CLAMP and Zelda function together as pioneer transcription factors to promote *Drosophila* zygotic genome activation

Jingyue Ellie Duan[1,*,†], Leila E. Rieder[2,*], Annie Huang[1], William T. Jordan, III[1], Mary McKenney[1], Scott Watters[3], Nicolas L. Fawzi[3], 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 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

Annie Huang, annie_huang@brown.edu

William T. Jordan, III, william_jordan@alumni.brown.edu

Mary McKenney mary_mckenney@brown.edu

Scott Watters scott_watters@brown.edu

Nicolas Fawzi nicolas_fawzi@brown.edu

Erica N. Larschan, erica_larschan@brown.edu

ABSTRACT

Because zygotic genome activation (ZGA) is an essential process across metazoans, it is key to evolve multiple pioneer transcription factors (TFs) to protect organisms from loss of a single factor. Pioneer TF Zelda (ZLD) is the only known factor which increases accessibility of chromatin to promote ZGA in the early *Drosophila* embryo. However, many genomic loci remain accessible without ZLD and have GA-rich motifs. Therefore, we hypothesized that other pioneer TFs that function with ZLD have not yet been identified in early embryos, especially those that bind to GA-rich motifs, such as CLAMP (Chromatin-linked adaptor for Male-specific lethal MSL proteins). Here, we determine that CLAMP is a novel pioneer TF which interacts directly with nucleosomes, regulates zygotic genome transcription, promotes chromatin accessibility, and facilitates the binding of ZLD to promoters. Thus, the maternal factor CLAMP functions with ZLD as a 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

INTRODUCTION

During zygotic genome activation (ZGA), dramatic reprogramming occurs in the zygotic nucleus to initiate global transcription and prepare the embryo for further development (Jukam et al., 2017). Chromatin changes that activate the zygotic genome during ZGA rely on cooperation among transcription factors (TFs) (Lee et al., 2014). However, only pioneer TFs (Cirillo and Zaret, 1999; Mayran and Drouin, 2018) can bind to ‘closed’ chromatin prior to ZGA because most TFs lack the ability to bind to nucleosomal DNA (Soufi et al., 2015).

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

GAGA-associated factor (GAF, Farkas et al., 1994) and Chromatin-linked adaptor for male-specific lethal (MSL) proteins (CLAMP, Soruco et al., 2013) are the

only two known TFs that are able to bind to GA-rich motifs and regulate transcriptional activation in *Drosophila* (Fuda et al., 2015; Kaye et al., 2018). GAF is known to perform several essential functions in early embryos, including chromatin remodeling (Leibovitch et al., 2002), nuclear divisions (Bhat et al., 1996) and RNA Pol II recruitment (Fuda et al., 2015). Recently, Harrison and colleagues demonstrated that GAF regulates ZGA and opens chromatin in the early embryo but functions largely independent from ZLD (companion submission).

CLAMP is essential for early embryonic development (Rieder et al., 2017) and plays several key roles including: 1) recruiting the MSL dosage compensation complex to the male X-chromosome prior to ZGA (Rieder et al., 2019); 2) activating coordinated regulation of the histone genes (Rieder et al., 2017). Therefore, we hypothesized that CLAMP as a new GA-binding pioneer TF which regulates ZGA either interdependently with or independent of ZLD.

Here, we identify the GA-binding factor CLAMP as a new pioneer transcription factor, one of only two known pioneer TFs in *Drosophila*. We combine genomic and biochemical approaches to demonstrate: 1) CLAMP is a novel pioneer factor which binds to nucleosomal DNA, activates zygotic transcription, and increases chromatin accessibility; 2) CLAMP and ZLD function interdependently to regulate transcription, chromatin accessibility and each other's occupancy at a subset of promoters; 3) When ZLD is bound to a locus but does not increase chromatin accessibility, CLAMP can often function redundantly to open the chromatin. Because ZGA is an essential process,

redundant pioneer TFs protect organisms from lethality that would be caused by mutation of a single non-redundant factor.

RESULTS

CLAMP binds to nucleosomal DNA and activates zygotic transcription

One of the intrinsic characteristics of pioneer transcription factors is their capacity to bind nucleosomal DNA and compacted chromatin (Cirillo and Zaret, 1999). To test the hypothesis that CLAMP is a novel pioneer factor, we performed electromobility shift assays (EMSAs) that directly test the intrinsic capacity of CLAMP to directly interact with nucleosomes *in vitro* **(Figure 1)**. First, we identified a 240 bp region of the X-linked 5C2 locus **(Figure 1A)** that CLAMP binds to *in vivo* and had decreased chromatin accessibility in the absence of CLAMP (J. Urban et al., 2017). This region is also normally occupied by nucleosomes **(Figure 1A)**, suggesting that CLAMP promotes accessibility of this region while binding to nucleosomes *in vivo*. Furthermore, CLAMP binding to 5C2 is important for transcriptional activation mediated by this locus (Alekseyenko et al., 2008).

Next, we performed *in vitro* nucleosome assembly using 240 bp of DNA from the 5C2 locus that contains three CLAMP-binding motifs and used 5C2 naked DNA as a control. We found that both the CLAMP DNA binding domain (DBD**, Figure 1B**) and full-length protein (F/L**, Figure 1C**) (both kept soluble by an N-terminal maltose binding

protein fusion) can bind and shift naked 5C2 DNA and nucleosomes assembled with 5C2 DNA. Increased protein concentration results in a secondary “super” shift species **(Figure 1B & 1C)**, indicating that the three CLAMP-binding motifs may be occupied by multiple CLAMP molecules. This is the first experiment that demonstrates the ability of CLAMP to directly bind nucleosomal DNA, the essential feature of a pioneer TF.

The second characteristic of a pioneer TF is its ability to directly target zygotic genes for activation (Zaret and Carroll, 2011). To define how CLAMP regulates transcription in early embryos, we examined the effect of maternal CLAMP depletion by RNAi on expression of maternally-deposited or zygotically-transcribed genes using mRNA-seq data (Rieder et al., 2017). We found that only the expression levels of zygotically-transcribed genes but not maternally-deposited genes were significantly ($p < 0.05$, Student’s t-test) downregulated in embryos lacking CLAMP **(Figure 1D).** We next asked whether the ability of CLAMP to bind to genes directly regulates zygotic gene activation by performing ChIP-seq on CLAMP in early embryos at two time ranges: 1) 0-2 hours after egg laying; 2) 2-4 hours after egg laying. Similar to a previous study (Harrison et al., 2011) that defined the role of ZLD in early embryos, we determined that genes strongly bound by CLAMP showed a higher level of gene expression reduction after *clamp* RNAi than weakly bound or unbound genes **(Figure 1E).** Furthermore, the relationship between CLAMP-binding and gene activation is stronger at zygotically-transcribed genes compared to genes encoding maternally-deposited mRNA **(Figure S1A-B)**. Therefore, the direct binding of CLAMP to genes regulates their transcriptional activation during ZGA, confirming the second key characteristic of pioneer TFs.

Next, we compared and contrasted the transcriptional roles of CLAMP and ZLD in early embryos. We found that binding of CLAMP was enriched at mid- and late-transcribed zygotic genes (categories defined in Li et al., 2014), while ZLD binds more strongly to early-transcribed zygotic genes **(Figure S1C & S1D)**. Overall, we demonstrate that like the pioneer factor ZLD, CLAMP also functions as a pioneer transcriptional activator that directly activates zygotically-transcribed genes during early development.

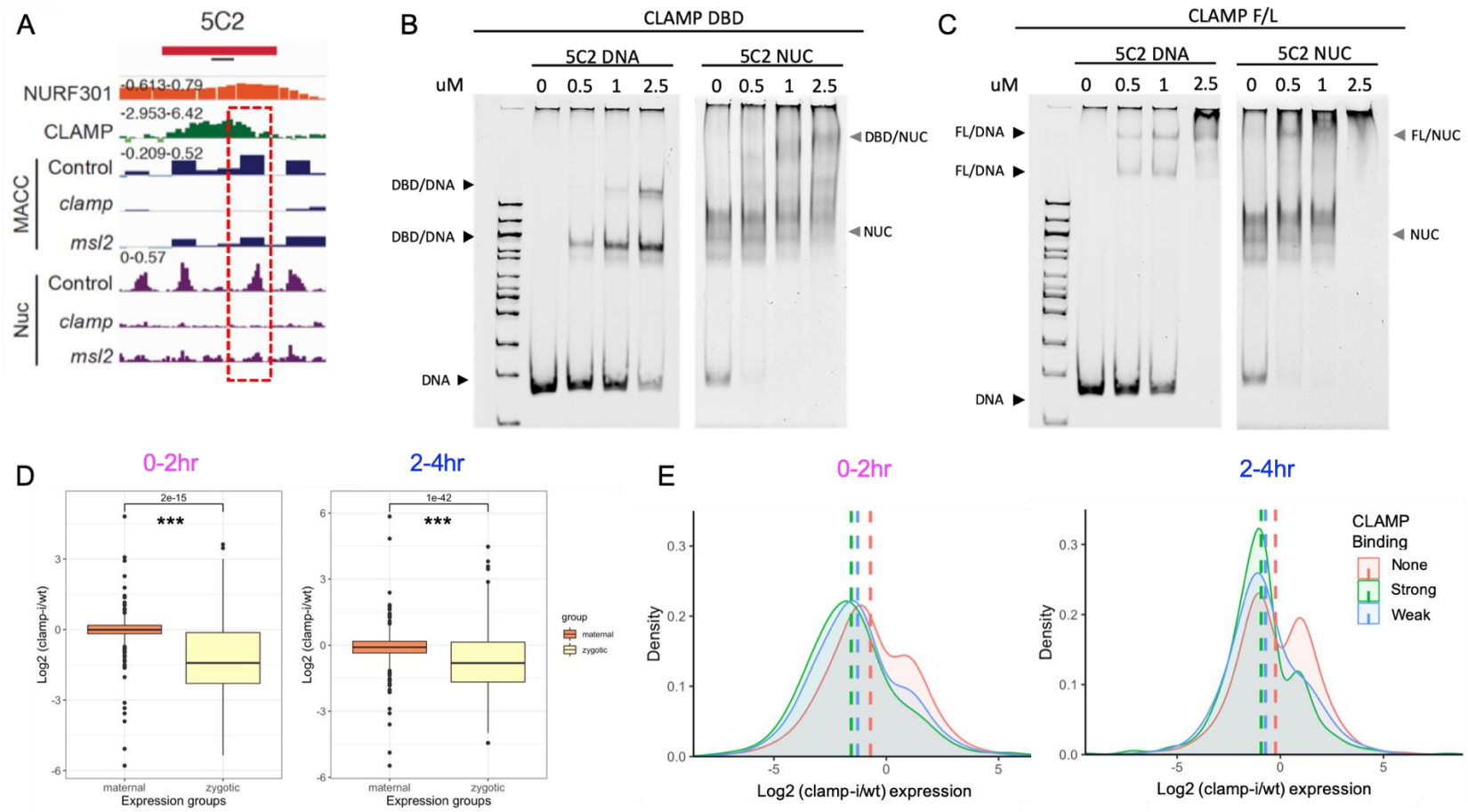

Figure 1. CLAMP exhibits two features of pioneer transcription factors: binding to nucleosomal DNA and activating zygotic transcription

A. Genome browser tracks are shown for a region of the CES 5C2 locus used to make *in vitro* reconstituted nucleosomes (Urban et al., 2017). CLAMP ChIP-seq normalized sequencing reads are shown in green. MNase-seq MACC scores from S2 cells or S2 cells depleted for maternal *clamp* are shown in dark blue. The nucleosome profile is shown in purple. The dashed rectangle highlights the genomic region used to reconstitute nucleosomes.

B. Electrophoretic mobility shift assay (EMSA) shows the binding of increasing amounts of CLAMP DNA-binding domain (fused to MBP) to 5C2 naked DNA or 5C2 *in vitro* reconstituted nucleosomes.

C. EMSA shows the binding of increasing amounts of full-length CLAMP (fused to MBP) to 5C2 DNA or 5C2 nucleosomes.

D. Effect of maternal CLAMP depletion 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).

E. Gene expression changes caused by maternal CLAMP depletion at genes with strong, weak and no CLAMP occupancy as measured by ChIP-seq in 0-2hr (left) or 2-4hr (right) embryos.

CLAMP regulates chromatin accessibility in early embryos

Another essential characteristic of pioneer transcription factors is that they are able to establish and maintain the accessibility of their DNA target sites, allowing other TFs to bind to DNA and activate transcription (Zaret and Carroll, 2011; Iwafuchi-Doi et al., 2016). We previously used MNase-seq (J. Urban et al., 2017) to determine that CLAMP guides MSL complex to GA-rich sequences by promoting an accessible chromatin environment on the male X-chromosomes in cell lines. Furthermore, GA-rich motifs enriched in regions that remain accessible in the absence of pioneer factor ZLD (Schulz et al., 2015; Sun et al., 2015). Therefore, we hypothesized that CLAMP regulates chromatin accessibility during ZGA.

To test our hypothesis, we performed Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) at 0-2hr and 2-4hr embryos with wild-type levels of CLAMP [maternal triple driver (MTD) alone] and embryos depleted for maternally contributed CLAMP using the MTD driver (*clamp-i*) as we performed previously (Rieder et al., 2017). Knockdown of CLAMP was validated by qPCR and western blot (see **Materials and Methods**).

Next, we defined differentially accessible (DA) regions **(Figure 2A)** by comparing ATAC-seq reads between MTD and *clamp-i* embryos using Diffbind (Stark and Brown, 2019). High Pearson correlation for DA regions among replicates indicate strong reproducibility of our data **(Figure S2A-B)**. There were hundreds of genomic

regions that had reduced chromatin accessibility in the absence of CLAMP (**Figure 2A,** 0-2hr: 636; 2-4hr: 419), indicating that CLAMP is required for the chromatin accessibility of these regions. In contrast, very few regions (0-2hr: 13; 2-4hr: 85) increased their accessibility in the absence of CLAMP **(Figure 2A)**. Gene Ontology (GO) analysis **(Figure S2C-D)** indicates that CLAMP increases accessibility of chromatin regions that are mainly within DNA-binding, RNA Pol II binding, and enhancer-binding TF encoding genes **(Figure S2C-D)**.

Moreover, a subset of DA regions were bound by CLAMP suggesting that CLAMP directly regulates chromatin accessibility [22% (138/636) at 0-2hr; 55% (229/419) at 2-4hr]. For example, a DA site with a cluster of CLAMP enrichment was located at the promoter of *cg11023* **(Figure 2B)**. We also determined how DA sites and CLAMP binding sites were distributed throughout the genome **(Figure 2C)**. While DA sites were highly enriched in promoter regions (81.2%), CLAMP binds to both promoters (36.8%) and other introns (23.9%). Therefore, CLAMP is required to establish or maintain open chromatin at promoters, but could also play other roles, such as regulating pre-mRNA splicing at intronic regions. Motif analysis also identified both GA-rich motifs and ZLD motifs enriched at regions which require CLAMP for their accessibility **(Figure 2D)**. These data suggest that CLAMP also regulates the accessibility of some ZLD binding sites, a hypothesis that we will discuss further below.

We further determined whether CLAMP-mediated accessibility could specifically drive early transcription by examining the relationship between the chromatin

accessibility changes (ATAC-seq) and gene expression of the nearest gene as measured by RNA-seq (Rieder et al., 2017). We obtained positive *R* values with significant Pearson correlation *p*-values between CLAMP-mediated chromatin accessibility changes and gene expression changes at both time points: 0-2hr (R = 0.24, p = 3.6e-08) and 2-4hr (R = 0.14, p = 0.0046, **Figure 2E)**. Therefore, CLAMP-mediated chromatin accessibility is positively correlated with gene expression. Overall, our ATAC-seq data indicate that CLAMP promotes accessibility of chromatin during ZGA, a key property of pioneer TFs.

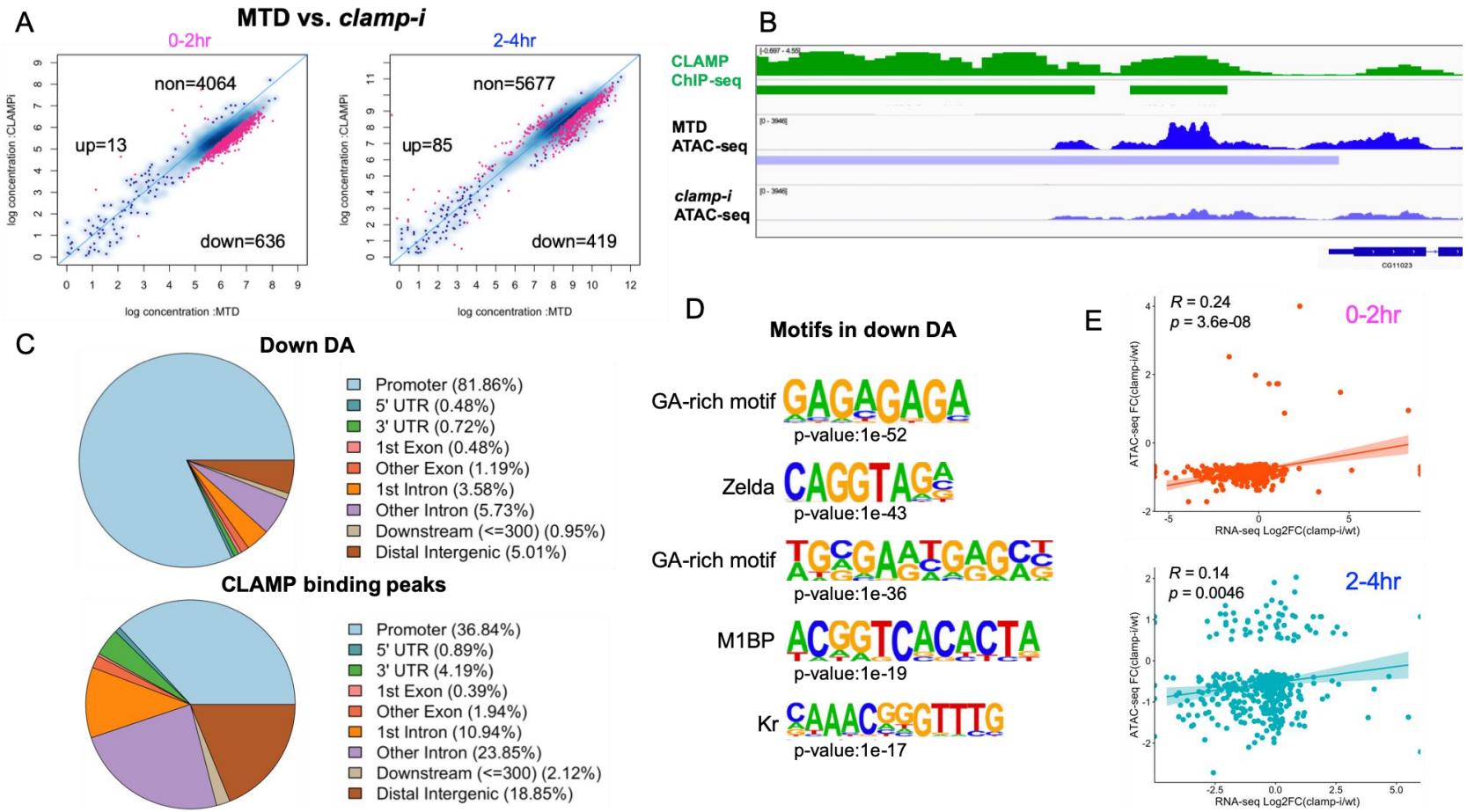

Figure 2. CLAMP regulates chromatin accessibility throughout ZGA

A. Differential accessibility (DA) analysis by ATAC-seq from MTD embryos versus *clamp-i* embryos in 0-2hr (left) or 2-4hr (right). Blue dots indicate non-differentially accessible sites. Pink dots indicate significant ($p < 0.05$) differential peaks after CLAMP depletion, identified by DiffBind (DESeq2). Number of peaks in each class is noted on the plot.

B. Example of genomic locus with CLAMP binding (ChIP-seq) which shows significant ATAC-seq signal reduction after *clamp* RNAi.

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

D. Motifs enriched in regions that require CLAMP for chromatin accessibility.

Pearson correlation between CLAMP-mediated changes in gene expression (mRNA-seq) and ATAC-seq signal.

Four classes of CLAMP-related genomic loci regulate zygotic transcription differently

In order to determine how the direct binding of CLAMP relates to CLAMP-mediated chromatin accessibility, we integrated CLAMP binding sites from ChIP-seq with ATAC-seq peaks. Inspired by a FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements) study which identified ZLD-mediated chromatin accessible regions (Schulz et al., 2015), we defined four classes of CLAMP-related peaks: 1) DA, CLAMP-bound (0-2hr: 138 peaks; 2-4hr: 229 peaks); 2) DA, CLAMP non-bound (0-2hr: 501 peaks; 2-4hr: 191 peaks); 3) Non-DA, CLAMP-bound (0-2hr: 427 peaks; 2-4hr: 1307 peaks); 4) Non-DA, CLAMP non-bound (0-2hr: 3641 peaks; 2-4hr: 4395 peaks) **(Figure 3A)**. Average profiles of ATAC-seq read counts validated these four classes of sites in control MTD and *clamp-i* embryos: DA regions had a significant decrease in accessibility in embryos lacking CLAMP, while Non-DA regions maintained accessibility in the absence of CLAMP **(Figure 3A).**

To compare CLAMP- and ZLD-mediated chromatin accessibility during ZGA, we also defined four ZLD-related classes of genomic loci using ZLD binding sites from ChIP-seq (generated in this study) and ATAC-seq datasets (Hannon et al., 2017) generated in wildtype (wt) and *zld-i* embryos at the NC14 (2-3hr) stage. Specifically, we defined four classes of genomic loci as described above for CLAMP-related classes: 1) DA, ZLD-bound (806 peaks); 2) DA, ZLD non-bound (426 peaks); 3) Non-DA, ZLD-bound (1,331 peaks); 4) Non-DA, ZLD non-bound (2,269 peaks) **(Figure 3A).** ATAC-

seq average profiles for each ZLD-related class also confirmed our classification **(Figure 3A)**.

We performed pairwise intersection of genomic regions among four classes of CLAMP-related peaks and four classes of ZLD-related peaks **(Table 1)**. We observed significant ($p < 0.05$, Hypergeometric test) overlap between the classes of sites that are bound by CLAMP and ZLD compared to classes of sites that are not bound by CLAMP or ZLD for both DA and Non-DA sites. This relationship among classes of sites suggests that CLAMP and ZLD act interdependently: CLAMP promotes chromatin accessibility at sites that are bound by ZLD but do not require ZLD for their accessibility and vice versa. Therefore, as Schulz et al. (2015) predicted a GA-binding TF can increase the accessibility of sites that are bound by ZLD but do not require ZLD for chromatin accessibility, we observed that CLAMP is one of these GA-binding TFs.

Table 1. Significance of pairwise overlaps among four classes of CLAMP-related peaks and four classes of ZLD-related peaks

	DA, w CLAMP 2-4hr	DA, wo CLAMP 2-4hr	Non-DA, w CLAMP 2-4hr	Non-DA, wo CLAMP 2-4hr
DA, w ZLD 2-3hr	0.015	-	1.43e-09	-
DA, wo ZLD 2-3hr	-	-	-	-
Non-DA, w ZLD 2-3hr	1.15e-37	-	1.61e-154	-
Non-DA, wo ZLD 2-3hr	-	-	-	6.96e-49

"-" not significant

w: with; wo: without

Next, we used ChIP-seq average profiles to measure the relative strength of CLAMP or ZLD binding over each class of the classes of sites in control MTD vs. *clamp-i* embryos **(Figure 3B)** and control MTD vs. *zld-i* embryos **(Figure 3C)**. First, we examined the binding of each factor at the four classes of sites defined based on regulation of chromatin accessibility by the same factor. Throughout ZGA, both DA and Non-DA sites with CLAMP or ZLD bound had significantly higher enrichment in MTD control embryos than in their RNAi embryos **(Figure 3B & 3C)**, as expected in ChIP-seq due to protein depletion in RNAi embryos **(Figure S3A)**. Furthermore, CLAMP binding is slightly enriched at DA sites compared with Non-DA sites in control MTD embryos suggesting that higher levels of CLAMP occupancy promote chromatin accessibility. **(Figure 3B)**. Similarly, ZLD binding is enriched at DA sites compared with non-DA sites in control MTD embryos **(Figure 3C)**, consistent with a previous study (Schulz et al., 2015). The opposite result was observed in *zld-i* embryos: ZLD is less enriched at DA sites than non-DA sites **(Figure 3C)**. It is important to note that ZLD levels also increased in *zld-i* embryos at 2-4 hours at the ChIP (ChIP-seq, **Figure S3B**), mRNA (qPCR, **Figure S3C**) and protein (western, **Figure S3D**) levels.

Prior work demonstrated that genes that require ZLD for chromatin accessibility were downregulated in the absence of ZLD (Schulz et al., 2015). To determine the functional impact of the four classes of CLAMP-related ATAC-seq sites on early zygotic transcription, we measured the expression of genes associated with each class of sites in *clamp-i* compared to control MTD embryos from RNA-seq data (Rieder et al., 2017) **(Figure 3D)**. Genes in the two CLAMP-bound classes (DA and non-DA) showed more

significant ($p < 0.05$, Student's t-test) downregulation in the absence of CLAMP compared to the two CLAMP non-bound classes (DA and non-DA) at the 0-2hr time point **(Figure 3D)**. In contrast, later in development (2-4hrs), genes associated with the two DA classes (CLAMP-bound and CLAMP non-bound) showed significant ($p < 0.05$, Student's t-test) reduction in expression compared to two non-DA classes (CLAMP-bound and CLAMP non-bound **Figure 3D**). Overall, these results indicate that CLAMP binding in the absence of chromatin accessibility changes can regulate gene expression early in development (0-2 hrs). However, later in development (2-4hrs) changes in chromatin accessibility are more important for regulating gene expression than CLAMP binding alone.

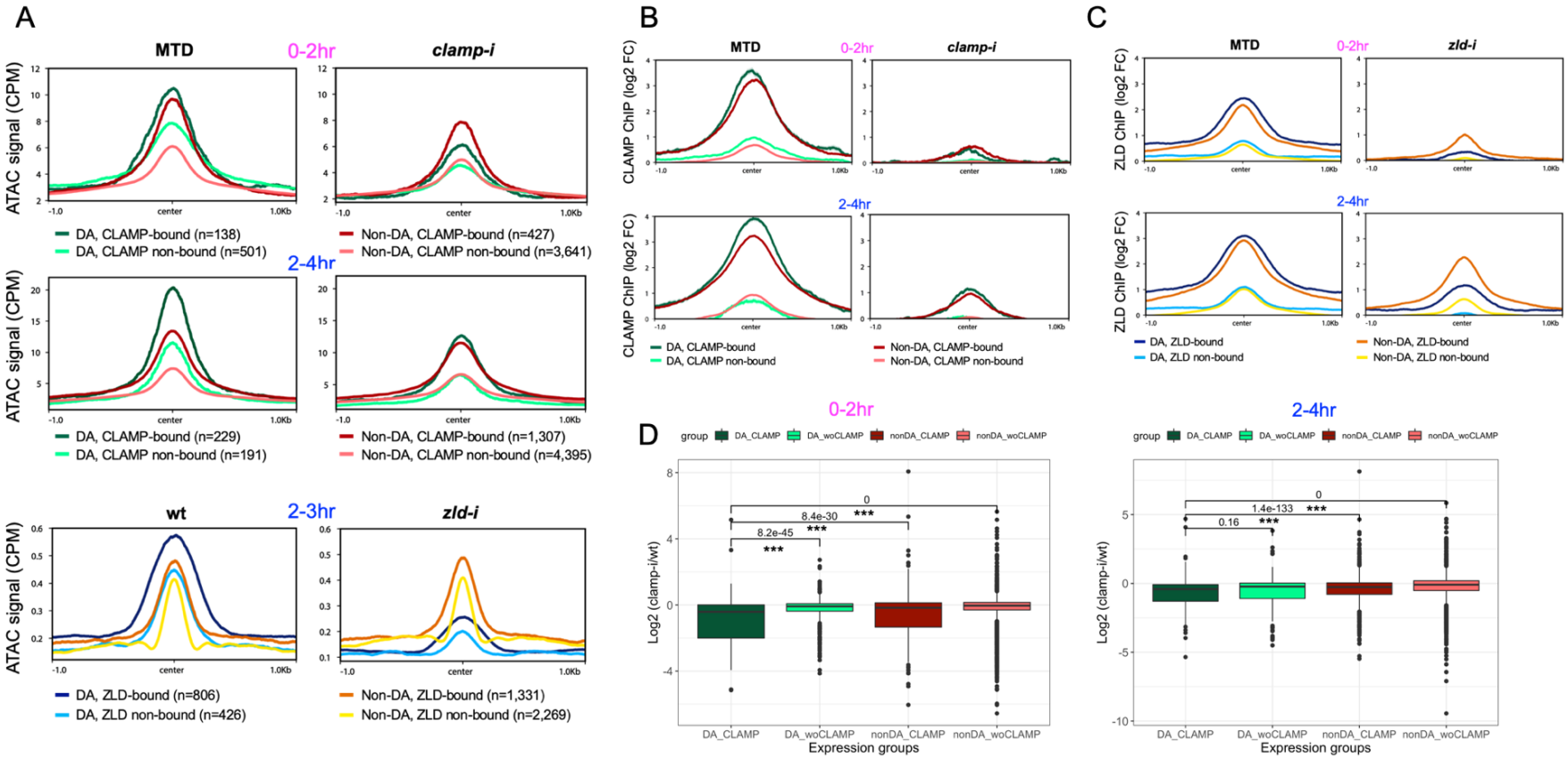

Figure 3. CLAMP-mediated chromatin accessibility is correlated with early CLAMP-dependent gene expression

A. Four classes of CLAMP-related peaks defined by combining ATAC-seq and CLAMP ChIP- seq peaks.
Dark green: DA (differentially accessible) CLAMP-bound; Light green: DA, CLAMP non-bound;
Dark red: Non-DA, CLAMP-bound; Light red: Non-DA, CLAMP non-bound.
Similarly, four classes of ZLD-related peaks were defined by combining ATAC-seq (Hannon et al., 2017) and ZLD ChIP-seq peaks (from this study) for further analysis:
Dark blue: DA ZLD-bound; Light blue: DA, ZLD non-bound;
Dark yellow: Non-DA, ZLD-bound; Light yellow: Non-DA, ZLD non-bound.

B. The enrichment of CLAMP ChIP-seq signals over four classes of CLAMP-related peaks.

C. The enrichment of ZLD ChIP-seq signals over four classes of ZLD-related peaks.

D. Gene expression differences caused by maternal CLAMP depletion among four classes of CLAMP-related peaks.

CLAMP and ZLD bind interdependently to many promoters

Due to the significant overlap between accessible regions that are both bound by CLAMP and ZLD **(Table 1)**, we hypothesized that these two pioneer TFs bind to chromatin interdependently. To test this hypothesis, we identified motifs enriched in the DNA sequences underlying the bound CLAMP-dependent ATAC-seq peaks (DA, CLAMP bound, **Figure 4A**) or bound ZLD-dependent ATAC-seq peaks (DA, ZLD bound, **Figure 4B**). As predicted, GA-rich motifs and ZLD motifs were enriched in CLAMP-dependent peaks at the 0-2hr time point and ZLD-dependent peaks at the 2-3hr time point. However, the ZLD motif is not present within CLAMP-dependent ATAC-seq peaks at the 2-4hr time point **(Figure 4A)**. This result suggests that an interdependent relationship between CLAMP and ZLD may be a developmental stage-specific event during ZGA.

To directly determine how CLAMP and ZLD impact each other's binding, we performed ChIP-seq **(Figure S4A-B)** for CLAMP and ZLD in control MTD embryos and embryos that were maternally depleted for each protein by RNA-i at two time points: 1) before (0-2hr) ZGA and 2) during and after (2-4hr) ZGA. Overall, ZLD has more peaks (0-2hr: 6,464; 2-4hr: 7,199) across the whole genome than CLAMP (0-2hr: 3,754, 2-4 hr: 6,071) in MTD embryos. As we hypothesized, CLAMP and ZLD peaks significantly ($p < 0.05$, Hypergeometric test) overlapped **(Figure S4C)**. Prior to ZGA (0-2hrs), ZLD showed a higher enrichment at promoters than CLAMP, while CLAMP had a similar distribution compared to ZLD at 2-4hr time point, indicating CLAMP binds to promoters

later than ZLD **(Figure 4C)**. Moreover, depletion of either maternal *zld* or *clamp* mRNA altered the genomic distribution of CLAMP and ZLD: both peaks shifted from promoters to introns. Interestingly, maternal *zld* RNAi no longer affects CLAMP binding to promoters at 2-4hr **(Figure 4C)**.

Next, we defined the differential binding (DB**, Figures 4D & 4E**) of CLAMP and ZLD in the absence of each other's maternally deposited mRNA using Diffbind (Stark and Brown, 2019). ZLD binding was significantly reduced in the absence of CLAMP. There were 274 (0-2hr) and 1,289 (2-4hr) down-DB sites where ZLD binding decreased in *clamp-i* compared to MTD controls **(Figures 4D & S4D)**. Fewer ZLD binding sites increased in occupancy after *clamp* RNAi: 8 sites (0-2hr) and 233 (2-4hr) up-DB sites. The majority of the ZLD binding sites were not affected (non-DB sites, 0-2h: 3,144; 2-4h: 5,672**, Figures 4D & S4D**). In contrast, loss of ZLD had a more minor impact on CLAMP binding, especially at 2-4hr: 390 (0-2 hr) and 30 (2-4 hr) down-DB sites **(Figures 4E & S4E)**. Very few up-DB sites were identified where CLAMP occupancy is increased after *zld* RNAi (0-2hr: 54, 2-4hr: 3). The majority of CLAMP binding sites remained unchanged after *zld* RNAi (non-DB, 0-2h: 4,184, 2-4h: 7,351, **Figures 4E & S4E**).

Therefore, we conclude that CLAMP impacts ZLD throughout ZGA and has major effects at the 2-4hr time point, while ZLD has a modest effect on CLAMP binding that occurs largely within the 0-2hr time window. An important caveat to note is that ZLD levels begin to recover **(Figure S3B-D)** in *zld-i* embryos by 2-4hr, likely due to

expression from the zygotic *zld* gene, which could influence our interpretation of these results. Moreover, CLAMP and ZLD down-DB sites or non-DB sites were also significantly ($p < 0.05$, Hypergeometric test) overlapped throughout ZGA **(Figure S4F)**, suggesting an interdependent relationship between CLAMP and ZLD in early development.

Next, we asked whether CLAMP chromatin accessibility could specifically drive ZLD binding and vice versa. Therefore, we measured the enrichment of CLAMP or ZLD binding by ChIP-seq four classes of regions defined based on RNAi for the opposite factor **(Figure 4F & 4G)**. For example, we examined CLAMP occupancy at sites defined based ZLD RNAi ATAC-seq data and vice versa. We found that ZLD binds preferentially to CLAMP-bound regions, independent of whether these loci depend on CLAMP for accessibility or not **(Figure 4F)**. ZLD enrichment was also significantly reduced upon CLAMP depletion throughout ZGA **(Figure 4F)**. In contrast, a different pattern was observed in CLAMP ChIP enrichment in ZLD-related regions: CLAMP also favors those that are ZLD-bound, but it is highly enriched in peaks that do not require ZLD for chromatin accessibility (non-DA, ZLD-bound, **Figure 4G**), consistent with the role of CLAMP in maintaining the accessibility of these sites **(Table 1)**. Moreover, ZLD depletion caused a significant reduction in CLAMP largely at ZLD-dependent ATAC-seq sites (DA, ZLD-bound) at 0-2hr, rather than at non-DA ZLD-bound sites **(Figure 4G)**. At 2-4hr, loss of ZLD was no longer able to reduce the CLAMP enrichment at non-DA ZLD-bound sites **(Figure 4G)**. Overall, CLAMP and ZLD exhibit inter-dependent binding that

correlates with high levels of enrichment of both factors, and CLAMP binding is mainly enriched at non-DA ZLD-bound groups, even after ZLD depletion.

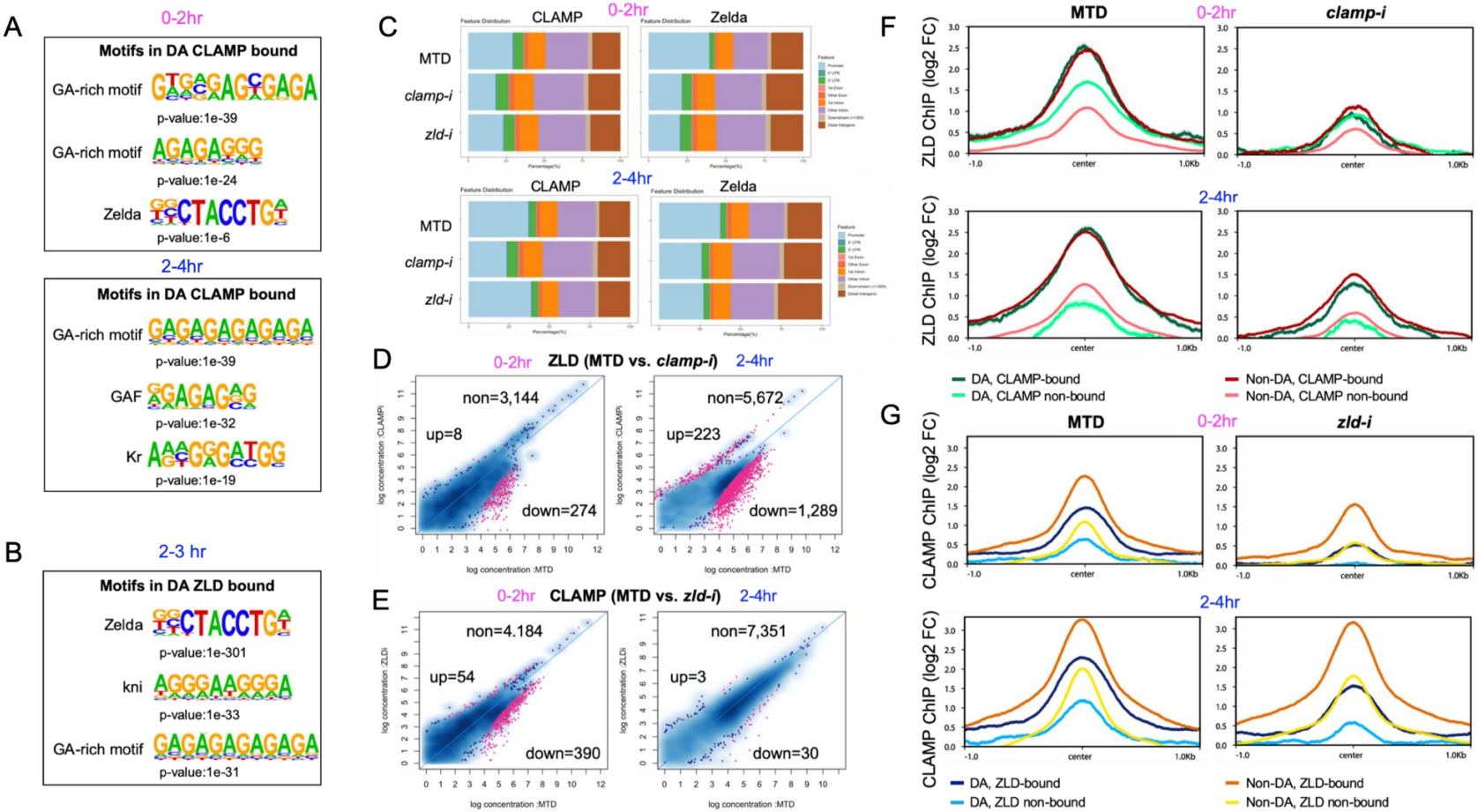

Figure 4. CLAMP and ZLD depend on each other for chromatin binding

A. Motifs enriched in regions that depend on CLAMP for accessibility and have CLAMP binding in 0-2hr and 2-4hr embryos.

B. Motifs enriched in regions that depend on ZLD for accessibility and have ZLD binding (2-3hr).

C. Genomic distribution fractions for CLAMP and ZLD peaks in the *Drosophila* genome in 0-2hr and 2-4hr embryos (MTD, *clamp-i* and *zld-i)*.

D. Scatter plots of ZLD peaks from 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 ($p < 0.05$) differential peaks identified by DiffBind (DESeq2). The number of peaks changed in each direction is noted in the plot.

E. Scatter 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 ($p < 0.05$) differential binding (DB) peaks identified by DiffBind (DESeq2). Number of peaks in each direction is noted in the plot.

F. The enrichment of ZLD ChIP-seq signals over four classes of CLAMP-related peaks.

G. The enrichment of CLAMP ChIP-seq signals over four classes of ZLD-related peaks.

CLAMP and ZLD interdependently regulate zygotic transcription

To determine how CLAMP and ZLD regulate each other's binding and transcription at sites where they bind dependently vs. independently, we further defined dependent sites as down-DB sites and independent sites as non-DB sites. Interestingly, the dependent sites for both CLAMP and ZLD showed a much broader binding pattern compared to independent sites **(Figures 5A & 5B)**. On average, the peak size of dependent sites (400-500bp) is almost double that of independent sites (200-250bp), with significant ($p < 0.001$, Mann-Whitney U-test) differences in peak size for both TFs at both time points **(Figure 5C)**. Moreover, dependent sites are enriched at promoters and TSS, while independent sites are mainly localized at introns **(Figure 5D)**. Overall, dependent sites are broad and localized at promoters while independent sites are narrower and located within introns.

Previous proteomic studies (J. A. Urban et al., 2017) found no evidence that CLAMP and ZLD could directly contact each other at the protein level, suggesting that CLAMP and ZLD might regulate each other via binding to their own motifs. Therefore, we asked whether the motifs enriched at dependent vs. independent sites differed from each other. We found that dependent sites are enriched for motifs specific for the protein required for the binding of the other factor **(Figures S5A & S5C)**, which are not present at independent sites **(Figures S5B & S5D).** Therefore, the presence of specific CLAMP and ZLD motifs correlates with the ability of CLAMP and ZLD to promote each other's binding.

To further understand how CLAMP and ZLD regulate each other's binding via their own motifs, we calculated the frequency of their motifs per peak at dependent and independent sites **(Figures S5E & S5F)**. Throughout ZGA, the number of binding motifs for the required protein is significantly ($p < 0.01$, Mann-Whitney U-test) higher at the dependent sites than at the independent sites for both TFs **(Figures S5E & S5F)**, explaining the broader binding pattern at dependent peaks compared with independent peaks.

Next, we asked whether CLAMP and ZLD regulate each other's binding to specifically drive transcription of target genes. Thus, we incorporated mRNA-seq data from embryos in which either maternal *zld* (Combs and Eisen, 2017, GSE71137) or *clamp* (Rieder et al., 2017, GSE102922) has been depleted. Absence of maternal *zld* significantly ($p < 0.05$, Mann-Whitney U-test) reduces the expression of genes at sites where CLAMP is dependent on ZLD more than independent sites at the 0-2hr but not the 2-4hr time point **(Figure 5E)**. Therefore, ZLD specifically regulates early genes where ZLD promotes CLAMP binding. Also, compared to independent genes, genes where ZLD binding is dependent on CLAMP had a significant ($p < 0.05$, Mann-Whitney U-test) expression reduction after *clamp* RNAi at both 0-2hr and 2-4hr time points **(Figure 5F)**. Thus, CLAMP also regulates genes targeted by ZLD. Overall, our analysis revealed that enrichment of ZLD and CLAMP motifs at promoters drives binding of both TFs such that they interdependently regulate transcriptional activation during ZGA.

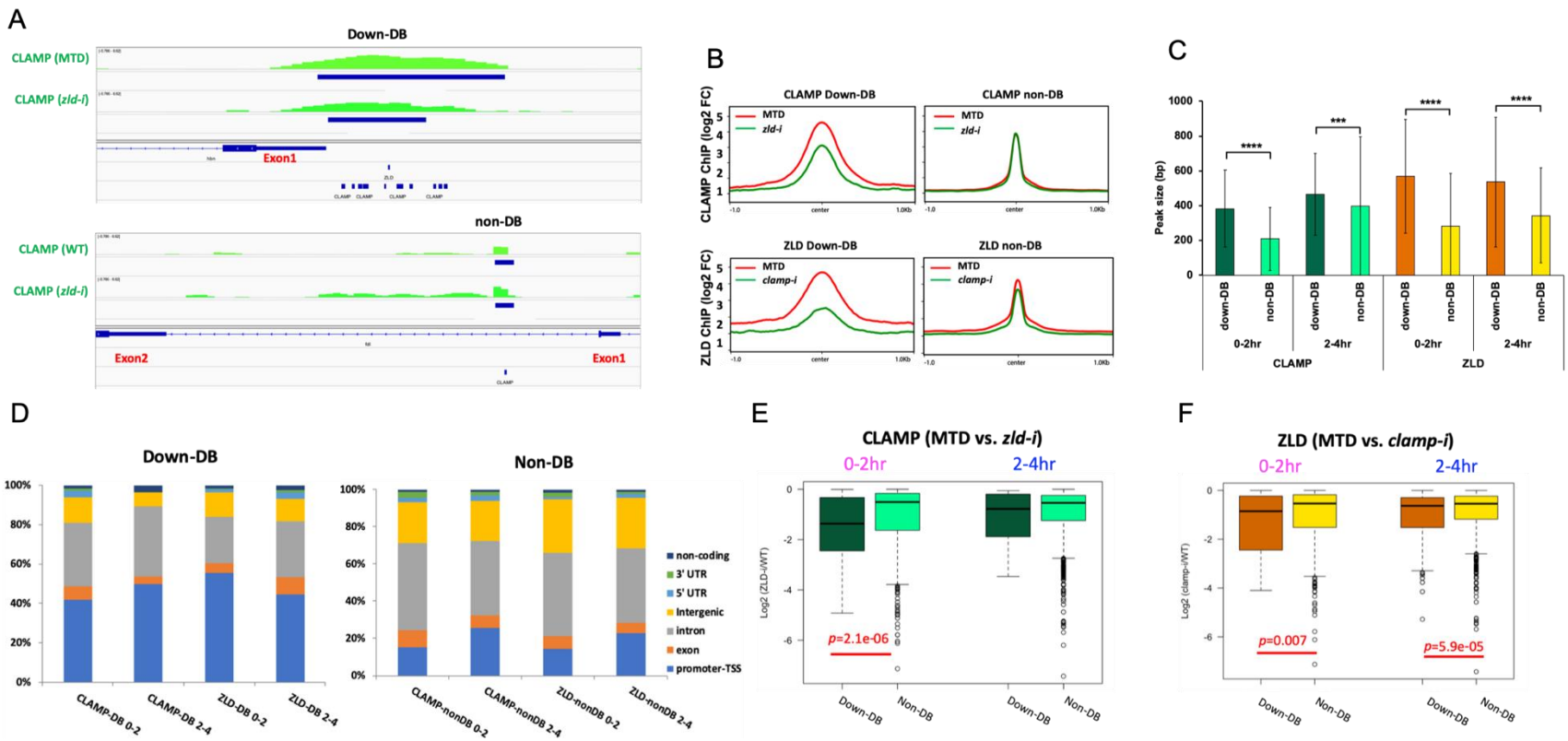

Figure 5. CLAMP and ZLD exhibit broad dependent binding sites at promoters and narrow independent binding sites at introns

A. ChIP-seq profile of an example of a CLAMP differential binding (DB) peak at the promoter region (upper) or non-differential binding (non-DB) peak at the intron (down) in MTD vs. *zld-i*. CLAMP and ZLD motifs are marked in blue.

B. Average profiles show the size of DB or non-DB peaks of CLAMP in MTD versus *zld-I* embryos and ZLD in MTD versus *clamp-i* embryos.

C. Bar plot of the size of DB and non-DB peaks. *** $p < 0.001$, **** $p < 0.0001$.

D. Stacked barplots of CLAMP and/or ZLD DB (left) or non-DB peaks (right) distribution fraction in the *Drosophila* genome in 0-2hr and 2-4hr.

E. Expression of CLAMP bound genes in DB and non-DB peaks in MTD vs. *zld-i* embryos.

Expression of ZLD bound genes in DB and non-DB peaks in MTD vs. *clamp-i embryos*.

Chromatin accessibility differs at CLAMP and ZLD dependent and independent sites

Given the strong interdependent relationship between CLAMP and ZLD binding to chromatin, we next asked whether chromatin accessibility at dependent and independent sites is changed by RNAi of required proteins. The average ATAC-seq signals are significantly reduced at sites where ZLD depends on CLAMP for binding in *clamp-i* embryos compared to MTD controls **(Figures 6A & 6B)**. Furthermore, the accessibility at independent sites is lower than that at dependent sites **(Figures 6A & 6B),** consistent with independent regions being mainly located in introns (**Figure 5D)** which usually have reduced chromatin accessibility compared to o promoters.

We also performed analysis of chromatin accessibility at sites where CLAMP is dependent or independent of ZLD for binding **(Figures 6C & 6D).** As expected, the largely intronic independent regions showed a very low level of chromatin accessibility. Interestingly, at sites where CLAMP depends on ZLD to bind, the accessibility was slightly increased upon the loss of ZLD, at 0-2hr **(Figure 6C).** This observation indicates that ZLD may reduce chromatin accessibility at those sites, consistent with the previous finding (Schulz et al. 2015) that over one hundred genomic loci had increased chromatin accessibility in the absence of ZLD. Moreover, at the 2-4hr time point **(Figure 6D)**, sites where CLAMP is dependent on ZLD to bind had reduced accessibility in *zld-i* embryos. However, there are very few sites (n=30) in this class that CLAMP still depends on ZLD

to bind at 2-4hr. Overall, these results provide new insight into a specific subclass of genomic loci where ZLD reduces chromatin accessibility.

Lastly, we further tested our hypothesis that CLAMP and ZLD regulate transcription interdependently by comparing the genes regulated by each factor with each other. Therefore, we overlapped the down-regulated genes ($\log_2$ fold change < 0) from embryos in which either maternal *zld* (Combs and Eisen, 2017, GSE71137) or *clamp* (Rieder et al., 2017, GSE102922) was depleted. We identified a significant ($p < 0.05$, Hypergeometric test) overlap between genes that are downregulated in the absence of either CLAMP or ZLD at the 0-2hr time point **(Figure 6E).** In contrast, CLAMP and ZLD down-regulated genes do not have a significant overlap at the 2-4hr time point **(Figure 6F)**, consistent with the observation that CLAMP binds independently from ZLD at this later time point. However, over three-hundred genes require both factors for expression at 2-4 hrs **(Figure 6F)**. Therefore,CLAMP and ZLD significantly co-regulate transcription early in development and co-regulate several hundred genes during and after ZGA. Taken together, our results are consistent with the direct action of CLAMP and ZLD on chromatin accessibility that influences the binding of both TFs to their binding sites and regulates target gene expression.

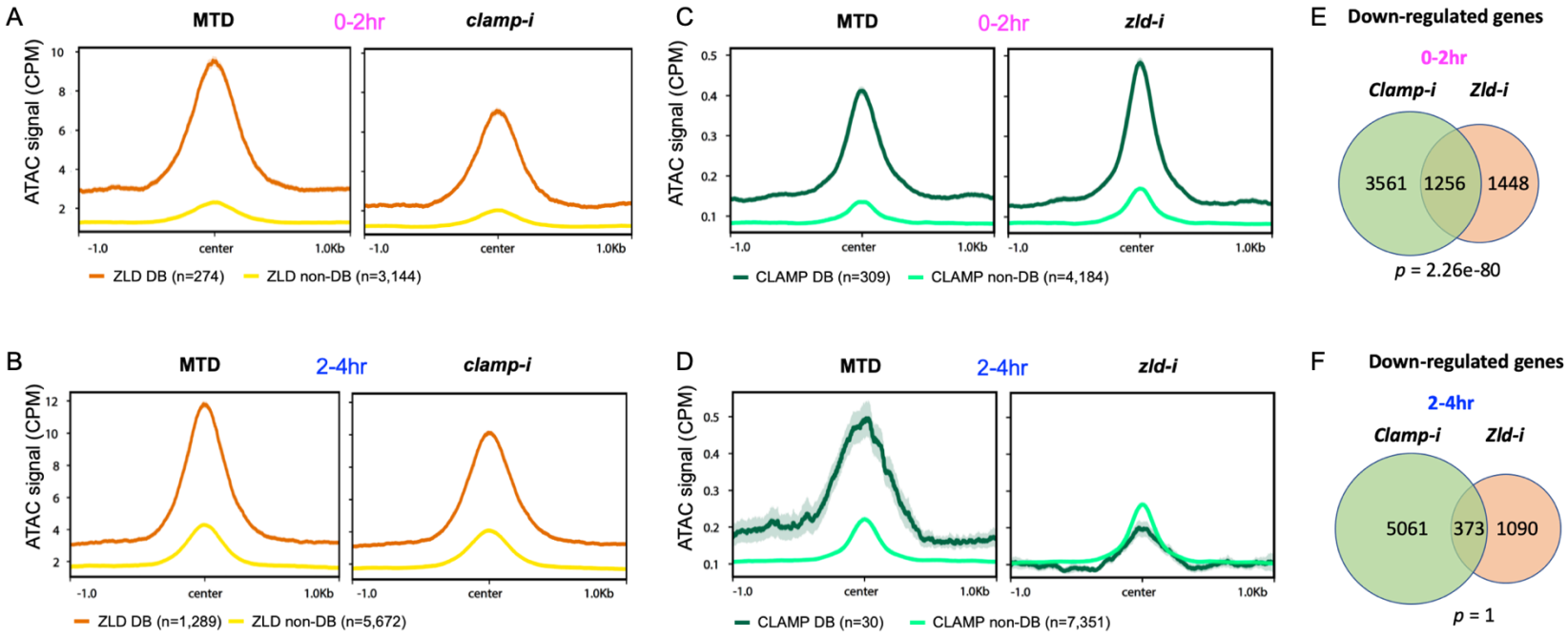

Figure 6. CLAMP and ZLD functioning together to mediate chromatin accessibility and zygotic genome activation

A. Average profiles show chromatin accessibility at ZLD-dependent and independent sites at the 0-2hr time point.

B. Average profiles show chromatin accessibility at ZLD-dependent and independent sites at the 2-4hr time point.

C. Average profiles show chromatin accessibility at CLAMP-dependent and independent sites at the 0-2hr time point.

D. Average profiles show chromatin accessibility at CLAMP-dependent and independent sites at the 2-4hr time point.

E. Down-regulated genes in the absence of CLAMP (*clamp-i*, green) and ZLD (*zld-i*, orange) and overlapped genes that are down-regulated after knockdown of both TFs at 0-2hr time points. *P*-value represents the significance (hypergeometric test) of their overlap.

F. Down-regulated genes in the absence of CLAMP (*clamp-i*, green) and ZLD (*zld-i*, orange) and overlapped genes that are down-regulated after knockdown of both TFs at 2-4hr time points. *P*-value represents the significance (hypergeometric test) of their overlap.

DISCUSSION

Two questions central to embryogenesis of all metazoans are how and where do early transcription factors work together to drive chromatin changes and zygotic genome activation. We identified CLAMP as a pioneer transcription factor that directly binds to nucleosomal DNA, regulates zygotic genome activation **(Figure 1)**, establishes and/or maintains chromatin accessibility **(Figure 2-3)**, and facilitates the binding of ZLD to promoters **(Figure 4)**. We further discovered that CLAMP and ZLD interdependently regulate each other's binding during ZGA **(Figure 5)**. Also, we identified the direct action of CLAMP and ZLD on chromatin accessibility that influences the binding of both TFs to their binding sites **(Figure 6)**. Overall, we identify a new pioneer TF and provide key insight into how CLAMP and ZLD function interdependently to remodel zygotic genome accessibility which drives zygotic genome activation.

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

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

McDaniel et al., 2019), we demonstrate that CLAMP also possesses these characteristics of a pioneer factor.

We defined four classes of CLAMP-related and ZLD-related peaks in early embryos, which reveal both interdependent and redundant roles of CLAMP and ZLD in defining chromatin accessibility during ZGA **(Figure 7)**. ZLD binding is enriched in both DA CLAMP-bound and non-DA CLAMP-bound groups, and CLAMP is required for ZLD binding at those two groups of sites. These results indicate that CLAMP could directly (DA) or indirectly (non-DA) mediate ZLD binding to DNA **(Figures 4F & 7)**.

In contrast, CLAMP binding is mainly enriched at non-DA ZLD-bound groups, even after ZLD depletion **(Figures 4G & 7).** Therefore, CLAMP functions redundantly with ZLD to maintain chromatin accessibility at regions that are bound by ZLD but that do not require ZLD for chromatin accessibility. TF redundancy is increased with organism complexity and it is a key to protect organisms from lethality that would be caused by loss of a single non-redundant factor (Rosanova et al., 2017). Moreover, a previous study found that GAF is also enriched at these non-DA ZLD-bound regions (Schulz et al., 2015). Both CLAMP and GAF are deposited maternally (Rieder et al., 2017; Hamm et al., 2017) and bind to similar GA-rich motifs (Kaye et al., 2018). Furthermore, Gaskill et al. (companion paper), demonstrate that GAF is also a key regulator of ZGA, but functions largely independent from ZLD unlike CLAMP.

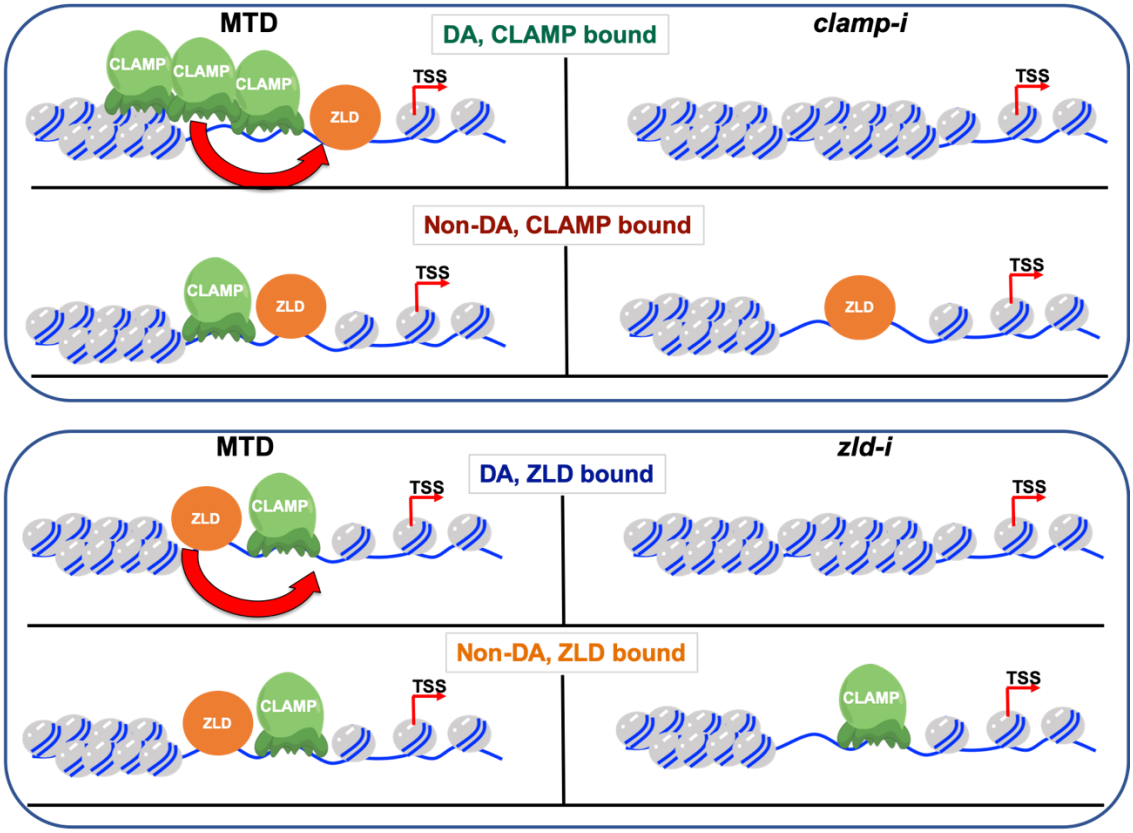

Figure 7. Model for integration of pioneer factor function in defining chromatin accessibility

ZLD binding regulates chromatin accessibility at early embryonic promoters and intronic regions, allowing CLAMP to access its binding sites. Also, CLAMP binding regulates chromatin accessibility at early embryonic promoters, allowing ZLD to access its binding sites. Therefore, CLAMP and ZLD are two pioneer TFs which function interdependently at promoters to open chromatin. At many loci, CLAMP and/or ZLD bind to chromatin but they are not required for accessibility. There is also functional redundancy between pioneer TFs because CLAMP can facilitate chromatin opening at sites bound by ZLD but which do not require ZLD for chromatin opening and vice versa.

Although we have demonstrated an instrumental role for CLAMP in defining open chromatin landscape in early embryos, our data show that CLAMP does not increase accessibility at all genomic loci. Therefore, other pioneer TFs, such as GAF, are likely to compensate for the depletion of CLAMP or ZLD. To test this hypothesis, we tried to perform GAF RNAi in the current study to prevent GAF from compensating for the loss of CLAMP. However, we could not achieve depletion of GAF in early embryos by RNAi, likely due to its prion-like self-perpetuating feature (Tariq et al., 2013). In the companion study, Gaskill et al. used a degron approach to deplete GAF and show that GAF is also critical for ZGA and functions independently of ZLD.

We previously demonstrated that competition between CLAMP and GAF at GA-rich binding sites is important for MSL complex recruitment in S2 cells (Kaye et al., 2018). However, we also observed interdependence between CLAMP and GAF at many additional binding sites not involved in MSL complex recruitment (Kaye et al., 2018). Yet, the relationship between CLAMP and GAF in early embryos remains unknown. It is very possible that the competitive relationship has not been established in early embryos, since dosage compensation has not yet been initiated (Prayitno et al., 2019). Moreover, the GA-rich sequences targeted by CLAMP and GAF are distinct *in vivo* and *in vitro*. GAF motifs (GAGAGAGAGA) show a uniform GA distribution with at least 5-bp of contiguous repeat, while CLAMP can bind to sequences that contain non-contiguous GA repeats (GA _ GAGA _)(Kaye et al., 2018). Therefore GAF and CLAMP may have overlapping and non-overlapping functions at different loci, tissues or developmental stages. In the future, an optogenetic inactivation approach could be used

to remove both CLAMP and GAF simultaneously in a spatial and temporal manner (McDaniel et al., 2019).

CLAMP and ZLD mediate each other's binding via their own motifs

ZLD is an essential TF that regulates activation of the very first set of zygotic genes during the minor wave of ZGA, as well as thousands of genes transcribed during the major wave of ZGA at NC14 (Liang et al., 2008; Harrison et al., 2011). ZLD also establishes and maintains chromatin accessibility of specific regions and facilitates transcription factor binding and early gene expression (Sun et al., 2015; Schulz et al., 2015). We previously demonstrated that maternally deposited CLAMP is also essential for early embryonic development (Rieder et al., 2017). CLAMP regulates histone gene expression (Rieder et al., 2017) and establishes/maintains chromatin accessibility at promoters genome-wide (J. Urban et al., 2017). Nonetheless, it remained unclear whether and how CLAMP and ZLD functionally interact during ZGA. Here, we demonstrate an interdependent relationship between CLAMP and ZLD at hundreds of promoters genome-wide. Furthermore, ZLD often regulates CLAMP earlier than CLAMP regulates ZLD occupancy.

Genomic loci at which CLAMP is dependent on ZLD early (0-2hr) in development often became ZLD-independent later (2-4hr) in development. Therefore, it is possible that CLAMP requires the pioneering activity of ZLD to access specific loci prior to ZGA, but ZLD is no longer required once the binding is established. Also, our results suggest

that CLAMP is a strong regulator ZLD binding, especially in 2-4hr embryos. Furthermore, ZLD is able to bind to many more promoter regions early in development, while CLAMP mainly binds to introns early in development but occupies promoters later in development. Therefore, CLAMP may require ZLD to open up the chromatin of these promoter regions (Schulz et al., 2015). In fact, ZLD is sufficient to activate a subgroup of early genes, although most ZLD bound regions are not active until NC14 (Bosch et al., 2006).

In addition to its role in early embryonic development, CLAMP also plays an essential role in targeting the MSL male dosage compensation complex to the X-chromosome (Soruco et al., 2013). *Drosophila* embryos initiate X chromosome counting in NC12 and start the sex determination cascade piror to the major wave of ZGA at NC14 (Gergen, 1987; Bosch et al., 2006). However, the majority of dosage compensation initiates much later in embryonic development (Prayitno et al., 2019). Therefore, our data support a model in which CLAMP functions early in the embryo prior to MSL complex assembly to open up specific chromatin regions for MSL complex recruitment later (J. Urban et al., 2017; Rieder et al., 2019). Moreover, ZLD likely functions primarily as an early pioneer factor whereas CLAMP has pioneering functions in both early and late embryos. Consistent with this hypothesis, CLAMP binding is enriched at early and late zygotic genes while ZLD binding is localized mainly to early zygotic genes, suggesting a sequential relationship between these two early TFs during ZGA.

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, suggesting that the required protein may multimerize at its binding sites. 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. Although the number of ZLD motifs is significantly enriched at dependent sites compared to independent sites, the motif count median in dependent sites was still close to one **(Figure S5E)**, further suggesting that ZLD 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. Mass spectrometry results of CLAMP-associated proteins did not reveal ZLD (J. A. Urban et al., 2017). Also, 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 in a sequential manner. Because ZGA is an essential process, it is key to have redundant TFs to protect organisms from lethality that would be caused by mutation of a single non-redundant factor.

MATERIALS AND METHODS

Recombinant Protein Expression and Purification of CLAMP

MBP-tagged CLAMP DBD was expressed and purified as described previously (Kaye et al., 2018). MBP-tagged (pTHMT, Peti and Page, 2007) full-length CLAMP protein was expressed in *Escherichia coli* BL21 Star (DE3) cells (Life Technologies). Bacterial cultures were grown to an optical density of 0.7 to 0.9 before induction with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) for 4 hrs at 37°C. Cell pellets were harvested by centrifugation and stored at -80°C. Cell pellets were resuspended in 20 mM Tris 1M NaCl 10 mM imidazole pH 8.0 with one EDTA-free protease inhibitor tablet (Roche) and lysed using an Emulsiflex C3 (Avestin). The lysate was cleared by centrifugation at 20,000 rpm for 50 min at 4°C, filtered using a 0.2 µm syringe filter, and loaded onto a HisTrap HP 5 mL column. The protein was eluted with a gradient from 10 to 300 mM imidazole in 20 mM Tris 1.0 M NaCl pH 8.0. Fractions containing MBP-CLAMP full-length were loaded onto a HiLoad 26/600 Superdex 200 pg column equilibrated in 20 mM Tris 1.0 M NaCl pH 8.0. Fractions containing full-length CLAMP were identified by SDS-PAGE and concentrated using a centrifugation filter with a 10 kDa cutoff (Amicon, Millipore) and frozen as aliquots.

In vitro assembly of nucleosome

The 240 bp 5C2 DNA fragment used for nucleosome *in vitro* assembly was amplified from 276 bp 5C2 fragments (50ng/ul, IDT gBlocks Gene Fragments) by PCR (see 276 bp 5C2 and primer sequences below) using OneTaq Hot Start 2X Master Mix (New

England Biolabs). The DNA was purified using the PCR clean-up kit (Qiagen) and concentrated to 1ug/ul by SpeedVac Vacuum (Eppendorf). The nucleosomes were assembled using the EpiMark® Nucleosome Assembly Kit (New England Biolabs) following the kit's protocol.

5C2 (276 bp), bold sequences are CLAMP-binding motifs, underlined sequences are primer binding sequences:

TCGACGACTAGTTTAAAGTTA<u>TTGTAGTTCTTAGAGCAGAATGT</u>ATTTTAAATATCAA
TGTTTCGATGTAGAAATTGAATGGTTTAAATCACGTTCACACAACTTA**GAAAGAGAT**
AGCGATGGCGGTGT**GAAAGAGAGCGAGATAG**TTGGAAGCTTCATG**GAAATGAAA**
GAGAGGTAGTTTTTGGAAATGAAAGTTGTACTAGAAATAAGTATTTTATGTATATAG
AATATCGAAGTACAG<u>AAATTCGAAGCGATCTCAAC</u>TTGAATATTATATCG

Primers (product is 240bp):

Forward: TTGTAGTTCTTAGAGCAGAATGT

Reverse: GTTGAGATCGCTTCGAATTT

Electromobility shift assays

DNA or nucleosome probe at 35nM (700fmol/reaction) was incubated with MBP-tagged CLAMP DBD protein or MBP-tagged full-length CLAMP protein in a binding buffer. The binding reaction buffer conditions are similar to conditions previously used to test ZLD nucleosome binding (McDaniel et al. 2019) in 20 ul total volume: 7.5ul BSA/HEGK buffer (12.5 mM HEPES, PH 7.0, 0.5 mM EDTA, 0.5 mM EGTA, 5% glycerol, 50 mM

KCl, 0.05 mg/ml BSA, 0.2 mM PMSF, 1 mM DTT, 0.25 mM ZnCl2, 0.006% NP-40,) 10 ul probe mix (5 ng poly[d-(IC)], 5 mM MgCl2, 700 fmol probe), and 2.5 ul protein dilution (0.5uM, 1uM, 2.5uM) at room temperature for 60 min. Reactions were loaded onto 6% DNA retardation gels (ThermoFisher) and run in 0.5X Tris–borate–EDTA buffer for 2 hours. Gels were post stained with Gelred Nucleic Acid Stain (Thermo Scientific) for 30 min and visualized using the ChemiDoc MP imaging system (BioRad).

Fly stocks and crosses

To deplete maternally deposited *clamp* or *zld* mRNA throughout oogenesis, we crossed a maternal triple driver (MTD-GAL4, Bloomington, #31777) line with a Transgenic RNAi Project (TRiP) *clamp* RNAi line (Bloomington, #57008) or a TRiP *zld* RNAi line (from C. Rushlow lab, Sun et al., 2015). We used the MTD-GAL4 line alone as the control line. We validated *clamp* or *zld* knockdown in early embryos by western blotting using the Western Breeze kit (Invitrogen) and qRT-PCR (Rieder et al., 2017).

ATAC-seq in embryos

We conducted ATAC-seq following the protocol from Blythe and Wieschaus (2016). 0-2hr or 2-4hr embryos were laid on grape agar plates, dechorionated them by 1 min exposure to 6% bleach (Clorox) and then washed them 3 times in deionized water. We homogenized 10 embryos and lysed them in 50 ul lysis buffer (10mM Tris 7.5, 10mM NaCl, 3mM MgCl2, 0.1% NP-40). We collected nuclei by centrifuging at 500 g at 4°C and resuspended nuclei in 5 ul TD buffer with 2.5 ul Tn5 enzyme (Illumina Tagment DNA TDE1 Enzyme and Buffer Kits). We incubated samples at 37°C for 30min at 800

rpm (Eppendorf Thermomixer) for fragmentation, and then purified samples with Qiagen Minelute columns before PCR amplification. We amplified libraries by adding 10 ul DNA to 25 ul NEBNext HiFi 2x PCR mix (New England Biolabs) and 2.5 ul of 25 uM each of Ad1 and Ad2 primers. We used 13 PCR cycles to amplify samples from 0-2hr embryos and 12 PCR cycles to amplify samples from 2-4hr embryos. Next, we purified libraries with 1.2x Ampure SPRI beads. We performed three biological replicates for each genotype (n=2) and time point (n=2). We measured the concentrations of 12 ATAC-seq libraries by Qubit and determined library quality by Bioanalyzer. We sequenced libraries on an Illumina Hi-seq 4000 sequencer at GeneWiz (South Plainfield, NJ) in 2x150-bp mode. ATAC-seq data is deposited at NCBI GEO and the accession number is GSE152596.

Chromatin Immunoprecipitation-sequencing (ChIP-seq)

We performed ChIP-seq as previously described (Blythe and Wieschaus, 2015). We collected and fixed 200-400 embryos from each MTD-GAL4 and RNAi cross 0-2hr or 2-4hr after fertilization. We used 3 ul of rabbit anti-CLAMP (Soruco et al., 2013) and 2 ul rat anti-ZLD (from C. Rushlow lab) per sample. We performed three biological ChIP replicates for each protein (n=2), genotype (n=3) and time point (n=2). In total, we prepared 36 libraries using the NEBNext Ultra ChIP-seq kit (New England Biolabs) and sequenced libraries on Illumina HiSeq 2500 sequencer in 2x150-bp mode. ChIP-seq data is deposited at NCBI GEO and the accession number is GSE152598.

Computational analyses

ATAC-seq analysis

Demultiplexed reads were trimmed of adapters using TrimGalore (Krueger, 2017) and mapped to the *Drosophila* genome dm6 version using Bowtie2 (v. 2.3.0) with option -X 2000. We used Picard tools (v. 2.9.2) and SAMtools (v.1.9, Li et al., 2009) to remove the reads that were unmapped, failed primary alignment, or duplicated (-F 1804), and retain properly paired reads (-f 2) with MAPQ >30. Peak regions for accessible regions were called using HMMRATAC (v1.2.10, Tarbell and Liu, 2019). ENCODE blacklist was used to filter out problematic regions in dm6 (Amemiya et al., 2019). DiffBind with the DESeq2 method (v. 3.10, Stark and Brown, 2019) was used to identify differentially accessible regions. We used DeepTools (version 3.1.0, Ramírez et al., 2014) and Homer (v 4.11, Givler and Lilienthal, 2005) to generate enrichment heatmaps (CPM normalization), average profiles, motif searches, peak overlap and peak annotation. Visualizations and statistical tests were conducted in R (R Core Team, 2014).

ChIP-seq analysis

We trimmed ChIP sequencing raw reads with Trim galore (v. 0.5.0, Krueger, 2017) with a minimal phred score of 20, 36 bp minimal read length and Illumina adaptor removal. We then mapped cleaned reads to the *D. melanogaster* genome (UCSC dm6) with Bowtie2 (v. 2.3.0) with the --very-sensitive-local flag feature. We used Picard tools (v. 2.9.2) and SAMtools (v.1.9, (Li et al., 2009) to remove the PCR duplicates. We used MACS2 (version 2.1.1, Zhang et al., 2008) to identify peaks with default parameters and MSPC (v.4.0.0, Jalili et al., 2015) to obtain consensus peaks from 3 replicates. ENCODE blacklist was used to filter out problematic regions in dm6 (Amemiya et al.,

2019). We identified differential binding and non-differential binding using DiffBind with the DESeq2 method (v. 3.10, Stark and Brown, 2019). We used DeepTools (version 3.1.0, Ramírez et al., 2014) and Homer (v 4.11, Givler and Lilienthal, 2005) to generate enrichment heatmaps (SES normalization), average profiles, motif searches, peak overlap and peak annotation. Visualizations and statistical tests were conducted in R (R Core Team, 2014).

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 mutations were from GSE71137 (Combs and Eisen, 2017). ATAC-seq data from wild type and *zld* germline mutations were from GSE86966 (Hannon et al., 2017).

ACKNOWLEDGEMENTS

This work was supported by NIH grant F32GM109663 and K99HD092625 to Dr. Leila Rieder and R35GM126994 to Dr. Erica Larschan, and in part by NSF grant 1845734 and NIGMS grant GM118530 (to N. L. F).

AUTHOR CONTRIBUTIONS

Conceptualization, L.E.R., J.E.D. and E.N.L.; Methodology, J.E.D., L.E.R. and E.N.L; ChIP-seq Experiment, L.E.R.; ATAC-seq Experiment, J.E.D.; Initial Analysis, W.J.III; Formal Analysis, J.E.D.; Protein Expression: M.M., S.W., and N.L.F.; Gel-shift

Experiment, A.H and J.E.D.; Investigation, J.E.D; Data Curation, J.E.D; Writing--Original Draft, J.E.D. and E.N.L.; Writing--Review & Editing, J.E.D., L.E.R. and E.N.L.; Visualization, J.E.D. and L.E.R.; Funding Acquisition, L.E.R. and E.N.L.

DECLARATION OF INTERESTS

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. 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

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

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

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

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

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

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

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

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

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

Rosanova A, Colliva A, Osella M, Caselle M. 2017. Modelling the evolution of transcription factor binding preferences in complex eukaryotes. *Sci Rep* **7**. doi:10.1038/s41598-017-07761-0

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

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

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

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

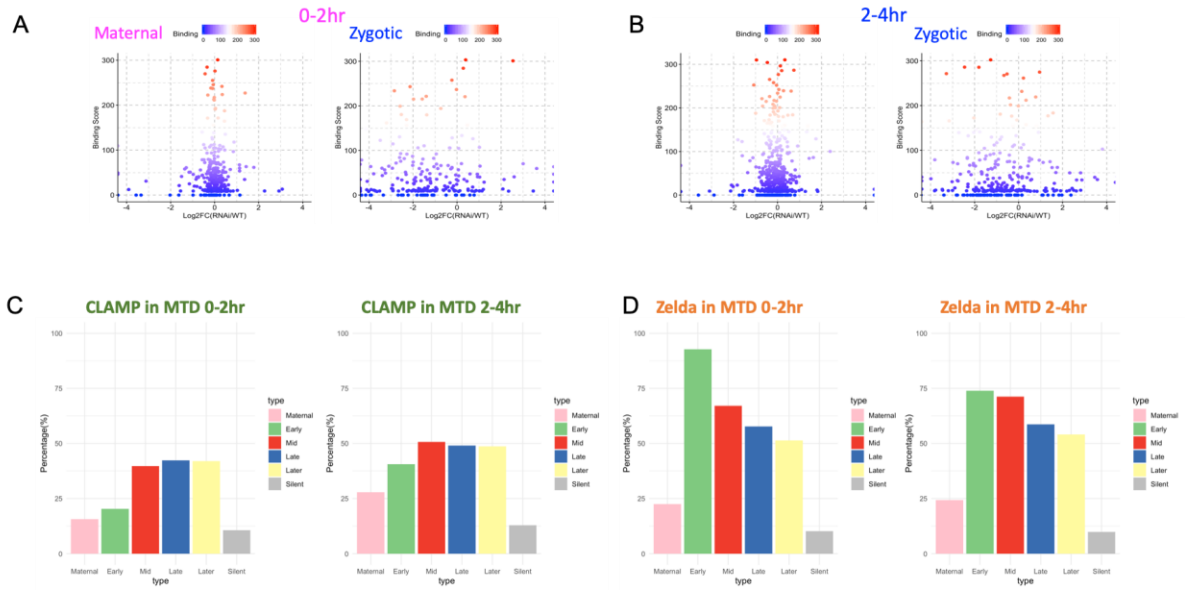

Figure S1. CLAMP and ZLD both activate transcription of zygotic genes

A-B. Effect of maternal CLAMP depletion and CLAMP binding on maternally deposited (left) and zygotically transcribed (right) gene expression: log2 (clamp-i/MTD) in 0-2hr and 2-4hr embryos. Gene categories were defined in Lott et al. (2011).

C-D. Percentage of CLAMP and ZLD binding sites 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 1kb promoter region and gene body). Gene categories were defined in Li et al. (2014).

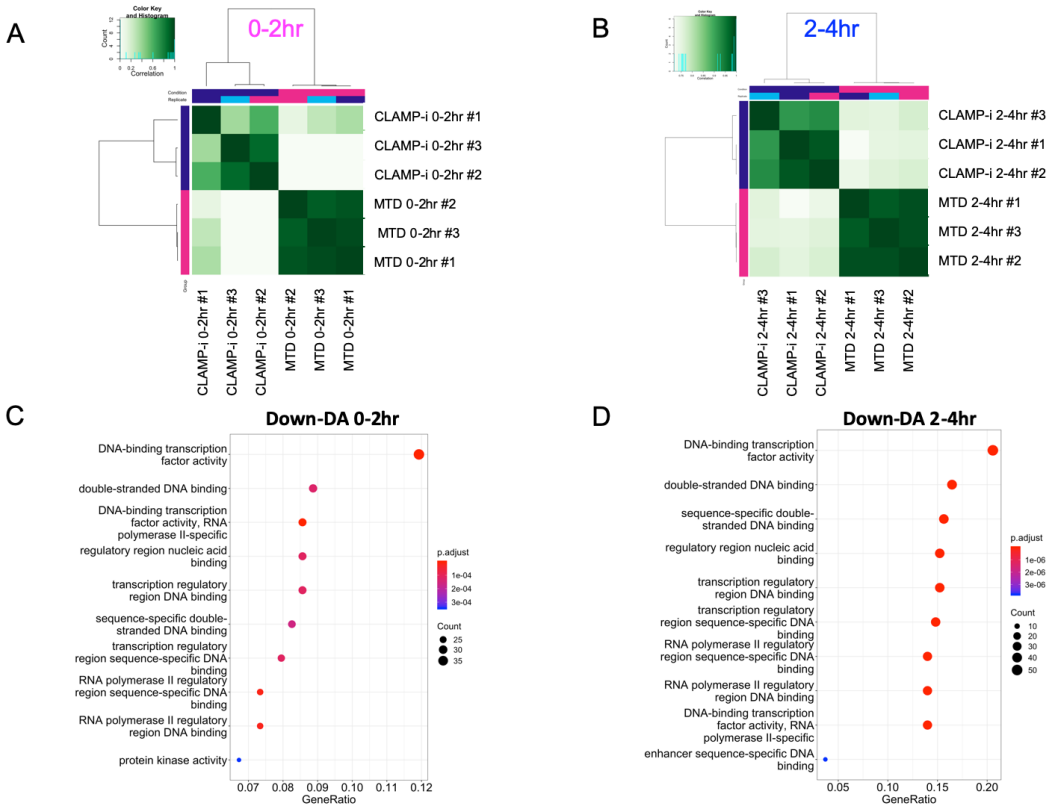

Figure S2. CLAMP regulates chromatin accessibility throughout ZGA

A-B. Pearson correlation of DA calls among replicates of peaks 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.

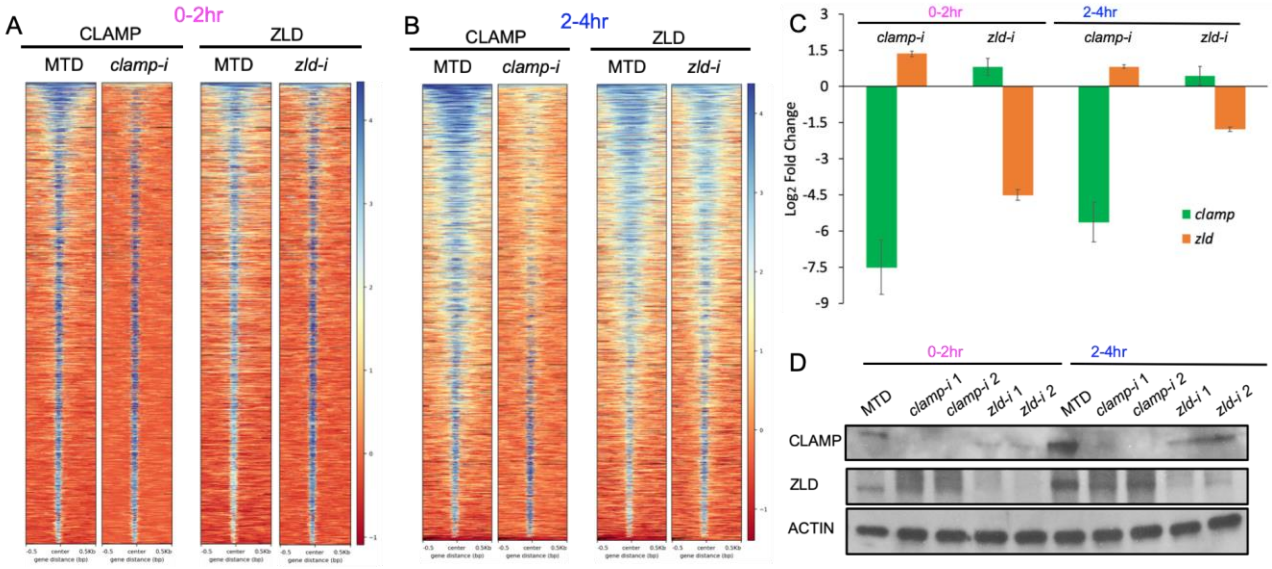

Figure S3. CLAMP-mediated chromatin accessibility is correlated with early CLAMP-dependent gene expression

A-B. Heatmap of 1 kb regions centered on CLAMP and/or ZLD ChIP-seq peaks at 0-2 and 2-4 hour time points. Blue represents above and red represents below background enrichment. Color key values represent the log2 ratio between peak vs. background normalized ChIP-seq signal.

C. Expression of mRNAs in MTD, *clamp-i* and *zld-i* embryos in 0-2hr and 2-4hr embryos. mRNA levels of *clamp* and *zld* were quantified by qRT-PCR. Log_2 Fold Change was calculated using the ΔΔCt method (Rao et al., 2013) and normalized to reference gene *pka*.

D. 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.

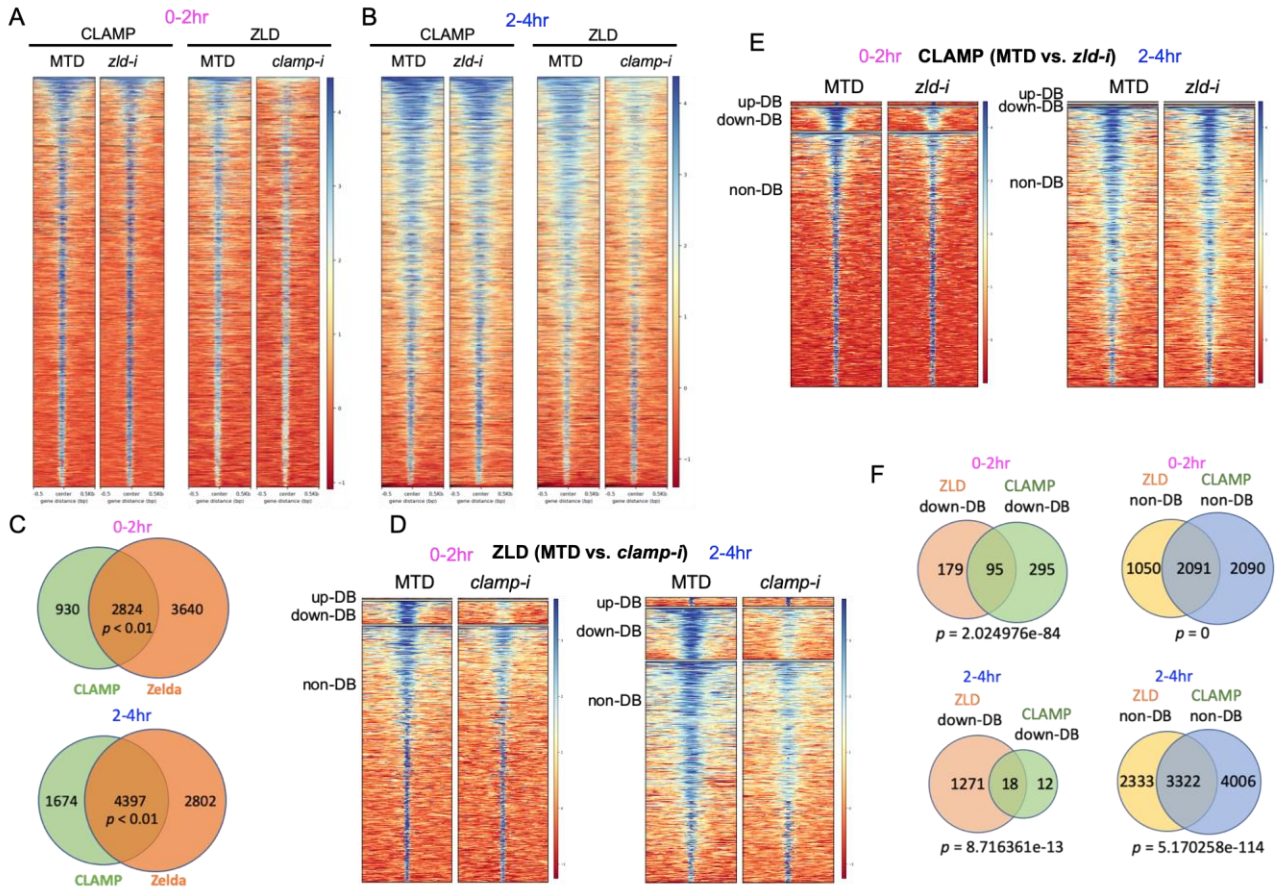

Figure S4. CLAMP and ZLD depend on each other for chromatin binding

A-B. Heatmaps of 1 kb regions centered on CLAMP and/or ZLD peaks in 0-2 and 2-4 hour samples. Blue represents above and red represents below background enrichment. Color key values represent the log2 ratio between peak vs. background normalized ChIP-seq signal.
C. CLAMP (green) and ZLD (orange) peaks and shared peaks where both CLAMP and ZLD are present in 0-2hr and 2-4hr embryos. *P*-values represent the significance (hypergeometric test) of overlap.
D. Heat maps of the ZLD peak enrichment from MTD embryos versus *clamp-i* embryos in 0-2hr (left) or 2-4hr (right). Blue represents above and red represents below background enrichment. Color key values represent the log2 ratio between peak vs. background.
E. Heat maps of the CLAMP peak enrichment from MTD embryos versus *zld-i* embryos in 0-2hr (left) or 2-4hr (right). Blue represents above and red represents below background enrichment. Color key values represent the log2 ratio between peak vs. background.
F. Venn diagram showing the number of overlap sites between ZLD and CLAMP down-DB or ZLD and CLAMP non-DB. *P*-values represent the significance (hypergeometric test) of overlap.

MTD: MTD-Gal4 line. *clamp-i:* MTD-Gal4-*clamp* mRNAi line, *zld-i:* MTD-Gal4-*zld* mRNAi line.

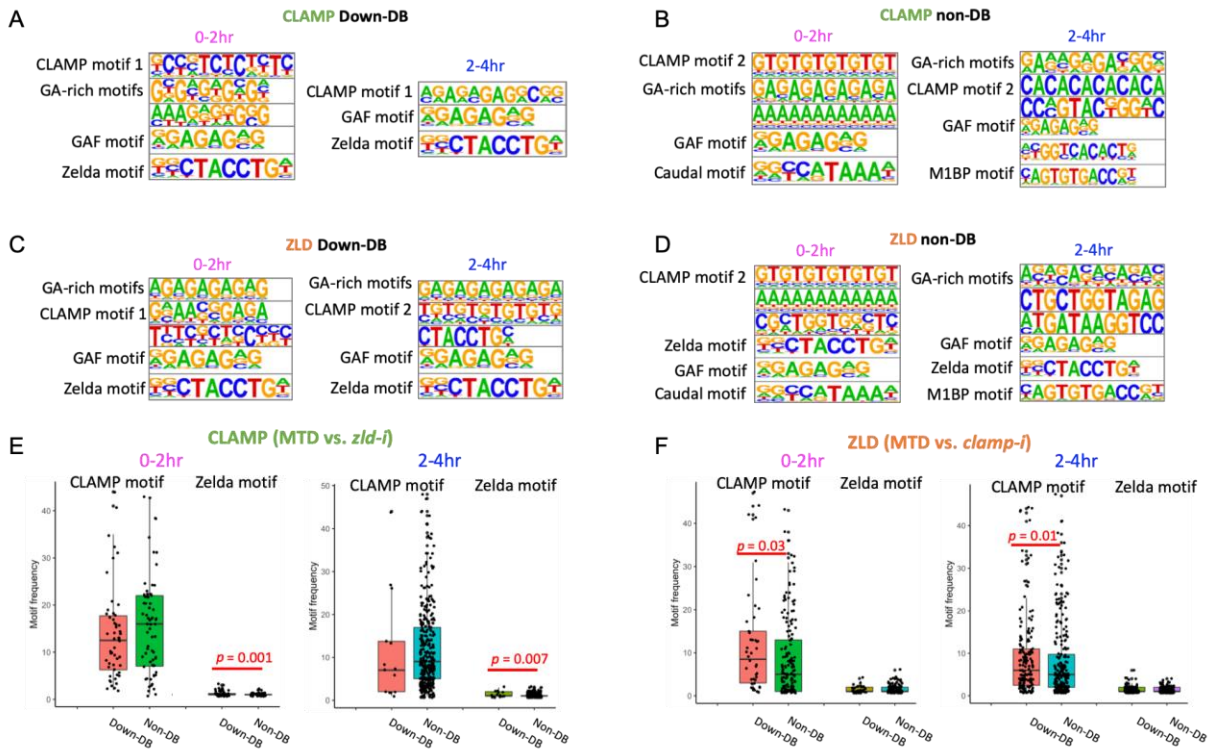

Figure S5. Transcription of dependent sites is regulated by the required protein via motifs.

A. Top *de novo* and known motifs for CLAMP DB peaks.
B. Top *de novo* and known motifs for non-DB peaks.
C. Top *de novo* and known motifs for ZLD DB peaks.
D. Top *de novo* and known motifs for ZLD non-DB peaks.

CLAMP motif 1: CLAMP unique motif; CLAMP motif 2: GA motifs that are recognized by CLAMP or GAF; unannotated motifs are novel/unknown motifs.

E. Boxplots for CLAMP or ZLD motif enrichment at CLAMP DB and non-DB peaks for MTD vs. *zld-i* embryos at 0-2hr and 2-4hr time points.

F. Boxplots for CLAMP or ZLD motif enrichment at ZLD DB and non-DB peaks of MTD vs. *clamp-i* embryos at 0-2hr and 2-4hr time points.