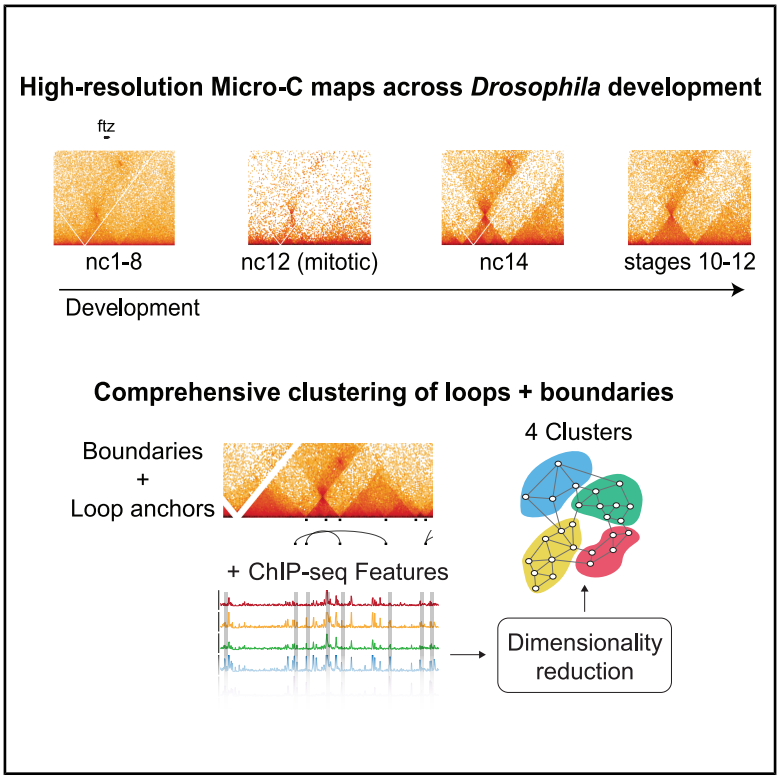
Cell Genomics

Resource

3D chromatin structures precede genome activation in *Drosophila* embryogenesis

Graphical abstract



Authors

Gabriel A. Dolsten, Evan M. Cofer, Xin Yang Bing, ..., Olga G. Troyanskaya, Michael S. Levine, Yuri Pritykin

Correspondence

xbing88@gmail.com (X.Y.B.), msl2@princeton.edu (M.S.L.), pritykin@princeton.edu (Y.P.)

In brief

Dolsten et al. map the 3D chromatin architecture across early *Drosophila* development using high-resolution Micro-C, and, by integrative analysis with 149 public ChIP-seq datasets, identify four classes of structural regulatory elements that form loops or insulate chromatin. Chromatin structures emerge earlier than previously described and are retained through mitosis.

Highlights

- High-resolution Micro-C maps across early *Drosophila* embryogenesis
- Chromatin structures form prior to zygotic genome activation
- ChIP-seq-based analysis defines four clusters of looping and boundary elements
- GAF identified as a mitotic bookmark conferring chromatin insulation

Dolsten et al., 2025, Cell Genomics *5*, 101002
November 12, 2025
https://doi.org/10.1016/j.xgen.2025.101002

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Resource

3D chromatin structures precede genome activation in *Drosophila* embryogenesis

Gabriel A. Dolsten,[1,2,6] Evan M. Cofer,[1,2,6] Xin Yang Bing,[1,6,*] Benjamin Brack,[1,3] Marcus Curlin,[1,3] Chandra L. Theesfeld,[1] Olga G. Troyanskaya,[1,4,5] Michael S. Levine,[1,3,*] and Yuri Pritykin[1,5,7,*]

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
[2]Graduate Program in Quantitative and Computational Biology, Princeton University, Princeton, NJ 08544, USA
[3]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA
[4]Flatiron Institute, Simons Foundation, New York, NY 10010, USA
[5]Department of Computer Science, Princeton University, Princeton, NJ 08540, USA
[6]These authors contributed equally
[7]Lead contact
*Correspondence: xbing88@gmail.com (X.Y.B.), msl2@princeton.edu (M.S.L.), pritykin@princeton.edu (Y.P.)
https://doi.org/10.1016/j.xgen.2025.101002

SUMMARY

3D chromatin structure is critical for the regulation of gene expression during development. Here we used Micro-C assays at 100-bp resolution to map genome organization in *Drosophila melanogaster* throughout the first half of embryogenesis. These high-resolution contact maps reveal fine-scale features such as loops and boundaries delineating topologically associating domains. Notably, we observe that 3D chromatin structures form prior to zygotic genome activation and persist during successive mitotic cycles. Integrative analysis with 149 public chromatin immunoprecipitation sequencing (ChIP-seq) datasets identifies four classes of chromatin structuring elements, including a distinct group enriched for GAGA-associated factor (GAF) and Zelda binding, associated with developmental-gene regulation. These elements are mitotically retained and exhibit sequence and structure similarity between *D. melanogaster* and *D. virilis*. We propose that 3D chromatin organization in the pre-cellular embryo facilitates deployment of developmentally regulated genes during *Drosophila* embryogenesis.

INTRODUCTION

The packaging of DNA into chromatin is a unifying feature of eukaryotic genomes,[1,2] and 3D chromatin structures play critical roles in a variety of cellular processes, including transcription and cell division.[3] In metazoans such as flies and humans, interphase chromatin structure is hierarchically organized at multiple scales.[2] At the highest level, different chromosomes are organized in distinct territories within the nucleus[4] and separated into extended A and B chromatin compartments.[5–7] Chromatin is further organized into contiguous self-interacting regions known as topologically associating domains (TADs), demarcated by TAD boundaries defined by strong local insulation.[8] In addition to TADs, chromatin can form long-range loops within TADs or spanning distant TADs.[9–14] Loops appear as distinct foci in contact-frequency maps.[9–12] These structures are critical for transcriptional regulation, and facilitate a number of key regulatory mechanisms, including enhancer-promoter contacts and Polycomb repression.[12,15–17] Perturbing these structures can trigger disease by dysregulating gene expression.[18–20]

In vertebrates, the formation of loops and TADs is thought to depend largely on loop extrusion, a process where elongating cohesin complexes are halted by convergently oriented CCCTC-binding factor (CTCF) pairs.[2,21–23] CTCF has also been implicated in boundary formation in *Drosophila* where it appears to interact with BEAF-32 and CP190.[24–29] However, although mammals employ CTCF for formation of most TADs and loops, *Drosophila* instead relies on a diverse set of insulators and looping factors such as GAGA-associated factor (GAF), Polycomb repressive complex 1 (PRC1), CTCF, BEAF-32, and CP190, each with locus-specific roles.[10,11,30–33] Understanding the grammar and syntax of fine-scale chromatin structures, such as loops and TADs, and the factors that control their formation, is critical for elucidating mechanisms of gene regulation.

Drosophila embryogenesis is a particularly relevant system for studying the relationship between chromatin architecture and gene regulation since *Drosophila* embryos must precisely pattern gene expression and perform rapid cell divisions. In particular, zygotic genome activation (ZGA)—the point at which the embryo begins independent transcription of mRNA—is a hallmark of early embryogenesis and a dramatic shift in gene regulation.[34] It is thought that chromatin structure is essentially disordered in *Drosophila* up to nuclear cycle (nc) 8, with a gradual emergence of structure throughout minor ZGA (nc9–13)[11,35–37] and rapid changes in structure at nc14 (the major wave of ZGA),[37,38] broadly matching observations in humans and mice.[34,39–42] However, previous studies of pre-ZGA chromatin structures relied on the Hi-C assay, which can be resolution

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

limited by restriction-enzyme cutting sites and is less effective at recovering fine-scale chromatin structures than Micro-C.[43,44] Additionally, chromatin structure has never been profiled in mitotic nuclei, raising the question of how chromatin structure is maintained across multiple rapid cell cycles. To further explore the role of chromatin-structure dynamics in *Drosophila* embryogenesis, we used Micro-C[43] to map chromatin dynamics across embryogenesis at single-nucleosome resolution.

High-resolution Micro-C identified hundreds of novel boundaries and loops, which we used to quantitatively measure the gradual refinement of chromatin structure during embryogenesis. Contrary to previous results, we identified an extensive degree of chromatin structure, including both loops and boundaries, prior to ZGA. These structures, which emerged as early as nuclear cycles 1–8, persisted through mitosis into ZGA and later stages of embryonic development. Leveraging 149 published chromatin immunoprecipitation sequencing (ChIP-seq) tracks, we identified four distinct classes of chromatin-structuring elements (CSEs). Transcriptionally active CSEs could be categorized into housekeeping and developmental-associated groups, each with distinct structural properties and transcription factor (TF)-binding profiles.[45] Notably, development-associated boundaries exhibited heightened structural similarity in *Drosophila virilis*, which may indicate a conserved role for their associated chromatin regulators.

Together, these results shed light on the dynamic positioning of CSEs prior to ZGA and suggest that evolutionarily conserved regulatory elements form poised chromatin contacts in both pre-ZGA and mitotic embryos to support the rapid deployment of chromatin contacts and establishment of functional gene regulation during ZGA.

RESULTS

Micro-C provides an unparalleled view of embryonic chromatin structure

We sought to exploit the improved resolution of Micro-C to study chromatin structure during *Drosophila* embryogenesis, with a specific focus on nc1–8 (pre-ZGA), nc14 (ZGA), and stages 10–12 (s10–12; post-gastrula germband stages) (Figure 1A; Table S1). In total, we obtained two biological replicates with two technical replicates for each stage (Figure S1A; Table S1). These data exhibited high resolution (Figure S1B; Table S1), reproducibility (Figures S1A and S1C), and fine-scale chromatin structures exceeding those of traditional Hi-C protocols (Figures 1A, 1B, and S1B). To identify these features genome-wide, we called loops and boundaries across all developmental stages. After merging these calls into a cross-developmental atlas (Figure S1D), we identified a total of 3,673 boundaries and 1,029 loops (Table S2), a sizable increase (e.g., 3.91× as many boundaries) compared to published Hi-C datasets from similar stages.[10,11,13,16,37] Thus we generated reproducible high-quality 3D maps at an unprecedented coverage and resolution for these three stages of *Drosophila* embryogenesis.

We sought to compare genomic regions that form loops and serve as insulators. The majority of loop anchors (57.3%) overlapped boundaries in our atlas (Figure 1C). Of the remaining loop anchors, most (71.98%) still overlapped "weak" boundaries that were not included in our atlas (STAR Methods). We identified far more (2.3-fold more) boundaries than loop anchors, and most (75.22%) did not overlap loop anchors. Notably, loop anchors that did not overlap boundaries typically fell into regions that were relatively distant from enhancers and promoters (STAR Methods). Of the boundaries in our atlas, the ones overlapping loop anchors were slightly but significantly stronger than the rest (Figure S1E). These observations suggest that boundary formation frequently occurs in the absence of loop formation in *Drosophila*. This is consistent with prior studies[46,47] suggesting that boundaries may form through different means than loops.

In sum, our replicated high-quality Micro-C data provide an unprecedented view of chromatin organization during *Drosophila* development.

Chromatin organization forms before ZGA

Published Hi-C datasets have uncovered little in the way of chromatin structure in pre-ZGA embryos, with most showing a near-uniform distribution of chromatin contacts and early structures emerging around nc12.[11,35–37,48] In contrast, our Micro-C data from nc1–8 embryos suggested a high degree of chromatin organization even in the earliest stages of embryogenesis, including definitive boundaries and loops as early as nc1–8 (Figure 1A). Furthermore, we have observed such pre-ZGA structures in previously published Hi-C data using aggregate pileup analysis of the features identified in our data (Figures S1F and S1G). We therefore conclude that high-resolution Micro-C maps identified reproducible pre-ZGA chromosomal structures.

We next asked whether these pre-ZGA structures were related to boundaries and loops seen at later stages of embryogenesis. Specifically, we quantified the observed strengths of loops and boundaries over time: the boundary strength being the degree to which a boundary inhibits spanning contacts, and the loop strength corresponding to the focal intensity in the contact-frequency matrix. Genome-wide boundary and loop strength appeared to increase significantly at each stage (Figures 1D and 1E). Boundary strength in pre-ZGA embryos was largely lower than in nc14 or s10–12, as ∼90% of boundaries increased substantially in strength (LFC > 0.4) across either comparison (Figures S2A and S2B). The dynamics of loop formation were generally consistent with the dynamics of boundary formation. Between nc14 and s10–12, loops mostly increased (n = 378) in strength (Figures S2C, S2D, and S3A). Between nc1–8 and s10–12, loops mostly increased in strength (n = 501 up; n = 24 down; Figures 1E, S2D, and S3B). Of the 24 loops with stronger contact in nc1–8 than in s10–12, the loops were still present in either s10–12 or nc14 (Figures S2D and S3C). Between nc1–8 and nc14, loops exhibited both increases (n = 283) and decreases (n = 159) in relative strength, suggesting dynamic regulation of loops at ZGA (Figures S2D and S3D–S3F). In sum, loops and boundaries present in nc1–8 persisted through developmental time.

Overall, this analysis identifies significant chromosomal boundaries and loops in pre-ZGA embryos and their relationship with structures observed at later developmental stages.

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

CellPress
OPEN ACCESS

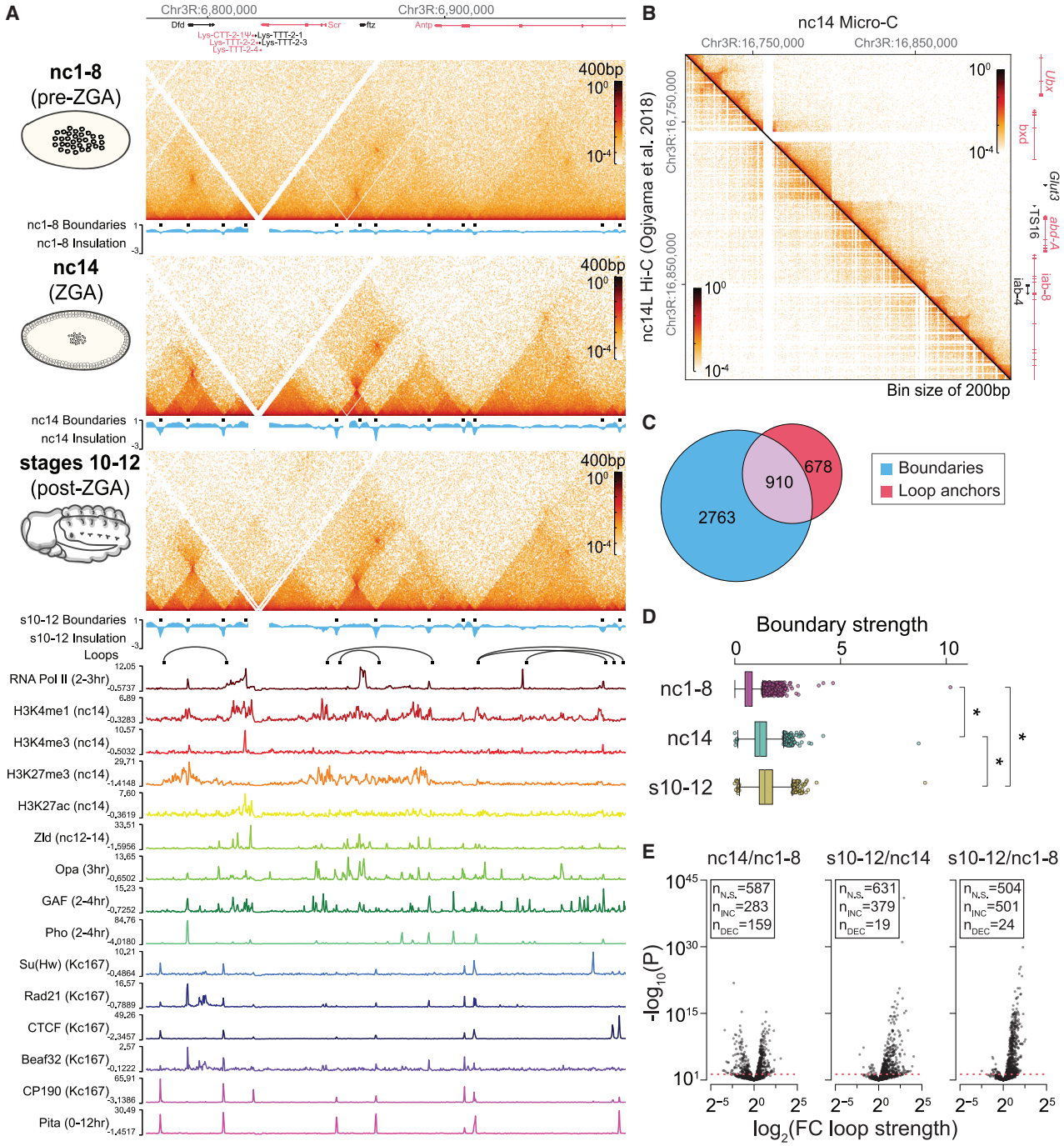


Figure 1. Micro-C uncovers chromatin structures with unparalleled resolution

(A) Micro-C contact-frequency maps surrounding the gene *ftz*. Contact-frequency maps for nc1–8 (top), nc14 (middle), and s10–12 (bottom) are shown alongside ChIP-seq tracks.

(B) Side-by-side comparison of Hi-C and Micro-C contact-frequency maps at the *abd-A* locus. The Micro-C data from this publication (upper triangle) are from nc14 embryos. The Hi-C data (Ogiyama et al.,[11] lower triangle) were collected from late nc14 embryos. The greater number of orange contact points to the right of the diagonal reflect more frequent contacts discovered/measured with Micro-C.

(C) The overlap between loop anchors and boundaries.

(D) Boxplot summary of boundary strengths in each embryonic stage. *$p < 0.001$, Bonferroni-corrected two-sided Wilcoxon signed-rank test, $n = 3,673$.

(E) Differential Hi-C signal between embryonic stages at loops (INC, increasing; DEC, decreasing; NS, not significant; see STAR Methods for details). The number of loops with significantly increasing, decreasing, or not significantly different strength are specified for each pairwise comparison.

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

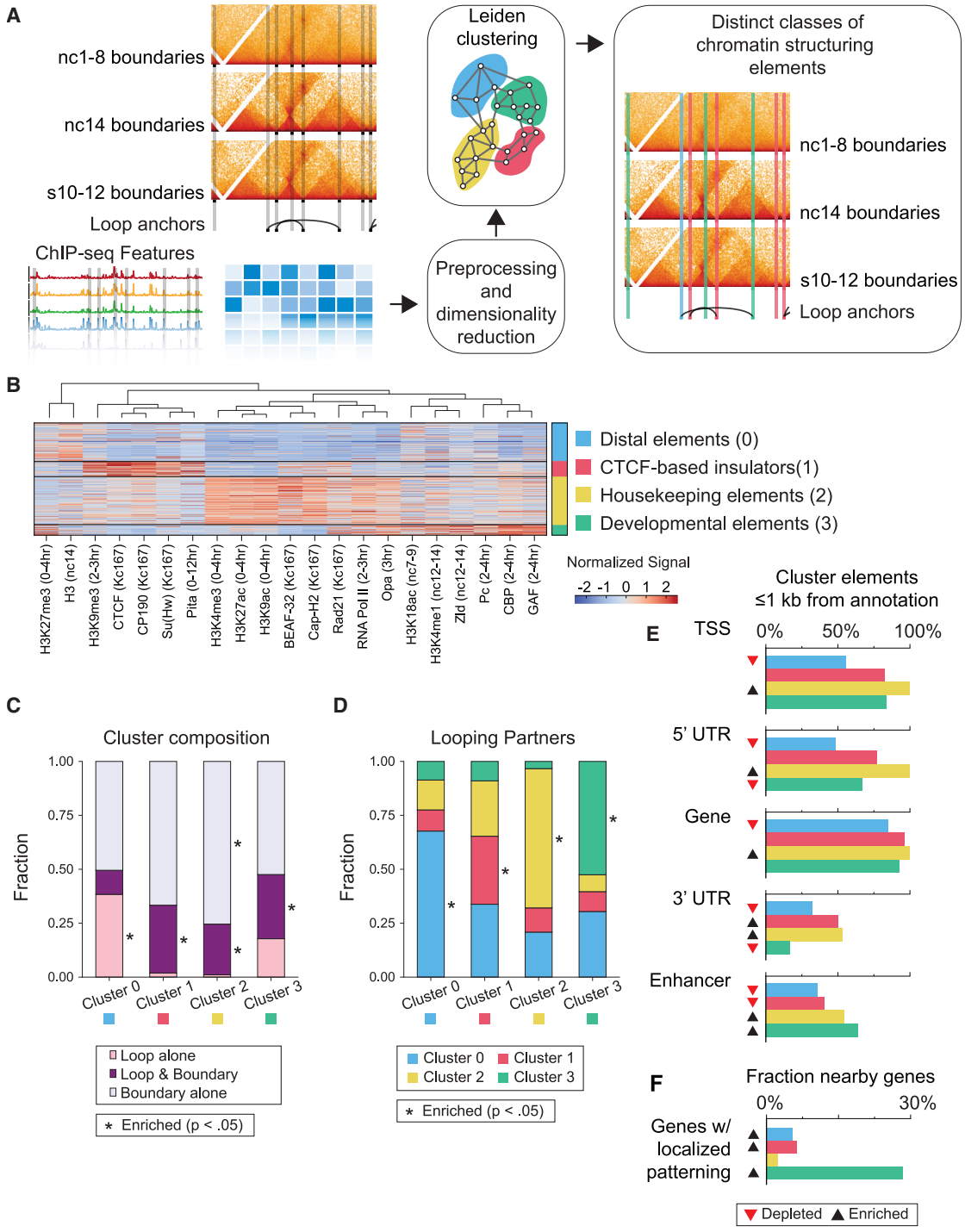


Figure 2. Joint clustering of 149 ChIP-seq datasets at chromatin boundaries and loop anchors defines four distinct classes of CSEs

(A) Schematic of the approach. The compendium of ChIP-seq data for various factors across biological contexts was collected and uniformly preprocessed. Loop anchors and boundaries were aggregated into an atlas of CSEs. The CSEs were clustered based on the vector of ChIP-seq signal across the compendium.

(B) Column-normalized heatmap of ChIP-seq signal at each CSE in the atlas. We visualized only ChIP-seq datasets that were significantly enriched and with a log fold change of at least two for at least one cluster relative to all other elements.

(C) Bar plots showing the structural composition of each CSE cluster. Enrichment was tested with a Bonferroni-corrected two-sided Fisher's exact test.

(D) Bar plots of loop anchor partnering preferences for each cluster of CSEs. Enrichment was tested with a Bonferroni-corrected two-sided Fisher's exact test.

(legend continued on next page)

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

Integration of Micro-C and ChIP-seq datasets defines four classes of CSEs

We next sought to determine which regulatory factors might underlie the formation of boundaries and loops and how different combinations of factors might give rise to unique structural properties and regulatory functions. To that end, we created a unified atlas of CSEs that combined both boundaries and loop anchors (STAR Methods; Figure S1D). We overlaid this atlas with 149 ChIP-seq datasets from varied biological contexts (Figures 2A and S4A; Table S3; STAR Methods). Of these datasets, 65% were from pre-ZGA, ZGA, or post-ZGA time points directly relevant to the time points profiled with Micro-C. Although many of the TFs in our atlas are bound to both CSEs and non-CSEs, we found a number of factors whose strongest peaks were enriched for CSE binding. Specifically, out of the strongest 10% of CTCF peaks, nearly 60% intersected boundaries, as opposed to just 8% of all CTCF peaks (Figure S4B). Likewise, at least 70% of the strongest peaks for BEAF-32, Pc, and DNA replication-related element factor (DREF) intersected CSEs (Figure S4B). Interestingly, only 14% of the strongest GAF peaks intersected CSEs. These results suggest that CTCF, BEAF-32, Pc, and DREF operate primarily at CSEs, with a broad correspondence between insulator occupancy and insulation function.[49] In contrast, GAF may perform additional functions at other sites, such as GA-rich heterochromatic regions.[49–52]

CSEs were further classified based on their epigenomic patterns. We formed a vector of normalized ChIP-seq signal across the compendium for each CSE and then clustered such vectors for all CSEs using the Leiden algorithm[53] (Table S2; STAR Methods). We identified four clusters of CSEs associated with different combinations of regulatory factors (Figure 2B): (cluster 0) distal elements (34% of all CSEs); (1) CTCF-mediated insulators (13%); (2) promoter-proximal elements (43%); and (3) developmental elements (9%). The explanations for these functional characterizations of the clusters are provided below.

Cluster 1 CSEs were associated with the classic insulator proteins CTCF and CP190 (Figure 2B) and were composed almost entirely of boundaries, including those that overlapped loop anchors (Figure 2C). Boundaries from cluster 1 were also the strongest at s10–12 and weaker at nc1–8 (Figures S5A–S5C). They did not exhibit strong differences between ZGA and s10–12,[54] consistent with the apparent dispensability of CTCF for the majority of chromatin architecture during embryogenesis (Figure S2).[29,55] We suggest that cluster 1 CSEs serve a general-purpose architectural role in *Drosophila* but are not essential for transcriptional regulation during early embryogenesis.

The promoter-proximal elements of cluster 2 comprised the largest cluster of CSEs. They were associated with BEAF-32 (Figure 2B) and almost always coincided with boundaries (Figure 2C). The few loops that were formed primarily linked promoters (Figure S5D). Cluster 2 boundaries were particularly strong in post-nc1–8 stages, suggesting that they functioned as insulators during and after ZGA (Figures S2A–S2C and S5A–S5C).

The developmental elements of cluster 3 showed an association with PRC1 members and developmental pioneer factors such as GAF and Zld (Figure 2B). Consistent with the idea that GAF associates with tethering elements,[16,52,56,57] we noted that CSEs in cluster 3 frequently formed loops and exhibited a strong preference to pair with other cluster 3 loop anchors (Figures 2C and 2D). These loops occurred over shorter genomic distances than other clusters (Figure S5E) and were especially strong during nc14 compared to all other loops (Figure S5F). These elements also frequently overlapped enhancers and sites of active transcription (Figure 2E). 28% of genes overlapping cluster 3 loop anchors intersected with the previously curated list of genes associated with localized developmental patterning, compared to <6% for all other clusters, confirming the association between cluster 3 CSEs and critical developmentally associated patterning genes (Figure 2F).[12] Cluster 3 boundaries were also frequently neighbored by other cluster 3 boundaries, consistent with enriched interaction and looping between these sites (Figure S5G). Altogether, this analysis suggests that cluster 3 CSEs facilitate enhancer-driven transcriptional regulation during ZGA and are likely to correspond to recently identified tethering elements that mediate enhancer-promoter and promoter-promoter interactions of developmental patterning genes.[16]

Unlike clusters 1, 2, and 3, cluster 0 was not obviously associated with specific transcription factors, and was only marginally associated with H3K27me3 and H3 (Figure 2B). Cluster 0 CSEs generally coincided with loop anchors, including those that overlapped boundaries (Figure 2C), but they rarely occur near sites of transcription (e.g., genes, enhancers) (Figure 2D). The function of cluster 0 CSEs remains uncertain.

In summary, our high-resolution Micro-C analysis coupled with analysis of a compendium of publicly available ChIP-seq datasets enabled a comprehensive characterization of distinct classes of chromosomal structural elements. Nearly 10% of these CSEs, those comprising cluster 3, appear to play an outsized role in the regulation of gene expression during ZGA, as we discuss below.

Cluster 2 elements are associated with BEAF-32 and the regulation of housekeeping genes

Clusters 1 and 2 were composed almost entirely of boundaries and exhibited binding of classic insulator factors, particularly CTCF and CP190.[28] They were also associated with Pita and Su(Hw), which have been shown to interact with CP190.[58,59] In keeping with previous work showing a division between BEAF-32 and CP190 boundaries,[29] we observed that cluster 2 CSEs had an elevated level of ChIP-seq signal for the known and conserved insulator BEAF-32.[60] We also noted that CSEs in cluster 2 were often located at traditional sites of transcription such as enhancers or transcription start sites (TSSs), whereas cluster 1 CSEs were depleted of overlap with enhancers relative to other CSEs (Figure 2D). Both clusters were associated with 3′ UTRs, where they presumably insulate genes from

(E) Bar plots showing the cluster-wise percentage of CSEs within 1 kb of a given annotation. Enrichment was tested with a Bonferroni-corrected two-sided Fisher's exact test.

(F) Bar plot of the cluster-wise fraction of genes within 1 kb of a CSE, which have been annotated as having localized patterning during embryogenesis (patterning gene list taken from Levo et al.[12]). Enrichment was tested with a Bonferroni-corrected two-sided Fisher's exact test.

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

downstream elements.[25,61] Although both cluster 2 and cluster 3 CSEs were enriched for RNA polymerase II (Figure 3A), cluster 2 boundaries were exclusively enriched for common markers of active transcription such as H3K4me3, H3K27ac, and H3K9ac[62] (Figures 3A and2B). We also noted an enrichment for Cap-H2 condensin subunit at cluster 2 CSEs—another canonical marker of promoter-proximal insulators[63]—and elevated levels of ChIP-seq signal for Chro, Dref, and Pzg, which have previously been shown to co-localize and physically interact with BEAF-32.[64,65]

Based on the molecular differences between cluster 2 and cluster 3 boundaries, we hypothesized that the two clusters might be associated with distinct patterns of transcriptional regulation. Previous reports have suggested that BEAF-32 is enriched at the promoters of housekeeping genes.[46,68] Indeed, 37% of cluster 2 boundaries overlapped housekeeping genes such as SppL and Pxd, compared to 5% for cluster 3 CSEs (Figures 3B and 3C), and Gene Ontology (GO) term enrichment analysis of CSE-flanking genes confirmed that cluster 2-associated genes were enriched for housekeeping functions (e.g., mitosis, translation) (Figure 3D; STAR Methods). In contrast, cluster 3 CSEs were more closely associated with developmental processes that varied over the course of embryogenesis (e.g., regulation of neurogenesis) (Figure 3D; STAR Methods). We also observed that the genes in cluster 3 CSEs rose sharply in expression at ZGA and steadily tapered off during the remainder of embryogenesis, while those flanking CSEs in cluster 2 showed relatively high levels of expression throughout development (Figure 3E). Together, these results indicate that cluster 2 CSEs were associated with insulation for housekeeping genes, whereas cluster 3 CSEs were associated with temporal coordination of developmentally associated gene-expression programs.

Cluster 3 elements are shared between *D. virilis* and *D. melanogaster*

To explore the role of cluster 3 CSEs, we further examined their molecular features. Zld and GAF are TFs that are typically found in developmental enhancers and have been linked to a multitude of regulatory functions in embryogenesis and particularly at ZGA.[11,32,37,50,69–71] Cluster 3 was strongly associated with both Zld and GAF (Figures 4A, S6A, and S6B) and showed a spike of chromatin accessibility just prior to ZGA (Figures 4B and S6C). In addition, cluster 3 loops were frequently differential in contact frequency between nc14 and other stages, suggesting that these loops are tightly regulated during ZGA (Figure S6D). Together, these results point to an important role of cluster 3 elements at ZGA.

Given the conserved roles of Zld[73] and potentially GAF,[74,75] we hypothesized that cluster 3 CSEs associated with these factors may have a conserved function in other drosophilids. To that end, we performed Micro-C on *D. virilis* embryos and identified CSEs shared with *D. melanogaster* (STAR Methods). Indeed, a significant fraction of cluster 3 CSEs shared similar sequence and 3D structure between *D. virilis* and *D. melanogaster* (Figures 4C and 4D). This was in contrast to cluster 0 boundaries, which were enriched for sequence similarity but not structure similarity. Thus, cluster 3 CSEs associated with developmental regulatory elements exhibited the highest levels of similarity of both sequence and structure among all chromosomal structural elements.

Next, we examined the PhyloP27 evolutionary conservation scores at individual motifs overlapping accessible peaks. PhyloP27 scores are derived from a multiple sequence alignment of 27 insect species.[76,77] We found a significant correlation ($r = 0.34$) between the average conservation score of a motif in CSEs and its relative frequency of being found in loop anchors rather than boundary CSEs (Figure 4E). Developmentally associated motifs such as GAF and Zld were among the highest-conserved motifs, with GAF also being one of the most enriched in loops (Figure 4E). Therefore, we speculated that the tendency for cluster 3 CSEs to form loops that regulate critical developmental processes, perhaps mediated by GAF motifs, may explain their increased sequence and function conservation. Altogether, cluster 3 CSEs show a pattern of sequence and structure conservation in accordance with their unique developmental functions.

Nucleosomal positioning reveals unique roles of GAF, Zelda, and BEAF-32

Since Micro-C employs the MNase enzyme for cutting, Micro-C data can be analyzed for nucleosomal footprints.[43,78,79] In mammals, CTCF is known for both phasing nucleosomes as well as forming boundaries.[78] Therefore, we sought to understand how nucleosomal positioning relates to looping and boundary formation in *Drosophila*. We found that CSEs were broadly depleted of nucleosomes, consistent with their high accessibility (Figures 5A and 4B). Intriguingly, cluster 3 CSEs showed only a modest depletion of MNase signal despite being the most accessible cluster of CSEs (Figures 5A and 4B). Examining looping and non-looping CSEs separately revealed that loop anchor CSEs in cluster 0 and cluster 3 displayed a peak in MNase coverage, suggestive of a TF footprint, whereas boundary CSEs were depleted of nucleosome coverage (Figures 4B and S7A). For cluster 3, the difference in MNase coverage was nc14 specific (Figure S7B). We confirmed that this footprint was due to TF binding, since no footprint was observed when examining H3 ChIP-seq (Figure S7C).

We hypothesized that differential binding and footprinting of ZGA-specific factors could explain the differences in MNase coverage between loops and boundaries in cluster 3. Indeed, boundary cluster 3 CSEs showed stronger binding by the pioneer factor Zld than looping cluster 3 CSEs, and Zld peak strength correlated with depletion of MNase coverage in nc14 (Figures S7D and S7E). Together, these results suggested that Zld displaces nucleosomes to help establish nc14-specific cluster 3 boundaries. In contrast, cluster 3 loop anchors showed stronger binding by an array of factors including GAF, Pc, and Mod(mdg4) (Figure S7D). This high level of combinatorial protein binding may confer the observed MNase footprint and would be consistent with previous studies describing a large ~1,000-kDa multiprotein complex at certain tethering elements.[14,80,81]

Next, we examined nucleosomal positioning at TSSs. TSSs with high levels of RNA polymerase II binding showed well-positioned nucleosome arrays, consistent with previous studies.[82–84] Even stronger positioning was observed at TSSs bound by BEAF-32 (Figure 5B). In contrast, the strongest GAF-bound TSSs showed poor positioning of nucleosomes and low levels

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

CellPress
OPEN ACCESS

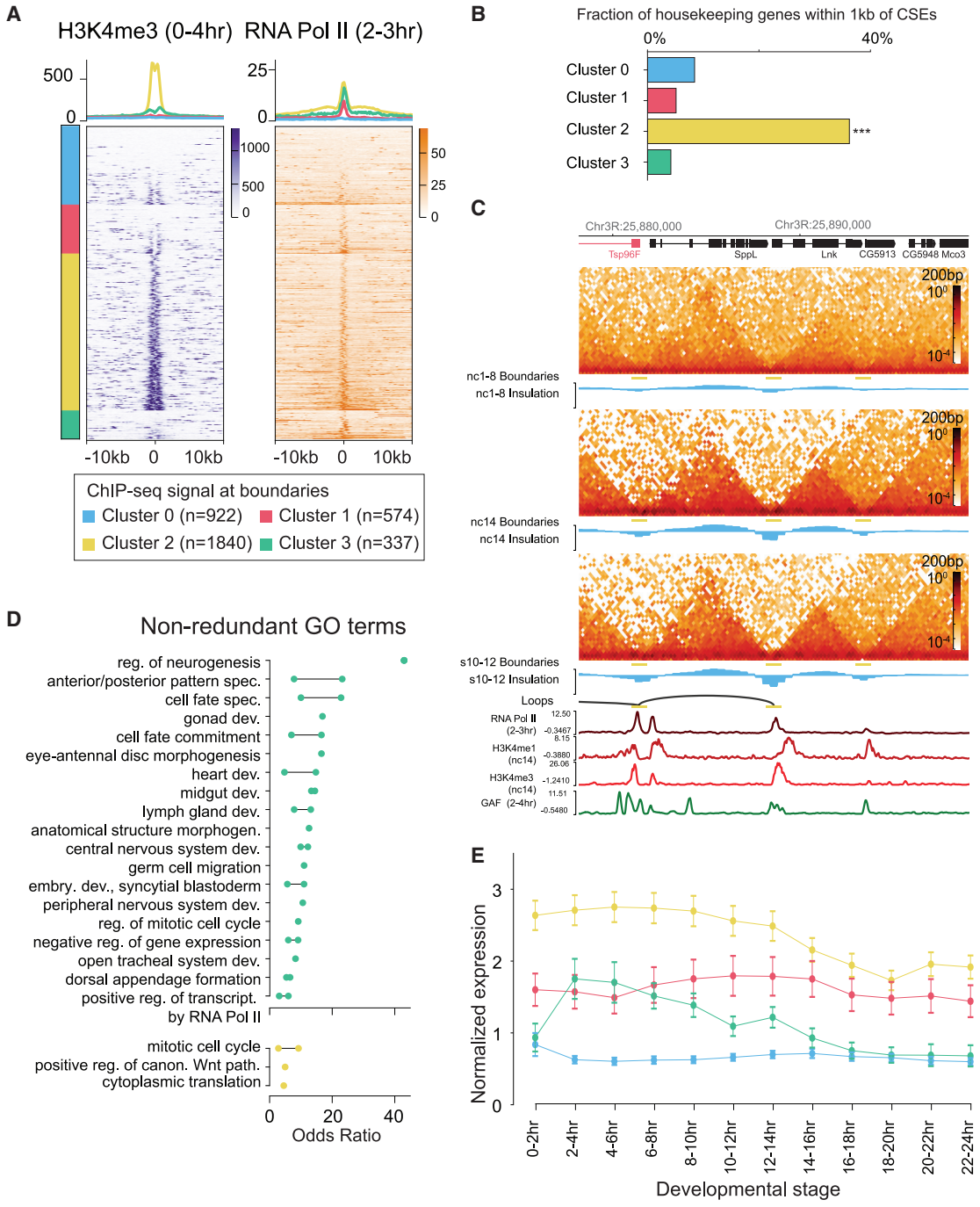


Figure 3. A distinct class of boundaries associated with early activation of housekeeping and cell-cycle genes

(A) Heatmaps showing H3K4me3 (0–4 h) and RNA Polymerase II (2–3 h) ChIP-seq signal at boundaries in each CSE cluster. Each row of the heatmap represents ChIP-seq signal centered at a CSE. Rows in each cluster are sorted by mean signal in the center 3.2 kb.

(B) Fraction of all housekeeping genes within 1 kb of a CSE of each cluster. (***$p < 1e{-}4$, Fisher's exact test).

(C) Genome Browser view of the *SppL* locus. Micro-C contact-frequency maps from nc1–8 (top), nc14 (middle), and s10–12 (bottom) are shown alongside ChIP-seq data. *SppL* and *Lnk* are linked to housekeeping functions such as proteolysis and core metabolic processes.[66,67]

(D) Non-redundant biological process GO terms that were significantly enriched among the set of genes flanking (≤5 kb) CSEs.

(E) Temporal trends in the normalized expression levels (±SEM) of genes flanking (≤500 bp) CSEs in each cluster. The mean normalized expression is plotted.

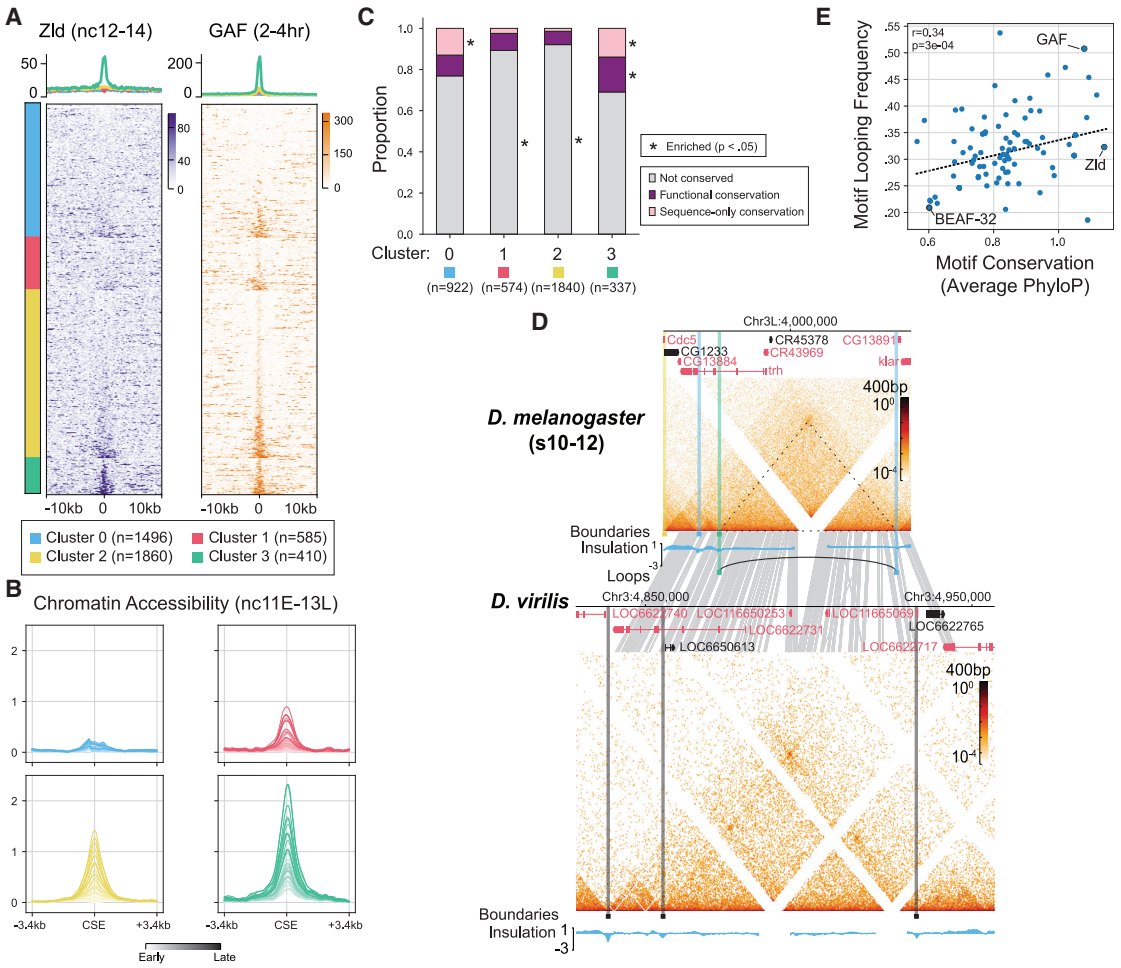


Figure 4. Regulatory elements associated with zygotic expression exhibit similarity of both sequence and structure

(A) Heatmaps showing Zld (nc12–14) and GAF (2–4 h) ChIP-seq signal at all CSEs in each cluster. Each row of the heatmap represents ChIP-seq signal centered at a CSE. Rows in each cluster are sorted by mean signal in the center 3.2 kb.

(B) Chromatin accessibility at CSEs as measured by ATAC-seq.[69] Accessibility is recorded in 3-min intervals from the beginning of nc11 to the end of nc13.

(C) Bar plots showing the cluster-wise similarity of CSEs in *D. virilis*. Enrichment was tested with a Bonferroni-corrected two-sided Fisher's exact test.

(D) Aligned multi-genome view of Micro-C contact-frequency maps for a locus *D. melanogaster* (top), which exhibits conserved structure in *D. virilis* (bottom). Note that LOC6622731 and LOC6622717 are orthologs of *trh* and *klar*, respectively.[72] The light gray bands between the two views indicate conserved sequences identified with lastz (STAR Methods).

(E) Scatterplot between motif looping frequency (fraction of motif instances in CSEs that overlap looping CSEs) and motif conservation (average PhyloP27 over all motif instances in ATAC-seq peaks in CSEs). r, Pearson correlation.

of H3K27ac despite high levels of RNA Pol2 binding, indicative of paused polymerases (Figure 5C).[85] Similar to the cluster 3 CSEs, we observed a strong MNase footprint directly upstream of the TSS, overlapping the GAF-binding site.[86–88] These TSSs included genes with known tethering elements such as beat-Ia and side-IV, as well as patterning genes such as engrailed and Krüppel. These TSSs were also enriched for cluster 3 CSEs (Figure S7F), and the corresponding genes were expressed at higher levels at later time points (Figure S7G). Therefore, we suggest that GAF binds directly upstream of the TSSs to facilitate pausing of Pol2 at the promoter. Abrogation of Pol2 pausing, perhaps through regulation by distal elements, could then enable rapid and temporal coordination of developmental-associated gene expression. These results reveal the broad diversity of functions by cluster 3 factors such as Zld and GAF, ranging from Zld-mediated insulation to GAF-mediated Pol2 pausing.

In sum, nucleosomal analysis using high-resolution Micro-C data reveals important patterns of transcription factor-driven regulation in the *Drosophila* embryogenesis.

GAF is associated with mitotic chromatin structures preceding ZGA

A GAF-associated MNase footprint was also observed in nc1–8 embryos, suggesting that GAF can bind even in prezygotic embryos undergoing rapid cell divisions (Figure S7H). Although chromatin dynamics during interphase differ radically from other

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

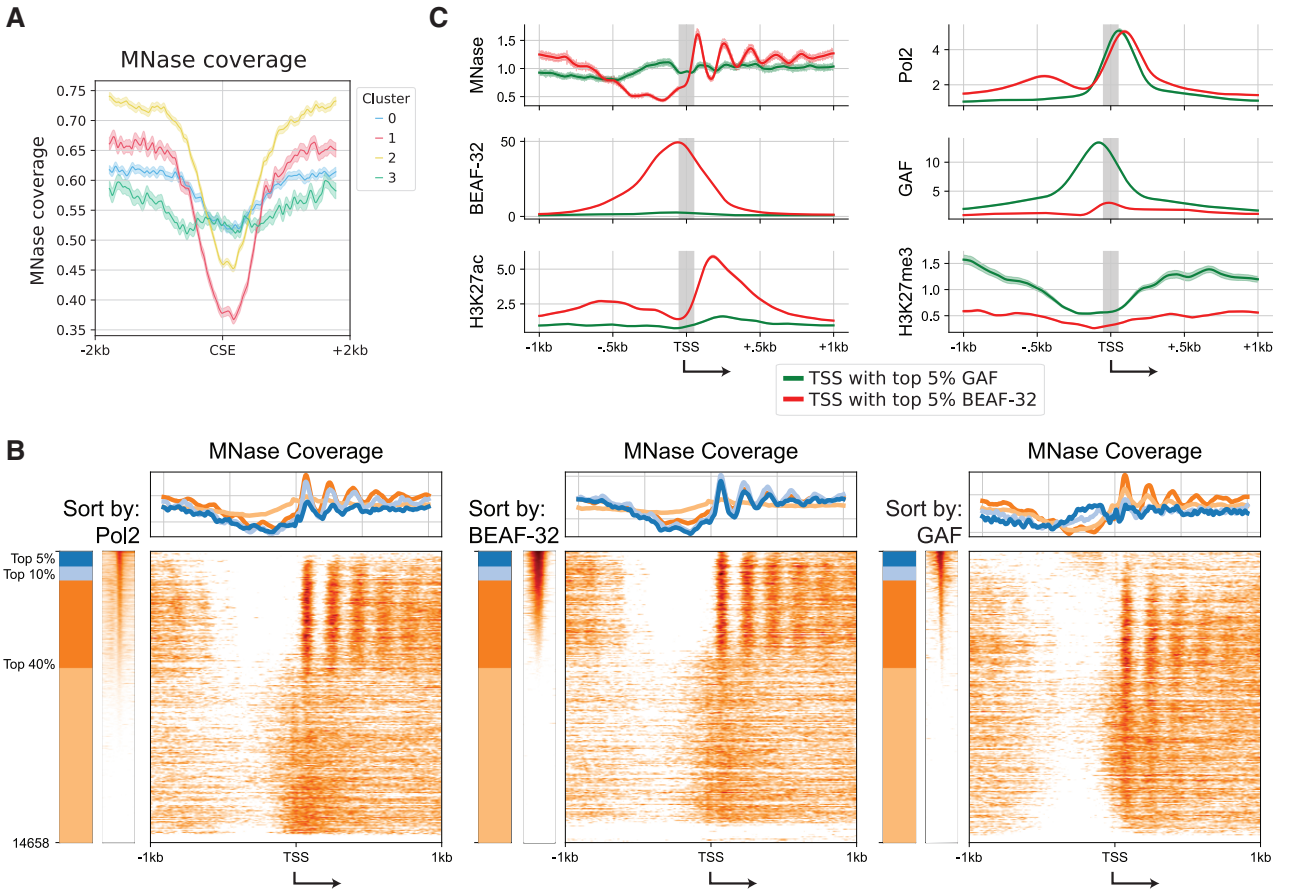


Figure 5. Nucleosomal analysis reveals distinct patterns of TF regulation

(A) MNase coverage at nc14 for all CSE clusters.

(B) MNase coverage for all TSSs, sorted by Pol2 (left), BEAF-32 (middle), or GAF (right) binding strength at the TSS. Clusters represent top 5%, 5%–10%, 10%–40%, and bottom 60% TSSs as ordered by the binding strength.

(C) MNase coverage or ChIP-seq data plotted around the TSSs with the top 5% of BEAF-32 or GAF binding. Outline represents ±1 SEM.

phases of the cell cycle,[89] a number of chromatin regulators remain localized to chromatin during mitosis,[71,90–92] and recent work has implicated GAF as a mitotic bookmark conferring transcriptional memory.[50,71] Thus, we sought to determine whether CSEs and their associated structures are preserved during mitosis. To do this, we performed Micro-C sequencing of mitotic embryos at nc12 (nc12M), and compared these with a second set of interphase embryos from nc14 (nc14v2) (Figures 6A and S8A). Consistent with previous work,[37] mitotic embryos exhibited a perturbed interaction-frequency decay curve with significant enrichment for long-range contacts beyond 150 kb (Figure S8B). Although large-scale organization is significantly altered in these embryos, we noticed a surprising degree of fine-scale structure (Figures S8A, S8C, and S8D). This is consistent with extensive work suggesting that the local structure of mitotic chromosomes is carefully regulated and structured despite a reduction in long-range loops.[9,93,94] In comparison to the other clusters, boundary strength at cluster 3 was preferentially preserved in mitotic embryos, although cluster 3 loops were weakened (Figures 6B, 6C, and S8E). ChIP-seq signal for several proteins implicated in developmental expression regulation such as GAF, Zld, and Opa was associated with boundaries that showed the smallest change in strength between the nc12M and nc14v2 stages (Figure 6D), and the GAF, Zld and Dref motif analysis at boundaries showed consistent results (Figure S8F). Next, we leveraged published GAF ChIP-seq data in mitotic embryos to determine whether mitotically retained GAF binding sites associated with stronger nc12M insulation.[71] Compared to interphase-only GAF-binding sites, mitotic GAF-binding sites were enriched for binding at CSEs, preferentially localized to cluster 3 CSEs, and showed no change in insulation score in nc12M (Figures 6E, S8G, and S8H). Combined with the known role of GAF as a stable mitotic bookmark conferring transcriptional memory,[71] the correspondence between mitotic GAF binding and mitotically stable boundaries suggests that GAF may maintain these structures during mitosis (Figures 6D and 6E).

DISCUSSION

Using systematic analysis of the reproducible high-resolution Micro-C data, we identified and comprehensively characterized thousands of fine-scale CSEs present as early as at nuclear cycles

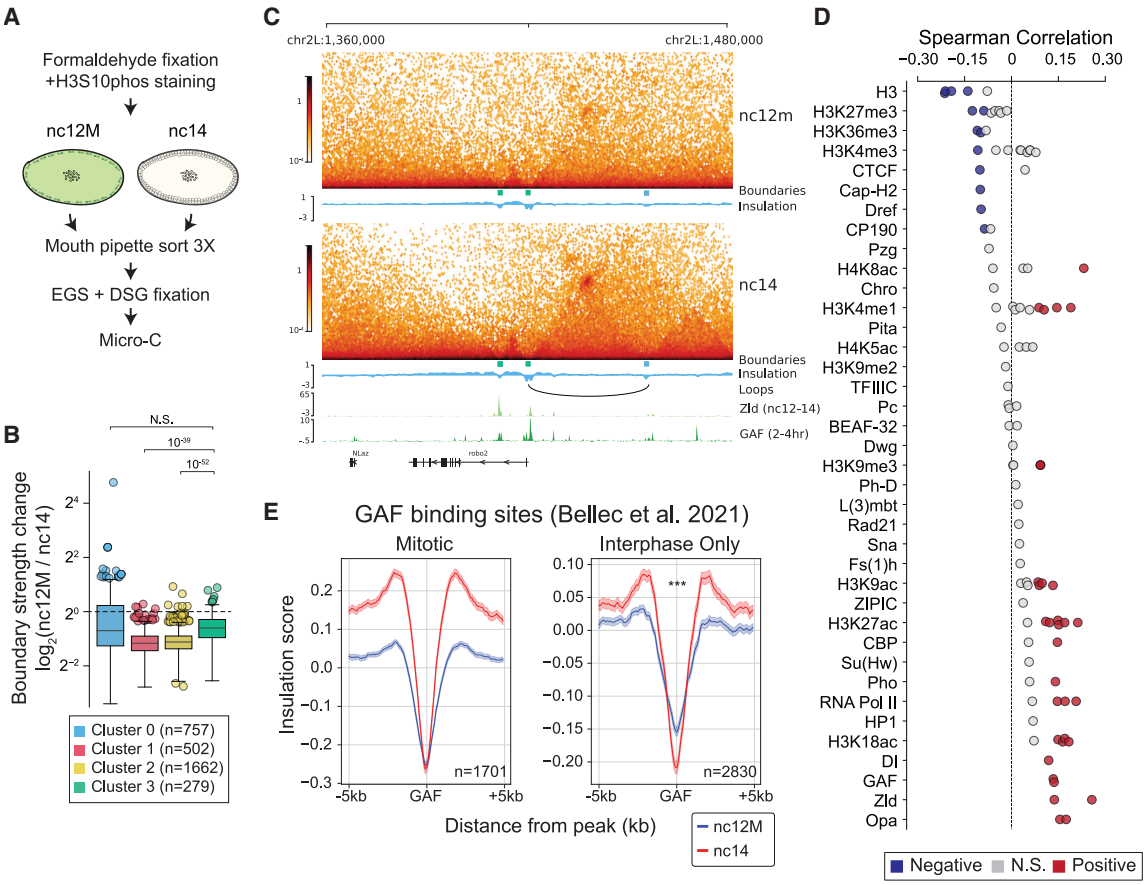


Figure 6. Developmental chromosomal structural elements are stably maintained throughout early mitoses

(A) Experimental method used to isolate mitotic embryos for the nc12 mitotic sample and the new nc14 sample.

(B) Boxplot summary of mitotic stability for CSEs separated by cluster. We excluded boundaries missing in nc1–8, nc12M, s10–12, or either of the nc14 samples. All pairwise comparisons of clusters were significant except where noted otherwise (Bonferroni-corrected two-sided Mann-Whitney U test; see supplemental table).

(C) Genome Browser view of the *robo2* locus. Micro-C contact-frequency maps from the nc12M (top) and nc14v2 (bottom) samples are shown alongside ChIP-seq data.

(D) Spearman's rank correlation between mitotic stability of boundaries and their signal in embryonic ChIP-seq datasets. Significant negative and positive correlations are colored blue and red, respectively (Table S2).

(E) Insulation in nc14v2 or nc12M at mitotically retained GAF binding sites (left) or interphase-only GAF binding sites (right) taken from Bellec et al..[71] Outline represents ±1 SEM. Statistical significance (***$p < 1e{-}4$; *$p < 0.05$) was tested using a Wilcoxon signed-rank test for insulation at the central value (corresponding to the CSE).

1–8 in *Drosophila* embryogenesis. These early chromatin structures persisted in mitosis and were continuously refined over the course of development, with a pronounced gain in organization at ZGA. Our observations are reproduced, in aggregate, in published data, and are consistent with results from orthogonal approaches—such as computational 3D modeling or oligopaint[35,36]—which have identified 3D structure as early as nc11.

Early studies of *Drosophila* chromatin architecture suggested that most chromatin structures emerged concurrently with gene activation during the minor and major waves of ZGA.[37,48] However, chromatin structures were largely preserved in the presence of transcription inhibition and did not vary in a tissue-specific manner, despite differences in gene expression between those tissues.[37,48] Together, these results suggested that transcription itself was not responsible for the onset of chromatin architecture, raising the question of how a rapid, global, and coordinated development occurs within the short time span of ZGA. Our work raises one possible answer: that chromatin structures are poised in the prezygotic embryo and preserved through mitosis to enable their rapid and faithful deployment during ZGA.

Through computational integration of diverse ChIP-seq datasets, we identified four distinct classes of CSEs that are present in the early embryo. These four classes of CSEs are associated with different regulatory factors and exhibit distinct characteristics, including genomic distribution and enrichment of genes associated with distinct biological processes. Our work extends early descriptions of active boundary clusters bound by combinations of CP190, CTCF, and BEAF-32.[95] Instead, we find that transcriptionally active boundaries are primarily bound by BEAF-32 and distinct from CTCF/CP190-associated boundaries.

Of our structural elements, the developmentally linked cluster 3 loops and boundaries were particularly stable across early mitotic cycles preceding ZGA. GAF binding in particular was found to be associated with mitotic structures, consistent with its role as a bookmark.[71] In addition, CSEs in cluster 3 were preferentially preserved—in both sequence and structure—in *D. virilis* compared to other CSEs. We therefore speculate that these regulatory elements are evolutionarily conserved mitotic bookmarks conferring timely establishment of chromatin structure and developmental-gene-expression programs.[71] We suggest that the precocious formation and retention of cluster 3 regulatory loops, possibly by GAF, may provide a 3D "pre-pattern" for the deployment of patterning genes during and after nc14.

Limitations of the study

Follow-up studies will be necessary to understand the changes in chromatin organization that occur during fertilization. This is necessary to understand the full continuum of chromatin structures in *Drosophila* and how they may be inherited from parental genomes, as well as contributions of parental macromolecules in establishing chromatin state in the embryo. While we did analyze a large number of ChIP-seq datasets for this study, additional datasets and continual improvements in ChIP-seq and related methods for assaying TF binding genome-wide will enable more complete characterization of the CSE cluster-specific regulation. Given the cell-cycle-dependent manner of chromatin-structure regulation, it may also be of interest to study whether mitotic kinases that target chromatin-regulating factors may modulate the establishment and dynamics of the chromatin structures identified in this work.[96]

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Yuri Pritykin (pritykin@princeton.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The code used for the analysis in this manuscript is available at GitHub (https://github.com/pritykinlab/drosophila-microc-zga) and Zenodo (https://doi.org/10.5281/zenodo.16621265).
- Raw sequencing data and processed mcool files generated in this study are deposited to NCBI GEO under accession number GSE265818.
- Processed ChIP-seq data (ChIP signal at all CSEs) have also been uploaded to GitHub (https://github.com/pritykinlab/drosophila-microc-zga) and Zenodo (https://doi.org/10.5281/zenodo.16621265).

ACKNOWLEDGMENTS

The authors acknowledge all members of the Levine, Pritykin, and Troyanskaya labs for their helpful discussions on the data analysis and manuscript. They also thank Michal Levo, Philippe Batut, João Raimundo, Xiao Li, and Pavan Choppakatla for many insightful suggestions over the years. The authors also thank Tsung-Han Stanley Hsieh for assistance with the Micro-C protocol and Nils Krietenstein for kind assistance with an early draft of the manuscript. This work was performed using the high-performance computing resources at Simons Foundation and the TIGRESS computer center at Princeton University. E.M.C. and G.A.D. were supported by grant T32 HG003284 from the National Institutes of Health. Y.P. was supported in part by NIH/NIAID grant DP2AI171161 and the Ludwig Institute for Cancer Research. E.M.C. was supported by the National Science Foundation Graduate Research Fellowship Program. X.Y.B was supported by an NIH grant to M.S.L. (GM118147).

AUTHOR CONTRIBUTIONS

Conceptualization, X.Y.B., M.S.L., Y.P., O.G.T., G.A.D., and E.M.C.; methodology, X.Y.B., E.M.C., G.A.D., Y.P., O.G.T., and M.S.L.; investigation, X.Y.B., G.A.D., E.M.C., B.B., and M.C.; formal analysis, E.M.C., G.A.D., and X.Y.B.; software, G.A.D. and E.M.C.; data curation, E.M.C., G.A.D., and X.Y.B.; writing – original draft, X.Y.B., E.M.C., and G.A.D.; writing – review & editing, G.A.D., E.M.C., X.Y.B., C.L.T., O.G.T., M.S.L., and Y.P.; funding acquisition, M.S.L., O.G.T., and Y.P.; supervision, Y.P., M.S.L., O.G.T., and C.L.T.

DECLARATION OF INTERESTS

E.M.C. is currently an employee at Asimov. X.Y.B. is currently an employee at BlueRock Therapeutics.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Fly husbandry
  - Embryo collection
- METHOD DETAILS
  - Embryo fixation
  - Embryo sorting
  - Mitotic embryo collection
  - Drosophila virilis embryo collection
  - Micro-C library preparation
  - Micro-C data processing
  - ChIP-seq data processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Chromatin boundary identification
  - Chromatin loop identification
  - Differential analysis of chromatin looping
  - Differential analysis of boundaries
  - Micro-C contact map pileup analysis
  - Construction of the chromatin structuring element atlas
  - Clustering analysis of the chromatin structuring element atlas
  - Genomic annotation analysis
  - Gene ontology term analysis
  - Gene expression analysis
  - Statistical analysis of loop anchor preferences
  - MNase nucleosomal coverage analysis
  - *Drosophila virilis* Micro-C data processing
  - Boundary similarity analysis
  - Motif analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2025.101002.

Received: May 1, 2024
Revised: May 9, 2025
Accepted: August 12, 2025

REFERENCES

1. Mattiroli, F., Bhattacharyya, S., Dyer, P.N., White, A.E., Sandman, K., Burkhart, B.W., Byrne, K.R., Lee, T., Ahn, N.G., Santangelo, T.J., et al.

(2017). Structure of histone-based chromatin in Archaea. Science *357*, 609–612.

2. Acemel, R.D., and Lupiáñez, D.G. (2023). Evolution of 3D chromatin organization at different scales. Curr. Opin. Genet. Dev. *78*, 102019.
3. Zheng, H., and Xie, W. (2019). The role of 3D genome organization in development and cell differentiation. Nat. Rev. Mol. Cell Biol. *20*, 535–550.
4. Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. *2*, 292–301.
5. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M. O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.
6. Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678.
7. Eagen, K.P. (2018). Principles of Chromosome Architecture Revealed by Hi-C. Trends Biochem. Sci. *43*, 469–478.
8. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.
9. Burke, L.J., Zhang, R., Bartkuhn, M., Tiwari, V.K., Tavoosidana, G., Kurukuti, S., Weth, C., Leers, J., Galjart, N., Ohlsson, R., and Renkawitz, R. (2005). CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. EMBO J. *24*, 3291–3300.
10. Eagen, K.P., Aiden, E.L., and Kornberg, R.D. (2017). Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. Proc. Natl. Acad. Sci. USA *114*, 8764–8769.
11. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G.L., Chang, J.-M., and Cavalli, G. (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. Mol. Cell *71*, 73–88.e5.
12. Levo, M., Raimundo, J., Bing, X.Y., Sisco, Z., Batut, P.J., Ryabichko, S., Gregor, T., and Levine, M.S. (2022). Transcriptional coupling of distant regulatory genes in living embryos. Nature *605*, 754–760.
13. Stadler, M.R., Haines, J.E., and Eisen, M.B. (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*, e29550. https://doi.org/10.7554/eLife.29550.
14. Bing, X., Ke, W., Fujioka, M., Kurbidaeva, A., Levitt, S., Levine, M., Schedl, P., and Jaynes, J.B. (2024). Chromosome structure in is determined by boundary pairing not loop extrusion. eLife *13*, RP94070. https://doi.org/10.7554/eLife.94070.
15. Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.
16. Batut, P.J., Bing, X.Y., Sisco, Z., Raimundo, J., Levo, M., and Levine, M. S. (2022). Genome organization controls transcriptional dynamics during development. Science *375*, 566–570.
17. Gisselbrecht, S.S., Palagi, A., Kurland, J.V., Rogers, J.M., Ozadam, H., Zhan, Y., Dekker, J., and Bulyk, M.L. (2020). Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. Mol. Cell *77*, 324–337.e8. https://doi.org/10.1016/j.molcel.2019.10.004.
18. Lupiáñez, D.G., Spielmann, M., and Mundlos, S. (2016). Breaking TADs: How Alterations of Chromatin Domains Result in Disease. Trends Genet. *32*, 225–237.
19. Long, H.K., Osterwalder, M., Welsh, I.C., Hansen, K., Davies, J.O.J., Liu, Y.E., Koska, M., Adams, A.T., Aho, R., Arora, N., et al. (2020). Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. Cell Stem Cell *27*, 765–783.e14.
20. Helmsauer, K., Valieva, M.E., Ali, S., Chamorro González, R., Schöpflin, R., Röefzaad, C., Bei, Y., Dorado Garcia, H., Rodriguez-Fos, E., Puiggròs, M., et al. (2020). Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. Nat. Commun. *11*, 5823.
21. Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA *112*, E6456–E6465.
22. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. *15*, 2038–2049.
23. Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573–1588.e28.
24. Mohan, M., Bartkuhn, M., Herold, M., Philippen, A., Heinl, N., Bardenhagen, I., Leers, J., White, R.A.H., Renkawitz-Pohl, R., Saumweber, H., and Renkawitz, R. (2007). The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. EMBO J. *26*, 4203–4214.
25. Nègre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A.H., et al. (2010). A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet. *6*, e1000814.
26. Schwartz, Y.B., Linder-Basso, D., Kharchenko, P.V., Tolstorukov, M.Y., Kim, M., Li, H.-B., Gorchakov, A.A., Minoda, A., Shanower, G., Alekseyenko, A.A., et al. (2012). Nature and function of insulator protein binding sites in the Drosophila genome. Genome Res. *22*, 2188–2198.
27. Avva, S.V.S.P., and Hart, C.M. (2016). Characterization of the Drosophila BEAF-32A and BEAF-32B Insulator Proteins. PLoS One *11*, e0162906.
28. Kaushal, A., Mohana, G., Dorier, J., Özdemir, I., Omer, A., Cousin, P., Semenova, A., Taschner, M., Dergai, O., Marzetta, F., et al. (2021). CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. Nat. Commun. *12*, 1011.
29. Cavalheiro, G.R., Girardot, C., Viales, R.R., Pollex, T., Cao, T.B.N., Lacour, P., Feng, S., Rabinowitz, A., and Furlong, E.E.M. (2023). CTCF, BEAF-32, and CP190 are not required for the establishment of TADs in early Drosophila embryos but have locus-specific roles. Sci. Adv. *9*, eade1085.
30. Tiwari, V.K., McGarvey, K.M., Licchesi, J.D.F., Ohm, J.E., Herman, J.G., Schübeler, D., and Baylin, S.B. (2008). PcG proteins, DNA methylation, and gene repression by chromatin looping. PLoS Biol. *6*, 2911–2927.
31. Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell *144*, 214–226.
32. Gutierrez-Perez, I., Rowley, M.J., Lyu, X., Valadez-Graham, V., Vallejo, D. M., Ballesta-Illan, E., Lopez-Atalaya, J.P., Kremsky, I., Caparros, E., Corces, V.G., and Dominguez, M. (2019). Ecdysone-Induced 3D Chromatin Reorganization Involves Active Enhancers Bound by Pipsqueak and Polycomb. Cell Rep. *28*, 2715–2727.e5.
33. Rhodes, J.D.P., Feldmann, A., Hernández-Rodríguez, B., Díaz, N., Brown, J.M., Fursova, N.A., Blackledge, N.P., Prathapan, P., Dobrinic, P., Huseyin, M.K., et al. (2020). Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. Cell Rep. *30*, 820–835.e10.
34. Vallot, A., and Tachibana, K. (2020). The emergence of genome architecture and zygotic genome activation. Curr. Opin. Cell Biol. *64*, 50–57.
35. Sun, Q., Perez-Rathke, A., Czajkowsky, D.M., Shao, Z., and Liang, J. (2021). High-resolution single-cell 3D-models of chromatin ensembles during Drosophila embryogenesis. Nat. Commun. *12*, 205.

36. Espinola, S.M., Götz, M., Bellec, M., Messina, O., Fiche, J.-B., Houbron, C., Dejean, M., Reim, I., Cardozo Gizzi, A.M., Lagha, M., and Nollmann, M. (2021). Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. Nat. Genet. *53*, 477–486.

37. Hug, C.B., Grimaldi, A.G., Kruse, K., and Vaquerizas, J.M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell *169*, 216–228.e19.

38. Hiraoka, Y., Dernburg, A.F., Parmelee, S.J., Rykowski, M.C., Agard, D.A., and Sedat, J.W. (1993). The onset of homologous chromosome pairing during Drosophila melanogaster embryogenesis. J. Cell Biol. *120*, 591–600.

39. Flyamer, I.M., Gassler, J., Imakaev, M., Brandão, H.B., Ulianov, S.V., Abdennur, N., Razin, S.V., Mirny, L.A., and Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature *544*, 110–114.

40. Chen, X., Ke, Y., Wu, K., Zhao, H., Sun, Y., Gao, L., Liu, Z., Zhang, J., Tao, W., Hou, Z., et al. (2019). Key role for CTCF in establishing chromatin structure in human embryos. Nature *576*, 306–310. https://doi.org/10.1038/s41586-019-1812-0.

41. Messina, O., Raynal, F., Gurgo, J., Fiche, J.-B., Pancaldi, V., and Nollmann, M. (2023). 3D chromatin interactions involving Drosophila insulators are infrequent but preferential and arise before TADs and transcription. Nat. Commun. *14*, 6678.

42. Ing-Simmons, E., Rigau, M., and Vaquerizas, J.M. (2022). Emerging mechanisms and dynamics of three-dimensional genome organisation at zygotic genome activation. Curr. Opin. Cell Biol. *74*, 37–46.

43. Hsieh, T.-H.S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell *162*, 108–119.

44. Akgol Oksuz, B., Yang, L., Abraham, S., Venev, S.V., Krietenstein, N., Parsi, K.M., Ozadam, H., Oomen, M.E., Nand, A., Mao, H., et al. (2021). Systematic evaluation of chromosome conformation capture assays. Nat. Methods *18*, 1046–1055.

45. Yáñez-Cuna, J.O., Arnold, C.D., Stampfel, G., Boryń, L.M., Gerlach, D., Rath, M., and Stark, A. (2014). Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. Genome Res. *24*, 1147–1156.

46. Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Grüning, B.A., Villaveces, J., Habermann, B., Akhtar, A., and Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat. Commun. *9*, 189.

47. Dorsett, D. (2019). The Many Roles of Cohesin in Drosophila Gene Transcription. Trends Genet. *35*, 542–551.

48. Ing-Simmons, E., Vaid, R., Bing, X.Y., Levine, M., Mannervik, M., and Vaquerizas, J.M. (2021). Independence of chromatin conformation and gene regulation during Drosophila dorsoventral patterning. Nat. Genet. *53*, 487–499.

49. Van Bortle, K., Nichols, M.H., Li, L., Ong, C.-T., Takenaka, N., Qin, Z.S., and Corces, V.G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol. *15*, R82.

50. Gaskill, M.M., Gibson, T.J., Larson, E.D., and Harrison, M.M. (2021). GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. eLife *10*, e66668. https://doi.org/10.7554/eLife.66668.

51. Chetverina, D., Erokhin, M., and Schedl, P. (2021). GAGA factor: a multifunctional pioneering chromatin protein. Cell. Mol. Life Sci. *78*, 4125–4141.

52. 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*, 1519–1526.e4.

53. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233.

54. Vian, L., Pękowska, A., Rao, S.S.P., Kieffer-Kwon, K.-R., Jung, S., Baranello, L., Huang, S.-C., El Khattabi, L., Dose, M., Pruett, N., et al. (2018). The Energetics and Physiological Impact of Cohesin Extrusion. Cell *173*, 1165–1178.e20.

55. Magbanua, J.P., Runneburger, E., Russell, S., and White, R. (2015). A variably occupied CTCF binding site in the ultrabithorax gene in the Drosophila bithorax complex. Mol. Cell Biol. *35*, 318–330.

56. Li, X., and Levine, M. (2024). What are tethering elements? Curr. Opin. Genet. Dev. *84*, 102151.

57. Mahmoudi, T., Katsani, K.R., and Verrijzer, C.P. (2002). GAGA can mediate enhancer function in trans by linking two separate DNA molecules. EMBO J. *21*, 1775–1781.

58. Sabirov, M., Kyrchanova, O., Pokholkova, G.V., Bonchuk, A., Klimenko, N., Belova, E., Zhimulev, I.F., Maksimenko, O., and Georgiev, P. (2021). Mechanism and functional role of the interaction between CP190 and the architectural protein Pita in Drosophila melanogaster. Epigenetics Chromatin *14*, 16.

59. Melnikova, L., Molodina, V., Erokhin, M., Georgiev, P., and Golovnin, A. (2019). HIPP1 stabilizes the interaction between CP190 and Su(Hw) in the Drosophila insulator complex. Sci. Rep. *9*, 19102.

60. Yang, J., Ramos, E., and Corces, V.G. (2012). The BEAF-32 insulator coordinates genome organization and function during the evolution of Drosophila species. Genome Res. *22*, 2199–2207.

61. Chetverina, D., Savitskaya, E., Maksimenko, O., Melnikova, L., Zaytseva, O., Parshikov, A., Galkin, A.V., and Georgiev, P. (2008). Red flag on the white reporter: a versatile insulator abuts the white gene in Drosophila and is omnipresent in mini-white constructs. Nucleic Acids Res. *36*, 929–937.

62. Gates, L.A., Shi, J., Rohira, A.D., Feng, Q., Zhu, B., Bedford, M.T., Sagum, C.A., Jung, S.Y., Qin, J., Tsai, M.-J., et al. (2017). Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. J. Biol. Chem. *292*, 14456–14472.

63. Rowley, M.J., Lyu, X., Rana, V., Ando-Kuri, M., Karns, R., Bosco, G., and Corces, V.G. (2019). Condensin II Counteracts Cohesin and RNA Polymerase II in the Establishment of 3D Chromatin Organization. Cell Rep. *26*, 2890–2903.e3.

64. Bohla, D., Herold, M., Panzer, I., Buxa, M.K., Ali, T., Demmers, J., Krüger, M., Scharfe, M., Jarek, M., Bartkuhn, M., and Renkawitz, R. (2014). A functional insulator screen identifies NURF and dREAM components to be required for enhancer-blocking. PLoS One *9*, e107765.

65. Melnikova, L.S., Molodina, V.V., Kostyuchenko, M.V., Georgiev, P.G., and Golovnin, A.K. (2021). The BEAF-32 Protein Directly Interacts with Z4/putzig and Chriz/Chromator Proteins in Drosophila melanogaster. Dokl. Biochem. Biophys. *498*, 184–189.

66. Narayanan, S., Sato, T., and Wolfe, M.S. (2007). A C-terminal region of signal peptide peptidase defines a functional domain for intramembrane aspartic protease catalysis. J. Biol. Chem. *282*, 20172–20179.

67. Slack, C., Werz, C., Wieser, D., Alic, N., Foley, A., Stocker, H., Withers, D. J., Thornton, J.M., Hafen, E., and Partridge, L. (2010). Regulation of lifespan, metabolism, and stress responses by the Drosophila SH2B protein, Lnk. PLoS Genet. *6*, e1000881.

68. Herman, N., Kadener, S., and Shifman, S. (2022). The chromatin factor ROW cooperates with BEAF-32 in regulating long-range inducible genes. EMBO Rep. *23*, e54720.

69. Blythe, S.A., and Wieschaus, E.F. (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. eLife *5*, e20148. https://doi.org/10.7554/eLife.20148.

70. Duan, J., Rieder, L., Colonnetta, M.M., Huang, A., Mckenney, M., Watters, S., Deshpande, G., Jordan, W., Fawzi, N., and Larschan, E. (2021). CLAMP and Zelda function together to promote Drosophila

zygotic genome activation. eLife *10*, e69937. https://doi.org/10.7554/eLife.69937.

71. Bellec, M., Dufourt, J., Hunt, G., Lenden-Hasse, H., Trullo, A., Zine El Aabidine, A., Lamarque, M., Gaskill, M.M., Faure-Gautron, H., Mannervik, M., et al. (2022). The control of transcriptional memory by stable mitotic bookmarking. Nat. Commun. *13*, 1176.
72. Kuznetsov, D., Tegenfeldt, F., Manni, M., Seppey, M., Berkeley, M., Kriventseva, E.V., and Zdobnov, E.M. (2023). OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. *51*, D445–D451.
73. Hamm, D.C., Bondra, E.R., and Harrison, M.M. (2015). Transcriptional activation is a conserved feature of the early embryonic factor Zelda that requires a cluster of four zinc fingers for DNA binding and a low-complexity activation domain. J. Biol. Chem. *290*, 3508–3518.
74. Petrascheck, M., Escher, D., Mahmoudi, T., Verrijzer, C.P., Schaffner, W., and Barberis, A. (2005). DNA looping induced by a transcriptional enhancer in vivo. Nucleic Acids Res. *33*, 3743–3750.
75. Matharu, N.K., Yadav, S., Kumar, M., and Mishra, R.K. (2021). Role of vertebrate GAGA associated factor (vGAF) in early development of zebrafish. Cells Dev. *166*, 203682.
76. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. *15*, 1034–1050.
77. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121.
78. Krietenstein, N., Abraham, S., Venev, S.V., Abdennur, N., Gibcus, J., Hsieh, T.-H.S., Parsi, K.M., Yang, L., Maehr, R., Mirny, L.A., et al. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. Mol. Cell *78*, 554–565.e7.
79. Hsieh, T.-H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O. J., Tjian, R., and Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. Mol. Cell *78*, 539–553.e8.
80. 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.
81. Kyrchanova, O., Ibragimov, A., Postika, N., Georgiev, P., and Schedl, P. (2023). Boundary bypass activity in the region of the bithorax complex is position dependent and regulated. Open Biol. *13*, 230035.
82. Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358–362.
83. Baldi, S., Jain, D.S., Harpprecht, L., Zabel, A., Scheibe, M., Butter, F., Straub, T., and Becker, P.B. (2018). Genome-wide Rules of Nucleosome Phasing in Drosophila. Mol. Cell *72*, 661–672.e4.
84. Chereji, R.V., Bryson, T.D., and Henikoff, S. (2019). Quantitative MNase-seq accurately maps nucleosome occupancy levels. Genome Biol. *20*, 198.
85. Tullius, T.W., Isaac, R.S., Dubocanin, D., Ranchalis, J., Churchman, L.S., and Stergachis, A.B. (2024). RNA polymerases reshape chromatin architecture and couple transcription on individual fibers. Mol. Cell *84*, 3209–3222.e5.
86. Li, J., and Gilmour, D.S. (2013). Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. EMBO J. *32*, 1829–1841.
87. Dollinger, R., and Gilmour, D.S. (2021). Regulation of Promoter Proximal Pausing of RNA Polymerase II in Metazoans. J. Mol. Biol. *433*, 166897.
88. 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 *2*, e00861.
89. Hiraoka, Y., Agard, D.A., and Sedat, J.W. (1990). Temporal and spatial coordination of chromosome movement, spindle formation, and nuclear envelope breakdown during prometaphase in Drosophila melanogaster embryos. J. Cell Biol. *111*, 2815–2828.
90. Oegema, K., Marshall, W.F., Sedat, J.W., and Alberts, B.M. (1997). Two proteins that cycle asynchronously between centrosomes and nuclear structures: Drosophila CP60 and CP190. J. Cell Sci. *110*, 1573–1583.
91. Deuring, R., Fanti, L., Armstrong, J.A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S.L., Berloco, M., et al. (2000). The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. Mol. Cell *5*, 355–365.
92. Gurudatta, B.V., Yang, J., Van Bortle, K., Donlin-Asp, P.G., and Corces, V.G. (2013). Dynamic changes in the genomic localization of DNA replication-related element binding factor during the cell cycle. Cell Cycle *12*, 1605–1615.
93. Belmont, A.S., Sedat, J.W., and Agard, D.A. (1987). A three-dimensional approach to mitotic chromosome structure: evidence for a complex hierarchical organization. J. Cell Biol. *105*, 77–92. https://doi.org/10.1083/jcb.105.1.77.
94. Keeney, S., and Kleckner, N. (1996). Communication between homologous chromosomes: genetic alterations at a nuclease-hypersensitive site can alter mitotic chromatin structure at that site both in cis and in trans. Genes Cells *1*, 475–489.
95. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.
96. Nigg, E.A. (2001). Mitotic kinases as regulators of cell division and its checkpoints. Nat. Rev. Mol. Cell Biol. *2*, 21–32.
97. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
98. Open2C; Abdennur, N., Fudenberg, G., Flyamer, I.M., Galitsyna, A.A., Goloborodko, A., Imakaev, M., and Venev, S.V. (2024). Pairtools: From sequencing data to chromosome contacts. PLoS Comput. Biol. *20*, e1012164.
99. Lee, S., Bakker, C.R., Vitzthum, C., Alver, B.H., and Park, P.J. (2022). Pairs and Pairix: a file format and a tool for efficient storage and retrieval for Hi-C read pairs. Bioinformatics *38*, 1729–1731.
100. Abdennur, N., and Mirny, L.A. (2020). Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics *36*, 311–316.
101. Open2C; Abdennur, N., Abraham, S., Fudenberg, G., Flyamer, I.M., Galitsyna, A.A., Goloborodko, A., Imakaev, M., Oksuz, B.A., Venev, S.V., and Xiao, Y. (2024). Cooltools: Enabling high-resolution Hi-C analysis in Python. PLoS Comput. Biol. *20*, e1012067.
102. Ursu, O., Boley, N., Taranova, M., Wang, Y.X.R., Yardimci, G.G., Stafford Noble, W., and Kundaje, A. (2018). GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs. Bioinformatics *34*, 2701–2707.
103. Xu, W., Zhong, Q., Lin, D., Zuo, Y., Dai, J., Li, G., and Cao, G. (2021). CoolBox: a flexible toolkit for visual analysis of genomics data. BMC Bioinf. *22*, 489.
104. Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., et al. (2021). Sustainable data analysis with Snakemake. F1000Res. *10*, 33.
105. Krueger, F. , and Simon, H. (2021). TrimGalore. https://github.com/FelixKrueger/TrimGalore. Accessed: July 31, 2025.
106. Broad Institute (2018). Picard Toolkit. http://broadinstitute.github.io/picard/; Accessed: July 31, 2025.
107. Gaspar, J.M. (2018). Improved peak-calling with MACS2. Preprint at bioRxiv. https://doi.org/10.1101/496521.

108. Kruse, K., Hug, C.B., and Vaquerizas, J.M. (2020). FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. Genome Biol. *21*, 303.

109. Roayaei Ardakany, A., Gezer, H.T., Lonardi, S., and Ay, F. (2020). Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. Genome Biol. *21*, 256.

110. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

111. Sande, M. , and Simon, H. (2021). qnorm. https://github.com/Maarten-vd-Sande/qnorm. Accessed: July 31, 2025.

112. Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations. Bioinformatics *37*, 1639–1643.

113. Suarez, H.G., Langer, B.E., Ladde, P., and Hiller, M. (2017). chainCleaner improves genome alignment specificity and sensitivity. Bioinformatics *33*, 1596–1603.

114. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinf. *10*, 421.

115. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

116. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

117. dos Santos, G., Schroeder, A.J., Goodman, J.L., Strelets, V.B., Crosby, M.A., Thurmond, J., Emmert, D.B., and Gelbart, W.M.; FlyBase Consortium (2015). FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res. *43*, D690–D697.

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

119. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci. Rep. *9*, 9354.

120. Larkin, A., Marygold, S.J., Antonazzo, G., Attrill, H., Dos Santos, G., Garapati, P.V., Goodman, J.L., Gramates, L.S., Millburn, G., Strelets, V.B., et al. (2021). FlyBase: updates to the Drosophila melanogaster knowledge base. Nucleic Acids Res. *49*, D899–D907.

121. Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074–1077.

122. Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature *512*, 91–95.

123. Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C., and Lempicki, R.A. (2003). Identifying biological themes within lists of genes with EASE. Genome Biol. *4*, R70.

124. Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800.

125. Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473–479.

126. Hemmer, L.W., Dias, G.B., Smith, B., Van Vaerenberghe, K., Howard, A., Bergman, C.M., and Blumenstiel, J.P. (2020). Hybrid dysgenesis in Drosophila virilis results in clusters of mitotic recombination and loss-of-heterozygosity but leaves meiotic recombination unaltered. Mob. DNA *11*, 10.

127. Flynn, J.M., Long, M., Wing, R.A., and Clark, A.G. (2020). Evolutionary Dynamics of Abundant 7-bp Satellites in the Genome of Drosophila virilis. Mol. Biol. Evol. *37*, 1362–1375.

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-phospho-Histone H3 (Ser10) antibody	Cell Signaling Technology	Cat #9701; RRID: AB_331535
Anti-mouse Alexa 488-conjugated secondary antibody	Life Technologies	A21202; RRID: AB_141607
Chemicals, peptides, and recombinant proteins		
DSG (disuccinimidyl glutarate)	Thermo Fisher Scientific	Cat #20593
EGS (ethylene glycol bis(succinimidyl succinate))	Thermo Fisher Scientific	Cat #21565
Deposited data		
Micro-C for D. melanogaster (nc1-8, nc12m, nc14, and s10-12 sample); Micro-C for D. Virilis	This manuscript	GEO: GSE265818
Mitotic GAF ChIP-seq	Bellec et al.[71]	GEO: GSE180812
Embryonic Hi-C	Ogiyama et al.[11]	GEO: GSE103625
Embryonic Hi-C	Hug et al.[37]	ArrayExpress: E-MTAB-4918
ChIP-seq compendium for CSE Clustering	Compiled in this manuscript	Table S3; Zenodo: 16621265
Experimental models: Organisms/strains		
D. virilis stock	National Drosophila Species Stock Center	SKU: 15010-1051.52
Software and algorithms		
BWA	Li and Durbin[97]	https://github.com/lh3/bwa
pairtools	Open2C et al.[98]	https://pairtools.readthedocs.io/en/latest/
pairix	Lee et al.[99]	https://github.com/4dn-dcic/pairix
cooler	Abdennur et al.[100]	https://cooler.readthedocs.io/en/latest/datamodel.html
Cooltools	Open2C et al.[101]	https://cooltools.readthedocs.io/en/latest/index.html
GenomeDISCO	Ursu et al.[102]	https://github.com/kundajelab/genomedisco
Coolbox	Xu et al.[103]	https://github.com/GangCaoLab/CoolBox
Snakemake	Mölder et al.[104]	https://snakemake.readthedocs.io/en/stable/
TrimGalore	https://github.com/FelixKrueger/TrimGalore[105]	https://github.com/FelixKrueger/TrimGalore
PicardTools	https://broadinstitute.github.io/picard/[106]	https://broadinstitute.github.io/picard/
MACS2	Gaspar[107]	https://github.com/macs3-project/MACS
FAN-C	Kruse et al.[108]	https://github.com/vaquerizaslab/fanc
Mustache	Ardakany et al.[109]	https://github.com/ay-lab/mustache
DESeq2	Love et al.[110]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Quantile Normalization (qnorm)	van der Sande[111]	https://github.com/Maarten-vd-Sande/qnorm
LiftOff	Shumate et al.[112]	https://github.com/agshumate/Liftoff
Make Lastz Chains	Suarez et al.[113]	https://github.com/hillerlab/make_lastz_chains?tab=readme-ov-file
BLASTN	NCBI[114]	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=GeoBlast&PAGE_TYPE=BlastSearch
FIMO (MEME-Suite)	Grant et al.[115]	https://meme-suite.org/meme/doc/fimo.html
Leiden	Traag et al.[53]	https://leidenalg.readthedocs.io/en/stable/intro.html

(*Continued on next page*)

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Scipy	Virtanen et al.[116]	https://scipy.org/
Code to reproduce analysis	This manuscript	GitHub (https://github.com/pritykinlab/drosophila-microc-zga); Zenodo (https://doi.org/10.5281/zenodo.16621265).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Fly husbandry

Stocks were maintained in a 25°C incubator with 12/12 light/dark cycle on standard cornmeal medium. All staged embryos from this study were collected from the wild-type y1w67c23 control line. The *Drosophila virilis* stock was obtained from the National Drosophila Species Stock Center (SKU: 15010-1051.52, originally from the U.S.S.R).

Embryo collection

Embryos were collected on yeasted apple juice plates in population cages. For nc1-8 embryos, plates were changed every 30 min, and subjected to immediate fixation (see below). For nc14 embryos, plates were laid for 1 h, then collected, incubated for 2 h at 25°C, and subjected to fixation. For stage 10–12 embryos, plates were laid for 2 h, incubated for 6 h at 25°C, and subjected to fixation.

METHOD DETAILS

Embryo fixation

Embryos were collected in nylon mesh sieves, dechorionated for 2 min in 3% sodium hypochlorite, rinsed with deionized water, and transferred to glass vials containing 5 mL PBST (0.1% Triton X- in PBS), 7.5 mL N-heptane, and 1.5 mL fresh 16% formaldehyde. Crosslinking was carried out at room temperature for exactly 15 min on an orbital shaker at 250 rpm, followed by addition of 3.7 mL 2M Tris-HCl pH7.5 and shaking for 5 min to quench the reaction. Embryos were washed twice with 15 mL PBST and stored at 4°C in PBST while additional rounds of embryo collections were done. Following initial fixation with formaldehyde, fixed embryos were combined and subject to secondary crosslinking.

For the secondary fixation, all collections from the same day were combined, and crosslinked in 10mL of freshly prepared 3mM final DSG and EGS in PBST for 45 min at room temperature with passive mixing. The reaction was quenched by addition of 3.7mL of 2M Tris-HCl pH7.5 for 5 min, washed twice with PBST, and contents were transferred to a clear dish for sorting.

Embryo sorting

Embryos were sorted by mouth pipetting visually confirmed embryos of the correct stage under a dissection microscope into a 1.5mL Eppendorf tube. In particular, nc1-8 embryos were collected by positive selection twice, using the visual marker of no nuclei having migrated to the cortex. For each biological replicate in a given stage, 500 embryos of nc1-8 were collected, 300 embryos for nc14, and 100 embryos for s10-12. Embryos were briefly spun down, and PBST was removed. Finally, samples were snap frozen with liquid nitrogen and stored at −80°C until Micro-C library preparation.

Mitotic embryo collection

For mitotic embryos, plates were laid for 1 h, collected, and incubated for 1 h at 25°C. Embryos were fixed in 1% formaldehyde in PBS, *not PBST*, for 15 min at room temperature shaking at 250 rpm. The lower layer was removed to stop the reaction, and then ice-cold methanol was added in 1:1 ratio to the remaining solution containing embryos. The solution was vortexed for 2 min, and the embryos were allowed to sink to the bottom. The remaining solution was removed and embryos were washed two times in methanol. Embryos were then resuspended in PBST and allowed to rehydrate for at least an hour, and stored at 4°C until all embryos were collected. Embryos were then stained with anti-phospho-Histone H3 (Ser10) antibody (Cell Signaling #9701) at a dilution 1:200 overnight at 4°C. Anti-mouse Alexa 488-conjugated (Life technologies, A21202) was used as a secondary antibody at a dilution 1:500, stained for 2 h at room temperature. All stainings were conducted in PBST with 3% BSA, and embryos were washed three times for 5 min in between stains. Finally, embryos were isolated by mouth pipetting for positively stained embryos under a fluorescent stereoscope. 300 embryos were collected for each biological replicate. Crosslinking was performed as normal, and samples were snap frozen and stored at −80°C until Micro-C library preparation.

Drosophila virilis embryo collection

Drosophila virilis embryos were prepared in the same manner as *D. melanogaster* embryos, but with the following differences. *D. virilis* flies were reluctant to lay eggs on fresh yeasted apple juice plates, so we extended laying time to an overnight collection of approximately 12 h. Aside from removing larvae, we made no attempt at sorting the *D. virilis* embryos.

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

Micro-C library preparation

Micro-C libraries were prepared as previously described.[16] Libraries were barcoded, pooled, and subjected to paired-end sequencing on an Illumina Novaseq S1 100 nt Flowcell (read length 50 bases per mate, 6-base index read).

Micro-C data processing

Micro-C data for *D. melanogaster* were aligned to the Berkeley Drosophila Genome Project (BDGP) Release 6 reference assembly[117] with BWA-MEM[97] using parameters -S -P −5 -M. The resultant BAM files were parsed, sorted, de-duplicated, filtered, and split with pairtools (https://github.com/mirnylab/pairtools). We removed pairs where only half of the pair could be mapped, or where the MAPQ score was less than three. The resultant files were indexed with pairix (https://github.com/4dn-dcic/pairix). Using Cooler,[100] we generated contact matrices for individual replicates (Table S1) at 10kb resolution, and analyzed the similarity of these contact matrices with GenomeDISCO.[102] The pairs files for individual replicates were merged with pairtools, and then used to generate contact matrices at 100bp resolution with Cooler. Finally, balancing and mcool file generation was performed using Cooler's zoomify tool. All processing pipelines were automated with Snakemake.[104] Visualizations of Micro-C maps were produced with CoolBox.[103] *Cis*-interacting read coverage was calculated using cooltools[101] at 100bp resolution.

ChIP-seq data processing

All ChIP-seq data (Table S3) were downloaded as FASTQ files from the Sequence Read Archive (SRA). After downloading the FASTQ files, we then filtered and clipped the reads using TrimGalore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).[105] Trimmed reads were aligned to the BDGP Release 6 reference assembly using BWA-MEM[97] with its default parameters. The resultant BAM files were sorted and indexed using SAMtools,[118] and filtered to remove unaligned reads, non-primary alignments, and supplementary alignments. Duplicate reads were removed from the BAM files using Picard Tools (https://broadinstitute.github.io/picard/).[106] We merged de-duplicated BAMs from individual experiments with matching genotypes and experimental conditions. For each of the merged BAM files, we calculated the genome-wide enrichment over matching control samples using MACS2.[107] Peaks were called by MACS2 using an effective genome size of 1.2×10^8, the nomodel option, zero shift, a *p*p-value cutoff of 1×10^{-2}, and no additional duplicate read filtering. Fragment length for peak calling was identified with MACS2's predictd command using a fold-enrichment range of 3–50.

QUANTIFICATION AND STATISTICAL ANALYSIS

Chromatin boundary identification

We calculated genome-wide normalized insulation scores and called insulatory chromatin boundaries using FAN-C.[108] Boundaries were called using a 400 bp bin size, a window size of 16, and a delta value of three. The resultant boundaries were filtered to retain only those with a boundary score of at least one. Note that even if a boundary did not meet our boundary strength threshold in all three stages, we still quantified its strength in all stages so long as it met the strength threshold in at least one stage. We removed boundaries that overlapped ENCODE V2 blacklist sites.[119] To enable cross-stage comparisons of boundaries, we also removed boundaries that fell within 1.2kb of an empty bin in any of the biological stages so as to ensure fair comparisons. If boundary strength could not be calculated in a stage due to missing data (n = 145), then its strength for that stage was set to NaN and not included in calculations.

Chromatin loop identification

We used Mustache[109] to perform genome-wide loop-calling for *D. melanogaster* embryos in the three primary stages (Table S2) at 800 bp resolution. Micro-C reads were downsampled using cooltools[101] to avoid depth-dependent differences in loop detection sensitivity. We further merged all biological replicates into one high-depth sample, and called loops on this merged dataset to produce a loop atlas that could be quantified across all three stages. Loops were called using Mustache with the following parameters: r = 800, p = 0.01, st = 0.85, and sz = 1. Adjacent loops were combined to avoid over-counting individual looping events. These collapsed loops were shifted to the pixel with the highest ratio of observed and expected counts in the merged sample. Finally, loop anchors were extended by 800 bp on either side.

Differential analysis of chromatin looping

Using the loop atlas described above, we aggregated read counts for each loop for each replicate (n = 2) in all conditions. Next, we used DESeq2[110] to identify loops that varied significantly in intensity across stages. Differences in the distance-decay curve between conditions meant that applying a single size factor to normalize read counts would be inappropriate and could lead to a distance bias across conditions. Instead, for each loop, we calculated the expected Hi-C contact frequency for all contacts of the same distance for each replicate. We then passed these in as normalization factors to DESeq2. Therefore, each loop is normalized according to the *expected contact* for all other contacts of a similar distance for that replicate. We then ran DESeq2 as normal. DESeq2 adjusts the read count for each loop in each condition for these normalization factors, correcting for both the sample-specific distance-decay and sequencing depth.

Differential analysis of boundaries

Differential boundaries were calculated by calculating the Log_2(fold change) in boundary strength between different conditions. Boundaries with an LFC >0.4 were considered higher in one condition than another. This calculation was performed for all pairs of conditions.

Micro-C contact map pileup analysis

To visualize trends in the contact frequency matrix near structures of interest (e.g., boundaries from a given cluster, loop foci), we plotted the mean contact frequency of 20 kb × 20 kb submatrices centered on each structure. Pixels were normalized based on the expected read count for their distance from the genome-wide diagonal. To enable visual comparisons between the pileups and prevent individual pileups from being dominated by only a few high-signal sites, we divided each sub-matrix by its mean intensity before taking the pixel-wise mean of the submatrices.

Construction of the chromatin structuring element atlas

Although loops and boundaries were identified in different manners, we identified significant overlap between the two classes early in the study. To allow comparisons both between and across these two classes of CSEs, we produced a unified atlas of CSEs. We first generated a minimal set of boundaries by merging adjacent or overlapping boundaries in the original boundary atlas. Merged boundaries were clipped down to their center 400 bp. We also generated a minimal set of loop anchors by merging adjacent or overlapping loop anchors found in the original loop atlas. Merged loop anchors were then clipped down to their center 800 bp. Loop anchors were then assigned to boundaries within 2400 bp. If there were multiple boundaries within 2400 bp of a loop anchor, we assigned the loop anchor to the closest boundary. For the three loop anchors with more than one equidistant boundary within this range, we assigned the boundary at random. Finally, all 4351 merged CSEs in the final atlas were clipped down to 800 bp intervals.

For a given CSE, the stage-specific boundary strength was calculated using the strongest stage-specific boundary score within 1 kb of the CSE's center. To mitigate subtle shifts (e.g., <1kb) in boundary positions over time and boundaries that show large differences in strength across stages, the strongest stage-specific boundary score for each CSE was identified using a set of boundaries that had not been filtered based on their boundary strength (i.e., but that had been filtered based on ENCODE v2 Blacklist sites and coverage). Thus, while the CSE positions must either be a loop anchor or a boundary that meets our strength thresholds in at least one stage, the stage-specific presence and boundary strength of those CSEs can be more subtle or even below this threshold. Indeed, 910 of our 1588 loop anchors in the CSE atlas overlapped a boundary from our original boundary atlas, and 488 of the remaining 678 loop anchors in the CSE atlas overlapped a subthreshold boundary.

Clustering analysis of the chromatin structuring element atlas

To identify sub-classes of CSEs associated with distinct regulatory signals, we clustered our atlas of CSEs using existing ChIP-seq data from *D. melanogaster* embryos and cell lines. First, we extracted the mean signal of each ChIP-seq experiment at each CSE location. In specific, the ChIP-seq signal at each CSE was measured with a normalized enrichment value X_{NORM},

$$X_{NORM} = log_2\left(\frac{X}{\mu_X} + 1\right)$$

where X is the fold enrichment of the ChIP-seq experiment relative to the control experiment at a given location, and μ_X is the genome-wide average of said fold enrichment. Prior to clustering, we also standardized these values and performed principal components analysis (PCA). Clustering was then performed on the first 67 principal components, which cumulatively explained at least 95% of the variance in the data. Clusters of chromatin structures were identified via Leiden clustering[53] of the Euclidean K-nearest neighbors graph (K = 1500).

Genomic annotation analysis

Genome annotations for D. melanogaster were downloaded from FlyBase (r6.39).[120] Enhancer annotations were sourced from previously published STARR-seq experiments[121] and the VT enhancer catalog.[122] We identified all annotations that fell within a 2.8kb region centered at the middle of the CSE (i.e., less than 1kb from a given 800bp CSE). The gene list for localized patterning was downloaded from Levo et al. 2022.[12] Housekeeping genes were annotated as described in Ing-Simmons et al. 2021.[48] We tested for associations between clusters of CSEs and specific annotations with a two-sided Bonferroni-corrected Fisher's exact test. That is, we tested if there is an association between whether a CSE belongs to a specific cluster, and whether it falls within 1kb of a given annotation. Specifically, we ran Fisher's exact test on the following contingency table for boundaries of Cluster "X".

Cluster X boundaries in annotation	Non-cluster X boundaries in annotation
Cluster X boundaries not in annotation	Non-cluster X boundaries not in annotation

Gene ontology term analysis

We downloaded gene ontology (GO) term annotations for *D. melanogaster* genes (FlyBase r6.39) from FlyBase (FB2021_03 release). For a given cluster of CSEs, genes were considered "proximal" genes if they fell within a 10.8kb region centered around each CSE in said cluster. All other genes were considered "distal" for said cluster. To identify significant associations between CSE clusters and the sets of proximal genes annotated with a given GO term, we used the one-sided Fisher's exact test with the EASE score modification.[123] A test result was considered significant if its Bonferroni-corrected *p*p-value was less than 1×10^{-4}. Finally, we used Revigo[124] to compress the set of significant GO terms for each cluster into a set of non-redundant GO terms.

Gene expression analysis

We downloaded the ModENCODE[125] developmental RNA-seq dataset from the FlyBase precomputed file database (FB2021_03 release). We imputed missing values with the mean expression level in the corresponding experiment. Since there were M = 128 total experiments, imputation was done only for 27 genes out of all 13,952 genes, and only 3 out of 2322 genes overlapping CSEs. To normalize expression data from different experiments, we used quantile normalization (https://github.com/Maarten-vd-Sande/qnorm)[111] and added a pseudocount of one. For all subsequent analysis, we used the $\log_{10}$-transformed ratio of gene expression to sample-wise mean expression value. To match genetic *cis*-regulatory logic to clusters of CSEs, we identified the set of genes that overlapped or nearly overlapped individual CSEs in that cluster. A gene was considered to be flanking a CSE if it fell within a 2.8kb region centered on the middle of the CSE. In this analysis, for a CSE cluster, we averaged the expression of all genes associated with these CSEs. If a gene was associated with CSEs from different clusters, we considered the expression of this gene for each of these clusters. If a gene overlapped multiple CSEs from the same cluster, we only included it once when calculating gene expression for that cluster.

Statistical analysis of loop anchor preferences

To determine loop anchor pairing preferences between clusters, we performed a Bonferroni-corrected two-sided Fisher's exact test. In specific, to test for an association between loop anchors in clusters *I* and *J* we have a typical contingency table.

|A|	|B|
|C|	|D|W

Where:

A = {observed loops with one loop anchor in *I* and one in *J*}.

B = {observed loops that are not between elements of *I* and *J*}.

C = {unobserved loops between elements of *I* and *J*}.

D = {unobserved loops that are not between elements of *I* and *J*}.

Note that the number of unobserved loops is calculated based on the number of total possible cis loops, which we calculate on a chromosome-by-chromosome basis (i.e., $\frac{N(N-1)}{2}$ where *N* is the number of loop anchors on a single chromosome).

MNase nucleosomal coverage analysis

One-dimensional MNase signal was calculated from Micro-C data as described previously[78] with minor modifications. First, all read pairs from the (+, +) orientations were considered. Then, all read pairs with a genomic distance <120bp were discarded since they represented nucleosome-free regions. Then, all reads were shifted by 73 base pairs to account for nucleosomal overlap positioning. Next, the MNase coverage of each genomic position was calculated as the number of reads whose 5′ position aligned to that genomic position. MNase coverage was then smoothed with a 1D Gaussian filter with sigma = 40. For 2D MNase heatmaps, plots were smoothed with a 2D Gaussian filter using sigma = 8 to improve ease of visualization.

Drosophila virilis Micro-C data processing

Micro-C data for *D. virilis* were aligned to the ASM798932v2 reference assembly[126] as described above. We called boundaries in the *D. virilis* data using the same approach we used for *D. melanogaster*. We retained *D. virilis* boundaries with a strength greater than one. Gene annotations for the *D. virilis* genome were mapped to ASM798932v2 with Liftoff 1.6.3.[112] We utilized a previously submitted version of the *Drosophila virilis* genome DvirRS2[127] and its annotations for that reference. Liftoff was run using the default parameters. The chain file mapping between loci in BDGP6 and ASM798932v2 was produced using whole-genome pairwise alignments (https://github.com/hillerlab/make_lastz_chains).[113]

Please cite this article in press as: Dolsten et al., 3D chromatin structures precede genome activation in *Drosophila* embryogenesis, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101002

Boundary similarity analysis

We checked for sequence similarity of boundaries by using BLASTN[114] with default parameters to search for aligning sequences in the genome of *D. virilis*. Sequences that did not return hits were labeled as "not conserved". If BLASTN identified hits for a given boundary in *D. virilis*, we checked if any of these hits fell within 1kb of a boundary in the *D. virilis* genome to determine if both the sequence and structure of the boundary were conserved in *D. virilis*. To test whether different clusters of boundaries were enriched or depleted for different similarity statuses, we used the Bonferroni-corrected two-sided Fisher's exact test. Specifically, we tested for whether there was an association between boundaries in a single cluster and between a single annotation status (i.e., not conserved, only sequence conserved, and both sequence and function conserved).

Motif analysis

Motifs were called jointly on all boundaries and loop anchors using the FIMO tool[115] from MEME-suite with default parameters and the JASPAR 2022 CORE insect database (https://jaspar.genereg.net/downloads/). Conservation scores for motifs were calculated by averaging PhyloP27 scores[77] for all motif instances overlapping ATAC-seq peaks in CSEs.

Cell Genomics, Volume *5*

Supplemental information

3D chromatin structures precede genome activation in *Drosophila* embryogenesis

Gabriel A. Dolsten, Evan M. Cofer, Xin Yang Bing, Benjamin Brack, Marcus Curlin, Chandra L. Theesfeld, Olga G. Troyanskaya, Michael S. Levine, and Yuri Pritykin

Supplementary Figures

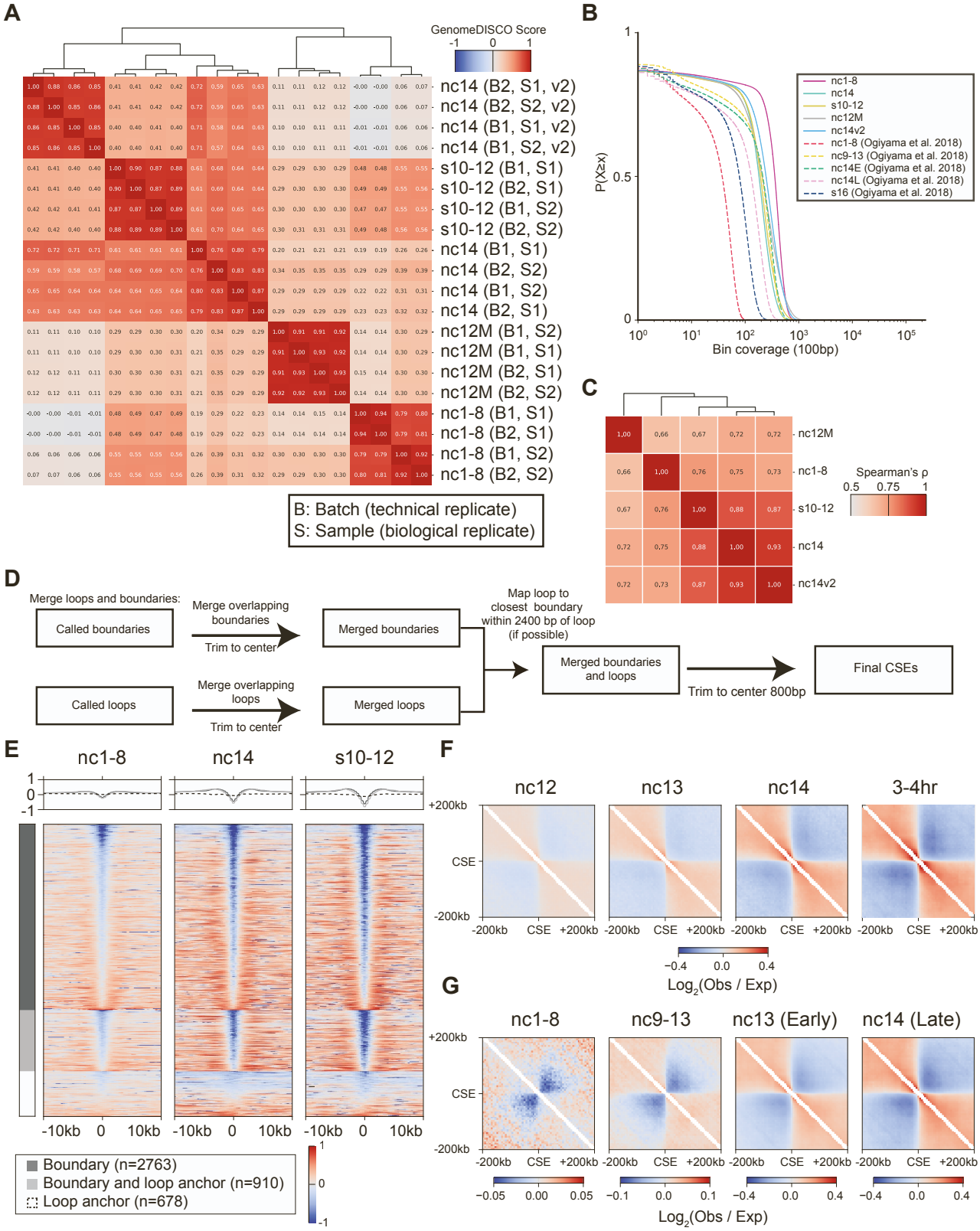


Figure S1: Summary of sequencing data and boundaries, related to Figure 1. (A) Replicate-wise GenomeDISCO similarity scores. Rows and columns are hierarchically clustered with complete linkage. **(B)** Empirical cumulative distribution functions (ECDF) of *cis*-interacting read coverage for all 100 bp bins. For this analysis, the genome was split into consecutive 100 bp bins. Bin coverage corresponds to the number of reads aligning to each bin, and the Y axis is the fraction of bins with at least that many reads. For comparison, we show ECDFs for the previously published Hi-C datasets (Ogiyama et al. 2018[1]). **(C)** Matrix showing Spearman's rank correlation in insulation scores from each sample. Insulation score is calculated as the depletion of contacts spanning a locus (Methods). Correlation was calculated based on 10M random locations in the genome. Rows and columns are hierarchically clustered with complete linkage. **(D)** Schematic of CSE merging protocol. **(E)** Profile heatmaps of insulation scores at all CSEs. Insulation scores are shown for nc1-8 (left), nc14 (middle), and s10-12 (right). For each class of CSE, rows are sorted by average signal in the center 3.2 kb. Boundaries that overlapped loop anchors were stronger than the rest (Bonferroni-corrected one-sided Mann-Whitney U test, $P_{nc1\text{-}8}=6\times10^{-4}$, $P_{nc14}=4\times10^{-9}$, $P_{s10\text{-}12}=1\times10^{-16}$, n=3673). **(F)** Contact map pileups for Micro-C boundary locations identified in this study, but using the previously published Hi-C contact frequency data (Hug et al. 2017[2]). Separate pileups are shown for nc12, nc13, nc14, and 3-4hr. The bin width is 5 kb. **(G)** Contact map pileups for our boundary locations, but using the previously published Hi-C contact frequency data (Ogiyama et al. 2018[1]). Separate pileups are shown for nc1-8, nc9-13, early nc14, and late nc14. The bin width is 3.2 kb.

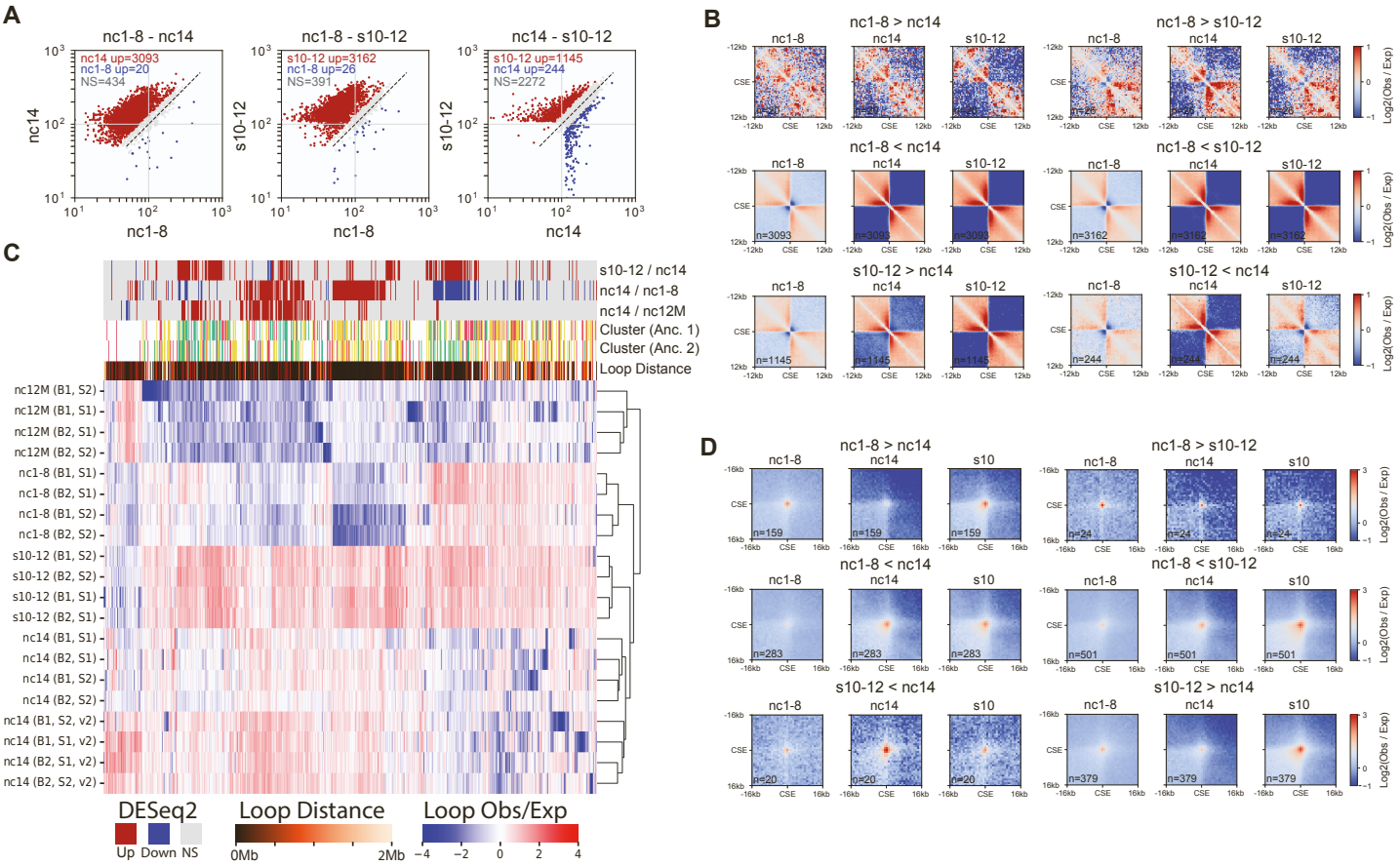


Figure S2: Identification and pileups of differential boundaries and loops, related to Figure 1. (A) Differential boundaries between conditions identified through changes in boundary score (LFC boundary score > .4 or < -.4; boundary score scaled by 1000). **(B)** Pileup of differential boundaries between all pairs of conditions. **(C)** Correlation of O/E scores at loop anchors in all sample replicates. O/E scores were Z-normalized by column. **(D)** Pileup of differential loops called between all pairs of conditions. *n*, the number of differential boundaries or loops.

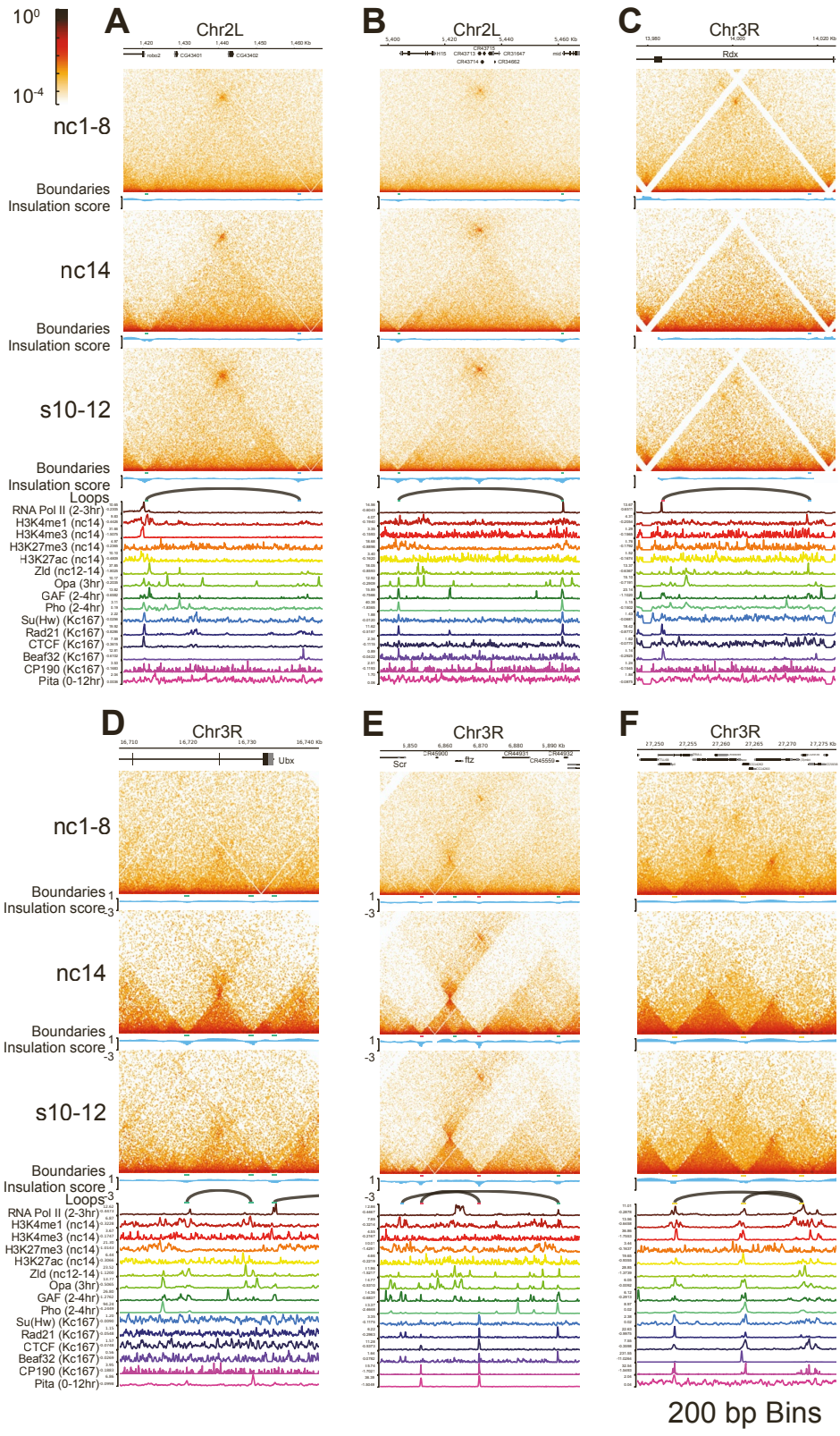


Figure S3: Plots of representative loops based on differential testing results, related to Figure 1. (A) A loop that increases in strength from nc14 to s10-12. **(B)** A loop that increases in strength from nc1-8 to s10-12. **(C)** A

loop that decreases in strength from nc1-8 to s10-12. **(D)** A loop that increases in strength from nc1-8 to nc14. **(E)** A loop that increases in strength from nc1-8 to nc14. **(F)** A loop that decreases in strength from nc1-8 to nc14.

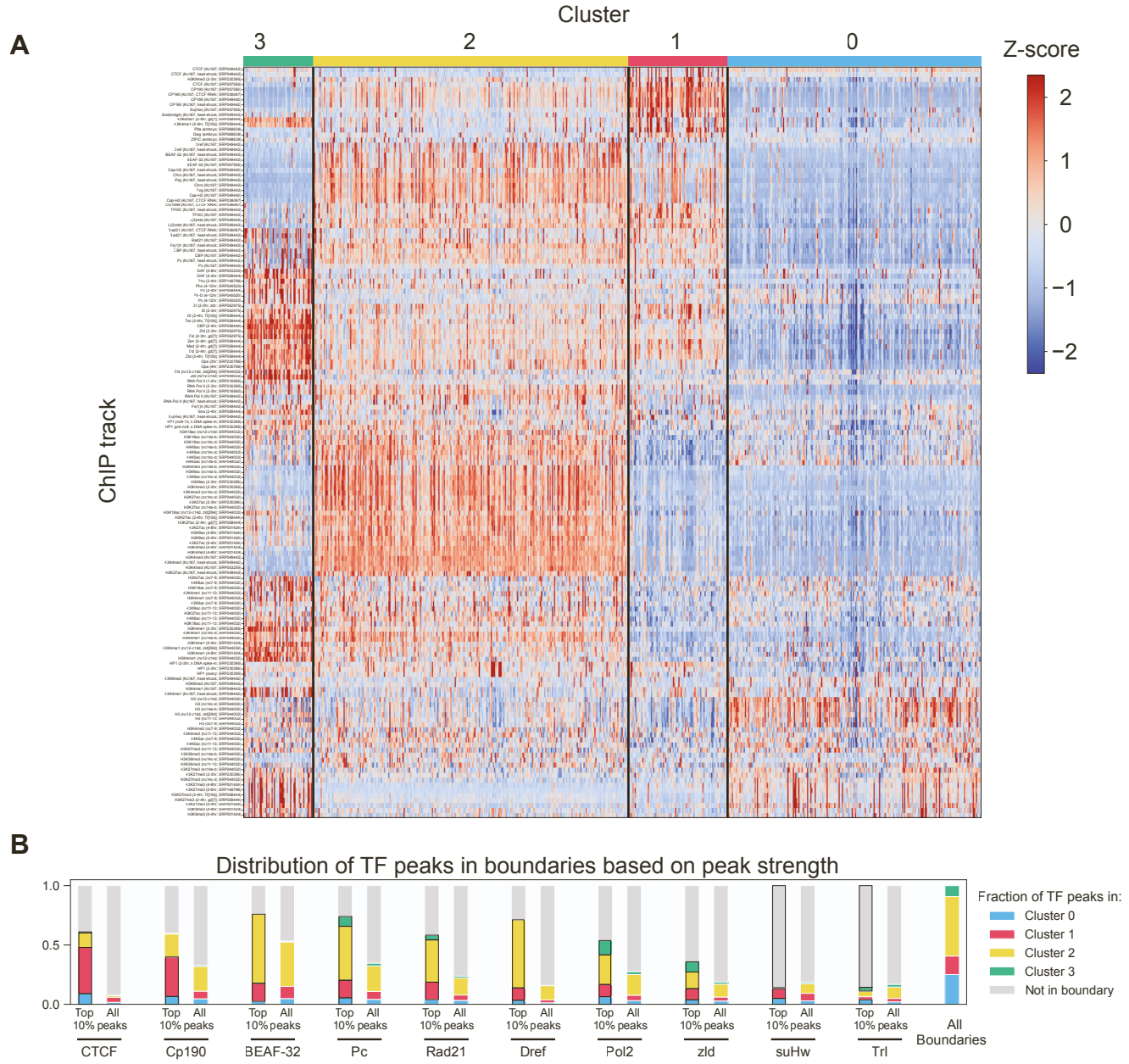


Figure S4: Characterization of ChIP-seq data at CSEs, related to Figure 2. (A) Heatmap of ChIP-seq data at CSEs for all 149 ChIP-seq tracks, grouped by CSE cluster. **(B)** Fraction of either all TF peaks or the strongest 10% of TF peaks overlapping boundaries.

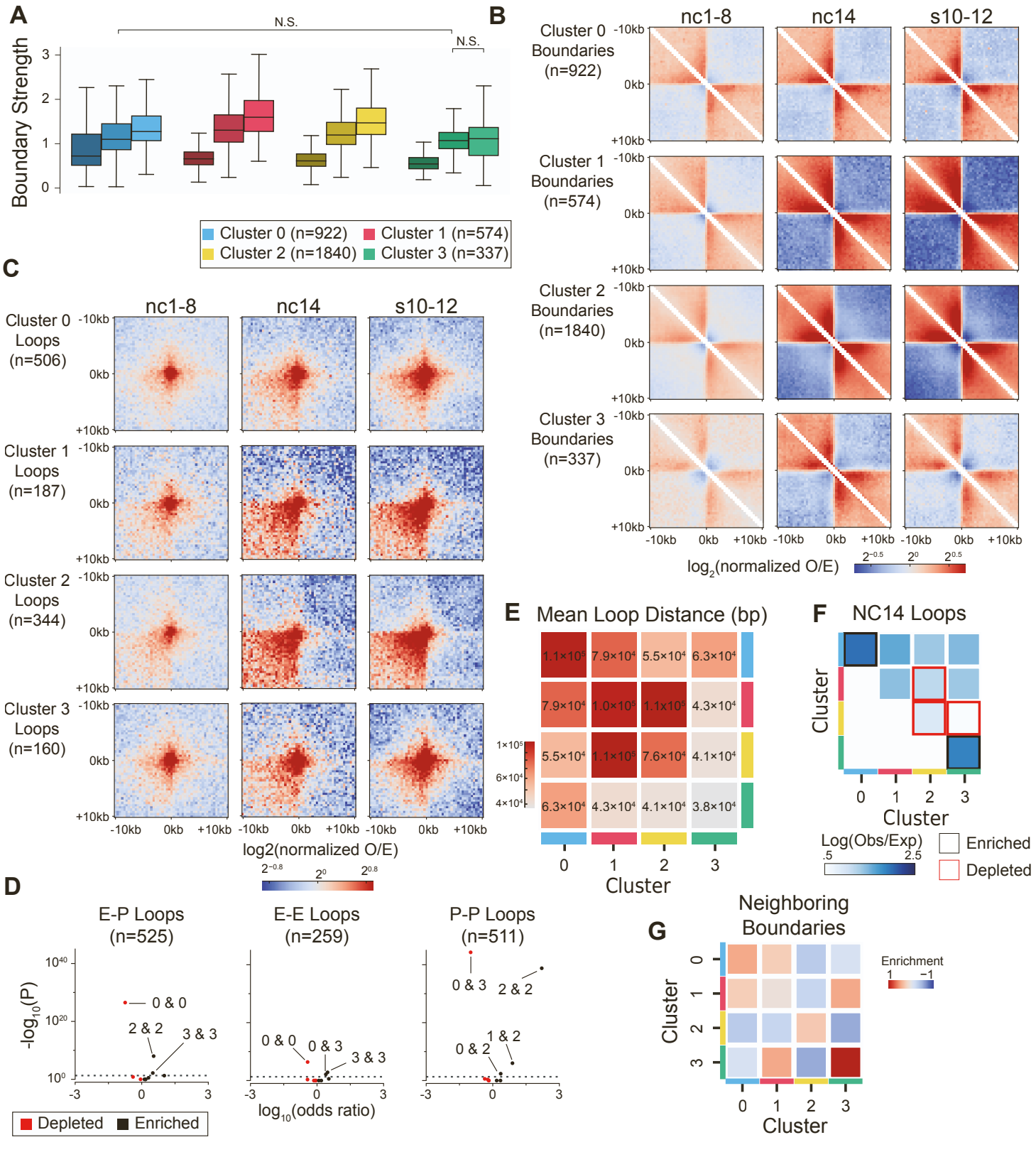


Figure S5: Cluster-specific differences in chromatin structures, related to Figure 2. (A) Boxplots summarizing boundary strength according to cluster. Bonferroni-corrected two-sided Mann-Whitney U test was

used to detect significant differences between clusters. Bonferroni-corrected two-sided Wilcoxon signed-rank test was used to detect significant changes over time within clusters. All comparisons were significant ($P<0.05$) unless marked N.S. above the boxplots. **(B)** Contact map pileups for boundaries, broken down by cluster. Contact maps are shown for nc1-8 (left column), nc14 (middle column), and s10-12 (right column). Note the slight light band in the upper diagonal of Cluster 2 boundaries. **(C)** Contact map pileups for loop foci, broken down by cluster. Contact maps are shown for nc1-8 (left column), nc14 (middle column), and s10-12 (right column). **(D)** Volcano plots showing enrichment for specific types of contacts for different cluster-to-cluster loop anchor pairings (e.g. loops between cluster 3 and itself are enriched for enhancer-promoter loops). Pairings that are depleted for a certain kind of contact are plotted in red. Pairings that are enriched are plotted in black. Note that some enhancers fall near or within promoters, and so some loops may be counted more than once among the three categories of contacts. **(E)** Heatmap of the mean genomic distance between loops in base pairs. **(F)** Hi-C signal (observed divided by expected) at loops between all pairs of clusters. Significance is demarcated by outlines (FDR < .05; Wilcoxon rank-sum test with Benjamini-Hochberg correction). **(G)** Heatmap plotting the cluster identities for pairs of adjacent boundaries. Enrichment of boundary pairs such as (3, 3) was observed, suggesting that Cluster 3 boundaries tend to co-occur in genomic coordinates. Enrichment was calculated relative to the null distribution consisting of random, rather than adjacent, boundary pairs. All comparisons were significant (FDR < .05; Permutation test with Benjamini-Hochberg correction).

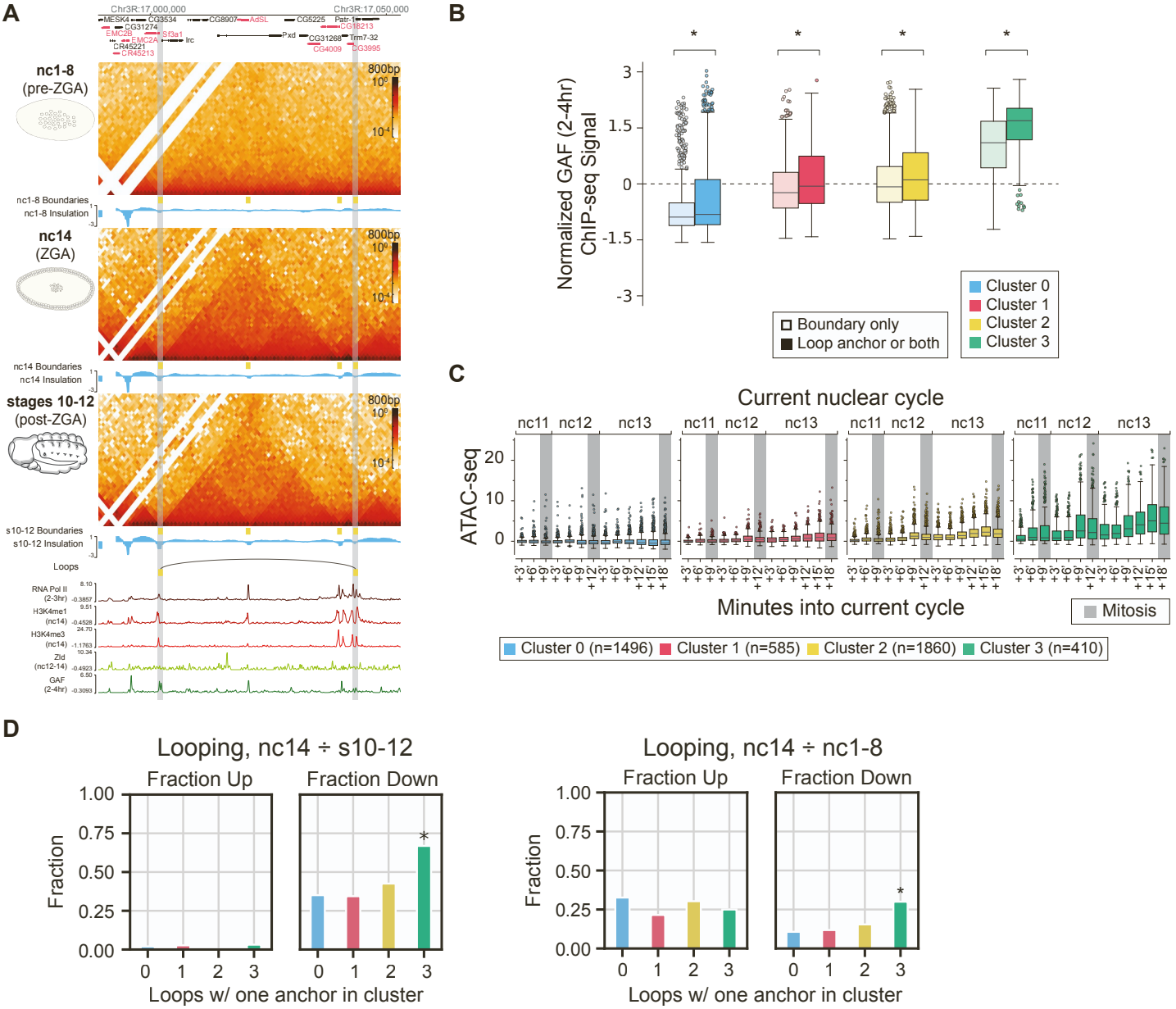


Figure S6: Overview of GAF and chromatin accessibility, related to Figure 4. (A) Genome browser view of the Pxd locus. Micro-C contact frequency maps from the nc1-8 (top), nc14 (middle), and s10-12 (bottom) are shown alongside ChIP-seq data. Note the number of genes linked to housekeeping functions such as transcription, splicing, and core metabolism (e.g. Sf3a1, Irc, AdSL, Trm7-32, Patr-1). Loop anchors are highlighted with gray for visibility. **(B)** Boxplots showing the z-score normalized ChIP-seq signal for GAF (2-4hr). For every cluster, loop anchors and loop anchors that overlapped boundaries exhibited elevated levels of GAF compared to boundaries that did not also function as loop anchors. (*) $P < 0.05$, Bonferroni-corrected two-sided Mann-Whitney U test). **(C)** Boxplots summarizing the chromatin accessibility at CSEs in each cluster. Accessibility for a given CSEs was quantified as the level of ATAC-seq signal in a 1 kb region centered around that CSE. ATAC-seq data

were sourced from (Blythe and Wieschaus 2016). Accessibility was recorded in 3 minute intervals from the beginning of nc11 to the end of nc13. Dark gray vertical bands indicate samples taken during mitosis. **(D)** CSE clusters for loop anchors associated with loops that are differential between nc14 relative and s10-12 or nc1-8. (*) FDR < 0.05, Fisher's exact test with Benjamini-Hochberg correction.

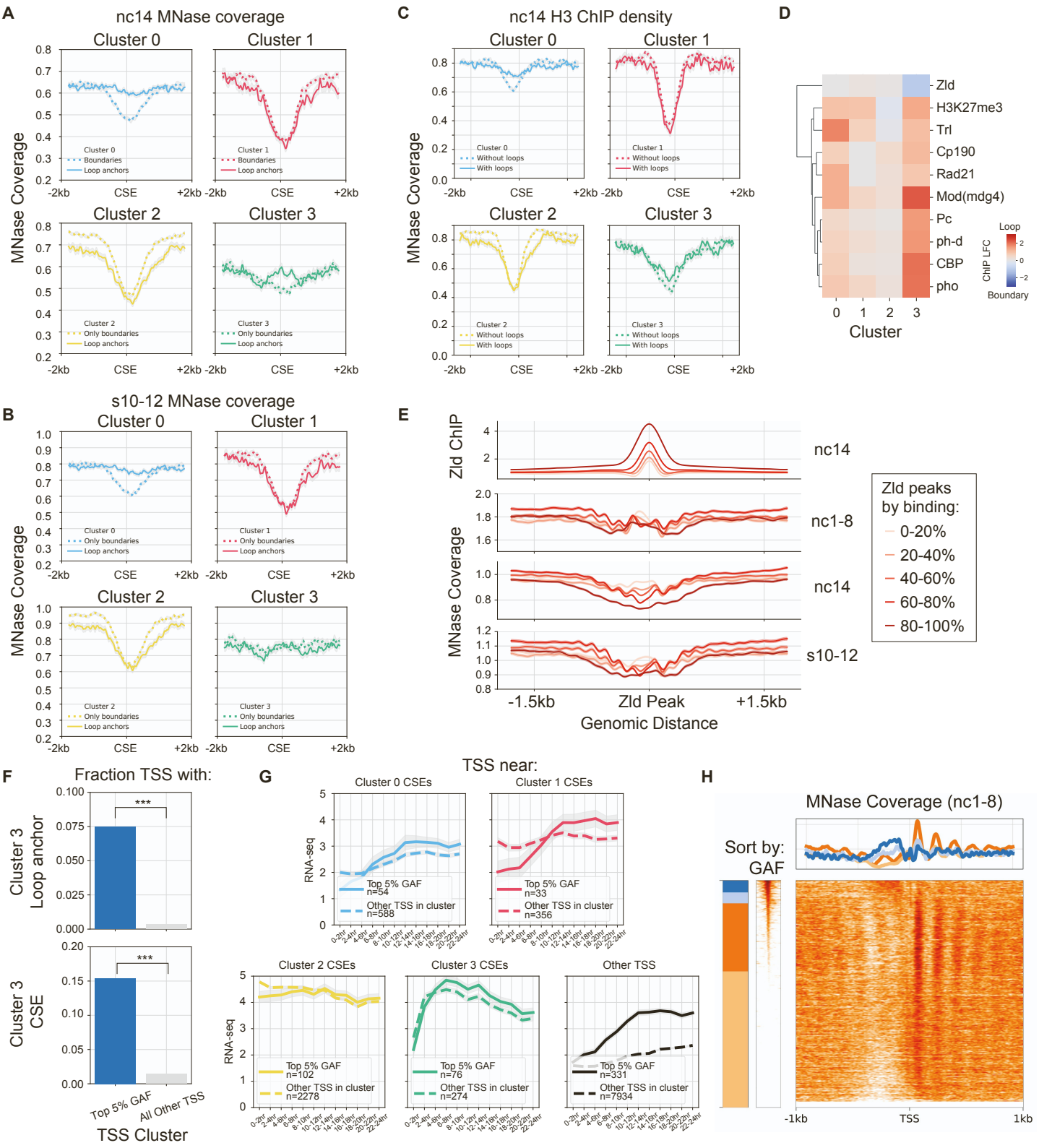


Figure S7: MNase coverage at CSEs and TSSs reveals patterns of TF binding and regulation, related to Figure 5. (A) MNase coverage at nc14 for CSE Clusters, split by whether they overlap a loop anchor or not. Gray outline represents ±SEM. **(B)** MNase coverage at s10-12 for CSE Clusters, split by whether they overlap a loop

anchor or not. Gray outline represents ±SEM. **(C)** H3 ChIP-seq density at nc14 for CSE Clusters, split by whether they overlap a loop anchor or not. Gray outline represents ±SEM. **(D)** LFC of ChIP-seq signal between CSEs annotated as loop anchors and CSEs annotated only as boundaries. **(E)** (Top row) Zelda ChIP-seq signal at Zelda peaks, split into quintiles by Zelda binding. (Second-fourth rows) MNase coverage in nc1-8, nc14, and s10-12 at Zelda peaks, split into quintiles by Zelda binding. **(F)** Fraction of TSSs overlapping a Cluster 3 loop anchor (top) or a Cluster 3 CSE (bottom). The bars represent two groups of TSSs: either TSSs with top 5% of GAF binding (the blue cluster in panel B), or all other TSSs. **(G)** RNA-seq over developmental time for TSSs near CSEs. TSSs with top 5% of GAF binding are plotted in a solid line and all other TSSs in the same CSE cluster are plotted in a dashed line. Gray outline represents ±SEM. **(H)** MNase coverage at nc1-8 at all TSSs, sorted by GAF binding at the TSS. Clusters represent top 5%, 5-10%, 10-40%, and bottom 60% TSSs as ordered by the GAF binding strength.

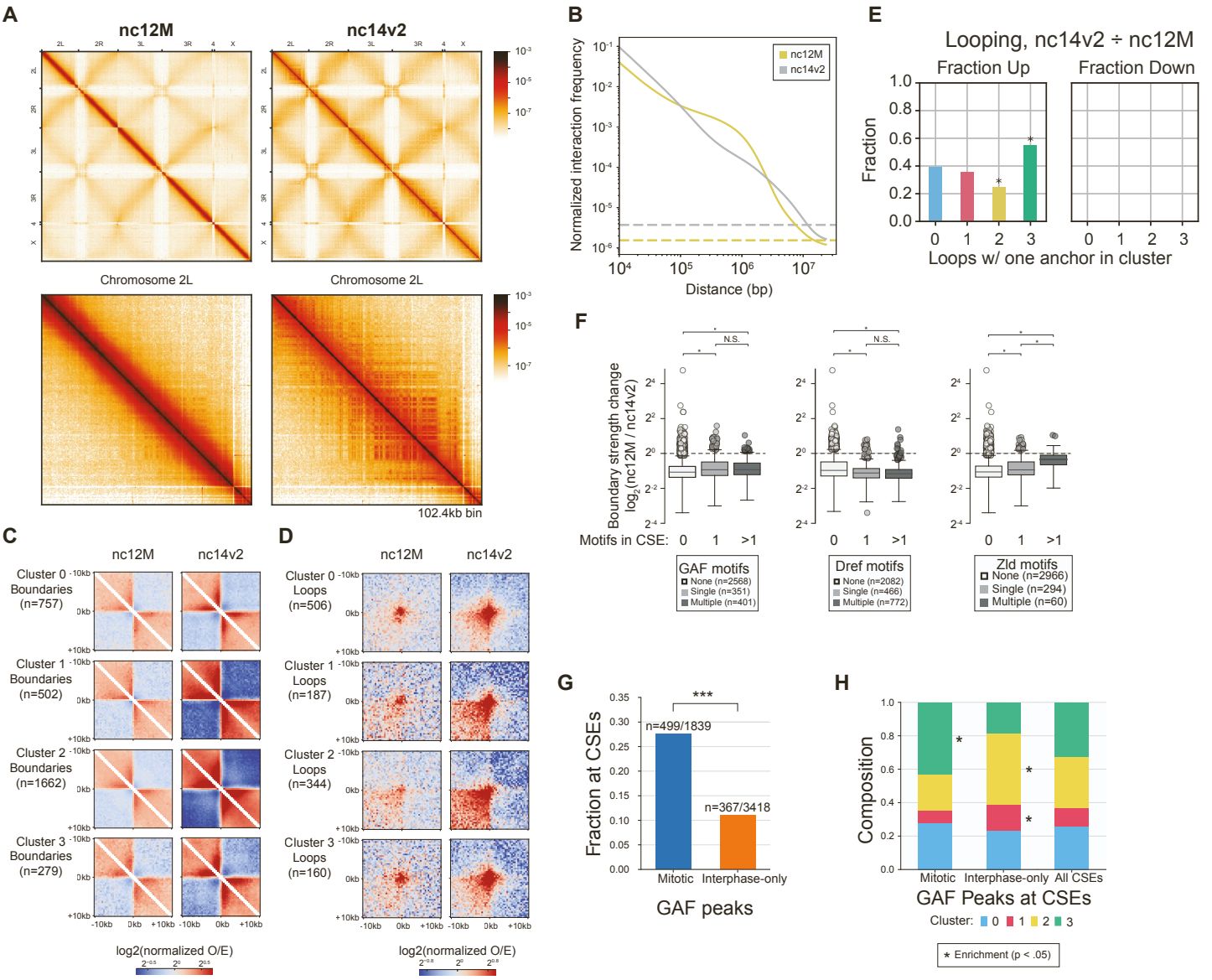


Figure S8: Summary of Mitotic Micro-C data, related to Figure 6. (A) Contact frequency maps for the nc12M and nc14v2 samples. The top row shows maps for the whole genome, and the bottom row contains maps for just Chromosome 2L. **(B)** Contact frequency decay curves for the nc12M and nc14v2 samples. **(C)** Boxplots showing the mitotic stability of boundaries based on the number of motifs that occur within the individual boundary. (*) P<0.05, Bonferroni-corrected two-sided Mann-Whitney U test. **(D)** Contact map pileups for boundaries, divided down by cluster. Contact maps are shown for the nc12M sample (left column) and nc14v2 sample (right column). Note the slight band in the upper diagonal of cluster 2 boundaries. **(E)** CSE clusters for loop anchors associated with loops that are differential between nc14 relative and s10-12 or nc1-8. (*) FDR < 0.05, Fisher's exact test with Benjamini-Hochberg correction. **(F)** Contact map pileups for loop foci, broken down by cluster. Contact

maps are shown for nc12M (left column) and nc14v2 (right column). **(G)** Fraction of GAF peaks overlapping CSEs, split by whether the GAF peaks are mitotic or interphase-only. Statistical enrichment was calculated using Fisher's exact test, (***; $p < 1e\text{-}4$). **(H)** CSE Cluster composition for mitotic or interphase-only GAF peaks that overlap CSEs. Statistical enrichment was calculated using Fisher's exact test (*; $p < .05$).

Supplemental References

1. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G.L., Chang, J.-M., and Cavalli, G. (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. Mol. Cell *71*, 73–88.e5.
2. Hug, C.B., Grimaldi, A.G., Kruse, K., and Vaquerizas, J.M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell *169*, 216–228.e19.