.L. Black, *Alternative RNA splicing in the nervous system.* Prog Neurobiol, 2001. **65**(3): p. 289-308.
32. Henkin, T.M., *Transcription termination control in bacteria.* Curr Opin Microbiol, 2000. **3**(2): p. 149-53.
33. Proudfoot, N., *New perspectives on connecting messenger RNA 3' end formation to transcription.* Curr Opin Cell Biol, 2004. **16**(3): p. 272-8.
34. West, S. and N.J. Proudfoot, *Transcriptional termination enhances protein expression in human cells.* Mol Cell, 2009. **33**(3): p. 354-64.
35. Wu, W.H., et al., *Mutational analysis of yeast TFIIB. A functional relationship between Ssu72 and Sub1/Tsp1 defined by allele-specific interactions with TFIIB.* Genetics, 1999. **153**(2): p. 643-52.
36. Dantonel, J.C., et al., *Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA.* Nature, 1997. **389**(6649): p. 399-402.
37. Ansari, A. and M. Hampsey, *A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping.* Genes Dev, 2005. **19**(24): p. 2969-78.
38. Vasudevan, S. and S.W. Peltz, *Nuclear mRNA surveillance.* Curr Opin Cell Biol, 2003. **15**(3): p. 332-7.
39. Hammell, C.M., et al., *Coupling of termination, 3' processing, and mRNA export.* Mol Cell Biol, 2002. **22**(18): p. 6441-57.
40. Carroll, K.L., et al., *Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts.* Mol Cell Biol, 2004. **24**(14): p. 6241-52.
41. Thiebaut, M., et al., *Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance.* Mol Cell, 2006. **23**(6): p. 853-64.
42. Kohler, A. and E. Hurt, *Exporting RNA from the nucleus to the cytoplasm.* Nat Rev Mol Cell Biol, 2007. **8**(10): p. 761-73.

43. Svejstrup, J., *Keeping RNA and DNA apart during transcription.* Mol Cell, 2003. **12**(3): p. 538-9.
44. Abruzzi, K.C., S. Lacadie, and M. Rosbash, *Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes.* EMBO J, 2004. **23**(13): p. 2620-31.
45. Rehwinkel, J., et al., *Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster.* Nat Struct Mol Biol, 2004. **11**(6): p. 558-66.
46. Huang, Y. and J.A. Steitz, *SRprises along a messenger's journey.* Mol Cell, 2005. **17**(5): p. 613-5.
47. Houseley, J., J. LaCava, and D. Tollervey, *RNA-quality control by the exosome.* Nat Rev Mol Cell Biol, 2006. **7**(7): p. 529-39.
48. Tourriere, H., K. Chebli, and J. Tazi, *mRNA degradation machines in eukaryotic cells.* Biochimie, 2002. **84**(8): p. 821-37.
49. Bousquet-Antonelli, C., C. Presutti, and D. Tollervey, *Identification of a regulated pathway for nuclear pre-mRNA turnover.* Cell, 2000. **102**(6): p. 765-75.
50. Galy, V., et al., *Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1.* Cell, 2004. **116**(1): p. 63-73.
51. Le Hir, H., A. Nott, and M.J. Moore, *How introns influence and enhance eukaryotic gene expression.* Trends Biochem Sci, 2003. **28**(4): p. 215-20.
52. Maquat, L.E., W.Y. Tarn, and O. Isken, *The pioneer round of translation: features and functions.* Cell, 2010. **142**(3): p. 368-74.
53. Chang, Y.F., J.S. Imam, and M.F. Wilkinson, *The nonsense-mediated decay RNA surveillance pathway.* Annu Rev Biochem, 2007. **76**: p. 51-74.
54. Hilleren, P., et al., *Quality control of mRNA 3'-end processing is linked to the nuclear exosome.* Nature, 2001. **413**(6855): p. 538-42.
55. Saguez, C., J.R. Olesen, and T.H. Jensen, *Formation of export-competent mRNP: escaping nuclear destruction.* Curr Opin Cell Biol, 2005. **17**(3): p. 287-93.
56. Chen, C.Y., et al., *AU binding proteins recruit the exosome to degrade ARE-containing mRNAs.* Cell, 2001. **107**(4): p. 451-64.

57. Wyers, F., et al., *Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase.* Cell, 2005. **121**(5): p. 725-37.

58. Doudna, J.A. and T.R. Cech, *The chemical repertoire of natural ribozymes.* Nature, 2002. **418**(6894): p. 222-8.

59. Kapranov, P., et al., *RNA maps reveal new RNA classes and a possible function for pervasive transcription.* Science, 2007. **316**(5830): p. 1484-8.

60. Amaral, P.P., et al., *The eukaryotic genome as an RNA machine.* Science, 2008. **319**(5871): p. 1787-9.

61. Ponjavic, J., C.P. Ponting, and G. Lunter, *Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs.* Genome Res, 2007. **17**(5): p. 556-65.

62. Prasanth, K.V. and D.L. Spector, *Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum.* Genes Dev, 2007. **21**(1): p. 11-42.

63. Costa, F.F., *Non-coding RNAs: Meet thy masters.* Bioessays, 2010. **32**(7): p. 599-608.

64. Payer, B. and J.T. Lee, *X chromosome dosage compensation: how mammals keep the balance.* Annu Rev Genet, 2008. **42**: p. 733-72.

65. Latos, P.A. and D.P. Barlow, *Regulation of imprinted expression by macro non-coding RNAs.* RNA Biol, 2009. **6**(2): p. 100-6.

66. Wassarman, D.A. and J.A. Steitz, *Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function.* Mol Cell Biol, 1991. **11**(7): p. 3432-45.

67. Hogg, J.R. and K. Collins, *RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation.* RNA, 2007. **13**(6): p. 868-80.

68. Nagano, T., et al., *The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin.* Science, 2008. **322**(5908): p. 1717-20.

69. Stricker, S.H., et al., *Silencing and transcriptional properties of the imprinted Airn ncRNA are independent of the endogenous promoter.* EMBO J, 2008. **27**(23): p. 3116-28.

70. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.* Cell, 2007. **129**(7): p. 1311-23.

71. Espinoza, C.A., et al., *B2 RNA binds directly to RNA polymerase II to repress transcript synthesis.* Nat Struct Mol Biol, 2004. **11**(9): p. 822-9.
72. Espinoza, C.A., J.A. Goodrich, and J.F. Kugel, *Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription.* RNA, 2007. **13**(4): p. 583-96.
73. Wang, X., et al., *Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription.* Nature, 2008. **454**(7200): p. 126-30.
74. Willingham, A.T., et al., *A strategy for probing the function of noncoding RNAs finds a repressor of NFAT.* Science, 2005. **309**(5740): p. 1570-3.
75. Pandey, R.R., et al., *Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation.* Mol Cell, 2008. **32**(2): p. 232-46.
76. Kotake, Y., et al., *Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene.* Oncogene, 2010.
77. Yap, K.L., et al., *Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a.* Mol Cell, 2010. **38**(5): p. 662-74.
78. Deng, X. and V.H. Meller, *roX RNAs are required for increased expression of X-linked genes in Drosophila melanogaster males.* Genetics, 2006. **174**(4): p. 1859-66.
79. Lanz, R.B., et al., *A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex.* Cell, 1999. **97**(1): p. 17-27.
80. Sanchez-Elsner, T., et al., *Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax.* Science, 2006. **311**(5764): p. 1118-23.
81. O'Gorman, W., et al., *Analysis of U1 small nuclear RNA interaction with cyclin H.* J Biol Chem, 2005. **280**(44): p. 36920-5.
82. Feng, J., et al., *The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator.* Genes Dev, 2006. **20**(11): p. 1470-84.
83. Shamovsky, I., et al., *RNA-mediated response to heat shock in mammalian cells.* Nature, 2006. **440**(7083): p. 556-60.

84. Lefevre, P., et al., *The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription.* Mol Cell, 2008. **32**(1): p. 129-39.
85. Tian, D., S. Sun, and J.T. Lee, *The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation.* Cell, 2010. **143**(3): p. 390-403.
86. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011.
87. Orom, U.A., et al., *Long noncoding RNAs with enhancer-like function in human cells.* Cell, 2010. **143**(1): p. 46-58.
88. Yang, P.K. and M.I. Kuroda, *Noncoding RNAs and intranuclear positioning in monoallelic gene expression.* Cell, 2007. **128**(4): p. 777-86.
89. Mattick, J.S. and I.V. Makunin, *Non-coding RNA.* Hum Mol Genet, 2006. **15 Spec No 1**: p. R17-29.
90. Mariner, P.D., et al., *Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock.* Mol Cell, 2008. **29**(4): p. 499-509.
91. Beniaminov, A., E. Westhof, and A. Krol, *Distinctive structures between chimpanzee and human in a brain noncoding RNA.* RNA, 2008. **14**(7): p. 1270-5.
92. Kanhere, A., et al., *Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2.* Mol Cell, 2010. **38**(5): p. 675-88.
93. Zhao, J., et al., *Genome-wide identification of polycomb-associated RNAs by RIP-seq.* Mol Cell, 2010. **40**(6): p. 939-53.
94. Kaneko, S., et al., *Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA.* Genes Dev, 2010. **24**(23): p. 2615-20.
95. Charlesworth, B., *The evolution of chromosomal sex determination and dosage compensation.* Curr Biol, 1996. **6**(2): p. 149-62.
96. Leeb, M., P.A. Steffen, and A. Wutz, *X chromosome inactivation sparked by non-coding RNAs.* RNA Biol, 2009. **6**(2): p. 94-9.
97. Plath, K., et al., *Role of histone H3 lysine 27 methylation in X inactivation.* Science, 2003. **300**(5616): p. 131-5.
98. Sparmann, A. and M. van Lohuizen, *Polycomb silencers control cell fate, development and cancer.* Nat Rev Cancer, 2006. **6**(11): p. 846-56.

99. Amrein, H. and R. Axel, *Genes expressed in neurons of adult male Drosophila.* Cell, 1997. **88**(4): p. 459-69.

100. Meller, V.H., et al., *roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system.* Cell, 1997. **88**(4): p. 445-57.

101. Franke, A. and B.S. Baker, *The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila.* Mol Cell, 1999. **4**(1): p. 117-22.

102. Gilfillan, G.D., et al., *Chromosome-wide gene-specific targeting of the Drosophila dosage compensation complex.* Genes Dev, 2006. **20**(7): p. 858-70.

103. Smith, E.R., et al., *The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation.* Mol Cell Biol, 2000. **20**(1): p. 312-8.

104. Kageyama, Y., et al., *Association and spreading of the Drosophila dosage compensation complex from a discrete roX1 chromatin entry site.* EMBO J, 2001. **20**(9): p. 2236-45.

105. Park, Y., et al., *Sequence-specific targeting of Drosophila roX genes by the MSL dosage compensation complex.* Mol Cell, 2003. **11**(4): p. 977-86.

106. Stuckenholz, C., V.H. Meller, and M.I. Kuroda, *Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in Drosophila melanogaster.* Genetics, 2003. **164**(3): p. 1003-14.

107. Park, S.W., M.I. Kuroda, and Y. Park, *Regulation of histone H4 Lys16 acetylation by predicted alternative secondary structures in roX noncoding RNAs.* Mol Cell Biol, 2008. **28**(16): p. 4952-62.

108. Turner, B., et al., *Induced fit of RNA on binding the L7Ae protein to the kink-turn motif.* RNA, 2005. **11**(8): p. 1192-200.

109. Meller, V.H., *Initiation of dosage compensation in Drosophila embryos depends on expression of the roX RNAs.* Mech Dev, 2003. **120**(7): p. 759-67.

110. Kelley, R.L., et al., *Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila.* Cell, 1995. **81**(6): p. 867-77.

111. Meller, V.H., *Dosage compensation: making 1X equal 2X.* Trends Cell Biol, 2000. **10**(2): p. 54-9.

112. Akhtar, A., D. Zink, and P.B. Becker, *Chromodomains are protein-RNA interaction modules.* Nature, 2000. **407**(6802): p. 405-9.

113. Meller, V.H., et al., *Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in Drosophila.* Curr Biol, 2000. **10**(3): p. 136-43.

114. Kuroda, M.I., et al., *The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila.* Cell, 1991. **66**(5): p. 935-47.

115. Lee, C.G., et al., *The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation.* EMBO J, 1997. **16**(10): p. 2671-81.

116. Richter, L., J.R. Bone, and M.I. Kuroda, *RNA-dependent association of the Drosophila maleless protein with the male X chromosome.* Genes Cells, 1996. **1**(3): p. 325-36.

117. Shimojo, H., et al., *Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain.* J Mol Biol, 2008. **378**(5): p. 987-1001.

118. Mendjan, S., et al., *Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila.* Mol Cell, 2006. **21**(6): p. 811-23.

119. Gu, W., et al., *Targeting the chromatin-remodeling MSL complex of Drosophila to its sites of action on the X chromosome requires both acetyl transferase and ATPase activities.* EMBO J, 2000. **19**(19): p. 5202-11.

120. Pannuti, A. and J.C. Lucchesi, *Recycling to remodel: evolution of dosage-compensation complexes.* Curr Opin Genet Dev, 2000. **10**(6): p. 644-50.

121. Meller, V.H. and B.P. Rattner, *The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex.* EMBO J, 2002. **21**(5): p. 1084-91.

122. Bhadra, U., M. Pal-Bhadra, and J.A. Birchler, *Role of the male specific lethal (msl) genes in modifying the effects of sex chromosomal dosage in Drosophila.* Genetics, 1999. **152**(1): p. 249-68.

123. Copps, K., et al., *Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly.* EMBO J, 1998. **17**(18): p. 5409-17.

124. Scott, M.J., et al., *MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in Drosophila.* EMBO J, 2000. **19**(1): p. 144-55.

125. Oh, H., Y. Park, and M.I. Kuroda, *Local spreading of MSL complexes from roX genes on the Drosophila X chromosome.* Genes Dev, 2003. **17**(11): p. 1334-9.

126. Deng, X., et al., *The severity of roX1 mutations is predicted by MSL localization on the X chromosome.* Mech Dev, 2005. **122**(10): p. 1094-105.

127. Kelley, R.L., et al., *Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin.* Cell, 1999. **98**(4): p. 513-22.

128. Fagegaltier, D. and B.S. Baker, *X chromosome sites autonomously recruit the dosage compensation complex in Drosophila males.* PLoS Biol, 2004. **2**(11): p. e341.

129. Demakova, O.V., et al., *The MSL complex levels are critical for its correct targeting to the chromosomes in Drosophila melanogaster.* Chromosoma, 2003. **112**(3): p. 103-15.

130. Gilfillan, G.D., I.K. Dahlsveen, and P.B. Becker, *Lifting a chromosome: dosage compensation in Drosophila melanogaster.* FEBS Lett, 2004. **567**(1): p. 8-14.

131. Mendjan, S. and A. Akhtar, *The right dose for every sex.* Chromosoma, 2007. **116**(2): p. 95-106.

132. Bell, O., et al., *Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster.* Mol Cell Biol, 2008. **28**(10): p. 3401-9.

133. Larschan, E., et al., *MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism.* Mol Cell, 2007. **28**(1): p. 121-33.

134. Sural, T.H., et al., *The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila melanogaster X chromosome.* Nat Struct Mol Biol, 2008. **15**(12): p. 1318-25.

135. Kind, J. and A. Akhtar, *Cotranscriptional recruitment of the dosage compensation complex to X-linked target genes.* Genes Dev, 2007. **21**(16): p. 2030-40.

136. Alekseyenko, A.A., et al., *A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome.* Cell, 2008. **134**(4): p. 599-609.

137. Kelley, R.L., O.K. Lee, and Y.K. Shim, *Transcription rate of noncoding roX1 RNA controls local spreading of the Drosophila MSL chromatin remodeling complex.* Mech Dev, 2008. **125**(11-12): p. 1009-19.

138. Park, S.W., et al., *An evolutionarily conserved domain of roX2 RNA is sufficient for induction of H4-Lys16 acetylation on the Drosophila X chromosome.* Genetics, 2007. **177**(3): p. 1429-37.

139. Izzo, A., et al., *Structure-function analysis of the RNA helicase maleless.* Nucleic Acids Res, 2008. **36**(3): p. 950-62.

140. Reenan, R.A., C.J. Hanrahan, and B. Ganetzky, *The mle(napts) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na+ channel transcript in a region of RNA editing.* Neuron, 2000. **25**(1): p. 139-49.

141. Lee, C.G., et al., *RNA helicase A is essential for normal gastrulation.* Proc Natl Acad Sci U S A, 1998. **95**(23): p. 13709-13.

142. Oh, H., et al., *Regulation of inter- and intramolecular interaction of RNA, DNA, and proteins by MLE.* Methods Mol Biol, 2010. **587**: p. 303-26.

143. Zhang, S. and F. Grosse, *Molecular characterization of nuclear DNA helicase II (RNA helicase A).* Methods Mol Biol, 2010. **587**: p. 291-302.

144. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. **17**(10): p. 1030-2.

145. Einhauer, A. and A. Jungbauer, *The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins.* J Biochem Biophys Methods, 2001. **49**(1-3): p. 455-65.

146. Field, J., et al., *Purification of a RAS-responsive adenylyl cyclase complex from Saccharomyces cerevisiae by use of an epitope addition method.* Mol Cell Biol, 1988. **8**(5): p. 2159-65.

147. Duncan, K., et al., *Sex-lethal imparts a sex-specific function to UNR by recruiting it to the msl-2 mRNA 3' UTR: translational repression for dosage compensation.* Genes Dev, 2006. **20**(3): p. 368-79.

148. Czaplinski, K., et al., *Identification of 40LoVe, a Xenopus hnRNP D family protein involved in localizing a TGF-beta-related mRNA during oogenesis.* Dev Cell, 2005. **8**(4): p. 505-15.

149. Robb, G.B. and T.M. Rana, *RNA helicase A interacts with RISC in human cells and functions in RISC loading.* Mol Cell, 2007. **26**(4): p. 523-37.

150. Herold, N., et al., *Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes.* Mol Cell Biol, 2009. **29**(1): p. 281-301.

151. Worringer, K.A., F. Chu, and B. Panning, *The zinc finger protein Zn72D and DEAD box helicase Belle interact and control maleless mRNA and protein levels.* BMC Mol Biol, 2009. **10**: p. 33.

152. Rajan, P., et al., *Regulation of gene expression by the RNA-binding protein Sam68 in cancer.* Biochem Soc Trans, 2008. **36**(Pt 3): p. 505-7.

153. Zhao, R., et al., *Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation.* J Cell Biol, 2008. **180**(3): p. 563-78.

154. Gonzales, F.A., et al., *Characterization of Saccharomyces cerevisiae Nop17p, a novel Nop58p-interacting protein that is involved in Pre-rRNA processing.* J Mol Biol, 2005. **346**(2): p. 437-55.

155. Hill, R.J. and G.T. Rudkin, *Polytene chromosomes: The status of the band–interband question.* Bioessays, 1987. **7**(1): p. 35-40.

156. Derry, J.J., et al., *Sik (BRK) phosphorylates Sam68 in the nucleus and negatively regulates its RNA binding ability.* Mol Cell Biol, 2000. **20**(16): p. 6114-26.

157. Zhuo, S., et al., *Expression, purification, crystallization, and biochemical characterization of a recombinant protein phosphatase.* J Biol Chem, 1993. **268**(24): p. 17754-61.

158. Straub, T., et al., *The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex.* PLoS Genet, 2008. **4**(12): p. e1000302.

159. Zobeck, K.L., et al., *Recruitment timing and dynamics of transcription factors at the Hsp70 loci in living cells.* Mol Cell, 2010. **40**(6): p. 965-75.

160. Cristofari, G. and J.L. Darlix, *The ubiquitous nature of RNA chaperone proteins.* Prog Nucleic Acid Res Mol Biol, 2002. **72**: p. 223-68.

161. Rajkowitsch, L., et al., *RNA chaperones, RNA annealers and RNA helicases.* RNA Biol, 2007. **4**(3): p. 118-30.

162. Herschlag, D., *RNA chaperones and the RNA folding problem.* J Biol Chem, 1995. **270**(36): p. 20871-4.

163. Fauth, T., et al., *The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome.* Nucleic Acids Res, 2010. **38**(10): p. 3209-21.

164. Fasken, M.B. and A.H. Corbett, *Process or perish: quality control in mRNA biogenesis.* Nat Struct Mol Biol, 2005. **12**(6): p. 482-8.

165. Shevchenko, A., M. Wilm, and M. Mann, *Peptide sequencing by mass spectrometry for homology searches and cloning of genes.* J Protein Chem, 1997. **16**(5): p. 481-90.

166. Selth, L.A., C. Gilbert, and J.Q. Svejstrup, *RNA immunoprecipitation to determine RNA-protein associations in vivo.* Cold Spring Harb Protoc, 2009. **2009**(6): p. pdb prot5234.

167. Johansen, K.M., et al., *Polytene chromosome squash methods for studying transcription and epigenetic chromatin modification in Drosophila using antibodies.* Methods, 2009. **48**(4): p. 387-97.

Acknowledgments

To begin with, I would like to express my gratitude to Asifa for giving me the opportunity to work in her lab. Her continuous support and foresight helped me to overcome many hurdles throughout this study .

I would also like to thank members of my TAC, Matthias Hentze, Juerg Mueller and Ed Hurt for their guidance and feedback.

Kent Duncan and Jan Medenbach shared their time and plasmids with me during the early parts of my PhD. They would take a break from their work and discuss my experiments with me, something I will never forget. I would also like to thank Maxim Nekrasov for teaching me how to handle millions of flies and Oguz Kanca for showing me how a curly wing looks like.

From my first day in the lab., until the day I submit my thesis there was always Herbert, relentlessly showing where T4 ligase is, how to purify a His-tagged protein, how to stain a gel etc etc… It is impossible to thank you enough Herbert.

Without the heated discussions in the lab. this thesis would lack a few experiments. I would like to thank all members of the Akhtar lab., but especially Thomas C., Erinc and Sunil.

Without the continuous support of my family, I would probably not be able carry out this work. I am especially grateful to my father who pushed me into biology when it was clear that I was not cut out for neither mathematics nor physics. It was my mother who made me culturally diverse and shaped me into the person I am today. Finally, I would like to thank my sister, Asli, who had to bear with my persona for so many years, for translating this part to my parents.

Last but not least, my wife Tugce gives meaning to this work and to my life. Hopefully, she will never stop dreaming and we will continue to exist in this world of figments.