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

KEYWORDS WORDS Zygotic genome activation, pioneer 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 targeting early zygotic genes (Harrison et al., 2011); and 3) modulating chromatin accessibility
56 to increase the ability of other non-pioneer TFs to bind to DNA (Schulz et al., 2015). However, a
57 large subset of ZLD binding sites (60%) are highly enriched for GA-rich motifs and have
58 constitutively open chromatin even in the absence of ZLD (Schulz et al., 2015). Therefore, we
59 and others (Schulz et al., 2015) hypothesized that other pioneer TFs which directly bind to GA-
60 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 (Leibovitch et al., 2002), and RNA Pol II recruitment (Fuda et al., 2015),
67 and is required for nuclear divisions (Bhat et al., 1996).

68

69 CLAMP is essential for early embryonic development (Rieder et al., 2017) and plays
70 several vital roles including: recruiting the MSL dosage compensation complex to the male X-

71 chromosome before ZGA (Rieder et al., 2019); activating coordinated regulation of the histone
72 genes (Rieder et al., 2017). Therefore, we hypothesized that CLAMP is a new pioneer TF that
73 regulates ZGA either cooperatively with ZLD or independent of ZLD.
74
75 Here, we identify the GA-binding factor CLAMP as a new pioneer transcription factor in
76 *Drosophila*. First, we determined that depleting maternal CLAMP and maternal ZLD cause
77 similar phenotypic changes in the early embryo, which we assessed using smFISH and
78 immunostaining for critical zygotic genes. Next, we combined genomic and biochemical
79 approaches to demonstrate that: 1) CLAMP is a novel pioneer factor which binds to
80 nucleosomal DNA that increases chromatin accessibility; 2) CLAMP and ZLD function
81 cooperatively to regulate transcription, chromatin accessibility and facilitate each other's
82 occupancy at promoters of genes which encode other TFs; 3) When ZLD is bound to a locus but
83 does not increase chromatin accessibility, CLAMP can often function to open chromatin.
84 Overall, we determined that CLAMP is a novel a pioneer transcription factor that plays an
85 essential role with ZLD in *Drosophila* zygotic genome activation.

86 **RESULTS**

87

88 **Depletion of maternal CLAMP disrupts of zygotic patterning gene expression and**
89 **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-supplementary file 1A**) that had
96 significantly (*p-value* < 0.05, DESeq2, RNA-seq) reduced expression in early embryos when
97 depleting maternal CLAMP (Rieder et al., 2017). We then used single-molecule fluorescent in
98 situ hybridization (smFISH) for *even-skipped* and *runt,* and immunostaining for *neurotactin* to
99 determine how the depletion of maternal *clamp* or *zld* alters phenotypic patterning and
100 cytoskeletal integrity in blastoderm stage embryos **(Figure 1)**. Knockdown of *clamp* or *zld* in
101 early embryos was validated by qRT-PCR and western blotting (**Figure 1 – figure supplement**
102 **1B, C**).

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. Similar, but slightly stronger, phenotypic changes
112 were observed in *zld* maternal depletion embryos **(Figure 1A, right)**, indicating that CLAMP and
113 ZLD have critical roles in establishing embryonic patterning in pre-cellular blastoderm embryos.

114 Moreover, all of embryos depleted for maternal *clamp* or maternal *zld* (n = 10) show defective
115 *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 WT pattern of cytoskeletal elements, which
122 can be seen in the e*gfp* RNAi control embryos **(Figure 1B, left)**. Embryos depleted for maternal
123 *zld* also reveal similar patterns of discordant nuclei. More than 50% of embryos depleted for
124 maternal *clamp* (n = 23) and 100% of embryos depleted for maternal *zld* (n=10) show NRT
125 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 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 factor, we performed electromobility shift assays (EMSAs) that test the
134 intrinsic capacity of CLAMP to directly interact with nucleosomes *in vitro*. First, we identified a
135 240 bp region of the X-linked 5C2 locus **(Figure 1C)** that CLAMP binds to in cultured S2 cells
136 and exhibited decreased chromatin accessibility in the absence of CLAMP (J. Urban et al.,
137 2017). This region is also occupied by a nucleosome **(Figure 1C)**, suggesting that CLAMP
138 promotes accessibility of this region while binding to nucleosomes in S2 cells.
139
140 We then performed *in vitro* nucleosome assembly using 240 bp of DNA from the 5C2
141 locus that contains three CLAMP-binding motifs and used 5C2 naked DNA as a control. We

142 found that both the CLAMP DNA binding domain (DBD**, Figure 1D**) and full-length protein (F/L**,**
143 **Figure 1E**) can bind and shift naked 5C2 DNA and nucleosomes assembled with 5C2 DNA.
144 Increased protein concentration results in a secondary "super" shift species **(Figure 1D & 1E)**,
145 indicating that multiple CLAMP molecules may occupy the three CLAMP-binding motifs. Both
146 full length CLAMP and CLAMP DBD are fused to the maltose binding protein (MBP), which we
147 previously demonstrated does not bind to DNA independent of CLAMP or alter the specificity of
148 CLAMP binding (Kaye et al., 2018). We also previously determined the specificity of CLAMP
149 binding to its GA-rich motif *in vivo* and *in vitro* (Soruco et al., 2013; Kaye et al., 2018).
150 Importantly, here we demonstrate that CLAMP can directly bind to nucleosomal DNA, a unique
151 feature of pioneer TFs that is not shared with the majority of TFs (Lai et al., 2018).
152
153 **CLAMP regulates chromatin accessibility in early embryos**
154
155 Another essential characteristic of pioneer transcription factors is that they can establish
156 and maintain the accessibility of their DNA target sites, allowing other TFs to bind to DNA and
157 activate transcription (Zaret and Carroll, 2011; Iwafuchi-Doi et al., 2016). We previously used
158 MNase-seq (J. Urban et al., 2017) to determine that CLAMP guides MSL complex to GA-rich
159 sequences by promoting an accessible chromatin environment on the male X-chromosome in
160 cultured S2 cells. Furthermore, GA-rich motifs are enriched in regions that remain accessible in
161 the absence of the pioneer factor ZLD (Schulz et al., 2015). Therefore, we hypothesized that
162 CLAMP regulates chromatin accessibility during ZGA.
163
164 To test our hypothesis, we performed Assay for Transposase-Accessible Chromatin
165 using sequencing (ATAC-seq, see **Figure 2-supplementary file 1** for analysis pipeline) on 0-
166 2hr (pre-ZGA) and 2-4hr (post-ZGA) embryos with wild-type levels of CLAMP [maternal triple
167 GAL4 driver (MTD) alone (Ni et al., 2011)] and embryos depleted for maternal CLAMP using the
168 MTD driver (*clamp-i*). We identified differentially accessible (DA) regions **(Figure 2A,** and **Table**
169 **1)** by comparing ATAC-seq reads between MTD and *clamp-i* embryos using Diffbind (Stark and

170 Brown, 2019). High Pearson correlation for DA regions between replicates indicates robust
171 reproducibility **(Figure 2-figure supplement 1A-B,** and **Figure 2- supplement table 1** for
172 mapping statistics**)**. Over a thousand genomic regions exhibited reduced chromatin accessibility
173 in the absence of CLAMP (**Figure 2A,** 0-2hr: 1,675; 2-4hr: 1,192), indicating that chromatin
174 accessibility of these regions requires CLAMP. In contrast, very few loci (0-2hr: 76; 2-4hr: 260)
175 increased their accessibility in the absence of CLAMP **(Figure 2A** and **Table 1)**. Gene Ontology
176 (GO) analysis **(Figure 2-figure supplement 1C-D)** indicates that CLAMP increases
177 accessibility of chromatin regions that are mainly within DNA-binding, RNA Pol II binding, and
178 enhancer-binding TF encoding genes **(Figure 2-figure supplement 1C-D)**.
179
180 **Table 1. The number of ATAC-seq peaks called in each replicate and the number of**
181 **differentially accessible peaks.**

Open chromatin peaks (read fragments <= 100bp)	MTD (caller: MACS2)			Maternal *Clamp*-i (caller: MACS2)			Differential accessible peaks (DiffBind, minimal overlap in 2 replicates)		
	#1	#2	#3	#1	#2	#3	Up	Down	None
0-2hr	7,951	13,947	10,417	8,509	8,908	2,157	76	1,675	9,465
2-4hr	26,280	20,896	14,131	14,697	10,884	18,159	260	1,192	18,851

182
183 Moreover, a subset of DA regions were bound by CLAMP suggesting that CLAMP
184 directly regulates their chromatin accessibility [17.5% (294/1,675) at 0-2hr; 20% (241/1,192) at
185 2-4hr] **(Table 2).** For example, the *mod(mdg4)* promoter is a DA site and shows a reduction of
186 CLAMP occupancy after *clamp* RNAi as expected **(Figure 2B)**. We also determined the
187 distribution of DA sites and CLAMP binding sites throughout the genome **(Figure 2C)**. While DA
188 sites were significantly ($p < 0.05$, Fisher's exact test) enriched in promoter regions (67.34%),
189 CLAMP binds most often to both promoters (29.08%) and introns that are not first introns
190 (27.72%). Therefore, CLAMP is required to establish or maintain open chromatin at promoters
191 but could also play other roles at intronic regions. Furthermore, motif analysis identified both

192 GA-rich motifs and ZLD motifs enriched at regions which require CLAMP for their accessibility in
193 0-2hr embryos **(Figure 2D)**. These data suggest that CLAMP also regulates the accessibility of
194 some ZLD binding sites, a hypothesis that we will discuss further below.
195
196 We further determined whether CLAMP-mediated chromatin accessibility could
197 specifically drive early transcription by examining the relationship between the chromatin
198 accessibility (DA, ATAC-seq) changes and gene expression of the nearest differential
199 expressed gene (DEG) as measured by RNA-seq (Rieder et al., 2017). We obtained significant
200 (*p*-values < 0.001, hypergeometric test, N= 15,682 total fly genes) overlap between CLAMP-
201 mediated down-regulate chromatin accessibility and down-regulate gene expression changes at
202 both time points: 0-2hr and 2-4hr (**Figure 2E)**. Overall, our ATAC-seq data indicate that CLAMP
203 promotes accessibility of chromatin during ZGA, a key function of pioneer TFs.
204
205 **CLAMP and ZLD binding mediate CLAMP-related chromatin accessibility**
206
207 To determine how direct binding of CLAMP relates to CLAMP-mediated chromatin
208 accessibility, we integrated CLAMP binding sites from ChIP-seq with ATAC-seq peaks. Inspired
209 by a FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements) study which
210 identified ZLD-mediated chromatin accessible regions (Schulz et al., 2015), we defined four
211 classes of CLAMP-related peaks: 1) DA with CLAMP binding, 2) DA without CLAMP binding, 3)
212 Non-DA with CLAMP binding, 4) Non-DA without CLAMP binding **(Figure 3A, Figure 3-figure**
213 **supplement 1A,** and **Table 2)**. Heatmap of z-score normalized ATAC-seq read counts (CPM) in
214 MTD and *clamp-i* embryos revealed that DA regions show a significant decrease in accessibility
215 in embryos lacking CLAMP. In contrast, non-DA regions remain accessible in the absence of
216 CLAMP **(Figure 3A,** and **Figure 3-figure supplement 1A).** In this way, we defined regions of
217 the genome where chromatin accessibility is dependent on CLAMP, which largely include
218 promoters of genes encoding transcriptional regulators.
219

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

Top 6,000 ATAC-seq peaks (CPM)	DA w/ CLAMP	DA w/o CLAMP	Non-DA w/ CLAMP	Non-DA w/o CLAMP
0-2hr	249	835	267	2,101
2-4hr	241	199	1,618	2,826
Top 6,000 ATAC-seq peaks in CPM (Hannon et al., 2017)	DA w/ ZLD	DA w/o ZLD	Non-DA w/ ZLD	Non-DA w/o ZLD
NC14+12min	806	426	1,331	2,269

221

222 Next, we generated a heatmap to examine the ChIP-seq occupancy of CLAMP and ZLD
223 at the four different classes of loci defined above, based on the effect of CLAMP on chromatin
224 accessibility (**Figure 3A** and **Figure 3-figure supplement 1A-B**). We found that highest
225 occupancy CLAMP sites are those at which CLAMP regulates chromatin accessibility (DA
226 w/CLAMP loci). Interestingly, DA w/CLAMP sites show higher occupancy of CLAMP compared
227 to Non-DA w/CLAMP sites, suggesting that enhanced CLAMP occupancy promotes the ability
228 of CLAMP to regulate chromatin accessibility **(Figure 3A** and **Figure 3-figure supplement 1A)**.
229

230 ZLD ChIP-seq and motif analysis revealed that ZLD and its motif are enriched at all of
231 the CLAMP-related classes of sites **(Figure 3A** and **Figure 3-figure supplement 1A)**.
232 Moreover, ZLD requires CLAMP for its occupancy at sites where CLAMP regulates chromatin
233 accessibility (DA regions) but binds independently of CLAMP to regions that do not depend on
234 CLAMP for accessibility (Non-DA regions) (**Figure 3A** and **Figure 3-figure supplement 1A).**
235 Together, these data are consistent with a model in which high levels of CLAMP occupancy
236 promote changes in chromatin accessibility. However, CLAMP-dependent changes in chromatin
237 accessibility promote the occupancy of ZLD even if there is not high CLAMP occupancy at these
238 loci.
239

240 To compare CLAMP- and ZLD-mediated chromatin accessibility during ZGA, we also
241 defined four ZLD-related classes of genomic loci using ZLD binding sites from ChIP-seq (this

242 study) and ATAC-seq datasets (Hannon et al., 2017; Soluri et al., 2020). The ZLD-related
243 ATAC-seq data were generated from embryos laid by wildtype (wt) mothers or mothers with *zld*
244 germline clone (*zld*-) at the nuclear cycle 14 (NC14) +12 min time point. Specifically, we defined
245 four classes of genomic loci for ZLD-related classes: 1) DA with ZLD, 2) DA without ZLD, 3)
246 Non-DA with ZLD, 4) Non-DA without ZLD **(Figure 3B** and **Table 2).** A heatmap of ATAC-seq
247 data (CPM) in wt and *zld*- embryos and ZLD ChIP-seq data in MTD and *zld-i* embryos in each
248 ZLD-related class also demonstrates that the read enrichment is consistent with our
249 classification **(Figure 3B** and **Figure 3-figure supplement 1C)**. We found the read enrichment
250 of ZLD ChIP-seq in MTD and *zld-i* embryos corresponds to our classification of ZLD binding
251 classes **(Figure 3B)**.
252
253 We also plotted CLAMP ChIP-seq at the four ZLD-related classes of sites. Strikingly,
254 CLAMP occupancy was highly enriched at sites where ZLD binds but does not regulate
255 chromatin accessibility (Non-DA w/ ZLD), even in the absence of maternal *zld* **(Figure 3B).**
256 Moreover, motif analysis also demonstrates that the CLAMP motif is significantly enriched in the
257 Non-DA with ZLD region**,** supporting our hypothesis that CLAMP binds to these regions and
258 keeps them open, even in the absence of ZLD **(Figure 3B)**.
259
260 Next, we determined the relationship between the four classes of CLAMP-related peaks
261 and four classes of ZLD-related peaks by performing a pairwise intersection of genomic regions
262 **(Table 3)**. At 0-2hrs, we observed that the sites which require CLAMP for their chromatin
263 accessibility but have low/no CLAMP occupancy (DA w/o CLAMP) significantly overlap with all
264 ZLD classes ($p < 0.05$, Hypergeometric test, N=15,682 total fly genes) **(Table 3)**. Among these
265 overlaps, the lowest *p*-value is between DA w/o CLAMP and DA w/ ZLD. In contrast, sites that
266 have ZLD occupancy but do not require ZLD for their chromatin accessibility (Non-DA w/ ZLD)
267 overlap significantly with all CLAMP classes. Interestingly, the lowest *p*-value is between DA w/
268 CLAMP and Non-DA w/ ZLD. At 2-4hr, we also found significant overlap between sites that are
269 bound by both CLAMP and ZLD (DA w/ CLAMP and Non-DA w/ ZLD; Non-DA w/ CLAMP and

DA w/ ZLD). This relationship between classes of sites suggests that CLAMP and ZLD act cooperatively: CLAMP promotes chromatin accessibility at locations bound by ZLD that do not depend on ZLD for their accessibility and vice versa. As Schulz et al. (2015) predicted, a GA-binding TF can increase the accessibility of sites bound by ZLD which do not require ZLD for chromatin accessibility. Here, we determined that CLAMP is one of the GA-binding TFs which can function cooperatively with ZLD.

Table 3. Significance of pairwise overlaps between four classes of CLAMP-related peaks and four classes of ZLD-related peaks in 0-2hr and 2-4hr embryos

Overlap peaks (peak number)	DA, w/ CLAMP 0-2hr (249)	DA, w/o CLAMP 0-2hr (835)	Non-DA, w/ CLAMP 0-2hr (267)	Non-DA, w/o CLAMP 0-2hr (2,101)
DA, w/ ZLD NC14+12min (806)	22 (p = 0.0006)	168 (p = 2.57e-59)	20 (p = 0.04)	157 (p = 9.13e-8)
DA, w/o ZLD NC14+12min (426)	0 (-)	62 (p = 2.13e-13)	0 (-)	55 (-)
Non-DA, w/ ZLD NC14+12min (1,331)	109 (p = 8.99e-53)	117 (p = 8.98e-09)	101 (p = 4.89e-12)	280 (p = 6.2e-17)
Non-DA, w/o ZLD NC14+12min (2,269)	9 (-)	176 (p = 2.4e-08)	11 (-)	579 (p = 1.4e-66)
Overlap peaks (peak number)	**DA, w/ CLAMP 2-4hr (241)**	**DA, w/o CLAMP 2-4hr (199)**	**Non-DA, w/ CLAMP 2-4hr (1,618)**	**Non-DA, w/o CLAMP 2-4hr (2,826)**
DA, w/ ZLD NC14+12min (806)	12 (-)	2 (-)	133 (p = 4.04-09)	113 (-)
DA, w/o ZLD NC14+12min (426)	2 (-)	2 (-)	4 (-)	48 (-)
Non-DA, w/ ZLD NC14+12min (1,331)	82 (p = 6.35e-30)	11 (-)	589 (p = 3.5e-263)	362 (p = 1.55e-19)
Non-DA, w/o ZLD NC14+12min (2,269)	25 (-)	37 (-)	179 (-)	1,044 (p = 2.87e-257)

"-" not significant p-value; w/: with; w/o: without
***p*-values are calculated by Hypergeometric test, N=15,682 total fly genes**

CLAMP and ZLD bind cooperatively to promoters

285 To directly determine how CLAMP and ZLD impact each other's binding, we performed
286 ChIP-seq (see **Figure 4-supplementary file 1** for analysis pipeline) for CLAMP and ZLD in
287 control MTD embryos and embryos that were maternally depleted for each factor with RNAi at
288 the same two time points we used for our ATAC-seq experiments: before ZGA (0-2hr) and
289 during and after ZGA (2-4hr) **(Figure 4A-B** and **Table 4)**. Overall, ZLD has more peaks (0-2hr:
290 6,974; 2-4hr: 8,035) across the whole genome than CLAMP (0-2hr: 4,962, 2-4hr: 7,564) in
291 control MTD embryos. As we hypothesized, CLAMP and ZLD peaks significantly overlap ($p <$
292 0.05, Hypergeometric test, N=15,682 total fly genes) **(Figure 4-figure supplement 1A)**.
293
294 Before ZGA (0-2hrs), ZLD showed a higher enrichment at promoters than CLAMP. In
295 contrast, CLAMP had a similar distribution to ZLD at the 2-4hr time point, indicating that CLAMP
296 binds to promoters later than ZLD **(Figure 4C)**. Depletion of either maternal *zld* or *clamp* mRNA
297 altered the genomic distribution of CLAMP and ZLD: both factors shifted their occupancy from
298 promoters to introns. Interestingly, maternal *zld* RNAi no longer affects CLAMP binding to
299 promoters at 2-4hr which may be because the maternal RNAi is not as effective at reducing ZLD
300 protein levels at this time point **(Figure 4C)**. Therefore, CLAMP and ZLD change the distribution
301 of each other's occupancy to increase their binding to promoters instead of introns.
302
303 **Table 4. The number of peaks for CLAMP and ZLD in MTD, *clamp-i* and *zld-i* embryos and**
304 **differential binding peaks**

ChIP-seq Peaks (consensus in at least 2 replicates)	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

305
306

307 Next, we defined the sites that show differential binding (DB**, Figures 4D-E** and **Table 4**)
308 of CLAMP and ZLD in the absence of each other's maternally deposited mRNA using Diffbind
309 (Stark and Brown, 2019). We found a significant reduction of ZLD binding in the absence of
310 CLAMP. There were 274 (0-2hr) and 1,289 (2-4hr) sites where ZLD binding decreased in
311 *clamp-i* compared to MTD controls (down-DB) **(Figures 4D, Figure 4-figure supplement 1B,**
312 and **Table 4)**. Fewer ZLD binding sites increased in occupancy after *clamp* RNAi: 8 (0-2hr) and
313 233 (2-4hr) up-DB sites. In contrast, loss of ZLD had a weaker impact on CLAMP binding: 390
314 (0-2 hr) and 30 (2-4 hr) down-DB sites **(Figures 4E, Figure 4-figure supplement 1C,** and
315 **Table 4)**. We identified very few sites where CLAMP occupancy increases after *zld* RNAi (up-
316 DB sites: 0-2hr: 54, 2-4hr: 3). Moreover, CLAMP and ZLD down-DB sites were also significantly
317 overlapped ($p < 0.05$, Hypergeometric test, N=15,682 total fly genes) throughout ZGA **(Figure**
318 **4-figure supplement 1D)**.
319
320 To further understand how CLAMP and ZLD bind to dependent (down-DB) and
321 independent (non-DB) sites, we determined the genomic distribution of these two types of sites.
322 We found that dependent sites are enriched at promoters and transcription start sites, while
323 independent sites are mainly located in introns **(Figure 4F)**. Overall, dependent peaks are much
324 broader in size and are located at promoters. In contrast, independent sites are narrower and
325 located within introns (**Figures 4G-H** and **Figure 4-figure supplement 1E)**. On average, the
326 peak size of dependent sites (400-500bp) is almost double that of independent sites (200-
327 250bp) with significant differences in peak size for both TFs at both time points ($p < 0.001$,
328 Mann-Whitney U-test) **(Figure 4I)**. In summary, these results support a model in which there is
329 a cooperative relationship between CLAMP and ZLD binding at promoters of genes which
330 encode other transcription factors.
331
332 **CLAMP and ZLD function together to regulate genes during zygotic genome activation by**
333 **mediating chromatin accessibility**
334

335 One essential function of a pioneer TF is its ability to activate zygotic genes (Zaret and
336 Carroll, 2011). To define how CLAMP regulates transcription in early embryos, we examined the
337 effect of maternal CLAMP depletion on the expression of maternally-deposited or zygotically-
338 transcribed genes using mRNA-seq data (Rieder et al., 2017). We found that the expression
339 levels of zygotically-transcribed genes but not maternally-deposited genes were significantly
340 downregulated in embryos lacking CLAMP ($p < 0.001$, Mann-Whitney U-test) **(Figure 5A).**
341 Therefore, CLAMP has a similar effect on zygotic transcription to that which has been previously
342 reported for ZLD (Liang et al., 2008; McDaniel et al., 2019).
343
344 We next asked whether the ability of CLAMP to bind to genes directly regulates zygotic
345 gene activation by integrating ChIP-seq data (this study) with RNA-seq data (Rieder et al., 2017;
346 Combs and Eisen, 2017). We determined that genes strongly bound by CLAMP showed a
347 significant ($p < 0.001$, Mann-Whitney U-test) level of gene expression reduction after *clamp*
348 RNAi (Rieder et al., 2017) compared to weakly bound or unbound genes **(Figure 5B** and **Figure**
349 **5-figure supplement 1A).** We also observed a significant ($p < 0.001$, Mann-Whitney U-test)
350 change in gene expression in *zld*- embryos (Combs and Eisen, 2017) for the genes which are
351 strongly bound by ZLD **(Figure 5C** and **Figure 5-figure supplement 1B)**. Furthermore, the
352 numbers of genes in the CLAMP or ZLD-regulated groups are similar size to each other. Also,
353 the magnitude of the transcriptional changes are similar for sites that directly bind CLAMP or
354 ZLD. Moreover, we compared and contrasted the transcriptional roles of CLAMP and ZLD in
355 early embryos at genes that have different temporal expression patterns. We found that CLAMP
356 binding was enriched at mid- and late-transcribed zygotic genes (categories defined in Li et al.,
357 2014), while ZLD binding is stronger in early-transcribed zygotic genes **(Figure 5-figure**
358 **supplement 1C-D)**.
359
360 We hypothesized that CLAMP and ZLD could regulate each other's binding to precisely
361 drive the transcription of target genes. To test this hypothesis, we plotted the gene expression
362 changes caused by depleting maternal ZLD at genes where CLAMP regulates ZLD binding. The

363 depletion of maternal *zld* significantly (p = 2.1e-6, Mann-Whitney U-test) reduces the expression
364 of genes where ZLD regulates CLAMP binding (down-DB) more than sites where CLAMP binds
365 independently of ZLD (non-DB) at the 0-2hr but not the 2-4hr (p = 0.53) time point **(Figure 5D)**.
366 Therefore, ZLD may specifically regulate early genes where ZLD promotes CLAMP binding.
367 Also, compared to genes where ZLD binds independent of CLAMP, genes where ZLD binding is
368 regulated by CLAMP had a significant (p < 0.001, Mann-Whitney U-test) expression reduction
369 after *clamp* RNAi at both 0-2hr and 2-4hr time points **(Figure 5D)**. Thus, CLAMP may regulate
370 the transcription of genes targeted by ZLD by promoting ZLD binding.
371
372 Previous proteomic studies (J. A. Urban et al., 2017; Hamm et al., 2017, 2015) found no
373 evidence that CLAMP and ZLD could directly contact each other at the protein level, suggesting
374 that CLAMP and ZLD may regulate each other via binding to their own DNA motifs. Therefore,
375 we analyzed the motifs enriched at dependent (down-DB) and independent (non-DB) sites. We
376 found that dependent sites are enriched for motifs specific for the required protein, which are not
377 present at the independent sites **(Figure 5-figure supplement 1E).** For example, the ZLD motif
378 is only enriched at sites where CLAMP requires ZLD for binding (CLAMP down-DB) but not at
379 sites where CLAMP binds independently of ZLD (CLAMP non-DB). Similarly, the CLAMP motif
380 is only enriched at sites where ZLD requires CLAMP for binding (ZLD down-DB) **(Figure 5-**
381 **figure supplement 1E).** Therefore, the presence of specific CLAMP and ZLD motifs correlates
382 with their ability to promote each other's binding which regulates the expression of each other's
383 target genes.
384
385 Lastly, given the cooperative relationship between CLAMP and ZLD binding to
386 chromatin, which regulates transcription, we measured chromatin accessibility changes at
387 dependent and independent sites **(Figures 5E-F).** We found that sites where ZLD regulates
388 CLAMP binding (CLAMP down-DB) have high chromatin accessibility while sites where CLAMP
389 binds independent (CLAMP non-DB) of ZLD showed low chromatin accessibility in both control
390 and *zld*- embryos. Interestingly, the accessibility slightly increases upon the loss of ZLD in

391 CLAMP down-DB at 0-2hr **(Figures 5E).** An active TF binding to DNA could prevent Tn5
392 cleavage at those regions (Yan et al., 2020). Therefore, loss of ZLD and CLAMP binding could
393 result in a perceived accessibility increase, as measured by ATAC-seq. Similarly, the average
394 ATAC-seq signals significantly reduced at site ZLD dependent on CLAMP to bind (ZLD down-
395 DB sites) in *clamp-i* embryos compared to MTD controls. Simultaneously, the accessibility at
396 where ZLD bind independently (ZLD non-DB) is much lower than that at ZLD DB sites **(Figures**
397 **5F)**. Overall, our data are consistent with the direct action of CLAMP and ZLD on chromatin
398 accessibility, which influences the binding of both TFs to their binding sites and regulates target
399 gene expression.

400 **DISCUSSION**

401

402 Two questions central to early embryogenesis of all metazoans are how and where do
403 early transcription factors work together to drive chromatin changes and zygotic genome
404 activation. Here, we identified CLAMP as a pioneer transcription factor that is essential for early
405 embryonic development, which directly binds to nucleosomal DNA **(Figure 1)**, establishes
406 and/or maintains chromatin accessibility **(Figure 2-3)**, facilitates the binding of ZLD to promoters
407 **(Figure 4)** and regulates zygotic genome activation **(Figure 5)**. We further discovered that
408 CLAMP and ZLD cooperatively regulate each other's binding during ZGA. Also, we determined
409 that both CLAMP and ZLD alter chromatin accessibility which influences the binding of both TFs
410 to their binding sites **(Figure 6)**. Overall, we identify a new pioneer TF, CLAMP, and provide
411 critical insight into how CLAMP and ZLD function cooperatively to remodel zygotic genome
412 accessibility, which drives zygotic genome activation.

413

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

416

417 Our ATAC-seq and EMSA data suggest that CLAMP is a novel pioneer transcription
418 factor in early *Drosophila* embryos. Pioneer transcription factors (Mayran and Drouin, 2018),
419 such as FoxA1 (Cirillo et al., 2002), have distinct characteristics from other TFs such as direct
420 biochemical interaction with nucleosomes, activating gene expression in the embryo,
421 establishing accessible chromatin domains, and facilitating binding of additional TFs. Analogous
422 to prior work on ZLD (Schulz et al., 2015; McDaniel et al., 2019), we demonstrate that CLAMP
423 also possesses these characteristics of a pioneer factor.

424

425 We defined CLAMP-dependent and ZLD-dependent peaks in early embryos which
426 reveals cooperative roles for CLAMP and ZLD in defining chromatin accessibility and zygotic
427 transcription during ZGA **(Figure 6)**: 1) CLAMP promotes ZLD enrichment at sites where

428 CLAMP binding increases chromatin accessibility (DA, w/ CLAMP) and further regulates ZLD
429 target gene expression. 2) ZLD increases chromatin accessibility at sites that are bound by
430 CLAMP but do not depend on CLAMP for chromatin accessibility (Non-DA, w/ CLAMP). 3) ZLD
431 facilitates CLAMP binding at sites where ZLD regulates chromatin accessibility (DA, w/ ZLD)
432 and promotes CLAMP target gene expression. 4) CLAMP opens chromatin at locations bound
433 by ZLD that do not depend on ZLD for their accessibility (Non-DA, w/ ZLD). Overall, our data
434 suggest that CLAMP functions cooperatively with ZLD to regulate chromatin accessibility and
435 gene expression of the early zygotic genome.
436
437 Although we have demonstrated an instrumental role for CLAMP in defining the open
438 chromatin landscape in early embryos, our data show that CLAMP does not increase chromatin
439 accessibility at all genomic loci. CLAMP mainly drives chromatin accessibility changes at
440 promoters of genes that encode transcription factors which are important for early development.
441 Moreover, ZLD binding and/or chromatin accessibility is not regulated by maternal depletion of
442 CLAMP at all GA-rich sites in the genome. A previous study found that GAF is also enriched at
443 these same ZLD-bound regions where ZLD is not required for chromatin accessibility (Schulz et
444 al., 2015). Both CLAMP and GAF are deposited maternally (Rieder et al., 2017; Hamm et al.,
445 2017) and bind to similar GA-rich motifs (Kaye et al., 2018). To test whether GAF compensates
446 for the depletion of CLAMP or ZLD, we tried to perform GAF RNAi in the current study to
447 prevent GAF from compensating for CLAMP depletion. However, we could not achieve
448 depletion of GAF in early embryos by RNAi, likely due to autoregulation of its own promoter and
449 its prion-like self-perpetuating function (Tariq et al., 2013).
450
451 We previously demonstrated that competition between CLAMP and GAF at GA-rich
452 binding sites is essential for MSL complex recruitment in S2 cells (Kaye et al., 2018).
453 Furthermore, CLAMP excludes GAF at the histone locus which co-regulates genes that encode
454 the histone proteins (Rieder et al., 2017). However, we also observed synergistic binding
455 between CLAMP and GAF at many additional binding sites (Kaye et al., 2018). Yet, the

456 relationship between CLAMP and GAF in early embryos remains unknown. It is very possible
457 that the competitive relationship has not been established in early embryos, since dosage
458 compensation has not yet been initiated (Prayitno et al., 2019).
459
460 The GA-rich sequences targeted by CLAMP and GAF are distinct from each other *in*
461 *vivo* and *in vitro*. GAF binding sites typically have 3.5 GA repeats; however, GAF is able to bind
462 to as few as three bases (GAG) within the *hsp70* promoter and *in vitro* (Wilkins and Lis, 1999).
463 In contrast, CLAMP binding sites contain an 8-bp core with a less well-conserved second GA
464 dinucleotide within the core (GA__GAGA) (Alekseyenko et al., 2008). CLAMP binding sites also
465 include a GAGAG pentamer at a lower frequency than GAF binding sites, and flanking bases
466 surrounding the 8-bp core are critical for CLAMP binding (Kaye et al., 2018).
467
468 Therefore, GAF and CLAMP may have overlapping and non-overlapping functions at
469 different loci, tissues, or developmental stages. Moreover, another TF Pipsqueak (Psq) also
470 targets sites containing the GAGAG motif, and has multiple functions during oogenesis,
471 embryonic pattern formation and functions with Polycomb in three-dimensional genome
472 organization (Lehmann et al., 1998; Gutierrez-Perez et al., 2019). In the future, an optogenetic
473 inactivation approach could be used to remove CLAMP, GAF and/or Psq simultaneously in a
474 spatial and temporal manner (McDaniel et al., 2019).
475
476 **CLAMP and ZLD regulate each other's binding via their own motifs**
477
478 ZLD is an essential TF that regulates activation of the first set of zygotic genes during
479 the minor wave of ZGA and thousands of genes transcribed during the major wave of ZGA at
480 NC14 (Liang et al., 2008; Harrison et al., 2011). ZLD also establishes and maintains chromatin
481 accessibility of specific regions and facilitates transcription factor binding and early gene
482 expression (Sun et al., 2015; Schulz et al., 2015). CLAMP regulates histone gene expression
483 (Rieder et al., 2017), X chromosome dosage compensation (Soruco et al., 2013), and

484 establishes/maintains chromatin accessibility (J. Urban et al., 2017). Nonetheless, it remained
485 unclear whether and how CLAMP and ZLD functionally interact during ZGA. Here, we
486 demonstrate a cooperative relationship between CLAMP and ZLD at hundreds of promoters
487 which encode other transcriptional regulators.
488
489 ZLD often regulates CLAMP occupancy earlier than CLAMP regulates ZLD occupancy.
490 Genomic loci at which CLAMP is dependent on ZLD early (0-2hr) in development often become
491 independent from ZLD later (2-4hr), with the caveat that ZLD depletion is not as effective later in
492 development. Therefore, CLAMP may require the pioneering activity of ZLD to access specific
493 loci before ZGA, but ZLD is no longer necessary once CLAMP binding is established. Also, our
494 results suggest that CLAMP is a potent regulator of ZLD binding, especially in 2-4hr embryos.
495 Furthermore, ZLD can bind to many more promoter regions at 0-2 hours, while CLAMP mainly
496 binds to introns early in development but occupies promoters later at 2-4 hours. Therefore,
497 CLAMP may require ZLD to increase chromatin accessibility of these promoter regions (Schulz
498 et al., 2015). In fact, ZLD is sufficient to activate a subgroup of early genes, although most ZLD
499 bound regions are not active until NC14 (Bosch et al., 2006).
500
501 In addition to its role in embryonic development, CLAMP also plays an essential role in
502 targeting the MSL male dosage compensation complex to the X-chromosome (Soruco et al.,
503 2013). *Drosophila* embryos initiate X-chromosome counting in NC12 and start the sex
504 determination cascade prior to the major wave of ZGA at NC14 (Gergen, 1987; Bosch et al.,
505 2006). However, most dosage compensation is initiated much later in embryonic development
506 (Prayitno et al., 2019). Therefore, our data support a model in which CLAMP functions early in
507 the embryo prior to MSL complex assembly to open up specific chromatin regions for MSL
508 complex recruitment later (J. Urban et al., 2017; Rieder et al., 2019). Moreover, ZLD likely
509 functions primarily as an early pioneer factor, whereas CLAMP has pioneering functions in both
510 early- and late-ZGA embryos. Consistent with this hypothesis, CLAMP binding is enriched at
511 both early and late zygotic genes. In contrast, ZLD binding is localized mainly to early zygotic

genes, suggesting a that there may be a sequential relationship between occupancy of these two TFs at some loci during early embryogenesis.

The different characteristics of dependent and independent CLAMP and ZLD binding sites also provide insight into how early transcription factors work together to regulate ZGA. At dependent sites, there are often relatively broad peaks of CLAMP and ZLD that are significantly enriched for clusters of motifs for the required protein. Our CLAMP EMSAs and those previously reported (Kaye et al., 2018) also show multiple shifted bands consistent with possible multimerization. CLAMP contains two central disordered prion-like glutamine-rich regions (Q domains) (Kaye et al., 2018), a type of domain that is critical for transcriptional activation and multimerization *in vivo* in several TFs, including GAF (Wilkins and Lis, 1999). Moreover, glutamine-rich repeats alone can be sufficient to mediate stable protein multimerization *in vitro* (Stott et al., 1995). Therefore, it is reasonable to hypothesize that the CLAMP glutamine-rich domain also functions in CLAMP multimerization.

In contrast, ZLD fails to form dimers or multimers (Hamm et al., 2015, 2017), indicating that ZLD most likely binds as a monomer. There is no evidence that CLAMP and ZLD have any direct protein-protein interaction at sites where they depend on each other to bind. For example, mass spectrometry results that identified dozens of CLAMP-associated proteins did not identify ZLD (J. A. Urban et al., 2017). No validated protein-protein interactions of ZLD with itself as a multimer or between ZLD and any other TFs have been identified to date using diverse methods (Hamm et al., 2017). Taken together, our study suggests that regulating the chromatin landscape in early embryos to drive ZGA requires the cooperation of multiple transcription factors.

536 **MATERIALS AND METHODS**

537

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

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

544 Cell pellets were harvested by centrifugation and stored at -80°C. Cell pellets were
545 resuspended in 20 mM Tris, 1M NaCl, 0.1mM ZnCl2, 10 mM imidazole pH 8.0 with one EDTA-
546 free protease inhibitor tablet (Roche) and lysed using an Emulsiflex C3 (Avestin). The lysate
547 was cleared by centrifugation at 20,000 rpm for 50 min at 4°C, filtered using a 0.2 µm syringe
548 filter, and loaded onto a HisTrap HP 5 mL column. The protein was eluted with a gradient from
549 10 to 300 mM imidazole in 20 mM Tris, 1.0 M NaCl pH 8.0, and 0.1mM ZnCl2. Fractions
550 containing MBP-CLAMP full-length were loaded onto a HiLoad 26/600 Superdex 200 pg column
551 equilibrated in 20 mM Tris 1.0 M NaCl pH 8.0. Fractions containing full-length CLAMP were
552 identified by SDS-PAGE and concentrated using a centrifugation filter with a 10 kDa cutoff
553 (Amicon, Millipore) and frozen as aliquots.

554

555 ***In vitro* assembly of nucleosomes**

556

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

563

564 5C2 (276 bp), **bold** sequences are CLAMP-binding motifs, underlined sequences are primer
565 binding sequences:
566 TCGACGACTAGTTTAAAGTTATTGTAGTTCTTAGAGCAGAATGTATTTTAAATATCAATGTTT
567 CGATGTAGAAATTGAATGGTTTAAATCACGTTCACACAACTTA**GAAAGAGATAG**CGATGGC
568 GGTGT**GAAAGAGAGCGAGATAG**TTGGAAGCTTCATG**GAAATGAAAGAGAGGTAG**TTTTT
569 GGAAATGAAAGTTGTACTAGAAATAAGTATTTTATGTATATAGAATATCGAAGTACAGAAATT
570 CGAAGCGATCTCAACTTGAATATTATATCG
571
572 **Primers for 5C2 region (product is 240bp):**
573 Forward: TTGTAGTTCTTAGAGCAGAATGT
574 Reverse: GTTGAGATCGCTTCGAATTT
575
576 **Electromobility shift assays**
577 DNA or nucleosome probes at 35nM (700fmol/reaction) were incubated with MBP-
578 tagged CLAMP DBD protein or MBP-tagged full-length CLAMP protein in a binding buffer. The
579 binding reaction buffer conditions are similar to conditions previously used to test ZLD
580 nucleosome binding (McDaniel et al. 2019) in 20 ul total volume: 7.5ul BSA/HEGK buffer (12.5
581 mM HEPES, PH 7.0, 0.5 mM EDTA, 0.5 mM EGTA, 5% glycerol, 50 mM KCl, 0.05 mg/ml BSA,
582 0.2 mM PMSF, 1 mM DTT, 0.25 mM ZnCl2, 0.006% NP-40,) 10 ul probe mix (5 ng poly[d-(IC)],
583 5 mM MgCl2, 700 fmol probe), and 2.5 ul protein dilution (0.5uM, 1uM, 2.5uM) at room
584 temperature for 60 min. Reactions were loaded onto 6% DNA retardation gels (ThermoFisher)
585 and run in 0.5X Tris–borate–EDTA buffer for 2 hours. Gels were post stained with Gelred
586 Nucleic Acid Stain (Thermo Scientific) for 30 min and visualized using the ChemiDoc MP
587 imaging system (BioRad).
588
589 **Fly stocks and crosses**
590 To deplete maternally deposited *clamp* or *zld* mRNA throughout oogenesis, we crossed
591 a maternal triple driver (MTD-GAL4, Bloomington, #31777) line (Ni et al., 2011) with a

592 Transgenic RNAi Project (TRiP) *clamp* RNAi line (Bloomington, #57008), a TRiP *zld* RNAi line
593 (from C. Rushlow lab) or *egfp* RNAi line (Bloomington, #41552). *egfp* RNAi line was used as
594 control in smFISH, Immunostaining and Imaging experiments. The MTD-GAL4 line alone was
595 used as the control line in ATAC-seq and ChIP-seq experiments.
596 Briefly, the MTD-GAL4 virgin females (5-7day age) were mated with TRiP UAS-RNAi
597 males to obtain MTD-Gal4/UAS-RNAi line daughters. The MTD drives RNAi during oogenesis in
598 these daughters: therefore, the targeted mRNA will be depleted in their eggs. Then MTD-
599 Gal4/UAS-RNAi daughters were mated with males to produce embryos with depleted maternal
600 *clamp* or *zld* mRNA targeted for ATAC-seq and ChIP-seq experiments. The embryonic
601 phenotypes of the maternal *zld*$^{-}$ TRiP RNAi line were confirmed previously (Sun et al., 2015).
602 Maternal *clamp*$^{-}$ embryonic phenotypes of the TRiP *clamp* RNAi line were confirmed by
603 immunofluorescent staining in our study (**Figure 1A-B**). Moreover, we validated CLAMP or ZLD
604 protein knockdown in early embryos by western blotting using the Western Breeze kit
605 (Invitrogen) and measured *clamp* and *zld* mRNA levels by qRT-PCR (**Figure 1 – figure**
606 **supplement 1B, C**).
607
608 **Embryo collections**
609 To optimize egg collections, young (5-7 day old) females and males were mated. To
610 ensure mothers do not lay older embryos during collections, we first starved flies for 2 hours in
611 the empty cages and discarded the first 2-hour grape agar plates with yeast paste (Plate set
612 #0). When we collected eggs for the experiments, we put flies in the cages with grape agar
613 plates (Plate set #1) with yeast paste for egg laying for 2 hours. Then, we replaced Plate set #1
614 with a new set of plates (Plate set #2) at the 2hr time point. We kept Plate set #1 embryos
615 (without any adult flies) to further develop for another 2 hours to obtain 2-4hr embryos. At the
616 same time, we obtained newly laid 0-2hr embryos from Plate set #2. Therefore, this strategy
617 successfully prevented cross contamination between 0-2hr embryos (Plate set #2) and the 2-4hr
618 embryos (Plate set #1).
619

620 **smFISH, Immunostaining and Imaging**

621 For whole embryo single molecule fluorescence *in situ* hybridization (smFISH) and
622 immunostaining and subsequent imaging, standard protocols were used (Little and Gregor,
623 2018). smFISH probes complementary to *run* were a gift from Thomas Gregor, and those
624 complementary to *eve* were a gift from Shawn Little. The concentrations of the different dyes
625 and antibodies were as follows: Hoechst (Invitrogen, 3µg/ml), Anti-NRT (Developmental Studies
626 Hybridoma Bank BP106, 1:10), AlexaFluor secondary antibodies (Invitrogen Molecular Probes,
627 1:1000). Imaging was done using either a Nikon A1 point-scanning confocal microscope with a
628 40X oil objective or the Leica SP8 confocal microscope system with a 40X oil objective (1.3N.A.)
629 Image processing and intensity measurements were done using Image J software (NIH).
630 Figures were assembled using Adobe Photoshop CS4.
631

632 **ATAC-seq in embryos**

633 We conducted ATAC-seq following the protocol from Blythe and Wieschaus, (2016). 0-
634 2hr or 2-4hr embryos were laid on grape agar plates, dechorionated by 1 min exposure to 6%
635 bleach (Clorox) and then washed 3 times in deionized water. We homogenized 10 embryos and
636 lysed them in 50 ul lysis buffer (10mM Tris 7.5, 10mM NaCl, 3mM MgCl2, 0.1% NP-40). We
637 collected nuclei by centrifuging at 500 g at 4°C and resuspended nuclei in 5 ul TD buffer with
638 2.5 ul Tn5 enzyme (Illumina Tagment DNA TDE1 Enzyme and Buffer Kits). We incubated
639 samples at 37°C for 30 min at 800 rpm (Eppendorf Thermomixer) for fragmentation, and then
640 purified samples with Qiagen Minelute columns before PCR amplification. We amplified libraries
641 by adding 10 ul DNA to 25 ul NEBNext HiFi 2x PCR mix (New England Biolabs) and 2.5 ul of a
642 25 uM solution of each of the Ad1 and Ad2 primers. We used 13 PCR cycles to amplify samples
643 from 0-2hr embryos and 12 PCR cycles to amplify samples from 2-4hr embryos. Next, we
644 purified libraries with 1.2x Ampure SPRI beads. We performed three biological replicates for
645 each genotype (n=2) and time point (n=2). We measured the concentrations of 12 ATAC-seq
646 libraries by Qubit and determined library quality by Bioanalyzer. We sequenced libraries on an

647 Illumina Hi-seq 4000 sequencer at GeneWiz (South Plainfield, NJ) using the 2x150-bp mode.
648 ATAC-seq data is deposited at NCBI GEO and the accession number is GSE152596.
649
650 **Chromatin Immunoprecipitation-sequencing (ChIP-seq) in embryos**
651 We performed ChIP-seq as previously described Blythe and Wieschaus, (2015). We
652 collected and fixed ~100 embryos from each MTD-GAL4 and RNAi cross 0-2hr or 2-4hr after
653 egg lay. We used 3 ul of rabbit anti-CLAMP (Soruco et al., 2013) and 2 ul rat anti-ZLD (from C.
654 Rushlow lab) per sample. We performed three biological ChIP replicates for each protein (n=2),
655 genotype (n=3) and time point (n=2). In total, we prepared 36 libraries using the NEBNext Ultra
656 ChIP-seq kit (New England Biolabs) and sequenced libraries on the Illumina HiSeq 2500
657 sequencer using the 2x150-bp mode. ChIP-seq data is deposited at NCBI GEO and the
658 accession number is GSE152598.
659
660 **Computational analyses**
661 ***ATAC-seq analysis***
662 Detailed ATAC-seq analysis pipeline code is provided as **Figure 2- supplementary file**
663 **1**. Prior to sequencing, the Fragment Analyzer showed the library top peaks were in the 180-
664 190bp range, which is comparable to the previously established embryo ATAC-seq protocol
665 (Haines, 2017). Demultiplexed reads were trimmed of adapters using TrimGalore (Krueger,
666 2017) and mapped to the *Drosophila* genome dm6 version using Bowtie2 (v. 2.3.0) with option -
667 X 2000. We used Picard tools (v. 2.9.2) and SAMtools (v.1.9, Li et al., 2009) to remove the
668 reads that were unmapped, failed primary alignment, or duplicated (-F 1804), and retain
669 properly paired reads (-f 2) with MAPQ >30. After quality trimming and mapping, the Picard tool
670 reported the mean fragment sizes for all ATAC-seq mapped reads is between 125-161bp. As
671 expected, we observed three classes of peaks: 1) A sharp peak at <100 bp (open chromatin); 2)
672 A peak at ~200bp (mono-nucleosome); 3) Other larger peaks (multi-nucleosomes).
673

674 ATAC-seq mapping stats are summarized in **Table 1 – Source Data 1.** After mapping,
675 we used Samtools to select a fragment size <= 100bp within the bam file to focus on open
676 chromatin. Peak regions for open chromatin regions were called using MACS2 (version 2.1.1,
677 Zhang et al., 2008). The peak numbers are summarized in **Table 1**. ENCODE blacklist was
678 used to filter out problematic regions in dm6 (Amemiya et al., 2019). DiffBind with the DESeq2
679 method (v. 3.10, Stark and Brown, 2019) was used to identify differentially accessible (DA)
680 regions **(Figure 2A)**. The DA and non-DA number were summarized in **Table 1.** We report
681 differentially accessible peak p-values from DiffBind in **Table 1 – Source Data 2.** We used
682 DeepTools (version 3.1.0, Ramírez et al., 2014) to generate enrichment heatmaps (CPM
683 normalization), and average profiles were generate in DeepStats (Gautier RICHARD, 2020).
684 Bigwig files generated by Deeptools bamCoverage (CPM normalization) is uploaded to IGV
685 (Robinson et al., 2011) for genomic track visualizations **(Figure 2B)**. Then, we used Homer (v
686 4.11, Givler and Lilienthal, 2005) for *de novo* motif searches **(Figure 2D)**. Visualizations and
687 statistical tests were conducted in R (R Core Team, 2014). Specifically, we annotated peaks to
688 their genomic regions using R packages Chipseeker **(Figure 2C,** Yu et al., 2015) and we
689 performed gene ontology enrichment analysis using clusterProfiler (Yu et al., 2012). Boxplot and
690 violin plots were generated using ggplot2 (Wickham, 2009) package. Intervene (Khan and
691 Mathelier, 2017) was used for intersection and visualization of multiple peak region sets **(Figure**
692 **2E)**.
693
694 ***ChIP-seq analysis***
695 Detailed ChIP-seq analysis pipeline code is provided as **Figure 4- supplementary file**
696 **1**. Briefly, we trimmed ChIP sequencing raw reads with TrimGalore (Krueger, 2017) with a
697 minimal phred score of 20, 36 bp minimal read length, and Illumina adaptor removal. We then
698 mapped cleaned reads to the *D. melanogaster* genome (UCSC dm6) with Bowtie2 (v. 2.3.0)
699 with the –very-sensitive-local flag feature. We used Picard tools (v. 2.9.2) and SAMtools (v.1.9,
700 Li et al., 2009) to remove the PCR duplicates. ChIP-seq mapping stats are summarized in **Table**
701 **4 – Source Data 1.** We used MACS2 (version 2.1.1, Zhang et al., 2008) to identify peaks with

702 default parameters and MSPC (v.4.0.0, Jalili et al., 2015) to obtain consensus peaks from 3
703 replicates. The peak number for each sample was summarized in Table 4. ENCODE blacklist
704 was used to filter out problematic regions in dm6 (Amemiya et al., 2019). We identified
705 differential binding (DB) and non-differential binding (non-DB) between MTD and RNAi samples
706 using DiffBind (v. 3.10, Stark and Brown, 2019) with the DESeq2 method **(Figure 4D-E).** The
707 DB and non-DB peak numbers are summarized in **Table 4**. Differential binding p-value from
708 DiffBind for all peaks were reported in **Table 4 – Source Data 2**. We used DeepTools (version
709 3.1.0, Ramírez et al., 2014) to generate enrichment heatmaps **(Figure 4A-B,** SES
710 normalization**)** and average profiles **(Figure 4H)**. Bigwig files were generated with Deeptools
711 bamCompare (SES normalization) and uploaded to IGV (Robinson et al., 2011) for genomic
712 track visualization **(Figure 4G).** We used Homer (v 4.11, Givler and Lilienthal, 2005) for *de novo*
713 motif searches and genomic annotation **(Figure 4F)**. Intervene (Khan and Mathelier, 2017) was
714 used for intersection and visualization of multiple peak region sets. Visualizations and statistical
715 tests were conducted in R (R Core Team, 2014). Specifically, we annotated peaks to their
716 genomic regions using the R package Chipseeker **(**Yu et al., 2015) and we did gene ontology
717 enrichment analysis using clusterProfiler (Yu et al., 2012). Boxplots and violin plots were
718 generated using the ggplot2 (Wickham, 2009) package **(Figure 4I and Figure 5)**.
719

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

721 We first selected the top 6,000 ATAC-seq peaks based on the count per million value
722 rank from our ATAC-seq data or the ATAC-seq in Hannon et al. (2017). Then, we used Bedtools
723 (Quinlan and Hall, 2010) intersection tool to intersect peaks in CLAMP ChIP-seq binding regions
724 with CLAMP DA or non-DA peaks. Based on the intersection of the peaks, we defined 4 types of
725 CLAMP related peaks: 1) DA with CLAMP, 2) DA without CLAMP, 3) Non-DA with CLAMP, 4)
726 Non-DA without CLAMP **(Figure 3A, Figure 3-figure supplement 1A,** and **Table 2).**
727 Similarly, we defined ZLD related peaks by intersecting ZLD DA or non-DA peaks and
728 ATAC-seq datasets (Hannon et al., 2017; Soluri et al., 2020) from wildtype (wt) and *zld* germline
729 clone (*zld-*) embryos at the NC14 +12 min stage. Specifically, we defined four classes of

genomic loci for ZLD-related classes: 1) DA with ZLD, 2) DA without ZLD, 3) Non-DA with ZLD, 4) Non-DA, without ZLD **(Figure 3B** and **Table 2).** We used DeepTools (version 3.1.0, Ramírez et al., 2014) to generate enrichment heatmaps (CPM normalization) and average profiles for each subclass of peaks **(Figure 3A-B).** We used Homer (v 4.11, Givler and Lilienthal, 2005) for *de novo* motif searches. Peaks locations in each CLAMP or ZLD-related category were summarized in **Table2 - Source Data 1**.

ATAC-seq and RNA-seq data integration

We annotated genes near differential (down-DA) ATAC-seq peaks in R using detailRanges function from the csaw package (Lun and Smyth, 2016). Then we used Intervene (Khan and Mathelier, 2017) to overlap the genes near differentially accessible (DA) peaks with differentially down-regulated genes identified by RNA-seq (Rieder et al., 2017).

ChIP-seq and RNA-seq data integration

To define strong, weak, and unbound genes close to peaks in CLAMP or ZLD ChIP-seq data, we used the peak binding score reported in MACS2 -log10(p-value) of 100 as a cut off value. We defined the following categories: 1) Strong binding peaks: score greater than 100; 2) Weak binding peak: score lesser than 100; 3) Unbound peak: the rest of the peaks that are neither strong or weak. Then, we annotated all peaks using Homer annotatePeaks (v 4.11, Givler and Lilienthal, 2005). We then obtained the log2 fold change (MTD/clamp-i or wt/zld-) of gene expression in the RNA-seq dataset for each protein binding group: CLAMP (Rieder et al., 2017) or ZLD (Combs and Eisen, 2017). Boxplots and violin plots were generated using the ggplot2 (Wickham, 2009) package **(Figure 4I** and **Figure 5)**.

Datasets

RNA-seq datasets from wild type and maternal *clamp* depletion by RNAi were from GSE102922 (Rieder et al., 2017). RNA-seq datasets from wild type and *zld* germline clones were from GSE71137 (Combs and Eisen, 2017). ATAC-seq data from wild type and *zld*

germline clones were from GSE86966 (Hannon et al., 2017). Processed ATAC-seq data identifying differential peaks between wild type and *zld* germline mutations were from Soluri et al., (2020).

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

767 **ACKNOWLEDGEMENTS**

768

769 This work was supported by NIH grant F32GM109663 and K99HD092625 to Dr. Leila Rieder
770 and R35GM126994 to Dr. Erica Larschan, and in part by NSF grant 1845734 and NIH grant
771 R01GM118530 to Dr. Nicolas L. Fawzi.

772

773 **AUTHOR CONTRIBUTIONS**

774

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

781

782 **DECLARATION OF INTERESTS**

783

784 The authors declare no competing interests.

REFERENCES

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, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**:279–289. doi:10.1016/s1097-2765(02)00459-8

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

Combs PA, Eisen MB. 2017. Genome-wide measurement of spatial expression in patterning mutants of Drosophila melanogaster. *F1000Research* **6**:41. doi:10.12688/f1000research.9720.1

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.

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

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/.

Lai X, Verhage L, Hugouvieux V, Zubieta C. 2018. Pioneer Factors in Animals and Plants—Colonizing Chromatin for Gene Regulation. *Molecules* **23**. doi:10.3390/molecules23081914

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

Leibovitch BA, Lu Q, Benjamin LR, Liu Y, Gilmour DS, Elgin SCR. 2002. GAGA Factor and the TFIID Complex Collaborate in Generating an Open Chromatin Structure at the Drosophila melanogaster hsp26 Promoter. *Molecular and Cellular Biology* **22**:6148–6157. doi:10.1128/MCB.22.17.6148-6157.2002

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 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

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. *Nature Biotechnology* **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

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* defected 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 defected embryos is on the right, *p-values* were calculated with the *Fisher's* Exact test; number of embryos is in parentheses.

C. 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*.

D. 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.

E. 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.

Figure 2. CLAMP regulates chromatin accessibility throughout early development

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 ($p < 0.05$) differential peaks after maternal *clamp* RNAi, identified by DiffBind (DESeq2) (DA peaks). The number of peaks in each class is noted on the plot. Average ATAC-seq signal (right) in counts per million (CPM) reads 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 encoding Mod(mdg4) 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 4 motifs enriched in regions that require CLAMP for chromatin accessibility. Enrichment *p*-value and percentage of sequences are noted.

E. Venn diagram comparing CLAMP-mediated changes in down-regulated differential gene expression (DEG, mRNA-seq) and down-regulated differential accessibility (DA, ATAC-seq). *p*-value represents the significance (hypergeometric test, N= 15,682 total fly genes) of the overlap.

Figure 3. CLAMP-mediated chromatin accessibility is correlated with CLAMP and ZLD binding

A. Four classes of CLAMP-related peaks defined by combining ATAC-seq and CLAMP ChIP-seq peaks in 0-2hr MTD and maternal *clamp* RNAi embryos. Data is displayed as a heatmap of z-score normalized ATAC-seq (CPM) 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. Top three motifs called from each peak type by Homer are listed on the right.
Four types of peaks and number of peaks are marked on plot:
DA w/ CLAMP: differentially accessible region depends on CLAMP and has CLAMP binding.
DA w/o CLAMP: differentially accessible region depends on CLAMP and has no CLAMP binding.
nonDA w/ CLAMP: non-differentially accessible region and has CLAMP binding.
nonDA w/o CLAMP: non-differentially accessible region and has no CLAMP binding.

B. Four classes of ZLD-related peaks defined by integrating ATAC-seq (Hannon et al., 2017; Soluri et al., 2020) in nuclear cycle 14 (NC14)+12min wildtype (wt) and maternal *zld* RNAi embryos and ZLD ChIP-seq peaks in 0-2hr MTD and maternal *zld* RNAi embryos. Data is displayed as a heatmap of *z*-score normalized ATAC-seq (CPM) 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 wildtype or MTD embryos. Top three motifs called from each peak type by Homer are listed on the right.
Four types of peaks and number of peaks are marked on plot:
DA w/ ZLD: differentially accessible region depends on ZLD and has ZLD binding.
DA w/o ZLD: differentially accessible region depends on ZLD and has no ZLD binding.
nonDA w/ ZLD: non-differentially accessible region and has ZLD binding.
nonDA w/o ZLD: non-differentially accessible region and has no ZLD binding.

Figure 4. 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 CLAMP and ZLD peaks in the *Drosophila* genome (dm6) in 0-2hr and 2-4hr embryos (MTD, *clamp*-i and *zld*-i).

D. Differential bing (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.

E. Differential bing (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.

F. 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.

G. 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.

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.

I. 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.

Figure 5. CLAMP and ZLD function together in zygotic genome activation by mediating chromatin accessibility

A. 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.

B. Gene expression changes caused by maternal *clamp* RNAi 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.

C. Gene expression changes caused by maternal *zld* RNAi (Combs and Eisen, 2017) 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.

D. Left: Gene expression changes caused by germline clone *zld*- (Combs and Eisen, 2017) 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.

   Right: 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.

E. Average profiles of ATAC-seq signal in CPM show chromatin accessibility at CLAMP down-DB (green line) and non-DB (grey line) sites in NC14+12min wt and *zld*- embryos (Hannon et al., 2017). Number of peaks are noted on the plot.

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

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

CLAMP and ZLD are two pioneer TFs which function cooperatively at promoters to open chromatin and regulate gene expression of genes encoding other key TFs. We defined CLAMP-related and ZLD-related peaks in early embryos, which revealed cooperative roles for CLAMP and ZLD in defining chromatin accessibility and zygotic transcription during ZGA:

DA, w/ CLAMP sites: CLAMP promotes ZLD enrichment at these sites where CLAMP binding increases chromatin accessibility and further regulates ZLD target gene expression. These sites are closed when maternal *clamp* is depleted.

Non-DA, w/ CLAMP sites: ZLD modulates chromatin opening at these sites that are bound by CLAMP but do not depend on CLAMP for chromatin accessibility. These sites remain open when maternal *clamp* is depleted.

DA, w/ ZLD sites: ZLD facilitates CLAMP binding to sites where chromatin accessibility is increased by ZLD and promotes CLAMP target gene expression. These sites are closed upon maternal *zld* depletion.

Non-DA, w/ ZLD sites: CLAMP opens chromatin at these locations which are bound by ZLD but do not depend on ZLD for their accessibility. These sites remain open upon maternal *zld* depletion.

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 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.

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

A-B. Pearson correlation heatmap of peaks dependent on CLAMP in MTD vs. *clamp-i* embryos at 0-2hr and 2-4hr time points.

C-D. GO terms for genes that require CLAMP for chromatin accessibility at 0-2hr and 2-4hr time points.

Figure2-supplementary file 1.html. ATAC-seq data analysis pipeline

Figure 3-figure supplement 1. CLAMP-mediated chromatin accessibility is correlated with CLAMP and ZLD binding

A. Four classes of CLAMP-related peaks defined by combining ATAC-seq and CLAMP ChIP-seq peaks in 2-4hr MTD and maternal *clamp* RNAi embryos. Data is displayed as a heatmap of z-score normalized ATAC-seq (CPM) 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. Top 3 motif called from each peak type by Homer was listed on the right. Four types of peaks and number of peaks are marked on plot:
DA w/ CLAMP: differentially accessible region depends on CLAMP and has CLAMP binding.
DA w/o CLAMP: differentially accessible region depends on CLAMP and has no CLAMP binding.
Non-DA w/ CLAMP: not differentially accessible region and has CLAMP binding.
Non-DA w/o CLAMP: not differentially accessible region and has no CLAMP binding.

B. 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.

C. 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.

Figure 4-figure supplement 1. CLAMP and ZLD depend on each other for chromatin binding

A. 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.

B. 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.

C. 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.

D. 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.

E. Average profiles of ChIP-seq signal in log2 (IP/input) show the size of DB or non-DB peaks of ZLD in MTD versus *clamp*-i embryos.

Figure4-supplementary file 1.html. ChIP-seq data analysis pipeline

Figure 5 figure supplement 1. CLAMP and ZLD functioning together in zygotic genome activation via mediating chromatin accessibility

A. Density plot of gene expression shifts caused by maternal *clamp* RNAi at genes with strong, weak and no CLAMP binding as measured by ChIP-seq in 0-2hr (left) or 2-4hr (right) embryos.

B. Density plot of gene expression shifts caused by maternal *zld* RNAi (Combs and Eisen, 2017) at genes with strong, weak and no ZLD binding as measured by ChIP-seq in 0-2hr (left) or 2-4hr (right) embryos.

C. 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).

D. 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).

E. 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.

Supplementary table captions

Table 1-Source Data1. Mapping statistics summary for ATAC-seq samples.

Table 1-Source Data2. 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.

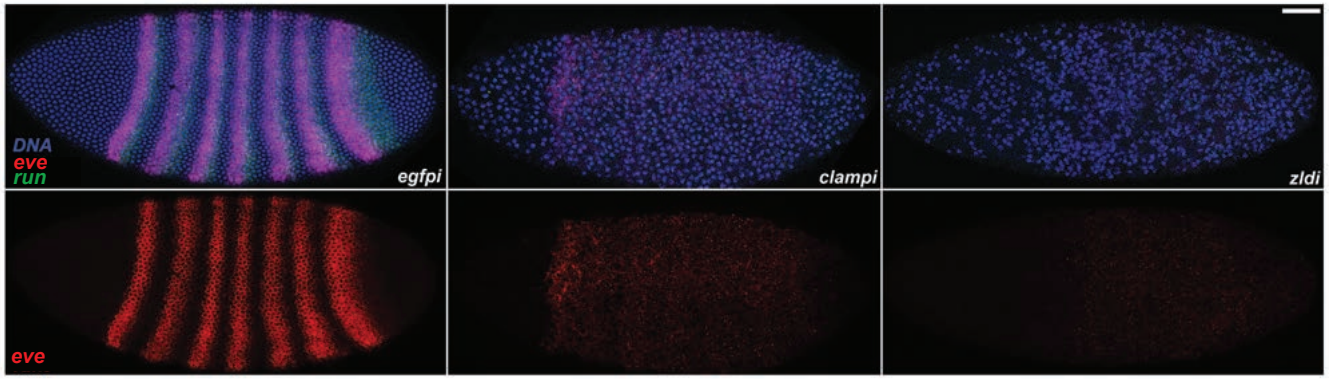
Table 2-Source Data1. Peaks locations in each CLAMP or ZLD-related category.
Page 1 DA with CLAMP 0-2hr;
Page 2 DA without CLAMP 0-2hr;
Page 3 Non-DA with CLAMP 0-2hr;
Page 4 Non-DA without CLAMP 0-2hr;
Page 5 DA with CLAMP 2-4hr;
Page 6 DA without CLAMP 2-4hr;
Page 7 Non-DA with CLAMP 2-4hr;
Page 8 Non-DA without CLAMP 2-4hr;
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.

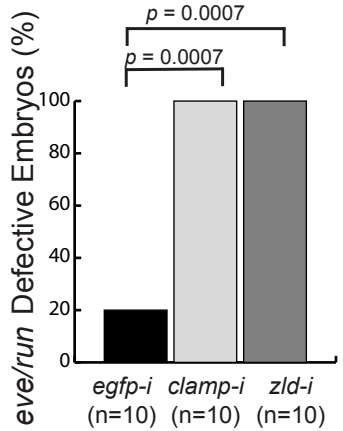
Table 4-Source Data1. Mapping statistics summary for ChIP-seq samples.

Table 4-Source Data2. ChIP-seq read counts in peak regions in replicates of MTD and RNAi samples (DiffBind analysis).
Page 1. CLAMP ChIP-seq in *zld*-i vs. MTD in 0-2hr embryos;
Page 2. CLAMP ChIP-seq in *zld-i* vs. MTD in 2-4hr embryos;
Page 3. ZLD ChIP-seq in *clamp-i* vs. MTD in 0-2hr embryos;
Page 4. ZLD ChIP-seq in *clamp-i* vs. MTD in 2-4hr embryos.

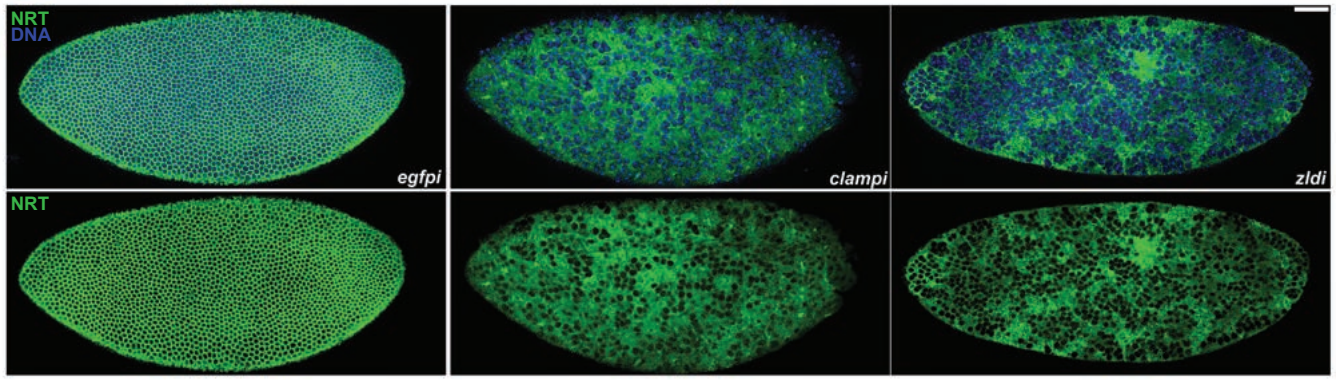
Fig.1

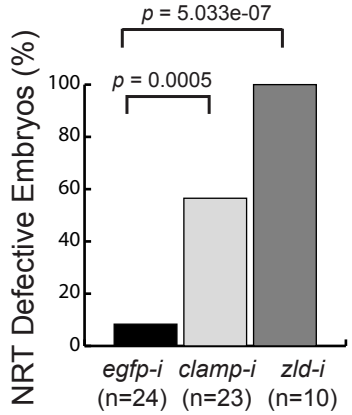
A


DNA
eve
run
egfpi
clampi
zldi
eve

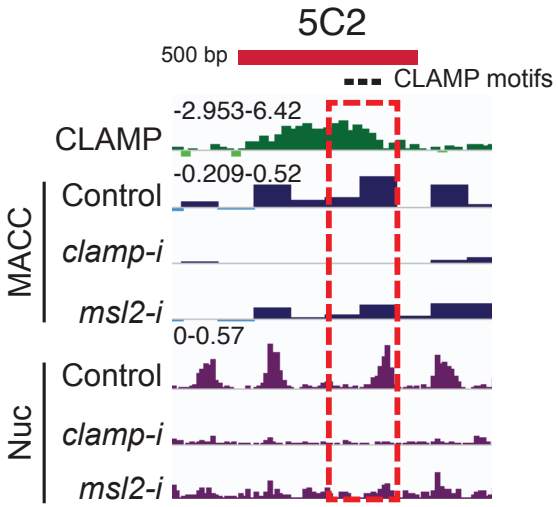

p = 0.0007
p = 0.0007
eve/run Defective Embryos (%)
100
80
60
40
20
0
egfp-i (n=10)
clamp-i (n=10)
zld-i (n=10)

B

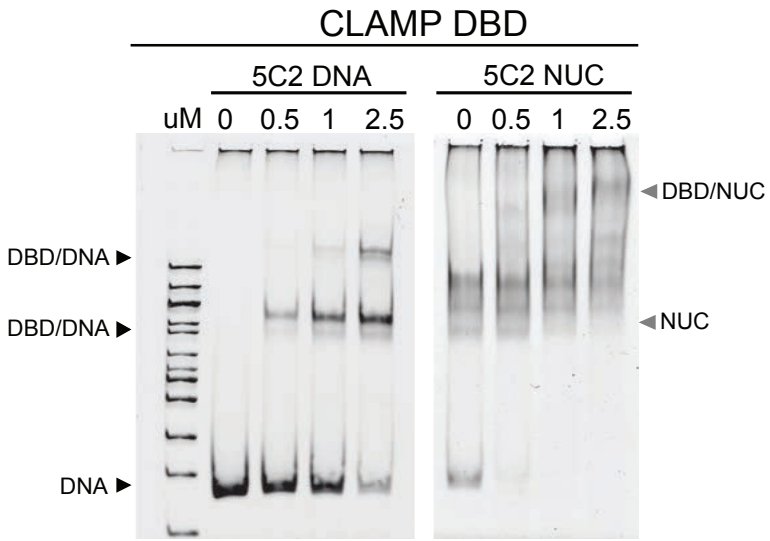

NRT
DNA
egfpi
clampi
zldi
NRT


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)

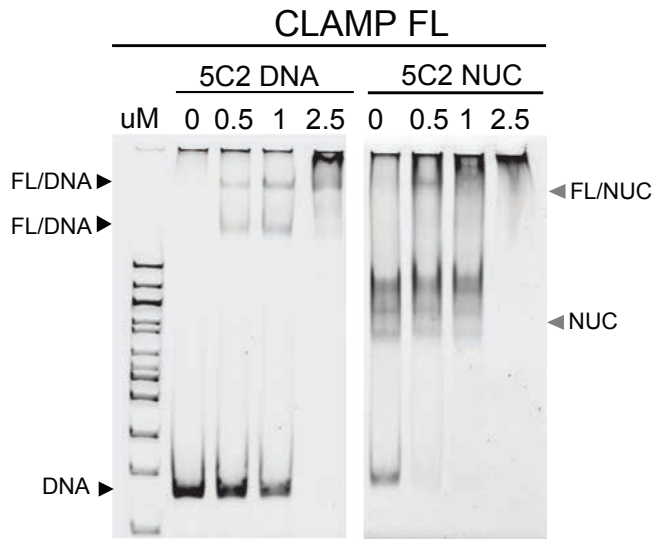
C

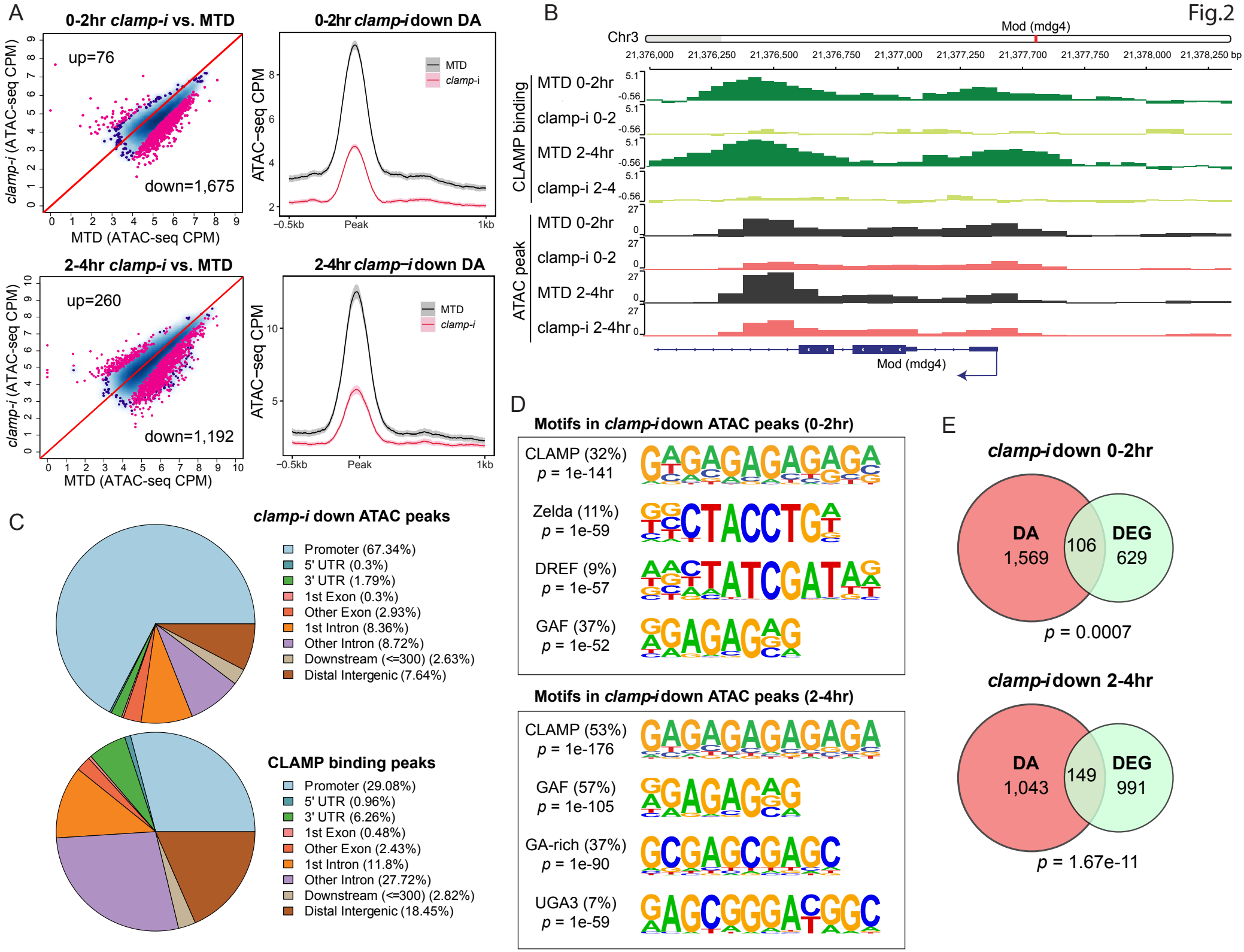

5C2
500 bp
CLAMP motifs
-2.953-6.42
CLAMP
-0.209-0.52
MACC
Control
clamp-i
msl2-i
0-0.57
Nuc
Control
clamp-i
msl2-i

D


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

E


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



A
0-2hr clamp-i vs. MTD
up=76
down=1,675
clamp-i (ATAC-seq CPM)
MTD (ATAC-seq CPM)
0-2hr clamp-i down DA
MTD
clamp-i
ATAC−seq CPM
−0.5kb
Peak
1kb
2-4hr clamp-i vs. MTD
up=260
down=1,192
clamp-i (ATAC-seq CPM)
MTD (ATAC-seq CPM)
2-4hr clamp−i down DA
MTD
clamp-i
ATAC−seq CPM
−0.5kb
Peak
1kb
B
Fig.2
Chr3
Mod (mdg4)
21,376,000
21,376,250
21,376,500
21,376,750
21,377,000
21,377,250
21,377,700
21,377,750
21,378,000
21,378,250 bp
CLAMP binding
MTD 0-2hr
clamp-i 0-2
MTD 2-4hr
clamp-i 2-4
ATAC peak
MTD 0-2hr
clamp-i 0-2
MTD 2-4hr
clamp-i 2-4hr
Mod (mdg4)
C
clamp-i down ATAC peaks
Promoter (67.34%)
5' UTR (0.3%)
3' UTR (1.79%)
1st Exon (0.3%)
Other Exon (2.93%)
1st Intron (8.36%)
Other Intron (8.72%)
Downstream (<=300) (2.63%)
Distal Intergenic (7.64%)
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 (0-2hr)
CLAMP (32%)
p = 1e-141
Zelda (11%)
p = 1e-59
DREF (9%)
p = 1e-57
GAF (37%)
p = 1e-52
Motifs in clamp-i down ATAC peaks (2-4hr)
CLAMP (53%)
p = 1e-176
GAF (57%)
p = 1e-105
GA-rich (37%)
p = 1e-90
UGA3 (7%)
p = 1e-59
E
clamp-i down 0-2hr
DA
1,569
106
DEG
629
p = 0.0007
clamp-i down 2-4hr
DA
1,043
149
DEG
991
p = 1.67e-11

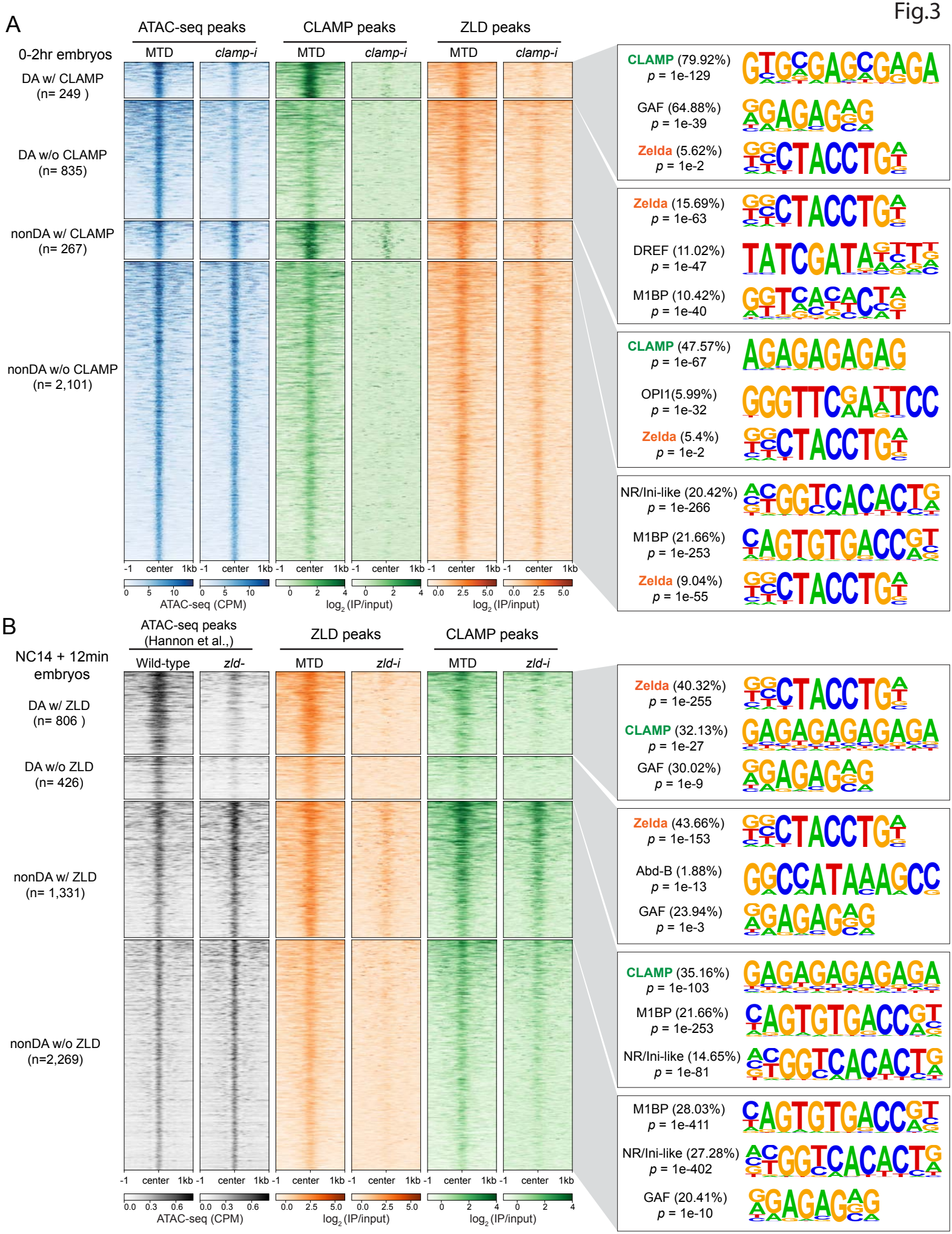


Fig.3
A
ATAC-seq peaks
CLAMP peaks
ZLD peaks
0-2hr embryos
MTD
clamp-i
DA w/ CLAMP (n= 249)
DA w/o CLAMP (n= 835)
nonDA w/ CLAMP (n= 267)
nonDA w/o CLAMP (n= 2,101)
-1 center 1kb
ATAC-seq (CPM)
log2 (IP/input)
CLAMP (79.92%) p = 1e-129
GAF (64.88%) p = 1e-39
Zelda (5.62%) p = 1e-2
Zelda (15.69%) p = 1e-63
DREF (11.02%) p = 1e-47
M1BP (10.42%) p = 1e-40
CLAMP (47.57%) p = 1e-67
OPI1(5.99%) p = 1e-32
Zelda (5.4%) p = 1e-2
NR/Ini-like (20.42%) p = 1e-266
M1BP (21.66%) p = 1e-253
Zelda (9.04%) p = 1e-55
B
ATAC-seq peaks (Hannon et al.,)
ZLD peaks
CLAMP peaks
NC14 + 12min embryos
Wild-type
zld-
MTD
zld-i
DA w/ ZLD (n= 806)
DA w/o ZLD (n= 426)
nonDA w/ ZLD (n= 1,331)
nonDA w/o ZLD (n=2,269)
ATAC-seq (CPM)
log2 (IP/input)
Zelda (40.32%) p = 1e-255
CLAMP (32.13%) p = 1e-27
GAF (30.02%) p = 1e-9
Zelda (43.66%) p = 1e-153
Abd-B (1.88%) p = 1e-13
GAF (23.94%) p = 1e-3
CLAMP (35.16%) p = 1e-103
M1BP (21.66%) p = 1e-253
NR/Ini-like (14.65%) p = 1e-81
M1BP (28.03%) p = 1e-411
NR/Ini-like (27.28%) p = 1e-402
GAF (20.41%) p = 1e-10

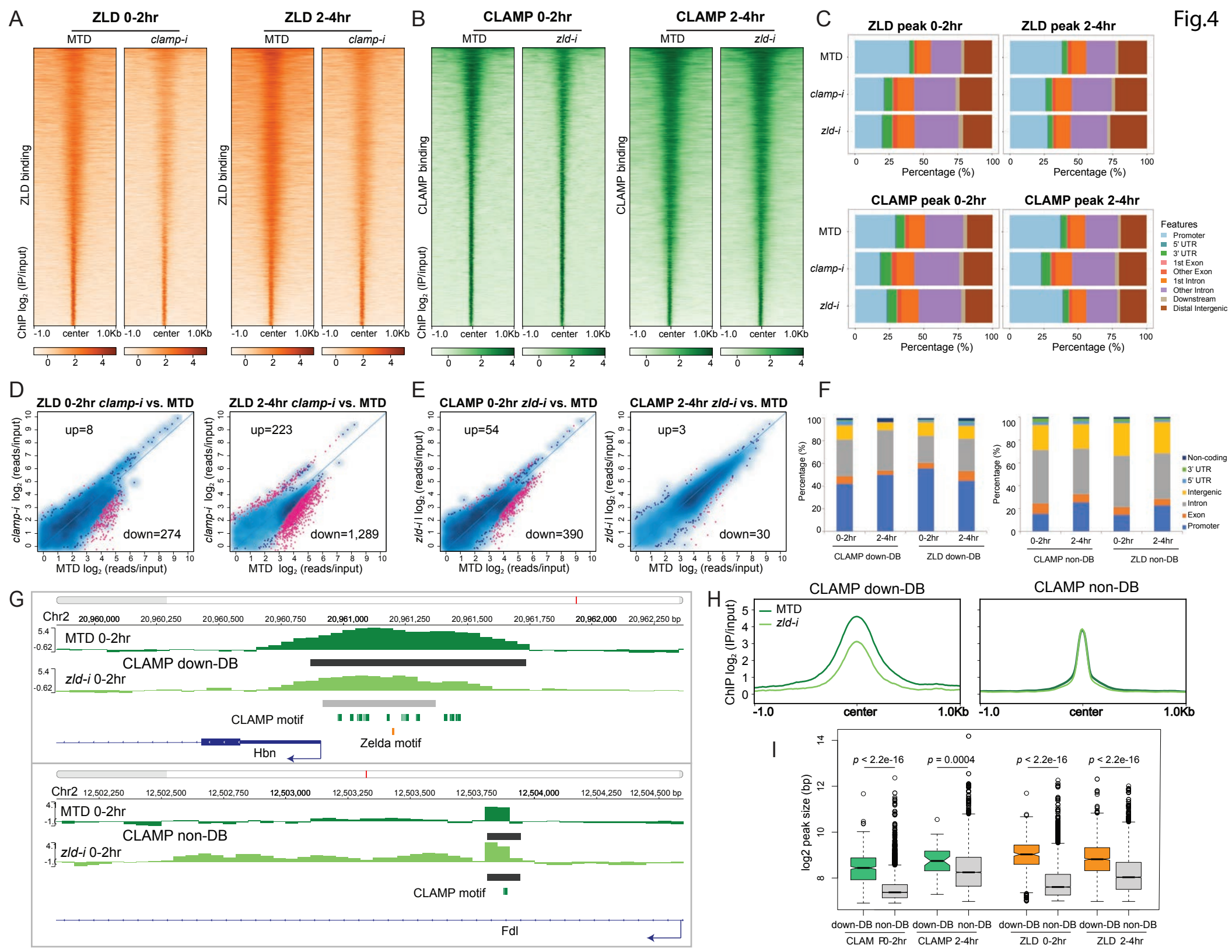


Fig.4
A
ZLD 0-2hr
ZLD 2-4hr
MTD
clamp-i
ZLD binding
ChIP log2 (IP/input)
-1.0 center 1.0Kb
B
CLAMP 0-2hr
CLAMP 2-4hr
zld-i
CLAMP binding
C
ZLD peak 0-2hr
ZLD peak 2-4hr
CLAMP peak 0-2hr
CLAMP peak 2-4hr
Percentage (%)
Features
Promoter
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream
Distal Intergenic
D
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)
E
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)
F
Percentage (%)
0-2hr
2-4hr
CLAMP down-DB
ZLD down-DB
CLAMP non-DB
ZLD non-DB
Non-coding
3' UTR
5' UTR
Intergenic
Intron
Exon
Promoter
G
Chr2
MTD 0-2hr
CLAMP down-DB
zld-i 0-2hr
CLAMP motif
Zelda motif
Hbn
CLAMP non-DB
Fdl
H
CLAMP down-DB
CLAMP non-DB
MTD
zld-i
ChIP log2 (IP/input)
-1.0 center 1.0Kb
I
log2 peak size (bp)
p < 2.2e-16
p = 0.0004
p < 2.2e-16
p < 2.2e-16
down-DB non-DB
CLAMP 0-2hr
CLAMP 2-4hr
ZLD 0-2hr
ZLD 2-4hr

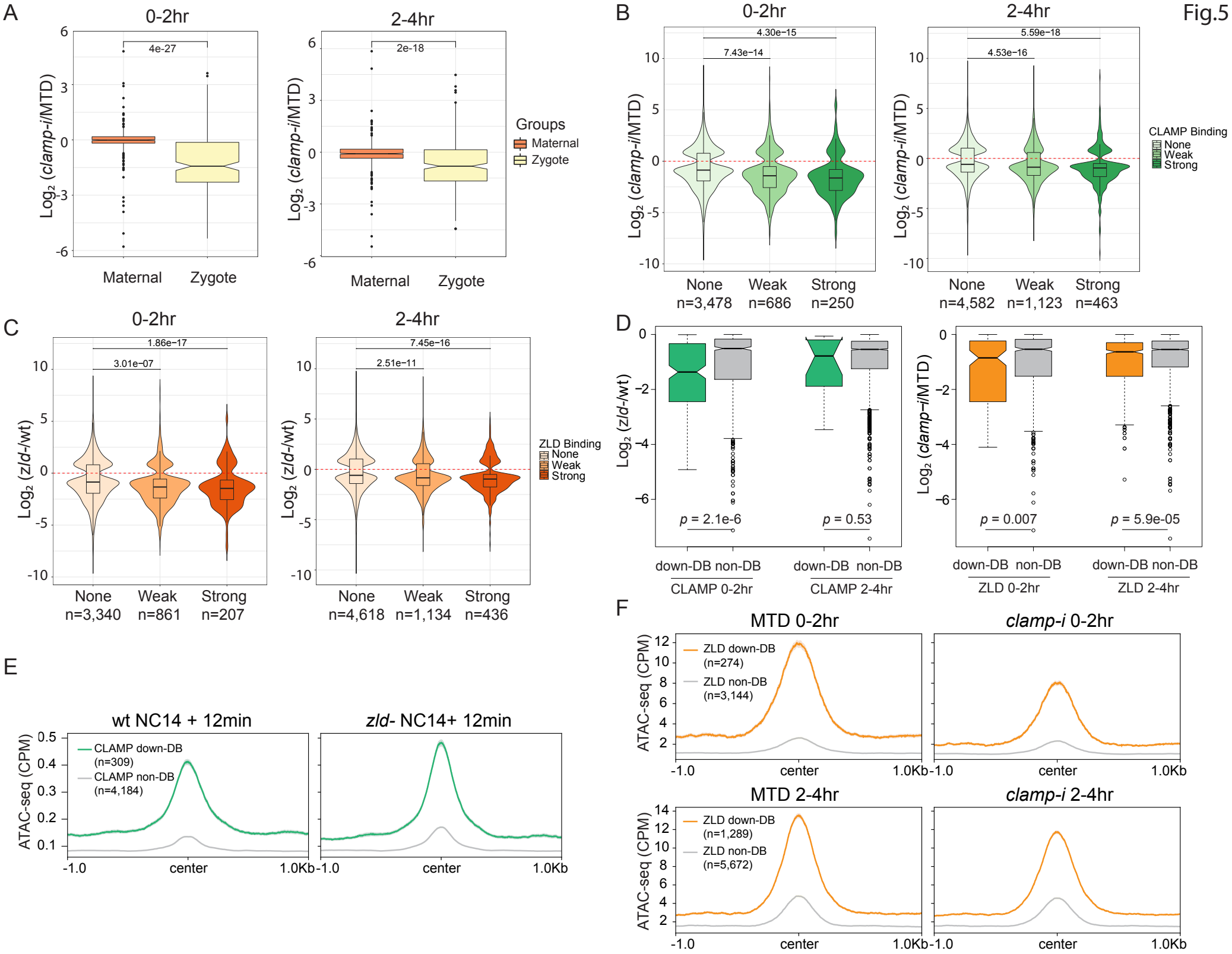


Fig.5
A
0-2hr
2-4hr
4e-27
2e-18
Log2 (clamp-i/MTD)
Maternal
Zygote
Groups
Maternal
Zygote
B
0-2hr
2-4hr
4.30e−15
7.43e−14
5.59e−18
4.53e−16
Log2 (clamp-i/MTD)
None n=3,478
Weak n=686
Strong n=250
None n=4,582
Weak n=1,123
Strong n=463
CLAMP Binding
None
Weak
Strong
C
0-2hr
2-4hr
1.86e−17
3.01e−07
7.45e−16
2.51e−11
Log2 (zld-/wt)
None n=3,340
Weak n=861
Strong n=207
None n=4,618
Weak n=1,134
Strong n=436
ZLD Binding
None
Weak
Strong
D
Log2 (zld-/wt)
p = 2.1e-6
p = 0.53
down-DB non-DB
CLAMP 0-2hr
down-DB non-DB
CLAMP 2-4hr
Log2 (clamp−i/MTD)
p = 0.007
p = 5.9e-05
down-DB non-DB
ZLD 0-2hr
down-DB non-DB
ZLD 2-4hr
E
wt NC14 + 12min
zld- NC14+ 12min
ATAC-seq (CPM)
CLAMP down-DB (n=309)
CLAMP non-DB (n=4,184)
-1.0
center
1.0Kb
F
MTD 0-2hr
clamp-i 0-2hr
ZLD down-DB (n=274)
ZLD non-DB (n=3,144)
MTD 2-4hr
clamp-i 2-4hr
ZLD down-DB (n=1,289)
ZLD non-DB (n=5,672)
ATAC-seq (CPM)
-1.0
center
1.0Kb

Fig.6

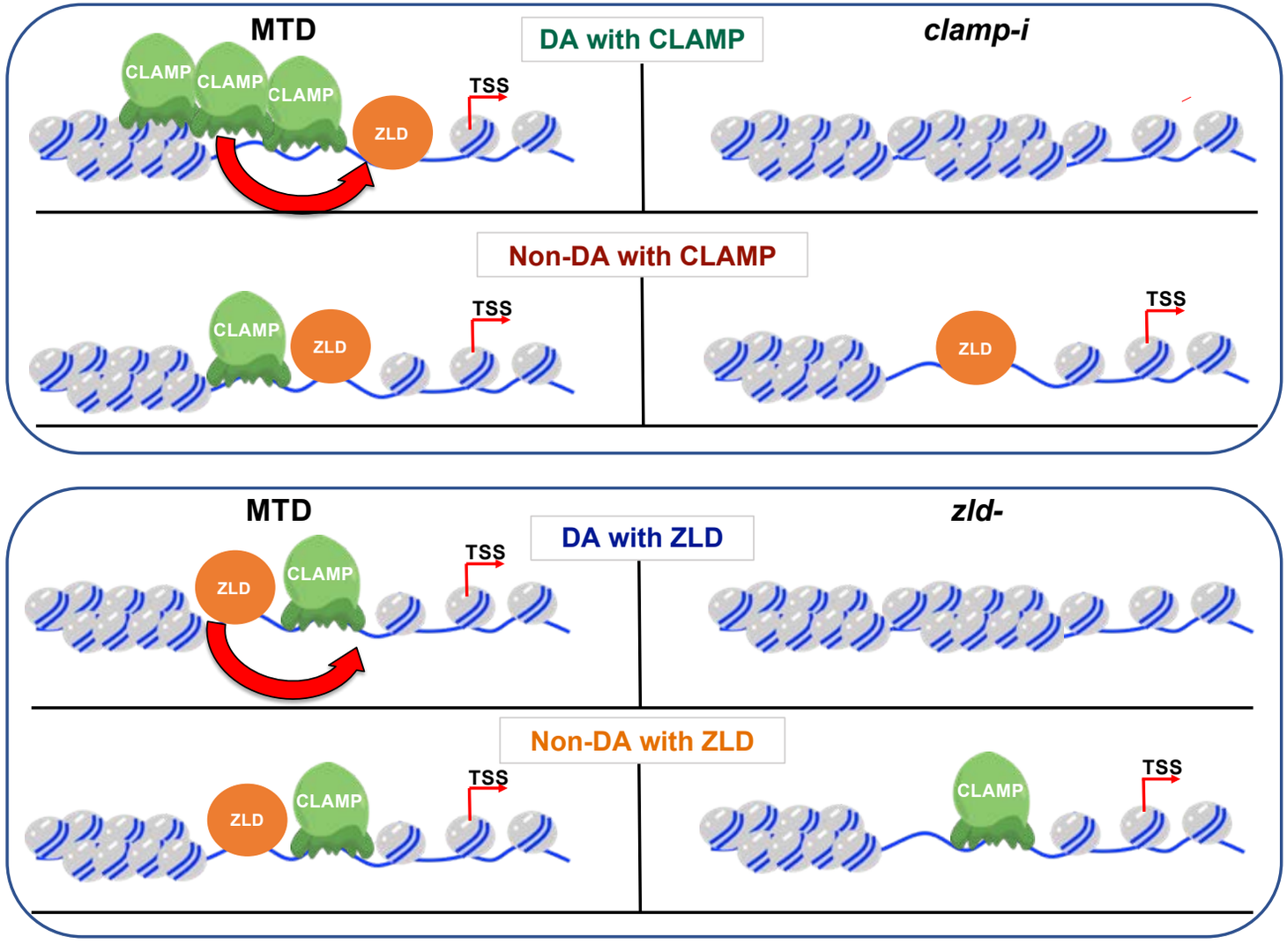

MTD
DA with CLAMP
clamp-i
CLAMP
CLAMP
CLAMP
ZLD
TSS
Non-DA with CLAMP
CLAMP
ZLD
TSS
ZLD
TSS
MTD
DA with ZLD
zld-
ZLD
CLAMP
TSS
Non-DA with ZLD
ZLD
CLAMP
TSS
CLAMP
TSS

Fig.1-figure supplement 1

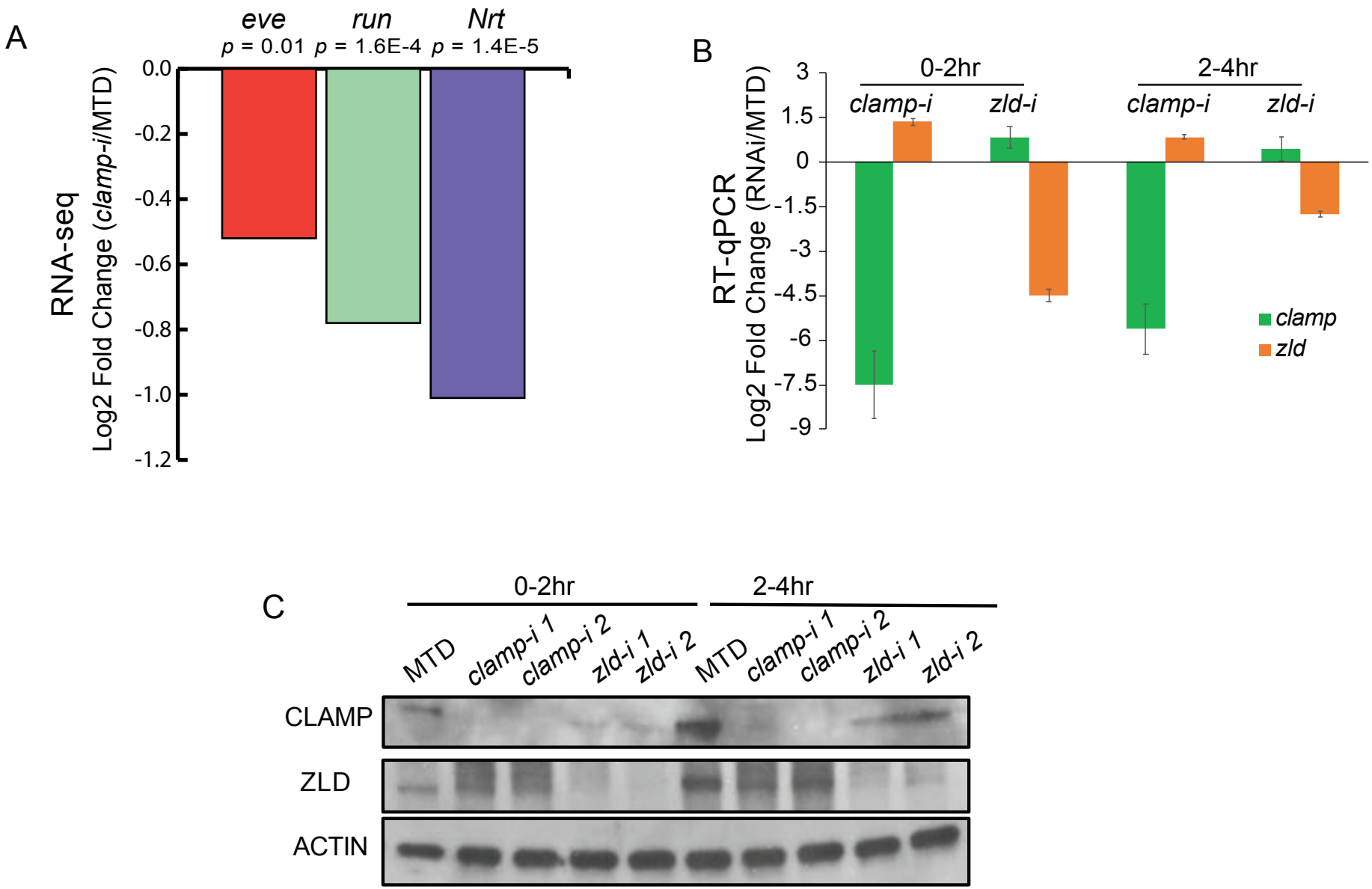


Fig.2- figure supplement 1

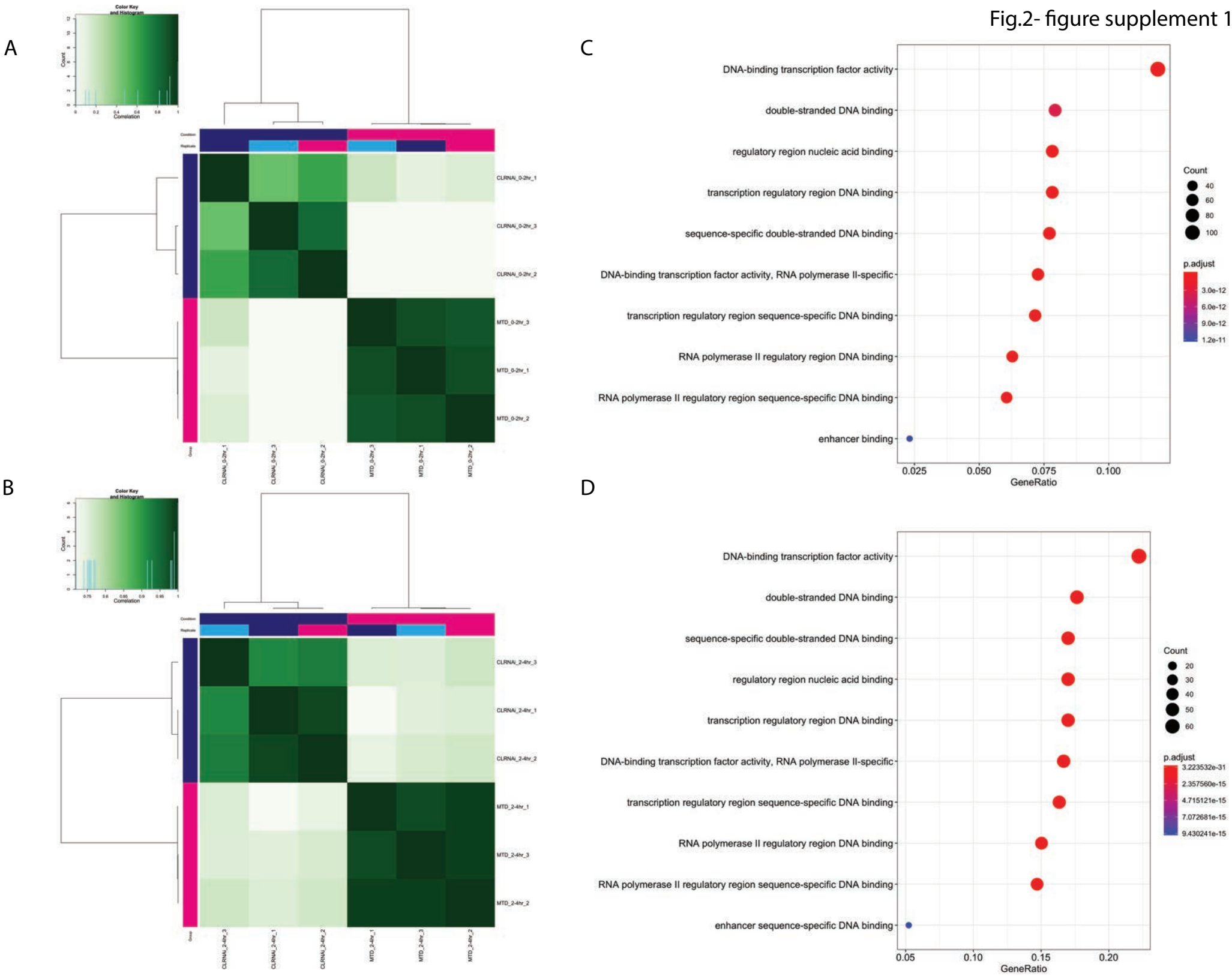


Fig.3-figure supplement 1

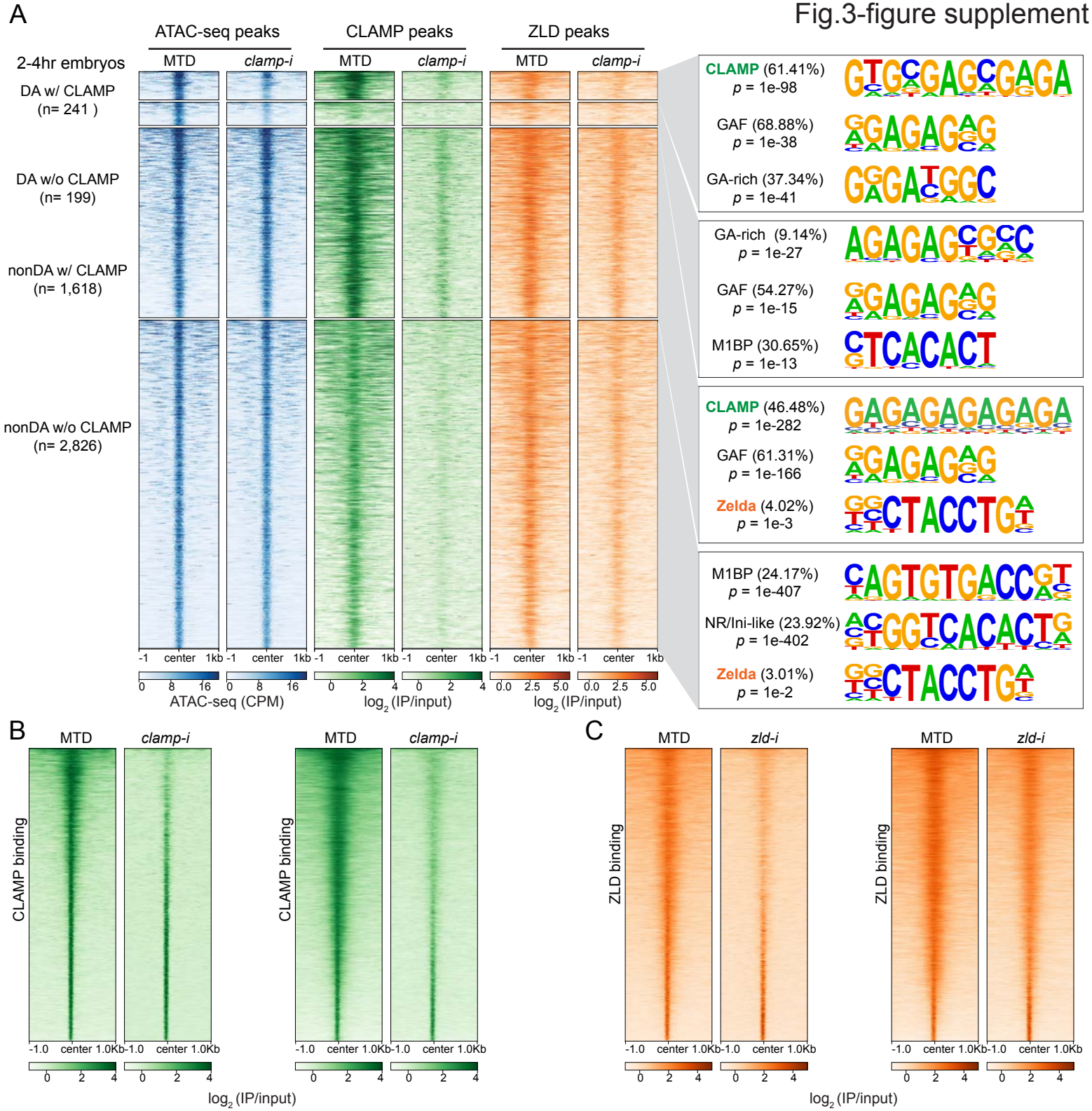


Fig.4-Figure supplement 1

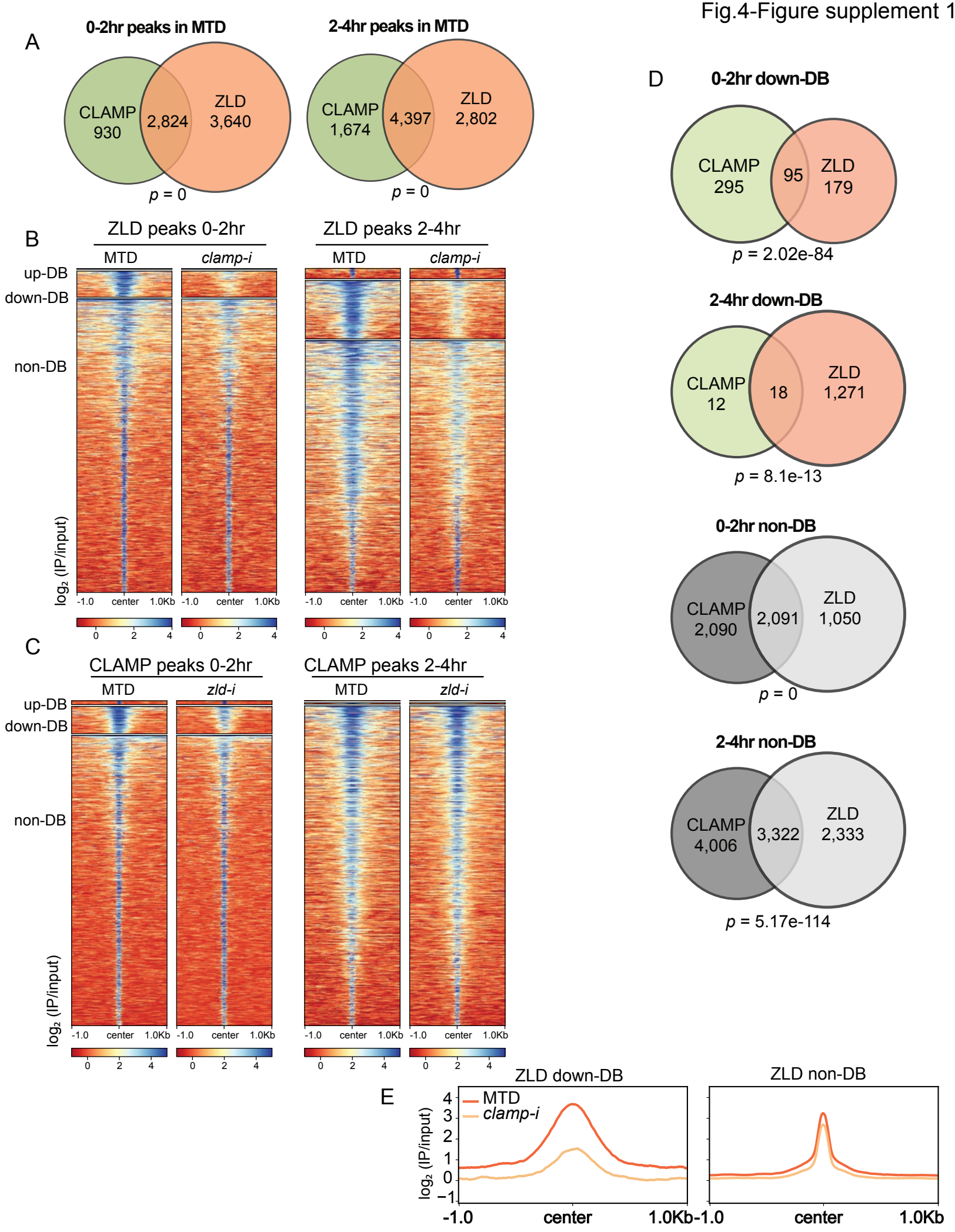
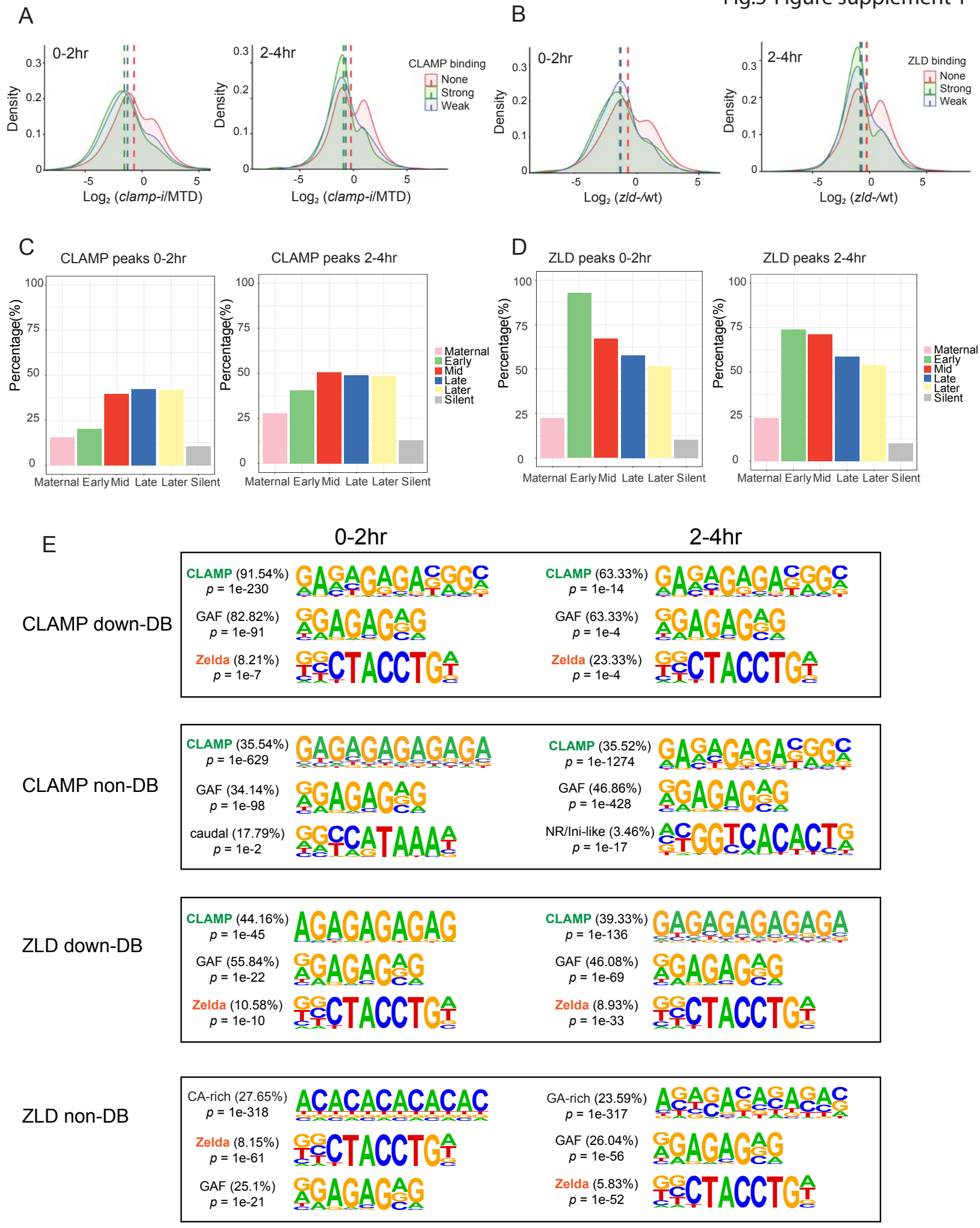



Fig.5-Figure supplement 1