The role of *roX* RNA

in Dosage Compensation during

Drosophila melanogaster embryogenesis

vorgelegt von
Khairunnadiya Prayitno
aus
Newcastle upon Tyne
2019

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Betreuer:	Prof. Dr. rer. nat. Peter B. Becker
Zweitgutachter:	Prof. Dr. rer. nat. Andreas Ladurner

Dekan:	Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung:	05.06.2020

I. Table of Contents

The role of *roX* RNA in Dosage Compensation during *Drosophila melanogaster* embryogenesis

II. Eidesstattliche Erklärung

Prayitno, Khairunnadiya

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Thema

**"The role of *roX* RNA in Dosage Compensation
during *Drosophila melanogaster* embryogenesis"**

selbstständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 10.06.2020   Khairunnadiya Prayitno
_____   _____
 Ort, Datum   Khairunnadiya Prayitno

VIII

III.   Preface

Part of this dissertation has been published with Tamas Schauer as a co-author in a research article, titled "Progressive dosage compensation during *Drosophila* embryogenesis is reflected by gene arrangement", in EMBO reports with PMID 31286660. This includes data presented in chapters 6.3.1 until 6.3.4, whereby ChIP-seq, and IF experiments were performed by me, and RNA-seq experiments were performed by Tamas Schauer. Tamas Schauer and I then performed the Bioinformatics analyses together.

x

IV. List of Figures

V. List of Tables

XIV

VI. Abbreviations

ac	acetylation
ael	after egg laying
AGO	Argonaute
ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
°C	degree Celsius
CBD	chromo barrel domain
CD	chromodomain
cDNA	complementary DNA
CES	chromatin entry site
CF	cytoplasmic fraction
Cl.8	*Drosophila* Clone 8 cell line, male karyotype
CLAMP	chromatin-linked adaptor for MSL proteins
DAPI	4',6-diamidino-2-phenylindole
DCC	dosage compensation complex
DCR	Dicer
DNA	deoxyribonucleic acid
DNase I	deoxyribonuclease I
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetate
EGTA	ethyleneglycol-bis-(2-aminoethyl)-N,N,N,N-tetraacetate
FCS	fetal calf serum
gDNA	genomic DNA
GFP	green fluorescent protein
gp	guinea pig
H3	histone 3
H4	histone 4
HAS	high affinity site
HAT	histone acetyltransferase
HDAC	histone deacetylase
HEPES	N-(2-hydroxyethyl)piperazine-H'-2-ethanesulfonic acid
HMT	histone methyltransferase
IF	immunofluorescence
IP	immunoprecipitation
K	lysine

Kc	*Drosophila* cell line, female karyotype
lncRNA	long non-coding RNA
me	methylation
mRNA	messenger RNA
miRNA	microRNA
MLE	maleless
MOF	males-absent-on-the-first
MRE	MSL response element
ms	mouse
MSL	male-specific-lethal
MZT	maternal-zygotic transition
nc	nuclear cycle
NDS	normal donkey serum
NE	nuclear extract
NGS	normal goat serum
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PionX	pioneering sites on the X
PMSF	phenylmethanesulfonyl fluoride
PEG	polyethylene glycol
PFA	paraformaldehyde
pre-miRNA	precursor microRNA
pri-miRNA	primary microRNA
PTM	post-translational modification
rb	rabbit
RBP	RNA-binding protein
RNA	ribonucleic acid
RNAi	RNA interference
RNAse H	ribonuclease H
RNP complex	ribonucleoprotein complex
roX	RNA-on-the-X
RT	room temperature
S2	*Drosophila* Schneider cell line, male karyotype
sd	standard deviation
SDS	sodium dodecyl sulfate
shRNA	short hairpin RNA
SL	stem-loop

SN	supernatant
SXL	sex lethal
TAD	topologically associating domain
TSS	transcription start site
TTS	transcription termination site
UAS	upstream activating sequence
UTR	untranslated region
WB	Western blot
ZGA	zygotic genome activation

Bismillah.

For Him, my family and friends.

xx

1 Summary

Dosage compensation (DC) in male *Drosophila melanogaster* flies is done through hypertranscription of the X chromosome. This involves the dosage compensation complex (DCC), a ribonucleoprotein complex of five protein subunits, Male-specific-lethal 1 (MSL1), MSL2, MSL3, Males-absent-on-the-first (MOF) and Maleless (MLE), and long noncoding RNA, *RNA-on-the-X* (*roX*), encoded by either *roX1* or *roX2* gene.

DC is interlinked with the process of sex determination. A hypothesis suggests that upon hybridization of *roX1* and *roX2* RNAs, a miRNA is produced that is implicated in a feedback mechanism of sex determination. Different approaches were used to reproduce hybridization and validate putative miRNA; however, such observations could not be seen.

As differential function of *roX* RNAs have been proposed, characterization of *roX1* and *roX2* RNAs in fractionated extracts were done by rt-qPCR. Long isoforms of *roX*, *roX1-RE* and *roX2-RB*, tended to be polyadenylated and enriched in the cytoplasm suggesting differential post-transcriptional processing and possible shuttling mechanism. A preliminary experiment of direct-RNA nanopore sequencing detected major parts of *roX* RNAs important for DC. With improved protocol of RNA preservation and library preparation, it may prove to be a potent tool to further characterize the lncRNAs and profile its isoforms.

Additionally, a detailed study on the establishment of dosage compensation during early embryogenesis was done. MSL2 binding to DNA was evident 4 hours after egg laying when least compensation of X-linked genes is observed. Concurrent detection of MOF on the X chromosome signified assembly of DCC in early development. This complex was active in its function to acetylate H4K16. Nevertheless, accumulation of H4K16ac on the X chromosome proceeded in a time- and space-dependent manner, coinciding with the progression of dosage compensation. Specifically, genes defined as constitutive were closer to DCC binding sites, more acetylated, and first compensated. Meanwhile, genes characterized as developmental were farther from DCC binding sites, lowly acetylated, and slowly compensated.

2 Zusammenfassung

Die Dosiskompensation bei männlichen *Drosophila melanogaster* Fliegen erfolgt durch Übertranskription des X-Chromosoms. Dies wird durch den Dosiskompensationskomplex (DCC) ermöglicht. Dieser Ribonukleoproteinkomplex besteht aus fünf Proteinuntereinheiten, Male-specific-lethal 1 (MSL1), MSL2, MSL3, Males-absent-on-the-first (MOF) und Maleless (MLE), und einer langen nichtkodierenden RNA, *RNA-on-the-X* (*roX*), die entweder durch das *roX1*- oder das *roX2*-Gen kodiert wird.

Die Dosiskompensation ist mit dem Prozess der Geschlechtsdetermination verbunden. Die Hypothese, dass bei Hybridisierung von *roX1*- und *roX2*-RNAs eine miRNA erzeugt wird, die an einem Rückkopplungsmechanismus der Geschlechtsdetermination beteiligt ist, wurde in dieser Arbeit getestet. Leider konnten vorangegangene, hypothesenstützende Beobachtungen nicht reproduziert werden.

Für die *roX*-RNAs wurde eine redundante Rolle in der Dosiskompensation sowie zusätzliche Funktionen außerhalb dieses Prozesses vorgeschlagen. Diese Dissertation umfasst die Charakterisierung von *roX1*- und *roX2*-RNAs in fraktionierten Embryoextrakten durch RT-qPCR. Die Ergebnisse deuten auf eine differenzielle posttranskriptionale Verarbeitung der RNAs hin. Lange Isoformen von *roX*, *roX1-RE* und *roX2-RB* sind polyadenyliert. Darüber hinaus sind sie im Zytoplasma angereichert, was auf einen möglichen Austausch mit dem Nucleus hindeutet. Definierende Abschnitte der für DC wichtigen *roX*-RNAs wurden in einem explorativen Experiment durch *direct-RNA nanopore* Sequenzieren nachgewiesen. Mit einem verbesserten Protokoll zur Extraktion der RNA, sowie deren Konservation und Bibliotheksvorbereitung könnte es sich als wirksames Instrument zur weiteren Charakterisierung der langen nichtkodierenden RNA und auch in Bezug auf die Selektion der RNA-Isoformen erweisen.

Zudem wurde im Rahmen dieser Arbeit eine detaillierte Studie zur Etablierung der Dosiskompensation während der frühen Embryogenese durchgeführt. Die Bindung von MSL2 an DNA war bereits 4 Stunden nach der Eiablage messbar. Zu diesem Zeitpinkt ist nur eine geringe Dosiskompensation von X-Chromosom gekoppelten Genen vorhanden. Gleichzeitig ließ sich ebenfalls MOF bereits in derselben Region auf dem X-Chromosom nachweisen. Dies zeigte die Bildung des DCC in der frühen Embryogenese. Der Komplex war bereits aktiv und acetylierte H4K16. Trotzdem verlief die Akkumulation von H4K16ac auf dem X-Chromosom nachfolgend zeit- und positionsabhängig, übereinstimmend mit dem Fortschreiten der Dosiskompensation. So wurden Gene, die als konstitutiv definiert wurden und näher an DCC-

Bindungsstellen lagen, stärker acetyliert und zuerst kompensiert. Als Entwicklungsgene definierte Gene hingegen, lagen weiter von DCC-Bindungsstellen entfernt, wurden nur schwach acetyliert und langsam kompensiert.

3 Introduction

3.1 Drosophila melanogaster

Drosophila, familiarly known as fruit flies, are human commensal species usually attracted to rotten fruits. The first reports on the use of *Drosophila melanogaster (D. melanogaster)* as a model organism emerged in the early 20th century. Due to the ease of chromosome analysis and the assortment of phenotypical markers available to follow crossing events, *Drosophila* proved to be a powerful organism to study genetics (1,2). Since then, *D. melanogaster* has established itself to be a versatile tool for a wide range of research from the most fundamental, such as molecular biology, to the most applicable, such as behavioural sciences and drug discovery (3). It presents simplified, robust, and expeditious variety of techniques while maintaining relevance to larger and more complex systems. For example, the genome editing tool CRISPR-Cas9 system that is first discovered as part of the prokaryotic immune system has been customized to the fly system extensively that nowadays there are 3699 guide RNA stocks combinable with 33 Cas9 or 47 UAS-dCas9 stocks available to create tissue-specific overexpression or knockout of genes. One of the greatest aspects of *D. melanogaster* is its detailed biology known to the scientific community and the vast adaptable methods available to track and manipulate it (3). Hence, it is inevitable that *D. melanogaster* is utilized as a model organism in many kinds of proof-of-principle studies.

3.1.1 Embryogenesis

D. melanogaster develops from a fertilized egg to an adult in a span of nine to ten days at a temperature of 25°C. During this time, embryogenesis takes ~24 hours (hrs) to accomplish, after which the animal progresses through three instar larval stages, termed first (~24 hrs), second (~24 hrs) and third (~48 hrs). In the next pupal stage (~five days) the animal undergoes intense metamorphosis followed by eclosion, where adult flies emerge and are ready to close the circle of life, i.e. mate, within ~eight to twelve hours (Fig. 1) (3).

The outward appearance of *D. melanogaster* embryo during development has been well described since 1970s as no special instrument other than a light microscope and a quick 5-minute preparation to remove the chorion suffice. To this end, Bownes published a detailed stage-by-stage description of the first 22 hours of development, which is still widely used as a reference to sort embryos into stages (4). To complement the information of external

characteristics, corresponding internal morphological changes has been described by Foe and colleagues (5).
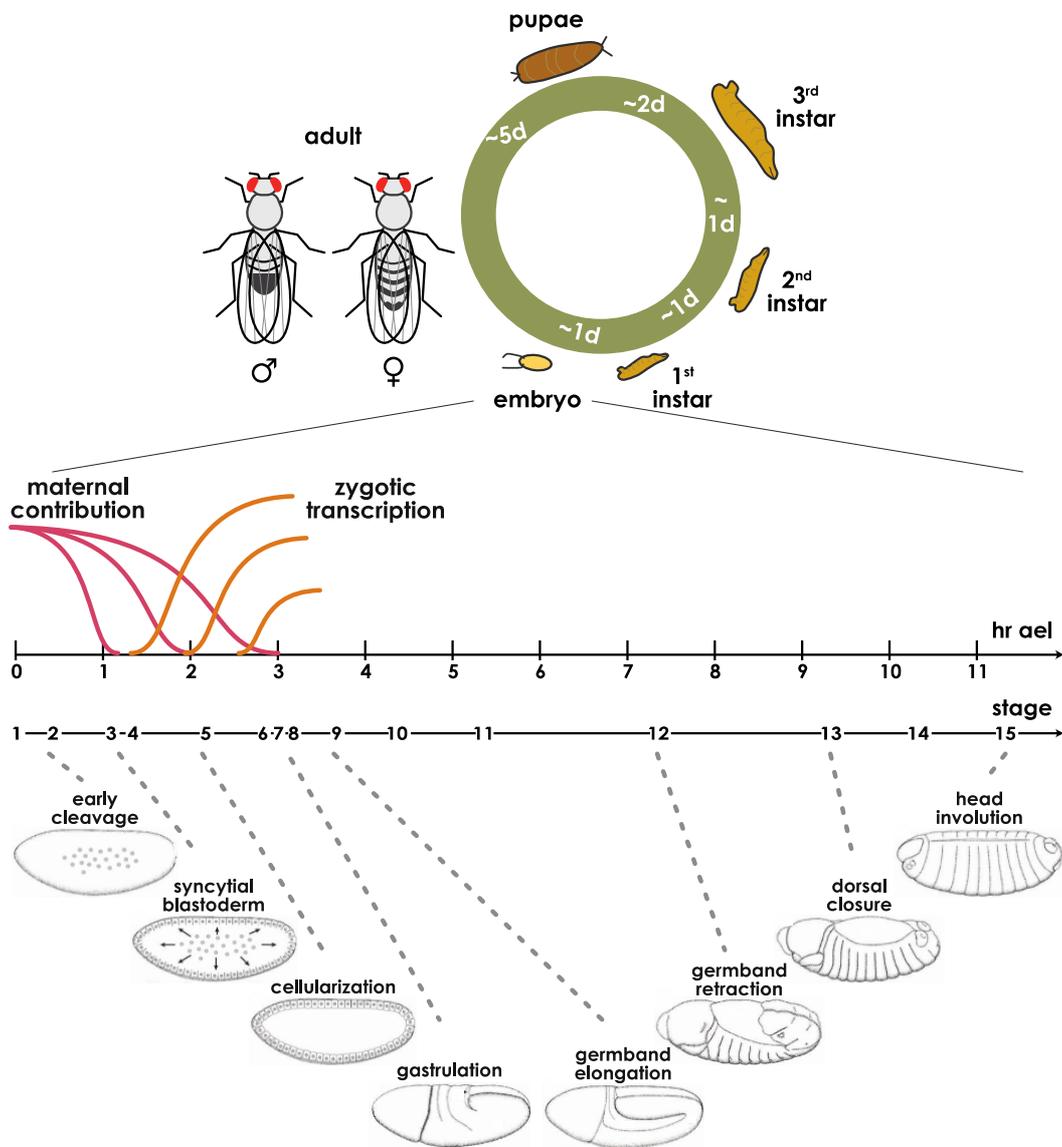
Figure 1. Life cycle of *Drosophila melanogaster* from embryogenesis to adulthood taking roughly 10 days to complete. Embryo images are adapted from the Atlas of *Drosophila* Development and used with permission (10).

Upon fertilization, maternal and paternal pronuclei merge and set off 13 nuclear cycles (nc) of synchronous division without cytokinesis. The nuclei are only subjected to S- and M-phases in these first two hours after egg laying (ael; stage 1-4) (6,7). ~6000 nuclei share a common cytoplasm to form a syncytial blastoderm before they undergo a final division (nc 14) accompanied by cellularization to form a cellular blastoderm (~2.5 hrs ael; stage 5). Gastrulation then occurs where the three layers of endoderm, mesoderm and ectoderm arise

(~3 hrs ael; stage 6-7). For the next 4 hours (stage 8-11), the embryo elongates its germ band to almost double the length to prepare for the next phase of segmentation that occurs as the germ band retracts (~7-10 hrs ael; stage 12-13). The next phase of development consists of head involution and dorsal closure (~10-13 hrs ael; stage 14-15), followed by small outward change, but intensive internal cellular differentiation to form the various organs (~13-24 hrs ael; stage 16-17) (Fig. 1).

Like many other organisms in which embryogenesis takes place *ex utero*, rapid development of *D. melanogaster* embryo is necessary for the survival of the animal. One solution to ensure sufficient resources is the deposition of maternal components into the developing oocyte. Maternal mRNAs, proteins, including translational machinery components, and nutrients are abundantly loaded and drive maturation of the egg and replication of the zygotic genome while it is still quiescent. It is roughly estimated that 65% of the *Drosophila* protein-coding transcriptome is supplied and post-transcriptionally regulated (7). Nonetheless, a transition must occur in which the maternal control on embryogenesis subsides and is replaced by zygotic components. This is a process termed Maternal-Zygotic-Transition (MZT) and includes a phenomenon named Zygotic Genome Activation (ZGA) (8,9). Although MZT and ZGA are gradual processes, zygotic transcription is in full gear by the time embryos reach gastrulation stage (~2.5 hrs ael; stage 5).

3.1.2 Maternal-Zygotic transition

During the initial phases of embryogenesis in which zygotic genome is still silenced, maternal mRNAs is tightly controlled by a set of RNA-binding proteins (RBPs). Directions include mRNA localization, translation efficiency, and poly(A)-tail lengths as well as mRNA stability, repression and degradation. This all translates to the patterning of the embryo that leads to its polarization required for further development. An example of a positive regulator is the protein Staufen. It is positively conserved and is involved in spatiotemporal control of maternal mRNA localization and protein production (11).

Two processes must conspire to achieve a successful transition from maternally controlled to zygotically dictated development. First is maternal clearance, i.e. the removal of maternal instructions, which is critical as many maternally deposited components are required to drive early growth but detrimental for later development of embryo (12). This process relies on tightly regulated activation of miRNA-mediated gene silencing, whereby RNA-binding proteins (RBPs) acting as translational repressor cooperatively invoke maternally and zygotically directed mRNA decay (13). Maternally directed decay is orchestrated by *Smaug*

(*Smg*), *Brain tumor* (*Brat*), and *Pumilio* (*Pum*). Their transcripts are maternally deposited and translated upon activation of the unfertilized egg (13). They initiate decay by recruiting deadenylation complexes and conserved degradation machinery to a subset of maternal transcripts (7). *Pum* and *Brat* also regulate zygotically directed mRNA decay, however, this phase accordingly relies on some zygotic transcription to produce a particular cluster of miRNA (*miR-309*) that instructs removal of maternal mRNAs specifically through the 3' UTR targeting (14). All in all, 25% and 35% of cleared transcripts are exclusive targets of the maternal and zygotic machineries, respectively, while 40% transcripts are cleared by both acting in concert (9,12).

The second determinant of a successful hand-over from maternal to zygotic control of development is the establishment of zygotic transcriptome (15). Two waves of ZGA set up its full activity, the minor and major waves (8,9). The minor wave of transcription occurs during the rapid nuclear replication of early embryos from nc 8 on. A small subset of short intronless genes of *Drosophila* are biasedly expressed during the limited time between each of the nuclear cycles (7,16). The expression profiles of these genes are thought to be driven by two factors, nuclear to cytoplasmic ratio (N:C) and time elapsed since fertilization (17,18). As nuclei amplify exponentially without an increase in embryo volume, the N:C ratio progressively changes. Experimentation with increased and decreased ratios suggest that with each division, maternal components are titrated down thereby relieving repression on transcriptional activity of target genes (19). Concurrently, N:C-independent genes rely on the maternal clock that times translation of their regulator. One such example is the maternally deposited and master regulator of ZGA *zelda* (*zld*). The transcription factor (TF) Zld is translated at nc 8, after which zygotic expression is observed for a small number of genes, whose enhancers and promoters are highly enriched with Zld binding sites (20-22). Early Zld binding increases chromatin accessibility specifically for Zld-dependent genes and promotes recruitment of Pol II (20). Interestingly, this includes components of the zygotic RNA degradation pathways, i.e. *miR-309*, thereby linking ZGA with events of zygotically directed maternal clearance (23). Although additional Zld binding sites are found on other regulatory regions of zygotic genes, their expression does not commence until the major wave of transcription is initiated pointing towards a concentration-dependent activation (7).

Several changes occur that signify the onset of the major wave of ZGA. In addition to Zld-dependent expression of genes, Zld appears to facilitate association of additional transcription and pioneering factors to maintain an open chromatin state. Among others, the GAGA Factor

(GAF) encoded by *Trithorax-like* (*Trl*) has been reported to cooperate with Zld in potentiating the major wave transcription (24). GAF maintains nucleosome-free regions and recruits Pol II to promoters that are both dependent and independent of Zld (25,26). Correspondingly, Pol II is majorly recruited at nc 13 to many Transcription Start Sites (TSSs). Nevertheless, their activity is kept poised to repress inappropriate expression of developmental genes yet prime them for expression at later stages (20,24,26,27). Accordingly, chromatin architecture that is largely loose and unstructured before ZGA undergoes progressive remodelling and maturation from being relatively decondensed with equal probability of contact across the genome to acquiring short-range topologically associating domain (TAD) structures and long-distance compartmentalization (28-30). Once established, these chromatin conformation and TAD boundaries are retained throughout development and in adulthood. Remarkably, regions consisting of house-keeping genes show pre-disposition for these boundaries even earlier and independent of transcription, although gene expression is still needed to refine organization (28). At the onset of nc 14, cellularization of nuclei and prolonged cell cycle length permit the introduction of a G2 gap phase, alleviating the disruptive forces of DNA replication on transcription (7). And so, transcription of the zygotic genome trickles in to replace maternal control over development.

The interplay of time and space coordination between destabilization of maternal mRNA and zygotic transcription of early developmental genes designates anteroposterior and dorsoventral axes necessary for proper patterning of the embryo. Indeed, this process involves the earliest regulator of mRNA decay, Smg, and the master regulator of ZGA, Zld, through their function in regulating a plethora of patterning genes (31-35).

3.1.3 Sex determination

In *D. melanogaster*, as is in *H. sapiens*, males are heterogametic for the sex chromosomes X and Y, whereas females are homogametic, XX. Unlike in human, however, sex determination is done early in embryogenesis and does not rely on the presence of a Y chromosome but rather on the number of X chromosome. With this distinction in mind, a mechanism to "count" the X chromosome must occur to determine sex specificity (3). Counting of X relies on the X-linked sex regulator gene, *Sex-lethal* (*Sxl*). In females, the double dose of X allows for enough SXL translation upon ZGA. In a positive feedback loop, SXL regulates its own alternative splicing to further stabilize female-specific transcripts (36), as well as splicing of additional factors involved in strengthening the female-specific gene expression, morphology, and behaviour, i.e. *transformer* (*tra*), *doublesex* (*dsx*) and *fruitless* (*fru*) (Fig. 2).

Measurable *Sxl* transcripts can be categorized into three classes: early-female, late-female, and late-male specific. The early promoter of *Sxl*, also known as the "establishment promoter" (*SxlPe*), transiently expresses early-female transcripts. This burst of expression commences at nc 12, when minor wave of ZGA allows X-linked signal elements (XSE), a group of four X-coded proteins (Ascute, sisA, Runt, Unpaired), to reach a threshold concentration to preclude a maternal Sxl negative regulator, Groucho (Gro), and activate *SxlPe* specifically in females (36-39). A switch in promoter choice occurs at the cellular blastoderm stage (~2.5 hrs), in parallel with the major wave of ZGA chain of events. The late promoter of *Sxl*, termed the "maintenance promoter" (*SxlPm*), produces *Sxl* mRNA that are spliced into either male- or female-specific isoforms. The male isoform includes exon 3, which contains a STOP codon, and therefore is translated into a truncated, non-functional protein (36). On the contrary, exon 3 is spliced out in the female isoform, allowing for functional production of Sxl (36) (Fig. 2).
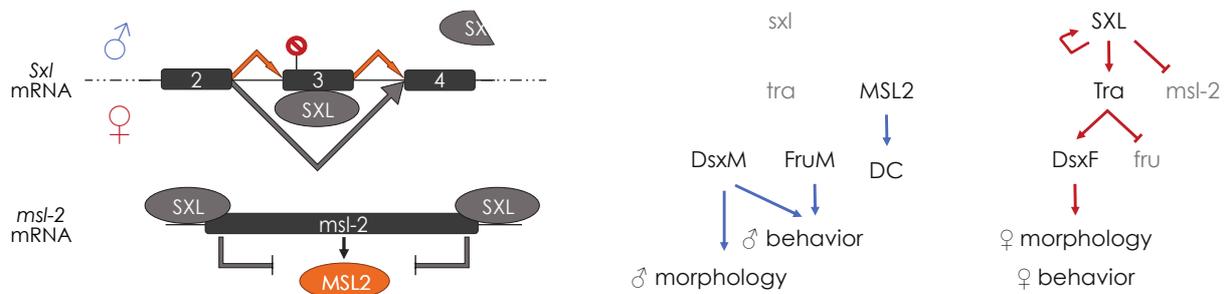
Figure 2. SXL is the master regulator of sex determination and regulates many genes to establish a female phenotype, and its absence drives the male development.

Sxl contains two highly conserved RNA-binding domains, which preferentially target long stretches of poly(U) interrupted with guanine (36). The *Sxl* pre-mRNA itself has such sequences upstream and downstream of exon 3, the male-specific exon, and interaction of functional Sxl with components of the splicing machinery at these sites ensures exon 3 skipping. In early female embryo, Sxl produced from *SxlPe* is sufficient to trigger the loop and the "late" form of female Sxl maintains a positive autoregulatory splicing control on *Sxl* expressed from the *SxlPm*, locking it ON. This regulation is absent in males due to lack of functional Sxl and so only noncoding *Sxl* mRNA with a premature STOP codon is ever produced (36,40). Once activated, Sxl initiates the female-determining program (Fig. 2). It first rescues *Tra* pre-mRNA, whose default splicing pattern encodes a short non-functional protein (41,42). Tra functions to strengthen the female phenotype by prompting female-specific splicing of *Dsx (DsxF)* and *Fru (FruF)* pre-mRNAs (42) (Fig. 2). DsxF is involved in female organ development and tissue differentiation, whereas *FruF* splicing results in a premature

STOP codon and accordingly non-functional protein (43-45). Concurrently, the absence of Tra in males leads to default splicing of *Dsx (DsxM)* and *Fru (FruM)*, where DsxM invokes male morphology and in concert with FruM determines male courtship behaviour in adults (43-46) (Fig. 2). Sxl sits at the top of the sex regulatory chain and its importance is clearly demonstrated by female-specific lethality upon loss of Sxl in XX animals and male-specific lethality upon inappropriate Sxl expression in XY (47,48).

3.1.4 Interlink of sex determination and dosage compensation pathways

Once sex has been determined, specific programs of development are initiated. A challenge unique to males is their state of heterogamy. As genes on the X are equally important for both sexes, dosage compensation of the single X is essential in male animals. The directive to initiate this mechanism is the outcome of the sex determination process. Therefore, as a link between the two pathways, Sxl orchestrates the expression of *male-specific-lethal 2* (*msl-2*), the core component of dosage compensation, multifacetedly (Fig. 2). The *msl-2* mRNA contains several putative Sxl binding sites in its 5' and 3' untranslated region (UTR), two of which are found within a 5' intron (49-52). In the nucleus, binding of Sxl to the 5' UTR of *msl-2* pre-mRNA ensures a specific splicing event in which the intron is retained (49,51,52). Once exported to the cytoplasm, the retained intron serves as a landing platform for Sxl. In combination with Sxl binding at 3'UTR, translational repression is exerted by steric inhibition of ribosomal initiation complex recruitment and start codon recognition (36,50). As females express fully functional SXL, this leak-proof repression continuously occurs. However, the male form of SXL is truncated therefore non-functional and so MSL2 is expressed at steady state and dosage compensation is maintained in males (50,53).

3.2 Dosage Compensation

Dosage compensation (DC) is a mechanism that has evolved to ensure balanced expression of sex-chromosomal gene products. Many species that are sexually dimorphic have adapted different approaches to address this imbalance, from the inactivation of one of two female Xs in *H. sapiens*, to halving the X expression in hermaphrodites of *C. elegans* (47). *D. melanogaster* achieves this balance by approximately increasing the male X expression two-fold (47,48). In cases where DC fails, male-specific lethality is observed and conversely, when DC is induced in females, low viability, sterility and developmental delay is described (54-56).

3.2.1 Dosage Compensation Complex and its components

Dosage compensation in flies is carried out by a complex termed Male-Specific-Lethal (MSL) or DC complex (Fig. 3). The first members of DCC discovered in 1980 were *msl-1*, *msl-2* and *mle* (54,57). A fourth gene, *msl-3*, was added in 1981 (58). And the final protein member of the complex, *males-absent-on-the-first* (*mof*) was linked to DCC much later in 1997 (59). Last but not least, a long non-coding RNA, *RNA-on-the-X (roX) RNA,* completes DCC (60). This ribonucleoprotein complex decorates the single male X chromosome exclusively and enriches it with acetylation at lysine 16 of histone H4 (H4K16ac) (61). The chromosome-wide histone modification enhances chromatin accessibility and supports hyper-transcription, thereby, increasing the production of X-linked genes (62,63) (Fig. 3).
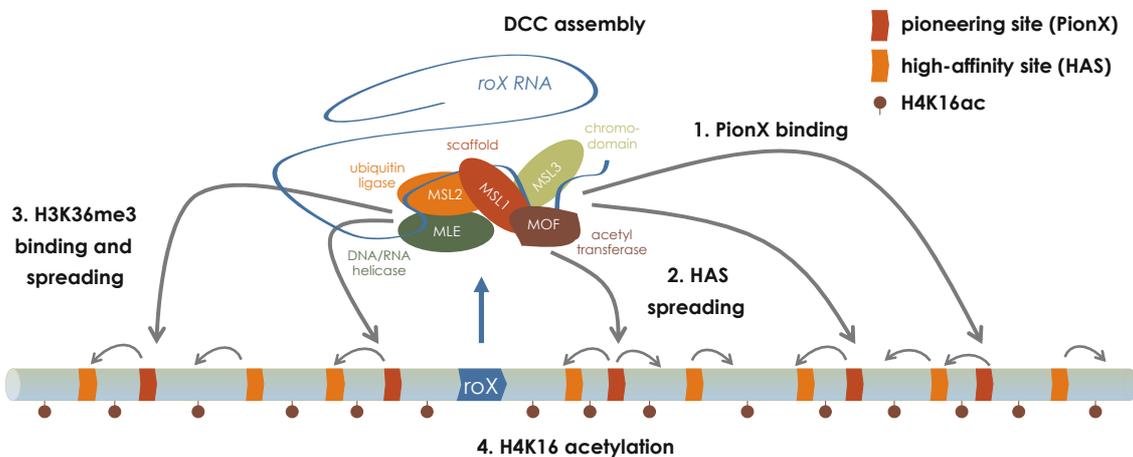
Figure 3. Dosage compensation complex targeting and spreading mechanism.

MSL2 is the only male-specific protein and is the decisive factor of DC. It also functions as an E3 ubiquitin ligase, shown to ubiquitinate itself as well as other Msl proteins to target them for degradation as means of maintaining stoichiometry (64-66). MSL2 interaction with the rest of the complex is mediated by MSL1, the scaffold (67). MSL1 and MSL2 make a core, whose interaction was revealed in a co-immunoprecipitation assay and its structure subsequently resolved (56,64). Whereas the MSL1 N-terminus provides MSL2 interaction, its C-terminus carry binding sites for MSL3 and MOF (64,68,69). MSL3 bears a chromodomain (CD) that can recognize methylated histone H3 at lysine 36 (H3K36me3), a mark associated with active transcription (70-72). An observation of male X-specificity of H4K16ac led to the discovery of MOF, the histone acetyltransferase, that works to ease transcription repression by chromatin (59,73-75). MLE is an RNA helicase with two double-stranded RNA binding motifs

specific for single-stranded RNA or DNA (76,77). It unwinds *roX* RNA to permit incorporation into DCC in an ATP-dependent manner (78,79).

3.2.2 Assembly and targeting of DCC

DCC is a solution to an issue at hand, namely X imbalance. Naturally, very specific mechanism had to be put in place so that an increase in gene expression only effects those that lacked in dose. Therefore, fine tuning of targeting by DCC is as necessary as its assembly.

The current state of literature agrees on an assembly of DCC centered around the transcription of *roX* RNA in a stepwise manner (80) (Fig. 3). As *roX* is being produced, MLE comes in and remodels it to reveal MSL2 binding sites (79,81). The core complex of MSL1/MSL2 joins in, bringing along MSL3 and MOF (64). MSL2's CXC and proline/basic-residue-rich domains equips DCC with targeted direct DNA binding to X-enriched sites termed chromosomal entry sites (CES) or more recently high affinity sites (HAS) (82-87). This binding occurs favourably *in vitro*, however, its placement *in vivo* requires the cooperation of a zinc-finger protein, chromatin-linked adaptor for MSL proteins (CLAMP), to compete with deposition of nucleosomes on sites termed MSL response elements (MREs) (88,89). Interestingly, two of these enhancer elements are within *roX* genes, strengthening the theory that *roX* genes serve as nucleation sites of the complex (82,90). Of the HAS are a subset found to be pioneering sites on the X (PionX) defined by refined DNA sequence and shape (91) (Fig. 3). Once bound to these sites, DCC spreads along the chromosome by various mechanisms. First, H3K36me3 recognition by MSL3 allows spreading of DCC to neighbouring genes that are active and require hyper-transcription (70-72). Second, established 3D chromosomal structure aids long-range interaction between active compartments that enhances spreading DCC (92). Third, by yet an unknown mechanism, *roX* RNA promotes efficient spreading along the chromosome, as evidenced by disrupted H4K16ac pattern in mutant males (93,94). A mapping of *roX* RNA-chromatin interaction has been published, and interestingly, among a majority of common *roX* binding to the X chromosome, there are evidently cell type specific interactions that is reflected in gene expression (95).

3.2.3 Mechanism of dosage compensation

Although many steps can be customized to increase expression, DCC appears to operate at the transcriptional level. Studies on Pol II in the context of DC suggests that enhancing recruitment of the transcription machineries led to a 1.2-fold increase in Pol II activity at promoters of X-linked hyperacetylated genes (96-99). Concurrently, multiple reports provided

evidence for a different mechanism showing that DCC improves transcriptional output by facilitating the progression of Pol II through a direct interaction between MSL1 and SPT5, a transcription elongation factor (100,101). Through improved genome-wide studies, 5' paused Pol II is detected to be equally present when normalized to the gene dose, although the elongating Pol II phosphorylated at serine 2 (Ser2P) is evidently increased over X-linked gene bodies (102,103). This suggests that the key rate-limiting step is the release of paused Pol II, which is overcome by the change in chromatin landscape facilitated by DCC. The 3'-biased H4K16ac, a histone modification that has been shown to decrease chromatin compaction, decreases steric hindrance for a transcribing Pol II, thereby allowing a coordinated increase of X-linked transcription (73,104-106). And so, with an increase in H4K16ac, an increase in Pol II-Ser2P follows.

3.2.4 During development: establishment of dosage compensation

As has been reviewed earlier, dosage compensation is initiated upon conclusion of being "male" in the sex determination pathway. Whereas MSL2 is stably expressed only upon ZGA, other members of DCC, i.e. MSL1, MSL3, MOF, and MLE are maternally contributed (Fig. 4). The assembly of DCC is attained and thereby single X targeting is feasible. Nonetheless, there is a disparity between onset of transcription and detection of DCC by means of immunofluorescence, which due to limited resolution may not show if MSLs are bound to HAS but have not spread onto active genes. Nuclear localization of MSLs is first observed at stage 6 of blastoderm (~3 hrs ael) and accumulation onto X territories only occurs at stage 9, about an hour later (107,108) (Fig. 4). As some X-linked genes are needed for development during this time, an MSL-independent mechanism is thought to persist to ensure that sex chromosome dose difference does not translate into disastrous outcome. An attractive candidate of such regulator turns out to be Sxl as many genes that harbor more than 3 Sxl binding sites in their 3' UTRs are X-linked (53,56,109,110). Indeed, several mRNAs of developmental regulators encoded on the X are present at equal levels in male and female embryos present in this time window (22).
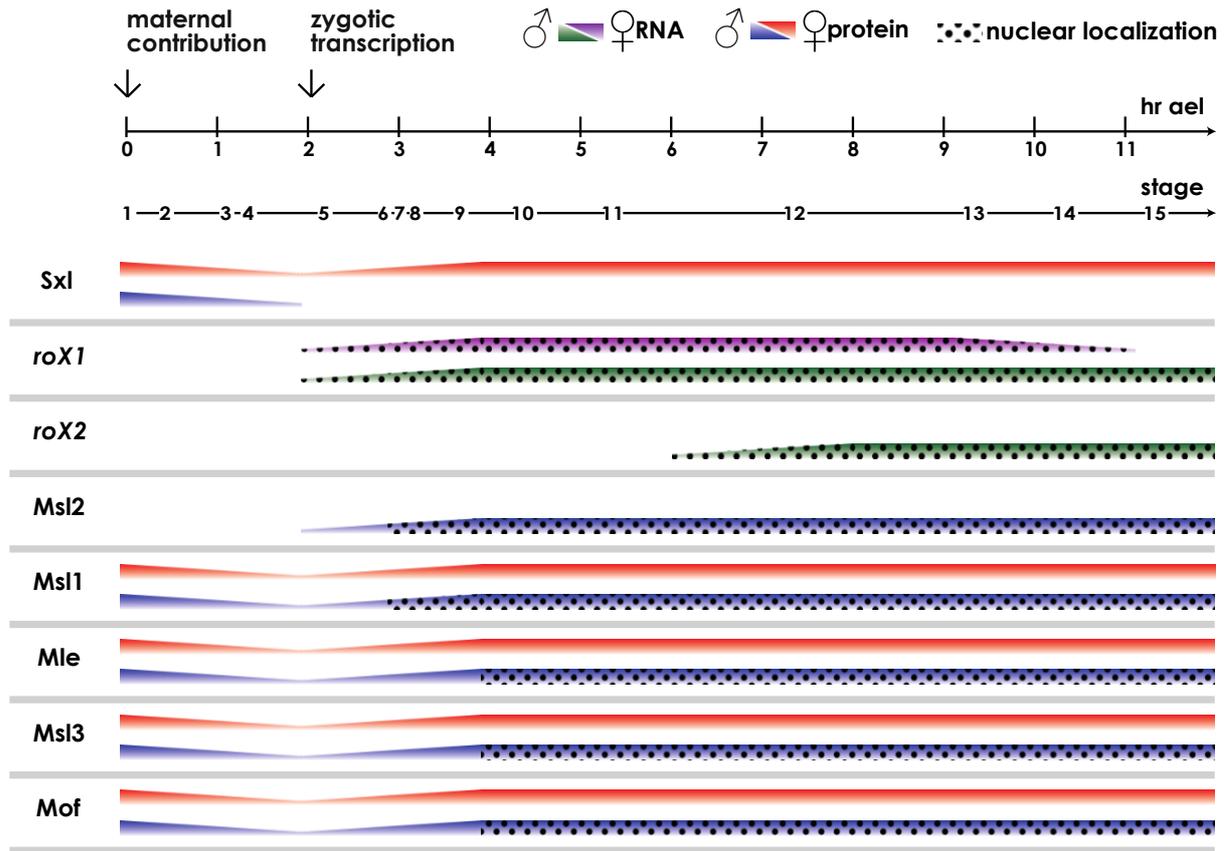
Figure 4. Summarized data of MSL protein and *roX* RNA expression in early embryogenesis.

3.3 lncRNA in Dosage Compensation

In 1972, Susumo Ohno coined the term "junk DNA" to sequences with "the importance of doing nothing" (111), a misnomer that has since been disproven in many studies across many different species (112,113). Although only a small percentage of the genome represents protein-coding sequences, the larger non-coding region is hardly "junk" and is as critical for the maintenance of life. As a matter of fact, these sequences that may seem to be repetitive elements enrich various organisms with adaptive tools in the process of evolution and highlights the dynamicity of genomes (113,114).

Annotation of non-coding elements of genomes has since revealed many hidden features that can act as switches and signals for protein-coding genes. One such class of regulatory elements encode for long non-coding RNA (lncRNA), a stretch of more than 200 nucleotides, that can modulate processes from chromatin organization all the way to post-translational modification (115).

3.3.1 roX RNA in dosage compensation

As the name suggests, *roX* RNA is encoded on the X chromosome and can be produced from either *roX1* or *roX2* gene. They were first discovered in studies of the *Drosophila* brain and were reported back-to-back in 1997 (116,117). *roX* RNA colocalization with MSLs on the X chromosome has been shown in various tissues through assorted methods that range from the most conventional, i.e. in-situ hybridization and immunostaining (116,118), to ones with high resolution and high-throughput, i.e. Chromatin Isolation by RNA purification sequencing (ChIRP-seq (78,119)), Chromatin-associated RNA sequencing (ChAR-seq (120)), and RNA-DamID (95). To elucidate whether the targeting of X chromosome is in part due to the X-linkage of the *roX* genes, autosomal integration of either *roX* genes was performed and showed that *roX* RNA is able to find the X chromosome (116,121,122). As additional *roX* and MSL spreading nearby the insertion site can be seen, it reiterates that *roX* RNA can work *in trans* to specifically target the X chromosome and *in cis* to spread DCC along the X chromosome (121,122). X chromosome composition wins over source of *roX* in the X targeting.

Early genetic studies on single and double mutants of *roX* RNA also reveal a functional redundancy within dosage compensation. Excision mutants of *roX1* did not affect X localization of MSLs and viability, whereas early design of *roX2* deletion did not disrupt MSL binding pattern but lowered viability through collateral deletion of neighbouring genes (116,118,121). Nevertheless, this indicates that only one of either *roX* RNAs is required and/or sufficient for proper DC. And so justifiably, as soon as both *roX* genes were altered, male double mutants do not survive past the third larval instar stage reminiscent of male-specific lethality, although some escaper males were detected (118,121,122). Examination of MSL distribution and transcriptional effect in these *roX* mutants reveal that the primary cause of lethality is the mislocalization of MSLs to ectopic sites on chromocenters and autosomes (121-123) that in turn fail to activate dosage compensation of the X (124).

Using a combination of deletion analysis and fly genetics, Stuckenholz, Park, Kelley, and colleagues identified the minimal requirement for their function to short conserved sequences termed *roX* boxes, GUUNUACG, in the 3' end of *roX* RNAs that are prevalent within stem-loops (SLs) (93,125-127) (Fig. 5). The structure, in addition to sequence, of these stem-loops coined SL*roX1* and SL*roX2* prove to be targets of MSL2 and MLE binding, necessary for the assembly of DCC (78,79). This specific targeting occurs in an ATP-dependent manner whereby the remodeling of *roX* RNA by MLE is required (78,79,128). Although they

seem to exhibit functional redundancy, interesting differences exist as outlined in the next two subchapters.

3.3.2 *roX RNAs and their isoforms*

roX1 and *roX2* RNAs exhibit similar characteristics, in addition to their X-chromosomal localization. Both RNAs only carry small open reading frames (ORFs), whereby prediction yields 59 and 45 amino acids, respectively (117). Their transcripts undergo post-transcriptional processing, including polyadenylation and alternative splicing, albeit lack of reports on nuclear export (117,129).
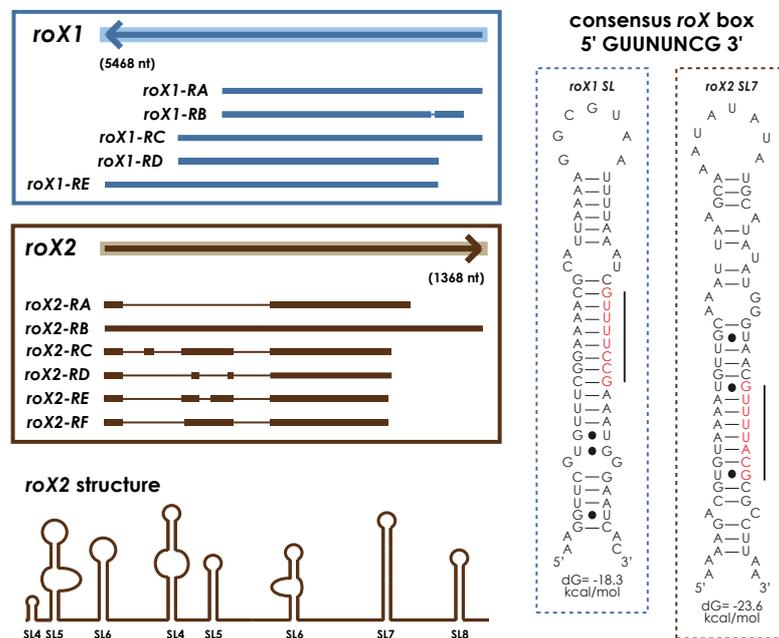
Figure 5. *roX* RNAs and their isoforms. *roX1* has five isoforms, while *roX2* has six. A predicted secondary structure has been hypothesized only for *roX2*; nonetheless, the main interacting stem loop structures have been mapped for both RNAs and consensus *roX* box sequence determined.

Interestingly, that is where the similarities end. *roX1* is a gene of length ~5400 nt (116), and five isoforms (*roX1-RA* to *-RE*), all containing *roX* boxes and roughly ~3700 nt in size, have been annotated (Fig. 5). The *roX2* gene is smaller than its counterpart with a sequence of ~1300 nt (117). And contrarily, *roX2* undergoes more extensive splicing to produce six isoforms (*roX2-RA* to *-RF*), also all containing *roX* boxes, that are largely of ~600 nt in size (Fig. 5). Although the significance and relevance of each individual *roX* isoforms is poorly understood, it has been shown that alternative splicing is crucial in DC as a plethora of *roX2*

isoforms was required to keep a steady-state level of RNA and an accumulation of MSL complex on the X chromosome (129).

Simple alignment of their primary sequence reveals that there is very low conservation between the two RNAs. A recent evolutionary study in diverse *Drosophilid* species across ~40 million years employed a strategy where synteny, microhomology, and secondary structures are integrated to identify orthologs of *roX1* and *roX2 RNA*. The findings reiterate the relevance of conserved structure over primary sequence homology as *roX* orthologs from distantly related species are functional in dosage compensation and are able to rescue failure to compensate across different species (114).

3.3.3 Expression of roX RNAs during development

The first publications on *roX* RNAs have already reported distinctions in profile of expression (Fig. 4). Puncta of *roX1* in nuclei are visible in both sexes during blastoderm formation (as early as 2 hrs ael) with the strongest signal being in neuronal cells (116). Upon germband retraction (~10 hrs ael), *roX1* diminishes specifically in females and its male specificity is evident by the time epidermal segmentation starts (~13 hrs ael) (130). By contrast, *roX2* transcription does not start until stomodeum invagination commences (~6 hrs ael) and its signal is much weaker than that of *roX1* (130). Nevertheless, it does so immediately in a male-specific manner. By the time male embryos reach the third instar larval stage, relative level of *roX1* and *roX2* is equalized (130). Two studies have described developmental transcriptome during embryogenesis and provided data for re-analysis (22,131). One of the studies looked into very early stages of synchronous rapid nuclei division, during which *roX* RNA expression is insignificant (22). The second study examined 2-hour window transcriptome of developing embryos from 0 to 24-hour ael; however, collection of mixed sex was used (131). Nevertheless, it recapitulates earlier reports that *roX1* RNA is transcribed first followed by *roX2* RNA transcription roughly a few hours later (131).

The peculiarities in expression profile impart a difference in early phenotype of single *roX* mutants. Although no delay in development is seen with either single mutants (116,121), *roX1* mutants suspend its MSL localization to the X chromosome until *roX2* is expressed (130). *roX2* mutants, on the other hand, is highly dependent on maternally deposited MLE to stabilize early *roX1* transcripts (130). To note, developmental delay is observed in mutants only expressing any one isoform of *roX2* at endogenous level that can partly be rescued through overexpression (129). This highlights the importance of steady state expression and alternative

splicing of a plethora of *roX* RNA; nonetheless, synthetic lethality of *roX1* and *roX2* demonstrate their redundant role in dosage compensation.

3.3.4 RNA and miRNA biogenesis

Another non-coding regulatory elements that have gained interest in recent decades are microRNA (miRNA). It belongs to the class of small non-coding RNA due to its size of only ~22 nt. Unlike lncRNA, miRNA tend to specifically regulate gene expression at the post-translational level by binding to the 3' UTR of their target mRNA. The base pairing of miRNA induces gene silencing by inhibiting further protein production (132,133). To date, there are 258 *Drosophila* miRNAs annotated and reported in the latest release of miRNA database (miRBase) that are implicated in various biological processes ranging from germline development to neuronal development (134).

miRNA is the final product of many steps of RNA maturation. Upon regulated transcription by RNA Polymerase II (Pol II), the long primary miRNA (pri-miRNA) transcript typically of size ~ 1 kb is processed to unearth the miRNA seed sequence that is embedded within a hairpin structure. To commence, a Microprocessor complex containing Drosha, a nuclear RNase III, crops a ~65 nt SLs out of the capped and polyadenylated pri-miRNA (132). Interestingly, the efficiency of this process can be altered by single nucleotide polymorphisms residing in the miRNA precursors (133). The resultant pre-miRNA is then exported into the cytoplasm for further processing by Dicer-1 (DCR1) to create an RNA duplex (132). Once the RNA duplex is loaded onto Argonaute 1 (AGO1) protein, an effector complex called RNA-induced Silencing Complex (RISC) is formed. Maturation of RISC is finalized with the unwinding and cleavage of the RNA duplex, and depending on strand selection, the guide strand is stabilized whilst the passenger degraded (132). At all stages from transcription to maturation, regulation of miRNA biogenesis occurs by means of processing, stability, feedback loops and modification of the RNA as well as proteins involved in the process (133).

The highly stabilized and mature miRNA-induced RISC complex (miRISC) scans the cytoplasm for complementary mRNA targets, where the seed sequence is of importance and confer specificity. Association of miRISC with mRNA induces gene silencing in two ways. The immediate response of miRISC:mRNA interaction is the inhibition of cap recognition of the eukaryotic initiation factor 4E (eIF4E). This leads to hindrance of the eIF4F complex assembly, of which eIF4E and eIF4G are subunits, at mRNA caps (133). eIF4F is required to recruit the pre-initiation complex and is therefore important for cap-dependent translation (135). The binding has also been shown to induce the disassembly of existing eIF4F complex

by releasing eIF4E and eIF4G from target mRNAs (136). Additionally, miRISC association prompts target mRNA decay. It recruits deadenylation complexes, PAN2-PAN3 and CCR4-NOT, which trim the poly(A) tail of the mRNA. The deadenylated mRNA then undergoes decapping, for which the enzyme Dcp1 is responsible for. Finally, deadenylated and decapped mRNA are degraded by a 5'-to-3' exoribonuclease (XRN1).

The miRNA-induced gene silencing is just another mechanism widely used in various organisms to maintain the steady-state equilibrium as well as to carry out responses to various stimuli, including but not limited to development.

3.4 Chromatin

Even the smallest and simplest organism has a genome longer in length than it can contain, necessitating the packaging of genetic information. Whereas this is done by supercoiling in prokaryotes, eukaryotes do so by packing it into structures called chromatin. The concept of chromatin was formed by cytologists and biochemists in the late 19th century (137,138). Although it was largely neglected in the first half of the 20th century, interests re-emerged with newfound discovery in genetics, specifically with the revelation of the polytene chromosome in *Drosophila* (139,140). Nowadays, it is irrefutable that chromatin biology is as important as the underlying blueprint of life. Every cell of an organism carries the same genetic code, yet these very same stretches of four different nucleotides give rise to many different types of cells. And so many great strides have been undertaken to better understand its role and regulation.

3.4.1 Chromatin structure and organization

The most fundamental unit of chromatin is the nucleosome, which consists of ~147 bp negatively charged DNA winding 1.65 turns around a positively charged histone protein core in a left-handed superhelical manner (141-143) (Fig. 6). The globular histone core is made up of pairs of four histones (H2A, H2B, H3 and H4), each having highly basic tails that protrude outward (141,144,145) (Fig. 6). The entire nucleosome structure is strengthened by histone/histone as well as histone/DNA electrostatic interactions (142,146), and therefore confers a physical barrier for any protein binding and processivity that require "reading" of the underlying genetic blueprint. Linker region of DNA bridges neighbouring nucleosomes, spaced on average ~200 bp apart, and is associated with the linker histone, H1 (147,148). Histone H1 interaction with the histone core around DNA entry and exit sites further stabilizes

the 10-nm nucleosomal array ("beads on a string" (149)) and allows its arrangement into higher-order architecture (148,150,151).
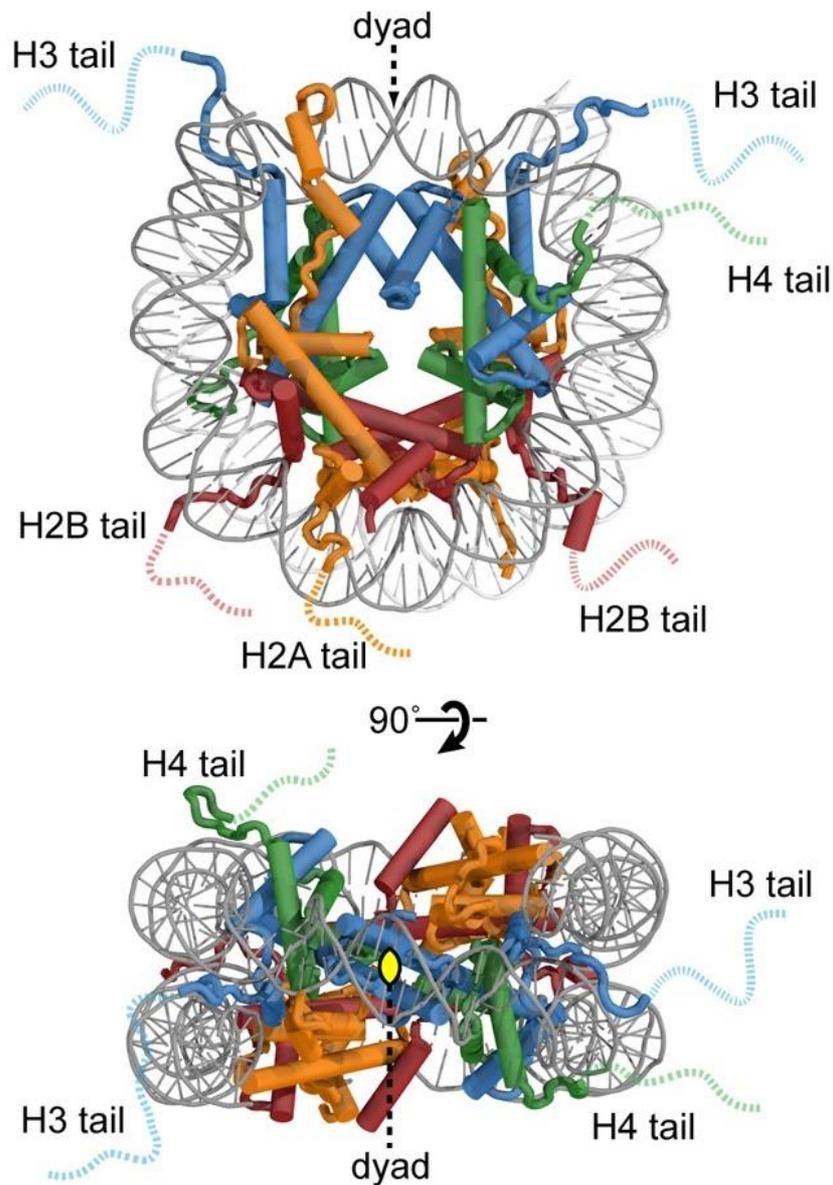
Figure 6. Architecture of the nucleosome core particle. Face and top view of the nucleosome structure upon a 90°C turn. Image is adapted from the "Post-Translational Modifications of Histones that Influence Nucleosome Dynamics" and used with permission (156).

Under low salt condition *in vitro*, nucleosome fibers fold into rods of 30-nm in diameter. *In vivo* studies, however, favour a more irregular model in which the 10-nm array can interdigitate as assisted by cellular cations into globular structures (152). This ensures a degree of dynamicity and accessibility allowing for the formation of TADs and 3D loop structures (152,153). Distant promoters and enhancers can be brought together into active chromatin loops to elicit regulation, e.g. Zelda-dependent transcription, and inactive loops congregate co-

regulated genes to silence them simultaneously, e.g. repressive Polycomb domain (29). Also responsible for the structuring of the chromatin as such are architectural proteins, such as mediator, cohesins and insulators, and potentially non-coding RNAs, which are all crucial for determining gene regulation and henceforth cell identity (154,155).

3.4.2 Histone post-transcriptional modifications (PTMs)

The degree of nucleosomal packing determines accessibility of genes and regulatory elements and therefore augments additional possibilities in refining gene regulation. This property is partly imparted by the largely unstructured and flexible C- and N-terminal tails of histones that carry more than 60 sites at which many types of PTMs can take place, the most common ones being acetylation, methylation, phosphorylation, and ubiquitination of residues arginine (R), lysine (K), serines (s), and threonines (T) (144,145,157,158). To note, limited histone core modification also exists (159). PTMs of the histone tails alter interactions between nucleosomes, DNA/histones, as well as the binding of chaperones, chromatin remodellers, and transcription factors. Deciphering the histone code has led to some modifications being associated with loosening, i.e. acetylation of H3/H4, and others tightening the chromatin compaction, i.e. methylation of H3K9/H3K27 (160-162). Moreover, combinatorial modifications can in some cases trump over individual ones (145,163)). Interestingly, PTMs are not completely independent of each other as deposition or removal of one can promote or inhibit the deposition or removal of another (162,164).

Regulation of PTMs involve a collection of histone modifiers that can be classified as writers, readers or erasers. Writers deposit, erasers remove, and readers recognize to elicit downstream effectors such as other modifiers and remodelling enzymes. Activity of histone modifiers is highly specific, i.e. it usually involves a certain modification of a specific amino acid residue (144,161,164). To add another layer of switch regulation, writers, readers and erasers are tightly controlled.

3.4.3 H4K16ac and H3K36me3 in dosage compensation

Naturally, chromatin structure and its compactness influence all stages of transcription from initiation to pre-mRNA processing. Histone acetylation is by and large associated with active transcription, but in the context of dosage compensation, male-specific H4K16ac is of importance. Acetylation of H4K16 abolishes folding of nucleosomal arrays (105). Indeed, contact was mapped between the basic patch of H4 N-terminus, i.e. residues K16 to R23, and the acidic patch surface of H2A/H2B dimer of a close-by nucleosome and this looping of the

histone tail is critical for chromatin folding (142,165). H4K16ac alters the net charge, weakening inter-/intra-nucleosomal interaction to confer easier displacement of nucleosomes (166-168). In the *Drosophila*, H4K16ac is deposited by DCC's HAT, MOF. The X-specific recruitment of MOF is imparted by MSL2, but its targeting towards active genes is thought to be afforded by DCC's PTM reader, MSL3. CD of MSL3 reads trimethylated H3K36, a signature of active transcription (70-72). The third methyl group is deposited by a histone methyltransferase (HMT), Set2, which directly interacts with the elongating Pol II, Pol II-Ser2P, as it travels across the gene body (169,170). Interestingly, H3K36me3 works to antagonize HATs by recruiting histone deacetylases (HDACs) to establish a hypoacetylated environment and prevent spurious intragenic transcription (171-173).

4 Aims

While it has been shown that the long non-coding *roX* RNA is essential for dosage compensation in fruit flies, the exact mechanism remains to be elucidated. Preliminary experiments by Sylvain Maenner suggest the hybridization of *roX1* and *roX2* RNAs leads to the production of a miRNA that targets *Sxl*. Biochemical analysis of *roX1* and *roX2* hybridization and cellular analysis of miRNA production were carried out to explore if *roX* RNAs are implicated in such a feedback mechanism to reinforce the sex determination.

Furthermore, to monitor the establishment of dosage compensation by the Dosage Compensation Complex (DCC) during embryonic development, transcriptome analyses of single-embryos were done concurrently with chromatin immunoprecipitation of MNase-digested chromatin coupled to next-generation sequencing for several DCC subunits, MSL2, MOF, MSL3, and MLE, as well as histone marks, H3K36me3 and H4K16ac. Genome-wide distribution profiles were correlated to tease out the characteristics of DCC targeting to high-affinity sites (HAS) and spreading on the male X chromosome. Datasets obtained from embryos and cell lines (Cl.8 and S2) enabled the comparison of HAS definitions important for targeting.

As both *roX1* and *roX2* RNAs are transcribed into different isoforms, quantitative RT-PCR was used to dissect their expression profiles in nuclear and cytoplasmic fractions of embryonic extracts, as well as cell lines with different *roX* RNA expressions. Finally, a new direct-RNA sequencing method using the Nanopore technology was tested.

5 Materials and Methods

5.1 Materials

5.1.1 Chemicals

Acetic Acid (CLN); Adenosine Triphosphate (ATP, Sigma); Agarose (Bio & Sell); Ammonium Sulfate (NH_4SO_2, Merck Millipore); Ampicillin (Carl Roth); AMPure XP DNA beads (Beckman Coulter); Agencourt RNAclean XP beads (Beckman Coulter); Amylose resin (Biolabs); Bovine Serum Albumin (BSA, Sigma); Bradford (Bio-Rad); Bromophenol blue (Sigma); Chloroform (NeoLab); cOmplete Protease inhibitor (PIC, absource); Coomassie Blue G250 (Serva); DAPI (Invitrogen); DMSO (Sigma); DTT (Roth); EDTA (Diagonal); EGTA (Carl Roth); Ethanol (VWR); Fetal Calf Serum (FCS, Sigma); Glycerol (VWR); Glycine (VWR); Heparin (Sigma); Hepes (Serva); n-heptane (Merck Millipore); Insulin (Sigma); IPTG (Carl Roth); KCl (VWR); Maltose (Sigma); 2-Mercaptoethanol (Sigma); Methanol (CLN); $MgCl_2$ (VWR); NaCl (Serva); Normal Donkey Serum (NDS, Merck Millipore); NP-40 (Sigma); Penicillin/Streptavidin (life technologies); 16% paraformaldehyde, methanol-free (PFA, life technologies); 37% PFA (Merck Millipore); Phenylmethylsulfonylfluoride (PMSF, Genaxxon); Phenol:Chloroform:Isoamyl-alcohol (Invitrogen); 2-Propanol (Sigma); Schneider's *Drosophila* Medium (life technologies); Sepharose protein A and G beads (Helmholtz Centre Munich, E. Kremmer); Shields and Sang M3 medium (Sigma); Sodium Acetate (Sigma); Sodium Azide (Merck); Sodium deoxycholate (Sigma); Sodium dodecyl sulfate (SDS, Serva); 6-14% Sodium Hypochlorite (Merck Millipore); Sodium metabisulfite (NaMBS); Sucrose (VWR); Tris (Diagonal); Triton X-100 (Sigma); Tween-20 (Sigma); tRNA (life technologies); Urea (AppliChem); VECTASHIELD (Vector Laboratories)

5.1.2 Enzymes, markers, and kits

100 bp and 1 kb DNA markers (NEB); DNA 1000/HS Kit (Agilent); Fast SYBR Green Master Mix (Applied Biosciences); MEGAscript T7 Transcription Kit (Thermo Fischer Scientific); MNase (Sigma); NucleoSpin Plasmid Mini Kit (Macherey-Nagel); NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel); Poly(A) Polymerase (NEB); Proteinase K (Qiagen); Restriction Enzymes (NEB); RNA Pico Kit (Agilent); RNase A (Sigma); RNase H (NEB); RNAsin (Promega); RNeasy Kit (Qiagen); rRNA Depletion Kit (NEB); SuperScript III First Strand Synthesis (Thermo Fischer Scientific); Qubit dsDNA HS assay kit (life technologies); Triple Colour Protein Standard III (Serva); TURBO DNase I (Sigma)

5.1.3 Antibodies

Table 1: Primary antibodies

Antigen	Species	Type	Application		Source
DCR2	rabbit	polyclonal	WB	1:250	Abcam, Cat. No. ab4732
GFP	rabbit	polyclonal	IF	1:2000	ChromoTek, Cat. No. 029762
H3K36me3	rabbit	polyclonal	ChIP	2 μL	Abcam, Cat No. ab9050
H4K16ac	rabbit	polyclonal	ChIP	2 μL	Abcam, Cat No. ab109463
H4K16ac	rabbit	polyclonal	ChIP IF	2 μL 1:100	Merck Millipore, Cat No. 07-329
lamin	mouse	monoclonal	WB	1:1000	H. Saumweber
MLE	rat	monoclonal (6E11)	ChIP WB	2 mL 1:500	E. Kremmer (Helmholtz)
MSL1	rabbit	polyclonal	WB	1:2000	E. Schulze
MSL2	guinea pig	polyclonal	ChIP IF WB	2 μL 1:1000 1:2000	C. Regnard (Pineda)
MSL2	rabbit	polyclonal (SA4868)	IF	1:1000	T. Fauth (Eurogentec)
MSL2	rabbit	polyclonal	IF	1:500	C. Regnard (Pineda)
MSL2	rat	monoclonal (1D6)	WB IF	1:500 2 mL	E. Kremmer (Helmholtz)
MSL3	goat	polyclonal	WB	1:1000	M. Kuroda
MSL3	rabbit	polyclonal	ChIP IF	2 μL 1:100	C. Regnard (Pineda)
MSL3	rat	monoclonal (1C9)	IF	1:2	E. Kremmer (Helmholtz)
MOF	rabbit	polyclonal (SA4897)	ChIP WB	2 μL 1: 2000	M. Prestel (Eurogentec)
SXL	rabbit	polyclonal	IF WB	1:500 1:1000	F. Gebauer
TUB	mouse	polyclonal	WB IF	1:2000 1:1000	Abcam, Cat. No. ab44928

Table 2: Secondary antibodies

Antigen	Conjugate	Dilution	Source
Goat, guinea pig, mouse, rabbit, rat IgG	HRP	1:20000	VWR
Goat, guinea pig, mouse, rabbit, rat IgG	IRDye 680RD	1:20000	LI-COR Biosciences
Goat, guinea pig, mouse, rabbit, rat IgG	IRDye 800CW	1:20000	LI-COR Biosciences
Rabbit Ig	AF555/AF647	1:400	Jackson Immunoresearch
Rabbit Ig	Cy3	1:250	Jackson Immunoresearch
Rat Ig	AF488	1:200	Jackson Immunoresearch

5.1.4 Consumables and instruments

Amicon Ultra-4 Centrifugal filter (Merck Millipore); Bioanalyzer (Agilent); Branson Sonicator (Thomas Scientific); CASY Cell Counter (OMNI Life Science); 25, 75, and 170 cm2 Cell Culture Flasks (Greiner); 12-mm round Coverslip (Paul Marienfeld); 24x60 mm Coverslip No. 1.5H (Carl Roth); Dialysis membranes (Spectrum Labs); Douncer (B. Braun); 1.5- and 2-mL Eppendorf tubes; 15- and 50-mL Falcon tubes; Genesys20 (Thermo Fisher Scientific); 3-well 14-mm Microscopy Slide (Thermo Fisher Scientific); Mr. Frosty Freezing Container (Thermo Fisher Scientific); Nanodrop (Thermo Fisher Scientific); 27G needle (B. Braun); 6-, 12-well plate (Sarstedt); Qubit Fluorometer (Thermo Fisher Scientific); table-top Centrifuge (Eppendorf); Thermocycler (Eppendorf); Yamato LH-21 homogenizer (Triad Scientific)

5.1.5 Oligonucleotides and in vitro transcripts

Oligonucleotides are purchased from Sigma-Aldrich. *In vitro* transcribed RNAs are produced using the MEGAscript T7 Transcription Kit and purified with the RNeasy Kit.

Table 3: Oligonucleotides for qPCR of *roX1* and *roX2* panels

Panel		Forward sequence (5' to 3')	Reverse sequence (5' to 3')
roX1	NP1	TAATCAGAGACCAGGGCACC	TCGAAAAAGCCGAAGGGGAC
	NP2	CTCTGGCAAGATGTAGCGTC	AAATGGCTTCTTTGGGATTG
	NP3	CTTCCCGGAGGAGTGTGGA	CATCTCGTTGCTGTTGCGTTA
	NP4	TCTCTTTCGGGACTTGCAGT	TGTAGACAAGGAGAGACGGC
	NP5	TCCGACCAGAAGTAGATCGTG	CAATTGACGAGGTGCTGAACA
	NP6	TCGTTTTCCGAAATGGGAATCA	AGGCTTTCAATACCGTTCCA
	NP7	TCTTCCCATTCTTGTGGATGCT	CTTTGCAGCATATGAACAACGA
roX2	NP1	GGCCTGGTCACACTAAGCTA	ACCTAAACGCTCGACTTATGAT
	NP2	CGCAGTGCAACGTATACACA	GGGAAAATTGAGCGGGTGTT
	NP3	TCTCCGAAGCAAAATCAAGCA	TCGTACTCATCTCACTGTCCG
	NP4	ATGCAATACAATACAATACAAGACAAA	TTCAGTTTGCATTGCGACTT
	NP5	GACGTGTAAAATGTTGCAAATTAAG	TGACTGGTTAAGGCGCGTA
	NP6	AGCGAGATGACAATAGAGAGGC	GCCATCTCTCTAAGCCAGC

Table 4: Oligonucleotides for qPCR of embryo staging

Name	Forward sequence (5' to 3')	Reverse sequence (5' to 3')	Trend
apc	CTGAGGGCACTTGCTGATGT	AAACGTCCTGGCTACGTCTG	
crag	ACAAGTTGGGCAGCATTGTG	CAACCCAGTGTTGTTGACCG	Up
cwo	GCACTGAACCCGAGATAGCA	CTGGACTCGGTGTGAACCTC	
socs16d	CAACGGACCCTTTTTCAGCG	CGAGGTCTCGCAGTTCGATT	
retn	CCTGAATCCACACAACATGG	ACCGTGAGCGTAGTCATCCT	
Sr-CII	GTGGTGGTGGTGCTATACCC	CACAATTCGGGGACATTGCC	Down
zld	GTACTCACCGGAGTGGAAGC	AAGTGTCTGGAGTGCGACAA	

Table 5: Oligonucleotides for qPCR of ChIP efficiency

Name	Forward sequence (5' to 3')	Reverse sequence (5' to 3')	Target
phlA	CTGTGCTATCTGCTCCCATGAT	CGTTCAGCTCTAATCGCAAGC	intron
phlD	GGCCACTGCGAAAACTCG	CGCCAACAGTTCGTACATCACG	gene body
phl hs	GAAGGGGACGGCTTTGGTT	GGTCGGCAGGCAACCC	HAS
roX2 SL3/4	TCTCCGAAGCAAAATCAAGCA	TCGTACTCATCTCACTGTCCG	gene body
roX2 hs	AGCGAGATGACAATAGAGAGGC	GCCATCTCTCTCTAAGCCAGC	HAS
set2	ATCTCGCGGT ACATCAACCA	CACGCTGAAG AAACCAATGC	gene body
set2 hs	GCGTACAGTAGCTGAGAGCTG	GCGCCTTTACTGCATGTTAG	HAS
ubx f4	TAGTCTTATCTGTATCTCGCTCTTA	CAGAACCAAAGTGCCGATAACTC	phantom peak
ubx f12	GCCGTGGAGCAGTTCAAAGTA	TCGTTGGTCGTGTCCTCTTAATT	gene body
runt	CACCAGATCTCAGCACGAACA	CCGTGATACTCCTGCAGCATC	ChIP efficiency

Table 6: Oligonucleotides for dsRNA production to perform RNAi treatment of cells

Name	Forward sequence (5' to 3')	Reverse sequence (5' to 3')	Template (length)
GFP	TTAATACGACTCACTATAGGGTGCTCAGGTAGTGGTTGTCG	TTAATACGACTCACTATAGGGCCTGAAGTTCATCTGCACCA	Plasmid (479 nt)
Jil-1	TTAATACGACTCACTATACGGCCGACATTTGGTTAGCTG	TTAATACGACTCACTATAGGGCGTGCTACCACA	g/cDNA (667 nt)
SXL I	TTAATACGACTCACTATAGGGAGAGATCACAGCCGCTGTCC	TTAATACGACTCACTATAGGGAGATACCGAATTAAGAGCAAATAATAA	gDNA (512 nt)
SXL II	TAATACGACTCACTATAGGGAGACCCTATTCAGAGCCATTGGA	TAATACGACTCACTATAGGGAGAGTTATGGTACGCGGCAGATT	gDNA (412 nt)

Table 7: Sequences of *in vitro* transcribed *roX* RNAs

Name	Sequence (5' to 3')
roX1-5' (518 nt)	ATTACGTTCGGAGTGGAAAATGGAAATTAAGTGAAATATCCAGTGATCGATCGGTAATAGTAAATTGTTTGATACGTTTAGGCCAGTTGGTAAAGCAAATTAACAGCAGTTGTAAGTAAATTTAATCAGAGACCAGGGCACCACACCCGAAAAGCGTGCAGATATTAGAAGACATGGGCGTAGTTTCATATACGAGCTGTCCCCTTCGGCTTTTTCGACAAGTGGCAGCCCTAATGGCCCTCGTTTTTTCGCCGACAAGCATTTAATGCGTAGTCACCGAAGAAAAGTGTTAGTTACCAGGGCCTGCCCTTTTAAAATTAAATTTAAATTGAAAAAAAAATCACCAAAAAATCGAAATCTCTGGCAAGATGTAGCGTCGAAAGAAAATTCATCAAACGGCATTGCCATCATCGTGCAGCAATCCCAAAGAAGCCATTTAGAATGCAGGCATCCAGGCAAAAACCAGAAACGTGCCTGTGGCAGCTGCACTTCGTGGCCTTGACGAGTCCGGACAAT
roX1-3' (535 nt)	ATCCGACCAGAAGTAGATCGTGTTCTGTGAACTAACCCCTTCAGTGTTCAGCACCTCGTCAATTGTTCTTAATTGTTCCTTTTATTTTATGTTGTGTTATCTAATTAACTTCCGTTGTATTTTACCCAGTCCCCTTCCTTGACTTTCTAATAATTTTCCATGTTTTGACATATCCTTTTTTGTCCCAGCCGAATAACCAACCATACTATTCCTATATAAGGTTCGTGTTTCGGAAAACGCATTAAAAGGCGTAATTTTAAATCGTTTTCCGAAATGGGAATCACATTTAAACAATATTTTGAACTGCGTAAAACGAATAAATGGAACGGTATTGAAAGCCTATGCATTCATTACGGTTCAAGAAGTATAACTAAAAAAAAGGAATGAAAAAAAAAAAGAAAAGA

	AAAACACATTTACTAACAAATAAAAACTTGCTGATCAACGTTCTACGCAGTTCTTAA AAAGATGTTGAAATGAACACAGCCAAAGCAAGTAAAAATGTGTGGAAACGTTATAC GAATCTTCACCAAGTGCC
roX2 **(552 nt)**	CCTGGTCACACTAAGCTAGGGCTACTTTTTATATCATAAGTCGAGCGTTTAGGTAGCT CGGATGGCCATCGAAAGGGTAAATTGGTGTTACATATAGCTTTAGAGATCGTTTCGA ATCACATTGATAATCGTTCGAAACGTTCTCCGAAGCAAAATCAAGCAAGAGTAACGA TTTCCGCATAGTCGAAAATGTTTAAGTTGAATTGTCTTACGGACAGTGAGATGAGTA CGACTATTTGGAAATCACAAACGAATTGTTTTCATGGTTGACGCGCTTGTCAAGCTAC AAAACAAAATGAATGATATACAATATACAATATACAATATGCAATACAATACAATAC AAGACAAAAAAATGTGTCTTGGAACGCAACATTGTACAAGTCGCAATGCAAACTGA AGTCTTAAAAGACGTGTAAAATGTTGCAAATTAAGCAAATATATATGCATATATGGG TAACGTTTTACGCGCCTTAACCAGTCAAAATACAAAATAAATTGGTAAATTTCATAT AACTAGTGAAATGTTATACGAAACTTAACAATTGCCAA

5.1.6 Cell lines

Table 8: Cell lines used in the study

Strain	Origin	Source
Clone-8 (Cl.8+ CME W1)	*D. melanogaster* wing disc, third instar larvae	DGRC
Kc	*D. melanogaster* embryo, dorsal closure stage	DGRC
Schneider's 2 (S2)	*D. melanogaster* embryo, late embryonic stage	DGRC

5.1.7 Fly strains

Fly strain used in most experiments is Oregon-R (wild type). A fly line expressing GFP under the *SxlPe* promoter is used for the analysis of protein expression in sex-sorted embryos (174).

5.1.8 Bacterial strains

Table 9: Bacterial strains used in the study

Strain	Genotype	Source
***E. coli* DH5a**	*fhuA2 (argF-lacZ) U169 phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB, Cat. No. c2987
***E. coli* BL21-Gold (DE3)**	*fhuA2 [lon] ompT gal [dcm] ΔhsdS*	Agilent, Cat. No. 230132

5.1.9 Standard buffers and solutions

Table 10: Standard buffers and solutions

Solution	Recipe
Agarose gel	TBE Buffer 1% Agarose Ethidium bromide

Ampicillin stock solution	100 mg/mL Ampicillin (1000x)
Buffer D	20 mM Hepes-KOH pH 7.9
	100 mM KCl
	3 mM MgCl$_2$
	20% glycerol
	0.2 mM EDTA
	0.5 mM DTT
Coomassie fixing solution	50% Ethanol
	10% Acetic Acid
Coomassie staining solution	5% Ethanol
	7.5% Acetic Acid
	0.0025% Coomassie Blue (w/v)
Dialysis buffer (MS2-MBP)	20 mM Hepes-KOH pH 7.9
	100 mM KCl
	20% Glycerol
	0.2 mM EDTA
	0.5 mM DTT
Elution buffer for MS2-MBP	Heparin column buffer 1
	10 mM Maltose
Ethidium bromide stock solution	10 mg/mL Ethidium bromide (20000x)
EX-50 buffer	10 mM Hepes-KOH pH 7.6
	10 mM KCl
	1.5 mM MgCl2
	10% Glycerol
	0.5 mM EGTA
	1 mM DTT
	0.2 mM PMSF
Extraction buffer for NE	30 mM Hepes-NaOH pH 7.6
	450 mM NaCl
	10% Glycerol
	1 mM EDTA
	1 mM DTT
	PIC
100x Fixing solution	500 mM Hepes pH 8.0
	1M NaCl
	10 mM EDTA
	5 mM EGTA
HEMG buffer	25 mM Hepes-KOH pH 7.6
	100 mM KCl
	2 mM MgCl2
	20% Glycerol
	0.1 mM EDTA
	1 mM DTT
	0.2 mM PMSF
	1 mM NaMBS
	PIC
Heparin column buffer	20 mM Tris-HCl pH 7.6

	buffer 1: no salt
	buffer 2: 50 mM NaCl
	buffer 3: + 100 mM NaCl
	buffer 4: + 200 mM NaCl
	buffer 5: + 1 M NaCl
Homogenization (NX-1) buffer	15 mM Hepes-KOH pH 7.6
	10 mM KCl
	2 mM MgCl2
	350 mM Sucrose
	0.1 mM EDTA
	0.5 mM EGTA
	1 mM DTT
	0.2 mM PMSF
	PIC
5x Laemmli loading buffer	250 mM Tris pH 6.8
	50% Glycerol (v/v)
	10% SDS (w/v)
	0.05% Blue Bromophenol (w/v)
	0.5 M DTT
LB agar plates	LB medium
	2% Agar (w/v)
LB medium	1% Pepton (w/v)
	0.5% Yeast extract (w/v)
	1% NaCl
Lysis buffer for NE	15 mM Hepes-KOH pH 7.6
	60 mM KCl
	15 mM NaCl
	10% Sucrose
	0.5 mM EGTA
	0.5 mM DTT
	PIC
Lysis buffer for WCE	50 mM Hepes-NaOH pH 7.6
	300 mM NaCl
	0.3% Tx-100
	PIC
MNase enzyme solution	500 U resuspended in 850 μL EX-50
5% native PAGE gel	5% Bis-Acrylamide
	0.05% APS
	0.05% TEMED
	TBE buffer
PBS(T) buffer	1.4 M NaCl
	27 mM KCl
	100 mM Na2HPO4
	18 mM KH2PO4
	(0.1% Tween-20)
RIPA (Hepes) buffer	25 mM Hepes-NaOH pH 7.6
	150 mM NaCl

	1% Triton-X-100
	0.1% SDS
	1 mM EDTA
	0.1% Na-deoxycholate
	1 mM PMSF
	PIC
RIPA (Tris) buffer	10 mM TRIS pH 8.0
	140 mM NaCl
	1% Triton-X-100
	0.1% SDS
	1 mM EDTA
	0.1% Na-deoxycholate
	1 mM PMSF
	PIC
SDS PAGE running buffer	25 mM Tris
	192 mM Glycine
	0.1% SDS
Sonication buffer (MS2-MBP)	Heparin column buffer 4
	PIC
Sucrose cushion	Lysis buffer for NE
	10% Sucrose
Transfer buffer	25 mM Tris
	192 mM Glycine
	20% Methanol
TRAX buffer I	same as Homogenization (NX-1) buffer
TRAX buffer II	15 mM Hepes-KOH, pH 7.6
	110 mM KCl
	2 mM $MgCl_2$
	0.1 mM EDTA
5% urea PAGE gel	8 M Urea
	5% Bis-Acrylamide
	0.05% APS
	0.05% TEMED
	TBE buffer
Urea sample buffer	9 M Urea
	25 mM Tris pH 6.8
	1% SDS
	1 mM EDTA
	100 mM DTT
TBE buffer	100 mM Tris base
	100 mM Boric acid
	2 mM EDTA
TE buffer	10 mM Tris pH 8
	1 mM EDTA

5.2 Drosophila studies

5.2.1 Drosophila husbandry

Population of Oregon-R flies are maintained at 25°C according to standard fly husbandry.

5.2.2 Transcriptionally active embryo extract (TRAX) preparation

Embryos of Oregon-R wild-type flies are collected every 12 hrs until a minimum of 100 g is reached. To stop them from developing further, embryos are kept at 4°C for a maximum of 72 hrs. Dechorionated embryos are then homogenized in 2 mL/g embryo of TRAX buffer I using the Yamato LH-21 homogenizer set to 1000 rpm in six passes. The homogenate is then filtered through a single layer of miracloth before further dilution to 5 mL/g embryo with TRAX buffer I. Nuclei are then pelleted by centrifugation at 10000 g, 4°C, for 15 min and resuspended in 1 mL/g embryos of TRAX buffer II without disturbing the yellow egg yolk pellet. Supernatant is kept as cytoplasmic fraction. Nuclei suspension is fully resuspended by 2 strokes of a B Dounce. 1/10 volume of room temperature 4 M $(NH_4)_2SO_4$ pH 7.9 is added quickly and mixed by rapid inversion. Viscous solution of nuclei is rotated at 4°C for 20 min and centrifuged at 150000 g, 4°C for 2 hrs to pellet out insoluble fraction. Supernatant below the lipid layer is aspirated and precipitated out using 0.3 g/mL finely ground $(NH_4)_2SO_4$ added steadily over a 5 min period, while stirring on ice. Proteins are pelleted by centrifugation at 27000 g, 4°C, for 20 min and resuspended in 0.2 mL/g embryo of HEMG-100. To remove excess salt, nuclear extract is dialysed against 2 L HEMG-100 for 4 hrs. Precipitate is spun out through centrifugation at 10000g, 4°C, for 5 min and extract is stored in small flash-frozen aliquots at -80°C.

5.2.3 Embryo collection for IF and chromatin preparation

Embryos are dechorionated using 25% commercial bleach and fixed in PBS/3.7% formaldehyde/heptane at 26°C for 20 min, agitated. After formaldehyde removal, embryos are shaken in heptane/methanol (1:1) solution to aid devitellinization. Embryos are then rinsed in methanol and stored at -20°C.

5.2.4 Immunofluorescence (IF)

For IF staining, embryos are first rehydrated and washed in PBSTx-0.1% (PBS/0.1% Triton X-100). Blocking is done twice for 20 min each in PBSTx-0.1%/5% NDS. Both primary and secondary antibodies are diluted accordingly in block solution. Primary staining is done overnight at 4°C with agitation. Upon four washes of 15 min each in PBSTx-0.1%, embryos

are stained with secondary antibodies in dark at room temperature for 2 hrs with agitation. Embryos are then rinsed and washed four times in PBSTx-0.1% for 15 min each, the last one with added DAPI to stain DNA. Upon a final wash in PBS, embryos are mounted onto a 3-well glass slide in VECTASHIELD mounting medium, covered, and sealed with nail polish.

5.2.5 Chromatin preparation from embryos

For chromatin preparation, staged mixed-sex embryos are dechorionated in 25% commercial bleach and weighted. 1 g embryos are washed in 50 mL PBS/0.01% Triton X-100 (PBSTx-0.01%) and resuspended in 10 mL 10x Fixing solution/3.7% formaldehyde. After addition of 30 mL n-heptane, embryos are rigorously shaken for 1 min and rotated at room temperature for 13.5 min. Following a spin at 2000 g for 1 min, crosslinking is halted by addition of 50 mL PBSTx-0.01%/125 mM Glycine. Washed embryos are flash-frozen and stored at -80°C until further use. Frozen embryos are resuspended in 10 mL Homogenization Buffer and dounced 20 times with a loose pestle and 20 times with a tight pestle before being spun down at 170 g for 10 min at 4°C. The nuclei pellet is resuspended in 4 mL RIPA (Hepes) Buffer. Fragmentation of chromatin is done either by MNase treatment or sonication shearing using the Covaris S220 instrument. To obtain similar digestion degree, stage 5-8 embryo are digested using 0.9 units MNase/g embryo and stage 13-15 embryo are digested using 2.9 units MNase/g embryo at 37°C for 30 min, shaking. Reaction is stopped by the addition of EDTA to a final concentration of 10 mM. Additional mechanical shearing is done by passing lysate through a 27G needle 15 times. Soluble chromatin is retrieved by centrifugation at 15000 g for 15 min at 4°C. Chromatin fragment size distribution is evaluated on a Bioanalyzer.

5.3 Cell biology methods

5.3.1 Cell counting

Cell concentration is determined using the CASY cell counter system as instructed. 50 μL cell suspension is mixed with 10 mL CasyTon solution and measured with the set program, where viable cells with median diameter of 8 µm are called between 6-20 µm and debris counted as those below 3 µm.

5.3.2 Maintenance of cell lines

Kc and S2 cell lines are kept at 26°C in Schneider's *Drosophila* medium enriched with 10% fetal calf serum (FCS) and 1% penicillin/streptomycin mix (P/S). Cl.8 cell line is maintained at 26°C in Shields and Sang M3 medium supplied with 2.5% fly extract, 2% FCS,

5 μg/mL insulin and 1% P/S. All components of cell culture are kept under sterile conditions and brought to room temperature before use. To maintain a 70% confluency, cell lines are diluted 1/10 every 72 hrs.

5.3.3 Fly extract for Clone 8 cells

Adult flies are collected in a 50-mL falcon tube and put in freezer for at least 1 hr. Frozen flies can also be stored for future use. 30 g frozen flies and 200 mL cold Shields and Sang M3 medium are blended in a cooled blender until all flies appear lysed. Homogenate is spun at 1500 g, 4°C for 15 min. Supernatant is filtered through miracloth to remove larger debris and placed in a 60°C water bath for 20 min to inactivate tyrosinase. Homogenate is spun again at 1500 g, 4°C for 90 min. Supernatant is filtered through a 0.22 μm filter, aliquoted and flash-frozen in liquid nitrogen. Fly extract is stored at -20°C.

5.3.4 Cryopreservation and thawing of cells

Cells are seeded to a density of $1x10^6$ cells/mL 3 days prior to freezing. Collected cells are spun at 500 g, room temperature for 5 min. Cell pellet is resuspended in freezing medium and aliquots of $10x10^6$ cells/mL are frozen gradually first at -20°C for 24 hrs then at -80°C for another 24 hrs in a freezing container before transfer to liquid nitrogen for long-term storage. Kc and S2 cells are frozen in FCS supplemented with 10% DMSO, whereas Cl.8 cells are cryopreserved in Shields and Sang M3 medium enriched with 2.5% fly extract, 10% FCS, 5 μg/mL insulin and 10% DMSO. To thaw, frozen cells are quickly warmed up in a water bath and transferred to 5 mL culture medium in a 75 cm2 culture flask for at least 1 hr. Once cells attach, medium is changed to remove DMSO. Slow freezing and quick thawing preserve cells best.

5.3.5 RNA interference in cells

Genomic DNA or plasmid containing cDNA template is used to amplify ~500 bp PCR product containing a T7 promoter site and sequence of target of interest. Purification of product is done using the PCR Clean-up Kit as instructed and is thereafter used as template in an *in vitro* transcription reaction using the MEGAscript T7 kit. Generated RNA is treated with DNase and precipitated using lithium chloride. For the formation of dsRNA, sample is denatured by heating at 85°C for 10 min and let cool to room temperature.

To carry out RNA interference (RNAi) experiments, $1x10^6$ cells in log growth phase are seeded in each well of a 6-well plate in 1 mL of serum-free medium. 10 μg of dsRNA is added per well. Plate is shaken for 10 min and incubated at 26°C for 50 min before 2 mL of fresh

medium is added. Cells are allowed to grow for an additional 4 to 14 days before analysis. To achieve better efficiency, fresh dsRNA can be supplied similarly between first treatment and analysis.

5.3.6 Immunofluorescence of cells

To start immunostaining, 0.5-0.8×10^6 cells are seeded onto coverslips treated with poly-lysine and let attach for 1 hr at 26°C. Carefully, cells are washed twice in PBS and fixed for 7.5 min in PBS/2% PFA on ice. To permeabilize, cells are treated with PBS/0.25% Triton-X-100/1% PFA for 7.5 min also on ice. Upon a PBS wash, cells are then blocked in PBS/3% BSA for 1 hr at 26°C. Thereafter, cells are incubated in diluted primary and subsequently secondary antibody in the blocking solution for 1 hr each with a PBS wash step in between.

5.3.7 Whole cell extract preparation

Whole cell extract is prepared from 2×10^8 cells in log growth phase. Cells are spun down and washed in PBS at 500 g for 5 min. Cell pellet is resuspended in 2 mL lysis buffer for WCE and passed through a 27G syringe 10 times. Cell suspension is rotated at 4°C for 10 min before a centrifugation at maximum speed for 10 min. Supernatant is collected and 100 μL aliquots of WCE is stored at -80°C. Concentration of extract is measured by Bradford.

5.3.8 Nuclear extract preparation from cells

Nuclear cell extract is prepared from 200×10^6 cells in log growth phase. Cells are spun down and washed in PBS at 500 g for 5 min. Cell pellet is resuspended in 3 packed cell volume (pcv) lysis buffer for NE and supplemented with 0.3% fresh NP-40. Cell suspension is rotated at 4°C for 8 min and slowly pipetted onto a sucrose cushion. Centrifugation at 1500 g, 4°C for 15 min is done to separate nuclei found below the sucrose cushion from the cytoplasmic fraction remaining above the cushion. Wash nuclei once in 3x pcv lysis buffer for NE and resuspend in 750 μL extraction buffer for NE. Extract concentration is measured by Bradford.

5.3.9 Chromatin preparation from cells

1/10 volume of 10x fixing solution and a final concentration of 1% PFA is added to cells growing in log growth phase. Fixing of cells is done for 8 min at 26°C and stopped by the addition of 125 mM glycine. Cells are spun down and washed twice in PBS at 500 g for 10 min. Nuclei isolation is thereafter done by resuspending cell pellet in PBS supplemented with 0.5% Triton-X-100 and PIC at a concentration of 70×10^6 cells/mL and rotated at 4°C for 15 min. Nuclei are collected by centrifugation at 1500 g, 4°C for 10 min and washed once in PBS.

To solubilize chromatin, nuclei is resuspended in RIPA (Tris) buffer at a concentration of 7×10^7 cells/mL, treated with 0.75 µL/mL MNase activated with 2 mM $CaCl_2$ and incubated at 37°C for 35 min. To stop reaction, 10 mM EGTA is added and sample is incubated on ice for 5 min. A sheering step using Covaris at 50W, 20% amplitude for 8 min follows to aid in releasing insoluble factors. Supernatant containing the soluble chromatin fraction is collected after centrifugation at max. speed, 4°C for 20 min and stored in aliquots of 100 µL at -80°C.

5.4 Molecular biology methods

5.4.1 General molecular biology methods

General molecular biology methods are done according to standard protocols.

5.4.2 DNA purification methods for NGS

Generally, DNA libraries are purified using AmpureXP beads for sequencing. In order to preserve small fragments that are otherwise selected against, DNA is precipitated after phenol/chloroform extraction by ethanol and sodium acetate (NaOAc). Briefly, equal volume of buffered phenol-chloroform-isoamyl alcohol (25:24:1) is added to sample. Upon a quick strong vortex, mixture is pipetted onto a MaXtract tube (Qiagen, Cat. No. 129046) and quickly spun as instructed. The aqueous phase remains on top, while the organic phase migrates below the gel. The aqueous phase is re-extracted once more with equal volume of chloroform in the same tube if volume allows. Supernatant is then transferred to a fresh tube containing glycogen at final concentration of 0.5 µg/µL as a carrier and 1/10 volume of 3 M NaOAc and 2.5x volume pure ethanol is added. Finally, DNA is precipitated by a minimum of 30 min incubation at -20°C and a 15 min centrifugation at maximum speed. DNA pellet is then washed thrice with cold 70% ethanol, air-dried and resuspended in TE buffer.

5.4.3 In vitro polyadenylation for NGS

RNA libraries for direct-RNA nanopore sequencing are prepared from 6 µg TRAX RNA first rRNA-depleted using the rRNA depletion kit and thereafter polyadenylated *in vitro* with 2.5 U Poly(A) Polymerase incubated for 10 min at 37°C. RNA is purified using the Agencourt RNAclean XP as per manufacturer's protocol.

5.5 Biochemistry methods

5.5.1 MS2-MBP purification

Bl21 cells are transformed with plasmid #1665 encoding for the MS2-MBP fusion protein and induced with 0.1 mM IPTG for expression upon reaching OD600 of 0.6-0.8. Cells are harvested after 2.5-hr incubation at 37°C, at which OD600 is around 1.6. To force cell lysis, bacteria pellet is resuspended in sonication buffer and sonicated using Branson Sonicator with setting 80 sec 10 sec ON and 20 sec OFF at 20% amplitude. Amylose beads are incubated with the homogenate at 4°C on a rotating wheel and collected by centrifugation at 1000 rpm, 4°C for 2 min. To elute MS2-MBP protein, beads are loaded onto a BioRad column with elution buffer for MS2-MBP and 0.5 mL fractions are collected. A heparin column is equilibrated using buffer 1 before loading fractions containing most MS2-MBP as analyzed on a Coomasie gel. Washes are done using buffer 2, and MS2-MBP is eluted using buffer 3 into 1 mL fractions. Fractions are analyzed once more on a Coomasie gel, and those containing the best MS2-MBP protein are dialyzed against Buffer D for 2 hrs and concentrated using a microcon to desired concentration. Flash frozen aliquots are stored at -80°C. Heparin column can be reused, if washed with buffer 4 and stored at 4°C in buffer 5 (175).

5.5.2 MS2-MBP affinity chromatography

100 pmol of MS2-tagged RNA diluted in Buffer D is incubated at 65°C for 10 min and let cool for 10 min at room temperature to allow denaturation and renaturation. Then, 5-fold molar excess of purified MS2-MBP fusion protein is added to bind RNA in a rotating incubation at 4°C for 20 min. To immobilize RNA-MS2/MS2-MBP, equilibrated amylose beads are added and rotated together at 4°C for 1 hr. Meanwhile, 1 mg extract is incubated at 30°C for 15 min to remove endogenous ATP. Once beads are washed twice with Buffer D, 1 mg extract, 1 mM ATP, 15 ug tRNA and RNAsin are added. RNP reconstitution is allowed to take place at 26°C for 10 min, rotating. Upon completion, beads are washed twice with Buffer D at 4°C for 10 min each, rotating, to remove excess unbound factors. Elution of RNP complexes is done using 10 mM maltose. RNA components are analyzed by PAGE bathed in ethidium bromide and protein components are analyzed by Western blot (WB).

5.5.3 Chromatin immunoprecipitation (ChIP)

1-2 µg soluble chromatin is used as input for each ChIP. Soluble chromatin diluted in RIPA buffer is first pre-cleared through incubation with protein A+G (1:1) beads, rotating at 4°C for

at least 2 hrs. RIPA buffer used in ChIP of cell chromatin is Tris based, whereas RIPA used for ChIP of embryo chromatin is HEPES based. Incubation of soluble chromatin with antibody of interest is done overnight and retrieved by incubation with 50% slurry mix of sepharose beads. For all ChIP using rabbit antibodies, slurry is made of protein A+G (1:1) beads, whereas rat antibodies are fished using only protein G beads. Reversal of crosslinking is done by an overnight incubation, shaking at 65°C. To rid of proteins and RNAs, sample is incubated with 1 µg RNase for 30 min at 37°C and with 0.1% SDS/1 µg Proteinase K for 1.5 hrs at 55°C. DNA is then purified either using 1.8x volume AMPure XP beads or by phenol-chloroform as described above. ChIP efficiency is evaluated by qPCR and libraries for sequencing are made using the NEBNext Ultra II DNA Library Prep Kit for Illumina.

5.5.4 Visualization of SDS PAGE gel

Upon completion of SDS PAGE, gel is fixed in 50% ethanol/10% acetic acid for 5 min at room temperature, shaking. Fixed gel is stained with 5% ethanol/7.5% acetic acid/0.0025% Coomassie Brilliant Blue G-250 for another 30 min at room temperature, shaking. Destaining in water is thereafter done until protein bands are visible. To improve staining, gel can be heat up. Alternatively, gel is submerged in TBE/ethidium bromide solution to stain for nucleic acids.

5.6 Data analysis

5.6.1 Image analysis

Images acquired on a Leica TCS SP5 confocal microscope are analyzed on software Fiji.

5.6.2 Sequence analysis

Sequence analysis is done using the software ApE (W. Davis) or SerialCloner (Serial Basics).

5.6.3 Statistical analysis

When possible, data are expressed as the mean ± standard error of mean. Statistical analysis is done using the Prism 8 (GraphPad Software) software. Gene expression analysis is calculated using the Pfaffl equation to determine relative difference in change of gene expression of a housekeeping gene and the gene of interest.

5.6.4 Genome-wide analysis

Alignment of MNase-based paired-end ChIP-seq reads to the reference genome (dmel release 6) are done using Bowtie2 (version 2.2.9). Single reads are filtered using samtools with

parameter -q 2, while paired-end reads are first subset into sub-nucleosomal (parameters: -I 10 -X 130) or mono-nucleosomal (-I 130 -X 220) fragments and filtered using samtools with parameter -q 12. Reads are processed and normalized (to total number of reads and subsequently to input) using the Homer Software package (176). To call peaks on pooled replicates, Homer findPeaks tool is utilized with parameters -style histone -size 1000 -F 2.5 for MNase MOF and H4K16ac; -size 2000 -F 2 for MNase MSL2 mono-nucleosomal fragments or with -style factor -size 200 -F 6 for MSL2 sub-nucleosomal fragments. Composite plots are averaged replicates from input normalized bedgraph coverages centered at sites. MNase MOF, H4K16ac and MSL2 sub- or mono-nucleosomal coverages are centered to HAS in a 4 kb window and mean +/- 95% confidence intervals across HAS are visualized. Sub- or mono-nucleosomal peak/region centered plots are generated in 4 kb or 10 kb windows, respectively. For correlation analysis at MSL2 peaks, ChIP enrichment is averaged in a 400 bp or 4 kb window for MSL2 sub- or mono-nucleosomal profiles, respectively, and Pearson correlation coefficient are calculated across samples and replicates. Meme (version 5.0.2) is used for *de novo* motif finding on MSL2 sub-nucleosomal peaks with parameters -nmotifs 1 -dna -revcomp -mod zoops -maxw 18. For genic analysis, MNase MOF and H4K16ac ChIP-seq and Input reads are counted over exons for each gene. Counts were normalized to TPM (similarly to RNA-seq). Finally, normalized counts were log2 transformed, Input is subtracted, and replicates were averaged. These genic ChIP enrichments were used for calculating Pearson´s correlation coefficients or for kmeans clustering (k = 3) to define high, moderate and low enrichment levels.

6   Results

6.1   *miRoX2* as a miRNA

6.1.1   *Hypothesis of a miRNA in development, miRoX2*

A series of pilot experiments from Sylvain Maenner led him to devise a hypothesis of a feedback loop that may act as a fail-safe mechanism in sex determination. The hypothesis states that *roX1* and *roX2* RNAs hybridize by base-pairing and the resulting hybrids are processed by the RNAi machinery into a miRNA, termed *miRoX2*, to induce the degradation of *Sxl* mRNA (Fig. 7). The *Sxl* mRNA is post-transcriptionally regulated to express a truncated, non-functional protein in male. Such a feedback mechanism would reinforce the negative regulation of *Sxl* in male that would otherwise negatively regulate the only male-specific protein in the DCC, MSL2. Interestingly, it would do so by involving another male-specific component of the DCC, *roX* RNAs.

Figure 7. A hypothetical, non-canonical pathway of miRNA production may be involved in a feedback loop mechanism of dosage compensation. A) *roX1-3'* and *roX2* RNAs complement to form a hybrid. B) Upon hybridization, the RNAi machinery produces a miRNA, *miRoX2*, from hybridized *roX*. C) *miRoX2* targets the 3' UTR of *Sxl* mRNA to induce degradation.

In these experiments, an MS2-MBP affinity chromatography was performed using various *in vitro* transcribed RNAs (Fig. 8A) to pull down proteins/complexes that interact with them upon addition of nuclear extracts (NE). The first set of results showed the hybridization of *roX2* RNA with specifically the 3' end of *roX1* (*roX1-3'*) in a Northern blot analysis (Fig.

8B). Additionally, both *roX1-3'* and *roX2* RNAs seemed to interact with proteins of the RNAi pathway, such as DCR2 and AGO2 (Fig. 8C). Lastly, degradations of both RNAs were observed when mixed in the presence of NE of cells (Fig. 8D).
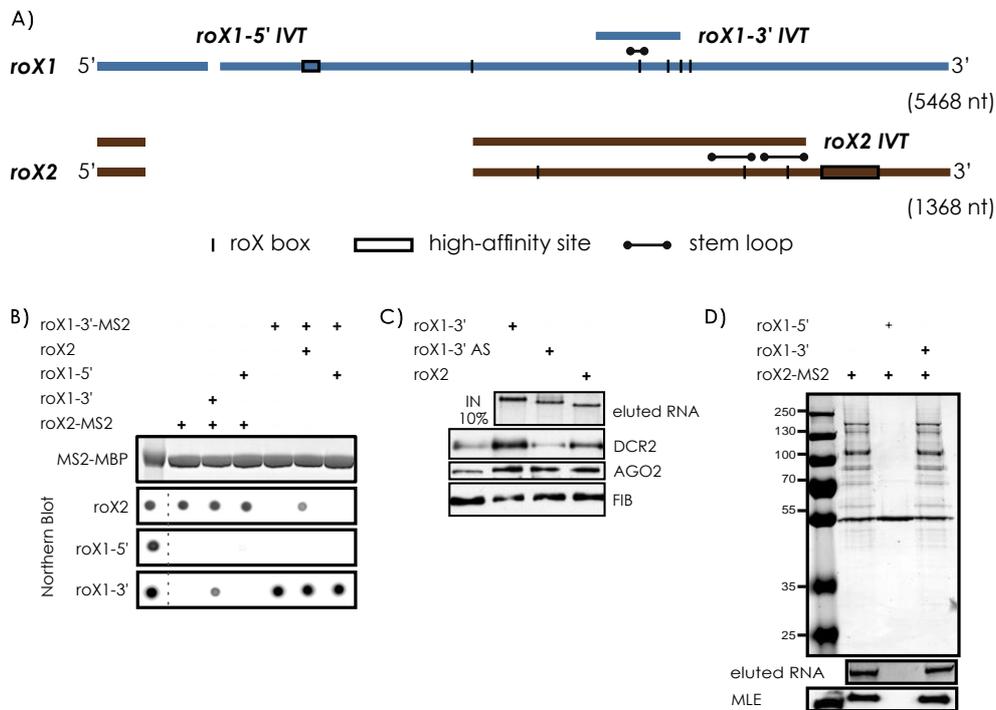

Figure 8. Pilot experimental data from S. Maenner hinted towards a new hypothesis upon *roX* RNA hybridization. A) *In vitro* transcribed *roX1-5'*, *roX1-3'*, and *roX2* RNAs used in pull-down experiments are denoted above schemes of *roX* genes. B) Northern blot suggested specific interaction of *roX1-3'* and *roX2* RNAs. C) *roX1-3'* and *roX2* RNAs interacted with DCR2 and AGO2. D) *roX1-3'* and *roX2* were degraded when incubated in NE of cells. Ticks represent *roX* boxes, boxes denote high-affinity sites on genes, and bulleted dashes indicate *roX1* SL and *roX2* SL7 and SL8. Introns are drawn in light shade, whereas exons in dark.

Coincidentally, an expression profile from a genome-wide survey of sexually dimorphic *Drosophila* miRNAs became available (177). Indeed, a sequence of 19 nt that could have originated from *roX2* RNA was only detected in ~24-hr old male embryos, although not enough reads were reported for it to warrant annotation (177). This 19-nt sequence, "CAATATACAATATGCAATA", was termed *miRoX2*, and through prediction algorithm searches on miRBase.org and microRNA.org, S. Maenner identified *Sxl* as one of its potential targets (Fig. 8).

The data of S. Maenner was considered preliminary, but worth following up given the interesting implications in case of its confirmation. The *roX* genes would play two roles in

establishing and carrying out DC, namely by producing a lncRNA and a miRNA, with very different modes of actions.

6.1.2 Revisiting miRoX2 hypothesis: do roX RNAs hybridize?

Several things must be confirmed in recapitulating previous results, in order to conclude that an alternative feedback mechanism exists as hypothesized. First, the interaction of *roX1-3'* and *roX2* RNAs; second, the binding of individual and hybrid RNAs with proteins of interest; and third, the degradation of the *roX* hybrid specifically. To probe for interaction between *roX1-3'* and *roX2* RNAs, hybridization experiments were carried out and analysed by electrophoresis. Included as control in the experiments were the antisense (AS) RNAs that should with no difficulties form hybrids with RNAs of interest. On a non-denaturing native PAGE, mixture of each RNA with its AS led to a different migration pattern, i.e. much faster, indicative of hybrid formation (Fig. 9A). Incubating *roX1-3'* and *roX2* RNAs together, however, did not (Fig. 9A). Each RNA migrated according to its own properties and the mixture of RNAs exhibited combined patterns where the two RNA bands were distinctly visible. As an unrelated positive control, dsRNA of Jil-1 3' (667 bp) was run simultaneously and migrated faster than the combined *roX1-3'* and *roX2* RNAs (Fig. 9B).
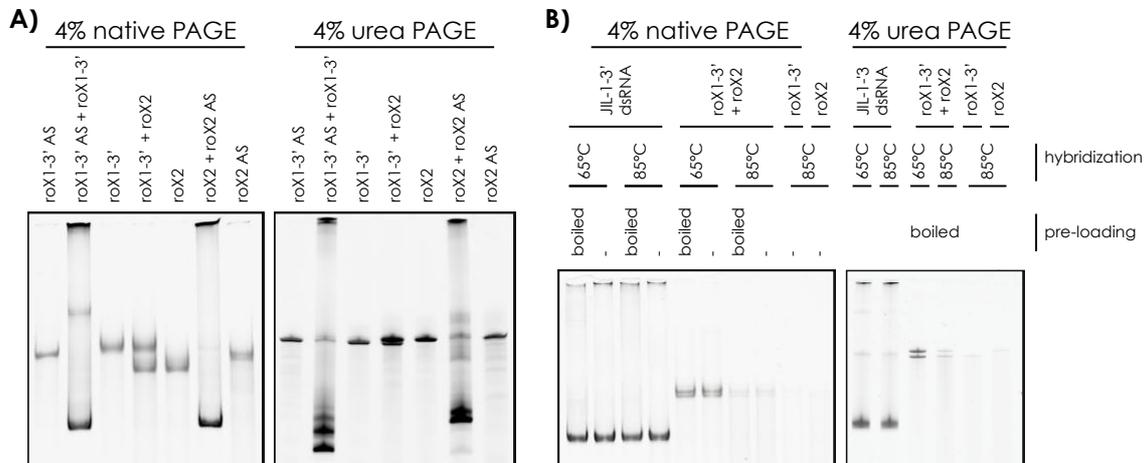

Figure 9. Attempts to recapitulate hybridization of *roX1-3'* and *roX2* RNAs failed. a) Mixing *roX1-3'* and *roX2* RNAs does not lead to a different migration pattern (lane 4) as seen in hybrids of either *roX* RNA with its AS RNA (lane 2, 6) on 4% native PAGE. As control, RNAs are run on 4% urea PAGE gel. b) As control, *JIL-1-3'* dsRNA of similar size runs faster upon hybridization, while *roX* RNAs do not after incubation at various temperature for hybridization.

Different parameters were tested to optimize the condition, including the temperature at which denaturation is done, the addition of a recombinant RNA helicase MLE at various

concentrations, and the salt molarity of the buffer (50 – 500 mM). Although positive controls worked, no hybrid formation could be observed between *roX1-3'* and *roX2* RNAs.

As these methods are rather low in resolution and may not report on low affinities between the RNAs, reverse-transcription quantitative PCR (rt-qPCR) was performed to measure pulled-down untagged RNA when a tagged RNA was used as bait. *roX1-MS2* RNAs were first incubated with MS2-MBP protein and amylose beads. Thereafter, untagged *roX2* RNA was added and let to hybridize. Upon several washes, RNAs bound to beads either through the MS2 tag or via hybridization to bait RNA were purified and reverse transcribed. qPCR indicated poor hybridization of *roX2* to *roX1*, regardless whether the 5' or 3' transcript was used, as pulled-down *roX2* RNA was miniscule (Fig. 10). In comparison, preliminary observation showed quantitative hybridization of input RNAs (Fig. 8D).
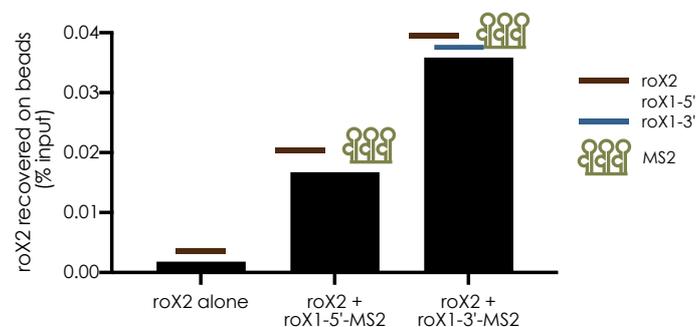

Figure 10. Miniscule untagged *roX2* RNA was pulled down using MS2-tagged *roX1-5'* and *roX1-3'* RNAs as quantified by rt-qPCR. Values were normalized to untagged RNA sticking to beads and represented as percentage of input.

6.1.3 miRoX2 degradation in NE of cells

Naturally, a reason why hybridization may not take place is the absence of an RNA binding protein or remodeller that was present in the initial experiment from the addition of nuclear extract. To see whether hybridization and degradation can be seen in the presence of proteins, various extracts were added upon hybridization and binding of MS2-tagged RNAs to amylose beads. In addition to NE, cytoplasmic extracts (CE) and whole cell extracts (WCE) were also used. Binding of RNAs to MLE and a protein of the miRNA degradation pathway, DCR2, were visible on a Western blot (WB) as shown before. However, degradation of RNAs were not observed upon the addition of NE (Fig. 11A) or WCE (Fig. 11B) of cell lines L24, S2, and KC.

The standard protocol dictates the denaturation and hybridization of both tagged and untagged RNAs first before binding to the MS2-MBP protein and immobilization onto amylose beads. Once incubated for 20 min, 1 mg extracts were added to allow for RNP reconstitution.

Another alteration introduced to improve the set-up was the order in which components were added to the reaction (reaction 2 vs. reaction 3 in Fig. 11). Assuming that the formation and subsequent degradation of hybrids might require the involvement of proteins in the extract, the untagged RNA was added simultaneously with extracts and incubated at room temperature to allow RNA processing to occur. Changing the order did not change the result and once again degradations of RNAs were not detected (Fig. 11).
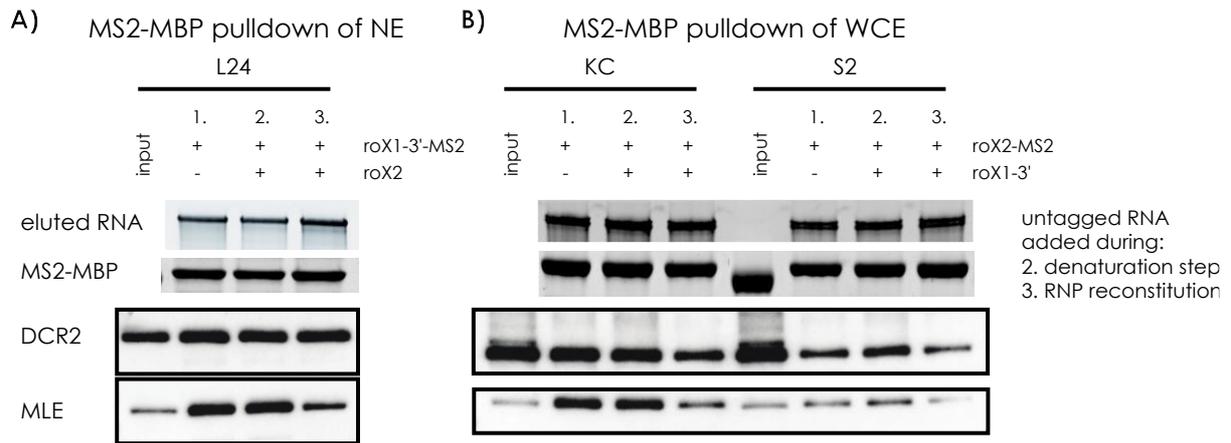

Figure 11. Incubation of *roX* RNAs in different nuclear (NE) or whole cell (WCE) extracts of various cell lines did not lead to degradation. Upon incubation in A) L24 NE and B) Kc or S2 WCL, *roX* RNAs recruited MLE and DCR2 as seen by Western blotting; however, no degradation was observed. Addition of untagged RNA to the reaction was either done during the denaturation step (reaction 2) or RNA reconstitution (reaction 3).

One last attempt to sort out parameters of the MS2-MBP affinity chromatography was to decrease the initial amount of RNAs while keeping all others the same. The standard protocol calls for an initial concentration of 333 nM individual RNA per reaction. In these experiments, a series of 1 in 5 dilution of input was done, with the lowest concentration being 2.7 nM of *roX1-3'-MS2* and untagged *roX2* RNAs each. Surprisingly, in all concentration tested, a decrease in RNA was observed upon the addition of 1 mg S2 NE (Fig. 12). Nevertheless, complete abolishment was seen when the least amount of RNAs were used, which represented 2.7% of the initial set parameters, and was likely occurring due to a low degree of RNAse activity of the NE. The observations could not explain why such clear-cut degradation in the initial experiments were not reproducible.
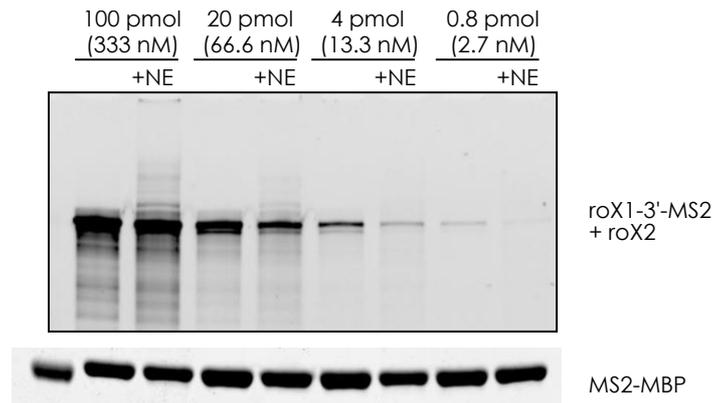
Figure 12. Titration of input *roX* RNAs into S2 NE did not lead to degradation. *roX1-3'-MS2* was used to bait untagged *roX2* RNA and analysis was done in an SDS PAGE.

6.1.4 miRoX2-expressing stable cell lines

For an *in vivo* exploration of the possibility that miRNA produced from *roX2* could directly target *Sxl* mRNA, a female Kc cell line (K9) was created bearing an integrated construct of a pri-miRNA termed *primiRoX2*, where *miRoX2* is packaged into the backbone of a ubiquitous miRNA, *dme-miR-1* (Fig. 13A) (178). A canonical miRNA is generated from a pri-miRNA whereby a hairpin structure becomes the source of an RNA duplex carrying the seed sequence. If the hypothesis were true, *miRoX2* miRNA would have to be produced by a noncanonical pathway, whereby the RNA duplex stems from the hybridization of two different lncRNAs. To simplify the system and possibly enhance expression and its consequence, a canonical scenario of *miRoX2* production was designed in a female setting. Expression of pri-miRNA is induced by $CuSO_4$ that activates the metallothionein promoter (Fig. 13B). Prior to experimentation, PCR was performed on isolated genomic DNA to ensure that the integrated construct had not been lost during maintenance, and indeed the construct was still fully integrated.

To probe for protein expression, K9 cells were treated with $CuSO_4$ for 10 days and protein extracts were analysed by Western blotting. SXL expression of K9 was compared to that of the parental Kc cell line, and a possible induction of MSL2 expression was analyzed against that of S2 cells. It can be concluded from a series of $CuSO_4$-treatment experiments that SXL decrease and concomitant increase in MSL2 did not occur (Fig. 13C). One hypothesis for the lack of effect may be that upon treatment, expression of *primiroX2* is not favourable to cell viability; therefore, those that do express are selected against and cells that do not express remain until the time of analysis. Shorter time-course experiments were performed to evaluate a more immediate effect, ranging from 24 to 96 hours; however, no observable difference in protein expression could be concluded.
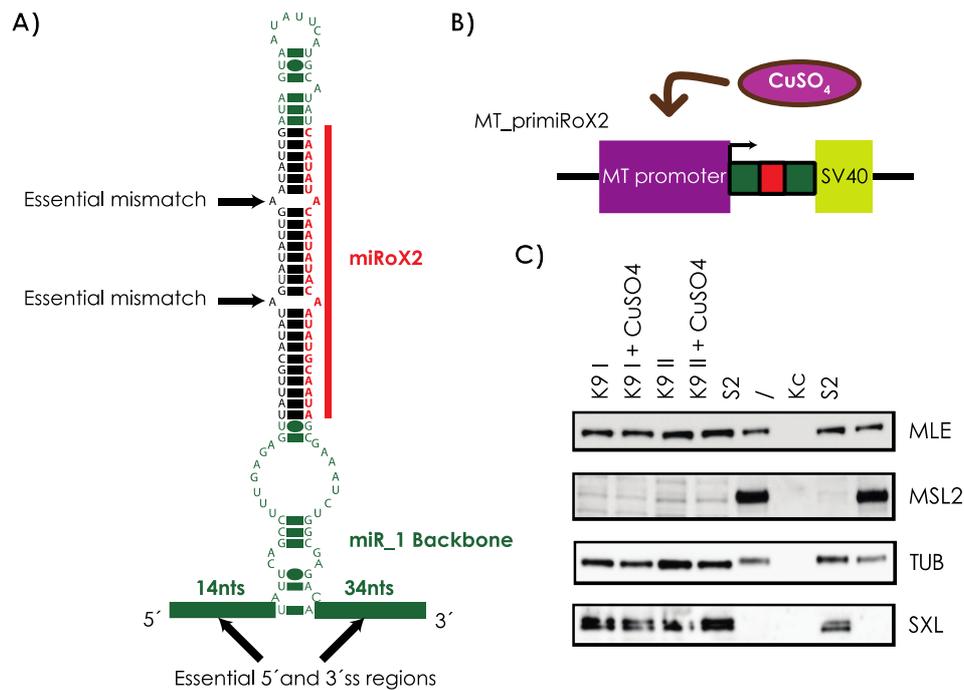
Figure 13. An *in vivo* system with inducible *primiRoX2* expression in K9 cells A) *primiRoX2*: *miRoX2* (red) is packaged into the miR_1 backbone (green). B) *MT_primiRoX2* construct in stable cell line K9 created by S. Maenner is inducible upon CuSO₄ treatment. B) Proteins of interest in K9 cells were probed upon a 10-day induction and compared to levels in Kc and S2 cells by WB.

To test for transcript changes that might not have affected protein level, TaqMan Advanced miRNA Assay was performed alongside conventional rt-qPCR. rt-qPCR allows for detection of the *primiRoX2* construct, whereas the qPCR-based TaqMan system enables highly sensitive quantification of mature *miRoX2* through a looped primer design that specifically recognizes specific target. Control amplicons behaved expectedly, whereby *GAPDH* and *dme-bantam* (179) featured similar Cp values of ~25 and ~18 in all samples, respectively. *primiroX2* expression, however, was not detected upon induction of K9 cells for7 days, and consequently, mature *miRoX2* was also not detected (Fig. 14). Cp values of either the pri-miRNA or the miRNA were well above 30, which is theoretically in the range of $2_{(35-30)}$ or 32 molecules per sample, if Cp of 35 is used as a cut-off of expression.

After multiple tries to recapitulate S. Maenner's observations, quantitative degradation of *roX* RNAs could not be observed in presence of extract. First and foremost, no hybridization was detected. Addition of various extracts did not lead to *roX* degradation. Finally, placing *miRoX2* sequence into a canonical miRNA backbone had no phenotype. If *miRoX2* exist, we were unable to detect it reproducibly.
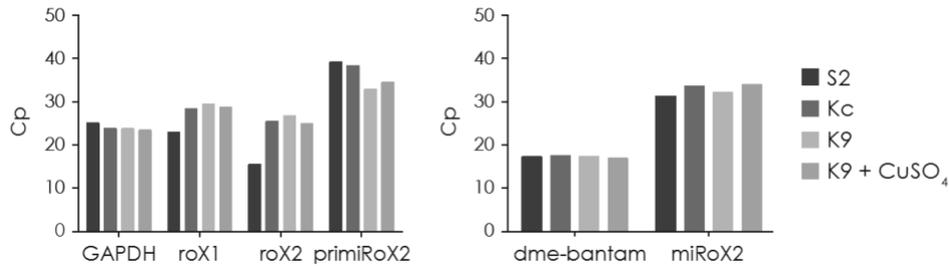
Figure 14. Expression of *primiRoX2* and *miRoX2* in K9 cells upon a 7-day $CuSO_4$ treatment were negligible as assessed by conventional qPCR and TaqMan Advanced miRNA assay, respectively.

6.2 Exploration into the biogenesis of functional *roX* RNAs

6.2.1 Processing of roX1 and roX2 RNAs

Redundancy of *roX* RNAs in DC has long been shown through various genetic experiments (116,118,121). The functional requirement of each *roX* RNAs is narrowed down to the *roX* boxes within SLs (93,125-127) and mutations (or lack thereof) of *roX* can be rescued by SLs from various related species (78,79,114). Interestingly, male cell lines that have been created from *D. melanogaster* preferentially express mainly one *roX* RNA, *roX2*. Reanalysis by T. Schauer of a recent transcriptomic study have also supported previous reports on heavily extensive post-transcriptional modification of *roX* RNA (117,129,131). Upon bulk RNA isolation from multiple stages of developing embryo, transcriptomic analysis of ribosomal RNA-depleted versus poly(A)-enriched samples pointed *roX2* RNA processing (Fig. 15). Furthermore, a switch in promoter choice seemed to occur as development progresses (Fig. 15).

Considering the complexity of *roX1* and *roX2* isoforms, characterization of the dynamics and stoichiometry of isoform expression becomes imperative as different isoforms may have different functions. A method to quantitatively examine expression of RNA is by means of rt-qPCR. A panel of primers was designed in such a way that combinations of reads could be used to deduce the isoforms. To do so, genomic sequence of *roX1* and *roX2* were first ascertained as qPCR is sensitive to single nucleotide differences in template. Sanger sequencing of gDNA retrieved from Oregon-R embryos and cell lines exposed annotated SNPs. The gDNA was thereafter used as standards in the quantitative determination of *roX1* and *roX2* in subsequent experiments.
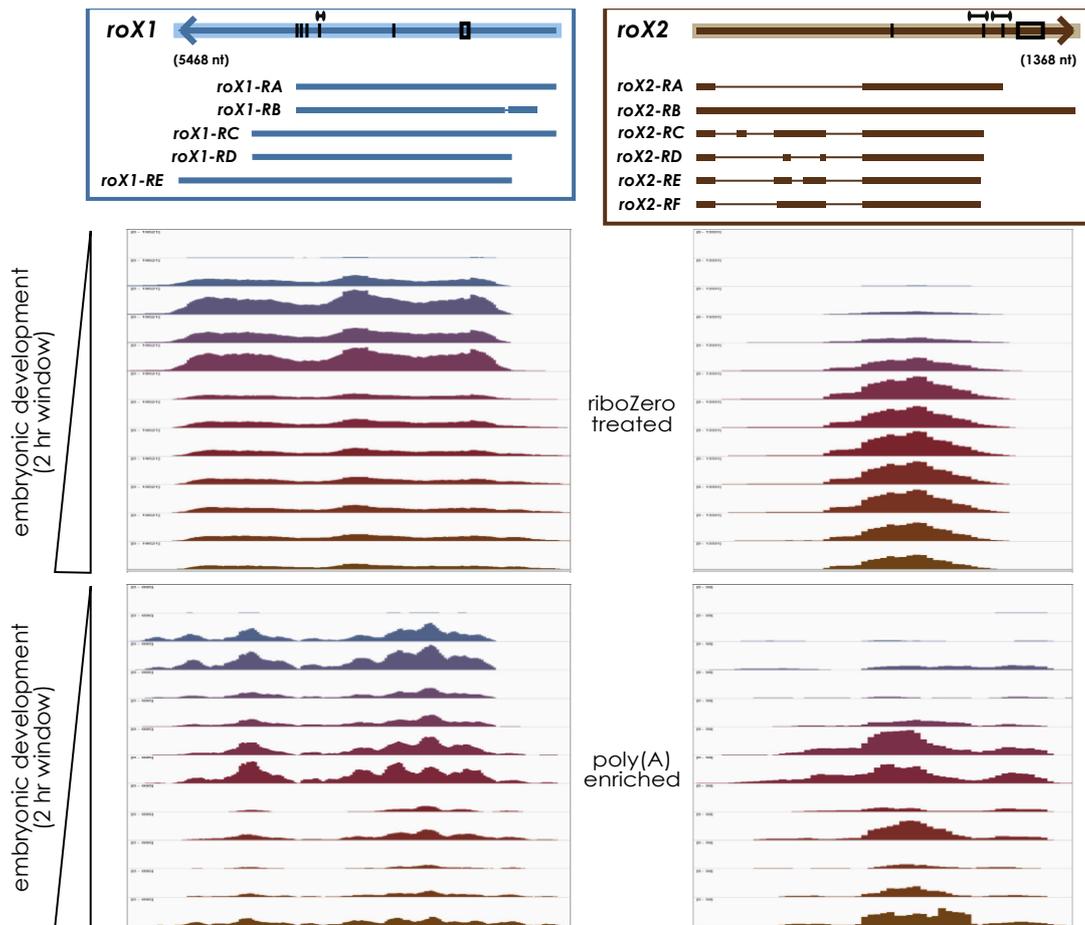
Figure 15. *roX*1 and *roX*2 RNAs in riboZero-treated and poly(A)-enriched transcriptomes during embryo development from Graveley et al., 2011, as reanalyzed by T. Schauer (131). *roX1* scale: 0-19521 and 0-695, *roX2* scale: 0-13331 and 0-58 in riboZero-treated and poly(A)-enriched transcriptomes, respectively.

6.2.2   *Quantifying isoforms of roX RNAs in embryo extracts*

To quantify and characterize early expression of *roX* RNAs, two types of embryo extracts were used. The first is a nuclear, transcriptionally active embryo extract (TRAX) prepared from 0-12 hrs old embryos (180). The second is a 0-90 min ael preblastoderm *Drosophila* embryo extract (DREX) that is rich in histones and factors of the chromatin assembly apparatus capable of reconstituting chromatin (181-183). Additional TRAX sample was kindly given by Andrea Lukacs and all DREX samples were gifts from Lisa Harpprecht.

Prior to *roX* analysis, nuclear enrichment during TRAX preparation was ascertained by WB (Fig. 16). Nuclear proteins, such as MSLs and lamin, were indeed enriched in TRAX and depleted from the cytoplasmic fraction (CF) collected during the preparation, indicating that the fractionation and solubilization have worked. Accordingly, maternally deposited proteins

were detectable in DREX, i.e. MOF, MSL1, MSL3, MLE, but those that require zygotic machinery to be translated were exclusively seen in TRAX, i.e. MSL2 and SXL.
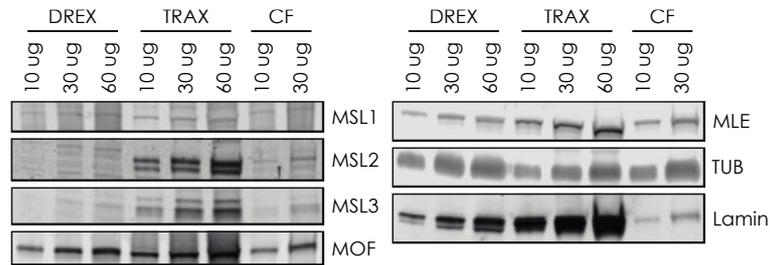

Figure 16. Fractionation analysis of DREX, TRAX and its cytoplasmic fraction (CF). Various amounts of extracts were loaded and DCC proteins were visualized. Tubulin and lamin served as controls. Tubulin is present in all extracts; however, lamin is highly enriched in TRAX compared to CF.

Next, the distribution of *roX* RNA in the TRAX, their corresponding CFs, and DREX were quantified using designed panel of primers (Fig. 17). Genomic DNA of embryos was used as standard to determine absolute amounts. First and foremost, it was surprising to detect both *roX* RNAs in DREX. Nonetheless, given the similar amounts of starting material, TRAX yielded much more *roX* RNAs, reflecting their steady-state transcription in late male embryos. A likely source of the measured *roX* RNAs in DREX may be contamination of older embryos during collection. A ratio analysis of *roX1* RNA in TRAX versus DREX revealed that there might have been a promoter switch between the 0-90 min ael and 0-12 hrs ael extracts, as amplicon 1 indicated a greater enrichment over amplicon 3 (Fig. 17A). Furthermore, the long isoform of *roX1*, *roX1-RE*, was less abundant as amplicon 7 indicated the least gain of all. These observations are in line with previous findings (see above) of *roX1* RNA undergoing a promoter switch and possibly alternative termination process through development (Fig. 17A). Concurrently, similar analysis of *roX2* RNA indicated that alternative poly(A) sites might be used as development proceeds, as significant drop was seen between amplicons 4 and 5 with amplicon 6 (Fig. 17B). This trend is also in line with previous data (see above) of *roX2* having alternative poly(A) sites that are functional at different stages of development (Fig. 17B).

Amplicon	1	2	3	4	5	6	7
TRAX	1,159	2,220	10,702	38,811	75,856	70,824	169
cytoplasmic	288	462	6,511	14,958	29,248	15,024	1,103
DREX	62	189	1,384	3,462	6,813	3,922	52

ratio	1	2	3	4	5	6	7
TRAX/DREX	19	12	8	11	11	18	3

fractionation %	1	2	3	4	5	6	7
TRAX	77	80	53	66	67	78	16
cytoplasmic	23	20	47	34	33	22	84

Amplicon	1	2	3	4	5	6
TRAX	11,319	1,089	257,849	105,773	29,401	1,146
cytoplasmic	221	27	2,728	1,003	449	96
DREX	190	14	2,379	581	462	36

ratio	1	2	3	4	5	6
TRAX/DREX	60	81	108	182	64	31

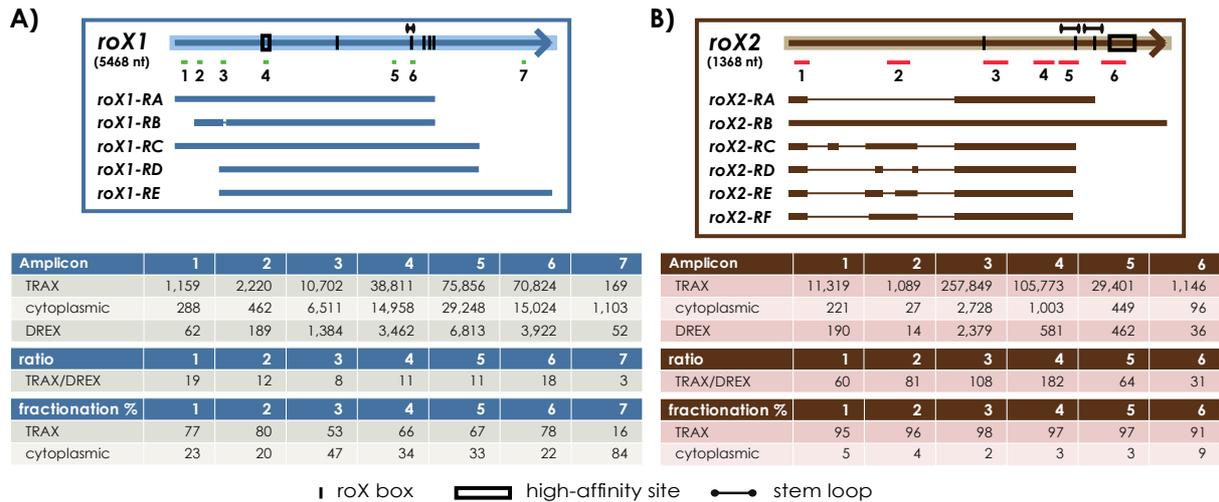| fractionation % | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| TRAX | 95 | 96 | 98 | 97 | 97 | 91 |
| cytoplasmic | 5 | 4 | 2 | 3 | 3 | 9 |

Figure 17. Quantifying A) *roX1* RNA and B) *roX2* RNAs in 1 ng of total RNA of extracts. Ticks represent *roX* boxes, boxes denote high-affinity sites, and bulleted dashes indicate *roX1* SL and *roX2* SL7 and SL8. Introns are drawn in light shade, whereas exons in dark. Numbered green and red dashes depict amplicons. TRAX and CF (n=3 each), DREX (n=2).

Another interesting aspect disclosed by the TRAX to DREX ratio was enrichment of *roX2* over *roX1* as development progresses. Taking into account amplicons specific for the major exons of *roX1* and *roX2* (amplicon 5 of *roX1* and amplicon 3 of *roX2*), there was roughly 3 *roX1* for every *roX2* RNA in DREX. On the other hand, 2 *roX2* was measured for every *roX1* RNA in TRAX (Fig. 17). Remarkably, quite a significant amount of *roX1* RNA were detected in the cytoplasm (Fig. 17). So far, such an export of *roX1* RNA to the cytoplasm has not been described. As *roX* RNAs are also known to undergo intensive post-transcriptional processing, another interesting aspect of fractionation is whether there is a bias towards certain compartments if RNAs were polyadenylated. Similar TRAX vs. CF analysis was done upon a poly(A)-selection protocol (Fig. 18). Interestingly, up to 40% of total *roX1* and 90% of polyadenylated *roX1* was cytoplasmic, whereas a mere 5% of total *roX2* and up to 60% of polyadenylated *roX2* was measured in CF. Specifically, cytoplasmic RNAs tended to be the long isoforms of *roX* RNA, *roX1-RE* and *roX2-RB*. Differences in the distribution of *roX* suggested differences in processing as well biogenesis of these RNAs, which may be interesting functionally. To note, *GAPDH* amplicon used as control behaved as expected, whereby the majority of total and polyadenylated mRNAs are cytoplasmic (Fig. 18).

fractionation %	roX1 amplicon 2	roX1 amplicon 5	roX2 amplicon 3	roX2 amplicon 5	roX2 amplicon 6	GAPDH
TRAX – total	60	85	96	95	70	10
cytoplasmic – total	40	15	4	5	30	90
TRAX – polyA	10	25	40	80	44	0.5
cytoplasmic – polyA	90	75	60	20	56	99.5

Figure 18. Percentage of total and polyadenylated *roX* RNAs in TRAX and the corresponding cytoplasmic fractions (CF) as quantified by qPCR. Two sets of primers were used for *roX1* (primer sets 2 and 5) and *roX2* (primer sets 3 and 5) quantification. As control, amplicon of a GAPDH was used. Both total and polyadenylated *GAPDH* were cytoplasmic.

6.2.3 Isoforms of roX RNAs in cell lines

To study the interplay of *roX* expression as well as isoform distribution in previously used Cl.8 and S2 cells, quantification of *roX* RNA was done by rt-qPCR using the same panel of primers as before. To start, the bulk population of *roX* RNAs was determined in Cl.8 cells and compared to that of S2. There was roughly 2600 *roX2* for every *roX1* RNA in S2 cells. On the other hand, Cl.8 equally expressed them (Fig. 19A). Fractionation and subsequent compartmental analysis showed contrastingly that *roX1* RNA of Cl.8 cells were mostly nuclear, having similarly small proportion of cytoplasmic fraction as that of *roX2* (Fig. 19B). For both *roX* RNAs, however, the most 3' amplicons showed a slight enrichment in CF compared to others, suggesting that at least some polyadenylated isoforms might shuttle in these cell lines as well.

A)

	roX1	roX2	total roX	roX1 : roX2
Cl.8+	421,111	392,911	814,022	1 : 1
S2	549	1,461,667	1,462,215	1 : 2,665

B)

% in Cl.8	1	2	3	4	5	6
NE	92	92	93	94	95	83
cytoplasmic	8	8	7	6	5	17

% in Cl.8	1	2	3	4	5	6
NE	95	89	95	98	95	78
cytoplasmic	5	11	5	2	5	22

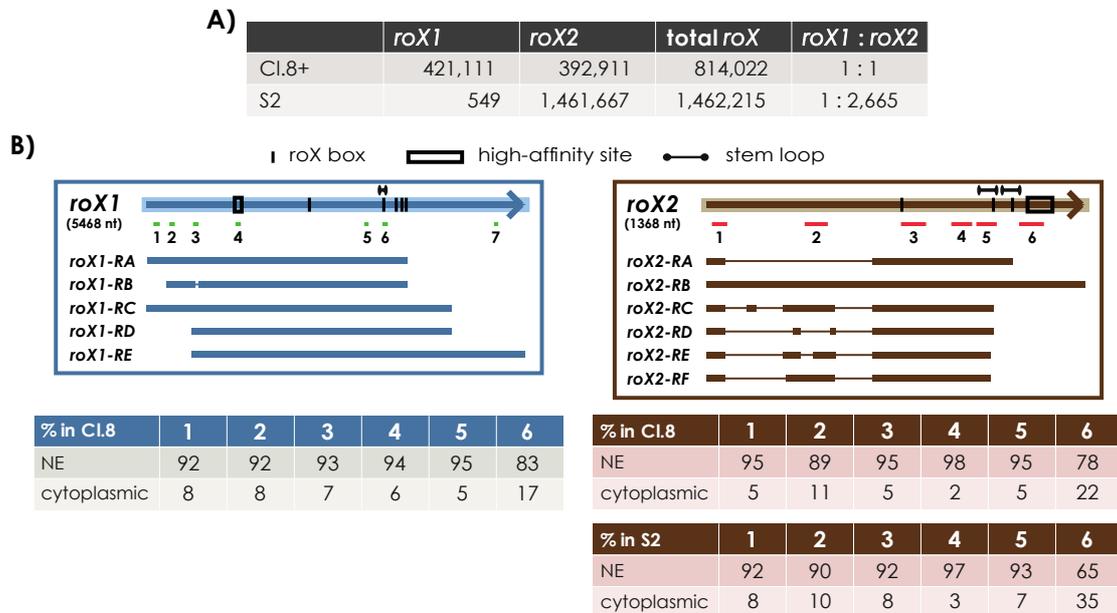| % in S2	1	2	3	4	5	6
NE	92	90	92	97	93	65
cytoplasmic	8	10	8	3	7	35

Figure 19. *roX* RNA expression in Cl.8 and S2 cells as quantified by rt-qPCR. A) Amounts of *roX1* and *roX2* RNA measured in bulk using amplicon 5 and amplicon 2, respectively.

6.2.4 Direct-RNA nanopore sequencing of total TRAX RNA

Naturally, limitations are present with the use of qPCR to identify and quantify isoforms of RNA. To resort to a better and more accurate method, a pilot experiment of direct-RNA nanopore sequencing was performed on isolated RNA from TRAX. Ionic current passes through the nanopore as measured by potential difference across a membrane (184). As each nucleotide of an RNA transcript is pulled through the nanopore, a change in current occurs specific for the given nucleotide (184). Through machine learning of base-calling, the sequence of the entire RNA can be deduced. The advantage of direct-RNA nanopore sequencing over conventional qPCR is the elimination of biases during PCR amplification and reverse transcription. It also enables long-read sequencing of potentially full-length transcripts up to ~2000 bases long (185). Furthermore, the possibility exists to identify any kind of post-transcriptional RNA modifications as the read-out can be sensitive enough to these changes, such as m6A (186). Sequencing of native *roX* RNAs by direct-RNA nanopore sequencing without fragmentation would reveal isoform identities and modifications.

A caveat of the method, however, is the need to pull RNA through a nanopore, which on its own has no specificity. It is equipped with a motor protein that feeds RNA into the nanopore and do so by recognizing poly(A) tails. In order to get an unbiased profile of all native RNA present in the TRAX, *in vitro* polyadenylation had to be done first. To assess efficiency, *in vitro* transcribed *roX2* RNA was used as a reference. The single *roX2* RNA population and the resolution of RNA analysis by Bioanalyzer enabled the determination of polyadenylation rate (Fig. 20). It is of importance that each RNA has a poly(A) tail, but not too long to avoid the mapping of mostly poly(A) tails. Thus, an optimization in the reaction setup was done to allow just enough polyadenylation.
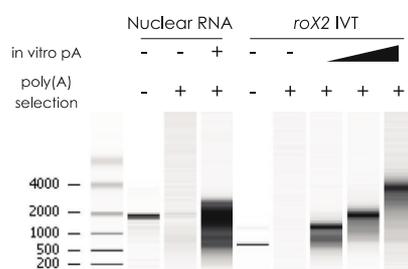

Figure 20. The progression of polyadenylation of nuclear RNA and *in vitro* transcribed *roX2* RNA were traced by Bioanalyzer analysis.

Once the *in vitro* polyadenylation step had been optimized, libraries were prepared on isolated TRAX RNA that had first been depleted of ribosomal RNA using the rRNA Depletion

Kit (Fig. 21A-B). Performing depletion once or twice greatly reduced the rRNA peaks at 2 kB; however, double depletion also removed additional RNA larger than 500 bp (Fig. 21A). At the fear of losing larger *roX* RNA fragments, the single rRNA depleted sample was *in vitro* polyadenylated. A shift of RNA population towards larger size indicated that polyadenylation had worked (Fig. 21B). Population of sequenced reads mimicked the library profile, in which majority of fragments were between 0 and 500 bases (Fig. 21B-C). Unfortunately, larger reads seemed to be scarce. Furthermore, up to 50% reads still originated from ribosomal RNA despite the rRNA depletion step. Nonetheless, pilot sequencing appeared to work as mRNA of genes that are abundant in cells, such as ribosomal protein RpL32, were well detected and its long isoform sequenced (Fig. 21D).
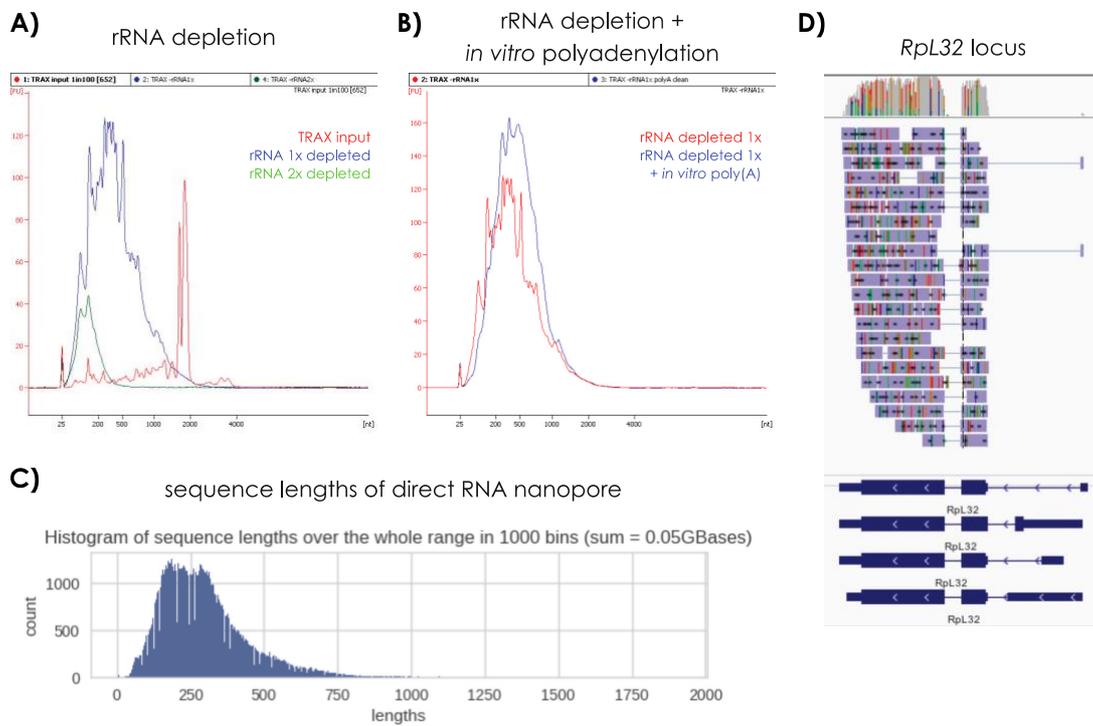

Figure 21. RNA library preparation for direct-RNA nanopore sequencing. A) Total RNA from TRAX was rRNA depleted once or twice. B) Once-depleted TRAX RNA was *in vitro* polyadenylated. C) Distribution of sequence lengths from Nanopore sequencing resembled the library profile in B). D) Reads aligning to the *RpL32* locus.

Only two sequences of *roX2* and one of *roX1* were obtained and mapped in this pilot experiment (Fig. 22). As expected, mapped reads stemmed from the most abundant parts of the *roX* RNAs. *roX2* reads aligned to the major exon that is present in all isoforms. In case of *roX1*, the only aligned read represented part of *roX1* sequence containing the *roX* box. Nanopore

sequencing proved to be a potential tool to study *roX1* and *roX2* heterogeneity; however, further optimization is needed.

Figure 22. Direct-RNA nanopore sequencing reads for *roX1* and *roX2*.

6.3 Establishment of dosage compensation

6.3.1 Transcriptomic measure as read-out of DC

The aftermath of DCC acting on the X chromosome is an upregulation of its transcription. This presents an opportunity to use transcriptomics of embryos in various early stages as a read-out of when dosage compensation starts and proceeds. Single-embryo RNA sequencing of embryos performed by Tamas Schauer indeed highlighted peculiarities of the initial timing of *roX* RNA expression and dosage compensation during embryogenesis that are in line with previous reports (116,130). It was possible to determine the sex of each embryo from the sex-specific gene expression in each dataset. Accordingly, a male balanced state can be construed as equal X expression relative to female embryos. Eight early stages were chosen for analysis based on phenotypical development as described by Bownes (4), the first of which was before ZGA (during formation of primordial germ cell) and therefore before sex determination. Phenotypical markers for latter stages included cellularization, early gastrulation, germ band elongation, stomodeum formation, germ band retraction, head involution, and dorsal closure.

As expected upon comparison of median transcript levels, female embryos maintained an X to A ratio of ~1 as soon as it took control of its own transcription. Male embryos, however, required some time to reach the same ratio and were there ~10 hrs later, surprisingly long in time for DC progression (Fig. 23A). Nonetheless, this type of comparison has its caveats, as different pools of genes with different types of regulation are weighted against one another. For a less biased comparison, a fold change expression of female over male was computed for every gene and the average value was compared for every chromosome. Conceivably, autosomal genes were expressed similarly in all stages of interest, whereas X-chromosomal genes acquired a female bias upon ZGA, which was slowly neutralized with development (Fig. 23B). Looking deeper into expression patterns, 20% of the least varied genes between stages were considered as constitutive, as supported by GO terms (metabolic processes, transcription,

translation) and acquired a balanced state rather quickly (Fig. 23C-D). Contrastingly, 20% of the most varied genes across stages were classified developmental as advocated by GO terms (anatomical and organ development) and maintained their female bias longer (Fig. 23C-D). To note, autosomal developmental and constitutive genes were equally expressed in male and female (not shown).
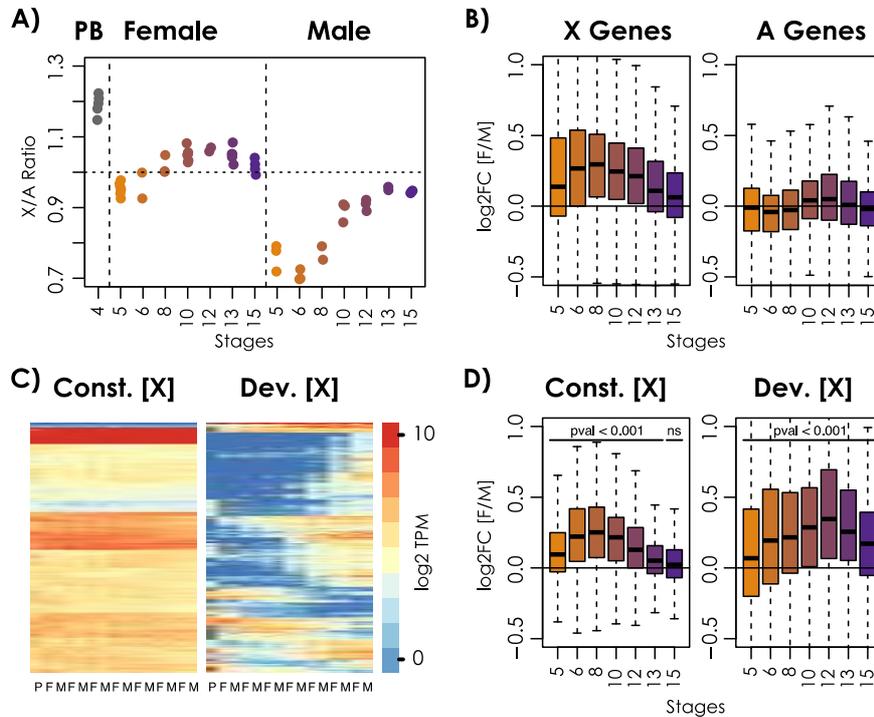

Figure 23. Single-embryo RNA-seq as a measure of compensation. A) Ratio of X-chromosomal over autosomal median transcript abundances (X/A) of preblastoderm (PB), female and male embryos in 7 developmental stages. Dotted horizontal line represents a ratio of 1, balanced transcript levels. B) Log2 Fold Change (log2FC) of RNA-seq between female (F) and male (M) embryos of expressed genes on X chromosome and autosomes in 7 developmental stages. C) Heatmaps of log2 TPM (RNA-seq) values for X-chromosomal constitutive, 20% least variant, (left) and developmental, 20% most variant (right) genes. D) log2FC of RNA-seq of X- chromosomal genes which are constitutively or developmentally expressed. Data was generated by T. Schauer.

According to Meller, *roX1* RNA is transcribed in all embryos very early in development, whereas *roX2* RNA is exclusively transcribed to a significant level in males beginning at stage 9 when stomodeum appears (~6 hrs ael) (130). Consistently, *roX* RNA expression was similar in the single-embryo RNA sequencing dataset (Fig. 24A). Examining the rest of the DCC components, single-embryo transcriptomics confirmed the maternal contribution of *msl-1*, *msl-3*, *mof*, and *mle* (Fig. 24A). As reviewed before, both *Sxl* and *msl-2* are regulated at the post-transcriptional level (50), explaining why their RNAs were detected in both sexes.

Nevertheless, *Sxl* transcript level was higher in females and *msl-2* transcript level was higher in males (Fig. 24A) and their protein expressions were sex-specific (Fig. 24B).
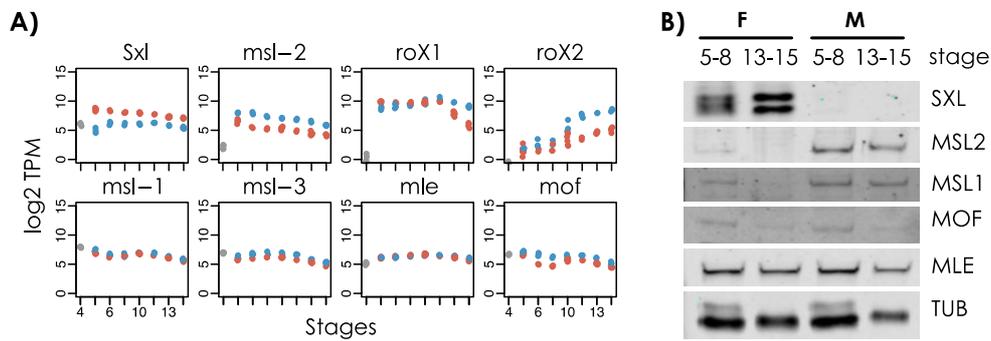

Figure 24. Expression of DCC components during timepoints of interest were determined in terms of A) transcript levels in single embryos and B) protein levels in sex-sorted embryos. GFP driven by the *SxlPm* promoter was used to mark female embryos.

6.3.2 DCC by IF in developing embryos

Previous immunofluorescence microscopy analyses described the emergence of MSL signal during the blastoderm stage of a male embryo, whereupon their colocalization on the X territory is only obvious by stage 8 (107,108). To see if with improved technology and sensitivity earlier observation of MSL staining was possible, embryos at various stages during development were stained with antibodies against various MSLs and H4K16ac. To set up the staining and imaging conditions, older embryos that were at least in stage 14, ~12 hrs ael, were used. Once optimized, male embryos were easily distinguishable from females as those having very clear confined signal of MSL2 and MSL3 in the X-territory of every nuclei (Fig. 25 panel 4 and 5). Furthermore, clear colocalization of MSL3 and H4K16ac was identifiable (not shown), indicating that at this stage of development, DCC is present and active in its function to specifically acetylate H4K16 of the X chromosome.

Delving a little earlier in development to embryos of stage 8, ~4 hrs ael, male embryos were also discernible through colocalization of MSL2 and MSL3 (Fig. 25 panel 2 and 3), as well as MSL3 and H4K16ac (not shown), at nuclei specifically undergoing mitosis. During gastrulation, certain localized cells of the embryo undergo division in order to support invagination and folding of tissue. Since nuclei compact during these divisions, concentrated amounts of MSL2, MSL3 and H4K16ac allow for signal detection on the X, which is absent in the rest of the nuclei.
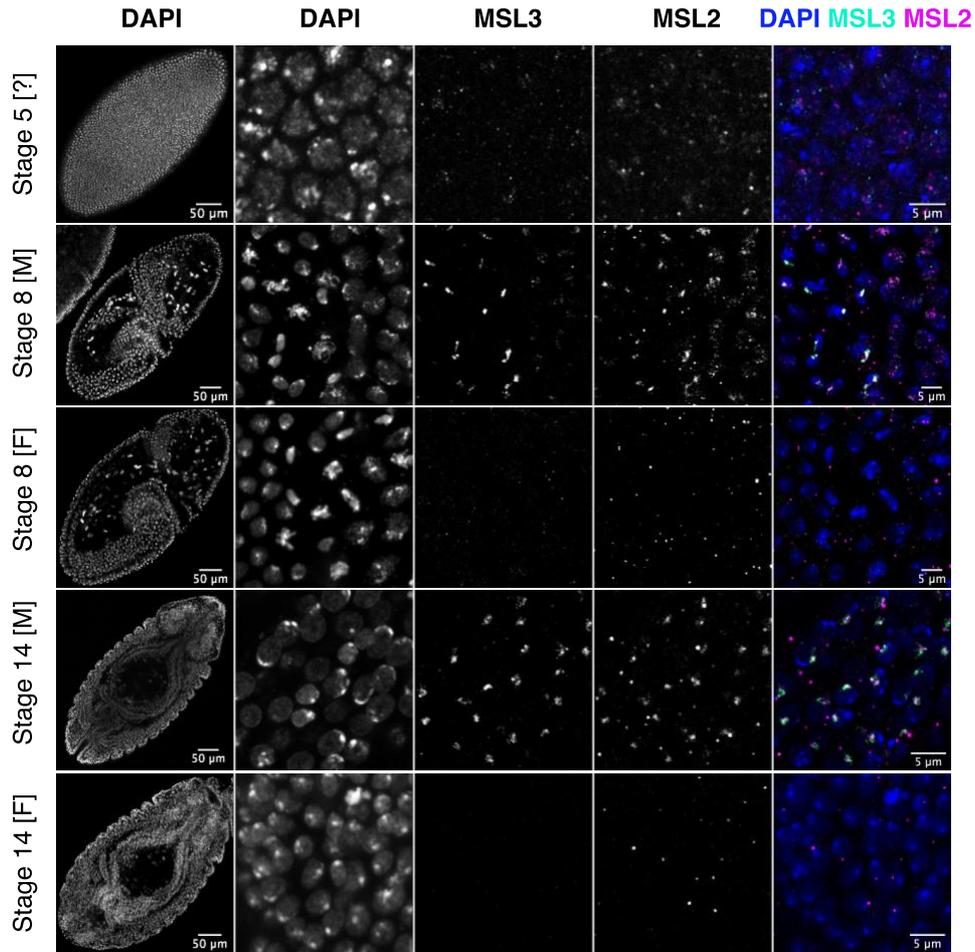
Figure 25. **DAPI MSL3 MSL2** IF staining on embryos at various stages of development. At stage 5, sex of embryos is not yet definable as obvious X-territories cannot yet be outlined. During gastrulation of male embryo [M] at stage 8, certain cells undergo mitosis, during which they condense and concentrate signals of MSL2 and MSL3. By stage 14, X territories are distinctly marked by MSL2 and MSL3 in all nuclei of a male embryo. Female embryos [F], whether in stage 8 or stage 14, do not attain MSL2 and MSL3 signal.

Looking into embryos in the blastoderm stage proved to be a bigger challenge, whereby signal and background were not discernible (Fig. 25 panel 1). This was likely due to the low amount of protein of interest that is present at this stage of development and its distribution over a larger area. The first seven of 14 nuclear divisions occur within a window of ~8 min (5). To allow for such speed, nuclei at this stage are present in a decondensed state compared to dividing cells at later stages. The compaction of nuclei coincides with longer cycle times and correlates with increasing concentration of H1. By the end of nc 14 nuclei sizes decrease by 5- to 20-fold (187). Therefore, it is likely that lower amount of protein distributed over a larger area brings the signal below threshold of detection and therefore cannot be observed. Nevertheless, what appeared to be "shadows" of signal might indicate the start of X-territory

appearance, which was most obvious for MSL2 (Fig. 25 panel 1), the pioneering factor of the complex, and least obvious for H4K16ac, the product of an active complex. As immunofluorescence microscopy was not sensitive enough to resolve subtle changes in DC progression during the time window of interest, genome-wide profiling of DCC was performed.

6.3.3 Genome-wide DCC distribution during embryonic development

Since all components of the DCC appear to be present in these early stages of development where compensation is still incomplete, chromatin interaction of DCC members were probed by chromatin immunoprecipitation followed by sequencing (ChIP-seq) to determine whether this discrepancy was due to ineffectiveness of DCC and to understand mechanistically the process of compensation. MSL2 and MOF were chosen as proxy of DCC presence, and H4K16ac reveals DCC activity. Two windows of time during development were selected, the first encompassing 3 to 4 hrs ael, where most female bias was observed, and the second between 11 to 12 hrs ael, where constitutive genes were fully compensated. qPCR of developmental genes upon an H3K36me3 ChIP confirmed the embryo staging (Fig. 26A) and bioinformatics analyses were performed by T. Schauer.

Past experience has proven ChIP to be quite sensitive to methods of solubilization and antibody sensitivity/cross-reactivity. In order to produce high-resolution chromatin interaction profiles, chromatin preparation was optimized, and antibody concentrations were titrated. Fragmentation through a combination of Micrococcal Nuclease (MNase) digestion and Branson sonication did not improve yield compared to MNase alone. An alternative shearing method by Covaris ultrasound increased DNA recovery but also destroyed the key protein, MSL2. Thus, the main method chosen was solubilization by MNase to preserve the integrity of factors. MNase, an endo-/exonuclease, digests nucleic acids until it hits an obstacle, e.g. a DNA-binding factor or a nucleosome. As a result, this solubilization method is also suitable to determine nucleosomal positioning and histone modification profiles. Nevertheless, it comes with its limitation: MNase largely acts based on the accessibility of nucleic acids; however, some sequence preference has been observed where it cleaves upstream of AT-rich region much more efficiently (188). On one hand, over-digestion will lead to digestion of histone-associated DNA and will destroy MNase-sensitive nucleosomes; on the other hand, under-digestion introduces noise during sequencing and will not truly reflect the nucleosomal landscape (189). To complicate matters, ChIP efficiency of various antibodies is affected by the degree of digestion. Hence, to lessen technicality differences, especially when using different input samples, a ratio of 4 to 1 of mono- to di-nucleosome was aimed for in every

MNase digest and a ChIP titration was done for every antibody. To evaluate the efficiency of pull-downs prior to sequencing, enrichment over X-chromosomal HAS for MSL2 and gene body enrichment of active X-linked genes for MOF and H4K16ac was determined by qPCR (Fig. 26A).

An additional bioinformatics tool utilized was the ability to deduce lengths of DNA fragments upon paired-end sequencing of a ChIP experiment (190). Direct DNA binding of a factor allows crosslinking between DNA and the binding factor, marking a footprint appropriate to the size of the binding site. Protein-protein crosslinking events can also occur such as between nucleosomes and the DNA binder, protecting fragments of nucleosomal size. MSL2 is the only DCC subunit able to bind DNA (82-87). To specifically consider direct contacts of MSL2 with DNA, small fragments of sub-nucleosomal lengths, i.e. of 10 to 130 bp, were subset. On the other hand, longer fragments may indicate contacts of MSL2 with nucleosomes neighbouring the DNA binding site directly or indirectly via the DCC complex. To analyse such secondary occurrences, fragments of nucleosomal lengths, i.e. of 130 to 220 bp, were referred to and termed chromatin interaction.

In the early time window, MSL2 was present as seen by WB and was bound to DNA (Fig. 25-26). Surprisingly, MOF and H4K16ac were also already abundantly present in the early time window and overlapped well with each other genome-wide (Fig. 26).

Naturally, MSL2 binding at PionX sites and HAS were of interest and indeed, appreciable signal corresponding to direct DNA binding and to chromatin interactions up to ~2 kb away from these binding sites were found enriched (Fig. 27A). Early and late interaction profiles of MSL2 were very similar. It appears that within few hours after egg laying, MSL2 finds its target site on the X chromosome and "spreads" to neighbouring chromatin and the spreading was not more extensive at the late time.

Similarly, MOF interaction was stably detected at and around PionX sites and HAS (Fig. 27A). Unlike MSL2, MOF does not have the capacity to bind DNA on its own, hence, its presence at PionX sites and HAS could only be mediated by its assembly into DCC even at the early time window. Chromatin binding of MOF also overlapped well with that of MSL2 around these sites and intensified between early and late time windows to indicate ongoing assembly.
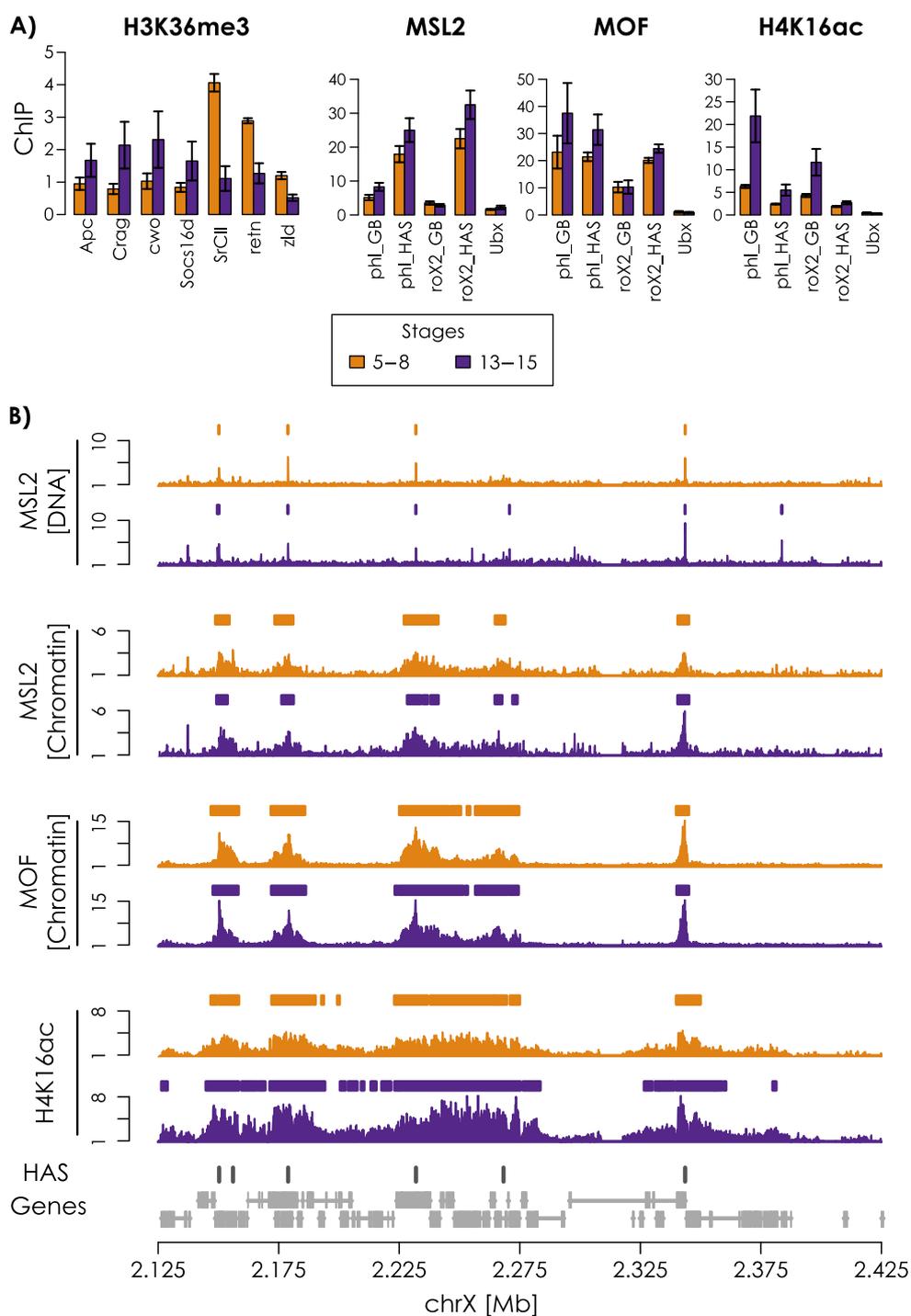
Figure 26. Genome-wide chromatin interaction profiles of MSL2, MOF and H4K16ac in early (orange) and late (purple) time windows. A) Stages were evaluated by qPCR upon an H3K36me3 ChIP. Amplicons chosen reflected up- or downregulated genes upon development, as described in Table. 4. Appropriately, *apc*, *crag*, *cwo*, and *socs16d* were upregulated, whereas *Sr-CII*, *retn*, and *zld* were downregulated. qPCR was also used to measure efficiency of MSL2, MOF, and H4K16ac ChIP, whereby MSL2 was better enriched at HAS, and MOF and H4K16ac at gene bodies (GB). Error bars represent SEM. B) Distribution profiles of MSL2, MOF and H4K16ac enrichment upon MNase-ChIP, normalized to input and total amount of reads. DNA and chromatin binding of MSL2 were differentiated through fragment length.

Called peaks/regions are represented as ticks/boxes above the tracks, respectively, and HAS below.

To expand the analysis genome-wide, peaks of sub-nucleosomal and regions of nucleosomal fragments for each time window were called for MSL2 and MOF. To correlate them to the functional status of DCC, H4K16ac regions were also called. Upon comparison of MSL2 DNA and chromatin interaction, three different patterns were unveiled (Fig. 27B-C). First, 139 MSL2 peaks resided within regions of MSL2 association. They featured the published MRE motif and were mostly known HAS that often reside within introns (86,191). These sites seemed to be robust initial DNA binding of MSL2 that led to an X-enriched chromatin "spreading" and were termed "functional" with respect to establishing DC. Regions around these sites were also rich in MOF, indicating the assembly of DCC. Second were 202 isolated DNA binding events that did not reside in regions of chromatin binding. Unlike the previous group, they lacked the GAGA-rich characteristic of MRE, and the majority were found on autosomal chromosomes. The lack of chromatin interaction suggests absent MSL3, leading to the conclusion that these binding events are rather "non-functional" in terms of DC. Indeed, almost no MOF were found at these sites. These newly identified sites might indicate transient interactions with accessible DNA detected through our optimized protocol. Last but not least, 199 chromatin binding regions were called that did not carry a direct DNA binding site, but nonetheless displayed binding of MSL2 and MOF. These regions without peaks did not have defined nucleosomal positioning and were not as intense in signal like the first group since they were plotted around the center of regions and not on the highest signal which would accumulate on a direct binding site.

X-enriched regions of MSL2 were further subdivided into two groups depending on whether they were present early or late. Upon distance analysis to the closest MSL2 DNA binding site, a correlation of time and space was revealed. Late appearing regions were farther than the average distance away from the nearest MSL2 DNA binding site (Fig. 28A). Furthermore, intensities of MSL2 and MOF in the early regions did not change much but slight accumulation in the late appearing ones were evident (Fig. 28B). Nonetheless, it appears that the entire DCC complex is present where they should be already at the earlier time window and the small differences in MSL2 and MOF profiles are inadequate to explain the disparity of dosage compensation between the two windows of time.
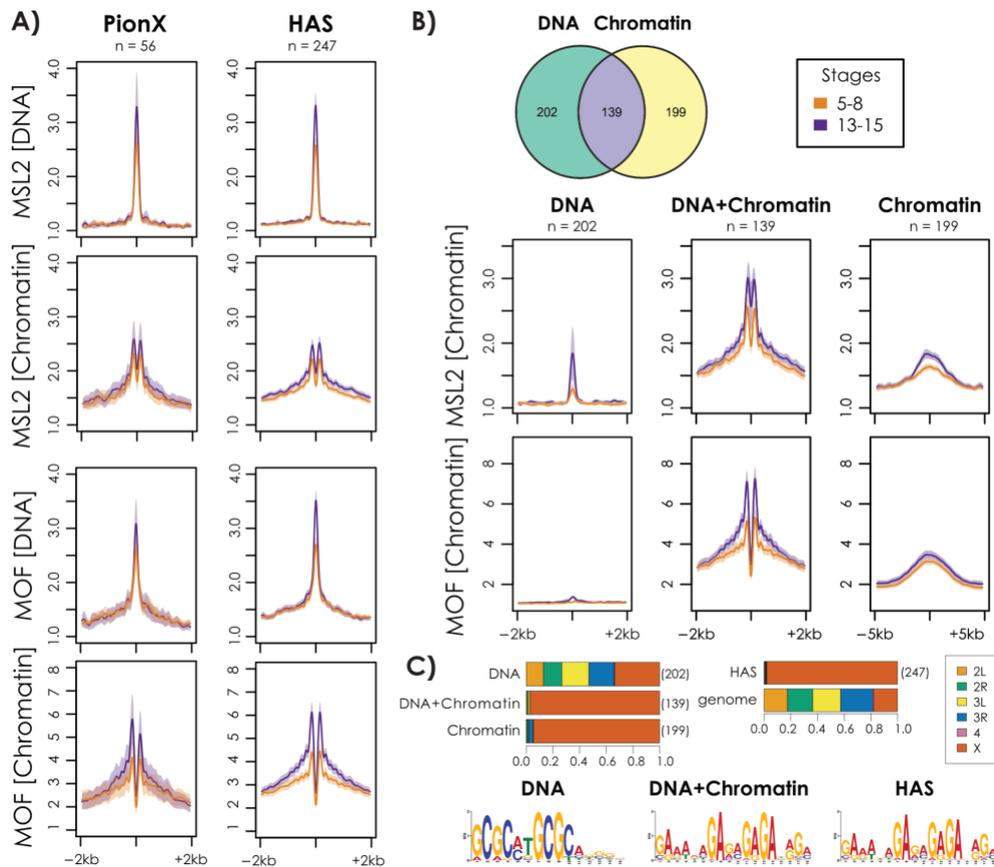
Figure 27. DCC was present on the X chromosome at the early time window. A) MSL2 and MOF ChIP-seq reads were subset to sizes of 10-130 bp [DNA] or 130-220 bp [Chromatin] and plotted over known PionX sites and HAS. B) Peaks and regions of MSL2 were grouped based on overlap as represented in the Venn diagram. Average composite plot of [Chromatin] were generated for each group for MSL2 and MOF. C) Relative genome distribution of peaks and regions grouped in B and the most significant *de novo* motif analysis of peaks grouped in B.

The delay of dosage compensation was clearly not due to the absence of MOF, but it was possible that MOF was not active as a HAT. Indeed, while the extent of MSL2 and MOF chromatin binding was unchanged, all averaging around 4 to 6 kb, H4K16ac regions extended from an average of 3 kb to 15 kb as development proceeded (Fig. 28C). Regions of acetylation extended far outside of MOF regions (Fig. 26B) specifically at X-chromosomal active gene bodies with time, while its writer, MOF, did so only to less degree (Fig. 28D). This suggests that MOF may reach out from its binding sites to acetylate remote chromatin through transient interactions.
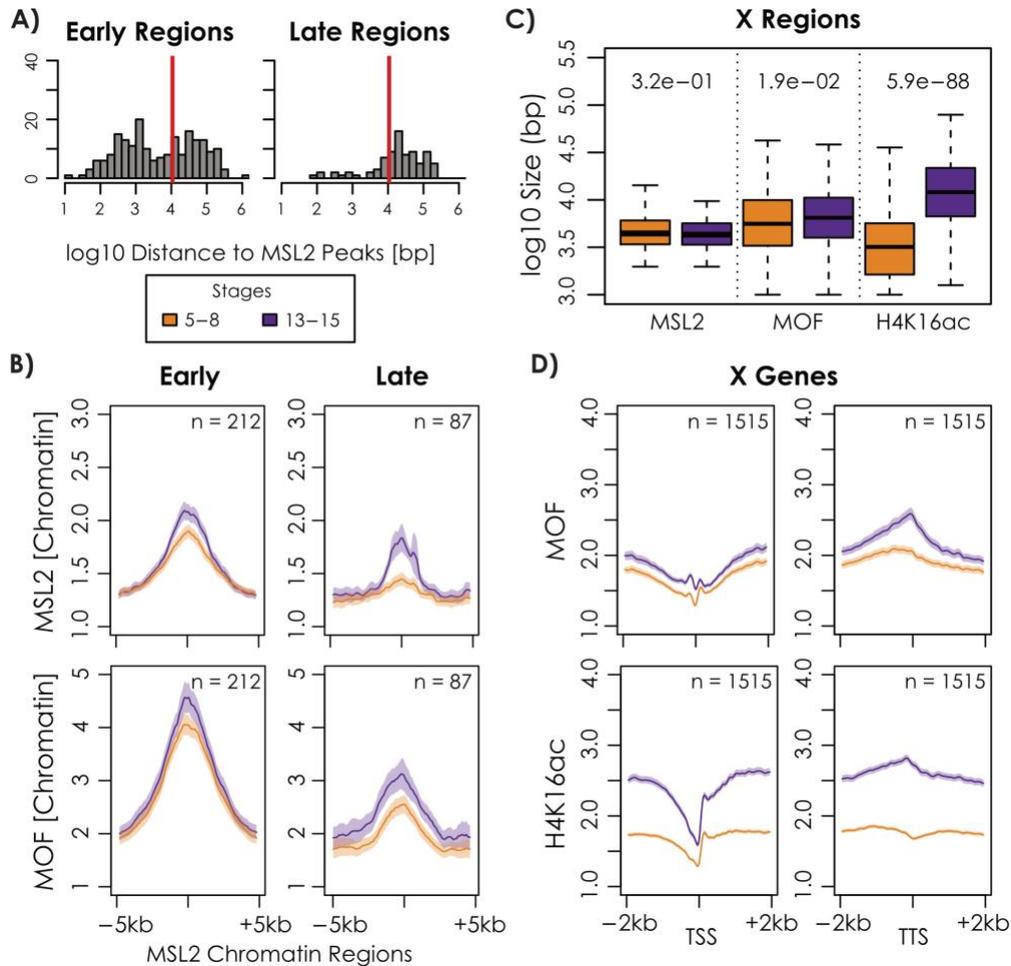
Figure 28. Chromosome-wide H4K16ac requires time. A) Distance from centers of MSL2 chromatin binding regions to nearest MSL2 DNA peak categorized whether they appeared early or late. Red lines indicate average distances. B) Average composite plot of MSL2 or MOF chromatin binding grouped by time of appearance. C) Distribution of MSL2, MOF and H4K16ac region sizes in the time windows. D) Average composite plots of MOF and H4K16ac ChIP-seq centered at Transcription Start Site (TSS) or Transcription Termination Site (TTS) in early or late embryos for X-chromosomal genes (total=1515).

6.3.4 DC progression in time and space

ChIP-seq of MSL2, MOF, and H4K16ac displayed the presence and activity of DCC, while single-embryo RNA-seq revealed the progression of DC. To correlate both events, X-chromosomal genes were plotted according to their acetylation levels and grouped by k-means clustering. Three clusters emerged that provided insights into the workings of dosage compensation during early embryogenesis (Fig. 29A). To note, all genes on the X gained acetylation between the time windows, revealing the progression of dosage compensation.
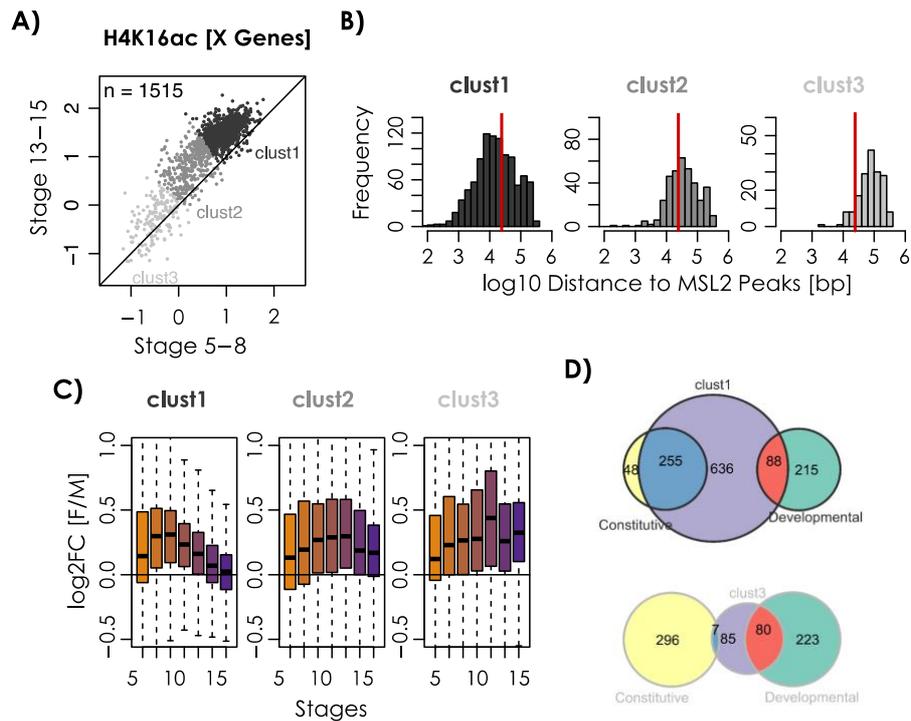
Figure 29. Dosage compensation is progressive. A) Scatterplot comparing X-chromosomal genic H4K16ac ChIP-seq levels in stage 5-8 and 13-15 of mixed-sex embryos. Genes were grouped by k-means clustering (k=3). B) Log10 distance between the middle of a gene and its nearest functional MSL2 DNA peak in cluster 1 to 3. Red line is median distance to the nearest peak. C) log2FC of RNA-seq between female (F) and male (M) embryos at genes as clustered in A). D) Comparison of cluster 1 (top) or cluster 3 (bottom) genes as defined by H4K16ac ChIP-seq levels with constitutive and developmental genes as defined in Fig. 23C.

The first cluster of genes were already highly acetylated at the early time window and were closer to MSL2 peaks compared to the median distance of X-chromosomal genes (Fig. 29A-B). As illustrated in the single-embryo RNA seq dataset, these genes reached full compensation by stage 15 and contained most of previously defined constitutive genes (Fig. 29C-D top). Contrastingly, the third cluster of genes were lowly acetylated and lay far from MSL2 peaks compared to the median distance of X-chromosomal genes (Fig. 29A-B). Unsurprisingly, these genes were nowhere near full compensation by stage 15 and contained only 7 of 333 constitutive genes (Fig. 29C-D bottom). Although only 172 genes were categorized into cluster 3, as compared to 979 in cluster 1, roughly 25% of developmental genes (88-80 out of 333) were found in each cluster 1 and cluster 3, respectively (red overlap in Fig. 29D).

6.3.5 Spreading of DCC

Several conclusions can be made by correlating the spreading of DCC and progression of dosage compensation. First, genes are by and large acetylated in the early time window and reach full compensation 12 hours into development. Second, developmental genes reflect a

lesser dependence on dosage compensation, because they may only be relevant for the stage at which they are needed. Last, evolution selected placement of MSL2 binding sites near genes that require faster compensation, namely the constitutive, house-keeping genes.

Long-range interaction of DCC with remote target genes is thought to be mediated by MSL3 contacting the transcription-associated H3K36me3 mark (70-72). In a final series of ChIP-seq experiments, MSL3 and H3K36me3 were included (Fig. 30). Profile of H4K16ac from the late time window in the establishment study was used for reference.
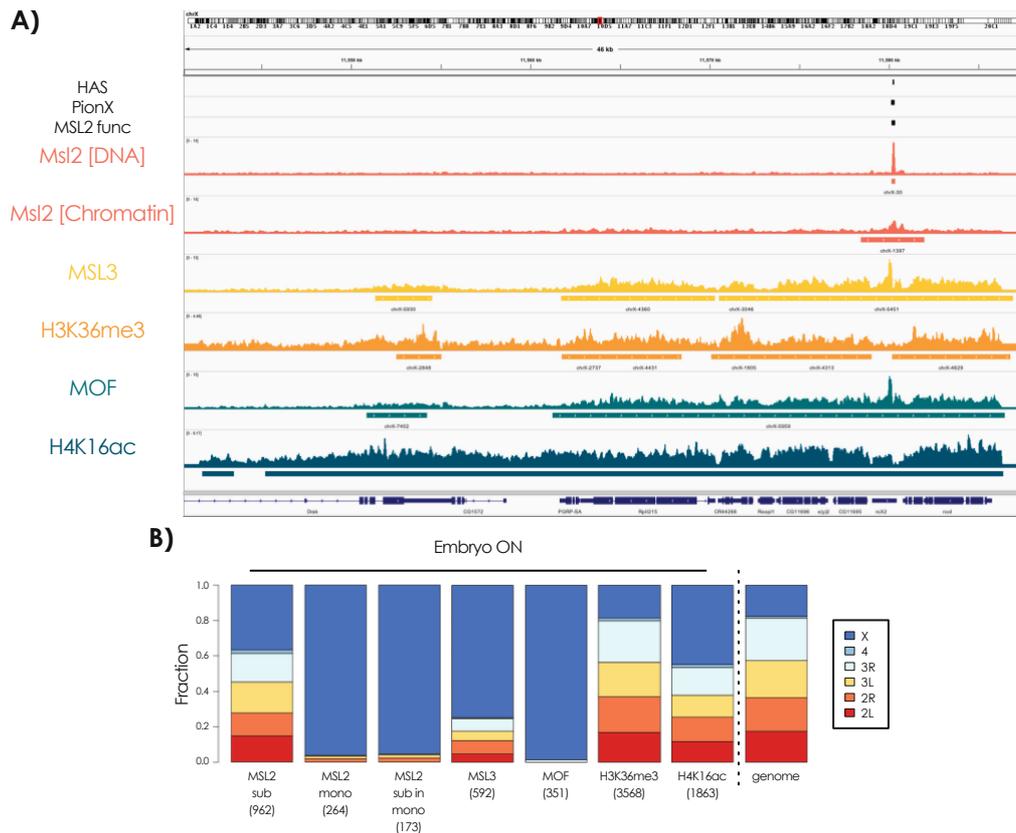

Figure 30. DCC spreading originate from an initial interaction at HAS. A) Genome-wide profiles of MSL2, MSL3, MOF, and H3K36me in embryo chromatin from an overnight collection. H4K16ac profile is the same as from the late time window in Fig. 26. B) Distribution of DNA binding peaks (sub) of MSL2, chromatin binding regions (mono) of MSL2, MSL3, and MOF, and H3K36me3 and H4K16ac marks from ChIP-seq.

Profiles of MSL2 and MOF from overnight embryos largely resembled that of the later time window obtained earlier, indicating that signal from an overnight collection was dominated by older embryos. As observed earlier, many MSL2 peaks were called on autosomes, but narrowing the list using the previous definition of functionality revealed their X-enrichment (Fig. 30B). Not surprisingly, H3K36me3 is found on all chromosomes. Nonetheless, MSL3 and MOF regions were clearly X-enriched, while autosomal distribution

of H4K16ac regions much smaller than those on X chromosome appeared quite significant (Fig. 30B).

The genome-wide profiles of DCC members painted a picture of how a chromosome-wide regulation can be achieved once the X chromosome has been identified. DCC interaction with the X chromosome is initiated at X-enriched HAS by MSL2 through its direct DNA binding capability. From there on, the complex spreads onto gene bodies by means of MSL3 which recognizes H3K36me3 on active genes. MOF travels along and deposits H4K16ac. Although regions of MSL3 and MOF binding closely correlate with gene bodies marked by H3K36me3, H4K16ac regions extend far and beyond those gene bodies, suggesting that the H3K36me3 is not an absolute requirement for H4K16 HAT activity.

6.3.6 *Genome-wide DCC distribution in cell lines*

A fraction of MSL2 binding also occurs on autosomes, many of which are likely non-functional. To narrow down the list of MSL2 peaks, a functional binding site had been defined as an MSL2 peak called within an MSL2 region. This definition differs from previous ones used under different experimental conditions. Straub et al. identify HAS as sites where MSL2 and MLE bind, whereas Villa et al. consider HAS as MSL2 *in vivo* peaks that overlap within two genome-wide profiles both performed in S2 cells (86,91). Evidently, different sites of HAS are obtained, depending on the experimental set-ups. It is also conceivable that the softer mode of solubilization by MNase allowed the detection of more transient binding of MSL2. In addition to the conventional S2 cells, another interesting *D. melanogaster* cell line is the Clone 8 (Cl.8) that originated from cells of the wing imaginal discs of a third instar larvae. It is one of the few male cell lines that expresses both *roX* RNAs (192). To compare the various HAS definitions, MNase-ChIP-seq of various DCC components were done in these cells and compared to the embryo profiles.

HAS that were defined as MSL2 peaks within MSL2 regions in the establishment study were shared among overnight embryo, Cl.8 and S2 cells (Fig. 31A). As it has always been observed, however, ChIP-seq in embryos was more challenging than that in cells because of technicality differences. Embryos consist of mixtures of cell types with different chromatin accessibilities and gene expression. For example, five HAS were defined in the *Tomosyn* gene by Straub et al., one of which was also identified as a PionX site (86,91). In chromatin of cells, three of the five HAS were called with the highest signal belonging to the PionX site (Fig. 31A). In chromatin of the overnight embryo collection, however, only one HAS that is also a PionX site was called (Fig. 31A). It is unlikely that the other HAS did not function, however,

crosslinking and subsequent immunoprecipitation might have been more efficient in cells, and therefore, better preserved weaker interactions. Interestingly, a gene that is highly relevant in neurons was marked by MSL2 and MSL3 in the Cl.8 cells but not S2, revealing that there may be some tissue-specific function of the DCC. The *dunce (dnc)* gene encodes a phosphodiesterase required for cAMP degradation. Its product is involved in neurological and behavioural plasticity including synaptic development and function. Conceivably, this gene is expressed in Cl.8 cells, but to a lesser extent in S2 cells (Fig. 31B).
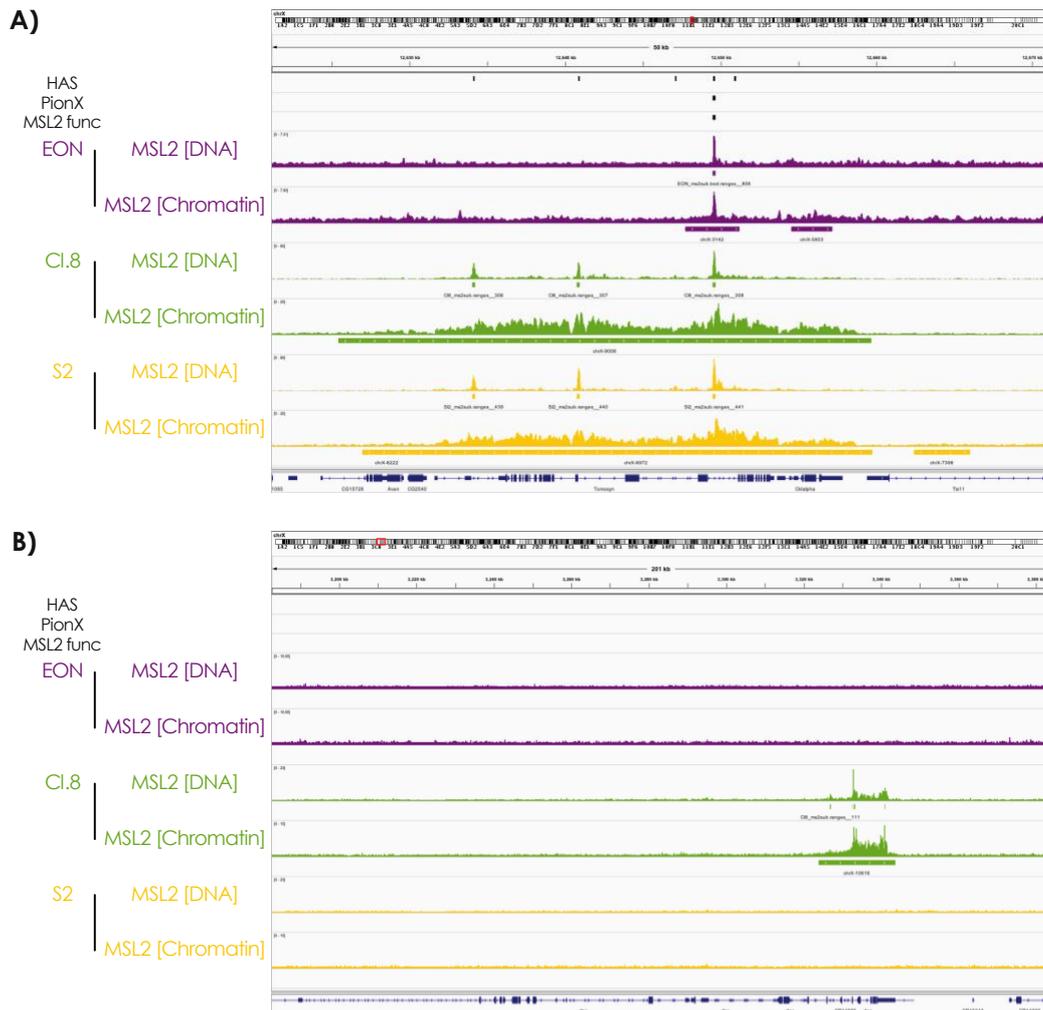

Figure 31. HAS in various tissues. MSL2 DNA and chromatin binding profile at the A) *Tomosyn* and B) *dnc* genes in Cl.8 and S2 cells as well as in chromatin of embryo from an overnight collection. HAS are as defined in Straub et al., 2013, PionX sites are as defined in Villa et al., 2016, and MSL2 functional sites are MSL2 peaks within MSL2 regions as defined in the establishment study.

As MLE was successfully immunoprecipitated in cell lines upon solubilization by MNase, definition of HAS by Straub et al. as colocalization sites of MSL2 and MLE could also be

tested. MSL2, MSL3, and MLE distribution were highly similar in Cl.8 and S2 cells, also at the *roX1* locus, which is only active in Cl.8 cells (Fig. 32A) (193,194). Genome-wide peak and region calling proved high overlap of MLE and MSL2 peaks that lay within regions of MSL2 and MSL3 in Cl.8 cells (Fig. 32B), indicating that these were robust and DC-relevant events. Although more MSL2 peaks were called in S2 cells, only a minority overlapped with MLE peaks or were within a larger MSL2 domain (Fig. 32C).
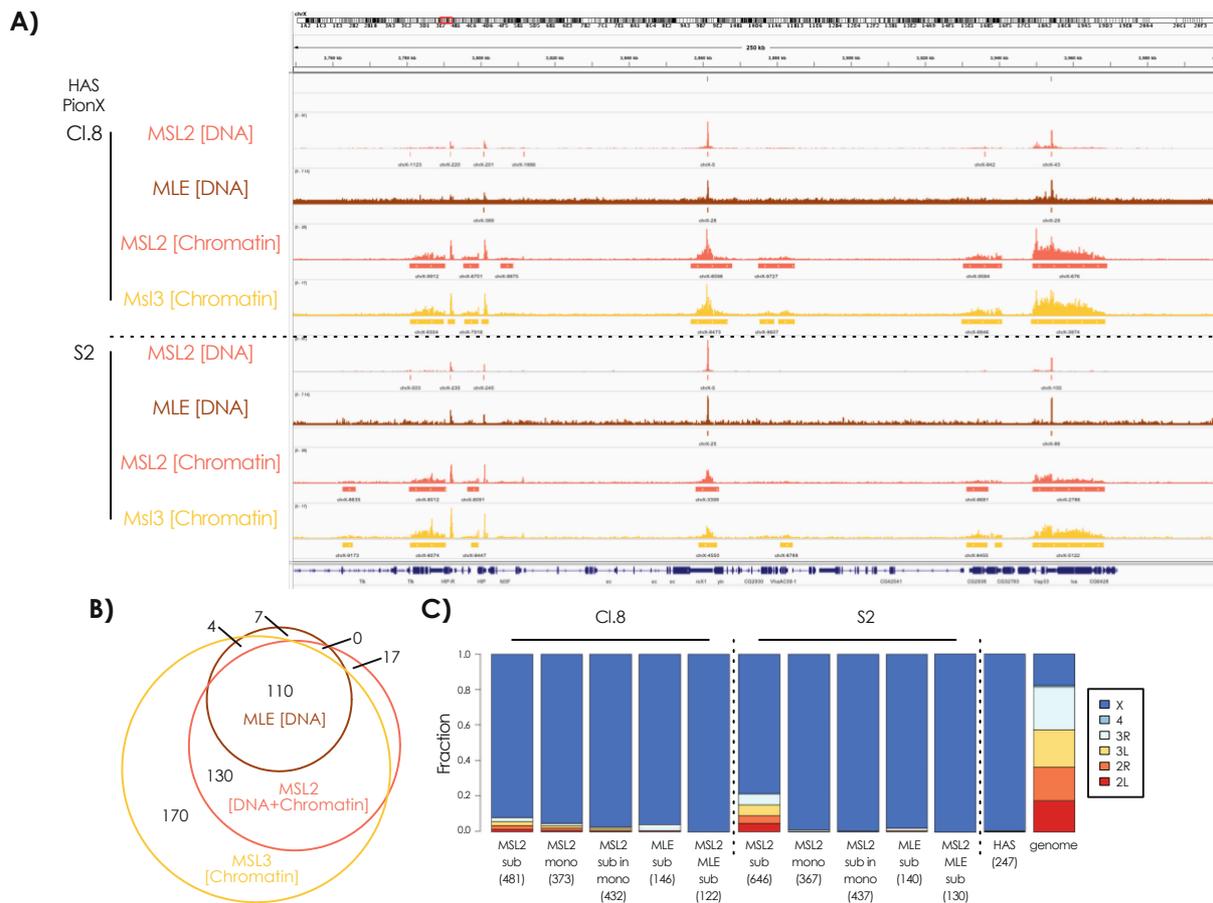

Figure 32. Cl.8 and S2 cells show similar DCC profiles. A) Genome-wide profiles of MSL2, MLE, and MSL3 from Cl.8 and S2 cells. The gene in the middle of the window is *roX1*. B) Overlap of MLE peaks, functional MSL2 peaks (MSL2 peak within an MSL2 region), and MSL3 regions called in Cl.8 cells. C) Distribution of peaks/regions called in ChIP-seq.

In summary, two definitions of HAS were compared, i.e. HAS as "MSL2 peak overlapping with MLE peak" or as "MSL2 peak within MSL2 region". Utilizing the Cl.8 or S2 datasets yielded similar results. Whereas "overlap of MSL2 and MLE peaks" yielded roughly 125 sites, 430 MSL2 peaks were identified within MSL2 regions. This suggests that "MSL2 peak within MSL2 region" is a better experimental classification of HAS than MSL2 peak alone or even "MSL2 peak overlapping with MLE peak". First, DNA binding that did not lead to spreading

were canceled out. Second, it did not rely on the mapping of MLE, which has been technically challenging due to its transient and RNA-dependent interaction.

6.3.7 Cell lines as model for establishing dosage compensation

To model dosage compensation establishment, female Kc cells were manipulated to induce SXL knockdown through RNAi. It has previously been shown that MSL2 is expressed upon removal of SXL in female cells (91,195), although this manipulation is not accompanied by massive changes in chromatin conformation (92). To determine if reducing SXL also alters *roX* RNA expression, *roX* RNAs were quantified under those conditions. Indeed, in addition to an increase in MSL2 protein expression (Fig. 33A-B), removing SXL induced both *roX* RNA expression (Fig. 33C-D). However, the extent of *roX* RNA expression in "sex-changed" Kc cells reached nowhere near that of S2 cells. In terms of *roX1*, RNAi-treated Kc attained half the expression, whereas *roX2* was only expressed at roughly an eighth of that in S2 (Fig. 33C-D).

Induced MSL2 level in Kc was comparable to S2 (Fig. 33A). However, it was still insufficient to induce a chromosome-wide targeting of MSL2 to the X, whereby only the PionX sites acquired strong MSL2 binding after 3 days (91). As such miniscule amounts of *roX* transcription were measured in "sex-changed" Kc cells, *roX* RNA may just be the limiting factor. Highlighting the importance of *roX* RNA in the establishment of DC, understanding its biogenesis into a functional DCC RNP complex is of interest.
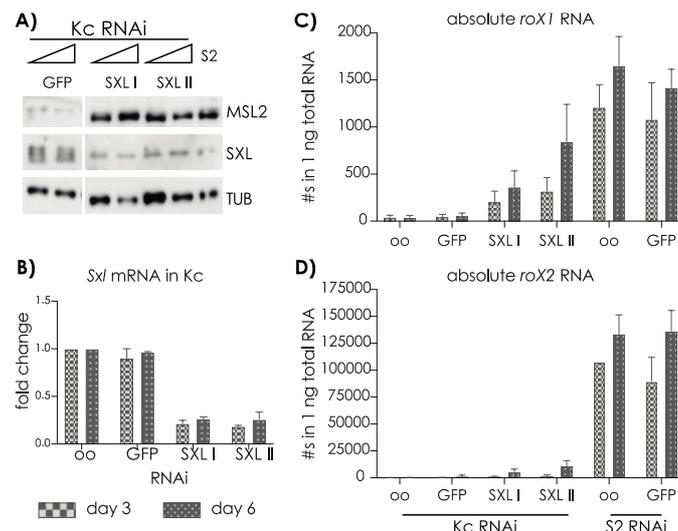

Figure 33. Sxl RNAi of Kc cells. A) MSL2 expression was greatly induced upon SXL repression in Kc cells as analyzed by WB. B-D) qPCR analysis confirmed decrease in *Sxl* mRNA and increase in both *roX* RNAs. *roX2* was more induced than *roX1*; however, the level reached in "sex-changed" Kc cells did not resemble S2 levels.

7 Discussion

7.1 miRoX2

The hypothesis by Sylvain Maenner that suggests *roX1* and *roX2* RNAs hybridize and by means of a non-canonical pathway produce a miRNA to negatively regulate *Sxl* could not be proven as his observations that would suggest so were difficult to reproduce. This raises the question whether such hybridization of two lncRNAs is possible and whether *miRoX2* exist.

7.1.1 lncRNA hybridization

Efforts to hybridize *roX* RNAs have been done in various conditions with RNAs in the absence or presence of extracts. As all experimental attempts have failed, *in silico* predictions of RNA hybrids between *roX1* and *roX2* up to 100 nt upstream and downstream of potential *miRoX2* interaction was done using the Unified Nucleic Acid Folding and hybridization package (UNAFOLD) tool. To note, any modelling of the *roX* RNA alone using similar prediction tools did not fit published experimental data on the secondary structures of *roX* RNAs (78,79), indicating that these predictions are of limited value. Sequences were fed into the algorithm, first starting with the region of interest containing *miRoX2* and the corresponding sequence on *roX1*, followed by incremental addition of 10 nt to each end. Modeling did not reveal any potential hybridization between *roX1* and *roX2*. With every addition, hypothetical secondary structure changed drastically and regions on corresponding *roX* RNAs thought to be the source of *miRoX2* did not necessarily hybridize in every simulation. If hybridization were to occur under physiological conditions, it would very likely involve an RNA helicase that can promote annealing of ssRNAs, such as MLE (76,196).

Interestingly, establishment of the mammalian system of dosage compensation through X inactivation also requires the working of two X-linked lncRNAs, *X Inactive Specific Transcript* (*XIST*) and its antisense *TSIX* (XIST in reverse), which is transcribed in the opposite direction. Contrastingly, while *XIST* silences the X chromosome from which it is transcribed by inducing chromosome-wide histone methylation through recruitment of the Polycomb repression complexes, *TSIX* acts to antagonize it on the active X chromosome by promoting *XIST* promoter methylation, indicative of silencing (197). Additionally, it is thought that *TSIX* transcription across the *XIST* promoter inhibits *XIST* expression, and as it is complementary to the *XIST* RNA, *TSIX* can repress *XIST* through antisense binding as a failsafe mechanism to prevent ectopic silencing (197). Therefore, it is not uncommon to find biological role in the

hybridization of two lncRNAs, and specifically in serving as "sponges" for miRNAs. Nevertheless, *XIST* and *TSIX* may work this way because they are expressed monoallelically on different chromosomes and have opposing effects in the final outcome of silencing. Furthermore, *XIST* and *TSIX* are perfectly complementary to each other, whereas *roX1* and *roX2* show only limited complementarity. What would be the biological context to which *roX1* and *roX2* RNAs, each conducive towards upregulation of the X on their own, require the presence of the other to carry out its role or to antagonize each other? And under which circumstances would such interaction be favourable?

7.1.2 Does miRoX2 exist?

In the original study by Fagegaltier et al., less than 50 short sequences corresponding to *miRoX2* were detected specifically in males upon miRNA sequencing (177). In comparison, a well-known and annotated miRNA, *dme-miR-1*, was found more than 450,000 times, pointing towards the possibility that *miRoX2* sequences may be degradation products. To explore potential targets of *miRoX2*, searches for miRNA family with similarity in seed sequence was done in the miRBase database (release 22.1, October 2018). The first family identified with the least mismatch was *dme-miR-2489* with an E-value of 17 that is equivalent to a p-value of 0.9999999586, indicating that a likeliness is quite improbable. Using the latest release of TargetScanFly (release 7.2, October 2018), *Sxl* was indeed identified as the second best hit of potential *dme-miR-2489* targets. Nonetheless, the low number of sequences found in the original small RNA dataset in flies (177) and the improbable likeness of *miRoX2* to *dme-miR-2489* rather suggest that the probability of such an event is very small.

Target prediction tools infer minimalistic condition, in which the sequence of a candidate miRNA is matched base-per-base with sequence of potential mRNA targets. When calculating minimal free energy values and probability of certain miRNA-mRNA interaction, the algorithms assume a linear unfolded free mRNA that float freely with an unlimited supply of miRNA. In reality, mRNAs are bound by RNA-binding proteins to ensure stability and avoid false interactions (133). Thus, successful interaction of miRNA with mRNA must in all cases win over any other competing transcripts and interaction partners and unless robust detection of a miRNA is seen, its source annotated and its target identified, it is difficult to ascertain the presence of one.

Considering the high variability of results, it is concluded that the system is very complex and may only happen under very specific condition. The hypothesis that *roX RNAs* can together produce a miRNA that leads to *Sxl* mRNA degradation would present a feedback loop that

reinforces the establishment of sex determination and dosage compensation by the boosting of SXL repression and MSL2 expression in males. This may explain why a stably sex-determined system such as a male cell line may not have much need for it and why tissue culture cells may not be a good model to study it. However, as both processes are already well defined without the involvement of *miRoX2*, why should such a system require *miRoX2*?

Transcriptomic study of single embryos during development illustrates the time window in which *roX1* and *roX2* RNAs may potentially meet. *roX2* is expressed later than *roX1*. Based on the compensation measurement and binding event of DCC to the X chromosome, by the time *roX2* is expressed, the male phenotype has already been established and thereby a potential function of *miRoX2* may simply be to strengthen the male line of development. Alternatively, hybridization of *roX1* and *roX2* RNAs could function in females as means of "sponging" out spurious, albeit low level, expression of *roX2*, which may otherwise negatively influence *Sxl* expression. In this case, *roX1* RNA negatively regulates by sequestering *miRoX2*, thus limiting their availability in the cell, and by extension, positively regulates *Sxl*. However, initial RNA concentrations used in the pilot experiments of *miRoX2* were much too high and so the question arises if interaction between *roX1* and *roX2* was observed because the system was pushed to increase the probability of *roX* RNAs finding each other and the original *miRoX2* sequence is simply a degradation product of *roX2*.

7.1.3 *Unique pathway of miRNA production*

As has been introduced, conventional miRNA production involves a pri-miRNA transcript that contains a stem-loop of hybridized sense and anti-sense sequences. If the hypothesis that the 3' end of *roX1* is able to hybridize with *roX2* to form a pri-miRNA construct were true, it would present a unique observation of a non-canonical processing of a miRNA. In this case, the sense and anti-sense sequences originate from two different sources and their processing into a functional miRNA would rely on various parameters including the transcription rate, stability, and secondary structure of each individual RNAs.

The common practice of identifying a miRNA is by a combination of computational, biochemistry, and genetic methods (198). Unfortunately, computational predictions yielded no convincing results, whether in finding a similar family of miRNA or identifying a potential target. All biochemical results that were initially observed were difficult to reproduce, although many types of experiments and methods were exhausted. As a final attempt, a female cell line was generated to express *miRoX2* in the context of a canonical RNAi pathway. To increase the production efficiency and enhance subsequent effect in an *in vivo* setting, an inducible system

to express *primiRoX2* was used, whereby *miRoX2* is packaged into a backbone of a well-known pri-miRNA that undergoes conventional miRNA processing (178). Unfortunately, neither the synthesis of *miRoX2* nor an effect on *Sxl* repression was obvious when the construct was induced for various timepoints. Although stimulation of *primiroX2* transcription could be improved, the question remains if *miRoX2* really exists.

The majority of miRNA arise from miRNA gene families, but it has been widely observed that their origins can be the 3' UTR of their own targets as well as pseudogenes that encode lncRNAs (133). These pre-miRNAs undergo nuclear processing by Drosha, followed by cytoplasmic processing by DCR to produce the final miRNA. Several non-canonical pathways of miRNA production have been reported in different kinds of species that are independent of Drosha (199). Mirtrons, which are pre-miRNA sequences embedded within introns of genes, are exported upon processing by spliceosomes for further cutting by DCR (199). Other sources of miRNA include small-nucleolar RNAs, short hairpin RNAs, as well as tRNAs, and are Drosha-independent but prove to be substrates for DCR cleavage (199). None of the reported examples, however, presented a case in which duplex of two lncRNAs produce a miRNA, except for one study of *XIST* and *TSIX* that has not been reproduced (200,201). Ogawa et al. reported that upon formation of a *XIST*:*TSIX* duplex on the active X chromosome *in vivo*, it is processed into small RNAs they termed *XiRNA* in a DCR-dependent manner and ablation of DCR leads to ineffective X silencing (200). Kenellopoulou et al., however, observed that establishing X inactivation does not require DCR, but it may influence the stability of *XIST* and its ability to coat the X (201). Although the experiments performed were quite different in design and the debate is still open whether *XiRNA* truly exist, and if so, to which extent they are required, i.e. establishment or maintenance of silencing (202), it is interesting to note that RNAi machinery and dosage compensation may intersect in different species and production of such small RNAs as feedback loops has been hypothesized before.

7.2 *roX* RNAs

Expression of *roX* RNA changes during development as has been observed since its first reports. *roX1* is present in male and female since onset of ZGA, whereas *roX2* is male-specific and more abundantly expressed roughly 2 hrs later (116,121). The switch in *roX1* and *roX2* stoichiometry in parallel with the progression of compensation proposes differential roles or efficiencies of their part in DCC.

7.2.1 Differential functionality of roX1 vs. roX2?

Many stable cell lines of the male phenotype from *D. melanogaster* express *roX2* much higher than *roX1*, which may be an adaptation to simplify DCC in a more homogenous system (193,194). For this reason, the Cl.8 cells that persist to express both *roX* RNAs pose an interesting biological setting to compare and contrast *roX1* and *roX2* RNAs.

Nuclear and cytoplasmic analyses of *roX* RNAs point toward contrasting distribution that may indicate distinct pathways of biogenesis. Although contamination may have introduced an artifact during the procedure, the presence of *roX* RNA in the cytoplasm may be biologically relevant. Poly(A)-enriched RNA profiles also suggest that *roX* RNAs undergo post-transcriptional processing that may involve its shuttling in and out of the nucleus. lncRNA are not largely known to shuttle between compartments unlike other non-coding RNAs such as miRNAs and tRNAs; however, a recent study in hepatocellular carcinoma cells proposes the shuttling of a lncRNA, MALAT1, as a crosstalk between the nucleus and mitochondria (203). And so, it is attractive to reason that *roX* RNAs, due to its polyadenylation, moves between compartments. A hypothetical function of such shuttling of *roX1* may be to assist the formation of DCC in the early stage of development, during which *roX1* is more prevalent. Although mRNAs of MSL1, MSL3, MOF, and MLE proteins are deposited into the egg, their translation to reach the homeostatic threshold may still require time. The shuttling of *roX1* RNA to the cytoplasm can therefore act as an assembly platform and facilitate faster formation of DCC, which would require the need to import sub- or full complexes of DC back into the nucleus. Contradicting reports on the associations of nuclear pore complex with MSL3 and MOF have been published (204,205), so it remains to be elucidated whether shuttling of *roX* RNAs is true and has a functional role in DC. Nevertheless, more and more evidence support the idea that *roX1* and *roX2* RNAs have separate roles and would explain their co-evolution through various *Drosophilid* species despite a functional redundancy in DC (114,206).

7.2.2 Quantification methods

Initial experiments to quantify *roX* RNAs and characterize their isoform expression were done through rt-qPCR. Although the method was well enough to indicate relative amounts of transcripts, it had drawbacks that led to absolute quantitative inaccuracy. First and foremost, a crucial part of the method is reverse transcription, which on its own introduces bias towards the 5' end of targets. Second, quantitation against a standard requires prior knowledge of the underlying sequence as the method is sensitive to single base mismatches between primer and target sequence due to SNPs present in both *roX* genes. Third, quantification of various regions

of an RNA species relies on the efficiency of various primer pairs and may not be analogous across the panel. The comparison of amplicons 3 and 4 of roX2, both of which should have detected all isoforms, showed variation by as much a 2-fold factor in the analysis of TRAX RNAs (Fig. 17). Although many improvements have been introduced to qPCR to allow for more accurate absolute quantification (207), more recent genome-wide methods of RNA sequencing provide an enticing platform to do so (208,209).

One such is direct-RNA sequencing using the Nanopore technologies. A pilot experiment has shown potential in characterizing *roX* isoform profiles. Interestingly, regions of both *roX* RNAs that are of great interest and relevant to DC were detected albeit at a very low level. In this experiment, the bait used to attract RNA to the nanopore was the poly(A) tail. In principle, custom adapters can be designed to target specific RNAs for sequencing (Oxford Nanopore Technologies), but such studies have not been published so far. Nevertheless, targeted RNA sequencing can perform simultaneously as an enrichment tool and might help in profiling isoforms of *roX* RNAs. A foreseeable caveat may be that 3' ends of *roX* isoforms differ, but comparison of expected major isoforms of *roX1* and *roX2* could already shed more light into the biogenesis of *roX* RNAs.

7.3 Dosage compensation

Dosage compensation presents an interesting biological phenomenon in which the assembly and targeting of a complex lead to a chromosome-wide regulation. Additionally, this complex requires the interplay of not just various proteins with enzymatic capabilities but also lncRNA to refine its placement and function within the nucleus. Although their importance is repeatedly reiterated, how they influence the workings of gene regulation mechanistically remains to be resolved.

7.3.1 Gradual acquirement of dosage compensation

Considering all the enormous changes happening during the early hours of embryogenesis, it is interesting to see that the establishment of dosage compensation occurs in a very progressive manner over a long period of time, up until at least 12 hrs ael (210). As the need to compensate arises, evolution must have retained an order of events that is most beneficial for the organism and as the mechanism to balance dose advances gradually, traces of evidence at how the solution came about are left behind.

The genome-wide analysis of MSL proteins reinforced the scheme that is used by DCC to perform its function in regulating X-chromosomal transcription. MSL2, as the pioneering

factor, identifies sites on the X chromosome that have accumulated in such a way that genes relying on DCC are closer to them than those that are rather less dependent. Upon recruitment of MSL2 to these high affinity sites, the rest of the complex is brought along via MSL1, the scaffold protein. MSL3 then identifies active gene bodies around these sites to allow specific H4K16 acetylation by MOF. *roX* RNA and its remodeller, MLE, refine the targeting of the entire complex and facilitate its spreading to blanket the entire X chromosome.

MSL2 binding as well as MOF recruitment to HAS are clearly evident 3 to 4 hours ael (210). These profiles do not widely change by 11 to 12 hours ael, indicating that most of the initial events in identifying binding sites on the X chromosome and active gene bodies have ensued. Nevertheless, gain in the acetylation level of H4K16 of X active genes indisputably occurs (210). The increase in H4K16ac level as well as its spatial distribution are very likely facilitated by the refinement of the chromosomal landscape during early embryogenesis (28,29). However, this does not rule out the possibility that a mechanistic switch occurs in the progression of time that increases the activity and/or efficiency of the complex. Such a switch could be the presence of *roX2*, as its expression occurs coincidentally with the gradual reversal of the female-biased X expression.

7.3.2 Spreading mechanism of DCC

HAS serve as nucleation sites, which allow DCC to concentrate on the X chromosome and spread either through an active mechanism or by diffusion (82,84,85,87,90,97,195,211,212). *roX* RNAs have been shown to promote efficient spreading of the DCC along the X chromosome (93,94,191). It may partly be due to the X-linkage of the *roX* genes, whereby autosomal integration of either *roX* genes has been shown to cause *in cis* spreading (82,116,121,122). Alternatively, active transcription through CES found in the main exon of *roX1* and at the 3' end of *roX2* may be relevant to the establishment of DC through some unknown mechanism.

Changes in *roX1* and *roX2* concentrations in parallel with the progression of compensation also propose differential roles or efficiencies. Given that MSL1-mediated dimerization has been implicated in targeting and spreading of DCC (64), a hetero-tetramer core complex of MSL1-MSL2 subunits would recruit two units of other DCC members, including *roX1* and/or *roX2* RNAs. Interestingly, Ilik et al. was unable to detect *roX1* and *roX2* RNAs within the same complex in a ChIRP study (78). Therefore, it is tempting to characterize distinctions of a DCC containing *roX1* RNA and a DCC containing *roX2* RNA. With the obvious difference in expression timing, it is appealing to posit that *roX1*-DCC is the major form in early

development and is sufficient for the recruitment and binding of DCC to the X chromosome. The later-appearing *roX2*-DCC, however, brings about a more efficient process of long-range DCC spreading onto the entire chromosome and/or more competent DCC in terms of hyperacetylation activity. Efficient spreading of DCC may also just be a product of *roX* abundance, whereby transcription of *roX2* in addition to *roX1* RNAs leads to greater number of associations between RNA and protein components of the complex, and therefore, more DCC are available to coat the chromosome.

The structure of MOF has been well characterized and, in addition to its HAT domain responsible for the deposition of H4K16ac, it carries a chromo barrel domain (CBD) (213). Unlike other CDs, such as that of MSL3 which binds H3K36me3, MOF's CBD lacks critical aromatic cage residues and thus is unable to bind methylated histones (213). Instead, it has been shown that MOF interacts with nucleic acids through the CBD with high preference of RNA over DNA *in vitro* and *in vivo* (214). Mutation or deletion of CBD leads to loss of *roX* RNA binding (214), which explains the loss of MOF recruitment in the absence of MLE (215) and loss of MOF association with the X upon RNAse treatment (214). It is appealing to think that maybe *roX* RNAs take part in the recruitment and stable association of MOF to the DCC and also induces MOF activity and in turn that of DCC.

Positive effects of an interaction between a reader domain and nucleic acids on the protein or an associated complex have been reported for other CD-containing proteins such as Mi-2, an ATP-dependent chromatin remodeler, and Chromoboxes (CBXs), members of the Polycomb repressive complex 1 (PRC1). Loss of Mi-2's CD leads to deficit DNA binding, decreased ATPase activity and impaired nucleosome sliding, while repressive activity of the PRC1 complex relies on RNA binding by hCBX7's CD (216,217). It has also been shown that enhancer RNAs stimulate the activity of histone acetylation by CREB-binding protein (CBP) at enhancer sites in the mammalian system (218). Considering not much is described regarding functionality and structure differences between *roX1* and *roX2*, it may just be the case that specific interaction of MOF and *roX2* RNA in the context of DCC leads to functional consequences such as increased chromatin affinity, thus DCC binding stability, facilitated diffusion across the spatial landscape, thus spreading, or a change in conformation that affects the efficiency of the HAT, thus activity.

Profiling of various MSL protein distribution of the same chromatin preparation of overnight embryos also exposed an interesting observation worth noting. Although X-enriched regions of MSL3, H3K36me3 and MOF closely resembled gene bodies, H4K16ac regions

extended far and beyond along the X chromosome, often times connecting multiple smaller regions into large ones encompassing many neighbouring gene bodies. H3K36me3 is thought to be the determining mark that recruits DCC to active genes along the X, however, discrepancy between H3K36me3 and H4K16ac distributions suggests that there may be other factors that contribute towards the spreading of acetylation (47,82,211,212,219,220). Recent experiments by Catherine Regnard saw a decrease in H3K36me3 upon knock-down of Set2, the HMT responsible for trimethylation of H3K36 from a dimethylated state. Nonetheless, distribution of MSL3 was largely unaffected. To investigate if there are other histone modification marks that also contribute towards DCC spreading, experiments using combinatorial modified histones to probe the recruitment of MOF that leads to H4K16ac spreading need to be performed.

One such candidate that may be worth further investigation is the monomethylated state of H4K20. Conflicting reports have shown the correlation of H4K20me1 with positive as well as negative gene regulation (221). Although this rather suggests a context-dependent signalling, two studies have observed DCC protein interaction with H4K20me. Moore et al. reported that *in vitro* binding studies show preferential binding of MSL3's CD to H4K20me1/2 and mutation to the binding cage of methylated lysine results in compromised survival of males (222). Furthermore, nucleic acid binding of the human MSL3 protein through its CD enhances H4K20me1 association (223). Several genome-wide distribution studies of H4K20me1 have been done in the context of female cells or mixed collection of embryos or larvae (224,225). It would be interesting to see how H4K20me1 is distributed in male population of cells.

7.3.3 *Dose imbalance of developmental genes*

Our study revealed that genes with the most variability in expression through the first 12 hours of development are rather farther away from sites upon which DCC interacts favourably (210). Therefore, it is attractive to posit that developmental genes are less dependent on DC mechanism. Several reasons could be attributed to this; either their expression is well buffered in that minimal amount of transcription is sufficient, the product of these genes function like a switch in a dose-independent manner, or dosage-compensated developmental genes are required later in embryogenesis that is outside of our study window. Additionally, cell-type specific expression can be dominated by more abundant, ubiquitous expression of genes, and is technically challenging to determine in such heterogenous setting.

A single embryo transcriptomic study by Lott et al. reported that many X-linked genes are equally expressed in male and female embryos at the onset of ZGA and indeed this list of genes

includes several key developmental regulators, such as *giant* (*gt*), *brinker* (*brk*), *buttonhead* (*btd*), and *short gastrulation* (*sog*) (22). Although the timepoints chosen by Lott et al. do not overlap with ours (210), the two transcriptomic studies complement one another in that both observe an X bias in maternally deposited genes. This may as well be a manifestation of a safety measure set in place to anticipate a possible deficit of X-linked genes, which are especially important for development at this early stage (22,226). As a matter of fact, the four genes mentioned above are expressed highest during the first 4 hrs ael as shown in *in situ* RNA studies of the Berkeley *Drosophila* Genome Project (227-230). Lott et al. also observed that the single X chromosome in the male organism expresses more transcripts than either one of the two X chromosomes in the female (22). Nevertheless, the total male transcript is still less than the female and requires compensation (22). They argue that a non-canonical pathway must have taken place prior to the canonical MSL-mediated dosage compensation that is very much in line with our observation, a 2-fold difference is never seen between male and female transcripts upon activation of the zygotic genome and the most bias observed is an average of 1.2-fold expression in the female (210). It is then crucial to acknowledge that reaching a balanced state of expression is not achieved through a single mechanism, but relies on multiple factors that in parallel activate and also fine-tune the extent of hypertranscription not only on a general and nonspecific level but also on a gene-by-gene specific scenario (231).

7.3.4 Are DCC, or any of its members, active on autosomes?

MOF is the main HAT of H4K16ac (59,73-75). It also participates in another complex, the Non-Specific-Lethal (NSL). Previous reports have shown that in the context of NSL, MOF interacts with MBD-R2, a Zinc-finger transcription factor, and localizes to 5' ends of gene bodies, specifically around promoters to mediate nucleosomal organization around TSSs (232,233). As the name suggests, the NSL complex is not specific towards any one sex as its disruption is lethal in both male and female (204). Furthermore, the complex functions in transcriptional regulation of housekeeping genes and is evolutionary conserved (234,235). Interestingly, MOF is highly X-enriched in our study and autosomal promoter peaks are rather negligible, in contrast to previous characterization of the protein (85,86) (210). Contrastingly, fairly even genome-wide distribution of H4K16ac regions emerge in the embryo dataset, although they are much larger on the X chromosome. It remains to be determined whether discrepancy of distribution is due to technical biases.

Recent published chromatin immunoprecipitation data of MSL2 appeared to suggest interaction of MSL2 with autosomes that have not been observed previously (236). In the study, flag-tagged transgene of *msl-2*, whose expression is driven by tub-Gal4:UAS system was generated, and MNase ChIP-seq was performed on L3 larvae. MSL2 peaks were found both in male and female at autosomal promoters of developmental genes, which they claimed have never been observed before due to the use of embryonic cell lines that do not express these developmental regulatory genes (236). Binding of MSL2 occurred specifically close to autosomal dosage-sensitive genes to evoke a gene-by-gene dosage compensation mechanism (236). Although they claimed that their *msl-2* transgene was expressed to the same level as the endogenous *msl-2*, such observation could arise due to the differential experimental design and subsequent analysis. Roughly 150 autosomal peaks of MSL2 were also observed in developing embryos, although many of them were disregarded as functional binding due to lack of chromatin interaction (Fig. 27B-C) (210). Nevertheless, this number is already much less than the 1684 peaks reported by Valsecchi et al. (236). The *msl-2* transgene lacked UTR and intron sequences, which are targets of *Sxl* in its negative regulation of the MSL2 (236). Thus, the female system was pushed to express MSL2 even in the presence of *Sxl*. The need for MSL2 to regulate developmental genes at such a late stage is inconsistent with the lack of MSL2 expression in female and their lethality when MSL2 is ectopically expressed (54-56).

7.3.5 Technical differences: ChIP-seq ≠ ChIP-seq?

Common to many ChIP protocols is the inclusion of formaldehyde crosslinking prior to chromatin solubilization. The small size of formaldehyde creates covalent linkages between amine groups that are ~2 Å apart (237). As these reactive groups can be part of an amino acid or a nucleobase, formaldehyde can capture nucleic acid-protein as well as protein-protein interactions that are in close proximity, but will not distinguish them from two molecules that are near each other but do not interact or any transient non-specific binding for that matter. At the same time, true interaction may be missed in over-crosslinking condition as it is deleterious to the solubilization of chromatin. This is a bias of crosslinking that is unavoidable; alternatively, ChIP protocols can be done under native conditions without crosslinking and gives much greater signal to noise ratio in cases where chromatin association is maintained. Whereas stable interactions and chromatin complexes are well preserved under these conditions, weaker and low abundant interactions, may not withstand subsequent steps of isolation. As MSL2 is rather difficult to work with, all genome-wide studies of DCC distribution were done under crosslinked conditions.

High shearing is commonly done to solubilize material for further assays. Past experiments to portray DCC generally rely on this to solubilize chromatin-bound proteins for ChIP (85,86,91). Recent results, however, have rather pointed towards MNase digestion as the better method to preserve interactions, especially for sensitive proteins such as MSL2 (210). Differences in observation clearly suggest that the various methods to probe chromatin context depict distinct facets of dynamic biological processes. Sonication increases the efficiency of immunoprecipitation by ridding of non-specific and retaining the strongest interaction in the soluble chromatin. Unfortunately, this method of solubilization fails proteins that are easily destroyed and/or take part in transient bindings. Increased shearing has the potential to also expose hidden epitopes and may present, with the increased sensitivity, binding patterns that are discrepant with MNase digestion (97). MNase digestion enables the portrayal of weaker interactions that may be lost through harsher, mechanical solubilization methods, but it usually results in worse signal to noise ratio (189). Another concern of the MNase digest is the possibility of bias introduction as the MNase enzyme has a sequence preference and relies on accessibility of DNA, thus chromatin structure, in its function to cut (188). Nonetheless, a recent study by Baldi et al., makes the argument that the latter is likely to be insignificant, whereby profiles from MNase-digested chromatin resemble that of MNase-digested gDNA and genome coverage is rather independent of digestion degree (238). Additionally, accessibility of chromatin probed via MNase-seq in human cells showed no differences in euchromatin and heterochromatin (239). Therefore, with the possibility to perform paired-end sequencing, MNase digestion provides an additional advantage in painting a genome-wide picture of DNA binders, as the various types of interaction, i.e. DNA-protein and protein-protein, can be deduced and allow for the study of larger assemblies indirectly bound to chromatin.

As genome-wide studies are becoming more and more prevalent in the attempt to better understand chromatin dynamics and higher order structures, it becomes imperative that appropriate methods and conditions are chosen. A recent development of the Cleavage Under Targets and Release Using Nuclease (CUT&RUN) method developed by the lab of Steven Henikoff presents another tool that can better probe the native state of chromatin (240). CUT&RUN profiles chromatin landscape by using the specificity of an antibody to target MNase cleavage to release protein-DNA complexes from intact nuclei. Combined with paired-end DNA sequencing, CUT&RUN has been used to successfully create profiles of transcription factors and histone proteins without crosslinking, although if needed, it can be included

(240,241). It may be worthwhile to look into DCC distribution through this method, especially in the case of MSL2, which has been one of the most fastidious proteins to work with.

8 References

1.   Bridges CB. Non-Disjunction as Proof of the Chromosome Theory of Heredity. Genetics. 1916 Jan;1(1):1–52.

2.   Morgan TH. SEX LIMITED INHERITANCE IN DROSOPHILA. Science. 1910 Jul 22;32(812):120–2.

3.   Hales KG, Korey CA, Larracuente AM, Roberts DM. Genetics on the Fly: A Primer on the DrosophilaModel System. Genetics. 2015 Nov 12;201(3):815–42.

4.   Bownes M. A photographic study of development in the living embryo of Drosophila melanogaster. J Embryol Exp Morphol. 1975 Jun;33(3):789–801.

5.   Foe VE, Alberts BM. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. J Cell Sci. 1983 May;61:31–70.

6.   Avilés-Pagán EE, Orr-Weaver TL. Activating embryonic development in Drosophila. Seminars in Cell and Developmental Biology. 2018 Dec;84:100–10.

7.   Hamm DC, Harrison MM. Regulatory principles governing the maternal-to-zygotic transition: insights from Drosophila melanogaster. Open Biol. 2018 Dec 12;8(12):180183.

8.   Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. Nat Rev Genet. 2018 Dec 20;20(4):221–34.

9.   Vastenhouw NL, Cao WX, Lipshitz HD. The maternal-to-zygotic transition revisited. Development. 2019 Jun 12;146(11):dev161471.

10.   Hartenstein V. Atlas of Drosophila Development. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1993.

11.   St Johnston D, Brown NH, Gall JG, Jantsch M. A conserved double-stranded RNA-binding domain. Proc Natl Acad Sci USA. 1992 Nov 15;89(22):10979–83.

12.   Thomsen S, Anders S, Janga SC, Huber W, Alonso CR. Genome-wide analysis of mRNA decay patterns during early Drosophila development. Genome Biol. 2010;11(9):R93.

13.   Barckmann B, Simonelig M. Control of maternal mRNA stability in germ cells and early embryos. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2013 Jun;1829(6-7):714–24.

14.   Bushati N, Stark A, Brennecke J, Cohen SM. Temporal Reciprocity of miRNAs and Their Targets during the Maternal-to-Zygotic Transition in Drosophila. Current Biology. 2008 Apr;18(7):501–6.

15.   Lee MT, Bonneau AR, Giraldez AJ. Zygotic Genome Activation During the Maternal-to-Zygotic Transition. Annu Rev Cell Dev Biol. 2014 Oct 11;30(1):581–613.

16.   Kwasnieski JC, Orr-Weaver TL, Bartel DP. Early genome activation in Drosophilais extensive with an initial tendency for aborted transcripts and retained introns. Genome Research. 2019 Jul 3;29(7):1188–97.

17.   Edgar BA, Kiehle CP, Schubiger G. Cell cycle control by the nucleo-cytoplasmic ratio in early Drosophila development. Cell. 1986 Jan;44(2):365–72.

18.   Lu X, Li JM, Elemento O, Tavazoie S, Wieschaus EF. Coupling of zygotic transcription to mitotic control at the Drosophila mid-blastula transition. Development. 2009 May 22;136(12):2101–10.

References

19. Blythe SA, Wieschaus EF. Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. Elife. 2016 Nov 23;5:e1003428.

20. Blythe SA, Wieschaus EF. Zygotic Genome Activation Triggers the DNA Replication Checkpoint at the Midblastula Transition. Cell. 2015 Mar;160(6):1169–81.

21. Harrison MM, Li X-Y, Kaplan T, Botchan MR, Eisen MB. Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. Copenhaver GP, editor. PLoS Genet. 2011 Oct 20;7(10):e1002266.

22. Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-seq. PLoS Biol. 2011 Feb 8;9(2):e1000590.

23. Liang H-L, Nien C-Y, Liu H-Y, Metzstein MM, Kirov N, Rushlow C. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature. 2008 Oct 19;456(7220):400–3.

24. Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, et al. Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophilaembryo. Genome Research. 2015 Nov 2;25(11):1715–26.

25. Fuda NJ, Guertin MJ, Sharma S, Danko CG, Martins AL, Siepel A, et al. GAGA Factor Maintains Nucleosome-Free Regions and Has a Role in RNA Polymerase II Recruitment to Promoters. Lieb JD, editor. PLoS Genet. 2015 Mar 27;11(3):e1005108.

26. Moshe A, Kaplan T. Genome-wide search for Zelda-like chromatin signatures identifies GAF as a pioneer factor in early fly development. Epigenetics & Chromatin. 2017 Jul 4;10(1):693.

27. Zeitlinger J, Stark A, Kellis M, Hong J-W, Nechaev S, Adelman K, et al. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet. 2007 Nov 11;39(12):1512–6.

28. Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell. 2017 Apr;169(2):216–9.

29. Ogiyama Y, Schuettengruber B, Papadopoulos GL, Chang J-M, Cavalli G. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. Molecular Cell. 2018 Jul;71(1):73–5.

30. van der Weide RH, de Wit E. Developing landscapes: genome architecture during early embryogenesis. Current Opinion in Genetics & Development. 2019 Apr;55:39–45.

31. Chen L, Dumelie JG, Li X, Cheng MH, Yang Z, Laver JD, et al. Global regulation of mRNA translation and stability in the early Drosophila embryo by the Smaug RNA-binding protein. Genome Biol. 2014;15(1):R4.

32. Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien C-Y, et al. Zelda Potentiates Morphogen Activity by Increasing Chromatin Accessibility. Current Biology. 2014 Jun;24(12):1341–6.

33. Sandler JE, Stathopoulos A. Stepwise Progression of Embryonic Patterning. Trends in Genetics. 2016 Jul;32(7):432–43.

34. Briscoe J, Small S. Morphogen rules: design principles of gradient-mediated embryo patterning. Development. 2015 Dec 1;142(23):3996–4009.

35. Götze M, Wahle E. Smaug destroys a huge treasure. Genome Biol. 2014;15(1):101.

36. Salz H, Erickson JW. Sex determination in Drosophila: The view from the top. Fly. 2014 Oct 27;4(1):60–70.

37. Estes PA, Keyes LN, Schedl P. Multiple response elements in the Sex-lethal early promoter ensure its female-specific expression pattern. Molecular and Cellular Biology. 1995 Feb;15(2):904–17.

38. Lu H, Kozhina E, Mahadevaraju S, Yang D, Avila FW, Erickson JW. Maternal Groucho and bHLH repressors amplify the dose-sensitive X chromosome signal in Drosophila sex determination. Developmental Biology. 2008 Nov;323(2):248–60.

39. Mahadeveraju S, Erickson JW. Evidence that Runt Acts as a Counter-Repressor of Groucho during Drosophila melanogasterPrimary Sex Determination. Vol. 17, bioRxiv. p. 643.

40. Salz HK. Sex determination in insects: a binary decision based on alternative splicing. Current Opinion in Genetics & Development. 2011 Aug;21(4):395–400.

41. Sosnowski BA, Davis DD, Boggs RT, Madigan SJ, McKeown M. Multiple Portions of a Small Region of the Drosophila transformer Gene Are Required for Efficient in Vivo Sex-Specific Regulated RNA Splicing and in Vitro Sex-Lethal Binding. Developmental Biology. 1994 Jan;161(1):302–12.

42. Yamamoto D, Koganezawa M. Genes and circuits of courtship behaviour in Drosophila males. Nat Rev Neurosci. 2013 Sep 20;14(10):681–92.

43. Arbeitman MN. A genomic analysis of Drosophila somatic sexual differentiation and its regulation. Development. 2004 Mar 31;131(9):2007–21.

44. Luo SD, Shi GW, Baker BS. Direct targets of the D. melanogaster DSXF protein and the evolution of sexual development. Development. 2011 Jun 7;138(13):2761–71.

45. Rideout EJ, Dornan AJ, Neville MC, Eadie S, Goodwin SF. Control of sexual differentiation and behavior by the doublesex gene in Drosophila melanogaster. Nat Neurosci. 2010 Mar 21;13(4):458–66.

46. Demir E, Dickson BJ. fruitless Splicing Specifies Male Courtship Behavior in Drosophila. Cell. 2005 Jun;121(5):785–94.

47. Kuroda MI, Hilfiker A, Lucchesi JC. Dosage Compensation in Drosophila-a Model for the Coordinate Regulation of Transcription. Genetics. 2016 Oct;204(2):435–50.

48. Samata M, Akhtar A. Dosage Compensation of the X Chromosome: A Complex Epigenetic Assignment Involving Chromatin Regulators and Long Noncoding RNAs. Annu Rev Biochem. 2018 Jun 20;87(1):323–50.

49. Bashaw GJ, Baker BS. The Regulation of the Drosophila msl-2 Gene Reveals a Function for Sex-lethal in Translational Control. Cell. 1997 May;89(5):789–98.

50. Beckmann K, Grskovic M, Gebauer F, Hentze MW. A Dual Inhibitory Mechanism Restricts msl-2 mRNA Translation for Dosage Compensation in Drosophila. Cell. 2005 Aug;122(4):529–40.

51. Gebauer F, Merendino L, Hentze MW, Valcárcel J. The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. RNA. 1998 Feb;4(2):142–50.

52. Kelley RL, Wang J, Bell L, Kuroda MI. Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism. Nature. 1997 May;387(6629):195–9.

References

53. Penalva LOF, Sanchez L. RNA Binding Protein Sex-Lethal (Sxl) and Control of Drosophila Sex Determination and Dosage Compensation. Microbiology and Molecular Biology Reviews. 2003 Sep 9;67(3):343–59.

54. Belote JM, Lucchesi JC. Male-specific lethal mutations of Drosophila melanogaster. Genetics. 1980 Sep;96(1):165–86.

55. Breen TR, Lucchesi JC. Analysis of the dosage compensation of a specific transcript in Drosophila melanogaster. Genetics. 1986 Mar;112(3):483–91.

56. Kelley RL, Solovyeva I, Lyman LM, Richman R, Solovyev V, Kuroda MI. Expression of Msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila. Cell. 1995 Jun;81(6):867–77.

57. Belote JM, Lucchesi JC. Control of X chromosome transcription by the maleless gene in Drosophila. Nature. 1980 Jun;285(5766):573–5.

58. Uenoyama T, Uchida S, Fukunaga A, Genetics KO. Studies on the sex-specific lethals of Drosophila melanogaster. IV. Gynandromorph analysis of three male-specific lethals, mle, msl-227 and mle(3)132. Genetics Soc America. 1982 Oct.

59. Hilfiker A. mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila. EMBO J. 1997 Apr 15;16(8):2054–60.

60. Richter L, Bone JR, Kuroda MI. RNA-dependent association of the Drosophila maleless protein with the male X chromosome. Genes Cells. 1996 Mar;1(3):325–36.

61. Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, et al. The Drosophila MSL Complex Acetylates Histone H4 at Lysine 16, a Chromatin Modification Linked to Dosage Compensation. Molecular and Cellular Biology. 2000 Jan 1;20(1):312–8.

62. Hamada FN. Global regulation of X chromosomal genes by the MSL complex in Drosophila melanogaster. Genes Dev. 2005 Oct 1;19(19):2289–94.

63. Chen G, Nguyen PH, Courey AJ. A Role for Groucho Tetramerization in Transcriptional Repression. Molecular and Cellular Biology. 1998 Dec 1;18(12):7259–68.

64. Hallacli E, Lipp M, Georgiev P, Spielman C, Cusack S, Akhtar A, et al. Msl1-Mediated Dimerization of the Dosage Compensation Complex Is Essential for Male X-Chromosome Regulation in Drosophila. Molecular Cell. 2012 Nov;48(4):587–600.

65. Schunter S, Villa R, Flynn V, Heidelberger JB, Classen A-K, Beli P, et al. Ubiquitylation of the acetyltransferase MOF in Drosophila melanogaster. PLoS ONE. 2017;12(5):e0177408.

66. Villa R, Forné I, Müller M, Imhof A, Straub T, Becker PB. MSL2 Combines Sensor and Effector Functions in Homeostatic Control of the Drosophila Dosage Compensation Machinery. Molecular Cell. 2012 Nov;48(4):647–54.

67. Scott MJ. MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in Drosophila. EMBO J. 2000 Jan 4;19(1):144–55.

68. Gu W, Szauter P, Lucchesi JC. Dev Genet. 1998;22(1):56–64.

69. Morales V, Straub T, Neumann MF, Mengus G, Akhtar A, Becker PB. Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. EMBO J. 2004 May 13;23(11):2258–68.

70. Alekseyenko AA. High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. Genes Dev. 2006 Mar 17;20(7):848–57.

71. Bell O, Wirbelauer C, Hild M, Scharf AND, Schwaiger M, MacAlpine DM, et al. Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila. EMBO J. 2007 Nov 15;26(24):4974–84.

72. Li B, Howe L, Anderson S, Yates JR III, Workman JL. The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. J Biol Chem. 2003 Mar 7;278(11):8897–903.

73. Akhtar A, Becker PB. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Molecular Cell. 2000 Feb;5(2):367–75.

74. Bone JR, Lavender J, Richman R, Palmer MJ, Turner BM, Kuroda MI. Acetylated histone H4 on the male X chromosome is associated with dosage compensation in Drosophila. Genes Dev. 1994 Jan;8(1):96–104.

75. Turner BM, Birley AJ, Lavender J. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. Cell. 1992 Apr;69(2):375–84.

76. Izzo A, Regnard C, Morales V, Kremmer E, Becker PB. Structure-function analysis of the RNA helicase maleless. Nucleic Acids Research. 2007 Dec 17;36(3):950–62.

77. Lee C-G, Reichman TW, Baik T, Mathews MB. MLE Functions as a Transcriptional Regulator of the roX2Gene. J Biol Chem. 2004 Nov 5;279(46):47740–5.

78. Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, et al. Tandem Stem-Loops in roX RNAs Act Together to Mediate X Chromosome Dosage Compensation in Drosophila. Molecular Cell. 2013 Jul;51(2):156–73.

79. Maenner S, Müller M, Fröhlich J, Langer D, Becker PB. ATP-Dependent roX RNA Remodeling by the Helicase maleless Enables Specific Association of MSL Proteins. Molecular Cell. 2013 Jul;51(2):174–84.

80. Lucchesi JC, Kuroda MI. Dosage Compensation in Drosophila. Cold Spring Harb Perspect Biol. 2015 May 1;7(5):a019398.

81. Fauth T, Müller-Planitz F, König C, Straub T, Becker PB. The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. Nucleic Acids Research. 2010 Feb 5;38(10):3209–21.

82. Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, Kuroda MI. Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell. 1999 Aug 20;98(4):513–22.

83. Alekseyenko AA, Ellison CE, Gorchakov AA, Zhou Q, Kaiser VB, Toda N, et al. Conservation and de novo acquisition of dosage compensation on newly evolved sex chromosomes in Drosophila. Genes Dev. 2013 Apr 29;27(8):853–8.

84. Lyman LM, Copps K, Rastelli L, Kelley RL, Kuroda MI. Drosophila male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association. Genetics. 1997 Dec;147(4):1743–53.

References

85.   Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB. The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex. PLoS Genet. 2008 Dec;4(12):e1000302.

86.   Straub T, Zabel A, Gilfillan GD, Feller C, Becker PB. Different chromatin interfaces of the Drosophila dosage compensation complex revealed by high-shear ChIP-seq. Genome Research. 2013 Mar 1;23(3):473–85.

87.   Zheng S, Villa R, Wang J, Feng Y, Wang J, Becker PB, et al. Structural basis of X chromosome DNA recognition by the MSL2 CXC domain during Drosophiladosage compensation. Genes Dev. 2014 Dec 1;28(23):2652–62.

88.   Soruco MML, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, et al. The CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage compensation. Genes Dev. 2013 Jul 19;27(14):1551–6.

89.   Albig C, Tikhonova E, Krause S, Maksimenko O, Regnard C, Becker PB. Factor cooperation for chromosome discrimination in Drosophila. Nucleic Acids Research. 2018 Dec 12;47(4):1706–24.

90.   Bai X, Alekseyenko AA, Kuroda MI. Sequence-specific targeting of MSL complex regulates transcription of the roX RNA genes. EMBO J. 2004 Jul 1;23(14):2853–61.

91.   Villa R, Schauer T, Smialowski P, Straub T, Becker PB. PionX sites mark the X chromosome for dosage compensation. Nature. 2016 Aug 31;537(7619):244–8.

92.   Schauer T, Ghavi Helm Y, Sexton T, Albig C, Regnard C, Cavalli G, et al. Chromosome topology guides the DrosophilaDosage Compensation Complex for target gene activation. EMBO Rep. 2017 Aug 14;18(10):1854–68.

93.   Park SW, Kuroda MI, Park Y. Regulation of Histone H4 Lys16 Acetylation by Predicted Alternative Secondary Structures in roX Noncoding RNAs. Molecular and Cellular Biology. 2008 Jul 28;28(16):4952–62.

94.   Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, et al. Direct Competition between hnRNP C and U2AF65 Protects the Transcriptome from the Exonization of Alu Elements. Cell. 2013 Jan;152(3):453–66.

95.   Cheetham SW, Brand AH. RNA-DamID reveals cell-type-specific binding of roX RNAs at chromatin-entry sites. Nat Struct Mol Biol. 2017 Dec 18;25(1):109–14.

96.   Conrad T, Akhtar A. Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nat Rev Genet. 2012 Jan 18;13(2):123–34.

97.   Straub T, Becker PB. Comment on "Drosophila Dosage Compensation Involves Enhanced Pol II Recruitment to Male X-Linked Promoters." Science. 2013 Apr 18;340(6130):273–3.

98.   Ferrari F, Jung YL, Kharchenko PV, Plachetka A, Alekseyenko AA, Kuroda MI, et al. Comment on "Drosophila Dosage Compensation Involves Enhanced Pol II Recruitment to Male X-Linked Promoters." Science. 2013 Apr 18;340(6130):273–3.

99.   Vaquerizas JM, Cavalli FMG, Conrad T, Akhtar A, Luscombe NM. Response to Comments on "Drosophila Dosage Compensation Involves Enhanced Pol II Recruitment to Male X-Linked Promoters." Science. 2013 Apr 18;340(6130):273–3.

100.   Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, et al. X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. Nature. 2011 Mar 2;471(7336):115–8.

101. Prabhakaran M, Kelley RL. Mutations in the Transcription Elongation Factor SPT5 Disrupt a Reporter for Dosage Compensation in Drosophila. Meller V, editor. PLoS Genet. 2012 Nov 29;8(11):e1003073.

102. Regnard C, Straub T, Mitterweger A, Dahlsveen IK, Fabian V, Becker PB. Global analysis of the relationship between JIL-1 kinase and transcription. PLoS Genet. 2011 Mar;7(3):e1001327.

103. Ferrari F, Plachetka A, Alekseyenko AA, Jung YL, Ozsolak F, Kharchenko PV, et al. "Jump Start and Gain" Model for Dosage Compensation in Drosophila Based on Direct Sequencing of Nascent Transcripts. Cell Reports. 2013 Nov;5(4):1157.

104. Smith ER, Allis CD, Lucchesi JC. Linking Global Histone Acetylation to the Transcription Enhancement of X-chromosomal Genes in DrosophilaMales. J Biol Chem. 2001 Aug 17;276(34):31483–6.

105. Shogren-Knaak M. Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. Science. 2006 Feb 10;311(5762):844–7.

106. Robinson PJJ, An W, Routh A, Martino F, Chapman L, Roeder RG, et al. 30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction. Journal of Molecular Biology. 2008 Sep;381(4):816–25.

107. Rastelli L, Richman R, Kuroda MI. The dosage compensation regulators MLE, MSL-1 and MSL-2 are interdependent since early embryogenesis in Drosophila. Mechanisms of Development. 1995 Oct;53(2):223–33.

108. Franke A, Dernburg A, Bashaw GJ, Baker BS. Evidence that MSL-mediated dosage compensation in Drosophila begins at blastoderm. Development. 1996 Sep;122(9):2751–60.

109. Gergen JP. Dosage Compensation in Drosophila: Evidence That daughterless and Sex-lethal Control X Chromosome Activity at the Blastoderm Stage of Embryogenesis. Genetics. 1987 Nov;117(3):477–85.

110. Bernstein M, Cline TW. Differential effects of Sex-lethal mutations on dosage compensation early in Drosophila development. Genetics. 1994 Mar;136(3):1051–61.

111. Ohno S. So much "junk" DNA in our genome. Brookhaven Symp Biol. 1972;23:366–70.

112. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet. 2015 Dec 15;17(1):47–62.

113. Jandura A, Krause HM. The New RNA World: Growing Evidence for Long Noncoding RNA Functionality. Trends in Genetics. 2017 Oct;33(10):665–76.

114. Quinn JJ, Zhang QC, Georgiev P, Ilik IA, Akhtar A, Chang HY. Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. Genes Dev. 2016 Jan 15;30(2):191–207.

115. Yao R-W, Wang Y, Chen L-L. Cellular functions of long noncoding RNAs. Nat Cell Biol. 2019 May 2;21(5):542–51.

116. Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL. roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. Cell. 1997 Feb 21;88(4):445–57.

117. Amrein H, Axel R. Genes Expressed in Neurons of Adult Male Drosophila. Cell. 1997 Feb;88(4):459–69.

References

118. Franke A, Baker BS. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. Molecular Cell. 1999 Jul;4(1):117–22.

119. Groth AC. Construction of Transgenic Drosophila by Using the Site-Specific Integrase From Phage phiC31. Genetics. 2004 Apr 1;166(4):1775–82.

120. Bell JC, Jukam D, Teran NA, Risca VI, Smith OK, Johnson WL, et al. Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. Elife. 2018 Apr 12;7:429.

121. Meller VH. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO J. 2002 Mar 1;21(5):1084–91.

122. Park Y. Extent of Chromatin Spreading Determined by roX RNA Recruitment of MSL Proteins. Science. 2002 Nov 22;298(5598):1620–3.

123. Deng X, Rattner BP, Souter S, Meller VH. The severity of roX1 mutations is predicted by MSL localization on the X chromosome. Mechanisms of Development. 2005 Oct;122(10):1094–105.

124. Deng X, Meller VH. roXRNAs Are Required for Increased Expression of X-Linked Genes in Drosophila melanogasterMales. Genetics. 2006 Dec 20;174(4):1859–66.

125. Stuckenholz C, Meller VH, Kuroda MI. Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in Drosophila melanogaster. Genetics. 2003 Jul;164(3):1003–14.

126. Park S-W, Kang YI, Sypula JG, Choi J, Oh H, Park Y. An Evolutionarily Conserved Domain of roX2RNA Is Sufficient for Induction of H4-Lys16 Acetylation on the Drosophila X Chromosome. Genetics. 2007 Nov 26;177(3):1429–37.

127. Kelley RL, Lee O-K, Shim Y-K. Transcription rate of noncoding roX1 RNA controls local spreading of the Drosophila MSL chromatin remodeling complex. Mechanisms of Development. 2008 Nov;125(11-12):1009–19.

128. Lv M, Yao Y, Li F, Xu L, Yang L, Gong Q, et al. Structural insights reveal the specific recognition of roX RNA by the dsRNA-binding domains of the RNA helicase MLE and its indispensable role in dosage compensation in Drosophila. Nucleic Acids Research. 2019 Jan 15;47(6):3142–57.

129. Park Y, Oh H, Meller VH, Kuroda MI. Variable splicing of non-coding roX2 RNAs influences targeting of MSL dosage compensation complexes in Drosophila. RNA Biology. 2005 Oct;2(4):157–64.

130. Meller VH. Initiation of dosage compensation in Drosophila embryos depends on expression of the roX RNAs. Mechanisms of Development. 2003 Jul;120(7):759–67.

131. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, et al. The developmental transcriptome of Drosophila melanogaster. Nature. 2010 Dec 22;471(7339):473–9.

132. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014 Jul 16;15(8):509–24.

133. Treiber T, Treiber N, Meister G. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. Nat Rev Mol Cell Biol. 2018 Sep 18;20(1):5–20.

134. Carthew RW, Agbu P, Giri R. MicroRNA function in Drosophila melanogaster. Seminars in Cell and Developmental Biology. 2017 May;65:29–37.

135. Chapat C, Jafarnejad SM, Matta-Camacho E, Hesketh GG, Gelbart IA, Attig J, et al. Cap-binding protein 4EHP effects translation silencing by microRNAs. Proc Natl Acad Sci USA. 2017 May 23;114(21):5425–30.

136. Fukaya T, Iwakawa H-O, Tomari Y. MicroRNAs Block Assembly of eIF4F Translation Initiation Complex in Drosophila. Molecular Cell. 2014 Oct;56(1):67–78.

137. Flemming W. Zellsubstanz, Kern und Zelltheilung. F.C.W. Vogel; 1882.

138. Olins DE, Olins AL. Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol. 2003 Oct;4(10):809–14.

139. Heitz E. Die somatische Heteropyknose bei Drosophila melanogaster und ihre genetische Bedeutung. ZZellforsch. 1933;20(1-2):237–87.

140. Painter TS. A NEW METHOD FOR THE STUDY OF CHROMOSOME REARRANGEMENTS AND THE PLOTTING OF CHROMOSOME MAPS. Science. 1933 Dec 22;78(2034):585–6.

141. Kornberg RD. Chromatin Structure: A Repeating Unit of Histones and DNA. Science. 1974 May 24;184(4139):868–71.

142. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature. 1997 Sep;389(6648):251–60.

143. Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. Nature. 2003 May;423(6936):145–50.

144. Jenuwein T. Translating the Histone Code. Science. 2001 Aug 10;293(5532):1074–80.

145. Huang H, Sabari BR, Garcia BA, Allis CD, Zhao Y. SnapShot: Histone Modifications. Cell. 2014 Oct;159(2):458–458.e1.

146. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. Journal of Molecular Biology. 2002 Jun;319(5):1097–113.

147. Noll M, Kornberg RD. Action of micrococcal nuclease on chromatin and the location of histone H1. Journal of Molecular Biology. 1977 Jan;109(3):393–404.

148. Fyodorov DV, Zhou B-R, Skoultchi AI, Bai Y. Emerging roles of linker histones in regulating chromatin structure and function. Nat Rev Mol Cell Biol. 2017 Oct 11;19(3):192–206.

149. Benbow RM. Chromosome structures. Sci Prog. 1992;76(301-302 Pt 3-4):425–50.

150. Garcia-Saez I, Menoni H, Boopathi R, Shukla MS, Soueidan L, Noirclerc-Savoye M, et al. Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation. Molecular Cell. 2018 Dec;72(5):902–7.

151. Hergeth SP, Schneider R. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. EMBO Rep. 2015 Sep 15;16(11):1439–53.

152. Maeshima K, Ide S, Babokhov M. Dynamic chromatin organization without the 30-nm fiber. Current Opinion in Cell Biology. 2019 Jun;58:95–104.

153. Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH, O'Shea CC. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science. 2017 Jul 27;357(6349):eaag0025.

References

154. Ricci MA, Manzo C, García-Parajo MF, Lakadamyali M, Cosma MP. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. Cell. 2015 Mar;160(6):1145–58.

155. Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. 2016 Oct 14;17(11):661–78.

156. Bowman GD, Poirier MG. Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. Chem Rev. 2014 Oct 20;115(6):2274–95.

157. Li B, Carey M, Workman JL. The Role of Chromatin during Transcription. Cell. 2007 Feb;128(4):707–19.

158. Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, et al. Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. Cell. 2011 Sep;146(6):1016–28.

159. Tessarz P, Kouzarides T. Histone core modifications regulating nucleosome structure and dynamics. Nat Rev Mol Cell Biol. 2014 Oct 15;15(11):703–8.

160. Eberharter A. Histone acetylation: a switch between repressive and permissive chromatin: Second in review series on chromatin dynamics. EMBO Rep. 2002 Mar 1;3(3):224–9.

161. Rando OJ. Global patterns of histone modifications. Current Opinion in Genetics & Development. 2007 Apr;17(2):94–9.

162. Boros IM. Histone modification in Drosophila. Briefings in Functional Genomics. 2012 Jul 16;11(4):319–31.

163. Bai X, Bi W, Dong H, Chen P, Tian S, Zhai G, et al. An Integrated Approach Based on a DNA Self-Assembly Technique for Characterization of Crosstalk among Combinatorial Histone Modifications. Anal Chem. 2018 Feb 21;90(6):3692–6.

164. Musselman CA, Lalonde M-E, Côté J, Kutateladze TG. Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol. 2012 Dec 5;19(12):1218–27.

165. Chen Q, Yang R, Korolev N, Liu CF, Nordenskiöld L. Regulation of Nucleosome Stacking and Chromatin Compaction by the Histone H4 N-Terminal Tail–H2A Acidic Patch Interaction. Journal of Molecular Biology. 2017 Jun;429(13):2075–92.

166. Reinke H, Hörz W. Histones Are First Hyperacetylated and Then Lose Contact with the Activated PHO5 Promoter. Molecular Cell. 2003 Jun;11(6):1599–607.

167. Zhao J, Herrera-Diaz J, Gross DS. Domain-Wide Displacement of Histones by Activated Heat Shock Factor Occurs Independently of Swi/Snf and Is Not Correlated with RNA Polymerase II Density. Molecular and Cellular Biology. 2005 Sep 30;25(20):8985–99.

168. Zhang R, Erler J, Langowski J. Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction. Biophysical Journal. 2017 Feb;112(3):450–9.

169. Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, et al. Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Molecular and Cellular Biology. 2003 Jun 15;23(12):4207–18.

170. Li B, Howe L, Anderson S, Yates JR III, Workman JL. The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. J Biol Chem. 2003 Mar 7;278(11):8897–903.

171. Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD. A Novel Domain in Set2 Mediates RNA Polymerase II Interaction and Couples Histone H3 K36 Methylation with Transcript Elongation. Molecular and Cellular Biology. 2005 Mar 29;25(8):3305–16.

172. Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, et al. Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. Cell. 2005 Nov;123(4):581–92.

173. Keogh M-C, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, et al. Cotranscriptional Set2 Methylation of Histone H3 Lysine 36 Recruits a Repressive Rpd3 Complex. Cell. 2005 Nov;123(4):593–605.

174. Thompson J, Schedl P, Drosophila RP4A. Sex-specific GFP-expression in Drosophila embryos and sorting by Copas flow cytometry technique. unionbiocom. 2014.

175. Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, Kastner B, et al. Protein Composition and Electron Microscopy Structure of Affinity-Purified Human Spliceosomal B Complexes Isolated under Physiological Conditions. Molecular and Cellular Biology. 2006 Jun 29;26(14):5528–43.

176. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell. 2010 May;38(4):576–89.

177. Fagegaltier D, König A, Gordon A, Lai EC, Gingeras TR, Hannon GJ, et al. A Genome-Wide Survey of Sexually Dimorphic Expression of DrosophilamiRNAs Identifies the Steroid Hormone-Induced miRNA let-7 as a Regulator of Sexual Identity. Genetics. 2014 Oct 14;198(2):647–68.

178. Haley B, Hendrix D, Trang V, Levine M. A simplified miRNA-based gene silencing method for Drosophila melanogaster. Developmental Biology. 2008 Sep;321(2):482–90.

179. Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM. bantam Encodes a Developmentally Regulated microRNA that Controls Cell Proliferation and Regulates the Proapoptotic Gene hid in Drosophila. Cell. 2003 Apr;113(1):25–36.

180. Becker PB, Rabindran SK, Wu C. Heat shock-regulated transcription in vitro from a reconstituted chromatin template. Proc Natl Acad Sci USA. 1991 May 15;88(10):4109–13.

181. Becker PB, Wu C. Cell-free system for assembly of transcriptionally repressed chromatin from Drosophila embryos. Molecular and Cellular Biology. 1992 May 1;12(5):2241–9.

182. Völker-Albert MC, Pusch MC, Fedisch A, Schilcher P, Schmidt A, Imhof A. A Quantitative Proteomic Analysis of In VitroAssembled Chromatin. Mol Cell Proteomics. 2016 Mar 1;15(3):945–59.

183. Baldi S, Jain DS, Harpprecht L, Zabel A, Scheibe M, Butter F, et al. Genome-wide Rules of Nucleosome Phasing in Drosophila. Molecular Cell. 2018 Nov;72(4):661–4.

184. Ayub M, Bayley H. Individual RNA Base Recognition in Immobilized Oligonucleotides Using a Protein Nanopore. Nano Lett. 2012 Oct 15;12(11):5637–43.

185. Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, et al. Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods. 2018 Jan 15;15(3):201–6.

186. Liu H, Begik O, Lucas MC, Ramirez JM, Mason CE, Wiener D, et al. Accurate detection of m6A RNA modifications in native RNA sequences. Nature Communications. 2019 Sep 9;10(1):1403.

References

187. Ner SS, Travers AA. HMG-D, the Drosophila melanogaster homologue of HMG 1 protein, is associated with early embryonic chromatin in the absence of histone H1. EMBO J. 1994 Apr 15;13(8):1817–22.

188. Dingwall C, Lomonossoff GP, Laskey RA. High sequence specificity of micrococcal nuclease. Nucleic Acids Research. 1981;9(12):2659–74.

189. Chereji RV, Kan T-W, Grudniewska MK, Romashchenko AV, Berezikov E, Zhimulev IF, et al. Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in Drosophila melanogaster. Nucleic Acids Research. 2016 Feb 17;44(3):1036–51.

190. Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. Epigenome characterization at single base-pair resolution. Proc Natl Acad Sci USA. 2011 Nov 8;108(45):18318–23.

191. Ramírez F, Lingg T, Toscano S, Lam KC, Georgiev P, Chung H-R, et al. High-Affinity Sites Form an Interaction Network to Facilitate Spreading of the MSL Complex across the X Chromosome in Drosophila. Molecular Cell. 2015 Oct;60(1):146–62.

192. Lee H, McManus C, Cho D-Y, Eaton M, Renda F, Somma M, et al. DNA copy number evolution in Drosophila cell lines. Genome Biol. 2014;15(8):R70.

193. Cherbas L, Willingham A, Zhang D, Yang L, Zou Y, Eads BD, et al. The transcriptional diversity of 25 Drosophila cell lines. Genome Research. 2011 Feb 1;21(2):301–14.

194. Johansson A-M, Allgardsson A, Stenberg P, Larsson J. msl2 mRNA is bound by free nuclear MSL complex in Drosophila melanogaster. Nucleic Acids Research. 2011 May 6;39(15):6428–39.

195. Alekseyenko AA, Ho JWK, Peng S, Gelbart M, Tolstorukov MY, Plachetka A, et al. Sequence-specific targeting of dosage compensation in Drosophila favors an active chromatin context. PLoS Genet. 2012;8(4):e1002646.

196. Lee CG. The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation. EMBO J. 1997 May 15;16(10):2671–81.

197. Dykes IM, Emanueli C. Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA. Genomics, Proteomics & Bioinformatics. 2017 Jun;15(3):177–86.

198. Liu B, Li J, Cairns MJ. Identifying miRNAs, targets and functions. Briefings in Bioinformatics. 2014 Jan 21;15(1):1–19.

199. Abdelfattah AM, Park C, Choi MY. Update on non-canonical microRNAs. Biomol Concepts. 2014 Aug;5(4):275–87.

200. Ogawa Y, Sun BK, Lee JT. Intersection of the RNA Interference and X-Inactivation Pathways. Science. 2008 Jun 6;320(5881):1336–41.

201. Kanellopoulou C, Muljo SA, Dimitrov SD, Chen X, Colin C, Plath K, et al. X chromosome inactivation in the absence of Dicer. Proc Natl Acad Sci USA. 2009 Jan 27;106(4):1122–7.

202. Collins LJ, Schönfeld B, Chen XS. The Epigenetics of Non-coding RNA. In: Handbook of Epigenetics. Elsevier; 2011. pp. 49–61.

203. Zhao Y, Liu S, Zhou L, Li X, Meng Y, Li Y, et al. Aberrant shuttling of long noncoding RNAs during the mitochondria-nuclear crosstalk in hepatocellular carcinoma cells. Am J Cancer Res. 2019;9(5):999–1008.

204. Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, et al. Nuclear Pore Components Are Involved in the Transcriptional Regulation of Dosage Compensation in Drosophila. Molecular Cell. 2006 Mar;21(6):811–23.

205. Grimaud C, Becker PB. The dosage compensation complex shapes the conformation of the X chromosome in Drosophila. Genes Dev. 2009 Nov 2;23(21):2490–5.

206. Kim M, Faucillion M-L, Larsson J. RNA-on-X 1 and 2 in Drosophila melanogaster fulfill separate functions in dosage compensation. Larschan E, editor. PLoS Genet. 2018 Dec 10;14(12):e1007842.

207. Taylor SC, Nadeau K, Abbasi M, Lachance C, Nguyen M, Fenrich J. The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. Trends Biotechnol. 2019 Jul;37(7):761–74.

208. Marinov GK. On the design and prospects of direct RNA sequencing. Briefings in Functional Genomics. 2017 Feb 17;16(6):326–35.

209. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. Trends in Genetics. 2018 Sep;34(9):666–81.

210. Prayitno K, Schauer T, Regnard C, Becker PB. Progressive dosage compensation during Drosophilaembryogenesis is reflected by gene arrangement. EMBO Rep. 2019 Jul 4;20(8):2751.

211. Oh H, Bone JR, Kuroda MI. Multiple Classes of MSL Binding Sites Target Dosage Compensation to the X Chromosome of Drosophila. Current Biology. 2004 Mar;14(6):481–7.

212. Dahlsveen IK, Gilfillan GD, Shelest VI, Lamm R, Becker PB. Targeting Determinants of Dosage Compensation in Drosophila. PLoS Genet. 2006;2(2):e5.

213. Nielsen PR, Nietlispach D, Buscaino A, Warner RJ, Akhtar A, Murzin AG, et al. Structure of the Chromo Barrel Domain from the MOF Acetyltransferase. J Biol Chem. 2005 Sep 9;280(37):32326–31.

214. Akhtar A, Zink D, Becker PB. Chromodomains are protein–RNA interaction modules. Nature. 2000 Sep;407(6802):405–9.

215. Gu W, Szauter P, Lucchesi JC. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of Drosophila melanogaster. Dev Genet. 1998;22(1):56–64.

216. Bouazoune K. The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. EMBO J. 2002 May 15;21(10):2430–40.

217. Weaver T, Morrison E, Musselman C. Reading More than Histones: The Prevalence of Nucleic Acid Binding among Reader Domains. Molecules. 2018 Oct;23(10):2614.

218. Bose DA, Donahue G, Reinberg D, Shiekhattar R, Bonasio R, Berger SL. RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell. 2017 Jan;168(1-2):135–149.e22.

219. Fagegaltier D, Baker BS. X Chromosome Sites Autonomously Recruit the Dosage Compensation Complex in Drosophila Males. R Scott Hawley, editor. PLoS Biol. 2004 Oct 5;2(11):e341.

220. Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, Yang P, et al. MSL Complex Is Attracted to Genes Marked by H3K36 Trimethylation Using a Sequence-Independent Mechanism. Molecular Cell. 2007 Oct;28(1):121–33.

References

221. van Nuland R, Gozani O. Histone H4 Lysine 20 (H4K20) Methylation, Expanding the Signaling Potential of the Proteome One Methyl Moiety at a Time. Mol Cell Proteomics. 2016 Mar 1;15(3):755–64.

222. Moore SA, Ferhatoglu Y, Jia Y, Al-Jiab RA, Scott MJ. Structural and Biochemical Studies on the Chromo-barrel Domain of Male Specific Lethal 3 (MSL3) Reveal a Binding Preference for Mono- or Dimethyllysine 20 on Histone H4. J Biol Chem. 2010 Dec 17;285(52):40879–90.

223. Kim D, Blus BJ, Chandra V, Huang P, Rastinejad F, Khorasanizadeh S. Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. Nat Struct Mol Biol. 2010 Jul 25;17(8):1027–9.

224. Li Y, Armstrong RL, Duronio RJ, MacAlpine DM. Methylation of histone H4 lysine 20 by PR-Set7 ensures the integrity of late replicating sequence domains in Drosophila. Nucleic Acids Research. 2016 Apr 29;:gkw333.

225. Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, et al. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Molecular Cell. 2017 Sep;67(5):837–7.

226. Wieschaus E, Nüsslein-Volhard C. The Heidelberg Screen for Pattern Mutants of Drosophila: A Personal Account. Annu Rev Cell Dev Biol. 2016 Oct 6;32(1):1–46.

227. Hammonds AS, Bristow CA, Fisher WW, Weiszmann R, Wu S, Hartenstein V, et al. Spatial expression of transcription factors in Drosophila embryonic organ development. Genome Biol. 2013;14(12):R140.

228. Tomancak P, Beaton A, Weiszmann R, Kwan E, Shu S, Lewis SE, et al. Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol. 2002;3(12):RESEARCH0088.

229. Tomancak P, Berman BP, Beaton A, Weiszmann R, Kwan E, Hartenstein V, et al. Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol. 2007;8(7):R145.

230. Weiszmann R, Hammonds AS, Celniker SE. Determination of gene expression patterns using high-throughput RNA in situ hybridization to whole-mount Drosophila embryos. Nat Protoc. 2009 Apr 9;4(5):605–18.

231. Lee H, Cho D-Y, Wojtowicz D, Harbison ST, Russell S, Oliver B, et al. Dosage-Dependent Expression Variation Suppressed on the DrosophilaMale XChromosome. G3. 2018 Feb 2;8(2):587–98.

232. Prestel M, Feller C, Straub T, Mitlöhner H, Becker PB. The Activation Potential of MOF Is Constrained for Dosage Compensation. Molecular Cell. 2010 Jun;38(6):815–26.

233. Lam KC, Chung H-R, Semplicio G, Iyer SS, Gaub A, Bhardwaj V, et al. The NSL complex-mediated nucleosome landscape is required to maintain transcription fidelity and suppression of transcription noise. Genes Dev. 2019 Apr 1;33(7-8):452–65.

234. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2008 Dec 18;4(1):44–57.

235. Lam KC, Mühlpfordt F, Vaquerizas JM, Raja SJ, Holz H, Luscombe NM, et al. The NSL complex regulates housekeeping genes in Drosophila. PLoS Genet. 2012;8(6):e1002736.

236. Valsecchi CIK, Basilicata MF, Semplicio G, Georgiev P, Gutierrez NM, Akhtar A. Facultative dosage compensation of developmental genes on autosomes in Drosophila and mouse embryonic stem cells. Nature Communications. 2018 Sep 7;9(1):189.

237. Hoffman EA, Frey BL, Smith LM, Auble DT. Formaldehyde Crosslinking: A Tool for the Study of Chromatin Complexes. J Biol Chem. 2015 Oct 30;290(44):26404–11.

238. Baldi S, Krebs S, Blum H, Becker PB. Genome-wide measurement of local nucleosome array regularity and spacing by nanopore sequencing. Nat Struct Mol Biol. 2018 Aug 20;25(9):894–901.

239. Schwartz U, Németh A, Diermeier S, Exler JH, Hansch S, Maldonado R, et al. Characterizing the nuclease accessibility of DNA in human cells to map higher order structures of chromatin. Nucleic Acids Research. 2019 Feb 20;47(3):1239–54.

240. Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife. 2017 Jan 16;6:576.

241. Skene PJ, Henikoff JG, Henikoff S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat Protoc. 2018 Apr 12;13(5):1006–19.

VII.   Acknowledgements

VIII. Curriculum Vitae