CLAMP and Zelda function together to promote *Drosophila* zygotic genome activation

Jingyue Ellie Duan[1,*,§], Leila E. Rieder[2,*], Megan M. Colonnetta[3], Annie Huang[1], Mary McKenney[1], Scott Watters[4], Girish Deshpande[3], William T. Jordan, III[1], Nicolas L. Fawzi[4], and Erica N. Larschan[1,§]

[1]Department of Molecular Biology, Cellular Biology, and Biochemistry, Brown University, Providence, RI, 02912, USA; [2]Department of Biology, Emory University, Atlanta, GA, 30322, USA; [3]Department of Molecular Biology, Princeton University, Princeton, USA 08540; [4]Department of Molecular Pharmacology, Physiology, and Biotechnology, Brown University, Providence, RI, 02912, USA

[*] These authors contributed equally

[§] Correspondence/Lead Contact: erica_larschan@brown.edu; jingyue_duan@brown.edu

AUTHOR INFORMATION

Jingyue E. Duan, jingyue_duan@brown.edu
Leila E. Rieder, leila.rieder@emory.edu
Megan M. Colonnetta, colonnetta@princeton.edu
Annie Huang, annie_huang@brown.edu
Mary McKenney mary_mckenney@brown.edu
Scott Watters scott_watters@brown.edu
Girish Deshpande, gdeshpande@princeton.edu
William T. Jordan, III, william_jordan@alumni.brown.edu
Nicolas Fawzi nicolas_fawzi@brown.edu
Erica N. Larschan, erica_larschan@brown.edu

ABSTRACT

During the essential and conserved process of zygotic genome activation (ZGA), chromatin accessibility must increase to promote transcription. *Drosophila* is a well-established model for defining mechanisms that drive ZGA. Zelda (ZLD) is a key pioneer transcription factor (TF) that promotes ZGA in the *Drosophila* embryo. However, many genomic loci that contain GA-rich motifs become accessible during ZGA independent of ZLD. Therefore, we hypothesized that other early TFs that function with ZLD have not yet been identified, especially those that are capable of binding to GA-rich motifs such as CLAMP. Here, we demonstrate that *Drosophila* embryonic development requires maternal CLAMP to: 1) activate zygotic transcription; 2) increase chromatin accessibility at promoters of specific genes that often encode other essential TFs; 3) enhance chromatin accessibility to facilitate ZLD occupancy at a subset of key embryonic promoters. Thus, maternal CLAMP functions with ZLD in a pioneer-like role to drive zygotic genome activation.

KEYWORDS WORDS Zygotic genome activation, pioneer-like transcription factors, CLAMP, Zelda, *Drosophila* embryo, nucleosomal gel shift

43 **INTRODUCTION**

44

45 During zygotic genome activation (ZGA), dramatic reprogramming occurs in the zygotic
46 nucleus to initiate global transcription and prepare the embryo for further development (Jukam
47 et al., 2017). Chromatin changes that activate the zygotic genome during ZGA rely on
48 cooperation among transcription factors (TFs) (Lee et al., 2014). However, only pioneer TFs
49 (Cirillo and Zaret, 1999; Mayran and Drouin, 2018) can bind to closed chromatin before ZGA
50 because most TFs cannot bind to nucleosomal DNA (Soufi et al., 2015).

51

52 In *Drosophila,* the pioneer TF Zelda (ZLD; Zinc-finger early *Drosophila* activator) plays a
53 key role during ZGA (Liang et al., 2008). ZLD exhibits several critical characteristics of pioneer
54 TFs, including: 1) binding to nucleosomal DNA (Sun et al., 2015; McDaniel et al., 2019); 2)
55 regulating transcription of early zygotic genes (Harrison et al., 2011); and 3) modulating
56 chromatin accessibility to increase the ability of other non-pioneer TFs to bind to DNA (Schulz et
57 al., 2015). However, a large subset of ZLD binding sites (60%) are highly enriched for GA-rich
58 motifs and have constitutively open chromatin even in the absence of ZLD (Schulz et al., 2015).
59 Therefore, we and others (Schulz et al., 2015) hypothesized that other pioneer TFs which
60 directly bind to GA-rich motifs work together with ZLD to activate the zygotic genome.

61

62 GAGA-associated Factor (GAF, Farkas et al., 1994) and Chromatin-Linked Adaptor for
63 Male-Specific Lethal (MSL) Proteins (CLAMP, Soruco et al., 2013) are two of few known TFs
64 that can bind to GA-rich motifs and regulate transcriptional activation in *Drosophila* (Fuda et al.,
65 2015; Kaye et al., 2018). GAF performs several essential functions in early embryos, including
66 chromatin remodeling (Shimojima et al., 2003; Judd et al., 2020; Gaskill et al., 2021) and RNA
67 Pol II recruitment (Li et al., 2013; Fuda et al., 2015; Duarte et al., 2016), and is required for
68 embryonic nuclear divisions (Bhat et al., 1996).

69

70 CLAMP is a GA-binding TF essential for early embryonic development (Rieder et al.,
71 2017) that binds to promoters and plays several vital roles including opening chromatin on the
72 male X-chromosome to recruit the MSL dosage compensation complex (J. Urban et al., 2017;
73 Rieder et al., 2019) and activating coordinated regulation of the histone genes at the histone
74 locus (Rieder et al., 2017). Therefore, we hypothesized that CLAMP functions with ZLD as a
75 pioneer-like factor to promote zygotic genome activation.
76
77 Here, we first demonstrate that depleting maternal CLAMP disrupts transcription of
78 critical early zygotic genes causing significant phenotypic changes in early embryos. Next, we
79 define several mechanisms by which CLAMP regulates zygotic genome activation: 1) CLAMP
80 activates zygotic transcription via direct binding to target genes; 2) CLAMP binds directly to
81 nucleosomal DNA and increases chromatin accessibility of promoters of a subset of genes that
82 often encode other essential TFs; 3) CLAMP and ZLD regulate each other's occupancy at
83 promoters. Overall, we determine that CLAMP is an essential pioneer-like factor that functions
84 with ZLD to regulate zygotic genome activation.
85

86 **RESULTS**

87

88 **Depletion of maternal CLAMP disrupts expression of genes that regulate zygotic**
89 **patterning and cytoskeletal organization in blastoderm embryos**

90

91 We previously reported that nearly 100% (99.87%) of maternal *clamp* RNAi embryos
92 never hatch and die at early embryonic stages (Rieder et al., 2017), demonstrating that
93 maternally-deposited CLAMP is critical for embryonic development. To assess embryonic
94 phenotypic patterning after maternal *clamp* depletion, we first identified three key early zygotic
95 genes (*even-skipped*, *runt,* and *neurotactin,* **Figure 1- figure supplement 1A**) that have
96 significantly (adjusted *p*-value < 0.05, DESeq2, RNA-seq) reduced expression in early embryos
97 when maternal CLAMP is depleted (Rieder et al., 2017). We then used single-molecule
98 fluorescent in situ hybridization (smFISH) for *even-skipped* and *runt,* and immunostaining for
99 Neurotactin to determine how the depletion of maternal *clamp* or *zld* alters phenotypic
100 patterning and cytoskeletal integrity in blastoderm stage embryos **(Figure 1)**. We validated
101 knockdown of *clamp* and *zld* in early embryos by qRT-PCR and western blotting (**Figure 1-**
102 **figure supplement 1B, C** and **Figure 1 Source Data 1-2**).

103

104 Both *even-skipped (eve)* and *runt (run)* play an important role in embryonic
105 segmentation (Manoukian and Krause, 1992). *Eve* also establishes sharp boundaries between
106 parasegments (Fujioka et al., 1995). Strikingly, when maternally-deposited *clamp* is depleted,
107 we observed the complete disruption of classic seven stripe pair-rule gene expression patterns
108 using smFISH **(Figure 1A, middle)**. Additionally, the nuclei in the embryonic syncytium were
109 disassociated compared to control *egfp* RNAi embryos **(Figure 1A, left)**. Furthermore, the
110 expression of *eve* and *run* was significantly reduced in *clamp* maternal depletion embryos that
111 also failed to form sharp stripe boundaries. We observed similar, but slightly stronger,
112 phenotypic changes in *zld* maternal depletion embryos **(Figure 1A, right)**, indicating that
113 CLAMP and ZLD have critical roles in establishing embryonic patterning in pre-cellular

114 blastoderm embryos. Moreover, all of embryos depleted for maternal *clamp* or maternal *zld* (n =
115 10) show defective *eve/run* localization ($p < 0.05$, *Fisher's* exact test).
116
117 Next, we used immunostaining to examine the localization of Neurotactin (NRT), a cell
118 adhesion glycoprotein that is expressed early during *Drosophila* embryonic cellularization in a
119 lattice surrounding syncytial blastoderm nuclei (Hortsch et al., 1990). In *clamp* maternal
120 depletion embryos **(Figure 1B, middle)**, we observed dramatically disrupted cellularization and
121 reduced NRT levels. These embryos fail to form the wild type pattern of cytoskeletal elements,
122 which can be seen in the e*gfp* RNAi control embryos **(Figure 1B, left)**. Embryos depleted for
123 maternal *zld* also reveal similar patterns of discordant nuclei. More than 50% of embryos
124 depleted for maternal *clamp* (n = 23) and 100% of embryos depleted for maternal *zld* (n=10)
125 show NRT disruption **(Figure 1B, right)**, ($p < 0.05$, *Fisher's* exact test). Overall, smFISH and
126 immunostaining results suggest that both maternally-deposited CLAMP and ZLD are essential
127 for early embryonic patterning and development.
128
129 **CLAMP binds to nucleosomal DNA *in vivo* and *in vitro***
130
131 One of the intrinsic characteristics of pioneer transcription factors is their capacity to bind
132 nucleosomal DNA and compacted chromatin (Cirillo and Zaret, 1999). To test the hypothesis
133 that CLAMP is a pioneer-like TF, we performed electromobility shift assays (EMSAs) that test
134 the intrinsic capability of CLAMP to directly interact with nucleosomes *in vitro*. First, we identified
135 a 240 bp region of the X-linked 5C2 locus **(Figure 1 – figure supplement 1D)** that CLAMP
136 binds to in cultured S2 cells and exhibited decreased chromatin accessibility in the absence of
137 CLAMP (J. Urban et al., 2017). This region is also occupied by a nucleosome **(Figure 1 – figure**
138 **supplement 1D)**, suggesting that CLAMP promotes accessibility of this region while binding to
139 nucleosomes.
140
141 We then performed *in vitro* nucleosome assembly using 240 bp of DNA from the 5C2

142 locus that contains three CLAMP-binding motifs, and we used 5C2 naked DNA as a control. We
143 found that both the CLAMP DNA binding domain (DBD**, Figure 1C** and **Figure 1 Source Data**
144 **3**) and full-length protein (FL**, Figure 1D** and **Figure 1 Source Data 4**) can bind and shift both
145 5C2 naked DNA and nucleosomes assembled with 5C2 DNA. Increased protein concentration
146 results in a secondary "super" shift species **(Figure 1C, 1D** and **Figure 1 Source Data 3-4)**,
147 indicating that multiple CLAMP molecules may occupy the three CLAMP-binding motifs. Both
148 full length CLAMP and CLAMP DBD are fused to the maltose binding protein (MBP), which we
149 previously demonstrated does not bind to DNA independent of CLAMP or alter the specificity of
150 CLAMP binding (Kaye et al., 2018). Previously, we determined that CLAMP binds specifically to
151 GAGA repeats *in vivo* and *in vitro* (Soruco et al., 2013; Kaye et al., 2018). Here we further
152 demonstrate that the zinc-finger protein CLAMP can directly bind to nucleosomal DNA and
153 generates multiple shift species consistent with the potential to bind to multiple binding sites
154 simultaneously.
155
156
157 **CLAMP regulates zygotic genome activation (ZGA)**
158
159 To define how CLAMP regulates early embryonic patterning, we examined the effect of
160 maternal CLAMP depletion on the expression of maternally-deposited or zygotically-transcribed
161 genes (Lott et al., 2011) using mRNA-seq data (Rieder et al., 2017). We found that the
162 expression levels of zygotically-transcribed genes but not maternally-deposited genes were
163 significantly downregulated in embryos lacking CLAMP ($p < 0.001$, Mann-Whitney U-test)
164 **(Figure 1E).** Therefore, CLAMP has a specific effect on transcription of zygotic genes similar to
165 that which has been previously reported for ZLD (Liang et al., 2008; Harrison et al., 2011;
166 McDaniel et al., 2019) and confirmed in this study **(Figure 1 – figure supplement 1E)** using
167 stage 5 embryos lacking maternal ZLD (GSE65837, Schulz et al., 2015).
168
169 **CLAMP regulates chromatin accessibility in early embryos**

170

171 An essential characteristic of pioneer transcription factors is that they can establish and
172 maintain the accessibility of their DNA target sites, allowing other TFs to bind to DNA and
173 activate transcription (Zaret and Carroll, 2011; Iwafuchi-Doi et al., 2016). We previously used
174 MNase-seq (J. Urban et al., 2017) to determine that CLAMP guides MSL complex to GA-rich
175 sequences by promoting an accessible chromatin environment on the male X-chromosome in
176 cultured S2 cells. Furthermore, GA-rich motifs are enriched in regions that remain accessible in
177 the absence of the pioneer factor ZLD (Schulz et al., 2015). Therefore, we hypothesized that
178 CLAMP regulates chromatin accessibility at some ZLD-independent GA-rich loci during ZGA.

179

180 To test our hypothesis, we performed Assay for Transposase-Accessible Chromatin
181 using sequencing (ATAC-seq) on 0-2hr (pre-ZGA) and 2-4hr (post-ZGA) embryos with wild-type
182 levels of CLAMP [maternal triple GAL4 driver (MTD) alone (Ni et al., 2011)] and embryos
183 depleted for maternal CLAMP using RNAi driven by the MTD driver (*clamp-i*). We identified
184 differentially accessible (DA) regions **(Figure 2A** and **Figure 2- √ table 1)** by comparing ATAC-
185 seq reads between MTD and *clamp-i* embryos using Diffbind (Stark and Brown, 2019). PCA
186 plots (**Figure 2-figure supplement 1A & D)** show that the first principal component (PC)
187 explains 86%-87% of the variation between MTD and *clamp-i* embryos. However, we also
188 observed that 5-6% of the variation among sample replicates is explained by PC2, suggesting
189 the presence of some developmental diversity within sample groups.

190

191 Despite some variation among replicates, the high Pearson correlation for DA regions
192 between replicates indicates robust reproducibility of these sites **(Figure 2-figure supplement**
193 **1B & E).** We identified a subset of genomic regions that exhibit significantly reduced chromatin
194 accessibility in the absence of CLAMP (**Figure 2A,** 0-2hr: 277; 2-4hr: 50 and **Figure 2-**
195 **supplementary table 1**), indicating that chromatin accessibility of these genomic loci (DA sites)
196 requires CLAMP. Moreover, DA sites include promoters of many genes essential for early
197 embryogenesis such as *Nrt, Abd-B, mod(mdg4),* and *zld,* which encodes the ZLD TF (**Figure**

198 **2A)**. A smaller number of loci (0-2hr: 20; 2-4hr: 36) increased their accessibility in the absence
199 of CLAMP **(Figure 2A)**. Gene Ontology (GO) analysis **(Figure 2-figure supplement 1C & F)**
200 indicates that CLAMP increases accessibility of chromatin regions that are mainly within DNA-
201 binding, RNA Pol II binding, and enhancer-binding TF encoding genes **(Figure 2-figure**
202 **supplement 1C & F)**. While CLAMP strongly regulates chromatin accessibility non-redundantly
203 with other factors at a subset of genomic loci, CLAMP target genes are key for early
204 development, consistent with the dramatic patterning defects observed after depleting maternal
205 CLAMP **(Figure 1A)**.
206
207 Furthermore, a subset [26.7% (74/277) at 0-2hrs; 90% (45/50) at 2-4hrs] of DA regions
208 are directly bound by CLAMP, suggesting that CLAMP directly regulates their chromatin
209 accessibility. For example, the *iab8* promoter, which is located within the essential *Drosophila*
210 *Hox* cluster that controls body plan patterning, is directly bound by CLAMP and shows a
211 reduction in chromatin accessibility after *clamp* RNAi **(Figure 2B)**. We also defined the
212 distribution of DA sites and CLAMP binding sites throughout the genome **(Figure 2C)**. While DA
213 sites were significantly ($p < 0.05$, *Fisher's* exact test) enriched at promoter regions (49.16%),
214 CLAMP binds almost equally frequently to both promoters (29.08%) and introns that are not first
215 introns (27.72%). Therefore, CLAMP is required to establish or maintain open chromatin largely
216 at promoters, but may also play other roles at intronic regions. Furthermore, motif analysis
217 identified both GA-rich motifs and ZLD motifs enriched at regions that require CLAMP for their
218 accessibility in 0-2hr embryos **(Figure 2D)**. These data suggest that CLAMP may also regulate
219 the accessibility of some ZLD binding sites, a hypothesis that we discuss further below.
220
221 We next determined whether CLAMP-mediated chromatin accessibility could specifically
222 drive early transcription by examining the relationship between the chromatin accessibility (DA,
223 ATAC-seq) changes and gene expression of the nearest gene as measured by RNA-seq
224 (Rieder et al., 2017). We observed significant (p-values < 0.05, Mann-Whitney U-test) reduction
225 in expression after maternal CLAMP depletion of genes at which CLAMP mediates chromatin

226 accessibility (DA genes) compared with genes at which CLAMP does not regulate chromatin
227 accessibility (non-DA genes) (**Figure 2E)**. Overall, our results indicate that CLAMP promotes
228 chromatin accessibility and transcription of a subset of other essential TF genes during ZGA,
229 which is consistent with the extensive developmental defects caused by maternal CLAMP
230 depletion.
231
232 **CLAMP and ZLD regulate each other's binding to a subset of promoters**
233
234 To directly determine how CLAMP and ZLD impact each other's binding, we performed
235 ChIP-seq for CLAMP and ZLD in control MTD embryos and embryos that were maternally
236 depleted for each factor with RNAi at the same two time points we used for our ATAC-seq
237 experiments: before ZGA (0-2hr) and during and after ZGA (2-4hr) **(Figure 3A-B, Figure 3-**
238 **figure supplement 1A-B** and **Table 1)**. Overall, there are more ZLD peaks (0-2hr: 6,974; 2-4hr:
239 8,035) across the whole genome than CLAMP peaks (0-2hr: 4,962, 2-4hr: 7,564) in control MTD
240 embryos. As we hypothesized, CLAMP and ZLD peaks significantly overlap ($p < 0.05$,
241 Hypergeometric test, N=15,682 total fly genes) **(Figure 3-figure supplement 1C)**.
242
243 Before ZGA (0-2hrs), ZLD showed higher enrichment at promoters than CLAMP. In
244 contrast, CLAMP had a similar distribution to ZLD at the 2-4hr time point, indicating that CLAMP
245 binds to promoters later than ZLD **(Figure 3C-D)**. Depletion of either maternal *zld* or *clamp*
246 mRNA altered the genomic distribution of CLAMP and ZLD: both factors shifted their occupancy
247 from promoters to introns. Interestingly, maternal *zld* RNAi no longer affects CLAMP binding to
248 promoters at the 2-4hr time point, which may be because the maternal RNAi is not as effective
249 at reducing ZLD protein levels at this time point **(Figure 3-figure supplement 1B)**. Therefore,
250 CLAMP and ZLD change the distribution of each other's occupancy to increase their binding to
251 promoters instead of introns.
252

253 **Table 1. The number of total and differentially bound peaks for CLAMP and ZLD in**

254 **control MTD, *clamp-i* and *zld-i* embryos**

ChIP-seq Peaks	CLAMP			ZLD		
	MTD	*clamp-i*	*zld-i*	*MTD*	*clamp-i*	*zld-i*
0-2hr	4,962	3,488	4,746	6,974	3,687	4,650
2-4hr	7,564	4,064	8,279	8,035	4,687	6,420
Differential Binding (DiffBind, DEseq2)	*MTD vs. zld-i*			*MTD vs. clamp-i*		
	up-DB	down-DB	non-DB	up-DB	down-DB	non-DB
0-2hr	54	390	4,184	8	274	3,144
2-4hr	3	30	7,351	223	1,289	5,672

255

256 Next, we defined the sites that showed differential binding (DB**, Figures 3E-F** and **Table**

257 **1**) of CLAMP and ZLD in the absence of each other's maternally-deposited mRNA using

258 Diffbind (Stark and Brown, 2019). We found a significant reduction of ZLD binding in the

259 absence of CLAMP: there were 274 (0-2hr) and 1,289 (2-4hr) sites where ZLD binding

260 decreased in *clamp-i* embryos compared to MTD controls (down-DB) **(Figures 3E, Figure 3-**

261 **figure supplement 1D,** and **Table 1)**. Fewer ZLD binding sites increased in occupancy after

262 *clamp* RNAi: 8 (0-2hr) and 233 (2-4hr) sites (up-DB). In contrast, loss of ZLD had a weaker

263 impact on CLAMP binding: 390 (0-2 hr) and 30 (2-4 hr) down-DB sites **(Figures 3F, Figure 3-**

264 **figure supplement 1E,** and **Table 1)**. We identified very few sites where CLAMP occupancy

265 increases after *zld* RNAi (up-DB sites: 0-2hr: 54, 2-4hr: 3).

266

267 Moreover, the CLAMP down-DB sites and the ZLD down-DB sites significantly overlap

268 with each other ($p < 0.05$, Hypergeometric test, N=15,682 total fly genes) at both timepoints

269 **(Figure 3-figure supplement 1F)**. For example, *iab-8*, an essential *Hox* cluster gene at which

270 CLAMP regulates chromatin accessibility (DA), is also one of the 95 genomic loci at which

271 CLAMP and ZLD promote each other's occupancy **(Figure 3-figure supplement 1F-G)**.

272

273 To further understand how CLAMP and ZLD bind to dependent (down-DB) and

274 independent (non-DB) sites, we determined the genomic distribution of these two types of sites.

275 Overall, dependent peaks are much broader in size and are located at promoters **(Figure 3G**,
276 **Figure 3-figure supplement 1H** and **Figure 3-figure supplement 2A)**. In contrast,
277 independent sites are narrower and located within introns **(Figure 3G, Figure 3-figure**
278 **supplement 1H** and **Figure 3-figure supplement 2A)**. On average, the peak size of dependent
279 sites (down-DB: 400-500bp) is almost double that of independent sites (non-DB: 200-250bp)
280 with significant differences in peak size for both TFs at both time points ($p < 0.001$, Mann-
281 Whitney U-test) **(Figures 3H)**.
282
283 Previous proteomic studies (J. A. Urban et al., 2017; Hamm et al., 2017, 2015) found no
284 evidence that CLAMP and ZLD directly contact each other at the protein level, suggesting that
285 CLAMP and ZLD regulate each other via binding to their DNA motifs. Therefore, we analyzed
286 the motifs enriched at dependent (down-DB) and independent (non-DB) sites. We found that
287 dependent sites are enriched for motifs specific for the required protein, which are not present at
288 the independent sites **(Figure 3-figure supplement 2B).** For example, the ZLD motif is only
289 enriched at sites where CLAMP requires ZLD for binding (CLAMP down-DB) but not at sites
290 where CLAMP binds independently of ZLD (CLAMP non-DB). Similarly, the CLAMP motif is only
291 enriched at sites where ZLD requires CLAMP for binding (ZLD down-DB) **(Figure 3-figure**
292 **supplement 2B).** Therefore, the presence of specific CLAMP and ZLD motifs correlates with
293 their ability to promote each other's binding.
294
295 Given the cooperative relationship between CLAMP and ZLD binding to chromatin, we
296 measured chromatin accessibility changes at their dependent and independent sites that we
297 defined from ChIP-seq data **(Figures 3I** and **Figure 3-figure supplement 2C).** We found the
298 average ATAC-seq signals were significantly reduced at sites where ZLD is dependent on
299 CLAMP to bind (ZLD down-DB sites) in *clamp-i* embryos compared to MTD controls **(Figure 3I)**.
300 Furthermore, the accessibility at sites where ZLD binds independently of CLAMP (ZLD non-DB)
301 is lower than that at ZLD DB sites but remains unchanged upon *clamp* RNA-i **(Figure 3I)**.

302 Therefore, the chromatin accessibility changes we observe over broader regions are enriched at
303 specific loci where CLAMP promotes ZLD binding.
304
305 Sites where ZLD regulates CLAMP binding (CLAMP down-DB) have high chromatin
306 accessibility while sites where CLAMP binds independently (CLAMP non-DB) of ZLD showed
307 low chromatin accessibility **(Figure 3-figure supplement 2C)**. Interestingly, accessibility slightly
308 increases upon the loss of ZLD at sites where CLAMP requires ZLD for binding at 0-2hr **(Figure**
309 **3-figure supplement 2C).** However, an active TF binding to DNA can prevent Tn5 cleavage at
310 genomic regions (Yan et al., 2020). Therefore, loss of ZLD and CLAMP binding could result in a
311 perceived accessibility increase, as measured by ATAC-seq, which does not necessarily reflect
312 a repressive function for ZLD.
313
314 In summary, CLAMP and ZLD increase each other's occupancy by binding to their
315 motifs and altering chromatin accessibility. These data support a model in which CLAMP and
316 ZLD increase each other's occupancy at promoters of a subset of genes that often encode other
317 transcription factors.
318
319 **CLAMP and ZLD function together to regulate genes during zygotic genome activation**
320
321 CLAMP and ZLD both specifically regulate zygotic transcription **(Figure 1E)** (Liang et al.,
322 2008; Harrison et al., 2011; McDaniel et al., 2019, Rieder et al., 2017). To further understand
323 how CLAMP and ZLD function to regulate ZGA, we compared the transcriptional roles of
324 CLAMP and ZLD in early embryos at genes that have different temporal expression patterns as
325 defined in Li et al. (2014). First, we found that both CLAMP and ZLD are present at genes
326 expressed throughout ZGA although CLAMP binding is more often present at mid- and late-
327 transcribed zygotic genes (categories defined in Li et al., 2014), while ZLD binding is more often
328 present at early-transcribed zygotic genes **(Figure 4A-B)**.
329

330 We next asked whether the ability of CLAMP to bind to genes directly regulates zygotic
331 gene activation by integrating ChIP-seq with RNA-seq data (Schulz et al., 2015; Rieder et al.,
332 2017). We determined that genes strongly bound or weakly bound by CLAMP showed a
333 significant ($p < 0.001$, Mann-Whitney U-test) level of gene expression reduction after *clamp*
334 RNAi (Rieder et al., 2017) compared to unbound genes **(Figure 4C).** We also observed a
335 significant ($p < 0.001$, Mann-Whitney U-test) change in gene expression in maternal *zld-i*
336 embryos (Schulz et al., 2015) for the genes that are strongly bound by ZLD **(Figure 4D)**. Also,
337 the magnitude of the transcriptional changes is similar for genes that are bound by CLAMP or
338 ZLD. Together, these data indicate that CLAMP regulates transcription of zygotic genes by
339 directly binding to target genes.
340
341 Next, we tested the hypothesis that CLAMP and ZLD could regulate each other's binding
342 to precisely drive the transcription of target genes. To test this hypothesis, we plotted the gene
343 expression changes caused by depleting maternal ZLD at genes where CLAMP regulates ZLD
344 binding **(Figure 4E)**. The depletion of maternal *zld* significantly ($p = 4.3e-5$, Mann-Whitney U-
345 test) reduces the expression of genes where ZLD regulates CLAMP binding (down-DB) more
346 than sites where CLAMP binds independently of ZLD (non-DB) at the 0-2hr and 2-4hr ($p =$
347 0.072) time points **(Figure 4E)**. Therefore, ZLD may specifically regulate zygotic genes at which
348 ZLD promotes CLAMP binding. Also, compared to genes where ZLD binds independent of
349 CLAMP, genes where ZLD binding is regulated by CLAMP had a significant ($p < 0.001$, Mann-
350 Whitney U-test) expression reduction after *clamp* RNAi at both 0-2hr and 2-4hr time points
351 **(Figure 4F)**. Thus, CLAMP may regulate the transcription of genes targeted by ZLD by
352 promoting ZLD binding.
353
354 Furthermore, sites where CLAMP and ZLD require each other for binding are enriched
355 for motifs specific for the required protein **(Figure 3-figure supplement 2B)**. Therefore, the
356 presence of specific CLAMP and ZLD motifs correlates with their ability to promote each other's
357 binding which further regulates the expression of each other's target genes.

358

359

360 **CLAMP and ZLD regulate gene expression via modulating chromatin accessibility**

361

362 To determine how direct binding of CLAMP and ZLD relates to zygotic chromatin
363 accessibility, we integrated ChIP-seq and ATAC-seq data. First, we defined four classes of
364 CLAMP-related peaks (DA with CLAMP, DA without CLAMP, Non-DA with CLAMP, Non-DA
365 without CLAMP in **Table 2** and **Table 2 - Source Data 1**). We also obtained ZLD-related ATAC-
366 seq data (Hannon et al., 2017; Soluri et al., 2020) that was generated from embryos laid by
367 wildtype (wt) mothers or mothers with *zld* germline clones (*zld*-) at the nuclear cycle 14 (NC14)
368 +12 min time point and integrated it with ChIP-seq data from the closest time point from this
369 study (0-2hr embryos). In this way, we defined four classes of genomic loci related to ZLD: DA
370 with ZLD, DA without ZLD, Non-DA with ZLD, Non-DA without ZLD **(Table 2** and **Table 2 -**
371 **Source Data 1)**.

372

373 **Table 2. The number of peaks in 4 types of CLAMP or ZLD mediated regions.**

ATAC-seq peaks (0-2hr)	DA w/ CLAMP	DA w/o CLAMP	Non-DA w/ CLAMP	Non-DA w/o CLAMP
16,597	74	203	1,239	15,081
ATAC-seq peaks (Hannon et al., 2017) (NC14+12min)	DA w/ ZLD	DA w/o ZLD	Non-DA w/ ZLD	Non-DA w/o ZLD
19,146	976	2,782	2,010	13,378
CLAMP ZLD co-bound open chromatin regions	Type I (Both DA)	Type II (CLAMP DA)	Type III (ZLD DA)	Type IV (Both non-DA)
525	5	23	123	374

374

375 Next, we generated heatmaps to visualize ATAC-seq read coverage and CLAMP and
376 ZLD ChIP-seq occupancy at their related classes of loci in MTD (wt), *clamp-i* and *zld-i* (*zld*-)
377 embryos **(Figure 5-figure supplement 1A)**. As expected, MTD and *clamp-i* embryo heatmaps
378 revealed that CLAMP-related DA regions (DA with CLAMP, DA without CLAMP) show a

379 significant decrease in accessibility in embryos lacking CLAMP. Regions dependent on ZLD to
380 open (DA with ZLD, DA without ZLD) also show a significant accessibility reduction in the
381 absence of ZLD. Moreover, ChIP-seq read enrichment for protein binding in each RNAi or
382 germline clone embryo class corresponds to our classification.
383
384 Interestingly, both CLAMP and ZLD protein binding to chromatin was reduced when the
385 other TF was depleted, especially at regions where the bound protein is not required for
386 chromatin accessibility (non-DA) **(Figure 5-figure supplement 1A)**. For example, ZLD binding
387 was reduced upon *clamp* RNAi at ZLD non-DA w/ ZLD regions. We also found that CLAMP is
388 enriched at these ZLD non-DA regions, supporting our hypothesis that CLAMP might facilitate
389 ZLD occupancy at some of these loci.
390
391 To assess the relationship between CLAMP and ZLD in regulating chromatin
392 accessibility at loci bound by both factors, we next identified the subset of genomic loci (n=525)
393 that co-bound both CLAMP and ZLD and have open zygotic chromatin **(Figure 5A, Table2** and
394 **Table 2 – Source Data 1).** Next, we divided these co-bound loci into four types: one percent
395 (n=5) of these loci show reduced accessibility after reducing levels of either CLAMP or ZLD
396 (Type I, **Figure 5A**); 23 and 123 loci are specifically dependent on either CLAMP or ZLD for
397 their accessibility, respectively **(**Type II & Type III**, Figure 5A)**; the majority (374 out of 525) of
398 CLAMP ZLD co-bound loci remain open when either protein is absent (Type IV, **Figure 5A**),
399 suggesting that CLAMP and ZLD function redundantly at some of these loci or that other TFs
400 regulate their accessibility.
401
402 Notably, at sites where CLAMP is required for chromatin accessibility (Type II, n=23),
403 ZLD occupancy is entirely ablated in *clamp-i* embryos **(Figure 5A)**. CLAMP binding levels are
404 also reduced after maternal *zld* RNAi at sites where ZLD is required for chromatin accessibility
405 (Type III, n=123). Overall, we observed that CLAMP and/or ZLD occupancy is reduced at most
406 of their co-bound regions when either one of the TFs is depleted, which is consistent with their

407 inter-dependent binding relationship**.** Moreover, *clamp-i* has a stronger impact on ZLD binding
408 than *zld-i* has on CLAMP binding.
409
410 To assess how CLAMP/ZLD-modulated chromatin accessibility impacts transcription, we
411 examined the effect of maternal *clamp* (Rieder et al., 2017) or *zld* (Schulz et al., 2015) depletion
412 on expression of genes that fall into the four types of CLAMP/ZLD co-occupied sites **(Figure**
413 **5B)**. We found that the expression levels of genes (Type II, n = 23) that require CLAMP for
414 chromatin accessibility are significantly ($p < 0.05$, Mann-Whitney U-test) downregulated in
415 embryos lacking CLAMP compared to the Type IV CLAMP and ZLD-independent group (n =
416 374) **(Figure 5B).** Genes (Type III, n = 123) dependent on ZLD for their accessibility also show
417 a significant ($p < 0.001$, Mann-Whitney U-test) reduction in expression upon maternal CLAMP
418 depletion, suggesting CLAMP also might contribute to the regulation of genes at which ZLD
419 regulates chromatin accessibility, likely by increasing ZLD binding.
420
421 In embryos depleted for maternal ZLD (Schulz et al., 2015), we found genes that fall into
422 the Type III ZLD-mediate chromatin accessibility group significantly ($p < 0.001$, Mann-Whitney
423 U-test) decreased in expression **(Figure 5B),** compared with the Type IV (n = 374) group.
424 Interestingly, genes within the CLAMP and ZLD-independent Type IV (n = 374) group do not
425 show significant expression fold changes after depleting either maternal *clamp* or *zld*, indicating
426 that CLAMP and ZLD function redundantly at these loci and/or other proteins play a major role
427 in regulating chromatin accessibility and transcription of these genes.
428
429 Motif analysis demonstrates that CLAMP and ZLD motifs are enriched at genomic loci
430 that are regulated by each factor as well as independent sites (Type IV), in addition to the motif
431 for another GA-binding protein, GAF **(Figure 5-figure supplement 1B**)**.** We next determined
432 whether GAF alters chromatin accessibility at loci at which depletion of CLAMP or ZLD
433 individually alters accessibility (Type IV) and are bound by all three factors. Indeed, we found
434 that approximately 10% of loci that require GAF for their chromatin accessibility (n=104) (Gaskill

435 et al., 2021) overlap with regions where depleting CLAMP or ZLD individually does not alter
436 accessibility **(Figure 5C upper panel)**. When we do not require occupancy of ZLD and CLAMP
437 at their non-DA sites, the overlap is approximately 97% with GAF-dependent loci **(Figure 5C**
438 **lower panel)**. These results suggesting GAF might function at these CLMAP/ZLD independent
439 sites, supporting a model in which multiple TFs coordinately regulate early zygotic chromatin
440 accessibility during ZGA.
441
442 Together, our results reveal the CLAMP and ZLD regulate chromatin accessibility, which
443 alters the occupancy of both factors and regulates zygotic transcription. Furthermore, GAF
444 and/or other TFs might function at sites that are not altered by depleting CLAMP or ZLD
445 individually, suggesting that multiple TFs promote chromatin accessibility during ZGA. It is also
446 possible that CLAMP and ZLD are functionally redundant at the subset of genomic loci at which
447 they regulate each other's occupancy, but depleting either factor individually is not sufficient to
448 alter chromatin and expression.

449 **DISCUSSION**

450

451 Two questions central to early embryogenesis of all metazoans are how and where do
452 early transcription factors work together to drive chromatin changes and zygotic genome
453 activation. Here, we identified CLAMP as a pioneer-like transcription factor that is essential for
454 early embryonic development, which directly binds to nucleosomal DNA **(Figure 1)**, establishes
455 and/or maintains chromatin accessibility at promoters of genes that often encode other TFs
456 **(Figure 2)**, facilitates the binding of ZLD to promoters **(Figure 3)** to regulate zygotic genome
457 activation **(Figure 4** and **Figure 5)**. We further discovered that CLAMP and ZLD regulate each
458 other's binding during ZGA to increase binding to promoters of genes which often encode other
459 early TFs. Also, we determined that both CLAMP and ZLD alter chromatin accessibility, which
460 influences the binding of both TFs to their binding sites **(Figure 6)**. Overall, we provide new
461 insight into how CLAMP and ZLD function together to enhance each other's occupancy and
462 increase chromatin accessibility, which drives zygotic genome activation.

463

464 **CLAMP and ZLD act together to define an open chromatin landscape and activate**
465 **transcription in early embryos**

466

467 We defined multiple classes of CLAMP-dependent and ZLD-dependent genomic loci in
468 early embryos, which provides insight into how CLAMP and ZLD regulate chromatin
469 accessibility and zygotic transcription during ZGA **(Figure 6)**: 1) CLAMP promotes ZLD
470 enrichment at sites where CLAMP increases chromatin accessibility and further regulates ZLD
471 target gene expression. These loci remain open and transcriptionally active even upon ZLD
472 depletion. 2) ZLD facilitates CLAMP occupancy at sites where ZLD regulates chromatin
473 accessibility and promotes CLAMP target gene expression. When maternal CLAMP is depleted,
474 these loci remain accessible and genes are actively transcribed. 3) GAF and/or other TFs could
475 play major roles in opening chromatin at locations co-bound by CLAMP and ZLD but that are
476 not altered in accessibility after depleting CLAMP or ZLD individually. CLAMP and ZLD could

477 also function redundantly at some of these loci because they alter each other's occupancy at
478 these loci but do not change accessibility or expression after depletion of either maternal
479 CLAMP or ZLD individually. Overall, our data suggest that CLAMP functions with ZLD to
480 regulate chromatin accessibility and gene expression of the early zygotic genome.
481
482 Although we have demonstrated an instrumental role for CLAMP in defining a subset of
483 the open chromatin landscape in early embryos, our data show that CLAMP does not increase
484 chromatin accessibility at promoters of all zygotic genes independent of ZLD. Consistent with
485 our results in the early embryo, CLAMP regulates chromatin accessibility at only a few hundred
486 genomic loci in male S2 (258 sites) and female Kc (102 sites) cell lines (Albig et al., 2019).
487 Unlike ZLD, which plays a global role in regulating chromatin accessibility at promoters
488 throughout the genome, depletion of CLAMP alone mainly drives changes at promoters of
489 specific genes that often encode transcription factors that are important for early development,
490 consistent with phenotypic data.
491
492 Moreover, ZLD binding and/or chromatin accessibility is not regulated by maternal
493 depletion of CLAMP at all GA-rich sites in the genome. GAF is also enriched at these same
494 ZLD-bound regions where ZLD is not required for chromatin accessibility (Schulz et al., 2015;
495 Gaskill et al., 2021). Both CLAMP and GAF are deposited maternally (Rieder et al., 2017;
496 Hamm et al., 2017) and bind to similar GA-rich motifs (Kaye et al., 2018). To test whether GAF
497 compensates for the depletion of CLAMP or ZLD, we tried to perform GAF RNAi in the current
498 study to prevent GAF from compensating for CLAMP depletion. However, we and other
499 laboratories could not achieve depletion of GAF in early embryos by RNAi, likely due to
500 autoregulation of its own promoter and its prion-like self-perpetuating function (Tariq et al.,
501 2013).
502
503 We previously demonstrated that competition between CLAMP and GAF at GA-rich
504 binding sites is essential for MSL complex recruitment in S2 cells (Kaye et al., 2018).

505 Furthermore, CLAMP excludes GAF at the histone locus which co-regulates genes that encode
506 the histone proteins (Rieder et al., 2017). However, we also observed synergistic binding
507 between CLAMP and GAF at many additional binding sites (Kaye et al., 2018). The relationship
508 between CLAMP and GAF in early embryos remains unclear. It is very possible that the
509 competitive relationship has not been established in early embryos, since dosage compensation
510 has not yet been initiated (Prayitno et al., 2019). Using GAF-dependent loci defined by Gaskill et
511 al. (2021), we found that genomic loci where GAF functions largely overlap with regions where
512 depletion of CLAMP or ZLD alone does not alter chromatin accessibility, indicating that GAF
513 may function independently of CLAMP or ZLD or is functionally redundant. Future studies are
514 required to distinguish between these models by examining how GAF and CLAMP affect each
515 other's binding to co-bound loci and simultaneously eliminating both factors.
516
517 The GA-rich sequences targeted by CLAMP and GAF are distinct from each other *in*
518 *vivo* and *in vitro*. GAF binding sites typically have 3.5 GA repeats; however, GAF is able to bind
519 to as few as three bases (GAG) within the *hsp70* promoter and *in vitro* (Wilkins and Lis, 1999).
520 In contrast, CLAMP binding sites contain an 8-bp core with a less well-conserved second GA
521 dinucleotide within the core (GA__GAGA) (Alekseyenko et al., 2008). CLAMP binding sites also
522 include a GAGAG pentamer at a lower frequency than GAF binding sites, and flanking bases
523 surrounding the 8-bp core are critical for CLAMP binding (Kaye et al., 2018). Therefore, GAF
524 and CLAMP may have overlapping and non-overlapping functions at different loci, tissues, or
525 developmental stages. Moreover, another TF, Pipsqueak (Psq) also binds to sites containing
526 the GAGAG motif, and has multiple functions during oogenesis andembryonic pattern formation
527 and functions with Polycomb in three-dimensional genome organization (Lehmann et al., 1998;
528 Gutierrez-Perez et al., 2019). In the future, an optogenetic inactivation approach could be used
529 to remove CLAMP, GAF, and/or Psq simultaneously in a spatial and temporal manner
530 (McDaniel et al., 2019).
531
532

533 **CLAMP and ZLD regulate each other's binding via their own motifs**

534

535 ZLD is an essential TF that regulates activation of the first set of zygotic genes during
536 the minor wave of ZGA and thousands of genes transcribed during the major wave of ZGA at
537 nuclear cycle 14 (Liang et al., 2008; Harrison et al., 2011). ZLD also establishes and maintains
538 chromatin accessibility of specific regions and facilitates transcription factor binding and early
539 gene expression (Sun et al., 2015; Schulz et al., 2015). CLAMP regulates histone gene
540 expression (Rieder et al., 2017), X chromosome dosage compensation (Soruco et al., 2013),
541 and establishes/maintains chromatin accessibility (J. Urban et al., 2017). Nonetheless, it
542 remained unclear whether and how CLAMP and ZLD functionally interact during ZGA. Here, we
543 demonstrate that CLAMP and ZLD function together at a subset of promoters that often encode
544 other transcriptional regulators.

545

546 ZLD regulates CLAMP occupancy earlier than CLAMP regulates ZLD occupancy.
547 Genomic loci at which CLAMP is dependent on ZLD early (0-2hr) in development often become
548 independent from ZLD later (2-4hr), with the caveat that ZLD depletion is not as effective later in
549 development. Therefore, CLAMP may require the pioneering activity of ZLD to access specific
550 loci before ZGA, but ZLD may no longer be necessary once CLAMP binding is established.
551 Also, our results suggest that CLAMP is a potent regulator of ZLD binding, especially in 2-4hr
552 embryos. ZLD can bind to many more promoter regions at 0-2 hours, while CLAMP mainly
553 binds to introns early in development but occupies promoters later at 2-4 hours. Therefore,
554 CLAMP may require ZLD to increase chromatin accessibility of these promoter regions (Schulz
555 et al., 2015).

556

557 In addition to its role in embryonic development, CLAMP also plays an essential role in
558 targeting the MSL male dosage compensation complex to the X-chromosome (Soruco et al.,
559 2013). *Drosophila* embryos initiate X-chromosome counting in nuclear cycle 12 and start the sex
560 determination cascade prior to the major wave of ZGA at nuclear cycle 14 (Gergen, 1987;

561 Bosch et al., 2006). However, most dosage compensation is initiated much later in embryonic
562 development (Prayitno et al., 2019). Our data support a model in which CLAMP functions early
563 in the embryo prior to MSL complex assembly to open up specific chromatin regions for MSL
564 complex recruitment (J. Urban et al., 2017; Rieder et al., 2019). Moreover, ZLD likely functions
565 primarily as an early pioneer factor, whereas CLAMP has pioneer-like functions in both early-
566 and late-ZGA embryos. Consistent with this hypothesis, CLAMP binding is enriched at both
567 early and late zygotic genes. In contrast, ZLD binding binds more frequently to early genes,
568 suggesting a that there may be a sequential relationship between occupancy of these two TFs
569 at some loci during early embryogenesis.
570
571 The different characteristics of dependent and independent CLAMP and ZLD binding
572 sites also provide insight into how early transcription factors work together to regulate ZGA. At
573 dependent sites, there are often relatively broad peaks of CLAMP and ZLD that are significantly
574 enriched for clusters of motifs for the required protein. Our CLAMP gel shift assays and those
575 previously reported (Kaye et al., 2018) also show multiple shifted bands consistent with possible
576 multimerization. CLAMP contains two central disordered prion-like glutamine-rich regions (Kaye
577 et al., 2018), a domain that is critical for transcriptional activation and multimerization *in vivo* in
578 several TFs, including GAF (Wilkins and Lis, 1999). Moreover, glutamine-rich repeats alone can
579 be sufficient to mediate stable protein multimerization *in vitro* (Stott et al., 1995). Therefore, it is
580 reasonable to hypothesize that the CLAMP glutamine-rich domain also functions in CLAMP
581 multimerization.
582
583 In contrast, ZLD fails to form dimers or multimers (Hamm et al., 2015, 2017), indicating
584 that ZLD most likely binds as a monomer. There is no evidence that CLAMP and ZLD have any
585 direct protein-protein interaction at sites where they depend on each other to bind. For example,
586 mass spectrometry results that identified dozens of CLAMP-associated proteins did not identify
587 ZLD (J. A. Urban et al., 2017). No data has validated any protein-protein interactions of ZLD
588 with itself as a multimer or between ZLD and any other TFs (Hamm et al., 2017). In the future,

589 simultaneous ablation of maternal CLAMP and ZLD will allow analysis of potential functional
590 redundancy at a subgroup of genomic loci. Our study suggests that regulating the chromatin
591 landscape in early embryos to drive ZGA requires the function of multiple pioneer transcription
592 factors.

593 **MATERIALS AND METHODS**

594

595 **Recombinant Protein Expression and Purification of CLAMP**

596 MBP-tagged CLAMP DBD was expressed and purified as described previously (Kaye et
597 al., 2018). MBP-tagged (pTHMT, Peti and Page, 2007) full-length CLAMP protein was
598 expressed in *Escherichia coli* BL21 Star (DE3) cells (Life Technologies). Bacterial cultures were
599 grown to an optical density of 0.7 to 0.9 before induction with 1 mM isopropyl-β-D-1-
600 thiogalactopyranoside (IPTG) for 4 hrs at 37°C.

601 Cell pellets were harvested by centrifugation and stored at -80°C. Cell pellets were
602 resuspended in 20 mM Tris, 1M NaCl, 0.1mM ZnCl2, 10 mM imidazole pH 8.0 with one EDTA-
603 free protease inhibitor tablet (Roche) and lysed using an Emulsiflex C3 (Avestin). The lysate
604 was cleared by centrifugation at 20,000 rpm for 50 min at 4°C, filtered using a 0.2 µm syringe
605 filter, and loaded onto a HisTrap HP 5 mL column. The protein was eluted with a gradient from
606 10 to 300 mM imidazole in 20 mM Tris, 1.0 M NaCl pH 8.0, and 0.1mM ZnCl2. Fractions
607 containing MBP-CLAMP full-length were loaded onto a HiLoad 26/600 Superdex 200 pg column
608 equilibrated in 20 mM Tris 1.0 M NaCl pH 8.0. Fractions containing full-length CLAMP were
609 identified by SDS-PAGE and concentrated using a centrifugation filter with a 10 kDa cutoff
610 (Amicon, Millipore) and frozen as aliquots.

611

612 ***In vitro* assembly of nucleosomes**

613 The 240 bp 5C2 DNA fragment used for nucleosome *in vitro* assembly was amplified
614 from 276 bp 5C2 fragments (50ng/ul, IDT gBlocks Gene Fragments) by PCR (see 276 bp 5C2
615 and primer sequences below) using OneTaq Hot Start 2X Master Mix (New England Biolabs).
616 The DNA was purified using the PCR clean-up kit (Qiagen) and concentrated to 1ug/ul by
617 SpeedVac Vacuum (Eppendorf). The nucleosomes were assembled using the EpiMark®
618 Nucleosome Assembly Kit (New England Biolabs) following the kit's protocol.

619

620 5C2 (276 bp), **bold** sequences are CLAMP-binding motifs, underlined sequences are primer
621 binding sequences:
622 TCGACGACTAGTTTAAAGTTATTGTAGTTCTTAGAGCAGAATGTATTTTAAATATCAATGTTT
623 CGATGTAGAAATTGAATGGTTTAAATCACGTTCACACAACTTA**GAAAGAGATAG**CGATGGC
624 GGTGT**GAAAGAGAGCGAGATAG**TTGGAAGCTTCATG**GAAATGAAAGAGAGGTAG**TTTTT
625 GGAAATGAAAGTTGTACTAGAAATAAGTATTTTATGTATATAGAATATCGAAGTACAGAAATT
626 CGAAGCGATCTCAACTTGAATATTATATCG
627
628 **Primers for 5C2 region (product is 240bp):**
629 Forward: TTGTAGTTCTTAGAGCAGAATGT
630 Reverse: GTTGAGATCGCTTCGAATTT
631
632 **Electromobility shift assays**
633 DNA or nucleosome probes at 35nM (700fmol/reaction) were incubated with MBP-
634 tagged CLAMP DBD protein or MBP-tagged full-length CLAMP protein in a binding buffer. The
635 binding reaction buffer conditions are similar to conditions previously used to test ZLD
636 nucleosome binding (McDaniel et al. 2019) in 20 ul total volume: 7.5ul BSA/HEGK buffer (12.5
637 mM HEPES, PH 7.0, 0.5 mM EDTA, 0.5 mM EGTA, 5% glycerol, 50 mM KCl, 0.05 mg/ml BSA,
638 0.2 mM PMSF, 1 mM DTT, 0.25 mM ZnCl2, 0.006% NP-40,) 10 ul probe mix (5 ng poly[d-(IC)],
639 5 mM MgCl2, 700 fmol probe), and 2.5 ul protein dilution (0.5uM, 1uM, 2.5uM) at room
640 temperature for 60 min. Reactions were loaded onto 6% DNA retardation gels (ThermoFisher)
641 and run in 0.5X Tris–borate–EDTA buffer for 2 hours. Gels were post stained with GelRed
642 Nucleic Acid Stain (Thermo Scientific) for 30 min and visualized using the ChemiDoc MP
643 imaging system (BioRad).
644
645 **Fly stocks and crosses**
646 To deplete maternally deposited *clamp* or *zld* mRNA throughout oogenesis, we crossed
647 a maternal triple driver (MTD-GAL4, Bloomington, #31777) line (Ni et al., 2011) with a

648 Transgenic RNAi Project (TRiP) *clamp* RNAi line (Bloomington, #57008), a TRiP *zld* RNAi line
649 (from C. Rushlow lab) or *egfp* RNAi line (Bloomington, #41552). The *egfp* RNAi line was used
650 as control in smFISH immunostaining and imaging experiments. The MTD-GAL4 line alone was
651 used as the control line in ATAC-seq and ChIP-seq experiments.
652 Briefly, the MTD-GAL4 virgin females (5-7 day old) were mated with TRiP UAS-RNAi
653 males to obtain MTD-Gal4/UAS-RNAi line daughters. The MTD drives RNAi during oogenesis in
654 these daughters. Therefore, the targeted mRNA is depleted in their eggs. Then MTD-Gal4/UAS-
655 RNAi daughters were mated with males to produce embryos with depleted maternal *clamp* or
656 *zld* mRNA and used for ATAC-seq and ChIP-seq experiments. The embryonic phenotypes of
657 the maternal *zld*$^{-}$ TRiP RNAi line were confirmed previously (Sun et al., 2015). Maternal *clamp*$^{-}$
658 embryonic phenotypes of the TRiP *clamp* RNAi line were confirmed by immunofluorescent
659 staining in our study. Moreover, we validated CLAMP or ZLD protein knockdown in early
660 embryos by western blotting using the Western Breeze kit (Invitrogen) and measured *clamp* and
661 *zld* mRNA levels by qRT-PCR (**Figure 1 – figure supplement 1B, C** and **Figure 1 Source**
662 **Data 1-2**).
663
664 **Embryo collections**
665 To optimize egg collections, young (5-7 day old) females and males were mated. To
666 ensure mothers do not lay older embryos during collections, we first starved flies for 2 hours in
667 the empty cages and discarded the first 2-hour grape agar plates with yeast paste (Plate set
668 #0). When we collected eggs for the experiments, we put flies in the cages with grape agar
669 plates (Plate set #1) with yeast paste for egg laying for 2 hours. Then, we replaced Plate set #1
670 with a new set of plates (Plate set #2) at the 2hr time point. We kept Plate set #1 embryos
671 (without any adult flies) to further develop for another 2 hours to obtain 2-4hr embryos. At the
672 same time, we obtained newly laid 0-2hr embryos from Plate set #2. Therefore, this strategy
673 successfully prevented cross contamination between 0-2hr embryos (Plate set #2) and the 2-4hr
674 embryos (Plate set #1).
675

676 **smFISH, Immunostaining and Imaging**

677 For whole embryo single molecule fluorescence *in situ* hybridization (smFISH) and
678 immunostaining and subsequent imaging, standard protocols were used (Little and Gregor,
679 2018). smFISH probes complementary to *run* were a gift from Thomas Gregor, and those
680 complementary to *eve* were a gift from Shawn Little. The concentrations of the different dyes
681 and antibodies were as follows: Hoechst (Invitrogen, 3µg/ml), Anti-NRT (Developmental Studies
682 Hybridoma Bank BP106, 1:10), AlexaFluor secondary antibodies (Invitrogen Molecular Probes,
683 1:1000). Imaging was done using a Nikon A1 point-scanning confocal microscope with a 40X oil
684 objective. Image processing and intensity measurements were done using Image J software
685 (NIH). Figures were assembled using Adobe Photoshop CS4.

686

687 **ATAC-seq in embryos**

688 We conducted ATAC-seq following the protocol from Blythe and Wieschaus, (2016). 0-
689 2hr or 2-4hr embryos were laid on grape agar plates, dechorionated by 1 min exposure to 6%
690 bleach (Clorox) and then washed 3 times in deionized water. We homogenized 10 embryos and
691 lysed them in 50 ul lysis buffer (10mM Tris 7.5, 10mM NaCl, 3mM MgCl2, 0.1% NP-40). We
692 collected nuclei by centrifuging at 500 g at 4°C and resuspended nuclei in 5 ul TD buffer with
693 2.5 ul Tn5 enzyme (Illumina Tagment DNA TDE1 Enzyme and Buffer Kits). We incubated
694 samples at 37°C for 30 min at 800 rpm (Eppendorf Thermomixer) for fragmentation, and then
695 purified samples with Qiagen MinElute columns before PCR amplification. We amplified libraries
696 by adding 10 ul DNA to 25 ul NEBNext HiFi 2x PCR mix (New England Biolabs) and 2.5 ul of a
697 25 uM solution of each of the Ad1 and Ad2 primers. We used 13 PCR cycles to amplify samples
698 from 0-2hr embryos and 12 PCR cycles to amplify samples from 2-4hr embryos. Next, we
699 purified libraries with 1.2x Ampure SPRI beads. We performed three biological replicates for
700 each genotype (n=2) and time point (n=2). We measured the concentrations of 12 ATAC-seq
701 libraries by Qubit and determined library quality by Bioanalyzer. We sequenced libraries on an
702 Illumina Hi-seq 4000 sequencer at GeneWiz (South Plainfield, NJ) using the 2x150-bp mode.
703 ATAC-seq data is deposited at NCBI GEO and the accession number is GSE152596.

704

705 **Chromatin Immunoprecipitation-sequencing (ChIP-seq) in embryos**

706 We performed ChIP-seq as previously described Blythe and Wieschaus, (2015). We
707 collected and fixed ~100 embryos from each MTD-GAL4 and RNAi cross 0-2hr or 2-4hr after
708 egg lay. We used 3 ul of rabbit anti-CLAMP (Soruco et al., 2013) and 2 ul rat anti-ZLD (from C.
709 Rushlow lab) per sample. We performed three biological ChIP replicates for each protein (n=2),
710 genotype (n=3) and time point (n=2). In total, we prepared 36 libraries using the NEBNext Ultra
711 ChIP-seq kit (New England Biolabs) and sequenced libraries on the Illumina HiSeq 2500
712 sequencer using the 2x150-bp mode. ChIP-seq data is deposited at NCBI GEO and the
713 accession number is GSE152598.

714

715 **Computational analyses**

716 ***ATAC-seq analysis***

717 Prior to sequencing, the Fragment Analyzer showed the library top peaks were in the
718 180-190bp range, which is comparable to the previously established embryo ATAC-seq protocol
719 (Haines, 2017). Demultiplexed reads were trimmed of adapters using TrimGalore (Krueger,
720 2017) and mapped to the *Drosophila* genome dm6 version using Bowtie2 (v. 2.3.0) with option -
721 -very-sensitive, --no-mixed, --no-discordant, --dovetail -X 2000 -k 2. We used Picard tools (v.
722 2.9.2) and SAMtools (v.1.9, Li et al., 2009) to remove the reads that were unmapped, failed
723 primary alignment, or duplicated (-F 1804), and retain properly paired reads (-f 2) with
724 MAPQ >30. After quality trimming and mapping, the Picard tool reported the mean fragment
725 sizes for all ATAC-seq mapped reads is between 125-161bp. As expected, we observed three
726 classes of peaks: 1) A sharp peak at <100 bp (open chromatin); 2) A peak at ~200bp (mono-
727 nucleosome); 3) Other larger peaks (multi-nucleosomes).

728

729 After mapping, we used Samtools to select a fragment size <= 100bp within the bam
730 files to focus on open chromatin. Peak regions for open chromatin regions were called using
731 MACS2 (version 2.1.1, Zhang et al., 2008) with parameters -f BAMPE -g dm --call-summits.

732 ENCODE blacklist was used to filter out problematic regions in dm6 (Amemiya et al., 2019).
733 Bam files and peak bed files were used in DiffBind v.3.12 (Stark and Brown, 2019) for count
734 reads (dba.count), library size normalization (dba.normalize) and calling (dba.contrast)
735 differentially accessible (DA) region with the DESeq2 method. Peak regions (201bp) were
736 centered by peak summits and extended 100bp on each side. Sites were defined as DA with
737 statistically significant differences between conditions using absolute cutoffs of FC > 0.5 and
738 FDR < 0.1. We report all accessible peaks from DiffBind in **Figure 2- supplementary table 1.**
739
740 We used DeepTools (version 3.1.0, Ramírez et al., 2014) to generate enrichment
741 heatmaps (CPM normalization), and average profiles were generated in DeepStats (Gautier
742 RICHARD, 2020). We used 1x depth (reads per genome coverage, RPGC) normalization in
743 Deeptools bamCoverage for making the coverage Bigwig files and uploaded to IGV (Robinson
744 et al., 2011) for genomic track visualizations. Homer (v 4.11, Givler and Lilienthal, 2005) was
745 used for *de novo* motif searches**.** Visualizations and statistical tests were conducted in R (R
746 Core Team, 2014). Specifically, we annotated peaks to their genomic regions using R packages
747 Chipseeker (Yu et al., 2015) and we performed gene ontology enrichment analysis using
748 clusterProfiler (Yu et al., 2012). Boxplot and violin plots were generated using ggplot2
749 (Wickham, 2009) package.
750
751 ***ChIP-seq analysis***

752 Briefly, we trimmed ChIP sequencing raw reads with TrimGalore (Krueger, 2017) with a
753 minimal phred score of 20, 36 bp minimal read length, and Illumina adaptor removal. We then
754 mapped cleaned reads to the *D. melanogaster* genome (UCSC dm6) with Bowtie2 (v. 2.3.0)
755 with the –very-sensitive-local flag feature. We used Picard tools (v. 2.9.2) and SAMtools (v.1.9,
756 Li et al., 2009) to remove the PCR duplicates. We used MACS2 (version 2.1.1, Zhang et al.,
757 2008) to identify peaks with default parameters and MSPC (v.4.0.0, Jalili et al., 2015) to obtain
758 consensus peaks from 3 replicates. The peak number for each sample was summarized in

759 Table 4. ENCODE blacklist was used to filter out problematic regions in dm6 (Amemiya et al.,
760 2019). We identified differential binding (DB) and non-differential binding (non-DB) between
761 MTD and RNAi samples using DiffBind (v. 3.10, Stark and Brown, 2019) with the DESeq2
762 method. Peak regions (501bp) were centered by peak summits and extended 250bp on each
763 side. The DB and non-DB peak numbers are summarized in **Table 1**. Differential binding was
764 defined with absolute FC > 0.5 and FDR < 0.05 (**Table 1 – Source Data 1**).

765 We used DeepTools (version 3.1.0, Ramírez et al., 2014) to generate enrichment
766 heatmaps and average profiles. Bigwig files were generated with DeepTools bamCompare
767 (scale factor method: SES; Normalization: log2) and uploaded to IGV (Robinson et al., 2011) for
768 genomic track visualization. We used Homer (v 4.11, Givler and Lilienthal, 2005) for *de novo*
769 motif searches and genomic annotation. Intervene (Khan and Mathelier, 2017) was used for
770 intersection and visualization of multiple peak region sets. Visualizations and statistical tests
771 were conducted in R (R Core Team, 2014). Specifically, we annotated peaks to their genomic
772 regions using the R package Chipseeker **(**Yu et al., 2015) and we did gene ontology enrichment
773 analysis using clusterProfiler (Yu et al., 2012). Boxplots and violin plots were generated using
774 the ggplot2 (Wickham, 2009) package**.**
775

776 ***ATAC-seq and ChIP-seq data integration***

777 We used Bedtools (Quinlan and Hall, 2010) intersection tool to intersect peaks in
778 CLAMP ChIP-seq binding regions with CLAMP DA or non-DA peaks. Based on the intersection
779 of the peaks, we defined 4 types of CLAMP related peaks: 1) DA with CLAMP, 2) DA without
780 CLAMP, 3) Non-DA with CLAMP, 4) Non-DA without CLAMP. Similarly, we defined ZLD related
781 peaks by intersecting ZLD DA or non-DA peaks and ATAC-seq datasets (Hannon et al., 2017;
782 Soluri et al., 2020) from wildtype (wt) and *zld* germline clone (*zld*-) embryos at the nuclear
783 cycle14 +12 min stage. Specifically, we defined four classes of genomic loci for ZLD-related
784 classes: 1) DA with ZLD, 2) DA without ZLD, 3) Non-DA with ZLD, 4) Non-DA, without ZLD. We
785 used DeepTools (version 3.1.0, Ramírez et al., 2014) to generate enrichment heatmaps for

786 each subclass of peaks. Peaks locations in each CLAMP or ZLD-related category were
787 summarized in **Table 2 - Source Data 1**.
788
789 ***ATAC-seq and RNA-seq data integration***
790 We annotated genes near differential (down-DA) ATAC-seq peaks in R using
791 detailRanges function from the csaw package (Lun and Smyth, 2016). Then we plotted
792 expression of genes using previously published RNA-seq data (Rieder et al., 2017).
793
794 ***ChIP-seq and RNA-seq data integration***
795 To define strong, weak, and unbound genes close to peaks in CLAMP or ZLD ChIP-seq
796 data, we used the peak binding score reported in MACS2 -log10(p-value) of 100 as a cut off
797 value. We defined the following categories: 1) Strong binding peaks: score greater than 100; 2)
798 Weak binding peak: score lesser than 100; 3) Unbound peaks: the rest of the peaks that are
799 neither strong or weak. Then, we annotated all peaks using Homer annotatePeaks (v 4.11,
800 Givler and Lilienthal, 2005). We then obtained the log2 fold change (*clamp-i/*MTD or *zld-i/yw*) of
801 gene expression in the RNA-seq dataset for each protein binding group: CLAMP (Rieder et al.,
802 2017) or ZLD (Schulz et al., 2015). Boxplots and violin plots were generated using the ggplot2
803 (Wickham, 2009) package.
804
805 ***Datasets***
806 RNA-seq datasets from wild type and maternal *clamp* depletion by RNAi were from
807 GSE102922 (Rieder et al., 2017). RNA-seq datasets from *yw* wild type and *zld* maternal RNAi
808 were from GSE65837 (Schulz et al., 2015). ATAC-seq data from wild type and *zld* germline
809 clones were from GSE86966 (Hannon et al., 2017). Processed ATAC-seq data identifying
810 differential peaks between wild type and *zld* germline mutations were from Soluri et al., (2020).
811
812 ***Data accessibility for reviewers only***

ATAC-seq (GSE152596) and ChIP-seq (GSE152598) are under GEO SuperSeries GSE152613. To review GEO accession GSE152613:

Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152613

Enter token ihevsmiqnxexrod into the box

ACKNOWLEDGEMENTS

We thank Melissa Harrison, Tyler Gibson and Marissa Gaskill for sending the GAF-dependent region bed file and helpful discussions. This work was supported by NIH grant F32GM109663, K99HD092625 and R00HD092625 to Dr. Leila Rieder and R35GM126994 to Dr. Erica Larschan, and in part by NSF grant 1845734 and NIH grant R01GM118530 to Dr. Nicolas L. Fawzi.

AUTHOR CONTRIBUTIONS

Conceptualization, L.E.R., J.E.D. and E.N.L.; Methodology, J.E.D., L.E.R. and E.N.L; ChIP-seq Experiment, L.E.R.; ATAC-seq Experiment, J.E.D.; Initial Analysis, W.J.III; Formal Analysis, J.E.D.; Microscopy, immunostaining and smFISH: M.M.C. and G.D. Protein Expression: M.M., S.W., and N.L.F.; Gel-shift Experiment, A.H and J.E.D.; Investigation, J.E.D; Data Curation, J.E.D; Writing--Original Draft, J.E.D. and E.N.L.; Writing--Review & Editing, J.E.D., L.E.R. and E.N.L.; Visualization, J.E.D.; Funding Acquisition, L.E.R., E.N.L. and N.L.F.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Albig C, Tikhonova E, Krause S, Maksimenko O, Regnard C, Becker PB. 2019. Factor cooperation for chromosome discrimination in Drosophila. *Nucleic Acids Research* **47**:1706–1724. doi:10.1093/nar/gky1238

Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee O-K, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ, Kuroda MI. 2008. A Sequence Motif within Chromatin Entry Sites Directs MSL Establishment on the Drosophila X Chromosome. *Cell* **134**:599–609. doi:10.1016/j.cell.2008.06.033

Amemiya HM, Kundaje A, Boyle AP. 2019. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports* **9**:9354. doi:10.1038/s41598-019-45839-z

Bhat KM, Farkas G, Karch F, Gyurkovics H, Gausz J, Schedl P. 1996. The GAGA factor is required in the early Drosophila embryo not only for transcriptional regulation but also for nuclear division. *Development* **122**:1113–1124.

Blythe SA, Wieschaus EF. 2016. Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. *eLife* **5**:e20148. doi:10.7554/eLife.20148

Blythe SA, Wieschaus EF. 2015. Zygotic Genome Activation Triggers the DNA Replication Checkpoint at the Midblastula Transition. *Cell* **160**:1169–1181. doi:10.1016/j.cell.2015.01.050

Bosch JR ten, Benavides JA, Cline TW. 2006. The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription. *Development* **133**:1967–1977. doi:10.1242/dev.02373

Cirillo LA, Zaret KS. 1999. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* **4**:961–969. doi:10.1016/s1097-2765(00)80225-7

Duarte FM, Fuda NJ, Mahat DB, Core LJ, Guertin MJ, Lis JT. 2016. Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes Dev* **30**:1731–1746. doi:10.1101/gad.284430.116

Farkas G, Gausz J, Galloni M, Reuter G, Gyurkovics H, Karch F. 1994. The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature* **371**:806–808. doi:10.1038/371806a0

Fuda NJ, Guertin MJ, Sharma S, Danko CG, Martins AL, Siepel A, Lis JT. 2015. GAGA Factor Maintains Nucleosome-Free Regions and Has a Role in RNA Polymerase II Recruitment to Promoters. *PLoS Genet* **11**. doi:10.1371/journal.pgen.1005108

Fujioka M, Jaynes JB, Goto T. 1995. Early even-skipped stripes act as morphogenetic gradients at the single cell level to establish engrailed expression. *Development* **121**:4371–4382.

Gaskill MM, Gibson TJ, Larson ED, Harrison MM. 2021. GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. *eLife* **10**:e66668. doi:10.7554/eLife.66668

Gautier RICHARD. 2020. gtrichard/deepStats: New tools and much needed fixes. Zenodo. doi:10.5281/zenodo.3668336

Gergen JP. 1987. Dosage Compensation in Drosophila: Evidence That daughterless and Sex-lethal Control X Chromosome Activity at the Blastoderm Stage of Embryogenesis. *Genetics* **117**:477–485.

Givler T, Lilienthal P. 2005. Using HOMER Software, NREL's Micropower Optimization Model, to Explore the Role of Gen-sets in Small Solar Power Systems; Case Study: Sri Lanka (No. NREL/TP-710-36774). National Renewable Energy Lab., Golden, CO (US). doi:10.2172/15016073

Gutierrez-Perez I, Rowley MJ, Lyu X, Valadez-Graham V, Vallejo DM, Ballesta-Illan E, Lopez-Atalaya JP, Kremsky I, Caparros E, Corces VG, Dominguez M. 2019. Ecdysone-Induced 3D Chromatin Reorganization Involves Active Enhancers Bound by Pipsqueak and Polycomb. *Cell Reports* **28**:2715-2727.e5. doi:10.1016/j.celrep.2019.07.096

Haines J. 2017. ATAC-seq on nuclei from frozen, sliced, Drosophila melanogaster embryo halves. doi:10.17504/protocols.io.kj5cuq6

Hamm DC, Bondra ER, Harrison MM. 2015. Transcriptional Activation Is a Conserved Feature of the Early Embryonic Factor Zelda That Requires a Cluster of Four Zinc Fingers for DNA Binding and a Low-complexity Activation Domain. *J Biol Chem* **290**:3508–3518. doi:10.1074/jbc.M114.602292

Hamm DC, Larson ED, Nevil M, Marshall KE, Bondra ER, Harrison MM. 2017. A conserved maternal-specific repressive domain in Zelda revealed by Cas9-mediated mutagenesis in Drosophila melanogaster. *PLoS Genet* **13**. doi:10.1371/journal.pgen.1007120

Hannon CE, Blythe SA, Wieschaus EF. 2017. Concentration dependent chromatin states induced by the bicoid morphogen gradient. *eLife* **6**:e28275. doi:10.7554/eLife.28275

Harrison MM, Li X-Y, Kaplan T, Botchan MR, Eisen MB. 2011. Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. *PLOS Genetics* **7**:e1002266. doi:10.1371/journal.pgen.1002266

Hortsch M, Patel NH, Bieber AJ, Traquina ZR, Goodman CS. 1990. Drosophila neurotactin, a surface glycoprotein with homology to serine esterases, is dynamically expressed during embryogenesis. *Development* **110**:1327–1340.

Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, Lee D, Kaestner KH, Zaret KS. 2016. The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell* **62**:79–91. doi:10.1016/j.molcel.2016.03.001

Jalili V, Matteucci M, Masseroli M, Morelli MJ. 2015. Using combined evidence from replicates to evaluate ChIP-seq peaks. *Bioinformatics* **31**:2761–2769. doi:10.1093/bioinformatics/btv293

Judd J, Duarte FM, Lis JT. 2020. Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription. *Genes Dev*. doi:10.1101/gad.341768.120

Jukam D, Shariati SAM, Skotheim JM. 2017. Zygotic Genome Activation in Vertebrates. *Dev Cell* **42**:316–332. doi:10.1016/j.devcel.2017.07.026

Kaye EG, Booker M, Kurland JV, Conicella AE, Fawzi NL, Bulyk ML, Tolstorukov MY, Larschan E. 2018. Differential Occupancy of Two GA-Binding Proteins Promotes Targeting of the Drosophila Dosage Compensation Complex to the Male X Chromosome. *Cell Rep* **22**:3227–3239. doi:10.1016/j.celrep.2018.02.098

Khan A, Mathelier A. 2017. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* **18**:287. doi:10.1186/s12859-017-1708-7

Krueger F. 2017. Trim Galore: a wrapper script to automate quality and adapter trimming as well as quality control. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Lee MT, Bonneau AR, Giraldez AJ. 2014. Zygotic genome activation during the maternal-to-zygotic transition. *Annu Rev Cell Dev Biol* **30**:581–613. doi:10.1146/annurev-cellbio-100913-013027

Lehmann M, Siegmund T, Lintermann K-G, Korge G. 1998. The Pipsqueak Protein of Drosophila melanogasterBinds to GAGA Sequences through a Novel DNA-binding Domain. *J Biol Chem* **273**:28504–28509. doi:10.1074/jbc.273.43.28504

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352

Li J, Liu Y, Rhee HS, Ghosh SKB, Bai L, Pugh BF, Gilmour DS. 2013. Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Mol Cell* **50**:711–722. doi:10.1016/j.molcel.2013.05.016

Li X-Y, Harrison MM, Villalta JE, Kaplan T, Eisen MB. 2014. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. *eLife* **3**:e03737. doi:10.7554/eLife.03737

Liang H-L, Nien C-Y, Liu H-Y, Metzstein MM, Kirov N, Rushlow C. 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**:400–403. doi:10.1038/nature07388

Little SC, Gregor T. 2018. Single mRNA molecule detection in Drosophila. *Methods Mol Biol* **1649**:127–142. doi:10.1007/978-1-4939-7213-5_8

Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. 2011. Noncanonical Compensation of Zygotic X Transcription in Early Drosophila melanogaster Development Revealed through Single-Embryo RNA-Seq. *PLoS Biol* **9**. doi:10.1371/journal.pbio.1000590

Lun ATL, Smyth GK. 2016. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res* **44**:e45. doi:10.1093/nar/gkv1191

Manoukian AS, Krause HM. 1992. Concentration-dependent activities of the even-skipped protein in Drosophila embryos. *Genes Dev* **6**:1740–1751. doi:10.1101/gad.6.9.1740

Mayran A, Drouin J. 2018. Pioneer transcription factors shape the epigenetic landscape. *J Biol Chem* **293**:13795–13804. doi:10.1074/jbc.R117.001232

McDaniel SL, Gibson TJ, Schulz KN, Fernandez Garcia M, Nevil M, Jain SU, Lewis PW, Zaret KS, Harrison MM. 2019. Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation. *Mol Cell* **74**:185-195.e4. doi:10.1016/j.molcel.2019.01.014

Ni J-Q, Zhou R, Czech B, Liu L-P, Holderbaum L, Yang-Zhou D, Shim H-S, Tao R, Handler D, Karpowicz P, Binari R, Booker M, Brennecke J, Perkins LA, Hannon GJ, Perrimon N. 2011. A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nat Methods* **8**:405–407. doi:10.1038/nmeth.1592

Peti W, Page R. 2007. Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. *Protein Expr Purif* **51**:1–10. doi:10.1016/j.pep.2006.06.024

Prayitno K, Schauer T, Regnard C, Becker PB. 2019. Progressive dosage compensation during Drosophila embryogenesis is reflected by gene arrangement. *EMBO Rep* **20**:e48138. doi:10.15252/embr.201948138

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. doi:10.1093/bioinformatics/btq033

R Core Team. 2014. R: A Language and Environment for Statistical Computing.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**:W187–W191. doi:10.1093/nar/gku365

Rieder LE, Jordan WT, Larschan EN. 2019. Targeting of the Dosage-Compensated Male X-Chromosome during Early Drosophila Development. *Cell Reports* **29**:4268-4275.e2. doi:10.1016/j.celrep.2019.11.095

Rieder LE, Koreski KP, Boltz KA, Kuzu G, Urban JA, Bowman SK, Zeidman A, Jordan WT, Tolstorukov MY, Marzluff WF, Duronio RJ, Larschan EN. 2017. Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP. *Genes Dev* **31**:1494–1508. doi:10.1101/gad.300855.117

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**:24–26. doi:10.1038/nbt.1754

Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, McKay DJ, Harrison MM. 2015. Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res* **25**:1715–1726. doi:10.1101/gr.192682.115

Shimojima T, Okada M, Nakayama T, Ueda H, Okawa K, Iwamatsu A, Handa H, Hirose S. 2003. Drosophila FACT contributes to Hox gene expression through physical and functional interactions with GAGA factor. *Genes Dev* **17**:1605–1616. doi:10.1101/gad.1086803

Soluri IV, Zumerling LM, Payan Parra OA, Clark EG, Blythe SA. 2020. Zygotic pioneer factor activity of Odd-paired/Zic is necessary for late function of the Drosophila segmentation network. *eLife* **9**:e53916. doi:10.7554/eLife.53916

Soruco MML, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, Sugden AU, Goebel K, Feng J, Xia P, Vedenko A, Bulyk ML, Park PJ, Larschan E. 2013. The CLAMP protein

links the MSL complex to the X chromosome during Drosophila dosage compensation. *Genes Dev* **27**:1551–1556. doi:10.1101/gad.214585.113

Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. 2015. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**:555–568. doi:10.1016/j.cell.2015.03.017

Stark R, Brown G. 2019. DiffBind: Differential Binding Analysis of ChIP-Seq Peak Data. Bioconductor version: Release (3.10). doi:10.18129/B9.bioc.DiffBind

Stott K, Blackburn JM, Butler PJ, Perutz M. 1995. Incorporation of glutamine repeats makes protein oligomerize: implications for neurodegenerative diseases. *PNAS* **92**:6509–6513. doi:10.1073/pnas.92.14.6509

Sun Y, Nien C-Y, Chen K, Liu H-Y, Johnston J, Zeitlinger J, Rushlow C. 2015. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. *Genome Res* **25**:1703–1714. doi:10.1101/gr.192542.115

Tariq M, Wegrzyn R, Anwar S, Bukau B, Paro R. 2013. Drosophila GAGA factor polyglutamine domains exhibit prion-like behavior. *BMC Genomics* **14**:374. doi:10.1186/1471-2164-14-374

Urban J, Kuzu G, Bowman S, Scruggs B, Henriques T, Kingston R, Adelman K, Tolstorukov M, Larschan E. 2017. Enhanced chromatin accessibility of the dosage compensated Drosophila male X-chromosome requires the CLAMP zinc finger protein. *PLOS ONE* **12**:e0186855. doi:10.1371/journal.pone.0186855

Urban JA, Urban JM, Kuzu G, Larschan EN. 2017. The Drosophila CLAMP protein associates with diverse proteins on chromatin. *PLoS ONE* **12**:e0189772. doi:10.1371/journal.pone.0189772

Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis, Use R! New York: Springer-Verlag.

Wilkins RC, Lis JT. 1999. DNA distortion and multimerization: novel functions of the glutamine-rich domain of GAGA factor. *Journal of Molecular Biology* **285**:515–525. doi:10.1006/jmbi.1998.2356

Yan F, Powell DR, Curtis DJ, Wong NC. 2020. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol* **21**:22. doi:10.1186/s13059-020-1929-3

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* **16**:284–287. doi:10.1089/omi.2011.0118

Yu G, Wang L-G, He Q-Y. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**:2382–2383. doi:10.1093/bioinformatics/btv145

Zaret KS, Carroll JS. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**:2227–2241. doi:10.1101/gad.176826.111

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**:R137. doi:10.1186/gb-2008-9-9-r137

Figure Legends

Figure 1. Novel pioneer factor CLAMP is essential for early embryonic development

A. Control maternal *egfp* depletion (left), maternal *clamp* depletion (middle) and maternal *zld* depletion (right) syncytial blastoderm stage embryos probed using smFISH for the pair-rule patterning genes *run* (green) and *eve* (red). Embryos were co-labeled with Hoechst (blue) to visualize nuclei. Scale bar represents 10 µm. Quantification (%) of *eve/run* defective embryos is on the right, *p-value*s were calculated with the *Fisher's* Exact test; number of embryos is in parentheses.

B. Control maternal *egfp* depletion (left), maternal *clamp* depletion (middle) and maternal *zld* depletion (right) syncytial blastoderm stage embryos were assessed for integrity of the developing cytoskeleton using anti-NRT antibody (green) and Hoechst (blue) to label nuclei. Scale bar represents 10 µm. Quantification (%) of NRT defective embryos is on the right, *p-values* were calculated with the *Fisher's* Exact test; number of embryos is in parentheses.

C. Electrophoretic mobility shift assay (EMSA) showing the binding of increasing amounts of CLAMP DNA-binding domain (DBD) fused to MBP to 5C2 naked DNA or 5C2 *in vitro* reconstituted nucleosomes (NUC). Concentrations (uM) of CLAMP DBD increase from left to right.

D. EMSA showing the binding of increasing amounts of full-length (FL) CLAMP (fused to MBP) to 5C2 DNA or 5C2 nucleosomes (NUC). Concentrations (uM) of CLAMP FL increase from left to right.

E. Effect of maternal *clamp* RNAi on maternally-deposited (orange) or zygotically-transcribed (yellow) gene expression log2 (*clamp*-i/MTD) in 0-2hr (left) or 2-4hr (right). Maternal vs. zygotic gene categories were as defined in Lott et al. (2011). *p*-values of significant expression changes between maternal and zygotic genes were calculated by Mann-Whitney U-test and noted on the plot.

Figure 2. CLAMP regulates chromatin accessibility of a subset early zygotic genome

A. Differential accessibility (DA) analysis (left) of ATAC-seq from MTD embryos versus *clamp-i* embryos in 0-2hr or 2-4hr. Blue dots indicate non-differentially accessible sites (non-DA). Pink dots indicate significant (FDR < 0.1) differential peaks after maternal *clamp* RNAi, identified by DiffBind (DESeq2) (DA peaks). The number of peaks and representative genes in each class is noted on the plot. Average ATAC-seq signal (right) in reads per genome coverage (RPGC) 1x normalization in 0-2hr or 2-4hr embryos after maternal *clamp* RNAi centered on open chromatin (<= 100bp) peaks identified significant changes upon maternal *clamp* RNAi.

B. Example of IGV views of genomic locus *iab-8* bound by CLAMP (ChIP-seq) which shows significantly decreased CLAMP binding and ATAC-seq signal after *clamp* RNAi.

C. Genomic features of regions that require CLAMP for chromatin accessibility (ATAC-seq) compared with all CLAMP binding sites (ChIP-seq).

D. Top motifs enriched in regions that require CLAMP for chromatin accessibility. Enrichment *p*-value and percentage of sequences are noted.

E. Violin plot comparing gene expression in CLAMP-mediated changes and unchanged differential accessibility regions in 0-2hr or 2-4hr embryos after maternal *clamp* RNAi. *p*-values of significant expression changes of CLAMP down-DA and non-DA were calculated by Mann-Whitney U-test and noted on the plot.

Figure 3. CLAMP and ZLD depend on each other for binding at a subset of sites

A. ZLD peaks in 0-2hr and 2-4hr MTD and maternal *clamp* RNAi embryos. Data is displayed as a heatmap of *z*-score normalized ChIP-seq (log2 IP/input) reads in a 2 Kb region centered at each peak. Peaks in each class are arranged in order of decreasing *z*-scores in control MTD embryos.

B. CLAMP peaks in 0-2hr and 2-4hr MTD and maternal *zld* RNAi embryos. Data is displayed as a heatmap of z-score normalized ChIP-seq (log2 IP/input) reads in a 2 Kb region centered around each peak. Peaks in each class are arranged in order of decreasing z-scores in control MTD embryos.

C. Genomic distribution fractions in promoters, UTRs, exon, intron and intergenic regions for ZLD peaks in the *Drosophila* genome (dm6) in 0-2hr and 2-4hr embryos (MTD, *clamp*-i and *zld*-i).

D. Genomic distribution fractions in promoters, UTRs, exon, intron and intergenic regions for CLAMP peaks in the *Drosophila* genome (dm6) in 0-2hr and 2-4hr embryos (MTD, *clamp*-i and *zld*-i).

E. Differential binding (DB) analysis of ZLD ChIP-seq. Mean difference (MA) plots of ZLD peaks in MTD embryos versus *clamp*-i embryos in 0-2hr (left) or 2-4hr (right). Blue dots indicate non-differential binding sites. Pink dots indicate significant (FDR < 0.05) differential peaks identified by DiffBind (DESeq2). The number of peaks changed in each direction is noted on the plot.

F. Differential binding (DB) analysis of CLAMP ChIP-seq. MA plots of CLAMP peaks from MTD embryos versus *zld*-i embryos in 0-2hr (left) or 2-4hr (right). Blue dots indicate non-differential binding (non-DB) sites. Pink dots indicate significant (FDR < 0.05) differential binding (DB) peaks identified by DiffBind (DESeq2). Number of peaks in each direction is noted on the plot.

G. Stacked bar plots of CLAMP and ZLD down-DB (left) and CLAMP and ZLD non-DB peaks (right) distribution fraction in the *Drosophila* genome (dm6) in 0-2hr and 2-4hr embryos.

H. Box plot of the peak sizes in CLAMP and ZLD down-DB and non-DB peaks in 0-2hr and 2-4hr embryos. *p*-values of significant size difference between down-DB and non-DB peaks were calculated by Mann-Whitney U-test and noted on the plot.

I. Average profiles of ATAC-seq signal coverage show chromatin accessibility at ZLD down-DB (orange line) and non-DB (grey line) sites in 0-2hr (left panel) or 2-4hr (right panel) MTD and *clamp*-i embryos. Number of peaks is noted on the plot.

Figure 4. CLAMP and ZLD function together in zygotic genome activation

A. Percentage of CLAMP binding sites in 0-2hr and 2-4hr embryos distributed in maternal (n = 646), early (n = 69), mid-(n = 73), late- (n = 104), later (n = 74), and silent (n = 921) genes (peaks within a 1 Kb promoter region and gene body). Gene categories were defined in Li et al. (2014).

B. Percentage of ZLD binding sites in 0-2hr and 2-4hr embryos distributed in maternal (n = 646), early (n = 69), mid-(n = 73), late- (n = 104), later (n = 74), and silent (n = 921) genes (peaks within a 1 Kb promoter region and gene body). Gene categories were defined in Li et al. (2014).

C. Gene expression changes caused by maternal *clamp* RNAi (Rieder et al., 2017) at genes with strong, weak, and no CLAMP binding as measured by ChIP-seq in 0-2hr (left) or 2-4hr (right) embryos. *p*-values of significant expression changes of CLAMP bindings were calculated by Mann-Whitney U-test and noted on the plot.

D. Gene expression changes caused by maternal *zld* RNAi (Schulz et al., 2015) at genes with strong, weak, and no ZLD binding as measured by ChIP-seq in 0-2hr (left) or 2-4hr (right) embryos. *p*-values of significant expression changes of ZLD bindings were calculated by Mann-Whitney U-test and noted on the plot.

E. Gene expression changes caused by maternal *zld* RNAi (Schulz et al., 2015) at genes with CLAMP down-DB and non-DB that defined in wt vs. *zld*- 0-2hr and 2-4hr embryos. *p*-values of significant expression changes of CLAMP down-DB and non-DB were calculated by Mann-Whitney U-test and noted on the plot.

F. Gene expression changes caused by maternal *clamp* RNAi (Rieder et al., 2017) at genes with ZLD down-DB and non-DB that defined in MTD vs. *clamp*-i 0-2hr and 2-4hr embryos. *p*-values of significant expression changes of ZLD down-DB and non-DB were calculated by Mann-Whitney U-test and noted on the plot.

Figure 5. CLAMP and ZLD regulate gene expression via modulating chromatin accessibility

A. Four classes of CLAMP and ZLD co-bound peaks defined by combining ATAC-seq (this study or Hannon et al., 2017; Soluri et al., 2020) and ChIP-seq peaks in 0-2hr MTD and RNAi embryos. Data is displayed as a heatmap of z-score normalized ATAC-seq and ChIP-seq reads in a 2 kb region centered around each peak. Peaks in each class are arranged in order of decreasing *z*-scores in control MTD embryos.
   - Type I (n=5): both DA, differentially accessible regions which depend on CLAMP or ZLD; has both proteins bound.
   - Type II (n=23): CLAMP DA and ZLD non-DA, differentially accessible regions which depend on CLAMP, not on ZLD; has both proteins bound.

- Type III (n=123): ZLD DA and CLAMP non-DA, differentially accessible regions which depend on ZLD, not on CLAMP; has both proteins bound.
- Type IV (n=374): both non-DA, accessibility independent from CLAMP or ZLD but has both proteins bound.

B. Left: Gene expression changes caused by maternal *clamp* RNAi (Rieder et al., 2017) in 0-2hr embryos at genes fall into four classes of CLAMP and ZLD co-bound peaks. *p*-values of significant expression changes among classes were calculated by Mann-Whitney U-test and noted on the plot.

   Right: Gene expression changes caused by maternal *zld* RNAi (Schulz et al., 2015) in stage5 embryos at genes fall into four classes of CLAMP and ZLD co-bound peaks. *p*-values of significant expression changes among classes were calculated by Mann-Whitney U-test and noted on the plot.

C. Upper panel: Venn diagram showing the number of overlapping sites between GAF-dependent DA sites (Gaskill et al., 2021), ZLD non-DA with ZLD bound, and CLAMP non-DA with CLAMP bound peaks.

   Lower panel: Venn diagram showing the number of overlapping sites between GAF-dependent DA sites (Gaskill et al., 2021), ZLD non-DA, and CLAMP non-DA peaks.

Figure 6. Model for how CLAMP and ZLD pioneer factor function together to define chromatin accessibility in early embryos.

CLAMP and ZLD function together at promoters to regulate each other's occupancy and gene expression of genes encoding other key TFs. We defined CLAMP and ZLD co-bound peaks in early embryos, which revealed roles for CLAMP and ZLD in defining chromatin accessibility and activating zygotic transcription at a subset of the zygotic genome.

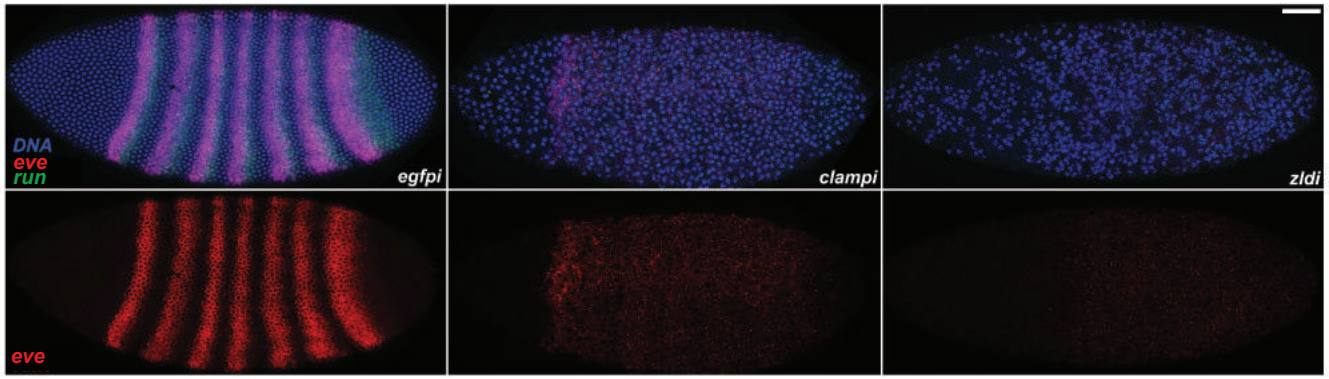
CLAMP-dependent regions: CLAMP promotes ZLD enrichment at these sites where CLAMP binding increases chromatin accessibility and regulates target gene expression. These sites are closed and lack binding of ZLD when maternal *clamp* is depleted, and they remain open and transcription is activated when maternal *zld* is depleted.

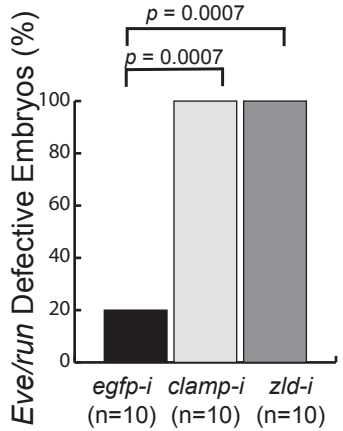
ZLD-dependent regions: ZLD modulates chromatin opening and transcription at these sites that are bound by CLAMP but do not depend on CLAMP for chromatin accessibility. These sites are closed and lack binding of CLAMP when maternal *zld* is depleted, and they remain open and active when maternal *clamp* is depleted.

CLAMP/ZLD-independent regions: GAF or other TFs open chromatin at locations co-bound by CLAMP and ZLD where chromatin accessibility is not altered when each factor is depleted individually. CLAMP and ZLD could also function redundantly at some of these loci. These sites remain accessible and transcriptionally active upon either maternal *zld* or *clamp* depletion.

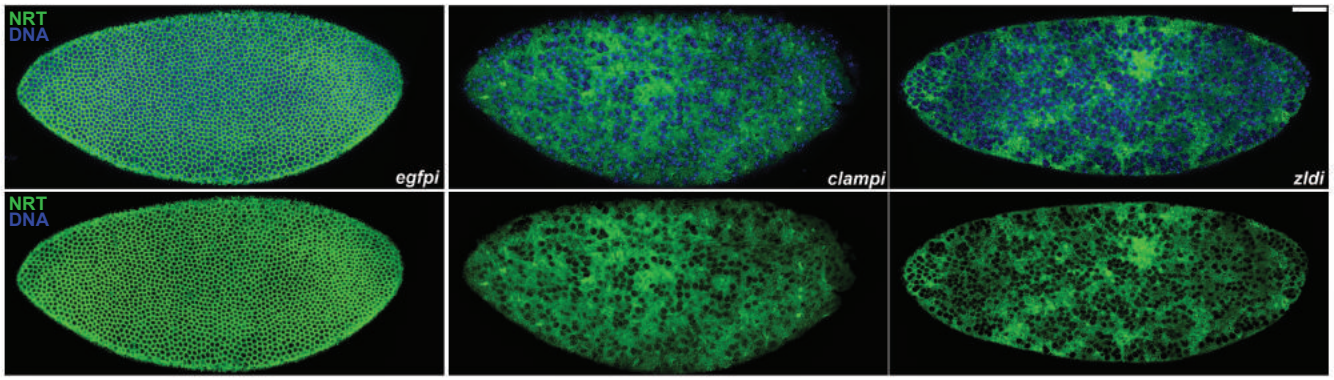
Fig.1

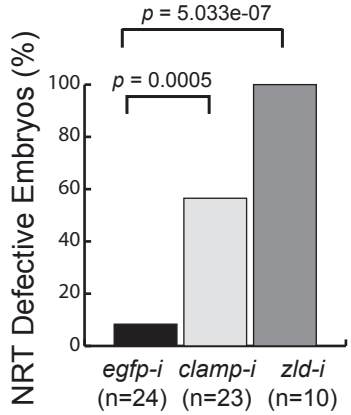
A


DNA
eve
run
egfpi
clampi
zldi
eve

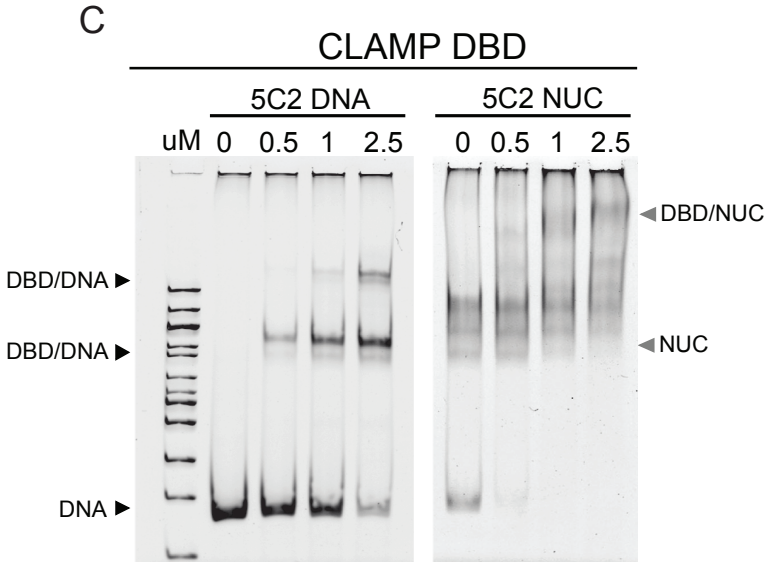

p = 0.0007
p = 0.0007
Eve/run Defective Embryos (%)
100
80
60
40
20
0
egfp-i (n=10)
clamp-i (n=10)
zld-i (n=10)

B

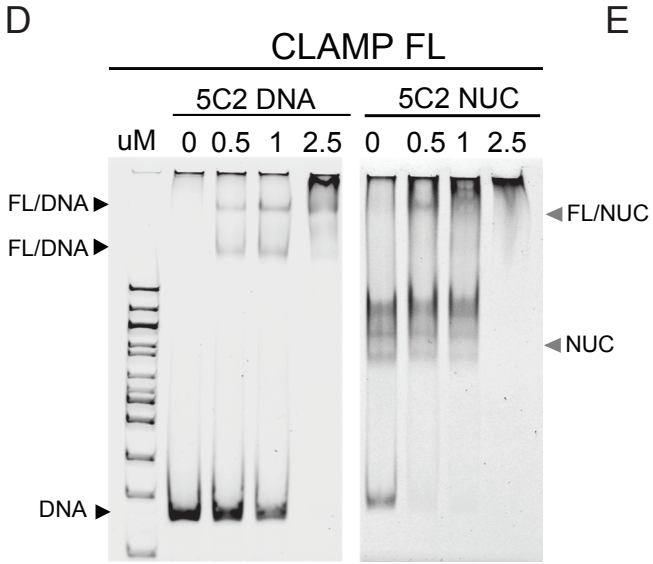

NRT
DNA
egfpi
clampi
zldi
NRT
DNA


p = 5.033e-07
p = 0.0005
NRT Defective Embryos (%)
100
80
60
40
20
0
egfp-i (n=24)
clamp-i (n=23)
zld-i (n=10)

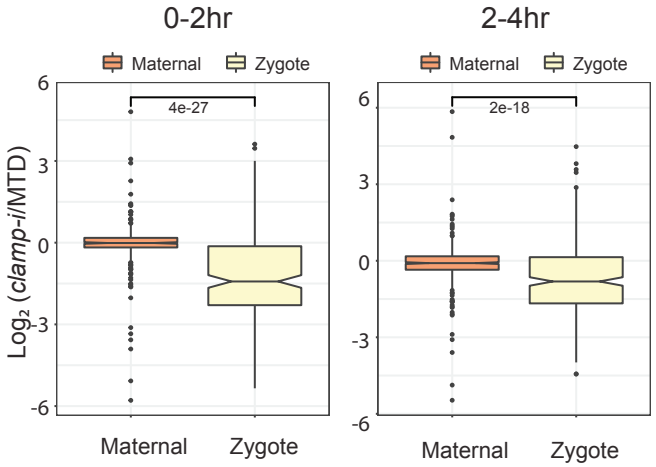
C


CLAMP DBD
5C2 DNA
5C2 NUC
uM 0 0.5 1 2.5
0 0.5 1 2.5
DBD/NUC
DBD/DNA
DBD/DNA
NUC
DNA

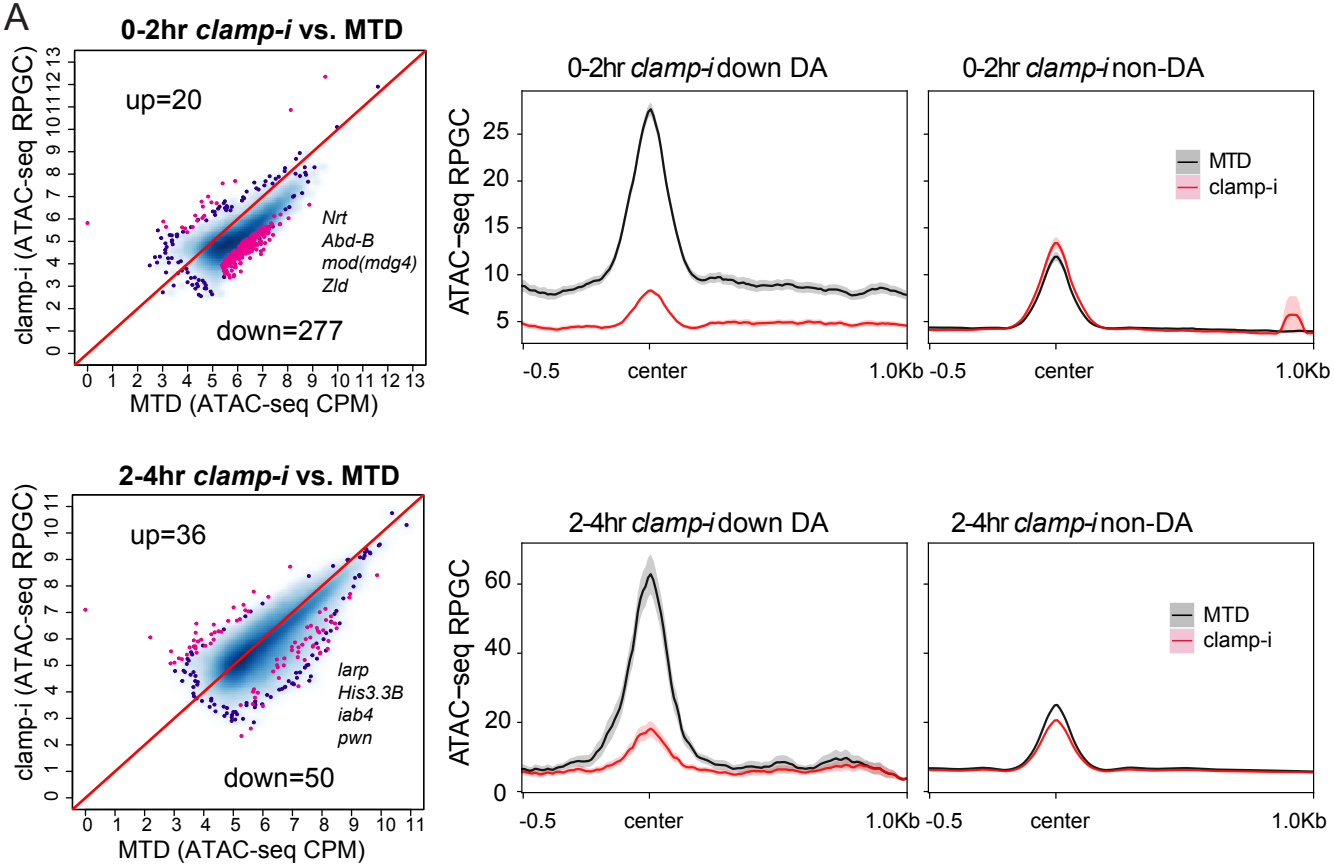
D


CLAMP FL
5C2 DNA
5C2 NUC
uM 0 0.5 1 2.5
0 0.5 1 2.5
FL/DNA
FL/DNA
FL/NUC
NUC
DNA

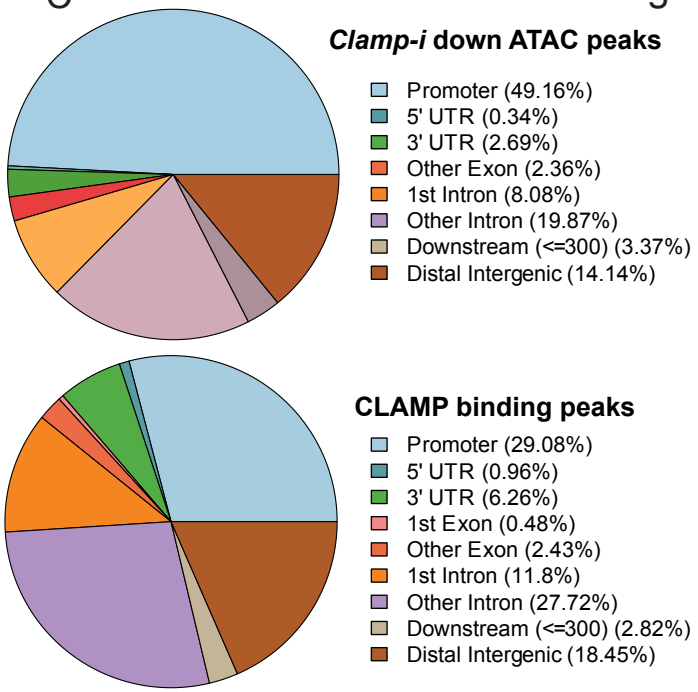
E


0-2hr
2-4hr
Maternal
Zygote
4e-27
2e-18
Log2 (clamp-i/MTD)
6
3
0
-3
-6
Maternal
Zygote

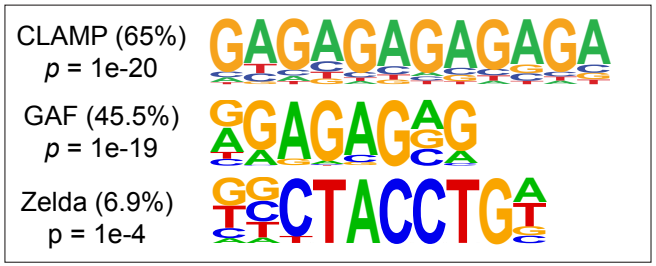
A


0-2hr clamp-i vs. MTD
up=20
down=277
Nrt
Abd-B
mod(mdg4)
Zld
clamp-i (ATAC-seq RPGC)
MTD (ATAC-seq CPM)
0-2hr clamp-i down DA
0-2hr clamp-i non-DA
ATAC−seq RPGC
MTD
clamp-i
-0.5
center
1.0Kb
2-4hr clamp-i vs. MTD
up=36
down=50
larp
His3.3B
iab4
pwn
2-4hr clamp-i down DA
2-4hr clamp-i non-DA

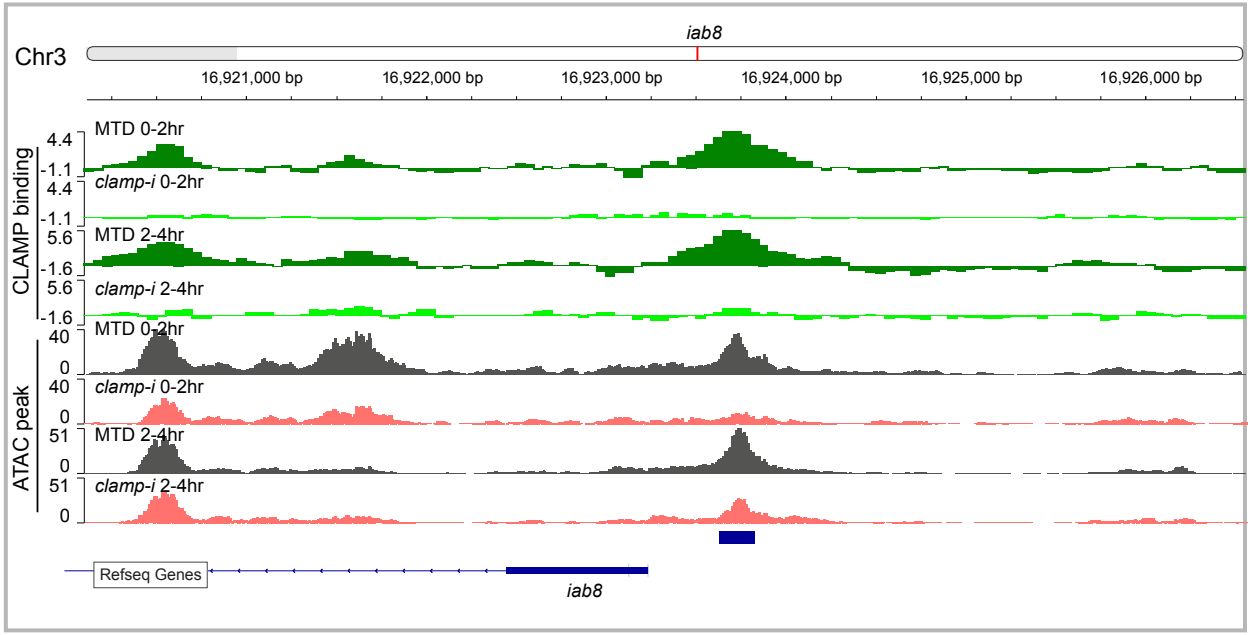
C


Clamp-i down ATAC peaks
Promoter (49.16%)
5' UTR (0.34%)
3' UTR (2.69%)
Other Exon (2.36%)
1st Intron (8.08%)
Other Intron (19.87%)
Downstream (<=300) (3.37%)
Distal Intergenic (14.14%)
CLAMP binding peaks
Promoter (29.08%)
5' UTR (0.96%)
3' UTR (6.26%)
1st Exon (0.48%)
Other Exon (2.43%)
1st Intron (11.8%)
Other Intron (27.72%)
Downstream (<=300) (2.82%)
Distal Intergenic (18.45%)

D


Motifs in clamp-i down ATAC peaks
CLAMP (65%)
p = 1e-20
GAF (45.5%)
p = 1e-19
Zelda (6.9%)
p = 1e-4

B


Chr3
iab8
16,921,000 bp
16,922,000 bp
16,923,000 bp
16,924,000 bp
16,925,000 bp
16,926,000 bp
CLAMP binding
ATAC peak
MTD 0-2hr
clamp-i 0-2hr
MTD 2-4hr
clamp-i 2-4hr
Refseq Genes

E

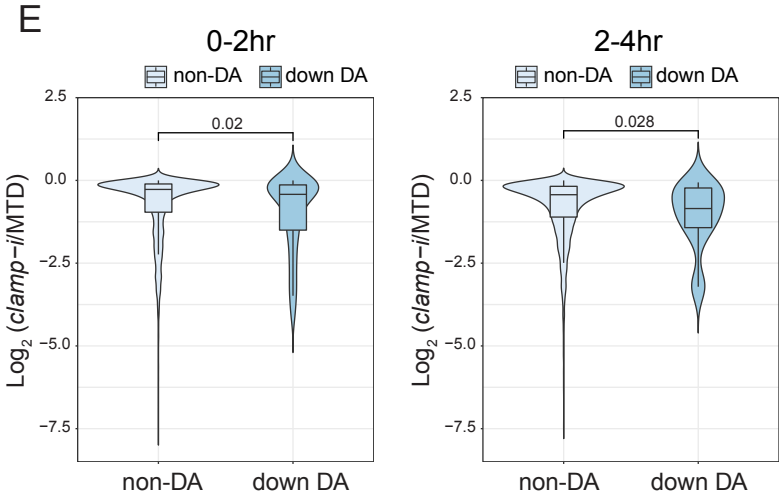
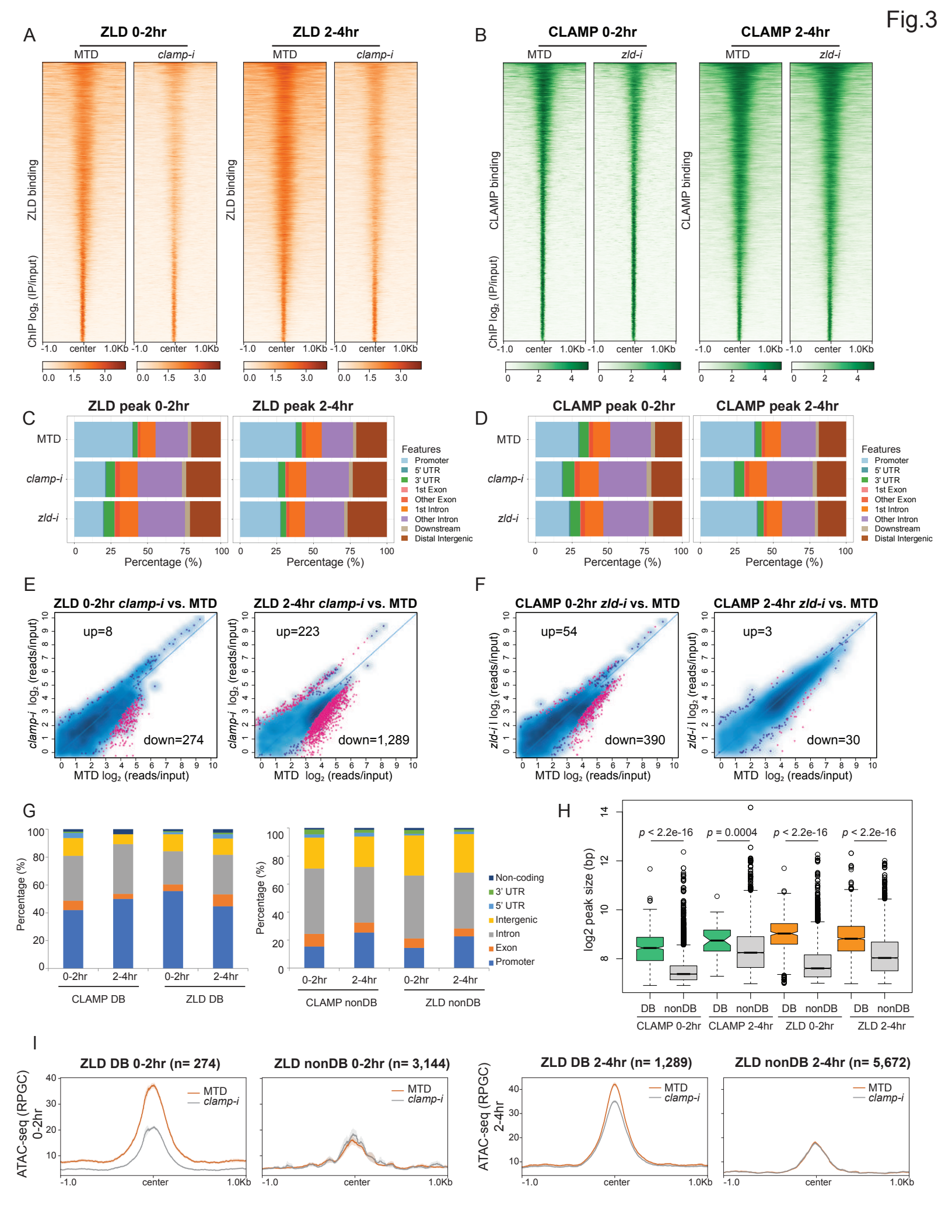

0-2hr
2-4hr
non-DA
down DA
0.02
0.028
Log2 (clamp−i/MTD)

Fig.3


A
ZLD 0-2hr
ZLD 2-4hr
MTD
clamp-i
ZLD binding
ChIP log2 (IP/input)
-1.0 center 1.0Kb
0.0 1.5 3.0
B
CLAMP 0-2hr
CLAMP 2-4hr
zld-i
CLAMP binding
0 2 4
C
ZLD peak 0-2hr
ZLD peak 2-4hr
Percentage (%)
Features
Promoter
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream
Distal Intergenic
D
CLAMP peak 0-2hr
CLAMP peak 2-4hr
E
ZLD 0-2hr clamp-i vs. MTD
ZLD 2-4hr clamp-i vs. MTD
up=8
down=274
up=223
down=1,289
clamp-i log2 (reads/input)
MTD log2 (reads/input)
F
CLAMP 0-2hr zld-i vs. MTD
CLAMP 2-4hr zld-i vs. MTD
up=54
down=390
up=3
down=30
zld-i log2 (reads/input)
G
Percentage (%)
0-2hr
2-4hr
CLAMP DB
ZLD DB
CLAMP nonDB
ZLD nonDB
Non-coding
3' UTR
5' UTR
Intergenic
Intron
Exon
Promoter
H
log2 peak size (bp)
p < 2.2e-16
p = 0.0004
p < 2.2e-16
p < 2.2e-16
DB nonDB
CLAMP 0-2hr
CLAMP 2-4hr
ZLD 0-2hr
ZLD 2-4hr
I
ZLD DB 0-2hr (n= 274)
ZLD nonDB 0-2hr (n= 3,144)
ZLD DB 2-4hr (n= 1,289)
ZLD nonDB 2-4hr (n= 5,672)
ATAC-seq (RPGC) 0-2hr
ATAC-seq (RPGC) 2-4hr
MTD
clamp-i
-1.0 center 1.0Kb

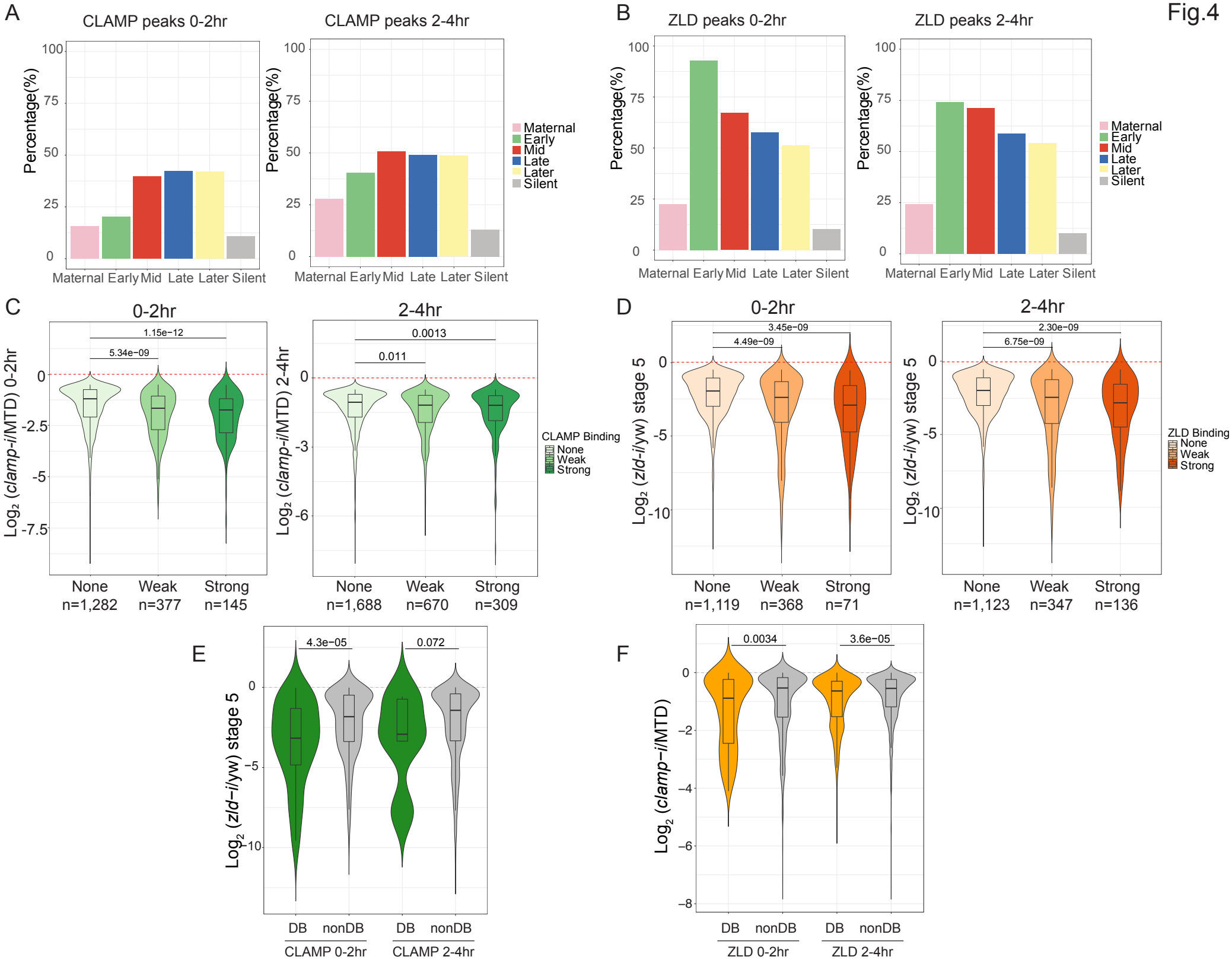


Fig.4
A
CLAMP peaks 0-2hr
CLAMP peaks 2-4hr
Percentage(%)
Maternal
Early
Mid
Late
Later
Silent
B
ZLD peaks 0-2hr
ZLD peaks 2-4hr
C
0-2hr
2-4hr
Log2 (clamp-i/MTD) 0-2hr
Log2 (clamp-i/MTD) 2-4hr
1.15e−12
5.34e−09
0.0013
0.011
CLAMP Binding
None
Weak
Strong
None n=1,282
Weak n=377
Strong n=145
None n=1,688
Weak n=670
Strong n=309
D
Log2 (zld-i/yw) stage 5
3.45e−09
4.49e−09
2.30e−09
6.75e−09
ZLD Binding
None n=1,119
Weak n=368
Strong n=71
None n=1,123
Weak n=347
Strong n=136
E
4.3e−05
0.072
DB
nonDB
CLAMP 0-2hr
CLAMP 2-4hr
F
0.0034
3.6e−05
Log2 (clamp−i/MTD)
ZLD 0-2hr
ZLD 2-4hr

Fig.5

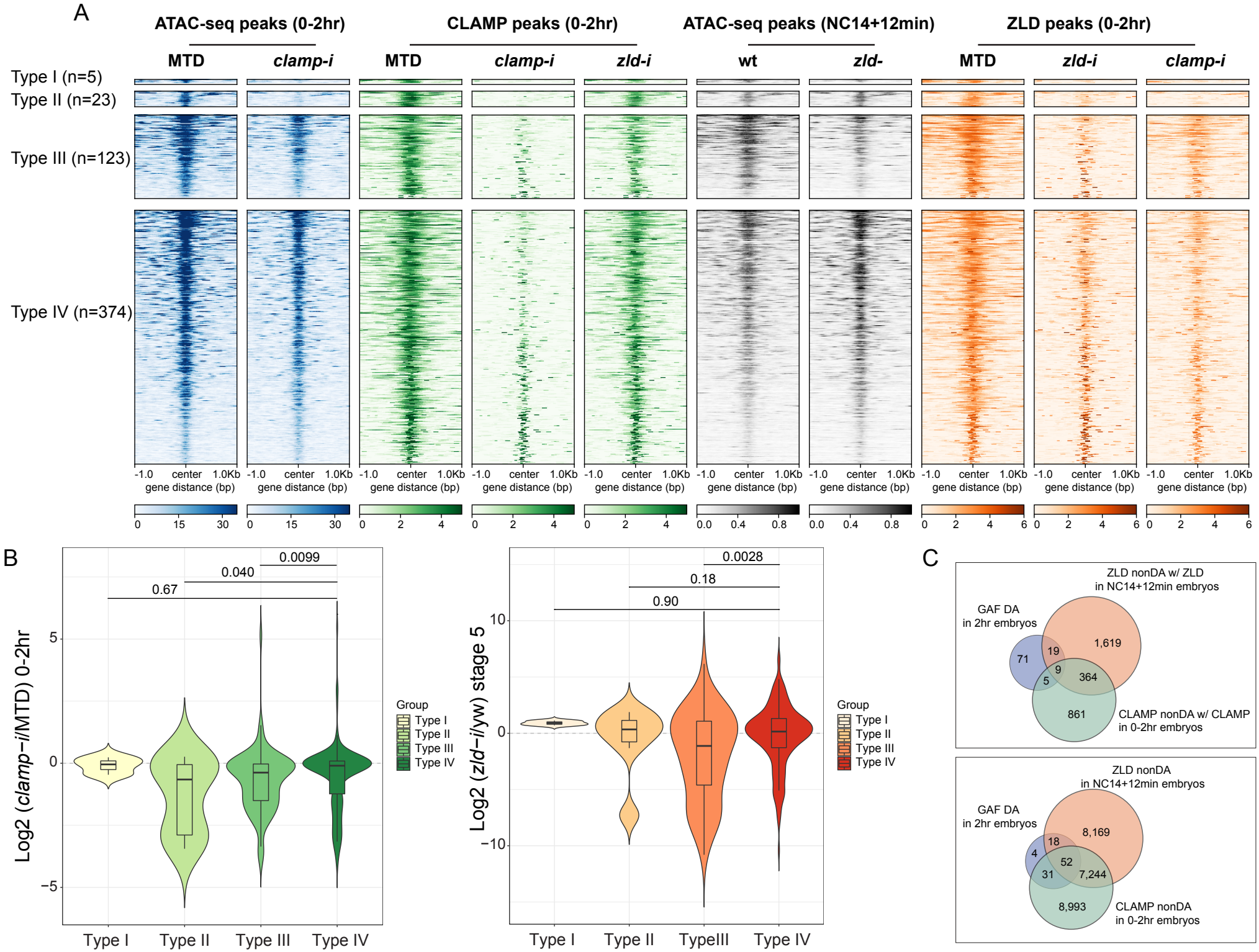
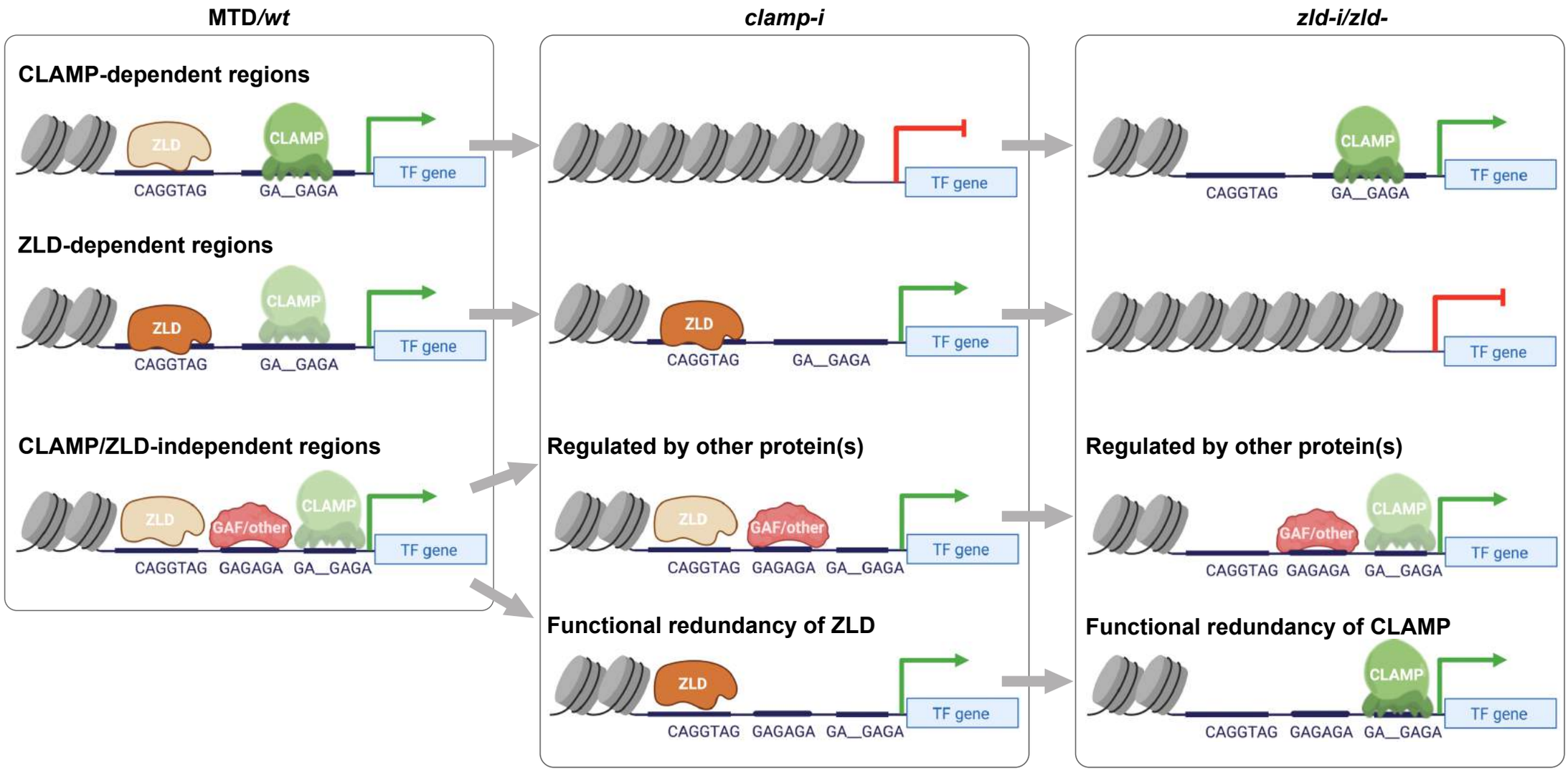

A
ATAC-seq peaks (0-2hr)
CLAMP peaks (0-2hr)
ATAC-seq peaks (NC14+12min)
ZLD peaks (0-2hr)
MTD
clamp-i
MTD
clamp-i
zld-i
wt
zld-
MTD
zld-i
clamp-i
Type I (n=5)
Type II (n=23)
Type III (n=123)
Type IV (n=374)
-1.0 center 1.0Kb
gene distance (bp)
B
0.0099
0.040
0.67
Log2 (clamp−i/MTD) 0-2hr
Type I
Type II
Type III
Type IV
Group
0.0028
0.18
0.90
Log2 (zld−i/yw) stage 5
TypeIII
C
ZLD nonDA w/ ZLD
in NC14+12min embryos
GAF DA
in 2hr embryos
1,619
71
19
9
364
5
861
CLAMP nonDA w/ CLAMP
in 0-2hr embryos
ZLD nonDA
in NC14+12min embryos
8,169
18
4
52
31
7,244
8,993
CLAMP nonDA
in 0-2hr embryos

Fig.6


MTD/wt
clamp-i
zld-i/zld-
CLAMP-dependent regions
ZLD
CLAMP
TF gene
CAGGTAG
GA__GAGA
ZLD-dependent regions
CLAMP/ZLD-independent regions
GAF/other
GAGAGA
Regulated by other protein(s)
Functional redundancy of ZLD
Functional redundancy of CLAMP

Supplemental Figure Legends

Figure 1-figure supplement 1. Novel pioneer factor CLAMP is essential for early embryonic development

A. Expression of early zygotic genes (*even-skipped*, *runt,* and *neurotactin*) in MTD and *clamp-i* embryos measured by RNA-seq (Rieder et al., 2017). Log_2 Fold Change and *p-value* were calculated using DESeq2.

B. Expression of mRNAs (*clamp* and *zld*) in MTD, *clamp-i* and *zld-i* embryos in 0-2hr and 2-4hr embryos measured by RT-qPCR. Log_2 fold change was calculated using the ΔΔCt method (Rao et al., 2013) and normalized to reference gene *pka*.

C. Western blot of CLAMP, ZLD and reference control Actin in MTD, *clamp-i* and *zld-i* embryos in 0-2hr and 2-4hr embryos. MTD: MTD-Gal4 line. *clamp-i:* MTD-Gal4-*clamp* mRNAi line, *zld-i:* MTD-Gal4-*zld* mRNAi line.

D. Genome browser tracks for a region of the CES 5C2 locus (red bar is 500bp) used to make *in vitro* reconstituted nucleosomes (Urban et al., 2017). Locations of three CLAMP binding motifs are marked as black dots. CLAMP ChIP-seq normalized sequencing reads are shown in green. MNase-seq MACC scores (dark blue) show chromatin accessibility in S2 cells in control (*egfp-i*), *clamp*, and *msl2* RNAi treatment. The nucleosome profile (Nuc) is shown in purple. The dashed red rectangle highlights the genomic region (240bp) used to reconstitute nucleosomes *in vitro*.

E. Effect of maternal *zld* RNAi (Schulz et al., 2015) on maternally-deposited (orange) or zygotically-transcribed (yellow) gene expression log2 (*clamp*-i/MTD) in 0-2hr (left) or 2-4hr (right). Maternal vs. zygotic gene categories were as defined in Lott et al. (2011). *p*-values of significant expression changes between maternal and zygotic genes were calculated by Mann-Whitney U-test and noted on the plot.

Figure 2-figure supplement 1. CLAMP regulates chromatin accessibility throughout ZGA

A. PCA plot for sample replicates in MTD and *clamp-i* embryos at 0-2hr time point.

B. Pearson correlation heatmap of peaks dependent on CLAMP in MTD vs. *clamp-i* embryos at 0-2hr time point.

C. GO terms for genes that require CLAMP for chromatin accessibility at 0-2hr time point.

D. PCA plots for sample replicates in MTD and *clamp-i* embryos at 2-4hr time point.

E. Pearson correlation heatmap of peaks dependent on CLAMP in MTD vs. *clamp-i* embryos at 2-4hr time point.

F. GO terms for genes that require CLAMP for chromatin accessibility at 2-4hr time point.

Figure 3-figure supplement 1. CLAMP and ZLD depend on each other for chromatin binding

A. CLAMP peaks in 0-2hr and 2-4hr MTD and maternal *clamp* RNAi embryos. Data is displayed as a heatmap of z-score normalized ChIP-seq (log2 IP/input) reads, in a 2 Kb region centered around each peak called in control MTD embryos. Peaks in each class are arranged in order of decreasing z-scores in control MTD embryos.

B. ZLD peaks in 0-2hr and 2-4hr MTD and maternal *zld* RNAi embryos. Data is displayed as a heatmap of z-score normalized ChIP-seq (log2 IP/input) reads, in a 2 Kb region centered around each peak called in control MTD embryos. Peaks in each class are arranged in order of decreasing z-scores in control MTD embryos.

C. CLAMP (green) and ZLD (orange) peaks and shared peaks where both CLAMP and ZLD are present in 0-2hr and 2-4hr embryos. *p*-values represent the significance (hypergeometric test, N= 15,682 total fly genes) of overlap.

D. ZLD up-DB, down-DB and non-DB peaks in 0-2hr and 2-4hr MTD and maternal *clamp* RNAi embryos. Data is displayed as a heatmap of *z*-score normalized ChIP-seq (log2 IP/input) reads in a 2 Kb region centered around each peak. Peaks in each class are arranged in order of decreasing *z*-scores in control MTD embryos.

E. CLAMP up-DB, down-DB and non-DB peaks in 0-2hr and 2-4hr MTD and maternal *zld* RNAi embryos. Data is displayed as a heatmap of z-score normalized ChIP-seq (log2 IP/input) reads in a 2 kb region centered around each peak. Peaks in each class are arranged in order of decreasing z-scores in control MTD embryos.

F. Venn diagram showing the number of overlapping sites between ZLD and CLAMP down-DB or ZLD and CLAMP non-DB sites in 0-2hr and 2-4hr. *p*-values represent the significance (hypergeometric test, N= 15,682 total fly genes) of overlap.

G. Example IGV views of genomic loci in *iab-8,* which CLAMP and ZLD were both down-DBs and dependent on each other to bind.

H. Average profiles of ChIP-seq signal in log2 (IP/input) show the size of down-DB or non-DB peaks of CLAMP in MTD versus *zld*-i embryos (upper panel) and ZLD in MTD versus *clamp*-i embryos (lower panel).

Figure 3-figure supplement 2. CLAMP and ZLD depend on each other for chromatin binding

A. Example IGV views of genomic loci: ChIP-seq in MTD and *zld*-i embryos at *Hbn* promoter region (upper panel) with CLAMP differential binding (down-DB) peak; *Fdl* (lower panel) with CLAMP non-differential binding (non-DB) peak at an intron. Peaks called by MACS2 are marked in dark grey (non-DB) and grey (down-DB). CLAMP and ZLD motifs are marked in green and orange respectively.

B. Top three motifs called by Homer for CLAMP down-DB, CLAMP non-DB, ZLD down-DB, and ZLD non-DB peaks sites in 0-2hr and 2-4hr embryos.

C. Average profiles of ATAC-seq signal coverage in NC14+12min wt and *zld*- embryos show chromatin accessibility at CLAMP down-DB (green line) and non-DB (grey line) sites defined in 0-2hr (left panel) or 2-4hr (right panel) MTD and *zld*-i embryos. Number of peaks is noted on the plot.

Figure 5-figure supplement 1. CLAMP-mediated chromatin accessibility is correlated with CLAMP and ZLD binding

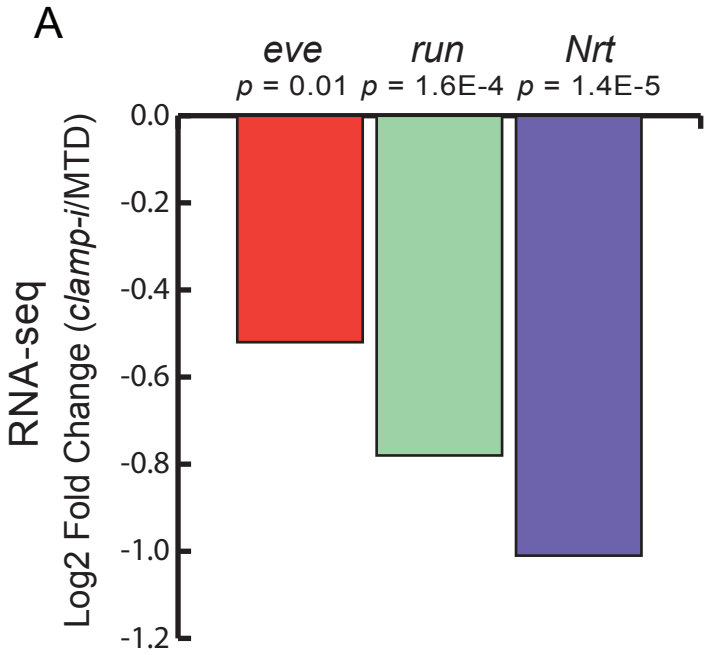
A. CLAMP-related and ZLD-related peaks defined by combining ATAC-seq (this study or Hannon et al., 2017; Soluri et al., 2020) and ChIP-seq peaks in 0-2hr MTD and RNAi embryos. Data is displayed as a heatmap of z-score normalized ATAC-seq and ChIP-seq (log2 IP/input) reads, in a 2 Kb region centered around each peak. Peaks in each class are arranged in order of decreasing z-scores in control MTD embryos. Peak types and numbers are marked on plot:

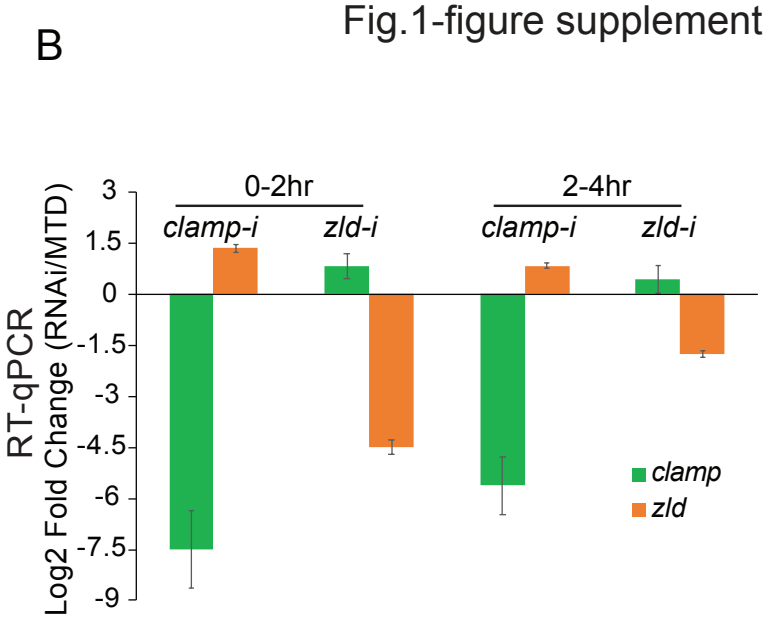
   - DA w/ CLAMP (n=74): differentially accessible regions that depend on CLAMP and has CLAMP binding.
   - DA w/o CLAMP (n=203): differentially accessible regions that depend on CLAMP and has no CLAMP binding.
   - Non-DA w/ CLAMP (n=1,238): not differentially accessible regions that have CLAMP binding.
   - DA w/ ZLD (n=986): differentially accessible regions that depends on ZLD and has ZLD binding.
   - DA w/o ZLD (n=2,797): differentially accessible regions that depends on ZLD and has no ZLD binding.
   - NonDA w/ ZLD (n=2,301): non-differentially accessible regions which have ZLD binding.

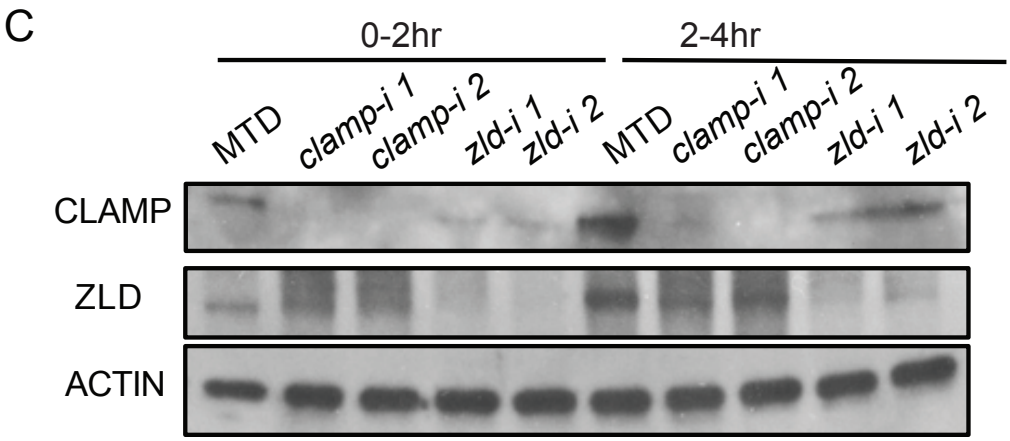
   Note that Non-DA w/o CLAMP (n=15,081) and Non-DA w/o ZLD(n=17,758) peaks are omitted from this plot.

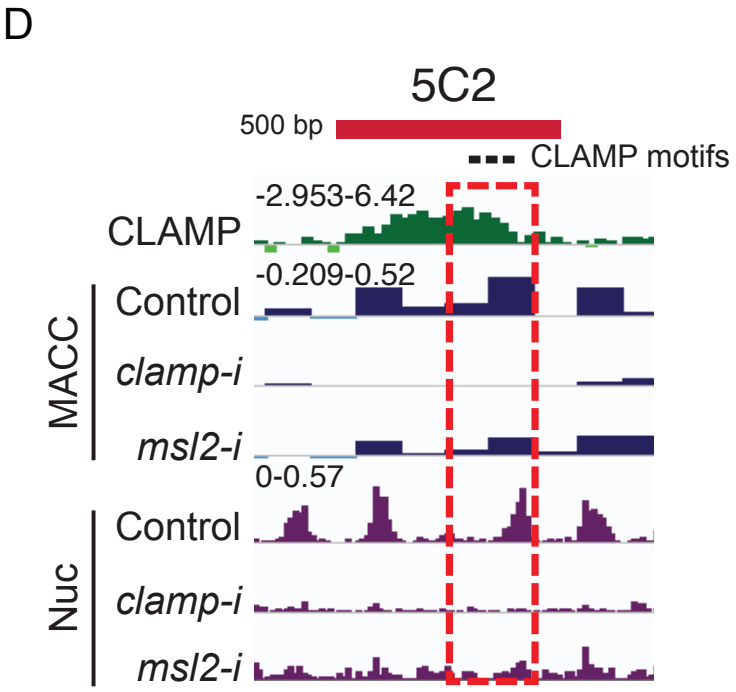
B. Top motifs enriched in each type of CLAMP and ZLD co-bound regions. Enrichment *p*-value and percentage of sequences are noted.

Fig.1-figure supplement 1









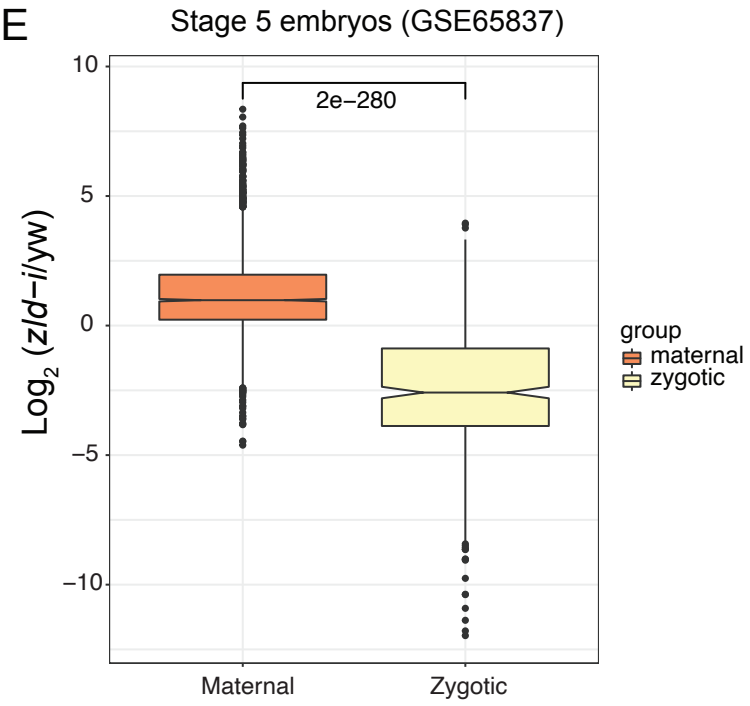


Fig.2-figure supplement 1

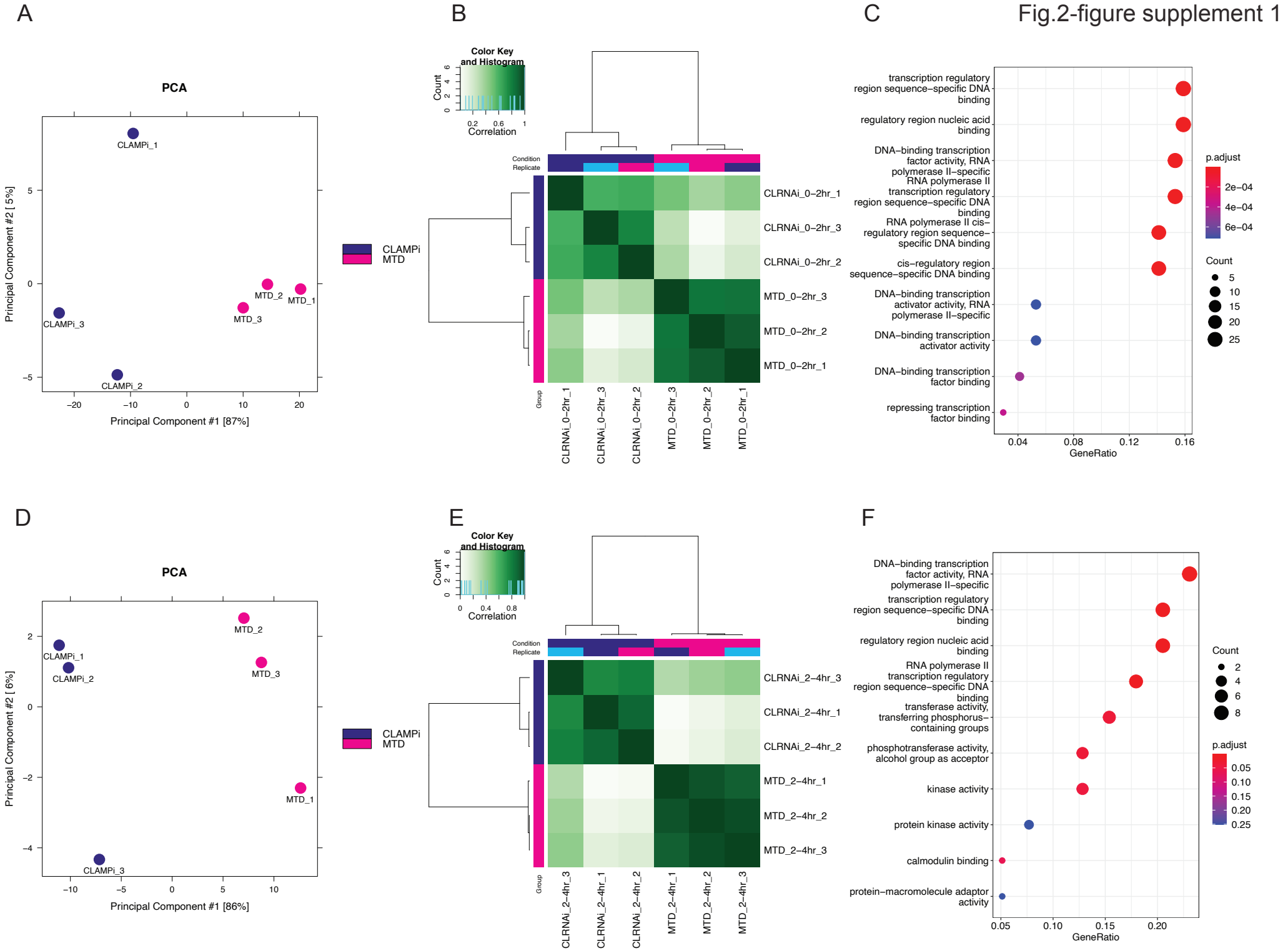
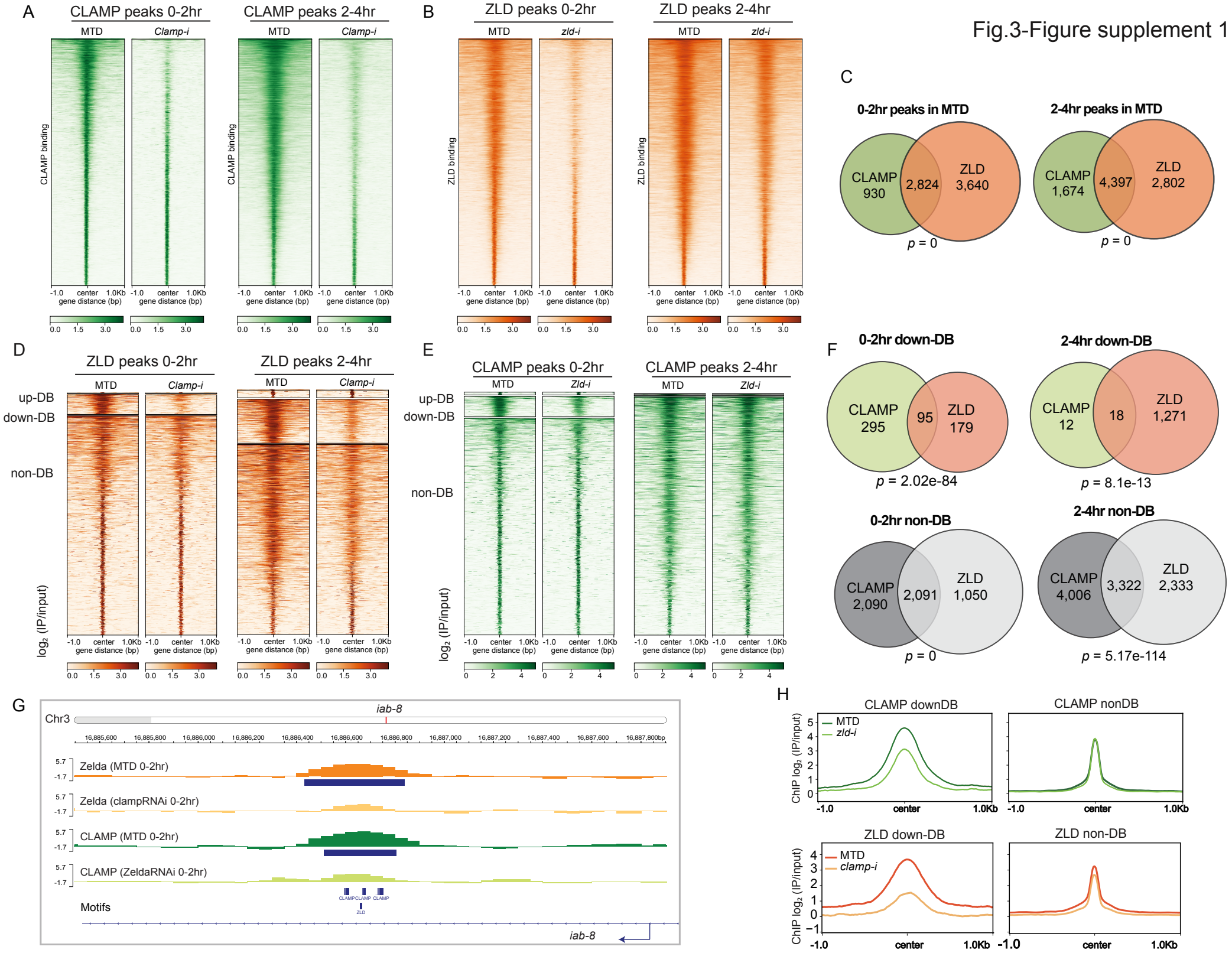
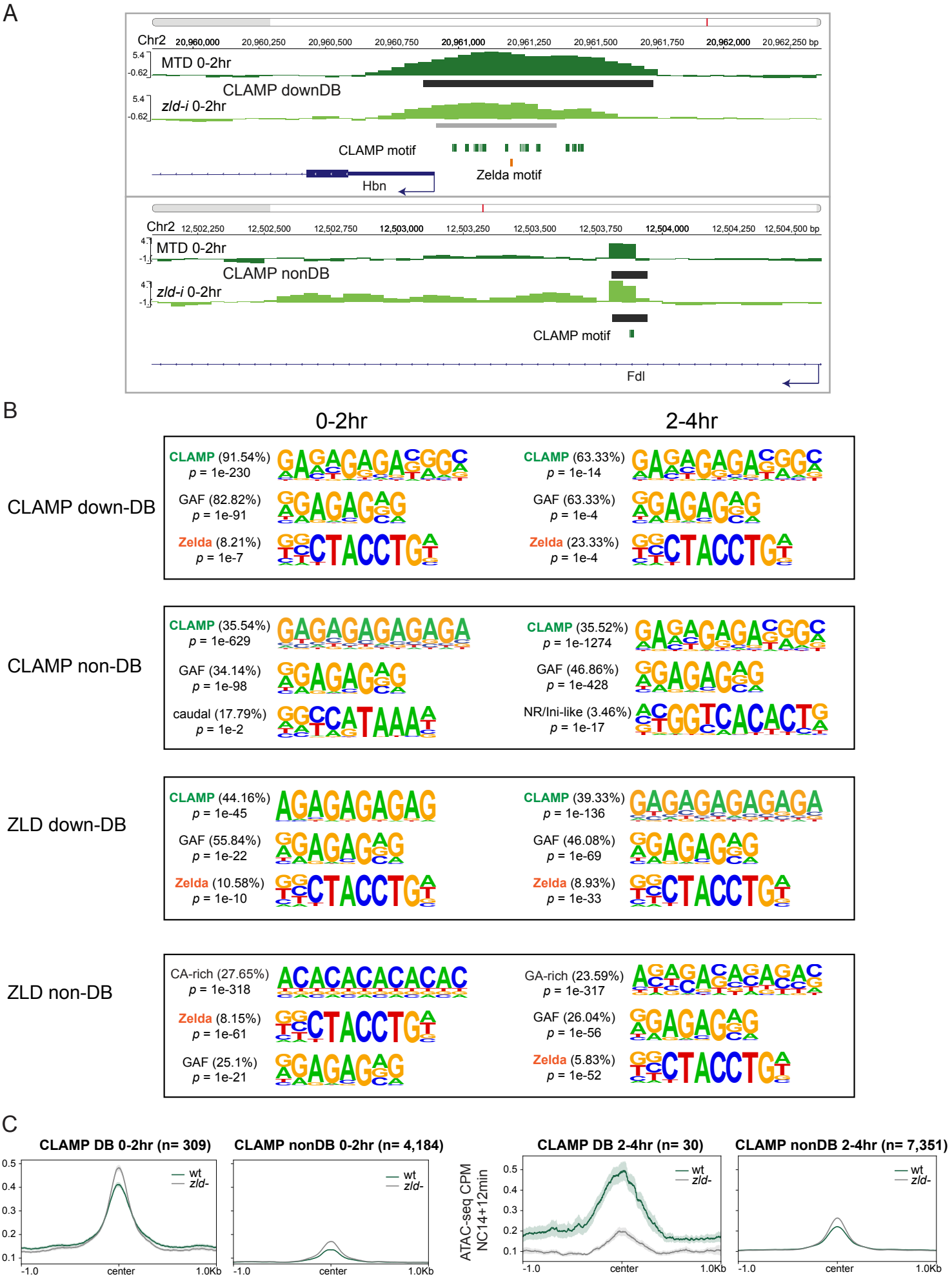
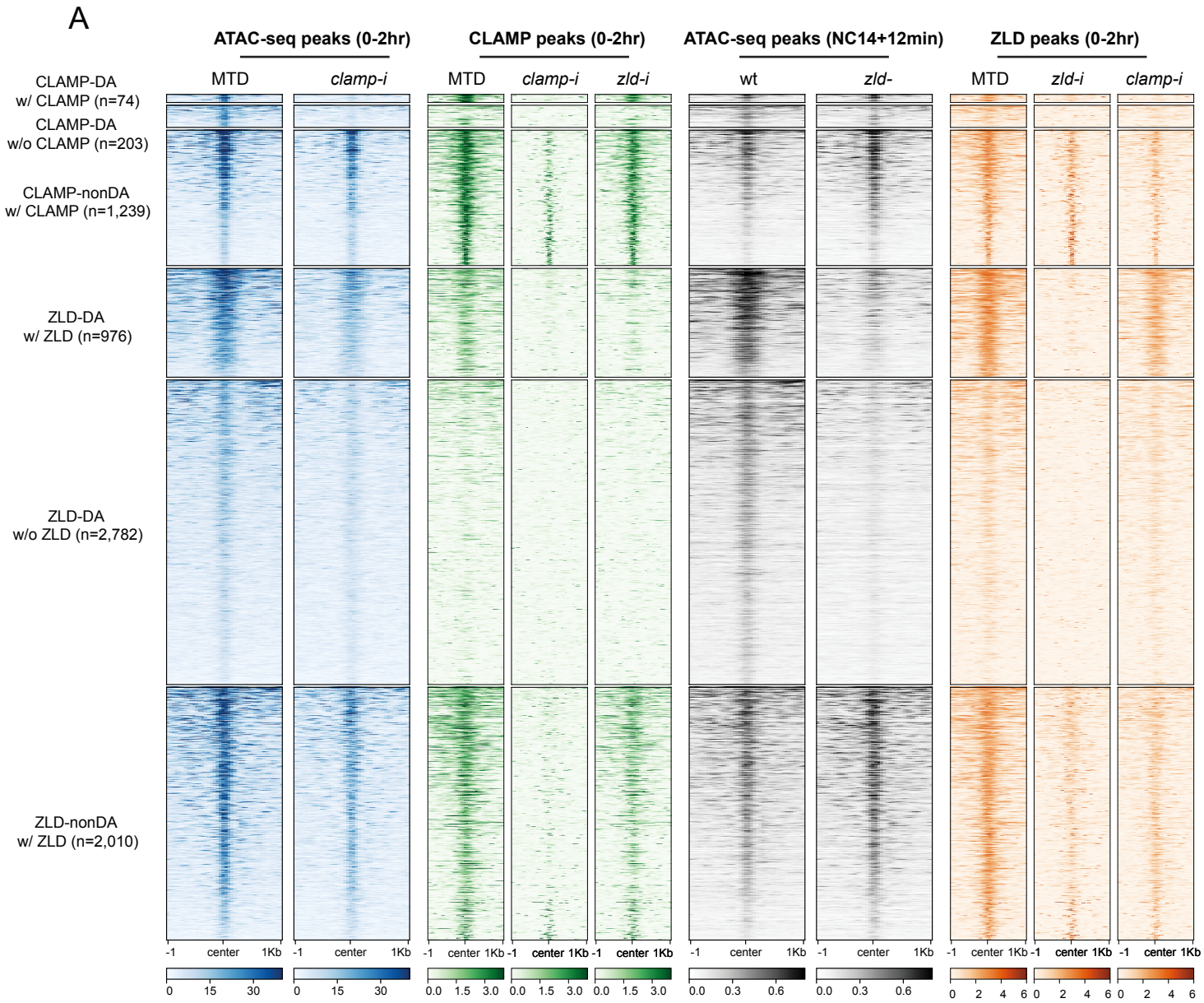
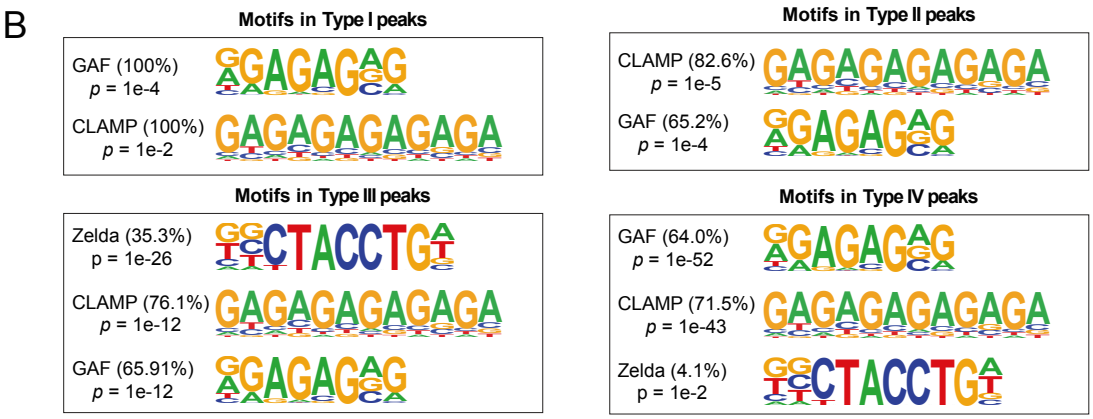


Fig.3-Figure supplement 1




A
Chr2
20,960,000
20,960,250
20,960,500
20,960,750
20,961,000
20,961,250
20,961,500
20,961,750
20,962,000
20,962,250 bp
MTD 0-2hr
CLAMP downDB
zld-i 0-2hr
CLAMP motif
Zelda motif
Hbn
Chr2
12,502,250
12,502,500
12,502,750
12,503,000
12,503,250
12,503,500
12,503,750
12,504,000
12,504,250
12,504,500 bp
MTD 0-2hr
CLAMP nonDB
zld-i 0-2hr
CLAMP motif
Fdl
B
0-2hr
2-4hr
CLAMP down-DB
CLAMP (91.54%) p = 1e-230
GAF (82.82%) p = 1e-91
Zelda (8.21%) p = 1e-7
CLAMP (63.33%) p = 1e-14
GAF (63.33%) p = 1e-4
Zelda (23.33%) p = 1e-4
CLAMP non-DB
CLAMP (35.54%) p = 1e-629
GAF (34.14%) p = 1e-98
caudal (17.79%) p = 1e-2
CLAMP (35.52%) p = 1e-1274
GAF (46.86%) p = 1e-428
NR/Ini-like (3.46%) p = 1e-17
ZLD down-DB
CLAMP (44.16%) p = 1e-45
GAF (55.84%) p = 1e-22
Zelda (10.58%) p = 1e-10
CLAMP (39.33%) p = 1e-136
GAF (46.08%) p = 1e-69
Zelda (8.93%) p = 1e-33
ZLD non-DB
CA-rich (27.65%) p = 1e-318
Zelda (8.15%) p = 1e-61
GAF (25.1%) p = 1e-21
GA-rich (23.59%) p = 1e-317
GAF (26.04%) p = 1e-56
Zelda (5.83%) p = 1e-52
C
CLAMP DB 0-2hr (n= 309)
CLAMP nonDB 0-2hr (n= 4,184)
CLAMP DB 2-4hr (n= 30)
CLAMP nonDB 2-4hr (n= 7,351)
ATAC-seq CPM NC14+12min
wt
zld-
-1.0
center
1.0Kb


A
ATAC-seq peaks (0-2hr)
CLAMP peaks (0-2hr)
ATAC-seq peaks (NC14+12min)
ZLD peaks (0-2hr)
MTD
clamp-i
MTD
clamp-i
zld-i
wt
zld-
MTD
zld-i
clamp-i
CLAMP-DA w/ CLAMP (n=74)
CLAMP-DA w/o CLAMP (n=203)
CLAMP-nonDA w/ CLAMP (n=1,239)
ZLD-DA w/ ZLD (n=976)
ZLD-DA w/o ZLD (n=2,782)
ZLD-nonDA w/ ZLD (n=2,010)
-1 center 1Kb
0 15 30
0.0 1.5 3.0
0.0 0.3 0.6
0 2 4 6


B
Motifs in Type I peaks
GAF (100%)
p = 1e-4
CLAMP (100%)
p = 1e-2
Motifs in Type II peaks
CLAMP (82.6%)
p = 1e-5
GAF (65.2%)
p = 1e-4
Motifs in Type III peaks
Zelda (35.3%)
p = 1e-26
CLAMP (76.1%)
p = 1e-12
GAF (65.91%)
p = 1e-12
Motifs in Type IV peaks
GAF (64.0%)
p = 1e-52
CLAMP (71.5%)
p = 1e-43
Zelda (4.1%)
p = 1e-2

Supplementary table captions

Table 1-Source Data1. ChIP-seq read counts in peak regions in replicates of MTD and RNAi samples (DiffBind analysis).

Page 1. ZLD ChIP-seq in *clamp-i* vs. MTD in 0-2hr embryos;

Page 2. ZLD ChIP-seq in *clamp-i* vs. MTD in 2-4hr embryos.

Page 3. CLAMP ChIP-seq in *zld*-i vs. MTD in 0-2hr embryos;

Page 4. CLAMP ChIP-seq in *zld-i* vs. MTD in 2-4hr embryos;

Figure 2-supplementary table 1. ATAC-seq read counts in peak region in replicates of MTD and RNAi samples (DiffBind analysis).

Page 1. *clamp*-i vs. MTD in 0-2hr embryos.

Page 2. *clamp*-i vs. MTD in 2-4hr embryos.

Table 2-Source Data1. Peaks locations in each CLAMP or ZLD-related category.

Page 1 Type I (n=5): both DA, CLAMP ZLD co-bound

Page 2 Type II (n=23): CLAMP DA and ZLD non-DA, CLAMP ZLD co-bound

Page 3 Type III (n=88): ZLD DA and CLAMP non-DA, CLAMP ZLD co-bound

Page 4 Type IV (n=434): both non-DA, CLAMP ZLD co-bound

Page 5 DA with CLAMP 0-2hr;

Page 6 DA without CLAMP 0-2hr;

Page 7 Non-DA with CLAMP 0-2hr;

Page 8 Non-DA without CLAMP 0-2hr;

Page 9 DA with ZLD NC14+12 min;

Page 10 DA without ZLD NC14+12 min;

Page 11 Non-DA with ZLD NC14+12 min;

Page 12 Non-DA, without ZLD NC14+12 min.