Localization of the *Drosophila* pioneer factor GAF to subnuclear foci is driven by DNA binding and required to silence satellite repeat expression

Marissa. M. Gaskill[1], **Isabella V. Soluri**[2], **Annemarie E. Branks**[1], **Alan P. Boka**[3,4], **Michael R. Stadler**[5], **Katherine Vietor**[1,6], **Hao-Yu S. Huang**[1,7], **Tyler J. Gibson**[1], **Apratim Mukherjee**[4,10], **Mustafa Mir**[4,8,9,10], **Shelby A. Blythe**[2], **Melissa M. Harrison**[1,*]

[1]Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA

[2]Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

[3]Biochemistry and Molecular Biophysics Graduate Group, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

[4]Center for Computational and Genomic Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[5]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720, USA

[6]Present Address: Lineberger Comprehensive Cancer Center, Department of Medicine, Division of Oncology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

[7]Present Address: Biology Department, Stanford University, Stanford, CA 94305, USA

[8]Epigenetics Institute, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

[9]Institute for Regenerative, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

[10]Department of Cell and Developmental Biology, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

Summary

The eukaryotic genome is organized to enable the precise regulation of gene expression. This organization is established as the embryo transitions from a fertilized gamete to the totipotent zygote. To understand the factors and processes that drive genomic organization, we focused on the pioneer factor GAGA factor (GAF) that is required for early development in *Drosophila.* GAF transcriptionally activates the zygotic genome and is localized to subnuclear foci. This non-uniform distribution is driven by binding to highly abundant GA-repeats. At GA-repeats, GAF is necessary to form heterochromatin and silence transcription. Thus, GAF is required to establish both active and silent regions. We propose that foci formation enables GAF to have opposing transcriptional roles within a single nucleus. Our data support a model in which the subnuclear concentration of transcription factors acts to organize the nucleus into functionally distinct domains essential for the robust regulation of gene expression.

eTOC blurb

Gaskill et al. demonstrate that in addition to the previously identified role of the pioneer factor GAF in activating the initial zygotic transcription in *Drosophila*, GAF is required to establish silencing of highly abundant AAGAG satellite repeats. This silencing occurs in subnuclear GAF foci established through DNA binding.

Graphical Abstract
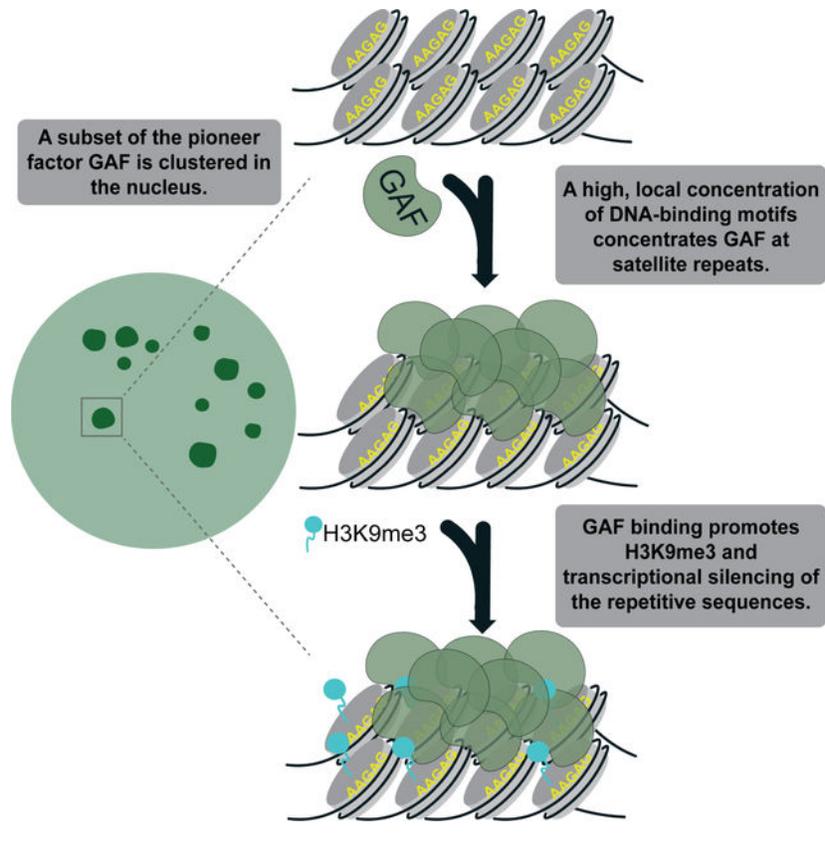
Introduction

Development requires the precise control of gene expression. Transitions in cell fate necessitate both the activation and silencing of transcription controlled by *trans*-acting factors that bind chromatin. However, chromatin presents a barrier to the binding of many transcription factors[1,2]. By contrast, pioneer transcription factors bind to sequence-specific target motifs when they are wrapped around a nucleosome[3–5]. These factors act at the top of gene regulatory networks to drive widespread changes in cell fate by creating local regions of chromatin accessibility that can be bound by additional transcription factors[4,5]. Silencing of gene expression is associated with a loss of chromatin accessibility at *cis*-regulatory regions and the deposition of histone modifications that inhibit transcription-factor binding[6]. Gene expression programs controlled by *trans*-acting factors must be carefully balanced to achieve the changes in cell fate required during development[4,6]. Nonetheless, much remains unknown about how this process is regulated to precisely control cell-fate changes.

Dramatic changes in cell fate and gene expression occur during the rapid reprogramming of the fertilized egg to the totipotent embryo. Initially following fertilization, the zygotic genome is transcriptionally quiescent, and development is controlled by maternally deposited mRNAs and proteins. The genome is devoid of most histone modifications and lacks features of mature heterochromatin domains. Thus, transcriptional activation and silencing need to be established *de novo* in the embryo during this maternal-to-zygotic transition (MZT)[7,8]. In early development, chromatin in the nucleus largely lacks long-range contacts and is not clearly organized into compartments. As the transcriptional program is established, active and silent genomic regions are segregated into distinct compartments, and some *trans*-acting factors become localized to discrete subnuclear domains[7–11]. Together these events restructure the genome and lead to the rapid reprogramming of cell fate.

Several regulatory proteins and transcription factors have been reported to have non-uniform distribution within the nucleus, visualized as foci and referred to as condensates or hubs. In many cases, this distribution is driven by multivalent interactions between the intrinsically disordered regions (IDRs) prevalent in eukaryotic transcription factors[12–15]. For example, Heterochromatin Protein 1a/α (HP1a/α) forms membraneless condensates thought to be formed by liquid-liquid phase separation (LLPS) in both flies and mammals[10,16,17]. In contrast to the stable condensates formed by HP1, the *Drosophila* transcription factor Zelda forms dynamic hubs that mediate DNA binding of additional transcription factors[18–21]. Functionally similar, high-local concentration microenvironments have also been reported for the transcription factor Ultrabithorax [22]. In zebrafish, the pioneer factor Nanog organizes a hub concentrating Sox19b and RNA Polymerase II at the highly transcribed *mir-430* locus[23]. These transcription factor condensates have been proposed to be important for a number of processes, including active transcription, and may contribute to partitioning the genome[24]. However, it has been challenging to test the functional significance of non-uniform subnuclear distribution of proteins within a biological system[25].

We employed the rapidly developing, highly genetically tractable model organism *Drosophila melanogaster* to study the impact of subnuclear domains during the dynamic reprogramming in the early embryo. Early *Drosophila* development is characterized by 13

synchronous nuclear divisions, which alternate between replication and mitosis with no gap phases. At nuclear cycle (NC) 8 transcription from some of the earliest zygotic genes can be detected, but widespread zygotic genome activation (ZGA) does not occur until NC14[26]. We recently showed that the pioneer factor, GAGA factor (GAF) is required for development through the MZT. GAF activates the zygotic genome, preferentially driving gene expression during the major wave of ZGA at NC14[27]. During early development, GAF forms subnuclear foci that are retained on chromatin during mitosis[27,28]. While these GAF foci were first observed decades ago, their contribution to transcriptional regulation remains unknown. Here we used GAF foci as a model to understand protein features that drive non-uniform distribution of pioneer factors and how this distribution contributes to transcription-factor function during cellular reprogramming.

Results

GAF forms foci during the MZT

GAF is non-uniformly distributed in the nucleus[28]. Imaging endogenous GAF tagged with super-folder GFP (sfGFP), we demonstrated that GAF forms robust, discrete foci during interphase of early embryogenesis and is mitotically retained[27] (Figure 1A). Using high resolution lattice light-sheet imaging on sfGFP-GAF embryos through NC13 and NC14, we identified that GAF formed on average 11 foci per nucleus (Figure S1A, B, C)[29]. In both NC13 and NC14, the average number of foci initially decreased, accompanied by an increase in the volume of GAF foci (Figure 1B, C, S1A, B, C). This could be explained by fusion events. Supporting this explanation, we observed fusion of GAF foci at NC14 (Figure 1D). The localization of GAF to concentrated foci that undergo fusion is reminiscent of phase-separated nuclear factors[10,11], or membraneless compartments formed by multivalent interactions between IDRs[12–14]. However, further into interphase of both NC13 and 14 we identified an increase in foci number and concordant decrease in total foci volume (Figure 1B, C, S1B, C). This suggests a subsequent splitting of foci, which we also observed (Figure 1D). This fusion and subsequent fission of the GAF foci is not consistent with LLPS and led us to investigate the features that caused the localization of GAF to discrete subnuclear foci and whether this localization was driven by IDRs, as has been reported for other factors.

The long GAF isoform is not required for viability

GAF has an intrinsically disordered poly-glutamine (poly-Q) enriched C-terminal domain, which drives multimerization *in vitro*[30]. There are two predominant isoforms of GAF that differ in the length and sequence of their poly-Q domains: a long isoform (582 aa) and a short isoform (519 aa) (Figure S2A). If the poly-Q IDR promotes localization to foci, it is possible that the isoform-specific domains promote distinct abilities of each isoform to localize to subnuclear foci and discrete *in vivo* functions. The GAF isoforms have developmentally distinct expression patterns[31]. The long isoform is not present in the 0–2-hour embryo (Figure S2B), and the protein encoded by this isoform is not detectable in the embryo at this time in development[27,31]. This distinct developmental expression pattern and the high level of conservation of the two GAF isoforms in distantly related *Drosophila* species suggests that the two isoforms may have separate *in vivo* functional roles[32]. By contrast, prior studies using transgene expression concluded that the two isoforms largely

overlap in their function[33]. Because transgenes do not always reflect endogenous expression levels and patterns, we made mutations in the endogenous GAF locus to establish the functional roles of each isoform.

To interfere with the long isoform, a stop codon was introduced at the beginning of the 6th exon, which is a coding region only in the long isoform (Figure 2A). This allele is referred to as GAF^L. To interfere with the short isoform, the 297 bp coding region unique to the short isoform, including the stop codon was deleted (424-519aa). This region is alternatively spliced out of the long isoform transcript (Figure 2A). This results in the deletion of the poly-Q domain from the short isoform and is referred to as $GAF^{S\ PQ}$. We demonstrated that GAF^L does not produce a detectable product for the long isoform, which is likely degraded by nonsense mediated decay (Figure 2A, S2C, S2D). Thus, flies homozygous for GAF^L lack the long isoform. $GAF^{S\ PQ}$ resulted in a stable transcript of the short isoform that encodes a truncated protein (Figure S2B–E). The long isoform was not detectable in these blots, likely because it is expressed at much lower levels than the short isoform[27]. Thus, flies homozygous for $GAF^{S\ PQ}$ express the long isoform and an altered version of the short GAF protein lacking the poly-Q domain.

Having generated isoform-specific alleles, we tested whether these alleles resulted in mutant phenotypes. GAF is essential for viability, and flies lacking zygotically expressed GAF die before the third instar larval stage[34]. Hypomorphic GAF alleles result in homeotic transformations[33,34]. By contrast, flies with the GAF^L or $GAF^{S\ PQ}$ allele over a GAF null allele were viable and fertile without any obvious mutant phenotypes. Quantification of adult viability confirmed that flies homozygous for the null allele failed to reach adulthood (Figure S2F). However, we identified no decrease in adult viability for either of the isoform-specific alleles in *trans* to a null allele (Figure S2F). Thus, the long isoform and the poly-Q domain of the short isoform are not individually required for adult viability.

The GAF poly-Q domain is not required for foci formation

Given the previously reported roles of the poly-Q domain in transcriptional activation, protein multimerization and DNA distortion *in vitro*[30,35,36], we were surprised that flies lacking this domain in the short isoform showed no defects in adult viability. To further investigate the functional relevance of this domain, we examined the first few hours of embryonic development when only the short isoform is present. At this time, development is controlled by maternally provided products, and we recently demonstrated that GAF is essential[27]. We tested the viability of embryos that inherited only $GAF^{S\ Q}$ from their mothers. Females heterozygous for the GAF null allele and either GAF^L, $GAF^{S\ PQ}$, or wild-type GAF were mated to w^{1118} males. While 94% of embryos inheriting mRNA encoding wild-type GAF hatched, only 41.6% of embryos inheriting $GAF^{S\ Q}$ hatched (χ^2, p= 2.2×10^{-16}) (Figure 2B). It is possible this effect is caused by additional amino acids translated after the deleted stop codon or a change in GAF protein level caused by the truncation. It is also possible that zygotically expressed wild-type GAF could provide some rescue. However, our results are consistent with the infertility reported in GAF mutants where both the long and short isoform lack the poly-Q domain[37]. Embryos inheriting only the short isoform (GAF^L) had a 92.1% hatching rate, similar to the wild-type control (Figure

2B). This suggests that the poly-Q domain supports, but is not absolutely required, for GAF function.

IDRs can drive protein aggregation and phase transitions, and the poly-Q domain of GAF has been implicated in such functions[12,15,30,36]. Thus, it is possible that the reduced hatching rate for embryos inheriting GAF$^{S \Delta PQ}$ might result from a failure of the mutant protein to concentrate in subnuclear foci. To identify the localization pattern of GAF lacking the poly-Q domain, we engineered the endogenous short isoform poly-Q deletion in the background of a line in which we previously tagged GAF with sfGFP to create *sfGFP-GAF$^{S \Delta PQ}$* (Figure S2G)[27]. In embryos inheriting this allele from their mothers, sfGFP-GAF$^{S \Delta PQ}$ continued to localize to discrete foci, and these foci completely overlapped with full-length GAF endogenously tagged with mCherry at NC14 (Figure 2C). Therefore, the poly-Q IDR is dispensable for the recruitment of GAF to subnuclear foci. However, GAF recruitment may be mediated by a protein-protein interaction between sfGFP-GAF$^{S \Delta PQ}$ and mCherry-GAF. To test if the poly-Q domain is required for foci formation, we imaged embryos in which all GAF protein lacks the poly-Q domain (laid by *sfGFP-GAF$^{S \Delta PQ}$/GAF$^{S \Delta PQ}$* females). When compared to controls laid by *sfGFP-GAF/+* females, we identified no difference in foci formation between wild-type and poly-Q deleted proteins (Figure 2D, S2H). Thus, the poly-Q domain of GAF is not required for foci formation. To determine if loss of the poly-Q domain affected genome occupancy, we performed chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) on stage 5 embryos laid by *sfGFP-GAF$^{S \Delta PQ}$/GAF$^{S \Delta PQ}$* females. (Table S1). Spike-in normalized ChIP-seq data demonstrated that the majority of regions bound by sfGFP-GAF was also occupied by sfGFP-GAF$^{S \Delta PQ}$ (Figure 2E). Together these data reveal that contrary to the expected role of the intrinsically disordered poly-Q domain in protein binding and localization, this domain is not required for formation of subnuclear foci or chromatin binding.

The DNA-binding domain of GAF is necessary and sufficient for localization to foci

Because the poly-Q IDR was dispensable for GAF subnuclear foci formation, we sought to systematically identify the protein domains responsible for driving this localization. We generated transgenic fly lines that drove the expression of sfGFP-tagged GAF in the germline and early embryo and assayed for colocalization with endogenous mCherry-GAF and mitotic retention during the MZT. We verified that transgenic expression of full-length GAF-sfGFP (1-519aa) recapitulated endogenous GAF localization both during mitosis and interphase (Figure 3A, S3A). As a control, we expressed sfGFP with the GAF nuclear localization signal (NLS) and confirmed that it was diffuse in the nucleus during interphase and not mitotically retained (Figure 3A, S3B). Expression of sfGFP-GAF without the poly-Q domain colocalized with endogenous GAF and was mitotically retained, as we observed for endogenous sfGFP-GAF$^{S \Delta PQ}$ (Figure 2C, 3A, S3A). These controls support our use of transgenic expression to identify the domains of GAF required for nonuniform subnuclear distribution.

Due to the reported role for IDRs in mediating protein aggregation, we used PONDR to identify other IDRs in the GAF protein outside of the poly-Q domain (Figure 3B). We identified an additional IDR N-terminal to the DNA-binding domain (DBD), suggesting that

regions outside of the poly-Q domain could facilitate the localization of GAF to subnuclear foci and compensate for the absence of the poly-Q IDR. We therefore systematically made transgenes to express sfGFP-tagged GAF lacking specific domains but retaining an NLS. All transgenes resulted in protein expression (Figure 3C–D, S3). We discovered that all truncated proteins containing the intact DBD localized to foci (Figure 3A, S3A). Expression of the DBD alone localized to a subset of foci enriched for endogenous GAF (Figure 3A, C), demonstrating this domain alone is sufficient to drive localization to many GAF foci. Expression of proteins with either the poly-Q domain or the unstudied IDR from 123aa-310aa (IDR2) along with the DBD were able to localize to all foci marked by full-length GAF (Figure 3A, S3A). Thus, these IDRs may be important in mediating interactions for recruitment to a subset of GAF foci. In addition to being sufficient, the DBD domain was also necessary for localization. Full-length GAF containing point mutations in the single, DNA-binding zinc finger was not localized to foci, but instead was diffusely distributed in the nucleus (Figure 3A, 3D). Although the zinc finger mutant GAF was expressed in a background with endogenous wild-type GAF, the mutant protein was not recruited to GAF foci, demonstrating that protein-protein interactions through other domains does not result in recruitment to foci. Our transgenic assays show that DNA binding rather than protein-protein interactions is necessary for GAF to localize to subnuclear foci.

GAF is retained on mitotic chromosomes and is enriched at pericentric heterochromatin during mitosis[28]. We assayed for mitotic retention and demonstrated that despite localization to foci, the DBD alone was not mitotically retained (Figure 3A, 3E). The full-length zinc finger mutant GAF was also not retained during division (Figure 3A, 3E). Therefore, DNA binding is necessary, but not sufficient for GAF localization to mitotic chromosomes. Addition of either the poly-Q domain or the entirety of the N-terminal portion of GAF to the DBD restored mitotic retention (Figure 3A, S3A). Expression of only the N-terminal IDR with the DBD resulted in a severe reduction in mitotic retention, despite the fact that this protein completely colocalized with endogenous GAF foci during interphase (Figure 3A, S3A). These data suggest that the N-terminal IDR and the poly-Q domain do not function equivalently to promote GAF binding during mitosis.

GAF binds AAGAG satellite repeats in the early embryo

To identify the genomic regions that underlie the GAF foci, we leveraged the sfGFP-tagged DBD, which was sufficient for GAF localization to foci. We performed ChIP-seq using an anti-GFP antibody on stage 5 embryos expressing DBD-sfGFP. Despite successful immunoprecipitation of our spike-in material, there was not sufficient enrichment to call peaks from this dataset, including at regions bound by full-length GAF (Figure 4A). Based on the necessity of the DBD for the localization of GAF to foci and our inability to detect enrichment in our DBD-sfGFP ChIP-seq dataset, we hypothesized that GAF foci might correspond to regions not represented in the reference genome, in particular the simple satellite AAGAG repeats enriched in *Drosophila* pericentric heterochromatin[38]. AAGAG repeats are abundant, comprising ~6% of the *Drosophila* genome and provide a highly concentrated region of the GA-repeat motif that GAF binds[39]. Indeed, in third instar larval brain tissue GAF localizes to these repeats during mitosis but not during interphase[40]. To identify whether GAF bound to these repetitive regions, we calculated the enrichment (IP/

Input) of the percentage of reads that contained the AAGAG repeat from our published GAF-sfGFP ChIP-seq data on stage 3 and stage 5 embryos[27]. The AAGAG repeat was enriched in GAF IPs at both stages when compared to another pioneer factor that functions in the early embryo, Zelda (ZLD), which we would not expect to bind the AAGAG repeat[27] (Figure 4B). At stage 5, the DBD-sfGFP IPs the AAGAG repeats at levels similar to full-length GAF (Figure 4B). We confirmed that the endogenous, full-length GAF foci localize to AAGAG repeats using DNA fluorescent in situ hybridization (FISH) on NC14 embryos using a $(AAGAG)_7$ probe while simultaneously immunostaining for sfGFP-GAF (Figure 4C). Similar experiments on embryos expressing DBD-sfGFP showed that the DBD-sfGFP signal overlapped with AAGAG repeats, further supporting the sufficiency of the DBD for recruitment to AAGAG repeats (Figure S4A,B). Together our data demonstrate that GAF binds to AAGAG repeats in the early embryo, and at these regions the high local concentration of the GAF motif drives the formation of GAF foci.

A subset of GAF foci colocalize with HP1a

Satellite repeats are often silenced in the genome, however AAGAG repeats are expressed in various tissues, and AAGAG RNA is required for viability, the nuclear matrix, and sperm maturation[41,42]. GAF has a variety of roles in transcriptional regulation, including activation, repression, insulator function, and chromatin organization[43,44]. To determine which of these many reported activities GAF may be employing at AAGAG repeats, we identified factors that colocalized with GAF at foci. It has been well established that several types of foci form during interphase in the nucleus, including transcriptionally active hubs, insulator bodies, and heterochromatin domains[13,14,45]. Since we had previously defined a role for GAF as a transcriptional activator during zygotic genome activation, we tested if GAF foci were sites of active transcription[27]. Robust transcription initiates at the histone locus body (HLB) in two early detectable foci in the embryo[11,46,47]. This phase-separated domain promotes high levels of histone gene expression and is marked by localization of the protein Multi Sex Combs (Mxc)[11]. Using a GFP-tagged Mxc, we demonstrated that GAF is not localized to the HLB, consistent with previous data from fixed embryos (Figure S4C)[48]. We then investigated if GAF foci were transcriptionally active hubs outside of the HLB. We imaged nascent transcription of a transgene driven by the regulatory region of the GAF-target gene *tailless* (*tll*)[49]. *tll* is zygotically expressed at NC14, bound by GAF at the promoter, and depends on GAF for transcription[27]. We failed to observe strong colocalization between nascent transcription of this transgenic reporter and GAF foci (Figure S4D). We propose that GAF foci are not transcription hubs, and that transcriptional activation is mediated by the population of GAF that is more diffuse in the nucleus.

In addition to its role in transcriptional activation, GAF functions as an insulator and interacts with several insulator proteins[50–53]. Despite prior reports showing that GAF is in the Large Boundary Complex (LBC) with the insulator protein Mod (mdg4), we did not detect robust colocalization of these proteins (Figure S4E). Nor did we identify colocalization with GAF and another insulator protein, CTCF (Figure S4F). GAF binds to Polycomb Response Elements (PREs) and interacts with subunits of the Polycomb repressive complex 2 (PRC2)[51,54–56]. As might be expected, we observed some

colocalization of small GAF foci with Polycomb (Pc). However, the majority of GAF and Pc foci did not overlap (Figure S4G).

Similar to other proteins classified as insulators, GAF is enriched at topologically associating domain (TAD) boundaries, suggesting GAF might regulate 3D chromatin structure[54,57]. GAF has been implicated in chromatin looping *in vitro* and *in vivo*[58–60]. To uncover changes in genomic contacts in the absence of GAF, we performed Hi-C on embryos in which GAF was degraded (GAF[deGradFP]) and control embryos at NC14[27]. The majority of 3D contacts were similar between the two conditions, indicating that GAF is not required for TAD formation during the MZT (Figure S5A). Chromatin is broadly divided into euchromatic (A) compartments, and heterochromatic (B) compartments[61], and we did not identify clear differences in this compartmentalization in the absence of GAF (Figure S5B). While the 3D organization of the genome remained largely unchanged in the absence of GAF, we identified a small subset of loops that were lost when GAF was degraded, including at the *Antp* locus (Figure S5C). GAF binds to the anchors of this loop and facilitates tethering the enhancer to the promoter at this locus[58,62]. Together our colocalization assays and Hi-C data show that GAF foci are unlikely to be either insulator or Polycomb bodies, and that GAF is not required for TAD formation or compartmentalization at NC14.

We next investigated whether GAF foci correspond to constitutive heterochromatin domains. Heterochromatin protein 1 (HP1a) forms phase separated domains in both *Drosophila* and mammals, and these domains contribute to the repression of heterochromatic regions through the selective concentration of silencing factors and exclusion of activating factors[10,17]. We identified robust colocalization of sfGFP-GAF and HP1a-RFP (Figure 5A), consistent with previous reports from fixed embryos[63,64]. Despite this clear overlap at a subset of foci, in each nucleus there were GAF foci that did not contain HP1a, and HP1a foci that did not contain GAF (Figure 5A). Given the limited resolution of our confocal images, we used lattice light-sheet microscopy to determine the degree of colocalization more robustly between these two proteins. We confirmed the colocalization of GAF and HP1a, but the increased resolution delineated subdomains within colocalized regions enriched for either GAF or HP1a (Supplemental movie S1).

During the MZT, mature, constitutive heterochromatin domains are progressively established. It is not until NC13–14 that repetitive elements accumulate the repressive histone modification H3K9me3 and display the late replication characteristic of silenced regions[65]. HP1a, which binds to H3K9me3, begins to form small foci at NC11, and these HP1a domains mature into robust phase-separated domains during NC14[10]. In intriguing contrast, GAF foci were clearly visualized starting at NC10. To determine the precise relationship between the timing of HP1a and GAF foci formation, we imaged sfGFP-GAF and HP1a-RFP in an embryo as it developed from NC10 through NC14. This analysis confirmed that GAF formed robust foci earlier than HP1a (Figure 5B, Supplemental movie S2). Because GAF is mitotically retained at pericentric heterochromatin while HP1a is not and forms foci prior to HP1a (Figure 5B, Supplemental movie S2), we hypothesized that GAF might be necessary for HP1a recruitment to foci. By imaging HP1a-RFP in GAF[deGradFP] embryos in which maternal GAF is degraded[27], we showed that HP1a was

not dependent on GAF for foci formation (Figure S5D). We then tested if HP1a might be important for GAF foci formation by imaging sfGFP-GAF in embryos with HP1a knocked down by RNAi, and again did not detect any notable differences in foci formation (Figure S5E,F). Thus, while GAF and HP1a colocalize, they can independently form subnuclear foci.

Given that GAF forms foci at AAGAG repeats, we wondered if GAF was required for HP1a recruitment specifically to these repeats. To determine whether GAF and HP1a colocalize at AAGAG satellite repeats, we performed DNA FISH on NC14 and immunostained for HP1a. At NC14, both HP1a and the AAGAG repeats colocalize in the apical domains of the nuclei, reflecting the stereotypical Rabl configuration (Figure 5C). Some HP1a-enriched regions do not overlap the AAGAG repeats, likely representing additional heterochromatic regions. We performed DNA FISH for AAGAG repeats and immunostained for HP1a in embryos lacking GAF. During the MZT, HP1a remained colocalized with AAGAG repeats in the absence of GAF (Figure 5C). Therefore, GAF is not absolutely required for HP1a localization to AAGAG repeats. Having established that many GAF foci correspond to HP1a-enriched AAGAG satellite repeats, we next wanted to determine what contribution, if any, GAF has to the formation of transcriptionally silent heterochromatin at these repeats.

GAF promotes H3K9me3 and represses transcription of AAGAG repeats

We tested if GAF regulated the establishment of H3K9me3, a histone modification instructive in the formation of HP1a-enriched constitutive heterochromatin. H3K9me3 is first detected at significant levels at NC14[65]. Thus, GAF binding to AAGAG repeats at NC10 might be instrumental to the formation of heterochromatin through promoting H3K9me3. We therefore investigated the enrichment of H3K9me3 at GAF-bound AAGAG repeats during NC14. We observed colocalization of many GAF foci with both AAGAG repeats and H3K9me3 signal at NC14 (Figure 6A). ChIP-seq for H3K9me3 at NC14 in control and GAF[deGradFP] embryos showed a robust decrease of H3K9me3 at pericentric heterochromatin when GAF was degraded. (Figure 6B). Consistent with our Hi-C results, in the GAF[deGradFP] embryos the Rabl conformation of the NC14 nucleus and the localization of AAGAG repeats to foci in the apical region was not disrupted (Figure S5G). The AAGAG repeats remain localized in discrete foci as determined by quantitative analysis of DNA FISH in the presence and absence of GAF, suggesting that GAF is not required for their compaction at NC14 (Figures S5H). To determine if the loss of H3K9me3 was specific to GAF-bound AAGAG repeats, we focused on the centromere of the 3rd chromosome which is enriched for the dodeca satellite and has relatively few AAGAG repeats[38]. In contrast to most of the genome in which there was a dramatic loss in H3K9me3, at this centromere H3K9me3 signal was largely unchanged between control and GAF[deGradFP] embryos. Nonetheless, loss of H3K9me3 was evident at regions enriched for the AAGAG motif (Figure 6B). Because simple satellite repeats are not well represented in the reference genome assembly, we also analyzed the raw reads from the H3K9me3 ChIP-seq data. In GAF[deGradFP] replicates there was a decrease in the enrichment of reads containing the AAGAG repeat compared to controls (Figure 6C). By contrast, when we performed the same analysis with the dodeca repeat, there was little difference between the GAF[deGradFP]

replicates and controls (Figure 6C). These data support a role for GAF in promoting the deposition of H3K9me3 at AAGAG repeats during the MZT.

Our data suggest that GAF is concentrated at AAGAG satellite repeats and is instrumental in promoting heterochromatin formation. We therefore directly tested whether GAF was essential to silence transcription from AAGAG repeats. Failure to silence transcription from repetitive elements or establish heterochromatin at repetitive elements during the MZT leads to developmental defects[66,67]. To test the necessity of GAF for silencing repeat expression, we performed RNA FISH using an (AAGAG)$_7$ repeat probe during NC14. We detected low levels of transcription from AAGAG repeats in control embryos, consistent with previous reports[41,42] (Figure 6D). By contrast, in GAF[deGradFP] embryos there was a robust increase in the RNA FISH signal from AAGAG repeats (Figure 6D). An increase in transcription of AAGAG repeats was not observed in embryos with only GAF[S PQ] maternally deposited, indicating that GAF does not require the poly-Q domain to repress transcription from AAGAG repeats (Figure 6D). Analysis of total RNA-seq from GAF[deGradFP] and control embryos at NC14 verified the increase in expression from AAGAG satellites in the absence of GAF[27] (Figure 6E). This effect was specific to GAF-bound repeats, as transcript abundance for several other simple repeats present in pericentromeric regions was unchanged in GAF[deGradFP] embryos compared to controls (Figure 6E). The increased levels of AAGAG repeats observed are likely indicative of nascent transcription as RNA Pol II is robustly localized with the transcripts in GAF[deGradFP] embryos (Figures S6A).

To test if the reduction in H3K9me3 levels observed in the GAF[deGradFP] embryos could promote this expression, we performed RNA FISH coupled with antibody staining for H3K9me3. H3K9me3 staining did not overlap with the RNA FISH signal from AAGAG repeats in NC14 GAF[deGradFP] embryos (Figure S6B). This suggests that loss of H3K9me3, which normally decorates these satellites, creates an environment permissive for transcription. Because H3K9me3 is not evident until NC14, we tested whether GAF was instrumental in repressing transcription of the AAGAG repeats prior to this time in development. Beginning at NC12, AAGAG expression is higher in GAF[deGradFP] embryos compared to controls, and AAGAG expression in GAF[deGradFP] embryos increases over successive nuclear cycles (Figure S6C). This is consistent with the more permissive transcriptional environment that is established at NC14 compared to earlier nuclear cycles[7,8] Together, our data demonstrate that GAF foci correspond to AAGAG satellites where GAF is concentrated by DNA binding to highly abundant GA-rich motifs and is required for robust methylation of H3K9 and transcriptional silencing during zygotic genome activation.

Discussion

GAF subnuclear domains are driven by DNA binding

Our data demonstrate that the formation of most high-concentration GAF foci is driven by sequence-specific DNA binding rather than IDR-mediated multivalent interactions. This result was unexpected because in eukaryotes IDRs are often instructive to *in vivo* transcription-factor binding[68–70]. IDRs are also important in the organization of subnuclear bodies, which can influence transcription-factor function by concentrating proteins at discrete genomic locations[23,24,71–74]. In contrast to these known roles for IDRs, we showed

that GAF IDRs are largely dispensable for concentration to subnuclear foci and that instead this localization was driven by the DBD. Because GAF foci correspond to GA-rich satellite repeats, we propose that the high density of the GAF DNA-binding motif concentrates GAF at these genomic loci. While multivalent interactions mediated by IDRs drive protein aggregation in many systems[12–14], recent work has begun to highlight the importance of additional domains to this process. For example, the DBD of the human reprogramming factor KLF4 is both necessary and sufficient to form condensates in the presence of DNA[75]. DNA promotes surface condensation of KLF4, which can occur with a low saturation of molecules due to the local high density created at the DNA surface[76]. Through similar but distinct mechanisms, nucleation at a specific DNA sequence is suggested to drive the formation of phase-separated domains, such as the histone locus body and nucleolus[11,77,78]. Our data support the importance of DNA binding in driving the formation of subnuclear regions of high protein concentration and highlights the complexity of mechanisms that can result in a nonuniform distribution of factors in the nucleus. This work reinforces that to determine if nonuniform protein distribution is driven by IDR-mediated phase separation it is essential to carefully analyze proteins expressed at endogenous levels and determine what portions of the protein are required for this distribution.

While we demonstrated that GAF DNA binding was required for localization to foci, we showed that not all endogenous GAF foci are occupied by the DBD alone and that the IDRs promoted localization to these foci. Recent single-molecule studies determined that regions outside the DBD, including the poly-Q IDR, are required for GAF to stably engage the genome[37]. Thus, we propose that IDRs are essential for stabilizing GAF binding at a subset of loci, potentially through multivalent protein interactions.

GAF supports heterochromatin establishment at AAGAG repeats

Despite the essential nature of heterochromatin, much remains enigmatic about how heterochromatin is established *de novo* during the MZT. The repressive histone modification H3K9me3 is not detectable until NC14, and HP1a does not form mature domains until the same time in development[10,65]. However, chromatin compaction is detected at satellite repeats beginning as early as NC8[79]. Furthermore, different satellite repeats accumulate H3K9me3 and HP1a at distinct time points in development and via different HP1a-recruitment mechanisms[65,80]. We demonstrated that GAF forms foci as early as NC10, prior to HP1a phase separation and H3K9me3 accumulation on chromatin. GAF foci are formed at AAGAG satellite repeats, and by NC14 these regions also colocalize with HP1a. This temporal relationship demonstrates that GAF binding to GA repeats precedes H3K9me3 deposition and HP1a enrichment. We propose that GAF binds GA-rich satellite repeats early in development and recruits heterochromatin factors to these regions. This is supported by the reduction in H3K9me3 and aberrant transcription of GA repeats in the absence of GAF. The recruitment of heterochromatin factors to GA-rich repeats is likely mediated through the BTB/POZ protein interaction domain of GAF, as we demonstrated that GAF continues to repress transcription from AAGAG repeats in the absence of the poly-Q domain. Intriguingly, HP1a recruitment to AAGAG repeats at NC14 is maintained in the absence of GAF, despite a reduction in H3K9me3 levels. It is possible that HP1a is recruited to these sites through a methylation-independent mechanism, or that the reduced levels of

H3K9me3 are sufficient for some HP1a recruitment. However, it is clear from the aberrant transcription of the AAGAG repeats that heterochromatin formation is compromised at these regions in the absence of GAF.

GAF binds AAGAG repeats during interphase in the embryo, and mitotic retention of GAF to pericentric heterochromatin regions requires DNA-binding activity. Together these data suggest that GAF is bookmarking AAGAG binding sites throughout mitosis. This is consistent with published data demonstrating that GAF mitotically bookmarks euchromatic binding sites in the early embryo[64]. Mitotic retention of GAF at AAGAG repeats requires regions outside of the DBD, and our data suggest that the IDR2, poly-Q domain, and N-terminal portion of the GAF protein are able to rescue mitotic retention to varying degrees. This is likely because the poly-Q domain and BTB/POZ domain of GAF promote GAF binding stability on DNA[37]. The BTB/POZ domain in particular strongly promotes stable GAF binding, which is likely why the complete N-terminus of GAF is required to achieve wild-type levels of GAF retention (Figure S3A)[37]. The ability of GAF to remain bound to AAGAG targets throughout mitosis may contribute to the recruitment of heterochromatic factors to these regions during the rapid mitotic cycles of early embryo development. A similar function was described for Prospero during neural differentiation[81]. Together with prior data investigating the mechanisms of heterochromatin establishment, our data support a model whereby the timing of heterochromatin formation at distinct satellites may depend on sequence-specific binding factors that can recruit silencing machinery.

The role of GAF in heterochromatin formation is consistent with the nuclear defects observed in GAF-depleted embryos during the MZT[27]. Anaphase bridges and micronuclei reminiscent of the division defects observed in GAF[deGradFP] embryos are reported when D1 chromosomal protein (D1) and Proliferation disrupter (Prod) are depleted[82,83]. D1 and Prod bind sequence specifically to repeats and function to spatially condense and organize repetitive regions in the genome, allowing for the formation of the chromocenter. Based on this evidence, it has been suggested that although there is little sequence conservation between closely related species, satellite repeats have a conserved function in recruiting proteins that facilitate the bundling of heterochromatin from multiple chromosomes into organized subnuclear domains[82,83]. AAGAG repeats continue to localize in foci at the heterochromatic apical regions of nuclei in NC14 GAF-depleted embryos, and our Hi-C and imaging data indicate that the loss of GAF does not affect organization of heterochromatin and euchromatin into distinct subnuclear compartments. This suggests that GAF is not essential for global genome organization. GAF might have a role at repeats separate from genome organization, or another factor, such as D1 and Prod, may function redundantly with GAF to preserve spatial organization of repetitive regions when GAF is depleted. This redundancy is supported by the fact that D1 and Prod are partially redundant to each other in chromocenter formation[82]. Alternatively, GAF knockdown may compromise spatial organization of repetitve regions in a subset of embryos, and this may cause catastrophic failure of nuclear divison and embryo death prior to NC14. We selected for NC14 embryos for analysis, as this is when distinctive marks of heterochromatin are clear. This resulted in the exclusion of dying embryos with dramatic defects in nuclear division, which might be caused by perturbed chromatin organization in the absence of GAF. Our data support an essential role for GAF in the formation of heterochromatin at AAGAG satellite repeats,

but additional studies will be required to determine the role of GAF in the maintenance of a prolonged silenced state. These future studies will have important implications for our understanding of heterochromatin formation more generally as the role of GAF in heterochromatin formation may be conserved. Notably, cKrox, the mouse ortholog of GAF, drives localization of GA-rich loci to heterochromatic domains on the nuclear periphery[84].

Mechanisms of distinct GAF functions

Here, we identified a previously unappreciated role for GAF in repression of satellite repeats. This was unexpected as our prior work, and that of others, had largely focused on the role of GAF as a pioneer factor essential for activating transcription[27,37,64,85–87]. Together our studies demonstrate that in a single nucleus at one point in development, GAF is functioning to establish both the silent and active transcriptional state. How GAF can perform these two opposing functions remains unclear but based on our data we propose that regions of high GAF density result in transcriptional silencing. By contrast, more diffuse populations of GAF may promote transcriptional activation. This model of protein density driving changes in transcription-factor function is exemplified by the oncogenic transcription factor EWS::FLI1. EWS::FLI1 can function as either an activator or a repressor depending on the level of low-complexity domain interaction[71]. At endogenous levels, EWS::FLI1 forms hubs at activated target genes. Induction of phase separation at EWS::FLI1 hubs by overexpression of low-complexity domains results in the silencing of genes normally activated by EWS::FLI1. The authors suggested a model where hub formation is precisely tuned for target-gene activation, with too few or too many IDR-mediated multivalent interactions resulting in gene repression. Therefore, the high concentration of GAF at satellite repeats may similarly result in repression, while lower concentrations of GAF promote transcription.

GAF interacts with a variety of co-factors that function in transcriptional activation and repression[51]. It is possible increased protein concentration results in repression via the selective trapping or exclusion of specific cofactors. For example, the high density of GAF at GA-repeats may result in the sequestration of repressive factors in these foci and the exclusion of activating factors. Indeed, phase separated domains, like heterochromatic HP1a, have been demonstrated to have this functionality[10]. GAF is also post-translationally modified, and this could provide another level of regulation to control GAF function[88–90]. These modifications might selectively regulate the ability of GAF to interact with co-factors or directly influence recruitment of GAF to specific subnuclear domains.

Along with prior studies, we show that GAF is distinctive as it functions broadly in both transcriptional activation and repression during the dynamic genomic reprogramming required for early development. We propose that the non-uniform distribution of GAF in the nucleus enables it to have opposing transcriptional roles within a single nucleus. Our data support a model in which a subset of transcription factors are important for organizing the nucleus into functionally distinct domains based on their subnuclear concentration, and this is essential for regulating gene expression during dramatic changes in cell fate.

Limitations of the study

We demonstrated that GAF is required for H3K9me3 and silencing of AAGAG satellite repeats, but the mechanism of silencing is unclear. Prior studies identified interactions between GAF and H3K9 methyltransferases Eggless and Su(var)3–9[44]. Nonetheless, our data does not demonstrate whether GAF is directly or indirectly promoting methylation and silencing. Similarly, while we showed that the poly-Q domain of the short isoform is important in promoting early embryo viability, it is not clear what function of GAF is inhibited in the absence of this domain. Based on single-molecule studies, it is likely that the poly-Q domain helps stabilize GAF genomic occupancy and this may be important for the pioneering role of GAF at euchromatin[37]. Identifying changes in chromatin structure and gene expression in the GAF poly-Q deletion should begin to address the role of this domain in GAF function in the early embryo. In addition to these compelling questions, the role of distinct GAF domains in promoting mitotic retention remains to be investigated.

STAR Methods

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Melissa Harrison (mharrison3@wisc.edu).

Materials availability—Reagents generated in this study, including *Drosophila* strains and plasmids, are available upon request to the lead contact.

Data and code availability—Sequencing data have been deposited in GEO under accession code GSE218020 and are publicly available.

Analysis code for lattice light-sheet microscopy analysis is available here: DOI:10.5281/zenodo.8089784 or https://gitlab.com/mir-lab/publications/gaf-llsm-analysis

Python scripts for all Hi-C analyses are available at DOI:10.5281/zenodo.8092663 or https://github.com/michaelrstadler/hic

Any additional information is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

***Drosophila* husbandry and housing**—All stocks were grown on standard molasses food at 25°C in an incubator with a 12 hour light/ 12 hour dark cycle.

***Drosophila* strains and genetics**—Fly strains used in this study: *w^1118^*, *w;His2Av-mRFP (III)* (Bloomington *Drosophila* Stock Center (BDSC) #23650), *w;mat-α-GAL4-VP16 (II)* (BDSC #7062), *w;Trl^s2325^/TM3* (BDSC #12088*), yw; P{CTCF-GFP.FPTB}* (BDSC #64810), *w;PBac{GFP-mod(mdg4).S}* (BDSC #51351), *yw;PBac{mxc-GFP.FPTB}* (BDSC # 84130), *w; P{Pc-EGFP}* (BDSC #9593), *His2AV-RFP(II);MCP-GFP (III)* [49], *UAS-dsRNA-Su(var)205* (BDSC #36792), *HP1a-RFP* (II)[91] (J. Lipsick, G. Karpen), *sfGFP-GAF*

(N) and *nos-degradFP (II)* were generated previously by our lab and are described in Gaskill et al.[27].

Transgenic lines: GAF truncation GFP-tagged transgenic flies were made using the sequence from the short GAF isoform (519aa). The NLS of GAF (209–217aa) was added to the C-terminus of sfGFP for the truncations that lacked the endogenous NLS. Full length (1–519aa)-sfGFP, DBD-PQ (311–519aa)-sfGFP, PQ (426–519aa)-sfGFP, sfGFP-BTB/POZ-DBD (1–391aa), sfGFP-BTB/POZ (1–122aa), IDR-DBD (123–391aa)-sfGFP, Full-length zinc finger mutant (1–519aa, C344S, C347S)-sfGFP were made by PhiC31 integrase-mediated transgenesis into the PBac{yellow[+]-attP-3B}VK00037 docking site (BDSC #9752) by BestGene Inc. DBD (310–391aa)-sfGFP was made by PhiC31 integrase-mediated transgenesis into attP40 (25C6) docking site by BestGene Inc. All transgenes were cloned using Gibson assembly into an attB vector with the *nanos* promoter and 5'UTR that was used to generate transgenes in Gaskill et al.[27].

To generate the MS2 transgene driven by the *tll* promoter, 3.3kb of the *tll* regulatory region was cloned upstream of 24x MS2 loops using Gibson assembly into an attB vector. Transgenes were made by PhiC-mediated Recombinase Mediated Cassette Exchange (RMCE) into the P{attP.w[+].attP}JB38F docking site (BDSC #27388) by BestGene Inc.

The following GAF mutant alleles were generated using Cas9-mediated genome engineering (outlined in detail below): GAF^L, GAF$^{S\ PQ}$, mCherry-GAF, and sfGFP-GAF$^{S\ PQ}$

To obtain the embryos for microscopy in a *Su(var)205* knockdown background, we crossed *mat-α-GAL4-VP16 (II)/CyO, sfGFP-GAF (N)(III)* to *UAS-dsRNA-Su(var)205 (*BDSC #36792) *(II).* Resulting *mat-α-GAL4-VP16/UAS-dsRNA-Su(var)205 (II), sfGFP-GAF (N)/ +(III)* females were crossed to their siblings, and their embryos were collected.

METHOD DETAILS

Cas9-genome engineering—Cas9-mediated genome engineering as previously described[92] was used to generate the N-terminal mCherry-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 mCherry tag and GAF N-terminal open reading frame. The mCherry sequence was placed downstream of the GAF start codon. A 3xP3-DsRed cassette flanked by the long-terminal repeats of PiggyBac transposase was placed in the second GAF intron for selection. The guide RNA sequence (TAAACATTAAATCGTCGTGT) was cloned into pBSK with U63 promoter using inverse PCR. Purified plasmid was injected into embryos of *yw; attP40{nos-Cas9}/CyO* 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.

To generate *GAF^L*, *GAF$^{S\ PQ}$*, *sfGFP-GAF$^{S\ PQ}$* lines, single-stranded oligodeoxynucleotide (ssODN) donors containing the desired mutations were produced by Integrated DNA Technologies. To generate *GAF^L*, a stop codon was introduced at the beginning of the 6th exon of the long isoform. Immediately downstream of the stop codon a HindIII site

(AAGCTT) was generated, and additional mutations were made in the seed region of the guide site in the 5th GAF intron to prevent Cas9 nuclease recutting. To generate *GAF$^{S\ PQ}$*, a 297bp deletion of exon 5 was created, removing sequence unique to the short isoform. To generate the deletion, two guide sites were used flanking the 297bp deletion. The deletion disrupted one guide site, and a mutation was created in the seed region of the other guide site in the 3'UTR of the short isoform. The guide RNA sequences (GAF^L - GCGGCAGTCTTCTCACCAGC) (GAF$^{S\ PQ}$ - AGCCTTCAATCATTCCAACG and ACGAGAGTGATATCGAATGC) were cloned into pBSK with the U63 promoter using inverse PCR. The ssODN and guide RNA plasmids were injected into embryos of *yw; attP40{nos-Cas9}/CyO* by BestGene Inc. Lines were screened using PCR and HindIII digestion for GAF^L, and PCR screening for the 297bp deletion for GAF$^{S\ PQ}$. The regions were then sequenced to confirm mutation without errors. *sfGFP-GAF$^{S\ PQ}$* was created identically to *GAF$^{S\ PQ}$* except the guide plasmids and ssODNs were injected into *nos-Cas9 (II)/sfGFP-GAF(N) (III)* embryos provided to BestGene Inc.

In both *sfGFP-GAF$^{S\ PQ}$* and *GAF$^{S\ PQ}$* lines, the 297bp deletion included the short isoform stop codon. Both mutant lines create a truncated protein product. Based on our sequencing of the mutant lines, we have predicted the additional amino acid sequence that would be translated from the transcript until the closet stop codon.

GAF$^{S\ PQ}$: 424 – PPPAEPSIIPTHQRHHHPHFQKNIKKKNITLTKTICK*

sfGFP-GAF$^{S\ PQ}$: 423 – THLQPSLQSFQRTNDTIIHISKKTLKKKT*

Confocal microscopy—Embryos were dechorionated in 50% bleach for 2 min and subsequently mounted on a hydrophobic membrane coated in heptane glue. Embryos were covered in halocarbon 27 oil prior to the addition of a coverslip. Embryos were imaged on a Nikon A1R+ confocal using a 100x objective at the University of Wisconsin-Madison Biochemistry Department Optical Core. Nuclear density, based on the number of nuclei/2500 μm^2, was used to determine the cycle of pre-gastrulation embryos. Nuclei were marked with His2AV-RFP. Image J[93] was used for post-acquisition image processing. For all images a single z-plane is shown, except Figure 5A and Figure 2D, which are maximum intensity projections of multiple z-stacks.

Lattice light-sheet microscopy data acquisition—Lattice light-sheet microscopy was performed as described previously[94]. A 5mm glass coverslip was rendered adhesive by deposition of a small drop of glue solution prepared by dissolving a roll of double-side scotch tape in heptane. The glue solution was allowed to completely dry before embryos were introduced. Embryos were collected from cages over a 90-minute laying period and arranged on a 5mm diameter glass cover slip. Lattice Light-Sheet Microscopy was performed using a home-built implementation of the instrument following designs from the Betzig Lab [29]. An 80 beam multi-bessel lattice was generated with inner and outer numerical apertures of 0.4 and 0.3 respectively. The sheet was dithered over a 10 μm range in 25 nm steps during each exposure to create a uniform excitation profile. 488 nm and 561 nm lasers were used to excite GFP and RFP respectively, with laser powers measured to be 925 μW for 488 nm and 120 μW for 561 nm at the back aperture of the detection objective.

Two Hamamatsu ORCA-FusionBT digital CMOS cameras (C15440-20UP) were used for detection. An image splitting long-pass dichroic (Semrock FF-560) was placed in between the two cameras to separate emission wavelengths of over and under 560 nm, bandpass filters Semrock FF01-525/50 for GFP and Semrock FF01-593/46 for RFP were placed in front of the cameras. Images at each excitation wavelength were acquired simultaneously at each z-plane with an exposure time of 70 ms for each channel and a 3 second pause between volume acquisitions, these settings resulting in an 8 second interval between volumes.

Chromatin immunoprecipitation—Chromatin immunoprecipitations were performed as described previously[95]. ChIP was performed using the anti-GFP antibody (abcam 290) on embryos from *sfGFP-GAF$^{S\ PQ}$/GAF$^{S\ P}$* and *sfGFP-GAF/+* females and *nos-DBD-sfGFP* heterozygous females. ChIP was performed using the anti-H3K9me3 antibody (Active motif 39161) on *sfGFP-GAF(N)* homozygous and GAF^{deGradFP} embryos.

Briefly, 200–400 embryos from 2–2.5 hr (*nos-DBD-sfGFP, sfGFP-GAF(N)* homozygous and GAF^{deGradFP}) or 2–3 hr (*sfGFP-GAF$^{S\ PQ}$/GAF$^{S\ P}$* and *sfGFP-GAF/+)* lays were collected, dechorionated in 50% bleach for 3 min, fixed for 15 min in 4% formaldehyde and hand-sorted by morphology to ensure they were stage 5. Embryos were 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 sonicated for 20 s 11 times at 20% output and full duty cycle (Branson Sonifier 250). At this point sheared spike-in chromatin from H3.3-GFP mouse cells (sfGFP-GAF$^{S\ PQ}$, sfGFP-GAF, and DBD-sfGFP ChIP) or *D. virilis* embryos (H3K9me3 ChIP) was added to the sonicated chromatin. Chromatin was incubated with 6 μg of anti-GFP antibody (abcam #ab290) or 10 μl anti-H3K9me3 antibody (Active Motif, # 39162) overnight at 4°C and then bound to 50 μl of Protein A magnetic beads (Invitrogen). The purified chromatin was 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. sfGFP-GAF$^{S\ PQ}$ ChIP-seq libraries were sequenced on the Illumina NextSeq 500 using 75bp single-end reads at the Northwestern Sequencing Core (NUSeq). DBD-sfGFP ChIP-seq libraries were sequenced on the Illumina HiSeq 4000 using 50bp single-end reads at the Northwestern Sequencing Core (NUSeq). H3K9me3 ChIP-seq libraries were sequenced on the Illumina NovaSeq 600 using 150bp paired-end reads at the UW Madison Biotechnology Center.

Adult phenotyping and viability assays—*GAF$^{S\ PQ}$* and *GAF^{L}* mutant alleles were assayed in *trans* to the GAF null mutation, *Trl^{s2325}*, to verify that any identified phenotypes were the result of a mutation in *Trl* and not a background mutation on the same chromosome. Trans-heterozygous adult progeny were checked for phenotypes and crossed to *w^{1118}* to determine fertility.

For the viability assays, three to five heterozygous males and five to 10 heterozygous females of the indicated genotypes were mated in standard molasses vials with dry yeast and flipped twice at 2-day intervals. Two days after the final flip, the adult flies were cleared from the vials, and their progeny were allowed to reach adulthood. The *Trl^{s2325}* allele was

used as the GAF null. Over 900 adults were counted for each cross. The ratio of TM3 and non-TM3 adults was determined and the $\chi 2$ value was calculated for each cross, correcting for the observed ratio from the GAF null/wild-type cross.

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 plate. Unhatched embryos were counted 26 hours or more after embryos were picked.

Antibody generation and purification—An N-terminal GAF antibody recognizing the first 130 amino acids of GAF was used for immunoblotting. To generate these antibodies, rabbits were immunized by Covance, Inc., with maltose binding protein (MBP) fused to amino acids 1–130 of GAF and purified against the same portion of the protein fused to glutathione S-transferase (GST). Similar to other anti-GAF antibodies[31,96], this antibody recognizes the short GAF isoform band at approximately 70 kD in an immunoblot on w^{1118} overnight embryo extract (Figure S2E).

Whole-embryo immunostaining—Embryos were dechorionated and added to 4% formaldehyde in 1x PBST (0.1% Triton-X) with an equal volume of heptanes. Embryos were rocked for 20 minutes at room temperature. The aqueous layer was removed, and methanol was added. Embryos were vortexed for 30 seconds, and all liquid was removed after the embryos settled. Embryos were washed 3x with methanol and 3x with 100% ethanol and stored in ethanol at −20C until use. To rehydrate embryos, they were washed sequentially for 10 min at room temp in 50% EtOH 50% PBST, 25% EtOH 75% PBST, and 100% PBST. They were then incubated overnight at 4C with the primary antibodies in PBST (Active motif (2AG-6F12-H4) anti-H3K9me3 (1:100)) (Abcam 290 anti-GFP (1:500)). The next day embryos were washed 3x for 5 minutes in PBST and incubated for 1.5 hours at room temp with secondary antibodies (Dylight 550 goat anti mouse 1:400 and Dylight 488 goat anti rabbit 1:400). Embryos were washed 3x for 5 minutes in PBST, then for 10 minutes in PBST + DAPI. Embryos were finally washed in for 1 minute in PBS and mounted in 70% glycerol and 1x PBS.

Immunoblotting—Proteins were transferred to 0.45 μm Immobilon-P PVDF membrane (Millipore) in transfer buffer (25 mM Tris, 200 mM Glycine, 20% methanol) for 60 min 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-GAF (1:250, this study), anti-GFP (1:2000, Abcam #ab290, #ab6556), anti-Tubulin (DM1A) (1:5000 Sigma #T6199), anti-HP1a (1:50, DSHB, C1A9) overnight at 4°C. The secondary incubation was performed with goat anti-rabbit IgG-HRP conjugate (1:3000 dilution, Bio-Rad #1706515) or anti-mouse IgG-HRP conjugate (1:3000 dilution, Bio-Rad # 1706516) for 1 hour at room temperature. Blots were treated with SuperSignal West Pico PLUS chemiluminescent substrate (Thermo-Scientific) and visualized using the Azure Biosystems c600 or Kodak/Carestream BioMax Film (VWR).

cDNA screening—Ten overnight embryos from the indicated genotype were picked into Trizol (Invitrogen #15596026) with 200 μg/ml glycogen (Invitrogen #10814010). RNA

was extracted and cDNA was generated using Superscript IV (Invitrogen). cDNA was diluted 1:10 and used for PCR. Two primer sets were used. Primer set 1 (F primer-CCTTTCTGCTGGACTTGCTAAAG, R primer- CGGATTGTGCCACCAGTT) amplified both the long and short isoform transcripts. With these primers, the long isoform transcript band is 1522 bp, the WT short isoform band is 2034 bp, and the truncated short isoform band is 1737 bp. Primer set 2 (F primer- CGACCAAGACCAACTGATTGC, R primer-GAACACAAATCATTCGATCAGATC) amplified only the short isoform transcript. With these primers the WT short isoform band is 1310 bp and the truncated short isoform band is 1013 bp. Bands marked with an asterisk were excised, purified, and sequenced to confirm they were the short or long transcripts.

Hi-C experimental procedure—500 *sfGFP-GAF(N)* homozygous or GAF[deGradFP] hand-sorted 2–2.5 hr AEL embryos were used as input for each replicate. Hi-C experiments and initial data processing were performed as described previously [97] with minor modifications: the restriction enzyme MluCI (ÂATT, NEB R0538L) was used, as we found that it gives more even coverage of the AT-rich *Drosophila* genome than GATC-cutters, and IDT xGen adaptors were used in place of Illumina adaptors.

DNA fluorescent in-situ hybridization

Fixation: DNA FISH was performed on 1.5–3 hr AEL embryos as described previously[98]. Briefly, embryos were dechorionated and transferred to buffer A (60mM KCl, 15 mM NaCl, 0.5 mM spermidine, 0.15 mM spermine, 2 mM EDTA, 0.5 mM EGTA and, 15 mM PIPES at pH7.4 made fresh) with 4% paraformaldehyde. Equal amount of heptane was added followed by 25 minutes of incubation on an orbital shaker at max speed. Fixed embryos were devitellinized, washed twice in 100% methanol, and stored at −20°C.

Hybridization: Embryos were rehydrated and incubated with 200 μg/ml RNAse A in PBT at 4°C on a rotating wheel overnight. Next day, embryos were incubated in PBS with 0.3% Triton X-100 (PBS-Tr) for 1 hour before being gradually acclimated to 100% pre-hybridization mixture (pHM) (50% Formamide, 4x SSC, 100 mM NaH_2PO_4, pH7.0, and 0.1% Tween 20). Once in 100% pHM, embryos were incubated for 15 min at 80°C. The DNA probe 5Cy5/(AAGAG)$_7$ was denatured in FISH hybridization buffer (FHB) (50% Formamide, 10% Dextran sulfate, 2x SSC, Salmon sperm DNA 0.5 mg/ml (0.05%)) for 10 minutes at 90°C. Samples hybridized to the probe overnight in a thermomixer set to 37°C with 450 rpm of agitation. Embryos were washed in (1) 50% Formamide, 2x SSC, and 0.3% CHAPS x 2; (2) 40% Formamide, 2x SSC, and 0.3% CHAPS; (3) 30% Formamide, 70% PBT; and (4) 20% Formamide, 80% PBT for 20 minutes each in a thermomixer set to 37°C and 850 rpm. Washes continued with 10% Formamide, 90% PBT, 100% PBT, and 100% PBS-Tr for 20 minutes each at room temperature on a rotating wheel.

Immunostraining: Embryos were processed for immunostaining by first incubating on a rotator in blocking solution of 3% BSA in PBS-Tr for two hours at room temperature. Primary antibodies were diluted in blocking solution (1:500 Rabbit anti-GFP Abcam 290; 1:50 Mouse anti-HP1a DSHB, C1A9) and incubated with embryos at 4°C overnight. Embryos were then washed in PBS-Tr 3x for 5 minutes and 3x for 20 minutes at room

temperature while rotating in between each wash. Secondary antibody was diluted at 1:1000 in blocking solution (Goat Anti-Rabbit IgG DyLight 488 conjugated; Goat Anti-Mouse IgG Dylight 550 conjugated) and allowed to incubate with the embryos for 1 hour at room temperature. Subsequent washing in PBS-Tr was the same as for the primary antibodies. DAPI (4',6-diamidino-2-phenylindole; Invitrogen REF: D1306) diluted at 1:1000 in PBT was added to the samples for 10 minutes at room temperature on a rotating wheel followed by a wash in PBT for 10 minutes. Embryos were mounted in 70% glycerol in PBS.

RNA fluorescent in-situ hybridization—RNA *in situ* hybridization was performed on embryos expressing an N-terminal EGFP-tagged GAF allele independently generated through CRISPR/Cas9 genome editing. Maternal EGFP-GAF was knocked down using deGradFP expressed via the Gal4-UAS system. Gal4 expression was driven by the second chromosome maternal alpha-tubulin Gal4-VP16 driver (*mat-alpha 4-Gal4-VP16*) "64" from Bloomington Drosophila Stock Center # 80361. DegradFP expression was through *UASP-NSlmb-vhhGFP4* "2" from Bloomington Drosophila Stock Center # 38422. Embryos from $GAF^{S\ PQ}/Df(3L)ED4543$ (BDSC 8073) females were used to determine if $GAF^{S\ PQ}$ can repress AAGAG RNA. Staged embryos were collected for *in situ* hybridization essentially as described above.

<u>Probe Synthesis:</u> The probe was synthesized as described in Mills et. al 2019. Template sequence listed below.

AAGAG(n)-T3as:

5'GAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAATCTCCCTTTAGTGAGGG TTAATT-3'

<u>Immunostaining:</u> Fixed embryos were washed 3× 10 minutes in PTx, then blocked for 1 hour at room temp. Embryos were incubated in primary antibody solution (1:1000 rabbit anti-GFP or 1:1000 rabbit anti-RNA Pol II CTD repeat YSPTSPS (phospho S5) in blocking buffer (1:10 Western Blocking Reagent (Sigma 11921673001) in PTx)) overnight at 4C, then washed 3× 10 minutes in PTx. The secondary antibody incubation (1:500 goat anti-rabbit 546 in blocking buffer) occurred for 2 hours at room temperature. Embryos were then washed 3× 10 minutes in PTx, post-fixed in a 4% paraformaldehyde solution for 20 minutes,

<u>Hybridization:</u> Fixed embryos were incubated in permeabilization solution (0.1% Triton X-100, 0.05% Igepal CA-630 (v/v), 500 ug/mL sodium deoxycholate, 500 ug/mL saponin, 2 mg/mL BSA Fraction V) for 2 hours at 4°C. Embryos were then post-fixed in a 4% paraformaldehyde solution for 20 minutes, washed 5x for 5 minutes in PTx (1x PBS, 0.1% Triton X-100), and incubated in a 1:1 Hybridization Buffer (50% deionized formamide (v/v), 25% 20x SSC (v/v), 50x) : PTx mix for 10 minutes. Prior to addition of the FISH probe, embryos were incubated in hybridization buffer for 1 hour at 55°C. A 1 ng/ul probe solution was prepared in hybridization buffer and probe hybridization occurred overnight at 55C. Post-hybridization, embryos were rinsed once in hybridization buffer, then incubated in fresh hybridization buffer for 1 hour at 55°C. Embryos were once again incubated in a 1:1 Hybridization Buffer: PTx mix for 10 minutes, followed by 5× 5 minute washes

in PTx. To detect the AAGAG probe, an anti-digoxigenin Alexa 488 conjugate (Jackson ImmunoResearch 200-541-156) was diluted to 2.5 ug/mL in blocking buffer (1:10 Western Blocking Reagent (Sigma 11921673001) in PTx). Embryos were blocked for 1 hour, then incubated in antibody solution for 2 hours at room temperature. Final washes performed 3× 10 minutes in PTw (1x PBS, 0.1% Tween-20).

RNA-FISH embryo imaging: Embryos were mounted in Prolong Gold Antifade Mountant (Invitrogen P10144) and imaged using Leica SP8 WLL confocal microscope. Images captured using 63× 1.3NA glycerol objective at 1024 × 512 pixels, 400 Hz scan rate. Alexa 488 and 546 were excited at 499 nm and 560 nm wavelengths respectively. Surface view images were collected with 0.30 um z-steps and max projected over a 15 μm range. RNA Pol II images captured under the same conditions described here, but with a 200 Hz scan speed and 3.0x zoom. Images shown at a single z slice.

QUANTIFICATION AND STATISTICAL ANALYSIS

Lattice light-sheet microscopy data analysis—Data were rendered using Imaris with no further processing for visualization and quantified as described below. Photobleaching was accounted for in two ways. LLSM imaging was optimized to minimize bleaching over time to the extent possible while not compromising temporal resolution or the ability to resolve GAF foci. Additionally, an exponential decay photobleaching model was fit to the nuclear intensities over time to generate an "intensity correction factor" for each time point to mitigate the loss of intensity due to bleaching.

The His2Av-RFP channel was used to segment nuclei in 3D. The nuclei images were convolved with a difference of Gaussians filter, followed by a threshold calculated using Otsu's method to generate a binary mask. The mask was cleaned to remove noise and any nuclei touching the image border.

To identify the volume occupied by GAF foci, a 93-percentile threshold calculated across all time points was applied to the GAF-GFP channel within the nuclei. Volumes were calculated from the volume of each image voxel ($0.1 \times 0.1 \times 0.2$ um^3) within the binary foci and nuclear masks.

To count discrete GAF puncta, the GAF-GFP channel was convolved with a Gaussian filter (sigma=1.2) and local maxima within a 5-voxel neighborhood were identified. These puncta were then filtered using the GAF foci volumetric mask. Images were manually inspected to ensure that this filter identified all visually apparent foci within each nucleus.

Quantification of foci volume from confocal microscopy—Max intensity projections of the first 10 layers, each a 0.5 μm step apart, in the z-plane of the apical part of the nuclei were generated. Binary masks of the fluorescent signal were created to define GAF foci and whole nuclei, using the threshold tool in ImageJ. Any nuclei on the periphery of the images were excluded from the analysis. GAF foci were defined as objects greater than $0.01 \mu m^2$. Mean fluorescent intensity was measured and multiplied by the area to generate the total fluorescent signal for each foci. Intensity was normalized to an average intensity for 10 different spots in the cytoplasm. Cytoplasmic signal proportional to that

of the object's size was then subtracted from the signal to account for background. Total signal for all foci in a nucleus was summed and divided by total nuclear signal to calculate percentage of GFP-GAF signal in foci.

ChIP-seq data analysis—Pearson's correlation coefficients were calculated for all ChIP-seq replicates using DeepTools and are reported in Table S2.

sfGFP-GAF^{S PQ} and DBD-sfGFP ChIP: ChIP-seq data was aligned to a combined *Drosophila melanogaster* reference genome (version dm6) using bowtie 2 v2.3.5 [99] 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[100] 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.

To compare the binding sites of sfGFP-GAF^{S PQ} to sfGFP-GAF, we used intersectBed from Bedtools[101] to compare different sets of peaks. Peaks overlapping by at least a 20bp overlap were considered to be shared. DeepTools[102] was used to generate read depth for 10 bp bins across the genome. A z-score was calculated for each 10 bp bin using the mean and standard deviation of read depth across all 10 bp bins. Z-score normalized read depth was used to generate heatmaps, metaplots, and genome browser tracks.

H3K9me3 ChIP: Reads were trimmed to remove adapter sequences using NGmerge[103]. ChIP-seq data was aligned to the *Drosophila melanogaster* reference genome (version dm6) using bowtie 2 v2.3.5 [99] without the following default parameters to retain multimapping reads: --no-mixed --no-discordant. All aligned reads were kept regardless of mapping quality to retain multimapping repeats. bamCompare was used with the following parameters --scaleFactorsMethod SES, --operation log2, to generate bigWig files of the H3K9me3 signal normalized to the input. bigWig files were visualized using IGV[104].

AAGAG repeat ChIP analysis: To analyze raw reads for simple satellite repeats raw fastq files were converted to fasta files. Fasta files were then searched for $(AAGAG)_5$ and the reverse complement sequence. To determine how well the repeat of interest was immunoprecipitated, we calculated the percentage of total reads that contained the repeat of interest in both the IP and paired input raw reads. The IP/input was then calculated. A scaling factor calculated from the spike-in normalization was applied to the IP/input for the H3K9me3 ChIP-seq replicates. To verify that the *D. virilis* or mouse spike in chromatin would not confound our analysis, we determined the amount of reads with $(AAGAG)_5$ in input samples from *D. virilis* ovaries[105] and MEF cells[106]. In contrast to the *D. melanogaster* ChIP-seq input samples analyzed for this study in which 0.39 – 3.28% of total reads contained the $(AAGAG)_5$ repeat, in *D. virilis* input material only 0.00059 – 0.00064% of the total reads contained the $(AAGAG)_5$ repeat, indicating this repeat is not abundant in the

D. virilis genome and would not impact our analysis. Similarly, in MEF input samples only 0.0032 – 0.0042% of total reads contain the $(AAGAG)_5$ repeat.

ChIP-seq spike-in normalization

sfGFP-GAF$^{S\ PQ}$ and DBD-sfGFP ChIP: Prior to addition of the anti-GFP antibody, mouse chromatin prepared from cells expressing an H3.3-GFP fusion protein was added to *Drosophila* chromatin at a 1:750 ratio. Following sequencing, reads were aligned to a combined reference genome containing both the *Drosophila* genome (version dm6) and the mouse genome (version mm39). Only reads that could be unambiguously aligned to one of the two reference genomes were retained. To control for any variability in the proportion of mouse chromatin in the input samples, the ratio of percentage of spike in reads in the IP relative to the input were used. A scaling factor was calculated by dividing one by this ratio. Z-score normalized read depth was adjusted by this scaling factor, and the resulting spike-in normalized values were used for heatmaps.

H3K9me3 ChIP: Prior to addition of the anti-H3K9me3 antibody, *D. virilis* chromatin prepared from stage 5 embryos was added to the *D. melanogaster* chromatin at a 1:25 ratio. Following sequencing, reads were aligned to a combined reference genome containing both the *D. melanogaster* (version dm6) and *D. virilis* genome. Reads were aligned to the combined reference genomes using parameters that retained multimapping reads. To control for any variability in the proportion of *D. virilis* chromatin in the input samples, the ratio of percentage of spike in reads in the IP relative to the input were used. A scaling factor was calculated by dividing one by this ratio. The scaling factor was used to adjust the IP/input value in the analysis of the raw reads.

Total RNA-seq analysis—To analyze raw reads for simple satellite repeats raw fastq files were converted to fasta files. Fasta files were searched for x5 repeats of common simple satellites and their reverse complement sequences. To determine if the repeat of interest changed in expression between GAF^{deGradFP} and control embryos, we calculated the percentage of total reads that contained the repeat of interest in each replicate. A two-tailed t-test was used to determine significance between the percentage of repeat reads in the GAF^{deGradFP} compared to control replicates.

Hi-C analysis—Insulation scores were computed by first computing the directionality (ratio of contacts with bins to the right vs. left) for each 500 bp bin in the genome, then performing a rolling difference calculation with window and step sizes of 16 bins (8 kb), followed by smoothing with a moving average with a 5 kb window size. Compartment scores were calculated for 25 kb bins of whole chromosomes by normalizing each bin for distance from the diagonal (i.e., observed/expected), calculating the covariance matrix, and taking the first eigenvector.

Quantification of DNA FISH—Images were max projected and the DAPI channel was processed with a Gaussian blur to denoise the nuclei. A 45% threshold was then used to segment nuclei. FISH foci were identified using a local maximal filter (radius 10 pixels), and additional random spots within each nucleus were generated as controls. Spots less than 5

pixels from the nuclear periphery or 10 pixels from the image border were discarded. All image channels were normalized to the nucleus mean.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Inclusion and Diversity

We support inclusive, diverse and equitable conduct of research.

References

1. Li XY, Thomas S, Sabo PJ, Eisen MB, Stamatoyannopoulos JA, and Biggin MD (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. Genome Biol. 12, R34. 10.1186/gb-2011-12-4-r34. [PubMed: 21473766]

2. Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, Dodonova SO, Nitta KR, Morgunova E, Taipale M, et al. (2018). The interaction landscape between transcription factors and the nucleosome. Nature 562, 76–81. 10.1038/s41586-018-0549-5. [PubMed: 30250250]

3. Zaret KS (2020). Pioneer Transcription Factors Initiating Gene Network Changes. Annu. Rev. Genet. 54, 367–385. 10.1146/annurev-genet-030220-015007. [PubMed: 32886547]

4. Larson ED, Marsh AJ, and Harrison MM (2021). Pioneering the developmental frontier. Mol. Cell 81, 1640–1650. 10.1016/j.molcel.2021.02.020. [PubMed: 33689750]

5. Balsalobre A, and Drouin J (2022). Pioneer factors as master regulators of the epigenome and cell fate. Nat. Rev. Mol. Cell Biol. 23. 10.1038/s41580-022-00464-z.

6. Allshire RC, and Madhani HD (2018). Ten principles of heterochromatin formation and function. Nat. Rev. Mol. Cell Biol. 19, 229–244. 10.1038/nrm.2017.119. [PubMed: 29235574]

7. Schulz KN, and Harrison MM (2019). Mechanisms regulating zygotic genome activation. Nat. Rev. Genet. 20, 221–234. 10.1038/s41576-018-0087-x. [PubMed: 30573849]

8. Vastenhouw NL, Cao WX, and Lipshitz HD (2019). The maternal-to-zygotic transition revisited. Development 146, dev161471. 10.1242/dev.161471. [PubMed: 31189646]

9. Zhang Y, and Xie W (2022). Building the genome architecture during the maternal to zygotic transition. Curr. Opin. Genet. Dev. 72, 91–100. 10.1016/j.gde.2021.11.002. [PubMed: 34896808]

10. Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, and Karpen GH (2017). Phase separation drives heterochromatin domain formation. Nature 547, 241–245. 10.1038/nature22989. [PubMed: 28636597]

11. Hur W, Kemp JP, Tarzia M, Deneke VE, Marzluff WF, Duronio RJ, and Di Talia S (2020). CDK-Regulated Phase Separation Seeded by Histone Genes Ensures Precise Growth and Function of Histone Locus Bodies. J. Clean. Prod. 54, 379–394.e6. 10.1016/j.devcel.2020.06.003.

12. Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, Schymkowitz J, Shorter J, Wolozin B, Van Den Bosch L, et al. (2018). Protein Phase Separation: A New Phase in Cell Biology. Trends Cell Biol. 10.1016/j.tcb.2018.02.004.

13. Rippe K (2022). Liquid–Liquid Phase Separation in Chromatin. Cold Spring Harb. Perspect. Biol. 14. 10.1101/cshperspect.a040683.

14. Peng L, Li EM, and Xu LY (2020). From start to end: Phase separation and transcriptional regulation. Biochim. Biophys. Acta - Gene Regul. Mech. 1863. 10.1016/j.bbagrm.2020.194641.

15. Ferrie JJ, Karr JP, Tjian R, and Darzacq X (2022). "Structure"-function relationships in eukaryotic transcription factors: The role of intrinsically disordered regions in gene regulation. Mol. Cell, 1–15. 10.1016/j.molcel.2022.09.021.

16. Strom AR, and Brangwynne CP (2019). The liquid nucleome – phase transitions in the nucleus at a glance. J. Cell Sci. 132, 1–7. 10.1242/jcs.235093.

17. Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, and Narlikar GJ (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature 547, 236–240. 10.1038/nature22822. [PubMed: 28636604]

18. Mir M, Reimer A, Haines JE, Li XY, Stadler M, Garcia H, Eisen MB, and Darzacq X (2017). Dense bicoid hubs accentuate binding along the morphogen gradient. Genes Dev. 31, 1784–1794. 10.1101/gad.305078.117. [PubMed: 28982761]

19. Mir M, Stadler MR, Ortiz SA, Hannon CE, Harrison MM, Darzacq X, and Eisen MB (2018). Dynamic multifactor hubs interact transiently with sites of active transcription in drosophila embryos. Elife 7. 10.7554/eLife.40497.

20. 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, 1–13. 10.1038/s41467-018-07613-z. [PubMed: 29317637]

21. Yamada S, Whitney PH, Huang SK, Eck EC, Garcia HG, and 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. 10.1016/j.cub.2019.03.019. [PubMed: 30982648]

22. Tsai A, Muthusamy AK, Alves MRP, Lavis LD, Singer RH, Stern DL, and Crocker J (2017). Nuclear microenvironments modulate transcription from low-affinity enhancers. Elife 6, 1–18. 10.7554/eLife.28975.

23. Kuznetsova K, Chabot NM, Ugolini M, Wu E, Lalit M, Oda H, Sato Y, Kimura H, Jug F, and Vastenhouw NL (2023). Nanog organizes transcription bodies. Curr. Biol. 33, 164–173.e5. 10.1016/j.cub.2022.11.015. [PubMed: 36476751]

24. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K, Manteiga JC, Hannett NM, et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell 175, 1842–1855.e16. 10.1016/j.cell.2018.10.042. [PubMed: 30449618]

25. Alberti S, Gladfelter A, and Mittag T (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell 176, 419–434. 10.1016/j.cell.2018.12.035. [PubMed: 30682370]

26. Hamm DC, and Harrison MM (2018). Regulatory principles governing the maternal-to-zygotic transition: Insights from Drosophila melanogaster. Open Biol. 8. 10.1098/rsob.180183.

27. Gaskill MM, Gibson TJ, Larson ED, and Harrison MM (2021). GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. Elife. 10.7554/eLife.66668.

28. Raff JW, Kellum R, and Alberts B (1994). The Drosophila GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle. EMBO J. 13, 5977–5983. 10.1002/j.1460-2075.1994.tb06943.x. [PubMed: 7813435]

29. Chen BC, Legant WR, Wang K, Shao L, Milkie DE, Davidson MW, Janetopoulos C, Wu XS, Hammer JA, Liu Z, et al. (2014). Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. Science (80-.). 346. 10.1126/science.1257998.

30. Wilkins RC, and Lis JT (1999). DNA distortion and multimerization: Novel functions of the glutamine-rich domain of GAGA factor. J. Mol. Biol. 285, 515–525. 10.1006/jmbi.1998.2356. [PubMed: 9878426]

31. Benyajati C, Mueller L, Xu N, Pappano M, Gao J, Mosammaparast M, Conklin D, Granok H, Craig C, and Elgin S (1997). Multiple isoforms of GAGA factor, a critical component of chromatin structure. Nucleic Acids Res. 25, 3345–3353. 10.1093/nar/25.16.3345. [PubMed: 9241251]

32. Lintermann KG, Roth GE, King-Jones K, Korge G, and Lehmann M (1998). Comparison of the GAGA factor genes of Drosophila melanogaster and Drosophila virilis reveals high conservation of GAGA factor structure beyond the BTB/POZ and DNA-binding domains. Dev. Genes Evol. 208, 447–456. 10.1007/s004270050202. [PubMed: 9799425]

33. Greenberg AJ, and Schedl P (2001). GAGA Factor Isoforms Have Distinct but Overlapping Functions In Vivo. Mol. Cell. Biol. 21, 8565–8574. 10.1128/mcb.21.24.8565-8574.2001. [PubMed: 11713290]

34. Farkas G, Gausz J, Galloni M, Reuter G, Gyurkovics H, and Karch F (1994). The Trithorax-like gene encodes the Drosophila GAGA factor. Nature 371, 806–808. 10.1038/371806a0. [PubMed: 7935842]

35. Vaquero A, Espinás ML, Azorín F, and Bernués J (2000). Functional mapping of the GAGA factor assigns its transcriptional activity to the C-terminal glutamine-rich domain. J. Biol. Chem. 275, 19461–19468. 10.1074/jbc.M000967200. [PubMed: 10764754]

36. Agianian B, Leonard K, Bonte E, Van Der Zandt H, Becker PB, and Tucker PA (1999). The glutamine-rich domain of the Drosophila GAGA factor is necessary for amyloid fibre formation in vitro, but not for chromatin remodelling. J. Mol. Biol. 285, 527–544. 10.1006/jmbi.1998.2355. [PubMed: 9878427]

37. Tang X, Li T, Liu S, Wisniewski J, Zheng Q, Rong Y, Lavis LD, and Wu C (2022). Kinetic principles underlying pioneer function of GAGA transcription factor in live cells. Nat. Struct. Mol. Biol. 29, 665–676. 10.1038/s41594-022-00800-z. [PubMed: 35835866]

38. Shatskikh AS, Kotov AA, Adashev VE, Bazylev SS, and Olenina LV (2020). Functional Significance of Satellite DNAs: Insights From Drosophila. Front. Cell Dev. Biol. 8, 1–19. 10.3389/fcell.2020.00312. [PubMed: 32117956]

39. Lohe AR, and Brutlag DL (1987). Identical satellite DNA sequences in sibling species of Drosophila. J. Mol. Biol. 194, 161–170. 10.1016/0022-2836(87)90365-2. [PubMed: 3112413]

40. Platero JS, Csink AK, Quintanilla A, and Henikoff S (1998). Changes in chromosomal localization of heterochromatin-binding proteins during the cell cycle in Drosophila. J. Cell Biol. 140, 1297–1306. 10.1083/jcb.140.6.1297. [PubMed: 9508764]

41. Mills WK, Lee YCG, Kochendoerfer AM, Dunleavy EM, and Karpen GH (2019). RNA from a simple-tandem repeat is required for sperm maturation and male fertility in drosophila melanogaster. Elife 8, 1–21. 10.7554/eLife.48940.

42. Pathak RU, Mamillapalli A, Rangaraj N, Kumar RP, Vasanthi D, Mishra K, and Mishra RK (2013). AAGAG repeat RNA is an essential component of nuclear matrix in Drosophila. RNA Biol. 10, 564–571. 10.4161/rna.24326. [PubMed: 23588056]

43. Adkins NL, Hagerman TA, and Georgel P (2006). GAGA protein: A multi-faceted transcription factor. Biochem. Cell Biol. 84, 559–567. 10.1139/O06-062. [PubMed: 16936828]

44. Chetverina D, Erokhin M, and Schedl P (2021). GAGA factor: a multifunctional pioneering chromatin protein. Cell. Mol. Life Sci. 78, 4125–4141. 10.1007/s00018-021-03776-z. [PubMed: 33528710]

45. Maeda RK, and Karch F (2007). Making connections: boundaries and insulators in Drosophila. Curr. Opin. Genet. Dev. 17, 394–399. 10.1016/j.gde.2007.08.002. [PubMed: 17904351]

46. Chen K, Johnston J, Shao W, Meier S, Staber C, and Zeitlinger J (2013). A global change in RNA polymerase II pausing during the Drosophila midblastula transition. Elife 2013, 1–19. 10.7554/eLife.00861.

47. Huang SK, Whitney PH, Dutta S, Shvartsman SY, and Rushlow CA (2021). Spatial organization of transcribing loci during early genome activation in Drosophila. Curr. Biol. 31, 5102–5110.e5. 10.1016/j.cub.2021.09.027. [PubMed: 34614388]

48. 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. 10.1101/gad.300855.117. [PubMed: 28838946]

49. Garcia HG, Tikhonov M, Lin A, and Gregor T (2013). Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning. Curr. Biol. 23, 2140–2145. 10.1016/j.cub.2013.08.054. [PubMed: 24139738]

50. Nègre N, Brown CD, Shah PK, Kheradpour P, Morrison CA, Henikoff JG, Feng X, Ahmad K, Russell S, White RAH, et al. (2010). A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet. 6. 10.1371/journal.pgen.1000814.

51. Lomaev D, Mikhailova A, Erokhin M, Shaposhnikov AV, Moresco JJ, Blokhina T, Wolle D, Aoki T, Ryabykh V, Yates JR, et al. (2017). The GAGA factor regulatory network: Identification of GAGA factor associated proteins. PLoS One 12, 1–20. 10.1371/journal.pone.0173602.

52. Wolle D, Cleard F, Aoki T, Deshpande G, Schedl P, and Karch F (2015). Functional Requirements for Fab-7 Boundary Activity in the Bithorax Complex. Mol. Cell. Biol. 35, 3739–3752. 10.1128/mcb.00456-15. [PubMed: 26303531]

53. Kyrchanova O, Kurbidaeva A, Sabirov M, Postika N, Wolle D, Aoki T, Maksimenko O, Mogila V, Schedl P, and Georgiev P (2018). The bithorax complex iab-7 Polycomb response element has a novel role in the functioning of the Fab-7 chromatin boundary. PLoS Genet. 14, 1–19. 10.1371/journal.pgen.1007442.

54. Ogiyama Y, Schuettengruber B, Papadopoulos GL, Chang JM, and Cavalli G (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. Mol. Cell 71, 73–88.e5. 10.1016/j.molcel.2018.05.032. [PubMed: 30008320]

55. Nègre N, Hennetin J, Sun LV, Lavrov S, Bellis M, White KP, and Cavalli G (2006). Chromosomal distribution of PcG proteins during Drosophila development. PLoS Biol. 4, 0917–0932. 10.1371/journal.pbio.0040170.

56. Strutt H, Cavalli G, and Paro R (1997). Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. EMBO J. 16, 3621–3632. 10.1093/emboj/16.12.3621. [PubMed: 9218803]

57. Hug CB, Grimaldi AG, Kruse K, and Vaquerizas JM (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell 169, 216–228.e19. 10.1016/j.cell.2017.03.024. [PubMed: 28388407]

58. Batut PJ, Bing XY, Sisco Z, Raimundo J, Levo M, and Levine MS (2022). Dynamics During Development. Science (80-.). 375, 566–570.

59. Mahmoudi T, Katsani KR, and Verrijzer CP (2002). GAGA can mediate enhancer function in trans by linking two separate DNA molecules. EMBO J. 21, 1775–1781. 10.1093/emboj/21.7.1775. [PubMed: 11927561]

60. Petrascheck M, Escher D, Mahmoudi T, Verrijzer CP, Schaffner W, and Barberis A (2005). DNA looping induced by a transcriptional enhancer in vivo. Nucleic Acids Res. 33, 3743–3750. 10.1093/nar/gki689. [PubMed: 16002789]

61. Maeshima K, Tamura S, Hansen JC, and Itoh Y (2020). Fluid-like chromatin: Toward understanding the real chromatin organization present in the cell. Curr. Opin. Cell Biol. 64, 77–89. 10.1016/j.ceb.2020.02.016. [PubMed: 32283330]

62. Li X, Tang X, Bing X, Catalano C, Li T, Dolsten G, Wu C, and Levine M (2023). GAGA-associated factor fosters loop formation in the Drosophila genome. Mol. Cell 83, 1–8. 10.1016/j.molcel.2023.03.011. [PubMed: 36608666]

63. Kellum R, Raff JW, and Alberts BM (1995). Heterochromatin protein 1 distribution during development and during the cell cycle in Drosophila embryos. J. Cell Sci. 108, 1407–1418. 10.1242/jcs.108.4.1407. [PubMed: 7615662]

64. Bellec M, Dufourt J, Hunt G, Lenden-Hasse H, Trullo A, Zine El Aabidine A, Lamarque M, Gaskill MM, Faure-Gautron H, Mannervik M, et al. (2022). The control of transcriptional memory by stable mitotic bookmarking. Nat. Commun. 13, 1–14. 10.1038/s41467-022-28855-y. [PubMed: 34983933]

65. Yuan K, and O'Farrell PH (2016). TALE-light imaging reveals maternally guided, H3K9me2/3-independent emergence of functional heterochromatin in Drosophila embryos. Genes Dev. 30, 579–593. 10.1101/gad.272237.115. [PubMed: 26915820]

66. Ferree PM, and Barbash DA (2009). Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. PLoS Biol. 7. 10.1371/journal.pbio.1000234.

67. Satyaki PRV, Cuykendall TN, Wei KHC, Brideau NJ, Kwak H, Aruna S, Ferree PM, Ji S, and Barbash DA (2014). The Hmr and Lhr Hybrid Incompatibility Genes Suppress a Broad Range of Heterochromatic Repeats. PLoS Genet. 10. 10.1371/journal.pgen.1004240.

68. Brodsky S, Jana T, Mittelman K, Chapal M, Kumar DK, Carmi M, and Barkai N (2020). Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. Mol. Cell 79, 459–471.e4. 10.1016/j.molcel.2020.05.032. [PubMed: 32553192]

69. Mazzocca M, Loffreda A, Colombo E, Fillot T, Gnani D, Falletta P, Monteleone E, Capozi S, Bertrand E, Legube G, et al. (2022). Chromatin organization drives the exploration strategy of nuclear factors. bioRxiv, 2022.12.29.522193.

70. Choi KJ, Quan MD, Qi C, Lee JH, Tsoi PS, Zahabiyon M, Bajic A, Hu L, Prasad BVV, Liao SCJ, et al. (2022). NANOG prion-like assembly mediates DNA bridging to facilitate chromatin reorganization and activation of pluripotency. Nat. Cell Biol. 24, 737–747. 10.1038/s41556-022-00896-x. [PubMed: 35484250]

71. Chong S, Graham TGW, Dugast-Darzacq C, Dailey GM, Darzacq X, and Tjian R (2022). Tuning levels of low-complexity domain interactions to modulate endogenous oncogenic transcription. Mol. Cell 82, 2084–2097.e5. 10.1016/j.molcel.2022.04.007. [PubMed: 35483357]

72. Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, Darzacq X, et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science (80-.). 361. 10.1126/science.aar2555.

73. Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science (80-.). 361. 10.1126/science.aar3958.

74. Wei MT, Chang YC, Shimobayashi SF, Shin Y, Strom AR, and Brangwynne CP (2020). Nucleated transcriptional condensates amplify gene expression. Nat. Cell Biol. 22, 1187–1196. 10.1038/s41556-020-00578-6. [PubMed: 32929202]

75. Sharma R, Choi KJ, Quan MD, Sharma S, Sankaran B, Park H, LaGrone A, Kim JJ, MacKenzie KR, Ferreon ACM, et al. (2021). Liquid condensation of reprogramming factor KLF4 with DNA provides a mechanism for chromatin organization. Nat. Commun. 12. 10.1038/s41467-021-25761-7.

76. Morin JA, Wittmann S, Choubey S, Klosin A, Golfier S, Hyman AA, Jülicher F, and Grill SW (2022). Sequence-dependent surface condensation of a pioneer transcription factor on DNA. Nat. Phys. 18, 271–276. 10.1038/s41567-021-01462-2.

77. Grob A, Colleran C, and McStay B (2014). Construction of synthetic nucleoli in human cells reveals how a major functional nuclear domain is formed and propagated through cell division. Genes Dev. 28, 220–230. 10.1101/gad.234591.113. [PubMed: 24449107]

78. Shevtsov SP, and Dundr M (2011). Nucleation of nuclear bodies by RNA. Nat. Cell Biol. 13, 167–173. 10.1038/ncb2157. [PubMed: 21240286]

79. Shermoen AW, McCleland ML, and O'Farrell PH (2010). Developmental control of late replication and s phase length. Curr. Biol. 20, 2067–2077. 10.1016/j.cub.2010.10.021. [PubMed: 21074439]

80. Seller CA, Cho CY, and O'Farrell PH (2019). Rapid embryonic cell cycles defer the establishment of heterochromatin by eggless/SetDB1 in Drosophila. Genes Dev. 33, 403–417. 10.1101/gad.321646.118. [PubMed: 30808658]

81. Liu X, Shen J, Xie L, Wei Z, Wong C, Li Y, Zheng X, Li P, and Song Y (2020). Mitotic Implantation of the Transcription Factor Prospero via Phase Separation Drives Terminal Neuronal Differentiation. Dev. Cell 52, 277–293.e8. 10.1016/j.devcel.2019.11.019. [PubMed: 31866201]

82. Jagannathan M, Cummings R, and Yamashita YM (2019). The modular mechanism of chromocenter formation in Drosophila. Elife 8, 1–16. 10.7554/eLife.43938.

83. Jagannathan M, Cummings R, and Yamashita YM (2018). A conserved function for pericentromeric satellite DNA. Elife 7, 1–19. 10.7554/eLife.34122.

84. Zullo JM, Demarco IA, Piqué-Regi R, Gaffney DJ, Epstein CB, Spooner CJ, Luperchio TR, Bernstein BE, Pritchard JK, Reddy KL, et al. (2012). DNA sequence-dependent

compartmentalization and silencing of chromatin at the nuclear lamina. Cell 149, 1474–1487. 10.1016/j.cell.2012.04.035. [PubMed: 22726435]

85. Fuda NJ, Guertin MJ, Sharma S, Danko CG, Martins AL, Siepel A, and Lis JT (2015). GAGA Factor Maintains Nucleosome-Free Regions and Has a Role in RNA Polymerase II Recruitment to Promoters. PLoS Genet. 11, e1005108. 10.1371/journal.pgen.1005108. [PubMed: 25815464]

86. Judd J, Duarte FM, and Lis JT (2021). Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription. Genes Dev. 35, 147–156. 10.1101/gad.341768.120. [PubMed: 33303640]

87. Tsukiyama T, Becker PB, and Wu C (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature 367, 525–532. 10.1038/367525a0. [PubMed: 8107823]

88. Bonet C, Fernández I, Aran X, Bernués J, Giralt E, and Azorín F (2005). The GAGA protein of Drosophila is phosphorylated by CK2. J. Mol. Biol. 351, 562–572. 10.1016/j.jmb.2005.06.039. [PubMed: 16023138]

89. Jackson SP, and Tjian R (1988). O-glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. Cell 55, 125–133. 10.1016/0092-8674(88)90015-3. [PubMed: 3139301]

90. Aran-Guiu X, Ortiz-Lombardía M, Oliveira E, Bonet Costa C, Odena MA, Bellido D, and Bernués J (2010). Acetylation of GAGA factor modulates its interaction with DNA. Biochemistry 49, 9140–9151. 10.1021/bi1004427. [PubMed: 20849112]

91. Wen H, Andrejka L, Ashton J, Karess R, and Lipsick JS (2008). Epigenetic regulation of gene expression by Drosophila Myb and E2F2-RBF via the Myb-MuvB/dREAM complex. Genes Dev. 22, 601–614. 10.1101/gad.1626308. [PubMed: 18316477]

92. Hamm DC, Larson ED, Nevil M, Marshall KE, Bondra ER, and Harrison MM (2017). A conserved maternal-specific repressive domain in Zelda revealed by Cas9-mediated mutagenesis in Drosophila melanogaster. PLoS Genet. 13, e1007120. 10.1371/journal.pgen.1007120. [PubMed: 29261646]

93. 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. 10.1038/nmeth.2019. [PubMed: 22743772]

94. Mir M, Reimer A, Stadler M, Tangara A, Hansen AS, Hockemeyer D, Eisen MB, Garcia H, and Darzacq X (2018). Single Molecule Imaging in Live Embryos Using Lattice Light-Sheet Microscopy. Nanoscale Imaging. Methods Mol. Biol. 1814, 541–559.

95. Blythe SA, and Wieschaus EF (2015). Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. Cell 160, 1169–1181. 10.1016/j.cell.2015.01.050. [PubMed: 25748651]

96. Bhat KM, Farkas G, Karch F, Gyurkovics H, Gausz J, and 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. [PubMed: 8620838]

97. Stadler MR, Haines JE, and Eisen MB (2017). Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early Drosophila melanogaster embryo. Elife 6, 1–29. 10.7554/eLife.29550.

98. Bantignies F, and Cavalli G (2014). Topological Organization of Drosophila Hox Genes Using DNA Fluorescent In Situ Hybridization. In Hox Genes: Methods and Protocols, Graba Y and Rezsohazy R, eds. (Springer New York), pp. 103–120. 10.1007/978-1-4939-1242-1_7.

99. Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. 10.1038/nmeth.1923. [PubMed: 22388286]

100. 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. 10.1186/gb-2008-9-9-r137. [PubMed: 18798982]

101. Quinlan AR, and Hall IM (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. 10.1093/bioinformatics/btq033. [PubMed: 20110278]

102. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, and Manke T (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165. 10.1093/nar/gkw257. [PubMed: 27079975]

103. Gaspar JM (2018). NGmerge: Merging paired-end reads via novel empirically-derived models of sequencing errors. BMC Bioinformatics 19, 536. 10.1186/s12859-018-2579-2. [PubMed: 30572828]

104. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, and Mesirov JP (2011). Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. 10.1038/nbt.1754. [PubMed: 21221095]

105. Le Thomas A, Marinov GK, and Aravin AA (2014). A transgenerational process defines piRNA biogenesis in Drosophila virilis. Cell Rep. 8, 1617–1623. 10.1016/j.celrep.2014.08.013. [PubMed: 25199836]

106. Rodier G, Kirsh O, Baraibar M, Houlès T, Lacroix M, Delpech H, Hatchi E, Arnould S, Severac D, Dubois E, et al. (2015). The transcription factor E4F1 coordinates CHK1-dependent checkpoint and mitochondrial functions. Cell Rep. 11, 220–233. 10.1016/j.celrep.2015.03.024. [PubMed: 25843721]

Highlights

The GAF unstructured polyQ domain is not required for localizing to nuclear foci.

The GAF DBD is necessary and sufficient for localization to condensed nuclear foci.

GAF subnuclear foci correspond to heterochromatic AAGAG satellite repeats.

GAF is required for heterochromatin formation and silencing of satellite repeats.
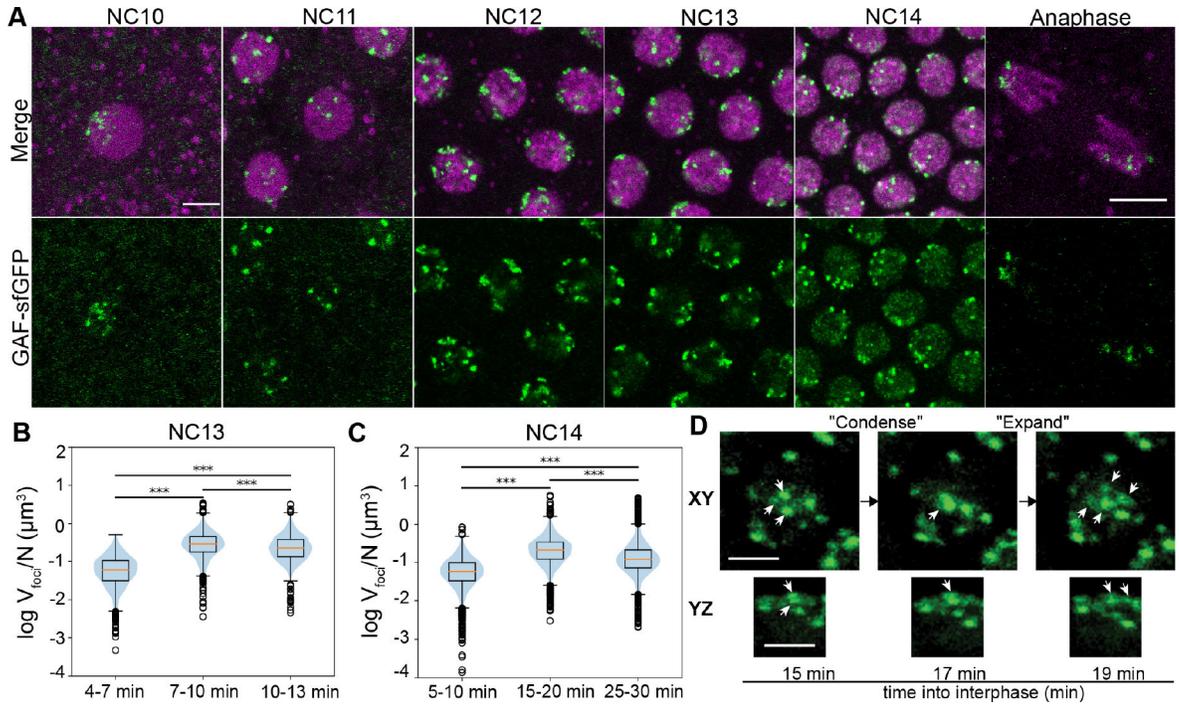
Figure 1: GAF forms multiple, stable nuclear foci during the MZT.

A. Embryos from NC10–14 (as indicated) laid by *His2Av-RFP; GAF-sfGFP* females. GAF-sfGFP is in green. His2Av-RFP is in magenta. Scale bars, 5μM. B. Average total volume of sfGFP-tagged GAF foci per nucleus in NC13. C. Average total volume of sfGFP-tagged GAF foci per nucleus in NC14. Asterisks indicate pairwise p-value thresholds. ** = 0.01, *** = 0.001, **** = 0.0001 (Tukey-Kramer test). n = 3 embryos. D. Orthogonal x-y and x-z views of sfGFP-tagged GAF foci condensing and expanding during NC14. Scale bar, 2.5μm. See also Figure S1.
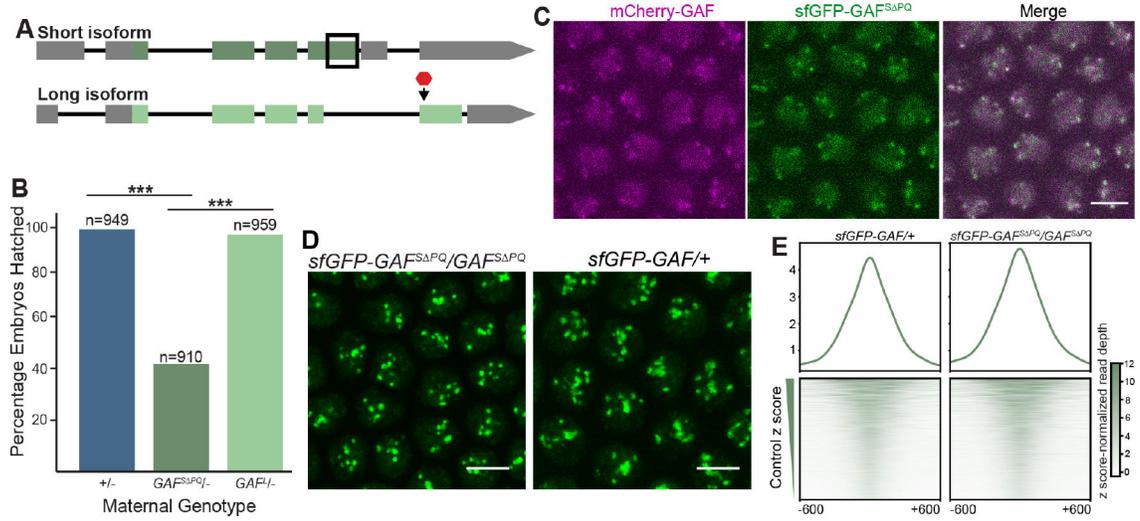
Figure 2: The intrinsically disordered poly-Q domain is not required for foci formation.
A. Model of two GAF splice isoforms. Coding regions are in green. Untranslated regions (UTRs) are in grey. Black lines indicate introns. The black box denotes the region deleted in the short-isoform specific deletion. The red octagon indicates the location of the stop codon introduced in the long isoform. B. Percentage of hatched embryos laid by the maternal genotypes indicated crossed to w^{1118} males (***, χ^2, p = 2.2 × 10^{-16}) n = total number of embryos assayed. C. Interphase NC14 embryos laid by *mCherry-GAF/sfGFP-GAF$^{S\Delta PQ}$* females. mCherry-GAF is in magenta. sfGFP-GAF$^{S\Delta PQ}$ is in green. Scale bar, 5μM. D. Maximum intensity z-stack projections of NC14 embryos laid by *sfGFP-GAF$^{S\Delta PQ}$/GAF$^{S\Delta PQ}$* females or control *sfGFP-GAF/+* females. Scale bars, 5μM E. Heatmaps of anti-GFP ChIP-seq from 2–3hr AEL embryos laid by *sfGFP-GAF$^{S\Delta PQ}$/GAF$^{S\Delta PQ}$* females and control *sfGFP-GAF/+* females. Heatmaps are ordered by z score-normalized signal from control embryos. n = 2. See also Figure S2 and Table S1.
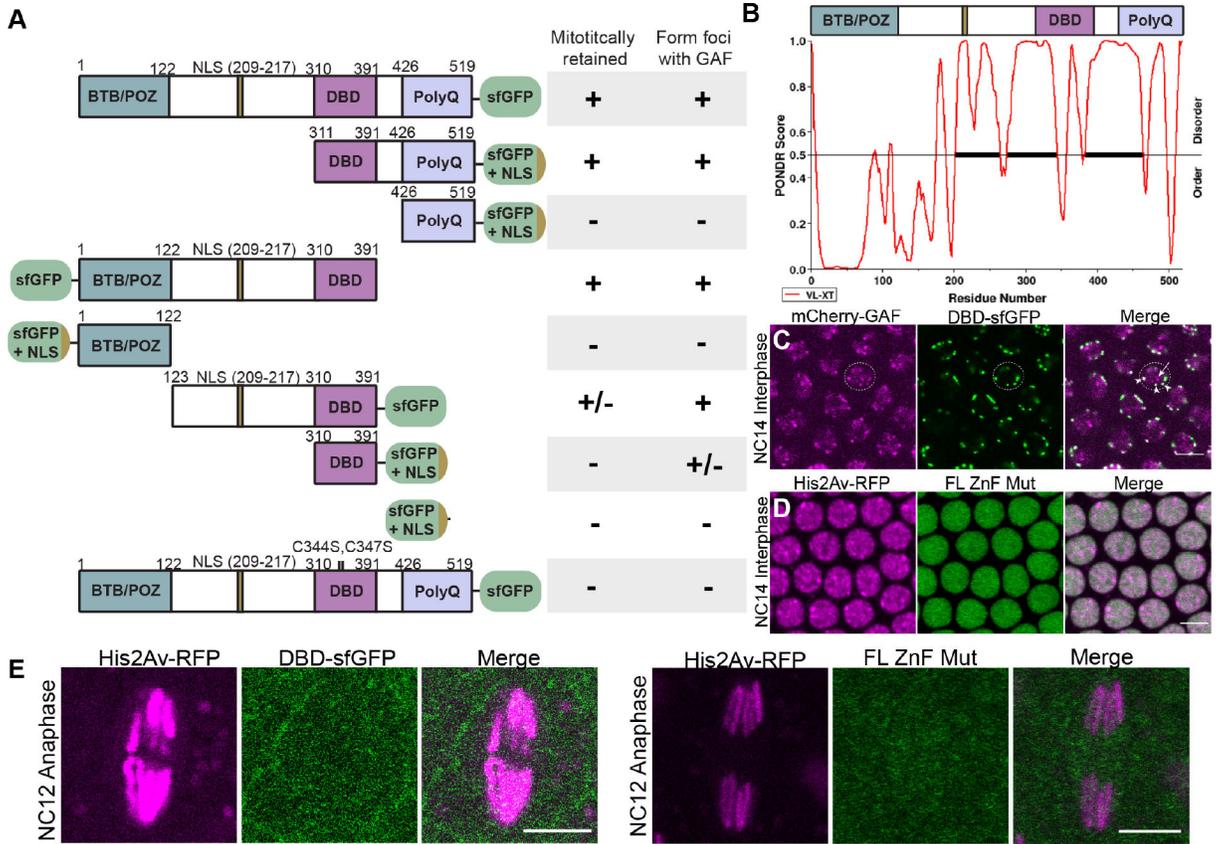
Figure 3: The DNA-binding domain of GAF is both necessary and sufficient for foci formation. A. Representations of tagged GAF truncations assayed (left) and whether or not those truncations were mitotically retained and localized to endogenous GAF foci (right). B. Prediction of intrinsically disordered regions of the short GAF isoform protein generated by PONDR. C. Interphase nuclei of an NC14 embryo expressing endogenously tagged mCherry-GAF and transgenically expressed DBD-sfGFP. mCherry-GAF is in magenta. DBD-sfGFP is in green. A dotted circle indicates a representative nucleus. DBD-sfGFP colocalizes with mCherry-GAF (arrowheads), but there is a subset of mCherry-GAF foci that do not colocalize with DBD-sfGFP (arrow). D. Interphase nuclei of an NC14 embryo expressing transgenic full length sfGFP-GAF with point mutations in the zinc finger DBD (FL ZnF Mut) and His2Av-RFP. His2Av-RFP is in magenta. FL ZnF Mut is in green. E. Anaphase of a NC12–13 embryo expressing His2Av-RFP and the DBD-sfGFP transgene (left) or the FL ZnF Mut (right). His2Av-RFP is in magenta. DBD-sfGFP is in green. Scale bars, 5μm. See also Figure S3.
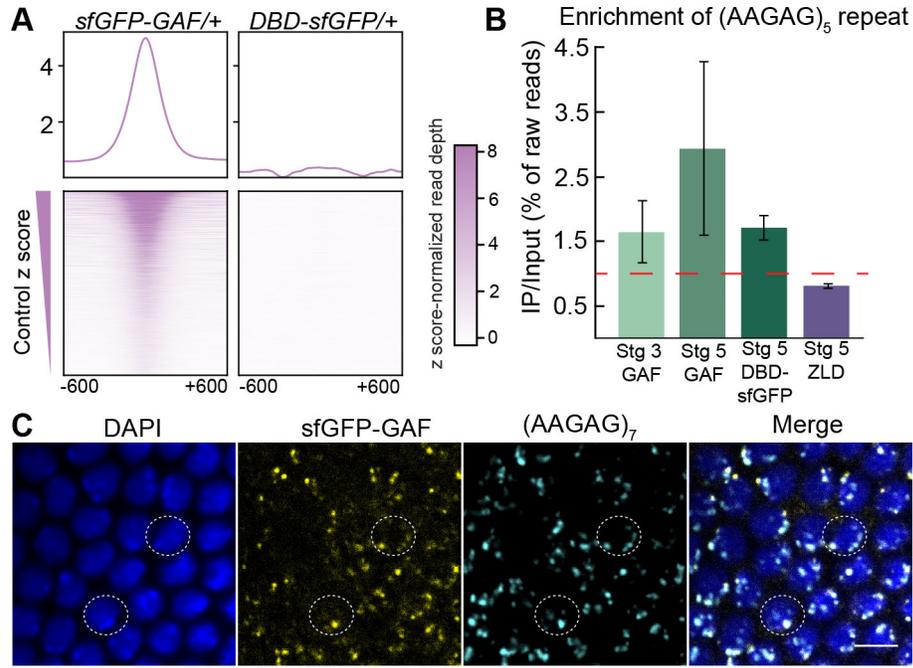
Figure 4. GAF foci correspond to the repetitive AAGAG elements.
A. Heatmaps of anti-GFP ChIP-seq performed on *sfGFP-GAF/+* embryos and embryos expressing transgenic DBD-sfGFP. n = 2 B. The percentage of the total raw ChIP-seq reads containing $(AAGAG)_5$ was determined and the ratio of the percentage of reads in the immunoprecipitation (IP) versus the input was plotted. Red line = IP/Input of 1. Error bars indicate the standard deviation of the two replicates tested. Stage 3 GAF-sfGFP, stage 5 GAF-sfGFP, and stage 5 ZLD ChIP-seq datasets were from Gaskill et al.[27]. C. DNA-FISH on NC14 *sfGFP-GAF(N)* homozygous embryos using an $(AAGAG)_7$ probe. Embryos were immunostained with an anti-GFP antibody and labelled with DAPI. Dotted circles indicate representative nuclei. Scale bar, 5μm. See also Figure S4.
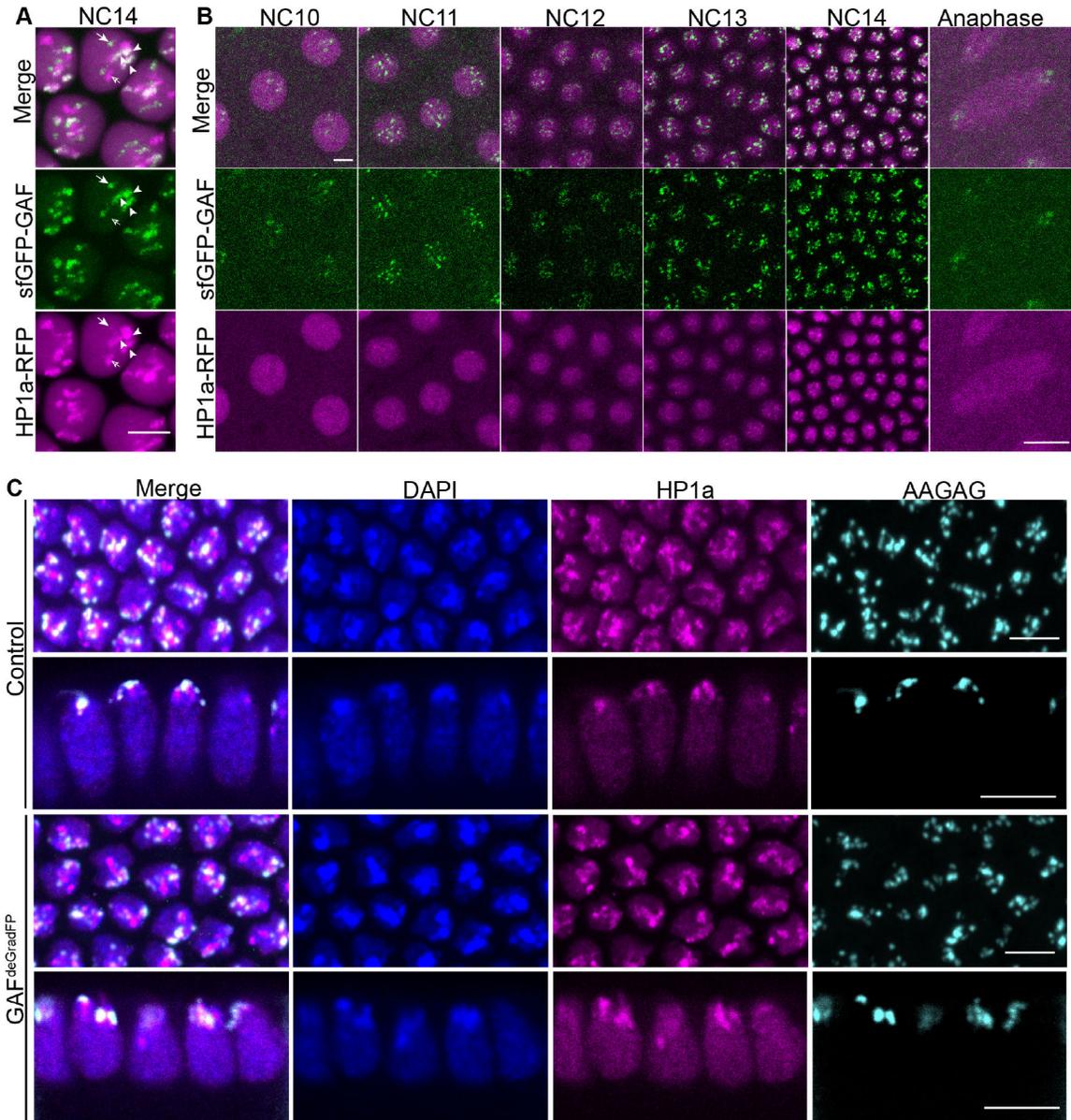
Figure 5: A subset of GAF foci localize with HP1a condensates.

A. Maximum intensity z-stack projection of interphase NC14 embryos laid by females expressing sfGFP-GAF and transgenic HP1a-RFP. HP1a-RFP is in magenta. sfGFP-GAF is in green. Arrowheads indicate regions of colocalization. Closed arrow indicates sfGFP-GAF only foci. Open arrow indicates HP1a-RFP only foci. B. Images of a single embryo from cycles NC10–14 (indicated) laid by a female expressing endogenous sfGFP-GAF and transgenic HP1a-RFP. HP1a-RFP is in magenta. sfGFP-GAF is in green. C. DNA-FISH performed on *sfGFP-GAF* homozygous embryos (top) and GAF[deGradFP] embryos (bottom) using an (AAGAG)$_7$ probe. Embryos were immunostained with an anti-HP1a antibody and labelled with DAPI. Scale bars, 5μM. See also Figure S5 and Movies S1 and S2
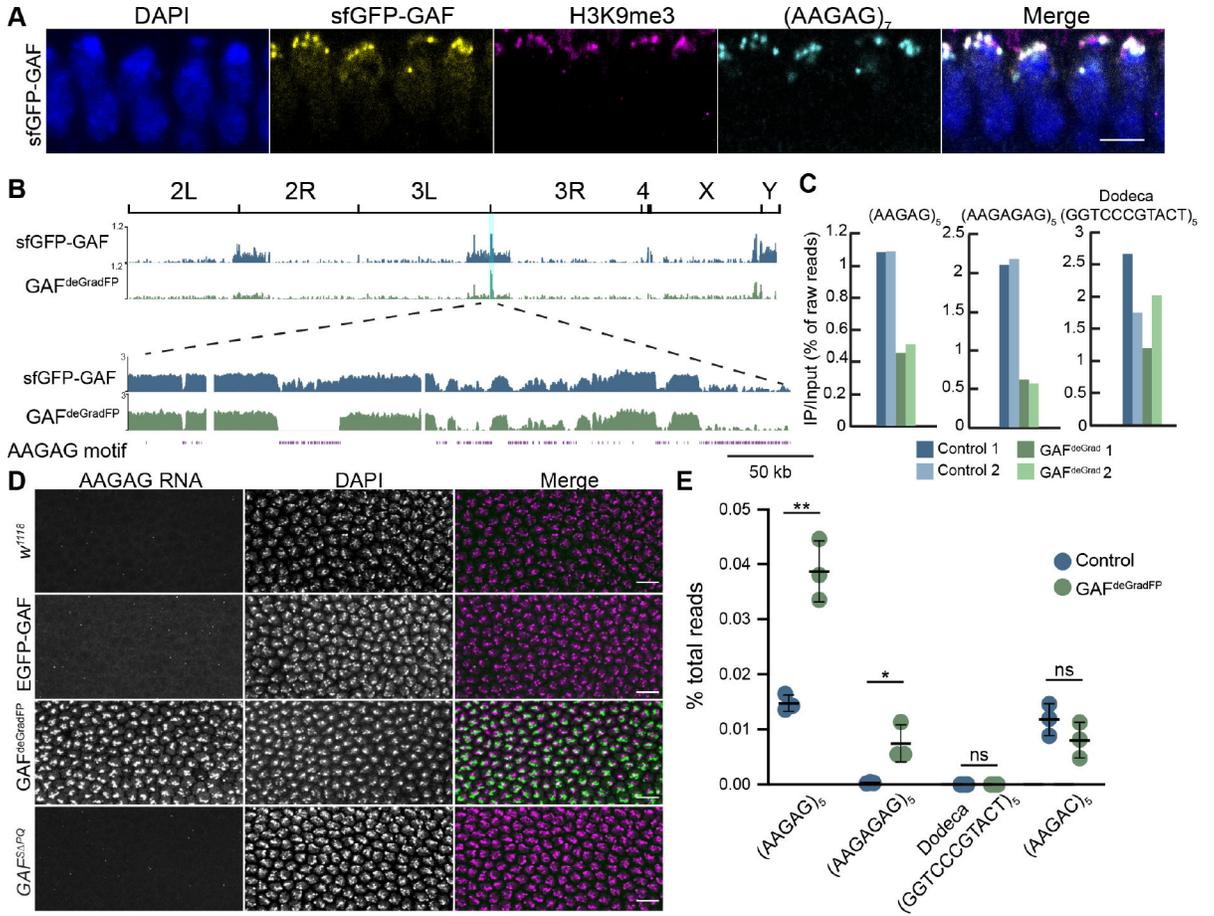
Figure 6: GAF is required to repress AAGAG satellite repeat expression during the MZT.
A. DNA-FISH on *sfGFP-GAF(N)* homozygous embryos at NC14 using an $(AAGAG)_7$ probe. Anti-GFP and anti-H3K9me3 antibodies were used for immunostaining. Scale bar, 5μM. B. Genome browser tracks of IP read depth normalized to input from anti-H3K9me3 ChIP-seq performed on 2–2.5hr AEL *sfGFP-GAF(N)* homozygous and GAF[deGradFP] embryos. The entire genome is shown. The region highlighted in blue from the 3rd chromosome centromere is shown in detail below to highlight specific reduction of signal at AAGAG repeats. n = 2. C. IP/Input of the percentage of raw reads that contain the indicated satellite repeat sequences from anti-H3K9me3 ChIP-seq on control (*sfGFP-GAF(N)* homozygous) and GAF[deGradFP] embryos at 2–2.5 hr AEL. D. RNA-FISH performed on *w[1118]*, *EGFP-GAF*, GAF[deGradFP], and *GAF[S ΔPQ]* embryos at NC14 using an $(AAGAG)_7$ probe. Scale bars, 10μm. E. The percentage of unaligned total RNA-seq reads that contained the satellite repeat listed. Total RNA-seq was performed on 2–2.25 hr AEL *sfGFP-GAF(N)* homozygous and GAF[deGradFP] embryos[27]. See also Figure S6.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
α-GFP	Abcam	ab290, ab6556
α-DM1A (tubulin)	Sigma	T6199
Goat α-mouse IgG-HRP	Bio-Rad	1706516
Goat α-rabbit IgG-HRP	Bio-Rad	1706515
α-digoxigenin Alexa 488	Jackson ImmunoResearch	200–541-156
Goat α-rabbit Alexa Fluor 546	Thermo Fisher Scientific	A11035
Goat α-mouse DyLight 550	Thermo Fisher Scientific	84540
Goat α-rabbit DyLight 488	Thermo Fisher Scientific	35552
α-RNA polymerase II CTD repeat YSPTSPS (phospho S5)	Abcam	ab5131
α-HP1a	DSHB	C1A9
α-H3K9me3 (rabbit -used for ChIP-seq)	Active motif	39161
α-H3K9me3 (mouse -used for immunostaining)	Active motif	2AG-6F12-H4
α-GAF (rabbit serum raised against 1–130aa of GAF)	This paper	N/a
Chemicals, peptides, and recombinant proteins		
16% Paraformaldehyde Aq solution		50980487
Axygen magnetic beads	Thermo Fisher Scientific	14–223-152
Glycogen	Invitrogen	10814010
DAPI	Thermo Fisher Scientific	D1306
Prolong Gold Antifade Mountant	Invitrogen	P10144
Western Blocking Reagent	Sigma	11921673001
Trizol	Invitrogen	15596026
Superscript IV	Invitrogen	18090050
RNAse A	Thermo Fisher Scientific	FEREN0531
SuperSignal West Pico PLUS chemiluminescent substrate	Thermo Fisher Scientific	34577
37% formaldehyde	Thermo Fisher Scientific	F79500
Proteinase K	Thermo Fisher Scientific	EO0491
NEBuilder® HiFi DNA Assembly Master Mix	New England Biolabs	E2621S
Critical commercial assays		
NEB Next Ultra II library kit	New England Biolabs	E7645S
Tn5 Transposase (Tagment DNA Enzyme)	Illumina	15027865

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Minielute Cleanup Kit	Qiagen	28204
Deposited data		
Raw and analyzed ChIP-seq data (H3K9me3)	This paper	GEO: GSE218020
Raw and analyzed ChIP-seq data (sfGFP-GAF$^{S\ PQ}$)	This paper	GEO: GSE218020
Raw and analyzed ChIP-seq data (DBD-sfGFP)	This paper	GEO: GSE218020
Raw and analyzed Hi-C data (control and GAF^{deGradFP})	This paper	GEO: GSE218020
Experimental models: Cell lines		
Mus musculus, H3.3-GFP	Gaskill et al. 27	N/a
Experimental models: Organisms/strains		
D. melanogaster, His2Av-RFP (III)	Bloomington	BDSC #23650
D. melanogaster, mat-α-GAL4-VP16 (II)	Bloomington	BDSC #7062
D. melanogaster, Trf^{s2325}	Bloomington	BDSC #12088
D. melanogaster, CTCF-GFP	Bloomington	BDSC #64810
D. melanogaster, GFP-modifier of mdg4(mod(mdg4))	Bloomington	BDSC #51351
D. melanogaster, Multi sex combs (Mxc)-GFP	Bloomington	BDSC # 84130
D. melanogaster, Polycomb (Pc)-EGFP	Bloomington	BDSC #9593
D. melanogaster, MS2 coat protein (MCP)-GFP	Garcia et al. 49	N/a
D. melanogaster, UAS-dsRNA-Su(var)205	Bloomington	BDSC #36792
D. melanogaster, HP1a-RFP (II)	Wen et al. 91	N/a
D. melanogaster, sfGFP-GAF (N)	Gaskill et al. 27	N/a
D. melanogaster, nos-deGradFP (II)	Gaskill et al. 27	N/a
D. melanogaster, GAF-sfGFP (C)	Gaskill et al. 27	N/a
D. melanogaster, nos-full-length GAF(1–519aa)-sfGFP (II)	This paper	N/a
D. melanogaster, nos-DBD-PQ (311–519aa)-sfGFP (II)	This paper	N/a
D. melanogaster, nos-PQ (426–519aa)-sfGFP (II)	This paper	N/a
D. melanogaster, nos-sfGFP-BTB/POZ-DBD (1–391aa) (II)	This paper	N/a
D. melanogaster, nos-sfGFP-BTB/POZ (1–122aa) (II)	This paper	N/a
D. melanogaster, nos-IDR-DBD (123–391aa)-sfGFP (II)	This paper	N/a
D. melanogaster, nos-full-length zinc finger mutant (519aa, C344S, C347S)-sfGFP (II)	This paper	N/a
D. melanogaster, nos-DBD (310–391aa)-sfGFP (II)	This paper	N/a
D. melanogaster, tll-MS2(24x) (II)	This paper	N/a
D. melanogaster, GAF^{L}	This paper	N/a
D. melanogaster, GAF$^{S\ PQ}$	This paper	N/a
D. melanogaster, mCherry-GAF	This paper	N/a
D. melanogaster, sfGFP-GAF$^{S\ PQ}$	This paper	N/a
D. virilis	Sean Carroll, UW-Madison	N/a

REAGENT or RESOURCE	SOURCE	IDENTIFIER
D. melanogaster, w¹¹¹⁸; Df(3L)ED4543, P{w[+mW.Scer\FRT.hs3]=3'.RS5+3.3'}ED4543/TM6C, cu¹Sb¹	Bloomington	BDSC #8073
D. melanogaster, y¹ w; P{UAS-Nslmb-vhhGFP4}2	Bloomington	BDSC #38422
D. melanogaster, w¹¹¹⁸	Bloomington	BDSC #3605
D. melanogaster, y¹ w; P{matalpha4-GAL-VP16}67; P{matalpha4-GAL-VP16}15	Bloomington	BDSC #80361
Oligonucleotides		
RNA-FISH probe 5'GAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAATCTCCCTTTAGTGAGGGTTAATT-3'		
DNA-FISH probe 5'/5Cy5/AAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAG-3'	Integrated DNA Technologies	N/a
gRNA to generate *GAF^Δ*: 5'GCGGCAGTCTTCTCACCAGC-3'	Integrated DNA Technologies	N/a
gRNAs to generate *GAF^{S PQ}*: 5'AGCCTTCAATCATTCCAACG-3' and 5'ACGAGAGTGATATCGAATGC-3'	Integrated DNA Technologies	N/a
gRNA to generate *mCherry-GAF*: 5'TAAACATTAAATCGTCGTGT- 3'	Integrated DNA Technologies	N/a
Software and algorithms		
ImageJ	Schindelin et al. 93	https://imagej.nih.gov/ij/
IGV	Robinson et al. 104	https://software.broadinstitute.org/software/igv/
bowtie 2 v2.3.5	Langmead and Salzberg 99	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
MACS v2	Zhang et al. 100	
DeepTools	Ramírez et al. 102	
NGMerge	Gaspar 103	
Bedtools	Quinlan and Hall 101	
PONDR		http://www.pondr.com/
Other		
Branson Sonifier 250	Branson Ultrasonics	
Nikon A1R+ confocal microscope	Nikon	
Leica SP8 WLL confocal microscope	Leica Microsystems	