A dual DNA/RNA-binding factor regulates co-transcriptional splicing through target RNA interaction and modulates splicing factor dynamics.

Mukulika Ray*, Julia Zaborowsky*, Pranav Mahableshwarkar, Smriti Vaidyanathan, Jasmine Shum, Annie Huang, Renjith Viswanathan, Isabelle Pilo, Victoria Chen, Kaitlyn Cortez, Ashley M. Conard, Szu-Huan Wang, Victoria Johnson, Noah Wake, Alexander E. Conicella, Ryan Puterbaugh, Nicolas L. Fawzi† and Erica Larschan†

* Co-first Authors

†Co-Corresponding authors

Summary

How RNA splicing events are targeted to the correct genomic locations in specific cellular contexts to generate context-specific transcript diversity remains very poorly understood. We show that a functionally conserved GA-rich DNA-binding transcription factor (TF), CLAMP, targets distinct RNA molecules in male and female cells to precisely regulate sex-specific splicing events through physical and functional interaction with RNA and RNA-binding proteins (RBPs). The prion-like domain of CLAMP (PrLD) and a stem-loop structure in the target RNA are essential for the CLAMP-RNA interaction. Moreover, the CLAMP PrLD domain regulates sex-specific splicing by modulating the dynamics of an hnRNPA2/B1 family protein that regulates alternative splicing. Thus, we demonstrate that a TF targets co-transcriptional splicing to the correct genomic locations by directly linking DNA binding sites to RNA targets and modulating the dynamics of RBP partners that drive alternative splicing.

Introduction

Mechanisms that drive precise co-transcriptional alternative splicing at thousands of genomic loci are critical to establishing sex and cell type-specific transcriptome diversity. Chromatin-bound transcription factors (TFs) and RNA binding proteins (RBPs) co-transcriptionally regulate alternative splicing[1-4]. Most TFs bind DNA and RNA[5-7] and interact with diverse RBPs[8,9]. However, the mechanisms by which TFs and RBPs coordinate transcription and alternative splicing at specific genomic locations in a particular cell type to drive specific alternative splicing events remain poorly understood. Also, how interactions between TFs, RNA, and RBPs generate the specific alternatively spliced transcripts essential for making developmental decisions remains unknown. We hypothesized that TFs are critical for targeting the correct co-transcriptional splicing events to specific genomic locations due to their unique ability to bind chromatin, RNA, and spliceosomal RBPs.

We test this hypothesis by defining the role of GA-binding TF CLAMP (chromatin-linked adapter for MSL Proteins) in targeting sex-specific splicing to the correct genomic locations on chromatin in *Drosophila* because we can simultaneously genetically manipulate the system, conduct high-throughput genomic experiments *in vivo,* and live imaging experiments in living tissue. CLAMP has several properties that suggest that it could target sex-specific alternative splicing events to chromatin: 1) CLAMP binds to specific GA-enriched DNA motifs via its mapped DNA binding domain, and its binding sites differ between males and females[10-13]; 2) CLAMP is a pioneer TF that regulates sex-specific splicing in embryos at genes where it does not regulate transcription[14]; 3) CLAMP is enriched at the intronic regions of CLAMP-dependent sex-specifically spliced genes[11,14]; 4) CLAMP is associated with RBPs that are spliceosome components[9].

Here, we combine diverse *in vivo* and *in vitro* approaches to define a new mechanism by which a TF can regulate co-transcriptional splicing. We define direct CLAMP-RNA interactions and compare them to CLAMP-DNA interactions to identify associations between TF-DNA and TF-RNA binding at genes where CLAMP regulates sex-specific alternative splicing. We have identified essential protein and RNA domains that promote direct CLAMP-RNA interactions using NMR (Nuclear Magnetic Resonance) spectroscopy and RNA-protein EMSAs (Electrophoretic

Mobility Shift Assays). Furthermore, we determined that CLAMP directly regulates the biophysical properties of Hrp38 its RBP interaction partner which is an hnRNPA2/B1 homolog known to regulate alternative splicing as part of nuclear splicing condensates. Thus, overall, our data support a model in which transcription factors with specific DNA and RNA binding capacity can target unique RNA molecules to specific genomic locations and alter the biophysical properties of RBP partner proteins to regulate context-specific alternative splicing.

Results

CLAMP binds RNA to chromatin

Because CLAMP is a DNA-binding protein enriched at the intronic regions of CLAMP-dependent sex-specifically spliced genes, which interacts with multiple splicing factors[11,14], we hypothesized that CLAMP functions as a regulator of co-transcriptional splicing. To test this hypothesis, we first investigated whether CLAMP regulates sex-specific splicing by directly interacting with target RNAs. Therefore, we defined RNA species interacting with CLAMP at high resolution across different cellular compartments: the nucleoplasm vs. chromatin. If we identified CLAMP-RNA interactions in the chromatin fraction involving RNAs that require CLAMP for alternative splicing, this would support a model in which CLAMP plays a role in co-transcriptional splicing. To further define a potential co-transcriptional function for CLAMP, we wanted to identify the spatial distribution of CLAMP RNA binding with respect to its DNA binding across the genome by comparing CLAMP-RNA binding sites with CLAMP DNA binding sites.

To achieve these objectives, we performed fractionation iCLIP (**i**ndividual-nucleotide resolution **C**ross**L**inking and **I**mmuno**P**recipitation) to identify CLAMP RNA targets using anti-CLAMP antibody in chromatin fractions (ChF) and nucleoplasmic fractions (NF) of male (S2) and female (Kc) *Drosophila* cell lines (Fig 1A). Kc and S2 cells are well-established cell lines for studying sex-specific processes such as dosage compensation [15,16].

93 We found that most CLAMP interactions with RNA occurs on chromatin (91.9% in males and
94 58.4% in females) with unique sex-specific targets, with more RNA transcripts binding to CLAMP
95 in males (N=654) than females (N=286) (Fig 1B, Table 1a). Interestingly, CLAMP also directly
96 interacts with spliceosomal RNAs sex-specifically (Table 1a). In the male chromatin fraction,
97 CLAMP interacts with the catalytic step 2 spliceosome consisting of U2, U5, and U6 snRNAs
98 (FDR:1.7E-3). In contrast, the female chromatin fraction is enriched for transcripts that encode
99 proteins that bind to the U1-U2 snRNAs (FDR:1.1E-2), suggesting that CLAMP may regulate
100 splicing differently in males and females by interacting with different spliceosomal RNAs.
101
102 Next, we asked how the CLAMP interaction with DNA correlates with its interaction with RNA
103 on chromatin. Therefore, we used Bindcompare, a tool we developed for comparing protein
104 binding to nucleic acids[52], to plot the frequency of CLAMP RNA binding peaks on chromatin
105 within a region ±1 kb of the closest CLAMP DNA binding peak. Using Bindcompare, we defined
106 the following categories of binding sites: a) complete overlap (CRO) of RNA peaks with DNA
107 peaks; b) partial overlap of RNA peak with the ends of a DNA peak (ORF or ORE); and c) RNA
108 peaks nearby (±1 kb) DNA peaks (PXP) (Fig 1C, Fig S1a-c, Table 1b).
109
110 We found that most of the total CLAMP RNA peaks identified (71.4% in males and 89.5% in
111 females) are either proximal to (~1 kb) or overlapping with CLAMP-DNA peaks (Fig 1D). It is
112 important to note that the number of CLAMP-RNA peaks identified (Fig 1D) is greater than the
113 number of RNA targets (Fig 1B) since many targets have multiple CLAMP-binding sites (Table
114 1b). Notably, almost 50% of these proximal or overlapping peaks were actually within 250 bp of
115 the middle of the nearest CLAMP DNA peak in both male and female cells, indicated by high
116 counts of CLAMP-RNA peaks plotted to the nearest CLAMP-DNA peak (Fig 1C, E), suggesting
117 that CLAMP links RNA to DNA during co-transcriptional splicing at a subset of its target genes
118 (Table 1c) identified in early embryos and cell lines[14]. Notably, only a subset of CLAMP-
119 dependent spliced genes in the embryo and cell lines are direct CLAMP-RNA targets. A subgroup
120 of the CLAMP-RNA targets overlaps with CLAMP-DNA targets in males (Table 1c). These
121 observations reveal that CLAMP directly contacts DNA and RNA at a subset of the target genes
122 regulating splicing, indicating a direct function in co-transcriptional splicing.
123

124

125 **The CLAMP prion-like domain (PrLD) is essential for CLAMP-RNA interaction.**

126

127 After identifying that CLAMP binds to RNA, we examined the CLAMP protein domain structure
128 to identify domains that may be responsible for its ability to bind RNA. CLAMP does not have a
129 canonical **R**NA **r**ecognition **m**otif (**RRM**). However, CLAMP contains a prion-like domain
130 (PrLD), a subclass of intrinsically disordered domains (IDR) enriched in polar residues similar to
131 those found in yeast prion proteins (Fig 1A). Many RBPs and TFs contain PrLDs [17,18] that can
132 promote RNA binding and drive the formation of phase-separated biomolecular condensates [19].
133 Some PrLDs in TFs promote co-assembly with RBPs to regulate transcription[20-22]. However, the
134 mechanisms by which the PrLD regions within TFs regulate alternative splicing remained
135 unknown. Therefore, we expressed and purified full-length CLAMP protein ($CLAMP^{WT}$-FL) and
136 CLAMP protein in which we deleted the PrLD domain ($CLAMP^{delPrLD}$-FL) (Fig 2A). We also
137 generated CLAMP1-300 aa, which removes the mapped DNA binding domain but includes the
138 PrLD domain (154-300 a.a). Next, we demonstrated the disordered nature of this domain (Fig 2B)
139 using NMR spectroscopy. We also generated an NMR spectrum of a mixture of total yeast RNA
140 extract and isolated CLAMP PrLD and found that CLAMP PrLD directly interacts with RNA (Fig
141 1C). Overall, we showed that the CLAMP 1-300 aa region is sufficient for direct interaction with
142 RNA.

143

144 Next, we further defined the role of the PrLD domain in promoting CLAMP-RNA interactions by
145 performing RNA-Protein EMSAs assays using specific RNA probes for a well-studied candidate
146 RNA that we identified as a CLAMP interactor by iCLIP (Fig. 2D). CLAMP is associated with
147 the Male Sex-lethal Complex (MSL), which regulates dosage compensation in males and contains
148 the *roX* (RNA on the X) long non-coding RNAs[12,23]. However, it was not known whether CLAMP
149 directly interacts with the *roX* RNAs. Because the *roX* RNAs have a defined structure and
150 function[24], we used *roX2,* identified by iCLIP as an *in vivo* male-specific CLAMP interactor (Table
151 1a), to determine whether CLAMP directly interacts with RNA *in vitro*. Therefore, we used a full-
152 length *roX2* probe (411 nt) for RNA-Protein EMSA gel shift assays (Fig 2D) with full-length
153 $CLAMP^{WT}$ and $CLAMP^{delPrLD}$ proteins that were expressed and purified from *E. coli*. At the same
154 protein and RNA concentrations, the $CLAMP^{delPrLD}$ protein binds to RNA less efficiently than

155 CLAMP^{WT} (Fig. 2E). Therefore, the CLAMP PrLD region promotes the binding of CLAMP to
156 RNA.
157
158 Next, we wanted to test how RNA sequence and known secondary structure regulate the CLAMP-
159 RNA interaction. In our iCLIP experiments, we identified the CLAMP-binding region on the *roX2*
160 RNA (23nt, Fig. 2D). This region is just 12 nt downstream of a functionally important stem-loop
161 structure (Fig. 2D), which is essential for *roX2* interaction with RNA helicase A, a component of
162 both the MSL complex and the spliceosome[24]. Therefore, we hypothesized that the *roX2* stem loop
163 promotes the interaction of *roX2* RNA with CLAMP. Hence, we used two additional *roX2* RNA
164 probes, which either completely delete both the stem-loop and the CLAMP binding region (rox2-
165 RNA^{mut}) or delete only the CLAMP binding region identified by iCLIP (rox2-RNA$^{delCLAMP\ binding}$)
166 (Fig. 2D) for CLAMP-RNA Gel-shift assays.
167
168 We found that: 1) the *roX2* probe (342nt) lacking both the stem-loop and the iCLIP CLAMP
169 binding region (69nt) reduced CLAMP binding to the *roX2* RNA (Fig 1F); 2) In contrast, the *roX2*
170 RNA$^{delCLAMP\ binding}$ probe (399nt) lacking only the iCLIP CLAMP binding region (25nt) but
171 retaining the stem-loop, did not qualitatively affect CLAMP binding to the *roX2* RNA *in vitro* (Fig
172 S2). Therefore, the presence of the *roX2* RNA stem-loop and the CLAMP PrLD increase the ability
173 of CLAMP to directly interact with RNA.
174
175
176 **The CLAMP PrLD domain is important for CLAMP-dependent sex-specific splicing.**
177
178 We combined *in vitro* approaches with our *in vivo* data to define the functional role of the CLAMP
179 PrLD domain in alternative splicing. We defined the role of the CLAMP PrLD by generating
180 complete (CLAMP Δ154-290 aa) and partial (CLAMP Δ160-290 aa) PrLD deleted mutant fly lines
181 at the endogenous CLAMP locus using CRISPR/Cas9 genomic mutagenesis and homologous
182 repair[39] (Fig 3A). Complete deletion of the PrLD (*clamp^{delPrLD}*) results in embryonic lethality. In
183 contrast, partial deletion (*clamp$^{delPrLD+6Q}$*) mutants that retain a stretch of 6 glutamines survive,
184 suggesting that these glutamines (aa154-160) are essential for viability (Fig 3A). Therefore, we
185 assayed sex-specific alternative splicing in trans-heterozygous mutants

186 (*clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$) to assure viability until it is possible to define the sex of animals at
187 the larval stage (Fig 3A, B). Analysis of RNA sequencing data using the time2splice pipeline[14]
188 from trans-heterozygous mutants (*clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$) identifies CLAMP PrLD-
189 dependent female- and male-specific splicing (Fig 3B, Table 2a, b). More than 70% of the CLAMP
190 prLD domain-dependent alternative splicing was sex-specific, with ~40-50% identified as new
191 splicing events (Fig 3B). Therefore, the CLAMP PrLD is essential for the viability and impacts
192 sex-specific alternative splicing.
193
194 To determine whether the CLAMP PrLD regulates sex-specific alternative splicing of RNAs that
195 are directly bound by CLAMP (Table 2a, b), we compared the RNAs bound to CLAMP *in vivo*
196 identified by iCLIP (Table 1a) with CLAMP PrLD-dependent sex-specifically spliced genes
197 (Table 2a, b). Due to the large sample required, performing iCLIP in mutant L3 larvae was not
198 possible, so we used cell line iCLIP targets. We determined that 14 CLAMP PrLD-dependent sex-
199 specifically spliced genes are direct CLAMP RNA interactors, identifying a smaller subset of key
200 genes for further analysis (Table 3). Moreover, most of these 14 direct targets are themselves
201 regulators of alternative splicing (Table 3). Therefore, it is possible that CLAMP, which is heavily
202 maternally deposited[25] and regulates splicing of 60% of all sex-specific spliced isoforms[14],
203 functions as an essential upstream splicing regulator by directly regulating the splicing of splicing
204 factors, which then regulates alternative splicing of additional target genes.
205
206 We further analyzed the functional role of the CLAMP PrLD domain in regulating alternative
207 splicing at the *hrp38* gene locus, the hnRNPA2B1 orthologue, which encodes a key regulator of
208 alternative splicing[26]. The hrp38 transcript is one of the 14 CLAMP PrLD-dependent spliced genes
209 that is a direct CLAMP RNA target (Table 2a, 3). Integrating iCLIP with alternative splicing
210 analysis suggested that CLAMP binds to exon 1 of the *hrp38* transcript in males on chromatin,
211 which undergoes splicing to remove exon 2 (Fig. 3C, D). Moreover, CLAMP binds DNA as well
212 as RNA at the *hrp38* gene locus (Fig. 3C), and CLAMP-RNA gel shift assays (Fig. 2) predict that
213 in the CLAMP *PrLD* mutant, CLAMP binding at the exon 1 of *the hrp38* transcript should be
214 affected (Fig. 3D). We found that in *clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$ mutants, splicing of intron
215 AF:r6:3R:28600568:28600718:+ between exon2 and exon4 (Fig 3D) is regulated by CLAMP as
216 predicted by genome-wide splicing analysis (Table 2a). In contrast, in *clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$

217 mutants, the *hrp38* transcript retains exon2 (Fig 3E, F). Next, we validated that the alternative
218 splicing of the *hrp38* gene is sex-specifically regulated by the CLAMP PrLD domain *in vivo* during
219 development at the third instar larval stage in males and females. In female and male mutants,
220 splicing of intron AF:r6:3R:28600568:28600718:+ occurs. However, in males, splicing was more
221 efficient than in females (Fig 3F). Our *in vivo* splicing analysis, thus, demonstrates that the
222 CLAMP PrLD domain is important for *hrp38* RNA splicing during male and female development.
223
224 Since our iCLIP data shows CLAMP binds to exon1 of *hrp38* RNA, we compared $CLAMP^{WT}$
225 binding to *hrp38* exon1 and exon2 probes using RNA EMSAs and found that at the same protein
226 and RNA concentrations, CLAMP binds strongly to *hrp38* exon1 RNA with no unbound RNA
227 remaining (Fig 3G). Furthermore, the PrLD domain of CLAMP is important for CLAMP-*hrp38*
228 RNA binding (Fig 3H). Interestingly, $CLAMP^{WT}$ still binds to *hrp38* exon2 *in vitro* even though it
229 does not bind *in vivo,* suggesting that other RBPs modulate the specificity of CLAMP-RNA
230 interactions *in vivo*.
231
232 Thus, we defined how CLAMP regulates the sex-specific splicing of the *hrp38* transcript.
233 Previously, we have demonstrated using proteomics that CLAMP sex-specifically interacts with
234 the Hrp38 protein, a component of highly mobile nuclear splicing speckles that regulate alternative
235 splicing[27-30]. Therefore, due to the functional interaction between CLAMP and Hrp38 at both the
236 RNA and protein levels, we determined how CLAMP regulates the sex-specific dynamics of
237 Hrp38, one of CLAMP's essential target gene and interactors, to define how CLAMP regulates
238 sex-specific splicing beyond directly interacting with the RNA of target genes.
239
240
241 **CLAMP regulates the nuclear dynamics of the alternative splicing regulator hnRNPA2/B1**
242 **orthologue Hrp38**
243
244 PrLD domains can drive the phase separation of proteins[19] and promote the formation of
245 subnuclear bio-condensates involved in transcription[20,31] and splicing[19,32]. Therefore, we
246 performed biochemical phase separation assays to determine how the CLAMP PrLD regulates the
247 phase separation behavior of CLAMP (Fig. 4A). We found that the CLAMP PrLD contributes to

248 phase separation of the N-terminal half of CLAMP (residues 1-300). Many RBPs involved in
249 splicing form splicing condensates, and TFs form transcription condensates, but the mechanisms
250 by which transcription and splicing condensates interact to promote the fidelity of co-
251 transcriptional splicing remain very poorly understood[33]. Although the PrLDs of TFs can promote
252 co-assembly with RBPs, current work does not yet explain how TF-RBP interactions regulate the
253 dynamics of splicing condensates, which is linked to their function.
254
255 The hnRNPA1/B2 orthologue Hrp38 is a strong candidate to function with CLAMP to promote
256 alternative splicing because it is associated with CLAMP in three ways: 1) CLAMP interacts with
257 the *hrp38* RNA transcript in our iCLIP experiments (Table 1a); 2) We demonstrate that CLAMP
258 regulates alternative splicing of the *hrp38* transcript (Table 2, Fig 3), and 3) CLAMP interacts with
259 Hrp38 protein[9]. Hrp38 forms dynamic nuclear splicing condensates or speckles[27-30]. Therefore, we
260 hypothesized that CLAMP PrLD-mediated phase separation affects the dynamics of Hrp38
261 condensates. We tested our hypothesis by combining *in vitro* and *in vivo* approaches to measure
262 condensate formation, stability, and dynamics.
263
264 In *vitro*, CLAMP 1-300^{WT} and Hrp38 co-phase-separate into liquid-like condensates (Fig 4B). To
265 complement *in vitro* studies, we performed *in vivo* analysis of Hrp38 nuclear speckles in fixed and
266 live conditions in *Drosophila* third instar larval tissue (Fig 4C). Salivary gland (SG) and
267 malpighian tubule (MT) tissue nuclei were used since they have large polyploid or polytene nuclei,
268 making it easier to visualize and track nuclear speckles. Prior studies[35-38] suggest that larger
269 aggregates of hnRNPs are more immobile and have decreased function in regulating alternative
270 splicing. Therefore, we did the following experiments in males and females to test our hypothesis
271 that the CLAMP PrLD domain regulates they dynamics of Hrp38 to drive changes in sex-specific
272 splicing: 1) We analyzed fixed SG nuclei expressing Hrp38-GFP samples to determine the effect
273 of the CLAMP PrLD domain on Hrp38 nuclear speckle size (Fig 4C-F); 2) We performed live
274 imaging to define how the CLAMP PrLD domain regulates the mobility of Hrp38-GFP nuclear
275 speckles (Fig 4C, G) 3) We performed FRAP analysis in MT nuclei to measure the mobility of
276 the Hrp38 protein in the nucleus by measuring the fraction of Hrp38 that remains bound (Fig 4C,
277 H). We also analyzed the co-localization of CLAMP and Hrp38 *in vivo* (Fig 4I-J) to determine
278 how frequently they interact and compare e with our *in vitro* interaction findings (Fig 4B).

279

280 Compared with wild type controls, in *clamp*^{delPrLD} mutants, we observe larger Hrp38 speckles (Fig
281 4D-F) with lower mobility (Figs. 4G,H). The Hrp38 protein forms nuclear speckles (Fig 4D),
282 which are very dynamic (Movie 1a); however, in *clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$ mutants, Hrp38
283 protein forms bigger nuclear aggregates in both females and males compared to controls (Fig 4E),
284 with median speckle size in males slightly higher than in females. Hrp38 nuclear size negatively
285 correlated with speckle dynamics (Movie 1b, Fig 4G, H). FRAP analysis shows that in
286 *clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$ mutant male and female individuals, Hrp38 remains more immobile
287 compared with matched wild-type controls (Fig. 4G, Fig. S2, Movie 2). Therefore, the CLAMP
288 PrLD domain reduces the size and increases the dynamics of the Hrp38 speckles.

289

290 Next, we analyzed the trajectories of moving Hrp38 speckles over time with live image analysis
291 (Movie 3) to determine whether CLAMP regulates their mobility. More restricted tracks denote
292 more immobile speckles, while freely moving particles represent more mobile speckles[34].
293 Interestingly, we found that in wild-type controls, female Hrp38 speckles are less mobile than male
294 Hrp38 speckles. Without the CLAMP PrLD domain, Hrp38 speckles in males become less mobile
295 (Fig 4H), therefore becoming more similar to those in females. Thus, the CLAMP PrLD regulates
296 the dynamics of Hrp38 speckles sex-specifically in live tissues, promoting their enhanced mobility
297 in males. We identified partial co-localization between CLAMP and Hrp38, which is significantly
298 increased in *clamp*^{delPrLD} mutant males (Fig. 4I- J). Therefore, the CLAMP PrLD domain in males
299 also prevents colocalization with Hrp38, supporting its differential role in influencing Hrp38
300 dynamics in males and females.

301

302 Thus, the CLAMP PrLD reduces the size of Hrp38 splicing condensates, which have been shown
303 to increase their ability to mediate alternative splicing with high fidelity. Therefore, our data
304 support a model in which the CLAMP PrLD modulates Hrp38 condensates' ability to regulate
305 alternative splicing. Furthermore, the CLAMP PrLD promotes the formation of CLAMP liquid
306 droplets (Fig. 4A), which can co-localize with Hrp38 splicing condensates (Fig. 2I- J). Deletion of
307 CLAMP PrLD promotes CLAMP phase separation *in vitro* (Fig. 2G) and sex-specifically regulates
308 the dynamics of Hrp38 speckles *in vivo* (Fig. 2K- O). Overall, our data suggest that the CLAMP

309 PrLD alters the dynamic properties of Hrp38, preventing it from forming large, nonfunctional
310 aggregates.
311
312
313 **Discussion**
314
315 Very little is understood about how alternative splicing is targeted to the right place at the right
316 time within a highly compacted genome. We hypothesized that transcription factors that interact
317 with RNA binding proteins would be strong candidate factors to target alternative splicing to the
318 correct genomic locations in a developmental context. To test our hypothesis, we examined the
319 CLAMP TF as a model to reveal new mechanisms by which the hundreds of TFs interacting with
320 RNA and RNA binding proteins regulate splicing because it regulates sex-specific splicing (Ray,
321 elife).
322
323 CLAMP is a pioneer TF that interacts with the male-specific MSL complex to regulate file dosage
324 compensation. It is essential for MSL targeting at GA-rich chromatin entry sites (CES) on
325 chromatin, thus helping to target the MSL complex to the X chromosome. Interestingly, CES
326 evolved from polypyrimidine tracts, suggesting a further link with splicing (**Quinn, Howard**
327 **Chang**). We found that, similar to its role in dosage compensation, where it recruits a
328 ribonucleoprotein complex containing the *roX* non-coding RNAs, CLAMP directly links RNA to
329 DNA at target genes, at which it regulates sex-specific splicing targets RNA, mainly on chromatin
330 at a subset of genomic locations where it binds DNA. We found that CLAMP regulates alternative
331 splicing at a subset of these sites. Because not all CLAMP-dependent sex-specific spliced genes
332 are both direct CLAMP-RNA and CLAMP-DNA targets, CLAMP regulates alternative splicing
333 through multiple mechanisms that may be regulated by the presence of the MSL complex in males
334 but not females. Interestingly, CLAMP regulates the splicing of many RBPs that are components
335 of alternative splicing complexes, including *hrp38.* The alternatively spliced forms of these
336 proteins can result in protein isoforms with different functions. CLAMP-regulated alternate first
337 (AF) alternative splicing of *hrp38* transcript gives rise to two different protein isoforms (Fig 3).
338 How these two protein isoforms differ in their physical and functional properties needs to be
339 determined.

340

341 We define CLAMP-regulated RBP protein dynamics as an alternative mechanism by which
342 CLAMP regulates splicing. We show that the CLAMP PrLD domain is essential for determining
343 the size of Hrp38 protein condensates, which in turn regulates their function. We hypothesize that
344 CLAMP affects Hrp38 splicing speckles by regulating the biophysical phase separation properties
345 of Hrp38 condensates. Our data support a model in which the CLAMP PrLD domain differentially
346 regulates CLAMP-Hrp38 interactions to regulate splicing condensate properties in males and
347 females. We propose that the interaction of CLAMP with the MSL complex in males but not
348 females promotes sex differences in Hrp38 function. The following evidence supports this model:
349 1) Hrp38 mobility is different between males and females (Fig 4J) as is the CLAMP-Hrp38 protein
350 interaction which we identified as female-enriched from proteomic analysis (Urban et al); 2) Hrp38
351 splicing condensate mobility is more dysregulated males than in females in *clamp PrLD* mutants
352 (Fig 4J); 3) The colocalization between CLAMP and Hrp38 is decreased in males versus females,
353 potentially due to competition from the CLAMP-MSL interaction. (Figs 4H, J). Overall, our data
354 support a model in which CLAMP sex-specifically regulates the dynamics of the hnRNP
355 orthologue Hrp38.

356

357 Thus, one determinant of the sex-specificity of CLAMP-dependent alternative splicing is sex-
358 dependent variability in Hrp38 condensate dynamics. CLAMP interacts with the Hrp38 protein
359 differentially between sexes [9]. Also, the PrLD domain of CLAMP regulates Hrp38 condensates
360 more strongly in males, predicting that the CLAMP PrLD domain is important for these male-
361 biased interactions. This might be due to CLAMP interacting with the MSL complex in males,
362 thereby inhibiting its interaction with other proteins, such as Hrp38. However, the absence of this
363 complex would make such associations possible. The biomolecular interactions by which the PrLD
364 domain of CLAMP regulates Hrp38 and determines CLAMP-Hrp38 interactions require further
365 study.

366

367 Another determinant of the sex-specificity of CLAMP-dependent alternative splicing is that
368 CLAMP has different RNA targets in males and females, with the PrLD domain again playing an
369 important role. We hypothesize that sex-specific CLAMP-RNA targets result from CLAMP
370 having different RBP interactors in males and females, especially closer to chromatin. Again, this

371 might result from differential interactions between CLAMP and different riboprotein complexes
372 like MSL2 complex (MSL proteins and rox RNA) and spliceosomal complex (RBPs and snRNAs)
373 on chromatin. In the future, new techniques like RD-SPRITE that identify the distribution of
374 different nuclear clusters, including splicing hubs on DNA, could be used to determine the global
375 as well as specific local role of CLAMP in establishing site-specific RNA splicing in male
376 compared with female cells.
377
378 We found substantial new splicing events occurring in the absence of CLAMP (Ref) (Fig 3B),
379 which indicates that at most genomic sites, CLAMP acts as an inhibitor of alternative splicing
380 events. Thus, overall, our data support a model where sex-specific differences in alternative
381 splicing arise from differential TF-RNA and TF-RBP interactions (Figs. 1 and 4). We show for
382 the first time that an intrinsically disordered domain (IDR) in a TF regulates alternative splicing
383 by regulating the dynamics of splicing condensates. Also, our data suggest that TFs that bind to
384 specific DNA motifs and have RNA binding properties are strong candidates for understanding
385 how specific alternative splicing events are regulated in specific cellular contexts.
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401

402
403
404
405
406
407
408
409 **Methods:**
410
411 **Cell culture:** Kc and S2 cells were maintained at 25°C in Schneider's media supplemented with
412 10% Fetal Bovine Serum and 1.4X Antibiotic-Antimycotic (Thermofisher Scientific, USA). Cells
413 were passaged every three days to maintain an appropriate cell density.
414
415 **Fly stocks and husbandry:** *Drosophila melanogaster* fly stocks were maintained at 24°C on
416 standard corn flour sucrose media. Fly strains with complete (CLAMP Δ154-290a.a) and partial
417 (CLAMP Δ160-290a.a) PrLD deleted mutant fly lines using CRISPR/Cas9 genomic mutagenesis
418 and homologous repair[39]. We used the flyCRISPR Optimal Target Finder tool from the University
419 of Wisconsin to design a CRISPR target sequence for *clamp*^{delPrLD} and *clamp*$^{delPrLD+6Q39}$. We
420 cloned target sequence oligonucleotides (one gRNA) for *clamp*^{delPrLD}
421 (sense:5'CTTCGGGTACAACGCCAAAGCGAG3';antisense:3'CCCATGTTGCGGTTTCGCTC
422 -CAAA5') and two gRNAs for *clamp*$^{delPrLD+6Q}$ (sense: 5'CTTCG-
423 AATAGAATCGCCGCCCGCT3';antisense:3'CTTATCTTAGCGGCGGGCGACAAA5')and(se
424 nse:5'CTTC-GTTGTGGCTGCACAGACTGG3';antisense:3'CAACACCGACGTGTCTGACC-
425 CAAA5') into the pU6-BbsI-chiRNA plasmid (Addgene no. 45946), following the protocol
426 outlined on the flyCRISPR website. We validated the correct ligation of the *clamp* CRISPR target
427 sequence into the pU6-BbsI-chiRNA plasmid by Sanger sequencing using universal M13 primers.
428 For homologous repair, we used
429 ssODN15'ACATAAGCTTTAAGTGTGACGTATGTTCAGATATGTTCCCTCATTTGGCAC
430 TTCTTAATGCTCATAGAGAGGCGAGCAGTGGGTCTGGCCATCATCCTGTGAAAAAAC
431 GAAATTCCCAGCAGATGACCAAAT3' for *clamp*^{delPrLD}and ssODN2
432 5'CTTCTTAATGCTCATAAGCGGATGCATACAGACGGGGAACAGCAGCAACAACAGC

433 AACATAACGCCCAAGCTGGCGGTACAACGCCAAAGCGAGAGGCGAGCAGTGGGTCT
434 GGCCATCATCCTGTGAAAAAA3' for *clamp*$^{delPrLD+6Q}$.The commercial service, BestGene Inc.,
435 microinjected the validated pU6-BbsI-chiRNA plasmid containing the *clamp* target sequence into
436 germline-expressing Cas9 flies Bloomington stock #51324. Flies containing a single mutation
437 were returned balanced over the *Curly of Oster* (*CyO)* second chromosome balancer. From these
438 progenies, we identified the CRISPR/Cas9-generated mutations by PCR across the target region
439 (forward:5′-GATATGTTCCCTCATTTGGCAC-3′, reverse:5′-
440 CACTCCCATGCTTCACACAG-3′). We isolated two independent clamp alleles from this
441 validation: (1) y^1, w^{1118}; *clamp*^{delPrLD}/*CyO*; and (2) y^1, w^{1118}; *clamp*$^{delPrLD+6Q}$/*CyO.* These were
442 crossed to obtain male and female *clamp*$^{delPrLD+6Q}$/*clamp*^{delPrLD} genotypes. Crisper/Cas9 generated
443 fly mutants and fly strain y^1 w^{1118}; P{w[+mC] =PTT-GC} Hrb98DE[ZCL0588] expressing
444 Hrp38GFP (Bloomington #6822) were crossed to obtain *clamp*$^{delPrLD+6Q}$/*clamp*^{delPrLD}male and
445 female expressing GFP tagged Hrp38 protein.
446
447 **CUT&RUN in cell lines:** Cells were allowed to grow to confluency and harvested. An equal
448 number of cells for each category were suspended in wash buffer and subjected to Cut&Run assay
449 according to Skene et al. 2018[40] using rabbit anti-CLAMP (5µg) to immunoprecipitate CLAMP
450 bound DNA fragments from male (S2) and female (Kc) cell lines. Three replicates for males and
451 females were run, but one female sample was dropped during later stages due to insufficient
452 starting material. Rabbit IgG was used as a control for each male and female cell line sample. 1ng
453 CUT&RUN DNA was used to generate libraries using the Kapa Hyper prep kit (Roche, USA) and
454 SeqCapAdapter Kit A (Roche, USA). 14 PCR cycles were used to amplify the libraries. AMPure
455 XP beads (Beckman Coulter, USA) were used for library purification, and fragment analysis was
456 performed to check the library quality. Paired-end 2x25 bp Illumina Hi-seq sequencing was
457 performed.
458
459 **RNA-seq in third instar larvae (L3):** Total RNA was extracted from control (y^1 w^{1118};
460 *hrp38GFP*) and *clamp* mutant (*yw, clamp*$^{delPrLD+6Q}$/*clamp*^{delPrLD}; *hrp38GFP*) male and female third
461 instar larvae (3 each) using Trizol (Invitrogen, USA). Four replicates were in each category
462 prepared. Messenger RNA was purified from total RNA using poly-T oligo-attached magnetic
463 beads. After fragmentation, the first strand of cDNA was synthesized using random hexamer

464 primers, followed by the second cDNA synthesis. The library was ready after end repair, A-tailing,
465 adapter ligation, size selection, amplification, and purification, followed by paired-end RNA-
466 sequencing in Illumina Novaseq 6000. The sequencing data was run through a SUPPA-based
467 time2splice pipeline[14] to identify CLAMP-dependent sex-specific splicing events. Data is to be
468 submitted to the GEO repository.
469
470 **iCLIP:** Cells were allowed to grow to confluency and UV crosslinked using 254 nm UV light in
471 Stratalinker 2400 on ice (Stratagene, USA). UV-treated cells were lysed to get different cellular
472 fractions (Cytoplasmic, Nucleoplasmic, and Chromatin) according to the Fr-iCLIP (fractionation-
473 iCLIP) protocol from Brugiolo et al 2017[42]. Chromatin and nucleoplasmic fractions were sonicated
474 with a Branson digital sonicator at 30% amplitude for 30 seconds (10 sec on and 20 sec off) to
475 disrupt DNA before IP. All three fractions were separately centrifuged at 20,000 xg for 5 min at
476 4°C. Fractions were tested by Western blotting using RNApolI for Chromatin Fraction and Actin
477 for Cytoplasmic Fraction. Protein quantification for each fraction was done using the
478 manufacturer's protocol for Pierce 660 nm protein assay reagent (Thermo Scientific, USA). Each
479 Fraction was subjected to iCLIP protocol as described in Huppertz et al. 2014[41] using rabbit-
480 CLAMP antibody to immunoprecipitate bound RNAs extracted using proteinase K and phenol:
481 chloroform. Custom cDNA libraries were prepared according to Huppertz et al. 2014[41] using
482 distinct primers Rt1clip-Rt16clip for separate samples containing individual 4nt-barcode
483 sequences that allow multiplexing of samples. cDNA libraries for each sample were amplified
484 separately using 31 cycles of PCR, mixed later, and sequenced using standard Illumina protocols.
485 Heyl et al. 2020[42] methods using the Galaxy CLIP-Explorer were followed to preprocess, perform
486 quality control, post-process, and peak calling.
487
488
489 **Validation of hrp38 splicing using RT-PCR assay:** Total RNA was extracted from 5 third instar
490 larvae (L3) female and male embryos expressing $clamp^{delPrLD+6Q}/clamp^{delPrLD}$and y^{1}, w^{1118}.
491 Following the manufacturer's protocol, we reverse-transcribed one microgram of total RNA using
492 the SuperScript VILO cDNA Synthesis Kit (Life Technologies, USA). We amplified target
493 sequences by PCR using primers designed to span alternatively spliced junctions (FP-
494 5'AGAACGGCAACTCCAATGGC3' and RP-5'GCCAGTCTCCTTGTCAATGA3') and Quick

495 load Taq 2X Master mix (#M0271L, NEB, USA) according to the manufacturer's protocol (28
496 cycles). 10ul of PCR product of each replicate for each gene was loaded in separate wells in 2%
497 agarose gels and imaged using a ChemiDoc^{TM} MP Imaging system (BioRad, USA). All replicates
498 for each gene were loaded on the same gel. The gel images were quantified using the densitometry
499 steps with the Fiji image analysis tool. Student's t-tests were performed to determine significant
500 differences between groups (two samples at a time). Three replicates for RT-PCR samples were
501 performed.
502
503 **CLAMP Protein Expression and Purification:** Vectors encoding maltose binding protein
504 (MBP)-tagged CLAMP 1-300 and MBP-tagged CLAMP 1-300 delPrLD were produced by
505 cloning into the pTHMT vector. Plasmids were transformed into BL21 Star (DE3) cells, and
506 bacteria cultures were grown in a minimal M9 medium supplemented with ^{15}N ammonium
507 chloride. Cultures were grown at 37°C and 200 RPM to an optical density of 0.6-0.8 and
508 subsequently induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) for 4 hours at
509 37°C. Cell pellets were harvested by centrifugation at 6,000 RPM, resuspended in lysis buffer (20
510 mM Tris pH 8, 1 M NaCl, 10 mM Imidazole, 1 mM DTT, and one EDTA-free protease inhibitor
511 tablet (Roche) per liter of culture), and lysed by sonication. The lysed cell suspension was cleared
512 by centrifugation at 20,000 RPM for 50 min. The supernatant was filtered using a 0.2 µm syringe
513 filter and loaded onto a HisTrap HP 5 ml column. The HisTrap column was first washed with 20
514 mM Tris pH 8, 1 M NaCl, 10 mM Imidazole, 1 mM DTT Buffer, and then the bound protein was
515 eluted with 20 mM Tris pH 8, 1 M NaCl, 500 mM Imidazole, 1 mM DTT buffer. Fractions
516 containing MBP-CLAMP 1-300 were collected and purified on a Superdex 200 (26/600) column
517 equilibrated in 20 mM Tris pH 8, 1.0 M NaCl buffer. As above, CLAMP PrLD (154-290) was
518 expressed and purified using HisTrap HP. Following that , CLAMP PrLD was collected from the
519 HisTrap HP, concentrated to 1 mL, and cleaved overnight with TEV protease to separate CLAMP
520 PrLD from the MBP in 20 mM Tris pH 8, 1 M NaCl, 10 mM Imidazole, and 1 mM DTT Buffer.
521 The cleaved protein was separated from the HisTag by a subtraction step, flowing the cleaved
522 protein over the HisTrap HP column using the same Tris buffer previously described and collecting
523 the flow-through. MBP-tagged Full-Length CLAMP was grown following the same methods
524 above. Once the Full-Length CLAMP cell lysate supernatant was loaded onto HisTrap HP 5 ml
525 column, the protein was eluted with a gradient from 10 to 500 mM imidazole in pH 5.5 20 mM

526 MES, and 361 mM $CaCl_2$. Fractions containing MBP-CLAMP FL were collected and purified on
527 a Superdex 200 (26/600) column equilibrated in pH 5.5, 20 mM MES, and 361 mM CaCl2 buffer.
528 For each CLAMP construct, fractions containing the desired protein were verified using SDS-
529 PAGE, concentrated using a 10 kDa centrifugation filter (Millipore), aliquoted, and flash frozen
530 in liquid nitrogen.
531
532 **Hrp38 protein purification and phase separation assay:** Maltose binding protein (MBP)-tagged
533 Hrp38 was expressed in BL21 Star (DE3) *E. coli* cells. The cells were resuspended in 20 mM NaPi
534 at pH 7.4 with 300 mM NaCl and 10 mM imidazole, and the lysate was cleared by centrifuging at
535 20,000 rpm for 60 mins at 4° C. The supernatant was filtered using 0.2 µM filters and loaded on a
536 5 ml HisTrap HP column. The protein was eluted using an imidazole gradient of 10-300 mM over
537 5-column volumes. The protein fractions were pooled and loaded on a Superdex 200 (26/600)
538 column for size exclusion. 20 mM NaPi with 300 mMNaCl at pH 7.4 was used for size exclusion
539 chromotography and storage. Protein was flash frozen as aliquots of 1 mM. Fluorescent labeling
540 of the CLAMP 1-300 was done using AlexaFluor 488 maleimide dye, while Hrp38 was labeled
541 using 555 maleimide. The protein was diluted to 100 µM in 20 mM Tris buffer at pH 7.4 with 50
542 mM NaCl, and a 5-fold concentration (500 µM) of the dye dissolved in DMSO was added to 5%
543 of the total volume. The reaction mixture was incubated for 1 hour, and then, unbound AlexaFluor
544 was removed using 1 ml Zeba spin desalting columns. The proteins were concentrated to 1 mM,
545 flash frozen, and stored. For phase separation assay and microscopy, Hrp38 was buffer exchanged
546 to 20 mM Tris buffer at pH 7.4 with 50 mM NaCl to a final concentration of 20 µM. Stock and
547 sample concentrations were measured using aNanoDrop spectrophotometer using the extinction
548 coefficient 130,000 M^{-1} cm^{-1} for Hrp38 and 78270 for MBP CLAMP 1-300. 1 µM of AlexaFluor
549 labeled CLAMP was added to Hrp38, and TEV protease was added to cleave the MBP tag. Less
550 than 1% fluorescent protein was used in all samples. A Nikon spinning-disc confocal microscope
551 (Nikon Eclipse Ti2) was used for imaging at 20X magnification with 1.5X zoom. The images were
552 processed using ImageJ.
553
554 **NMR Sample Preparation and Spectroscopy:** ^{15}N isotopically labeled samples of CLAMP 1-
555 300 were prepared at 50 µM in a buffer containing 20 mM MES pH 5.5, 100 mM NaCl, 1 mM
556 DTT, and 5% D_2O and then moved into a 5 mm NMR tube using a glass pipette. CLAMP

557 concentration was determined by measuring absorbance at 280 nm (and then dividing absorbance
558 by the extinction coefficient estimated by the Expasy ProtParam). NMR spectra were recorded on
559 Bruker Avance 850 MHz ^{1}H Larmor frequency spectrometer with HCN TCI z-gradient
560 cryoprobes. A two-dimensional $^{1}H^{15}N$ HSQC was acquired using spectral widths of 10.5 ppm and
561 30.0 ppm in the direct and indirect dimensions, with 3072 and 512 total points and acquisition time
562 of 172 ms and 99.0 ms, respectively. Samples of isotopically labeled CLAMP PrLD were prepared
563 at 20 μM in a buffer containing 50 mM MES pH 5 and 5% D_2O. Because the PrLD domain of
564 CLAMP contains no tyrosines or tryptophans, the A_{280} absorbance could not be measured to
565 determine the protein concentration. Instead, sample concentration was estimated by measuring
566 the absorbance at 230 nm and calculating the extinction coefficient to be 300 $M^{-1}cm^{-1}$ per peptide
567 bond (40.5 $mM^{-1}cm^{-1}$ for CLAMP PrLD). Torula yeast was added to one PrLD sample at a 1:1
568 Protein to RNA ratio by weight, and NMR spectra were acquired in 5 mm NMR tubes on a Bruker
569 Avance 850 MHz ^{1}H Larmor frequency spectrometer. Both two-dimensional $^{1}H^{15}N$ HSQC spectra
570 were acquired using spectral widths of 14.0 ppm and 25.0 ppm in the direct and indirect
571 dimensions, with 2048 and 256 total points and acquisition time of 86.0 ms and 59.4 ms,
572 respectively. All data was processed and analyzed using NMRPipe and CCPNMR Analysis 2.5
573 software[50,51].
574
575 **RNA Electrophoretic Mobility Shift Assays:** *rox2*RNA probes at 100 nM and *hrp38* RNA
576 probes at 50nM were incubated with MBP-tagged FL $CLAMP^{WT}$ protein or MBP-tagged FL
577 $CLAMP^{delPrLD}$ protein in REMSA binding buffer provided with the LightShift Chemiluminescent
578 RNA EMSA kit (Thermo Scientific, USA) at room temperature for 30 min according to
579 manufacturer's protocol. Reactions were loaded onto 6% TBE retardation gels (Thermo Fisher
580 Scientific) and run in 0.5× Tris–borate–EDTA buffer for one hour. RNA-Protein complex was
581 transferred to the Nylon membrane using the iBlot transfer system (ThermoFisher Scientific), and
582 the probe signal was detected using a Chemiluminescent Nucleic acid detection module (#80880,
583 ThermoFisher Scientific).
584
585 **Immunostaining of *Drosophila* tissue**:
586 Third instar larval salivary gland tissue from control and $clamp^{delPrLD+6Q}/clamp^{delPrLD}$ male and
587 female dissected in 1XPBS (Phosphate Buffer Saline, 10X PBS, Boston Bioproducts Inc, USA)

588 and fixed in 4% PFA in 1XPBS (Paraformaldehyde, Sigma Aldrich, USA) for 20 min. Incubated
589 in 1% PBST (Triton-X in 1XPBS) for 10 minutes and washed twice in 0.1% PBST for 10 minutes
590 each. They were incubated in a moist chamber with a blocking solution (3% BSA, bovine serum
591 albumin in 0.1%PBST) for 2 hours at RT. Tissues were then incubated in rabbit anti-CLAMP
592 (1:500, SDIX) overnight at 4 degrees Celsius in a blocking solution in a moist chamber. Washed
593 tissues in 0.1% PBST three times for 15 minutes each. Secondary anti-rabbit Alexa fluor 568 at
594 1:200 dilution was used to incubate tissue for 2 hrs at RT in a blocking solution in a moist chamber
595 and again, washed in 0.1% PBST three times for 15 minutes each, followed by DAPI staining
596 (1ug/ml in 1XPBS) for 10 min at RT in the dark. After two washes in 0.1% PBST for 10 min each,
597 tissues were mounted on a slide in Vectashield (Vector Laboratories Inc, USA).
598
599 **Fluorescence recovery after photobleaching (FRAP)**:
600 *In vivo* (Malphighian tubule principal cell nucleus expressing Hrp38GFP in control and
601 $clamp^{delPrLD+6Q}/clamp^{delPrLD}$male and female third instar larvae) FRAP was performed on a Nikhon
602 Spin-disc Confocal Microscope with a 488 nm laser on a 60× objective taking frames without
603 delay (short time course and fast-recovering control) with 3-sec acquisition pre-bleaching,
604 bleaching using 473 nm laser at 100%- and 2-minutes acquisition post-bleach. Malphighian
605 tubules (MTs) were dissected in Grace's Insect media with 1:50 dilution of ProlongLive
606 (antifading agent). A 6% slurry of low melting agarose (A9414-5G) in Grace's Insect media was
607 used to stabilize the MTs during imaging on a bridge slide with a cavity to pour the agarose, which
608 was allowed to solidify, forming a soft base to mount the tissue. The intensities recorded on
609 selected regions of interest were obtained using NIS element software. Data fitting and immobile
610 fraction analysis were obtained with the NIS element software FRAP analysis module.
611
612 ***In-vivo* live imaging:** Salivary gland expressing Hrp38GFP (nuclear) in control and
613 $clamp^{delPrLD+6Q}/clamp^{delPrLD}$ male and female third instar larvae were dissected in Grace's Insect
614 media with 1:50 dilution of ProlongLive (antifading agent) and mounted in the dissecting media
615 on a bridge slide with low melting agarose (A9414-5G) in Grace's Insect media as a base. Moving
616 Hrp38GFP condensates in the nucleus were imaged using a 488 nm laser on a Nikon Spin-disc
617 Confocal Microscope at 60X magnification for 2 minutes each without any delay in acquisition.
618

619

620 **Computational Methods**

621 **CUT&RUN Data analysis:** Sequenced reads were run through FASTQC[43](fastqc
622 replicate_R1_001.fastq.gz replicate_R2_001.fastq.gz) with default parameters to check the quality
623 of raw sequence data and filter out any sequences flagged for poor quality. Sequences were
624 trimmed and reassessed for quality using TrimGalore
625 (https://github.com/FelixKrueger/TrimGalore/issues/25) and FastQC[43], respectively. All Illumina
626 lanes of the same flow cell. fastq files were merged, and sequenced reads were mapped to release
627 6 *Drosophila melanogaster* genome (dm6). We compared Bowtie2[44], HISAT2[45], and BWA[46]. We
628 found the best alignment quality with BWA and thus used the results of this method downstream.
629 Next, we performed conversion to bam and sorting (e.g., using: bowtie2 -x dm6_genome -1
630 replicate_R1_001.fastq.gz -2 replicate_R2_001.fastq.gz -S out. sam> stout.txt 2>
631 alignment_info.txt; samtools view -bSout.sam>out.bam; rm -rfout.sam; samtools sort out.bam -o
632 out.sorted.bam). We removed reads (using samtools) with a MAPQ less than 30 and any reads
633 with PCR duplicate reads (identified using MarkDuplicates Picard -2.20.2). Peaks identified using
634 MACS2[47](macs2 callpeak -t out.sorted.bam -B -f BAM --nomodel --SPMR --keep-dup all -g dm
635 --trackline -n outname --cutoff-analysis --call-summits -p 0.01 --outdiroutdir) and keep duplicates
636 separate**.** To calculate fold-enrichment macs2 is rerun (macs2 bdgcmp -t $treat -c $control -o
637 $out.sorted.bam_FE.bdg -m FE 2> $ out.sorted.bam_FE.log; macs2 bdgcmp -t $treat -c $control
638 -o $out.sorted.bam_logLR.bdg -m logLR -p 0.00001 2). For motif analysis, the MEME[48] suite was
639 used. Data was submitted to the GEO repository (#GSE174781, #GSE220981 and #GSE220053).

640

641 **iCLIP Data analysis:** The method from Heyl et al. 2020[42] using the Galaxy CLIP-Explorer was
642 followed to preprocess, perform quality control, post-process and perform peak calling. UMI tools
643 were used for preprocessing, and then UMI tools and Cutadapt were used for Adapter, Barcode,
644 and UMI removal. Cutadapt (Galaxy version 3.5) was used for filtering with a custom
645 adaptersequenceAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTAT
646 GCCGTCTTCTGCTTG. All other settings followed the Heyl et al. 2020 Galaxy iCLIP-explorer
647 workflow. UMI-Tools Extract (Galaxy Version 1.1.2+galaxy2) was then used with a barcode
648 pattern of NNNXXXXNN. No unpaired reads were allowed. The barcode was on the 3' end. Je-
649 Demultiplex (Galaxy Version 1.2.1) was then used for demultiplexing. FastQC was used for

650 quality control. Mapping was done by RNA STAR (Galaxy version 2.5.2b-2) using dm6. All
651 settings were chosen based on the existing parameters from the iCLIP-explorer settings. We
652 selected FALSE for the option to use end-to-end read alignments with no soft-clipping. bedtools
653 used for Read-Filtering, and UMI-Tools (Galaxy version 0.5.3.0) for de-duplication. PEAKachu
654 was used for Peak Calling to generate bed files. The PEAKachu settings were followed using the
655 Galaxy CLIP-explorer workflow. The maximum insert size was set to 150, the minimum cluster
656 expression fraction was set to 0.01, the minimum block overlap was set to 0.5, and the minimum
657 block expression was set to 0.1. The Mad Multiplier was set to 0.0, the Fold Change Threshold
658 was set to 2.0, and the adjusted p-value threshold was set to 0.05. Peaks were annotated using
659 RCAS[49] (RNA Centric Annotation System), an R package using Rstudio. MEME Suite is used for
660 motif detection. RCAS was used for functional analysis of the transcriptomes isolated by iCLIP,
661 such as transcript features. Data was submitted to the GEO repository (#GSE205987).
662
663 **Integrating CUT&RUN and iCLIP data:** A Python script was created that iterates through all
664 of the DNA peak bed files for CLAMP DNA binding sites in Kc and S2 cell lines (CUT&RUN
665 data, #GSE220053) as a reference and tests for overlap with CLAMP-bound RNA peaks (each
666 sequence is between 25-50bp in size) in the Kc and S2 (iCLIP data, (#GSE205987). The overlaps
667 are categorized into four main categories based upon the location of the overlap: 1) completely
668 overlapping (purple lines in frequency plot), 2) partially overlapping at the DNA peak start site
669 (red lines in frequency plot); 3) partially overlapping at the DNA peak end site (blue lines in
670 frequency plot) and 4) non-overlapping, i.e. when there is an overlap in a region outside the DNA
671 binding site (yellow lines in frequency plot). This extended region is defined by the *scope* variable
672 in the script, allowing the overlap to look for binding sites near the DNA binding site (this scope
673 is 2 kb, including the DNA binding site). We note that multiple RNA peaks can be found on one
674 DNA peak. These overlaps are placed onto a [-scope, scope] region. Then, each type of overlap
675 shown with a different color is overlaid and plotted onto a frequency plot. So, if the frequency at
676 a given base pair is 5, five overlaps contain that base pair within the region defined by the scope.
677
678 **Analysis of Imaging data:**
679

680 ***In vivo* co-localization of CLAMP and Hrp38** Images of the fixed salivary glands
681 immunostained for CLAMP (magenta, anti-CLAMP antibody) and Hrp38 (green, GFP) were
682 acquired as z-stacks and saved as nd2 files. The raw image data from each nd2 file was converted
683 to 3D numpy arrays for each specified channel (magenta and green) employing the nd2 reader
684 Python package. Colocalization analysis was then performed to determine the Pearson Correlation
685 Coefficient (PCC) and corresponding p-value between the two channels of each nd2 file using the
686 scikit-image Python package (scikit-image 0.22.0), indicating the degree of overlap between the
687 proteins (CLAMP and Hrp38). The PCC and p-value were obtained for 20 male control, 38 male
688 $clamp^{delPrLD+6Q}/clamp^{delPrLD}$, 22 female control, and 28 female $clamp^{delPrLD+6Q}/clamp^{delPrLD}$ nd2
689 files. The mean PCC and standard error of the mean (calculated using SciPy) for each background
690 across all nd2 files were then plotted using the matplotlib Python package. Statistical analysis was
691 performed comparing the colocalization coefficients across different experimental conditions,
692 including pairwise p-values and t-values for each experimental background against the others.

693 All image processing and analysis for colocalization was done through a custom Python script.
694 (https://github.com/smritivaidyanathan/3D_Speckle_Volume_Quantifier/blob/main/scripts/coloc
695 alizer.py).

696

697 **Quantification of speckle dynamics; Bound Fraction calculations** Live-imaging movies (2 min
698 long) of dynamically moving GFP tagged Hrp38 protein were pre-processed using Fiji (ImageJ)
699 Python macro script, involving Fiji Plugins and built-in software on the Hrp38 condensate (green
700 channel) to distinguish Hrp38 phase condensates from the cell background. Simple Ratio bleach
701 correction was applied, and the minimum and maximum intensity adjustment was executed based
702 on the mean intensity and standard deviation of intensity. The specific adjustment method varied
703 depending on the image quality, ensuring optimal visibility of Hrp38 condensate speckles.
704 Afterward, background subtraction was performed with a rolling ball radius of 120 pixels. The
705 conversion to a binary image involved max entropy thresholding, although the minimum
706 thresholding method was employed for some images to accommodate diverse lighting conditions.
707 The thresholded images were saved as TIFF files for use in the TrackIt tracking software[34], a
708 program for tracking and analyzing single-molecule experiments developed by the Gebhardt lab.
709 In total, 29 male control movies, 55 male $clamp^{delPrLD+6Q}/clamp^{delPrLD}$ mutant movies, 23 female

control movies, and 65 female *clamp*$^{delPrLD+6Q}$/*clamp*^{delPrLD} mutant movies were utilized in subsequent analyses.

For tracking analysis, TIFF movies from each experimental condition were loaded into TrackIt[34], and analysis was performed using MATLAB on a Windows 10 PC, utilizing default settings with specific parameter adjustments: threshold (4), tracking radius (8), minimum track length (2), one gap frame allowed, and minimum track length before the gap (3). Each experimental condition was analyzed using TrackIt's MATLAB-based data visualization tool, with settings further edited for formatting in a MATLAB script. Tracking data outputs were used for bound fraction analysis according to TrackIt package[34]. TrackIt provided information on tracked events, including long, short, linked, and nonlinked events. The calculation of bound fractions by TrackIt involved determining the ratio of the sum of short and long events to non-linked events. Movie-wise bound fraction values, and the pooled fraction were plotted (sum over all movies for each type of event). Statistical analysis comparing the bound fraction of Hrp38 speckles across different experimental conditions was performed using paired student's t-test for each experimental background against the others. Error bars for the pooled fraction values were calculated using linear error propagation, following the formula indicated in the TrackIt Manual[34] (https://gitlab.com/GebhardtLab/TrackIt/-/blob/master/TrackIt_manual.pdf?ref_type=heads).

Quantification of speckles volume Z-stack confocal images from fixed third instar salivary gland tissue of male and female wildtype control and *clamp*$^{delPrLD+6Q}$/*clamp*^{delPrLD} mutant sample expressing GFP tagged Hrp38 protein in the nucleus were used to determine the size of Hrp38 nuclear speckles in the presence of CLAMP^{WT} and CLAMP^{delPrLD}. We wrote Python script for pre-processing step to create a binary image, with all areas covered by a fluorescently tagged Hrp38 condensate to be represented by a 1 and all others to be marked as 0 using the following steps - a. Conversion of the z-stacks to a 3D array using the nd2reader python package. b. Smoothing the image to reduce noise using a Gaussian filter with a sigma value of 1.0. c. Thresholding the image using the formula, where the threshold was calculated as follows:

Each pixel in the smoothed image is represented by an intensity value ranging from 0 (black) to 255 (white). Following the application of a smoothing algorithm, the arithmetic average (μ) and standard deviation (σ) are calculated across all pixel intensities in the smoothed image.

Thresholding was performed to generate a binary image using these calculated parameters. The threshold value (T) is defined as:

$$T = \mu + (n \times \sigma)$$

where n is a user-defined constant, with n≥0, for each pixel in the smoothed image, if the intensity was greater than T, the corresponding pixel in the binary image was assigned a value of 1 (white). Conversely, if the intensity was less than or equal to T, the corresponding pixel in the binary image was assigned a value of 0 (black). The resulting binary image retained the same dimensions and pixel count as the original smoothed image.

d. Storing and sending to the next step of the pipeline. In the above function, the value of n varied between datasets. It was refined through a visual inspection of the thresholded 3D image using the Napari Viewer python package (https://napari.org), and n was set to equal to 7. In total, 17 male control, 15 male $clamp^{delPrLD+6Q}/clamp^{delPrLD}$ mutant, 17 female control, and 25 female $clamp^{delPrLD+6Q}/clamp^{delPrLD}$ mutant z-stack movies were processed and analyzed. To account for irregular shapes of speckles, the segmentation algorithm union-find (van der Walt et al., 2014) with modified connectivity parameter set as None, meaning that the program will not limit the number of jumps, set to input.ndim, or the size of the array itself (Skimage. measure, 2024) was used. The mean speckle size for each category was plotted with the 95% confidence intervals, calculated using a Z-score of 1.96. The same was done for medians, except using bootstrapping with 1000 samples to calculate the error bars, again calculated using a Z-score of 1.96. The pairwise p-values and t-values were calculated using an independent sample t-test (student t-test). Details of the 3D volume quantification scripts can be found on the Github page: https://github.com/smritivaidyanathan/3D_Speckle_Volume_Quantifier.

Competing Interest Statement

The authors declare no competing interests.

Acknowledgments

This work was supported by R35GM126994 to E.N.L. from NIH, and by National Science Foundation (Directorate for Biological Sciences) 1845734 and R01GM147677 to N.L.F. A.M.C is funded by the NSF Graduate Research Fellowship and CCMB, Brown University. We thank Bloomington Stock Center for fly lines. We thank Steve Henikoff for the pAMNase protein and spike-in DNA for Cut and Run.

Author Contributions

M.R., **N. L. F,** and **E.N.L**. planned experiments, analyzed results, and wrote the manuscript. **J.Z** did all protein purifications and the NMR and phase separation assays for CLAMP. **M.R.** conducted the experimental work and collected data for Cut and Run, iCLIP, FRAP, Live imaging, RNA-EMSA, RNA isolation for RNA-sequencing, and generated fly mutants. **P.M** analyzed third instar larval RNA-seq data using the time2splice pipeline to identify CLAMP-dependent splicing events, integrated CLAMP iCLIP data with CLAMP CUT&RUN in cell lines, and performed the splicing assay for *hrp38*. **S.V.** analyzed all the imaging data sets. **R.V.** performed the co-LLPS assay with Hrp38 and CLAMP. **J.S** cloned CLAMP1-300 and CLAMP1-300del PrLD clones. **J.S.**, **V.C., I.P.,** and **K.C.** helped with RNA-EMSA assays. **A.H.** analyzed the iCLIP-seq data. **A.M.C** analyzed CUT&RUN data. **S.H.W.** helped with NMR analysis. **V.J.** and **N.W.** helped with CLAMP Protein Expression and Purification. **A.E.C** and **R.P** helped with the protein and NMR assay

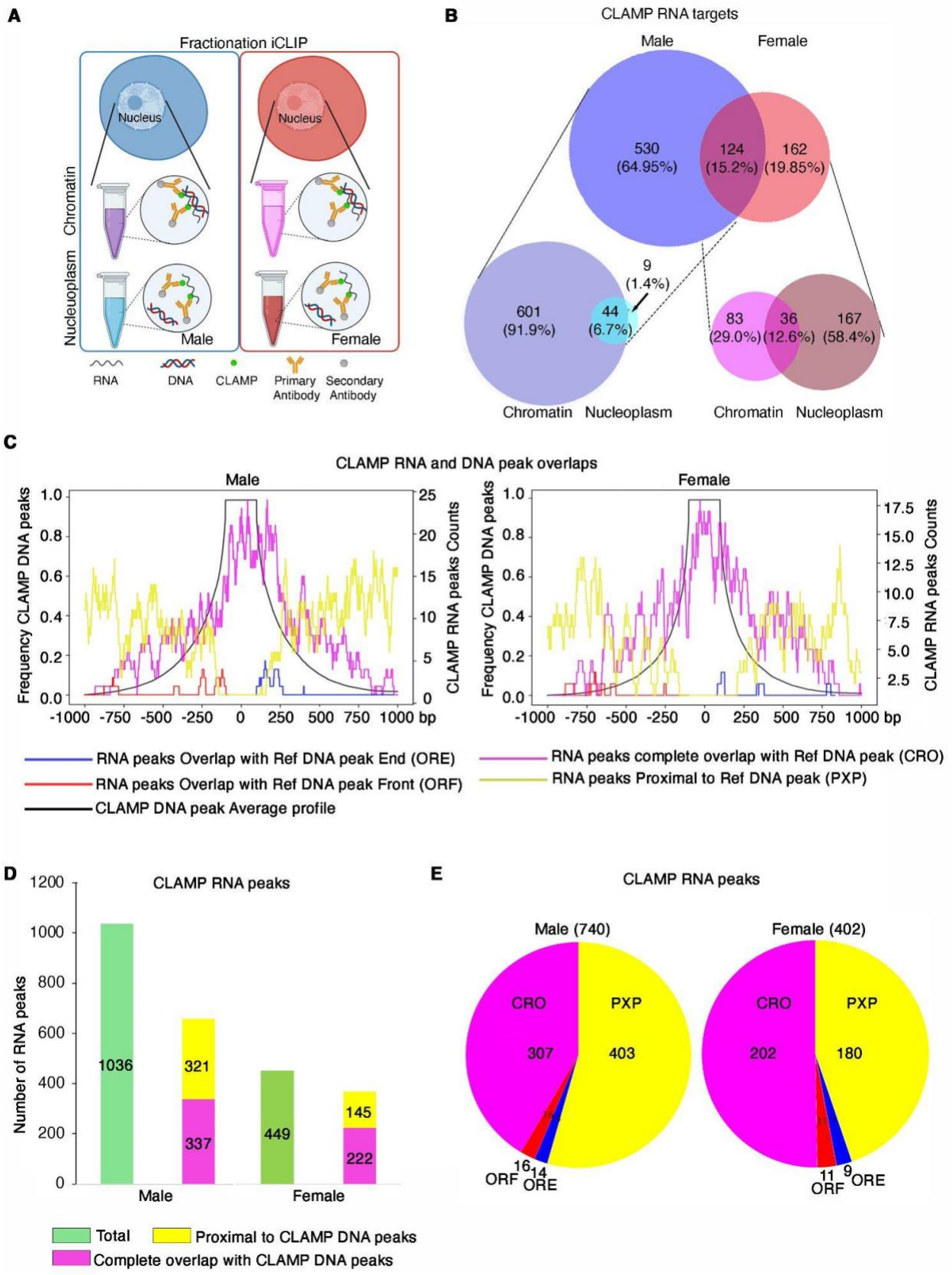
References

1. Boumpas, P., Merabet, S. & Carnesecchi, J. Integrating transcription and splicing into cell fate: transcription factors on the block. *Wiley Interdisciplinary Reviews: RNA***14**, e1752 (2023).
2. Han, H. et al. Multilayered control of alternative splicing regulatory networks by transcription factors. *Molecular cell***65**, 539-553. e7 (2017).
3. Saulnier, O. et al. ERG transcription factors have a splicing regulatory function involving RBFOX2 that is altered in the EWS-FLI1 oncogenic fusion. *Nucleic Acids Research***49**, 5038-5056 (2021).
4. Jabre, I. et al. Does co-transcriptional regulation of alternative splicing mediate plant stress responses? *Nucleic Acids Research***47**, 2716-2726 (2019).
5. Oksuz, O. et al. Transcription factors interact with RNA to regulate genes. *Molecular Cell***83**, 2449-2463. e13 (2023).
6. He, C. et al. High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Molecular cell***64**, 416-430 (2016).
7. Cassiday, L.A. & Maher III, L.J. Having it both ways: transcription factors that bind DNA and RNA. *Nucleic acids research***30**, 4118-4126 (2002).
8. Göös, H. et al. Human transcription factor protein interaction networks. *Nature communications***13**, 766 (2022).
9. Urban, J.A., Urban, J.M., Kuzu, G. & Larschan, E.N. The Drosophila CLAMP protein associates with diverse proteins on chromatin. *PloS one***12**, e0189772 (2017).
10. Kuzu, G. et al. Expansion of GA dinucleotide repeats increases the density of CLAMP binding sites on the X-chromosome to promote Drosophila dosage compensation. *PLoS genetics***12**, e1006120 (2016).
11. Kaye, E.G. et al. Differential occupancy of two GA-binding proteins promotes targeting of the drosophila dosage compensation complex to the male X chromosome. *Cell reports***22**, 3227-3239 (2018).
12. Soruco, M.M. et al. The CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage compensation. *Genes & development***27**, 1551-1556 (2013).
13. Albig, C. et al. Factor cooperation for chromosome discrimination in Drosophila. *Nucleic acids research***47**, 1706-1724 (2019).
14. Ray, M. et al. Sex-specific splicing occurs genome-wide during early Drosophila embryogenesis. *eLife***12**, e87865 (2023).
15. Straub, T., Zabel, A., Gilfillan, G.D., Feller, C. & Becker, P.B. Different chromatin interfaces of the Drosophila dosage compensation complex revealed by high-shear ChIP-seq. *Genome research***23**, 473-485 (2013).
16. Alekseyenko, A.A. et al. A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell***134**, 599-609 (2008).
17. March, Z.M., King, O.D. & Shorter, J. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain research***1647**, 9-18 (2016).
18. Harrison, A.F. & Shorter, J. RNA-binding proteins with prion-like domains in health and disease. *Biochemical Journal***474**, 1417-1438 (2017).
19. Gotor, N.L. et al. RNA-binding and prion domains: the Yin and Yang of phase separation. *Nucleic acids research***48**, 9491-9504 (2020).

20. Wang, Y. et al. A prion-like domain in transcription factor EBF1 promotes phase separation and enables B cell programming of progenitor chromatin. *Immunity***53**, 1151-1167. e6 (2020).
21. Yang, J. et al. A prion-like domain of TFEB mediates the co-aggregation of TFEB and mHTT. *Autophagy***19**, 544-550 (2023).
22. Brodsky, S., Jana, T. & Barkai, N. Order through disorder: The role of intrinsically disordered regions in transcription factor binding specificity. *Current opinion in structural biology***71**, 110-115 (2021).
23. Eggers, N., Gkountromichos, F., Krause, S., Campos-Sparr, A. & Becker, P.B. Physical interaction between MSL2 and CLAMP assures direct cooperativity and prevents competition at composite binding sites. *Nucleic Acids Research***51**, 9039-9054 (2023).
24. Ilik, I.A. et al. A mutually exclusive stem–loop arrangement in roX2 RNA is essential for X-chromosome regulation in Drosophila. *Genes & development***31**, 1973-1987 (2017).
25. Duan, J. et al. CLAMP and Zelda function together to promote Drosophila zygotic genome activation. *Elife***10**, e69937 (2021).
26. Li, S. et al. Genetic interaction of hnRNPA2B1 and DNAJB6 in a Drosophila model of multisystem proteinopathy. *Human molecular genetics***25**, 936-950 (2016).
27. Blanchette, M. et al. Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. *Molecular cell***33**, 438-449 (2009).
28. Borah, S., Wong, A.C. & Steitz, J.A. Drosophila hnRNP A1 homologs Hrp36/Hrp38 enhance U2-type versus U12-type splicing to regulate alternative splicing of the prospero twintron. *Proceedings of the National Academy of Sciences***106**, 2577-2582 (2009).
29. Herold, N. et al. Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes. *Molecular and cellular biology* (2009).
30. Singh, A.K. & Lakhotia, S.C. Dynamics of hnRNPs and omega speckles in normal and heat shocked live cell nuclei of Drosophila melanogaster. *Chromosoma***124**, 367-383 (2015).
31. Mann, R. & Notani, D. Transcription factor condensates and signaling driven transcription. *Nucleus***14**, 2205758 (2023).
32. Hennig, S. et al. Prion-like domains in RNA binding proteins are essential for building subnuclear paraspeckles. *Journal of Cell Biology***210**, 529-539 (2015).
33. Guo, Y.E. et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature***572**, 543-548 (2019).
34. Kuhn, T., Hettich, J., Davtyan, R. & Gebhardt, J.C.M. Single molecule tracking and analysis framework including theory-predicted parameter settings. *Scientific Reports***11**, 9465 (2021).
35. Forman-Kay, J.D., Ditlev, J.A., Nosella, M.L. & Lee, H.O. What are the distinguishing features and size requirements of biomolecular condensates and their implications for RNA-containing condensates? *RNA***28**, 36-47 (2022).
36. Deshaies, J.-E. et al. TDP-43 regulates the alternative splicing of hnRNP A1 to yield an aggregation-prone variant in amyotrophic lateral sclerosis. *Brain***141**, 1320-1333 (2018).
37. Shin, Y. et al. Spatiotemporal control of intracellular phase transitions using light-activated optoDroplets. *Cell***168**, 159-171. e14 (2017).

38. Shin, Y. & Brangwynne, C.P. Liquid phase condensation in cell physiology and disease. *Science***357**, eaaf4382 (2017).
39. Gratz, S.J., Rubinstein, C.D., Harrison, M.M., Wildonger, J. & O'Connor-Giles, K.M. CRISPR-Cas9 genome editing in Drosophila. *Current protocols in molecular biology***111**, 31.2. 1-31.2. 20 (2015).
40. Skene, P.J., Henikoff, J.G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature protocols***13**, 1006-1019 (2018).
41. Huppertz, I. et al. iCLIP: protein–RNA interactions at nucleotide resolution. *Methods***65**, 274-287 (2014).
42. Heyl, F., Maticzka, D., Uhl, M. & Backofen, R. Galaxy CLIP-Explorer: a web server for CLIP-Seq data analysis. *GigaScience***9**, giaa108 (2020).
43. Andrews, S. FastQC: a quality control tool for high throughput sequence data. 2010. (2017).
44. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods***9**, 357-359 (2012).
45. 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).
46. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics***25**, 1754-1760 (2009).
47. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome biology***9**, 1-9 (2008).
48. Bailey, T.L., Johnson, J., Grant, C.E. & Noble, W.S. The MEME suite. *Nucleic acids research***43**, W39-W49 (2015).
49. Uyar, B. et al. RCAS: an RNA centric annotation system for transcriptome-wide regions of interest. *Nucleic acids research***45**, e91-e91 (2017).
50. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR***6**, 277-293 (1995).
51. Skinner, S.P. et al. CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *Journal of biomolecular NMR***66**, 111-124 (2016).
52. Mahableshwarkar, P., Shum, J., Ray, M., Larschan, E., Bind Compare: a novel integrated protein–nucleic acid binding analysis platform, *Bioinformatics*, 40 (11, (2024) btae668.

Figure 1 Ray, Zaborowsky et al 2024



920

921 **Figure 1**. **CLAMP is a dual DNA/RNA-binding protein with sex-specific RNA targets. A.**
922 Schematic showing the design of the fractionation iCLIP experiment to determine CLAMP RNA
923 targets in different nuclear fractions of male and female *Drosophila* cell lines. **B**. Venn diagrams
924 show the distribution of CLAMP RNA targets between male (S2) and female (Kc) cell types and
925 between chromatin (ChF) and nucleoplasm (NF) fractions (**four replicates** for each category
926 performed except **three replicates** for the Kc nucleoplasm fraction). The number within each
927 circle/overlap region denotes the number of RNA targets/genes, and the percentage of RNA
928 targets/gene is shown within the parentheses in each category. **C.** Frequency distribution of
929 CLAMP RNA binding peaks represented as peak counts (right side secondary Y-axis) from iCLIP
930 data (**four replicates** male and female chromatin fraction) plotted over a region (±1kb across the
931 middle of the DNA peak) spanning CLAMP DNA binding peaks (black line, CUT&RUN data,
932 **three replicates** for males, **two replicates** for females). Complete overlaps (CRO) are denoted by
933 magenta, proximal (within ±1kb across the DNA peak),non-overlaps (PXP) in yellow, partial
934 overlaps in red (near the starting boundary of DNA peaks, ORF), and blue (near the ending
935 boundary of DNA peaks, ORE). **D-E.** Bar plots (**D**) and pie-chart (**E**) show the distribution of the
936 number of CLAMP-RNA peaks that overlap with CLAMP-DNA peaks or are within a ±1kb region
937 of CLAMP-DNA peaks (proximal peaks, PXP). Overlapping peaks are sub-categorized into
938 complete RNA peaks overlapping with DNA peaks (CRO), overlapping RNA peaks with DNA
939 peak front (5' end, ORF), and overlapping RNA peaks with DNA peak 3' end (ORE).
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957

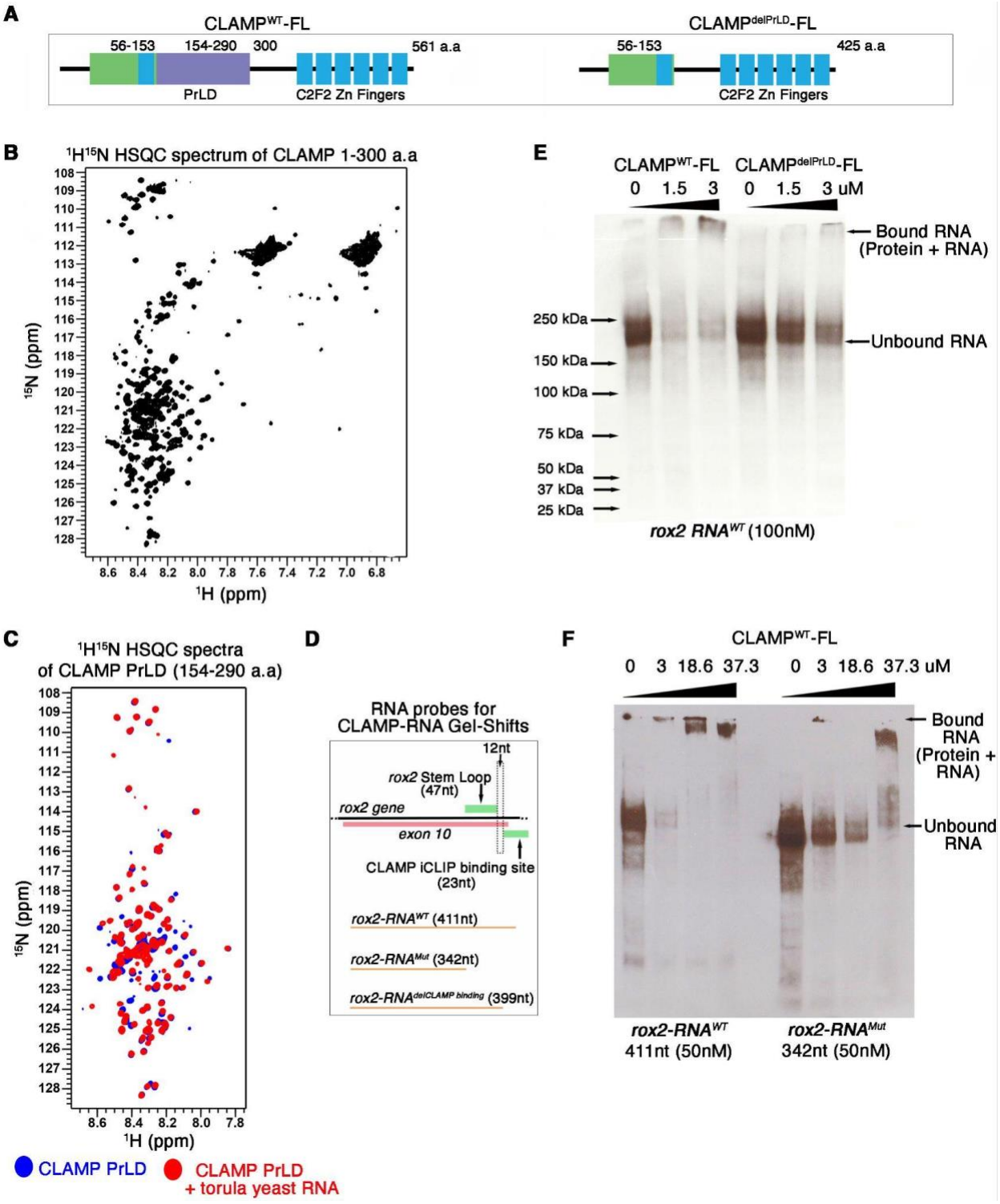
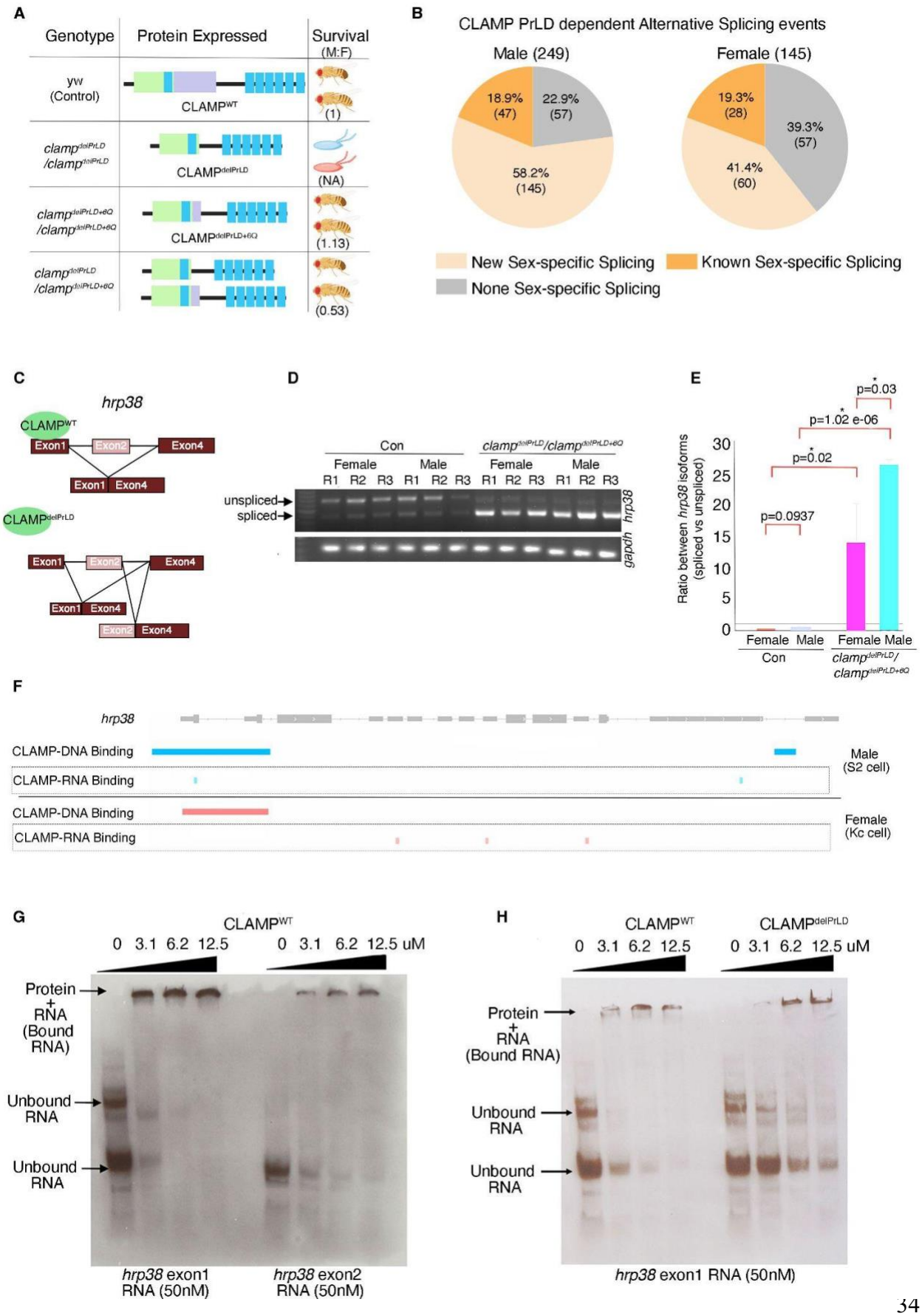


Figure 2. **The PrLD domain of CLAMP is important for Clamp-RNA interaction A.** Schematic of full-length (FL) CLAMP depicting the location of the PrLD (residues 154-290 amino

963 acids) and other essential features in the presence (left) and absence (right) of the PrLD **B.** The
964 $^{1}H^{15}N$ HSQC spectrum of 1-300aa CLAMP at 50 μM demonstrates the largely disordered nature
965 of the protein that lacks the DNA binding domain. Dispersed resonances in the region from 7.9 –
966 7.0 ppm in the ^{1}H and 115-122 ppm in the ^{15}N likely arise from the folded N-terminal zinc finger
967 **C.** $^{1}H^{15}N$ HSQC spectra of CLAMP PrLD at 20 μM with (red) and without (blue) torula yeast
968 RNA, showing changes consistent with binding. **D.** Schematic shows the length of *roX2* RNA
969 probes used to study the role of RNA sequence/structure in CLAMP-RNA interaction with RNAl
970 shift assays (RNA-EMSA). **E.** RNA-EMSA showing the binding of increasing amounts of MBP
971 fusion CLAMP-Full length and CLAMP PrLD domain deleted protein to *roX2* RNA biotinylated
972 probes (100 nM). Concentrations (μM) of CLAMP protein increase from left to right. **F.** RNA-
973 EMSA showing difference in binding of full-length (FL) CLAMP (fused to MBP) to *rox2-RNA*
974 probe with (411nt) and without (342nt) stem-loop and CLAMP-roX2 RNA binding region
975 identified from iCLIP. Concentrations (μM) of CLAMP FL increase from left to right.
976 Concentrations (μM) of CLAMP FL and CLAMPdel PrLD increase from left to right.
977
978


A
Genotype
Protein Expressed
Survival (M:F)
yw (Control)
CLAMPWT
(1)
clampdelPrLD /clampdelPrLD
CLAMPdelPrLD
(NA)
clampdelPrLD+6Q /clampdelPrLD+6Q
CLAMPdelPrLD+6Q
(1.13)
clampdelPrLD /clampdelPrLD+6Q
(0.53)
B
CLAMP PrLD dependent Alternative Splicing events
Male (249)
18.9% (47)
22.9% (57)
58.2% (145)
Female (145)
19.3% (28)
39.3% (57)
41.4% (60)
New Sex-specific Splicing
Known Sex-specific Splicing
None Sex-specific Splicing
C
hrp38
CLAMPWT
Exon1
Exon2
Exon4
Exon1 Exon4
CLAMPdelPrLD
Exon1
Exon2
Exon4
Exon1 Exon4
Exon2 Exon4
D
Con
clampdelPrLD/clampdelPrLD+6Q
Female
Male
Female
Male
R1 R2 R3 R1 R2 R3 R1 R2 R3 R1 R2 R3
unspliced
spliced
hrp38
gapdh
E
Ratio between hrp38 isoforms (spliced vs unspliced)
p=0.03
p=1.02 e-06
p=0.02
p=0.0937
30
25
20
15
10
5
0
Female Male Female Male
Con
clampdelPrLD/ clampdelPrLD+6Q
F
hrp38
CLAMP-DNA Binding
CLAMP-RNA Binding
Male (S2 cell)
CLAMP-DNA Binding
CLAMP-RNA Binding
Female (Kc cell)
G
CLAMPWT
0 3.1 6.2 12.5 0 3.1 6.2 12.5 uM
Protein + RNA (Bound RNA)
Unbound RNA
Unbound RNA
hrp38 exon1 RNA (50nM)
hrp38 exon2 RNA (50nM)
H
CLAMPWT
CLAMPdelPrLD
0 3.1 6.2 12.5 0 3.1 6.2 12.5 uM
Protein + RNA (Bound RNA)
Unbound RNA
Unbound RNA
hrp38 exon1 RNA (50nM)

979

980 **Figure 3. The CLAMP PrLD domain regulates sex-specific splicing in *Drosophila* third instar**
981 **larvae A.** Region of CLAMP-Full length protein deleted to create fly mutants expressing full
982 (*clamp*^{delPrLD}) and partial (*clamp*$^{delPrLD+6Q}$) PrLD deleted CLAMP protein using Crisper-Cas
983 genome editing. Schematic shows corresponding stages till they survive and within parenthesis the
984 Male vs female ratio in adults of those that survive. **B.** Pie chart showing proportion of CLAMP
985 PrLD domain dependent alternative splicing events that is sex-specific in male and female third
986 instar larvae. The number of splicing events are noted within parenthesis. **C.** Schematic of *hrp38*
987 transcript showing CLAMP binding at exon1 and alternative splicing of intron
988 AF:r6:3R:28600568:28600718:+ in *clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$ mutant male. **D-E.** The agarose
989 electrophoretic gel image is shown in **D** and the corresponding bar plot in **E** showing the change
990 in levels of intron AF:r6:3R:28600568:28600718:+ spliced isoforms resulting from alternative
991 splicing events in male and female L3 larvae under control *clamp*^{WT}(green) and
992 *clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$ (orange) conditions. The isoform transcript levels are normalized by
993 the levels of *gapdh* housekeeping gene transcript. p-values (paired student's t-test) for groups
994 showing significant differences are noted at the top of the line connecting the compared groups
995 (**three replicates** for each category). **F.** CLAMP-DNA binding (CUT&RUN data) and CLAMP-
996 RNA binding in chromatin fraction (iCLIP data) peaks visualized in the IGV genome browser at
997 the *hrp38* gene location in female (Kc cells, red) and male (S2 cells, blue). **G-H.** RNA EMSA
998 mobility shift showing a difference in the efficiency of CLAMP full-length protein-*hrp38* exon1
999 and exon2 binding (**G**) and that of CLAMP-PrLD domain deleted and CLAMP full-length protein
1000 with *hrp38* exon1(**H**). CLAMP^{delPrLD} and CLAMP^{WT}-Full length MBP fusions were used at
1001 increasing concentrations of 0, 3.1, 6.2, and 12.5 µM with 50 nM biotinylated RNA probes.
1002
1003
1004

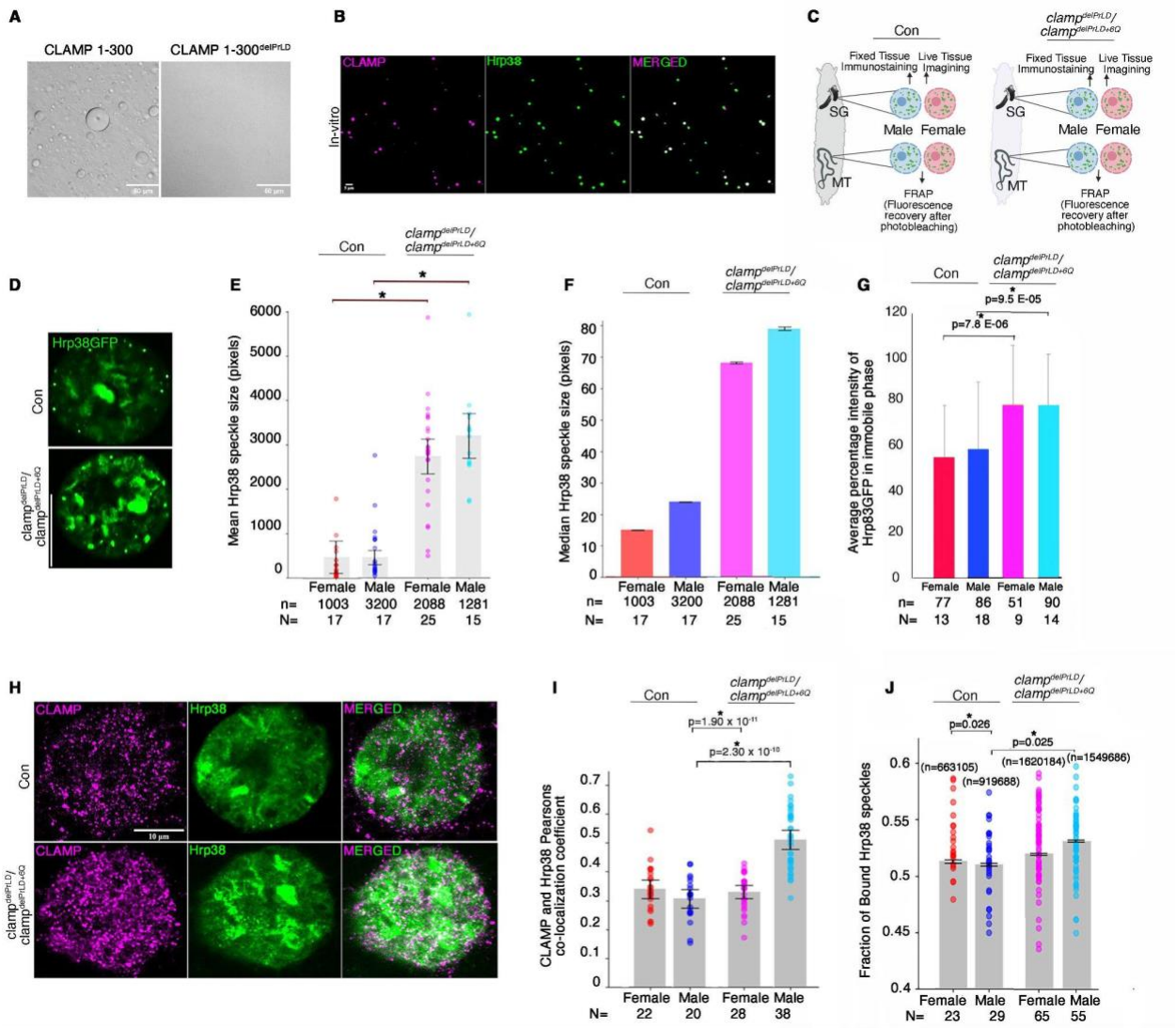


Figure 4. The CLAMP PrLD domain regulates the dynamics of hnRNPA2 homolog Hrp38 biomolecular condensatezs. A. DIC micrograph of CLAMP 1-300 with or without PrLD domain (100 µM) at 150 mM salt concentration with 5% PEG. **B**. Single plane confocal images showing the co-phase separation of Alexa Fluor 488 maleimide dye-labeled Hrp38 with a single cysteine added for labeling (green) and Alexa Fluor 555 maleimide dye-labeled 1-300 CLAMP (pseudo color, magenta) under phase separating conditions for Hrp38. **C**. Schematic showing experimental design setup for studying Hrp38 nuclear condensates in-vivo using fixed and live tissue imaging techniques. **D-F.** Confocal image of a single plane of third instar larval salivary gland nuclei expressing Hrp83-GFP in *clamp*^{WT}(green) and *clamp*^{delPrLD}/*clamp*$^{\text{delPrLD+6Q}}$ conditions in **D** with bar plots show mean (**E**) and median (**F**) Hrp38 speckle/condensate size in control and *clamp*^{delPrLD}/*clamp*$^{\text{delPrLD+6Q}}$ male and females. P-values (paired student's t-test) for groups

showing significant differences in mean speckle size are noted at the top of the line connecting the compared groups. **G**. Bar plots showing the percentage of Hrp38-GFP in the immobile phase (FRAP analysis) in male and female L3 larvae under control *clamp*^{WT} and *clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$ conditions. p-values (paired student's t-test) for groups showing significant differences marked by asterisk are noted at the top of the line connecting the compared groups—**n**=Number of Malphighian tubule principal cell nuclei, **N**= Number of individuals. **H**. Mean bound fraction of Hrp38GFP speckles (Number of Hrp38GFP speckles bound, i.e., immobile track lengths/the total number of Hrp38GFP speckles with all types of track lengths over a fixed time-period) from live imaging of salivary gland nuclei expressing Hrp38-GFP in male and female *clamp*^{WT}and *clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$ individuals. Each dot **N** represents each individual (movie), and '**n**' denotes the number of Hrp38 speckles identified in each category. p-values (paired student's t-test) for groups showing significant differences in the mean fraction of bound Hrp38 speckles are noted at the top of the line connecting the compared groups. **I-J.** Confocal single plane images of immune-stained third instar larval salivary gland nuclei showing in vivo partial co-localization of CLAMP (magenta) and Hrp38 (green) speckles in control and mutant (*clamp*^{delPrLD}/*clamp*$^{delPrLD+6Q}$) males in **I** with corresponding bar plots in **J** showing the Pearson's colocalization coefficient between CLAMP (magenta) and Hrp38 (green) in both males and females. p-values (paired student's t-test) for groups showing significant differences are noted at the top of the line connecting the compared groups.

1054 **Table legends**
1055
1056 **Table 1a:** CLAMP RNA targets identified in the nuclear fractions of male (S2) and female (Kc)
1057 cells.
1058
1059
1060 **Table 1b:** List of genes and genomic locations where CLAMP RNA peaks completely overlap
1061 with CLAMP DNA peaks (CRO), CLAMP RNA peaks overlap with 5' and 3' ends of DNA peaks
1062 (ORF and ORE), and CLAMP RNA peaks are proximal to DNA peaks (PXP) in male (S2) and
1063 female (Kc) cells.
1064
1065 **Table 2a:** List of all CLAMP PrLD-dependent differentially spliced genes in *Drosophila* female
1066 third instar larvae (L3).
1067
1068 **Table 2b:** List all CLAMP PrLD-dependent differentially spliced genes in *Drosophila* male third
1069 instar larvae (L3).
1070
1071 **Table 3:** List of CLAMP PrLD-dependent male and female specifically spliced genes whose RNA
1072 isoforms are direct targets of the CLAMP protein.
1073
1074
1075