Sex-specific transcript diversity is regulated by a maternal pioneer factor in early *Drosophila* embryos

Mukulika Ray[*#], Ashley Mae Conard[*ψ], Jennifer Urban[ξ], Erica Larschan[#Ω]

[*]contributed equally
[Ω]Corresponding Author
[#] MCB department, Brown University, Providence, RI, USA
[ψ] CCMB department, Brown University, Providence, RI, USA
[ξ] Biology department, Johns Hopkins University, Baltimore, Maryland, USA

Abstract

Maternally deposited RNAs and proteins play a crucial role in initiating zygotic transcription during early embryonic development. However, the mechanisms by which maternal factors regulate zygotic transcript diversity early in development remain poorly understood. Furthermore, how early in development sex-specific transcript diversity occurs is not known genome-wide in any organism. Here, we determine that sex-specific transcript diversity occurs much earlier in development than previously thought in *Drosophila*, concurrent with **Z**ygotic **g**enome **a**ctivation (ZGA). We use genetic, biochemical, and genomic approaches to demonstrate that the essential maternally-deposited pioneer factor CLAMP (**C**hromatin **l**inked **a**dapter for **M**SL **p**roteins) is a key regulator of sex-specific transcript diversity in the early embryo via the following mechanisms: 1) In both sexes, CLAMP directly binds to the gene bodies of female and male sex-specifically spliced genes. 2) In females, CLAMP modulates chromatin accessibility of an alternatively-spliced exon within Sex-lethal, the master regulator of sex determination, to promote protein production. 3) In males, CLAMP regulates **M**ale**less** (MLE) distribution, a spliceosome component to prevent aberrant sex-specific splicing. Thus, we demonstrate for the first time how a maternal factor regulates early zygotic transcriptome diversity sex-specifically. We also developed a new tool to measure how splicing changes over time called time2splice.

Introduction

One of the greatest challenges in modern biology is understanding the mechanism and significance of widespread transcript diversity between sexes and different developmental stages, tissues, and

cell-types. In addition to basic development and physiology, transcriptome diversity is critical for disease biology, especially in neurodegenerative diseases and developmental disorders that often show sex or tissue-specific difference in progression and severity (Faustino and Cooper 2003; Wang and Cooper 2007; Ober et al. 2008; Mayne et al. 2016). Across species, precise regulation of genes to produce specific splice variants is critical for all developmental decisions, including sex determination. A key to understanding how gene regulation drives biological processes during an organism's lifetime lies in the events that shape the initial few hours of its existence.

During early development, protein and RNA from the mother shape early embryonic milestones across metazoans (Schulz et al. 2015; Schulz and Harrison 2019). Initially, all embryos increase their cell number, followed by cellular differentiation into specific cell types. Sexual identity is then established, driving the fundamental physiological differences between sexes. However, how maternally deposited proteins and RNAs regulate this process of sexual differentiation is understood very little. Moreover, maternal factors are often vital epigenetic regulators in the initial stages of development and can have a lasting impact on gene regulation later in the life of an organism. Thus, it is essential to define the influence of maternal factors on transcriptome diversity during the early stages of embryonic development. Therefore, the key question is: How do maternally deposited products regulate transcript diversity, including sex-specific splice variants?

The *Drosophila* embryo is an excellent tool to study the role of maternally deposited proteins and RNA in early development as it is easy to perform genetic manipulation to remove maternal factors and define the effect on splicing and transcription. Also, embryos can be sexed before zygotic genome activation due to the recent application of a meiotic drive system (Rieder et al. 2017).

During *Drosophila* embryogenesis, **Z**ygotic **G**enome **A**ctivation (**ZGA**) occurs shortly after the first two hours of development. Concurrently, maternal transcripts gradually decrease, and zygotic transcripts increase, a process called **M**aternal to **Z**ygotic **T**ransition (**MZT**). Multiple variants of the same gene transcripts are produced, resulting in transcript diversity. **A**lternative **S**plicing (**AS**), a mechanism of selective inclusion or exclusion of introns and exons, is one factor which drives extensive transcript diversity (Revil et al. 2010; Aanes et al. 2013). Although the earliest genes transcribed from the zygotic genome are mainly intron-less, almost 30% of early zygotic

transcripts do have introns (De Renzis et al. 2007; Guilgur et al. 2014). Furthermore, genes involved in sex determination use AS to drive male versus female-specific development (Förch and Valcárcel 2003). Hence, during early embryonic development, AS is important for shaping cell and tissue-specific transcriptomes and sexual differentiation.

The master regulator of *Drosophila* sex determination is the *sxl* gene, which undergoes alternative splicing such that exon three is retained in males but not in females (Salz and Erickson 2010; Haussmann et al. 2016). Therefore, functional Sxl protein is made only in the female embryo. Early expression of the Sxl protein made from a transcript driven by Sxl_{Pe}, an early *sxl* promoter (Estes et al. 1995), autoregulates splicing of *sxl*. Next, the Sxl protein causes splicing of exon three from pre-mRNA for *sxl* transcripts initiated from the late maintenance promoter (Sxl_{Pm}) only in females (Keyes et al. 1992; González et al. 2008). The early activation of the Sxl_{Pe} promoter is regulated by a group of X-chromosome genes called X signal elements (XSE). A higher dosage of XSE gene products drives Sxl_{Pe} expression in XX females and not XY males (Salz 2007).

The Sxl protein is an RNA binding protein that binds to *sxl* pre-mRNA and pre-mRNA of downstream target genes such as *transformer* (*tra*) and *male-specific lethal protein 2* (*msl-2*) to regulate their splicing and stability, giving rise to female-specific splice variants (Penalva and Sánchez 2003). The female-specific Tra isoform then regulates sex-specific splicing of the *doublesex* (*dsx*) gene resulting in different Dsx protein isoforms in females and males which directly regulate sexual dimorphism in both sexes (Salz and Erickson 2010; Haussmann et al. 2016).

Sxl regulation of *msl-2* RNA splicing and stability links the sex determination pathway to another important early embryonic event: dosage compensation, which equalizes gene dosage of X-linked genes between females with two X-chromosome and males with a single X-chromosome. MSL-2 is an essential core component of the MSL complex, present only in the male embryo and responsible for dosage compensation (Prayitno et al. 2019; Rieder et al. 2019). The binding of Sxl to the *msl-2* transcript specifically in females results in skipping of the first exon in females, thereby hindering translation of the protein, which prevents MSL complex formation in females and allows it to be formed only in males (Gebauer et al. 1998).

Once formed in males, the MSL complex targets the male X-chromosome at high-affinity binding sequences called **C**hromatin **E**ntry **S**ites (**CES**) rich in multiple GA-repeats (Alekseyenko et al. 2008; Straub et al. 2013). The maternally deposited Chromatin linked adapter for MSL complex (CLAMP) protein opens chromatin (Urban et al. 2017a; Duan et al. 2020) and directly contacts MSL-2 (Albig et al. 2019) to promote MSL complex recruitment. Subsequently, the MSL complex increases transcript levels on the male X-chromosome by promoting transcription elongation (Larschan et al. 2011). Interestingly, the MSL complex shares a key highly conserved component MLE (RNA helicase A) with the spliceosome complex (Kuroda et al. 1991; Cugusi et al. 2015; Ankush Jagtap et al. 2019), suggesting an interplay between dosage compensation and sex-specific splicing that has not yet been defined.

CLAMP is an essential pioneer **T**ranscription **F**actor (**TF**) that opens chromatin (Urban et al. 2017a; Duan et al. 2020). Loss of CLAMP results in complete loss of MSL complex from the male X-chromosome causing male lethality (Soruco et al. 2013). Unlike MSL complex component mutants, which cause only male-specific lethality, *clamp* mutants show lethality in both males and females. All embryos die early in embryogenesis after losing maternal CLAMP (Duan et al. 2020). Furthermore, CLAMP is bound to both intronic and promoter regions on chromatin (Kaye et al. 2018). These intronic regions are rich in polypyrimidine tracts, which have similar GA-rich sequences to CES and contain CLAMP binding sites (Quinn et al. 2016). Another feature of these polypyrimidine tracts is that they are present near the splice junctions of a large number of transcript variants. Importantly, MALDI-mass spectrometry data identifying putative CLAMP interactors predicted its binding to RNA binding proteins, including several which are involved in splicing (Urban et al. 2017c). Based on these lines of evidence, we hypothesized that CLAMP is a maternally deposited factor that shapes transcriptome diversity in early embryos by regulating RNA transcript splicing.

To date, there are no reports defining transcriptome diversity due to alternative splicing in early *Drosophila melanogaster* embryos. Moreover, how maternal TFs regulate alternative splicing in any organism is not known. Here, we determine that 16-18% of transcripts undergo alternative splicing during the first four hours of development in *Drosophila melanogaster* embryos. Although

sex-specific isoforms are known for several genes in early development (Telonis-Scott et al. 2009; Lott et al. 2014; Paris et al. 2015), a comprehensive analysis of all sex-specific spliced forms has not yet been performed. Towards that end, we used a meiotic drive system to sex embryos before MZT, and computationally defined all sex-specific isoforms and determined the extent of alternative splicing. Next, we identified male-specific and female-specific genes whose splicing required maternal CLAMP during the first few hours of development. We found that maternal CLAMP influences **m**utually e**x**clusive **e**xon (MXE) splicing, mainly in males and mostly during the Pre-MZT stage in both males and females.

Moreover, we defined multiple mechanisms by which an essential maternal factor (CLAMP) drives sex-specific transcript diversity during early embryonic development: 1) In both males and females, CLAMP is specifically enriched at gene bodies of sex-specifically spliced genes supporting a direct role for CLAMP in splicing; 2) In females, CLAMP regulates chromatin accessibility at *sxl* exon three and regulates its splicing; 3) In males, our data are consistent with a model in which CLAMP alters the relative distribution of MLE between MSL complex and the spliceosome complex by influencing its association with chromatin to promote male-specific splicing and inhibit aberrant female-specific isoforms. Overall, we demonstrate that sex-specific transcriptome variation is established very early in development and is regulated by maternally deposited CLAMP using diverse mechanisms.

Results

1. Alternative splicing plays an essential role in shaping the early embryonic transcriptome.

To determine when during early *Drosophila melanogaster* development sex-specific splicing originates, we analyzed RNA-seq data generated from sexed embryos at two-time points: 1) 0-2 hours (pre-MZT); 2) 2-4 hours (post-MZT stages) (Rieder et al. 2017). Embryos were sexed before MZT (#GSE102922) using a meiotic drive system that produces sperm with either only X or only Y chromosomes (Rieder et al. 2017), resulting in progeny of either only female or only male genotypes. Next, we measured the amount of AS in early *Drosophila* development using a tool we developed called, time2Splice (https://github.com/ashleymaeconard/time2splice). AS events can

be classified into seven types: 1) Alternative Last Exon (AL); 2) Skipping Exon (SE); 3) Alternative 5' Splice Site (A5SS); 4) Alternative 3' Splice Site (A3SS); 5) Mutually Exclusive Exon (MXE) 6) Retained Intron (RI) and 7) Alternative First Exon (AF, **Fig1A**). We measured the total number of splicing events that belonged to each of the seven different categories at pre- and post-MZT embryonic stages in both female and male embryos. We found that 16-18% of total transcripts are alternatively spliced in early embryos. **A**lternative **F**irst Exon (AF) is the most common alternative splice type, constituting almost one-fourth of total AS. In contrast, **A**lternative **L**ast Exon (AL) is the least common group. The AS transcript distribution across types was similar between the two time points and sexes (**Fig 1B**).

We hypothesized that CLAMP regulates AS in early embryos because it is a maternally deposited pioneer factor enriched at intronic regions. Its binding sites evolved from polypyrimidine tracts that regulate splicing (Quinn et al. 2016). Furthermore, proteomic data identified a physical association between spliceosome components and CLAMP (Urban et al. 2017c). We tested our hypothesis in early staged and sexed embryos by measuring differences in splicing within RNA-sequencing data generated from male and female pre- and post-MZT embryos with and without maternal CLAMP (Rieder et al. 2017). Maternal triple driver GAL4, *MTD-GAL4* was used to drive *UAS-CLAMPRNAi[val22]* to lower maternal CLAMP levels (Rieder et al. 2017).

First, we asked which type of alternative splicing is most affected by CLAMP and calculated the percentage of these events affected by the absence of maternal CLAMP. The overall distribution of AS transcripts into the seven splicing types remained mostly unaffected in the absence of maternal CLAMP. However, at the pre-MZT stage, loss of maternal CLAMP resulted in a more substantial decrease in **M**utually E**x**clusive **E**xon (**MXE**) splicing in both males and females compared with all of the other types of splicing (males: p-value < 3.21e-21; females: p-value < 6.26e-87 chi-squared test) (**Fig 1B**). At the post-MZT stage, only male embryos had a significant percentage of MXE splicing affected in the absence of maternal CLAMP (p-value < 1.95e-137 chi-squared test) (**Fig 2B**). During MXE splicing one isoform of the transcript retains one of the alternative exons and excludes another exon, which is retained by another isoform (schematic **Fig 1A**). Furthermore, MXE alternative splicing occurs in many transcripts that encode components of the sex determination pathway (Brooks et al. 2015). Sex determination pathway components

undergo alternative splicing and influence splicing of downstream genes to drive transcriptome variation and sexual dimorphism. Therefore, we next asked whether CLAMP regulates sex-specific splicing.

2. Maternal CLAMP regulates sex-specific alternative splicing in early *Drosophila* embryos

To determine whether CLAMP-dependent alternative splicing events are enriched for **s**ex-**s**pecific **s**plicing (SSS) events, we first identified all of the transcripts undergoing female and male sex-specific splicing. Then, we measured the percentage of total alternatively spliced and sex-specifically spliced transcripts that are CLAMP-dependent in males and females at both pre- and post-MZT stages. While only 2-3% of total Alternative splicing is CLAMP-dependent, ~30-60% of sex-specific splicing is CLAMP-dependent (**Fig 2A**). Therefore, CLAMP-dependent alternative splicing events are highly enriched for sex-specific splicing events.

We then divided all CLAMP-dependent alternatively spliced genes into two categories: 1) sex-specifically spliced (SSS) genes; 2) non-sex specifically spliced (non-SSS) genes (Fig 2B). Next, we further classified the sex-specifically spliced genes as sex-specifically-spliced genes previously identified (known SSS) and newly identified sex-specifically spliced genes (new SSS) (**Fig 2C**). The majority of genes show CLAMP dependent sex-specific splicing, especially in female embryos (Fig 2B). Interestingly, the majority of CLAMP-dependent SSS events are new sex-specific splicing events that did not occur in the presence of maternal CLAMP (**Fig 2C, D**). Therefore, we hypothesized that maternal CLAMP prevents the formation of aberrant sex-specific splicing events.

Next, we quantified the amount of alternative splicing using an exon-centric approach to quantify individual splice junctions by measuring **p**ercent **s**pliced **i**n (**PSI**) for a particular exon using SUPPA within time2splice (Trincado et al. 2018). PSI measures the proportion of a specific exon's inclusion in a particular transcript. We quantified CLAMP-regulated splicing of sex-specifically spliced genes by comparing the change in PSI (ΔPSI) for both known and novel female and male sex-specifically spliced CLAMP-dependent genes (Fig. S1). Exon inclusion is represented as positive PSI, and exclusion events are defined as negative PSI. We found that before MZT, the

CLAMP-dependent known male sex-specifically spliced genes have more exon inclusion events than their female counterparts (**Fig S1**). In contrast, at the post-MZT stage, both CLAMP-dependent exon inclusion and exclusion were significantly enriched in male new SSS genes compared to their female-specific counterparts (**Fig S1**). Because in the absence of CLAMP, new sex-specific spliced isoforms are generated, these data support the hypothesis that CLAMP normally inhibits aberrant splicing in male post-MZT embryos. Overall, more genes show CLAMP-dependent splicing in females than males (**Fig 2B, C**). However, the magnitude of the effect of CLAMP on splicing was stronger in males than in females (**Fig S1**).

During the first few hours of their development, *Drosophila* embryos have predominantly maternal transcripts. Therefore, we asked whether CLAMP-dependent female and male specifically-spliced genes are maternally deposited or zygotically transcribed. We overlapped our list of CLAMP-dependent sex-specifically spliced genes with known maternally expressed genes (Atallah and Lott 2018; Kwasnieski et al. 2019). We found very low levels of overlap with maternally deposited transcripts (**Fig 2E**). Therefore, most of the sex-specifically spliced genes we observed are likely to be zygotic transcripts, consistent with the function of CLAMP as a pioneer TF in the early embryo (Duan et al. 2020).

Next, we defined two groups of sex-specifically spliced genes: 1) female-specifically spliced genes, which are spliced CLAMP-dependently in females (known female sex-specifically spliced genes and novel female sex-specifically spliced genes) and 2) male-specifically spliced genes which are spliced CLAMP-dependently in males (known male sex-specifically spliced genes and novel male sex-specifically spliced genes). Then, we performed Gene Ontology analysis to determine which biological processes are enriched within CLAMP-dependent sex-specifically spliced genes at pre- and post-MZT time points, using time2splice. At the pre-MZT time point, female-specifically spliced genes are primarily mRNA binding factors such as spliceosome components and TFs (**Fig 2F**). In contrast, the male specifically-spliced genes pre-MZT are not enriched for any specific biological function or process. In females, CLAMP alters the splicing of genes that can regulate the transcription and splicing of other genes to amplify its regulatory role. At the post-MZT stage in both sexes, CLAMP regulates the splicing of genes that drive

development, such as organogenesis, morphogenesis, cell proliferation, signaling, and neurogenesis (**Fig 2F**).

3. CLAMP is highly enriched along gene bodies of sex-specifically spliced genes after MZT

Because maternal CLAMP regulates sex-specific alternative splicing, we next asked: How does CLAMP regulate female and male sex-specifically spliced genes? To address this question, we first defined the binding pattern of CLAMP at the CLAMP-dependent female and male sex-specifically spliced genes in sexed embryos (#GSE133637). We generated average profiles for CLAMP at genes showing CLAMP-dependent splicing in females and males at pre- and post-MZT time points in males and females (**Fig 3A, B**).

First, we compared and contrasted CLAMP binding patterns at sex-specifically spliced genes with genes whose expression but not splicing is sex-biased and dependent on CLAMP. We defined CLAMP-dependent sex-biased genes as differentially expressed genes in males in females and dependent on CLAMP. In both sexes, post-MZT (2-4 hours), we observed that CLAMP is enriched at the gene bodies of CLAMP-dependent sex-specifically spliced genes compared with randomized controls. In contrast, CLAMP occupancy shows a very different pattern at CLAMP-dependent genes expressed but not spliced in a CLAMP-dependent and sex-biased manner: CLAMP is enriched at the TSS and TES instead of at gene bodies (**Fig 3C and D**, rectangular box in **Fig 3A-D**). Furthermore, CLAMP binding was also modestly enriched at the TSS of female-biased expressed genes in females, consistent with enhanced CLAMP occupancy at the TSS of expressed genes (Soruco et al. 2013). As a control, we used a random set of active genes which were not regulated by CLAMP (green lines in **Fig 3A-D**) and we observed lower occupancy than at target genes. Therefore, we demonstrate that CLAMP binds specifically to gene bodies of sex-specifically spliced genes, consistent with a direct role in alternative splicing. Overall, we found preferential binding of CLAMP along the gene body of genes showing CLAMP-dependent splicing events in both females and males.

4. CLAMP regulates chromatin opening at the *sxl* gene which governs splicing of other sex determination pathway components

Because Sex-lethal (Sxl) is the master regulator of sex determination (Salz and Erickson 2010) and drives subsequent sex-specific splicing in females (Bell et al. 1991), we asked whether CLAMP regulates *sxl* gene splicing and Sxl protein levels. Functional Sxl protein is produced specifically in females because exon three in the *sxl* transcript contains a premature stop codon which is spliced out in females but retained in males (Haussmann et al. 2016). Therefore, the Sxl protein is non-functional in males and becomes degraded (Salz and Erickson 2010; Moschall et al. 2019). In females, Sxl protein inhibits the splicing of the *msl2* transcript which prevents aberrant Msl2 expression and MSL complex formation in females. Furthermore, Sxl promotes female specific-splicing of the downstream effector called transformer (*tra*). The female-specific Transformer protein regulates splicing of another downstream transcript doublesex (*dsx*), which is sex-specifically spliced. In contrast, the male-specific Transformer protein promotes male-specific splicing of doublesex (*dsx*) and the male-specific Doublesex protein isoform (**Fig 4A**). The sex-specific isoforms of the Doublesex protein drive sex-specific splicing of many downstream genes, shaping sexual variation between males and females (Nagoshi and Baker 1990; Coschigano and Wensink 1993).

Because CLAMP regulates sex-specific splicing and is maternally deposited, we hypothesized that CLAMP regulates the canonical sex determination pathway described above. To test this hypothesis, we used our previously described recessive *clamp* null mutant *clamp*^{2} line (Urban et al. 2017b), the heterozygous mutant *clamp*^{2}*/CyOGFP* line, and a rescue line which is homozygous for the *clamp*^{2} allele and contains an insertion of the wild type *CLAMP* gene. Sxl is strongly regulated via alternative splicing of its mRNA (Bell et al. 1988; Salz et al. 1989; Samuels et al. 1991). Therefore, we designed an RT-PCR assay to distinguish between the female (excluding exon 3) and male (including exon 3) versions of the *sxl* transcript (**Fig 4B**). We used this assay to measure CLAMP-dependent changes in alternative splicing and found that in homozygous *clamp*^{2} female animals, there is a small but detectable amount of the longer male-specific *sxl* transcript (**Fig 4B**, lane c). This mis-regulation of *sxl* splicing is rescued by our *CLAMP* containing rescue construct (**Fig 4B**, lane d). To test whether defects in splicing altered Sxl protein levels, we conducted western blot experiments to quantify Sxl protein in wild type, *clamp*^{2} heterozygous, and homozygous females and males (**Fig 4C**). We observed a reduction in Sxl protein levels in females

in the *clamp*^{2} null background when compared with controls. Also, homozygous *clamp*^{2} mutant males die before the third instar larval stage, and therefore it was not possible to measure the splicing of transcripts in male *clamp*^{2} mutant larvae. Overall, we determined that CLAMP promotes female-specific splicing of the *sxl* transcript to ensure that normal Sxl protein levels are produced.

To determine a possible mechanism by which CLAMP regulates splicing of *sxl*, we examined Micrococcal Nuclease sequencing data (Urban et al. 2017a), which measures chromatin accessibility in S2/male and Kc/female cells in the presence and absence of CLAMP. We found that after the loss of CLAMP in female Kc cells, chromatin accessibility around exon 3 of *sxl* increases significantly (**Fig 4D**). These data suggest that CLAMP regulates the splicing of *sxl* by decreasing chromatin accessibility of the alternatively spliced exon in females.

Next, we examined the splicing of other components of the sex determination pathway for defects in alternative splicing in the absence of CLAMP (**Fig 4E-F**). In embryos which lack CLAMP (**Fig 4E**, lane 2), the *dsx* female-specific transcript is aberrantly produced in males (**Fig 4E**, lanes 4 and 5). In contrast, the male-specific *dsx* transcript is not expressed in male embryos which lack CLAMP similar to wild type female embryos (**Fig 4E**, lane7-10). We also observed male-specific *dsx* transcripts in female *clamp*^{2} mutant larvae (**Fig 4E**, lane c).

We also found that CLAMP regulates splicing of the male-specific lethal-2 (*msl-2*) transcript, which is usually present only in males because Sxl regulates its splicing and stability. In the *clamp*^{2} mutants: 1) levels of the nonfunctional female *msl-2* isoform (**Fig 4F**, lane 5) increased in male embryos (**Fig 4F**, lane 2); 2) the functional male-specific *msl-2* isoform increased in levels in females (**Fig 4F**, lane c). To determine whether these splicing defects also cause dysregulation of MSL2 protein and localization on chromatin, we performed polytene immunostaining from female *clamp*^{2} mutant salivary glands. Ectopic MSL2 protein (in red) is present at several locations on female chromatin compared with controls (*clamp*^{2}*/CyOGFP* heterozygous female) where MSL2 protein is not present on chromatin (**Fig 4G**). Therefore, CLAMP functions to regulate splicing of *sxl* which indirectly regulates splicing of Sxl targets to promote sexual differentiation and prevent aberrant localization of MSL complex to chromatin in females. Overall, CLAMP promotes

appropriate MSL complex localization in males (Soruco et al., 2013) and prevents aberrant MSL complex localization in females.

5. CLAMP interacts with components of the spliceosome complex and influences their occupancy on chromatin

CLAMP occupancy on the gene bodies of sex-specifically spliced genes (**Fig 3**) and its physical association with spliceosome components from mass spectrometry analysis (Urban et al. 2017c) lead us to hypothesize that CLAMP regulates recruitment of components of the splicing machinery to chromatin. To test this hypothesis, we first examined how CLAMP regulates the occupancy of the MLE RNA helicase, a component of both the MSL complex and the spliceosome (Kuroda et al. 1991; Reenan et al. 2000; Herold et al. 2009; Cugusi et al. 2015; Cugusi et al. 2016) for two reasons: 1) We and others have previously demonstrated a physical association between CLAMP and MLE consistent with direct binding of CLAMP to the MSL complex in males (Quinn et al. 2014; Lindehell et al. 2015; Albig et al. 2019); 2) We hypothesized that CLAMP could regulate the distribution of MLE between the spliceosome and MSL complex to regulate MLE function in sex-specific alternative splicing in males.

To determine whether CLAMP modulates MLE distribution on chromatin, we measured MLE distribution at the genomic level using Cut-and-Run (Skene et al. 2018; Uyehara and McKay 2019). We performed Cut-and-Run in the presence (*MTD-GAL4>UAS-GFPRNAi*) and absence (*MTD-GAL4>UAS-CLAMP[val22]RNAi*) of maternal CLAMP at the pre-MZT and post-MZT embryonic stages in males. First, we compared the distribution of CLAMP peaks with MLE peaks. Almost 50% of CLAMP peaks on the X-chromosome overlap with MLE peaks (**Fig 5A** and **Fig S2A**), consistent with their association with MSL complex. After the loss of maternal CLAMP, ~23% of these MLE peaks are lost at the pre-MZT stage, which increases to ~51% the post-MZT stage. We found that overall ~60% of MLE peaks were lost by the 2-4 hr post-MZT stage in the absence of CLAMP. Moreover, ~25% (pre-MZT) and ~35% (post-MZT) of the MLE peaks observed in the absence of CLAMP were new and not present in control embryos (**Fig 5A** and **Fig S2A**). Furthermore, we hypothesized that MLE at the new peaks is part of the spliceosome complex and not MSL complex because MSL complex is not present on chromatin in the absence

of CLAMP (Soruco et al. 2013). Overall, these data suggest that CLAMP regulates the distribution of MLE on chromatin.

Next, we classified MLE peaks into two classes: 1) MLE peaks overlapping with CLAMP peaks (**Fig 5B**) and 2) unique MLE peaks which do not overlap with CLAMP peaks (**Fig 5C**) at 0-2 Hr pre-MZT and 2-4 Hr post-MZT stages. Independent of stage, MLE peaks which overlap with CLAMP peaks are mainly distributed along promoters with a lower number of peaks localizing to introns. In contrast, unique MLE peaks are primarily localized to introns (**Fig 5B, C**). This distribution is similar for X-chromosomal and autosomal peaks. The loss of CLAMP results in a considerable loss of both types of overlapping and non-overlapping MLE peaks. However, the number of new unique MLE peaks which appear in the absence of CLAMP is greater than the number of MLE peaks overlapping with CLAMP are X-enriched and most often located within introns. (**Fig 5D, E**, and **Fig S2A**). Overall, our data suggest that MLE is redistributed in the absence of CLAMP suggesting that CLAMP normally prevents aberrant recruitment of MLE in addition to the formation of aberrant splice isoforms (**Fig. 2,4**).

To provide insight into the differences between MLE peaks which overlap with CLAMP and those which do not, we identified sequence motifs which are enriched within each class of peaks using MEME within time2splice. The known CLAMP motif (Soruco et al. 2013), a stretch of (GA)n repeats, was enriched at MLE peaks overlapping with CLAMP independent of stage and chromosome type. In contrast, MLE peaks which do not overlap with CLAMP have motifs with stretches of GTs, GCTs, and GTAs but not (GA)n repeats (**Fig 5B, C**). In the absence of CLAMP, the remaining MLE peaks were most enriched for (GT)n motifs. These results support a role for CLAMP in sequestering MLE at (GA)n rich sequences within promoters which acts in opposition to the ability of MLE to bind to intronic sequences independent of CLAMP. Thus, in the absence of CLAMP, MLE is redistributed, aberrantly binding to a new set of locations which are largely within introns.

To determine how MLE redistribution could alter sex-specific splicing, we plotted the distribution of MLE binding on CLAMP-dependent female and male-specifically spliced genes (**Fig S2B**). Pre-MZT MLE binds near the TSS of male-specifically spliced genes independent of maternal

CLAMP. At the post-MZT stage, loss of maternal CLAMP causes MLE to change its binding distribution along the gene body (dotted lines: **Fig S2B**) of CLAMP-dependent males specifically spliced genes (blue line). It is possible that this change in MLE distribution alters sex-specific splicing.

Next, we asked whether CLAMP associates with the spliceosome complex components other than MLE, which is also a component of the MSL complex. We previously reported that CLAMP physically associates with several spliceosome complex components from mass spectrometry analysis (Urban et al. 2017c). To validate these associations, we performed co-immunoprecipitation (coIP) experiments to assess interactions between CLAMP and spliceosome components with known functions in sex-specific splicing, hrb27C and Squid (Blanchette et al. 2009; Herold et al. 2009; Hartmann et al. 2011). We found that in both S2 (male) and Kc (female) cells, CLAMP interacts with hrb27C. In contrast, CLAMP only associates with Squid in female Kc cells and not in male cells (**Fig S2C, D**), consistent with mass spectrometry data. Unlike MLE and CLAMP, Squid occupancy on polytene chromosomes is decreased on the male X chromosome compared with the female X chromosome (**Fig S2E**). Thus, CLAMP differentially associates with spliceosome components in males and females, providing a potential mechanism by which CLAMP can regulate sex-specific splicing. Therefore, it is possible that CLAMP influences the distribution of spliceosome components like MLE between different protein complexes such as MSL complex and the spliceosome itself. We hypothesize that differential binding of CLAMP with components of spliceosome like Squid and MLE in females and males respectively confirm sex-specific functionality of the spliceosome, thus generating transcriptome diversity between female and male sex.

Discussion:

Alternative splicing (AS) is a highly conserved mechanism for gene regulation, generating protein diversity (Blencowe 2006; Venables et al. 2012; Mayne et al. 2016). Several studies have reported highly dynamic RNA bound proteomes (RBPs) during the Maternal Zygotic Transition (MZT) across diverse phyla, with widespread alternative splicing events occurring during early embryonic development (Revil et al. 2010; Aanes et al. 2013; Paris et al. 2015; Gibilisco et al. 2016;

Kwasnieski et al. 2019). Furthermore, different isoforms are present in maternal and zygotic transcripts (Atallah and Lott 2018; Kwasnieski et al. 2019). However, the mechanisms that regulate RNA metabolism during the early hours of an organism's life remain elusive.

Maternally-deposited pioneer transcription factors drive zygotic genome activation, but their role in generating transcription diversity in the early embryo was unknown. Here, we define alternative spliced isoforms in pre- and post- MZT *Drosophila melanogaster* female and male embryos genome-wide for the first time. Furthermore, we identify a maternally-deposited pioneer TF, CLAMP, which regulates sex-specific transcript diversity in early embryos. Prior work on sex-specific transcript diversity (Ranz et al. 2003; Arbeitman et al. 2004; Zhang et al. 2007; Alekseyenko et al. 2008; Telonis-Scott et al. 2009; Hartmann et al. 2011; Lott et al. 2014; Paris et al. 2015; Sun et al. 2015; Gibilisco et al. 2016) either examined sex-biased variation in expression or sex-specific variation resulting from alternative splicing in gonads or the brain in adults. To overcome the challenge of sexing early embryos before zygotic genome activation, we used a meiotic drive system that generates sperm with either X or only Y chromosomes (Rieder et al. 2017) and focused on sex specific transcript diversity caused by alternative splicing.

Splice variants occur in different proportions over time and between sexes. To date, we lack methods to characterize these how isoforms change over time. Therefore, we developed time2splice, which identifies mechanisms to regulate temporal and sex-specific alternative splicing by combining RNA-seq and protein-DNA interaction data from CUT&RUN and ChIP-seq experiments. Time2splice has three parts: 1) temporal expression analysis; 2) temporal protein-DNA analysis, and 3) temporal multi-omics integration. The tool and analysis steps can be accessed at https://github.com/ashleymaeconard/time2splice.

Even following the initial few hours of its existence, there is a clear difference between a male and female *Drosophila* embryo's transcriptome (**Fig1,2**). Because the RNA transcript variants in both males and females encode genes involved in developmental processes, sex-specific development may occur earlier than previously thought. We demonstrate that a fundamental developmental trajectory differs between males and females from the initial hours of their existence long before

gonad formation. Such early sex-specific gene expression provides insight into how developmental anomalies that originate before gonad formation can exhibit variable penetrance between sexes.

We defined groups of genes in both males and females that undergo alternative splicing events which are regulated by maternally deposited CLAMP. Thus, the maternal environment regulates transcription initiation and shapes RNA processing by regulating post-transcriptional modifications. To provide mechanistic insight into how a pioneer factor like CLAMP regulates sex-specific splicing, we first examined the splicing of sex determination pathway components. In *Drosophila*, the master regulator Sxl initiates sex-specific splicing cascade very early in development to drive sex determination (Salz et al. 1989; Penalva and Sánchez 2003; Haussmann et al. 2016). Decreased Sxl protein levels in female *clamp*^{2} mutant s and mis-expression of female and male-specific *dsx* transcripts suggest that CLAMP may regulate sexual differentiation because sex-specific Dsx protein isoforms are known determinants of sexual dimorphism (Salz and Erickson 2010).

The key question is: How does CLAMP, a ubiquitously expressed pioneer TF, regulate sex-specific splicing? We identified several mechanisms by which CLAMP could regulate sex-specific splicing. CLAMP binds directly to intronic regions of the sex-specifically spliced genes it regulates in both males and females. Therefore, it may recruit spliceosome complex components to regulate their splicing (**Fig 6A**). In females and not in males, CLAMP binds near the early promoter of the *sxl* gene (SxlPe) and governs the chromatin environment at exon three of *sxl*, normally spliced out in females (**Fig 6B**). In the absence of CLAMP, exon three is retained even in females similar to males, resulting in the formation of a non-functional Sxl protein, which gets degraded, thus reducing Sxl protein levels in females and dysregulating downstream splicing events (**Fig 6B**). Because CLAMP binding sites are present near the promoter region of the *sxl* gene, we hypothesize that CLAMP regulates chromatin at exon three from a distance. CLAMP can mediate long-range chromatin interactions (Bag et al. 2019; Jordan and Larschan 2020) and act on chromatin accessibility at a distance (Urban et al. 2017a).

However, the regulation of Sxl provides insight into how CLAMP regulates sex-specific splicing in females but not in males. Based on our results, we hypothesize a mechanism by which CLAMP

can regulate splicing in males (**Fig 6C**). Proteomic analysis (Urban et al. 2017c) and coIPs (**Fig S2C**) show that CLAMP is associated with spliceosome complex components, including Squid and Hrb27C, further supporting a direct role for CLAMP in splicing. Interestingly, CLAMP associates with Squid, specifically in females. In contrast to Squid, the MLE component of both the spliceosomes and MSL complex (Cugusi et al. 2015) physically interacts with CLAMP in males. Because Squid is already known to regulate sex-specific splicing (Hartmann et al. 2011), these data support a model in which differential CLAMP binding to spliceosome complex components in males and females regulates sex-specific splicing.

Our results demonstrate that CLAMP inhibits aberrant splicing events in males, especially at the post-MZT stage (**Fig 2**). In males, we found that MLE associates with CLAMP at promoter regions. In contrast, MLE shows more affinity towards intronic sequences, a prerequisite for its splicing function, when it does not associate with CLAMP (**Fig 6C**). However, in the absence of CLAMP, the promoter bound MLE is reduced, and MLE re-localizes from its normal intronic binding sites to new intronic regions. Therefore, we hypothesize that CLAMP regulates the localization of MLE to prevent the aberrant formation of female-specific transcripts in males that we observed in the absence of CLAMP.

In males, CLAMP recruits the MSL complex to the X chromosome, thus regulating dosage compensation. CLAMP physically interacts directly with MSL2 (Albig et al. 2019) and associates with MLE, an RNA helicase that acts on the *roX* non-coding RNA components of MSL complex (Lindehell et al. 2015; Cugusi et al. 2016; Urban et al. 2017b; Ankush Jagtap et al. 2019; Lv et al. 2019). Furthermore, RNA transcript splicing is regulated by the spliceosome, another ribonucleoprotein complex, consisting of many protein and RNA components (Herold et al. 2009). Because MLE is part of both the MSL complex and the spliceosome complex, we hypothesized that CLAMP influences the relative distribution of MLE between the MSL complex and the spliceosome (**Fig 6C**). Without CLAMP, the MSL complex does not localize to the X-chromosome (Soruco et al. 2013); thus, MLE is no longer part of the MSL complex and is available to redistribute to new spliceosomes binding sites. Therefore, we provide evidence to support a model in which CLAMP inhibits aberrant binding of MLE to other genes as part of spliceosome, which regulates their splicing.

Overall, we hypothesize that both different spliceosome compositions and differential recruitment to chromatin drive developmental changes in splicing, such as sex-specific splice variants. We identify CLAMP as a maternal factor that regulates sex-specific alternative splicing through its sex-biased association with spliceosome components and its ability to influence the sex determination pathway. Identifying the factors that regulate this sex-biased CLAMP association with spliceosome complex components will be a key future direction.

Here, we show for the first time that a maternal factor controls sex-specific splicing during early embryonic development, highlighting how the maternal environment influences transcript diversity in the zygote from activation of the zygotic genome to the processing of zygotic RNA products. Furthermore, our results suggest that CLAMP could be one example of a more general splicing regulatory mechanism controlled by the interaction between pioneer TFs directly binding the chromatin and components of the RNA processing machinery, which generates spatio-temporal diversity in the transcriptome. While we analyzed sex-specific transcriptome diversity in this study, similar mechanisms could drive cell-type specific variation. For example, cell fate-determining transcription factors could regulate the chromatin occupancy of splicing complex components to promote the formation of cell-type-specific isoforms. We also present time2splice, a new tool to uncover mechanisms which drive such spatio-temporal transcript diversity by integrating expression and chromatin occupancy data.

Material and Methods:

Fly strains and rearing

Drosophila melanogaster fly stocks were maintained at 24°C on standard corn flour sucrose media. Fly strains used: *MTD-GAL4* (Bloomington, #31777), *UAS-CLAMPRNAi[val22]* (Bloomington, #57008), Meiotic drive fly stocks +; SD72/CyO and 19-3, yw, Rsp[s]-B[s]/Dp(2:y)CB25-4, y+, Rsp[s]B[s]; SPSD/CyO (Bloomington, #64332) (both gifts from Cynthia Staber). These were crossed to obtained male and female embryo of desired genotypes according to Rieder et al 2017.

Sample collection and Western blotting

Salivary glands from third instar larvae were dissected in cold PBS and samples frozen in liquid nitrogen. Total protein from the samples was extracted by homogenizing tissue in the lysis buffer (50mM Tris-HCl pH 8.0, 150mM NaCl, 1% SDS, 0.5X protease inhibitor) using a small pestle. After a five-minute incubation at room temperature, cleared the samples by centrifuging at room temperature for 10 minutes at 14,000xg. To blot for CLAMP and Actin, 5 micrograms of total protein was run on a Novex 10% Tris-Glycine precast gel (Life technologies). To measure Sex-lethal protein levels, 20 micrograms of total protein was run on a Novex 12% Tris-Glycine precast gel (Life technologies). Protein was transferred to PVDF membranes using the iBlot transfer system (ThermoFisher Scientific) and probed the membranes for CLAMP (1:1000, SDIX), Actin (1:400,000, Millipore), and SXL (1:500, a gift from Fatima Gebauer) antibodies using the Western Breeze kit following the manufacturer's protocol (ThermoFisher Scientific). We quantified the relative expression of protein for SXL using the gel analysis tool in ImageJ software following the website's guidelines (Schneider et al. 2012). For each genotype, we first internally normalized the amount of SXL protein to Actin. Next, we determined the protein's relative expression by comparing the Actin normalized quantities to y[1], w[1118] female samples.

Polytene chromosome squashes and immunostaining

Polytene chromosome squashes were prepared as previously described in Reider et al. 2017. We stained polytene chromosomes with rabbit anti-CLAMP (1:1000, SDIX), mouse anti-Squid (1:50, 1B11, DSHB), rabbit anti-MLE (1:1000, gift from Mitzi Kuroda), rat anti-MSL2 (1:500, gift from Peter Becker) antibodies. For detection, we used all Alexa Fluor secondary antibodies against rabbit and mouse at a concentration of 1:1000 and visualized slides at 40X on a Zeiss Axioimager M1 Epifluorescence upright microscope with the AxioVision version 4.8.2 software.

Splicing assay for male and female-specific transcripts

To test for the male and female splice forms of sex-lethal, transformer, doublesex, and msl2, total RNA was extracted from ten third instar larvae from each genotype. We reverse-transcribed two micrograms of total RNA using the SuperScript VILO cDNA Synthesis Kit (Life Technologies) following the manufacturer's protocol. We amplified target sequences by PCR using primers

designed to span Alternatively spliced junctions. Alternative splicing primer sequences for sxl FP-TGCAACTCACCTCATCATCC, sxl RP- GATGGCAGAGAATGGGACAT, for tra FP-TGAAAATGGATGCCGACAG, tra RP- CTCTTTGGCGCAATCTTCTC, for dsx female transcript dsxFFP-CTATCCTTGGGAGCTGATGC, dsxF RP-TCGGGGCAAAGTAGTATTCG, for dsx male transcript dsxM FP-CAGACGCCAACATTGAAGAG, dsxM RP- CTGGAGTCGGTGGACAAATC, for msl2 FP-GTCACACTGGCTTCGCTCAG and msl2 RP- CCTGGGCTAGTTACCTGCAA were used.

Immunoprecipitation

Nuclear extract preparation: Male (S2) and female (Kc) cells were grown to a cell concentration of 2X 10^6 cells/mL in T25 tissue culture flasks. Cells were scraped from the flask, centrifuged for 5min at 2500rpm at 4°C. Supernatant was removed and cell pellets were washed twice in 5ml of cold PBS. The washed cell pellets were then resuspended in 5X volume of Buffer A (10mM HEPES pH7.9, 1.5mM $MgCl_2$, 10mM KCl, 0.5mMDTT, 1X Protease inhibitors). Cells were incubated on ice for 15 minutes before dounce homogenization with an A pestle. Cytoplasmic fraction was collected after centrifugation at 4°C for 20 min at 700xg. The remaining nuclear pellet was re-suspended in 3 times volume in Buffer B (20mM HEPES pH7.9, 20%Glycerol, 0.5%NP 40, 200mMKCl, 0.5mMEDTA, 1mMEGTA, 1X protease inhibitors). Nuclei after re-suspension were dounce homogenized with a B pestle. Nuclear debris was then pelleted by centrifugation at 10,000xg for 10 min at 4°C. 1ml aliquots of both cytoplasmic and nuclear fractions were prepared in 1.5mL Protein LoBind Eppendorf tubes and flash frozen in liquid nitrogen for storage at -80 °C.

Immunoprecipitation: Magnetic anti-CLAMP beads were prepared to a final concentration of 10mg/mL by coupling rabbit anti-CLAMP antibody (SDIX) to magnetic beads, according to Dynabeads Antibody coupling kit (ThermoFisher Scientific) instructions. Both prepared anti-CLAMP and purchased anti-IgG (anti-rabbit IgG M-280 Dynabeads) were blocked to reduce background the night before the immunoprecipitation. First, the beads were washed 3 times for 5 minutes in 500L Tris-NaCl Wash (50mM Tris, 500mM NaCl, 0.1% NP-40) by rotating at 4C. The beads were next suspended in block buffer (3.3mg/mL of yeast tRNA extract prepared in 20mM HEPES, pH7.9, 20% Glycerol, 0.5% NP-40, 200mM KCl, 1mM EDTA, and 2mM EGTA) and rotated overnight at 4C. The next day, beads were washed 3 times for 5 minutes in the block buffer

without yeast tRNA by rotating at 4°C. After the final wash, beads were resuspended in the same amount of block buffer as the starting volume.

To 1mL of previously prepared nuclear extract, 100uL of blocked anti-CLAMP or anti-IgG magnetic Dynabeads were added. The nuclear extracts and beads were then rotated for 1 hour at 4°C. Afterward, the beads were collected and the supernatant discarded. The beads were then washed three times in Tris-NaCl wash (50mM Tris, 500mM NaCl, 0.1% NP-40) by rotating for 5 minutes at 4°C and clearing by using a magnetic rack. To elute proteins from the beads, 100uL of 1% SDS was added, and the beads were boiled for 10 minutes at 95C. To the eluate, 300uL of ultrapure water was added, and the tubes gently vortexed. After collecting the beads on a magnetic rack, the eluate was saved in a clean Protein LoBind Eppendorf tube.

Western blotting: Squid and Hrb27C detected in IP-CLAMP and IGG-rabbit protein samples using mouse anti-Squid (1:500, 1B11, DSHB) and rabbit anti-Hrb27C (1:5000, Fatima Gebauer), performed as mentioned above under western blotting protocol.

Cut and Run

0-2 hr and 2-4 hr male and female embryos of desired genotypes (~50 each) collected on standard grape juice agar medium and washed with water. The embryos were dechorionated in 6% bleaching solution for 2 min and washed twice in ice cold 1XPBS. Centrifuged at 12,000g for 10 min at 4°C. Supernatant discarded and embryos resuspended in 200µl Dig-Wash buffer with EDTA (20mM HEPES-NaOH, 150mM NaCl, 2mM EDTA, 0.5mM Spermidine, 10mM PMSF, 0.05% digitonin) and washed twice. Embryos were incubated in 200µl primary antibody overnight at 4°C on a tube rotator. Centrifuged at 12,000g for 10 min at 4°C and liquid removed, embryos washed twice in Dig-Wash buffer with EDTA. Incubated for 3 hours at 4°C in ~700 ng/ml pAMNase solution in Dig-Wash buffer with EDTA. Washed twice in Dig-Wash buffer without EDTA and resuspended in 150µl of Dig-Wash buffer without EDTA. Samples were equilibrated to 0°C on a heat block maintained on ice-bath. 2µl of 100mm $CaCl_2$ added to each sample to initiate MNase activity and digestion was allowed till 30 min before adding 150µl of 2X RSTOP Buffer (200mM NaCl, 20mM EDTA, 4mM EGTA, 50ug/ml RNase, 40ug/ml glycogen, 10pg/ml

yeast spike-in DNA) to stop the reaction. Incubated at 37°C for 10 minutes to release the DNA fragments. Samples spun at 12,000g for 10 minutes and aqueous layer transferred to a fresh 1.5 ml microfuge tube and centrifuged at 16,000g for 5 minutes. Cleared liquid again transferred to a fresh tube, 1µl of 20% SDS and 2.5µl proteinase K (20ng/ml) added, incubated at 70°C for 10 minutes. 300µl PCI added to each tube, mixed and total solution transferred to phase lock tubes and centrifuged at 16,000g for 5 minutes. After adding 300µl of chloroform and mixing gently, centrifuged at 16,000g for 5 minutes at RT. Aqueous layer transferred to DNA low binding tube. 1µl glycogen (5mg/ml) and 750µl ethanol added to precipitate DNA at -80°C. Centrifuged at 16,000g for 10 min at 4°C and washed in ethanol twice. Pellet air dried and dissolved in 15µl of 1mM TrisHCl + 0.1mM EDTA pH 8.0 (Skene et al. 2018; Uyehara and McKay 2019). 1ng of cut and Run DNA was used to make libraries using KAPA Hyper prep kit and SeqCap adapters A &B (Roche) according to manufacturer's protocol. For library amplification 14 cycles used and a 1.0X SPRI bead cleanup was performed using Agencourt Ampure XP beads. The following antibody concentrations were used: rabbit anti-CLAMP (5µg/sample, SDIX); 1:200 anti-rabbit (MilliporeSigma); rat anti-MLE (1:50, 6E11); 700ng/ml pA-MNase (from Steven Henikoff).

Computational Methods:

Time2splice tool: see Supplementary method

Sex biased genes and sex-specific spliced genes

RNA sequencing data from Rieder et al 2017 (#GSE102922) was analyzed using time2splice to determine sex-specifically spliced genes and sex-biased genes. dmel-all-r6.29.gtf from BDGP6 in genomes (DePristo et al. 2011) was used to map each transcript identifier (ID) to gene ID and symbol, for .bed creation data for the associated chromosome, transcription start site (TSS) and transcription end site (TES), and strand information were imported from Illumina (https://support.illumina.com/sequencing/sequencing_software/igenome.html). From the raw data after quality control i.e, FastQC (Andrews 2010), Salmon (Patro et al. 2017) was used to quantify transcript expression for treatment and control samples. Calculated transcripts per million (TPM) values from SUPPA (Trincado et al. 2018) were used for all four replicates of female and male controls at both time points (before and after MZT). Each sample was filtered to include transcripts

where the mean value is less than or equal to 3 TPMs per gene. The number of transcripts included at various thresholds were plotted from 1 to 10 and the fraction of genes filtered out begins to plateau around threshold 3. The percent of spliced in (PSI) transcripts between females and males were compared at both pre-MZT and post-MZT (p-value of 0.05), thereby resulting in delta PSI values and p-values for each transcription in each experimental condition comparison. Given these resulting delta transcript PSI values, significantly alternatively splice genes (p-value 0.05) were found between females vs. males pre-MZT controls to show which genes are normally sex-specifically spliced pre-MZT. These enabled us to determine the sex-biased genes, that is, the genes that are uniquely expressed in males or females. The same process was followed at post-MZT. To then determine the sex-specifically spliced genes, the female RNAi experiment compared with the control delta PSI gave the number of total alternative spliced transcripts pre-MZT, then considering those that are not shared with males, and are only expressed in females normally, this defined our sex specifically spliced set of genes for females pre-MST. This process was also performed for males pre-MZT. The same process was followed at post-MZT.

ChIP-seq: Data analysis

We used preprocessed ChIP-seq data from Rieder et al 2019 (#GSE133637), specifically the .bw and .broadPeak.gz files in our analysis using ChIPseeker (Yu et al. 2015) and deeptools (Ramírez et al. 2014). Specifically, when plotting the average profiles using deeptools, we achieved a baseline signal representing genome-wide binding taking into consideration the number of genes in other groups by the following procedure: of all genes that are on (no zero read-count genes), we sampled the number of the largest other group (to which we are comparing), and ran computeMatrix on that subset. This process was repeated 500 times and the resulting 500 matrices were averaged to produce a representative signal. For motif analysis MEME (Bailey et al. 2015) suite was used.

Cut and Run: Data analysis

Sequenced reads were run through FASTQC (Andrews 2010)(fastqc replicate_R1_001.fastq.gz replicate_R2_001.fastq.gz) with default parameters to check the quality of raw sequence data and

filter out any sequences flagged for poor quality. Sequences were trimmed and reassessed for quality using TrimGalore (https://github.com/FelixKrueger/TrimGalore/issues/25) and FastQC (Andrews 2010), respectively. All Illumina lanes of the same flow cell .fastq files were merged, and sequenced reads were then mapped to release 6 *Drosophila melanogaster* genome (dm6). We compared Bowtie2 (Langmead and Salzberg 2012), HISAT2 (Kim et al. 2019), and BWA (Li and Durbin 2009). We found the best alignment quality with BWA and thus used this method's results downstream. Next, we performed conversion to bam and sorting (e.g. using: bowtie2 -x dm6_genome -1 replicate_R1_001.fastq.gz -2 replicate_R2_001.fastq.gz -S out.sam > stout.txt 2> alignment_info.txt; samtools view -bS out.sam > out.bam; rm -rf out.sam; samtools sort out.bam -o out.sorted.bam). We removed reads (using samtools) with a MAPQ less than 30 and any reads with PCR duplicate reads (identified using MarkDuplicates Picard -2.20.2). Peaks identified using MACS2 (Zhang et al. 2008)(macs2 callpeak -t out.sorted.bam -B -f BAM --nomodel --SPMR --keep-dup all -g dm --trackline -n outname --cutoff-analysis --call-summits -p 0.01 --outdir outdir) and keep duplicates separate**.** To calculate fold-enrichment macs2 is run again (macs2 bdgcmp -t $treat -c $control -o $out.sorted.bam_FE.bdg -m FE 2> $ out.sorted.bam_FE.log; macs2 bdgcmp -t $treat -c $control -o $out.sorted.bam_logLR.bdg -m logLR -p 0.00001 2). For motif analysis MEME (Bailey et al. 2015) suite was used. Data submitted in GEO repository (awaiting accession number).

Competing Interest Statement

Authors declare no conflicting interests

Acknowledgments

This work and funding to M.R. was supported by R35GM126994 to E.N.L. from NIH. A.M.C funded by NSF Graduate Research Fellowship and CCMB, Brown University. We thank Bloomington stock center for fly lines. We thank Leila Rieder for SD and Rsp^s stocks, Peter Becker for MSL2 and MLE antibodies, Steve Heinkoff for pAMNase antibody and spike-in DNA for Cut and Run, Daniel J McKay for sharing Cut and Run protocol for tissues.

Author Contributions

M.R., A.M.C. and E.N.L. planned experiments, analyzed results and wrote the manuscript. A.M.C did all the computational analysis. M.R. carried out the experimental work and collected data for Cut and Run, Polytene squashes and IF, splicing assay and IP. J.U. carried out splicing assay and WB. E.N.L acquired the funding.

References

Aanes H, Østrup O, Andersen IS, Moen LF, Mathavan S, Collas P, Alestrom P. 2013. Differential transcript isoform usage pre-and post-zygotic genome activation in zebrafish. *BMC genomics* **14**: 331.

Albig C, Tikhonova E, Krause S, Maksimenko O, Regnard C, Becker PB. 2019. Factor cooperation for chromosome discrimination in Drosophila. *Nucleic acids research* **47**: 1706-1724.

Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee O-K, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ. 2008. A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* **134**: 599-609.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Ankush Jagtap PK, Müller M, Masiewicz P, von Bülow S, Hollmann NM, Chen P-C, Simon B, Thomae AW, Becker PB, Hennig J. 2019. Structure, dynamics and roX2-lncRNA binding of tandem double-stranded RNA binding domains dsRBD1, 2 of Drosophila helicase Maleless. *Nucleic acids research* **47**: 4319-4333.

Arbeitman MN, Fleming AA, Siegal ML, Null BH, Baker BS. 2004. A genomic analysis of Drosophila somatic sexual differentiation and its regulation. *Development* **131**: 2007-2021.

Atallah J, Lott SE. 2018. Evolution of maternal and zygotic mRNA complements in the early Drosophila embryo. *PLoS genetics* **14**: e1007838.

Bag I, Dale RK, Palmer C, Lei EP. 2019. The zinc-finger protein CLAMP promotes gypsy chromatin insulator function in Drosophila. *Journal of cell science* **132**.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. *Nucleic acids research* **43**: W39-W49.

Bell LR, Horabin JI, Schedl P, Cline TW. 1991. Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in Drosophila. *Cell* **65**: 229-239.

Bell LR, Maine EM, Schedl P, Cline TW. 1988. Sex-lethal, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55**: 1037-1046.

Blanchette M, Green RE, MacArthur S, Brooks AN, Brenner SE, Eisen MB, Rio DC. 2009. Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. *Molecular cell* **33**: 438-449.

Blencowe BJ. 2006. Alternative splicing: new insights from global analyses. *Cell* **126**: 37-47.

Brooks AN, Duff MO, May G, Yang L, Bolisetty M, Landolin J, Wan K, Sandler J, Booth BW, Celniker SE. 2015. Regulation of alternative splicing in Drosophila by 56 RNA binding proteins. *Genome research* **25**: 1771-1780.

Coschigano KT, Wensink PC. 1993. Sex-specific transcriptional regulation by the male and female doublesex proteins of Drosophila. *Genes & development* **7**: 42-54.

Cugusi S, Kallappagoudar S, Ling H, Lucchesi JC. 2015. The Drosophila helicase maleless (MLE) is implicated in functions distinct from its role in dosage compensation. *Molecular & Cellular Proteomics* **14**: 1478-1488.

Cugusi S, Li Y, Jin P, Lucchesi JC. 2016. The Drosophila helicase MLE targets hairpin structures in genomic transcripts. *PLoS genetics* **12**: e1005761.

De Renzis S, Elemento O, Tavazoie S, Wieschaus EF. 2007. Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS Biol* **5**: e117.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**: 491.

Duan JE, Rieder LE, Huang A, Jordan WT, McKenney M, Watters S, Fawzi NL, Larschan EN. 2020. CLAMP and Zelda function together as pioneer transcription factors to promote Drosophila zygotic genome activation. *bioRxiv*.

Estes PA, Keyes LN, Schedl P. 1995. Multiple response elements in the Sex-lethal early promoter ensure its female-specific expression pattern. *Molecular and Cellular Biology* **15**: 904-917.

Faustino NA, Cooper TA. 2003. Pre-mRNA splicing and human disease. *Genes & development* **17**: 419-437.

Förch P, Valcárcel J. 2003. Splicing regulation in Drosophila sex determination. in *Regulation of Alternative Splicing*, pp. 127-151. Springer.

Gebauer F, Merendino L, Hentze MW, Valcárcel J. 1998. The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. *Rna* **4**: 142-150.

Gibilisco L, Zhou Q, Mahajan S, Bachtrog D. 2016. Alternative splicing within and between Drosophila species, sexes, tissues, and developmental stages. *PLoS genetics* **12**: e1006464.

González AN, Lu H, Erickson JW. 2008. A shared enhancer controls a temporal switch between promoters during Drosophila primary sex determination. *Proceedings of the National Academy of Sciences* **105**: 18436-18441.

Guilgur LG, Prudêncio P, Sobral D, Liszekova D, Rosa A, Martinho RG. 2014. Requirement for highly efficient pre-mRNA splicing during Drosophila early embryonic development. *Elife* **3**: e02181.

Hartmann B, Castelo R, Miñana B, Peden E, Blanchette M, Rio DC, Singh R, Valcárcel J. 2011. Distinct regulatory programs establish widespread sex-specific alternative splicing in Drosophila melanogaster. *Rna* **17**: 453-468.

Haussmann IU, Bodi Z, Sanchez-Moran E, Mongan NP, Archer N, Fray RG, Soller M. 2016. m 6 A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* **540**: 301-304.

Herold N, Will CL, Wolf E, Kastner B, Urlaub H, Lührmann R. 2009. Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes. *Molecular and cellular biology* **29**: 281-301.

Jordan W, Larschan E. 2020. The zinc finger protein CLAMP promotes long-range chromatin interactions that mediate dosage compensation of the Drosophila male X-chromosome. *bioRxiv*.

Kaye EG, Booker M, Kurland JV, Conicella AE, Fawzi NL, Bulyk ML, Tolstorukov MY, Larschan E. 2018. Differential occupancy of two GA-binding proteins promotes targeting of the drosophila dosage compensation complex to the male X chromosome. *Cell reports* **22**: 3227-3239.

Keyes LN, Cline TW, Schedl P. 1992. The primary sex determination signal of Drosophila acts at the level of transcription. *Cell* **68**: 933-943.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37**: 907-915.

Kuroda MI, Kernan MJ, Kreber R, Ganetzky B, Baker BS. 1991. The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila. *Cell* **66**: 935-947.

Kwasnieski JC, Orr-Weaver TL, Bartel DP. 2019. Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns. *Genome research* **29**: 1188-1197.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357.

Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, Kuroda MI. 2011. X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. *Nature* **471**: 115-118.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**: 1754-1760.

Lindehell H, Kim M, Larsson J. 2015. Proximity ligation assays of protein and RNA interactions in the male-specific lethal complex on Drosophila melanogaster polytene chromosomes. *Chromosoma* **124**: 385-395.

Lott SE, Villalta JE, Zhou Q, Bachtrog D, Eisen MB. 2014. Sex-specific embryonic gene expression in species with newly evolved sex chromosomes. *PLoS Genet* **10**: e1004159.

Lv M, Yao Y, Li F, Xu L, Yang L, Gong Q, Xu Y-Z, Shi Y, Fan Y-J, Tang Y. 2019. 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* **47**: 3142-3157.

Mayne BT, Bianco-Miotto T, Buckberry S, Breen J, Clifton V, Shoubridge C, Roberts CT. 2016. Large scale gene expression meta-analysis reveals tissue-specific, sex-biased gene expression in humans. *Frontiers in genetics* **7**: 183.

Moschall R, Rass M, Rossbach O, Lehmann G, Kullmann L, Eichner N, Strauss D, Meister G, Schneuwly S, Krahn MP. 2019. Drosophila Sister-of-Sex-lethal reinforces a male-specific gene expression pattern by controlling Sex-lethal alternative splicing. *Nucleic acids research* **47**: 2276-2288.

Nagoshi RN, Baker BS. 1990. Regulation of sex-specific RNA splicing at the Drosophila doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes & development* **4**: 89-97.

Ober C, Loisel DA, Gilad Y. 2008. Sex-specific genetic architecture of human disease. *Nature Reviews Genetics* **9**: 911-922.

Paris M, Villalta JE, Eisen MB, Lott SE. 2015. Sex bias and maternal contribution to gene expression divergence in Drosophila blastoderm embryos. *PLoS Genet* **11**: e1005592.
Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* **14**: 417-419.
Penalva LO, Sánchez L. 2003. RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation. *Microbiology and molecular biology reviews* **67**: 343-359.
Prayitno K, Schauer T, Regnard C, Becker PB. 2019. Progressive dosage compensation during Drosophila embryogenesis is reflected by gene arrangement. *EMBO reports* **20**: e48138.
Quinn JJ, Ilik IA, Qu K, Georgiev P, Chu C, Akhtar A, Chang HY. 2014. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nature biotechnology* **32**: 933-940.
Quinn JJ, Zhang QC, Georgiev P, Ilik IA, Akhtar A, Chang HY. 2016. Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. *Genes & development* **30**: 191-207.
Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**: W187-W191.
Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the Drosophila transcriptome. *Science* **300**: 1742-1745.
Reenan RA, Hanrahan CJ, Ganetzky B. 2000. The mlenapts RNA helicase mutation in Drosophila results in a splicing catastrophe of the para Na+ channel transcript in a region of RNA editing. *Neuron* **25**: 139-149.
Revil T, Gaffney D, Dias C, Majewski J, Jerome-Majewska LA. 2010. Alternative splicing is frequent during early embryonic development in mouse. *BMC genomics* **11**: 399.
Rieder LE, Jordan III WT, Larschan EN. 2019. Targeting of the dosage-compensated male X-chromosome during early Drosophila development. *Cell reports* **29**: 4268-4275. e4262.
Rieder LE, Koreski KP, Boltz KA, Kuzu G, Urban JA, Bowman SK, Zeidman A, Jordan WT, Tolstorukov MY, Marzluff WF. 2017. Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP. *Genes & development* **31**: 1494-1508.
Salz H, Erickson JW. 2010. Sex determination in Drosophila: The view from the top. *Fly* **4**: 60-70.
Salz HK. 2007. Male or female? The answer depends on when you ask. *PLoS Biol* **5**: e335.
Salz HK, Maine EM, Keyes LN, Samuels ME, Cline TW, Schedl P. 1989. The Drosophila female-specific sex-determination gene, Sex-lethal, has stage-, tissue-, and sex-specific RNAs suggesting multiple modes of regulation. *Genes & development* **3**: 708-719.
Samuels ME, Schedl P, Cline TW. 1991. The complex set of late transcripts from the Drosophila sex determination gene sex-lethal encodes multiple related polypeptides. *Molecular and cellular biology* **11**: 3584-3602.
Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**: 671-675.
Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, McKay DJ, Harrison MM. 2015. Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome research* **25**: 1715-1726.
Schulz KN, Harrison MM. 2019. Mechanisms regulating zygotic genome activation. *Nature Reviews Genetics* **20**: 221-234.

Skene PJ, Henikoff JG, Henikoff S. 2018. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature protocols* **13**: 1006.

Soruco MM, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, Sugden AU, Goebel K, Feng J, Xia P. 2013. The CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage compensation. *Genes & development* **27**: 1551-1556.

Straub T, Zabel A, Gilfillan GD, Feller C, Becker PB. 2013. Different chromatin interfaces of the Drosophila dosage compensation complex revealed by high-shear ChIP-seq. *Genome research* **23**: 473-485.

Sun X, Yang H, Sturgill D, Oliver B, Rabinow L, Samson M-L. 2015. Sxl-Dependent, tra/tra2-Independent Alternative Splicing of the Drosophila melanogaster X-Linked Gene found in neurons. *G3: Genes, Genomes, Genetics* **5**: 2865-2874.

Telonis-Scott M, Kopp A, Wayne ML, Nuzhdin SV, McIntyre LM. 2009. Sex-specific splicing in Drosophila: widespread occurrence, tissue specificity and evolutionary conservation. *Genetics* **181**: 421-434.

Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyras E. 2018. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome biology* **19**: 1-11.

Urban J, Kuzu G, Bowman S, Scruggs B, Henriques T, Kingston R, Adelman K, Tolstorukov M, Larschan E. 2017a. Enhanced chromatin accessibility of the dosage compensated Drosophila male X-chromosome requires the CLAMP zinc finger protein. *PloS one* **12**: e0186855.

Urban JA, Doherty CA, Jordan WT, Bliss JE, Feng J, Soruco MM, Rieder LE, Tsiarli MA, Larschan EN. 2017b. The essential Drosophila CLAMP protein differentially regulates non-coding roX RNAs in male and females. *Chromosome Research* **25**: 101-113.

Urban JA, Urban JM, Kuzu G, Larschan EN. 2017c. The Drosophila CLAMP protein associates with diverse proteins on chromatin. *PloS one* **12**: e0189772.

Uyehara CM, McKay DJ. 2019. Direct and widespread role for the nuclear receptor EcR in mediating the response to ecdysone in Drosophila. *Proceedings of the National Academy of Sciences* **116**: 9893-9902.

Venables JP, Tazi J, Juge F. 2012. Regulated functional alternative splicing in Drosophila. *Nucleic acids research* **40**: 1-10.

Wang G-S, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics* **8**: 749-761.

Yu G, Wang L-G, He Q-Y. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**: 2382-2383.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**: 1-9.

Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased gene expression in the genus Drosophila. *Nature* **450**: 233-237.

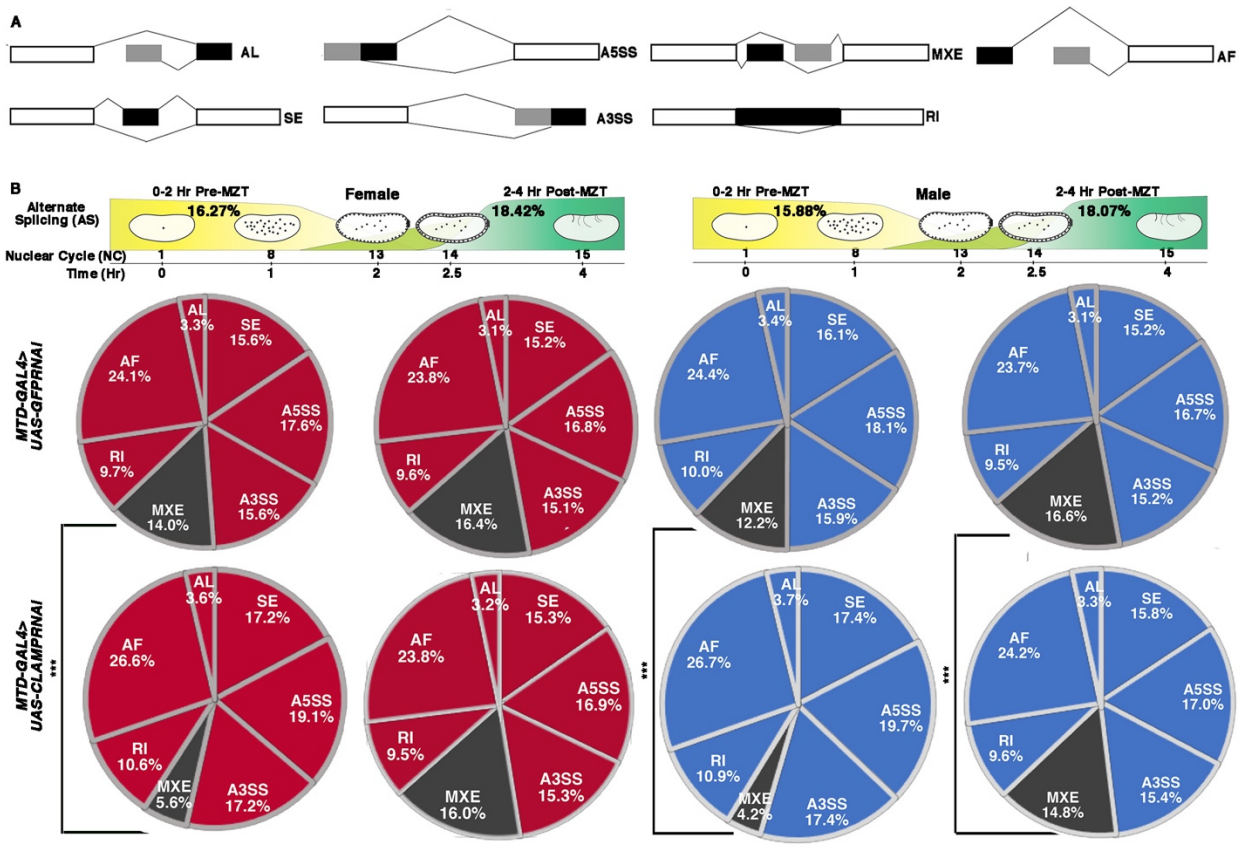

Fig 1. Alternative splicing during early Drosophila melanogaster embryonic development

A Schematic diagram is showing 7 different categories of Alternative splicing (AS).

B Pie chart showing the distribution of various categories of AS at 0-2 Hr pre-MZT and 2-4 Hr post-MZT female (red) and male (Blue) embryos in the presence (top row) and absence (bottom row) of maternal CLAMP. The total percentage of AS in each category in control embryos mentioned at the top. A Chi-square test was performed to check the significance between the percentage of MXE splicing (grey region) in the presence and absence of CLAMP in each category, females/males 0-2, and 2-4 Hr embryo. A significant difference ($p<0.001$ marked by ***) was found between categories connected by solid black lines.

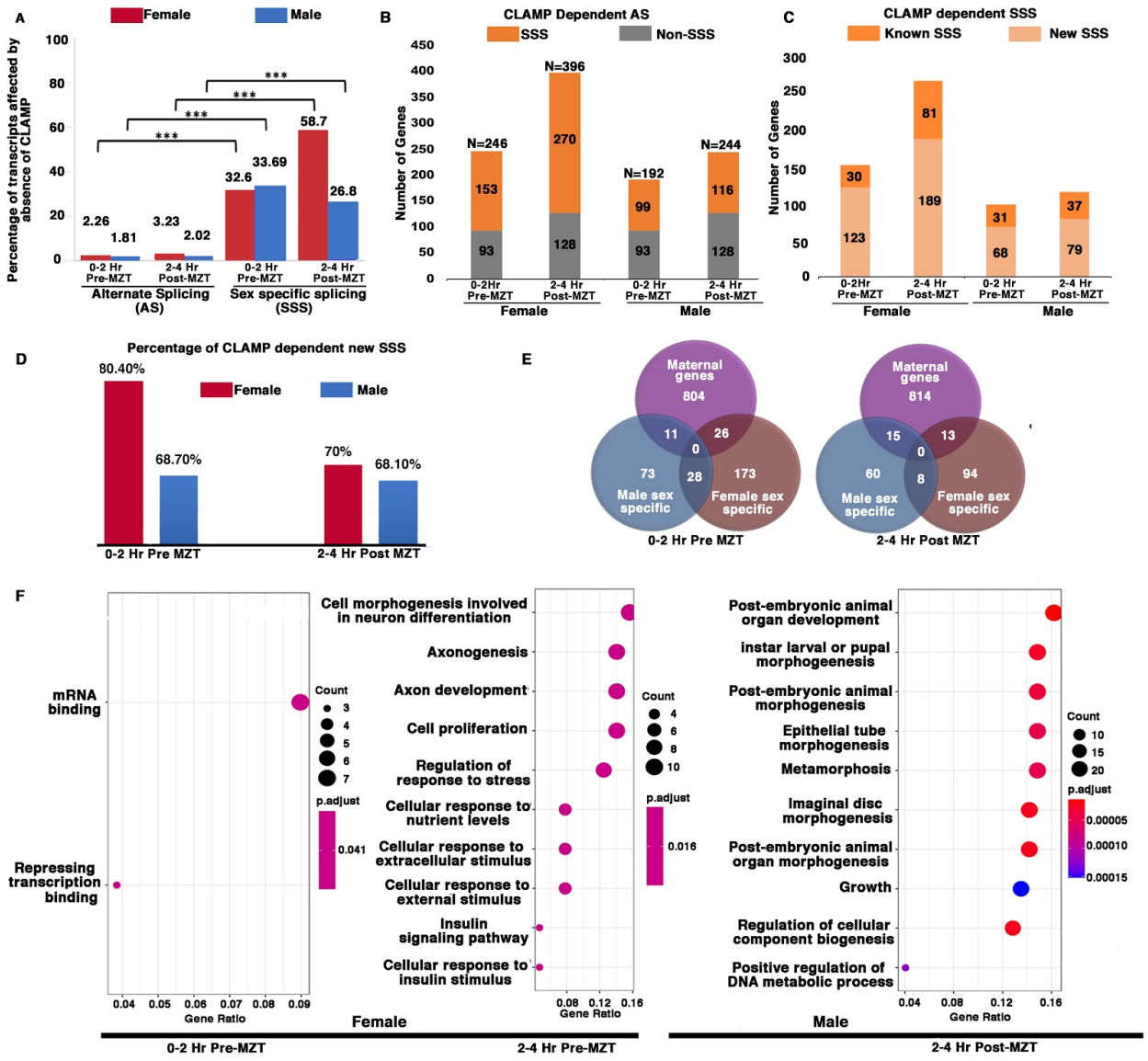
A
Female
Male
Percentage of transcripts affected by absence of CLAMP
2.26
1.81
3.23
2.02
32.6
33.69
58.7
26.8
0-2 Hr Pre-MZT
2-4 Hr Post-MZT
Alternate Splicing (AS)
Sex specific splicing (SSS)
B
CLAMP Dependent AS
SSS
Non-SSS
Number of Genes
N=246
N=396
N=192
N=244
153
93
270
128
99
93
116
128
Female
Male
C
CLAMP dependent SSS
Known SSS
New SSS
30
123
81
189
31
68
37
79
D
Percentage of CLAMP dependent new SSS
80.40%
68.70%
70%
68.10%
0-2 Hr Pre MZT
2-4 Hr Post MZT
E
Maternal genes
804
11
26
0
73
28
173
Male sex specific
Female sex specific
0-2 Hr Pre MZT
Maternal genes
814
15
13
0
60
8
94
Male sex specific
Female sex specific
2-4 Hr Post MZT
F
mRNA binding
Repressing transcription binding
Count
p.adjust
0.041
Gene Ratio
Cell morphogenesis involved in neuron differentiation
Axonogenesis
Axon development
Cell proliferation
Regulation of response to stress
Cellular response to nutrient levels
Cellular response to extracellular stimulus
Cellular response to external stimulus
Insulin signaling pathway
Cellular response to insulin stimulus
0.016
Female
Post-embryonic animal organ development
instar larval or pupal morphogeenesis
Post-embryonic animal morphogenesis
Epithelial tube morphogenesis
Metamorphosis
Imaginal disc morphogenesis
Post-embryonic animal organ morphogenesis
Growth
Regulation of cellular component biogenesis
Positive regulation of DNA metabolic process
Male
0-2 Hr Pre-MZT
2-4 Hr Pre-MZT
2-4 Hr Post-MZT

Fig 2.

Fig 2. Maternal CLAMP regulates sex-specific alternative splicing during early embryonic development in Drosophila.

A Bar graph showing the percentage of transcripts out of total Alternative splicing and sex-specific splicing events affected by a decrease in maternal CLAMP, at pre-MZT and post-MZT stages, in females (red bars) and males (blue bars). A Fischer's Exact Test was performed, significance at $p<0.001$.

B Bar plot showing the total number of genes undergoing CLAMP-dependent Alternative splicing (AS) in females and males at 0-2 Hr pre-MZT and 2-4 Hr post-MZT embryonic stages. Spliced genes are divided into non-sex specific (grey) and sex-specific (orange) sub-category of genes.

C CLAMP dependent female and male sex-specifically Spliced (SSS) genes are divided into known and new sub-categories identified in 0-2 Hr pre-MZT and 2-4 Hr post-MZT embryos.

D Percentage of new female (red) and male (blue) CLAMP dependent sex specifically spliced genes in 0-2 Hr pre-MZT and 2-4 Hr post-MZT embryos.

E Male (blue) and females (brown) CLAMP-dependent sex-specific spliced genes compared with maternal genes (**N=841**, magenta) at 0-2 Hr pre-MZT (female, **N=153** and male, **N=99**) and 2-4 Hr post-MZT stages (female, **N= 270** and male, **N=116**).

F Gene ontology results for genes showing CLAMP-dependent female sex-specific splicing in the embryo at 0-2 Hr pre-MZT stage and for genes exhibiting CLAMP-dependent female as well as male sex-specific splicing in the embryo at 2-4 Hr post-MZT stage. The size of the circle increases as the number of the genes in that category increases. The color of the circle represents significance (p-value).

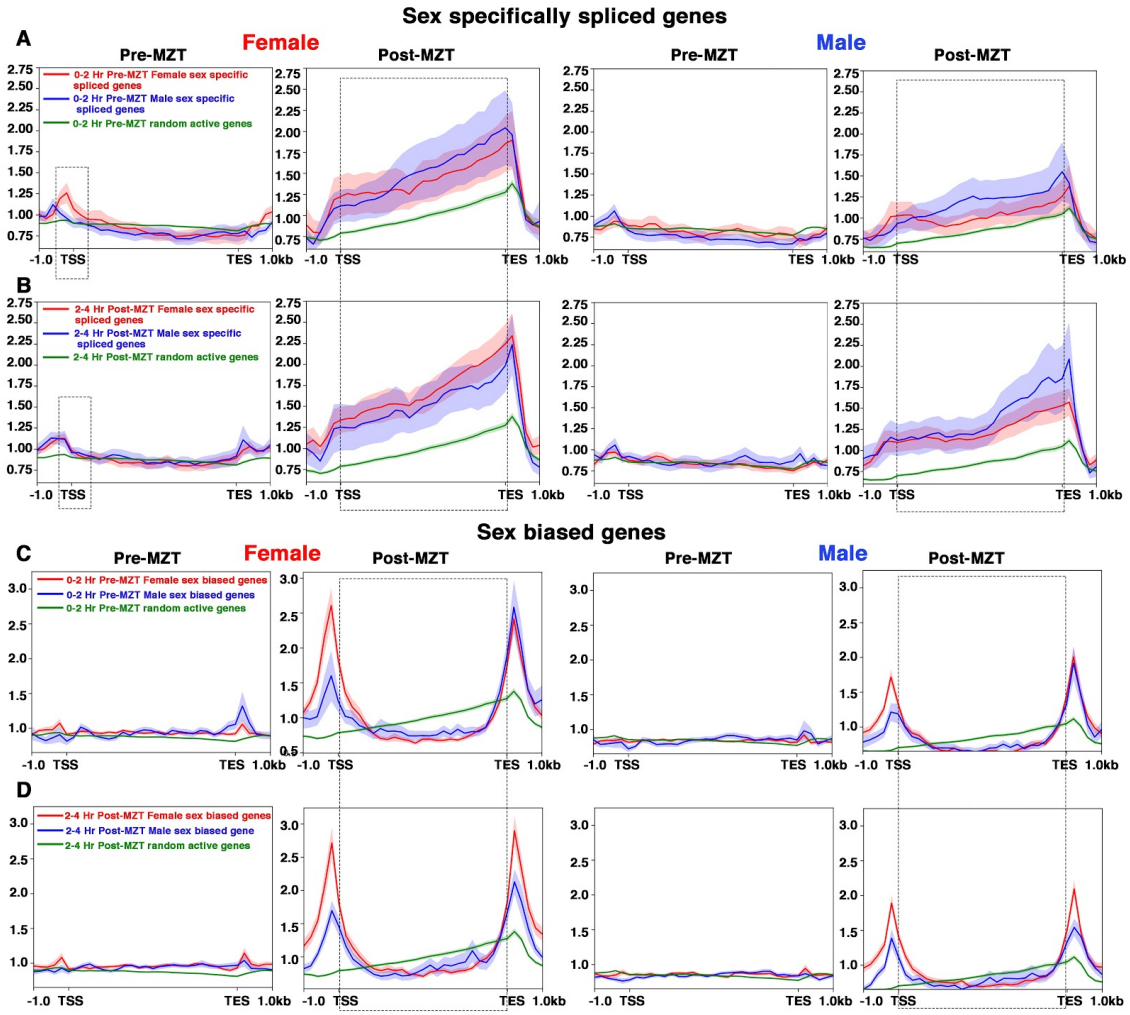
Sex specifically spliced genes
A
Pre-MZT
Female
Post-MZT
Pre-MZT
Male
Post-MZT
0-2 Hr Pre-MZT Female sex specific spliced genes
0-2 Hr Pre-MZT Male sex specific spliced genes
0-2 Hr Pre-MZT random active genes
B
2-4 Hr Post-MZT Female sex specific spliced genes
2-4 Hr Post-MZT Male sex specific spliced genes
2-4 Hr Post-MZT random active genes
-1.0 TSS
TES 1.0kb
Sex biased genes
C
0-2 Hr Pre-MZT Female sex biased genes
0-2 Hr Pre-MZT Male sex biased genes
0-2 Hr Pre-MZT random active genes
D
2-4 Hr Post-MZT Female sex biased genes
2-4 Hr Post-MZT Male sex biased gene
2-4 Hr Post-MZT random active genes

Fig 3.

Fig 3. Transcription factor CLAMP binds to introns along the gene body of female and male sex, specifically spliced genes at the post-MZT embryonic stage

A-B Average profiles for CLAMP binding at pre-MZT and post-MZT embryonic stages in females and males for genes spliced female-specifically (red line) and male-specifically (blue line) during the pre-MZT (**A**) and post-MZT (**B**) stages.

C-D Average profiles for CLAMP binding to genes expressed in a sex-biased manner in females (red line) and males (blue line) during pre-MZT (**C**) and post-MZT (**D**) stage.

Green lines in **A-D** represent CLAMP binding at a random set of active genes used as a control (see **Material and Methods** for details). Stippled regions in **A, B** (female, 0-2 hr pre-MZT) denote chromatin around TSS with more CLAMP binding in female spliced genes vs. male spliced genes. The dotted box in **A-D** marks the gene body region in sex-specifically spliced and biased expressed genes.

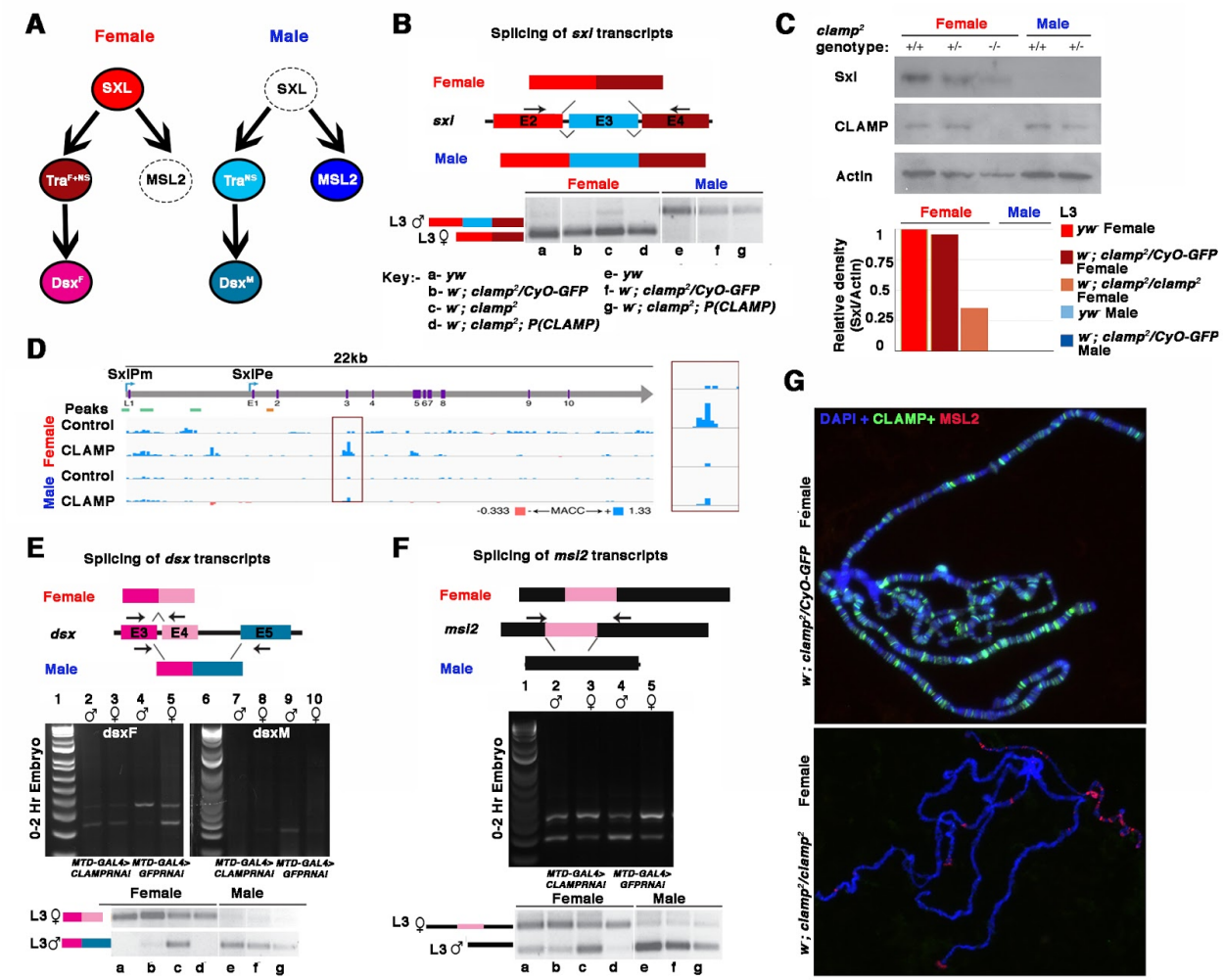
A
Female
Male
SXL
SXL
MSL2
MSL2
Dsx^F
Dsx^M
B
Splicing of sxl transcripts
Female
sxl
E2
E3
E4
Male
Female
Male
L3 ♂
L3 ♀
a b c d e f g
Key:- a- yw
b- w; clamp²/CyO-GFP
c- w; clamp²
d- w; clamp²; P(CLAMP)
e- yw
f- w; clamp²/CyO-GFP
g- w; clamp²; P(CLAMP)
C
clamp² genotype:
Female
Male
+/+ +/- -/- +/+ +/-
Sxl
CLAMP
Actin
Relative density (Sxl/Actin)
1
0.75
0.5
0.25
0
L3
yw Female
w; clamp²/CyO-GFP Female
w; clamp²/clamp² Female
yw Male
w; clamp²/CyO-GFP Male
D
22kb
SxlPm
SxlPe
Peaks
Female
Control
CLAMP
Male
Control
CLAMP
-0.333
MACC
1.33
E
Splicing of dsx transcripts
Female
dsx
E3
E4
E5
Male
1 2 3 4 5 6 7 8 9 10
♂ ♀ ♂ ♀ ♂ ♀ ♂ ♀
dsxF
dsxM
0-2 Hr Embryo
MTD-GAL4> CLAMPRNAI
MTD-GAL4> GFPRNAI
MTD-GAL4> CLAMPRNAI
MTD-GAL4> GFPRNAI
Female
Male
L3 ♀
L3 ♂
a b c d e f g
F
Splicing of msl2 transcripts
Female
msl2
Male
1 2 3 4 5
♂ ♀ ♂ ♀
0-2 Hr Embryo
MTD-GAL4> CLAMPRNAI
MTD-GAL4> GFPRNAI
Female
Male
L3 ♀
L3 ♂
a b c d e f g
G
DAPI + CLAMP+ MSL2
Female
w; clamp²/CyO-GFP
Female
w; clamp²/clamp²

Fig 4.

Fig 4. Alternative splicing of components of sex determination pathway is regulated by CLAMP

A The sex determination pathway in Drosophila regulated by master regulator SXL.

B Electrophoresis gel image (inverted) showing splicing of *sxl* transcripts in third instar larvae of females and males of genotypes mentioned in the key (**a-g**) with a representative schematic at the top of the gel image.

C Western blot showing the level of SXL and CLAMP protein in genotypes mentioned at the top of each lane. Actin levels indicate equal loading of protein. Below the blot is the relative density of SXL protein over Actin, with each genotype representing separate colored bars.

D Chromatin accessibility measured by the MNase Accessibility (MACC) score is shown across the sxl gene in male (S2) and female (Kc) cells under control and CLAMP RNAi conditions. The MACC score is a quantification measure of nucleosome accessibility measuring both each genomic locus and its accessibility. Positive/high accessibility values (blue) indicate high MNase, and negative/low (red) accessibility values indicate low MNase. Each window covers MACC values in the range -0.333 - 1.33. MACC values increase in females after CLAMP RNAi, specifically at exon 3 (red box), and can be seen inset to the right. Green boxes represent CLAMP binding peaks in the *sxl* gene just below the schematic for the *sxl* gene itself.

E-F Electrophoresis gel image at 0-2 Hr embryonic (lane 2-5 & 7-10) and third instar larval stages (a-g) showing splicing of *dsx* (E) and *msl2* (F) transcripts in female (lane 3,5,8,10, a-d) and male (lane 2,4,7,9, e-g). Embryos were from *MTD-GAL4>GFPRNAi* control (lane 4,5,9,10) and *MTD-GAL4>CLAMPRNAi* (lane 2,3,7,8) females. a-g is the same as in B. Schematic at the top of each gel image shows female and male splice variants of *dsx* (E) and *msl2* (F) transcripts. G Fluorescent microscope images of polytene chromosomes from third instar salivary gland in genotypes mentioned to the left of each panel showing CLAMP (green) and MSL2 (red) distribution on chromatin (blue, DAPI).

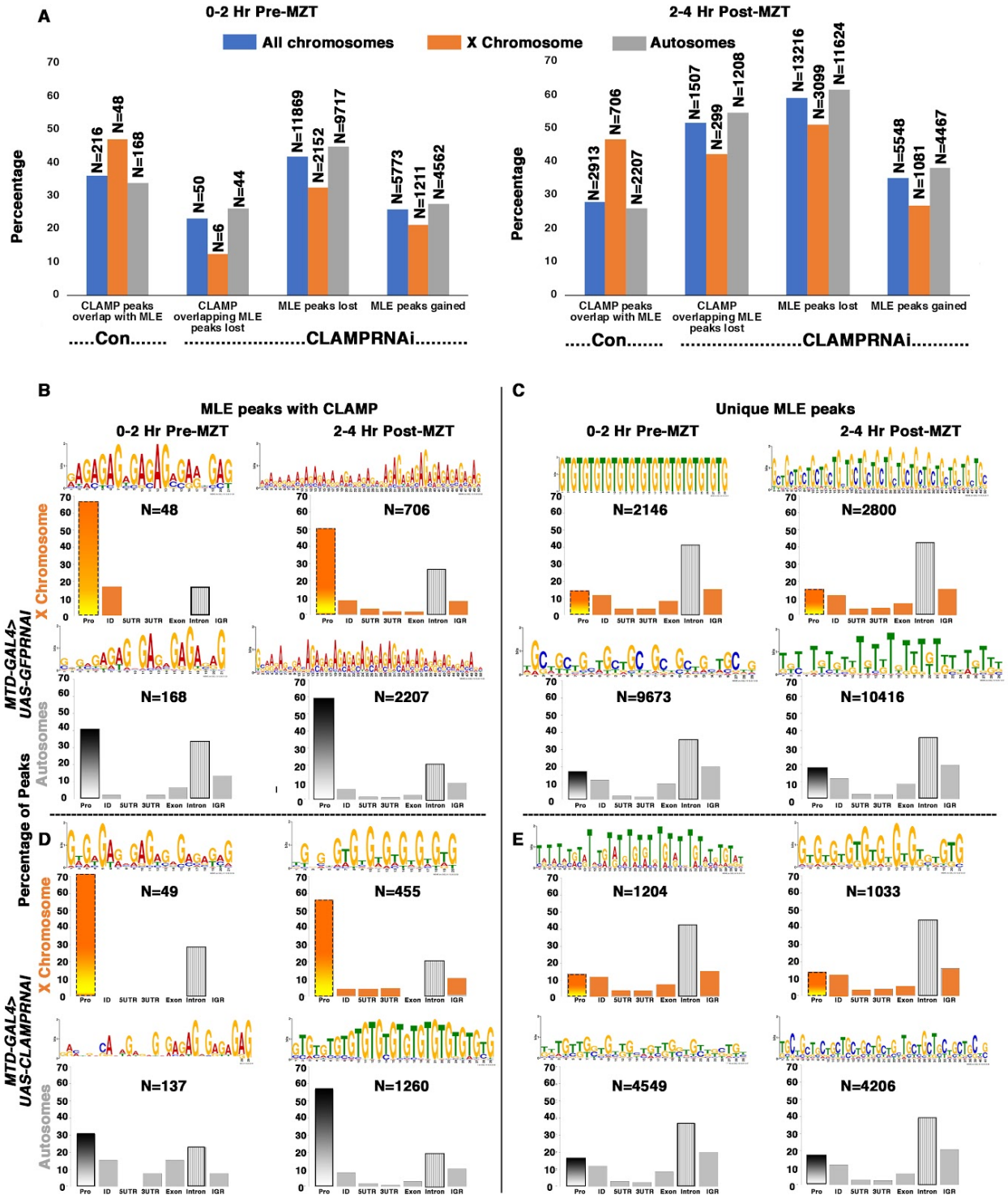
A
0-2 Hr Pre-MZT
2-4 Hr Post-MZT
All chromosomes
X Chromosome
Autosomes
Percentage
N=216
N=48
N=168
N=50
N=6
N=44
N=11869
N=2152
N=9717
N=5773
N=1211
N=4562
N=2913
N=706
N=2207
N=1507
N=299
N=1208
N=13216
N=3099
N=11624
N=5548
N=1081
N=4467
CLAMP peaks overlap with MLE
CLAMP overlapping MLE peaks lost
MLE peaks lost
MLE peaks gained
.....Con.....
.....CLAMPRNAi.....
B
MLE peaks with CLAMP
C
Unique MLE peaks
MTD-GAL4> UAS-GFPRNAI
MTD-GAL4> UAS-CLAMPRNAI
X Chromosome
Autosomes
Percentage of Peaks
N=48
N=706
N=168
N=2207
N=2146
N=2800
N=9673
N=10416
D
E
N=49
N=455
N=137
N=1260
N=1204
N=1033
N=4549
N=4206
Pro
ID
5UTR
3UTR
Exon
Intron
IGR

Fig 5.

Fig 5. CLAMP regulates the distribution of MLE on chromatin

A Bar plots showing the percentage of MLE peak distribution over all chromosomes (blue), on the X chromosome (orange), and autosomes (grey) in male 0-2 Hr pre-MZT and 2-4 Hr post-MZT embryos after the loss of maternal CLAMP. 'N' at the top of each bar indicates the number of peaks in each category.

B-E Distribution of MLE peak percent overlap with CLAMP peaks (B, D) and unique MLE peaks (**C, E**) over different gene regions- 1) Pro=Promoter, 2) ID= Immediate downstream, 3) 5UTR=5' untranslated region, 4) 3UTR= 3' untranslated region, 5) Exon, 6) Intron and 7) IGR=Intergenic region on the X chromosome and autosomes in male embryos of 0-2 hr pre-MZT and 2-4 Hr post-MZT stages under normal conditions (B, C) and after the loss of maternal CLAMP (D, E). 'N' is the total number of peaks in each category. Most probable motif according to MEME for MLE peaks overlapping with CLAMP (B, D) and unique MLE peaks (C, E) in each category shown at the top of the relevant bar plot.

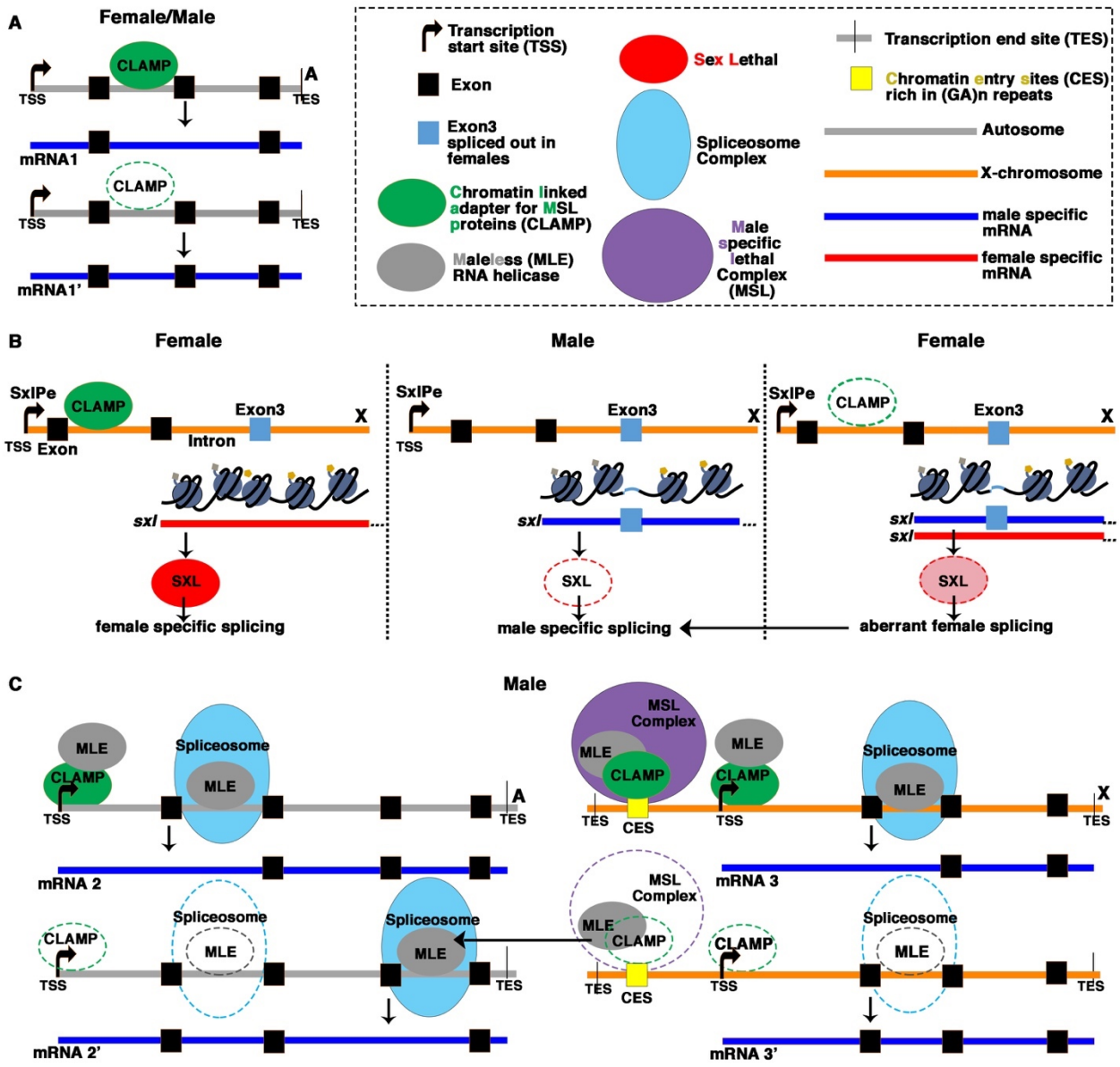
A
Female/Male
CLAMP
TSS
A
TES
mRNA1
CLAMP
TSS
TES
mRNA1'
Transcription start site (TSS)
Exon
Exon3 spliced out in females
Chromatin linked adapter for MSL proteins (CLAMP)
Maleless (MLE) RNA helicase
Sex Lethal
Spliceosome Complex
Male specific lethal Complex (MSL)
Transcription end site (TES)
Chromatin entry sites (CES) rich in (GA)n repeats
Autosome
X-chromosome
male specific mRNA
female specific mRNA
B
Female
Male
Female
SxlPe
CLAMP
Exon3
X
TSS
Exon
Intron
sxl
SXL
female specific splicing
male specific splicing
aberrant female splicing
C
Male
MLE
CLAMP
Spliceosome
TSS
TES
A
X
mRNA 2
mRNA 2'
MSL Complex
CES
mRNA 3
mRNA 3'

Fig 6.

Fig 6. Mechanisms by which CLAMP regulates sex-specific splicing in females and males

A CLAMP regulates splicing in both males and females via directly binding to intronic sequences of CLAMP-dependent sex-specifically spliced genes.

B In females, binding of CLAMP near the SxlPe promoter regulates chromatin at exon three (blue square) of the *sxl* gene, such that it is spliced out and functional Sxl protein is formed, which drives female-specific splicing events. In males, the absence of CLAMP near the SxlPe promoter changes chromatin structure at exon three, which is retained in the *sxl* transcript, resulting in the lack of functional Sxl protein. The absence of CLAMP in females thus results in non-functional male-specific *sxl* transcripts, which retain exon three, reducing levels of function Sxl protein.

C CLAMP regulates MLE distribution between the spliceosome and the X-chromosome specific MSL complex in males. CLAMP increases the occupancy of MLE at promoter and CES sequences. In the absence of CLAMP, MLE is lost from many sites, including CES and promoters, and is gained at new intronic sequences, resulting in aberrant sex-specific splicing in males. The three mechanisms proposed are not mutually exclusive.