The pioneer factor GAF is essential for zygotic genome activation and chromatin accessibility in the early *Drosophila* embryo

Marissa M. Gaskill*, Tyler J. Gibson*, Elizabeth D. Larson, and Melissa M. Harrison

*These authors contributed equally to this work

Department of Biomolecular Chemistry, University of Wisconsin School of Medicine and Public Health, Madison, WI

Abstract
Following fertilization, the genomes of the germ cells are reprogrammed to form the totipotent embryo. Pioneer transcription factors are required for remodeling the chromatin and driving the initial wave of zygotic gene expression. In *Drosophila melanogaster*, the pioneer factor Zelda is essential for development through this dramatic period of reprogramming, known as the maternal-to-zygotic transition (MZT). However, it was unknown whether additional pioneer factors were necessary for this transition. We identified an additional maternally encoded factor required for development through the MZT, GAGA Factor (GAF). GAF is needed to activate widespread zygotic transcription and to remodel the chromatin accessibility landscape. We demonstrated that Zelda preferentially controls expression of the earliest transcribed genes, while genes expressed during widespread activation are predominantly dependent on GAF. Thus, progression through the MZT requires coordination of multiple pioneer factors, and we propose that as development proceeds transcriptional control is gradually transferred from Zelda to GAF.

Introduction

Pronounced changes in cellular identity are driven by pioneer transcription factors that act at the top of gene regulatory networks. While nucleosomes present a barrier to the DNA binding of many transcription factors, pioneer factors can bind DNA in the context of nucleosomes. Pioneer-factor binding establishes accessible chromatin domains, which serve to recruit additional transcription factors that drive gene expression (Zaret and Mango 2016; Iwafuchi-Doi and Zaret 2014; Zaret and Carroll 2011). These unique characteristic of pioneer factors enable them to facilitate widespread changes in cell identity. Nonetheless, cell-fate transitions often require a combination of pioneering transcription factors to act in concert to drive the necessary transcriptional programs. Indeed, the reprogramming of a specified cell type to an induced pluripotent stem cell requires a cocktail of transcription factors, of which Oct4, Sox2, and Klf4 function as pioneer factors (Takahashi and Yamanaka 2006; Takahashi et al. 2007; Chronis et al. 2017; Soufi et al. 2012, 2015). Despite the many examples of multiple pioneer factors functioning together to drive reprogramming, how these factors coordinate gene expression changes within the context of organismal development remains unclear.

Pioneer factors are also essential for the reprogramming that occurs in the early embryo. Following fertilization, specified germ cells must be rapidly and efficiently reprogrammed to generate a totipotent embryo capable of differentiating into all the cell types of the adult organism. This reprogramming is initially driven by mRNAs and proteins that are maternally deposited into the oocyte. During this time, the zygotic genome remains transcriptionally quiescent. Only after cells have been reprogrammed is the zygotic genome gradually activated. This maternal-to-zygotic transition (MZT) is broadly conserved among metazoans and essential for future development (Schulz and Harrison 2019; Vastenhouw et al. 2019). Activators of the zygotic genome have been identified in a number of species (zebrafish – Pou4f3, Sox19b, Nanog; mice – DUX4, NFY; humans – DUX4, OCT4; fruit flies – Zelda), and all share essential features of pioneer factors (Schulz and Harrison 2019; Vastenhouw et al. 2019).

Early *Drosophila* development is characterized by thirteen, rapid, synchronous nuclear divisions. Zygotic transcription gradually becomes activated starting about the eighth nuclear division, and widespread transcription occurs at the fourteenth nuclear division when the division cycle slows and the nuclei are cellularized (Schulz and Harrison 2019; Vastenhouw et al. 2019). The transcription factor Zelda (Zld) is required for transcription of hundreds of genes throughout zygotic genome activation (ZGA) and was the first identified global genomic activator (Liang et al. 2008; Harrison et al. 2011; Nien et al. 2011). Embryos lacking maternally encoded Zld die before completing the MZT (Liang et al. 2008; Harrison et al. 2011; Nien et al. 2011; Fu et al. 2014; Staudt et al. 2006). Zld has the defining features of a pioneer transcription factor: it binds to nucleosomal DNA (McDaniel et al. 2019), facilitates chromatin accessibility (Schulz et al. 2015; Sun et al. 2015) and this leads to subsequent binding of additional transcription factors (Yáñez-Cuna et al. 2012; Xu et al. 2014; Foo et al. 2014).

By contrast to the essential role for Zld in flies, no single global activator of zygotic transcription has been identified in other species. Instead multiple transcription factors function together to activate zygotic transcription (Schulz and Harrison 2019; Vastenhouw et al. 2019). Work from our lab and others has implicated additional factors in regulating reprogramming in the *Drosophila* embryo. Specifically, the enrichment of GA-dinucleotides in regions of the genome that remain accessible in the absence of Zld and at loci that gain accessibility late in ZGA, suggest that a protein that binds to these regions functions with Zld to define cis-regulatory regions during the initial stages of development (Schulz et al. 2015; Sun et al. 2015; Blythe and Wieschaus 2016). Two proteins, GAGA Factor (GAF) and Chromatin-linked adaptor for MSL proteins (CLAMP), are known to bind to GA-dinucleotide repeats and are expressed in the early embryo, implicating one

or both proteins in reprogramming the embryonic transcriptome (Rieder et al. 2017; Soruco et al. 2013; Kuzu et al. 2016; Biggin and Tjian 1988; Bhat et al. 1996; Soeller et al. 1993).

CLAMP was first identified based on its role in targeting the dosage compensation machinery, and it preferentially localizes to the X chromosome (Larschan et al. 2012; Soruco et al. 2013). GAF, encoded by the *Trithorax-like* (*Trl*) gene, has broad roles in transcriptional regulation, including functioning as a transcriptional activator (Farkas et al. 1994; Bhat et al. 1996), repressor (Mishra et al. 2001; Busturia et al. 2001; Bernués et al. 2007; Horard et al. 2000) and insulator (Ohtsuki and Levine 1998; Wolle et al. 2015; Kaye et al. 2017). Through interactions with chromatin remodelers, GAF is instrumental in driving regions of accessible chromatin both at promoters and distal *cis*-regulatory regions (Okada and Hirose 1998; Tsukiyama et al. 1994; Xiao et al. 2001; Tsukiyama and Wu 1995; Fuda et al. 2015; Judd et al. 2020). Analysis of hypomorphic alleles has suggested an important function for GAF in the early embryo in both driving expression of *Ubx*, *Abd-B*, *en*, and *ftz* and in maintaining normal embryonic development (Farkas et al. 1994; Bhat et al. 1996). Given these diverse functions for GAF, and that it shares many properties of a pioneer transcription factor, we sought to investigate whether it has a global role in reprogramming the zygotic genome for transcriptional activation.

Investigation of the role of GAF in the early embryo necessitated the development of a system to robustly eliminate GAF, which had not been possible since GAF is essential for maintenance of the maternal germline and is resistant to RNAi knockdown in the embryo (Bhat et al. 1996; Bejarano and Busturia 2004; Rieder et al. 2017). For this purpose, we generated endogenously GFP-tagged GAF, which provided the essential functions of the untagged protein, and used the deGradFP system to deplete GFP-tagged GAF in early embryos expressing only the tagged construct (Caussinus et al. 2012). Using this system, we identified an essential function for GAF in driving chromatin accessibility and gene expression during the MZT. Thus, at least two pioneer transcription factors, Zld and GAF, must cooperate to reprogram the zygotic genome of *Drosophila* following fertilization.

Results

GAF binds the same loci throughout ZGA

To investigate the role of GAF during the MZT, we used Cas9-mediated genome engineering to tag the endogenous protein with super folder Green Fluorescent Protein (sfGFP) at either the N- or C-termini, sfGFP-GAF(N) and GAF-sfGFP(C), respectively. There are two protein isoforms of GAF (Benyajati et al. 1997). Because the N-terminus of GAF is shared by both isoforms (Supplemental Fig. S1A), the sfGFP tag on the N-terminus labels all GAF protein. By contrast, the two reported isoforms differ in their C-termini, and thus the C-terminal sfGFP labels only the short isoform (Supplemental Fig. S1B). Whereas null mutants in *Trithorax-like* (*Trl*), the gene encoding GAF, are lethal (Farkas et al. 1994), both sfGFP-tagged lines are homozygous viable and fertile. Additionally, in embryos from both lines sfGFP-labelled GAF is localized to discrete nuclear puncta and is retained on the mitotic chromosomes in a pattern that recapitulates what has been previously described for GAF based on antibody staining in fixed embryos (Fig. 1A, S1C,D) (Raff et al. 1994). Together, these data demonstrate that the sfGFP tag does not interfere with essential GAF function and localization.

To begin to elucidate the role of GAF during early embryogenesis, we determined the genomic regions occupied by GAF during the MZT. We hand sorted homozygous *GAF-sfGFP(C)* stage 3 and stage 5 embryos and performed chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) using an anti-GFP antibody. While alternative splicing generates two GAF protein isoforms that differ in their C-terminal polyQ domain, in the early embryo only the short isoform is detectable (Benyajati et al. 1997). We confirmed the expression of the short isoform in the early embryo by blotting extract from 0-4 hrs AEL (after egg laying) N- and C-terminally tagged GAF embryos with an anti-GFP antibody (Supplemental Fig. S1E). The long isoform was undetectable in extract from embryos of both sfGFP-tagged lines harvested 0-4 hrs AEL, but was detectable at low levels in extract from *sfGFP-GAF(N)* embryos harvested 13-16 hrs AEL. Thus, we conclude that during the MZT the C-terminally sfGFP-labelled short isoform comprises the overwhelming majority of GAF present in the *GAF-sfGFP(C)* embryos. We identified 3,391 GAF peaks at stage 3 and 4,175 GAF peaks at stage 5 (Supplemental Table S1). To control for possible cross-reactivity with the anti-GFP antibody, we performed ChIP-seq on w^{1118} stage 3 and stage 5 embryos in parallel. No peaks were called in the w^{1118} dataset for either stage, confirming the specificity of the peaks identified in the *GAF-sfGFP(C)* embryos (Supplemental Fig. S2A,B). Further supporting the specificity of the ChIP data, the canonical GA-rich GAF-binding motif was the most highly enriched sequence identified in ChIP peaks from both stages, and these motifs were centrally located in the peaks (Fig. 1B). Peaks were identified in the regulatory regions of several previously identified GAF-target genes, including the *heat shock* promoter (*hsp70)*, *Ultrabithorax (Ubx)*, *engrailed (en)*, *even-skipped (eve),* and *Kruppel (Kr)* (Fig. 1C) (Biggin and Tjian 1988; Gilmour et al. 1989; Soeller et al. 1988; Lee et al. 1992; Read et al. 1990; Kerrigan et al. 1991). Peaks were enriched at promoters, which fits with the previously defined role of GAF in establishing paused RNA polymerase (Supplemental Fig. S2C) (Lee et al. 2008; Fuda et al. 2015)

There was a substantial degree of overlap between GAF peaks identified at both stage 3 and stage 5. 2,955 peaks were shared between the two time points, representing 87% of total stage 3 peaks and 71% of total stage 5 peaks (Fig. 1C, D, S2D). This demonstrates that GAF binding is established prior to widespread ZGA and remains relatively unchanged during early development, similar to what has been shown for Zld, the major activator of the zygotic genome, (Harrison et al. 2011). We compared GAF-binding sites we identified in the early embryo to previously identified GAF-bound regions in S2 cells, which are derived from 20-24-hour old embryos (Fuda et al. 2015). Despite the difference in cell-type and antibody used, 56% (1,906) of stage 3 peaks and 43% (1,794) of stage 5 peaks overlapped peaks identified in S2 cells (Fig. 1C, D, Supplemental Fig. S2D). It was previously noted that GAF binding in 8-16 hr embryos was

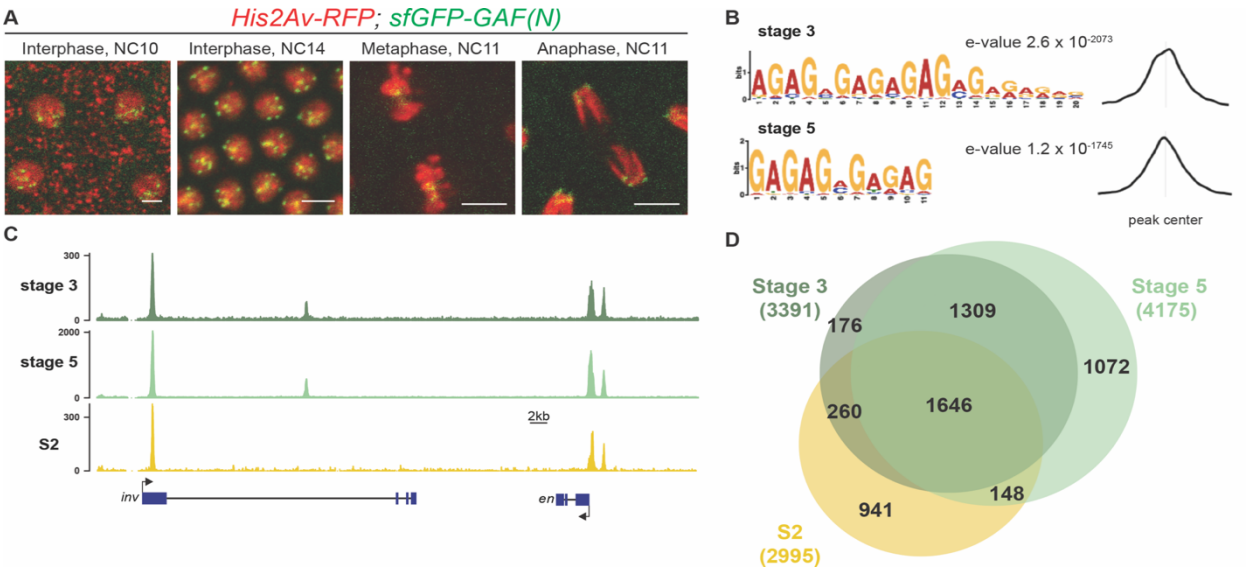

Figure 1: GAF binds thousands of loci throughout the MZT. A. Images of *His2Av-RFP; sfGFP-GAF(N)* embryos at the nuclear cycles (NC) indicated above. sfGFP-GAF(N) localizes to puncta during interphase and is retained on chromosome during mitosis. Scale bars, 5 µm. B. Binding motif enrichment of GA-dinucleotide repeats at GAF peaks identified at stage 3 and stage 5 determined by MEME-suite (*left*). Distribution of the GA-repeat motif within peaks (*right*). Gray line indicates peak center. C. Representative genome browser tracks of ChIP-seq peaks for GAF-sfGFP(C) from stage 3 and stage 5 embryos and GAF ChIP-seq from S2 cells (Fuda et al. 2015). D. Venn diagram of the peak overlap for GAF as determined by ChIP-seq for GAF-sfGFP(C) from sorted stage 3 embryos and stage 5 embryos and by ChIP-seq for GAF from S2 cells (Fuda et al. 2015). Total number of peaks identified at each stage is indicated in parentheses.

highly similar to GAF occupancy in the wing imaginal disc harvested from the larva (Slattery et al. 2014), and our data indicate that this binding is established early in development prior to activation of the zygotic genome.

Maternal GAF is required for embryogenesis and progression through the MZT

Early embryonic GAF is maternally deposited in the oocyte, and this maternally deposited mRNA can sustain development until the third instar larval stage (Farkas et al. 1994). To investigate the role of GAF during the MZT therefore necessitated a system to eliminate this maternally encoded GAF. RNAi failed to successfully knockdown GAF in the embryo (Rieder et al. 2017), and female germline clones cannot be generated as GAF is required for egg production (Bhat et al. 1996; Bejarano and Busturia 2004). To overcome these challenges, we leveraged our N-terminal sfGFP-tagged allele and the previously developed deGradFP system to target knockdown at the protein level (Caussinus et al. 2012). The deGradFP system uses a genomically encoded F-box protein fused to a nanobody recognizing GFP, which recruits GFP-tagged proteins to a ubiquitin ligase complex. Ubiquitination of the GFP-tagged protein subsequently leads to efficient degradation by the proteasome. To adapt this system for efficient use in the early embryo, we generated transgenic flies in which the deGradFP nanobody fusion was driven by the *nanos (nos)* promoter for strong expression in the embryo 0-2 hrs after fertilization (Wang and Lehmann 1991). In embryos laid by *nos-deGradFP; sfGFP-GAF(N)* females all maternally encoded GAF protein is tagged with GFP and thus subject to degradation by the deGradFP nanobody fusion. These embryos are hereafter referred to as $GAF^{deGradFP}$.

We verified the efficiency of the deGradFP system by imaging living embryos in which nuclei were marked by His2Av-RFP. $GAF^{deGradFP}$ embryos lack the punctate, nuclear GFP signal identified in

control embryos that do not carry the deGradFP nanobody fusion, indicating efficient depletion of sfGFP-GAF(N) (Fig. 2A). This knockdown was robust, as we failed to identify any NC10-NC14 embryo with nuclear GFP signal, and none of the embryos carrying both the deGradFP nanobody and the sfGFP-tagged GAF hatched (Fig. 2B). Based on live embryo imaging, the majority of embryos died prior to NC14, indicating that maternal GAF is essential for progression through the MZT. We identified a small number of GFP-expressing, gastrulating escapers. It is unclear if these embryos had an incomplete knockdown of maternally encoded sfGFP-GAF(N), or if a small percentage of embryos survived until gastrulation in the absence of GAF and that the GFP signal was the result of zygotic gene expression. Nonetheless, none of these embryos survived until hatching. Despite being able to maintain a strain homozygous of the N-terminal, sfGFP-tagged GAF, quantitative analysis revealed an effect on viability. Embryos homozygous for *sfGFP-GAF(N)* had only a 30% hatching rate (Fig. 2B). Therefore, all future experiments controlled for the effect of the N-terminal tag by using these homozygous embryos as paired controls with GAF^{deGradFP} embryos.

Having identified a dramatic effect of eliminating maternally expressed GAF in the early embryo, we used live imaging to investigate the developmental defects in our GAF^{deGradFP} embryos. GAF^{deGradFP} embryos in which nuclei were marked by a fluorescently labelled histone (His2Av-RFP) were imaged through several rounds of mitosis. We observed defects such as asynchronous mitosis, anaphase bridges, disordered nuclei, and nuclear dropout (Fig. 2C, D) (Supplemental Video 1,2). Nevertheless, embryos were able to complete several rounds of mitosis without GAF before arresting. Nuclear defects became more pronounced as GAF^{deGradFP} embryos developed and approached NC14. The arrested/dead embryos were often arrested in mitosis or had large, irregular nuclei (Fig. 2C), similar to the nuclear "supernovas" in GAF deficient nuclei reported in Bhat et al, 1996. Our live imaging allowed us to detect an additional nuclear defect: that in the absence of maternal GAF nuclei become highly mobile and, in some cases, adopt a "swirling" pattern (Supplemental Video 2). As a control, we imaged *sfGFP-GAF(N)* homozygous embryos through several rounds of mitosis. These embryos proceeded normally through NC10-14, demonstrating that the mitotic defects in the GAF^{deGradFP} embryos were caused by the absence of GAF (Fig. 2D) (Supplemental Video 3). The defects in GAF^{deGradFP} embryos are similar to those previously reported for maternal depletion of *zld* (Liang et al. 2008; Staudt et al. 2006) and identify a fundamental role for GAF during the MZT.

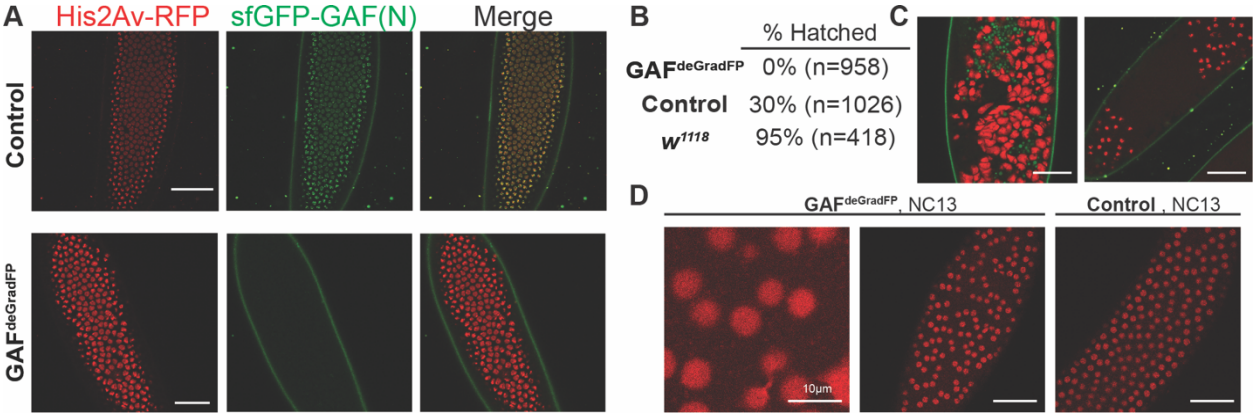

Figure 2: Embryos lacking maternal GAF die during early embryogenesis with nuclear and mitotic defects. A. Images of control (maternal genotype: *His2Av-RFP; sfGFP-GAF(N)*) and GAF^{deGradFP} (maternal genotype: *His2Av-RFP/nos-deGradFP; sfGFP-GAF(N)*) embryos at NC14, demonstrating loss of nuclear GFP signal specifically in GAF^{deGradFP} embryos. His2Av-RFP marks the nuclei. B. Hatching rates after >24 hours for control, GAF^{deGradFP}, and *w^{1118}* embryos. C. Confocal images of His2Av-RFP in arrested/dying GAF^{deGradFP} embryos with blocky nuclei, mitotic arrest, and nuclear fallout. D. Confocal images of His2Av-RFP in NC13 control and GAF^{deGradFP} embryos. Scale bars, 50µm except where indicated.

GAF is required for the activation of hundreds of zygotic genes

During the MZT, there is a dramatic change in the embryonic transcriptome as developmental control shifts from mother to offspring. Having demonstrated that maternal GAF was essential for development during the MZT, we investigated the role of GAF in regulating these transcriptional changes. We performed total-RNA seq on bulk collections of $GAF^{deGradFP}$ and control (*sfGFP-GAF(N)*) embryos harvested 2-2.5 hrs AEL, during the beginning of NC14 when the widespread genome activation has initiated. Our replicates were reproducible (Supplemental Fig. S3A), allowing us to identify 1,452 transcripts that were misexpressed in $GAF^{deGradFP}$ embryos as compared to controls. Importantly, by using *sfGFP-GAF(N)* homozygous embryos as a control we have excluded from our analysis any transcripts misexpressed as a result of the sfGFP tag on GAF. Of the misexpressed transcripts 884 were down-regulated and 568 were up-regulated in the absence of GAF (Fig. 3A). The gene encoding GAF, *Trithorax-like,* was named because it was required for expression of the homeotic genes *Ubx* and *Abd-B* (Farkas et al. 1994). Our RNA-seq analysis identified 7 of the 8 *Drosophila* homeotic genes (*Ubx*, *Abd-B*, *adb-A*, *pb*, *Dfd*, *Scr*, and *Antp*) down-regulated in GAF^{deGrad} embryos. Additionally, many of the gap genes, essential regulators of anterior-posterior patterning, are down-regulated: *giant* (*gt*), *knirps* (*kni*), *huckebein* (*hkb*), *Krüppel* (*Kr*), and *tailless* (*tll*) (Supplemental Table S2). Gene ontology (GO)-term analysis showed down-regulated genes were enriched for functions in system development and developmental processes as would be expected for essential genes activated during ZGA (Supplemental Fig. S3B). GO-term analysis of the up-regulated transcripts showed weak enrichment for response to stimulus and metabolic processes (Supplemental Fig. S3C).

To determine whether GAF is functioning predominantly in transcriptional activation or repression, we used our stage 5 ChIP data to determine the likely direct targets of GAF. We found that 45% (397) of the down-regulated transcripts were proximal to a GAF peak. By contrast, only 17% (99) of up-regulated transcripts are near GAF peaks (Fig. 3A). The significant enrichment for GAF-binding sites proximal to down-regulated transcripts as compared to up-regulated ($p < 2.2 \times 10^{-16}$, two-tailed Fisher's exact test) supports a role for GAF specifically in transcriptional activation. Because zygotic gene expression is required for degradation of a subset of maternal mRNAs, if the genome is not activated maternal transcripts are not properly degraded and therefore are increased in RNA-seq data (Harrison et al. 2011; Hamm et al. 2017; Liang et al. 2008). Indeed, down-regulated transcripts were enriched for zygotically expressed genes, and up-regulated transcripts were largely maternally contributed ($p < 2.2 \times 10^{-16}$, two-tailed Fisher's exact test; Fig. 3B). Thus, GAF is essential for transcriptional activation during the MZT and likely functions along with Zld to drive zygotic genome activation.

Previous data defined a role for the additional GA-dinucleotide binding protein, CLAMP, in activating zygotic gene expression (Larschan et al. 2012; Urban et al. 2017b; Rieder et al. 2017; Soruco et al. 2013). Therefore, we compared transcripts down-regulated in $GAF^{deGradFP}$ embryos to transcripts down-regulated in *clamp*-RNAi embryos 2-4 hrs AEL (Rieder et al. 2017). 174 transcripts are down-regulated in both datasets, comprising 19.7% of total down-regulated GAF targets and 50.1% of total down-regulated CLAMP targets (Supplemental Fig. S4A). While this demonstrates that a subset of transcripts requires both GAF and CLAMP for proper expression during ZGA, a majority of GAF-regulated transcripts only depend on GAF for expression, independent of CLAMP. GAF and CLAMP have similar, but not identical binding preferences (Kaye et al. 2018), which is reflected in their partially, but not completely overlapping genome occupancy (Figure S4B). These binding site differences likely explain their differential requirement during ZGA.

Having identified that GAF was required for ZGA, we wanted to ensure that the effects were not due to changes in the levels of Zld, the previously identified activator of the zygotic genome. Immunoblots for Zld on extract from GAF^{deGradFP} and control (*sfGFP-GAF(N)*) 2-2.5 hrs AEL (stage 5) embryos confirmed that Zld levels were consistent between extracts (Supplemental Fig. S4C). Therefore, the effects of GAF on genome activation are not due to a loss of Zld. Based on the roles for Zld and GAF in activating the zygotic genome, we investigated whether these proteins were required to activate distinct or overlapping target genes. We compared transcripts down-regulated in GAF^{deGradFP} embryos to genes down-regulated when Zld was inactivated

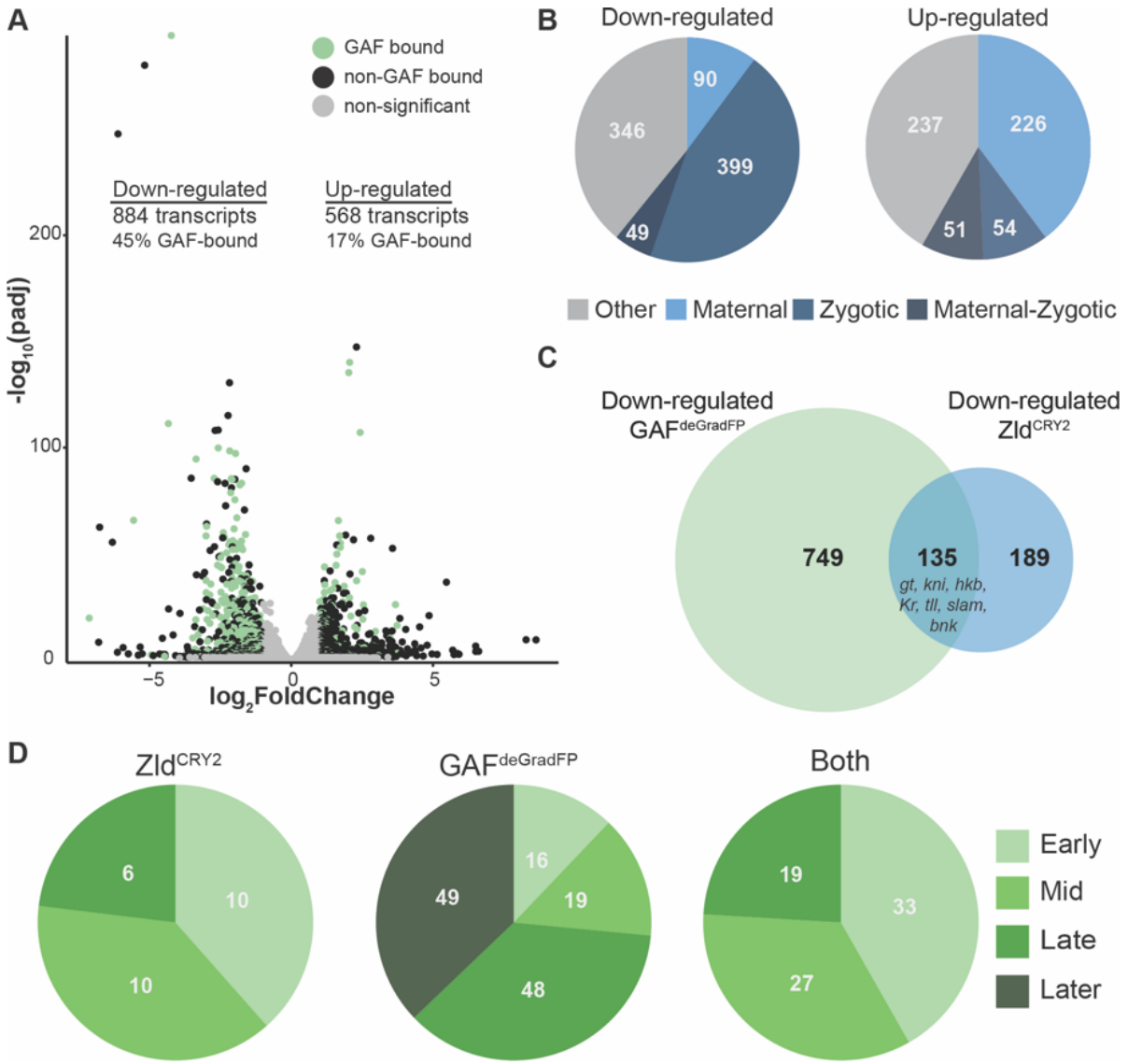

Figure 3: GAF is required for zygotic genome activation. A. Volcano plot of transcripts mis-expressed in GAF^{deGradFP} embryos as compared to *sfGFP-GAF(N)* controls. Stage 5 GAF-sfGFP(C) ChIP-seq was used to identify GAF-bound target genes. B. The percentage of up-regulated and down-regulated transcripts in GAF^{deGradFP} embryos classified as maternal, zygotic, or maternal-zygotic based on Lott et al. 2011. C. Overlap of down-regulated embryonic transcripts in the absence of GAF or Zld activity ($p < 2.2 \times 10^{-16}$, two-tailed Fisher's exact test). Down-regulated genes for Zld are from McDaniel et al., 2019. D. Transcripts down-regulated in both GAF^{deGradFP} and Zld^{CRY2} embryos or in either condition alone were classified based on temporal expression during ZGA (Li et al. 2014). Early = NC10-11, Mid = NC12-13, Late = early NC14, Later = late NC14. Only genes that were assigned to one of the four classes are shown.

optogenetically throughout zygotic genome activation (NC10-14)(McDaniel et al. 2019). We identified 135 genes down-regulated in both datasets, comprising 42% of the total number of down-regulated genes dependent on Zld and 15% of the total down-regulated transcripts dependent on GAF (Fig. 3C). An even lower degree of overlap is observed when only direct targets are considered with 49 down-regulated targets shared between Zld and GAF, which have 232 and 397 direct targets, respectively. By contrast only 29 up-regulated genes were shared between the two datasets (Supplemental Fig. S4D). Genes that required both factors for activation include the gap genes previously mentioned (*gt, kni, hkb, Kr, tll*) as well as genes involved in cellular blastoderm formation such as *slow as molasses* (*slam*) and *bottleneck* (*bnk*). While Zld and GAF share some targets, they each are required for expression of hundreds of individual genes.

Activation of the zygotic genome is a gradual process that initiates with transcription of a small number of genes around cycle 8. Transcripts can therefore be divided into categories based on the timing of their initial transcription (Li et al. 2014; Lott et al. 2011). Previous data suggested that while Zld was required for activation of genes throughout the MZT, early genes were particularly sensitive to loss of Zld and that GAF might be functioning later (Schulz et al. 2015; Blythe and Wieschaus 2016). To determine when GAF-dependent genes were expressed during ZGA, we took all of the genes that could be classified based on their timing of activation during the MZT (as determined in Li et al. 2014) and divided them based on their dependence on Zld and GAF for activation: those down-regulated in both $GAF^{deGradFP}$ and Zld^{CRY2} embryos (Both), those down-regulated only in Zld^{CRY2} embryos (Zld^{CRY2}), and those down-regulated only in $GAF^{deGradFP}$ embryos ($GAF^{deGradFP}$) (Fig. 3D). Genes activated by Zld, both those regulated by Zld alone (Zld^{CRY2}) and those regulated by both Zld and GAF (Both) were enriched for genes expressed early (NC10-11) and Mid (NC12-13). By contrast, 73% of genes activated by GAF alone ($GAF^{deGradFP}$) were expressed either late (early NC14) or later (late NC14) ($p < 2 \times 10^{-6}$, two-tailed Fisher's exact test). This analysis supports a model in which GAF and Zld are essential activators for the initial wave of ZGA and that GAF has an additional role, independent of Zld, in activating widespread transcription during NC14.

GAF and Zld bind the genome independently

To further delineate the relationship between Zld- and GAF-mediated transcriptional activation during the MZT, we determined whether GAF binding was dependent on Zld. Not only are a large subset of genes dependent on both Zld and GAF for wild-type levels of expression (Fig. 3C), but 42% of GAF-binding sites at stage 5 are also occupied by Zld (Harrison et al. 2011) (Fig. 4A,B). The GAF peaks that are co-bound with Zld include many of the highest GAF peaks identified at stage 5 (Supplemental Fig. S4E). Zld is known to facilitate the binding of multiple different transcription factors (Twist, Dorsal, and Bicoid), likely by forming dynamic subnuclear hubs (Yáñez-Cuna et al. 2012; Xu et al. 2014; Foo et al. 2014; Mir et al. 2018; Dufourt et al. 2018; Yamada et al. 2019). We therefore examined GAF localization upon depletion of maternal *zld* using RNAi driven in the maternal germline by *matα*-GAL4-VP16 in a background containing sfGFP-GAF(N) (Sun et al. 2015). Immunoblot confirmed a nearly complete knockdown of Zld (Fig. 4C), and we verified that RNAi-treated embryos failed to hatch. To assess the depletion of Zld on the subnuclear localization of GAF, we imaged *sfGFP-GAF(N); zld*-RNAi and control embryos using the identical acquisition settings. We observed no difference in puncta formation of GAF at the beginning of ZGA (NC10) or late ZGA (NC14) (Fig. 4D). We conclude that Zld is not required for GAF to form subnuclear puncta during the MZT.

To more specifically determine the impact of loss of Zld on GAF chromatin occupancy, we performed ChIP-seq with an anti-GFP antibody on embryos expressing *zld*-RNAi and sfGFP-GAF(N) and compared the occupancy to our ChIP-seq data from *GAF-sfGFP(C)* embryos. The

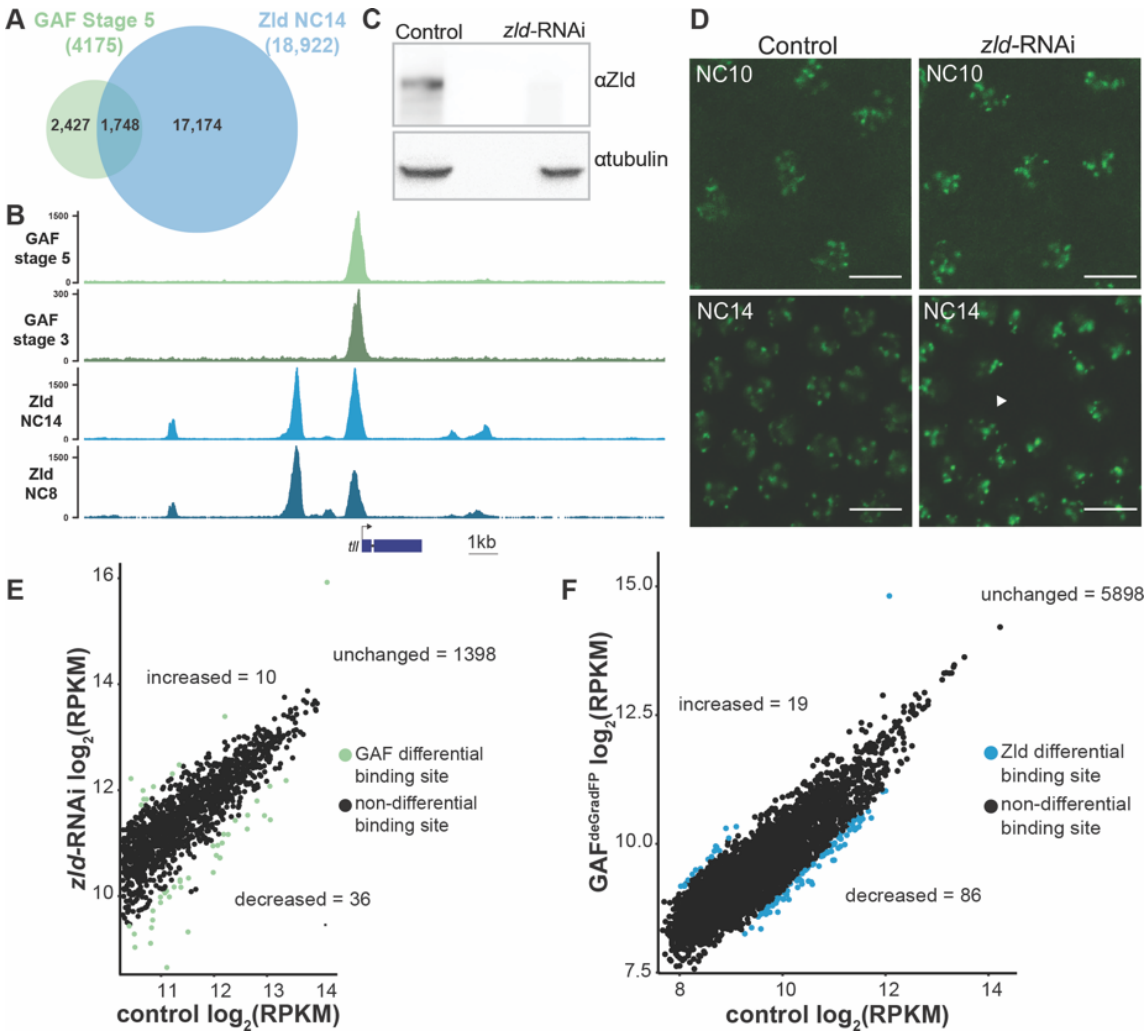

Figure 4: GAF and Zld bind chromatin independently. A. Overlap of Zld- and GAF-binding sites determined by GAF-sfGFP(C) stage 5 ChIP-seq and Zld NC14 ChIP-seq (Harrison et al., 2011). B. Representative genome browser tracks of Zld and GAF ChIP-seq peaks at the *tailless* locus. C. Immunoblot for Zld on embryo extracts from *zld*-RNAi; *sfGFP-GAF(N)* and control embryos harvested 2-3 hrs AEL. αtubulin was used as a loading control. D. Images of *zld*-RNAi; *sfGFP-GAF(N)* embryos and control *sfGFP-GAF(N)* embryos at NC10 and NC14 as marked. Arrowhead shows nuclear dropout, a phenotype indicative of *zld* loss-of-function. Scale bar, 10 µm. E. Correlation between $\log_2$(RPKM) of ChIP peaks for GAF from *GAF-sfGFP(C)* stage 5 embryos (control) and *zld*-RNAi; *sfGFP-GAF(N)* embryos (*zld*-RNAi). Peaks that differ between the two datasets are indicated in green. F. Correlation between $\log_2$(RPKM) of ChIP peaks for Zld from *sfGFP-GAF(N)* embryos (control) and $GAF^{deGradFP}$ embryos fixed 2-2.5 hrs AEL. Peaks that differ between the two datasets are indicated in blue.

majority of peaks called in the GAF-sfGFP(C) dataset are maintained in the *zld*-RNAi dataset (Supplemental Fig. S5A). Nonetheless, the global peak strength is lower in the *zld*-RNAi background (Supplemental Fig. S5A). While we cannot rule out that *zld* knock down causes a global reduction in GAF binding, it is possible that the overall decrease in peak height is due to technical, rather than biological differences (see Materials and Methods). To begin to distinguish between these possibilities, we determined whether the overall GAF occupancy was similar between the two data sets. For this purpose, we ranked peaks in each dataset based on the number of reads per kilobase per million mapped reads (RPKM) and determined if the peak RPKM values were correlated between the two sets. We identified a high degree of correlation in peak rank between two datasets (Pearson correlation, r = 0.89; Supplemental Fig. S5B), indicating that

the strongest peaks in the GAF-sfGFP(C) dataset are also the strongest peaks in the *zld*-RNAi dataset. To identify any specific peaks that differed between the two conditions we confined our analysis to the top third of the GAF peaks (1444 peaks), restricting analysis to only high-confidence peaks and limiting differences that might be due to variability in detection of low-occupancy peaks. We identified only 46 differential GAF peaks between *zld*-RNAi and control embryos: 36 peaks were decreased (21, Zld-bound) and 10 peaks were increased (5, Zld-bound) (Fig. 4E). Thus, very few specific GAF peaks are changed upon *zld*-RNAi. While we cannot rule out the possibility that in the RNAi embryos residual Zld remains at levels that we cannot detect by immunoblot, the high degree of rank correlation between the datasets and the low number of peaks that are significantly different between the RNAi and control datasets together suggest that GAF does not broadly depend on Zld for chromatin occupancy in the early embryo.

To assess the impact of the loss of GAF on Zld binding, we completed the reciprocal experiment and performed ChIP-seq for Zld on GAF^{deGradFP} and control (*sfGFP-GAF(N)*) embryos at 2-2.5 hrs AEL (stage 5). Similar to GAF binding when Zld is depleted, there was a global decrease in ChIP signal (Supplemental Fig. S5C), but a high degree of correlation of Zld peak rank in GAF^{deGradFP} and control embryos (Pearson correlation, r = 0.85) (Supplemental Fig. S5D). To establish a set of high-confidence peaks, we overlapped the peaks from the control embryos to the previously published data for Zld at NC14 and identified a set of 6,003 shared peaks that we further analyzed (Harrison et al. 2011). Among these high-confidence peaks, DESeq identified 105 Zld-binding sites that differed in occupancy between GAF^{deGradFP} embryos and controls: 86 peaks were decreased (1, GAF-bound) and 19 peaks increased (5, GAF-bound) (Fig. 4F). Globally, there was not a substantial effect of GAF depletion on Zld binding at GAF and Zld co-bound sites, as would be expected if the two factors were binding cooperatively (Supplemental Fig. S5C). We conclude that though some individual Zld-binding sites are affected by the loss of GAF, Zld binding to target loci is largely unperturbed in the absence of GAF. Altogether, our data indicate the depletion of either GAF or Zld does not specifically influence binding of the other factor to the vast majority of target sites. Thus, although GAF and Zld are co-bound at many sites during ZGA, these two factors do not depend on each other for chromatin occupancy.

GAF is essential for accessibility at hundreds of loci during ZGA

GAF interacts with chromatin remodelers and is required to maintain chromatin accessibility in tissue culture (Okada and Hirose 1998; Tsukiyama et al. 1994; Tsukiyama and Wu 1995; Xiao et al. 2001; Judd et al. 2020; Fuda et al. 2015). Furthermore, GAF-binding motifs are enriched at regions of open chromatin that are established at NC13 and that do not depend on Zld for accessibility (Schulz et al. 2015; Sun et al. 2015; Blythe and Wieschaus 2016). To directly test the function of GAF in determining accessible chromatin domains during the MZT, we performed the assay for transposase-accessible chromatin (ATAC)-seq on nuclei isolated from a bulk collection of 150-200 GAF^{deGradFP} and control (*sfGFP-GAF(N)*) embryos at 2-2.5 hrs AEL. In contrast to our control, replicates from the GAF^{deGradFP} nuclei showed variability, suggesting that developmental defects caused by the lack of GAF might have lowered our ability to detect subtle changes in accessibility. Nonetheless, we identified 509 regions with significant changes in accessibility in GAF^{deGradFP} embryos as compared to controls (Fig. 5A, Supplemental Table S3), and these regions were among those with the highest ATAC-seq signal (Supplemental Fig. S6A). 97% (494) of these differentially accessible regions lose accessibility in the absence of GAF, and of these, 28% (137) overlap with a GAF-binding site (Fig. 5A). Consistent with the enrichment of GAF-binding sites in promoters, half of all regions that depend on GAF for accessibility are in promoters (Fig. 5B). These regions that depend on GAF for accessibility are associated with genes that have decreased expression in the absence of GAF (Supplemental Fig. S6B), indicating that GAF-mediated chromatin accessibility promotes gene expression. Motif analysis identified binding motifs for both Zld and GAF underlying those regions that depended on GAF for

accessibility (Fig. 5C). Those regions that were directly bound by GAF, as determined by ChIP-seq, were preferentially enriched for GAF-binding motifs (Fig. 5D). Regions that lost accessibility but were not bound by GAF were enriched for the both the GAF- and Zld-binding motifs (Fig. 5E). Thus, GAF is essential for driving chromatin accessibility at hundreds of regions during ZGA.

Previous work from our lab and others suggested that during ZGA GAF may be responsible for maintaining chromatin accessibility at Zld-bound regions in the absence of Zld (Schulz et al. 2015; Sun et al. 2015). To directly test this hypothesis, we divided up the GAF-dependent accessible regions based on the classes of Zld-dependent and Zld-independent accessible regions previously identified in embryos depleted for maternal *zld* (Schulz et al. 2015). Regions bound by Zld in wild-type embryos that maintained accessibility in the absence of Zld were classified as constitutive Zld-bound regions and were previously shown to be enriched for GA-dinucleotide motifs (Schulz et al. 2015). These constitutive, Zld-bound regions accounted for 48% (245) of the 509 differentially accessible regions in GAF^{deGradFP} embryos (Fig. 5F), supporting the model that GAF maintains accessibility at these regions in the absence of Zld. Similar results were found when comparisons were made to another dataset analyzing chromatin accessibility in the presence and absence of Zld (Supplemental Fig. S6C) (Hannon et al. 2017). A majority of GAF-dependent regions do not depend on Zld for accessibility (Fig. 5F), demonstrating that Zld and GAF are independently required for chromatin accessibility at hundreds of loci during the MZT.

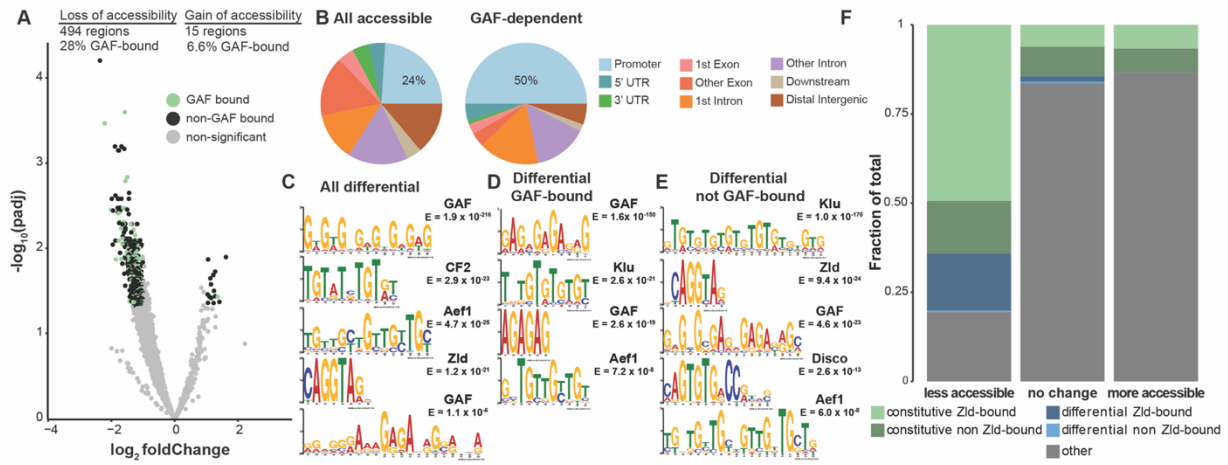

Figure 5: GAF is required for chromatin accessibility. A. Volcano plot of regions that change in accessibility in GAF^{deGradFP} embryos as compared to *sfGFP-GAF(N)* controls, stage 5 GAF-sfGFP(C) ChIP-seq was used to identify GAF-bound target regions. B. Genomic distribution of all accessible regions identified by ATAC-seq (all accessible) and genomic distribution of regions that are dependent on GAF for accessibility (GAF-dependent). C-E. Motif analysis using MEME-suite (*left*). Predicted factor binding to the enriched motif (*right*). C. Motifs enriched in all regions that depend on GAF for accessibility. D. Motifs enriched in GAF-bound regions that depend on GAF for accessibility. E. Motifs enriched at all non-GAF bound regions that depend on GAF for accessibility. F. Accessible regions were classified based on their change in GAF^{deGradFP} embryos (less accessible, more accessible, no change). The fraction of each GAF class that belong to specific sets of accessible regions as defined in Schulz et al., 2015 are shown. Constitutive, Zld-bound regions are bound by Zld at stage 5, but do not depend on this Zld binding for chromatin accessibility. Differential, Zld-bound regions are bound by Zld and lose accessibility in the absence of Zld. Differential, non Zld-bound regions also lose accessibility in the absence of Zld, but are not bound by Zld at stage 5. Other indicates regions that are accessible, but were not in the top 5000 accessible regions identified by FAIRE-seq in Schulz et al., 2015.

Discussion

Through a combination of Cas9-genome engineering and the deGradFP system, we have depleted maternally encoded GAF and demonstrated that GAF is required for progression through the MZT. Along with Zld, GAF is broadly required for activation of the zygotic genome and for shaping chromatin accessibility. During the major wave of ZGA, when thousands of genes are transcribed, Zld and GAF are largely independently required for both transcriptional activation and chromatin accessibility (Fig. 6). Thus, in *Drosophila*, as in mice, zebrafish, and humans, transcriptional activation during the MZT is driven by multiple factors with pioneering characteristics (Schulz and Harrison 2019; Vastenhouw et al. 2019). Together our system has enabled us to begin to determine how these two powerful factors collaborate to ensure the rapid, efficient transition from specified germ cells to a pluripotent cell population.

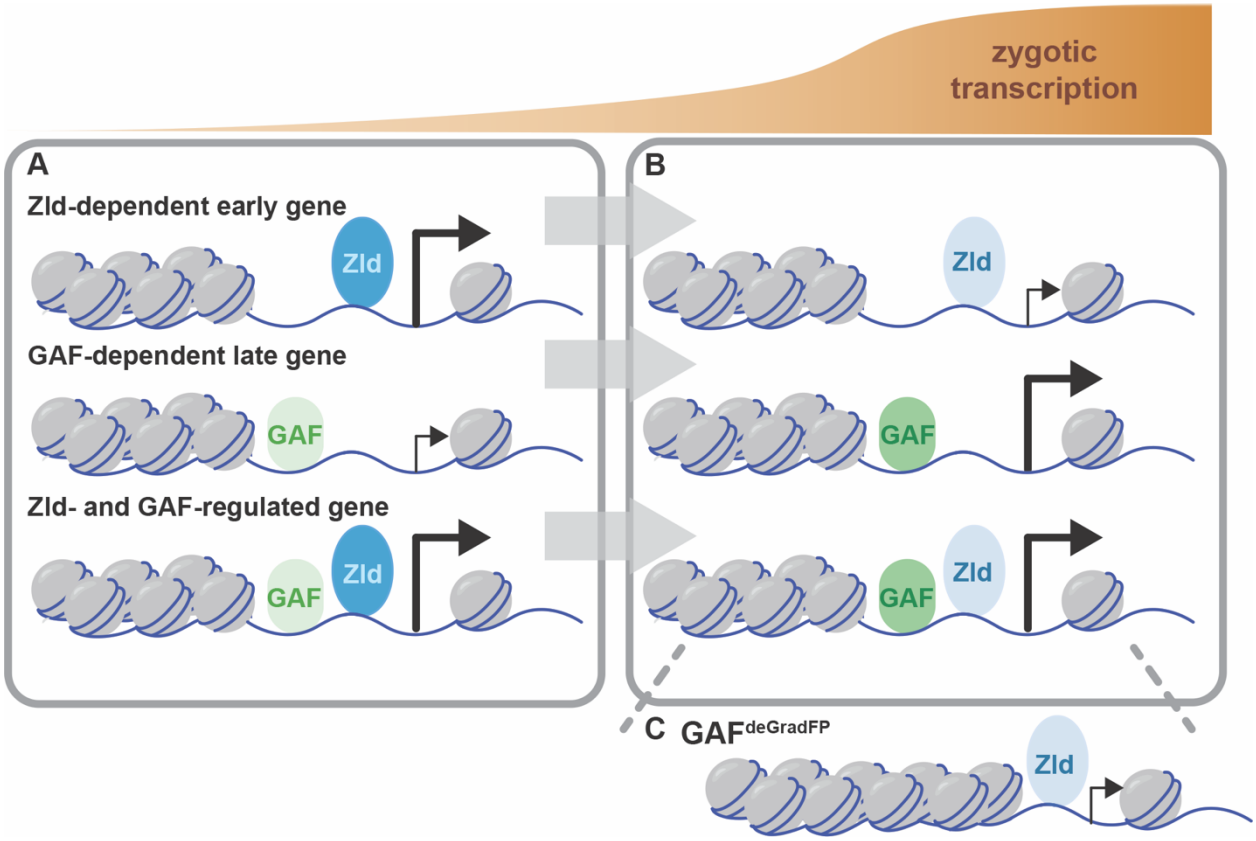

Figure 6: Zld and GAF independently regulate embryonic reprogramming. A. During the minor wave of ZGA (NC10-13), Zld is the predominant factor required for driving expression of genes bound by Zld alone and genes bound by both Zld and GAF. B. As the genome is more broadly activated during NC14, GAF becomes the major factor in driving zygotic transcription. C. GAF maintains chromatin accessibility at many Zld-bound regions. In GAF^{deGradFP} embryos accessibility is lost, but Zld binding remains unchanged.

Maternally encoded GAF is essential for development beyond the MZT

Using the deGradFP system, we demonstrated that maternally encoded GAF is required for embryogenesis. In contrast to zygotic GAF null embryos, which survive until the third instar larval stage, GAF^{deGradFP} embryos do not hatch (Farkas et al. 1994). Our imaging showed that depletion of GAF in the early embryo resulted in severe defects, including asynchronous mitosis, anaphase

bridges, nuclear dropout, and high nuclear mobility (Fig. 2C,D, Supplemental Video 1,2). Similar defects are seen when *zld* is maternally depleted from embryos (Liang et al. 2008; Staudt et al. 2006), and like embryos lacking maternal *zld,* GAF^{deGradFP} embryos died before the completion of the MZT. We identified defects in a subset of GAF^{deGradFP} embryos as early as NC10, which is consistent with our genomics data indicating that GAF is required for proper gene expression of some of the earliest genes expressed during ZGA (Fig. 3D).

Live imaging of GFP-tagged, endogenously encoded GAF demonstrated that GAF is mitotically retained in small foci. This is similar to what was reported for antibody staining on fixed embryos, which showed GAF localized at pericentric heterochromatin regions of GA-rich satellite repeats (Raff et al. 1994; Platero et al. 1998). Because this prior imaging necessitated fixing embryos, it was unclear if mitotic retention of GAF was required for mitosis. Our system enabled us to determine that in the absence of GAF nuclei can undergo several rounds of mitosis albeit with noticeable defects (Supplemental Video 1,2). We conclude that GAF is not strictly required for progression through mitosis. However, the nuclear defects observed in GAF^{deGradFP} embryos support the model that GAF is broadly required for nuclear division and chromosome stability, in addition to its role in transcriptional activation during ZGA (Bhat et al. 1996). Our imaging also identified high nuclear mobility in GAF^{deGradFP} embryos as compared to control embryos; nuclei of a subset of embryos showed a dramatic “swirling” pattern of movement (Supplemental Video 2). This defect is likely due to disorder in the cytoskeletal network that is responsible for nuclear migration and division in the syncytial embryo (Sullivan and Theurkauf 1995). Altogether, our phenotypic analysis of GAF^{deGradFP} embryos shows for the first time that maternal GAF is required for progression through the MZT, and suggests GAF has an early, global role in nuclear division and chromosome stability.

In addition to GAF and Zld, the transcription factor CLAMP is expressed in the early embryo and functions in chromatin accessibility and transcriptional activation (Rieder et al. 2017; Soruco et al. 2013; Urban et al. 2017a, 2017b; Rieder et al. 2019). While CLAMP, like GAF, binds GA-dinucleotide repeats, the two proteins preferentially bind to slightly different GA-repeats (Kaye et al. 2018). GAF and CLAMP can compete for binding sites *in vitro*, and, in cell culture, when one factor is knocked down the occupancy of the other increases, suggesting that CLAMP and GAF compete for a subset of binding sites and may have partial functional redundancy (Kaye et al. 2018). We demonstrate that GAF, like CLAMP, is essential in the early embryo, indicating that these two GA-dinucleotide binding proteins cannot completely compensate for each other *in vivo* during early development (Rieder et al. 2017). Furthermore, our sequencing analysis identified that a majority of genes that require GAF for expression are distinct from those that are regulated by CLAMP. Thus, while GAF and CLAMP may have some overlapping functions, they are independently required to regulate embryonic development during the MZT.

GAF is necessary for widespread zygotic genome activation and chromatin accessibility

GAF is a multi-purposed transcription factor with known roles in transcriptional regulation at promoters and enhancers as well as additional suggested roles in high-order chromatin structure. Our analysis showed that during the MZT GAF acts largely as an activator, directly binding and activating hundreds of zygotic transcripts during ZGA (Fig. 3). We identified thousands of regions bound by GAF throughout the MZT, and these regions were preferentially associated with genes whose transcription decreased when maternally encoded GAF was degraded. This function may be driven, in part, through GAF-mediated chromatin accessibility as we identified hundreds of regions that depend on GAF for accessibility. This activity, in both transcriptional activation and mediating open chromatin, is similar to Zld, the only previously identified essential activator of the zygotic genome in *Drosophila,* and is shared with genome activators in other species, such as

Pou5f3 and Nanog (zebrafish) and Dux4 (mammals) (Schulz and Harrison 2019; Vastenhouw et al. 2019).

Given that GAF directly recruits chromatin remodelers, it was surprising that the majority of regions that changed in accessibility in the absence of GAF were not directly bound by GAF as assayed by ChIP-seq (Fig. 5A) (Okada and Hirose 1998; Tsukiyama et al. 1994; Xiao et al. 2001; Tsukiyama and Wu 1995; Fuda et al. 2015; Judd et al. 2020). Nonetheless, this is consistent with previous data demonstrating a role for GAF in targeting the Male Specific Lethal (MSL) complex to binding sites on the X-chromosome, despite GAF not occupying these sites as determined by ChIP (Kaye et al. 2018). It is possible that, as is proposed in Kaye et al. 2018, regions that require GAF for accessibility in the embryo may experience transient GAF binding not captured by ChIP-seq, but nevertheless required to maintain accessibility. Indeed, in addition to the Zld-binding motif, we identified an enrichment of degenerate GAF motifs in these differential, non GAF-bound peaks (Fig. 5D,E). These motifs diverge from the canonical GAGAG pentamer motif enriched in differential, GAF-bound regions and are similar to those motifs preferentially bound by CLAMP as compared to GAF (Kaye et al. 2018). Thus, at these regions GAF may have more transient binding than at the canonical GAGAG motif, and this binding was not captured in the ChIP. Alternatively, GAF may indirectly affect chromatin accessibility in these regions through a role in shaping three-dimensional chromatin structure. GAF is capable of forming enhancer-promoter loops *in vitro* and *in vivo* (Mahmoudi et al. 2002; Melnikova et al. 2004; Petrascheck et al. 2005). In the embryo, GAF-binding motifs are enriched at TAD boundaries which form during the MZT (Hug et al. 2017) and GAF binding is enriched at Polycomb group dependent repressive loops that form following NC14 (Ogiyama et al. 2018). Whether GAF has a broader role in global chromatin architecture in the early embryo needs to be further investigated.

GAF and Zld are uniquely required to reprogram the zygotic genome

We had previously identified the GAF-binding motif enriched in accessible, *cis*-regulatory regions of the early embryo in which Zld was bound, but not required for accessibility (Schulz et al. 2015). This suggested that GAF might function to maintain accessibility at these Zld-bound regions even in the absence of Zld, and our ATAC-seq data support this model. We identified regions highly dependent on GAF for accessibility and showed that half of these regions overlap regions that are bound by Zld, but not dependent on Zld for accessibility. Thus, GAF mediates chromatin accessibility at a subset of Zld-bound regions independent of Zld activity (Fig. 6C).

Those regions that depended solely on Zld for accessibility were associated with genes expressed during the earliest part of ZGA, while the GAF motif was enriched in regulatory regions of genes expressed during widespread genome activation (Schulz et al. 2015; Blythe and Wieschaus 2016). Furthermore, these regions became accessible later in the MZT than regions that were enriched only for the canonical Zld motif (Blythe and Wieschaus 2016), suggesting the earliest wave of gene expression might be predominantly dependent on Zld for activation while GAF might be more important later. GAF and Zld are both required for activation of a subset of the earliest expressed genes. Nonetheless, those genes dependent on Zld alone are enriched for early expressed genes. By contrast, genes that are activated by GAF alone are preferentially expressed later, during widespread ZGA at nuclear cycle 14. Based on this evidence, we propose that there is a gradual handoff in transcriptional activation from Zld to GAF as ZGA progresses (Fig. 6). GAF interacts with chromatin remodelers and this activity may have a more substantial impact later in ZGA when the division cycle slows as compared to early in development when the division cycle is a series of alternating rapid synthesis and mitotic phases. Thus, Zld and GAF collaborate during ZGA to promote chromatin accessibility and drive gene expression.

Some pioneer factors work together to stabilize their interaction on chromatin (Chronis et al. 2017; Donaghey et al. 2018; Liu and Kraus 2017; Swinstead et al. 2016). Therefore, we tested whether Zld or GAF depend on the other for binding and whether this might be specific to the thousands of regions bound by both factors. Our data indicate that there is a small subset of Zld and GAF binding sites that are lost when the other factor is knocked down. However, overall binding to target sites of both factors is maintained in the absence of the other, indicating that the two factors bind targets in chromatin independently of one another. This is consistent with our results showing GAF and Zld largely regulate transcriptional activation and chromatin accessibility independently. Therefore, although GAF and Zld both function as transcriptional activators during ZGA, they function independently and preferentially at temporally distinct time points.

Together our data support the requirement for at least two pioneer transcription factors, Zld and GAF, to sequentially reprogram the zygotic genome following fertilization and allow for future embryonic development. It is likely that there are additional factors, as in a companion paper Duan et al. identify an essential role for CLAMP in directing Zld binding. Future studies will enable more detailed mechanistic insights into how multiple pioneer factors work together to reshape the transcriptional landscape and transform cell fate in the early embryo.

Materials and Methods

Drosophila strains and genetics

All stocks were grown on molasses food at 25ºC. Fly strains used in this study: *w*^{1118} , *His2Av-RFP (II)* (Bloomington *Drosophila* Stock Center (BDSC) #23651), *mat-α-GAL4-VP16* (BDSC #7062), *UAS-shRNA-zld* (Sun et al. 2015). *sfGFP-GAF(N)* and *GAF-sfGFP(C)* mutant alleles were generated using Cas9-mediated genome engineering (see below).

nos-degradFP (II) transgenic flies were made by PhiC31 integrase-mediated transgenesis into the *PBac{yellow[+]-attP-3B}VK00037* docking site (BDSC #9752) by BestGene Inc. The sequence for *NSlmb-vhhGFP4* was obtained from Caussinus et al. 2012 and amplified from genomic material from *UASp-Nslmb.vhhGFP4* (BDSC #58740). The *NSlmb-vhhGFP4* sequence was cloned using Gibson assembly into pattB (DGRC #1420) with the *nanos* 5'UTR.

To obtain the embryos for ChIP-seq and live embryo imaging in a *zld* knockdown background, we crossed *mat-α-GAL4-VP16 (II)/Cyo; sfGFP-GAF N)(III)* flies to *UAS-shRNA-zld (III)* flies and took *mat-α-GAL4-VP16/+ (II); sfGFP-GAF(N)/ UAS-shRNA-zld (III)* females. These females were crossed to their siblings, and their embryos were collected. Embryos from *mat-α-GAL4-VP16 (II)/CyO; sfGFP-GAF(N) (III)* females were used as controls.

To obtain embryos for live imaging, hatching rate assays, RNA-seq, ATAC-seq, and ChIP-seq in a GAF knockdown background we crossed *nos-degradFP (II); sfGFP-GAF(N)/TM6c (III)* flies to *His2Av-RFP (II); sfGFP-GAF(N) (III)* flies and selected females that were *nos-degradFP/His2Av-RFP (II); sfGFP-GAF(N) (III).* These females were crossed to their siblings of the same genotype, and their embryos were collected. Embryos from *His2Av-RFP (II); sfGFP-GAF(N) (III)* females were used as paired controls.

Cas9-genome engineering

Cas9-mediated genome engineering as described in (Hamm et al. 2017) was used to generate the N-terminal and C-terminal super folder Green Fluorescent Protein (sfGFP)-tagged GAF. The double-stranded DNA (dsDNA) donor was created using Gibson assembly (New England BioLabs, Ipswich, MA) with 1-kb homology arms flanking the sfGFP tag and GAF N-terminal or C-terminal open reading frame. sfGFP sequence was placed downstream of the GAF start codon (N-terminal) or just upstream of the stop codon in the 5th GAF exon, coding for the short isoform (C-terminal). Additionally, a 3xP3-DsRed cassette flanked by the long-terminal repeats of PiggyBac transposase was placed in the second GAF intron (N-terminal) or fourth GAF intron (C-terminal) for selection. The guide RNA sequences (N-terminal- TAAACATTAAATCGTCGTGT), (C-terminal- AAATGAATACTCGATTA) were cloned into pBSK under the U63 promoter using inverse PCR. Purified plasmid was injected into embryos of *yw; attP40{nos-Cas9}/CyO* for the N-terminal line and *y*^{1} *M{vas-Cas9.RFP-}ZH-2A w*^{1118} (BDSC#55821) for the C-terminal line by BestGene Inc. Lines were screened for DsRed expression to verify integration. The entire 3xP3-DsRed cassette was cleanly removed using piggyBac transposase, followed by sequence confirmation of precise tag integration.

Live embryo imaging

Embryos were dechorionated in 50% bleach for 2 minutes and subsequently mounted in halocarbon 700 oil. Due to the fragility of the GAF^{degradFP} embryos, embryos used for videos were mounted in halocarbon 700 oil without dechorionation. The living embryos were imaged on a Nikon A1R+ confocal at the University of Wisconsin-Madison Biochemistry Department Optical Core. Nuclear density, based on the number of nuclei/2500 µm2, was used to determine the cycle of pre-gastrulation embryos. Nuclei were marked with His2AV-RFP. Image J (Schindelin et al.

2012) was used for post-acquisition image processing. Videos were acquired at 1 frame every 10 seconds. Playback rate is 7 frames/second.

Hatching rate assays

A minimum of 50 females and 25 males of the indicated genotypes were allowed to mate for at least 24 hours before lays were taken for hatching rate assays. Embryos were picked from overnight lays and approximately 200 were lined up on a fresh molasses plate. Unhatched embryos were counted 26 hours or more after embryos were selected.

Immunoblotting

Proteins were transferred to 0.45 µm Immobilon-P PVDF membrane (Millipore, Burlington, MA) in transfer buffer (25 mM Tris, 200 mM Glycine, 20% methanol) for 60 min (75 min for Zld) at 500mA at 4°C. The membranes were blocked with blotto (2.5% non-fat dry milk, 0.5% BSA, 0.5% NP-40, in TBST) for 30 min at room temperature and then incubated with anti-GFP (1:2000, #ab290) (Abcam, Cambridge, United Kingdom) anti-Zld (1:750) (Harrison et al. 2011), or anti-tubulin (DM1A, 1:5000) (Sigma, St. Louis, MO), overnight at 4°C. The secondary incubation was performed with goat anti-rabbit IgG-HRP conjugate (1:3000) (Bio-Rad, Hercules, CA) or anti-mouse IgG-HRP conjugate (1:3000) (Bio-Rad) for 1 hour at room temperature. Blots were treated with SuperSignal West Pico PLUS chemiluminescent substrate (Thermo Fisher Scientific, Waltham, MA) and visualized using the Azure Biosystems c600 or Kodak/Carestream BioMax Film (VWR, Radnor, PA).

Chromatin Immunoprecipitation

ChIP was performed as described previously (Blythe and Wieschaus 2015) on hand selected embryos: stage 3 and stage 5 *GAF-sfGFP(C)* homozygous embryos, stage 3 and stage 5 w^{1118} embryos, stage 5 embryos from *mat-α-GAL4-VP16/+ (II); sfGFP-GAF(N)/ UAS-shRNA-zld (III)* females, stage 5 embryos from *nos-degradFP/His2Av-RFP (II); sfGFP-GAF(N) (III)* females, and stage 5 embryos from *His2Av-RFP (II); sfGFP-GAF(N) (III)* females. Briefly, 1000 stage 3 embryos or 400-500 stage 5 embryos were collected, dechorionated in 50% bleach for 3 min, fixed for 15 min in 4% formaldehyde and then lysed in 1 mL of RIPA buffer (50 mM Tris-HCl pH 8.0, 0.1% SDS, 1% Triton X-100, 0.5% sodium deoxycholate, and 150 mM NaCl). The fixed chromatin was then sonicated for 20 s 11 times at 20% output and full duty cycle (Branson Sonifier 250). Chromatin was incubated with 6 µg of anti-GFP antibody (Abcam #ab290) or 8µl of anti-Zld antibody (Harrison et al. 2011) overnight at 4°C , and then bound to 50 µl of Protein A magnetic beads (Dynabeads Protein A, Thermo Fisher Scientific). The purified chromatin was then washed, eluted, and treated with 90 µg of RNaseA (37°C, for 30 min) and 100 µg of Proteinase K (65°C, overnight). The DNA was purified using phenol/chloroform extraction and concentrated by ethanol precipitation. Each sample was resuspended in 25 µl of water. Sequencing libraries were made using the NEB Next Ultra II library kit and were sequenced on the Illumina Hi-Seq4000 using 50bp single-end reads at the Northwestern Sequencing Core (NUCore).

ChIP-seq Data Analysis

ChIP-seq data was aligned to the *Drosophila melanogaster* reference genome (version dm6) using bowtie 2 v2.3.5 (Langmead and Salzberg 2012) with the following non-default parameters: -k 2, --very-sensitive. Aligned reads with a mapping quality < 30 were discarded, as were reads aligning to scaffolds or the mitochondrial genome. To identify regions that were enriched in immunoprecipitated samples relative to input controls, peak calling was performed using MACS v2 (Zhang et al. 2008) with the following parameters: -g 1.2e8, --call-summits. To focus analysis on robust, high-quality peaks, we used 100 bp up- and downstream of peak summits, and retained only peaks that were detected in both replicates and overlapped by at least 100 bp. All downstream analysis focused on these high-quality peaks. Peak calling was also performed for

control ChIP samples performed on w^{1118} with the α-GFP antibody. No peaks were called in any of the w^{1118} controls, indicating high specificity of the α-GFP antibody. To compare GAF-binding sites at stage 3, stage 5 and in S2 cells, and to compare GAF, Zld and CLAMP binding, we used GenomicRanges R package (Lawrence et al. 2013) to compare different sets of peaks. Peaks overlapping by at least 100 bp were considered to be shared. To control for differences in data processing and analysis between studies, previously published ChIP-seq datasets for Zld (Harrison et al. 2011, GSE30757), GAF (Fuda et al. 2015, GSE40646), and CLAMP (Rieder et al. 2019, GSE133637) were processed in parallel with ChIP-seq data sets generated in this study. DeepTools (Ramírez et al. 2016) was used to calculate counts per million (CPM)-normalized read depth used for all heatmaps and metaplots. *De novo* motif discovery was performed using MEME-suite (Bailey et al. 2009).

Analysis of differential binding in GAF- and Zld-depleted embryos

Considering the global decrease in ChIP signal in GAF- and Zld-depleted embryos (Fig. S5A,C), we considered technical factors that may have affected these experiments. Technical reasons global GAF binding signal may have been lower in the *zld* RNAi background:
(1) In the *zld* RNAi background the *sfGFP-GAF(N)* allele was heterozygous while in the control it was homozygous. (2) *zld* RNAi embryos die around stage 5, therefore a portion of these embryos collected for ChIP-seq may have been dead. (3) In the *zld* RNAi background the sfGFP tag on GAF is on the N-terminus, while in the control dataset the sfGFP tag is on the C-terminus. The N-terminal tag may interfere with GAF binding efficiency in a manner the C-terminal tag does not. (4) The IP efficiency may have been lower in the *zld* RNAi ChIP.

Technical reasons global Zld binding signal may have been lower in the $GAF^{deGradFP}$ background: (1) $GAF^{deGradFP}$ embryos die at variable times around the MZT, therefore a portion of these embryos collected for ChIP-seq may have been dead. (2) The IP efficiency may have been lower in the $GAF^{deGradFP}$ ChIP.

To control for these technical factors, we performed an analysis based on peak rank, rather than peak intensity. The number of reads aligning within each peak was quantified using featureCounts from the Subread package (v1.6.4) (Liao et al. 2014). Peaks were then ranked based on the mean RPKM-normalized read count between replicates, allowing comparison of peak rank between different conditions.

DESeq2 (Love et al. 2014) was used to identify potential differential binding sites in a more statistically rigorous way. To control for variable detection of low-intensity peaks, only the top third of GAF Stage 5 peaks (1,444 peaks) were analyzed. For Zld ChIP-seq, analysis was restricted to 6,003 high-confidence peaks shared between our control dataset and previously published Zld ChIP (Harrison et al. 2011). A table of read counts for these peaks, generated by featureCounts as described above, was used as input to DESeq2. Peaks with an adjusted p-value < 0.05 and a fold change > 2 were considered to be differentially bound.

Total RNA-seq

150-200 embryos from *His2Av-RFP/nos-degradFP (II); sfGFP-GAF(N) (III)* and *His2Av-RFP (II); sfGFP-GAF(N) (III)* females were collected from a half hour lay and aged for 2 hours. Embryos were then picked into Trizol (Invitrogen, Carlsbad, CA) with 200 µg/ml glycogen (Invitrogen). RNA was extracted and RNA-seq libraries were prepared using the Universal RNA-Seq with NuQuant, *Drosophila* AnyDeplete Universal kit (Tecan, Männedorf, Switzerland). Samples were sequenced on the Illumina NextSeq500 using 75bp single-end reads at the Northwestern Sequencing Core (NUCore).

RNA-seq analysis
RNA-seq data was aligned to the *Drosophila melanogaster* genome (dm6) using HISAT v2.1.0 (Kim et al. 2015 Nature Methods). Reads with a mapping quality score < 30 were discarded. The number of reads aligning to each gene was quantified using featureCounts, generating a read count table that was used to analyze differential expression with DESeq2. Genes with an adjusted p-value < 0.05 and a fold change > 2 were considered statistically significant. To identify GAF-target genes, GAF ChIP peaks were assigned to the nearest gene (this study). Zygotically and maternally expressed genes (Lott et al. 2011, GSE25180) zygotic gene expression onset (Li et al. 2014, GSE58935), Zld-dependent genes (McDaniel et al. 2019, GSE121157), CLAMP-dependent genes (Reider et al. 2017, GSE102922), and Zld targets (Harrison et al. 2011, GSE30757) as previously defined.

Assay for Transposase-Accessible Chromatin
150-200 embryos from *His2Av-RFP/nos-degradFP (II); sfGFP-GAF(N) (III)* and *His2Av-RFP (II); sfGFP-GAF(N) (III)* females were collected from a half hour lay and aged for 2 hours. Embryos were dechorionated in 50% bleach for 2 minutes. Embryos were then dissociated on ice in HL buffer (25mM KCl; 90mM NaCl; 4.8mM NaHCO3; 80 mM D-glucose; 5mM Trehalose; 5mM L-glutamine; 10mM HEPES; 5mM EDTA; pH to 6.9) with 5 stokes of a loose pestle with a 1 mL dounce. The cells suspension was filtered through miracloth (Thermo Fisher Scientific) and pelleted for 10 minutes at 300g at 4ºC. Cells were resuspended in cell lysis buffer (10mM Tris 7.5, 10mM NaCl, 3mM MgCl2, 0.1% NP-40), pelleted, and washed with cell lysis buffer. Nuclei were then pelleted for 10 minutes at 1000g at 4ºC and resuspended in 22.5 µl of water before adding 25 µl of buffer TD (Illumina, San Diego, CA) and 2.5 µl Tn5 transposase (Tagment DNA Enzyme, Illumina). Reactions were incubated for 30 min at 4ºC then immediately purified with the MinElute Cleanup Kit (Qiagen, Hilden, Germany) and eluted in 10 µL buffer EB. Libraries were amplified for 12 PCR cycles with unique dual index primers using the NEBNext Hi-Fi 2X PCR Master Mix (New England Biolabs). Amplified libraries were purified using a 1.2X ratio of Axygen magnetic beads. Libraries were submitted to the University of Wisconsin-Madison Biotechnology Center for 150 bp, paired-end sequencing on the Illumina NovaSeq 6000.

ATAC-seq analysis
Adapter sequences were removed from raw sequence reads using NGMerge (Gaspar 2018). ATAC-seq reads were aligned to the *Drosophila melanogaster* (dm6) genome using bowtie2 with the following parameters: --very-sensitive, --no-mixed, --no-discordant, -X 5000, -k 2. Reads with a mapping quality score < 30 were discarded, as were reads aligning to scaffolds or the mitochondrial genome. Analysis was restricted to fragments < 100 bp, which, as described previously, are most likely to originate from nucleosome-free regions (Buenrostro et al. 2013). To maximize the sensitivity of peak calling, reads from all replicates of $GAF^{deGradFP}$ and control embryos were combined. Peak calling was performed on combined reads using MACS2 with parameters -f BAMPE --keep-dup all -g 1.2e8 --call-summits. This identified 64,133 accessible regions. 201 bp peak regions (100 bp on either side of the peak summit) were used for downstream analysis. Reads aligning within accessible regions were quantified using featureCounts, and differential accessibility analysis was performed using DESeq2 with an adjusted p-value < 0.05 and a fold change > 2 as thresholds for differential accessibility. Regions depending on Zld for accessibility (as assayed by FAIRE-seq) were previously defined (Schulz et al. 2015). For comparison to ATAC-seq data from Zld-depleted embryos (Fig S6C), previously published data (Hannon et al. 2017, GSE86966) were re-analyzed in parallel with ATAC-seq data from this study to ensure identical processing of data sets. Heatmaps and metaplots of CPM-normalized read depth were generated with DeepTools. MEME-suite was used to for *de novo* motif discovery for differentially accessible ATAC peaks and to match discovered motifs to previously known motifs from databases.

Data availability
All raw sequencing reads are available at GEO GSE152773.

Acknowledgments
We thank Erica Larschan and Jingyue Duan for helpful discussions. We also thank Christine Rushlow, the Bloomington Stock Center, and the Drosophila Genome Resource Center for providing reagents and fly lines. We acknowledge the University of Wisconsin-Madison Biochemistry Department Optical Core for access to microscopes for imaging and the University of Wisconsin-Madison Biotechnology Center and the NUSeq Core Facility for sequencing. M.M.G. and T.J.G. were supported by National Institutes of Health (NIH) National Research Service Award T32 GM007215. Experiments were supported by a R01 GM111694 and R35 GM136298 from the National Institutes of Health (NIH) and a Vallee Scholar Award (M.M.H).

Author contributions
M.M.G., T.J.G, and M.M.H. conceived the study. M.M.G., T.J.G. and E.D.L. performed the experiments and analysis. M.M.G. and M.M.H. wrote the original manuscript. M.M.G., T.J.G, E.D.L. and M.M.H. read and edited the manuscript.

References

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202–W208.

Bejarano F, Busturia A. 2004. Function of the Trithorax-like gene during Drosophila development. *Dev Biol* **268**: 327–341.

Benyajati C, Mueller L, Xu N, Pappano M, Gao J, Mosammaparast M, Conklin D, Granok H, Craig C, Elgin S. 1997. Multiple isoforms of GAGA factor, a critical component of chromatin structure. *Nucleic Acids Res* **25**: 3345–3353.

Bernués J, Piñeyro D, Kosoy A. 2007. General, negative feedback mechanism for regulation of Trithorax-like gene expression in vivo: New roles for GAGA factor in flies. *Nucleic Acids Res* **35**: 7150–7159.

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.

Biggin MD, Tjian R. 1988. Transcription factors that activate the Ultrabithorax promoter in developmentally staged extracts. *Cell* **53**: 699–711.

Blythe SA, Wieschaus EF. 2016. Establishment and maintenance of heritable chromatin structure during early drosophila embryogenesis. *Elife* **5**: e20148.

Blythe SA, Wieschaus EF. 2015. Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. *Cell* **160**: 1169–1181.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–1218.

Busturia A, Lloyd A, Bejarano F, Zavortink M, Xin H, Sakonju S. 2001. The MCP silencer on the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. *Development* **128**: 2163–2173.

Caussinus E, Kanca O, Affolter M. 2012. Fluorescent fusion protein knockout mediated by anti-GFP nanobody. *Nat Struct Mol Biol* **19**: 117–122.

Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K. 2017. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* **168**: 442-459.e20.

Donaghey J, Thakurela S, Charlton J, Chen JS, Smith ZD, Gu H, Pop R, Clement K, Stamenova EK, Karnik R, et al. 2018. Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat Genet* **50**: 250–258.

Dufourt J, Trullo A, Hunter J, Fernandez C, Lazaro J, Dejean M, Morales L, Nait-Amer S, Schulz KN, Harrison MM, et al. 2018. Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nat Commun* **9**: 5194.

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.

Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien CY, Kirov N, Shvartsman SY, Rushlow CA. 2014. Zelda potentiates morphogen activity by increasing chromatin accessibility. *Curr Biol* **24**: 1341–1346.

Fu S, Nien CY, Liang HL, Rushlow C. 2014. Co-activation of microRNAs by Zelda is essential for early drosophila development. *Dev* **141**: 2108–2118.

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**: e1005108.

Gaspar JM. 2018. NGmerge: Merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics* **19**: 536.

Gilmour DS, Thomas GH, Elgin SCR. 1989. Drosophila nuclear proteins bind to regions of alternating C and T residues in gene promoters. *Science (80-)* **245**: 1487–1490.

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**: e1007120.
Hannon CE, Blythe SA, Wieschaus EF. 2017. Concentration dependent chromatin states induced by the bicoid morphogen gradient. *Elife* **6**: e28275.
Harrison MM, Li XY, 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 Genet* **7**: e1002266.
Horard B, Tatout C, Poux S, Pirrotta V. 2000. Structure of a Polycomb Response Element and In Vitro Binding of Polycomb Group Complexes Containing GAGA Factor. *Mol Cell Biol* **20**: 3187–3197.
Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. 2017. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**: 216-228.e19.
Iwafuchi-Doi M, Zaret KS. 2014. Pioneer transcription factors in cell reprogramming. *Genes Dev* **28**: 2679–2692.
Judd J, Duarte FM, Lis JT. 2020. Pioneer factor GAF cooperates with PBAP and NURF to regulate transcription. *bioRxiv* 2020.05.10.087262. https://www.biorxiv.org/content/10.1101/2020.05.10.087262v1.
Kaye EG, Booker M, Kurland J V., 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.
Kaye EG, Kurbidaeva A, Wolle D, Aoki T, Schedl P, Larschan E. 2017. Drosophila Dosage Compensation Loci Associate with a Boundary-Forming Insulator Complex. *Mol Cell Biol* **37**: e00253-17.
Kerrigan LA, Croston GE, Lira LM, Kadonaga JT. 1991. Sequence-specific transcriptional antirepression of the Drosophila Kruppel gene by the GAGA factor. *J Biol Chem* **266**: 574–582.
Kuzu G, Kaye EG, Chery J, Siggers T, Yang L, Dobson JR, Boor S, Bliss J, Liu W, Jogl G, et al. 2016. Expansion of GA Dinucleotide Repeats Increases the Density of CLAMP Binding Sites on the X-Chromosome to Promote Drosophila Dosage Compensation. *PLoS Genet* **12**: e1006120.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.
Larschan E, Soruco MML, Lee OK, Peng S, Bishop E, Chery J, Goebel K, Feng J, Park PJ, Kuroda MI. 2012. Identification of chromatin-associated regulators of MSL complex targeting in Drosophila dosage compensation. *PLoS Genet* **8**: e1002830.
Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* **9**: e1003118.
Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, Jiang C, Li J, Pugh BF, Gilmour DS. 2008. NELF and GAGA Factor Are Linked to Promoter-Proximal Pausing at Many Genes in Drosophila. *Mol Cell Biol* **28**: 3290–3300.
Lee HS, Kraus KW, Wolfner MF, Lis JT. 1992. DNA sequence requirements for generating paused polymerase at the start of hsp 70. *Genes Dev* **6**: 284–295.
Li XY, 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.
Liang HL, Nien CY, Liu HY, 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.
Liao Y, Smyth GK, Shi W. 2014. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

Liu Z, Kraus WL. 2017. Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. *Mol Cell* **65**: 589-603.e9.
Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. 2011. Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-Seq. *PLoS Biol* **9**: e1000590.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
Mahmoudi T, Katsani KR, Verrijzer CP. 2002. GAGA can mediate enhancer function in trans by linking two separate DNA molecules. *EMBO J* **21**: 1775–1781.
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.
Melnikova L, Juge F, Gruzdeva N, Mazur A, Cavalli G, Georgiev P. 2004. Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. *Proc Natl Acad Sci U S A* **101**: 14806–14811.
Mir M, Stadler MR, Ortiz SA, Hannon CE, Harrison MM, Darzacq X, Eisen MB. 2018. Dynamic multifactor hubs interact transiently with sites of active transcription in drosophila embryos. *Elife* **7**.
Mishra RK, Mihaly J, Barges S, Spierer A, Karch F, Hagstrom K, Schweinsberg SE, Schedl P. 2001. The iab-7 Polycomb Response Element Maps to a Nucleosome-Free Region of Chromatin and Requires Both GAGA and Pleiohomeotic for Silencing Activity. *Mol Cell Biol* **21**: 1311–1318.
Nien CY, Liang HL, Butcher S, Sun Y, Fu S, Gocha T, Kirov N, Manak JR, Rushlow C. 2011. Temporal coordination of gene networks by Zelda in the early Drosophila embryo. *PLoS Genet* **7**: e1002339.
Ogiyama Y, Schuettengruber B, Papadopoulos GL, Chang JM, Cavalli G. 2018. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Mol Cell* **71**: 73-88.e5.
Ohtsuki S, Levine M. 1998. GAGA mediates the enhancer blocking activity of the eve promoter in the Drosophila embryo. *Genes Dev* **12**: 3325–3330.
Okada M, Hirose S. 1998. Chromatin Remodeling Mediated by Drosophila GAGA Factor and ISWI Activates fushi tarazu Gene Transcription In Vitro. *Mol Cell Biol* **18**: 2455–2461.
Petrascheck M, Escher D, Mahmoudi T, Verrijzer CP, Schaffner W, Barberis A. 2005. DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Res* **33**: 3743–3750.
Platero JS, Csink AK, Quintanilla A, Henikoff S. 1998. Changes in chromosomal localization of heterochromatin-binding proteins during the cell cycle in Drosophila. *J Cell Biol* **140**: 1297–1306.
Raff JW, Kellum R, Alberts B. 1994. The Drosophila GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle. *EMBO J* **13**: 5977–5983.
Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**: W160–W165.
Read D, Nishigaki T, Manley JL. 1990. The Drosophila even-skipped promoter is transcribed in a stage-specific manner in vitro and contains multiple, overlapping factor-binding sites. *Mol Cell Biol* **10**: 4334–4344.
Rieder LE, Jordan WT, Larschan EN. 2019. Targeting of the Dosage-Compensated Male X-Chromosome during Early Drosophila Development. *Cell Rep* **29**: 4268-4275.e2.
Rieder LE, Koreski KP, Boltz KA, Kuzu G, Urban JA, Bowman SK, Zeidman A, Jordan WT, Tolstorukov MY, Marzluff WF, et al. 2017. Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP. *Genes Dev* **31**: 1494–1508.
Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S,

Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: An open-source platform for biological-image analysis. *Nat Methods* **9**: 676–682.
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*.
Schulz KN, Harrison MM. 2019. Mechanisms regulating zygotic genome activation. *Nat Rev Genet* **20**: 221–234.
Slattery M, Ma L, Spokony RF, Arthur RK, Kheradpour P, Kundaje A, Nègre N, Crofts A, Ptashkin R, Zieba J, et al. 2014. Diverse patterns of genomic targeting by transcriptional regulators in Drosophila melanogaster. *Genome Res* **24**: 1224–1235.
Soeller WC, Oh CE, Kornberg TB. 1993. Isolation of cDNAs encoding the Drosophila GAGA transcription factor. *Mol Cell Biol* **13**: 7961–7970.
Soeller WC, Poole SJ, Kornberg T. 1988. In vitro transcription of the Drosophila engrailed gene. *Genes Dev* **2**: 68–81.
Soruco MML, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, Sugden AU, Goebel K, Feng J, Xia P, et al. 2013. The CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage compensation. *Genes Dev* **27**: 1551–1556.
Soufi A, Donahue G, Zaret KS. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**: 994–1004.
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.
Staudt N, Fellert S, Chung HR, Jäckle H, Vorbrüggen G. 2006. Mutations of the Drosophila zinc finger-encoding gene vielfältig impair mitotic cell divisions and cause improper chromosome segregation. *Mol Biol Cell* **17**: 2356–2365.
Sullivan W, Theurkauf WE. 1995. The cytoskeleton and morphogenesis of the early Drosophila embryo. *Curr Opin Cell Biol* **7**: 18–22.
Sun Y, Nien CY, Chen K, Liu HY, 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.
Swinstead EE, Miranda TB, Paakinaho V, Baek S, Goldstein I, Hawkins M, Karpova TS, Ball D, Mazza D, Lavis LD, et al. 2016. Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell* **165**: 593–605.
Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**: 861–872.
Takahashi K, Yamanaka S. 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**: 663–676.
Tsukiyama T, Becker PB, Wu C. 1994. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**: 525–532.
Tsukiyama T, Wu C. 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**: 1011–1020.
Urban J, Kuzu G, Bowman S, Scruggs B, Henriques T, Kingston R, Adelman K, Tolstorukov M, Larschan E. 2017a. Enhanced chromatin accessibility of the dosage compensated Drosophila male X-chromosome requires the CLAMP zinc finger protein. *PLoS One* **12**: e0186855.
Urban JA, Doherty CA, Jordan WT, Bliss JE, Feng J, Soruco MM, Rieder LE, Tsiarli MA, Larschan EN. 2017b. The essential Drosophila CLAMP protein differentially regulates non-coding roX RNAs in male and females. *Chromosom Res* **25**: 101–113.
Vastenhouw NL, Cao WX, Lipshitz HD. 2019. The maternal-to-zygotic transition revisited. *Development* **146**: dev161471.

Wang C, Lehmann R. 1991. Nanos is the localized posterior determinant in Drosophila. *Cell* **66**: 637–647.
Wolle D, Cleard F, Aoki T, Deshpande G, Schedl P, Karch F. 2015. Functional Requirements for Fab-7 Boundary Activity in the Bithorax Complex . *Mol Cell Biol* **35**: 3739–3752.
Xiao H, Sandaltzopoulos R, Wang HM, Hamiche A, Ranallo R, Lee KM, Fu D, Wu C. 2001. Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol Cell* **8**: 531–543.
Xu Z, Chen H, Ling J, Yu D, Struffi P, Small S. 2014. Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila. *Genes Dev* **28**: 608–621.
Yamada S, Whitney PH, Huang SK, Eck EC, Garcia HG, Rushlow CA. 2019. The Drosophila Pioneer Factor Zelda Modulates the Nuclear Microenvironment of a Dorsal Target Enhancer to Potentiate Transcriptional Output. *Curr Biol* **29**: 1387-1393.e5.
Yáñez-Cuna JO, Dinh HQ, Kvon EZ, Shlyueva D, Stark A. 2012. Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. *Genome Res* **22**: 2018–2030.
Zaret KS, Carroll JS. 2011. Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev* **25**: 2227–2241.
Zaret KS, Mango SE. 2016. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev* **37**: 76–81.
Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.