Chromosomes and Gene Expression

Chromosome Structure I: Loop extrusion or boundary:boundary pairing?

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed preprint version 1
February 5, 2024 (this version)

Posted to preprint server
November 17, 2023

Sent for peer review
November 12, 2023

Xinyang Bing, Wenfan Ke, Miki Fujioka, Amina Kurbidaeva, Sarah Levitt, Mike Levine, Paul Schedl, James B. Jaynes

Lewis Sigler Institute, Princeton University, Princeton, NJ. 08544 • Present address: BlueRock Therapeutics, Cambridge, Ma 20142 • Department of Molecular Biology, Princeton University, Princeton NJ 08544 • Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA 19107 • Present address: Center for Genomics and Systems Biology, New York University, NY 10003

Abstract

Two different models have been proposed to explain how the endpoints of chromatin looped domains ("TADs") in eukaryotic chromosomes are determined. In the first, a cohesin complex extrudes a loop until it encounters a boundary element roadblock, generating a stem-loop (and an unanchored loop). In this model, boundaries are functionally autonomous: they have an intrinsic ability to halt the movement of incoming cohesin complexes that is independent of the properties of neighboring boundaries. In the second, loops are generated by boundary:boundary pairing. In this model, boundaries are functionally non-autonomous, and their ability to form a loop depends upon how well they match with their neighbors. Moreover, unlike the loop-extrusion model, pairing interactions can generate both stem-loops and circle-loops. We have used a combination of MicroC to analyze how TADs are organized and experimental manipulations of the *even skipped* TAD boundary, *homie*, to test the predictions of the "loop-extrusion" and the "boundary-pairing" models. Our findings are incompatible with the loop-extrusion model and instead suggest that endpoints of TADs in flies are determined by a mechanism in which boundary elements physically pair with their partners, either head-to-head, or head-to-tail, with varying degrees of specificity. How the partners find each other is not clear but is unlikely to require a loop extrusion mechanism.

eLife assessment

This **valuable** work presents elegant experimental data from the Drosophila embryo supporting the notion that interactions among specific loci, called boundary elements, contribute to topologically associated domain (TAD) formation and gene regulation. Although the evidence supporting boundary elements as determinants of 3D structures is **compelling**, the evidence rejecting loop extrusion is **incomplete**. This study will be of interest to the nuclear structure community, particularly those using Drosophila as a model.

Introduction

The chromosomes of multicellular animals have a regular and inheritable physical organization. This was first recognized in studies on the lampbrush chromosomes in amphibian oocytes and the polytene chromosomes of insects (Callan, 1963; Nora et al., 2020; Zhimulev and Belyaeva, 1991; Zhimulev et al., 1983). Subsequent research has shown that the architectural features inferred from analysis of lampbrush and polytene chromosomes are present in chromosomes throughout much of the animal kingdom. The key organizational principle is the subdivision of the chromatin fiber into a series of independent looped domains, commonly called "TADs" (Cavalheiro et al., 2021; Chetverina et al., 2017; Jerković et al., 2020; Matthews and White, 2019; Rowley and Corces, 2018). With important exceptions, the arrangement of TADs along a given chromosome tend to be invariant and are largely independent of the cell type or developmental stage. This regular and inheritable organization is a reflection of the underlying mechanism of TAD formation. TADs are separated from each other by special elements called boundaries or insulators. While these elements have been found in many different species, they have been most fully characterized in *Drosophila* (Cavalheiro *et al*., 2021; Chetverina *et al*., 2017). Fly boundaries span DNA sequences of 150 bp to 1.5 kb in length and contain one or more nucleosome-free nuclease-hypersensitive regions. These nuclease-hypersensitive regions are targets for a large collection of DNA binding proteins that have been implicated in boundary function. Some of the fly boundary factors are widely conserved (CTCF, BEN family proteins, GAF), while others appear to be restricted to insect linages (Su(Hw), Pita, Zw5, Pita, Mod(mdg4), BEAF) (Heger et al., 2013; Heger and Wiehe, 2014; Schoborg and Labrador, 2010).

Boundary elements in flies are not only responsible for organizing the chromatin fiber, they also have genetic activities. When interposed between enhancers or silencers and target promoters, boundary elements block regulatory interactions (Bell et al., 2001; Chetverina et al., 2014; Chetverina *et al*., 2017). This insulating activity provides a mechanism for delimiting units of independent gene activity: genes located between a pair of compatible boundaries are subject to regulatory interactions with enhancers/silencers present in the same chromosomal interval, while they are insulated from the effects of enhancers/silencers located beyond either boundary in adjacent regulatory neighborhoods. Genetic studies suggest that the insulating activity of boundary elements is a consequence of subdividing the chromosome into a series of topologically independent domains (Cai and Shen, 2001; Gohl et al., 2011; Muravyova et al., 2001). Organizing the chromatin fiber into looped domains enhances contacts between sequences within the loop, while it suppresses contacts with sequences outside of the loop. While it is not known whether a similar insulation mechanism is at play in mammals, the fact that mammalian chromosomes are also subdivided into TADs by boundary-like elements suggests that it might be.

A critical question in chromosome biology is the mechanism(s) responsible for determining the endpoints of the loop domains, the TADs. In mammals, a novel loop-extrusion model has been proposed to not only generate chromosomal loops but also determine the endpoints of those loops (Alipour and Marko, 2012; Fudenberg et al., 2016; Guo et al., 2012; Guo et al., 2015; Nuebler et al., 2018; Sanborn et al., 2015). In this model, a cohesin complex initiates loop formation at a loading site within the "loop-to-be" and then extrudes a chromatin loop until it bumps into CTCF-dependent roadblocks, one on each side of the extruding loop. The location of these roadblocks combined with the processive action of the cohesin complex determines the endpoints of each TAD (Davidson and Peters, 2021; Ghosh and Meyer, 2021; Mirny and Dekker, 2022; Perea-Resa et al., 2021). A key assumption of the loop-extrusion model is that mammalian boundaries are fully autonomous: they are roadblocks, and their physical presence in and of itself is sufficient to define the loop endpoint, independent of the functional properties of neighboring boundaries. In a more specific variant of this model, the relative orientation of the paired CTCF roadblocks is also important—in order to halt cohesin-mediated extrusion, the CTCF

sites on each side of the loop must have a convergent orientation. In this model the CTCF roadblocks are able to block an oncoming cohesin complex in only one direction; however, their intrinsic blocking activity is independent of the orientation of other CTCF sites in their neighborhood.

While loop extrusion is widely thought to be operative in mammals, the evidence in flies is inconsistent with this mechanism. Genetics studies have shown that fly boundaries are functionally non-autonomous and that their activities in both loop formation and gene regulation depend upon their ability to engage in direct physical interactions with other boundaries (Chetverina *et al*., 2014; Chetverina *et al*., 2017). That fly boundaries might function by physical pairing was first suggested by studies which showed that the *gypsy* transposon *su(Hw)* boundary and the bithorax complex (BX-C) *Mcp* boundary can mediate regulatory interactions (enhancer/silencer: reporter) between transgenes inserted at sites separated by 15-32 MB or more (Muller et al., 1999; Sigrist and Pirrotta, 1997; Vazquez et al., 1993). Consistent with direct physical interactions, these distant transgenes co-localize *in vivo*, and in living tissue remain in contact for extended periods of time (Chen et al., 2018; Li et al., 2011; Vazquez et al., 2006).

Some of the parameters that govern pairing interactions have been defined in two different transgene assays, insulator bypass and boundary competition. In one version of the bypass assay, a set of enhancers are placed upstream of two different reporters (Cai and Shen, 2001; Kyrchanova et al., 2008a; Muravyova *et al*., 2001). When a single boundary is introduced between the enhancers and the closest reporter, the enhancers are unable to activate either reporter. However, if the same boundary is placed downstream of the closest reporter, bypass is observed. In this case the closest reporter, which is bracketed by the two boundaries, is still insulated from the enhancers; however, the downstream reporter is activated (Muravyova *et al*., 2001). Kyrchanova *et al*. (Kyrchanova *et al*., 2008a) showed that bypass activity is, in most cases, orientation-dependent. When the same boundary is placed in the upstream and downstream position, they typically need to be introduced in the opposite orientation. The reason for this is that self-pairing is head-to-head and this configuration generates a stem-loop topology, bringing the upstream enhancers in close proximity to the downstream reporter. When the two boundaries are introduced in the same orientation, the enhancers do not activate the downstream reporter because a circle-loop rather than a stem-loop is formed (Chetverina *et al*., 2017; Kyrchanova *et al*., 2008a). Head-to-head pairing seems to be a common feature of self-pairing interactions between fly boundaries and was observed for *scs*, *scs'*, *iA2* and *wari* (Kyrchanova *et al*., 2008a). A preference for head-to-head interactions is also observed for heterologous pairing interactions between different boundaries in the *Abdominal-B* (*Abd-B*) region of the BX-C (Kyrchanova et al., 2011; Kyrchanova et al., 2008b). There are of course some exceptions. The BX-C *Fab-7* boundary pairs with itself and with its neighbor *Fab-8* in both orientations (Kyrchanova *et al*., 2011; Kyrchanova *et al*., 2008b). Bypass activity also depends upon finding the proper match. For example, when multimerized dCTCF or Zw5 binding sites are placed both upstream and downstream of the closest reporter, bypass is observed. However, there is no bypass when multimerized dCTCF sites are tested with multimerized Zw5 sites (Kyrchanova *et al*., 2008a). Similar results have been reported for the Zipic and Pita factors (Zolotarev et al., 2016). Boundary:boundary pairing preferences are also observed in competition assays in which different boundaries are introduced into the same transgene (Gohl *et al*., 2011). In these assays, the insulating activity of a boundary in the blocking position (between an enhancer and a reporter that is flanked by a 3' boundary) is challenged by placing a heterologous boundary upstream of the enhancer. If the boundary upstream of the enhancer is a better match for the boundary located downstream of the reporter, then the insulating activity of the boundary in the blocking position can be compromised or lost altogether. These and other experiments argue that fly boundaries are functionally non-autonomous—that is, their activities depend upon their ability to physically interact with other boundaries.

However, the differences between the two models are not limited to the mechanisms for determining the endpoints of TADs. The models also differ in the possible topologies of the chromatin loops that form TADs. Only two loop topologies can be generated by the cohesin-extrusion mechanism. One is a stem-loop while the other is an unanchored loop. Depending upon the location of the road-blocks that halt cohesin extrusion and the cohesin loading sites, the chromatin fiber would be organized into a series of stem-loops. These stem-loops will be separated from each other by unanchored loops (see **Fig. 1B**). This is the configuration that is often illustrated in articles discussing the loop-extrusion model. If the relevant road-blocks are very closely juxtaposed, then the unanchored connecting loops will disappear and be replaced by a series of stem-loops which point in opposite orientations, as illustrated in **Fig. 1C**. Stem-loops are also possible in the boundary-pairing model, and they will be formed when two heterologous boundaries pair with each other head-to-tail (see **Fig 1D**) in *cis*. Like loop-extrusion, the stem-loops could be separated from each other by unanchored loops (**Fig. 1E**). If the boundaries in the neighborhood pair with each other head-to-tail, a series of connected stem-loops pointing in opposite orientations will be generated (**Fig. 1D**). In this case, the main axis of the chromosome would be a series of paired boundaries (**Fig. 1F**).

Stem-loops are not, however, the only loop topology that can be generated in the boundary-pairing model. If the neighboring boundaries pair with each other head-to-head instead of head-to-tail, a circle-loop structure will be generated (**Fig 1G**). Like stem-loops, the circle-loops could be connected by unanchored loops (**Fig. 1H**). Alternatively, if boundaries pair with both neighbors head-to-head this will generate a series of linked circle-loops oriented in (more or less) the same direction (see **Fig. 1I**).

In the studies reported here we have critically evaluated these two models. We have used a combination of MicroC to analyze how TADs are organized and experimental manipulations to test the predictions of the "loop-extrusion" and the "boundary-pairing" models. For these experimental manipulations, we have used the well-characterized *homie* boundary from the *even skipped* (*eve*) locus. *homie* together with *nhomie* flank the 16 kb *eve* locus, and these two elements share many of the properties of other fly boundaries (Fujioka et al., 2013; Fujioka et al., 2009). Like the *gypsy* and *Mcp* boundaries, *nhomie* and *homie* can mediate long-distance regulatory interactions (Fujioka et al., 2016). For example, a *lacZ* reporter transgene containing *homie* (or *nhomie*) inserted on one homolog can be activated by an *eve* enhancer in a *homie*- (or *nhomie*-) containing transgene inserted on the other homolog over a distance of 2 Mb. Long-distance activation is also observed in heterologous combinations of *nhomie* and *homie*, but not in combinations with phage *lambda* DNA. Similarly, regulatory interactions are observed when a *homie*- or *nhomie*- containing reporter is inserted at an attP landing site in the *hebe* gene, 142 kb upstream of the *eve* locus (Chen *et al*., 2018; Fujioka *et al*., 2016; Fujioka *et al*., 2009). In this case, the *eve* enhancers drive reporter expression during development in a pattern that mimics the stage-and tissue-specific expression of the endogenous *eve* gene. Activation also depends upon the orientation of the boundary (5'→3') relative to the reporter in the transgene. For *homie*, the reporter must be located "upstream" of the boundary (5'→3'), just like the *eve* enhancers (and *eve* gene) are "upstream" of endogenous *homie* (see **Fig. 2B**). Little or no activation is observed when the relative orientation is reversed so that the reporter is located "downstream" of *homie*. For *nhomie*, the reporter must be located "downstream", in the same relative position as that in which the *eve* enhancers and *eve* gene are located with respect to endogenous *nhomie* (5'→3': **Fig. 2B**).

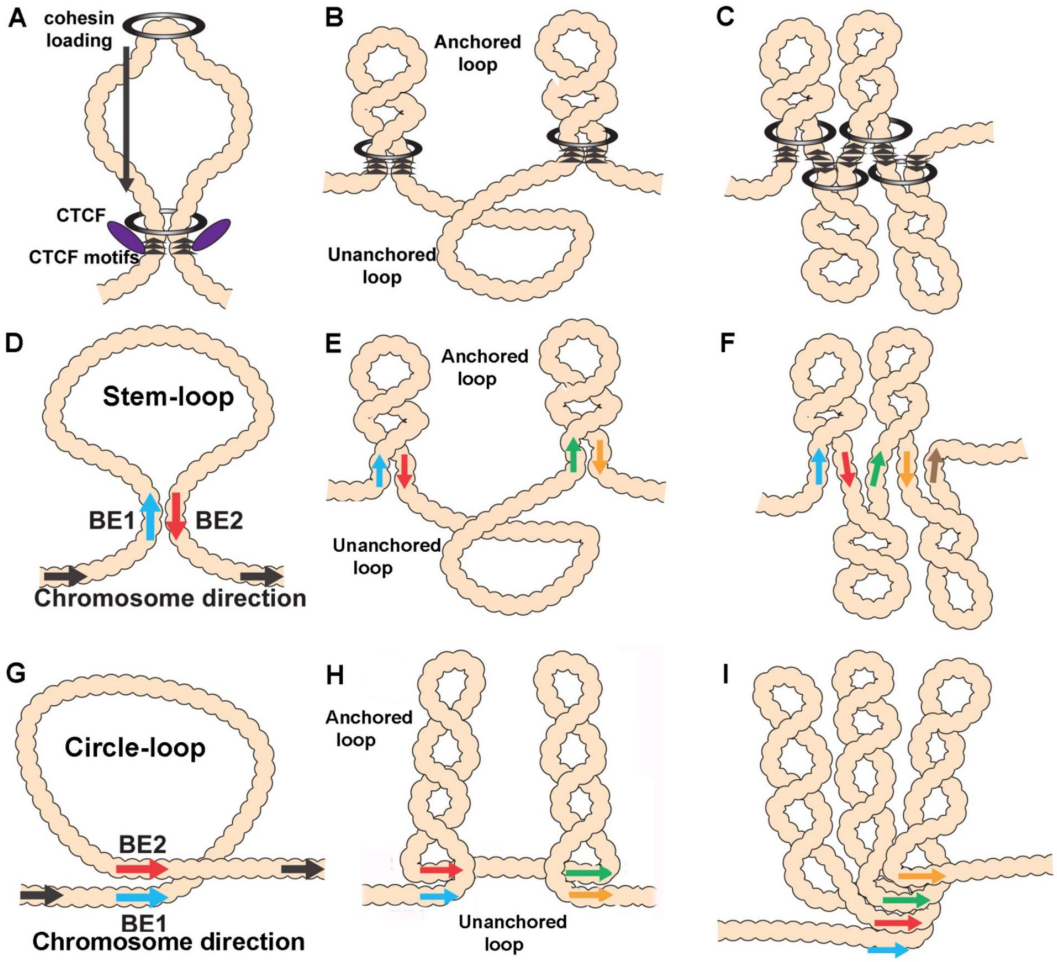

Figure 1.

Diagram of the possible loop topologies generated by loop-extrusion and boundary:boundary pairing.

A) Cohesin embraces a loop at a loading site somewhere within the TAD-to-be and then extrudes a loop at an equal rate on both strands. Extrusion continues until cohesin encounters boundary roadblocks on both strands. B) In one model, the orientation of the roadblock is important. As a consequence, the chromatin fiber will be organized into a series of stem-loops separated from each other by unanchored loops. C) If the presence of a boundary roadblock, but not its orientation, is sufficient to halt extrusion, the chromatin fiber will be organized into a series of linked stem-loops. The main axis of the chromosome will be defined by the cohesin:CTCF roadblocks. D) The pairing of two boundaries head-to-tail generates a stem-loop. E) If boundary pairing interactions are strictly pairwise, head-to-tail pairing will generate a series of stem-loops separated from each other by unanchored loops. F) If boundaries can engage in multiple head-to-tail pairing interaction, the chromosome will be organized into a series of linked stem-loops. The main axis of the chromosome will be defined by a series paired boundaries. G) The pairing of two boundaries head-to-head generates a circle-loop. H) If boundary pairing interactions are strictly pairwise there will be a series of circle-loops separated from each other by unanchored loops. I) If boundaries can engage in multiple head-to-head pairing interactions, the chromatin fiber will be organized into a series of circle-loos connected to each other at their base. Pairing interactions between boundaries #1 and #2 need not be in register with pairing of boundaries #2 and #3. In this case, the main axis of the chromosome may bend and twist and this could impact the relative orientation of the circle-loops.

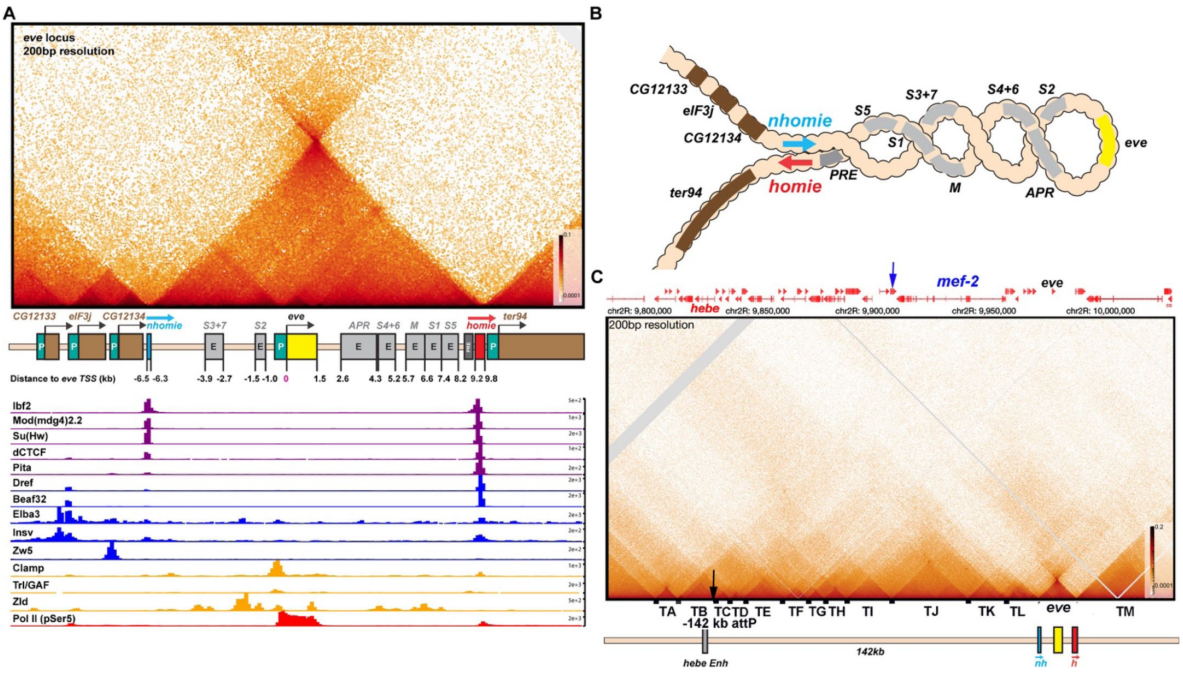

Figure 2.

TAD organization of the *even-skipped* locus and neighboring sequences.

A) The *eve* skipped TAD is a volcano with a plume that is anchored by *nhomie* and *homie*. ChIP-seq data below the MicroC map indicate many of the known fly chromosomal architectural proteins proteins are associated with the two *eve* boundary elements *in vivo* (Cuartero et al., 2014; Duan et al., 2021; Gaskill et al., 2021; Li et al., 2015; Sun et al., 2015; Ueberschär et al., 2019; Zolotarev *et al*., 2016). B) The *eve* locus forms a stem-loop structure. In this illustration *nhomie* pairs with *homie* head-to-tail, and this configuration forms a stem-loop which brings sequences upstream *nhomie* and downstream of *homie* in contact as is observed in insulator bypass assays (Kyrchanova *et al*., 2008a). The *eve* locus is shown assembled into a coiled "30 nM" chromatin fiber. C) MicroC contact pattern for the chromosomal region spanning the attP site at –142 kb and the *eve* locus. Like the *eve* volcano, this contact pattern was generated using aggregated previously published NC14 data (Batut *et al*., 2022; Levo *et al*., 2022). A black arrow indicates the –142kb locus where transgenes are integrated into the genome. A blue arrow indicates position of the *Etf-QO* gene. Note the numerous TADs separating the –142kb *hebe* locus from the *eve* "volcano TAD". Individual TADs are labeled TA—TM. Black boxes indicate positions of sequences that, based on ChIP experiments, are bound by one or more known insulator proteins *in vivo*.

Results

The eve TAD

We used MicroC to probe the TAD organization of the *eve* locus and the genomic region that extends from *eve* to just beyond the attP site at –142 kb. For this purpose, we chose nuclear cycle 14 embryos. At this stage of development, most nuclei in our sample will have completed replication and will be arrested in G2. As sister chromosomes in flies pair with each other in register, there will be two juxtaposed copies of this region of the chromosome. The sister chromosomes should be linked to each other by cohesin in preparation for the asynchronous nuclear divisions at the end of NC14 (Collier and Nasmyth, 2022; Gligoris et al., 2014). While homolog pairing is limited in pre-cellular blastoderm embryos, there are regions of the chromosome that already show evidence of pairing (Fung et al., 1998). In fact, transvection between *gypsy* insulator-containing transgenes has been observed in live imaging studies of nuclear cycle 14 embryos (Lim et al., 2018). Since the NC14 nuclei are just emerging from an earlier mitosis and a round of DNA synthesis, one would imagine that other key features of chromosome organization might also be in the process of being established. Thus, it should be possible to capture evidence of intermediates like those expected to be generated as the extruding cohesin complexes move through the *eve* TAD and other nearby loci at this particular stage of development.

Shown in **Fig. 2A** is a blow-up of the TAD organization of the *eve* locus. As expected, the left and right boundaries of the *eve* TAD correspond to *nhomie* and *homie*, respectively, and sequences containing these two boundaries converge at the apex of the *eve* interaction triangle. All of the sequences within the TAD appear to come in contact with each other and define a high density of internal contact domain (HDIC). These contacts are expected to be generated, at least in part, by the sliding of the chromatin fiber within the TAD against itself, much like sequences in circular/supercoiled DNA can interact with each other (**Fig. 1E**). There are also sequences within the *eve* TAD that show enhanced interactions. For example, there is a darker interaction triangle connecting *nhomie* and the promoter region, while another triangle connects *homie* (and the PRE adjacent to *homie*) to the promoter region. These enhanced internal contacts would be expected to help insulate the *eve* gene from external regulatory elements and at the same time facilitate interactions between *eve* and its upstream and downstream enhancers.

As predicted by both the loop-extrusion and boundary-pairing models, the *eve* TAD is a stem-loop, not a circle-loop (**Fig. 2B**). This is evident from the prominent “volcano plume” at the apex of the *eve* TAD (volcano triangle). A volcano plume is observed because sequences flanking *eve* are brought into close proximity when *homie* and *nhomie* are linked to each other by either cohesin or by head-to-tail pairing (see **Fig.1**: see also Ke et al.). Fig. Supplemental 1 shows the hierarchy of interactions between TADs to either side of *eve* (L-M, K-M and J-M in Sup. Fig. 1). This conclusion is supported by previous transgene studies. As first described in the boundary bypass experiments of Cai and Shen (Cai and Shen, 2001) and Muravyova et al. (Muravyova *et al*., 2001), the formation of a stem-loop structure when two boundaries are linked together brings sequences immediately beyond the linked boundaries into close proximity with each other. In this assay, a stem-loop structure is required, as it enables enhancers flanking the “upstream” boundary to interact with reporters flanking the downstream boundary (Kyrchanova *et al*., 2008a). In contrast, a circle-loop configuration does not bring the neighbors to either side of *eve* into close proximity.

While a stem-loop is the topology expected for the *eve* TAD in both models, one feature of the TAD is inconsistent with the loop-extrusion model. According to this model, loops (TADs) are generated by a cohesin complex that forms a small bubble at a loading site somewhere in the TAD-to-be and then extrudes both strands of the growing loop at an equal rate until it bumps into the boundary road-blocks (Davidson and Peters, 2021; Ghosh and Meyer, 2021; Mirny and Dekker, 2022;

Perea-Resa *et al*., 2021). If the loading site were located in the middle of the TAD-to-be, and the extrusion intermediates in a population of nuclei were captured by crosslinking, one should observe a crosslink-generated stripe that begins at the loading site and extends perpendicular to the chromosomal DNA until it reaches the apex that links the two TAD boundaries with imaginary lines of 45° (Fig. Supplemental 2B). This is the pattern that would be generated by a "broken zipper" when it transiently links the zipper teeth equidistant from the initial "loading site". If the loading site is off center to the left, the crosslinking of the passing cohesin complex in a population of nuclei will generate a perpendicular stripe until cohesin come to a halt when it encounters the roadblock on the left (Fig. Supplemental 2C). Assuming that the complex continues to extrude the right strand, it will then generate a stripe at 45° degrees that extends to the apex of the TAD. If the loading site corresponds to one of the CTCF roadblocks or is located close to the roadblock, the cohesin complex will generate a 45° stripe that comes to a halt at the apex of the triangle when it encounters the neighboring CTCF roadblock (Fig. Supplemental 2D and E). If there are multiple loading sites within the TAD, there should be a series of perpendicular stripes that generate and sequentially reinforce the stripe(s) at 45° (Fig. Supplemental 2F). However, in spite of the fact that NC14 fly embryos should provide by far the best prospect of actually capturing loop extrusion intermediates, there are no vertical stripes in the *eve* TAD, nor are there stripes at 45° that extend to the apex of the *eve* TAD. Nor do we observe a population of perpendicular stripes that convert into a series of reinforced 45° stripes. There is also no evidence of stripes that are initially tilted to the left or right of a "loading site", as might be observed if the rate of extrusion were unequal on the two strands, and then subsequently convert to a 45° stripe as cohesin comes to a halt at the first boundary encountered.

TADs in the eve environs

The attP insertion site at –142 kb is near the end of the first exon of the *hebe* transcription unit (**Fig. 2C**). It is located just within a ~14 kb TAD, TB, that has a high density of internal contacts, and extends from the *hebe* promoter to the 3' end of the neighboring *dila* gene. In between *hebe* and the *eve* locus there at least are ten distinct TADs, TC-TL (**Fig. 2C**). The endpoints of most of these TADs correspond to sequences that are associated with one or more known chromosomal architectural factors (dots in **Fig. 2C**).

These TADs correspond to the fundamental building blocks for organizing the 3D architecture of this chromosomal segment. Superimposed upon these TADs are regions that exhibit lower density of contacts (LDC). For example, the *mef2* gene spans two TADs, TJ and TK. TK contains the *mef2* distal promoter region and extends to the internal *mef2* promoter, while TJ extends to near the promoter of the divergently transcribed *Etf-QO* gene (blue arrow). Both of these TADs are linked together by a rectangular LDC domain J-K (see Fig. Supplemental 2). Likewise, TJ is linked to its immediate neighbor TI by the LDC domain I-J. In the next level, TADs separated by a single TAD are linked together. In the region immediately above TJ, the LDC domain I-K links TI to TK (Fig. Supplemental 2). Similarly, TJ is linked to TL by J-L while TJ is also linked to the large TAD TM (which contains *TER94* and the neighboring gene, *pka-R2*) on the other side of *eve* by J-M. The same pattern of a hierarchical series of LDC domains linking TAD neighbors, TAD next-door neighbors and next-next-door neighbors is observed in the DNA segment that includes the *hebe* gene.

--cohesin-mediated loop extrusion

Like *eve*, the TADs in the region between *eve* and the attP site in the *hebe* locus are defined by right angle triangles (volcanos) with a high density of internal contacts. However, unlike *eve,* these volcano triangles do not have a plume, and instead are surrounded by a series of LDC rectangular clouds. This is not the contact pattern that one might expect for either a series of stem-loops connected by unanchored loops or a series of stem-loops connected to each other. Moreover, there are no perpendicular or angled crosslinking stripes (broken zippers) emanating from the base of these triangles, nor are their stripes along their 45° legs. Thus, in spite of the fact that the NC14

nuclei used in this analysis are just emerging from an earlier mitosis and a round of DNA synthesis and should be in the process of assembling TADs, the crosslinking signatures that should be generated by the embrace of passing cohesin complexes are completely absent.

In the loop extrusion model, the LDC domains would arise in a subset of nuclei because the extruding cohesin complex breaks through one or more boundary roadblocks before its progression is halted (Fig. Supplemental 3A versus B, C and D) (Hsieh et al., 2020; Krietenstein et al., 2020). Breakthroughs could occur at only one roadblock on one side of the extruding loop (Fig. Supplemental 3B and C), or at multiple roadblocks on one or both sides (Fig. Supplemental 3D), giving rise to a set of overlying LDC domains like that evident in the region between *hebe* and *eve*. However, while there would have to be multiple breakthrough events to account for the hierarchical array of LDC domains observed in these NC14 embryos, there is no evidence of the 45° crosslinking stripes that would be expected to mark the legs of the LDC domains following the breakthrough. For example, the I-J LDC domain (Fig. Supplemental 2) could be generated by cohesin complexes that initiated extrusion in either TI or TJ and then failed to halt at the TI:TJ boundary (green arrowhead). In either case, there should be a 45° stripe of crosslinking that marks (at least part of) the left (TI) and/or the right (TJ) leg of the I-J LDC domain and extends to the apex of the TI-TJ LDC triangle; however, there is no evidence of such stripes. Nor is there evidence of crosslinking stripes marking one or both legs of the J-K or I-K LDC domains (Fig. Supplemental 2).

--boundary:boundary pairing

At least two different mechanisms are expected to account for the TADs and LDC domains in the region between *eve* and the attP insertion site at -142 kb. First, since none of the TADs in this DNA segment are topped by volcano plumes, they could correspond to circle-loops. As circle-loops are generated by head-to-head pairing, the boundaries in the region between *nhomie* and *hebe* would be expected to pair with each other head-to-head. (The TAD to TAD contact maps generated by circle-loops and stem-loops are considered further in the accompanying paper, Ke *et al.*)

As illustrated in Fig. Supplemental 3E-G, the coiled circle-loop TADs, though topologically independent, are expected to be in relatively close proximity to each other. This means that in addition to crosslinking events within the TADs, there will be crosslinking events between neighboring TADs. These crosslinking events could generate the LDCs seen in **Fig. 2C** and Fig. Supplemental 2. Since TADs next to each other (Fig. Supplemental 3E and F; Fig. Supplemental 2 I-J or J-K) would typically be expected to interact more frequently than TADs separated by one, two or more TADs (Fig. Supplemental 3G; Fig. Supplemental 2 I-K or J-L), there should be a progressive reduction in contact frequency at each step. This is generally what is seen (see Fig. Supplemental 2). While connected stem-loops generated by cohesin extrusion should also bump into one another, contacts between next-door neighbors might be expected to be less frequent than contacts between next-next-door neighbors (**Fig. 1B, C** and Fig. Supplemental 3A). (TAD-to-TAD crosslinking and the impact of loop topology is further examined in Ke et al., 2023.)

The second mechanism would be switching and/or combining pairing partners (Fig. Supplemental 3H-J). Though imaging studies suggest that pairing interactions can be of relatively long duration (>30 min: Chen *et al*., 2018; Vazquez *et al*., 2006)), they are not permanent, and other nearby boundaries can compete for pairing interactions (Gohl *et al*., 2011). If switching/combining occurs in NC14 embryos (see Fig. Supplemental 3H-J), then the precise pattern of TADs and LDC domains could also be impacted by the relative avidity of potential pairing partners in the neighborhood and also the distances separating potential partners. For example, the rectangle linking TH and TI (H-I in Fig. Supplemental 2) could represent partner switching so that the boundary between TI and TJ (green arrowhead in Fig. Supplemental 2) sometimes pairs with the boundary separating TH and TI (blue arrowhead) and sometimes with the boundary separating TG and TH (red arrowhead in Fig. Supplemental 1). Distant boundary elements could also

preferentially interact with each other, generating a supra-TAD that contains multiple TAD domains. For example, at the apex of the LDC domain B-I there is an interaction dot (purple arrow, Fig. Supplemental 2) that links the TA-TB boundary to the TI-TJ boundary (green arrowhead).

Activation of a distant reporter by the eve enhancers

In previous studies we found that a minimal 367 bp *homie* fragment can orchestrate regulatory interactions between the *eve* enhancers and reporters inserted at an attP site in the first intron of the *hebe* transcription unit, 142 kb upstream of the *eve* gene (Fujioka *et al*., 2016). This minimal fragment lacks the single *homie* CTCF site, but has sites for several other generic boundary proteins including BEAF, Zipic, Su(Hw), Pita and Ibf2. In blastoderm stage embryos, *lacZ* reporter expression was observed in a 7-stripe pattern that coincided with the stripes of the endogenous gene. However, at this stage only a subset of *eve*-expressing nuclei expressed the reporter. Later in development, during mid-embryogenesis, reporter expression was observed in cells in the dorsal mesoderm, the ventral CNS, and in the anal plate region, in a pattern that recapitulated *eve* gene expression (Fujioka *et al*., 2016). At these later stages, most of the *eve*- expressing cells also appear to express the reporter. The *eve*-dependent activation of the reporter in a stripe pattern has also been visualized in live imaging of pre-cellular blastoderm-stage embryos (Chen *et al*., 2018). Within the 7 *eve* stripes, the majority of the nuclei (~80-85%) did not express the *lacZ* reporter. In these nuclei, the transgene was found to be at a considerable average physical distance from the *eve* locus (~700 nM). In the remaining nuclei, the transgene was located in closer proximity to the *eve* locus (~350 nM, on average). Within this subset, the transgene was expressed in most of the nuclei. Moreover, in those cases in which the transgene was active, it remained in close proximity and continued to express *lacZ* for the duration of the experiment (~30 min).

In the two models for TAD formation, quite different mechanisms must be invoked to account for activation of the reporter at –142 kb by the *eve* enhancers. In the boundary-pairing model, the transgene *homie* boundary at –142 kb loops over the intervening TADs and pairs with the *nhomie:homie* complex flanking the *eve* TAD (c.f., **Fig. 3B and C**). In the loop-extrusion model, a cohesin complex initiating loop extrusion in the *eve* TAD must break through the *nhomie* roadblock at the upstream end of the *eve* TAD. It must then make its way past the boundaries that separate *eve* from the *attP* site in the *hebe* gene, and come to a halt at the *homie* boundary associated with the *lacZ* reporter. This would generate a novel TAD, *eveMammoth* (*eveMa*), that extends from the *eve homie* all the way to the *homie* fragment at –142 kb, and encompasses both the reporter and the *eve* gene, including its enhancers (c.f., **Fig. 3D**). Of course, the *eveMa* TAD could also be generated by a cohesin complex that initiated in, for example, TF. However, in this case, the runaway cohesin complex would have to break through the intervening boundary roadblocks in both directions. In both the boundary-pairing and loop-extrusion models, the configuration of the chromatin fiber would lead to the activation of *lacZ* expression by the *eve* enhancers, while the reporter would be protected from the *hebe* enhancers by the *homie* boundary.

To test these two models, we used a transgene that has two divergently transcribed reporters, *lacZ* and *gfp*, that are each under the control of an *eve* promoter (**Fig. 3A**). Two different versions of the transgene were generated. In the first, *LhomieG*, the *eve-lacZ* reporter is located “upstream” of *homie* (with respect to the 5’→3’orientation of *homie* in the *eve* locus), while *eve-gfp* is located downstream (**Fig. 3A**: top). In the second, the orientation of *homie* within the transgene is reversed, so that *gfp* is upstream of *homie*, while *lacZ* is downstream (*GhomieL*) (**Fig. 3A**: bottom). These two transgenes were then individually inserted into the *attP* site at –142 kb, in either the Z5 or Z3 orientation. In the Z5 orientation, the *eve-lacZ* reporter is on the *eve* side of the *homie* boundary, so that it is closer to the *eve* locus (**Fig. 3A**). The *eve-gfp* reporter is on the opposite side of the *homie* boundary and is farther away from the *eve* locus. This places the *eve-gfp* reporter next to a series of *hebe* enhancers located farther upstream, in the intron of the *hebe* gene, and thus it should be subject to regulation by these enhancers (**Fig. 2C**). In the Z3 orientation, the entire transgene is inserted in the opposite orientation so that *eve-gfp* is on the *eve*

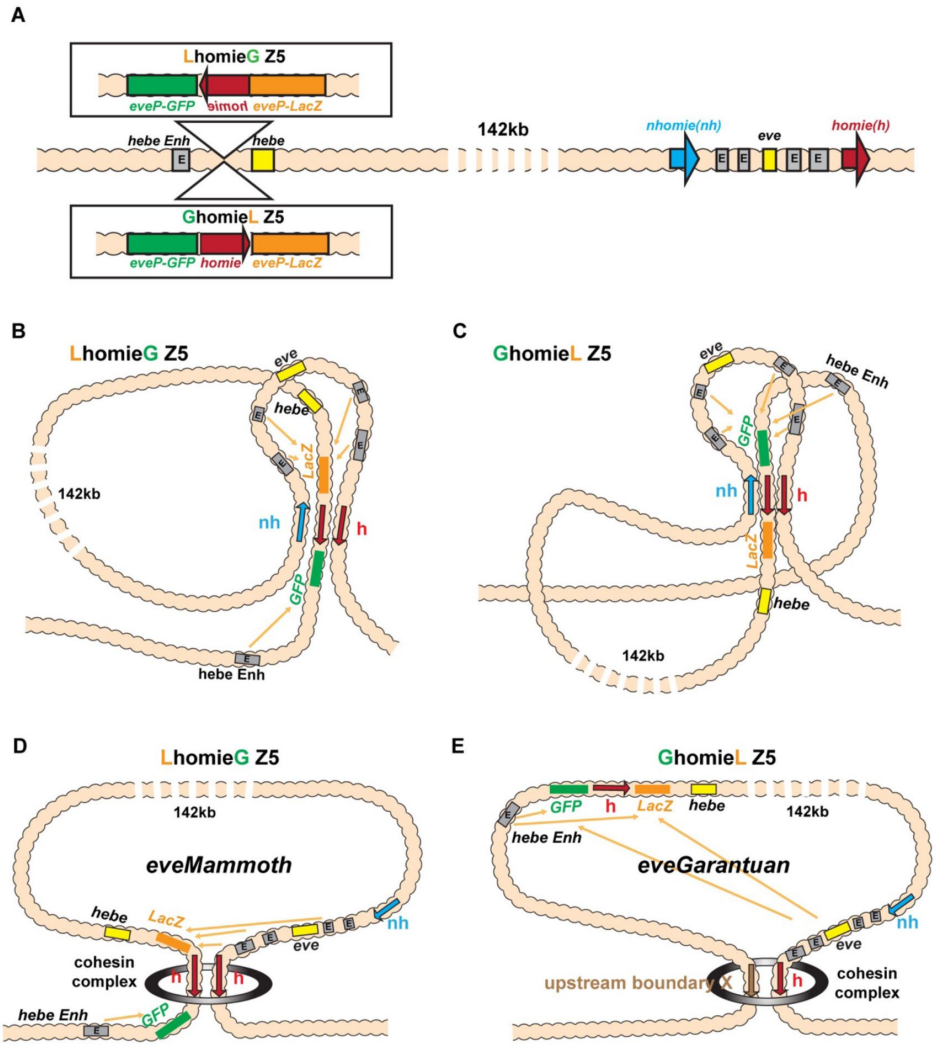

Figure 3.

Schematics of boundary pairing and loop extrusion.

A) *LhomieG Z5* and *GhomieL Z5* transgenes and the *eve* locus on a linear map. The same color codes are used throughout. In the *eve* locus, *nhomie* (blue arrow) and *homie* (red arrow) are oriented, by convention, so that they are pointing towards the right. This convention is maintained in the two transgenes. In *LhomieG Z5 homie* (top) is in the opposite orientation from *homie* in the *eve* locus is pointing away from the *eve* locus. In *GhomieL Z5 homie* (bottom) *homie* is in the same orientation as the *homie* in the *eve* locus and is pointing towards the *eve* locus. B) Predicted boundary pairing interactions between *LhomieG Z5* and the *eve* locus. *homie* in the transgene pairs with *homie* in the *eve* locus head-to-head. Since *homie* in the transgene is pointing in the opposite orientation from *homie* in the *eve* locus, a stem-loop will be generated. If *homie* in the transgene also pairs with *nhomie* in the *eve* locus head-to-tail, a loop structure like that shown in B will be generated. In this topology, *eve-lacZ* is in close proximity to *eve* enhancers and *eve-gfp* is in contact with the *hebe* enhancer. C) Predicted boundary pairing between *GhomieL Z5* and the *eve* locus. The *GhomieL Z5* transgene is inserted in the same chromosomal orientation as *LhomieG Z5*; however, the *homie* boundary is inverted so that it is pointing towards the *eve* in the same orientation as the *homie* in the *eve* locus. *Homie* in the transgene will pair with *homie* in the *eve* locus head-to-head, and this generates a circle-loop. If *homie* in the transgene also pairs with *nhomie* in the *eve* loucs a loop structure like that shown in C will be generated. In this topology *eve-gfp* will be activated by both the *eve* and *hebe* enhancers. D) Loop extrusion model for *LhomieG Z5*. Transgene *homie* and endogenous *homie* determine the endpoints of the extruded *eveMammoth* (*eveMa*) loop. In this topology, *eve-lacZ* is in close proximity to the *eve* enhancers, while *eve-gfp* is close to the *hebe* enhancers. E) Loop extrusion model for *GhomieL Z5*. The cohesin complex bypasses transgenic *homie* and is stopped at an upstream boundary X to generate a novel *eveGarantuan* (*eveGa*) loop. Both *eve-gfp* and *eve-lacZ* are located within the same *eveGa* TAD, and thus should interact with *eve*. Since the *hebe* enhancers are within this TAD as well, they would also be able to activate both reporters.

side of the *homie* boundary, so that it is closer to the *eve* locus (**Fig. 3A**). The *eve-gfp* reporter is on the opposite side of the *homie* boundary and is farther away from the *eve* locus. This places the *eve-gfp* reporter next to a series of *hebe* enhancers located farther upstream, in the intron of the *hebe* gene, and thus it should be subject to regulation by these enhancers (**Fig. 2C**). In the Z3 orientation, the entire transgene is inserted in the opposite orientation so that *eve-gfp* is on the *eve* side of *homie*, while *eve-lacZ* is next to the *hebe* enhancers (**Fig. 6A**). As a control, we also generated a Z5 transgene in which *lambda* DNA instead of *homie* was inserted between the *eve-gfp* and *eve-lacZ* reporters.

LhomieG Z5 and GhomieL Z5

The results for transgenes *LhomieG Z5* and the control *LlambdaG Z5* are most straightforward and will be considered first. In both the loop-extrusion and boundary-pairing models, the *LlambdaG Z5* control is not expected to display any regulatory interactions with the *eve* locus. **Fig. 4C** shows that this is the case: the *eve-lacZ* and *eve-gfp* reporters in *LlambdaG Z5* embryos are silent at the blastoderm stage, and transcripts are only very rarely detected. Later in development, both reporters are expressed in a repeating pattern along the dorsal midline (**Fig. 6E**). This expression is driven by the *hebe* gene enhancers located just beyond the *gfp*-reporter (Fujioka *et al*., 2009).

In the case of the *LhomieG Z5* insert, regulatory interactions between the transgene and the *eve* locus will be observed if a stem-loop is established that links the transgene to *eve*. In the loop-extrusion model, the breakthrough cohesin complexes would generate a stem-loop, *eveMammoth* (*eveMa*), by coming to a halt at the *homie* boundary in the transgene and the *homie* boundary in *eve*, as illustrated in **Fig. 3D**. This will place the *lacZ* reporter in close proximity to the *eve* enhancers. It will also disrupt the *eve* TAD, as *nhomie* is no longer linked to the *homie* boundary by cohesin. In the boundary-pairing model, the *lacZ* reporter in the transgene is “upstream” of *homie* just like the *eve* gene and its enhancers are “upstream” of *homie* in the *eve* locus. Since *homie* pairs with itself head-to-head, pairing between *homie* in the transgene and *homie* in the *eve* locus will generate the same stem-loop as that generated by loop-extrusion. However, because *homie* pairs with *nhomie* head-to-tail, a more complex structure like that in **Fig. 3B** would be expected if the pairing interactions with the transgene do not disrupt the *eve* TAD. Consistent with a tripartite structure of the sort shown here, previous studies have shown that three *Mcp* transgenes on three different chromosomes can interact with each other genetically (Muller *et al*., 1999). Subsequent direct visualization of *Mcp*-mediated pairing interactions in imaginal discs by Vazquez et al. (Vazquez *et al*., 2006) showed that four *Mcp*- containing transgene separated by Mbs (at Bridges cytogenetic intervals ~65 on the left arm of chromosome 3, and ~83 and ~95 on the right arm of chromosome 3) and/or located on different homologs, clustered in the same nuclear foci in 94% of the nuclei.

As predicted from the formation of a stem-loop linking the transgene to the *eve* locus, *lacZ* is expressed in blastoderm-stage embryos in 7 stripes that coincide with the 7 stripes of the endogenous *eve* gene. However, unlike *eve*, not all nuclei in the seven stripes express *lacZ* (compare with the *eve* control in **Fig. 4A**). Later, during mid-embryogenesis, *lacZ* is expressed in the mesoderm, the CNS, and the anal plate in a pattern that mimics the endogenous *eve* gene. At this stage almost all *eve* positive cells are also positive for *lacZ*. Unlike the *LlambdaG Z5* transgene, the *hebe* enhancers do not drive *eve-lacZ* expression, as they are beyond the *homie* boundary (**Fig. 4E**). The *eve-gfp* reporter in the *LhomieG Z5* transgene responds differently. As was observed for the *LlambdaG Z5* control, the *eve* enhancers drive little if any *gfp* expression, either at the blastoderm stage or later in development (**Fig. 5C, E**). Instead, *gfp* is expressed in the midline during mid-embryogenesis under the control of the *hebe* enhancers (**Fig. 5E**), which should be in the same TAD as *eve-gfp*. These results would be consistent with both the loop-extrusion (*eveMa*) and boundary-pairing models (**Fig. 3B and D**).

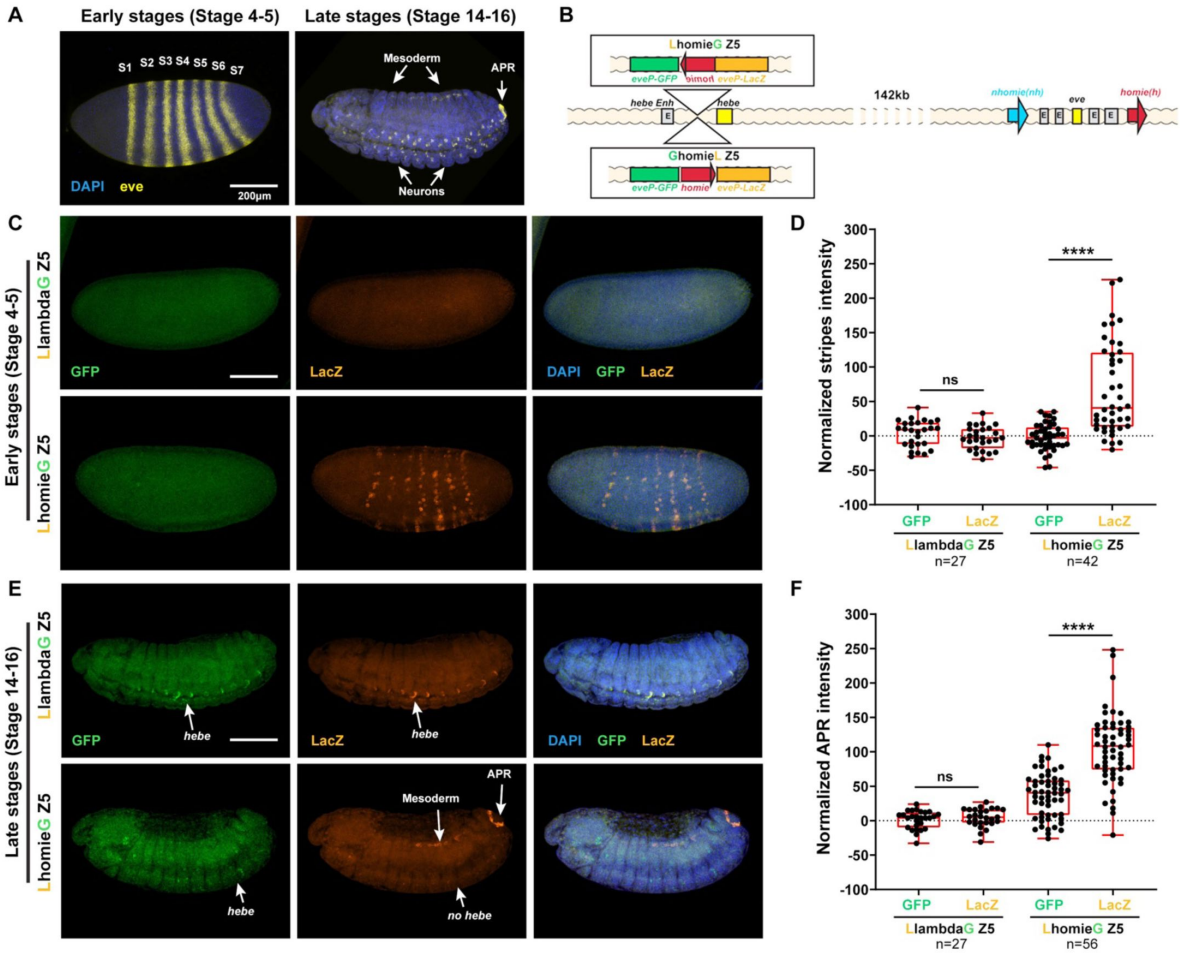

Figure 4.

The *lacZ* reporter is activated by *eve* enhancers in the *LhomieG Z5* transgene.

A) Expression patterns of endogenous *eve* in early (stage 4-5) and late (stage 14-16) embryos by smFISH. *eve* (Atto 633) probes were used to hybridize with *eve* mRNA. *eve* is shown in yellow and DAPI in blue. N>3. B) Schematics of transgenes. C) Expression patterns of transgenic reporters in early stages embryos. Top: *LlambdaG Z5*. Bottom: *LhomieG Z5*. *GFP* is in green, *lacZ* is in orange, and DAPI is in blue, here and in E. D) Quantification of normalized stripe signals for transgenes shown in C. The reporter smFISH intensity (signal to background), here and in F, was measured, normalized, and plotted as described in Methods. N>3. n=27 for *LlambdaG Z5* and n=42 for *LhomieG Z5*. E) Expression patterns of transgenic reporters in late stages embryos. Top: *LlambdaG Z5*. Bottom: *LhomieG Z5*. F) Quantification of normalized stripe signals for transgenes shown in E. N>3. n=27 for *LlambdaG Z5* and n=56 for *LhomieG Z5*. Scale bars = 200µm. N = # independent biological replicates. n = # animals. The paired two-tailed t-test were used for statistical analysis. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001, ns: not significant. Raw measurements are available in the Source Data files.

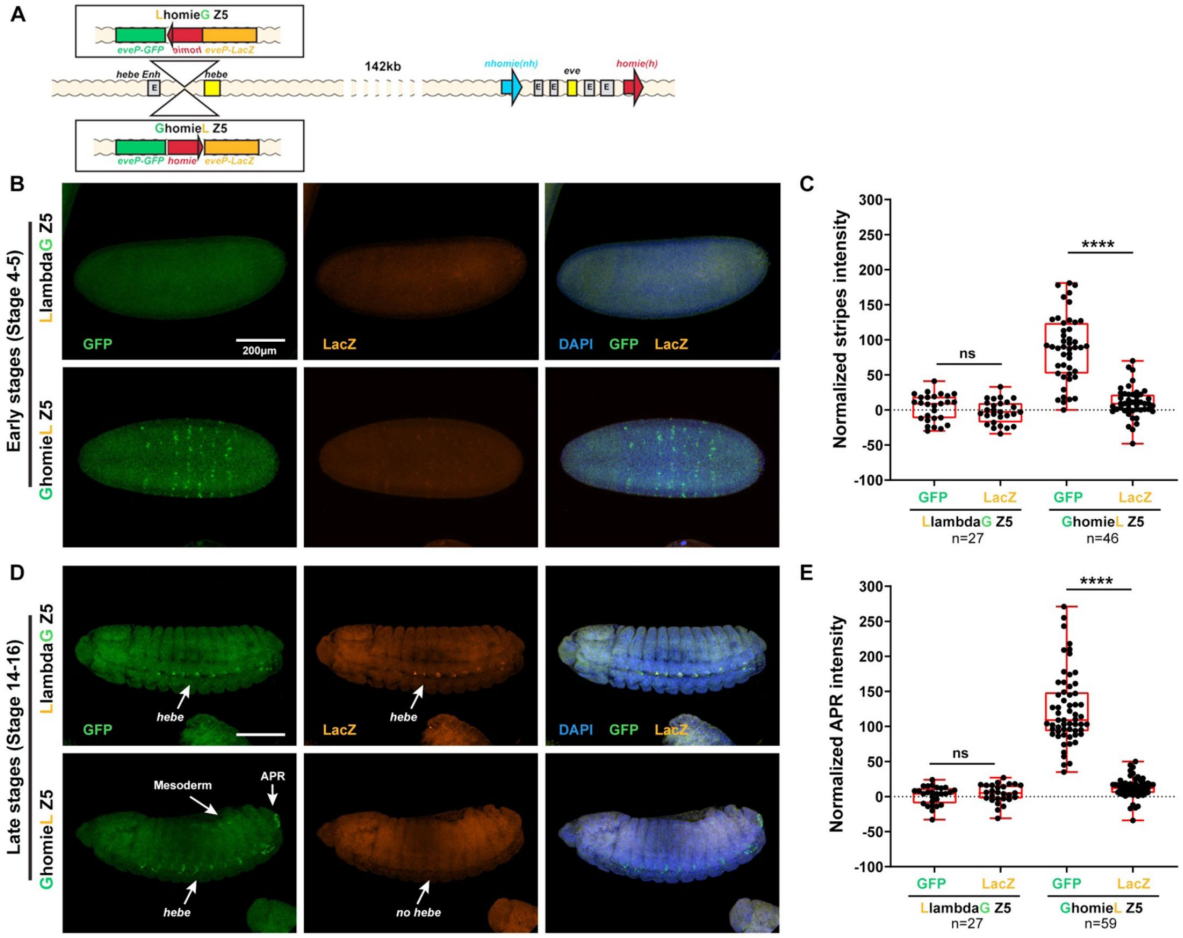

Figure 5.

The *GFP* reporter is activated by both *eve* and *hebe* enhancers in the *LhomieG Z5* transgene.

A) Schematics of transgenes. B) Expression patterns of transgenic reporters in early stage embryos. Top: *LlambdaG Z5*. Bottom: *GhomieL Z5*. *GFP* is in green, *lacZ* is in orange, and DAPI is in blue, here and in D. C) Quantification of normalized stripes signal as represented in B. The reporter smFISH intensity (signal to background) was measured, normalized, and plotted as described in Methods, here and in E. N>3. n=27 for *LlambdaG Z5,* and n=46 for *LhomieG Z5*. D) Expression patterns of transgenic reporters in late stage embryos. Top: *LlambdaG Z5*. Bottom: *GhomieL Z5*. E) Quantification of normalized stripe signals for transgenes shown in D. N>3. n=27 for *LlambdaG Z5*, and n=59 for *LhomieG Z5*. Scale bars = 200µm. N = # independent biological replicates. n = # animals. The paired two-tailed t-test were used for statistical analysis. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001, ns: not significant. Raw measurements are available in the Source Data files.

GhomieL Z5

The situation is more complicated for the *GhomieL Z5* transgene. This transgene is inserted in the same chromosomal orientation as *LhomieG Z5*: *eve-lacZ* is on the *eve* side of the *homie* boundary and *eve-gfp* is on the *hebe* enhancer side of the boundary. However, the orientation of the boundary within the transgene is switched so that *eve-gfp* rather than *eve-lacZ* is "upstream" of *homie*. In the boundary-pairing model, flipping the orientation of *homie* in the transgene but keeping the same orientation of the transgene in the chromosome will have two consequences. First, as illustrated in **Fig. 3C**, the topology of the chromatin loop connecting the *homie* boundary in the transgene and the *eve* locus will change from a stem-loop to a circle-loop. The reason for the change in loop topology is that *homie* pairs with *homie* head-to-head, and with *nhomie* head-to-tail (Fujioka *et al*., 2016). As before, the transgene *homie* will pair with *eve homie* head-to-head to generate a simple circle-loop; however, if, as expected, it also pairs with *nhomie* so that the *eve* TAD remains intact, a more complicated loop structure would be generated (**Fig. 3C**). Second, the reporter that is preferentially activated by the *eve* enhancers will switch from *eve-lacZ* to *eve-gfp*. In order to be activated by the *eve* enhancer, the reporter must be "upstream" of the transgene *homie*. This physical relationship means that the *eve-gfp* reporter will be activated by the *eve* enhancers independent of the orientation of the transgene in the chromosome. On the other hand, the *eve-lacZ* reporter will not be activated by the *eve* enhancers. In addition, it will still be insulated from the *hebe* enhancers by the *homie* boundary.

It is not possible to form a circle-loop in the loop extrusion model. Instead, the breakthrough cohesin complex will come to a halt as before when it encounters the inverted *homie* boundary at –142 kb (**Fig. 3D**). This means that the *eve-lacZ* reporter will be in the same TAD (*eveMa*) as the *eve* enhancers and will be activated by the *eve* enhancers. In contrast, the *eve-gfp* reporter will be in the neighboring TAD and will not be activated by the *eve* enhancers. It will, however, be regulated by the nearby *hebe* enhancers.

Fig. 5 shows that the four predictions of the boundary-pairing model are correct: a) the *eve* enhancers drive *eve-gfp* expression, b) the *hebe* enhancers drive *eve-gfp* expression, c) *eve-lacZ* is not subject to regulation by the *eve* enhancers, and d) *eve-lacZ* is insulated from the *hebe* enhancers. This would also mean that pairing of the transgene *homie* in the *GhomieL Z5* insert with *homie* (and *nhomie*) in the *eve* locus generates a circle-loop, not a stem-loop, as was the case for *LhomieG Z5*. In contrast, since loop-extrusion cannot generate circle-loops, the key predictions of this model, namely that *eve-gfp* will be silent while *eve-lacZ* will be regulated by the *eve* enhancers, are not satisfied. Instead, *eve-gfp* is activated by the *eve* enhancers, while *eve-lacZ* is not.

While the results described above for the *GhomieL Z5* transgene are inconsistent with a "simple" loop-extrusion model, there is one potential caveat: the 5'→3' orientation of the *homie* boundary is inverted so that it is "pointing" towards the *eve* locus rather than away from it like the endogenous *homie* boundary. In this orientation, it is possible that *homie* no longer functions as an effective roadblock, and instead the cohesin complex emanating from the *eve* locus (or from a loading site somewhere in between) slips past this *homie* fragment and continues until it is blocked by a properly oriented boundary further upstream. In this case the newly formed stem-loop, *eveGargantuan* (*eveGa*) (**Fig. 3E**), would include the *eve-gfp* reporter, and it would then be activated by the *eve* enhancers. However, there are two problems with postulating this novel *eveGa* TAD. First, *eve-lacZ* would be in the same *eveGa* TAD as *eve-gfp*, and it should also be activated by the *eve* enhancers. Second, since the *homie* boundary was bypassed by the cohesin complex to form the larger *eveGa* TAD, it should not be able to block the *hebe* enhancers from activating *eve-lacZ*. Neither of these predictions is correct.

GhomieL Z3 and *LhomieG Z3*

To confirm these conclusions, we examined the expression patterns of the *GhomieL* and *LhomieG* transgenes when they are inserted in the opposite orientation (Z3) in the –142 kb attP site. In the boundary-pairing model, the orientation of *homie* in the transgene determines which reporter will be preferentially activated, while the orientation of the transgene *homie* in the chromosome determines whether a stem-loop or a circle loop will be formed. In *GhomieL Z3* a stem-loop will be formed, and *eve-gfp* will be activated by the *eve* enhancers; however, since *eve-lacZ* is adjacent to the *hebe* enhancers, they will activate it rather than the *eve-gfp* reporter (**Fig. 6B**). In *LhomieG Z3* the topology of the loop will switch to a circle-loop, and *eve-lacZ* will be activated by both the *eve* and *hebe* enhancers (**Fig. 6C**). In the first version of the loop-extrusion model, the *eve-gfp* reporter in both *GhomieL Z3* (**Fig. 6D**) and *LhomieG Z3* (not shown), should be activated by the *eve* enhancers, since it is included in the *eveMa* TAD when the transgenes are inserted in the Z3 orientation, while *eve-lacZ* will be regulated by the *hebe* enhancers (**Fig. 6D**). In the second version of the loop-extrusion model, the functioning of the *homie* roadblock depends upon its orientation relative to endogenous *homie*. In this model, *GhomieL Z3* is expected to form the *eveMa* TAD (**Fig. 6D**), while *LhomieG Z3* will form the larger *eveGa* TAD (**Fig. 6E**).

As would be predicted by all three models, the *eve* enhancers activate the *eve-gfp* reporter in the *GhomieL Z3* insert (**Fig. 6F and G**). Likewise, *homie* blocks the *hebe* enhancers from activating the *eve-gfp* reporter, while they turn on *eve-lacZ* expression instead. As predicted by the boundary-pairing model, *eve*-dependent expression of the reporters switches from *gfp* to *lacZ* in the *LhomieG Z3* insert (**Fig. 6F and G**). This would not be expected in the first version of the loop-extrusion model, but it is predicted in the second version. However, as indicated in the diagram in **Fig. 6E**, the two reporters, along with the *hebe* enhancers, are included in the larger *eveGa* TAD, and for this reason both reporters should be activated equally by both the *eve* and *hebe* enhancers. This is not the case. The *hebe* enhancers are blocked by the intervening *homie* boundary, while there is little or no *eve*-dependent *gfp* expression.

TAD formation by transgenes inserted at–142 kb

We used MicroC to examine the TAD organization of the genomic region extending from upstream of the attP site through the *eve* locus in 12-16 hr embryos carrying different transgene insertions. At this stage, *eve* is expressed in only a small number of cells in the CNS, mesoderm and anal plate, and this is also true for *hebe*. This means that most of the interactions detected by MicroC are in cells in which the *eve* gene and the two reporters are inactive. In addition, while the transgene co-localizes with *eve* in less than about a fifth of blastoderm stage nuclei, the frequency of physical contact is expected to be much higher in 12-16 hr embryos (c.f., Vazquez et al. Vazquez *et al*., 2006). This is because at this stage of development, cell cycles are much longer, and homologs are fully paired. The key findings are summarized below.

LlambdaG Z5

While there is no evidence of contact between sequences in *eve* and sequences around the attP site at –142 kb in wild-type embryos, this not true for embryos carrying the dual reporter transgene. Perhaps the most unexpected finding is a weak, but clearly discernable pattern of interaction linking the *LlambdaG Z5* transgene to the *eve* locus (**Fig. 7A**). However, since the *LlambdaG Z5* transgene does not respond to the *eve* enhancers, this physical interaction is not sufficient to drive detectable expression of the reporters. Though our experiments do not allow us to unequivocally identify which sequences in the transgene are responsible for this long-distance interaction, the most likely candidates are the two *eve* promoters in the *LlambdaG Z5* transgene. As noted above, interaction triangles are observed within the *eve* TAD spanning the regions between *nhomie* and *homie* (and/or the *eve* 3' PRE) and the *eve* promoter. It seems possible that the factors supporting

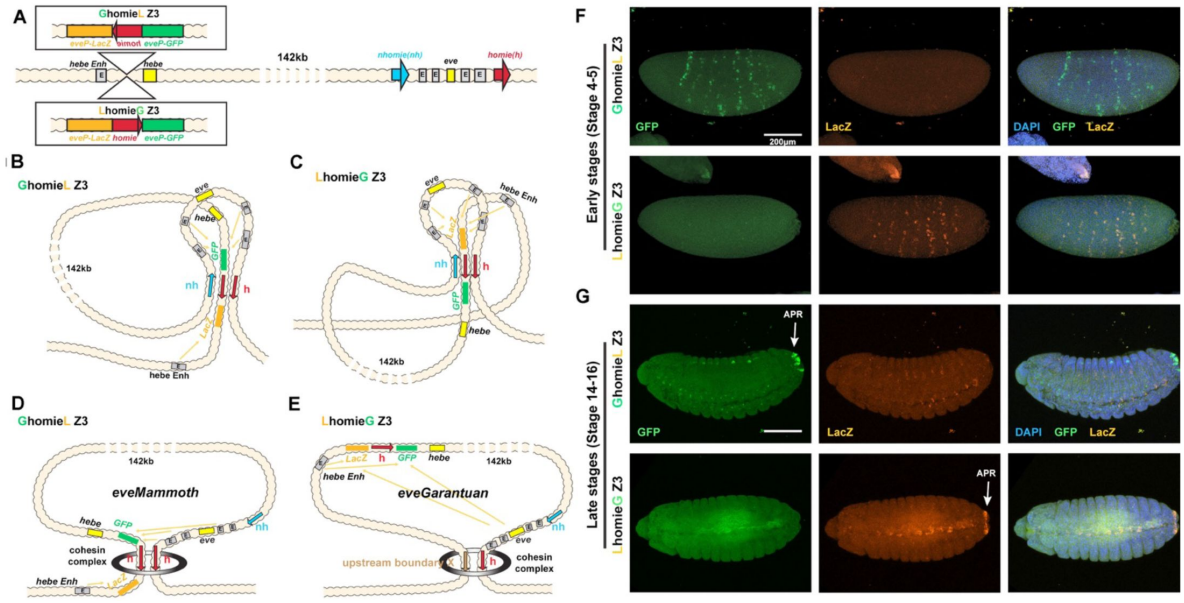

Figure 6.

Expression of reporters in Z3 transgenes.

A) Schematics of Z3 transgenes. B) Boundary pairing for *GhomieL Z3* transgene long-range interactions. In this topology, *GFP* is in close proximity to *eve* enhancers, and *lacZ* is in contact with the *hebe* enhancer region. C) Boundary pairing for the *LhomieG Z3* transgene. In this topology, *lacZ* is in close proximity to both the *eve* enhancers and the *hebe* enhancers. D) Loop extrusion model for *GhomieL Z3*. Transgenic *homie* and endogenous *homie* flank the *eveMammoth* loop. In this topology, *GFP* is in close proximity to *eve* enhancers, and *lacZ* is close to the *hebe* enhancers. E) Loop extrusion model for the *LhomieG Z3* transgene. The cohesin complex bypasses transgenic *homie* and is stopped at an upstream boundary X to create the *eveGarantuan* loop. Both *GFP* and *lacZ* are at a similar physical distance from *eve* enhancers and *hebe* enhancers in this topology. F) Expression patterns of transgenic reporters in early stage embryos. Top: *GhomieL Z3*. Bottom: *LhomieG Z3*. *GFP* is in green, *lacZ* is in orange, and DAPI is in blue. N=3, n=24 for both *GhomieL Z3* and *LhomieG Z3*. **G)** Expression patterns of transgenic reporters in late stage embryos. Top: *GhomieL Z3*. Bottom: *LhomieG Z3*. N=3, n=24 for both *GhomieL Z3* and *LhomieG Z3*. Scale bars = 200µm, N = # independent biological replicates, n = # animals.

these intra-TAD interactions are also responsible for connecting *LlambdaG Z5* to the *eve* locus. In this respect it is interesting to note that *homie* and *nhomie* stimulate enhancer-promoter interactions in a transgene assay (Yokoshi et al., 2020).

LhomieG Z5

Unlike *LlambdaG Z5*, the *eve-lacZ* reporter in the *LhomieG Z5* insert is activated by the *eve* enhancers, and this regulatory interaction is perfectly paralleled by the strong pattern of crosslinking between the transgene and the *eve* locus (**Fig. 7B**). On the *eve* side, the interactions between the transgene and the *eve* locus generate a heavily populated band of crosslinked sequences that span the 16 kb *eve* TAD. On the transgene side, there is an unequal distribution of crosslinked sequences to either side of the transgene *homie* boundary. As shown in the inset, the heaviest density of crosslinked sequences is on the *eve-lacZ* side of the transgene, consistent with the activation of this reporter by *eve* enhancers. To confirm that the interaction bias mimics the differences in activity of the two reporters in cells in which the reporters are expressed, we analyzed the interaction profiles from viewpoints within either the *eve-lacZ* or *eve-gfp* genes. **Fig. 7D** shows that *eve-lacZ* interacts with sequences extending across the *eve* TAD, while *eve-gfp* interactions are largely restricted to the *eve homie* element. While these results would be seemingly be consistent with both the boundary-pairing and loop-extrusion models, there is no evidence of cohesin breaking through multiple intervening boundaries into order to establish the novel *eve(Mam)* TAD. Instead, the organization of the TADs and LDC domains in the interval in between the transgene and *eve* (**Fig. 7B**) resemble that seen in wild-type NC14 embryos (see **Fig. 2**) (with the caveat that there are significantly fewer reads in the transgene experiment). Equally important, the same is true for the *eve* TAD. In order to generate the *eve*(*Mam*) TAD, the cohesin complex must break through the *nhomie* boundary and, in so doing, disrupt the *eve* TAD. However, there are no obvious alterations in either the *eve* volcano or plume with respect to their structure or relative contact density compared to the *LlambdaG Z5* and wild type NC14 embryos. Likewise, there is no evidence of novel vertical and 45° crosslinking stripes that would reflect the boundary breakthrough events that are expected in the loop-extrusion model to accompany the formation of a physical connection between the transgene and the *eve* locus.

We next used the transgene *homie* as the viewpoint to test the predictions of the boundary-pairing and loop-extrusion models. Since *homie* can pair with itself (head-to-head) and with *nhomie* (head-to-tail) in transvection assays (Fujioka *et al*., 2016), a prediction of the boundary-pairing model is that there will be physical interactions between the transgene *homie* and both *nhomie* and *homie* in the *eve* locus (c.f., Vazquez et al., Vazquez *et al*., 2006)). Also based on what is known about how fly boundaries pair with each other (Chetverina *et al*., 2014; Chetverina *et al*., 2017; Kyrchanova *et al*., 2008a), self-interactions are expected to be stronger than heterologous interactions. In contrast, in the loop-extrusion model, the cohesin complex forms the *eveMa* TAD by linking *homie* in the transgene to *homie* in the *eve* locus, and this requires the cohesin complex to break through *nhomie* and disrupt the connections between *nhomie* and *homie* in the *eve* locus. In this case, *transgene:homie*← →*eve:homie* interactions should be observed, while *transgene:homie*← →*-eve:nhomie* should not. **Fig. 8A** shows that, as predicted by the boundary-pairing model, the transgene *homie* interacts with both *homie* and *nhomie* in the *eve* locus, and the cross-linking frequency is greater for *homie:homie* interactions.

Since the *eve* TAD does not appear to be disturbed by the presence of the *LhomieG Z5* transgene, these findings would suggest that the transgene *homie* likely pairs with both *nhomie* and *homie* in a pattern somewhat like that illustrated in **Fig. 3B**. While the interactions with the transgene are inconsistent with the loop-extrusion model, one could make the *ad hoc* assumption that runaway cohesin complexes form not one but two different novel TADs. One would be *eveMa*. In the other, *eveElephant* (*eveEl*), a cohesin complex that initiates in, for example, the TE TAD would come to a halt on one side at the *homie* boundary in the transgene, and on the other side at the *nhomie* boundary in the *eve* locus (Fig. Supplemental 4). However, this configuration would not explain

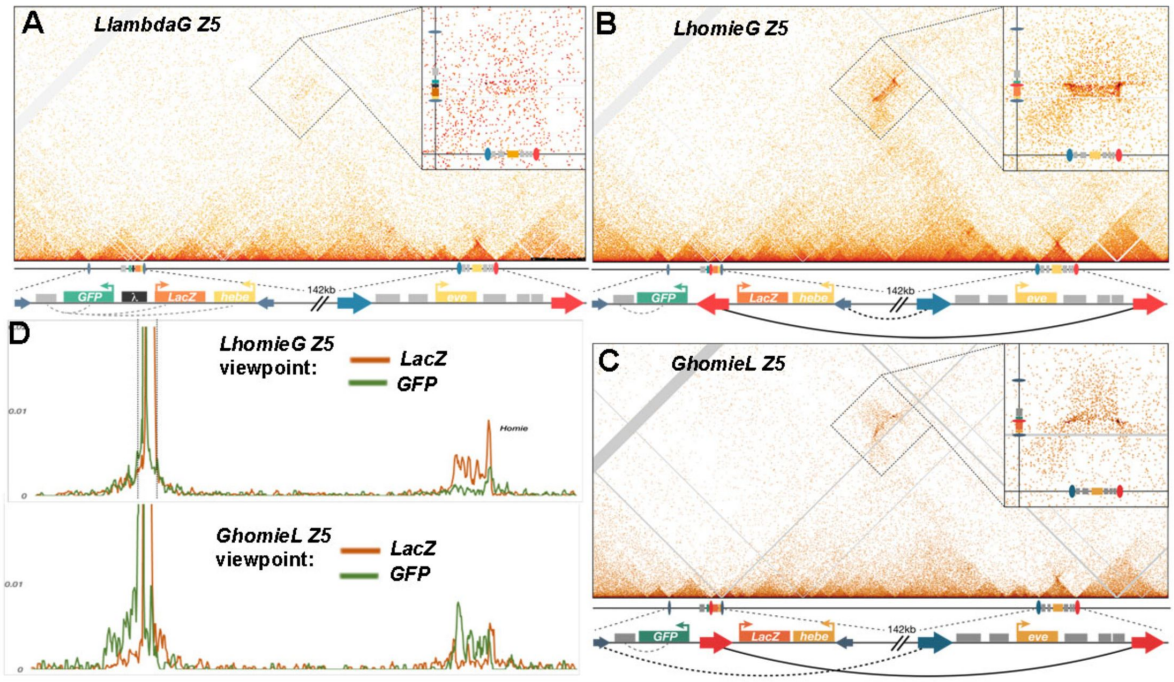

Figure 7.

homie in Z5 transgene *LhomieG* and *GhomieL* Z3 mediates long-range interactions with the *eve* locus

Small dark blue arrows – insulators upstream and downstream of *hebe* that appear important to demarcate interaction domains between –142kb and *eve*. *Large blue arrow* – endogenous *nhomie*. *Large red arrow* – *homie*, either endogenous or in the transgene. The direction of arrows follow established convention on *nhomie/homie*, and do not reflect orientation of insulator protein binding motifs *per se*. *Gray boxes* – enhancers. A) MicroC map of the control line *LlambdaG Z5*. Scaled cartoons of the two loci of interest are shown directly below the MicroC map, and an unscaled blowup of the elements of interest at each locus is provided. Within the MicroC map of the entire locus, a zoom-in of the off-diagonal interaction between the –142kb insertion cassette and the endogenous *eve* locus is shown. Note a slight increase in interaction frequency (compare to Figure 1A). B, C) MicroC map of *LhomieG Z5* and *GhomieL Z5*, respectively. The only difference between the two lines is the orientation of transgenic *homie*, as indicate below each MicroC map. Blow up of the off-diagonal interaction between –142kb and *eve*, including scaled cartoons denoting features of interest in the top right corner. Note the changed pattern of interaction with the endogenous *eve* locus due to the orientation switch. D) "Virtual 4C" maps obtained from MicroC maps of "B" (top panel), and "C" (bottom panel). Viewpoints are shown in both panels from either the *lacZ* gene (orange) or the *GFP* gene (green). Note in each cassette the altered interaction of either gene with endogenous *eve*, depending on the change of orientation of *homie*.

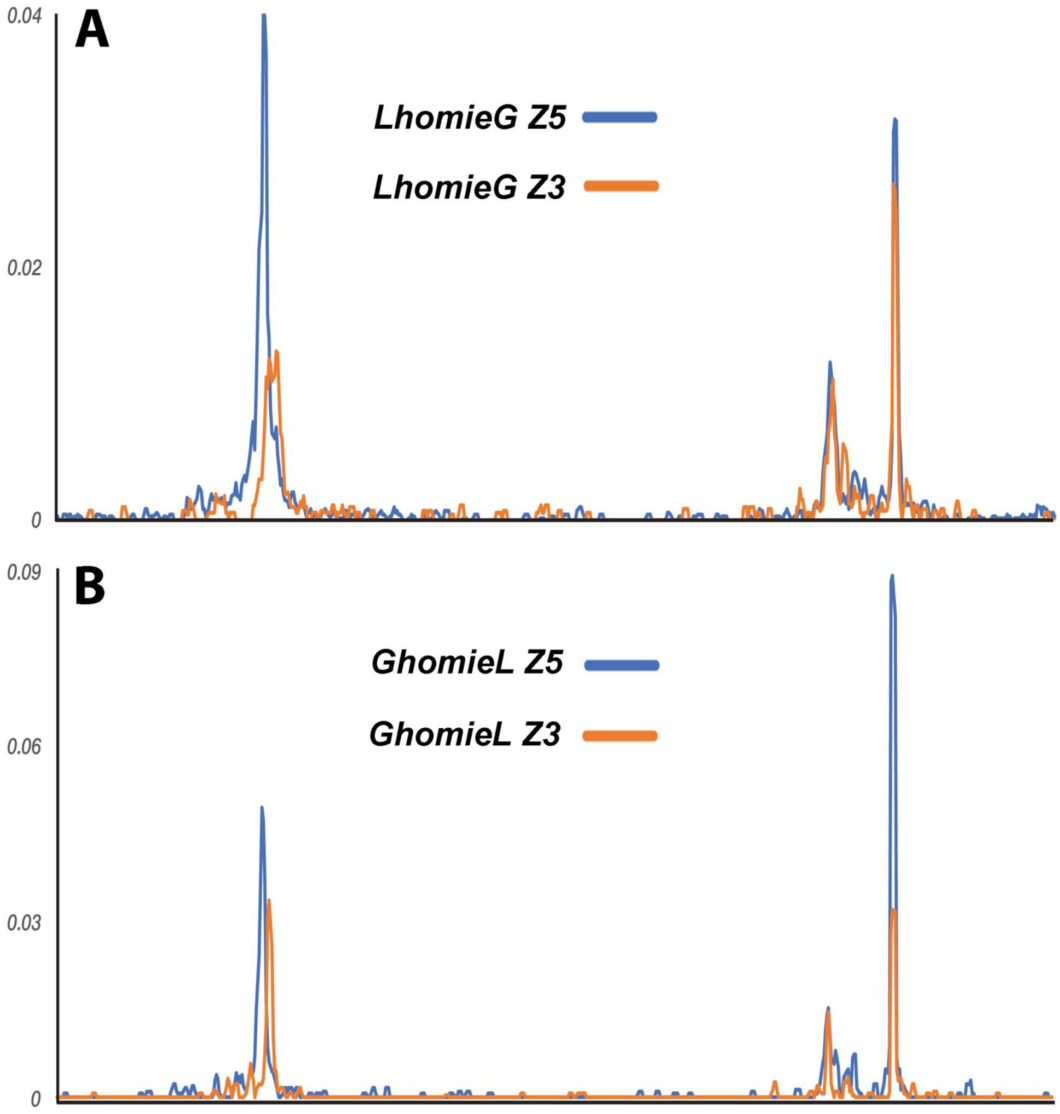

Figure 8.

***homie* in the transgene interacts with *eve homie* and *nhomie*.** "Virtual 4C" maps of the *LhomieG* (A) or *GhomieL* (B) transgenes, for the Z5 (blue) and Z3 (orange) orientations. Viewpoints are taken from the transgenic *homie* sequence. Importantly, note that transgenic *homie* preferentially interacts with endogenous *homie* no matter the orientation. Intriguingly, the orientation switch of transgenic *homie* leads to an apparent change in its degree of preference for endogenous *homie* over *nhomie*.

how the transgene reporter is activated by *eve* enhancers. Since the *eve* enhancers in *eveEl* are located in a different TAD than the TADs containing *eve-lacZ* and *eve-gfp*, they would not be able to activate either reporter.

GhomieL Z5

When *homie* is reversed in the *GhomieL Z5* transgene, the *eve-gfp* reporter responds to the *eve* enhancers instead of *eve-lacZ*. In the boundary-pairing model, *eve-gfp* is activated because the self- and heterologous-pairing interactions between *nhomie* and *homie* are orientation-dependent. As a result, the topology of the loop formed between the transgene and *eve* switches from a stem-loop to a circle-loop, and this brings the *eve-gfp* reporter in close proximity to the *eve* enhancers instead of the *eve-lacZ* reporter. Since circle-loops cannot be formed in the loop-extrusion model, activation of the distal *eve-gfp* reporter would require the formation of a larger *eveGa* TAD in order to encompass the reporters in the transgene and the *eve* enhancers in the same looped domain (**Fig. 3E**). However, as shown in **Fig. 7C**, we do not observe the cross-linking pattern expected for a new *eveGa* TAD that includes the entire transgene at –142 kb and then extends to an even more distal boundary. Nor are there any novel vertical/45° crosslinking stripes or obvious perturbations in the basic TAD organization in the region beyond –142 kb, or between –142 kb and the *eve* locus, as might be expected for cohesin complexes breaking through multiple intervening roadblocks. Instead, as was the case for *LhomieG Z5*, the *GhomieL Z5* transgene is physically linked to the *eve* locus, and there is a prominent band of crosslinked transgene sequences that spans the *eve* TAD (**Fig. 7C**). Consistent with the expression patterns of the two reporters, the heaviest density of crosslinked sequences is on the *eve-gfp* side of the transgene (see inset). This is confirmed by the interaction profiles generated using either the *eve-gfp* or *eve-lacZ* gene body as the viewpoint: the most frequent contacts are between sequences within *eve* and the *eve-gfp* reporter (**Fig. 7D**). In addition, as was the case for *LhomieG Z5*, when the transgene *homie* is used as the viewpoint, interactions are observed between both *homie* and *nhomie* in the *eve* locus, with the most frequent corresponding to self-pairing interactions (**Fig. 8B**). Other than postulating the *ad hoc eveEl* TAD (Fig. Supplemental 4), this would not be expected for either version of the loop-extrusion model.

LhomieG Z3 and *GhomieL Z3*

To confirm the findings for transgenes inserted in the Z5 orientation, we used MicroC to analyze the TAD organization of *LhomieG Z3* and *GhomieL Z3*. (Sup. Figs. 5 and 6). Like their Z5 counterparts, both show a strong band of interaction between the transgene and sequences spanning the *eve* locus. In the case of *GhomieL Z3*, which is predicted to form a stem-loop, bringing *eve-gfp* in contact with the *eve* enhancers, the *eve-gfp* sequences interact more frequently with the *eve locus* (see inset in Fig. Supplemental 5A). *homie* is in the opposite orientation in the transgene and in the chromosome in *LhomieG Z3*, and in the boundary pairing model is expected to generate a circle-loop. While the sample preparation for MicroC was of poorer quality than the others, the inset in Sup. Fig.6A shows that, as expected, *eve-lacZ* interacts more frequently with sequences in the *eve* locus than *eve-gfp*. On the other hand, there is no evidence of a larger *evaGa* TAD even though *eve-lacZ* is activated by the *eve* enhancers. These interactions are confirmed when either *eve-lacZ* or *eve-GFP* are used as viewpoints to plot the physical contacts between the *eve* TAD and the two reporters in *LhomieG Z3* and *GhomieL Z3* (Sup. Figs. 5B and 6B). In addition, as was the case for the two Z5 transgenes, when the transgene *homie* is used as the viewpoint, interactions are observed between *homie* in the transgene and *both homie* and *nhomie* in the *eve* locus (**Fig. 8A and B**). Again, the most frequent interactions are self-pairing interactions between *homie* in the transgene and the *eve homie*.

Loop topology and interactions with neighbors

The results in the previous sections demonstrate that the topology of the loop formed between the transgene and the *eve* locus depends upon the orientation of *homie* in the transgene *and* the orientation of the transgene in the chromosome. When *homie* in the transgene is pointing away from the *eve* locus, the topology is a stem-loop. When *homie* in the transgene is pointing towards the *eve* locus, the topology is a circle-loop.

Changing the topology also impacts how sequences flanking the transgene and flanking *eve* interact with each other. With the important caveat that the depth of our MicroC sequencing is limited, the interaction profiles reflect the topology of the transgene-induced loop. For *homie* pointing away from *eve* (*LhomieG Z5* and *GhomieL Z3*), sequences to the *TER94* side of *eve* and to the *hebe* enhancer side of the transgene are more frequently crosslinked with each other (Fig. Supplemental 7). These interactions would be expected for a stem-loop structure (**Fig. 3B**). For *homie* pointing towards *eve*, sequences on the *TER94* side of *eve* interact with sequences on the *eve* side of the transgene. This would be expected for a circle-loop structure (**Fig. 3C**).

Discussion

Two different mechanisms have been proposed to explain how TADs are formed and their endpoints determined. In the first, a cohesin complex first induces a small loop at a loading site within the TAD-to-be and then progressively extrudes one or both strands until the complex encounters boundary roadblocks. If the loading site is located asymmetrically within the TAD-to-be, extrusion will continue on the unblocked strand until a roadblock is encountered. The endpoints of the loops were initially thought to be determined by the location of the first two roadblocks encountered by the cohesin complex (Alipour and Marko, 2012). In a refinement of this model, the orientation of a given roadblock relative to the direction of movement of the cohesin complex determines whether it will come to a halt (Davidson and Peters, 2021; Fudenberg *et al*., 2016; Guo *et al*., 2012; Guo *et al*., 2015; Sanborn *et al*., 2015). It has also been suggested that cohesin complexes can break through a roadblock and come to a halt at more distant roadblocks, so that in a population of cells analyzed by Hi-C or MicroC, the basic units of organization, the TADs, are overlaid by a hierarchy of LDC domains (Hsieh *et al*., 2020; Krietenstein *et al*., 2020). However, the only loop topologies that are possible in the loop-extrusion model are a stem-loop and an unanchored loop.

In the other mechanism, TAD formation is governed by boundary:boundary pairing interactions. In this case, the critical factors are the pairing and orientation preferences of potential partners, together with proximity. Since boundary pairing interactions often exhibit an orientation preference, the topology of the loop formed by paired boundaries can be either a stem-loop or a circle-loop, depending on their relative orientation in the chromosome. In this respect it differs from the loop-extrusion model in that switching the orientation of the boundary does not preclude the formation of a TAD, whereas it can in some versions of the loop-extrusion model.

In the studies reported here, we have used MicroC to analyze the TAD organization of a ~150 kb chromosomal segment flanking the *eve* locus, and have experimentally tested the predictions of these two models for the mechanisms of TAD formation. Our studies do not support a loop-extrusion mechanism, and are inconsistent with this model on multiple levels.

Manipulating "TAD" formation

In our experimental paradigm, a transgene containing two divergently transcribed reporters and the *eve* boundary *homie* is inserted 142 kb upstream of the *eve* TAD. Depending upon the orientation of the *homie* boundary within the transgene, either *eve-gfp* or *eve-lacZ* is activated by developmentally regulated enhancers in the *eve* TAD. Since there are a series of well-defined TADs and their associated boundary elements separating the transgene from *eve*, a version of the loop-extrusion model in which the cohesin complex invariably arrests upon encountering *nhomie* and *homie* at the ends of the *eve* TAD cannot account for this long-distance regulation. Instead, the extruding cohesin complex must be able to break through the boundaries located between *homie* in the transgene and *homie* in the *eve* TAD. This would generate a novel TAD, *eveMa*, with endpoints corresponding to the two *homie* boundaries (**Fig. 3D**). In this case, the reporter on the *eve* side of the transgene *homie* would be in the same domain as the *eve* enhancers, and thus potentially subject to regulation. While *eveMa* could account for reporter expression in inserts in which the transgene *homie* is pointing away from *eve* (and also away from the reporter on the *eve* side of the transgene insert), it does not explain reporter expression when the transgene *homie* is pointing towards *eve* (and in this case towards the reporter on the *eve* side of the transgene insert). In the latter case, the reporter activated by the *eve* enhancers lies outside of the *eveMa* (cf., **Fig. 5D**), and thus is not included in the same TAD as the *eve* enhancers. In order to account for activation of the distal reporter, one must imagine that the cohesin complex breaks through the *homie* boundary in the transgene and continues on until extrusion is halted by an even more distal boundary. In this revision of the loop-extrusion model, both reporters would be included in the same TAD, *eveGa* (c.f., **Fig. 5E**), as the *eve* enhancers, and both reporters should be activated. However, this is not observed.

TAD organization: loop-extrusion versus boundary pairing

Analysis of the TAD organization in the region spanning the transgene insertion site and *eve* also does not fit with the expectations of the loop-extrusion model. To account for reporter activation, the cohesin complex must be able to break through the *nhomie* boundary. This breakthrough could be engineered by a cohesin complex that initiated loop extrusion either from within the *eve* TAD or from one of the TADs located between the *nhomie* boundary and the transgene at –142 kb. However, in the absence of the transgene, there is no indication that *nhomie* is particularly prone to break-through events. In fact, the *eve* TAD differs from most of the TADs in between *eve* and the transgene in that it is incorporated into larger LDC interaction triangles much less frequently. There is also no reason to imagine that the presence of a *homie* (or *nhomie*: see Fujioka *et al.*, 2016)) transgene at –142 kb would somehow induce *nhomie* break-through events.

As expected from the regulatory interactions, *homie*-containing transgenes at –142 kb generate a strong band of crosslinking events that physically link the transgene to the *eve* locus. In the loop-extrusion model, this strong band of interaction should generate a novel interaction triangle (LDC or HDIC) with endpoints corresponding to the *homie* elements in the transgene and *eve*. Contrary to this expectation, a triangle of increased interaction frequencies spanning the entire region between –142 kb and *homie* is not observed for any of the transgene inserts. In viewpoints from the *homie* element in the *LhomieG Z5* and *GhomieL Z3* transgenes, there are two peaks in the *eve* TAD. The more prominent peak maps to *homie*, while the less prominent peak maps to *nhomie*. To account for these interaction peaks, the loop-extrusion model would have to assume that the presence of the *LhomieG Z5* and *GhomieL Z3* transgenes induces the formation of two novel TADs, the *eveMa* (*homie:homie*) TAD in **Fig. 5** and another TAD, *eveElephant*, that links the transgene *homie* to the *nhomie* element in the *eve* locus (Fig. Supplemental 4). However, while the *eveMa* TAD could potentially explain the activation of reporters on the proximal side of the transgene *homie* by the *eve* enhancers, these TADs do not explain how the *eve* enhancers are able to activate reporters on the distal (*hebe* enhancer) side of the *GhomieL Z5* and *LhomieG Z3* transgene. This would require the formation of yet another TAD, *eveGa*, that encompasses *eve*, the transgene, and one or more TADs beyond the *hebe* gene (**Fig. 3E**). However, there is no indication of physical

interactions between the *homie* boundary in *eve* and a boundary located beyond the *hebe* gene. Moreover, like *eveMa*, *eveGa* would also have to be accompanied by the formation of *eveEl*, since the *homie* element in *GhomieL Z5* and *LhomieG Z3* interacts with both *nhomie* and *homie*.

While these results are inconsistent with the expectations of the loop-extrusion model, they dovetail nicely with many of the predictions of the boundary-pairing model. First, in all four transgenes, *homie* interacts with both *homie* and *nhomie* in the *eve* locus, but does not interact with the multiple boundary elements located in between. This is consistent with other studies which show that boundaries often have strong partner preferences (Gohl *et al*., 2011; Kyrchanova *et al*., 2011; Kyrchanova *et al*., 2008b). Second, interaction frequencies are greater for self-pairing than for heterologous pairing. This fits with what is known about how partner preferences are determined (Blanton et al., 2003; Chetverina *et al*., 2017; Erokhin et al., 2021; Kyrchanova *et al*., 2008a; Zolotarev *et al*., 2016). Third, the pattern of reporter activation in the four transgenes is precisely as expected from previous studies on the orientation dependence of *homie* pairing with itself (head-to-head) and with *nhomie* (head-to-tail) (Chen *et al*., 2018; Fujioka *et al*., 2016). This orientation dependence means that the reporter located "upstream" of *homie* is the one that is activated, independent of the orientation of the transgene in the chromosome. Fourth, the pattern of physical interactions between the *eve* locus and the transgenes recapitulates the pattern of reporter activation. That is, when *eve-lacZ* sequences are preferentially crosslinked to sequences in the *eve* locus, then the *eve-lacZ* reporter is activated by the *eve* enhancers. Likewise, *eve-gfp* is preferentially crosslinked to sequences in *eve* when it is activated by the *eve* enhancers. Again this is dependent on the orientation of *homie* in the transgene, but not on the orientation of the transgene in the chromosome. Fifth, unlike the loop-extrusion model where only stem-loops or disconnected loops can be generated, both stem-loops and circle-loops can be, and, as we have shown here, actually are generated by boundary:boundary pairing. The topology of the loop formed between the transgene and *eve* depends on the orientation of the transgene *homie* in the chromosome. When *homie* in the transgene insert is pointing away from the *eve* locus, a stem-loop is generated. When *homie* in the transgene insert is pointing towards the *eve* locus, a circle-loop is generated. That both loop topologies can be generated by boundary:boundary pairing is confirmed by not only by the pattern of expression of the two reporters in the four transgene inserts but also by their pattern of physical interactions evident in our MicroC experiments.

Are the pairing interactions of homie and nhomie unusual?

There are two reasons to think that the properties of most fly boundaries are similar to those of *homie* and *nhomie*. First, all of the fly boundaries that have been tested in assays that are expected to require pairing do have this activity (Blanton *et al*., 2003; Cai and Shen, 2001; Erokhin et al., 2011; Gohl *et al*., 2011; Kyrchanova *et al*., 2008a; Kyrchanova *et al*., 2011; Kyrchanova et al., 2007; Kyrchanova *et al*., 2008b; Muravyova *et al*., 2001). Second, multimerized binding sites for generic polydactyl zinc finger proteins like Su(Hw), Pita, dCTCF, Zw5, and Zipic that are found associated with many fly boundaries are not only able to block enhancer/silencer action but can also mediate pairing interactions in insulator bypass and transvection assays (Erokhin *et al*., 2021; Kyrchanova *et al*., 2008a; Zolotarev *et al*., 2016). Moreover, as is the case for *homie* and *nhomie*, pairing interactions require the appropriate partners: multimerized sites for Zw5 do not pair with multimerized sites for dCTCF (Erokhin *et al*., 2021; Kyrchanova *et al*., 2008a; Zolotarev *et al*., 2016). This makes good sense, since these proteins form dimers/multimers, as do other fly boundary factors like BEAF, Mod(mdg4), the BEN-domain protein Insensitive and the Elba complex (Avva and Hart, 2016; Bonchuk et al., 2021; Bonchuk et al., 2011; Bonchuk et al., 2020; Fedotova et al., 2018; Fedotova et al., 2019; Fedotova et al., 2017; Hart et al., 1997). The ability to multimerize means that boundaries that share binding sites for the same protein, i.e., for Pita, Zipic or CTCF, can be physically linked to each other by a single protein complex.

Boundary pairing orientation and loop topology

Our analysis of the TAD organization in the *eve* neighborhood suggests that the predominant loop topology may be a circle-loop, not a stem-loop. Only the *eve* TAD (and perhaps the very small TAD, TA) is clearly a stem-loop. It is topped by a distinctive volcano plume and is not overlaid by a hierarchy of LDC domains, indicating that it must be "insulated" from crosslinkable interactions with other nearby TADs. By contrast, none of the TADs in the region spanning *eve* and the *hebe* gene have volcano plumes at their apex. Instead, all of these TADs are overlaid by a hierarchal organization of LDC domains (clouds) that diminish in crosslinking intensity as the number of in-between TADs increases. As discussed above, the contact pattern of the TADs to the left of *eve* are likely to arise from partner switching and/or head-to-head boundary pairing interactions that generate circle-loops. Pairing orientation versus loop topology is considered further in the accompanying paper (Ke *et al*.).

Orientation and long-distance regulatory interactions

Our experiments highlight an important constraint on long-distance regulatory interactions that are mediated not only by boundary elements like *homie* and *nhomie* but also by the GAF-dependent tethering and bypass elements (Batut et al., 2022; Kyrchanova et al., 2023; Kyrchanova et al., 2018; Kyrchanova et al., 2019a; Kyrchanova et al., 2019b; Levo et al., 2022; Li et al., 2023; Mohana et al., 2023), namely the orientation of the physical interactions that bring distant enhancers (or silencers) into close proximity with their gene targets. As we have shown here, the orientation of the physical interactions that bring distant sequences close together determines whether the enhancer will be able to activate its target gene. In principle the effects of orientation could be mitigated by increasing the distance between the boundary/tethering elements and either the enhancers (silencers) or the promoter. However, distance can be a critical if not a rate limiting step in transcriptional activation. For example, live imaging experiments have shown that the frequency and amplitude of transcriptional bursts depend on the distance between the enhancer and target promoter. Nearly continuous bursting and high amplitude transcription was observed when the enhancer was located close the promoter (Fukaya et al., 2016). However, at distances as little as 6 kb, transcription became discontinuous with discrete bursts, while at 9 kb, the bursts were much less frequent, and the amplitude/output was substantially reduced (Yokoshi *et al*., 2020). If similar distance-dependent reductions in transcriptional efficiency are common, then it would not be possible to overcome the topological constraints on regulatory interactions by simply increasing the distance between the boundary/tethering elements and the enhancers/promoters. While our studies do not directly address how chromosome architecture impacts regulatory interactions in vertebrates, it seems clear that long-distance regulatory interactions in vertebrates (c.f., Tan et al., 2023) will be subject to the same types of topological constraints as those that apply in flies.

Acknowledgements

We would like to thank the Gordon Grey for running the fly food facility at Princeton, members of the Lewis Sigler Genomics Core facility for their invaluable assistance with DNA sequencing and Sergey Ryabichko for his help with smFISH probes. We would also like to thank members of MOL431 for creative input. Special thanks to Olga Kyrchanova, Daria Chetverina, Maksim Erokhin, Pavel Georigev, Tsutomu Aoki, Girish Deshpande, Airat Ibragimov, Sergey Ryabichko, Yuri Pritykin and Alex Ostrin for stimulating discussions and sharing unpublished data.

Materials and methods

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
n-Heptane	Fisher Chemical	O3008-4
Paraformaldehyde 20% solution, EM Grade	Electron Microscopy Sciences	15713-S
Formaldehyde, 16%, methanol free, Ultra Pure	Polysciences Inc	18814-10
PBS - Phosphate-Buffered Saline (10X) pH 7.4, RNase-free	Thermal Fisher	AM9624
Tween 20	Sigma	P1379
Triton X-100	Bio-Rad	161-0407
Tris base	Sigma	11814273001
Methanol	Fisher Chemical	203403
SSC, 20X	Thermal Fisher	15557044
Formamide	Thermal Fisher	17899
Dextran sulfate	Sigma	D8906
Salmon Sperm DNA	Thermal Fisher	AM9680
Ribonucleoside Vanadyl Complex	NEB	S1402S
Nuclease-free BSA	Sigma	126609
Triethylammonium Acetate	Sigma	625718
dGTP (100 MM)	VWR	76510-208
dTTP (100 MM)	VWR	76510-224
Lonza NuSieve 3:1 Agarose	Thermal Fisher	BMA50090
T4 DNA ligase	NEB	M0202L
Biotin-11-dCTP	Jen Bioscience	NU-809-BIOX

Biotin-14-dATP	Jen Bioscience	NU-835-BIO14
Qubit dsDNA HS Assay Kit	Life Technologies Corporation	Q32851
Atto 633 NHS ester	Sigma	01464
Phase Lock Gel™, QuantaBio - 2302830, Phase Lock Gel Heavy	VMR	10847-802
NEBNext Ultra II DNA Library Prep Kit for Illumina	NEB	E7645S
Ampure Xp 5ml Kit	Thermal Fisher	NC9959336
Hifi Hotstart Ready Mix	Thermal Fisher	501965217
Dynabeads MyOne Streptavidin C1	Life Technologies Corporation	65001
cOmplete™, EDTA-free Protease Inhibitor Cocktail	Sigma	11873580001
N,N-Dimethylformamide	Sigma	227056
Potassium acetate solution	Sigma	95843
DSG (disuccinimidyl glutarate)	Thermal Fisher	PI20593
T4 Polynucleotide Kinase - 500 units	NEB	M0201S
DNA Polymerase I, Large (Klenow) Fragment - 1000 units	NEB	M0210L
End-it DNA End Repair Kit	Thermal Fisher	NC0105678
Proteinase K recomb. 100 mg	Sigma	3115879001
Nuclease Micrococcal (s7)	Thermal Fisher	NC9391488
EGS (ethylene glycol bis(succinimidyl succinate))	Thermal Fisher	PI21565
Atto 565 NHS ester	Sigma	72464
Oligonucleotides		
WK_eve_probeset Sequence see “Supplemental Data S1”	Biosearch Technologies	

WK_LacZ_probeset Sequence see “Supplemental Data S1”	Biosearch Technologies	
WK_GFP_probeset Sequence see “Supplemental Data S1”	Biosearch Technologies	
NEBNext Multiplex Oligos for Illumina	NEB	E7335S
Software and algorithms		
Fiji (ImageJ)	Schindelin et al., 2012	fiji.sc
NIS element	Nikon	microscope.healthcare.nikon.com/products/software/nis-elements
Graphpad Prism 8	GraphPad Software	graphpad.com
HiGlass	Kerpedjiev et al., 2018	higlass.io
bwa	Li H. and Durbin R., 2009	bio-bwa.sourceforge.net/
samtools	Github/open source	samtools.github.io
pairsamtools	Github/open source	github.com/mirnylab/pairsamtools
pairix	Github/open source	github.com/4dn-dcic/pairix
cooler	Abdennur, N., and Mirny, L., 2019	github.com/mirnylab/cooler
Miniconda	Anaconda	docs.conda.io/en/latest/miniconda.html
Snakemake	Github/open source	snakemake.github.io

Plasmid construction and transgenic lines

The dual reporters construct was described previously (Fujioka *et al.*, 2016). In short, both reporters contain the *eve* basal promoter (-275 to +106 bp relative to *eve* start site), the *lacZ* (*eve-lacZ*) or *EGFP* (*eve-GFP*) coding region, and the *eve* 3’ UTR (+1300 to +1525 bp). These two reporters are divergently transcribed. The 367 bp *homie* fragment, or a 500 bp fragment from *lambda* phage DNA, was placed between the two reporters. The *homie* fragment was placed in both orientations: in *LhomieG*, the *eve-lacZ* reporter is located upstream of the *homie* fragment, while in *GhomieL*, *eve-GFP* is located upstream of the *homie* fragment (see **Fig 2A**).

The -142 kb attP landing site was described previously (Fujioka *et al.*, 2009). The -142 kb site contains two attP target sites for phiC31 recombinase-mediated cassette exchange (RMCE) (Bateman et al., 2006) and *mini-white* as a marker. RMCE can result in either orientation of an

inserting transgene (Z5 and Z3; see text). RMCE events were identified by loss of *mini-white*, and the orientation of each insert was determined by PCR.

smFISH probe preparation

The sequences of target genes were obtained from Flybase (*flybase.org*)(Gramates et al., 2022). To design probes, the target gene sequences were submitted to the Biosearch Technologies Stellaris RNA FISH probe designer tool (*biosearchtech.com*), with parameters set to 20nt probes, 48 probes per gene, and mask level 5 for *Drosophila melanogaster*. The sequence of the probes used are listed in Supplemental Data S1. The probes with 3' MdC(TEG-Amino) modifications were then ordered from Biosearch Technologies in 96-well plate format, cartridge purification and salt free conditions and in 5 nmol scale. To prepare the fluorophore for probe labeling, the Sigma Atto NHS ester fluorophore (Atto 565 NHS or Atto 633 NHS) were dissolved in dimethylformamide at 5mg/mL as the stock solution. For probe labeling, 1 nmol of each probe was combined and mixed with Atto NHS ester fluorophore at 1mg/mL in 500µL of 0.1M sodium bicarbonate (pH 8.0-8.5) for 2 hours, and the probes were then precipitated with 0.5M potassium acetate (pH 5.2) in 75% EtOH. After that, the probes were dried and resuspend in 50µL of DePC-treated H2O. From there, the probes were injected into HPLC in which the percentage 0.1M triethylammonium acetate varied from 7% to 30% with a flow rate of 1mL/min. The detector to monitor coupled probes during the HPLC purification were set to 260nm for DNA and the absorption wavelength of the coupled fluorophore. The coupled probes were collected when both 260nm and the absorption wavelength channel were detected. The coupled probes were dried and diluted into appropriate concentration in DePC-treated H2O for smFISH experiments.

smFISH

smFISH methods were adopted and modified from previous studies (Little and Gregor, 2018; Trcek et al., 2017). 100-200 flies were placed in a cage with an apple juice plate at the bottom of the cage. Early stages of embryos were for 7hours, while for later stages embryos, collection was overnight. Embryos from each plates were washed into collection mesh and dechorionated in bleach for 2min, then fixed in 5mL of 4% paraformaldehyde in 1X PBS and 5mL of heptane for 15min with horizontal shaking. The paraformaldehyde were then removed and replaced with 5mL methanol. The embryos were then devitellinized by vortexing for 30s, and washed in 1mL of methanol twice. Methanol were then removed and replaced by PTw(1X PBS with 0.1% Tween-20) through serial dilution as 7:3, 1:1, and 3:7 methanol:PTw. The embryos were washed twice in 1mL of PTw and then twice in 1mL smFISH wash buffer (4X SSC, 35% formamide and 0.1% Tween-20). Embryos were incubated with ~5nM coupled smFISH probes in hybridization buffer (0.1g/mL dextran sulfate, 0.1mg/mL salmon sperm ssDNA, 2mM ribonucleoside vanadyl complex, 20µg/mL RNAse-free BSA, 4X SSC, 1% Tween-20 and 35% formamide) for 12-16h. Embryos were then washed twice for 2 hours in 1mL smFISH wash buffer, followed by 4X washing in 1mL PTw. For DAPI/Hoechst staining, the embryos were stained with 1ug/mL DAPI or Hoechst in PTw, and washed 3X with 1mL PTw. Finally, the embryos were mounted on microscope slides with Aqua PolyMount and #1.5 coverslip for imaging.

Imaging, image analysis and statistics

Embryos from smFISH were imaged by using a Nikon A1 confocal microscope system, Plan Apo 20X/0.75 DIC objective. Z-stack images were taken with interval of 2µm, 4X average, 1024X1024 resolution, and the appropriate laser power and gain were set for 405, 561, and 640 channels to avoid overexposure. Images were processed by ImageJ and the maximum projection was applied to each of the stack images. To measure the stripe intensity of early embryos, multi-channel images were first split into single channels and the stripe signal was highlighted and selected by the MaxEntropy thresholding method. For APR intensity measurement, the ROI tool was used to crop out the APR region of late-stage embryos. The cells with APR signal were also highlighted and selected by MaxEntropy thresholding. The particle measurement tool was used to measure the

average intensity of all cells that have the signal. At the same time, the background signal (average intensity) was measured on cells without a signal in the same embryo. The relative intensity (signal to background) for each embryo was calculated by using the stripe signal and the background from the same embryo. To make comparisons between independent biological replicates, the average background signal of all embryos from each replicate was calculated. The relative intensity of each embryo from each replicate were normalized based on the average background signal of all embryos from same replicates. GraphPad Prism was used for data visualization and statistical analysis. To compare intensity from embryos in different groups, different signals from the same embryo (e.g., *lacZ* and *GFP*) were paired and paired two-tailed t-tests were used to calculate p-values. All raw measurements and normalized data are included in Supplemental Data S2.

MicroC library construction

Embryos were collected on yeasted apple juice plates in population cages for 4 hours, incubated for 12 hours at 25°C, then subjected to fixation as follows. Embryos were dechorionated for 2 mins in 3% sodium hypochlorite, rinsed with deionized water, and transferred to glass vials containing 5 mL PBST (0.1% Triton-X100 in PBS), 7.5 mL n-heptane, and 1.5mL fresh 16% formaldehyde. Crosslinking was carried out at room temperature for exactly 15 mins on an orbital shaker at 250rpm, followed by addition of 3.7 mL 2M Tris-HCl pH7.5 and shaking for 5 mins to quench the reaction. Embryos were washed twice with 15 mL PBST and subjected to secondary crosslinking. Secondary crosslinking was done in 10mL of freshly prepared 3mM final DSG and ESG in PBST for 45 mins at room temperature with passive mixing. The reaction was quenched by addition of 3.7mL 2M Tris-HCl pH7.5 for 5 mins, washed twice with PBST, snap-frozen, and stored at -80°C until library construction. Micro-C libraries were prepared as previously described (Batut *et al*., 2022) with the following modifications: 50uL of 12-16hr embryos were used for each biological replicate. 60U of MNase was used for each reaction to digest chromatin to a mononucleosome:dinucleosome ratio of 4. 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

MicroC data for *D. melanogaster* were aligned to custom genomes edited from the Berkeley Drosophila Genome Project (BDGP) Release 6 reference assembly (dos Santos et al., 2015) with BWA-MEM (Li and Durbin, 2009) using parameters **-S -P -5 -M**. Briefly, the custom genomes are simply insertions of the transgenic sequence into the –142kb integration site, as predicted from perfect integration. These events were confirmed using PCR post-integration. 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*). The files from replicates were merged with Pairtools before generating 100bp contact matrices using Cooler (Abdennur and Mirny, 2020). Finally, balancing and Mcool file generation was performed with Cooler's Zoomify tool.

Virtual 4C profiles were extracted from individual replicates using FAN-C (Kruse et al., 2020) at 400bp resolution. The values were summed across replicates and smoothed across three bins (1.2kb). Viewpoints were determined based on the most informative region for interpretation. Ultimately we decided to use 800bp regions in the gene body of either *GFP* or *lacZ*, moving downstream from the *eve* promoter. However, similar results were obtained using viewpoints between transgenic *homie* and the promoters of either gene. For *homie* viewpoints, due to transgenic *homie* being masked, some loss of signal is expected at the viewpoint (**Figure 6**).

Supplemental Figures

Figure Sup. 1. TAD organization of the *even-skipped* locus and neighboring sequences MicroC contact pattern for the chromosomal region spanning the attP site at –142 kb and the *eve* locus. Like the *eve* volcano, this contact pattern was generated using aggregated previously published NC14 data (Batut *et al.*, 2022; Levo *et al.*, 2022). Arrow indicates the –142kb locus where transgenes are integrated into the genome. LDC domains are labeled according to the interacting TADs: for example, H-I indicates crosslinking events generated by physical interactions between sequences in two neighboring TADs, TH and TI, while I-J indicates interactions between sequences in TI and TJ. Likewise, G-I indicates crosslinking events generated by physical interactions between sequences TADs TG and TI, which are separated from each other by TAD TH. Note that unlike the TADs to the left of *eve*, the *eve* TAD does not interact frequently with its neighbors. Purple arrow indicates potential interaction dot between the boundary marking the border of TA and TB and the boundary marking the border of TI and TJ. Red, blue and green arrowheads mark boundaries that could engage in partner switching: see text.

Figure Sup. 2. Trouble ahead, trouble behind. A) Illustration of a broken zipper—zipping in front but not closing behind. B) Crosslinking pattern expected for cohesin loading at a site very close to the middle of the TAD-to-be: The vertical line of linked DNA is generated when the two chromatin fibers in the TAD-to-be are transiently linked together as cohesin passes. When cohesin bumps into the CTCF roadblocks (BE) at the boundaries of the TAD, the vertical line stops. C) Cohesin loading site is offset from center of TAD-to-be to the left. The crosslinking initiates at the loading site and extends vertically until the cohesin complex bumps into the CTCF roadblock on the left. At that point extrusion of the left chromatin fiber halts, while it continues to extrude the chromatin fiber on the right. The line of crosslinked sequences then extends at a 45° angle until cohesin comes to a halt at the CTCF roadblock on the right. D and E) The loading site is at or close to the boundary on the left (D) or the right (E). In this case extrusion will be unidirectional and the DNAs crosslinked as cohesin passes will form a line that extends at a 45° angle until cohesin comes to a halt at the CTCF roadblock. F). Cohesin is able to load at multiple sites in the TAD-to-be. In this case there will be a series of vertical lines whose intensity will be correlated with the relative frequency that a give loading site is used. The vertical lines will extend in each case until cohesin comes to a halt at the closest CTCF roadblock. At that point, the line of crosslinked DNA will extend at a 45° angle until that particular cohesin complex comes to halt at the CTCF roadblock that is farther away. The 45° line of crosslinked DNA will become progressive more intense as it approaches the apex of the triangle.

Figure Sup. 3. Models for the formation of LDC domains. Several different mechanisms could explain the LDC domains. A-D). In the loop extrusion model, the LDC domains arise because cohesin sometimes breaks through a CTCF roadblock. In A) there are three primary TADs, TAD1, TAD2, and TAD3. B and C) Occasional breakthrough events can generate a secondary TAD, LDC1-2 (TAD1+TAD2) or LDC2-3 (TAD2+TAD3). D) Even less frequent breakthrough events can generate a tripartite TAD (TAD1+TAD2+TAD3) that will correspond to LDC1-3. E-G). LDC domains arise because TADs sometimes bump into their neighbors. In E, TAD1 bumps into TAD2 to give LDC1-2, while in F TAD2 bumps into TAD3 to give LDC2:3. G) Less frequently TAD1 bumps into TAD3 to give LDC1:3. H-J) LDC domains arise because boundaries can switch partners. H and I) In H boundary A (dark green) skips boundary B (red) and pairs with boundary C (tan). This generates a new TAD, TAD1+TAD2 which corresponds to LDC domain 1-2. In I, boundary B (red) skips boundary C (tan) and pairs with D (blue). This generates a new TAD, TAD2+TAD3 which corresponds to LDC domain 2-3. J) In J, boundary A (dark green) skips boundary B (red) and boundary C (tan) and pairs with boundary D (blue). This generates a new TAD, TAD1+TAD2+TAD3 which corresponds to LDC domain 1-3.

Figure Sup. 4. Loop extrusion: *eveElephant*. A) Loop extrusion orientation dependent model for LhomieG Z5 transgene as the sketched in **Figure 2E**. This configuration generates *eveMa*. B) Instead running through *nhomie*, the cohesin complex comes to a halt at the *nhomie* boundary,

generating *eveElephant*. A second cohesin complex generates the *eve* TAD by loading at a site within the *eve* locus. Both *eveElephant* and the *eve* TAD are stem-loops. However, while this double stem-loop configuration would explain why the transgene *homie* makes contact with both *nhomie* and *homie*, the *eve* enhancers are isolated in their own loop and would not make contact with the two reporters in the transgene.

Figure Sup. 5. MicroC of *GhomieL Z3*. A) MicroC map of *GhomieL Z3*. Gray lines are masked regions due to repeated sequences. In this case, focusing on the zoom-in map from top to bottom, most likely P-element remnants 3' of the *lacZ* gene, transgenic *homie*, and P-element remnants 3' of the *GFP* gene. Compare directly to *GhomieL Z5* for changes in interaction preference due to change of *homie* orientation (**Figure 4C**). B) "Virtual 4C" map from either *lacZ* or *GFP* gene body viewpoints.

Figure Sup. 6. Micro C of *LhomieG Z3*. A) MicroC map of *LhomieG Z3*. Gray lines are masked regions due to repeated sequences. In this case, most possibly *P-element* remnants 3' of the *lacZ* gene. Compare directly to *GhomieL Z3* for changes in interaction preference due to change of homie orientation (**Figure 4B**). B) "Virtual 4C" map from either *lacZ* or *GFP* gene body viewpoints.

Figure Sup. 7. Blowup of interactions generated by the transgene with sequences neighboring –142 kb and the *eve* TAD. The interaction pattern indicates that the loop topology changes as a result of the orientation of transgenic *homie*. Blow-up of off-diagonal interactions, focused on the interaction between the endogenous *eve* locus and the –142kb transgenic insert with A) *LhomieG Z3*, B) *LhomieG Z5*, C) *GhomieL Z3*, or D) *GhomieL Z5*. Note the switch in preferential interaction between regions adjacent to transgenic *homie* and endogenous *homie* (indicated by arrow brackets). In cassettes where transgenic *homie* is in the same orientation in the genome as endogenous *homie* (A and D), regions upstream of transgenic *homie* preferentially interact with regions downstream of endogenous *homie*. In contrast, when transgenic *homie* is in the opposite orientation in the genome as endogenous *homie* (B and C), regions downstream of transgenic *homie* preferentially interact with regions downstream of endogenous *homie*. This suggests a different loop topology when the orientation of the transgene *homie* element in the chromosome is flipped.

References

Abdennur N., Mirny L.A (2020) **Cooler: scalable storage for Hi-C data and other genomically labeled arrays** *Bioinformatics* **36**:311–316 https://doi.org/10.1093/bioinformatics/btz540

Alipour E., Marko J.F. (2012) **Self-organization of domain structures by DNA-loop-extruding enzymes** *Nucleic Acids Res* **40**:11202–11212 https://doi.org/10.1093/nar/gks925

Avva S.V., Hart C.M (2016) **Characterization of the Drosophila BEAF-32A and BEAF- 32B Insulator Proteins** *PLoS One* **11** https://doi.org/10.1371/journal.pone.0162906

Bateman J.R., Lee A.M., Wu C.T (2006) **Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange** *Genetics* **173**:769–777

Batut P.J., Bing X.Y., Sisco Z., Raimundo J., Levo M., Levine M.S (2022) **Genome organization controls transcriptional dynamics during development** *Science* **375**:566–570 https://doi.org/10.1126/science.abi7178

Bell A.C., West A.G., Felsenfeld G (2001) **Insulators and boundaries: versatile regulatory elements in the eukaryotic genome** *Science* **291**:447–450 https://doi.org/10.1126/science.291.5503.447

Blanton J., Gaszner M., Schedl P (2003) **Protein:protein interactions and the pairing of boundary elements in vivo** *Genes Dev* **17**:664–675

Bonchuk A., Boyko K., Fedotova A., Nikolaeva A., Lushchekina S., Khrustaleva A., Popov V., Georgiev P (2021) **Structural basis of diversity and homodimerization specificity of zinc-finger-associated domains in Drosophila** *Nucleic Acids Res* **49**:2375–2389 https://doi.org/10.1093/nar/gkab061

Bonchuk A., Denisov S., Georgiev P., Maksimenko O (2011) **Drosophila BTB/POZ domains of "ttk group" can form multimers and selectively interact with each other** *J. Mol. Biol* **412**:423–436 https://doi.org/10.1016/j.jmb.2011.07.052

Bonchuk A., Kamalyan S., Mariasina S., Boyko K., Popov V., Maksimenko O., Georgiev P (2020) **N-terminal domain of the architectural protein CTCF has similar structural organization and ability to self-association in bilaterian organisms** *Sci Rep* **10** https://doi.org/10.1038/s41598-020-59459-5

Cai H.N., Shen P (2001) **Effects of cis arrangement of chromatin insulators on enhancer-blocking activity** *Science* **291**:493–495

Callan H.G (1963) **THE NATURE OF LAMPBRUSH CHROMOSOMES** *Int. Rev. Cytol* **15**:1–34 https://doi.org/10.1016/s0074-7696(08)61114-6

Cavalheiro G.R., Pollex T., Furlong E.E (2021) **To loop or not to loop: what is the role of TADs in enhancer function and gene regulation?** *Curr. Opin. Genet. Dev* **67**:119–129 https://doi.org/10.1016/j.gde.2020.12.015

Chen H., Levo M., Barinov L., Fujioka M., Jaynes J.B., Gregor T (2018) **Dynamic interplay between enhancer–promoter topology and gene activity** *Nat. Genet* **50**

Chetverina D., Aoki T., Erokhin M., Georgiev P., Schedl P (2014) **Making connections: insulators organize eukaryotic chromosomes into independent cis-regulatory networks** *Bioessays* **36**:163–172

Chetverina D., Fujioka M., Erokhin M., Georgiev P., Jaynes J.B., Schedl P (2017) **Boundaries of loop domains (insulators): Determinants of chromosome form and function in multicellular eukaryotes** *Bioessays* **39**

Collier J.E., Nasmyth K.A (2022) **DNA passes through cohesin's hinge as well as its Smc3-kleisin interface** *Elife* **11** https://doi.org/10.7554/eLife.80310

Cuartero S., Fresan U., Reina O., Planet E., Espinas M.L (2014) **Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function** *EMBO J* **33**:637–647

Davidson I.F., Peters J.M (2021) **Genome folding through loop extrusion by SMC complexes** *Nat Rev Mol Cell Biol* **22**:445–464 https://doi.org/10.1038/s41580-021-00349-7

dos Santos G., Schroeder A.J., Goodman J.L., Strelets V.B., Crosby M.A., Thurmond J., Emmert D.B., Gelbart W.M., FlyBase C. (2015) **FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations** *Nucleic Acids Res* **43**:D690–697

Duan J., Rieder L., Colonnetta M.M., Huang A., McKenney M., Watters S., Deshpande G., Jordan W., Fawzi N., Larschan E (2021) **CLAMP and Zelda function together to promote Drosophila zygotic genome activation** *Elife* **10** https://doi.org/10.7554/eLife.69937

Erokhin M., Davydova A., Kyrchanova O., Parshikov A., Georgiev P., Chetverina D (2011) **Insulators form gene loops by interacting with promoters in Drosophila** *Development* **138**:4097–4106

Erokhin M. *et al.* (2021) **Boundaries potentiate polycomb response element-mediated silencing** *BMC Biol* **19** https://doi.org/10.1186/s12915

Fedotova A., Aoki T., Rossier M., Mishra R.K., Clendinen C., Kyrchanova O., Wolle D., Bonchuk A., Maeda R.K., Mutero A (2018) **The BEN domain protein insensitive binds to the Fab-7 chromatin boundary to establish proper segmental identity in Drosophila** *Genetics* **210**:573–585

Fedotova A., Clendinen C., Bonchuk A., Mogila V., Aoki T., Georgiev P., Schedl P (2019) **Functional dissection of the developmentally restricted BEN domain chromatin boundary factor Insensitive** *Epigenetics & chromatin* **12**

Fedotova A.A., Bonchuk A.N., Mogila V.A., Georgiev P.G (2017) **C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors** *Acta Naturae* **9**:47–58

Fudenberg G., Imakaev M., Lu C., Goloborodko A., Abdennur N., Mirny L.A (2016) **Formation of Chromosomal Domains by Loop Extrusion** *Cell Rep* **15**:2038–2049 https://doi.org/10.1016/j.celrep.2016.04.085

Fujioka M., Mistry H., Schedl P., Jaynes J.B (2016) **Determinants of Chromosome Architecture: Insulator Pairing in cis and in trans** *PLoS Genet* **12**

Fujioka M., Sun G., Jaynes J.B (2013) **The Drosophila eve insulator Homie promotes eve expression and protects the adjacent gene from repression by Polycomb spreading** *PLoS Genet* **9**

Fujioka M., Wu X., Jaynes J.B (2009) **A chromatin insulator mediates transgene homing and very long-range enhancer-promoter communication** *Development* **136**:3077–3087

Fukaya T., Lim B., Levine M (2016) **Enhancer control of transcriptional bursting** *Cell* **166**:358–368

Fung J.C., Marshall W.F., Dernburg A., Agard D.A., Sedat J.W (1998) **Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations** *J. Cell Biol* **141**:5–20

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

Ghosh R.P., Meyer B.J (2021) **Spatial Organization of Chromatin: Emergence of Chromatin Structure During Development** *Annu. Rev. Cell. Dev. Biol* **37**:199–232 https://doi.org/10.1146/annurev-cellbio-032321-035734

Gligoris T.G., Scheinost J.C., Bürmann F., Petela N., Chan K.L., Uluocak P., Beckouët F., Gruber S., Nasmyth K., Löwe J (2014) **Closing the cohesin ring: structure and function of its Smc3-kleisin interface** *Science* **346**:963–967 https://doi.org/10.1126/science.1256917

Gohl D., Aoki T., Blanton J., Shanower G., Kappes G., Schedl P (2011) **Mechanism of chromosomal boundary action: roadblock, sink, or loop?** *Genetics* **187**:731–748

Gramates L.S. *et al.* (2022) **FlyBase: a guided tour of highlighted features** *Genetics* **220** https://doi.org/10.1093/genetics/iyac035

Guo Y., Monahan K., Wu H., Gertz J., Varley K.E., Li W., Myers R.M., Maniatis T., Wu Q (2012) **CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice** *Proc Natl Acad Sci USA* **109**:21081–21086 https://doi.org/10.1073/pnas.1219280110

Guo Y., Xu Q., Canzio D., Shou J., Li J., Gorkin D.U., Jung I., Wu H., Zhai Y., Tang Y (2015) **CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function** *Cell* **162**:900–910

Hart C.M., Zhao K., Laemmli U.K (1997) **The scs' boundary element: characterization of boundary element-associated factors** *Mol Cell Biol* **17**:999–1009

Heger P., George R., Wiehe T (2013) **Successive gain of insulator proteins in arthropod evolution** *Evolution* **67**:2945–2956

Heger P., Wiehe T (2014) **New tools in the box: an evolutionary synopsis of chromatin insulators** *Trends Genet* **30**:161–171

Hsieh T.S., Cattoglio C., Slobodyanyuk E., Hansen A.S., Rando O.J., Tjian R., Darzacq X (2020) **Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding** *Mol. Cell* **78**:539–553 https://doi.org/10.1016/j.molcel.2020.03.002

Jerković I., Szabo Q., Bantignies F., Cavalli G (2020) **Higher-Order Chromosomal Structures Mediate Genome Function** *J. Mol. Biol* **432**:676–681 https://doi.org/10.1016/j.jmb.2019.10.014

Ke W., Fujioka M., Schdle P., Jaynes J.B. **Ke, W., Fujioka, M., Schdle, P., and Jaynes, J.B. Chromosome Structure II: Stem-loops and circle-loops.** *Chromosome Structure II: Stem-loops and circle-loops*

Krietenstein N. *et al.* (2020) **Ultrastructural Details of Mammalian Chromosome Architecture** *Mol. Cell* **78**:554–565 https://doi.org/10.1016/j.molcel.2020.03.003

Kruse K., Hug C.B., Vaquerizas J.M (2020) **FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data** *Genome Biol* **21** https://doi.org/10.1186/s13059-020-02215-9

Kyrchanova O., Chetverina D., Maksimenko O., Kullyev A., Georgiev P (2008) **Orientation-dependent interaction between Drosophila insulators is a property of this class of regulatory elements** *Nucleic Acids Res* **36**:7019–7028

Kyrchanova O., Ibragimov A., Postika N., Georgiev P., Schedl P (2023) **Boundary bypass activity in the abdominal-B region of the Drosophila bithorax complex is position dependent and regulated** *Open Biol* **13** https://doi.org/10.1098/rsob.230035

Kyrchanova O., Ivlieva T., Toshchakov S., Parshikov A., Maksimenko O., Georgiev P (2011) **Selective interactions of boundaries with upstream region of Abd-B promoter in Drosophila bithorax complex and role of dCTCF in this process** *Nucleic Acids Res* **39**:3042–3052

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

Kyrchanova O., Sabirov M., Mogila V., Kurbidaeva A., Postika N., Maksimenko O., Schedl P., Georgiev P (2019) **Complete reconstitution of bypass and blocking functions in a minimal artificial Fab-7 insulator from Drosophila bithorax complex** *Proceedings of the National Academy of Sciences* **201907190**

Kyrchanova O., Toshchakov S., Parshikov A., Georgiev P (2007) **Study of the functional interaction between Mcp insulators from the Drosophila bithorax complex: effects of insulator pairing on enhancer-promoter communication** *Mol. Cell. Biol* **27**:3035–3043

Kyrchanova O., Toshchakov S., Podstreshnaya Y., Parshikov A., Georgiev P (2008) **Functional interaction between the Fab-7 and Fab-8 boundaries and the upstream promoter region in the Drosophila Abd-B gene** *Mol. Cell. Biol* **28**:4188–4195

Kyrchanova O., Wolle D., Sabirov M., Kurbidaeva A., Aoki T., Maksimenko O., Kyrchanova M., Georgiev P., Schedl P (2019) **Distinct Elements Confer the Blocking and Bypass Functions of the Bithorax Fab-8 Boundary** *Genetics, genetics* **302694**

Levo M., Raimundo J., Bing X.Y., Sisco Z., Batut P.J., Ryabichko S., Gregor T., Levine M.S (2022) **Transcriptional coupling of distant regulatory genes in living embryos** *Nature* **605**:754–760 https://doi.org/10.1038/s41586-022-04680-7

Li H., Durbin R (2009) **Fast and accurate short read alignment with Burrows-Wheeler transform** *Bioinformatics* **25**:1754–1760 https://doi.org/10.1093/bioinformatics/btp324

Li H.-B., Muller M., Bahechar I.A., Kyrchanova O., Ohno K., Georgiev P., Pirrotta V (2011) **Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in Drosophila melanogaster** *Mol. Cell. Biol* **31**:616–625

Li L. *et al.* (2015) **Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing** *Mol. Cell* **58**:216–231 https://doi.org/10.1016/j.molcel.2015.02.023

Li X., Tang X., Bing X., Catalano C., Li T., Dolsten G., Wu C., Levine M (2023) **GAGA-associated factor fosters loop formation in the Drosophila genome** *Mol. Cell* **83**:1519–1526 https://doi.org/10.1016/j.molcel.2023.03.011

Lim B., Heist T., Levine M., Fukaya T (2018) **Visualization of transvection in living Drosophila embryos** *Mol. Cell* **70**:287–296

Little S.C., Gregor T (2018) **Single mRNA Molecule Detection in Drosophila** *Methods Mol Biol* **1649**:127–142 https://doi.org/10.1007/978-1-4939-7213-5_8

Matthews N.E., White R. (2019) **Chromatin Architecture in the Fly: Living without CTCF/Cohesin Loop Extrusion?**

Mirny L., Dekker J (2022) **Mechanisms of chromosome folding and nuclear organization: their interplay and open questions** *Cold Spring Harbor Perspectives in Biology* **14**

Mohana G. *et al.* (2023) **Chromosome-level organization of the regulatory genome in the Drosophila nervous system** *Cell* **186**:3826–3844 https://doi.org/10.1016/j.cell.2023.07.008

Muller M., Hagstrom K., Gyurkovics H., Pirrotta V., Schedl P (1999) **The mcp element from the Drosophila melanogaster bithorax complex mediates long-distance regulatory interactions** *Genetics* **153**:1333–1356

Muravyova E., Golovnin A., Gracheva E., Parshikov A., Belenkaya T., Pirrotta V., Georgiev P (2001) **Loss of insulator activity by paired Su(Hw) chromatin insulators** *Science* **291**:495–498

Nora E.P. *et al.* (2020) **Molecular basis of CTCF binding polarity in genome folding** *Nat Commun* **11** https://doi.org/10.1038/s41467-020-19283-x

Nuebler J., Fudenberg G., Imakaev M., Abdennur N., Mirny L.A (2018) **Chromatin organization by an interplay of loop extrusion and compartmental segregation** *Proc Natl Acad Sci USA* **115**:E6697–e6706 https://doi.org/10.1073/pnas.1717730115

Perea-Resa C., Wattendorf L., Marzouk S., Blower M.D (2021) **Cohesin: behind dynamic genome topology and gene expression reprogramming** *Trends Cell Biol* **31**:760–773 https://doi.org/10.1016/j.tcb.2021.03.005

Rowley M.J., Corces V.G (2018) **Organizational principles of 3D genome architecture** *Nature Reviews Genetics* **19**:789–800

Sanborn A.L., Rao S.S., Huang S.-C., Durand N.C., Huntley M.H., Jewett A.I., Bochkov I.D., Chinnappan D., Cutkosky A., Li J (2015) **Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes** *Proceedings of the National Academy of Sciences* **112**:E6456–E6465

Schoborg T.A., Labrador M (2010) **The phylogenetic distribution of non-CTCF insulator proteins is limited to insects and reveals that BEAF-32 is Drosophila lineage specific** *J. Mol. Evol* **70**:74–84 https://doi.org/10.1007/s00239-009-9310-x

Sigrist C.J., Pirrotta V (1997) **Chromatin insulator elements block the silencing of a target gene by the Drosophila polycomb response element (PRE) but allow trans interactions between PREs on different chromosomes** *Genetics* **147**:209–221

Sun Y., Nien C.Y., Chen K., Liu H.Y., Johnston J., Zeitlinger J., Rushlow C (2015) **Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation** *Genome Res* **25**:1703–1714 https://doi.org/10.1101/gr.192542.115

Tan L. *et al.* (2023) **Lifelong restructuring of 3D genome architecture in cerebellar granule cells** *Science* **381**:1112–1119 https://doi.org/10.1126/science.adh3253

Trcek T., Lionnet T., Shroff H., Lehmann R. (2017) **mRNA quantification using single-molecule FISH in Drosophila embryos** *Nat Protoc* **12**:1326–1348 https://doi.org/10.1038/nprot.2017.030

Ueberschär M., Wang H., Zhang C., Kondo S., Aoki T., Schedl P., Lai E.C., Wen J., Dai Q (2019) **BEN-solo factors partition active chromatin to ensure proper gene activation in Drosophila** *Nat Commun* **10** https://doi.org/10.1038/s41467-019-13558-8

Vazquez J. *et al.* (1993) **Genetic and molecular analysis of chromatin domains** *Cold Spring Harb Symp Quant Biol* **58**:45–54 https://doi.org/10.1101/sqb.1993.058.01.008

Vazquez J., Muller M., Pirrotta V., Sedat J.W (2006) **The Mcp element mediates stable long-range chromosome-chromosome interactions in Drosophila** *Molecular Biology of the Cell* **17**:2158–2165

Yokoshi M., Segawa K., Fukaya T (2020) **Visualizing the Role of Boundary Elements in Enhancer-Promoter Communication** *Mol. Cell* **78**:224–235 https://doi.org/10.1016/j.molcel.2020.02.007

Zhimulev I.F., Belyaeva E.S. (1991) **Chromomeric organization of polytene chromosomes** *Genetica* **85**:65–72 https://doi.org/10.1007/bf00056107

Zhimulev I.F., Pokholkova G.V., Bgatov A.V., Semeshin V.F., Umbetova G.H., Belyaeva E.S (1983) **Genetic interpretation of polytene chromosomes banding pattern** *Mol. Biol. Rep* **9**:19–23 https://doi.org/10.1007/bf00777469

Zolotarev N., Fedotova A., Kyrchanova O., Bonchuk A., Penin A.A., Lando A.S., Eliseeva I.A., Kulakovskiy I.V., Maksimenko O., Georgiev P (2016) **Architectural proteins Pita, Zw5, and ZIPIC contain homodimerization domain and support specific long-range interactions in Drosophila** *Nucleic Acids Res* **44**:7228–7241 https://doi.org/10.1093/nar/gkw371

Article and author information

Xinyang Bing

Lewis Sigler Institute, Princeton University, Princeton, NJ. 08544, Present address: BlueRock Therapeutics, Cambridge, Ma 20142
ORCID iD: 0000-0001-6789-1918

Wenfan Ke
Department of Molecular Biology, Princeton University, Princeton NJ 08544
ORCID iD: 0000-0002-7047-5445

Miki Fujioka
Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA 19107

Amina Kurbidaeva
Department of Molecular Biology, Princeton University, Princeton NJ 08544, Present address: Center for Genomics and Systems Biology, New York University, NY 10003

Sarah Levitt
Department of Molecular Biology, Princeton University, Princeton NJ 08544

Mike Levine
Lewis Sigler Institute, Princeton University, Princeton, NJ. 08544

Paul Schedl
Department of Molecular Biology, Princeton University, Princeton NJ 08544
For correspondence: pschedl@princeton.edu
ORCID iD: 0000-0001-5704-2349

James B. Jaynes
Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA 19107

Copyright

Editors

Reviewing Editor
Mia Levine
University of Pennsylvania, Philadelphia, United States of America

Senior Editor
Claude Desplan
New York University, New York, United States of America

Reviewer #1 (Public Review):

Summary:
The authors addressed how long-range interactions between boundary elements are established and influence their function in enhancer specificity. Briefly, the authors placed two different reporters separated by a boundary element. They inserted this construct ectopically ~140 kb away from an endogenous locus that contains the same boundary

element. The authors used expression patterns driven by nearby enhancers as an output to determine which enhancers the reporters interact with. They complemented this analysis with 3D DNA contact mapping. The authors found that the orientation of the boundary element determined which enhancers each reporter interacted with. They proposed that the 3D interaction topology, whether being circular or stem configuration, distinguished whether the interaction was cohesin mediated or through an independent mechanism termed pairing.

Strengths:
The transgene expression assays are built upon prior knowledge of the enhancer activities. The 3D DNA contacts confirm that transgene expression correlates with the contacts. Using 4 different orientations covers all combinations of the reporter genes and the boundary placement.

Weaknesses:
The interpretation of the data as a refusal of loop extrusion playing a role in TAD formation is not warranted, as the authors did not deplete the loop extruders to show that what they measure is independent. As the authors show, the single long DNA loop mediated by cohesin loop extrusion connecting the ectopic and endogenous boundary is clearly inconsistent with the results, therefore the main conclusion of the paper that the 3D topology of the boundary elements a consequence of pairing is strong. However, the loop extrusion and pairing are not mutually exclusive models for the formation of TADs. Loop-extruding cohesin complexes need not make a 140 kb loop, multiple smaller loops could bring together the two boundary elements, which are then held together by pairing proteins that can make circular topologies.

https://doi.org/10.7554/eLife.94070.1.sa2

Reviewer #2 (Public Review):

In Bing et al, the authors analyze micro-C data from NC14 fly embryos, focusing on the eve locus, to assess different models of chromatin looping. They conclude that fly TADs are less consistent with conventional cohesin-based loop extrusion models and instead rely more heavily on boundary-boundary pairings in an orientation-dependent manner.

Overall, I found the manuscript to be interesting and thought-provoking. However, this paper reads much more like a perspective than a research article. I strongly suggest the authors spend some time editing their introduction to the most salient points as well as organizing their results section in a more conventional way with conclusion-based titles. It was very difficult to follow the authors' logic throughout the manuscript as written. It was also not clear as written which experiments were performed as part of this study and which were reanalyzed but published elsewhere. This should be made clearer throughout.

It has been shown several times that Drosophila Hi-C maps do not contain all of the features (frequent corner peaks, stripes, etc.) observed when compared to mammalian cells. Considering these features are thought to be products of extrusion events, it is not an entirely new concept that Drosophila domains form via mechanisms other than extrusion. That being said, the authors' analyses do not distinguish between the formation and the maintenance of domains. It is not clear to this reviewer why a single mechanism should explain the formation of the complex structures observed in static Hi-C heatmaps from a population of cells at a single developmental time point. For example, how can the authors rule out that extrusion initially provides the necessary proximity and possibly the cis preference of contacts required for boundary-boundary pairing whereas the latter may more reflect the structures observed at maintenance? Future work aimed at analyzing micro-C data in cohesin-depleted cells might shed additional light on this.

Additional mechanisms at play include compartment-level interactions driven by chromatin states. Indeed, in mammalian cells, these interactions often manifest as a "plume" on Hi-C maps similar to what the authors attribute to boundary interactions in this manuscript. How do the chromatin states in the neighboring domains of the eve locus impact the model if at all?

How does intrachromosomal homolog pairing impact the models proposed in this manuscript (Abed et al. 2019; Erceg et al., 2019). Several papers recently have shown that somatic homolog pairing is not uniform and shows significant variation across the genome with evidence for both tight pairing regions and loose pairing regions. Might loose pairing interactions have the capacity to alter the cis configuration of the eve locus?
In summary, the transgenic experiments are extensive and elegant and fully support the authors' models. However, in my opinion, they do not completely rule out additional models at play, including extrusion-based mechanisms. Indeed, my major issue is the limited conceptual advance in this manuscript. The authors essentially repeat many of their previous work and analyses. The authors make no attempt to dissect the mechanism of this process by modifying extrusion components directly. Some discussion of Rollins et al., 1999 on the discovery of Nipped-B and its role in enhancer-promoter communication should also be made to reconcile their conclusions in the proposed absence of extrusion events.

https://doi.org/10.7554/eLife.94070.1.sa1

Reviewer #3 (Public Review):

Bing et al. attempt to address fundamental mechanisms of TAD formation in Drosophila by analyzing gene expression and 3D conformation within the vicinity of the eve TAD after insertion of a transgene harboring a Homie insulator sequence 142 kb away in different orientations. These transgenes along with spatial gene expression analysis were previously published in Fujioka et al. 2016, and the underlying interpretations regarding resulting DNA configuration in this genomic region were also previously published. This manuscript repeats the expression analysis using smFISH probes in order to achieve more quantitative analysis, but the main results are the same as previously published. The only new data are the Micro-C and an additional modeling/analysis of what they refer to as the 'Z3' orientation of the transgenes. The rest of the manuscript merely synthesizes further interpretation with the goal of addressing whether loop extrusion may be occurring or if boundary:boundary pairing without loop extrusion is responsible for TAD formation. The authors conclude that their results are more consistent with boundary:boundary pairing and not loop extrusion; however, most of this imaging data seems to support both loop extrusion and the boundary:boundary models. This manuscript lacks support, especially new data, for its conclusions. Furthermore, there are many parts of the manuscript that are difficult to follow. There are some minor errors in the labelling of the figures that if fixed would help elevate understanding. Lastly, there are several major points that if elaborated on, would potentially be helpful for the clarity of the manuscript.

Major Points:
1. The authors suggest and attempt to visualize in the supplemental figures, that loop extrusion mechanisms would appear during crosslinking and show as vertical stripes in the micro-C data. In order to see stripes, a majority of the nuclei would need to undergo loop extrusion at the same rate, starting from exactly the same spots, and the loops would also have to be released and restarted at the same rate. If these patterns truly result from loop extrusion, the authors should provide experimental evidence from another organism undergoing loop extrusion.
2. On lines 311-314, the authors discuss that stem-loops generated by cohesin extrusion would possibly be expected to have more next-next-door neighbor contacts than next-door neighbor

contacts and site their models in Figure 1. Based on the boundary:boundary pairing models in the same figure would the stem-loops created by head-to-tail pairing also have the same phenotype? Making possible enrichment of next-next-door neighbor contacts possible in both situations? The concepts in the text are not clear, and the diagrams are not well-labeled relative to the two models.
3. The authors appear to cite Chen et al., 2018 as a reference for the location of these transgenes being 700nM away in a majority of the nuclei. However, the exact transgenes in this manuscript do not appear to have been measured for distance. The authors could do this experiment and include expression measurements.
4. The authors discuss the possible importance of CTCF orientation in forming the roadblock to cohesin extrusion and discuss that Homie orientation in the transgene may impact Homie function as an effective roadblock. However, the Homie region inserted in the transgene does not contain the CTCF motif. Can the authors elaborate on why they feel the orientation of Homie is important in its ability to function as a roadblock if the CTCF motif is not present? Trans-acting factors responsible for Homie function have not been identified and this point is not discussed in the manuscript.
5. The imaging results seem to be consistent with both boundary:boundary interaction and loop extrusion stem looping.
6. The authors suggest that the eveMa TAD could only be formed by extrusion after the breakthrough of Nhomie and several other roadblocks. Additionally, the overall long-range interactions with Nhomie appear to be less than the interactions with endogenous Homie (Figures 7, 8, and supplemental 5). Is it possible that in some cases boundary:boundary pairing is occurring between only the transgenic Homie and endogenous Homie and not including Nhomie?
7. In Figure 4E, the GFP hebe expression shown in the LhomieG Z5 transgenic embryo does not appear in the same locations as the LlambdaG Z5 control. Is this actually hebe expression or just a background signal?
8. Figure 6- The LhomieG Z3 late-stage embryo appears to be showing the ventral orientation of the embryo rather than the lateral side of the embryo as was shown in the previous figure. Is this for a reason? Additionally, there are no statistics shown for the Z3 transgenic images. Were these images analyzed in the same way as the Z5 line images?
9. Do the Micro-C data align with the developmental time points used in the smFISH probe assays?

https://doi.org/10.7554/eLife.94070.1.sa0