The *homie* insulator has sub-elements with different insulating and long-range pairing properties

Miki Fujioka[1], Wenfan Ke[2], Paul Schedl[2], and James B Jaynes[1]

1: Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA 19107

2: Department of Molecular Biology, Princeton University, Princeton NJ 08544

ABSTRACT

Chromatin insulators are a major determinant of chromosome architecture. Specific architectures induced by insulators profoundly influence nuclear processes, including how enhancers interact with promoters over long distances and between homologous chromosomes. Insulators can pair with copies of themselves in *trans*, thereby facilitating homolog pairing. They can also pair with other insulators, sometimes with great specificity, inducing long-range chromosomal loops in *cis*. Contrary to their canonical function of enhancer blocking, these loops can bring distant enhancers and promoters together to activate gene expression, while at the same time blocking other interactions in *cis*. The details of these effects depend on the choice of pairing partner, and on the orientation specificity of pairing, implicating the 3-dimensional architecture as a major determinant of function. Here we dissect the *homie* insulator from the Drosophila *even skipped* (*eve*) locus, to understand its complex substructure. We test pairing function in *cis* based on *homie*-carrying transgenes interacting with endogenous *eve*. The assay is sensitive to both pairing strength and orientation specificity. Using this assay, we found that a consensus Su(Hw) binding site in *homie* is required for efficient long-range interaction, although some activity remains without it. This binding site also contributes to the canonical insulator activities of enhancer blocking and barrier function. Based on this and other results from our functional dissection, enhancer blocking and barrier activities appear to be partially separable. Overall, our results show the complexity inherent in insulator functions, which can be provided by an array of proteins with both shared and distinct properties.

Introduction

Genomic elements called boundaries or insulators function to separate genes from each other, or, in some cases, to separate distinct regulatory domains within a gene (the fly Hox genes). However, they have properties that have proven to be difficult to reconcile mechanistically. For example, while these elements can block enhancer-promoter interactions when inserted between the two, they have also been shown to mediate long-distance regulatory interactions (CHETVERINA *et al.* 2017; SCHWARTZ AND CAVALLI 2017; ÖZDEMIR AND GAMBETTA 2019).

One of the first insulator elements discovered is located within the *Drosophila gypsy* retrotransposon (MODOLELL *et al.* 1983; PEIFER AND BENDER 1988). The *gypsy* transposon was found to be responsible for many spontaneous mutations in flies. In many cases, the mutant phenotypes arose because the *gypsy* transposon inserted between a gene and its regulatory elements, blocking regulatory interactions. The insulator properties of *gypsy* have been attributed to the binding of a set of chromatin proteins, the sequence-specific component of which is the protein Suppressor of Hairy Wing (Su(Hw) (PARKHURST *et al.* 1988; HOLDRIDGE AND DORSETT 1991; GEYER AND CORCES 1992). Several insulators found in the bithorax complex (BxC) of Drosophila have also been extensively studied, including *Mcp*, *Fab7*, and *Fab8* (KARCH *et al.* 1994; ZHOU *et al.* 1999; BARGES *et al.* 2000). Some but not all of these bind Su(Hw), while an array of other proteins, including Cp190, are also known to contribute to the activities of these and other insulators (ZHAO *et al.* 1995; GASZNER *et al.* 1999; GHOSH *et al.* 2001; PAI *et al.* 2004; MOON *et al.* 2005; RATH *et al.* 2006; ROY *et al.* 2007; MAKSIMENKO *et al.* 2008; JIANG *et al.* 2009; AOKI *et al.* 2012; LI AND GILMOUR 2013; CUARTERO *et al.* 2014; MAKSIMENKO *et al.* 2015; ZOLOTAREV *et al.* 2016; FEDOTOVA *et al.* 2018; RAMÍREZ *et al.* 2018; BAG *et al.* 2021).

The Drosophila *even skipped* (*eve*) gene is flanked by two insulators, *nhomie* (*neighbor of homie*) and *homie* (*homing insulator at eve*), at its 5'- and 3'-ends, respectively. These two elements define the *eve* TAD (topologically associated domain) (FUJIOKA *et al.* 2009; FUJIOKA *et al.* 2016; BING *et al.* 2024) (see map in Fig. 1A). The properties of the 3' insulator *homie* have been well-characterized. Like other insulators, it has enhancer blocking activity and an ability to prevent the spread of Polycomb dependent silencing (FUJIOKA *et al.* 2013; FUJIOKA *et al.* 2021). *homie* abuts the promoter of an essential housekeeping gene, *TER94* (PINTER *et al.* 1998; LEON AND MCKEARIN 1999; RUDEN *et al.* 2000). Since the *eve* TAD is assembled into a Polycomb group (PcG) silenced domain in most cells during all but the early stages of development (SCHWARTZ *et al.* 2006; TOLHUIS *et al.* 2006), this barrier activity is thought to be a critically important function. In *eve-TER94* "pseudo-locus" transgenes, *homie* is required to keep the *eve* Polycomb response elements (PREs) from shutting down *TER94* through the spreading of the repressive Pc histone modification H3K27me3 (trimethylation of histone H3 residue lysine-27) (FUJIOKA *et al.* 2013).

In addition to these activities, *homie* and *nhomie* share another characteristic property of fly insulators, namely, an ability to physically pair with themselves and with each other (FUJIOKA *et al.* 2009; FUJIOKA *et al.* 2016; CHEN *et al.* 2018). These pairing interactions have two important consequences. First, pairing between heterologous boundaries in *cis* generates a loop or TAD. Depending on the orientation of the pairing interaction (head-to-head or head-to-tail), this TAD

can be a circle-loop (head-to-head) or a stem-loop (head-to-tail). Since *homie* and *nhomie* pair with each other head-to-tail, the *eve* TAD has a stem-loop topology. As we have shown in Ke *et al.* (KE *et al.* 2024), this loop topology enhances the physical isolation of the *eve* TAD from its neighbors. Second, homologous chromosomes in flies are paired in register throughout much of the life cycle, and the available evidence suggests that self-pairing interactions between TAD boundary elements facilitate this (VIETS *et al.* 2019). While *homie:nhomie* pairing is head-to-tail, *homie* and *nhomie* self-pairing is head-to-head, like that of other fly boundaries that have been tested (e.g., (KYRCHANOVA *et al.* 2008). This orientation preference has been shown to favor transvection (regulatory cross-talk) between paired homologous chromosomes (FUJIOKA *et al.* 2016), a phenomenon that is widespread in *Drosophila*.

As has been observed for the *gypsy* insulator and for boundaries from the Bithorax complex (SIGRIST AND PIRROTTA 1997; MULLER *et al.* 1999; VAZQUEZ *et al.* 2006; LI *et al.* 2011), both *homie* and *nhomie* can also mediate long-distance (100 kb to Mbs) regulatory interactions (FUJIOKA *et al.* 2009; FUJIOKA *et al.* 2016; CHEN *et al.* 2018; BING *et al.* 2024). As an example, when a *homie*-containing reporter transgene is inserted in an attP site located in the 1st exon of the *hebe* gene, 142 kb upstream of the *eve* gene, enhancers in the *eve* TAD can drive reporter expression (FUJIOKA *et al.* 2009; FUJIOKA *et al.* 2016; CHEN *et al.* 2018). Reporter activation depends on the relative orientation of the *homie* element in the transgene, but not on the orientation of the transgene in the chromosome. This constraint arises because *homie* in the transgene pairs with *homie* in the *eve* locus head-to-head, and it pairs with *eve*-locus *nhomie* head-to-tail (FUJIOKA *et al.* 2016; BING *et al.* 2024).

Here, we have undertaken a functional dissection of *homie*, focusing on three important activities of this class of elements: enhancer blocking, PRE blocking and long-range pairing. Detailed slicing and dicing of *homie* reveals a general correlation between these three activities. However, the correlation is not perfect, indicating that while the activities are closely related, there are some mechanistic differences. For example, we show that a minimal <300 bp sequence retains all three activities. Within the minimal element there is a single consensus recognition sequence for the *gypsy* insulator-binding protein Su(Hw). While mutation of this sequence results in a near-complete loss of long-range pairing activity, it only partially compromises enhancer blocking activity, and has a minimal effect on PRE blocking.

Results

Multiple insulator sub-elements contribute to long-range pairing

In previous studies on long-range pairing interactions, we used a "minimal" *homie*, CDEF, which consisted of four contiguous ~100 bp sub-elements, as it appeared to have nearly fully activity (FUJIOKA *et al.* 2016; CHEN *et al.* 2018). To better understand which of these *homie* sequences contribute to long-range pairing, we examined the long-distance pairing activity of different combinations of sub-elements from a larger 800 bp fragment, ABCDEFGH, that was previously identified as the "*homie"* insulator (FUJIOKA *et al.* 2009). As indicated in Fig. 1B, this larger fragment extends from just downstream of the *eve* PRE to just beyond the first exon of the *TER94-RA* transcript (GRAMATES *et al.* 2022), and it spans a DNase hypersensitive region (SACKERSON *et al.* 1999). To test for long-range (LR) pairing interactions, we inserted *homie*-containing reporter gene constructs into an attP site in the 1st exon of the *hebe* gene, 142 kb upstream of *eve* (FUJIOKA *et al.* 2009; FUJIOKA *et al.* 2016; CHEN *et al.* 2018). For the first set of experiments, we used the single reporter transgene shown in Fig. 1C, while in experiments described below we used the dual reporter transgene shown in Fig. 1D. When *homie* has LR pairing activity (and is inserted in the correct orientation relative to the reporter: see Fig. 1C, D) the reporter gene is subject to regulation by enhancers in the *eve* TAD and is expressed in an *eve*-like pattern (Fig. 2). By contrast, no reporter expression is observed when the transgenes have an equal-length DNA segment from phage λ ("λ DNA" in Fig. 2, Suppl. Fig. 2). Also, as described further below in the section on enhancer blocking activity, a *hebe* enhancer is located upstream of the attP site at –142 kb (Suppl. Fig. 1B). When *homie* is interposed between this enhancer and the transgene reporter, it blocks the enhancer from activating the reporter.

In the experiments shown in Fig. 2, we compared the LR pairing activity of different constructs using two criteria: the relative number of the cells expressing the reporter gene in an *eve* pattern and the relative level of expression as judged by the intensity of staining. These two measures generally correlated with each other. We first extended the minimal CDEF *homie* endpoint by ~200 bp downstream into the beginning of the *TER94* transcription unit to give CDEFGH (see map, Fig. 1B). The addition of the GH sub-elements appeared to increase the LR pairing frequency and/or stability (compare CDEF with CDEFGH in Fig. 2). This is reflected in a greater number of cells expressing the reporter in early stripes, and in the stronger expression at later stages, most noticeably in the CNS and mesoderm of stage 13 embryos (Suppl. Fig. 1A). To further test the contribution of the GH region to LR pairing activity, we combined GH with EF. As shown below (see Fig. 3B), EF on its own has no LR pairing activity. In contrast, an *eve*-like pattern of *lacZ* expression is observed for EFGH throughout embryogenesis (Fig. 2). However, the intensity of staining is considerably weaker than CDEF, and fewer cells express *lacZ*. This shows that GH can partially substitute for the CD sub-elements. On the other hand, as shown in Fig. 2, the AB sub-elements cannot replace EF, as the ABCD combination has little or no LR pairing activity.

Next, we assayed the LR pairing activity of the four possible combinations of three different CDEF sub-elements. Of these sub-element combinations, only DEF has LR pairing activity comparable to that of CDEF (Fig. 3A). It seems that C facilitates LR pairing by DEF, since CDEF

shows slightly stronger expression in more cells than does DEF (Fig. 3A). This can be seen as weaker expression with DEF in the mesoderm and the CNS (Suppl. Fig. 1A). However, the reduction is small and somewhat variable. F facilitates LR pairing more than does C, since expression with DEF is considerably higher than that with CDE at all stages, and the number of cells expressing *lacZ* is also greater (Fig. 3A). Both D and E are required for LR pairing activity, since neither CDF nor CEF show any *eve*-patterned expression (Fig. 3A).

Since DEF has significantly more LR pairing activity than any of the other tripartite combinations, we tested the relative contributions of its three sub-elements. As shown in Fig. 3B, DF and EF have minimal LR pairing activity, as with each, there are at most only a few cells in the embryo in which the reporter responds to *eve* enhancers. By contrast, the DE sub-element combination clearly retains LR pairing activity. However, as evident from the number of cells in which the *eve* enhancers activate reporter expression (compare DEF and DE) and level of expression (particularly at later stages), the F sub-element clearly bolsters the LR pairing activity of the DE combination.

While EF does not have LR pairing activity on its own (Fig. 3B), the EFGH combination shows weak activity (Fig. 2). This fits with our finding that GH also increases the activity of CDEF. In other experiments, we further tested the AB region by creating ABDE in the context of the double reporter transgene (see Fig. 1D). The AB region is unable to enhance the activity of DE, since *lacZ* expression in an *eve* pattern is not increased over that with DE alone (Suppl. Fig. 3).

Insulator protein binding sites are required for long-range pairing

The experiments in the previous section show that D and E harbor most of *homie*'s LR pairing activity. Analysis of the *homie* sequence identified a Su(Hw) consensus site (RAMOS *et al.* 2006; BAXLEY *et al.* 2017) in D, and a Cp190-associated sequence (CUARTERO *et al.* 2014) in E (Suppl. Fig. 7). We tested whether these sites are involved in LR pairing. We mutated the Su(Hw) site in the context of DEF. As shown in Fig. 4, this results in a loss of most but not all of the LR pairing activity. Mutating the Cp190-associated sequence in the context of DEF reduced LR pairing, although the effect is not as drastic as that of the Su(Hw) site mutation (Fig. 4).

To further analyze the effects of the Su(Hw) site mutation, we used smFISH to probe for LR pairing in stage 15 DEF and DEF∆SH dual reporter transgene embryos. In the experiment in Fig. 5, DEF and DEF∆SH dual reporters were inserted into the attP site at –142 kb in two different orientations. In the Z5 orientation, the *LacZ* reporter is closer to the *eve* TAD, while in the 3Z orientation, the *GFP* reporter is closer to the *eve* TAD (see Fig. 1D for diagrams). We then probed for *LacZ* and *GFP* expression in the anal plate region of the embryo. As expected from studies on larger *homie* sub-fragments, the relative orientation of the DEF element in the transgene, but not in the chromosome, determines which reporter is activated: *eve* dependent *LacZ* expression is observed in the anal plate, while *GFP* expression is not detected (Fig. 5). Consistent with the digoxigenin experiments, mutation of the Su(Hw) recognition sequence in the D sub-element impairs LR pairing activity, resulting in a substantial drop in *LacZ* expression (Fig. 5).

The orientation of the DEF element in the transgene determines which of the two reporters is activated, and this is independent of the orientation of the transgene in the chromosome.

On the other hand, our quantitation of *LacZ* mRNA in the anal plate reveals that for DEF, the anal plate enhancer drives a higher level of reporter expression when the dual reporter transgene is inserted in the 3Z orientation (Fig. 5). In this configuration, the topology of the loop linking the reporter to the *eve* TAD is a circle-loop.

Next, we tested whether the Su(Hw) site is fully responsible for the activity of D in the context of our double reporter construct. Here, we added back the 3' half of D, which contains the Su(Hw) site, to EF, which has no LR pairing activity on its own (Fig. 3B), to make D3EF. Fig. 6 shows that the addition of the D3 fragment restores LR pairing activity, and now the *lacZ* reporter is expressed in an *eve* pattern. However, since the activity of D3EF is noticeably weaker than DEF, it would appear that sequences in the 5' half of D must interact with a factor(s) that is able to contribute to LR pairing.

Finally, we tested whether the order of sequences within *homie* is important for its activity. To do this, we reversed the order of EF adjacent to the Su(Hw) site-containing 3' half of D, to make D3FE. As shown in Fig. 6, with the D3FE transgene, the *eve* enhancers do not activate *lacZ* expression. It seemed possible that the change in sub-element order might alter the orientation dependence of D3FE. Since D3FE is inserted in the dual reporter, we were able to test this possibility by assaying GFP expression. Suppl. Fig. 4 shows that GFP is not expressed in the D3FE dual reporter, indicating that LR pairing activity has indeed been lost with the change in sub-element order, not switched in its orientation dependence. Consistent with the maintenance of orientation specificity with these constructs whenever there is significant LR pairing, we found that none of the other reporters tested expresses GFP in an *eve* pattern to any significant degree (DEF, DEFΔSu(Hw): Suppl. Fig. 4).

Contribution of *TER94*-associated elements to LR pairing activity

In the experiments described above, GH showed LR pairing activity in the context of either CDEF or EF *homie*. G contains the *TER94-RA* transcription start site (which is transcribed away from the *eve* locus), while H contains a non-coding lncRNA start site (CR45324, which is transcribed toward the *eve* locus) and the *TER94-RD* transcript initiation site (GRAMATES *et al.* 2022) (see Fig. 1B diagram). Because D3EF showed weak but clear-cut LR pairing activity (Fig. 6), we tested whether the adjacent GH region enhances the LR pairing activity of D3EF. Indeed, D3EFGH shows considerably stronger activity than does D3EF (Suppl. Fig. 5). We then trimmed back the added region to exclude the non-coding RNA start site, which resulted in no apparent loss of LR pairing activity (D3EFGH5, Suppl. Fig. 5). We further trimmed it to just downstream of the *TER94* start site, again with no apparent loss of activity (D3EFG5, Suppl. Fig. 5). This suggests that it is the ~50 bp region just downstream of the F region that harbors most, if not all, of the LR pairing activity of GH. This facilitating region includes the *TER94* start site (to +2 nt), and so may contain a basal promoter element from this housekeeping gene. This increased activity is comparable to, but not quite as strong as, that of CDEF, as evidenced by the more complete *eve* pattern in the late-stage mesoderm in the latter. It is also worth noting our previous finding that *TER94* promoter activity is enhanced by sequences downstream of its start site, which are not included in this insulator-facilitating fragment. So, the sequences facilitating *homie*'s insulating activity and those responsible for *TER94*'s transcriptional activity seem to be mostly, but perhaps not entirely, separable.

Enhancer blocking activity only roughly correlates with other insulator activities

We tested many of the same *homie* fragments described above for their ability to block the *hebe* enhancer upstream of the –142 kb attP site from activating transgene reporters (Fig. 7). Enhancer blocking activity was tested at the two stages of embryogenesis when the *hebe* enhancer is active in cells in the CNS midline (stages 12 and 13). One set of fragments was tested in the context of our single reporter transgene (the eZ vector, Fig. 7A), and another, overlapping set was tested in the double reporter transgene (the eZ-eG vector, Fig. 7B). The two assays gave a consistent order of activities with all of the constructs common to the two assays. However, the enhancer blocking activity of DE in the eZ-eG vector is reduced relative to that seen with the eZ vector, suggesting that eZ-eG provides a better context for distinguishing among stronger enhancer blocking activities, while eZ is better for distinguishing among weaker activities. The double reporter context also allowed us to see the effectiveness of blocking expression from one reporter, located on the opposite side of the *homie* element from the *hebe* enhancer (Suppl. Fig. 1B), at the same time that this enhancer is seen to activate the proximal reporter, providing an internal control (CNS expression in Suppl. Fig. 4).

When CDEFGH is located between the reporter and the *hebe* enhancer, it blocks the enhancer from activating reporter expression (Suppl. Fig. 1A, CDEFGH: only *eve*-like expression is seen in the CNS). While similar results are observed for both CDEF and EFGH, the ABCD combination, like λ DNA, has little or no blocking activity, and so the *hebe* enhancers are able to activate *LacZ* expression in CNS midline cell clusters, mimicking *hebe* expression at this stage (Fig. 2, arrowheads, Suppl. Fig. 2). For the tripartite combinations CDE, DEF, and CDF, blocking activity follows a pattern roughly similar to LR pairing activity (Fig. 7A). DEF is the most effective enhancer blocker, followed by CDE, and then CDF. Because we obtained a transgenic line carrying only one of the two possible orientations of CEF (the 3Z orientation), we were not able to determine its enhancer blocking activity. For the three 2-element combinations DE, DF, and EF, DE appears to be almost as effective in stage 13 embryos as DEF, while both DF and EF have only partial blocking activity (Fig. 3B), roughly paralleling their LR pairing activities.

Since the Su(Hw) and Cp190 sites contribute to the LR pairing activity of DEF, we tested the effects of mutations in these recognition sequences. As was the case for LR pairing activity, mutations in the Cp190 site had a modest effect on DEF's blocking activity. In contrast, mutating the Su(Hw) site led to a substantial loss in blocking activity. Again, these effects on enhancer blocking activity parallel the effects of these alterations on LR-pairing activity. However, the correlation is not perfect for all constructs (summarized in Fig. 10), suggesting that while the two activities are closely related, there are some mechanistic differences.

The clearest example is provided by a comparison of the two *homie* derivatives D3EF and D3FE (in which the order of the E and F regions of DEF *homie* are reversed – neither is inverted in the chromosome, their order is simply switched). While D3FE is unable to mediate LR pairing with the *eve* TAD (Fig. 6), it has more enhancer blocking activity (showing 45% and 40% *hebe*-like midline expression at stages 12 and 13, respectively, Fig. 7B) than D3EF (with 70% and 54% *hebe*-like midline expression, Fig. 7B). This suggests that, at least in the context of the *hebe* locus, LR pairing activity is much more sensitive to the order of binding sites for insulator proteins than is enhancer blocking activity.

Another example is provided by the quantitative effect of the Su(Hw) binding site in DEF *homie*. DEFΔSH has a very low level of enhancer blocking activity, but still a clearly measurable amount of LR pairing activity, while the *gypsy* insulator shows no LR pairing activity, but has strong enhancer blocking activity (Suppl. Fig. 3, Fig. 7B). This suggests that in the context of DEF *homie*, Su(Hw) is particularly important for enhancer blocking, while other *homie*-binding factors that are sufficient for LR pairing contribute little to enhancer blocking. And while Su(Hw) contributes a significant fraction of DEF *homie*'s LR pairing activity, it contributes even more of its enhancer blocking activity. This again speaks to the specificity of our LR pairing assay, which reflects the ability of *homie* to specifically pair with copies of itself when the two are located at considerable chromosomal distances from each other. Enhancer blocking in this assay is a more local phenomenon, one that may not depend on specific pairing with another insulator.

We used the PRE blocking and enhancer blocking assays to further compare the effects of the Su(Hw) binding site in fragment D. D alone, but not E or F alone, has measurable activity in the enhancer blocking assay (Fig. 7C). When the Su(Hw) site is mutated, it abolishes activity in the context of D alone (Fig. 7C). Since E alone has no clearly measurably activity in the enhancer blocking assay, mutating the Cp190-associated site shows no effect there.

PRE blocking activity

We showed previously that *homie* has the ability to block the spread of repressive chromatin (barrier activity) from an *eve* pseudo-locus transgene into the neighboring gene, *TER94* (FUJIOKA *et al.* 2013). *TER94* promoter activity is repressed when *homie* is removed, and this correlates with the spreading of the histone modification H3K27me3, characteristic of Polycomb- (PcG-) dependent repressive chromatin. This loss of expression is not due to reduced *TER94* promoter activity caused by removing *homie*, since removing the *eve* PRE in addition to *homie* fully restores expression (FUJIOKA *et al.* 2013). In this transgenic context, when *homie* is present, *TER94* promoter-driven GFP expression is high in ovaries dissected from adult females.

When full-length *homie* (ABCDEF) is replaced by CDEF, or any of the tripartite combinations (DEF, CEF, CDF, or DEF) *TER94*-driven GFP expression is undiminished (Suppl. Fig. 6A, B). This is illustrated for DEF and "wt" in Fig. 8A. Furthermore, DE alone appears to have full activity (Suppl. Fig. 6A, B), while AB has no barrier activity (Fig. 8A, B). As shown in Fig. 8A and B, we also tested individual *homie* sub-elements. Of these, C has no barrier activity, while E alone is sufficient to completely block PcG spreading. D and F both have partial barrier activity: GFP expression is diminished, but not as much as when *homie* is completely replaced by λ DNA (compare λ DNA with D and E in Fig. 8A). It is also worth noting that this barrier activity does not depend on the orientation of *homie* (Suppl. Fig. 6A, B, CDEF vs. FEDC).

The single *homie* sub-elements that retain barrier activity show some interesting distinctions in their PRE blocking activities. D has strong activity in the somatic follicle cells that form the outer layer of the ovariole, but clearly has less activity in the inner, germline cells (nurse cells and oocyte). In contrast, E seems to have full activity in both the germline cells and the follicle cells. F has somewhat less activity than E overall, and seems to have more activity in

germline cells than in follicle cells (compare D and F to E in Fig. 8A, B). These results show that there can be tissue specificity in barrier activity, at least when somatic follicle cells are compared to ovarian germline cells.

Above, we showed that the Su(Hw) binding site in D is required for most, but not all, of the activity of the D region in the LR pairing assay. We tested the same mutation here, and found that this site is required for the D region to show any barrier activity (Fig. 9A, B). We also tested the requirement for the Cp190-associated site in region E. However, it does not seem to have any effect on the barrier activity of E (Fig. 9A, B). This last set of experiments was performed in a different context than the one shown in Fig. 8. Instead of the full-length *eve* locus in the transgene, we used only the 3' end of the locus, which contains the *eve* 3' PRE, *homie*, and the *TER94* promoter driving GFP (diagrammed at the top of Fig. 9A). Despite the absence of most of the *eve* locus from the transgene, the results for regions D and E (Fig. 9) are very similar to those shown in Fig. 8. However, the germline specificity seen for the F region in Fig. 8 is not seen here. Instead of having less activity in follicle cells, here it shows the opposite specificity, with stronger activity in follicle cells than in germline cells. We note that in addition to the transgene construct being different, the insertion site of the transgene is different. It will require further study to determine why the barrier activity of region F is so sensitive to the chromosomal context, while that of regions D and E do not show this sensitivity. As expected, when the PRE is not present in the transgene, there is no repression of *TER94* promoter-driven GFP expression, even in the absence of any insulator sequences ("no PRE, no *homie*", Fig. 9; here, both the PRE and the starting *homie* fragment are replaced by equal-length stretches of λ DNA).

Discussion

A primary function of boundary elements/insulators in *Drosophila* is the subdivision of the chromosome into a series of looped domains, or TADs. TAD formation is thought to depend upon the physical pairing of neighboring boundary elements. In most cases that have been examined in detail, these pairing interactions are orientation-dependent (KYRCHANOVA *et al.* 2008; FUJIOKA *et al.* 2016). Orientation dependence determines the topology of the TAD. When boundaries pair with their neighbors head-to-head, a circle-loop is generated, while head-to-tail pairing generates a stem-loop. Since the two boundaries that delimit the *eve* TAD, *homie* and *nhomie*, pair with each other head-to-tail, the *eve* TAD is a stem-loop (KE *et al.* 2024). Like other fly boundaries, *homie* and *nhomie* also engage in self-pairing interactions, and in these cases, pairing is head-to-head (FUJIOKA *et al.* 2016). Unlike heterologous pairing, which occurs in *cis*, self-pairing interactions take place between the same boundary element in *trans* and are thought to be important for mediating the juxtapositioning and precise alignment of homologs.

In their endogenous location, *homie* and *nhomie* are separated by only 16 kb; however, both elements can pair with themselves and with each other when separated by multiple TADs and TAD boundaries (FUJIOKA *et al.* 2009; FUJIOKA *et al.* 2016; CHEN *et al.* 2018; BING *et al.* 2024). In addition to this long-range (LR) pairing activity, *homie* can block enhancer promoter interactions (FUJIOKA *et al.* 2009; FUJIOKA *et al.* 2016) and act as a barrier to the spread of PcG silencing (FUJIOKA *et al.* 2013). In the studies reported here, we have undertaken a functional dissection of the *homie* boundary and identified sequences/sub-elements that are important for these three activities.

homie was originally defined as an 800 bp fragment that extended from a site close to the *eve* PRE to just beyond the 1st exon of the neighboring *TER94* gene, a region associated with DNase hypersensitivity (SACKERSON *et al.* 1999). For the purpose of analysis, this 800 bp fragment was subdivided into eight ~100 bp sequences or sub-elements, ABCDEFGH. As shown in Fig. 1, G extends from upstream of the *TER94* transcription start site into the first *TER94-RA* mRNA exon, while H includes sequences from the first exon and the beginning of the first intron, including the start site for a *lncRNA* and the *TER94-RD* mRNA.

Barrier activity: Of the three activities we assayed, the least demanding is blocking PcG silencing. While both LR pairing and enhancer blocking activity require a combination of several sub-elements, single sub-elements are able to block PcG silencing to a greater or lesser extent. Of the sub-elements tested, D, E, and F are able to inhibit the spread of PcG silencing completely on their own (Figs. 8 and 9) . In contrast, AB and C are unable to function as barriers against PcG silencing. Since the *TER94* reporter in the mini-locus that contains a deletion of ABCDEF is also silenced by the *eve* PRE, the GH elements are unable to prevent PcG silencing (Fig. 9). While the E sub-element is as effective in protecting the *TER94* reporter as the larger DEF element or the *homie* element in the intact pseudo-locus (ABCDEFGH), this is not the case for either D of F. The barrier activity of F is about half that of E or DEF and appears to be position-dependent, while the barrier activity of D is tissue specific (Figs. 8 and 9): it is greatly reduced in germline nurse cells, while it is close to that of DEF in somatic follicle cells. Interestingly, in spite of the clear tissue-specific difference in barrier activity, the D sub-element

requires the Su(Hw) binding site in both cell types. We also tested whether the barrier activity of the E sub-element requires the Cp190-associated sequence; it does not.

Enhancer blocking/LR pairing: The requirements for enhancer blocking and LR pairing activity are more stringent than for PcG barrier function. Among the single sub-elements, only D has significant enhancer blocking activity (Fig. 7C), and none of them can mediate LR pairing. While the DE combination has both activities, its LR pairing activity is very weak (Fig. 3B) and its blocking activity is incomplete (Fig. 7B). Of the tripartite sub-element combinations, robust enhancer blocking and LR pairing activity are only observed with the DEF combination. Consistent with a link between enhancer blocking and LR pairing activity, we find that in most but not all cases, the different combinations of sub-elements exhibit a similar trend with respect to their blocking and LR pairing activity. For example, among the *homie* combinations with four sub-elements that we tested (Figs. 2 and 7), CDEF has robust blocking and LR pairing activity, and while EFGH blocks the *hebe* enhancers, blocking is incomplete and LR pairing activity is only modest. The ABCD combination has only very little blocking activity and no LR pairing activity. The same is true for the four tripartite combinations. The blocking activity of DEF is equivalent to that of CDEF, while LR pairing activity is at most only slightly reduced. At the other end of the spectrum, CDF has only weak blocking activity and virtually no LR pairing activity. CDE has somewhat weaker blocking activity , and modest LR pairing. This is likely because the DE element has some blocking and LR pairing activity on its own. The other combinations, DF and EF, retain a low level of blocking activity, but have at most a very low level of LR pairing activity.

Su(Hw) and Cp190 motifs: D has a sequences motif for the Su(Hw) protein while E has a Cp190-associated sequence. Both of these proteins have been implicated in the functioning of other fly boundaries. Consistent with an important role in *homie* activity, mutations in the Su(Hw) site in the DEF combination led to a substantial drop in both blocking and LR pairing activity (Figs. 4 and 7). In contrast, Cp190 appears to have a much less important role, as disruption of the Cp190-associated sequence had only a minor effect on blocking activity and caused just a modest reduction in LR pairing. This was also true for barrier activity. Mutations in the Su(Hw) recognition motif eliminated the barrier activity of the D element, while mutations in the Cp190-associated sequence had minimal if any effect (Figs. 7 and 9).

A previous study of the effects of a mutation in the *Cp190* gene on insulator activities concluded that it affects enhancer blocking but not LR pairing by *homie* (KAUSHAL *et al.* 2022). While our results with the Cp190-associated sequence mutation also implicate Cp190 in enhancer blocking, the effects are not as dramatic as the *Cp190* gene mutation. Likewise, while Kaushal et al. (KAUSHAL *et al.* 2022) observed little if any effect on LR pairing activity, our results suggest that Cp190 contributes to *homie*'s LR pairing activity. It is worth noting here that Cp190 may not bind directly to DNA, but may instead function through interactions with DNA binding proteins, including one that binds to the consensus "Cp190-associated" site (CUARTERO *et al.* 2014) in *homie*.

The D sub-element: Both the enhancer blocking and LR pairing activities of DEF are nearly equivalent to those of CDEF. While mutations in the D sub-element Su(Hw) recognition sequence largely eliminates both blocking and LR pairing activity (Fig. 4, Fig. 7C), combining a 50

bp sequence from the 3' half of D (D3), which contains the Su(Hw) site, with EF (to give D3EF) does not restore the blocking and LR pairing activities of DEF (Fig. 6, Fig. 7B). The blocking activity of the D3EF fragment is close to that of DEF∆SH, while the LR pairing activity is only marginally better than EF by itself. These findings argue that in addition to the Su(Hw) recognition sequence, sequences elsewhere in D are required for both its enhancer blocking and LR pairing activities.

Order of sub-elements is important for LR activity: Our analysis of the D3EF combination uncovered another important feature of *homie* boundary activity. We found that the order of the E and F sub-elements in D3 combination is important for LR pairing interactions, but not for enhancer blocking (Fig. 6, Fig. 7B). Thus, while D3EF mediates LR interactions, changing the order of the EF sub-elements to give D3FE completely disrupts LR pairing activity.

This loss of LR pairing activity makes sense, as it is thought that the order of binding sites for chromosomal architectural proteins like Su(Hw), CTCF and Pita is likely to play an important role in generating orientation-dependent boundary:boundary pairing interactions (MAKSIMENKO *et al.* 2015). Members of this particular class of polydactyl zinc finger proteins have self-interaction domains that can generate multimeric complexes and potentially directly link boundary elements that have binding sites in common (FEDOTOVA *et al.* 2017; BONCHUK *et al.* 2021).

However, it has been suggested that the consensus site should be extended beyond the core, because Su(Hw) contacts sequences flanking the core consensus binding site using a part of the protein that contributes to insulator activity (BAXLEY *et al.* 2017). Since the Su(Hw) site is very near the D-E junction, swapping E and F separates this extended region from the core site, which might cause a reduction in Su(Hw) binding or insulator activity, and thereby contribute to the observed loss of LR pairing. On the other hand, the D region without this extended portion of the Su(Hw) site has both enhancer blocking (Fig. 7) and PRE blocking activities (Fig. 9) that are totally dependent on the core Su(Hw) binding site (D vs. D∆SH), suggesting that the site has strong activity without the extended region. Therefore, it seems unlikely that the large difference in LR pairing activity between D3EF and D3FE is due solely, or even primarily, to a loss of Su(Hw) binding to the core consensus site that is present in both.

Changing the order of E and F in D3EF (which already has rather modest LR pairing activity) to give the D3FE element could interfere with LR pairing activity at several levels. In order for the *eve* enhancers to activate reporter expression at a distance of almost 150 kb and across nearly a dozen intervening TADs, the transgene *homie* must recognize potential partners (*homie* and/or *nhomie*) in the *eve* TAD and "initiate" physical pairing (BING *et al.* 2024). Changing the order of E and F may make it impossible for proteins associate with each sub-element to simultaneously interact with their partners in the endogenous *homie* and *nhomie*, and the initiation step could fail. It is also possible that the initial interaction might involve only a subset of proteins associated with each boundary. In this case, factors associated with a single sub-element might be sufficient to initiate pairing, but steric hinderance arising from the mis-ordered sub-elements could prevent the formation of a sufficient number of physical links (a zippering up) between the boundary elements to generate a stable pairing interaction. Live imaging experiments indicate that after pairing interactions are established and stabilized, activation of the transgene reporter by the *eve* enhancers requires a further interaction that

appears to involve forming contacts between the *eve* enhancers and the promoter of the reporter (CHEN *et al.* 2018). Though less likely, this step might also be perturbed by steric hinderance arising from misaligned sub-element-mediated contacts.

While D3FE is defective in LR pairing, this rearranged boundary nevertheless retains enhancer blocking activity (Fig. 7B). Several mechanisms have been proposed to explain how boundaries block enhancers. Boundaries could act as roadblocks or sinks (GEYER 1997; BLACKWOOD AND KADONAGA 1998; BI AND BROACH 1999; BLANTON *et al.* 2003; GOHL *et al.* 2011), in which case blocking activity is autonomous and will depend on functional properties of the proteins bound to D3EF and D3FE. Alternatively, insulation could be achieved by organizing the chromatin fiber into looped domains (BING *et al.* 2024). In this case, the blocking activity of D3EF and D3FE would depend upon whether boundaries in the neighborhood can pair with the two *homie* derivatives. A 2-dimensional roadblock model cannot explain the pattern of MicroC interactions within TADs, nor would it explain how E-P interactions between sequences that are 142 kb apart result in one reporter being activated and the other nearby reporter not (BING *et al.* 2024). So it seems likely that blocking activity must be envisioned as a process in 3 dimensions that builds on the TAD organization of the chromosome.

Fig. 10 summarizes all our results comparing the *gypsy* insulator and the various *homie* subelements and their derivatives in our assays for LR pairing and enhancer blocking. The constructs are listed in the order of their relative activities in each assay, highest to lowest, from top to bottom. The 3 lines between the two columns connect constructs which are most clearly "in different places" in the two lists. This indicates that their activities do not follow the same trend in the two assays, suggesting that these two insulator activities have some mechanistic distinctions. In particular, *gypsy* does not show LR pairing with the *eve* locus, but is capable of enhancer blocking in our assay, which measures enhancer blocking at short range in *cis*. This distinction may arise from the strong requirement for pairing specificity in the LR pairing assay. The ability to pair with other insulators in the vicinity of the insertion site may be sufficient for enhancer blocking, but not for LR pairing. Alternatively, enhancer blocking may not require insulator pairing at all, even though it is clear that pairing can affect which interactions are blocked and which are facilitated by insulators. Similar considerations may explain the disparate positions of D3FE in the two lists, as discussed above. The discrepancy in the positions of D3EF between the two lists is harder to explain. This activity, harbored within the 5' portion of region D, appears to be more important for enhancer blocking in the context of DEF than it is for LR pairing. Further investigation will be required to understand the basis of this activity, but its existence suggests that there are yet to be uncovered mechanisms underlying these distinctive insulator activities.

Enhancer blocking/PRE blocking: We also compared the enhancer blocking and PRE blocking activities of an extensive set of *homie* subelements. These comparisons provide the clearest indication yet of a mechanistic distinction between enhancer blocking and PRE blocking: as Fig. 7C shows, D has clearly more enhancer blocking activity than E (which has little to none), while the opposite is true in the PRE blocking assay, where E has strong activity, and D's is considerably weaker (particularly in the germline; Figs. 8 and 9). One possible explanation is that PRE blocking only requires the introduction of an extended nucleosome-

depleted region in 1-dimensional space (along the chromosome), while enhancer blocking may require an additional ability to suppress looping in 3 dimensions.

Materials and Methods

Plasmid construction and transgenic production

Construction of eZ (used for Figs. 2, 3, 7A, 7C and Suppl. Fig. 1) and eZ-eG (used for Figs. 4, 5, 6, 7B and Suppl. Figs. 2 – 5) vectors used for the LR interaction and enhancer blocking assays was described previously (FUJIOKA *et al.* 2009; FUJIOKA *et al.* 2016). These vectors carrying modified *homie* sequences were inserted into the attP site at –142kb relative to the *eve* locus (FUJIOKA *et al.* 2009). Analysis of *hebe* enhancer activity (Suppl. Fig. 1B) was done using the eZ vector, inserted into the attP site at cytological location 74A2 (FUJIOKA *et al.* 2013).

The *eve* pseudo-locus construct used for analyzing PRE blocking activity was described previously (FUJIOKA *et al.* 2013). The attP sites at 74A2 (used for Fig. 8) and 95E5 (for Suppl. Fig. 6) were used. The smaller construct used to analyze PRE blocking activity (used for Fig. 9) is the same as the pseudo-locus, except that the sequence upstream of +8.4 kb (relative to the *eve* transcription start site) was removed. The constructs using this vector were inserted into the miMIC 24B1 attP site (VENKEN *et al.* 2011; FUJIOKA *et al.* 2021). In both constructs, the *homie* region from +9.2 to +9.8 kb (relative to *eve* +1) were replaced with modified *homie* and control sequences.

Analysis of transgenic lines

In situ hybridization was performed as described previously (FUJIOKA *et al.* 2013). RNA was visualized using alkaline phosphatase conjugated anti-DIG antibody (Roche), using CBIP and NBT as substrates (Roche). Each set of experiments was carried out with control and experimental samples in parallel to minimize experimental variation. Representative expression patterns are shown in each figure. The same embryo populations were also subjected to counting of CNS midline cell clusters. Visible clusters were counted in about 20 embryos for each construct.

Ovaries were dissected from adult female flies aged 18-22 hr. Live images of ovaries were obtained using the same camera settings throughout, except as noted in the figures. In order to quantify GFP expression, ovaries were subjected to RT-PCR. Total RNA was extracted from 5-8 pairs of ovaries using an RNA extraction kit (Roche). 50 ng of total RNA was used to make cDNA using the Transcriptor First Strand cDNA Synthesis kit (Roche) with random primer. One μl out of 50 μl cDNA was used for each quantitative PCR reaction, each time performed in triplicate. Data are shown as GFP expression normalized to endogenous *TER94* RNA expression, similarly quantified in each sample. Each set of lines was analyzed in three independent assays.

Binding site analysis

Binding site searches was done using MacVector software (MacVector Inc). The Su(Hw) consensus sequences YWGCATACTTTT (RAMOS *et al.* 2006) and NWWWWNYRTWGCATACTTTTKGSDB (BAXLEY *et al.* 2017), and the Cp190-associated consensus sequence GGTTBDWRWMYYNGCTD (CUARTERO *et al.* 2014), were used. The results are shown in Suppl. Fig. 6.

Acknowledgements

This work was supported by NIH grants to Paul Schedl (5R35GM126975) and James B. Jaynes (1R01GM137062). We thank Qing Liu for excellent technical assistance.

References

Aoki, T., A. Sarkeshik, J. Yates and P. Schedl, 2012 Elba, a novel developmentally regulated chromatin boundary factor is a hetero-tripartite DNA binding complex. eLife 1**:** e00171.

Bag, I., S. Chen, L. F. Rosin, Y. Chen, C. Y. Liu *et al.*, 2021 M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity. Nat Commun 12**:** 4170.

Barges, S., J. Mihaly, M. Galloni, K. Hagstrom, M. Muller *et al.*, 2000 The Fab-8 boundary defines the distal limit of the bithorax complex iab-7 domain and insulates iab-7 from initiation elements and a PRE in the adjacent iab-8 domain. Development 127**:** 779-790.

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

Baxley, R. M., J. D. Bullard, M. W. Klein, A. G. Fell, J. A. Morales-Rosado *et al.*, 2017 Deciphering the DNA code for the function of the Drosophila polydactyl zinc finger protein Suppressor of Hairy-wing. Nucleic Acids Res 45**:** 4463-4478.

Bi, X., and J. R. Broach, 1999 UASrpg can function as a heterochromatin boundary element in yeast. Genes Dev 13**:** 1089-1101.

Bing, X., W. Ke, M. Fujioka, A. Kurbidaeva, S. Levitt *et al.*, 2024 Chromosome Structure I: Loop extrusion or boundary: boundary pairing? eLife.

Blackwood, E. M., and J. T. Kadonaga, 1998 Going the distance: a current view of enhancer action. Science 281**:** 60-63.

Blanton, J., M. Gaszner and P. Schedl, 2003 Protein:protein interactions and the pairing of boundary elements in vivo. Genes & Development 17**:** 664-675.

Bonchuk, A., K. Boyko, A. Fedotova, A. Nikolaeva, S. Lushchekina *et al.*, 2021 Structural basis of diversity and homodimerization specificity of zinc-finger-associated domains in Drosophila. Nucleic Acids Res 49**:** 2375-2389.

Chen, H., M. Levo, L. Barinov, M. Fujioka, J. B. Jaynes *et al.*, 2018 Dynamic interplay between enhancer–promoter topology and gene activity. Nature genetics 50**:** 1296.

Chetverina, D., M. Fujioka, M. Erokhin, P. Georgiev, J. B. Jaynes *et al.*, 2017 Boundaries of loop domains (insulators): Determinants of chromosome form and function in multicellular eukaryotes. Bioessays 39**:** 1600233.

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

Fedotova, A., T. Aoki, M. Rossier, R. K. Mishra, C. Clendinen *et al.*, 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. A., A. N. Bonchuk, V. A. Mogila and P. G. Georgiev, 2017 C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors. Acta Naturae 9**:** 47-58.

Fujioka, M., Y. Emi-Sarker, G. L. Yusibova, T. Goto and J. B. Jaynes, 1999 Analysis of an *even-skipped* rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. Development 126**:** 2527-2538.

Fujioka, M., H. Mistry, P. Schedl and J. B. Jaynes, 2016 Determinants of Chromosome Architecture: Insulator Pairing in *cis* and in *trans*. PLoS Genetics 12**:** e1005889.

Fujioka, M., A. Nezdyur and J. B. Jaynes, 2021 An insulator blocks access to enhancers by an illegitimate promoter, preventing repression by transcriptional interference. PLoS Genet 17**:** e1009536.

Fujioka, M., G. Sun and J. B. Jaynes, 2013 The Drosophila *eve* insulator Homie promotes *eve* expression and protects the adjacent gene from repression by Polycomb spreading. PLoS Genetics 9**:** e1003883.

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

Gaszner, M., J. Vazquez and P. Schedl, 1999 The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. Genes & Development 13**:** 2098-2107.

Geyer, P. K., 1997 The role of insulator elements in defining domains of gene expression. Current Opinion in Genetics & Development 7**:** 242-248.

Geyer, P. K., and V. G. Corces, 1992 DNA position-specific repression of transcription by a Drosophila zinc finger protein. Genes & Development 6**:** 1865-1873.

Ghosh, D., T. I. Gerasimova and V. G. Corces, 2001 Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. Embo J. 20**:** 2518-2527.

Gohl, D., T. Aoki, J. Blanton, G. Shanower, G. Kappes *et al.*, 2011 Mechanism of chromosomal boundary action: roadblock, sink, or loop? Genetics 187**:** 731-748.

Gramates, L. S., J. Agapite, H. Attrill, B. R. Calvi, M. A. Crosby *et al.*, 2022 FlyBase: a guided tour of highlighted features. Genetics 220.

Holdridge, C., and D. Dorsett, 1991 Repression of hsp70 heat shock gene transcription by the suppressor of hairy-wing protein of Drosophila melanogaster. Molecular and cellular biology 11**:** 1894-1900.

Jiang, N., E. Emberly, O. Cuvier and C. M. Hart, 2009 Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. Molecular & Cellular Biology 29**:** 3556-3568.

Karch, F., M. Galloni, L. Sipos, J. Gausz, H. Gyurkovics *et al.*, 1994 Mcp and Fab-7: molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of Drosophila melanogaster. Nucleic Acids Research 22**:** 3138-3146.

Kaushal, A., J. Dorier, B. Wang, G. Mohana, M. Taschner *et al.*, 2022 Essential role of Cp190 in physical and regulatory boundary formation. Sci Adv 8**:** eabl8834.

Ke, W., M. Fujioka, P. Schedl and J. B. Jaynes, 2024 Chromosome Structure II: Stem-loops and circle-loops. bioRxiv**:** 2024.2001. 2002.573928.

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

Leon, A., and D. McKearin, 1999 Identification of TER94, an AAA ATPase protein, as a Bam-dependent component of the Drosophila fusome. Molecular Biology of the Cell 10**:** 3825-3834.

Li, H.-B., M. Muller, I. A. Bahechar, O. Kyrchanova, K. Ohno *et al.*, 2011 Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in Drosophila melanogaster. Molecular & Cellular Biology 31**:** 616-625.

Li, J., and D. S. Gilmour, 2013 Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. Embo j 32**:** 1829-1841.

Maksimenko, O., M. Bartkuhn, V. Stakhov, M. Herold, N. Zolotarev *et al.*, 2015 Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. Genome research 25**:** 89-99.

Maksimenko, O., A. Golovnin and P. Georgiev, 2008 Enhancer-promoter communication is regulated by insulator pairing in a Drosophila model bigenic locus. Molecular & Cellular Biology 28**:** 5469-5477.

Modolell, J., W. Bender and M. Meselson, 1983 Drosophila melanogaster mutations suppressible by the suppressor of Hairy-wing are insertions of a 7.3-kilobase mobile element. Proceedings of the National Academy of Sciences of the United States of America 80**:** 1678-1682.

Moon, H., G. Filippova, D. Loukinov, E. Pugacheva, Q. Chen *et al.*, 2005 CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. EMBO Rep 6**:** 165-170.

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

Özdemir, I., and M. C. Gambetta, 2019 The Role of Insulation in Patterning Gene Expression. Genes (Basel) 10.

Pai, C. Y., E. P. Lei, D. Ghosh and V. G. Corces, 2004 The centrosomal protein CP190 is a component of the gypsy chromatin insulator. Mol Cell. 16**:** 737-748.

Parkhurst, S. M., D. A. Harrison, M. P. Remington, C. Spana, R. L. Kelley *et al.*, 1988 The Drosophila su(Hw) gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative DNA-binding protein. Genes Dev. 2**:** 1205-1215.

Peifer, M., and W. Bender, 1988 Sequences of the gypsy transposon of Drosophila necessary for its effects on adjacent genes. Proceedings of the National Academy of Sciences of the United States of America 85**:** 9650-9654.

Pinter, M., G. Jekely, R. J. Szepesi, A. Farkas, U. Theopold *et al.*, 1998 TER94, a Drosophila homolog of the membrane fusion protein CDC48/p97, is accumulated in nonproliferating cells: in the reproductive organs and in the brain of the imago. Insect Biochemistry & Molecular Biology 28**:** 91-98.

Ramírez, F., V. Bhardwaj, L. Arrigoni, K. C. Lam, B. A. Grüning *et al.*, 2018 High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nature communications 9**:** 189.

Ramos, E., D. Ghosh, E. Baxter and V. G. Corces, 2006 Genomic organization of gypsy chromatin insulators in Drosophila melanogaster. Genetics 172**:** 2337-2349.

Rath, U., Y. Ding, H. Deng, H. Qi, X. Bao *et al.*, 2006 The chromodomain protein, Chromator, interacts with JIL-1 kinase and regulates the structure of Drosophila polytene chromosomes. Journal of Cell Science 119**:** 2332-2341.

Roy, S., M. K. Gilbert and C. M. Hart, 2007 Characterization of BEAF mutations isolated by homologous recombination in Drosophila. Genetics 176**:** 801-813.

Ruden, D. M., V. Sollars, X. Wang, D. Mori, M. Alterman *et al.*, 2000 Membrane fusion proteins are required for oskar mRNA localization in the Drosophila egg chamber. Developmental Biology 218**:** 314-325.

Sackerson, C., M. Fujioka and T. Goto, 1999 The *even-skipped* locus is contained in a 16-kb chromatin domain. Dev. Biol. 211**:** 39-52.

Schwartz, Y. B., and G. Cavalli, 2017 Three-Dimensional Genome Organization and Function in Drosophila. Genetics 205**:** 5-24.

Schwartz, Y. B., T. G. Kahn, D. A. Nix, X. Y. Li, R. Bourgon *et al.*, 2006 Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nature Genetics 38**:** 700-705.

Sigrist, C. J., and V. Pirrotta, 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.

Small, S., A. Blair and M. Levine, 1992 Regulation of even‐skipped stripe 2 in the Drosophila embryo. The EMBO journal 11**:** 4047-4057.

Small, S., A. Blair and M. Levine, 1996 Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. Developmental Biology 175**:** 314-324.

Tolhuis, B., E. de Wit, I. Muijrers, H. Teunissen, W. Talhout *et al.*, 2006 Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster.[erratum appears in Nat Genet. 2006 Jul;38(7):850]. Nature Genetics 38**:** 694-699.

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

Venken, K. J. T., K. L. Schulze, N. A. Haelterman, H. Pan, Y. He *et al.*, 2011 MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. Nature Methods 8**:** 737-743.

Viets, K., M. E. Sauria, C. Chernoff, R. R. Viales, M. Echterling *et al.*, 2019 Characterization of button loci that promote homologous chromosome pairing and cell-type-specific interchromosomal gene regulation. Developmental cell 51**:** 341-356. e347.

Zhao, K., C. M. Hart and U. K. Laemmli, 1995 Visualization of chromosomal domains with boundary element-associated factor BEAF-32. Cell. 81**:** 879-889.

Zhou, J., H. Ashe, C. Burks and M. Levine, 1999 Characterization of the transvection mediating region of the abdominal-B locus in Drosophila. Development 126**:** 3057-3065.

Zolotarev, N., A. Fedotova, O. Kyrchanova, A. Bonchuk, A. A. Penin *et al.*, 2016 Architectural proteins Pita, Zw5,and ZIPIC contain homodimerization domain and support specific long-range interactions in Drosophila. Nucleic Acids Res 44**:** 7228-7241.

Figure legends

Fig. 1. Maps of the *eve* TAD and transgenes inserted at –142kb from the *eve* locus.
A: map of the *eve* locus. Blue: early stripe enhancers (Small *et al.* 1992; Small *et al.* 1996; Fujioka *et al.* 1999; Sackerson *et al.* 1999). Pink: neuronal enhancers (Fujioka *et al.* 1999; Sackerson *et al.* 1999). Orange: tissue-specific enhancers (Fujioka *et al.* 1999; Sackerson *et al.* 1999). Yellow: PRE. Red: insulators *homie* (H) and *nhomie* (N) (Fujioka *et al.* 2009; Fujioka *et al.* 2016). Green arrowheads: DNase I hypersensitive sites (Sackerson *et al.* 1999). **B: blowup of the *homie* region, subelements ABCDEFGH.** The *TER94-RA* transcript starts in G. Both the *TER94-RD* transcript and the *lncRNA:CR45324* start in H (Gramates *et al.* 2022). The positions of the Su(Hw) binding site and the Cp190-associated site (Ramos *et al.* 2006; Cuartero *et al.* 2014; Baxley *et al.* 2017) are shown as red ovals. **C, D. the eZ-CDEF and eZ-CDEF-eG transgenes.** The –142kb attP site is located in the 5' UTR of the *hebe* transcription unit. Because the attP site was originally inserted via P-element transposition, the 5'- and 3'-P-element ends flank it in the genome. The 3'-P-end ("3'P") is located closer to the CNS midline enhancer of the *hebe* gene, while the 5'-P-end ("5'P") is closer to the *eve* locus. The *hebe* transcription unit (all major transcripts start at the same site, according to Flybase) (GRAMATES *et al.* 2022) is shown as a blue arrow. Insertion of a transgene via RMCE then results in two recombined versions of the original attP and attB sites, which become the two "attR" sites (BATEMAN *et al.* 2006). With RMCE, transgene insertion can occur in either orientation: "Z5", where *lacZ* is closer to 5'P, or "3Z", where *lacZ* is closer to 3'P. In the case of Z5, a modified *homie* is located between 3'P and the *eve-lacZ* reporter gene (direction of transcription is shown as a black arrow). In the case of 3Z, modified *homie* is located between *eve-lacZ* and 5'P, as illustrated. The orientation of *homie* is indicated by the order of its subelements: either "CDEF" for the same orientation in the chromosome as endogenous *homie*, or "FEDC", for the opposite orientation. The eZ-CDEF-eG vector is the same, except for the additional presence of an *eve-GFP* reporter (green arrow), transcribed in the opposite direction as *eve-lacZ*. Each reporter is driven by the *eve* basal promoter, which confers no *eve*-like expression on its own (FUJIOKA *et al.* 2016).

Fig. 2. 400 bp of *homie* (CDEF) contains virtually all of its long-range pairing activity.
The previously described 800bp "full-length" *homie* (FUJIOKA *et al.* 2009) was further dissected, by dividing it into roughly 100bp segments (A – H) and testing them in our functional assays. Each tested region is shown as black lines on the map. "λ DNA" is 500bp of phage λ DNA, which serves as a negative control. *lacZ* expression from eZ-vector transgenes (in the Z5 orientation) is shown at embryonic stages 5, 7, 11, and 13 (from top to bottom in each column). In stage 13 embryos (bottom row) with both ABCD and λ DNA, midline cell clusters in the CNS are seen, which are easily distinguished from smaller *eve*-positive neurons, seen in the 1st two columns. Some of these are indicated with arrowheads for ABCD.

Fig. 3. Functional dissection of the CDEF fragment of *homie*.
A. *lacZ* expression of eZ-vector transgenes is shown, in either the Z5 (top panel) or 3Z orientation (bottom panel), at embryonic stages 5, 7, 11, and 13 (from top to bottom in each column). In stage 13 CDF embryos, clusters of midline cells are indicated by arrowheads. Note

that DEF contains most of the activity of CDEF. **B.** Further dissection of the DEF subfragment. Transgenes in the Z5 orientation are shown, arranged as in A. Note that only DEF and DE contains significant elements of expression in an *eve* pattern, as well as the ability to block expression driven by the *hebe* CNS midline enhancer, as indicated by the absence of large midline clusters of expression (bottom row).

Fig. 4. A consensus binding site for Su(Hw) and a Cp190-associated site each contribute to long-range pairing of *homie*.
lacZ expression is shown from eZ-eG transgenes in the Z5 orientation at embryonic stages 5, 7, 11, and 13 (from top to bottom in each column). "ΔSu(Hw)" and "ΔCp190" indicate versions with the single consensus sites for those proteins point-mutated. Note that mutating the Su(Hw) site in the context of DEF *homie* causes a severe loss of its LR pairing activity, while mutating the Cp190 site causes a clear but incomplete loss of LR pairing (see text).

Fig. 5. The single consensus Su(Hw) binding site in *homie* contributes to its LR pairing activity.
A, B. smFISH analysis was performed on stage 15 DEF and DEFΔSH dual reporter transgene embryos. DEF and DEFΔSH dual reporters were inserted into the attP site at –142 kb in two different orientations. In the Z5 orientation, the *LacZ* reporter is closer to the *eve* TAD, while in the 3Z orientation, the *GFP* reporter is closer to the *eve* TAD. We then probed for *LacZ* and *GFP* expression in the anal plate region of the embryo (A). Note that *eve*-like *LacZ* expression is observed in the anal plate, while *GFP* expression is not detected, and that mutation of the Su(Hw) recognition sequence in the D sub-element impairs LR pairing activity, resulting in a substantial drop in *LacZ* expression, quantified in B ($P < 0.001$).

Fig. 6. The order of insulator elements is important for LR pairing activity. *lacZ* expression is shown from eZ-eG transgenes in the Z5 orientation, at embryonic stages 5, 7, 11, and 13 embryos (from top to bottom in each column). In D3EF, the 5' half of D was removed, while in D3FE, the orientation of section EF was reversed relative to that in D3EF. The orientation of EF is indicated as blue arrows in the diagram at the top. Note that removing the 5' half of section D causes a partial loss of LR pairing, although not as much as mutating the Su(Hw) site in the 3' half of D (Fig. 3B), and that reversing the EF region abolishes detectable LR pairing.

Fig. 7. Enhancer blocking activity of *homie* and its derivatives. Average numbers of midline clusters visible at stages 12 and 13 in transgenic embryos are shown as box-and-whiskers plots. Those with a higher number of visible clusters have less enhancer blocking activity. The *homie* subelements present in the transgene are listed below each pair of plots: names are as in the text; e.g., "ΔSH" has the Su(Hw) consensus binding site mutated and "ΔCp" has the Cp190-associated site mutated. Constructs were ordered based on their relative levels of activity in the enhancer blocking and LR pairing assays. **A:** eZ vector transgenes. **B: Strength of enhancer blocking does not always correlate with LR pairing activity:** eZ-eG vector transgenes. Images are reproduced from Fig. 4 for comparison. **C: Strength of enhancer blocking does not always correlate with PRE blocking activity.** Analysis of enhancer blocking by individual *homie* subelements and their derivatives in eZ vector transgenes shows that only D, and neither DΔSH, E, EΔCp, nor F has significant enhancer blocking activity (top graph; DE is

reproduced from part A for comparison), while E has stronger PRE blocking activity than does D (bottom graph, reproduced from Fig. 8B for comparison).

Fig. 8. Multiple, non-overlapping *homie* subelements are sufficient for PRE blocking activity. A. Top: map of the *eve* pseudo-locus (FUJIOKA *et al.* 2013). Below: live images of *TER94-GFP* in ovaries from transgenic flies carrying the pseudo-locus or modified versions, as follows: **wt**: *homie* intact; **DEF, D, E, F, AB, C**: ABCDEF *homie* replaced with these subelements. In each case, λ DNA was used to make their spacing similar to that in intact *homie*; **λ DNA**: the same sequence used in Fig. 2 replaced ABCDEF. For λ DNA, AB, and C, longer exposures of the same ovaries are shown at the right. **B.** *GFP* RNA expression in dissected ovaries, normalized to endogenous *TER94* RNA in the same samples, was quantified using RT-PCR. Top panel: 3 independent RT-PCR results are shown. Bottom panel: averages and standard deviations of the same 3 results.

Fig. 9. The Su(Hw) binding site, but not the Cp190-associated site, contributes to PRE blocking. A. Top: map of the PRE blocking assay construct, which consists of the 3' end of the *eve* pseudo-locus, starting just upstream of the PRE. Below: live images of *TER94-GFP* in ovaries from transgenic flies carrying the blocking assay construct or modified versions, as follows: **wt**, **D, E, F, λ DNA**: same as in Fig. 7C. **D-ΔSH, E-ΔCP**: the same point-mutations of the Su(Hw) and Cp190 binding sites in *homie* used in Fig. 7C; **no PRE, no *homie***; the *eve* PRE and ABCDEF *homie* together were replaced by 1.3 kb of phage λ DNA. For λ DNA and D-ΔSH, longer exposures of the same ovaries are shown at the right. **B.** *GFP* RNA expression in dissected ovaries, normalized to endogenous *TER94* RNA in the same samples, was quantified using RT-PCR. Top panel: 3 independent RT-PCR results are shown. Bottom panel: averages and standard deviations of the same 3 results.

Fig. 10. Relative strengths of enhancer blocking and LR pairing activities do not always correlate. Constructs were ordered based on their relative levels of activity in the enhancer blocking and LR pairing assays. Lines connect those (in bold) that clearly differ in their order of activity in the two assays. Note: those shown in green lettering were tested in the eZ vector, while those in black lettering were tested in the eZ-eG vector; CDEF, DEF, and DE were tested in both.

Supplemental Figure legends

Suppl. Fig. 1. A. Expression of *eZ-lacZ* transgenes inserted at –142 kb in *eve*(+) embryonic mesoderm and CNS. *lacZ* RNA expression in transgenic embryos carrying the indicated *homie* fragments. Left panel: dorsal view of stage 13 embryos. Arrowheads indicate mesodermal cells. Right panel: ventral view of stage 13 embryos. Arrowheads indicate *lacZ* expression in EL (black) and CQ (red) neurons. **B. Mapping of *hebe* enhancers located near the transgene insertion site at –142 kb from endogenous *eve*.** The 1st intron of *hebe* was analyzed for enhancer activity using the eZ vector. *lacZ* expression from eZ-vector transgenes carrying

fragments from the *hebe* gene: *hebe*-1 (2.3 kb), *hebe*-2 (2.5 kb), and *hebe*-3 (2.8 kb), each inserted at the 74A2 attP site (on chromosome 3, where there is no interaction with the *eve* locus). Note that the *hebe* CNS midline enhancer is located in *hebe*-2, and that *hebe*-1 contains an anal plate ring enhancer active only after stage 16.

Suppl. Fig. 2. Negative control for transgenes inserted at –142 kb from endogenous *eve*. Reporter gene RNA expression from eZ-eG vector transgenes inserted in the indicated orientations and carrying 500 bp of λ DNA, in embryos at stages 5, 7, 11, and 13 (from top to bottom), used as a negative control for eZ-eG vector transgenes carrying *homie* derivatives.

Suppl. Fig. 3. LR pairing activities of *homie* sub-elements and a *gypsy* element. *lacZ* RNA expression from eZ-eG transgenes in the Z5 orientation at embryonic stages 5, 7, 11, and 13 (from top to bottom in each column). **CDEF, DEF, DE, ABDE**: map at top shows the analyzed regions of *homie* as black lines. ***gypsy***: a 451 bp *gypsy* transposable element was tested in the same context.

Suppl. Fig. 4. *GFP* RNA expression from transgenes used in Fig. 4 and 6. Embryonic stages 5, 7, 11, and 13 (from top to bottom in each column) are shown. Note that no *eve*-like expression is discernible, and that expression is strong in *hebe* CNS midline cell clusters.

Suppl. Fig. 5. Sequences within *TER94* do not contribute to LR pairing by *homie*. Diagram (top): the green and red arrows show the directions of transcription of *TER94-RA* and non-coding *CR45324* RNAs, respectively. Shown is *lacZ* expression from eZ-eG transgenes in the Z5 orientation at embryonic stages 5, 7, 11, and 13 (from top to bottom in each column). CDEF is the same as in Suppl. Fig. 3, and D3EF is the same as in Fig. 5. In D3EFG5, D3EF is extended to just downstream of the *TER94* start site; in D3EFGH5, it is extended to just downstream of the *CR45324* start site; in D3EFGH, it is extended through the end of *homie* subregion H. Note that each of the 3'-extended versions of D3EF have about the same activity, which is noticeably stronger than that of D3EF, but not quite as strong as that of CDEF, especially in the stage 13 mesoderm.

Suppl. Fig. 6. A. Live images of *TER94-GFP* expression in ovaries carrying pseudo-locus transgenes. The ABCDEF region of the pseudo-locus inserted at 95E5 (2013) was replaced with *homie* elements, as follows. **wt**: intact pseudo-locus; **ΔIns**: ABCDEF *homie* deleted; **CDEF, FEDC, DEF, CDE, CDF, CEF, DE**: these regions of *homie* replaced ABCDEF. In DEF, CDE, CDF, CEF, and DE, DNA from phage λ was inserted to make the position of the remaining region similar to that in the intact pseudo-locus. Note that all except ΔIns retain full PRE blocking activity. **B. *GFP* RNA expression in transgenic ovaries measured using RT-PCR.** *GFP* RNA expression was normalized to endogenous *TER94* RNA expression. **Top panel**: 3 independent RT-PCR results are graphed. **Bottom panel**: averages and standard deviations of the same 3 results are shown. Labeling is the same as in A, except that "Δhomie" is used in place of "ΔIns".

Suppl. Fig. 7. DNA and primer sequences. See the main text and figure legends for how these sequences were used.

Fig. 1

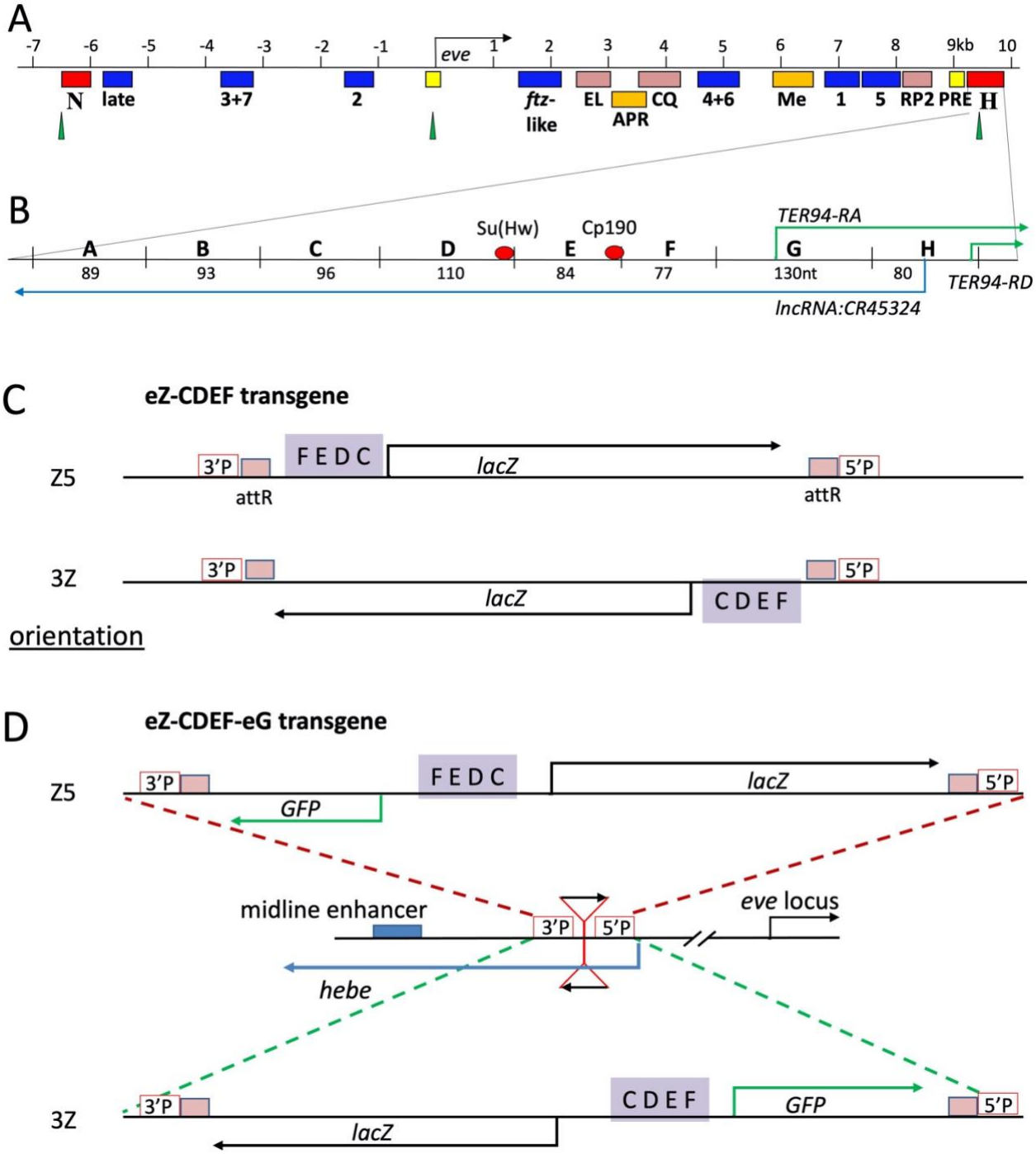

A
-7
-6
-5
-4
-3
-2
-1
eve
1
2
3
4
5
6
7
8
9kb
10
N
late
3+7
2
ftz-like
EL
APR
CQ
4+6
Me
1
5
RP2
PRE
H
B
A
B
C
D
E
F
G
H
89
93
96
110
84
77
130nt
80
Su(Hw)
Cp190
TER94-RA
TER94-RD
lncRNA:CR45324
C
eZ-CDEF transgene
Z5
3'P
attR
F E D C
lacZ
5'P
attR
3Z
3'P
5'P
lacZ
C D E F
orientation
D
eZ-CDEF-eG transgene
Z5
3'P
F E D C
lacZ
5'P
GFP
midline enhancer
3'P
5'P
eve locus
hebe
3Z
3'P
C D E F
GFP
5'P
lacZ

Fig. 2

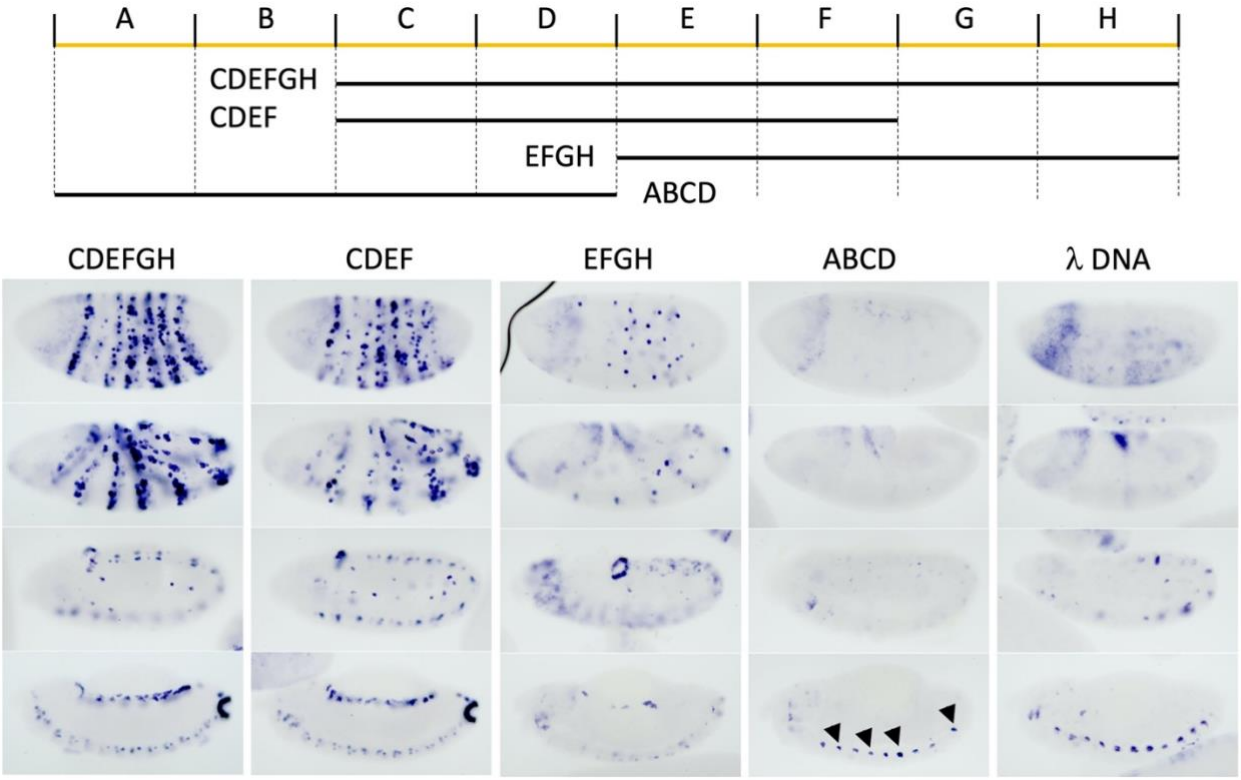


Fig. 3

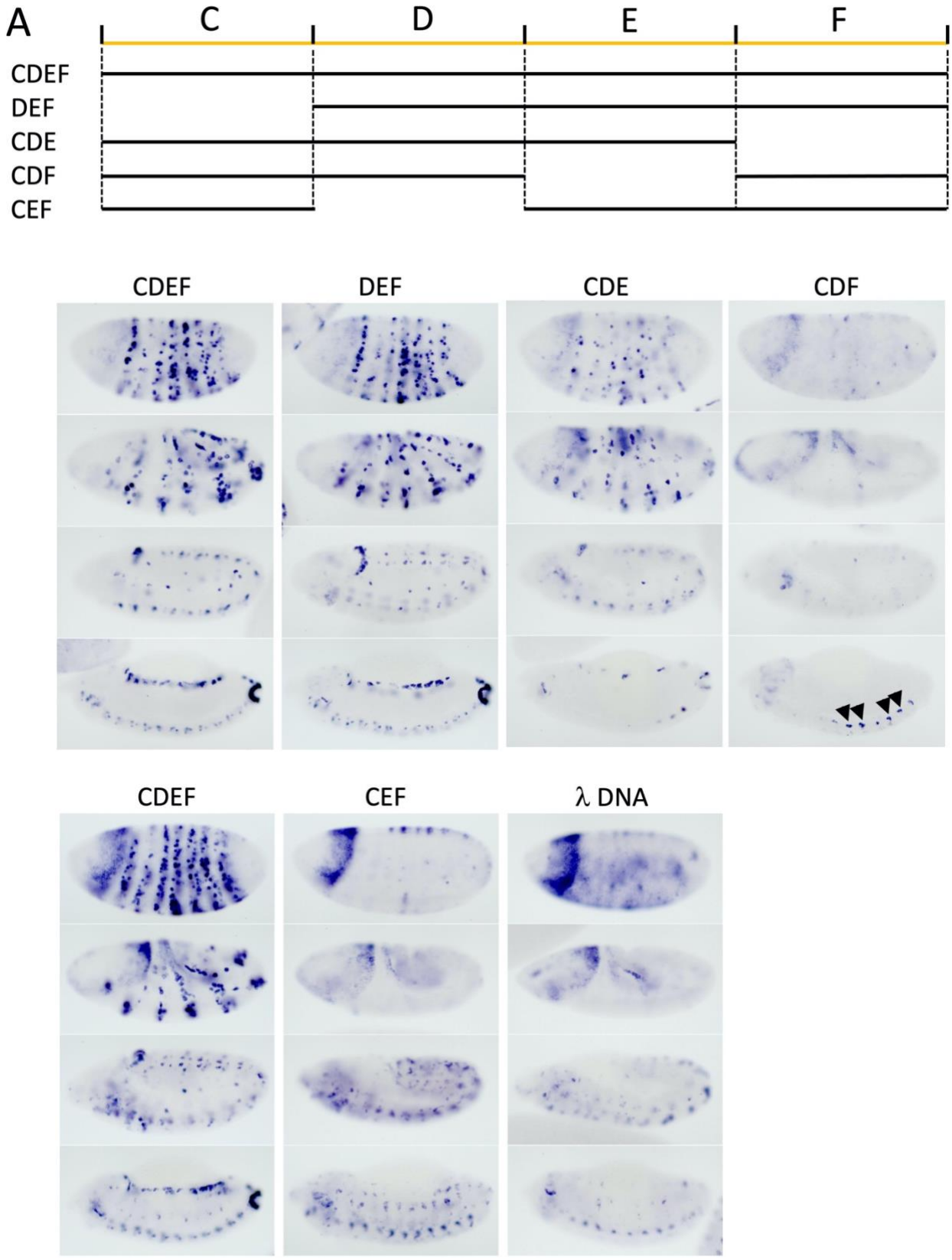


Fig. 3 (cont.)

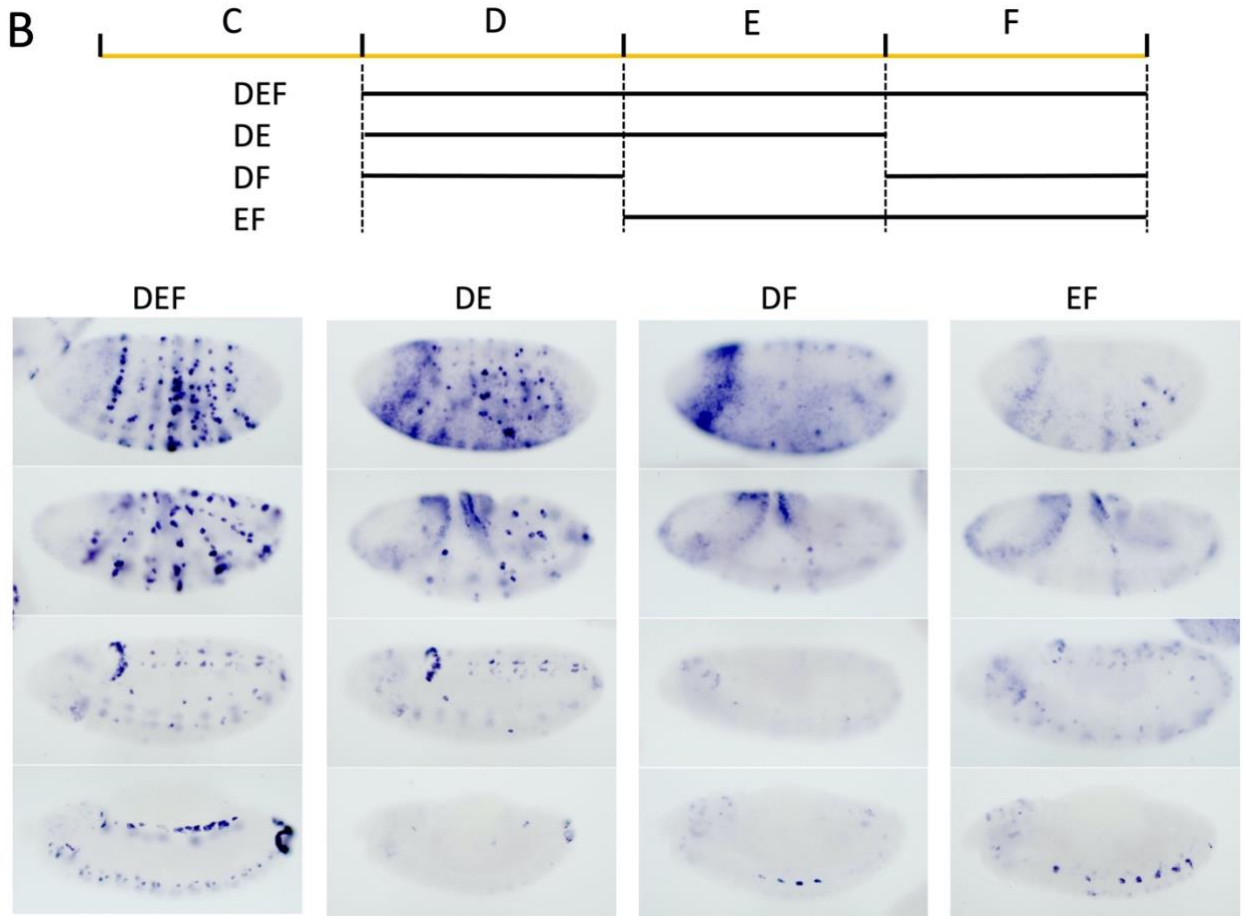


Fig. 4

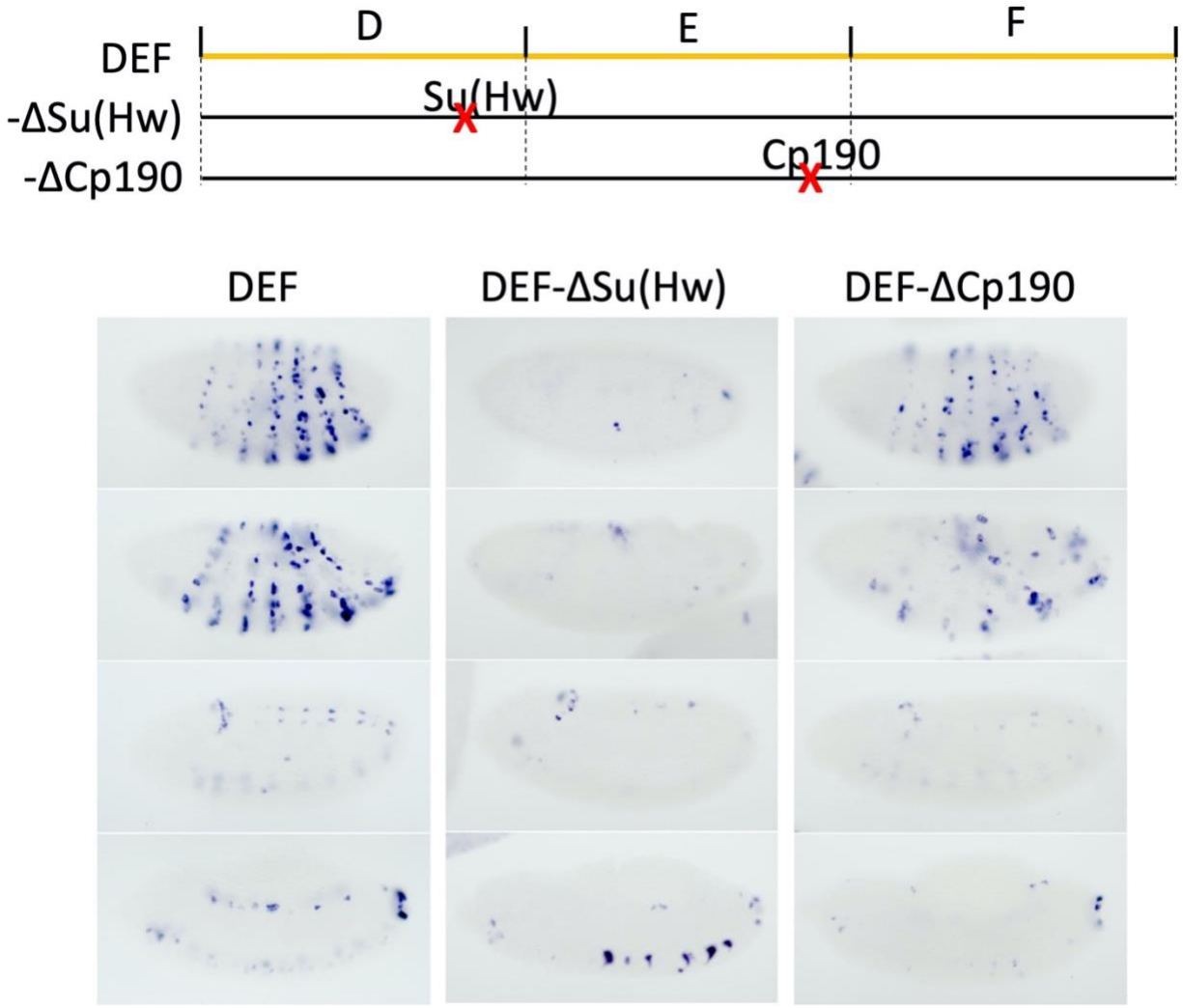


Fig. 5

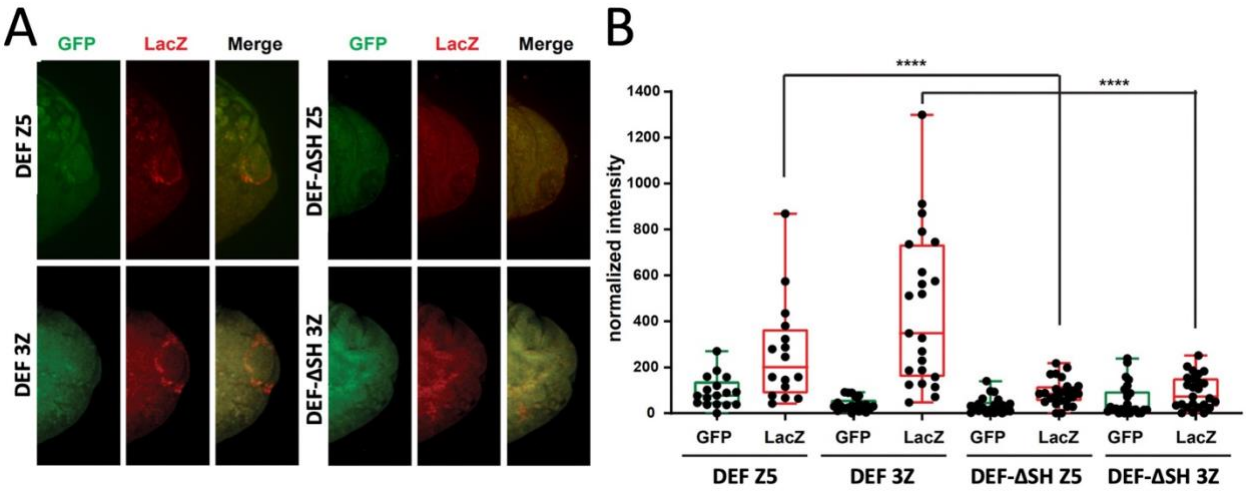

A
GFP
LacZ
Merge
GFP
LacZ
Merge
DEF Z5
DEF-ΔSH Z5
DEF 3Z
DEF-ΔSH 3Z
B
****
****
normalized intensity
1400
1200
1000
800
600
400
200
0
GFP
LacZ
GFP
LacZ
GFP
LacZ
GFP
LacZ
DEF Z5
DEF 3Z
DEF-ΔSH Z5
DEF-ΔSH 3Z

Fig. 6

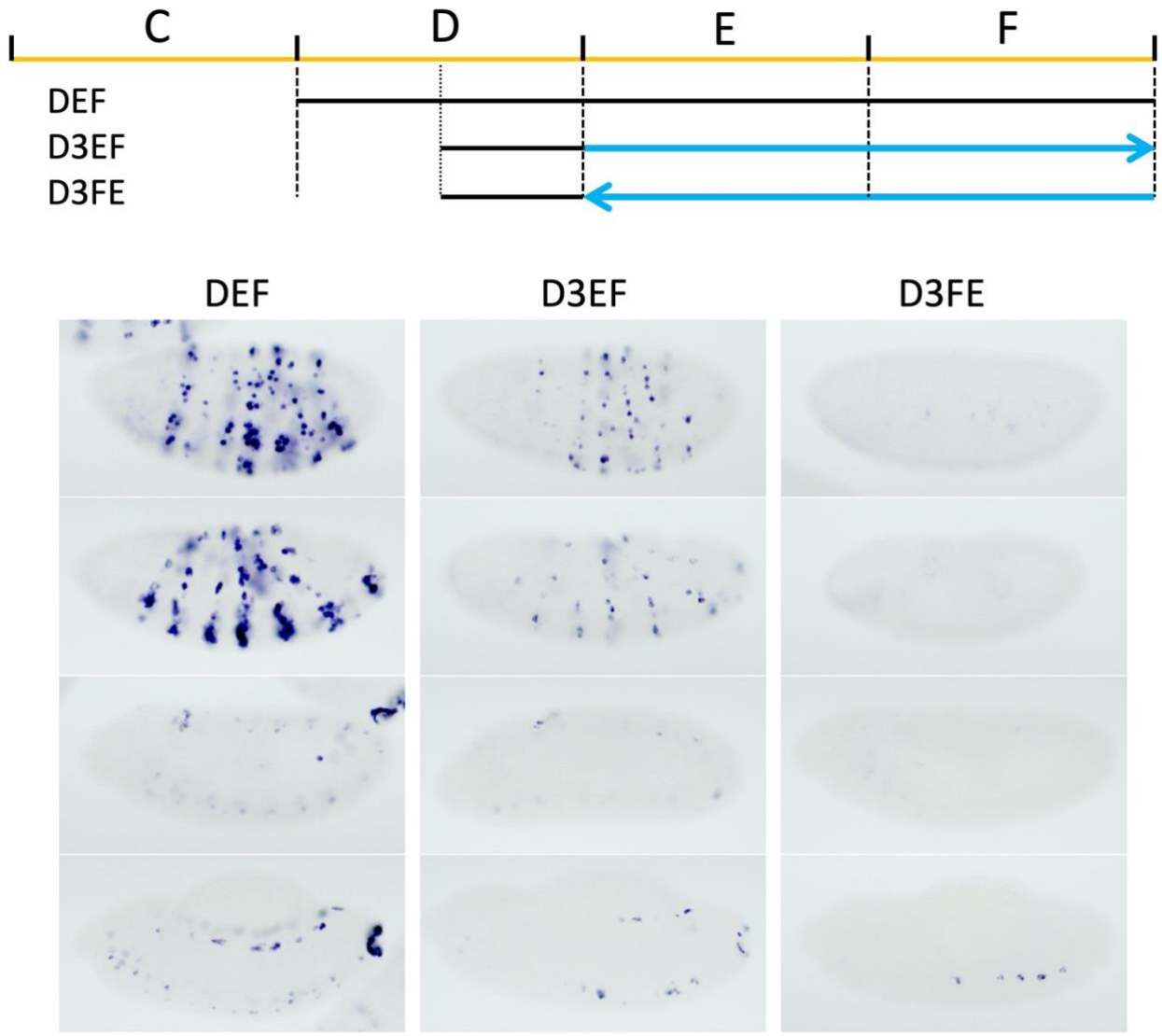


Fig. 7

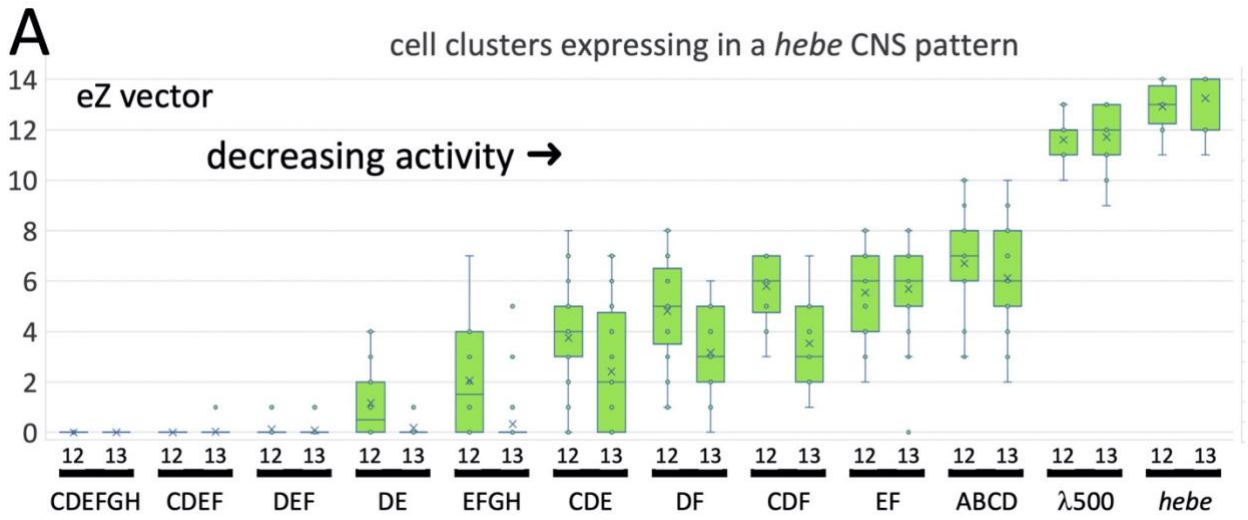
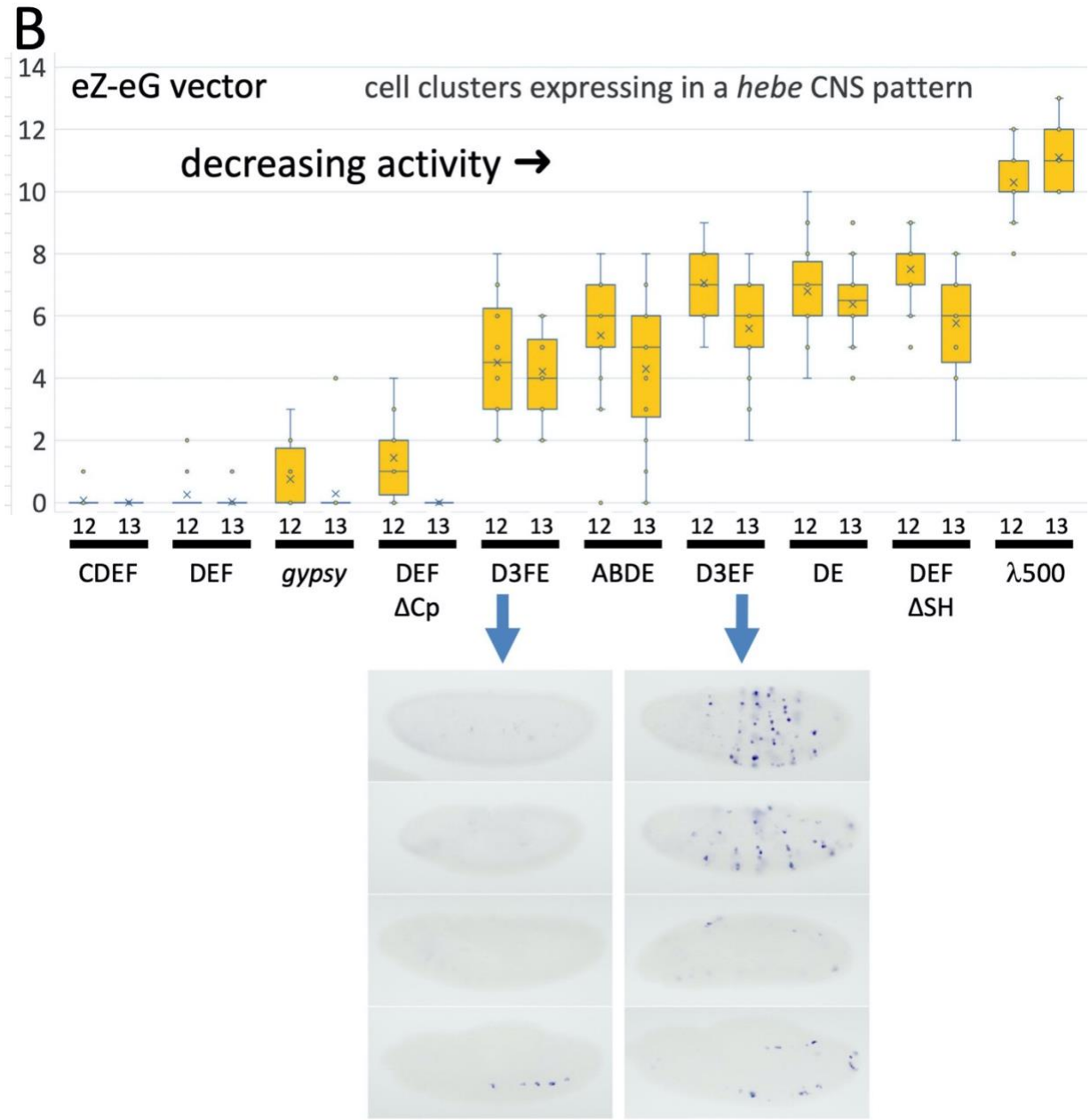
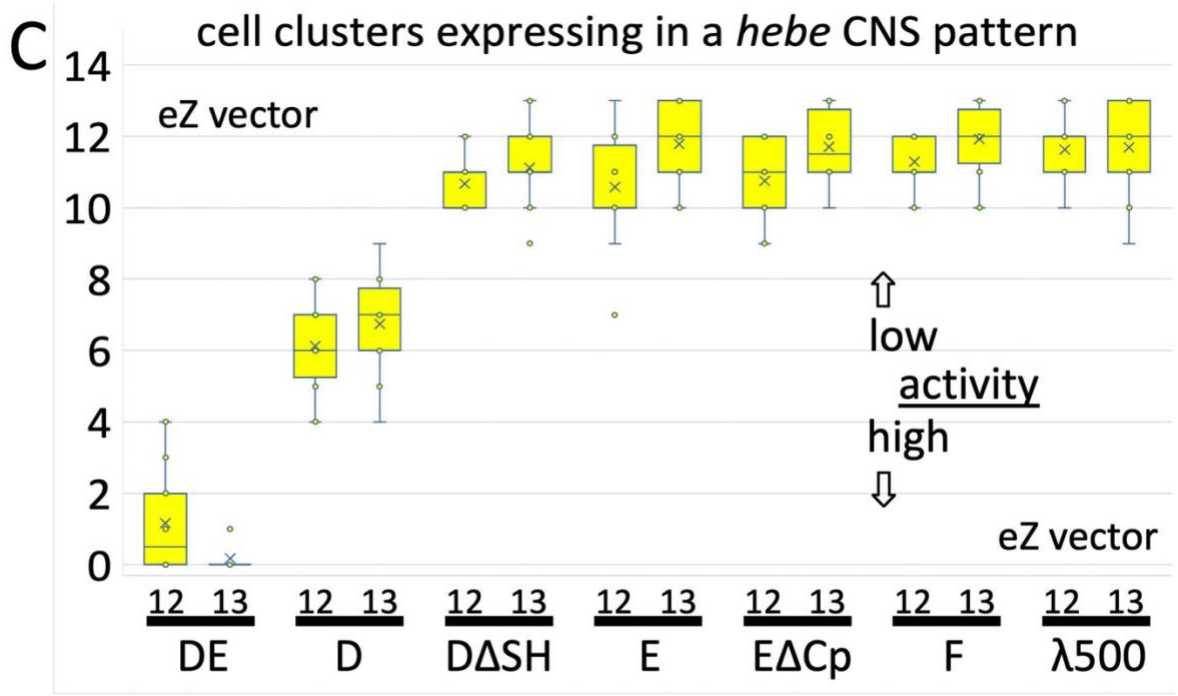

A
cell clusters expressing in a *hebe* CNS pattern
eZ vector
decreasing activity →
14
12
10
8
6
4
2
0
12 13
CDEFGH
CDEF
DEF
DE
EFGH
CDE
DF
CDF
EF
ABCD
λ500
hebe

Fig. 7 (cont.)



D3FE enhancer blocking is a bit better, while LR pairing by D3EF is clearly better

Fig. 7 (cont. 2)



enhancer blocking:
D > E

PRE blocking:
E > D

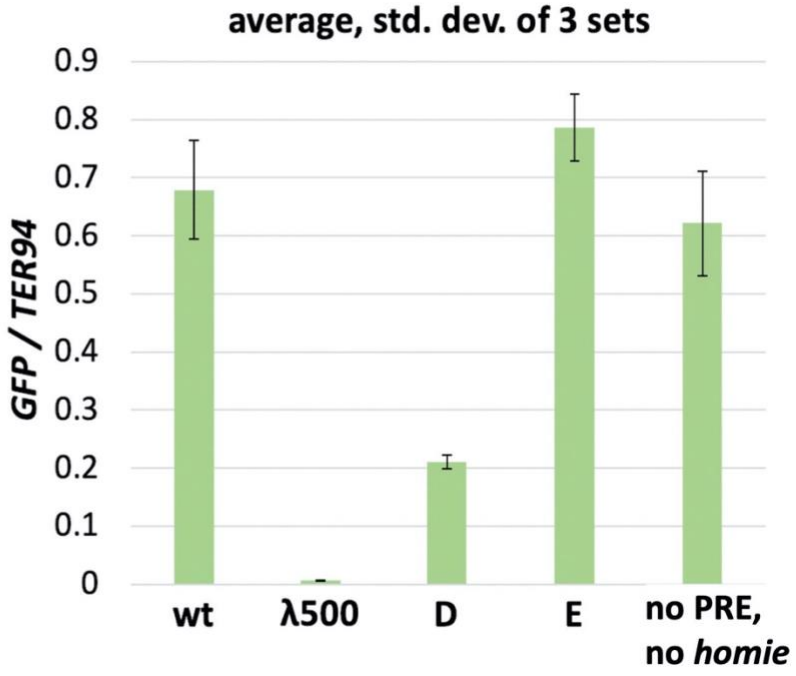


Fig. 8

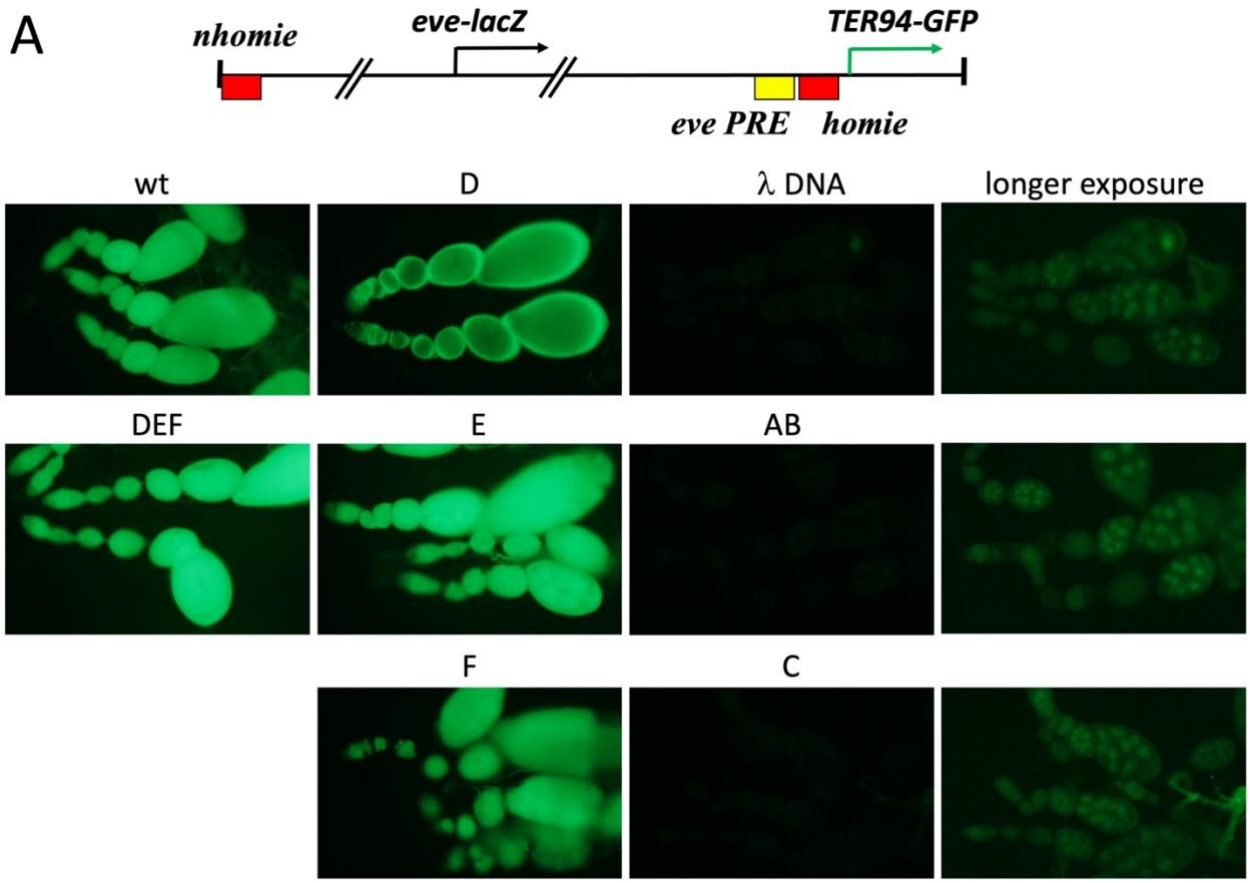
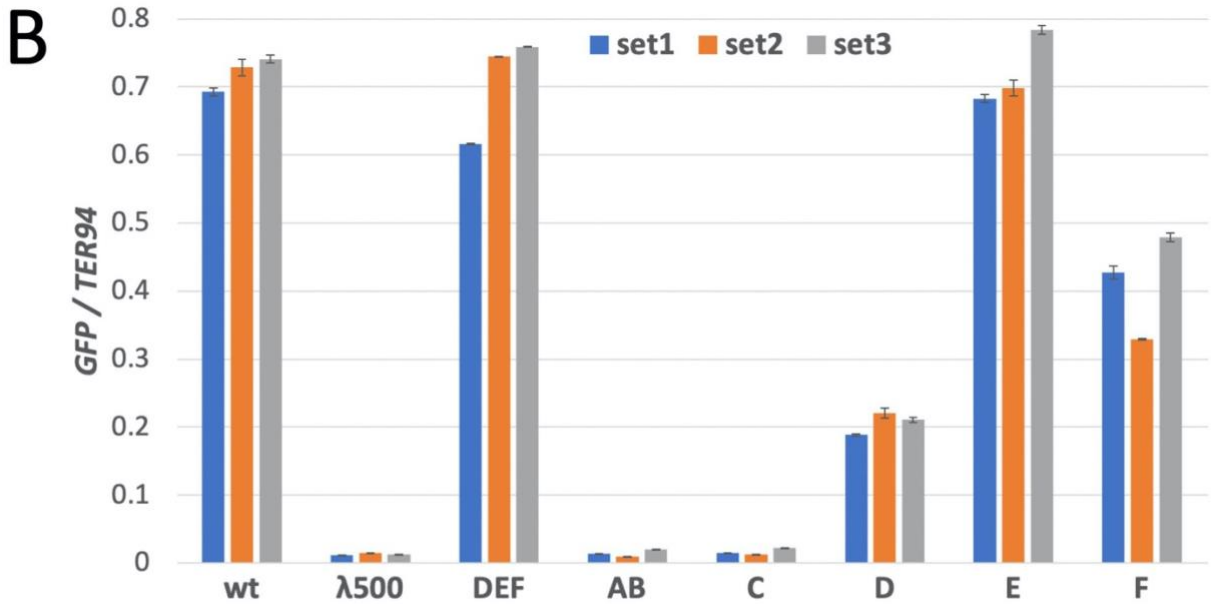

A
nhomie
eve-lacZ
TER94-GFP
eve PRE
homie
wt
D
λ DNA
longer exposure
DEF
E
AB
F
C

Fig. 8 (cont.)


B
set1
set2
set3
GFP / TER94
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
wt
λ500
DEF
AB
C
D
E
F

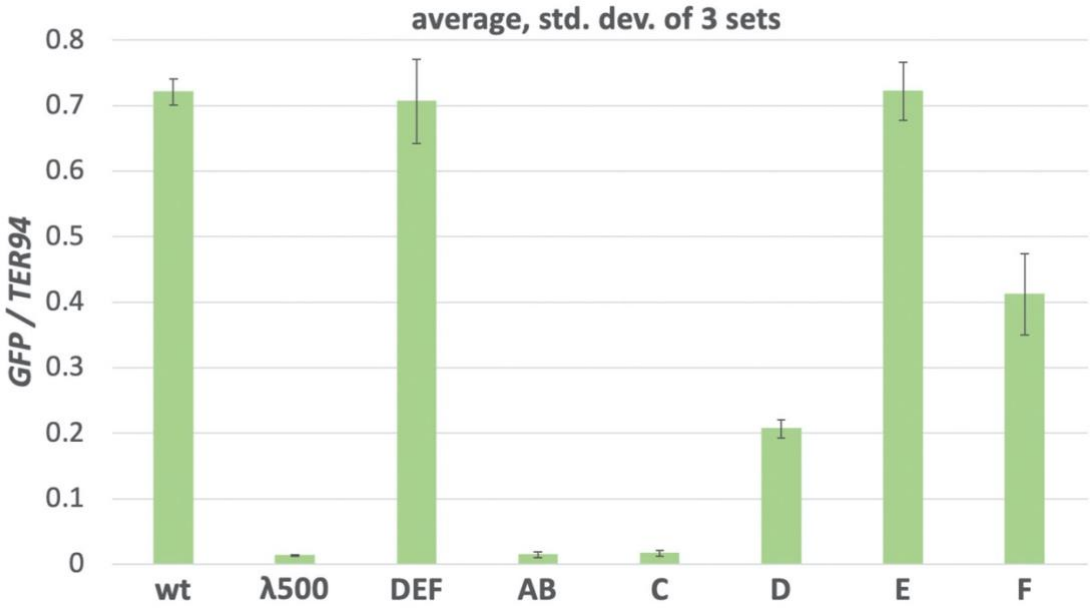

average, std. dev. of 3 sets
GFP / TER94
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
wt
λ500
DEF
AB
C
D
E
F

Fig. 9

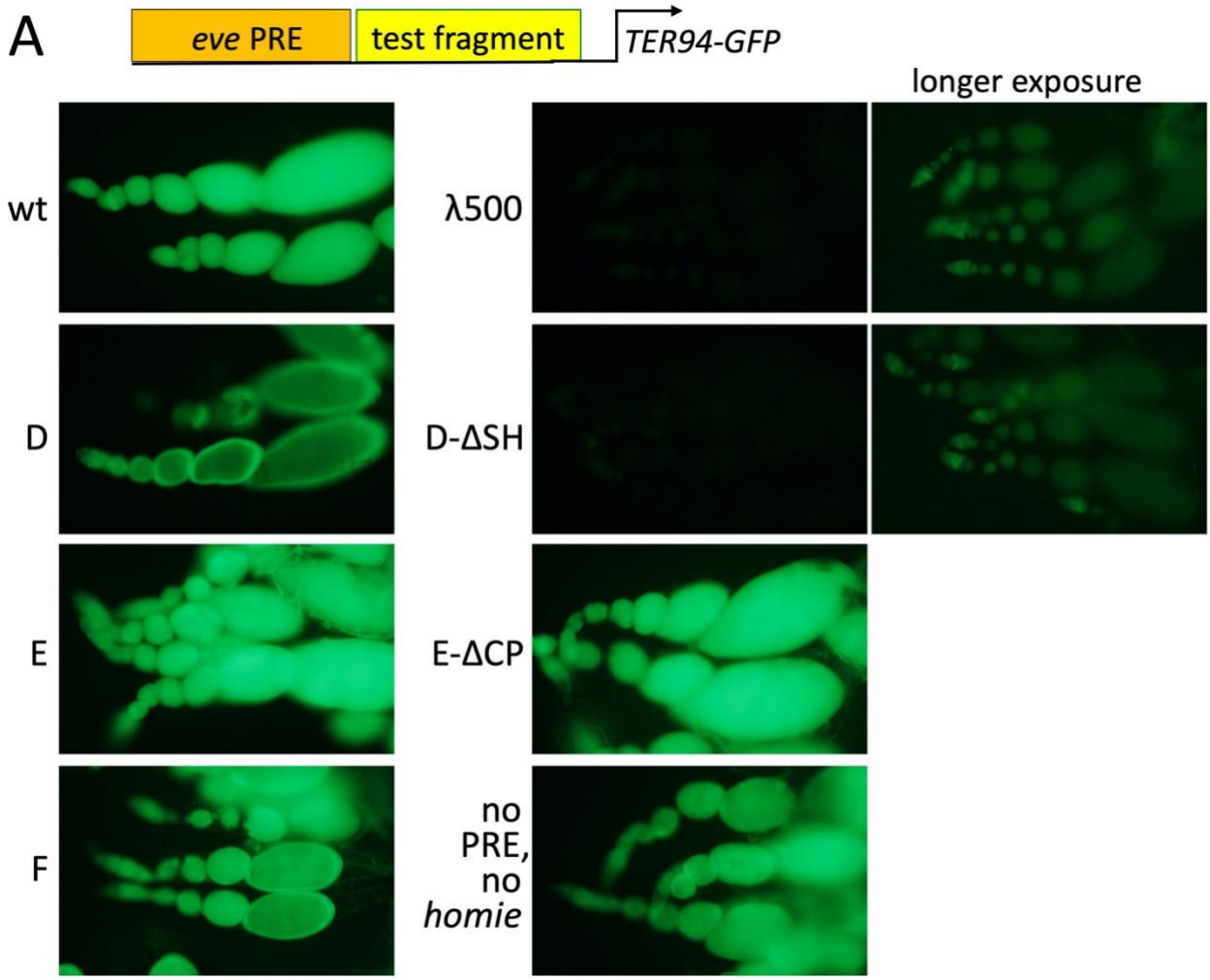

A
eve PRE
test fragment
TER94-GFP
longer exposure
wt
λ500
D
D-ΔSH
E
E-ΔCP
F
no PRE, no homie

Fig. 9 (cont.)

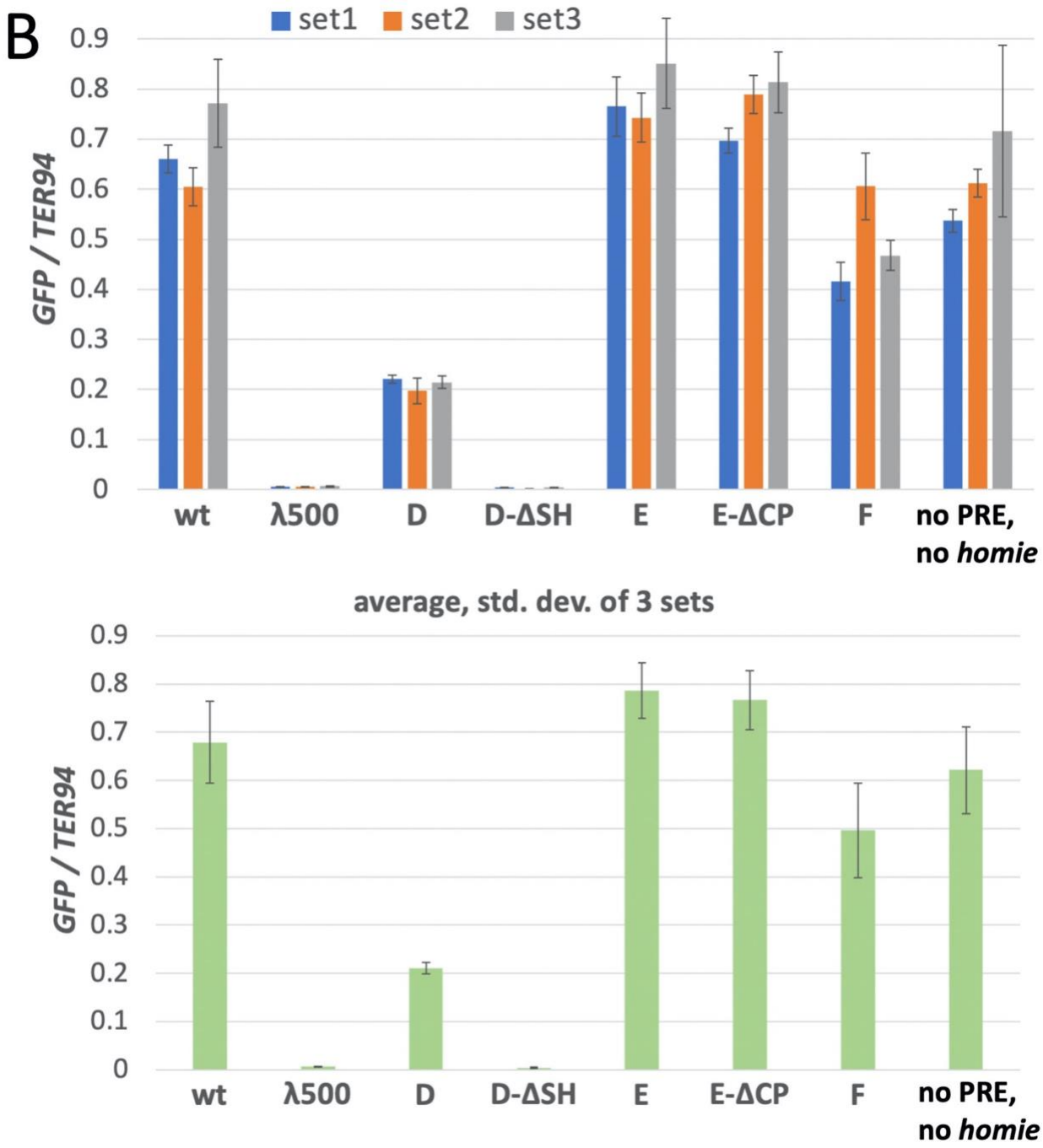
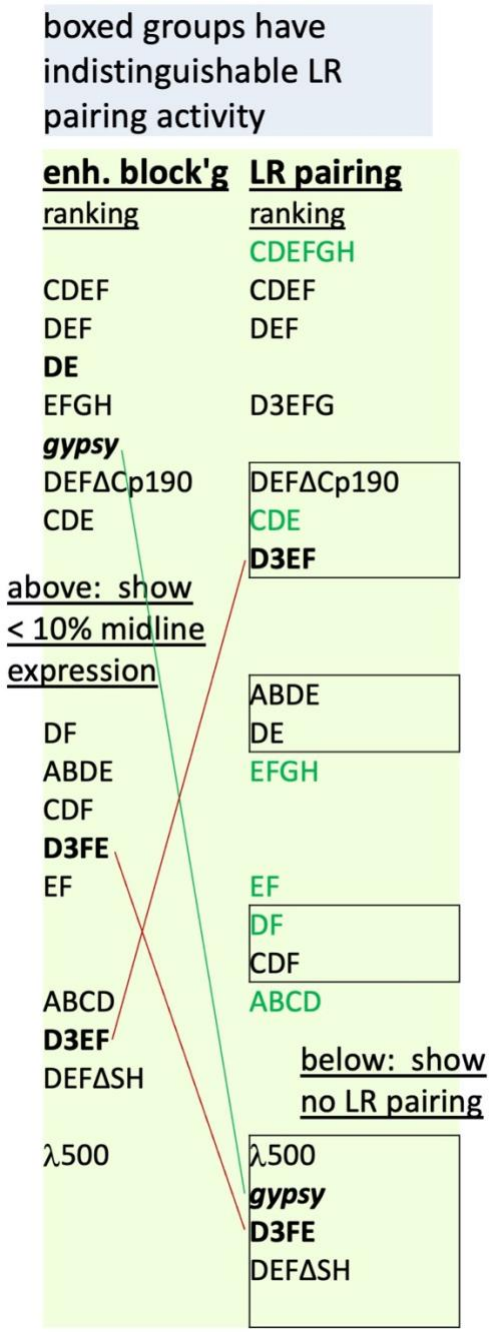
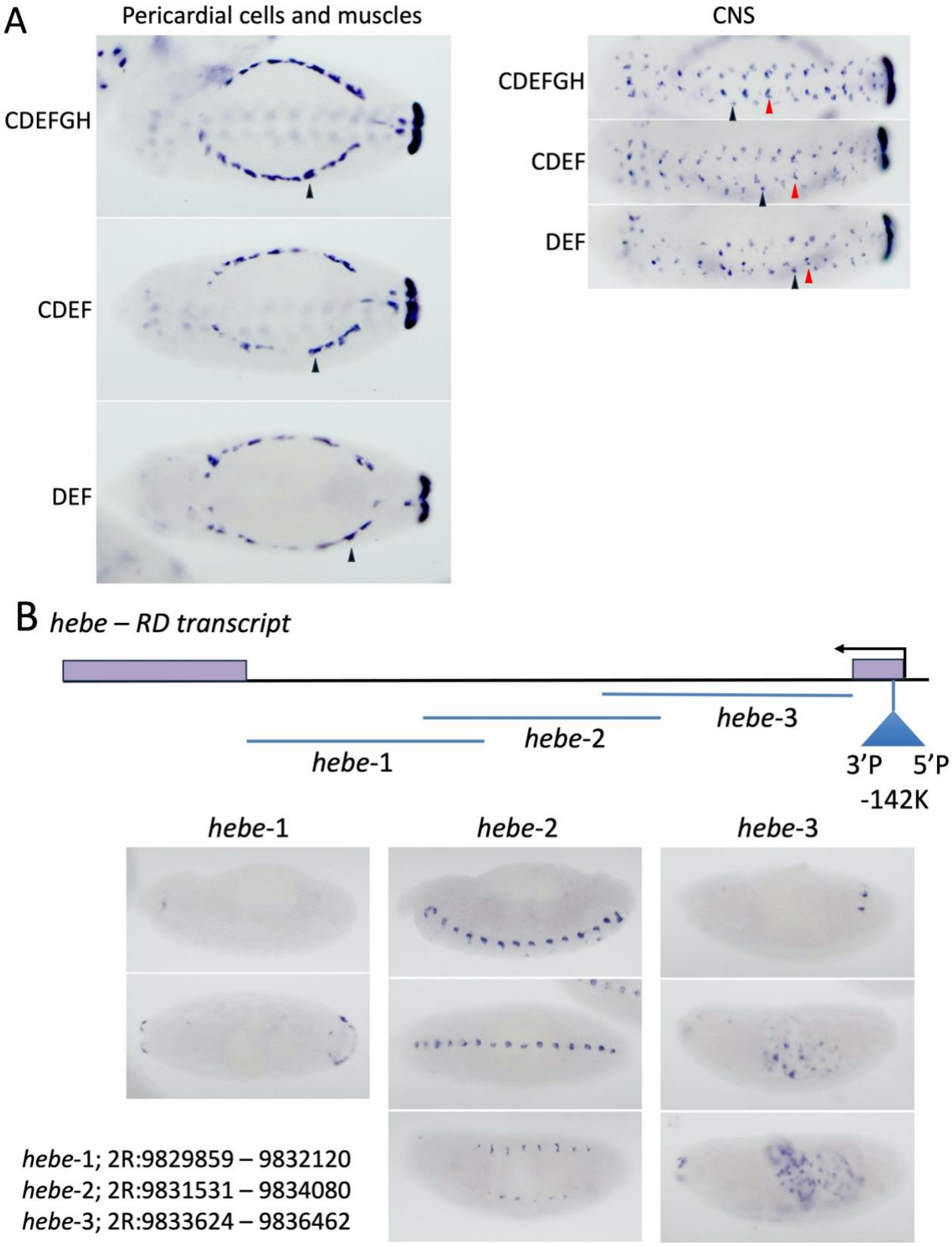

B
set1
set2
set3
GFP / TER94
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
wt
λ500
D
D-ΔSH
E
E-ΔCP
F
no PRE, no homie
average, std. dev. of 3 sets
GFP / TER94
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
wt
λ500
D
D-ΔSH
E
E-ΔCP
F
no PRE, no homie

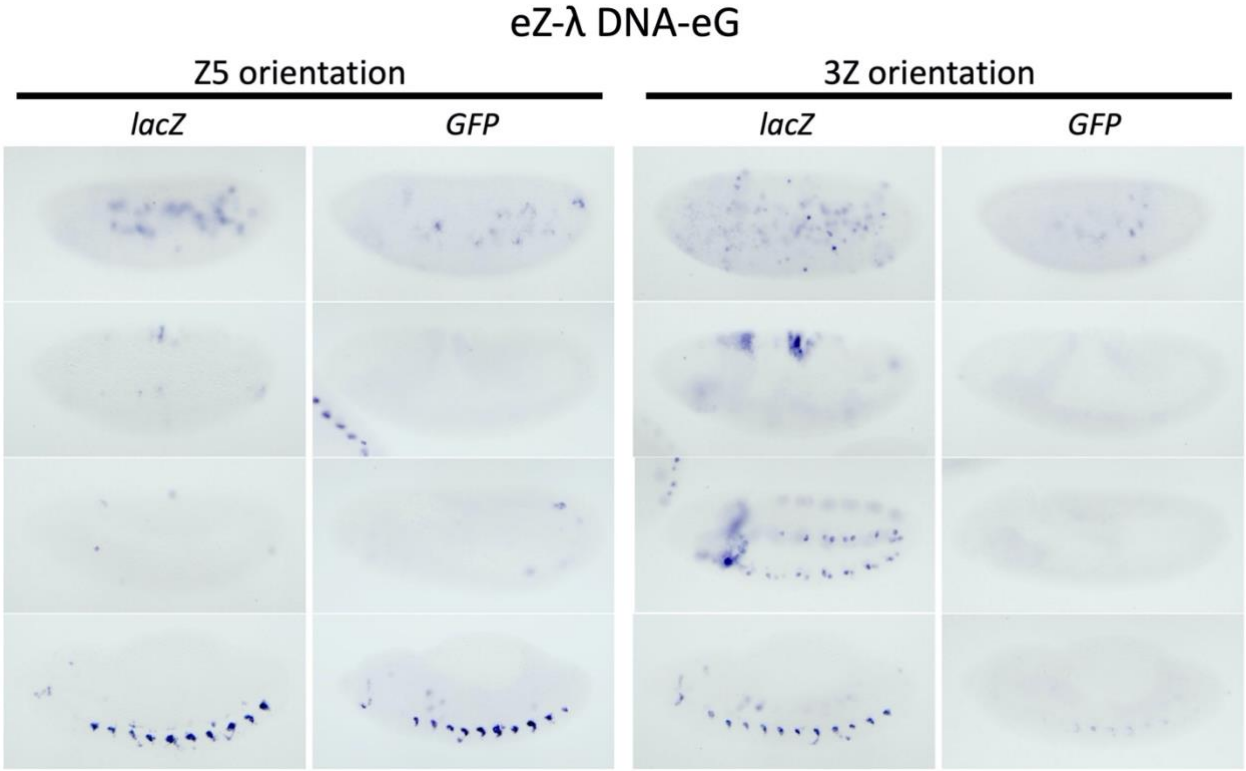
Fig. 10


boxed groups have indistinguishable LR pairing activity
enh. block'g ranking
LR pairing ranking
CDEFGH
CDEF
DEF
DE
EFGH
gypsy
DEFΔCp190
CDE
above: show < 10% midline expression
DF
ABDE
CDF
D3FE
EF
ABCD
D3EF
DEFΔSH
λ500
CDEF
DEF
D3EFG
DEFΔCp190
CDE
D3EF
ABDE
DE
EFGH
EF
DF
CDF
ABCD
below: show no LR pairing
λ500
gypsy
D3FE
DEFΔSH

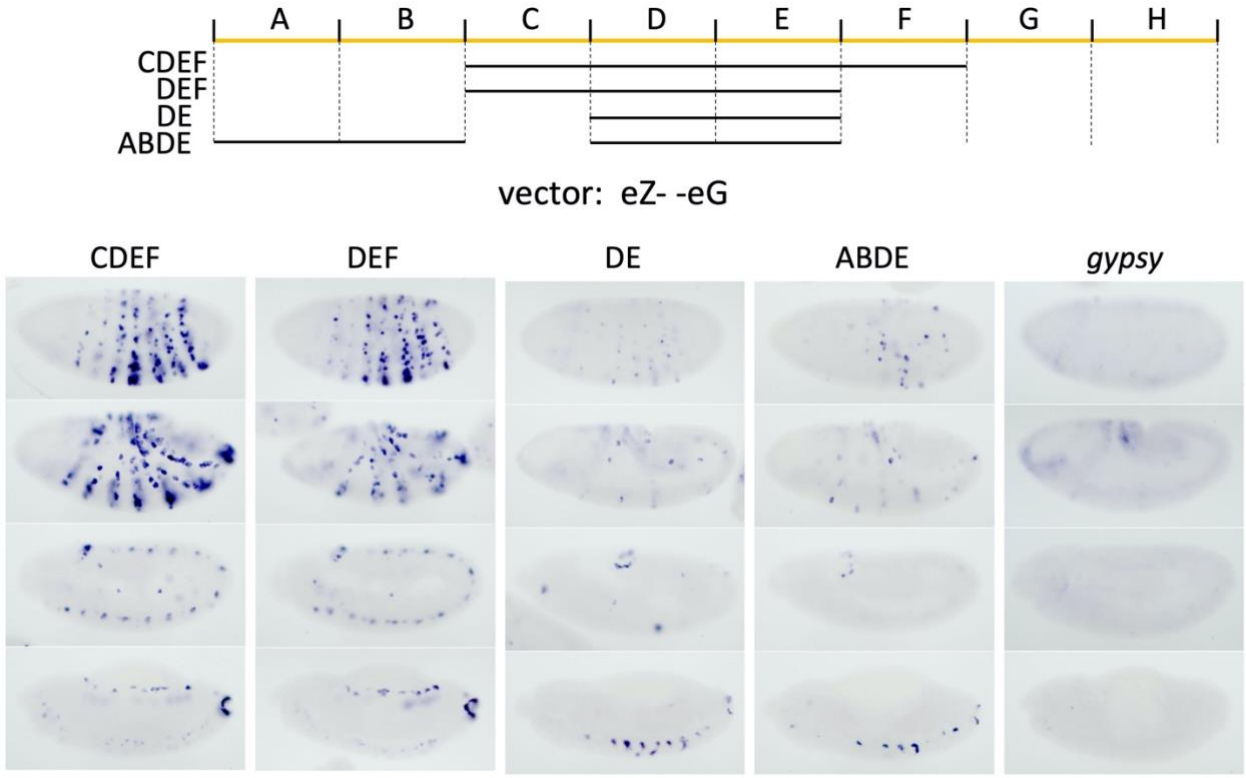
Supplemental Fig. 1



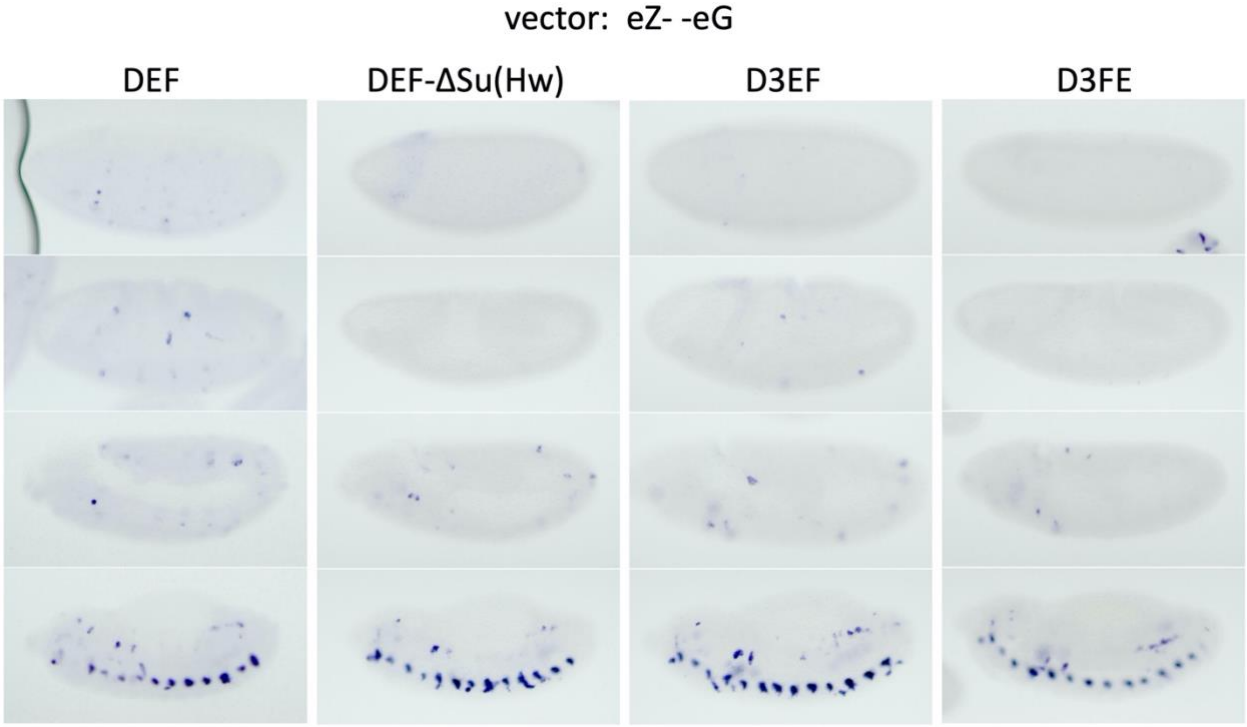
Supplemental Fig. 2



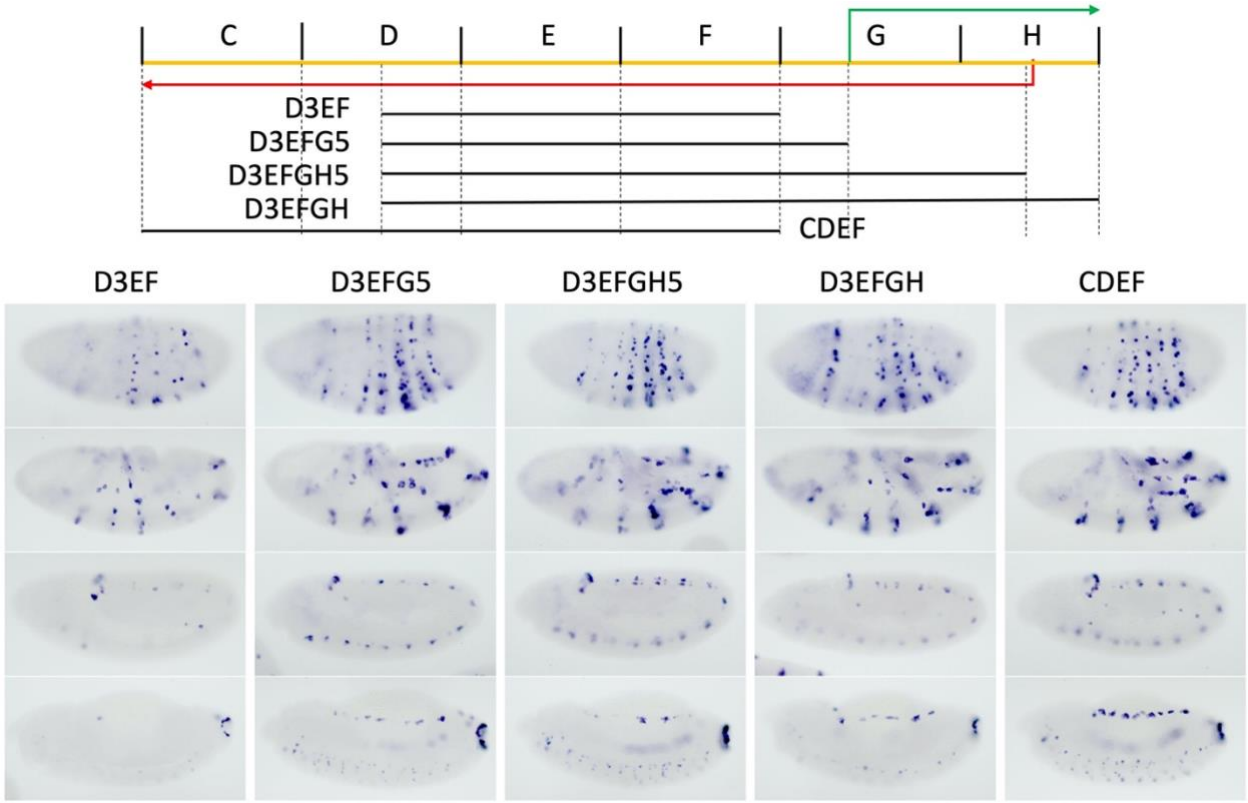
Supplemental Fig. 3



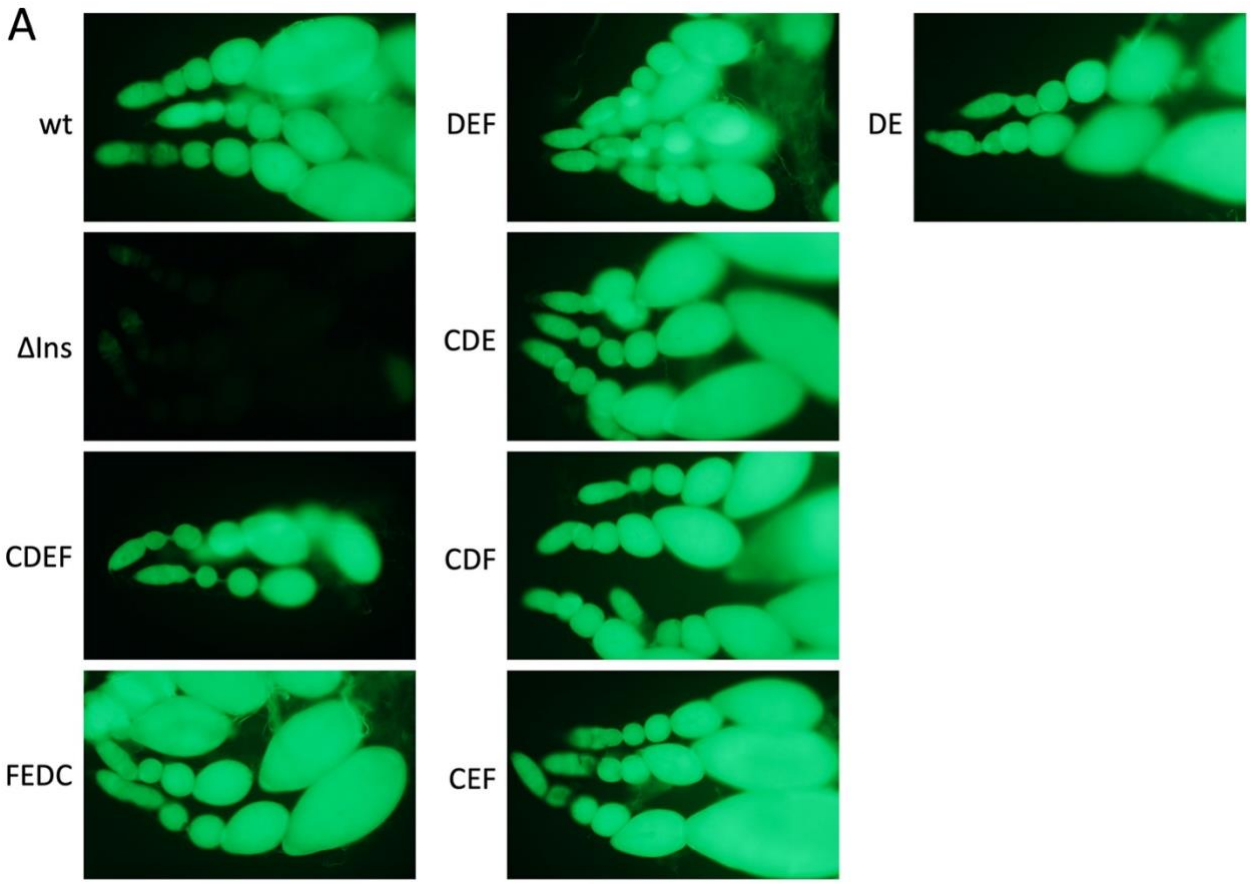
Supplemental Fig. 4



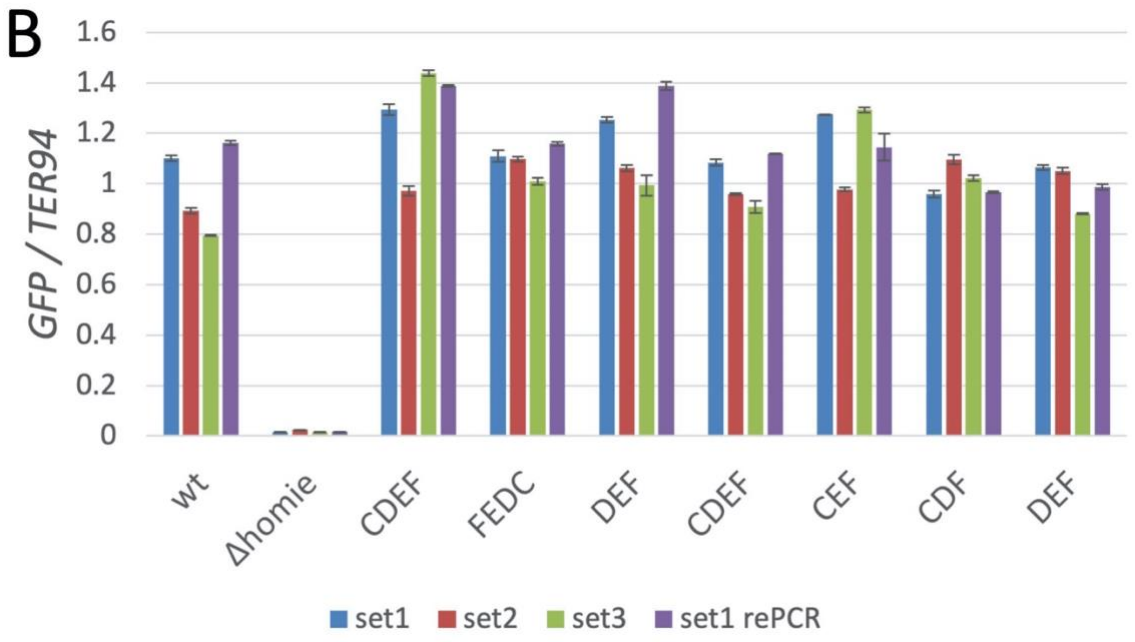
Supplemental Fig. 5

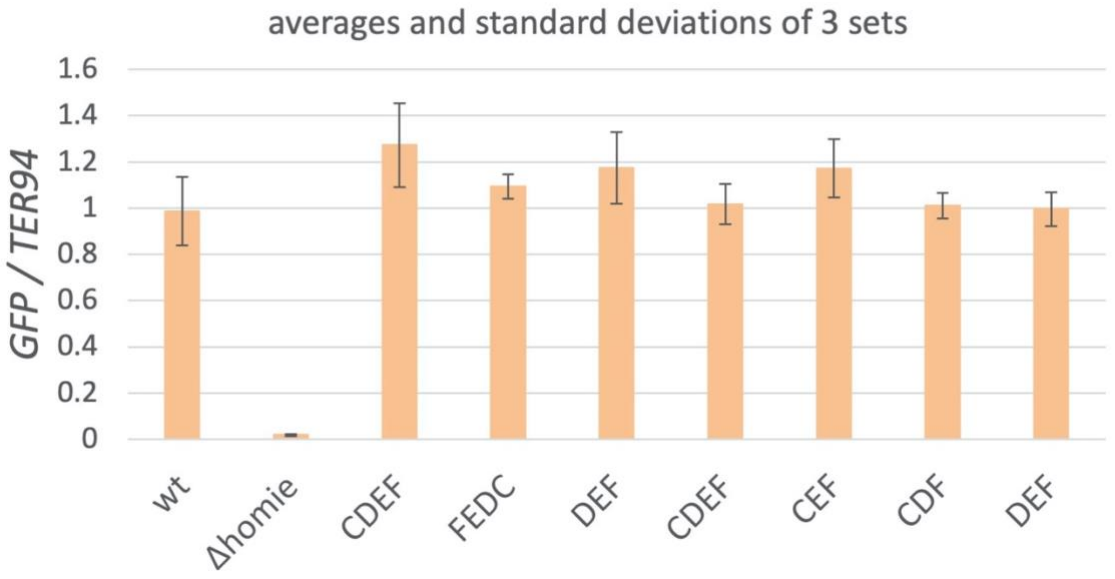


Supplemental Fig. 6



Supplemental Fig. 6 (cont.)





Supplemental Fig. 7

CDEF *homie* sequence

AATACTAAAAAGTTTTTACTAGCATAAGCTGCGATTGAGCAGTTATTGTAGGGACAGGTTGATGGCCGA
GTGTTCTAGGGAAATGAATGAATGAAG**ATTTTTTCTTACTACCAAAACTTCTACAACTAAGTGAAATAG**
ACGGATTGAGTTCTACATACTTCCTACATATATTTTAGATATAGACATAGCCGAAAAGTATGCTGCAC*T*
*TTTT*CAATAAAAACGTAAGCAGCTAAGCAGCGTAAGGCAGCTTCATAGCAGCGATCATCTGCCAGCGAG
CA*TAGCCAAGTTAG*GCACCTGCACACATTGCCACTCTTTCTTTTAGCGTTGCCACTTCAATTTCTTTTACAA
ACCATCGCAGCGTGTAATC

C: plain text
D: **bold**
E: *italics*
F: underlined
Red: Su(Hw) binding site
Green: Cp190-associated site

Su(Hw) consensus binding site (on complementary strand)

```
Searched:          YWGCATACTTTT
Found:             CAGCATACTTTT  (red = matches)

Searched:   NWWWWNYRTWGCATACTTTTKGSDB
Found:      AAAAAGTGCAGCATACTTTTCGGCT  (red = matches)
Mutated:    AAAAAGAACAGGATCCGAACATCCT
```

Cp190-associated site (complementary strand)

```
Searched:   GGTTBDWRWMYYNGCTD
Found:      GGTGCCTAACTTGGCTA   (green = matches)
Mutated:    GTTGCAAAGCTTGTCGA
```

D3EFGH sequence

ATATAGACATAGCCGAAAAGTATGCTGCACTTTTTCAATAAAAACGTAAGCAGCTAAGCAGCGTAAGGCA
GCTTCATAGCAGCGATCATCTGCCAGCGAGCATAGCCAAGTTAGGCACCTGCACACATTGCCACTCTTTC
TTTTAGCGTTGCCACTTCAATTTCTTTTACAAACCATCGCAGCGTGTAATCGATATTTTCTAAGCCC
AATCGATGAATGTCCGGTAACGCACATC**ACCAATCTGTTTCTGCTGCCTCATCTTTGAA**
TTTGACTAAAAGTGCCGATTTCCAGCTCAACGAATAGTCCAAGTCGCGCCGAATTGTAC
AATTAAGTTAAAAAGTTCAACGTCAACTAGTTTATCATGGCAGATTCCAAGGGGTAAGT
GAACTGGACGCCGAAATCTAGAGGCAGCAGGAGAAAT

D3EF: smaller, green font
D3EFG5: D3EF was extended by the underlined sequence

D3EFGH5: D3EFG5 was extended by the **bold** sequence
D3EFGH: D3EFGH5 was extended to the end of the sequence

Primers used for RT-PCR

GFP

GGGCACAAGCTGGAGTACAACTACAA
TGGCGGATCTTGAAGTTCACCTTG

TER94

TGAAGCCACCGCGTGGTATTCTTA
TTTGGACATGATCTCCGGTCCGTT