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], Joseph Aguilera[#], Annie Huang[#], 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) In both males and females, GA-binding pioneer factor CLAMP (**C**hromatin **l**inked **a**dapter for **M**SL **p**roteins) links the DNA of gene bodies of sex-specifically spliced genes directly to the RNA of target genes and physically interacts with snRNA and protein components of the splicing machinery; 2) In males, CLAMP regulates the distribution of the spliceosome component **M**ale**less** (MLE) to prevent aberrant sex-specific splicing; 3) In females, CLAMP binds to the DNA and RNA encoded by the *sxl* gene, the master regulator of sex determination, to directly regulate its splicing which also modulates downstream targets. Overall, we provide key insight into how maternal factors influence sex-specific transcript diversity.

34

35 **Introduction**

36

37 One of the greatest challenges in modern biology is understanding the mechanism and significance
38 of widespread transcript diversity between sexes and different developmental stages, tissues, and
39 cell-types. **A**lternative **S**plicing (**AS**), a mechanism of selective inclusion or exclusion of introns
40 and exons, drives widespread transcript diversity[1,2]. In addition to basic development and
41 physiology, transcriptome diversity is critical for disease biology, especially in neurodegenerative
42 diseases and developmental disorders that often show sex or tissue-specific differences in
43 progression and severity[3-6]. Across species, precise regulation of genes to produce specific splice
44 variants is critical for all developmental decisions, including sex determination. A key to
45 understanding how transcript diversity drives biological processes lies in the events that shape the
46 initial few hours of an organism's existence.

47

48 During early development, protein and RNA deposited by the mother into the embryo shape early
49 embryonic milestones across metazoans[7,8]. Initially, cell number increases, followed by cellular
50 differentiation into specific cell types. Sexual identity is then established, driving the fundamental
51 physiological differences between sexes. However, the mechanisms by which maternally
52 deposited proteins and RNAs regulate this process of sexual differentiation remains poorly
53 understood. Moreover, maternal factors are often essential regulators that can have a lasting impact
54 on gene regulation later in the life of an organism. Thus, it is essential to define the influence of
55 maternal factors on transcriptome diversity during the early stages of embryonic development.
56 Therefore, the key question is: How do maternally deposited products regulate transcript diversity
57 including sex-specific splice variants?

58

59 The *Drosophila* embryo is an excellent tool to study the role of maternally deposited proteins and
60 RNA in early development as it is easy to perform genetic manipulation to remove maternal factors
61 to define how they regulate splicing and transcription. Also, embryos can be sexed before zygotic
62 genome activation due to the recent application of a meiotic drive system [9].

63

64 During *Drosophila* embryogenesis, **Z**ygotic **G**enome **A**ctivation (**ZGA**) occurs shortly after the
65 first two hours of development. Concurrently, maternal transcripts gradually decrease in
66 abundance, and zygotic transcription increases, a process called the **M**aternal to **Z**ygotic **T**ransition
67 (**MZT**). ZGA starts approximately 80 min after egg laying and most maternal transcripts are
68 degraded by 180 min after egg laying[10]. Even at these early stages of development, AS generates
69 multiple variants of the same gene transcripts resulting in transcript diversity. Although the earliest
70 genes transcribed from the zygotic genome are mainly intron-less, almost 30% of early zygotic
71 transcripts do have introns[11,12]. Furthermore, genes involved in sex determination use AS to drive
72 male versus female-specific development [13]. Hence, during early embryonic development, AS is
73 important for shaping cell and tissue-specific transcriptomes and sexual differentiation.
74
75 The master regulator of *Drosophila* sex determination is the *sxl* gene, which undergoes alternative
76 splicing such that exon three is retained in males but not in females[14,15]. Therefore, functional Sxl
77 protein is made only in the female embryo. Early expression of the Sxl protein made from a
78 transcript driven by Sxl_{Pe}, an early *sxl* promoter[16], autoregulates splicing of *sxl*. Next, the Sxl
79 protein causes splicing of exon three from pre-mRNA of *sxl* transcripts initiated from the late
80 maintenance promoter (Sxl_{Pm}) specifically in females[17,18]. The early activation of the Sxl_{Pe}
81 promoter is regulated by a group of X-linked genes called X signal elements (XSE). A higher
82 dosage of XSE gene products drives Sxl_{Pe} expression in XX females and not XY males[19].
83
84 The Sxl protein is an RNA binding protein that binds to *sxl* pre-mRNA and the pre-mRNA of
85 downstream target genes such as *transformer* (*tra*) and *male-specific lethal protein 2* (*msl-2*) to
86 regulate their splicing and stability, giving rise to female-specific splice variants[20]. The female-
87 specific Tra isoform then regulates sex-specific splicing of the *doublesex* (*dsx*) gene resulting in
88 different Dsx protein isoforms in females and males which directly regulate sexual dimorphism in
89 both sexes[14,15].
90
91 Sxl regulation of *msl-2* RNA splicing and stability links the sex determination pathway to another
92 important early embryonic event: dosage compensation, which equalizes gene dosage of X-linked
93 genes between females with two X-chromosome and males with a single X-chromosome. MSL-2
94 is an essential core component of the MSL complex, present only in the male embryo and

95 responsible for dosage compensation[21,22]. The binding of the Sxl protein to the *msl-2* transcript
96 specifically in females results in skipping of the first exon, thereby hindering translation of the
97 protein and preventing MSL complex formation in females[23].
98
99 Once formed specifically in males, the MSL complex targets the male X-chromosome at high-
100 affinity binding sequences called **C**hromatin **E**ntry **S**ites (**CES**) that are enriched for clustered long
101 GA-repeat sequences[24,25]. The maternally deposited **C**hromatin **l**inked **a**dapter for **M**SL **p**roteins
102 (CLAMP) opens chromatin[26,27] and directly contacts MSL-2[28] to promote MSL complex
103 recruitment. Subsequently, the MSL complex increases transcript levels on the male X-
104 chromosome by promoting transcription elongation[29]. Interestingly, the MSL complex shares a
105 key highly conserved component MLE (RNA helicase A) with the spliceosome complex[30-32],
106 suggesting that there is likely to be an interplay between dosage compensation and splicing that
107 has not yet been defined.
108
109 CLAMP is an essential pioneer **T**ranscription **F**actor (**TF**) that opens chromatin[26,27]. Loss of
110 CLAMP results in complete loss of MSL complex from the male X-chromosome causing male
111 lethality[33]. Unlike MSL complex component mutants, which cause only male-specific lethality,
112 *clamp* mutants are lethal in both males and females early in embryogenesis after losing maternal
113 CLAMP[27]. Several lines of evidence lead us to hypothesize that CLAMP regulates alternative
114 splicing: 1) CLAMP is bound to both intronic and promoter regions on chromatin[34], 2) Intronic
115 regions are rich in polypyrimidine tracts, which have similar GA-rich sequences to CES and often
116 contain CLAMP binding sites[35], 3) GA-rich repeat sequences within CES are thought to have
117 evolved from polypyrimidine tracks that regulate splicing[35], and 4) MALDI-mass spectrometry
118 data identifying putative CLAMP interactors identified association with 33 RNA binding proteins,
119 including 6 that regulate alternative splicing[36]. Therefore, we hypothesized that CLAMP is a
120 maternally deposited factor that shapes transcriptome diversity in early embryos by regulating
121 RNA transcript splicing.
122
123 To date, there are no reports defining alternative splicing globally in early *Drosophila*
124 *melanogaster* sexed embryos. Moreover, how maternal TFs regulate alternative splicing in any
125 organism is not known. Here, we determine that 16-18% of transcripts undergo alternative splicing

126 during the first four hours of development in *Drosophila melanogaster* embryos. Although sex-
127 specific isoforms have been identified for several candidate genes in early development[37-40], a
128 comprehensive analysis of all sex-specific isoforms forms has not yet been performed in early
129 embryos and is only available in fully developed tissues like ovaries, testis and brain[38,41].
130 Therefore, we used a meiotic drive system to sex embryos and measured total AS and sex-specific
131 splicing genome-wide before (0-2 Hrs) and after (2-4 Hrs) the major wave of MZT. For ease of
132 understanding, we have defined 0-2 Hr embryos as pre-MZT and 2-4 Hr embryos as post-MZT
133 even though a small number of early zygotic genes are activated during the 0-2 Hr time period.
134
135 To reveal new mechanisms by which maternal factors shape early embryonic transcriptome
136 diversity, we have focused on analyzing how the maternal transcription factor CLAMP regulates
137 sex-specific alternative splicing during the first few hours of development. We have identified
138 male-specific and female-specific genes whose splicing requires maternal CLAMP and are
139 important regulators of early development with diverse functions form including chromatin
140 regulation and splicing.
141
142 Moreover, we defined multiple mechanisms by which sex-specific transcript diversity is regulated
143 during early embryonic development: 1) In both males and females, CLAMP links the DNA of
144 gene bodies of sex-specifically spliced genes directly to the RNA of target genes and physically
145 interacts with U1-U6snRNAs and protein components of the splicing machinery; 2) In males,
146 CLAMP regulates the distribution of the spliceosome component **M**ale**less** (MLE) to prevent
147 aberrant sex-specific splicing; 3) In females, CLAMP binds to the DNA and RNA encoded by the
148 *sxl* gene, the master regulator of sex determination, to directly regulate its splicing which also
149 modulates downstream targets. Overall, we demonstrate that sex-specific transcriptome variation
150 is established very early in development and is regulated by maternally deposited CLAMP via
151 diverse mechanisms including direct binding to the spliceosome and the DNA and RNA of sex-
152 specifically spliced genes.
153
154
155
156

157 **Results**

158

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

160

161 To determine when during early *Drosophila melanogaster* development sex-specific splicing
162 begins, we analyzed RNA-sequencing data generated from sexed embryos at two-time points: 1)
163 0-2 Hrs (pre-MZT); 2) 2-4 Hrs (post-MZT)[9] (#GSE102922). Embryos were sexed using a meiotic
164 drive system that produces sperm with either only X or only Y chromosomes[9], resulting in progeny
165 of either only female or only male genotypes. Next, we measured the amount of AS in early
166 *Drosophila* development using a pipeline that we developed and made publicly available called
167 time2Splice (https://github.com/ashleymaeconard/time2splice). Time2Splice implements the
168 commonly used SUPPA2 algorithm to identify splice variants and provides additional modules to
169 integrate time, sex, and chromatin localization data (See **Supplemental method**) (Fig S1).
170 SUPPA2 measures the percent spliced in (PSI) for each exon to calculate the differential alternative
171 splicing between samples, represented as ΔPSI[42].

172

173 AS events can be classified into seven categories: 1) Alternative Last Exon (AL); 2) Skipping
174 Exon (SE); 3) Alternative 5' Splice Site (A5SS); 4) Alternative 3' Splice Site (A3SS); 5) Mutually
175 Exclusive Exon (MXE) 6) Retained Intron (RI) and 7) Alternative First Exon (AF) (**Fig1A**). We
176 measured the total number of splicing events that belong to each of the seven different categories
177 at the 0-2 Hr and 2-4 Hr embryonic stages in both female and male embryos. We found that 16-
178 18% of total transcripts are alternatively spliced in early embryos. **A**lternative **F**irst Exon (AF) is
179 the most common type of alternative splicing, constituting almost one-fourth of total AS. In
180 contrast, **A**lternative **L**ast Exon (AL) is the least common alternative splice type. The AS transcript
181 distribution across categories was similar between the two time points and sexes (**Fig 1B**).

182

183 We hypothesized that CLAMP regulates AS in early embryos for the following reasons: 1)
184 CLAMP is a maternally deposited pioneer factor enriched at intronic regions ; 2) Proteomic data
185 identified a physical association between spliceosome components and CLAMP[36]; 3) CLAMP
186 binding sites evolved from polypyrimidine tracts that regulate splicing[35]. We tested our hypothesis
187 in early staged and sexed embryos by measuring differences in splicing in RNA-sequencing data

188 generated from male and female 0-2 Hr and 2-4 Hr (pre- and post-MZT) embryos with and without
189 maternal CLAMP[9]. The maternal triple driver GAL4, *MTD-GAL4* was used to drive *UAS-*
190 *CLAMPRNAi[val22]* that strongly reduce maternal CLAMP levels, validated by qPCR and
191 Western blot [9].
192
193 First, we asked whether CLAMP alters AS and we found 200-400 transcripts where AS is regulated
194 by CLAMP depending on the time point and sex (**Fig 2A, B** and **Fig S2B, C, D, E**). Next, we
195 asked which type of alternative splicing is most affected by depleting CLAMP. The overall
196 distribution of AS transcripts into the seven splicing types remains mostly unaffected in the
197 absence of maternal CLAMP. However, at the 0-2 Hr (pre-MZT) stage, loss of maternal CLAMP
198 results in a more substantial decrease in **M**utually E**x**clusive **E**xon (**MXE**) splicing in both males
199 and females compared with all of the other types of splicing (males: p-value < 3.21e-21; females:
200 p-value < 6.26e-87 chi-squared test) (**Fig 1B**). At the 2-4 Hr (post-MZT) stage, only male embryos
201 have a significant percentage of MXE splicing affected in the absence of maternal CLAMP (p-
202 value < 1.95e-137 chi-squared test) (**Fig 1B**). Therefore, CLAMP alters AS and has a stronger
203 effect on MXE splicing than other types of splicing.
204
205 During MXE splicing one isoform of the transcript retains one of the alternative exons and
206 excludes another exon, which is retained by another isoform (schematic **Fig 1A**). Interestingly,
207 MXE alternative splicing occurs in many transcripts that encode components of the sex
208 determination pathway[43]. Sex determination pathway components undergo MXE alternative
209 splicing that regulates translation and influences the splicing of downstream genes to drive
210 transcriptome variation and sexual dimorphism. Furthermore, CLAMP has a sex-specific role in
211 dosage compensation[21,44]. Therefore, we next asked whether CLAMP regulates sex-specific
212 splicing.
213
214 **2. Maternal CLAMP regulates sex-specific alternative splicing in early *Drosophila* embryos**
215
216 To determine whether CLAMP-dependent alternative splicing events are enriched for **s**ex-**s**pecific
217 **s**plicing (SSS) events, we first identified all of the transcripts undergoing female and male sex-
218 specific splicing (**Fig S2A, D**) (Supplementary **Method**). We measured alternative splicing using

219 an exon-centric approach to quantify individual splice junctions by measuring **p**ercent **s**pliced **i**n
220 (**PSI**) for a particular exon using the established SUPPA algorithm within the time2splice
221 pipeline[42]. Exon inclusion is represented as positive PSI, and exclusion events are defined as
222 negative PSI (equation in **Supplementary Method**). In this way, we identified sex-specific
223 splicing events (**Fig S2A, D**) and CLAMP-dependent splicing events (**Fig S2B-C, E-F**) in female
224 and male 0-2 Hr and 2-4 Hr embryos. We did not see any statistically significant enrichment for
225 MXE within the CLAMP-dependent sex-specifically spliced events (see Supplementary **Table**
226 **S1**). Although we found substantial evidence of CLAMP-dependent sex-specific splicing, we think
227 the gene list is likely to be too small to determine significance for a particular class of splicing
228 event.
229
230 Next, we measured the percentage of total alternatively spliced and sex-specifically spliced
231 transcripts that are CLAMP-dependent in males and females at both pre- and post-MZT stages.
232 While only 2-3% of total AS is CLAMP-dependent, ~30-60% of sex-specific splicing is CLAMP-
233 dependent (**Fig 2A**). Therefore, CLAMP regulates sex-specific splicing more strongly than total
234 AS. Overall, more genes show CLAMP-dependent splicing in females (~150-250) than males
235 (~100) (**Fig 2B and Supplementary Table S1**). However, the magnitude of the effect of CLAMP
236 on splicing was stronger in males than in females (**Fig S3**). We found that before MZT, the
237 CLAMP-dependent known male sex-specifically spliced genes have more exon inclusion events
238 than their female counterparts since ΔPSI values for males is significantly higher than in females
239 (**Fig S3**). This was further supported by validating a randomly selected set of sex-specific splicing
240 events we identified as described in more detail below (**Fig S4; Table S2**).
241
242 We then divided all CLAMP-dependent alternatively spliced events into two categories: 1) sex
243 specifically spliced (SSS) events; and 2) non-sex specifically spliced (non-SSS) events (**Fig 2B**).
244 We quantified CLAMP-dependent sex-specifically spliced events by comparing the change in PSI
245 (ΔPSI) for two classes of transcripts which we defined as follows: 1) **known** sex-specifically
246 spliced isoforms are different from each other in control samples ($p<0.05$) (**Fig. S2A, D**); 2) **new**
247 sex-specifically spliced isoforms are those not present in control samples (**Fig 2B**). We identified
248 widespread CLAMP-dependent sex-specific splicing, especially in female embryos (**Fig 2B**).
249 Interestingly, the majority of CLAMP-dependent SSS events are **new** sex-specific splicing events

250 that did not occur in the presence of maternal CLAMP (~70%) (**Fig 2C**). Furthermore, at the post-
251 MZT stage, both CLAMP-dependent exon inclusion and exclusion were significantly enriched in
252 male new SSS genes compared to their female-specific counterparts (**Fig S3**). Thus, in the absence
253 of CLAMP new aberrant sex-specific spliced isoforms are generated. Therefore, we hypothesized
254 that CLAMP normally inhibits aberrant sex-specific splicing events.
255
256 During the first few hours of their development, *Drosophila* embryos have predominantly maternal
257 transcripts. Therefore, we asked whether CLAMP-dependent female and male specifically-spliced
258 genes are maternally deposited or zygotically transcribed. We compared our list of CLAMP-
259 dependent sex-specifically spliced genes with known maternally expressed genes from previous
260 studies [45,46]. We found very low levels of overlap with maternally deposited transcripts (**Fig 2D**).
261 Therefore, most of the sex-specifically spliced genes we observed are likely to be zygotic
262 transcripts, consistent with the function of CLAMP as a pioneer TF in the early embryo [27].
263
264 Next, we divided our CLAMP-dependent sex-specifically spliced genes into female-specific or
265 male-specific categories. Here, it is important to note that many of these genes have multiple AS
266 events, and thus multiple isoforms regulated by CLAMP (**Table S1**). Then, we performed Gene
267 Ontology (GO) analysis to determine which biological processes are enriched within CLAMP-
268 dependent sex-specifically spliced genes at pre- and post-MZT time points using time2splice. At
269 the pre-MZT time point, female-specifically spliced genes are primarily mRNA regulatory factors
270 such as spliceosome components and TFs (**Fig 2E**). Therefore, in females CLAMP alters the
271 splicing of genes that can regulate the transcription and splicing of other genes to amplify its
272 regulatory role. In contrast, the male specifically-spliced genes pre-MZT are not enriched for any
273 specific biological function or process likely due to the small number of genes in the gene list. At
274 the post-MZT stage in both sexes, CLAMP regulates the splicing of genes that drive development
275 including organogenesis, morphogenesis, cell proliferation, signaling, and neurogenesis (**Fig 2E**).
276
277 We also randomly selected four genes for validation that we identified computationally as having
278 CLAMP-dependent splicing events of diverse types (**Fig S4**) using qRT-PCR or RT-PCR. These
279 genes perform diverse biological functions and provide new insights into how CLAMP regulates
280 early development (**Fig. 2E**). We have summarized the functions of the validated target genes at

281 which splicing is regulated by CLAMP (**Table S2**). *iab4,* one of the target genes that we validated,
282 has known functional links to CLAMP suggesting that we have identified relevant target
283 genes[36,47,48]. Furthermore, many genes that are sex-specifically spliced by CLAMP are themselves
284 involved in splicing and chromatin regulation (**Table S2**).
285
286 **3. CLAMP is highly enriched along gene bodies of sex-specifically spliced genes after MZT**
287
288 Because maternal CLAMP regulates sex-specific alternative splicing, we next asked: How does
289 CLAMP regulate female and male sex-specific splicing? If CLAMP directly regulates sex-specific
290 splicing, we hypothesized that it would directly bind to the genes that it regulates. Therefore, we
291 defined the binding pattern of CLAMP at the CLAMP-dependent female and male sex-specifically
292 spliced genes in sexed embryos using CLAMP ChIP-seq data (#GSE133637).
293
294 We generated average profiles for CLAMP occupancy at genes showing CLAMP-dependent
295 splicing in females and males at pre- and post-MZT time points in males and females (**Fig 3A,**
296 **B**). We found that CLAMP occupies the gene bodies of many sex-specifically spliced genes that
297 require CLAMP for their splicing. Specifically, 43.8% percent of all CLAMP-dependent sex-
298 specifically spliced genes are bound by CLAMP: 21.9% in 0-2 Hr female embryos, 8.2% in 0-2
299 Hr male embryos, 65.2% in 2-4 hr female embryos and 59.43% in 2-4 hr male embryos (**Table**
300 **S3**). The increase in percentage of genes bound in 2-4 Hr embryos compared with 0-2 Hr embryos
301 is consistent with an increased number of CLAMP binding sites at the later time point[48]. Overall,
302 these data are consistent with a direct role for CLAMP in regulating splicing of sex-specifically
303 spliced genes.
304
305 Next, we compared the average CLAMP binding pattern at sex-specifically spliced genes to the
306 CLAMP binding pattern at genes whose expression but not splicing is both sex-biased and
307 dependent on CLAMP. We defined CLAMP-dependent sex-biased genes as differentially
308 expressed genes in males in females that are dependent on CLAMP. At sex-specifically spliced
309 genes, CLAMP occupancy is present over gene bodies which is a dramatically different binding
310 profile from genes that are expressed but not spliced in a CLAMP-dependent and sex-biased
311 manner where CLAMP is enriched at the TSS and TES instead of at gene bodies (**Fig 3C and D**,

312 rectangular box in **Fig 3A-D**). Furthermore, CLAMP binding is also modestly enriched at the TSS
313 of female-biased expressed genes in females, consistent with enhanced CLAMP occupancy at the
314 TSS of expressed genes [33]. As a control, we used a random set of active genes that are not regulated
315 by CLAMP (green lines in **Fig 3A-D**) and we observed lower occupancy than at CLAMP-
316 dependent genes. Overall, we found preferential binding of CLAMP along the gene bodies of
317 genes that have CLAMP-dependent splicing in both females and males in contrast to TSS and TES
318 binding at genes where expression requires CLAMP.
319
320 To determine whether the binding of CLAMP to gene bodies occurs close to splice junctions, we
321 measured the distance of CLAMP peaks from the nearest splice junction (**Fig S5**). We found that
322 CLAMP peaks are most frequently within 200-400bp of either the start or the end of a splice
323 junction, especially in sex-specifically spliced genes. The resolution of these measurements is also
324 limited by sonication and therefore it is possible that binding occurs even closer to splice junctions.
325 We also found that CLAMP binds to chromatin closer to splice junctions at sex-specifically spliced
326 genes compared to sex-biased genes in 2-4 hr female embryo samples that have the most target
327 genes and CLAMP binding events. The results were similar for all CLAMP peaks (**Fig S5C**)
328 compared to peaks only present in introns (**Fig S5G**). We did not observe a difference in proximity
329 of CLAMP peaks to the nearest splice junction between sex-specifically spliced genes and sex-
330 biased genes in other samples but this negative result is not conclusive due to the decreased number
331 of peaks and target genes in the other samples. Together these data support a direct role for
332 CLAMP in co-transcriptional RNA processing that we hypothesize is due to direct contact with
333 target RNA transcripts and altering the recruitment of spliceosome components.
334
335 **4. CLAMP binds to RNA on chromatin of many sex-specifically spliced genes and is a**
336 **component of the mature spliceosome complex specifically in males**
337
338 To test our hypothesis that CLAMP directly regulates co-transcriptional RNA splicing by
339 contacting both the DNA and RNA of sex-specifically spliced genes and altering spliceosome
340 recruitment, we first asked whether to CLAMP directly binds to RNA. Although CLAMP do not
341 have a canonical **R**NA **r**ecognition **m**otifs (**RRM**), it has a prion-like intrinsically disordered
342 domain, present typically in many RNA binding proteins[49,50]. Using individual nucleotide

343 resolution UV crosslinked immunoprecipitation (iCLIP) which defines direct protein-RNA
344 interactions[51], we determined that most CLAMP RNA binds directly to hundreds of RNAs and
345 most targets are sex-specific with only 15% of the targets RNAs shared between males and females
346 (**Fig S6A, Table S5**). iCLIP was conducted on fractionated embryonic male and female cultured
347 cells because it was not possible to generate enough starting material from sexed embryos.
348 Interestingly, most CLAMP interaction with RNA occurs on chromatin (91.9% of male RNA
349 targets; 58.4% of female RNA targets) (**Fig S6A**). Even though iCLIP was conducted in embryos
350 cell lines and not embryos, we still found 47 target genes where CLAMP regulates sex-specific
351 splicing and interacts with both the DNA and RNA including but not only the key regulator of sex-
352 specific splicing *sxl*. Therefore, CLAMP sex-specifically and directly interacts with RNA targets
353 on chromatin including the RNA encoded by genes at which it regulates sex-specific splicing.
354
355 In addition to target RNAs, CLAMP also directly interacts with spliceosomal RNAs. Specifically,
356 we found that CLAMP binds to U1-U6 snRNAs (**Table S5**) more strongly in males (N=23)
357 compared to females (N=9) (**Fig S6B, C**). In the male chromatin fraction CLAMP interacts with
358 the catalytic step 2 spliceosome consisting of U2, U5, U6 snRNAs (FDR:1.7E-3). In contrast, the
359 female chromatin fraction is enriched for transcripts that encode proteins that bind to the U1-U2
360 snRNAs (FDR:1.1E-2) suggesting a different type of regulation of splicing in males and females.
361 Although these data establish that CLAMP interacts directly with some of its RNA targets to
362 regulate co-transcriptional splicing and interacts with the spliceosomes, it still does not explain
363 how CLAMP regulates splicing in a sex-specific manner.
364
365
366 **4. CLAMP interacts with components of the spliceosome complex and influences their**
367 **occupancy on chromatin**
368
369 CLAMP occupies the gene bodies of many CLAMP-dependent sex-specifically spliced genes
370 (**Fig 3**), physically associates with protein and RNA spliceosome components[36] , and directly
371 interacts with the RNA of a subset of target genes. Therefore, we hypothesized that CLAMP
372 regulates recruitment of components of the splicing machinery to chromatin differentially in males
373 and females. To test this hypothesis, we first examined how CLAMP regulates the occupancy of

374 the MLE RNA helicase, a component of both the MSL complex and the spliceosome[30,32,52-54] for
375 two reasons: 1) MLE is a component of both the MSL complex which is present only in males
376 (ref) and the spliceosome in both sexes[54] and 2) CLAMP physically associates with MLE[28,55,56].
377 Therefore, we hypothesized that CLAMP regulates the distribution of MLE between the
378 spliceosome and MSL complex to modulate sex-specific alternative splicing in males.
379
380 To determine whether CLAMP modulates MLE distribution on chromatin, we measured MLE
381 distribution at the genomic level using CUT&RUN[57,58]. We performed CUT&RUN in the presence
382 (*MTD-GAL4>UAS-GFPRNAi*) and absence (*MTD-GAL4>UAS-CLAMP [val22] RNAi*) of
383 maternal CLAMP at the pre-MZT and post-MZT embryonic stages in males. Next, we identified
384 MLE peaks in the presence or absence of maternally deposited CLAMP. MLE binds to chromatin
385 both in males and females, with stronger binding in males (**Fig 4A**). The absence of CLAMP
386 results in loss of MLE male peaks in males compared to no change in female MLE peaks (**Fig 4A**).
387 This supports our hypothesis that CLAMP regulates MLE recruitment to chromatin in males.
388
389 We next compared the distribution and location of MLE peaks with that of CLAMP peaks
390 previously identified in control embryos at the same time points[21] and classified MLE peaks into
391 two groups: 1) MLE peaks overlapping with CLAMP peaks (**Fig 4B,C and Fig S7**) and 2) unique
392 MLE peaks which do not overlap with CLAMP peaks (**Fig 4B,C and Fig S7**) at both 0-2 Hr pre-
393 MZT and 2-4 Hr post-MZT stages. Also, independent of developmental stage, MLE peaks which
394 overlap with CLAMP peaks are largely at promoters with fewer peaks localizing to introns. In
395 contrast, unique MLE peaks that do not overlap with CLAMP peaks are primarily localized to
396 introns (**Fig S7**). This distribution is similar for X-chromosomal and autosomal peaks.
397
398 In the absence of CLAMP, there is in a considerable loss and redistribution of both overlapping
399 and non-overlapping MLE peaks. We found that overall ~60% of MLE peaks were lost at the 2-4
400 Hr post-MZT stage in the absence of CLAMP. Moreover, 26% (pre-MZT) and ~35% (post-MZT)
401 of the MLE peaks observed in the absence of CLAMP were new and not present in control embryos
402 (**Fig 4B-C**). After the loss of maternal CLAMP, ~23% of MLE peaks overlapping with CLAMP
403 are also lost at the pre-MZT stage, which increases to ~51% at the post-MZT stage (**Fig 4B-C**).
404 Overall, our data suggest that MLE is redistributed in the absence of CLAMP suggesting that

405 CLAMP normally prevents aberrant recruitment of MLE in addition to the formation of aberrant
406 splice isoforms (**Fig 2**). Furthermore, we hypothesize that MLE at the new peaks is part of the
407 spliceosome complex and not MSL complex because MSL complex is not present on chromatin
408 in the absence of CLAMP[33].
409
410 To provide insight into the differences between MLE peaks which overlap with CLAMP and those
411 which do not, we identified sequence motifs which are enriched within each class of peaks using
412 MEME within the time2splice pipeline. The known CLAMP motif [33], a stretch of (GA)n repeats,
413 is enriched at regions that are bound by both MLE and CLAMP independent of stage and
414 chromosome type as expected. In contrast, MLE peaks which do not overlap with CLAMP have
415 motifs with stretches of GTs, GCTs, and GTAs but not (GA)n repeats (**Fig S7**). In the absence of
416 CLAMP, the remaining MLE peaks (red circle) were most enriched for (GT)n motifs (**Fig S7C,**
417 **D**) which have known roles in splicing through forming secondary RNA structures[59-61]. Therefore,
418 CLAMP prevents MLE from redistributing to sequence motifs that are known regulators of
419 splicing.
420
421 We also found that CLAMP alters the distribution of MLE from introns to promoters. MLE peaks
422 that overlap with CLAMP (intersection between green circle with red and grey circles, **Fig S7**) on
423 the X-chromosome (**Fig S7A, C**) or autosomes (**Fig S7B, D)** are enriched at promoters (**Fig S7A,**
424 **C**(X-chromosome), **Fig S7B, D** (Autosomes). In contrast, **new** unique MLE peaks not overlapping
425 with CLAMP (grey area in Venn diagrams, **Fig S7**) and those that are gained after CLAMP RNAi
426 (red area in Venn diagrams, **Fig S7**) are enriched at introns (**Fig S7A, C** (X-chromosome), **Fig**
427 **S7B, D** (Autosomes). These results support a role for CLAMP in sequestering MLE at (GA)n rich
428 sequences within promoters that prevents it from binding to GT motifs within introns that are
429 known regulators of splicing[59-61]. Thus, in the absence of CLAMP, MLE is redistributed and
430 aberrantly binds to intronic sequences with known motifs that regulate splicing irrespective of
431 whether present on X-chromosome or autosomes.
432
433 To determine how MLE redistribution could alter sex-specific splicing, we plotted the distribution
434 of MLE binding on CLAMP-dependent female and male-specifically spliced genes in the presence
435 and absence of CLAMP (**Fig S8A, B**). Pre-MZT, MLE binds near the TSS of male-specifically

436 spliced genes independent of maternal CLAMP (**Fig S7A**). At the post-MZT stage, loss of
437 maternal CLAMP in male embryos causes MLE to change its binding distribution along the gene
438 body (rectangle with dotted lines: **Fig S8B**) of CLAMP-dependent male-specifically spliced genes
439 (blue line) relative to CLAMP-dependent female-specifically spliced genes (red line). These
440 profiles are consistent with a model in which CLAMP is essential for MLE distribution at male-
441 specifically spliced genes which regulates male sex-specific splicing. In males, increased MLE
442 binding at female sex-specifically spliced genes (red line, enclosed within rectangle with dotted
443 lines: **Fig S8B**) in absence of CLAMP may result in differential splicing of these genes. Thus, we
444 hypothesize that CLAMP inhibits mis-localization of MLE to female sex-specifically spliced
445 genes in males.
446
447 Next, we asked whether CLAMP associates with spliceosome complex protein components other
448 than MLE, which is also a component of the MSL complex. We have shown that CLAMP directly
449 binds to snRNAs (**Fig S6**) and previously reported that CLAMP physically associates with several
450 spliceosome complex components based on mass spectrometry analysis[36]. To validate these
451 associations, we performed co-immunoprecipitation (coIP) experiments to assess association
452 between CLAMP and two spliceosome components with known functions in sex-specific splicing,
453 the conserved hrb27C and Squid proteins [54,62,63]. We found that in both S2 (male) and Kc (female)
454 cells, CLAMP interacts with hrb27C (**Fig S9A, B**). In contrast, CLAMP only associates with Squid
455 in female Kc cells and not in male S2 cells (**Fig S9A, B**), consistent with mass spectrometry data.
456 In contrast to MLE and CLAMP which are enriched on the male X-chromosome, Squid occupancy
457 on polytene chromosomes is decreased on the male X chromosome compared with the female X
458 chromosome (**Fig S9C-E**). Therefore, it is possible that there is a competition between CLAMP
459 recruitment of MSL complex to the male X chromosome and CLAMP recruitment of the
460 spliceosome complex containing Squid that contributes to sex-specific splicing.
461
462 Overall, CLAMP differentially associates with spliceosome components in males and females,
463 providing a potential mechanism by which CLAMP can regulate sex-specific splicing.
464 Specifically, it is possible that CLAMP influences the distribution of spliceosome components like
465 MLE between different protein complexes such as MSL complex and the spliceosome itself. We
466 hypothesize that differential binding of CLAMP to components of spliceosome like Squid in

467 females and MLE in males confers sex-specific recruitment of the spliceosome, thus generating
468 transcriptome diversity between female and male animals.

469 **5. CLAMP regulates the chromatin accessibility and splicing of the *sxl* gene and directly**
470 **interacts with the *sxl* RNA and alters splicing of other sex determination pathway component**
471 **genes**
472

473 In *Drosophila***,** sex-specific alternative splicing is regulated by the sex-determination pathway.
474 Sex-lethal (Sxl) is the master regulator of sex determination[15] and drives subsequent sex-specific
475 splicing in females[64]. Therefore, we asked whether CLAMP regulates *sxl* gene splicing and Sxl
476 protein levels in addition to directly regulating sex-specific splicing by binding to the DNA and
477 RNA of target genes as we have demonstrated (**Fig 3, S6 and Table S3, S5**).
478

479 Functional Sxl protein is produced specifically in females because exon three in the *sxl* transcript
480 contains a premature stop codon which is spliced out in females but retained in males [14]. Therefore,
481 the Sxl protein is non-functional in males and becomes degraded[15,65]. In females, Sxl protein
482 inhibits the splicing of the *msl-2* transcript which prevents aberrant MSL-2 expression and MSL
483 complex formation in females (**Fig 5A**). Furthermore, Sxl promotes female specific-splicing of the
484 downstream effector called transformer (*tra*). The female-specific Transformer protein regulates
485 splicing of another downstream transcript doublesex (*dsx*), which is sex-specifically spliced. In
486 contrast, the male-specific *transformer* transcript is not translated into functional Transformer
487 protein in males. Therefore, in the absence of any functional male-specific Transformer protein,
488 the *doublesex* transcript undergoes male-specific splicing (**Fig 5A**). The sex-specific isoforms of
489 the Doublesex protein drive sex-specific splicing of many downstream genes, shaping sexual
490 variation between males and females [66,67].
491

492 Because CLAMP regulates sex-specific splicing and is maternally deposited, we hypothesized that
493 CLAMP regulates the canonical sex determination pathway described above. To test this
494 hypothesis, we used our previously described recessive *clamp* null mutant $clamp^2$ line [44], the
495 heterozygous mutant $clamp^2/CyO\text{-}GFP$ line, and our previously reported rescue line which is
496 homozygous for the $clamp^2$ allele and contains an insertion of the wild type *CLAMP* gene. Sxl is
497 strongly regulated via alternative splicing of its mRNA[68-70]. Therefore, we designed an RT-PCR
498 assay to distinguish between the female-specific (excluding exon 3) and male-specific (including
499 exon 3) versions of the *sxl* transcript (**Fig 5B**). We used this assay to measure CLAMP-dependent
500 changes in alternative splicing and found that in homozygous $clamp^2$ female animals, there is a

501 small but detectable amount of the longer male-specific *sxl* transcript (**Fig 5B**, lane c). This mis-
502 regulation of *sxl* splicing is rescued by our *CLAMP* containing rescue construct (**Fig 5B**, lane d).
503 Our iCLIP data shows that CLAMP directly binds to *sxl* transcripts in females and not males
504 (**Table S5**) emphasizing CLAMP's direct role in *sxl* splicing in females.
505
506 To test whether defects in splicing altered Sxl protein levels, we performed western blots to
507 quantify Sxl protein in wild type females and males and *clamp*^{2}null females (**Fig 5C**). We
508 observed a reduction in Sxl protein levels in females in the *clamp*^{2} null background when compared
509 with controls. Also, homozygous *clamp*^{2} mutant males die before the third instar larval stage, and
510 therefore it was not possible to measure the splicing of transcripts in male *clamp*^{2} mutant larvae.
511 Overall, we determined that CLAMP promotes female-specific splicing of the *sxl* transcript to
512 ensure that normal Sxl protein levels are produced.
513
514 To determine a possible mechanism by which CLAMP regulates splicing of *sxl*, we mined our
515 previously generated Micrococcal Nuclease sequencing data[26], which measures chromatin
516 accessibility in S2(male) and Kc(female) cells in the presence and absence of CLAMP. Regions
517 of the genome that are accessible have a positive MACC score and are shown in blue and regions
518 of the genome that are inaccessible have a negative score and are shown in red (range is between
519 -0.33(red) and +1.33 (blue). We found that after the loss of CLAMP in female Kc cells, chromatin
520 accessibility at exon 3 of *sxl* increases significantly (**Fig 5D**). As shown in the Fig 5D boxed
521 rectangular inset, *sxl* exon3 shows a strong and statistically significant peak[26] indicative of open
522 chromatin specifically in females in which CLAMP has been depleted compared to control
523 females. Therefore, CLAMP normally promotes a closed chromatin environment at exon 3 in
524 females but not males.
525
526 Our results suggest that increased chromatin accessibility in males compared to females results in
527 retention of exon3 in the male *sxl* transcript. Consistent with our results, recent reports provide
528 strong evidence that increased chromatin accessibility contributes substantially to the retention of
529 introns during AS[71]. In addition, splicing-associated chromatin signatures have recently been
530 identified[72]. Consistent with our results in males, open chromatin marks such as H3K4me1&
531 H3K4me2 are enriched just upstream of the start site of retained exons[72]. In contrast, histone marks

532 associated with condensed chromatin such as H4K20me1&2, H3K9me3, and H3K27me3 are
533 highly enriched at excluded exons[72], consistent with our results in females.
534
535 Next, we examined the splicing of other components of the sex determination pathway for defects
536 in alternative splicing in the absence of CLAMP for which we were able to design effective RT-
537 PCR primers (**Fig 5E-F**). In embryos which lack CLAMP (**Fig 5E**, lane 2), the *dsx* female-specific
538 transcript is aberrantly produced in males (**Fig 5E**, lanes 4 and 5). In contrast, the male-specific
539 *dsx* transcript is not expressed in male embryos which lack CLAMP, similar to wild type female
540 embryos (**Fig 5E**, lane7-10). We also observed male-specific *dsx* transcripts in female *clamp*^{2}
541 mutant larvae (**Fig 5E**, lane c). Therefore, *dsx* splicing is regulated by CLAMP and CLAMP binds
542 to the *dsx* gene (**Fig S10**). However, *dsx* splicing is also regulated by Sxl whose protein levels are
543 regulated by CLAMP (**Fig 5C**).
544
545
546 Furthermore, we found that CLAMP regulates splicing of the male-specific lethal-2 (*msl-2*)
547 transcript, which is present only in males because Sxl regulates its splicing and stability. In the
548 *clamp*^{2} mutants: 1) levels of the nonfunctional female *msl-2* isoform (**Fig 5F**, lane 5) increased in
549 male embryos (**Fig 5F**, lane 2); 2) levels of the functional male-specific *msl-2* isoform increased
550 in females (**Fig 5F**, lane c). To determine whether these splicing defects also cause dysregulation
551 of MSL-2 protein expression and localization, we performed polytene immunostaining from
552 female *clamp*^{2} mutant salivary glands. In the absence of CLAMP, ectopic MSL2 protein (in red) is
553 present at several locations on female chromatin compared with controls (*clamp*^{2}*/CyOGFP*
554 heterozygous females) where MSL-2 protein is not present on chromatin (**Fig 5G**). Similar to *dsx*,
555 the gene is also bound by CLAMP (**Fig S11**) and regulated by Sxl and therefore could be regulated
556 through both direct and indirect mechanisms. Since CLAMP does not binds to the *dsx* or *msl-2*
557 transcripts directly it is possible that it indirectly regulates *dsx* and *msl-2* splicing through Sxl.
558 CLAMP does binds to the RNA that encodes another sex-specific splicing regulator and Sxl target
559 gene *fru* (Fruitless) in both females and males (**Table S5**). Together, these data reveal that CLAMP
560 regulates the splicing and protein expression of multiple components of the sex determination
561 pathway.
562

563 To further define how CLAMP and Sxl function together to regulate sex-specific splicing, we
564 asked whether CLAMP nuclear RNA targets in males and females have known Sxl binding motifs
565 (**Fig S11A**). Therefore, we compared CLAMP RNA targets from female embryonic cell lines with
566 available data identifying Sxl RNA targets in adult female heads (#**GSE98187**). Even though the
567 data sets are from different cell types, we still identified 61 overlapping targets in females (**Fig**
568 **S11B, Table S6**) which include the *sxl* transcripts and snRNAU5, a component of the U5snRNP
569 complex involved in splicing. Interestingly, 153 CLAMP RNA targets from males overlapped with
570 Sxl female RNA targets, indicating that CLAMP may function in males at genes that are Sxl targets
571 in females (**Fig S11B, C Table S6**). Overall, we found ~20% of CLAMP RNA targets have Sxl
572 motifs and similarly ~25% of CLAMP targets are Sxl targets as well. Also, 55 (19.6%) of the
573 CLAMP-dependent female sex-specific spliced genes are Sxl targets (**Table S7**). These results
574 further support our hypothesis that CLAMP functions together with Sxl to regulate sex-specific
575 splicing both via directly interacting with the DNA and RNA of sex-specifically spliced genes and
576 indirectly through regulating Sxl.
577
578
579 **Discussion:**
580
581 Alternative splicing (AS) is a highly conserved mechanism that generates transcript and protein
582 diversity[3,73,74]. Several studies have reported highly dynamic RNA bound proteomes (RBPs)
583 during the Maternal Zygotic Transition (MZT) across diverse phyla, with widespread alternative
584 splicing events occurring during early embryonic development[1,2,39,41,46]. Furthermore, different
585 isoforms are present in maternal and zygotic transcripts[45,46]. However, the mechanisms that
586 regulate transcript diversity during the early hours of an organism's life remain elusive.
587
588 Maternally-deposited pioneer transcription factors drive zygotic genome activation, but their role
589 in generating transcription diversity in the early embryo was unknown. Here, we define sex-
590 specific alternative spliced isoforms in pre- and post- MZT *Drosophila melanogaster* female and
591 male embryos genome-wide for the first time. Furthermore, we identify a maternally-deposited
592 pioneer TF, CLAMP, which regulates sex-specific transcript diversity in early embryos. Prior work
593 on sex-specific transcript diversity[24,37-39,41,62,75-78] either examined sex-biased differences in gene

594 expression or sex-specific transcript diversity much later in development in adult gonads or brain.
595 To overcome the challenge of sexing early embryos before zygotic genome activation, we used a
596 meiotic drive system that generates sperm with either X or only Y chromosomes[9] and measured
597 sex specific transcript diversity caused by alternative splicing.
598
599 Splice variants occur at different frequencies over time and between sexes. To date, we lack
600 pipelines to characterize these how isoforms change over time. Therefore, we developed
601 time2splice, which identifies mechanisms to regulate temporal and sex-specific alternative
602 splicing by combining RNA-seq and protein-DNA interaction data from CUT&RUN and ChIP-
603 seq experiments. Time2splice has three parts: 1) temporal splicing analysis based on the SUPPA
604 algorithm; 2) temporal protein-DNA analysis, and 3) temporal multi-omics integration. The
605 pipeline and analysis steps can be accessed at https://github.com/ashleymaeconard/time2splice.
606
607 Even following the initial few hours of its existence, there is a clear difference between a male and
608 female *Drosophila* embryo's transcript variation that was not previously identified (**Fig1, 2**).
609 Because the RNA transcript variants in both males and females encode genes involved in
610 developmental processes, sex-specific developmental distinctions may occur earlier than
611 previously thought. We demonstrate that a fundamental developmental trajectory differs between
612 males and females from the initial hours of their existence long before gonad formation. Such early
613 sex-specific transcript diversity provides insight into how developmental disorders that originate
614 before gonad formation can exhibit variable penetrance between sexes.
615
616 We defined groups of genes in both males and females that undergo alternative splicing events
617 which are regulated by maternally deposited CLAMP. Thus, the maternal environment regulates
618 both transcription initiation and shapes RNA processing. The key question is: How does CLAMP,
619 a ubiquitously expressed pioneer TF, regulate sex-specific splicing? We identified several
620 mechanisms by which CLAMP regulates sex-specific splicing.
621
622 CLAMP binds directly to intronic regions of approximately half of the sex-specifically spliced
623 genes that it regulates in both males and females. CLAMP also directly binds to different RNA
624 transcripts on chromatin in males and females and associates with U1-U6 snRNAs providing a

625 mechanism by which it regulates, co-transcriptional splicing. Therefore, our data supports a model
626 in which direct CLAMP binding to DNA and RNA regulates splicing of its target genes.
627
628 Furthermore, CLAMP regulates chromatin as a pioneer TF[26,48] and recent literature links
629 chromatin and splicing[71,72]. For example, closed chromatin marks have recently been linked to
630 exon exclusion and open chromatin has been linked to exon inclusion[71,72]. Proteomic analysis [36]
631 and coIPs (**Fig S9 A,B**) show that CLAMP is associated with spliceosome complex components,
632 including Squid and Hrb27C, further supporting a direct role for CLAMP in splicing. Interestingly,
633 CLAMP associates with Squid specifically in females. In contrast to Squid, MLE, a component of
634 both the spliceosomes and MSL complex[32] physically interacts with CLAMP only in males.
635 Because Squid is already known to regulate sex-specific splicing[62], these data support a model in
636 which differential association between CLAMP and spliceosome complex components in males
637 and females regulates sex-specific splicing. Our results also indicate that CLAMP is present in
638 functional spliceosome complex in males but not in females (**Table S5**). Thus, we hypothesize that
639 CLAMP may recruit spliceosome complex components to regulate splicing by altering the
640 chromatin environment or/and directly binding to target RNA transcripts (**Fig 6A**).
641
642 Based on our results, we hypothesize that there are two separate mechanisms by which CLAMP
643 can regulate splicing in males and females (**Fig 6).** We show that CLAMP inhibits aberrant splicing
644 events in males, especially at the post-MZT stage (**Fig 2**). In males, we found that MLE distribution
645 on chromatin is CLAMP dependent (**Fig 4A**) Also, MLE colocalizes with CLAMP at promoter
646 regions but in contrast, MLE alone is more often bound to intronic sequences, a prerequisite for
647 its splicing function, (**Fig 6B**). However, in the absence of CLAMP, the promoter bound MLE is
648 reduced, and MLE re-localizes from its normal intronic binding sites to new intronic regions that
649 contain GT sequence motifs (**Fig S7**) that are known to regulate splicing[59-61]. Therefore, we
650 hypothesize that CLAMP regulates the localization of MLE to prevent the aberrant formation of
651 female-specific transcripts in males that we observed in the absence of CLAMP.
652 In males, another function of CLAMP is to recruit the MSL ribonucleoprotein complex to the X
653 chromosome to regulate dosage compensation. CLAMP physically interacts directly with MSL-
654 2[28] and associates with MLE, an RNA helicase that acts on the *roX* non-coding RNA components
655 of MSL complex[31,44,52,55,79]. Furthermore, RNA transcript splicing is regulated by the spliceosome,

656 another ribonucleoprotein complex, consisting of many protein and RNA components[54]. Because
657 MLE is part of the MSL complex only in males and the spliceosome complex in both sexes, we
658 hypothesize that CLAMP influences the relative distribution of MLE between the MSL complex
659 and the spliceosome to co-regulate sex-specific splicing and male dosage compensation (**Fig 6B**).
660 Without CLAMP, the MSL complex does not localize to the X-chromosome and becomes
661 destabilized [33]; thus, MLE is no longer part of the MSL complex and is available to redistribute to
662 new spliceosome binding sites. Therefore, we provide evidence to support a model in which
663 CLAMP sex-specifically inhibits aberrant binding of MLE to motifs that regulate splicing which
664 alters sex-specific transcript diversity.
665
666 To provide mechanistic insight into how a pioneer factor like CLAMP regulates sex-specific
667 splicing, we also examined the splicing of sex determination pathway components. In females but
668 not in males, CLAMP binds near the early promoter of the *sxl* gene (SxlPe) and regulates the
669 chromatin environment at exon 3 of *sxl* which is normally spliced out in females (**Fig 6C**). Also,
670 only in females and not in males CLAMP directly binds to *sxl* transcript. Consistent with recent
671 literature[71,72], we hypothesize that closed chromatin at exon 3 induces exclusion of this exon from
672 female *sxl* transcripts whereas more open chromatin in males result in its inclusion in the male *sxl*
673 transcript. In the absence of CLAMP in females, the chromatin becomes more open, and *sxl*
674 transcript is not bound by CLAMP. Therefore, exon 3 is included in *sxl* transcripts which prevents
675 translation due to the incorporation of a stop codon. In the absence of CLAMP, exon 3 is retained
676 in females similar to wild type males, resulting in the formation of untranslatable *sxl* transcript in
677 females, significantly reducing levels of Sxl protein, thus dysregulating downstream splicing
678 events (**Fig 6C**). Because CLAMP binding sites are present near the promoter region of the *sxl*
679 gene, we hypothesize that CLAMP regulates chromatin at exon 3 from a distance consistent with
680 literature. CLAMP can mediate long-range chromatin interactions [80,81] and act on chromatin
681 accessibility at a distance [26].
682
683 In *Drosophila*, the master regulator Sxl initiates the sex-specific splicing cascade very early in
684 development to drive sex determination [14,20,69]. Decreased Sxl protein levels in female *clamp*^{2}
685 mutants and mis-expression of female and male-specific *dsx* transcripts suggest that CLAMP may
686 regulate sexual differentiation because sex-specific Dsx protein isoforms are known determinants

687 of sexual dimorphism[15]. Also, CLAMP directly binds to the DNA of *sxl, dsx,* and *msl-2* target
688 genes. Furthermore, CLAMP binds to the DNA and RNA of *sxl* and *fru*. Fruitless (*fru*) encodes a
689 BTB zinc finger transcription factor that contributes to sexual differentiation of the neural circuits
690 [82,83] and many of the CLAMP dependent sex-specifically spliced genes belong to neural
691 development pathway (**Fig 2E**). Since CLAMP and Sxl have both overlapping and distinct targets,
692 we hypothesize that CLAMP regulates sex-specific splicing both via the Sxl-mediated sex-
693 determination pathway as well as independent of it.
694
695
696 Therefore, CLAMP may regulate splicing through at least two possible mechanisms that are not
697 mutually exclusive: 1) CLAMP directly regulates the splicing of many sex-specifically spliced
698 genes by linking the RNA to the DNA and altering the recruitment of the spliceosome; 2) CLAMP
699 regulates the sex-specific of genes such as *msl-2* and *dsx* indirectly by regulating Sxl which is a
700 known regulator of their splicing and thus downstream splicing events.
701
702
703 Overall, we hypothesize that both different composition of the spliceosome and differential
704 recruitment to chromatin drive sex-specific changes in splicing. We identify CLAMP as a maternal
705 factor that regulates sex-specific alternative splicing through its sex-biased association with the
706 DNA and RNA of target genes and recruitment of spliceosome components and its ability to
707 influence the sex determination pathway. Identifying the factors that regulate this sex-biased
708 association of CLAMP with spliceosome complex components will be a key future direction.
709
710 Here, we show for the first time that a maternal factor controls sex-specific splicing during early
711 embryonic development, highlighting how the maternal environment influences transcript
712 diversity in the zygote from activation of the zygotic genome to the processing of zygotic RNA
713 products. Consistent with recent literature linking chromatin accessibility and splicing, our results
714 suggest that CLAMP could be one example of a more general splicing regulatory mechanism
715 controlled by the interaction between pioneer TFs that alter chromatin accessibility and
716 components of the RNA processing machinery to generate spatial-temporal transcript diversity.
717 While we analyzed sex-specific transcriptome diversity in this study and linked it to the sex-

718 specific dosage compensation process, similar mechanisms could drive cell-type specific variation.
719 For example, cell fate-determining transcription factors could regulate the chromatin occupancy
720 of splicing complex components to promote the formation of cell-type-specific isoforms. We also
721 present time2splice, a new pipeline to uncover mechanisms which drive such spatial-temporal
722 transcript diversity by integrating splicing and chromatin occupancy data.
723
724 **Materials and Methods:**
725
726 **Fly strains and rearing**
727 *Drosophila melanogaster* fly stocks were maintained at 24°C on standard corn flour sucrose
728 media. Fly strains used: *MTD-GAL4* (Bloomington, #31777), *UAS-CLAMPRNAi[val22]*
729 (Bloomington, #57008), Meiotic drive fly stocks +; SD72/CyO and 19-3, yw, Rsp[s]-
730 B[s]/Dp(2:y)CB25-4, y+, Rsp[s]B[s]; SPSD/CyO (Bloomington, #64332) (both gifts from
731 Cynthia Staber). These were crossed to obtained male and female embryo of desired genotypes
732 according to Rieder et al 2017.
733
734 **Sample collection and Western blotting**
735 Salivary glands from third instar larvae were dissected in cold PBS and samples frozen in liquid
736 nitrogen. Total protein from the samples was extracted by homogenizing tissue in the lysis buffer
737 (50mM Tris-HCl pH 8.0, 150mM NaCl, 1% SDS, 0.5X protease inhibitor) using a small pestle.
738 After a five-minute incubation at room temperature, cleared the samples by centrifuging at room
739 temperature for 10 minutes at 14,000xg. To blot for CLAMP and Actin, 5 micrograms of total
740 protein was run on a Novex 10% Tris-Glycine precast gel (Life technologies). To measure Sex-
741 lethal protein levels, 20 micrograms of total protein was run on a Novex 12% Tris-Glycine precast
742 gel (Life technologies). Protein was transferred to PVDF membranes using the iBlot transfer
743 system (ThermoFisher Scientific) and probed the membranes for CLAMP (1:1000, SDIX), Actin
744 (1:400,000, Millipore), and SXL (1:500, a gift from Fatima Gebauer) antibodies using the Western
745 Breeze kit following the manufacturer's protocol (ThermoFisher Scientific). We quantified the
746 relative expression of protein for SXL using the gel analysis tool in ImageJ software following the
747 website's guidelines [84]. For each genotype, we first internally normalized the amount of SXL

748 protein to Actin. Next, we determined the protein's relative expression by comparing the Actin
749 normalized quantities to y[1], w[1118] female samples.
750
751 **Polytene chromosome squashes and immunostaining**
752 Polytene chromosome squashes were prepared as previously described in Reider et al. 2017. We
753 stained polytene chromosomes with rabbit anti-CLAMP (1:1000, SDIX), mouse anti-Squid (1:50,
754 1B11, DSHB), rabbit anti-MLE (1:1000, gift from Mitzi Kuroda), rat anti-MSL2 (1:500, gift from
755 Peter Becker) antibodies. For detection, we used all Alexa Fluor secondary antibodies against
756 rabbit and mouse at a concentration of 1:1000 and visualized slides at 40X on a Zeiss Axioimager
757 M1 Epifluorescence upright microscope with the AxioVision version 4.8.2 software.
758
759 **Splicing assays for male and female-specific transcripts**
760 To test for the male and female splice forms of *sex-lethal, transformer, doublesex*, and *msl2*, total
761 RNA was extracted from ten third instar larvae from each genotype. We reverse-transcribed two
762 micrograms of total RNA using the SuperScript VILO cDNA Synthesis Kit (Life Technologies)
763 following the manufacturer's protocol. We amplified target sequences by PCR using primers
764 designed to span Alternatively spliced junctions. Alternative splicing primer sequences for sxl FP-
765 TGCAACTCACCTCATCATCC, sxl RP- GATGGCAGAGAATGGGACAT, for tra FP-
766 TGAAAATGGATGCCGACAG, tra RP- CTCTTTGGCGCAATCTTCTC, for dsx female
767 transcript dsxFFP-CTATCCTTGGGAGCTGATGC, dsxF RP-
768 TCGGGGCAAAGTAGTATTCG, for dsx male transcript dsxM FP-
769 CAGACGCCAACATTGAAGAG, dsxM RP- CTGGAGTCGGTGGACAAATC, for msl2 FP-
770 GTCACACTGGCTTCGCTCAG and msl2 RP- CCTGGGCTAGTTACCTGCAA were used.
771
772 **Immunoprecipitation**
773 *Nuclear extract preparation:* Male (S2) and female (Kc) cells were grown to a cell concentration
774 of 2X 10^6 cells/mL in T25 tissue culture flasks. Cells were scraped from the flask, centrifuged for
775 5min at 2500rpm at 4°C. Supernatant was removed and cell pellets were washed twice in 5ml of
776 cold PBS. The washed cell pellets were then resuspended in 5X volume of Buffer A (10mM
777 HEPES pH7.9, 1.5mM $MgCl_2$, 10mM KCl, 0.5mMDTT, 1X Protease inhibitors). Cells were
778 incubated on ice for 15 minutes before dounce homogenization with an A pestle. Cytoplasmic

779 fraction was collected after centrifugation at 4°C for 20 min at 700xg. The remaining nuclear pellet
780 was re-suspended in 3 times volume in Buffer B (20mM HEPES pH7.9, 20%Glycerol, 0.5%NP
781 40, 200mMKCl, 0.5mMEDTA, 1mMEGTA, 1X protease inhibitors). Nuclei after re-suspension
782 were dounce homogenized with a B pestle. Nuclear debris was then pelleted by centrifugation at
783 10,000xg for 10 min at 4°C. 1ml aliquots of both cytoplasmic and nuclear fractions were prepared
784 in 1.5mL Protein LoBind Eppendorf tubes and flash frozen in liquid nitrogen for storage at -80 °C.
785
786 *Immunoprecipitation:* Magnetic anti-CLAMP beads were prepared to a final concentration of
787 10mg/mL by coupling rabbit anti-CLAMP antibody (SDIX) to magnetic beads, according to
788 Dynabeads Antibody coupling kit (ThermoFisher Scientific) instructions. Both prepared anti-
789 CLAMP and purchased anti-IgG (anti-rabbit IgG M-280 Dynabeads) were blocked to reduce
790 background the night before the immunoprecipitation. First, the beads were washed 3 times for 5
791 minutes in 500L Tris-NaCl Wash (50mM Tris, 500mM NaCl, 0.1% NP-40) by rotating at 4C. The
792 beads were next suspended in block buffer (3.3mg/mL of yeast tRNA extract prepared in 20mM
793 HEPES, pH7.9, 20% Glycerol, 0.5% NP-40, 200mM KCl, 1mM EDTA, and 2mM EGTA) and
794 rotated overnight at 4C. The next day, beads were washed 3 times for 5 minutes in the block buffer
795 without yeast tRNA by rotating at 4°C. After the final wash, beads were resuspended in the same
796 amount of block buffer as the starting volume.
797
798 To 1mL of previously prepared nuclear extract, 100uL of blocked anti-CLAMP or anti-IgG
799 magnetic Dynabeads were added. The nuclear extracts and beads were then rotated for 1 hour at
800 4°C. Afterward, the beads were collected and the supernatant discarded. The beads were then
801 washed three times in Tris-NaCl wash (50mM Tris, 500mM NaCl, 0.1% NP-40) by rotating for 5
802 minutes at 4°C and cleared by using a magnetic rack. To elute proteins from the beads, 100uL of
803 1% SDS was added, and the beads were boiled for 10 minutes at 95C. To the eluate, 300uL of
804 ultrapure water was added, and the tubes gently vortexed. After collecting the beads on a magnetic
805 rack, the eluate was saved in a clean Protein LoBind Eppendorf tube.
806
807 *Western blotting:* Squid and Hrb27C were detected in IP-CLAMP and IGG-rabbit protein samples
808 using mouse anti-Squid (1:500, 1B11, DSHB) and rabbit anti-Hrb27C (1:5000, Fatima Gebauer),
809 performed as mentioned above under western blotting protocol.

810

811 **CUT&RUN**

812 0-2 hr and 2-4 hr male and female embryos of desired genotypes (~50 each) were collected on
813 standard grape juice agar medium and washed with water. The embryos were dechorionated in 6%
814 bleaching solution for 2 min and washed twice in ice cold 1XPBS. Centrifuged at 12,000g for 10
815 min at 4°C. Supernatants were discarded and embryos resuspended in 200µl Dig-Wash buffer with
816 EDTA (20mM HEPES-NaOH, 150mM NaCl, 2mM EDTA, 0.5mM Spermidine, 10mM PMSF,
817 0.05% digitonin) and washed twice. Embryos were incubated in 200µl primary antibody overnight
818 at 4°C on a tube rotator. Next, embryos were centrifuged at 12,000g for 10 min at 4°C and liquid
819 removed and embryos were washed twice in Dig-Wash buffer with EDTA. Then, embryos were
820 incubated for 3 hours at 4°C in ~700 ng/ml pAMNase solution in Dig-Wash buffer with EDTA.
821 Embryos were washed twice in Dig-Wash buffer without EDTA and resuspended in 150µl of Dig-
822 Wash buffer without EDTA. Samples were equilibrated to 0°C on a heat block maintained on ice-
823 bath. 2µl of 100mm $CaCl_2$ added to each sample to initiate MNase activity and digestion was
824 performed for 30 min before adding 150µl of 2X RSTOP Buffer (200mM NaCl, 20mM EDTA,
825 4mM EGTA, 50ug/ml RNase, 40ug/ml glycogen, 10pg/ml yeast spike-in DNA) to stop the
826 reaction. Incubated at 37°C for 10 minutes to release the DNA fragments. Samples were spun at
827 12,000g for 10 minutes and aqueous layer transferred to a fresh 1.5 ml microfuge tube and
828 centrifuged at 16,000g for 5 minutes. Cleared liquid was again transferred to a fresh tube, 1µl of
829 20% SDS and 2.5µl proteinase K (20ng/ml) added, incubated at 70°C for 10 minutes. 300µl PCI
830 was added to each tube, mixed and total solution was transferred to phase lock tubes and
831 centrifuged at 16,000g for 5 minutes. After adding 300µl of chloroform and mixing gently, samples
832 were centrifuged at 16,000g for 5 minutes at RT. The aqueous layer was transferred to a DNA low
833 binding tube. 1µl glycogen (5mg/ml) and 750µl ethanol added to precipitate DNA at -80°C.
834 Samples were centrifuged at 16,000g for 10 min at 4°C and washed in ethanol twice. Pellet air
835 dried and dissolved in 15µl of 1mM TrisHCl + 0.1mM EDTA pH 8.0 [57,58]. 1ng of Cut and Run
836 DNA was used to make libraries using the KAPA Hyper prep kit and SeqCap adapters A &B
837 (Roche) according to manufacturer's protocol. For library amplification 14 cycles were used and
838 a 1.0X SPRI bead cleanup was performed using Agencourt Ampure XP beads. The following

839 antibody concentrations were used: rabbit anti-CLAMP (5µg/sample, SDIX); 1:200 anti-rabbit
840 (MilliporeSigma); rat anti-MLE (1:50, 6E11); 700ng/ml pA-MNase (from Steven Henikoff).
841
842 **iCLIP**

843 Kc and S2 cells were maintained at 25°C in Schneider's media supplemented with 10% Fetal
844 Bovine Serum and 1.4X Antibiotic-Antimycotic. Cells were passaged every 3 days to maintain an
845 appropriate cell density. Cells were allowed to grow to confluency and UV crosslinked using
846 254nm UV light in Stratalinker 2400 on ice (Stratagene, USA). UV treated cells were lysed to get
847 different cellular fractions (Cytoplasmic, Nucleoplasmic and Chromatin) according to Fr-iCLIP
848 (fractionation-iCLIP) protocol from Brugiolo et al 2017[85]. Chromatin and Nucleoplasmic fractions
849 were sonicated with a Branson digital sonicator at 30% amplitude for 30 s total (10 sec on and 20
850 sec off) to disrupt DNA before IP. All three fractions were separately centrifuged at 20,000 xg for
851 5 min at 4°C. Fractions were tested by Western blotting using RNApolI for Chromatin Fraction,
852 Actin for Cytoplasmic Fraction. Protein quantification for each fraction was done using
853 manufacturer's protocol for Pierce 660nm protein assay reagent (Thermo Scientific, USA). Each
854 Fraction was subjected to iCLIP protocol as described in Huppertz et al 2014[51] using rabbit-
855 CLAMP antibody to immuniprecipitate bound RNAs which were extracted using proteinase K and
856 phenol:chloroform. Custom cDNA libraries prepared according to Huppertz et al 2014[51] using
857 distinct primers Rt1clip-Rt16clip for separate samples containing individual 4nt-barcode
858 sequences that allow multiplexing of samples. cDNA libraries for each sample amplified
859 separately using 31 cycles of PCR, mixed together later and sequenced using standard illumina
860 protocols. Heyl et al. 2020[86] methods using the Galaxy CLIP-Explorer were followed to
861 preprocess, perform quality control, post-process and perform peak calling.

862 **Computational Methods:**

863 **Time2splice tool: see Supplementary methods**
864
865 **Sex-specific splicing event analysis**

866 RNA sequencing data from Rieder et al 2017 (#GSE102922) was analyzed using time2splice to
867 determine sex-specifically splicing events dmel-all-r6.29.gtf from BDGP6 in genomes [87] was used
868 to map each transcript identifier (ID) to gene ID and symbol, for .bed creation data for the
869 associated chromosome, transcription start site (TSS) and transcription end site (TES), and strand
870 information were imported from Illumina
871 (https://support.illumina.com/sequencing/sequencing_software/igenome.html). From the raw data
872 after quality control i.e, FastQC[88], Salmon[89] was used to quantify transcript expression for
873 treatment and control samples. Calculated transcripts per million (TPM) values from SUPPA[42]
874 were used for all four replicates of female and male controls at both time points (before and after
875 MZT). Each sample was filtered to include transcripts where the mean value is less than or equal
876 to 3 TPMs per gene. The number of transcripts included at various thresholds were plotted from 1
877 to 10 and the fraction of genes filtered out begins to plateau around threshold 3. The percent of
878 spliced in (PSI) transcripts between females and males were compared at both 0-2 Hr (pre-MZT)
879 and 2-4 Hr (post-MZT) (p-value of 0.05), thereby resulting in delta PSI values and p-values for
880 each transcription in each experimental condition comparison. Given these resulting delta
881 transcript PSI values, significantly alternatively splice genes (p-value 0.05) were found between
882 females vs. males 0-2 Hr (pre-MZT) controls to show which genes are normally sex-specifically
883 spliced pre-MZT. The same process was followed at 2-4 Hr (post-MZT). To then determine the
884 sex-specifically spliced genes, the female RNAi experiment compared with the control delta PSI
885 gave the number of total alternative spliced transcripts pre-MZT, then considering those that are
886 not shared with males, and are only expressed in females normally, this defined our sex specifically
887 spliced set of genes for females pre-MST. This process was also performed for males pre-MZT.
888 The same process was followed for post-MZT samples.
889
890 **ChIP-seq: Data analysis**
891
892 We used preprocessed ChIP-seq data from Rieder et al 2019 (#GSE133637), specifically the .bw
893 and .broadPeak.gz files in our analysis using ChIPseeker [90] and deeptools [91]. Specifically, when
894 plotting the average profiles using deeptools, we achieved a baseline signal representing genome-
895 wide binding taking into consideration the number of genes in other groups by the following
896 procedure: of all genes that are on (no zero read-count genes), we sampled the number of the

897 largest other group (to which we are comparing), and ran computeMatrix on that subset. This
898 process was repeated 500 times and the resulting 500 matrices were averaged to produce a
899 representative signal. For motif analysis MEME [92] suite was used.
900
901 **Cut and Run: Data analysis**
902
903 Sequenced reads were run through FASTQC[88](fastqc replicate_R1_001.fastq.gz
904 replicate_R2_001.fastq.gz) with default parameters to check the quality of raw sequence data and
905 filter out any sequences flagged for poor quality. Sequences were trimmed and reassessed for
906 quality using TrimGalore (https://github.com/FelixKrueger/TrimGalore/issues/25) and FastQC [88],
907 respectively. All Illumina lanes of the same flow cell .fastq files were merged, and sequenced reads
908 were then mapped to release 6 *Drosophila melanogaster* genome (dm6). We compared Bowtie2[93],
909 HISAT2[94], and BWA[95]. We found the best alignment quality with BWA and thus used this
910 method's results downstream. Next, we performed conversion to bam and sorting (e.g. using:
911 bowtie2 -x dm6_genome -1 replicate_R1_001.fastq.gz -2 replicate_R2_001.fastq.gz -S out.sam >
912 stout.txt 2> alignment_info.txt; samtools view -bS out.sam > out.bam; rm -rf out.sam; samtools
913 sort out.bam -o out.sorted.bam). We removed reads (using samtools) with a MAPQ less than 30
914 and any reads with PCR duplicate reads (identified using MarkDuplicates Picard -2.20.2). Peaks
915 identified using MACS2[96](macs2 callpeak -t out.sorted.bam -B -f BAM --nomodel --SPMR --
916 keep-dup all -g dm --trackline -n outname --cutoff-analysis --call-summits -p 0.01 --outdir outdir)
917 and keep duplicates separate**.** To calculate fold-enrichment macs2 is run again (macs2 bdgcmp -t
918 $treat -c $control -o $out.sorted.bam_FE.bdg -m FE 2> $ out.sorted.bam_FE.log; macs2 bdgcmp
919 -t $treat -c $control -o $out.sorted.bam_logLR.bdg -m logLR -p 0.00001 2). For motif analysis
920 MEME [92] suite was used. Data submitted in GEO repository (#GSE174781).
921
922 **iCLIP: Data analysis**
923
924 Heyl et al. 2020[86] methods using the Galaxy CLIP-Explorer were followed to preprocess, perform
925 quality control, post-process and perform peak calling. For preprocessing UMI-Tools was used,
926 and then UMI-tools and Cutadapt used for Adapter, Barcode and UMI-removal. Cutadapt (Galaxy
927 version 3.5) was used for filtering with a custom adapter sequence

928 AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCT
929 GCTTG. All other settings followed the Heyl et al 2020 Galaxy iCLIP-explorer workflow. UMI-
930 Tools Extract (Galaxy Version 1.1.2+galaxy2) was then used with a barcode pattern of
931 NNNXXXXNN. No unpaired reads were allowed. The barcode was on the 3' end. Je-Demultiplex
932 (Galaxy Version 1.2.1) was then used for demultiplexing. FastQC was used for quality control.
933 Mapping was done by RNA STAR (Galaxy version 2.5.2b-2) using dm6. All settings were chosen
934 based on the existing parameters from the iCLIP-explorer settings. We selected FALSE for the
935 option to use end-to-end read alignments with no soft-clipping. bedtools used for Read-Filtering,
936 and UMI-Tools (Galaxy version 0.5.3.0) for de-duplication. PEAKachu was used for Peak Calling
937 to generate bed files. The PEAKachu settings were followed using the Galaxy CLIP-explorer
938 workflow. The maximum insert size was set to 150, the minimum cluster expression fraction was
939 set to 0.01, the minimum block overlap set to 0.5, the minimum block expression set to 0.1. The
940 Mad Multiplier was set to 0.0, the Fold Change Threshold was set to 2.0, and the adjusted p-value
941 threshold was set to 0.05. Peaks were annotated using RCAS[97] (RNA Centric Annotation System),
942 a R package using Rstudio. MEME Suite used for motif detection. RCAS was used for functional
943 analysis of the transcriptomes isolated by iCLIP, such as transcript features. ShinyGO 0.76[98] was
944 used to perform Gene Ontology Analysis of the iCLIP data. Data submitted in GEO repository
945 (#GSE205987).
946
947 **Competing Interest Statement**
948
949 The authors declare no conflicting interests.
950
951 **Acknowledgments**
952
953 This work and funding to M.R. was supported by R35GM126994 to E.N.L. from NIH. A.M.C is
954 funded by the NSF Graduate Research Fellowship and CCMB, Brown University. We thank
955 Bloomington stock center for fly lines. We thank Leila Rieder for SD and Rsp^{s} stocks, Peter Becker
956 for MSL2 and MLE antibodies, Steve Henikoff for pAMNase protein and spike-in DNA for Cut
957 and Run, and Daniel J McKay for sharing his Cut and Run protocol for tissues.
958

959 **Author Contributions**

960

961 M.R., A.M.C. and E.N.L. planned experiments, analyzed results and wrote the manuscript. A.M.C
962 did all the computational analysis regarding time2splice pipeline. M.R. carried out the
963 experimental work and collected data for Cut and Run, iCLIP, Polytene squashes and IF, splicing
964 assays and IP. J.U. carried out sex determination pathway splicing assays and WB. JA analyzed
965 the MLE cut and run data. A.H analyzed the iCLIP-seq data.

966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989

990 **References**

991 1. Aanes, H. *et al.* Differential transcript isoform usage pre-and post-zygotic genome
992 activation in zebrafish. *BMC genomics* **14**, 331 (2013).
993 2. Revil, T., Gaffney, D., Dias, C., Majewski, J. & Jerome-Majewska, L.A. Alternative splicing
994 is frequent during early embryonic development in mouse. *BMC genomics* **11**, 399
995 (2010).
996 3. Mayne, B.T. *et al.* Large scale gene expression meta-analysis reveals tissue-specific, sex-
997 biased gene expression in humans. *Frontiers in genetics* **7**, 183 (2016).
998 4. Ober, C., Loisel, D.A. & Gilad, Y. Sex-specific genetic architecture of human disease.
999 *Nature Reviews Genetics* **9**, 911-922 (2008).
1000 5. Wang, G.-S. & Cooper, T.A. Splicing in disease: disruption of the splicing code and the
1001 decoding machinery. *Nature Reviews Genetics* **8**, 749-761 (2007).
1002 6. Faustino, N.A. & Cooper, T.A. Pre-mRNA splicing and human disease. *Genes &*
1003 *development* **17**, 419-437 (2003).
1004 7. Schulz, K.N. *et al.* Zelda is differentially required for chromatin accessibility, transcription
1005 factor binding, and gene expression in the early Drosophila embryo. *Genome research*
1006 **25**, 1715-1726 (2015).
1007 8. Schulz, K.N. & Harrison, M.M. Mechanisms regulating zygotic genome activation. *Nature*
1008 *Reviews Genetics* **20**, 221-234 (2019).
1009 9. Rieder, L.E. *et al.* Histone locus regulation by the Drosophila dosage compensation
1010 adaptor protein CLAMP. *Genes & development* **31**, 1494-1508 (2017).
1011 10. Artieri, C.G. & Fraser, H.B. Transcript length mediates developmental timing of gene
1012 expression across Drosophila. *Molecular biology and evolution* **31**, 2879-2889 (2014).
1013 11. De Renzis, S., Elemento, O., Tavazoie, S. & Wieschaus, E.F. Unmasking activation of the
1014 zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS Biol* **5**,
1015 e117 (2007).
1016 12. Guilgur, L.G. *et al.* Requirement for highly efficient pre-mRNA splicing during Drosophila
1017 early embryonic development. *Elife* **3**, e02181 (2014).
1018 13. Förch, P. & Valcárcel, J. Splicing regulation in Drosophila sex determination. in
1019 *Regulation of Alternative Splicing* 127-151 (Springer, 2003).
1020 14. Haussmann, I.U. *et al.* m 6 A potentiates Sxl alternative pre-mRNA splicing for robust
1021 Drosophila sex determination. *Nature* **540**, 301-304 (2016).
1022 15. Salz, H. & Erickson, J.W. Sex determination in Drosophila: The view from the top. *Fly* **4**,
1023 60-70 (2010).
1024 16. Estes, P.A., Keyes, L.N. & Schedl, P. Multiple response elements in the Sex-lethal early
1025 promoter ensure its female-specific expression pattern. *Molecular and Cellular Biology*
1026 **15**, 904-917 (1995).
1027 17. González, A.N., Lu, H. & Erickson, J.W. A shared enhancer controls a temporal switch
1028 between promoters during Drosophila primary sex determination. *Proceedings of the*
1029 *National Academy of Sciences* **105**, 18436-18441 (2008).
1030 18. Keyes, L.N., Cline, T.W. & Schedl, P. The primary sex determination signal of Drosophila
1031 acts at the level of transcription. *Cell* **68**, 933-943 (1992).

1032 19. Salz, H.K. Male or female? The answer depends on when you ask. *PLoS Biol* **5**, e335
1033 (2007).
1034 20. Penalva, L.O. & Sánchez, L. RNA binding protein sex-lethal (Sxl) and control of Drosophila
1035 sex determination and dosage compensation. *Microbiology and molecular biology*
1036 *reviews* **67**, 343-359 (2003).
1037 21. Rieder, L.E., Jordan III, W.T. & Larschan, E.N. Targeting of the dosage-compensated male
1038 X-chromosome during early Drosophila development. *Cell reports* **29**, 4268-4275. e2
1039 (2019).
1040 22. Prayitno, K., Schauer, T., Regnard, C. & Becker, P.B. Progressive dosage compensation
1041 during Drosophila embryogenesis is reflected by gene arrangement. *EMBO reports* **20**,
1042 e48138 (2019).
1043 23. Gebauer, F., Merendino, L., Hentze, M.W. & Valcárcel, J. The Drosophila splicing
1044 regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. *Rna* **4**,
1045 142-150 (1998).
1046 24. Alekseyenko, A.A. *et al.* A sequence motif within chromatin entry sites directs MSL
1047 establishment on the Drosophila X chromosome. *Cell* **134**, 599-609 (2008).
1048 25. Straub, T., Zabel, A., Gilfillan, G.D., Feller, C. & Becker, P.B. Different chromatin
1049 interfaces of the Drosophila dosage compensation complex revealed by high-shear ChIP-
1050 seq. *Genome research* **23**, 473-485 (2013).
1051 26. Urban, J. *et al.* Enhanced chromatin accessibility of the dosage compensated Drosophila
1052 male X-chromosome requires the CLAMP zinc finger protein. *PloS one* **12**, e0186855
1053 (2017).
1054 27. Duan, J.E. *et al.* CLAMP and Zelda function together as pioneer transcription factors to
1055 promote Drosophila zygotic genome activation. *bioRxiv* (2020).
1056 28. Albig, C. *et al.* Factor cooperation for chromosome discrimination in Drosophila. *Nucleic*
1057 *acids research* **47**, 1706-1724 (2019).
1058 29. Larschan, E. *et al.* X chromosome dosage compensation via enhanced transcriptional
1059 elongation in Drosophila. *Nature* **471**, 115-118 (2011).
1060 30. Kuroda, M.I., Kernan, M.J., Kreber, R., Ganetzky, B. & Baker, B.S. The maleless protein
1061 associates with the X chromosome to regulate dosage compensation in Drosophila. *Cell*
1062 **66**, 935-947 (1991).
1063 31. Ankush Jagtap, P.K. *et al.* Structure, dynamics and roX2-lncRNA binding of tandem
1064 double-stranded RNA binding domains dsRBD1, 2 of Drosophila helicase Maleless.
1065 *Nucleic acids research* **47**, 4319-4333 (2019).
1066 32. Cugusi, S., Kallappagoudar, S., Ling, H. & Lucchesi, J.C. The Drosophila helicase maleless
1067 (MLE) is implicated in functions distinct from its role in dosage compensation. *Molecular*
1068 *& Cellular Proteomics* **14**, 1478-1488 (2015).
1069 33. Soruco, M.M. *et al.* The CLAMP protein links the MSL complex to the X chromosome
1070 during Drosophila dosage compensation. *Genes & development* **27**, 1551-1556 (2013).
1071 34. Kaye, E.G. *et al.* Differential occupancy of two GA-binding proteins promotes targeting
1072 of the drosophila dosage compensation complex to the male X chromosome. *Cell*
1073 *reports* **22**, 3227-3239 (2018).
1074 35. Quinn, J.J. *et al.* Rapid evolutionary turnover underlies conserved lncRNA–genome
1075 interactions. *Genes & development* **30**, 191-207 (2016).

1076 36. Urban, J.A., Urban, J.M., Kuzu, G. & Larschan, E.N. The Drosophila CLAMP protein
1077 associates with diverse proteins on chromatin. *PloS one* **12**, e0189772 (2017).
1078 37. Lott, S.E., Villalta, J.E., Zhou, Q., Bachtrog, D. & Eisen, M.B. Sex-specific embryonic gene
1079 expression in species with newly evolved sex chromosomes. *PLoS Genet* **10**, e1004159
1080 (2014).
1081 38. Telonis-Scott, M., Kopp, A., Wayne, M.L., Nuzhdin, S.V. & McIntyre, L.M. Sex-specific
1082 splicing in Drosophila: widespread occurrence, tissue specificity and evolutionary
1083 conservation. *Genetics* **181**, 421-434 (2009).
1084 39. Paris, M., Villalta, J.E., Eisen, M.B. & Lott, S.E. Sex bias and maternal contribution to gene
1085 expression divergence in Drosophila blastoderm embryos. *PLoS Genet* **11**, e1005592
1086 (2015).
1087 40. Lott, S.E. *et al.* Noncanonical compensation of zygotic X transcription in early Drosophila
1088 melanogaster development revealed through single-embryo RNA-seq. *PLoS Biol* **9**,
1089 e1000590 (2011).
1090 41. Gibilisco, L., Zhou, Q., Mahajan, S. & Bachtrog, D. Alternative splicing within and
1091 between Drosophila species, sexes, tissues, and developmental stages. *PLoS genetics* **12**,
1092 e1006464 (2016).
1093 42. Trincado, J.L. *et al.* SUPPA2: fast, accurate, and uncertainty-aware differential splicing
1094 analysis across multiple conditions. *Genome biology* **19**, 1-11 (2018).
1095 43. Brooks, A.N. *et al.* Regulation of alternative splicing in Drosophila by 56 RNA binding
1096 proteins. *Genome research* **25**, 1771-1780 (2015).
1097 44. Urban, J.A. *et al.* The essential Drosophila CLAMP protein differentially regulates non-
1098 coding roX RNAs in male and females. *Chromosome Research* **25**, 101-113 (2017).
1099 45. Atallah, J. & Lott, S.E. Evolution of maternal and zygotic mRNA complements in the early
1100 Drosophila embryo. *PLoS genetics* **14**, e1007838 (2018).
1101 46. Kwasnieski, J.C., Orr-Weaver, T.L. & Bartel, D.P. Early genome activation in Drosophila is
1102 extensive with an initial tendency for aborted transcripts and retained introns. *Genome*
1103 *research* **29**, 1188-1197 (2019).
1104 47. Gutierrez-Perez, I. *et al.* Ecdysone-induced 3D chromatin reorganization involves active
1105 enhancers bound by Pipsqueak and Polycomb. *Cell reports* **28**, 2715-2727. e5 (2019).
1106 48. Duan, J.E. *et al.* CLAMP and Zelda function together to promote Drosophila zygotic
1107 genome activation. *elife* **10**, 2020.07. 15.205054 (2021).
1108 49. Harrison, A.F. & Shorter, J. RNA-binding proteins with prion-like domains in health and
1109 disease. *Biochemical Journal* **474**, 1417-1438 (2017).
1110 50. March, Z.M., King, O.D. & Shorter, J. Prion-like domains as epigenetic regulators,
1111 scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain*
1112 *research* **1647**, 9-18 (2016).
1113 51. Huppertz, I. *et al.* iCLIP: protein–RNA interactions at nucleotide resolution. *Methods* **65**,
1114 274-287 (2014).
1115 52. Cugusi, S., Li, Y., Jin, P. & Lucchesi, J.C. The Drosophila helicase MLE targets hairpin
1116 structures in genomic transcripts. *PLoS genetics* **12**, e1005761 (2016).
1117 53. Reenan, R.A., Hanrahan, C.J. & Ganetzky, B. The mlenapts RNA helicase mutation in
1118 Drosophila results in a splicing catastrophe of the para Na+ channel transcript in a
1119 region of RNA editing. *Neuron* **25**, 139-149 (2000).

1120 54. Herold, N. *et al.* Conservation of the protein composition and electron microscopy
1121 structure of Drosophila melanogaster and human spliceosomal complexes. *Molecular*
1122 *and cellular biology* **29**, 281-301 (2009).
1123 55. Lindehell, H., Kim, M. & Larsson, J. Proximity ligation assays of protein and RNA
1124 interactions in the male-specific lethal complex on Drosophila melanogaster polytene
1125 chromosomes. *Chromosoma* **124**, 385-395 (2015).
1126 56. Quinn, J.J. *et al.* Revealing long noncoding RNA architecture and functions using domain-
1127 specific chromatin isolation by RNA purification. *Nature biotechnology* **32**, 933-940
1128 (2014).
1129 57. Skene, P.J., Henikoff, J.G. & Henikoff, S. Targeted in situ genome-wide profiling with high
1130 efficiency for low cell numbers. *Nature protocols* **13**, 1006 (2018).
1131 58. Uyehara, C.M. & McKay, D.J. Direct and widespread role for the nuclear receptor EcR in
1132 mediating the response to ecdysone in Drosophila. *Proceedings of the National*
1133 *Academy of Sciences* **116**, 9893-9902 (2019).
1134 59. Gabellini, N. A polymorphic GT repeat from the human cardiac Na+ Ca2+ exchanger
1135 intron 2 activates splicing. *European Journal of Biochemistry* **268**, 1076-1083 (2001).
1136 60. Hefferon, T.W., Groman, J.D., Yurk, C.E. & Cutting, G.R. A variable dinucleotide repeat in
1137 the CFTR gene contributes to phenotype diversity by forming RNA secondary structures
1138 that alter splicing. *Proceedings of the National Academy of Sciences* **101**, 3504-3509
1139 (2004).
1140 61. Lin, C.-L. *et al.* RNA structure replaces the need for U2AF2 in splicing. *Genome research*
1141 **26**, 12-23 (2016).
1142 62. Hartmann, B. *et al.* Distinct regulatory programs establish widespread sex-specific
1143 alternative splicing in Drosophila melanogaster. *Rna* **17**, 453-468 (2011).
1144 63. Blanchette, M. *et al.* Genome-wide analysis of alternative pre-mRNA splicing and RNA-
1145 binding specificities of the Drosophila hnRNP A/B family members. *Molecular cell* **33**,
1146 438-449 (2009).
1147 64. Bell, L.R., Horabin, J.I., Schedl, P. & Cline, T.W. Positive autoregulation of sex-lethal by
1148 alternative splicing maintains the female determined state in Drosophila. *Cell* **65**, 229-
1149 239 (1991).
1150 65. Moschall, R. *et al.* Drosophila Sister-of-Sex-lethal reinforces a male-specific gene
1151 expression pattern by controlling Sex-lethal alternative splicing. *Nucleic acids research*
1152 **47**, 2276-2288 (2019).
1153 66. Coschigano, K.T. & Wensink, P.C. Sex-specific transcriptional regulation by the male and
1154 female doublesex proteins of Drosophila. *Genes & development* **7**, 42-54 (1993).
1155 67. Nagoshi, R.N. & Baker, B.S. Regulation of sex-specific RNA splicing at the Drosophila
1156 doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing
1157 patterns. *Genes & development* **4**, 89-97 (1990).
1158 68. Bell, L.R., Maine, E.M., Schedl, P. & Cline, T.W. Sex-lethal, a Drosophila sex
1159 determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to
1160 RNA binding proteins. *Cell* **55**, 1037-1046 (1988).
1161 69. Salz, H.K. *et al.* The Drosophila female-specific sex-determination gene, Sex-lethal, has
1162 stage-, tissue-, and sex-specific RNAs suggesting multiple modes of regulation. *Genes &*
1163 *development* **3**, 708-719 (1989).

1164 70. Samuels, M.E., Schedl, P. & Cline, T.W. The complex set of late transcripts from the
1165 Drosophila sex determination gene sex-lethal encodes multiple related polypeptides.
1166 *Molecular and cellular biology* **11**, 3584-3602 (1991).
1167 71. Petrova, V. *et al.* Chromatin accessibility regulates intron retention in a cell type-specific
1168 manner. *bioRxiv* (2021).
1169 72. Agirre, E., Oldfield, A., Bellora, N., Segelle, A. & Luco, R. Splicing-associated chromatin
1170 signatures: a combinatorial and position-dependent role for histone marks in splicing
1171 definition. *Nature communications* **12**, 1-16 (2021).
1172 73. Venables, J.P., Tazi, J. & Juge, F. Regulated functional alternative splicing in Drosophila.
1173 *Nucleic acids research* **40**, 1-10 (2012).
1174 74. Blencowe, B.J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37-47
1175 (2006).
1176 75. Ranz, J.M., Castillo-Davis, C.I., Meiklejohn, C.D. & Hartl, D.L. Sex-dependent gene
1177 expression and evolution of the Drosophila transcriptome. *Science* **300**, 1742-1745
1178 (2003).
1179 76. Zhang, Y., Sturgill, D., Parisi, M., Kumar, S. & Oliver, B. Constraint and turnover in sex-
1180 biased gene expression in the genus Drosophila. *Nature* **450**, 233-237 (2007).
1181 77. Sun, X. *et al.* Sxl-Dependent, tra/tra2-Independent Alternative Splicing of the Drosophila
1182 melanogaster X-Linked Gene found in neurons. *G3: Genes, Genomes, Genetics* **5**, 2865-
1183 2874 (2015).
1184 78. Arbeitman, M.N., Fleming, A.A., Siegal, M.L., Null, B.H. & Baker, B.S. A genomic analysis
1185 of Drosophila somatic sexual differentiation and its regulation. *Development* **131**, 2007-
1186 2021 (2004).
1187 79. Lv, M. *et al.* Structural insights reveal the specific recognition of roX RNA by the dsRNA-
1188 binding domains of the RNA helicase MLE and its indispensable role in dosage
1189 compensation in Drosophila. *Nucleic acids research* **47**, 3142-3157 (2019).
1190 80. Bag, I., Dale, R.K., Palmer, C. & Lei, E.P. The zinc-finger protein CLAMP promotes gypsy
1191 chromatin insulator function in Drosophila. *Journal of cell science* **132**(2019).
1192 81. Jordan, W. & Larschan, E. The zinc finger protein CLAMP promotes long-range chromatin
1193 interactions that mediate dosage compensation of the Drosophila male X-chromosome.
1194 *bioRxiv* (2020).
1195 82. Vernes, S.C. Genome wide identification of Fruitless targets suggests a role in
1196 upregulating genes important for neural circuit formation. *Scientific Reports* **4**, 4412
1197 (2014).
1198 83. Jai, Y.Y., Kanai, M.I., Demir, E., Jefferis, G.S. & Dickson, B.J. Cellular organization of the
1199 neural circuit that drives Drosophila courtship behavior. *Current biology* **20**, 1602-1614
1200 (2010).
1201 84. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image
1202 analysis. *Nature methods* **9**, 671-675 (2012).
1203 85. Brugiolo, M., Botti, V., Liu, N., Müller-McNicoll, M. & Neugebauer, K.M. Fractionation
1204 iCLIP detects persistent SR protein binding to conserved, retained introns in chromatin,
1205 nucleoplasm and cytoplasm. *Nucleic acids research* **45**, 10452-10465 (2017).
1206 86. Heyl, F., Maticzka, D., Uhl, M. & Backofen, R. Galaxy CLIP-Explorer: a web server for
1207 CLIP-Seq data analysis. *GigaScience* **9**, giaa108 (2020).

87. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491 (2011).
88. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, 2010).
89. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* **14**, 417-419 (2017).
90. 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).
91. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**, W187-W191 (2014).
92. Bailey, T.L., Johnson, J., Grant, C.E. & Noble, W.S. The MEME suite. *Nucleic acids research* **43**, W39-W49 (2015).
93. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357 (2012).
94. Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37**, 907-915 (2019).
95. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754-1760 (2009).
96. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, 1-9 (2008).
97. Uyar, B. *et al.* RCAS: an RNA centric annotation system for transcriptome-wide regions of interest. *Nucleic acids research* **45**, e91-e91 (2017).
98. Ge, S.X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628-2629 (2020).

1246 **Fig 1. Alternative splicing during early *Drosophila melanogaster* embryonic development**
1247 **A** Schematic diagram showing 7 different types of Alternative splicing (AS).
1248 **B** Pie chart showing the distribution of different types of AS at 0-2 Hr pre-MZT and 2-4 Hr post-
1249 MZT for female (red) and male (Blue) embryos in the presence (top row) and absence (bottom
1250 row) of maternal CLAMP. The total percentage of AS in each category in control embryos is noted
1251 at the top of the schematic. A Chi-square test was performed to determine if there is a significant
1252 difference between the percentage of each type of AS including MXE splicing (grey region) in the
1253 presence vs. absence of CLAMP in each class of sample: females/males 0-2 Hr, and 2-4 Hr
1254 embryos. A significant difference ($p<0.001$ marked by ***) was found between categories
1255 connected by solid black lines.
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276

1277

1278 **Fig 2. Maternal CLAMP regulates sex-specific alternative splicing during early embryonic**
1279 **development.**

1280 **A** Bar graph showing the percentage of transcripts (values noted at the top of each bar) out of total
1281 AS events or sex-specific splicing events (number of events noted within parentheses at the top of
1282 each bar) regulated by maternal CLAMP, at the pre-MZT and post-MZT stages, in females (red
1283 bars) and males (blue bars). A Fischer's Exact Test was performed, with significance at $p<0.001$.

1284 **B** Bar plot showing the total number of splicing events undergoing CLAMP-dependent AS (**N**) in
1285 females and males at 0-2 Hr pre-MZT and 2-4 Hr post-MZT embryonic stages. Spliced genes are
1286 divided into non-sex specific (grey) and sex-specific (orange shades) sub-categories of
1287 genes. CLAMP-dependent female and male sex-specifically spliced (SSS) genes are divided into
1288 known (darker orange) and new (lighter orange) sub-categories identified in 0-2 Hr pre-MZT and
1289 2-4 Hr post-MZT embryos.

1290 **C** Percentage of **new** female (red) and male (blue) CLAMP-dependent sex-specifically spliced
1291 genes in 0-2 Hr pre-MZT and 2-4 Hr post-MZT embryos that were not identified in control
1292 samples.

1293 **D** Male (blue) and female (brown) CLAMP-dependent sex-specific spliced genes compared with
1294 maternal genes (**N=841**, magenta) at 0-2 Hr pre-MZT (female, **N=153** and male, **N=99**) and 2-4
1295 Hr post-MZT stages (female, **N= 270** and male, **N=116**).

1296 **E** Gene ontology results for genes showing CLAMP-dependent female sex-specific splicing in the
1297 embryo at 0-2 Hr pre-MZT stage and for genes exhibiting CLAMP-dependent female as well as
1298 male sex-specific splicing in the embryo at the 2-4 Hr post-MZT stage. The size of the circle
1299 increases as the number of the genes in that category increases. The color of the circle represents
1300 significance (p-value).

1301

1302

1303

1304

1305

1306

1307

1308 **Fig 3. CLAMP binds along the gene body of female and male sex-specifically spliced genes**
1309 **at the post-MZT embryonic stage**
1310 **A-B** Average profiles for CLAMP binding at pre-MZT and post-MZT embryonic stages in females
1311 and males for genes spliced female-specifically (red line) and male-specifically (blue line) during
1312 the pre-MZT (**A**) and post-MZT (**B**) stages.
1313 **C-D** Average profiles for CLAMP binding to genes expressed in a sex-biased manner in females
1314 (red line) and males (blue line) during pre-MZT (**C**) and post-MZT (**D**) stage.
1315 Green lines in **A-D** represent CLAMP binding at a random set of active genes used as a control
1316 (see **Material and Methods** for details). Stippled regions in **A, B** (female, 0-2 Hr pre-MZT) denote
1317 chromatin around the TSS with more CLAMP binding in female sex-specifically spliced genes vs.
1318 male sex-specifically spliced genes. The dotted box in **A-D** marks the gene body region in sex-
1319 specifically spliced and biased expressed genes.
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338

1339 **Fig 4. CLAMP regulates the distribution of MLE on chromatin in males**

1340 **A-B.** Heat maps showing the distribution of MLE at the male (**A**) and female (**B**) specific control
1341 MLE peaks on the X chromosome and autosomes in male and female 0-2 Hr pre-MZT embryos
1342 with maternal CLAMP (*MTD-GAL4>GFPRNAi*) and after the loss of maternal CLAMP (*MTD-*
1343 *GAL4>CLAMPRNAi)*.

1344 **C-D.** Venn diagrams and bar plots showing loss and gain of MLE peaks in the presence and
1345 absence of maternal CLAMP in male 0-2 Hr pre-MZT (**B**) and 2-4 Hr post-MZT (**C**) embryos.
1346 CLAMP peaks were identified only under control conditions (green circle), whereas MLE peaks
1347 were identified in the presence (grey circle) and absence (red circle) of maternal CLAMP protein
1348 depleted using the *MTD-GAL4>CLAMPRNAi* system.

1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368

1369 **Fig 5. Alternative splicing of components of the sex determination pathway is regulated by**
1370 **CLAMP in females**
1371 **A** The sex determination pathway in *Drosophila* is regulated by master regulator SXL.
1372 **B** Electrophoresis gel image (inverted colors) showing splicing of *sxl* transcripts in third instar
1373 larvae of females and males of genotypes listed in the key (**a-g**) with a representative schematic at
1374 the top of the gel image.
1375 **C** Western blot showing the level of SXL protein in genotypes (3 replicates for each) mentioned
1376 below each lane. Tubulin levels were used for protein loading control. Below the blot is the relative
1377 density of SXL protein compared with Tubulin, with each genotype represented by separate
1378 colored bars.
1379 **D** Chromatin accessibility measured by the MNase Accessibility (MACC) score is shown across
1380 the *sxl* gene in male (S2) and female (Kc) cells under control and CLAMP RNAi conditions. The
1381 MACC score is a previously reported (Urban et al 2017) quantification of chromatin accessibility
1382 at each locus in the genome. Positive/high accessibility values (blue) indicate high chromatin
1383 accessibility, and negative/low (red) accessibility values indicate low chromatin accessibility.
1384 Each window covers MACC values ranging from -0.333 to +1.33. MACC values increase in
1385 females after CLAMP RNAi, specifically at exon 3 (red box) and are shown in the inset to the
1386 right. Green boxes represent CLAMP binding peaks in the *sxl* gene just below the schematic for
1387 the *sxl* gene itself.
1388 **E-F** Electrophoresis gel image from 0-2 Hr embryonic (lane 2-5 & 7-10) and third instar larval
1389 samples (a-g) showing splicing of *dsx* (E) and *msl2* (F) transcripts in females (lane 3,5,8,10, a-d)
1390 and males (lane 2,4,7,9, e-g). Embryos were from *MTD-GAL4>GFPRNAi* control (lane 4,5,9,10)
1391 and *MTD-GAL4>CLAMPRNAi* (lane 2,3,7,8) females. a-g genotypes are the same as in B. The
1392 schematic at the top of each gel image shows female and male splice variants of *dsx* (E) and *msl2*
1393 (F) transcripts.
1394 **G** Fluorescent microscopy images of polytene chromosomes from third instar salivary gland in the
1395 genotypes listed to the left of each panel show the distribution of CLAMP (green) and MSL2 (red)
1396 on chromatin (blue, DAPI)
1397
1398
1399

1400 **Fig 6. Mechanisms by which CLAMP regulates sex-specific splicing in females and males**
1401 **A** CLAMP regulates splicing in both males and females via directly binding to intronic DNA
1402 sequences of CLAMP-dependent sex-specifically spliced genes and interaction with some sex-
1403 specifically spliced genes and spliceosomal RNAs.
1404 **B** CLAMP regulates MLE distribution between the spliceosome and the male X-chromosome
1405 specific MSL complex in males. CLAMP increases the occupancy of MLE at promoters and CES.
1406 In the absence of CLAMP, MLE is lost from many sites, including CES and promoters, and is
1407 gained at new intronic sequences which contain motifs that regulate splicing, resulting in aberrant
1408 sex-specific splicing in males.
1409 **C** In females, CLAMP binds near the SxlPe promoter and regulates chromatin accessibility at exon
1410 three (blue square) of the *sxl* gene and to the *sxl* mRNA. In this way, CLAMP promotes the
1411 splicing out of exon3 such that functional Sxl protein is formed, which drives female-specific
1412 splicing events. The absence of CLAMP in females thus results in the aberrant production of non-
1413 functional male-specific *sxl* transcripts which retain exon3, reducing levels of functional Sxl
1414 protein. CLAMP and Sxl have shared and distinct RNA targets suggesting that they function by
1415 both dependent and independent mechanisms.
1416 The three mechanisms proposed in parts **A**, **B**, and **C** are not mutually exclusive and are likely to
1417 occur simultaneously.
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428

Fig 1. Ray, Conard et al 2022

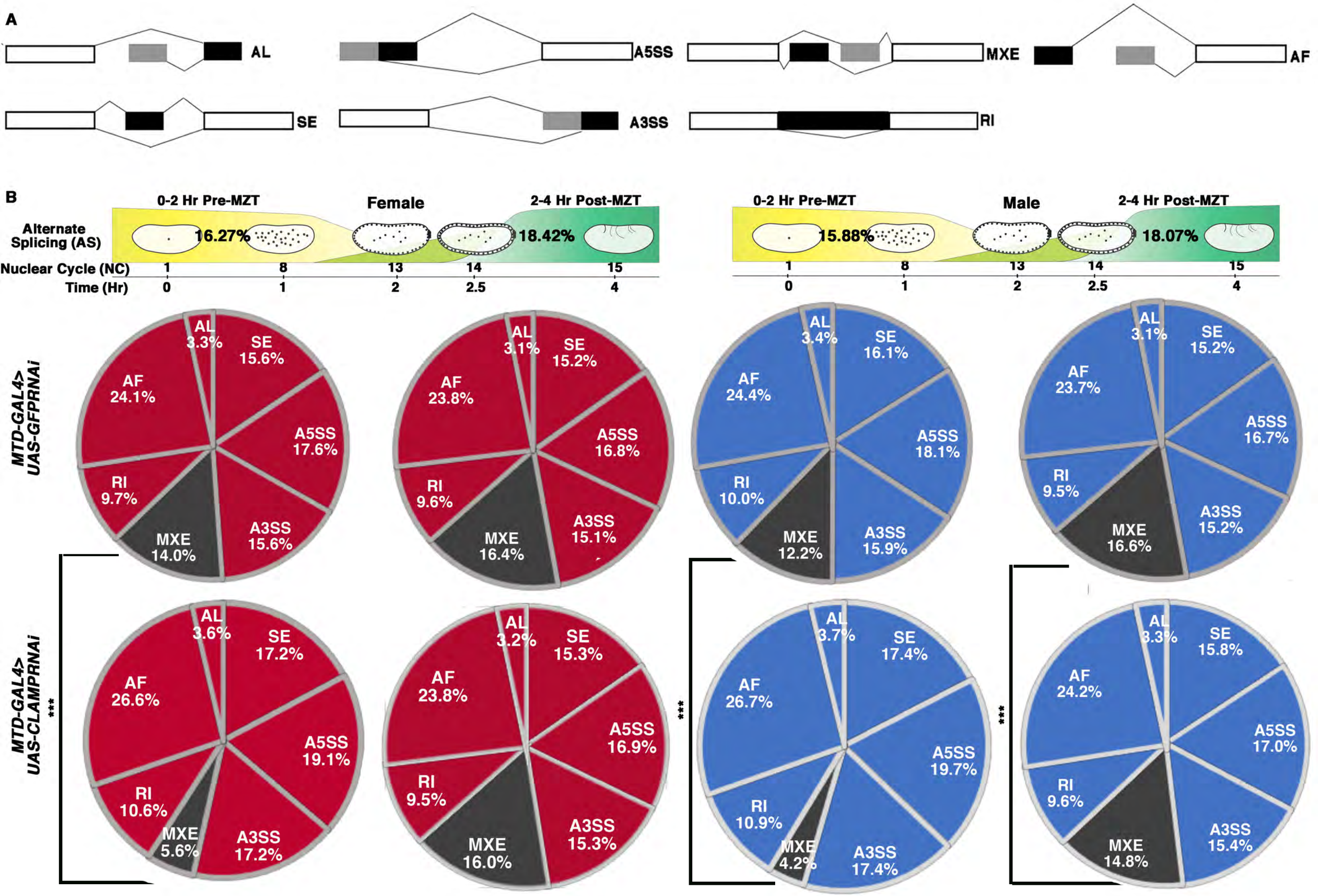

Fig 2. Ray, Conard et al 2022

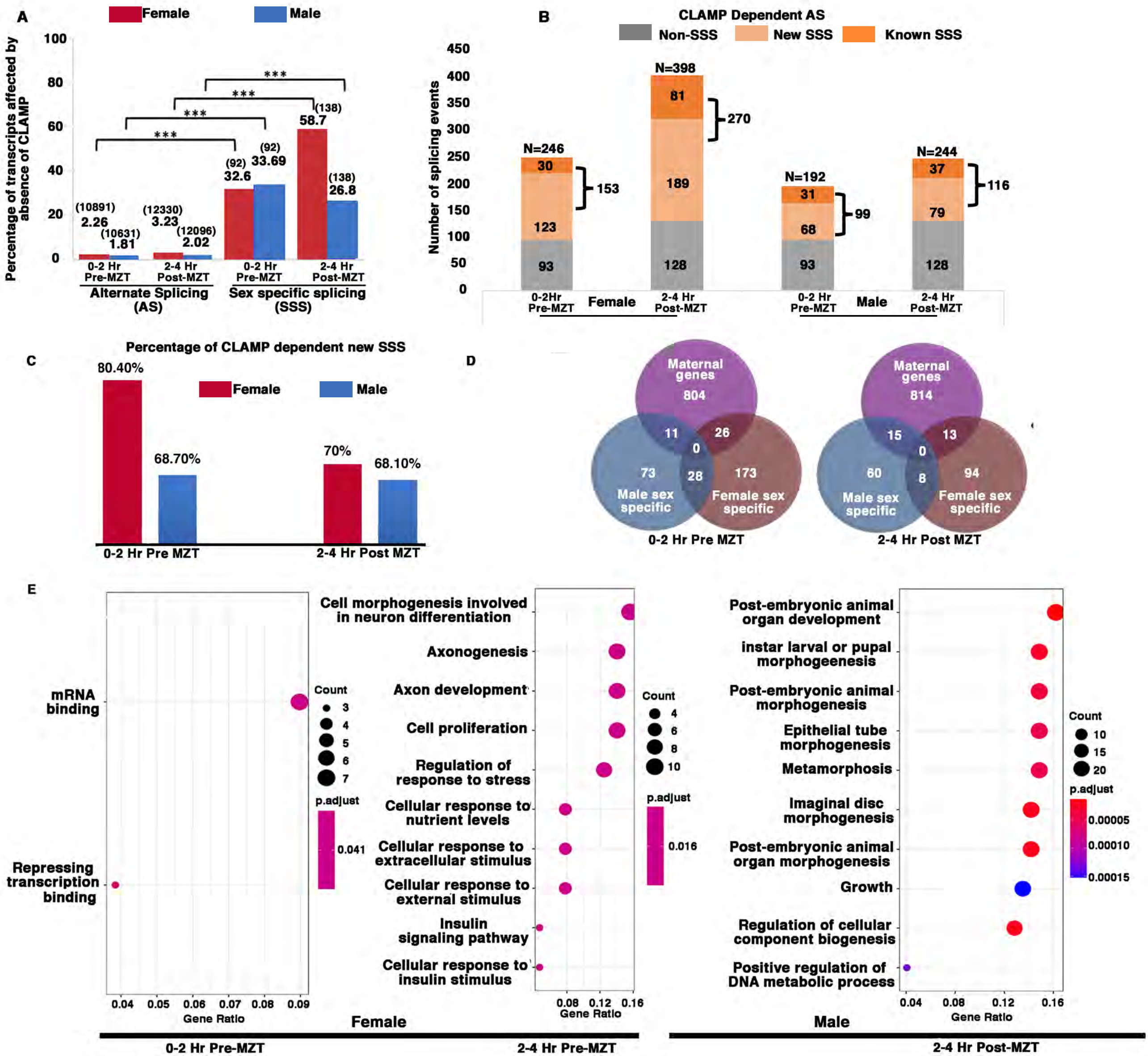

Fig 3. Ray, Conard et al 2022

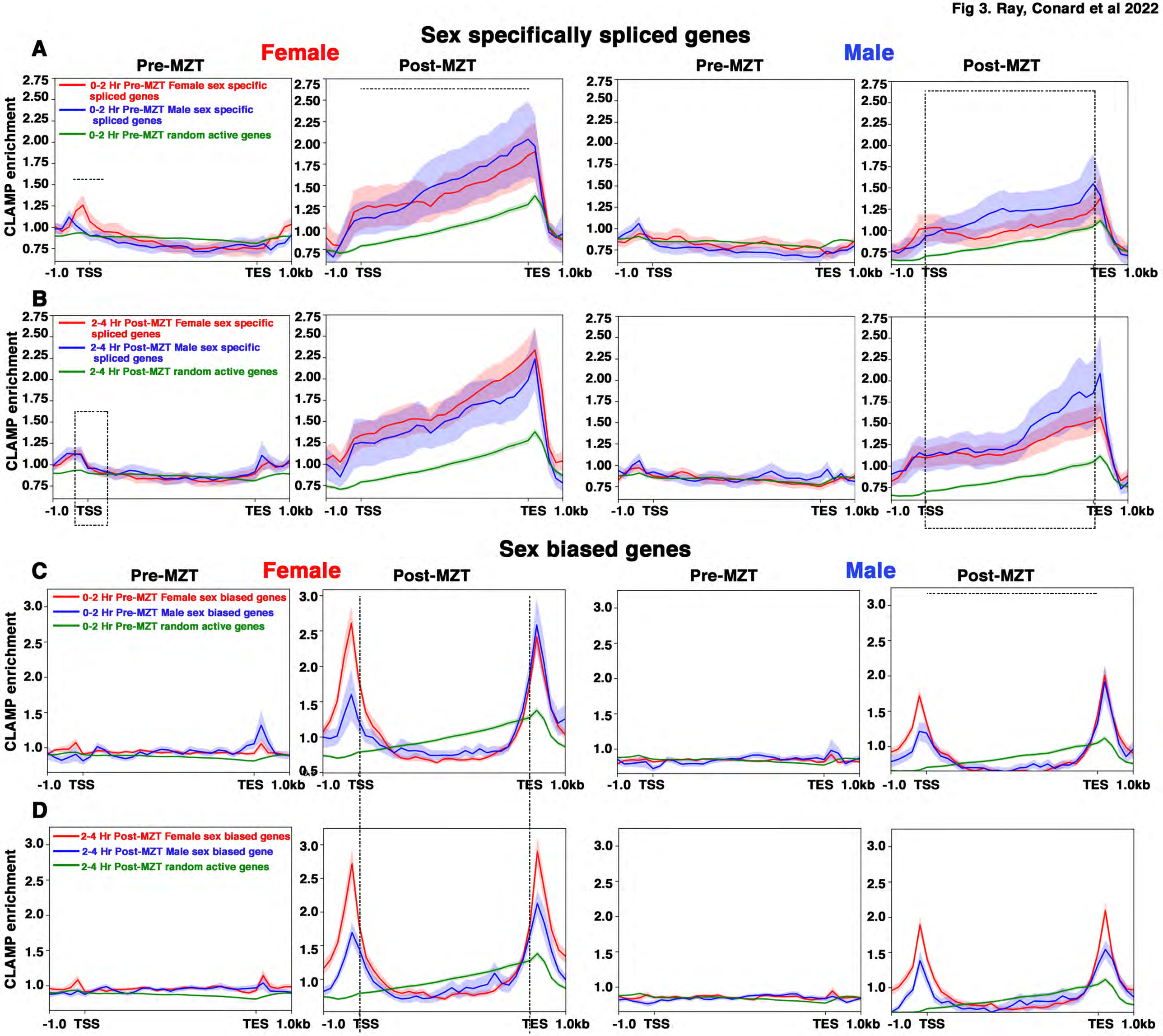

Fig 4. Ray, Conard et al 2022

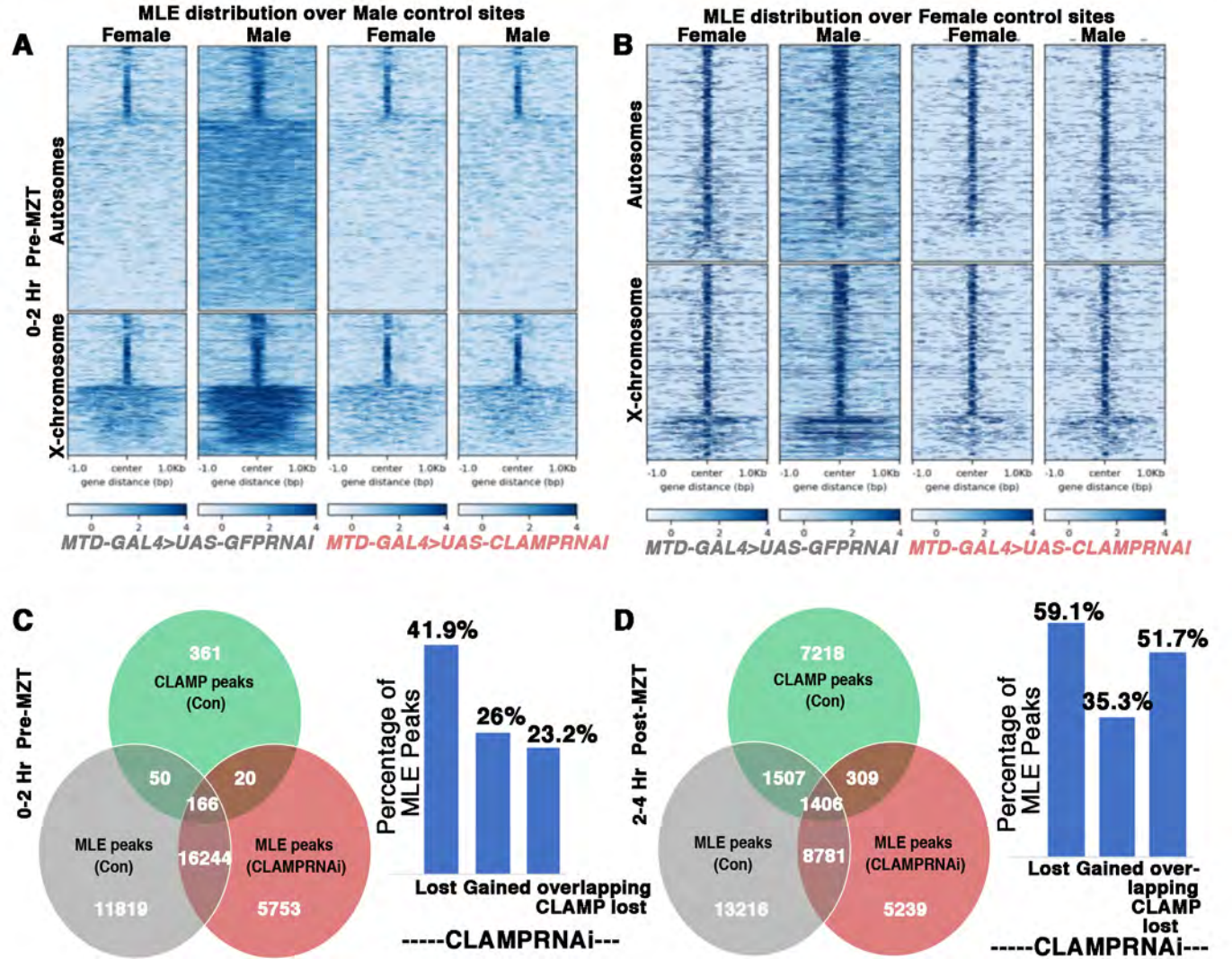

Fig 5. Ray, Conard et al 2022

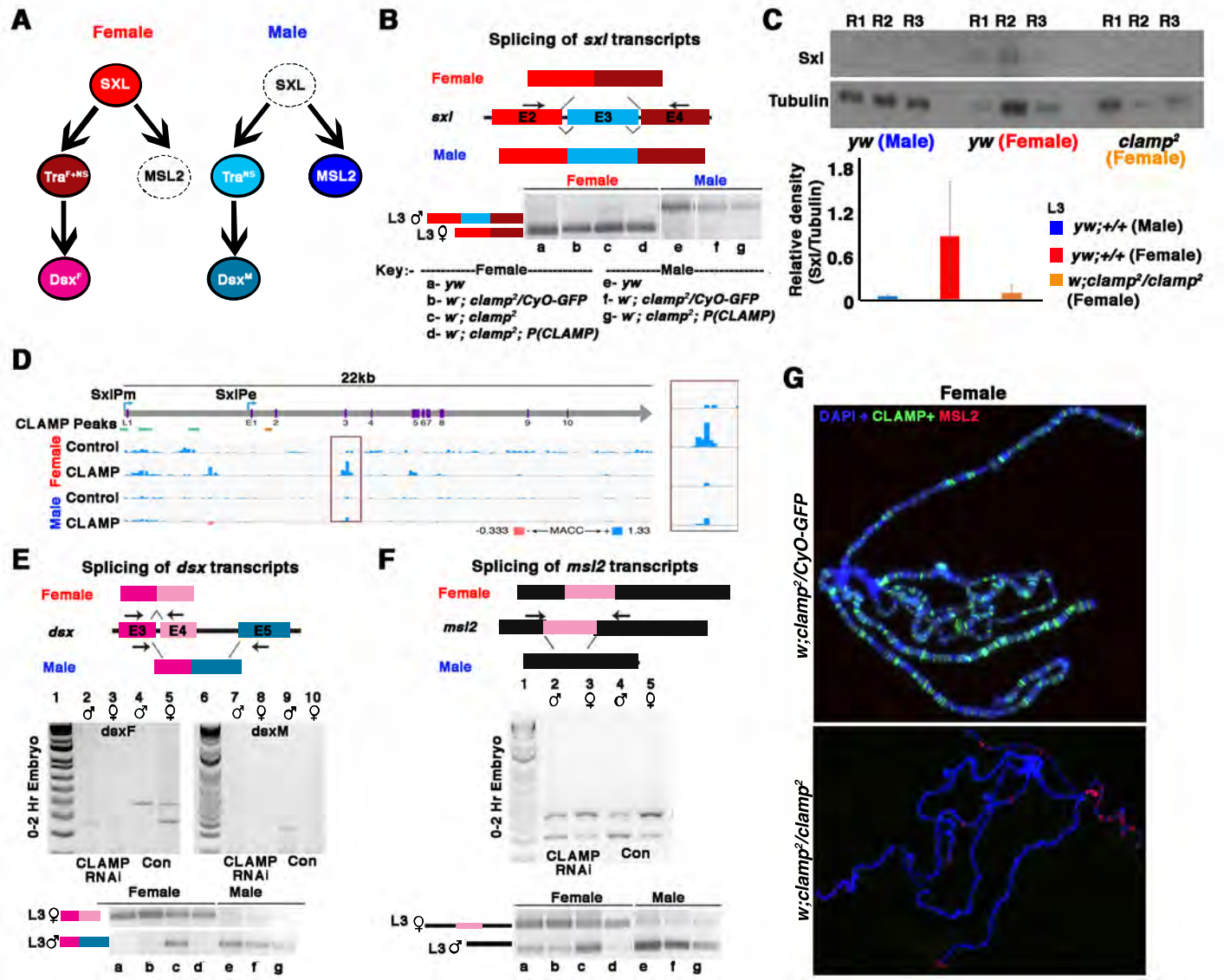
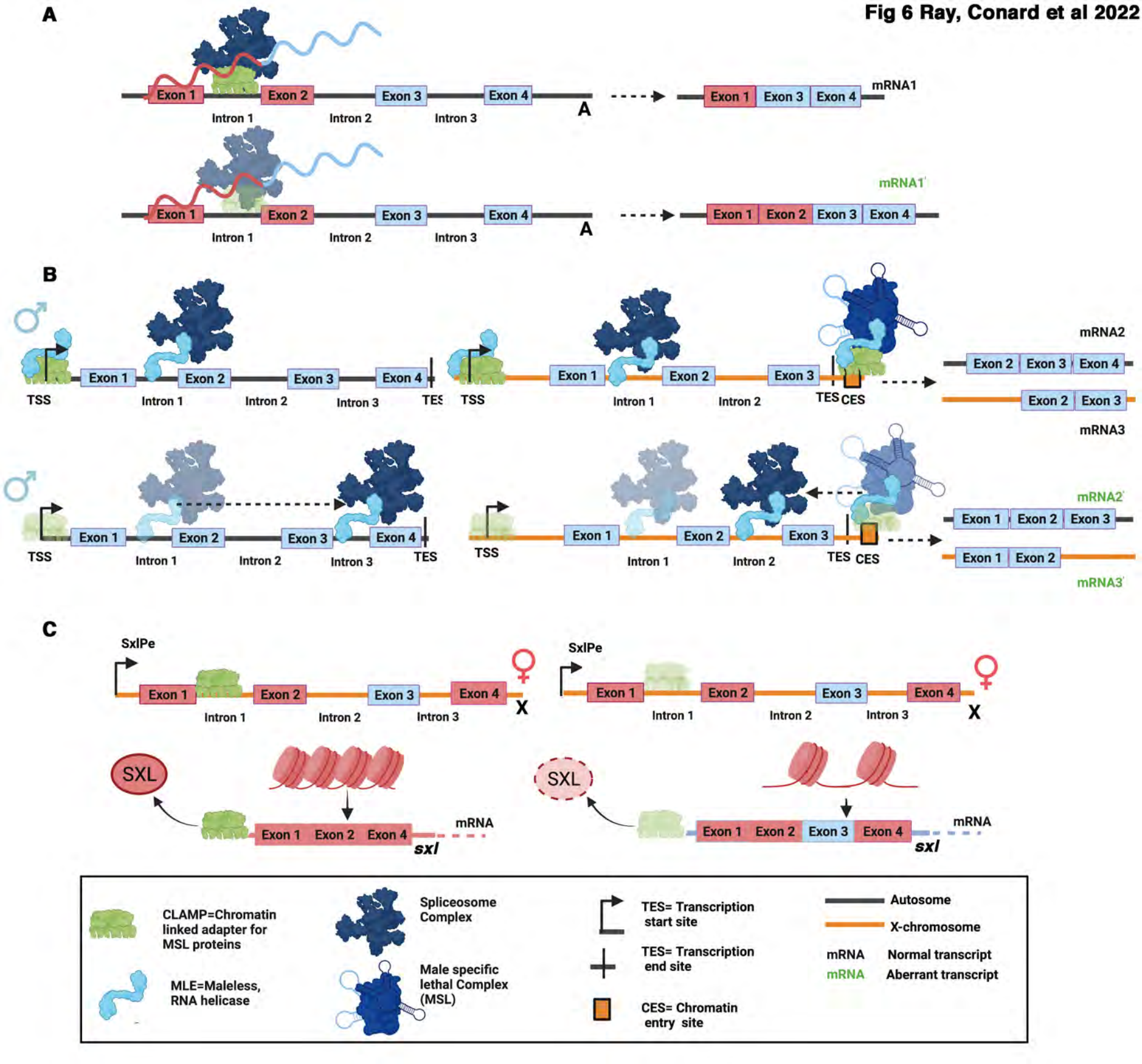
Fig 6 Ray, Conard et al 2022