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

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

[*] Corresponding Author
[Ω] contributed equally
[#] 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 factors play a crucial role in initiating zygotic transcription. However, the mechanisms by which maternal factors regulate early zygotic transcript diversity through alternative splicing remain unclear. Furthermore, how early in development widespread sex-specific transcript diversity occurs is not known. We show that widespread sex-specific transcript diversity occurs much earlier than previously thought and present a new pipeline called time2splice to quantify splicing changes over time. Using the powerful *Drosophila* model, we define several mechanisms by which a maternal factor regulates sex-specific zygotic transcriptome diversity: 1) GA-binding pioneer factor CLAMP (**C**hromatin **l**inked **a**dapter for **M**SL **p**roteins) binds to the gene bodies of sex-specifically spliced genes to alter the recruitment of spliceosome components; 2) CLAMP modulates chromatin accessibility of an alternatively-spliced exon within *sxl,* the master regulator of sex determination; 3) CLAMP regulates the distribution of the spliceosome component **M**ale**le**ss (MLE) to prevent aberrant sex-specific splicing. Overall, we provide key insight into how maternal factors influence sex-specific transcript diversity.

## 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. **A**lternative **S**plicing (**AS**), a mechanism of selective inclusion or exclusion of introns and exons drives widespread transcript diversity[1, 2]. 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 differences in progression and severity[3, 4, 5, 6]. 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 transcript diversity drives biological processes lies in the events that shape the initial few hours of an organism's existence.

During early development, protein and RNA deposited by the mother into the embryo shape early embryonic milestones across metazoans[7, 8]. Initially, cell number increases, 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 remains poorly understood. Moreover, maternal factors are often essential regulators that 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 [9].

During *Drosophila* embryogenesis, **Z**ygotic **G**enome **A**ctivation (**ZGA**) occurs shortly after the first two hours of development. Concurrently, maternal transcripts gradually decrease in abundance, and zygotic transcription increases, a process called the **M**aternal to **Z**ygotic **T**ransition (**MZT**). ZGA starts approximately 80 min after egg laying and most maternal transcripts are degraded by 180 min after egg laying[10]. Even at these early stages of development, AS generates multiple variants of the same gene transcripts resulting in transcript diversity. Although the earliest genes transcribed from the zygotic genome are mainly intron-less, almost 30% of early zygotic transcripts do have introns[11, 12]. Furthermore, genes involved in sex determination use AS to drive male versus female-specific development [13]. 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[14, 15]. Therefore, functional Sxl protein is made only in the female embryo. Early expression of the Sxl protein made from a transcript driven by $\mathrm{Sxl_{Pe}}$, an early *sxl* promoter [16], autoregulates splicing of *sxl*. Next, the Sxl protein causes splicing of exon three from pre-mRNA of *sxl* transcripts initiated from the late maintenance promoter ($\mathrm{Sxl_{Pm}}$) specifically in females [17, 18]. The early activation of the $\mathrm{Sxl_{Pe}}$ promoter is regulated by a group of X-linked genes called X signal elements (XSE). A higher dosage of XSE gene products drives $\mathrm{Sxl_{Pe}}$ expression in XX females and not XY males [19].

The Sxl protein is an RNA binding protein that binds to *sxl* pre-mRNA and the 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[20]. 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 [14, 15].

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 [21, 22]. The binding of the Sxl protein to the *msl-2* transcript specifically in females results in skipping of the first exon, thereby hindering translation of the protein and preventing MSL complex formation in females[23].

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**) that are enriched for clustered long GA-repeat sequences [24, 25]. The maternally deposited **C**hromatin **l**inked **a**dapter for **M**SL **p**roteins (CLAMP) opens chromatin [26, 27] and directly contacts MSL-2[28] to promote MSL complex recruitment. Subsequently, the MSL complex increases transcript levels on the male X-chromosome by promoting transcription elongation[29]. Interestingly, the MSL complex shares a key highly conserved component MLE (RNA helicase A) with the spliceosome complex [30, 31, 32], suggesting that there is likely to be an interplay between dosage compensation and splicing that has not yet been defined.

CLAMP is an essential pioneer **T**ranscription **F**actor (**TF**) that opens chromatin [26, 27]. Loss of CLAMP results in complete loss of MSL complex from the male X-chromosome causing male lethality [33]. 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 [27]. Furthermore, CLAMP is bound to both intronic and promoter regions on chromatin[34]. These intronic regions are rich in polypyrimidine tracts, which have similar GA-rich sequences to CES and contain CLAMP binding sites[35]. Interestingly, GA-rich repeat sequences within CES are thought to have evolved from polypyrimidine tracks that regulate splicing[35]. Importantly, MALDI-mass spectrometry data identifying putative CLAMP interactors identified association with 33 RNA binding proteins, including 6 that regulate alternative splicing[36]. 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 alternative splicing globally in early *Drosophila melanogaster* sexed 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 have been identified for several candidate genes in early development[37, 38, 39, 40], a comprehensive analysis of all sex-specific isoforms forms has not yet been performed in early embryos and is only available in fully developed tissues like ovaries, testis and brain[38, 41]. Therefore, we used a meiotic drive system to sex embryos and measured total AS and sex-specific splicing genome-wide before (0-2 Hrs) and after (2-4 Hrs) the major wave of MZT. For ease of understanding, we have defined 0-2 Hr embryos as pre-MZT and 2-4 Hr embryos as post-MZT even though a small number of early zygotic genes are activated during the 0-2 Hr time period.

To reveal new mechanisms by which maternal factors shape early embryonic transcriptome diversity, we have focused on analyzing how the maternal transcription factor CLAMP regulates sex-specific alternative splicing during the first few hours of development. We have identified male-specific and female-specific genes whose splicing requires maternal CLAMP and are important regulators of early development with diverse functions form including chromatin regulation and splicing. We found that maternal CLAMP most frequently influences splicing during the 0-2 Hr embryonic time period in both males and females.

Moreover, we defined multiple mechanisms by which CLAMP drives sex-specific transcript diversity during early embryonic development: 1) In both males and females, CLAMP is specifically enriched at gene bodies of approximately half of the sex-specifically spliced genes that it regulates, supporting a direct role for CLAMP in regulating splicing through altering recruitment of spliceosome components with which it associates; 2) In females, CLAMP also regulates the *sxl* pathway through chromatin accessibility, splicing, and protein levels which is likely to alter the splicing of several downstream targets; 3) In males, our data suggest that 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 via diverse mechanisms including direct binding to sex-specifically spliced genes.

## 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-sequencing data generated from sexed embryos at two-time points: 1) 0-2 hours (pre-MZT); 2) 2-4 hours (post-MZT)[9] (#GSE102922). Embryos were sexed using a meiotic drive system that produces sperm with either only X or only Y chromosomes[9], resulting in progeny of either only female or only male genotypes. Next, we measured the amount of AS in early *Drosophila* development using a pipeline that we developed and made publicly available called time2Splice (https://github.com/ashleymaeconard/time2splice). Time2Splice implements the commonly used SUPPA2 algorithm to identify splice variants and provides additional modules to integrate time, sex, and chromatin localization data (See Supplemental methods) (Fig S1). SUPPA2 measures the percent spliced in (PSI) for each exon to calculate the differential alternative splicing between samples, represented as ΔPSI[42].

AS events can be classified into seven categories: 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 belong to each of the seven different categories at the 0-2 Hr and 2-4 Hr 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 type of alternative splicing, constituting almost one-fourth of total AS. In contrast, **A**lternative **L**ast Exon (AL) is the least common alternative splice type. The AS transcript distribution across categories was similar between the two time points and sexes (**Fig 1B**).

We hypothesized that CLAMP regulates AS in early embryos for the following reasons: 1) CLAMP it is a maternally deposited pioneer factor enriched at intronic regions ; 2) Proteomic data identified a physical association between spliceosome components and CLAMP[36]; 3) CLAMP binding sites evolved from polypyrimidine tracts that regulate splicing[35]. We tested our hypothesis in early staged and sexed embryos by measuring differences in splicing in RNA-sequencing data generated from male and female 0-2 Hr and 2-4 Hr (pre- and post-MZT) embryos with and without maternal CLAMP[9]. The maternal triple driver GAL4, *MTD-GAL4* was used to drive *UAS-CLAMPRNAi[val22]* that strongly reduce maternal CLAMP levels, validated by qPCR and Western blot [9].

First, we asked whether CLAMP alters AS and we found 200-400 transcripts where AS is regulated by CLAMP depending on the time point and sex (**Fig 2A**). Next, we asked which type of alternative splicing is most affected by depleting CLAMP. The overall distribution of AS transcripts into the seven splicing types remains mostly unaffected in the absence of maternal CLAMP. However, at the 0-2 Hr (pre-MZT) stage, loss of maternal CLAMP results 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 2-4 Hr (post-MZT) stage, only male embryos have a significant percentage of MXE splicing affected in the absence of maternal CLAMP (p-value < 1.95e-137 chi-squared test) (**Fig 1B**). Therefore, CLAMP alters AS and has a stronger effect on MXE splicing than other types.

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**). Interestingly, MXE alternative splicing occurs in many transcripts that encode components of the sex determination pathway[43]. Sex determination pathway components undergo MXE alternative splicing that influences the splicing of downstream genes to drive transcriptome variation and sexual dimorphism. Furthermore, CLAMP has a sex-specific role in dosage compensation[21, 44]. 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 (**Fig S2**) (Details described in Supplementary methods **S1**). We measured alternative splicing using an exon-centric approach to quantify individual splice junctions by measuring **p**ercent **s**pliced **in** (**PSI**) for a particular exon using the established SUPPA algorithm within the time2splice pipeline[42]. Exon inclusion is represented as positive PSI, and exclusion events are defined as negative PSI (equation in **Supplementary Methods S1**). First, we identified sex-specific splicing events (**Fig S2A, D**) and CLAMP-dependent splicing events (**Fig S2B-C, E-F**) in female and male 0-2 Hr and 2-4 Hr embryos.

Next, 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 AS is CLAMP-dependent, ~30-60% of sex-specific splicing is CLAMP-dependent (**Fig 2A**). Therefore, CLAMP regulates sex-specific splicing more strongly than total AS. Overall, more genes show CLAMP-dependent splicing in females (~150-250) than males (~100) (**Fig 2B, C and Supplementary Table S1**). However, the magnitude of the effect of CLAMP on splicing was stronger in males than in females (**Fig S3**). We found that before MZT, the CLAMP-dependent known male sex-specifically spliced genes have more exon inclusion events than their female counterparts since ΔPSI values for males is significantly higher than in females (**Fig S3**). This was further supported by validating a randomly selected set of 12 sex-specific splicing events we identified as described in more detail below (**Fig S4&5; Table S2**).

We then divided all CLAMP-dependent alternatively spliced events into two categories: 1) sex specifically spliced (SSS) events; and 2) non-sex specifically spliced (non-SSS) events (**Fig 2B**). We quantified CLAMP-dependent sex-specifically spliced events by comparing the change in PSI (ΔPSI) for two classes of transcripts which we defined as follows: 1) **known** sex-specifically spliced isoforms are different from each other in control samples ($p<0.05$) (**Fig. S2A, D**); 2) **new** sex-specifically spliced isoforms are those not present in control samples (**Fig 2C and S3**). We identified widespread 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 (~70%) (**Fig 2C, D**). 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 S3**). In the absence of CLAMP new aberrant sex-specific spliced isoforms are generated. Therefore, we hypothesized that CLAMP normally inhibits aberrant sex-specific splicing events.

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 from previous studies [45, 46]. 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 [27].

Next, we divided our CLAMP-dependent sex-specifically spliced genes into female-specific or male-specific categories. Here, it is important to note that many of these genes have multiple AS events, and thus multiple isoforms regulated by CLAMP (**Table S1**). 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 regulatory factors such as spliceosome components and TFs (**Fig 2F**). Therefore, in females CLAMP alters the splicing of genes that can regulate the transcription and splicing of other genes to amplify its regulatory role. In contrast, the male specifically-spliced genes pre-MZT are not enriched for any specific biological function or process likely due to the small number of genes in the gene list. At the post-MZT stage in both sexes, CLAMP regulates the splicing of genes that drive development including organogenesis, morphogenesis, cell proliferation, signaling, and neurogenesis (**Fig 2F**).

We also selected 12 random genes for validation that we identified computationally as having CLAMP-dependent splicing events of diverse types such as alternative 5' splice site (A5) or alternative first exon (AF) (**Fig S4&5**) using qRT-PCR or RT-PCR. These genes perform diverse biological functions and provide new insights into how CLAMP regulates early development (**Fig. 2F**). We have summarized the results and functions of the validated target genes at which splicing is regulated by CLAMP (**Table S2**). Several of the target genes have functional links to CLAMP including *iab4*, *psq* and *pep* suggesting that we have identified relevant target genes[36, 47, 48]. Furthermore, many genes that are sex-specifically spliced by CLAMP are themselves involved in splicing and chromatin regulation (**Table S2**).

### 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-specific splicing? If CLAMP directly regulates sex-specific splicing, we hypothesized that it would directly bind to the genes that it regulates. Therefore, we defined the binding pattern of CLAMP at the CLAMP-dependent female and male sex-specifically spliced genes in sexed embryos using CLAMP ChIP-seq data (#GSE133637).

We generated average profiles for CLAMP occupancy at genes showing CLAMP-dependent splicing in females and males at pre- and post-MZT time points in males and females (**Fig 3A, B**). We found that CLAMP occupies the gene bodies of many sex-specifically spliced genes that require CLAMP for their splicing. Specifically, 43.8% percentage of all CLAMP-dependent sex-specifically spliced genes are bound by CLAMP: 21.9% in 0-2 Hr female embryos, 8.2% in 0-2 Hr male embryos, 65.2% in 2-4 hr female embryos and 59.43% in 2-4 hr male embryos (**Table S3**). The increase in percentage of genes bound in 2-4 hr embryos compared with 0-2 hr embryos is consistent with an increased number of CLAMP binding sites at the later time point[48]. Overall, these data are consistent with a direct role for CLAMP in regulating splicing of sex-specifically spliced genes.

Next, we compared the average CLAMP binding pattern at sex-specifically spliced genes to the CLAMP binding pattern at genes whose expression but not splicing is both sex-biased and dependent on CLAMP. We defined CLAMP-dependent sex-biased genes as differentially expressed genes in males in females that are dependent on CLAMP. In contrast to sex-specifically spliced genes, CLAMP occupancy shows a very different pattern at 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 is also modestly enriched at the TSS of female-biased expressed genes in females, consistent with enhanced CLAMP occupancy at the TSS of expressed genes [33]. As a control, we used a random set of active genes that are not regulated by CLAMP (green lines in **Fig 3A-D**) and we observed lower occupancy than at target genes. Overall, we found preferential binding of CLAMP along the gene bodies of genes that have CLAMP-dependent splicing in both females and males.

To determine whether the binding of CLAMP to gene bodies occurs close to splice junctions, we measured the distance of CLAMP peaks from the nearest splice junction (**Fig S6**). We found that CLAMP peaks are most frequently within 200-400bp of either the start or the end of a splice junction, especially in sex-specifically spliced genes. We also found that CLAMP binds to chromatin closer to splice junctions at sex-specifically spliced genes compared to sex-biased genes in 2-4 hr female embryo samples that have the most target genes and CLAMP binding events. The results were similar for all CLAMP peaks (**Fig S6C**) compared to peaks only present in introns (**Fig S6G**). We did not observe a difference in proximity of CLAMP peaks to the nearest splice junction between sex-specifically spliced genes and sex-biased genes in other samples but this negative result is not conclusive due to the decreased number of peaks and target genes in the other samples. Together these data support a direct role for CLAMP in co-transcriptional RNA processing that we hypothesize is due to altering the recruitment of spliceosome components.

## 4. CLAMP regulates chromatin accessibility and splicing of the *sxl* gene, and binds to and alters splicing of other sex determination pathway component genes

Because Sex-lethal (Sxl) is the master regulator of sex determination[15] and drives subsequent sex-specific splicing in females[49], 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 [14]. Therefore, the Sxl protein is non-functional in males and becomes degraded [15, 50]. In females, Sxl protein inhibits the splicing of the *msl-2* transcript which prevents aberrant MSL-2 expression and MSL complex formation in females (**Fig 4A**). 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* transcript is not translated into functional Transformer protein in males. Therefore, in the absence of any functional male-specific Transformer protein, the *doublesex* transcript undergoes male-specific splicing (**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 [51, 52].

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 [44], the heterozygous mutant *clamp*$^2$*/CyO-GFP* line, and our previously reported 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 [53, 54, 55]. Therefore, we designed an RT-PCR assay to distinguish between the female-specific (excluding exon 3) and male-specific (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 performed western blots 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 mined our previously generated Micrococcal Nuclease sequencing data [26], which measures chromatin accessibility in S2(male) and Kc(female) cells in the presence and absence of CLAMP. Regions of the genome that are accessible have a positive MACC score and are shown in blue and regions of the genome that are inaccessible have a negative score and are shown in red (range is between -0.33(red) and +1.33 (blue). We found that after the loss of CLAMP in female Kc cells, chromatin accessibility around exon 3 of *sxl* increases significantly (**Fig 4D**). As shown in the Fig 4D boxed rectangular inset, *sxl* exon3 shows a strong and statistically significant blue peak[26] indicative of open chromatin specifically in females in which CLAMP has been depleted compared to control females. Therefore, CLAMP normally promotes a closed chromatin environment at exon 3 in females but not males.

Our results suggest that increased chromatin accessibility in males compared to females results in retention of exon3 in the male *sxl* transcript. Consistent with our results, recent reports provide strong evidence that increased chromatin accessibility contributes substantially to the retention of introns during AS[56]. In addition, splicing-associated chromatin signatures have recently been identified[57]. Consistent with our results in males, open chromatin marks such as H3K4me1& H3K4me2 are enriched just upstream of the start site of retained exons[57]. In contrast, histone marks associated with condensed chromatin such as H4K20me1&2, H3K9me3, and H3K27me3 are highly enriched at excluded exons[57], consistent with our results 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 for which we were able to design effective RT-PCR primers (**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). Therefore, *dsx* splicing is regulated by CLAMP and CLAMP binds to the *dsx* gene (**Fig S7**). However, *dsx* splicing is also regulated by Sxl whose protein levels are regulated by CLAMP (**Fig 4C**). Therefore, it is possible that CLAMP directly regulates *dsx* splicing, indirectly regulates it through Sxl or both mechanisms could be occurring simultaneously.

Furthermore, we 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) levels of the functional male-specific *msl-2* isoform increased in females (**Fig 4F**, lane c). To determine whether these splicing defects also cause dysregulation of MSL-2 protein expression, we performed polytene immunostaining from female *clamp*$^2$ mutant salivary glands. In the absence of CLAMP, ectopic MSL2 protein (in red) is present at several locations on female chromatin compared with controls (*clamp*$^2$*/CyOGFP* heterozygous females) where MSL-2 protein is not present on chromatin (**Fig 4G**). Similar to *dsx*, *msl-2* is also bound by CLAMP (Fig S7) and regulated by Sxl and therefore could be regulated through both direct and indirect mechanisms. Together, these data reveal that CLAMP regulates the splicing and protein expression of multiple components of the sex determination pathway.

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

CLAMP occupancy on the gene bodies of hundreds of sex-specifically spliced genes (**Fig 3**) and its physical association with spliceosome components from mass spectrometry analysis[36] 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 [30, 32, 58, 59, 60] 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 [28, 61, 62]; 2) We hypothesized that CLAMP could regulate the distribution of MLE between the spliceosome and MSL complex to regulate MLE targeting and alter sex-specific AS in males.

To determine whether CLAMP modulates MLE distribution on chromatin, we measured MLE distribution at the genomic level using Cut-and-Run[63, 64]. 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. Next, we identified MLE peaks at the 0-2 Hr and 2-4 Hr embryonic stages in males in the presence or absence of maternally deposited CLAMP.

We previously identified CLAMP peaks at the same time points[21] and therefore we first compared the distribution and location of MLE and CLAMP peaks in control embryos. We classified MLE peaks into two groups: 1) MLE peaks overlapping with CLAMP peaks (**Fig 5A**) and 2) unique MLE peaks which do not overlap with CLAMP peaks (**Fig 5B**) at both 0-2 Hr pre-MZT and 2-4 Hr post-MZT stages. Independent of stage, MLE peaks which overlap with CLAMP peaks are largely at promoters with fewer peaks localizing to introns. In contrast, unique MLE peaks that do not overlap with CLAMP peaks are primarily localized to introns (**Fig 5A, B**). This distribution is similar for X-chromosomal and autosomal peaks.

The loss of CLAMP results in a considerable loss and redistribution of both overlapping and non-overlapping MLE peaks. We found that overall ~60% of MLE peaks were lost at 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 S8A-C**). Also, almost 50% of CLAMP peaks on the X-chromosome overlap with MLE peaks (**Fig 5A** and **Fig S8B**), consistent with their association with MSL complex. After the loss of maternal CLAMP, ~23% of these MLE peaks overlapping with CLAMP are also lost at the pre-MZT stage, which increases to ~51% at the post-MZT stage (**Fig S8A-C**). 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**). Furthermore, we hypothesize 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[33].

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 the time2splice pipeline. The known CLAMP motif [33], a stretch of (GA)n repeats, is enriched at regions that are bound by both MLE and CLAMP independent of stage and chromosome type as expected. 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 5A, B**). In the absence of CLAMP, the remaining MLE peaks were most enriched for (GT)n motifs (**Fig 5C, D**) which have known roles in splicing through forming secondary RNA structures[65, 66, 67]. Therefore, CLAMP prevents MLE from redistributing to sequence motifs that are known regulators of splicing.

We also found that CLAMP alters the distribution of MLE from introns to promoters. MLE peaks that **overlap** with CLAMP on the X-chromosome (orange) or autosomes (grey) are enriched at promoters (Fig 5C, **Fig S8E** (X-chromosome), **Fig S8F** (Autosomes)). In contrast, **new** unique

MLE peaks that are gained after CLAMP RNAi on the X-chromosome (orange) or autosomes (grey) are enriched at introns (Fig 5D, **Fig S8E** (X-chromosome), **Fig S8F** (Autosomes). These results support a role for CLAMP in sequestering MLE at (GA)n rich sequences within promoters that prevents it from binding to GT motifs within introns that are known regulators of splicing[65, 66, 67]. Thus, in the absence of CLAMP, MLE is redistributed and aberrantly binds to intronic sequences with known motifs that regulate splicing irrespective of whether present on X-chromosome or autosomes.

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 in the presence and absence of CLAMP (**Fig S8G, H**). Pre-MZT, MLE binds near the TSS of male-specifically spliced genes independent of maternal CLAMP (**Fig S8G**). At the post-MZT stage, loss of maternal CLAMP in male embryos causes MLE to change its binding distribution along the gene body (rectangle with dotted lines: **Fig S8H**) of CLAMP-dependent male-specifically spliced genes (blue line) to CLAMP-dependent female-specifically spliced genes (red line). These profiles are consistent with a model in which CLAMP is essential for MLE distribution at male-specifically spliced genes which regulates male sex-specific splicing. In males, more MLE binding at female sex-specifically spliced genes (red line, enclosed within rectangle with dotted lines: **Fig S8H**) in absence of CLAMP may result in differential splicing of these genes. Thus, we hypothesis that under normal condition in males, CLAMP inhibits mis-localization of MLE to female sex-specifically spliced genes.

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 based on mass spectrometry analysis[36]. To validate these associations, we performed co-immunoprecipitation (coIP) experiments to assess association between CLAMP and spliceosome components with known functions in sex-specific splicing, the conserved hrb27C and Squid proteins [60, 68, 69]. We found that in both S2 (male) and Kc (female) cells, CLAMP interacts with hrb27C (**Fig S9A, B**). In contrast, CLAMP only associates with Squid in female Kc cells and not in male S2 cells (**Fig S9A, B**), consistent with mass spectrometry data. In contrast to MLE and CLAMP which are enriched on the male X-chromosome, Squid occupancy on polytene chromosomes is decreased on the male X chromosome compared with the female X chromosome (**Fig S9C-E**). Therefore, it is possible that there is a competition between CLAMP recruitment of MSL to the male X chromosome and CLAMP recruitment of the spliceosome complex that contains Squid that contributes to sex-specific splicing.

Overall, CLAMP differentially associates with spliceosome components in males and females, providing a potential mechanism by which CLAMP can regulate sex-specific splicing. Specifically, 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 in females and MLE in males confers sex-specific recruitment of the spliceosome, thus generating transcriptome diversity between female and male animals.

## Discussion:

Alternative splicing (AS) is a highly conserved mechanism that generates transcript and protein diversity [3, 70, 71]. 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 [1, 2, 39, 41, 46]. Furthermore, different isoforms are present in maternal and zygotic transcripts [45, 46]. However, the mechanisms that regulate transcript diversity 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 sex-specific 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 [37, 38, 39, 41, 68, 72, 73, 74, 75] either examined sex-biased differences in gene expression or sex-specific transcript diversity much later in development in adult gonads or brain. 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 [9] and measured sex specific transcript diversity caused by alternative splicing.

Splice variants occur at different frequencies over time and between sexes. To date, we lack pipelines 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 splicing analysis based on the the SUPPA algorithm; 2) temporal protein-DNA analysis, and 3) temporal multi-omics integration. The pipeline 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 transcript variation that was not previously identified (**Fig1, 2**). Because the RNA transcript variants in both males and females encode genes involved in developmental processes, sex-specific developmental distinctions 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 transcript diversity provides insight into how developmental disorders 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 both transcription initiation and shapes RNA processing. 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. 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 the sex-specific splicing cascade very early in development to drive sex determination [14, 20, 54]. Decreased Sxl protein levels in female *clamp*$^2$

mutants 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 [15]. Also, CLAMP binds to the *sxl, dsx,* and *msl-2* target genes. Therefore, CLAMP may regulate splicing through at least two possible mechanisms that are not mutually exclusive: 1) CLAMP directly regulates the splicing of both genes by altering the recruitment of the spliceosome; 2) CLAMP regulates *msl-2* and *dsx* splicing indirectly by regulating Sxl which is a known regulator of their splicing (**Fig 4**).

CLAMP binds directly to intronic regions of approximately half of the sex-specifically spliced genes that it regulates in both males and females and associates with spliceosome components. Therefore, our data supports a model in which direct CLAMP binding regulates splicing of its target genes. Furthermore, CLAMP regulates chromatin as a pioneer TF[26, 48] and recent literature links chromatin and splicing[56, 57]. For example, closed chromatin marks have recently been linked to exon exclusion and open chromatin has been linked to exon inclusion[56, 57]. Thus, we hypothesize that CLAMP may recruit spliceosome complex components to regulate their splicing by altering the chromatin environment similar to the regulation that occurs at the *sxl* locus (**Fig 6A**).

In females but 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**). Consistent with recent literature[56, 57], we hypothesize that closed chromatin at exon3 induces exclusion of this exon from female *sxl* transcripts whereas more open chromatin in males result in its inclusion in the male *sxl* transcript. In the absence of CLAMP in females, the chromatin becomes more open, and therefore exon 3 is included in *sxl* transcripts which prevents translation due to the incorporation of a stop codon. In the absence of CLAMP, exon 3 is retained in females similar to wild type males, resulting in the formation of untranslatable *sxl* transcript in females, significantly down-regulating levels of Sxl protein, thus 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 consistent with literature. CLAMP can mediate long-range chromatin interactions [76, 77] and act on chromatin accessibility at a distance [26].

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 [36] and coIPs (**Fig S9A,B**) 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[32] physically interacts with CLAMP only in males. Because Squid is already known to regulate sex-specific splicing[68], these data support a model in which differential association between CLAMP and spliceosome complex components in males and females regulates sex-specific splicing.

Furthermore, we show that CLAMP inhibits aberrant splicing events in males, especially at the post-MZT stage (**Fig 2**). In males, we found that MLE colocalizes 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 that contain GT sequence motifs that are known to regulate splicing[65, 66, 67]. 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, another function of CLAMP is to recruit the MSL ribonucleoprotein complex to the X chromosome to regulate dosage compensation. CLAMP physically interacts directly with MSL-2[28] and associates with MLE, an RNA helicase that acts on the *roX* non-coding RNA components of MSL complex[31, 44, 58, 61, 78]. Furthermore, RNA transcript splicing is regulated by the spliceosome, another ribonucleoprotein complex, consisting of many protein and RNA components[60]. Because MLE is part of the MSL complex only in males and the spliceosome complex in both sexes, we hypothesize that CLAMP influences the relative distribution of MLE between the MSL complex and the spliceosome to co-regulate sex-specific splicing and male dosage compensation (**Fig 6C**). Without CLAMP, the MSL complex does not localize to the X-chromosome and becomes destabilized [33]; 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 sex-specifically inhibits aberrant binding of MLE to motifs that regulate splicing which alters sex-specific transcript diversity.

Overall, we hypothesize that both different composition of the spliceosome and differential recruitment to chromatin drive sex-specific changes in splicing. We identify CLAMP as a maternal factor that regulates sex-specific alternative splicing through its sex-biased association and recruitment of spliceosome components and its ability to influence the sex determination pathway. Identifying the factors that regulate this sex-biased association of CLAMP 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. Consistent with recent literature linking chromatin accessibility and splicing, our results suggest that CLAMP could be one example of a more general splicing regulatory mechanism controlled by the interaction between pioneer TFs that alter chromatin accessibility and components of the RNA processing machinery to generate spatial-temporal transcript diversity. While we analyzed sex-specific transcriptome diversity in this study and linked it to the sex-specific dosage compensation process, 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 pipeline to uncover mechanisms which drive such spatial-temporal transcript diversity by integrating splicing and chromatin occupancy data.

## Materials 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 [79]. 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 assays 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 cleared 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 were 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) were 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. Supernatants were 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. Next, embryos were centrifuged at 12,000g for 10 min at 4°C and liquid removed and embryos were washed twice in Dig-Wash buffer with EDTA. Then, embryos were incubated for 3 hours at 4°C in ~700 ng/ml pAMNase solution in Dig-Wash buffer with EDTA. Embryos were 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 performed for 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 were 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 was 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 was added to each tube, mixed and total solution was transferred to phase lock tubes and centrifuged at 16,000g for 5 minutes. After adding 300µl of chloroform and mixing gently, samples were centrifuged at 16,000g for 5 minutes at RT. The aqueous layer was transferred to a DNA low binding tube. 1µl glycogen (5mg/ml) and 750µl ethanol added to precipitate DNA at -80°C. Samples were 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 [63, 64]. 1ng of Cut and Run DNA was used to make libraries using the KAPA Hyper prep kit and SeqCap adapters A &B (Roche) according to manufacturer's protocol. For library amplification 14 cycles were 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 methods

### Sex-specific splicing event analysis

RNA sequencing data from Rieder et al 2017 (#GSE102922) was analyzed using time2splice to determine sex-specifically splicing events. dmel-all-r6.29.gtf from BDGP6 in genomes [80] 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[81], Salmon[82] was used to quantify transcript expression for treatment and control samples. Calculated transcripts per million (TPM) values from SUPPA[42] 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 0-2 Hr (pre-MZT) and 2-4 Hr (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 0-2 Hr (pre-MZT) controls to show which genes are normally sex-specifically spliced pre-MZT. The same process was followed at 2-4 Hr (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 for post-MZT samples.

## 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 [83] and deeptools [84]. 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 [85] suite was used.

## Cut and Run: Data analysis

Sequenced reads were run through FASTQC[81](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 [81], 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[86], HISAT2[87], and BWA[88]. 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[89]( 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 [85] suite was used. Data submitted in GEO repository (#GSE174781).

## Competing Interest Statement

The 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 is funded by the 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 Henikoff for pAMNase protein and spike-in DNA for Cut and Run, and Daniel J McKay for sharing his 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 assays and IP. J.U. carried out sex determination pathway splicing assays and WB.

## References

1. Aanes H*, et al.* Differential transcript isoform usage pre-and post-zygotic genome activation in zebrafish. *BMC genomics* **14**, 331 (2013).

2. Revil T, Gaffney D, Dias C, Majewski J, Jerome-Majewska LA. Alternative splicing is frequent during early embryonic development in mouse. *BMC genomics* **11**, 399 (2010).

3. Mayne BT*, et al.* Large scale gene expression meta-analysis reveals tissue-specific, sex-biased gene expression in humans. *Frontiers in genetics* **7**, 183 (2016).

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

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

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

7. Schulz KN*, et al.* Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome research* **25**, 1715-1726 (2015).

8. Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. *Nature Reviews Genetics* **20**, 221-234 (2019).

9. Rieder LE*, et al.* Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP. *Genes & development* **31**, 1494-1508 (2017).

10. Artieri CG, Fraser HB. Transcript length mediates developmental timing of gene expression across Drosophila. *Molecular biology and evolution* **31**, 2879-2889 (2014).

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

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

13. Förch P, Valcárcel J. Splicing regulation in Drosophila sex determination. In: *Regulation of Alternative Splicing*). Springer (2003).

14. Haussmann IU*, et al.* m 6 A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* **540**, 301-304 (2016).

15. Salz H, Erickson JW. Sex determination in Drosophila: The view from the top. *Fly* **4**, 60-70 (2010).

16. 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* **15**, 904-917 (1995).

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

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

19. Salz HK. Male or female? The answer depends on when you ask. *PLoS Biol* **5**, e335 (2007).

20. Penalva LO, Sánchez L. RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation. *Microbiology and molecular biology reviews* **67**, 343-359 (2003).

21. Rieder LE, Jordan III WT, Larschan EN. Targeting of the dosage-compensated male X-chromosome during early Drosophila development. *Cell reports* **29**, 4268-4275. e4262 (2019).

22. Prayitno K, Schauer T, Regnard C, Becker PB. Progressive dosage compensation during Drosophila embryogenesis is reflected by gene arrangement. *EMBO reports* **20**, e48138 (2019).

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

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

25. 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* **23**, 473-485 (2013).

26. Urban J*, et al.* Enhanced chromatin accessibility of the dosage compensated Drosophila male X-chromosome requires the CLAMP zinc finger protein. *PloS one* **12**, e0186855 (2017).

27. Duan JE*, et al.* CLAMP and Zelda function together as pioneer transcription factors to promote Drosophila zygotic genome activation. *bioRxiv*, (2020).

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

29. Larschan E*, et al.* X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. *Nature* **471**, 115-118 (2011).

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

31. Ankush Jagtap PK*, et al.* 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 (2019).

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

33. Soruco MM*, et al.* The CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage compensation. *Genes & development* **27**, 1551-1556 (2013).

34. Kaye EG*, et al.* 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 (2018).

35. Quinn JJ, Zhang QC, Georgiev P, Ilik IA, Akhtar A, Chang HY. Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. *Genes & development* **30**, 191-207 (2016).

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

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

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

39. Paris M, Villalta JE, Eisen MB, Lott SE. Sex bias and maternal contribution to gene expression divergence in Drosophila blastoderm embryos. *PLoS Genet* **11**, e1005592 (2015).

40. 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* **9**, e1000590 (2011).

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

42. Trincado JL*, et al.* SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome biology* **19**, 1-11 (2018).

43. Brooks AN*, et al.* Regulation of alternative splicing in Drosophila by 56 RNA binding proteins. *Genome research* **25**, 1771-1780 (2015).

44. Urban JA*, et al.* The essential Drosophila CLAMP protein differentially regulates non-coding roX RNAs in male and females. *Chromosome Research* **25**, 101-113 (2017).

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

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

47. Gutierrez-Perez I*, et al.* Ecdysone-induced 3D chromatin reorganization involves active enhancers bound by Pipsqueak and Polycomb. *Cell reports* **28**, 2715-2727. e2715 (2019).

48. Duan JE*, et al.* CLAMP and Zelda function together to promote Drosophila zygotic genome activation. *elife* **10**, 2020.2007. 2015.205054 (2021).

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

50. Moschall R*, et al.* Drosophila Sister-of-Sex-lethal reinforces a male-specific gene expression pattern by controlling Sex-lethal alternative splicing. *Nucleic acids research* **47**, 2276-2288 (2019).

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

52. Nagoshi RN, Baker BS. 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 (1990).

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

54. Salz HK, Maine EM, Keyes LN, Samuels ME, Cline TW, Schedl P. 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 (1989).

55. Samuels ME, Schedl P, Cline TW. 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 (1991).

56. Petrova V*, et al.* Chromatin accessibility regulates intron retention in a cell type-specific manner. *bioRxiv*, (2021).

57. Agirre E, Oldfield A, Bellora N, Segelle A, Luco R. Splicing-associated chromatin signatures: a combinatorial and position-dependent role for histone marks in splicing definition. *Nature communications* **12**, 1-16 (2021).

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

59. Reenan RA, Hanrahan CJ, Ganetzky B. 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 (2000).

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

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

62. Quinn JJ*, et al.* Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nature biotechnology* **32**, 933-940 (2014).

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

64. Uyehara CM, McKay DJ. 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 (2019).

65. Gabellini N. A polymorphic GT repeat from the human cardiac Na+ Ca2+ exchanger intron 2 activates splicing. *European Journal of Biochemistry* **268**, 1076-1083 (2001).

66. Hefferon TW, Groman JD, Yurk CE, Cutting GR. A variable dinucleotide repeat in the CFTR gene contributes to phenotype diversity by forming RNA secondary structures that alter splicing. *Proceedings of the National Academy of Sciences* **101**, 3504-3509 (2004).

67. Lin C-L*, et al.* RNA structure replaces the need for U2AF2 in splicing. *Genome research* **26**, 12-23 (2016).

68. Hartmann B*, et al.* Distinct regulatory programs establish widespread sex-specific alternative splicing in Drosophila melanogaster. *Rna* **17**, 453-468 (2011).

69. Blanchette M*, et al.* 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 (2009).

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

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

72. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. Sex-dependent gene expression and evolution of the Drosophila transcriptome. *Science* **300**, 1742-1745 (2003).

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

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

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

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

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

78. Lv M*, 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* **47**, 3142-3157 (2019).

79. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**, 671-675 (2012).

80. DePristo MA*, et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491 (2011).

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

82. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* **14**, 417-419 (2017).

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

84. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**, W187-W191 (2014).

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

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

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

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

89. Zhang Y*, et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, 1-9 (2008).

**Fig 1. Alternative splicing during early *Drosophila melanogaster* embryonic development**

**A** Schematic diagram showing 7 different types of Alternative splicing (AS).

**B** Pie chart showing the distribution of different types 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 is noted at the top of the schematic. A Chi-square test was performed to determine if there is a significant difference between the percentage of each type of AS including MXE splicing (grey region) in the presence vs. absence of CLAMP in each class of sample: females/males 0-2 Hr, and 2-4 Hr embryos. A significant difference ($p<0.001$ marked by ***) was found between categories connected by solid black lines.

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

**A** Bar graph showing the percentage of transcripts (values noted at the top of each bar) out of total AS events or sex-specific splicing events (number of events noted within parentheses at the top of each bar) regulated by 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 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 that were not identified in control samples.

**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).

**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 the TSS with more CLAMP binding in female sex-specifically spliced genes vs. male sex-specifically spliced genes. The dotted box in **A-D** marks the gene body region in sex-specifically spliced and biased expressed genes.

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

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

**B** Electrophoresis gel image (inverted colors) 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 previously reported (Ref) quantification of chromatin accessibility at each locus in the genome. Positive/high accessibility values (blue) indicate high chromatin accessibility, and negative/low (red) accessibility values indicate low chromatin accessibility. Each window covers MACC values ranging from -0.333 to +1.33. MACC values increase in females after CLAMP RNAi, specifically at exon 3 (red box), and can be seen in the 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 from 0-2 Hr embryonic (lane 2-5 & 7-10) and third instar larval samples (a-g) showing splicing of *dsx* (E) and *msl2* (F) transcripts in females (lane 3,5,8,10, a-d) and males (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 genotypes are the same as in B. The schematic at the top of each gel image shows female and male splice variants of *dsx* (E) and *msl2* (F) transcripts.

**G** Fluorescent microscopy images of polytene chromosomes from third instar salivary gland in the genotypes mentioned to the left of each panel showing the distribution of CLAMP (green) and MSL2 (red) on chromatin (blue, DAPI).

**Fig 5. MLE binds to different motifs and to different chromatin regions when colocalizing with CLAMP compared to binding alone**

**A-D** Distribution of MLE peak percent overlap with CLAMP peaks (**A, C**) and unique MLE peaks (**B, D**) at different genomic 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. MLE distribution was measured on the X chromosome and autosomes in male embryos at 0-2 hr pre-MZT and 2-4 Hr post-MZT stages under normal conditions (**A, B**) and after the loss of maternal CLAMP (**C, D**). '**N**' denotes the total number of peaks in each category. The most frequently identified sequence motif (MEME) for MLE peaks overlapping with CLAMP (**A, C**) and unique MLE peaks (**B, D**) in each category is shown at the top of the relevant bar plot.

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

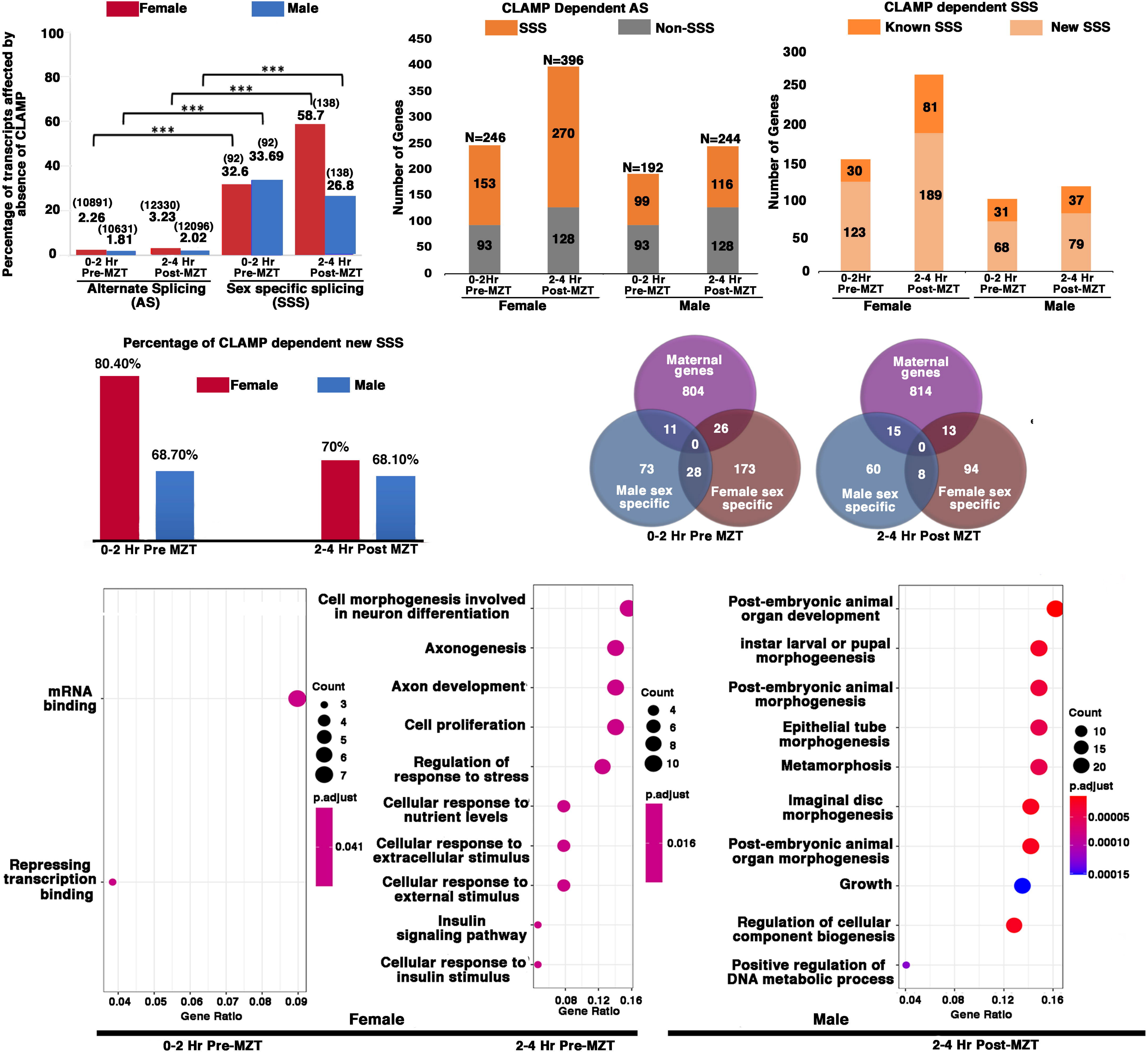
**A** CLAMP regulates splicing in both males and females via directly binding to intronic sequences of CLAMP-dependent sex-specifically spliced genes.

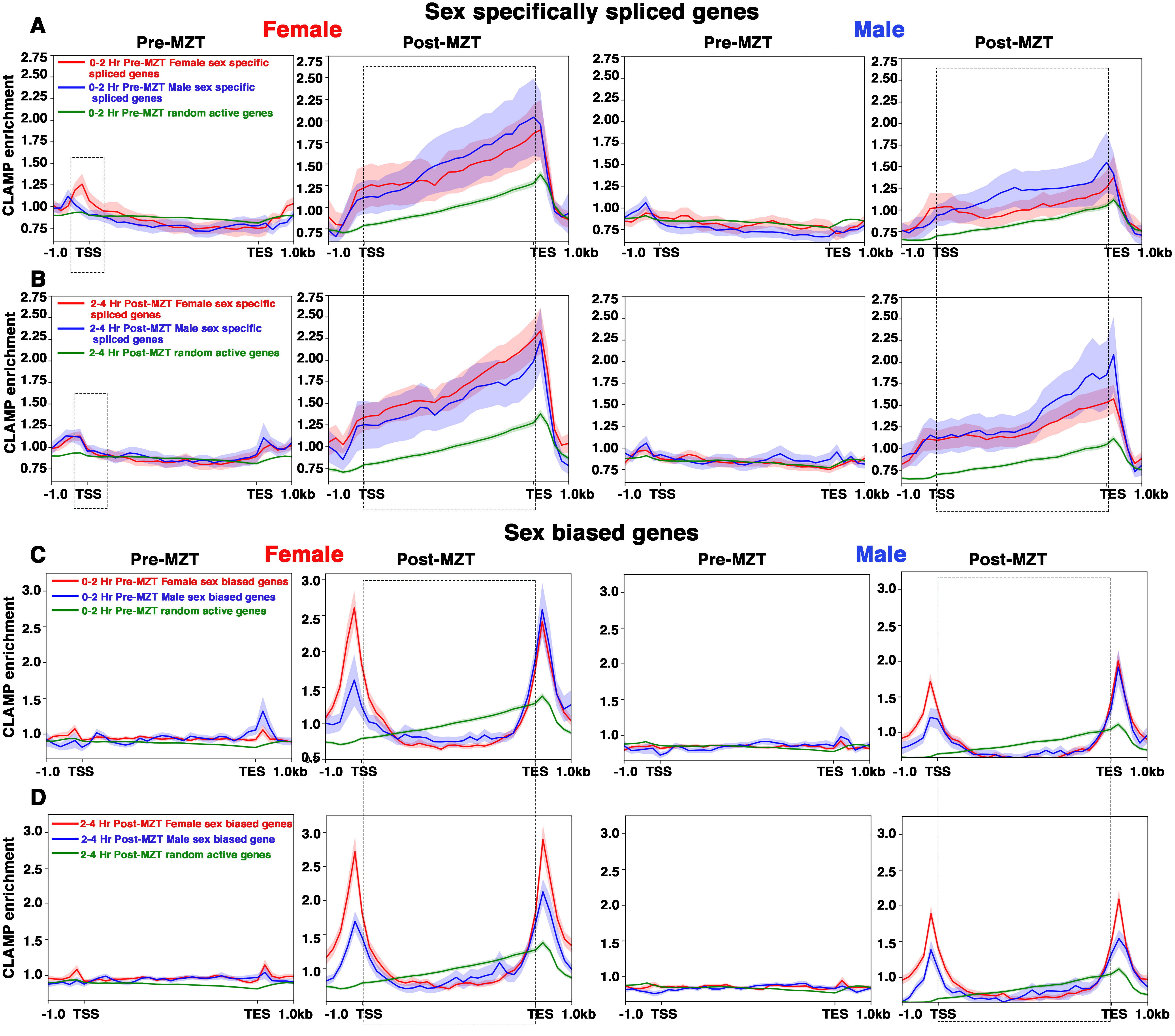
**B** In females, binding of CLAMP near the SxlPe promoter regulates chromatin at exon three (blue square) of the *sxl* gene, such that it is spliced out and functional Sxl protein is formed, which drives female-specific splicing events. In males, the absence of CLAMP near the SxlPe promoter changes chromatin structure at exon three, which is retained in the *sxl* transcript, reducing levels of functional Sxl protein. The absence of CLAMP in females thus results in the aberrant production of non-functional male-specific *sxl* transcripts which retain exon three, reducing levels of functional 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 promoters and CES. In the absence of CLAMP, MLE is lost from many sites, including CES and promoters, and is gained at new intronic sequences which contain motifs that regulate splicing, resulting in aberrant sex-specific splicing in males.

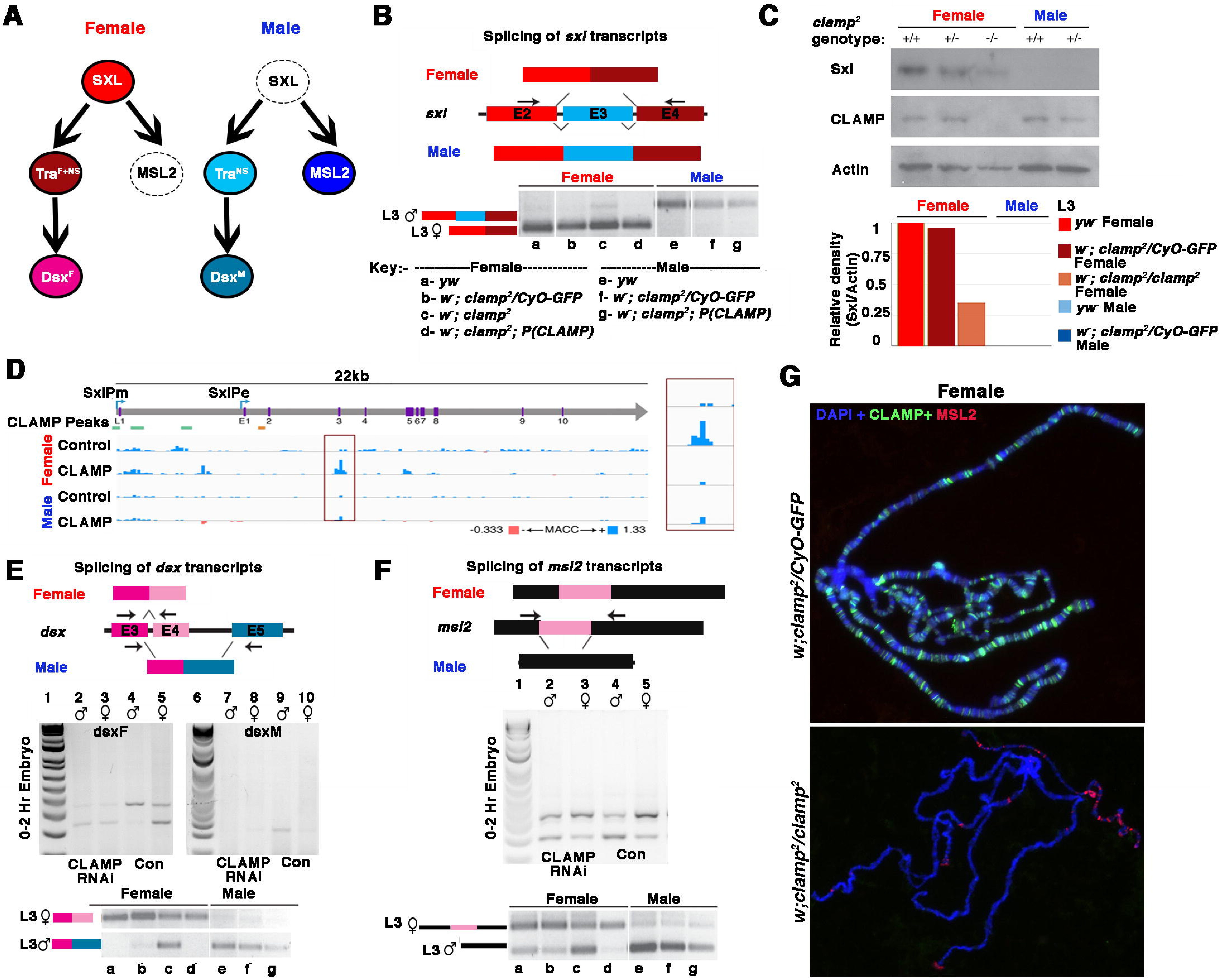
The three mechanisms proposed in parts **A**, **B**, and **C** are not mutually exclusive and are likely to occur simultaneously.

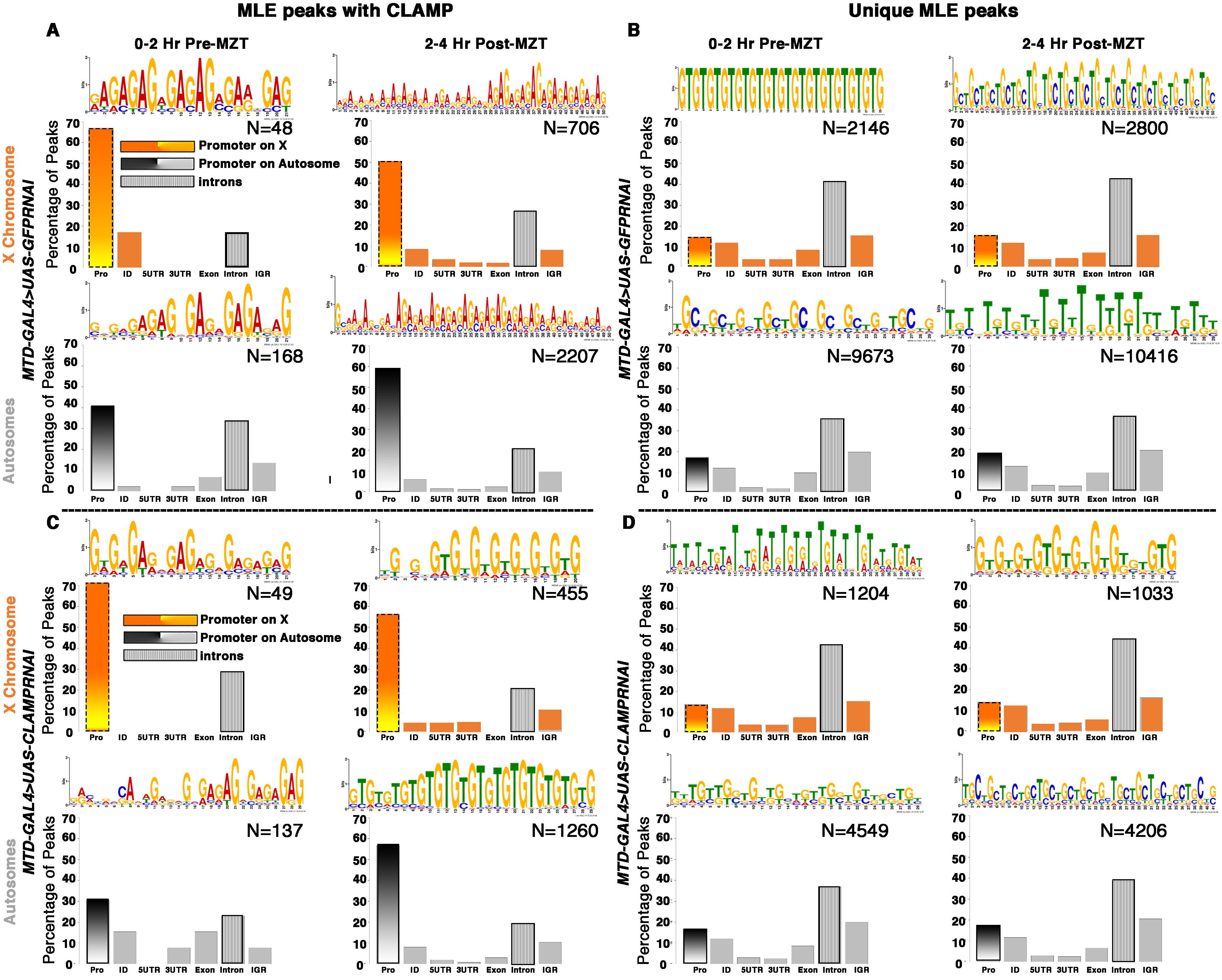
A
AL
A5SS
MXE
AF
SE
A3SS
RI
B
Female
Male
0-2 Hr Pre-MZT
2-4 Hr Post-MZT
Alternate Splicing (AS)
16.27%
18.42%
15.88%
18.07%
Nuclear Cycle (NC)
1
8
13
14
15
Time (Hr)
0
1
2
2.5
4
MTD-GAL4> UAS-GFPRNAi
MTD-GAL4> UAS-CLAMPRNAi
AL 3.3%
SE 15.6%
A5SS 17.6%
A3SS 15.6%
MXE 14.0%
RI 9.7%
AF 24.1%
AL 3.1%
SE 15.2%
A5SS 16.8%
A3SS 15.1%
MXE 16.4%
RI 9.6%
AF 23.8%
AL 3.4%
SE 16.1%
A5SS 18.1%
A3SS 15.9%
MXE 12.2%
RI 10.0%
AF 24.4%
AL 3.1%
SE 15.2%
A5SS 16.7%
A3SS 15.2%
MXE 16.6%
RI 9.5%
AF 23.7%
AL 3.6%
SE 17.2%
A5SS 19.1%
A3SS 17.2%
MXE 5.6%
RI 10.6%
AF 26.6%
AL 3.2%
SE 15.3%
A5SS 16.9%
A3SS 15.3%
MXE 16.0%
RI 9.5%
AF 23.8%
AL 3.7%
SE 17.4%
A5SS 19.7%
A3SS 17.4%
MXE 4.2%
RI 10.9%
AF 26.7%
AL 3.3%
SE 15.8%
A5SS 17.0%
A3SS 15.4%
MXE 14.8%
RI 9.6%
AF 24.2%
***
***
***


Female
Male
Percentage of transcripts affected by absence of CLAMP
100
80
60
40
20
0
***
***
***
***
(10891)
2.26
(10631)
1.81
(12330)
3.23
(12096)
2.02
(92)
32.6
(92)
33.69
(138)
58.7
(138)
26.8
0-2 Hr Pre-MZT
2-4 Hr Post-MZT
0-2 Hr Pre-MZT
2-4 Hr Post-MZT
Alternate Splicing (AS)
Sex specific splicing (SSS)
CLAMP Dependent AS
SSS
Non-SSS
Number of Genes
450
400
350
300
250
200
150
100
50
0
N=246
153
93
N=396
270
128
N=192
99
93
N=244
116
128
0-2Hr Pre-MZT
2-4 Hr Post-MZT
0-2 Hr Pre-MZT
2-4 Hr Post-MZT
Female
Male
CLAMP dependent SSS
Known SSS
New SSS
Number of Genes
300
250
200
150
100
50
0
30
123
81
189
31
68
37
79
0-2Hr Pre-MZT
2-4 Hr Post-MZT
0-2 Hr Pre-MZT
2-4 Hr Post-MZT
Female
Male
Percentage of CLAMP dependent new SSS
Female
Male
80.40%
68.70%
70%
68.10%
0-2 Hr Pre MZT
2-4 Hr Post MZT
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
mRNA binding
Repressing transcription binding
Count
3
4
5
6
7
p.adjust
0.041
0.04
0.05
0.06
0.07
0.08
0.09
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
Count
4
6
8
10
p.adjust
0.016
0.08
0.12
0.16
Gene Ratio
Female
0-2 Hr Pre-MZT
2-4 Hr Pre-MZT
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
Count
10
15
20
p.adjust
0.00005
0.00010
0.00015
0.04
0.08
0.12
0.16
Gene Ratio
Male
2-4 Hr Post-MZT

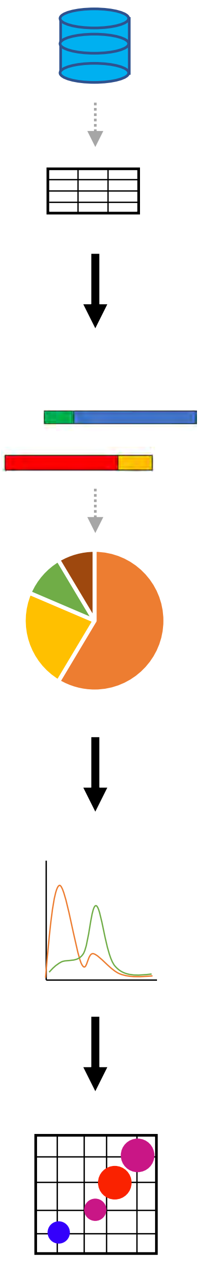


Sex specifically spliced genes
A
Female
Male
Pre-MZT
Post-MZT
Pre-MZT
Post-MZT
CLAMP enrichment
0-2 Hr Pre-MZT Female sex specific spliced genes
0-2 Hr Pre-MZT Male sex specific spliced genes
0-2 Hr Pre-MZT random active genes
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
-1.0
TSS
TES
1.0kb
B
2-4 Hr Post-MZT Female sex specific spliced genes
2-4 Hr Post-MZT Male sex specific spliced genes
2-4 Hr Post-MZT random active genes
Sex biased genes
C
Female
Male
Pre-MZT
Post-MZT
Pre-MZT
Post-MZT
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
3.0
2.5
2.0
1.5
1.0
0.5
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


A
Female
Male
SXL
Tra^F+NS
MSL2
Dsx^F
SXL
Tra^NS
MSL2
Dsx^M
B
Splicing of sxl transcripts
Female
sxl
E2
E3
E4
Male
L3 ♂
L3 ♀
Female
Male
a b c d e f g
Key:- ------------Female------------- ------------Male------------
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
Female
Male
L3
Relative density (Sxl/Actin)
1
0.75
0.5
0.25
0
yw⁻ Female
w⁻; clamp²/CyO-GFP Female
w⁻; clamp²/clamp² Female
yw⁻ Male
w⁻; clamp²/CyO-GFP Male
D
22kb
SxlPm
SxlPe
L1
E1
2
3
4
5 6 7
8
9
10
CLAMP 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
CLAMP RNAi
Con
CLAMP RNAi
Con
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
CLAMP RNAi
Con
Female
Male
L3 ♀
L3 ♂
a b c d e f g
G
Female
DAPI + CLAMP+ MSL2
w;clamp²/CyO-GFP
w;clamp²/clamp²


MLE peaks with CLAMP
Unique MLE peaks
A
B
C
D
0-2 Hr Pre-MZT
2-4 Hr Post-MZT
X Chromosome
Autosomes
MTD-GAL4>UAS-GFPRNAi
MTD-GAL4>UAS-CLAMPRNAi
Percentage of Peaks
Promoter on X
Promoter on Autosome
introns
Pro
ID
5UTR
3UTR
Exon
Intron
IGR
N=48
N=706
N=2146
N=2800
N=168
N=2207
N=9673
N=10416
N=49
N=455
N=1204
N=1033
N=137
N=1260
N=4549
N=4206
bits

A
Exon 1
Exon 2
Exon 3
Exon 4
Intron 1
Intron 2
Intron 3
A
Exon 1
Exon 3
Exon 4
mRNA1
Exon 1
Exon 2
Exon 3
Exon 4
Intron 1
Intron 2
Intron 3
A
Exon 1
Exon 2
Exon 3
Exon 4
mRNA1'
CLAMP=Chromatin linked adapter for MSL proteins
MLE=Maleless, RNA helicase
Spliceosome Complex
Male specific lethal Complex (MSL)
TES= Transcription start site
TES= Transcription end site
CES= Chromatin entry site
Autosome
X-chromosome
mRNA Normal transcript
mRNA Aberrant transcript
B
SxlPe
Exon 1
Exon 2
Exon 3
Exon 4
Intron 1
Intron 2
Intron 3
X
♀
SxlPe
Exon 1
Exon 2
Exon 3
Exon 4
Intron 1
Intron 2
Intron 3
X
♂
SXL
sxl
Exon 1
Exon 2
Exon 4
mRNA
SXL
sxl
Exon 1
Exon 2
Exon 3
Exon 4
mRNA
SXL
Exon 1
Exon 2
Exon 3
Exon 4
Intron 1
Intron 2
Intron 3
X
♀
C
♂
TSS
Exon 1
Exon 2
Exon 3
Exon 4
Intron 1
Intron 2
Intron 3
TES
TSS
Exon 1
Exon 2
Exon 3
Intron 1
Intron 2
TES
CES
mRNA2
Exon 2
Exon 3
Exon 4
Exon 2
Exon 3
mRNA3
♂
TSS
Exon 1
Exon 2
Exon 3
Exon 4
Intron 1
Intron 2
Intron 3
TES
TSS
Exon 1
Exon 2
Exon 3
Intron 1
Intron 2
TES
CES
mRNA2'
Exon 1
Exon 2
Exon 3
Exon 1
Exon 2
mRNA3'



## A. Retrieve raw data, quality control, trim, and alignment

Run as needed. Note enumeration follows script names.

### 1. Retrieve data

Creates time2splice/ folder structure, metadatafile.csv and SraAccList.txt. .fastqs retrieved using SraAccList.txt.

### 2. Run quality control

Run ***FastQC*** for all .fastq files in time2splice/ directory.

### 3. Trim data

Run ***Trim Galore!***, then ***FastQC*** to trim reads below quality threshold. Merge lanes of the same flow cell .fastq files.

### 4/5. Run & plot alignment

Run ***Bowtie2***, ***BWA***, or ***HISAT2*** on .fastq data in time2splice directory. Plot alignment for one or two aligners.

## B. Temporal expression analysis

Run as needed. Note enumeration follows script names.

### 1. Run transcript quantification

Quantify transcript treatment and control expression with ***Salmon***.

### 2. Run differential splicing analysis

Run ***SUPPA*** differential splicing analysis across case and controls.

### 3. Format results

Converts NM_ gene names to flybase name. Merges outputs.

### 4. Identify differential splicing forms

***SUPPA*** identifies forms of differential splicing (e.g. using PSI and DTU).

### 5. Calculate total control alternative splicing

Calculate and plot the proportions of alternative splicing in control samples.

### 6. Calculate total case alternative splicing

Calculate and plot the proportions of alternative splicing in case samples.

### 7. Get bias genes

Retrieve male and female biased genes. Create .beds to plot average profiles.

### 8/9. Plot splicing events

Plots alternative splicing (PSI and DTU), and events in categories (e.g. female sex specific, male new sex specific).

## C. Temporal protein-DNA analysis

Run as needed. Note enumeration follows script names.

### 1. Mark duplicate reads

Run Run ***Picard's MarkDuplicates*** in for all .sorted.bam files in a given directory.

### 2. Call peaks

Run ***MACS2*** to call peaks for all .sorted.bam files in a given directory.

### 3. Find fold enrichment

Generate signal track using ***MACS2*** to profile transcription factor modification enrichment levels genome-wide.

time2splice

## D. Temporal multi-omics integration

Run as needed. Miscellaneous tests at README end plot peak intensity and perform chi-squared test

### Peak intersections

Run ***Intervene*** to view intersection of each narrowpeak file.

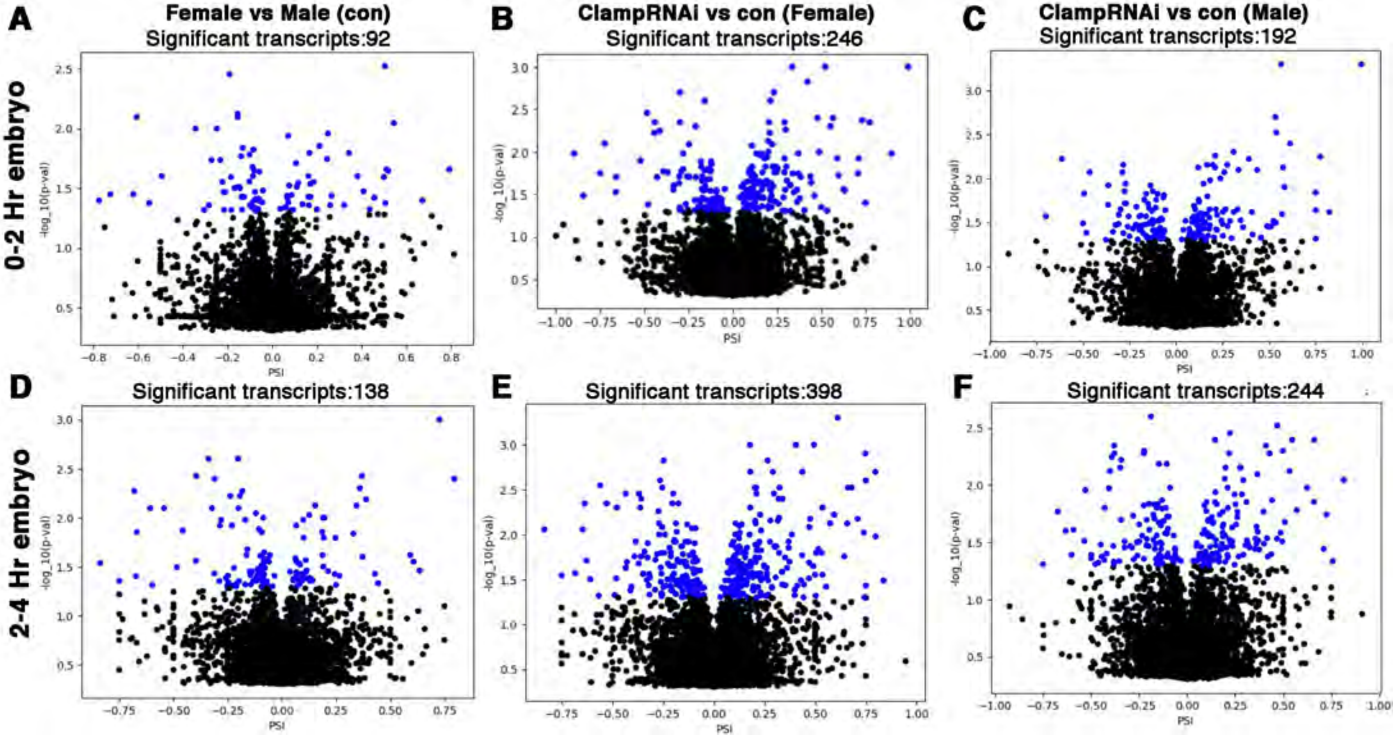
### Gene ontology

Perform gene ontology analysis with ***ClusterProfiler*** given a list of genes.

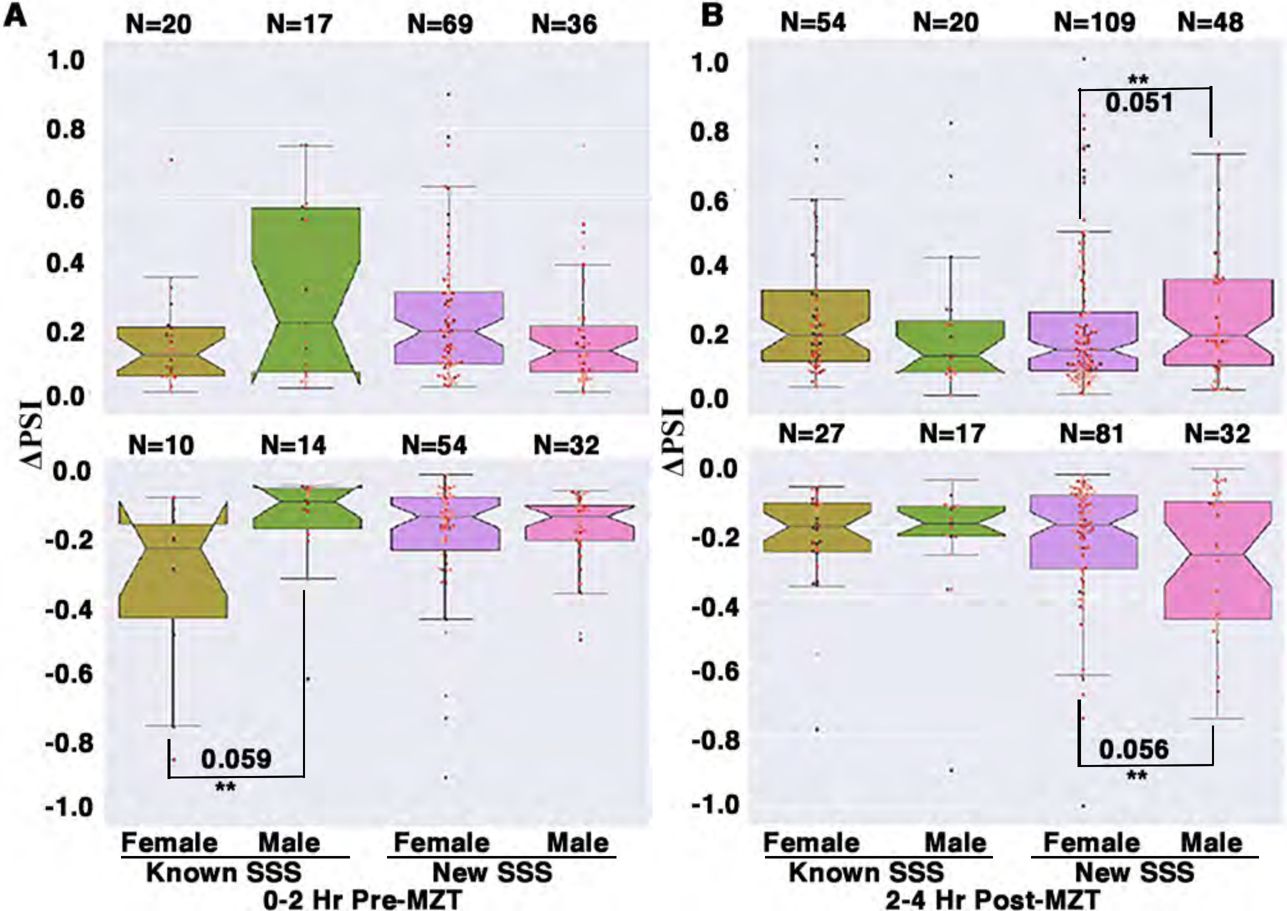
### Find motifs *denovo*

Get coordinates of bed file and run through ***MEME***.

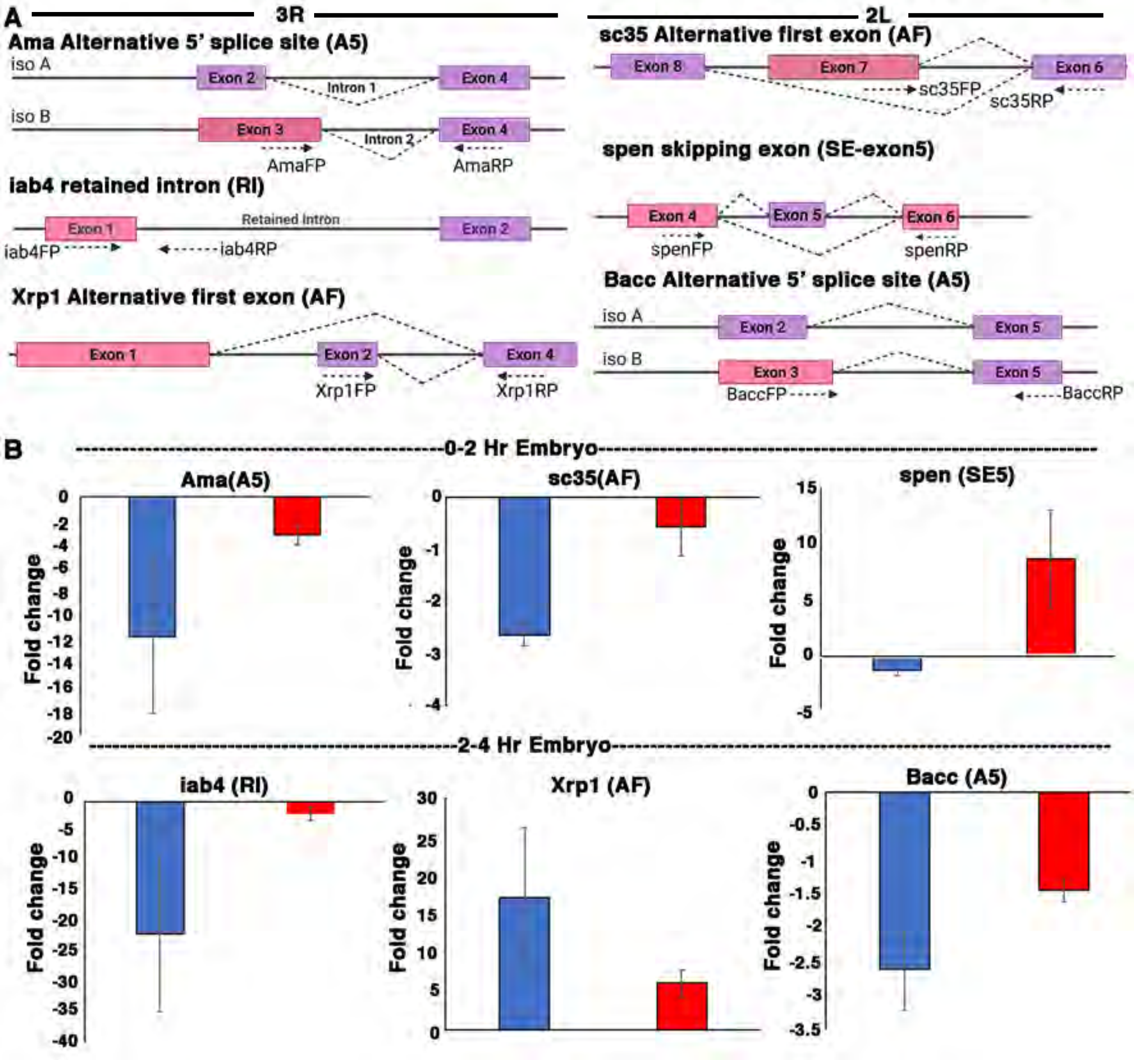
Ray, Conard 2021_Fig S2



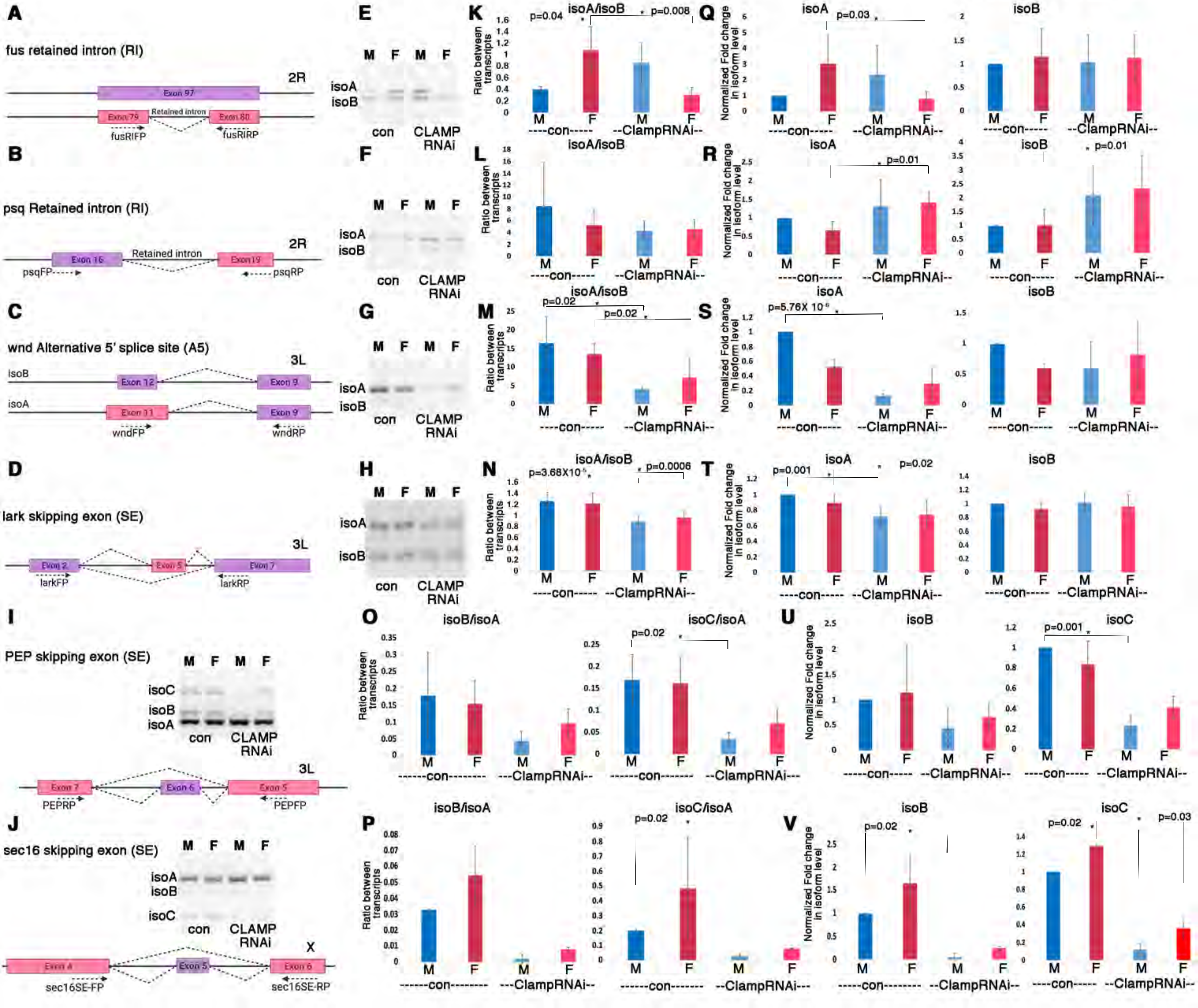
Ray, Conard 2021_Fig S3



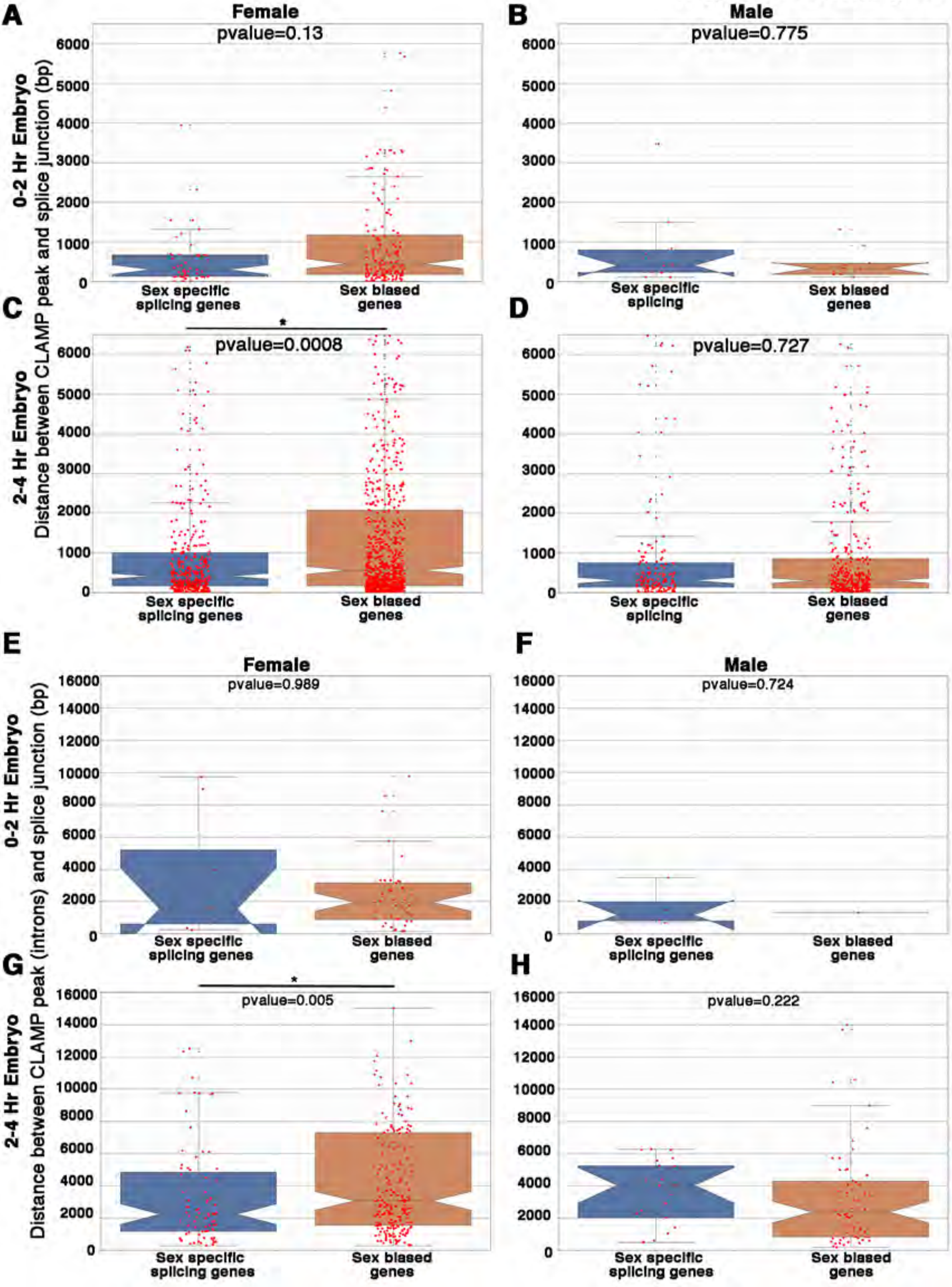
Ray, Conard 2021_Fig S4



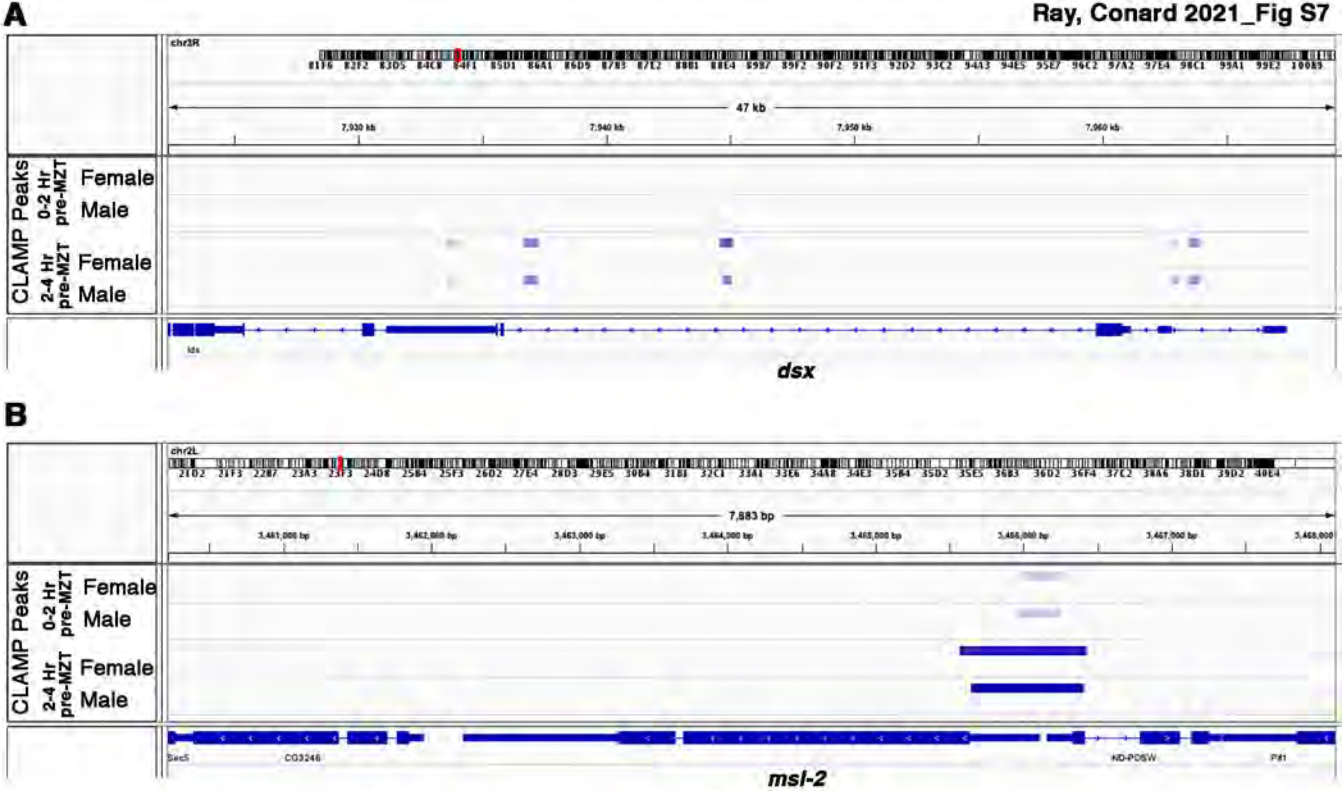
**Ray, Conard 2021_Fig S5**



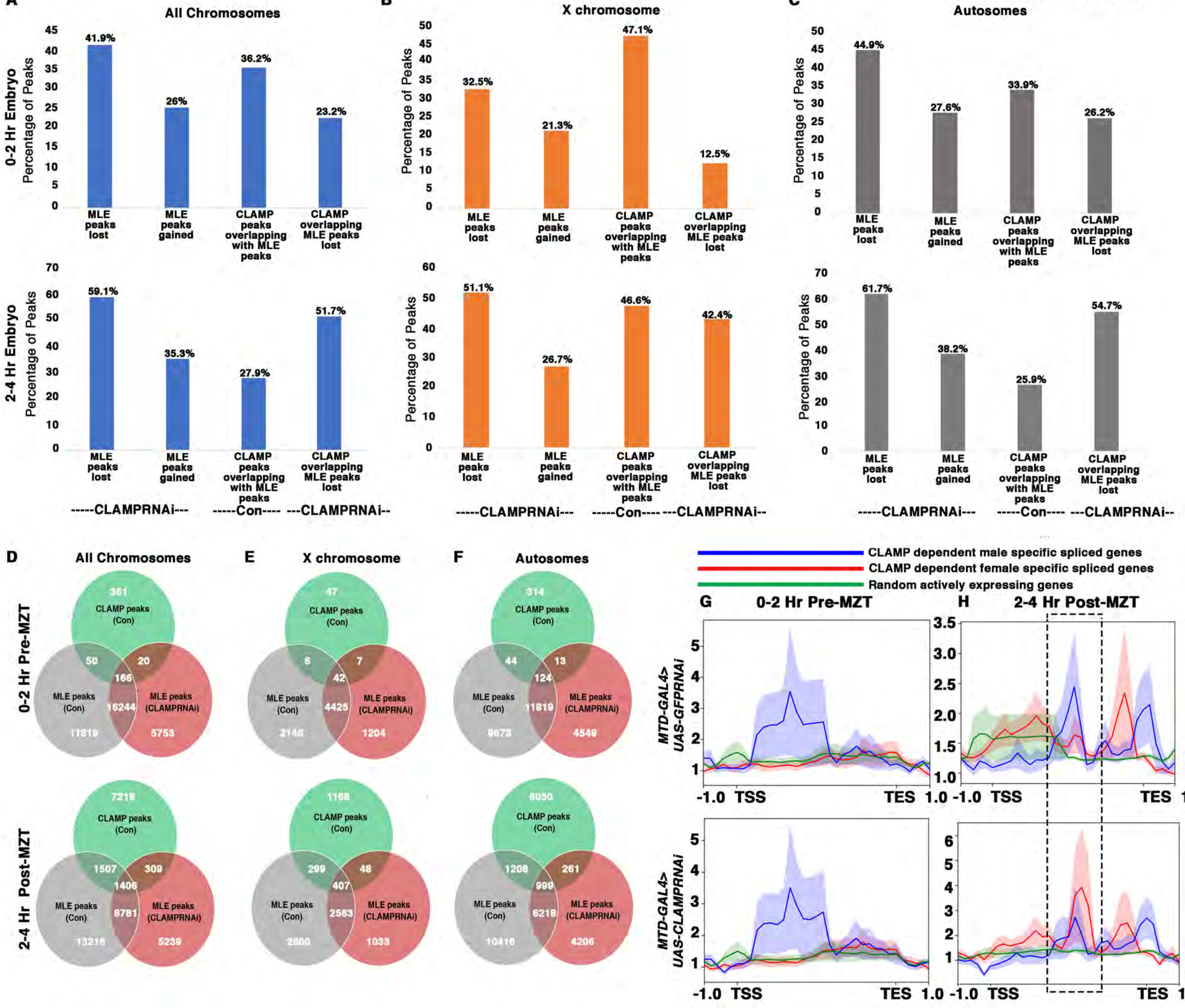
Ray, Conard 2021_Fig S6

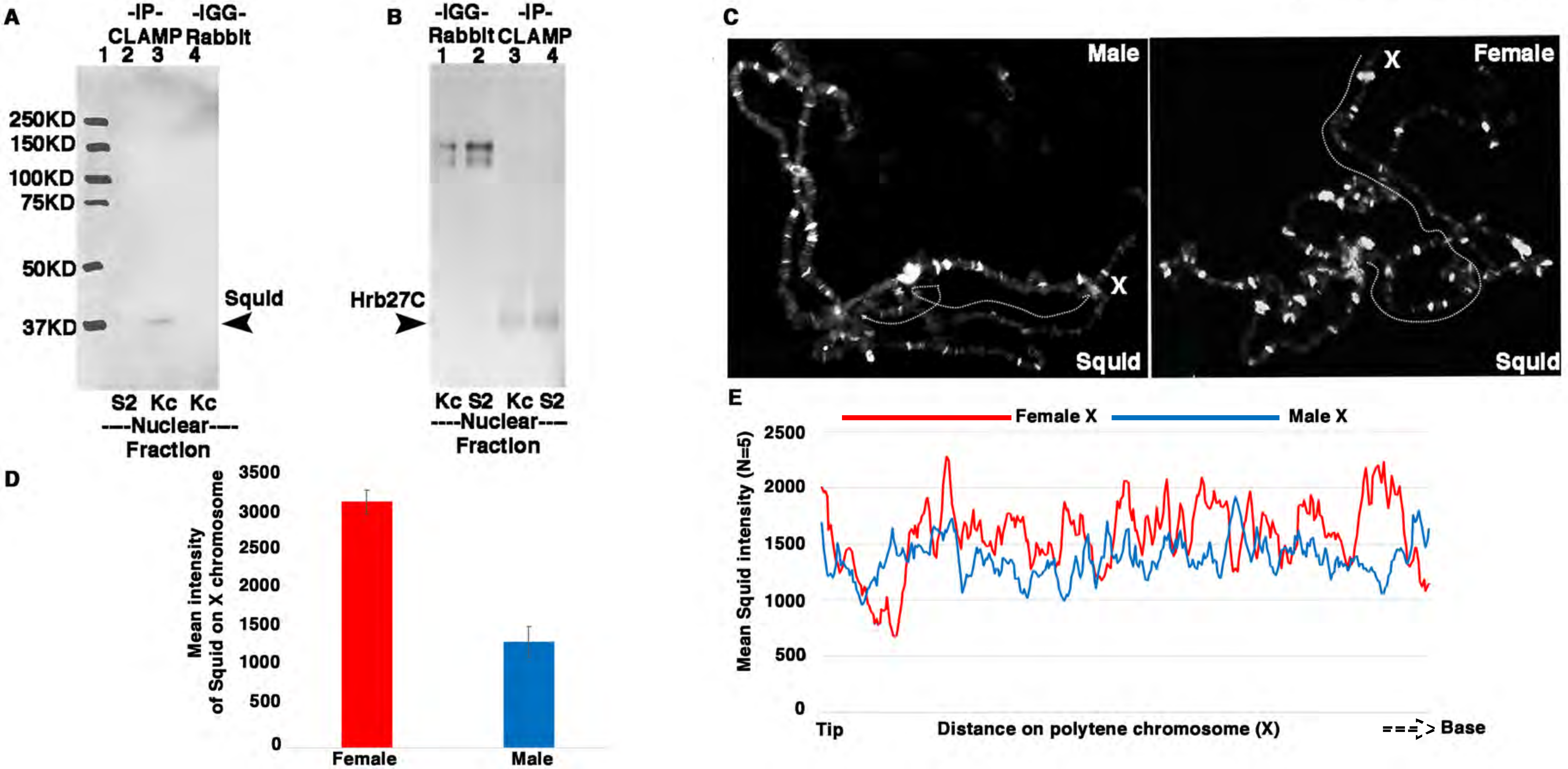


Ray, Conard 2021_Fig S7



Ray, Conard 2021_Fig S8



Ray, Conard 2021_Fig S9