TADs pair homologous chromosomes to promote interchromosomal gene regulation

Kayla Viets[1], Michael Sauria[1], Chaim Chernoff[1], Caitlin Anderson[1], Sang Tran[1], Abigail Dove[1], Raghav Goyal[1], Lukas Voortman[1], James Taylor[1], and Robert J. Johnston Jr.[1]

[1]Department of Biology, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218

Abstract

Homologous chromosomes colocalize to regulate gene expression in processes including genomic imprinting and X-inactivation, but the mechanisms driving these interactions are poorly understood. In *Drosophila*, homologous chromosomes pair throughout development, promoting an interchromosomal gene regulatory mechanism called transvection. Despite over a century of study, the molecular features that drive chromosome-wide pairing and transvection are unknown. Here, we find that the ability to pair with a homologous sequence is not a general feature of all loci, but is specific to "button" loci interspersed across the genome. Buttons are characterized by topologically associated domains (TADs), which drive pairing with their endogenous loci from multiple genomic locations. Using a button spanning the *spineless* gene as a paradigm, we find that pairing is necessary but not sufficient for transvection. *spineless* pairing and transvection are cell-type-specific, suggesting that local buttoning and unbuttoning regulates transvection efficiency between cell types. Together, our data support a model in which button loci bring homologous chromosomes together to facilitate cell-type-specific interchromosomal gene regulation.

Introduction

Chromosomes are organized in a complex manner in the nucleus. In higher organisms, they localize to distinct territories (*1*). Regions of chromosomes interact to form compartments, which are segregated based on gene expression states (*2*). Chromosomes are further organized into TADs, regions of self-association that are hypothesized to isolate genes into regulatory domains and ensure their activation by the correct *cis*-regulatory elements (*2*). TADs vary in size from ~100 kb in *Drosophila melanogaster* to ~1 Mb in mammals *(2, 3)*. Disruptions of nuclear organization, such as alteration of TAD structure and localization of genes to incorrect nuclear compartments, have major effects on gene expression (*4-7*). However, it is unclear how elements within the genome interact between chromosomes to organize chromatin and regulate gene expression.

One key aspect of nuclear architecture involves the tight colocalization, or "pairing," of homologous chromosomes to facilitate regulatory interactions between different alleles of the same gene (*8*). In *Drosophila melanogaster*, homologous chromosomes are paired throughout interphase in somatic cells (*9*). This stable pairing provides an excellent paradigm to study the mechanisms driving interactions between chromosomes.

Despite over a century of study, it is poorly understood how homologous chromosomes come into close physical proximity. Two main models have been proposed. In the "zipper" model, all regions of the genome have an equal ability to pair based on sequence homology, and chromosome pairing begins at the centromere and proceeds distally to the telomeres (**Fig. 1A**) (*10*). The "button" model proposes that regions of high pairing affinity are interspersed along the chromosome arms and come together through a random walk to initiate pairing (**Fig. 1A**) (*11-13*). A handful of DNA transgenes are known to drive pairing with their endogenous loci at a relatively low frequency (*14-17*), but the general sequence and structural features that contribute to stable, chromosome-wide pairing are unknown.

Pairing of homologous chromosomes is required for a gene-regulatory mechanism known as transvection, in which two different mutant alleles interact between chromosomes to rescue gene expression (**Fig. 1B**) (*10*). Transvection has been described for a number of *Drosophila* genes (*18*). Generally, the efficiency of transvection decreases in the presence of chromosome rearrangements, which are assumed to disrupt chromosome pairing (*10, 18*). However, it is unclear if the same DNA elements are required for both homologous chromosome pairing and transvection and whether pairing and transvection are mechanistically separable.

Homologous chromosome pairing occurs more strongly in some cell types than in others (*11-13*). Similarly, transvection efficiency varies widely between cell types (*19-21*). However, a direct link between the level of pairing in a given cell type and the strength of transvection in that cell type has not been established.

Here, we develop a method to screen for DNA elements that pair and find that regions interspersed across the genome drive pairing, supporting the button hypothesis. A subset of TADs are associated with buttons and can drive pairing from different positions in the genome. By testing mutant alleles and transgenes of the *spineless* gene, we find that pairing and transvection are mechanistically separable and cell-type-specific. Together, our data suggest that buttons drive homologous chromosome pairing, promoting cell-type-specific interchromosomal gene regulation.

Results

"Button" loci are interspersed along chromosome arms

To distinguish between the zipper and button models of pairing, we tested whether transgenes inserted on different chromosomes were sufficient to drive pairing with their endogenous loci. We hypothesized that if chromosomes come together through a zipper mechanism, all transgenes would drive pairing, whereas if chromosomes are buttoned together, only a subset of transgenes would drive pairing.

We first screened a set of ~80-110 kb transgenes tiling a ~1 Mb region on chromosome 3R (**Fig. 1G**). We inserted individual transgenes into a site on chromosome 2L (site 1; **Fig. 1C**) and visualized their nuclear position using Oligopaints DNA FISH (*22*). As the endogenous and transgenic sequences were identical, we distinguished between them by labeling the sequence neighboring the endogenous locus with red probes and the sequence neighboring the transgene insertion site with green probes (**Fig. 1C**). We examined pairing in post-mitotic larval photoreceptors to avoid disruptions caused by cell division.

Only a subset (5/17) of transgenes drove pairing between chromosomes 2L and 3R, bringing the average distances between the red and green FISH signals significantly closer than in a control with no transgene (**Fig. 1C-E, G-H; Fig. S1A**). The red and green signals did not completely overlap, likely because they did not directly label the paired sites (**Fig. S2A-C**). For the remaining 12/17 transgenes, the distances between the red and green signals were not significantly different from the no transgene control, indicating that they did not drive pairing (**Fig. 1C-D, F-H; Fig. S1B**). Thus, remarkably, a subset of transgenes overcame endogenous nuclear architecture to drive pairing between non-homologous chromosomes, supporting the button model.

The pairing observed between transgenes on chromosome 2L and endogenous sequences on chromosome 3R could be affected by the transgene insertion site. To test the position independence of button pairing, we inserted *Transgene E* onto chromosome 3L (site 3; **Fig. S3A**) and found that it paired with its homologous endogenous locus on chromosome 3R (**Fig. S3B-D**), showing that buttons can drive pairing from different sites in the genome.

Our data suggest that pairing initiates through a button mechanism, in which specific loci interspersed along chromosomes drive homologous chromosomes together.

TADs are features of buttons

Only a subset of transgenes drove pairing, suggesting that unique chromatin structures within these transgenes might enable button function. We examined 14 publicly available Hi-C datasets to determine the relationship between buttons and topologically associated domains (TADs), genomic regions of self-association. We defined TADs using directionality indices, which measure the bias of a genomic region towards upstream or downstream interactions along the chromosome (*23*). TADs on a directionality index are read from the beginning of a positive peak, which indicates downstream interactions, to the end of a negative peak, which indicates upstream interactions (**Fig. 2A; Fig. S4A; Fig. S5A-E; Fig. S6A-E**).

We found that 60% of transgenes that drove pairing ("pairers") encompassed a complete TAD, including both TAD boundaries, compared to only 8% of transgenes that did not drive pairing ("non-pairers") (**Fig. 1H; Fig. S4A; Fig. S7A-B**). In a striking example, *Transgenes E* and *F* overlapped significantly, but only *Transgene E,* which contained a full TAD, drove pairing (**Fig. 2A; Fig. S1A-B**). Together, these data suggest that specific TADs contribute to button function.

To test the hypothesis that TADs are a feature of buttons, we examined additional transgenes spanning regions on chromosomes X, 2L, 2R, and 3R. Six of these transgenes encompassed entire TADs, while the remaining four did not span TADs (**Fig. 2F; Fig. S1E; Fig. S5A-E; Fig. S6A-E; Fig. S7A-B**). Based on the availability of Oligopaints probes, we used an alternative FISH strategy for a subset of these transgenes, in which the identical transgene and endogenous sequences were labeled with the same red fluorescent probes (**Fig. 2B**). Half of these transgenes drove pairing with their endogenous loci (**Fig. 2C-F; Fig. S1C-E**). All of the pairers spanned a TAD, whereas only one of five non-pairers spanned a TAD (**Fig. 2F; Fig. S1E; Fig. S5A-E; Fig. S6A-E; Fig. S7A-B**), further supporting the importance of TADs for button activity.

In total for all transgenes tested in **Fig. 1H**, **Fig. 2F**, and **Fig. S1E**, 80% of pairers spanned a TAD (8/10) while only 12% of non-pairers spanned a TAD (2/17) (**Fig 2G**), indicating that specific TADs contribute to button activity and drive pairing.

The ~80-110 kb size limitation of publicly available transgenes prevented testing larger TADs for pairing with our transgene assay. Transgenes that covered only parts of a large TAD on chromosome 3R did not drive pairing (**Fig. S8A**). To test this large TAD for pairing, we utilized a 460-kb duplication of chromosome 3R onto chromosome 2R (**Fig. S8B**), which encompassed the entire TAD (**Fig. S8A**). We found that the duplication drove pairing with its homologous endogenous site (**Fig. S8C-E**), further supporting a role for TADs in pairing.

Because TAD boundaries are often enriched for insulator protein binding sites (*3*), we hypothesized that pairers might contain a higher number of insulator protein binding sites than non-pairers. We examined modENCODE ChIP data and found that pairers were enriched for insulator binding sites (**Fig. 2H**), consistent with the conclusion that TADs contribute to button function to drive homologous chromosome pairing.

One prediction of the button model for chromosome pairing is that the content of a transgene (i.e. TADs), not the length of DNA homology, determines pairing. We found no relationship between transgene length and ability to drive pairing (**Fig. 2I**). Indeed, transgenes of near identical lengths had different pairing abilities (**Fig. 2I**), further indicating that buttons have distinct features beyond DNA sequence homology.

To identify additional genomic elements that contribute to pairing, we further examined modENCODE ChIP data and found that activating H3K4me3 marks positively correlated with pairing (**Fig. 2J**). As pairing was not associated with Polycomb Group (PcG) binding sites, repressing epigenetic marks, or non-coding RNAs (ncRNAs) (**Fig. S9A-F**), we hypothesized that active transcription plays a role in pairing. To test this possibility, we performed RNA-seq on larval eye discs, the same tissue we used in our pairing experiments. We found that pairing positively correlated with gene expression (**Fig. 2K**), suggesting that gene activity, in addition to TADs, is a feature of buttons that drives pairing.

Together, our data indicate that buttons, characterized by TADs and gene activity, drive pairing of homologous chromosomes.

Pairing and transvection occur despite chromosomal rearrangements

We next interrogated the relationship between button pairing and the gene regulatory process of transvection. Chromosomal rearrangements have been shown to disrupt pairing of genes located near rearrangement breakpoints (*10, 18*). However, we observed pairing of ~100 kb transgenes with their endogenous loci, suggesting that intact homologous chromosomes are not required for pairing and that pairing tolerates nearby breakpoints. We therefore reexamined how rearrangements affect pairing, focusing on a button defined by a TAD spanning the *spineless (ss)* locus ("*ss* button"; **Fig. 3A**).

To assess the effects of local rearrangements on *ss* button pairing, we examined a naturally occurring chromosomal inversion with a breakpoint in the gene immediately upstream of *ss* ($ss^{inversion}$) and a duplication with a breakpoint immediately downstream of *ss* (**Fig. 3E**). Both $ss^{inversion}$ and the duplication paired with endogenous *ss* (**Fig. S8C-E; Fig. S10A-B**), showing that *ss* button pairing occurs despite chromosomal rearrangements. Consistent with these findings, pairing also occurred at the *ss* locus in flies with balancer chromosomes containing numerous large inversions and rearrangements (**Fig. S10F-J**). Thus, in some cases, pairing occurs despite chromosomal rearrangements, consistent with the button model.

Pairing is required for the genetic phenomenon of transvection, in which DNA elements on a mutant allele of a gene act between chromosomes to rescue expression of a different mutant allele (**Fig. 1B**). In cases where chromosomal rearrangements perturb pairing, transvection is also disrupted (*10, 18*). Since chromosomal rearrangements did not ablate pairing at the *ss* button, we hypothesized that transvection would occur at the *ss* locus in these genetic conditions.

In the fly eye, Ss is normally expressed in ~70% of R7 photoreceptors to activate expression of Rhodopsin 4 (Rh4) and repress Rhodopsin 3 (Rh3; **Fig. 3B-D**). Ss is absent in the remaining 30% of R7s, allowing Rh3 expression (**Fig. 3B-D**) (*24*). Regulatory mutations in the *ss* gene cause decreases or increases in the ratio of Ss^{ON}: Ss^{OFF} cells. When two *ss* alleles with different ratios are heterozygous, transvection between chromosomes (also known as Interchromosomal Communication) determines the final ratio of Ss^{ON}: Ss^{OFF} R7s (*25*). Thus, the Ss^{ON}: Ss^{OFF} ratio is a phenotype that allows for quantitative assessment of transvection. Throughout our *ss* transvection experiments, we evaluated Rh3 and Rh4 expression, as they faithfully report Ss expression in R7s (i.e. Ss^{ON} = Rh4; Ss^{OFF} = Rh3). We previously observed transvection at the *ss* locus for the duplication and balancer chromosome alleles (*25*). We similarly observed transvection at the *ss* locus for the $ss^{inversion}$ allele (**Fig. S10C-E**). Together, these data suggested that buttons can drive pairing and transvection despite chromosomal rearrangements.

Pairing is necessary but not sufficient for transvection

As chromosomal rearrangements did not impair *ss* pairing or transvection, we further investigated the relationship between pairing and transvection using *ss* transgenes. Both *Transgene S* and *Transgene T* are expressed in 100% of R7s (**Fig. S11A-J**) because they lack a silencer DNA element, but do not produce functional Ss protein because they lack critical coding exons (**Fig. 3E**) (*25*). *Transgene T* differs from *Transgene S* in that it lacks 6 kb at its 5' end (**Fig. 3E**). We predicted that if *Transgenes S* and *T* performed transvection, they would upregulate expression of endogenous *ss*.

When inserted onto chromosomes 2L or 3L (sites 1 and 3), *Transgenes S* and *T* did not drive pairing with the endogenous *ss* locus on chromosome 3R (**Fig. 3F-G, I, K; Fig. S12A-C, E, Q**). At these sites, *Transgenes S* and *T* did not upregulate *ss* expression, indicating that they could not perform transvection when unpaired (**Fig. 3G-L; Fig. S12A-F**).

We next wondered whether *Transgenes S* and *T* could perform transvection if we mimicked pairing by forcing them into close physical proximity with endogenous *ss*. We performed a FISH screen to identify genomic sites that naturally loop to endogenous *ss* (**Fig. 3M**) and identified three such sites, located 4.8 Mb upstream of *ss*, 0.4 Mb upstream of *ss*, and 4.6 Mb downstream of *ss* (sites 2, 4, and 5; **Fig. 3N-O; Fig. S12G-H, M-N, Q**).

When we inserted *Transgene S* at these sites, it was forced into close proximity with endogenous *ss* (**Fig. 3Q; Fig. S12I, O, Q**) and upregulated Ss (Rh4) into nearly 100% of R7s (**Fig. 3P, R; Fig. S12J, P**) (*25*). Thus, natural chromosome looping can force loci into proximity and, like pairing, facilitate transvection. In contrast, when we forced *Transgene T* into close proximity with endogenous *ss*, it did not upregulate Ss (Rh4) expression, indicating that it could not perform transvection even when paired (**Fig. 3P, S-T; Fig. S12K-L, Q**). Thus, pairing is necessary but not sufficient for transvection.

We compared the DNA sequences of *Transgene T*, which does not perform transvection, to *Transgene S,* the duplication, and the $ss^{inversion}$, which perform transvection. An upstream region of ~1.6 kb is present in *Transgene S,* the duplication, and the $ss^{inversion}$, but missing from *Transgene T*, suggesting that this region contains a critical element for transvection (**Fig. 3E**). ModENCODE ChIP-seq data showed that this region was bound by the *Drosophila* insulator proteins CTCF, BEAF, Mod(Mdg4), and Cp190. Additionally, this DNA sequence performed P-element homing (*25*), an indicator of insulator activity. Together, these data suggested that the DNA element required for transvection is an insulator.

To further test whether this insulator was required for transvection, we examined *Transgene E*, which drove pairing and contained the complete *ss* locus, except for the insulator element (**Fig. 1H; Fig. 3E; Fig. S1A; Fig. S3B-D**). We utilized genetic backgrounds in which *Transgene E* was the only source of Ss protein, so that any changes in Ss (Rh4) expression would indicate transvection effects on *Transgene E*. As a control, we examined *Transgene E* expression when the endogenous *ss* locus was hemizygous for a protein null allele ($ss^{protein\ null}$) that did not perform transvection (**Fig. S13A-B**). In this background, *ss* on *Transgene E* was expressed in 52% of R7s (**Fig. S13A-B**). We next tested *Transgene E* for transvection with a high-frequency protein null allele ($ss^{high\ freq\ null}$), which can perform transvection to increase *ss* expression (**Fig. S10D**) (*25*). When the endogenous *ss* locus was hemizygous for the $ss^{high\ freq\ null}$, we observed no increase in *Transgene E* expression, indicating that it did not perform transvection (51% Ss (Rh4); **Fig. S13A, C**). Moreover, *Transgene E* did not perform transvection in other genetic conditions (**Fig. S13D-E**). Thus, *Transgene E* paired with the endogenous *ss* locus but failed to perform transvection. These data show that an insulator is required

for transvection but not for pairing, indicating that transvection and pairing are mechanistically separable.

ss pairing and transvection are cell-type-specific

It is poorly understood how pairing impacts transvection in a cell-type-specific manner. We propose two models: constitutive and cell-type-specific buttoning. In the constitutive model, all buttons drive pairing in all cell types, and differences in transvection would occur due to variation in transcription factor binding or chromatin state between cell types (**Fig. 4A**). In the cell-type-specific model, different buttons drive pairing in each cell type, bringing different regions into physical proximity to control transvection efficiency (**Fig. 4A**).

We tested these models by investigating pairing and transvection of *ss* in two different tissues. In addition to its role in R7 photoreceptors, *ss* is required for the development of the arista, a structure on the antenna (**Fig. 4C-D**) (*24, 26*). *Transgene E,* which contains the *ss* button, drove pairing in the eye (**Fig. 1H; Fig. 4B; Fig. S1A; Fig. S3B-D**) but not the antenna from two different insertion sites (sites 1 and 3; **Fig. 4B; Fig. S14A-H**), suggesting that button pairing is cell-type-specific.

As pairing is required for transvection and the *ss* button pairs in a cell-type-specific manner, we hypothesized that transvection at the *ss* locus is cell-type-specific. To test this hypothesis, we examined an allele of *ss* that specifically affects arista development ($ss^{arista\ 1}$) (**Fig. 4G-J; Fig. S15A-F**). In flies transheterozygous for $ss^{arista\ 1}$ and a *ss* deficiency (ss^{def}), aristae were transformed into legs (i.e. aristapedia) (**Fig. 4G-H; Fig. S15A, C**). Aristapedia was also observed for $ss^{protein\ null}$ / ss^{def} flies (**Fig. 4E-F**). In the eye, $ss^{protein\ null}$ performed transvection to rescue *ss* expression (**Fig. S16A-D**). However, the aristapedia mutant phenotype persisted in $ss^{arista\ 1}$ / $ss^{protein\ null}$ flies (**Fig. 4I-J; Fig. S15D, F**), suggesting that, unlike in the eye, transvection does not rescue *ss* expression in the arista. Cell-type-specific transvection of the *ss* gene in the eye but not the arista was also observed in other genetic conditions (**Fig. S15G-L; S17A-L**).

As *ss* button pairing and transvection are cell-type-specific and pairing is required for transvection, our data support the cell-type-specific model, in which local buttoning and unbuttoning occur in a cell-type-specific manner to determine transvection efficiency (**Fig. 4A**).

Discussion

Despite the discovery of homologous chromosome pairing in flies over 100 years ago (*9*), the mechanisms that facilitate pairing have remained unclear. We find that the ability to pair with a homologous sequence is not a general feature of all loci, but is specific to a subset of loci (buttons) interspersed across the genome. Specific TADs drive button activity and can pair from multiple locations in the genome. Individual TADs may take on unique chromatin conformations or bind unique combinations of proteins to create nuclear microcompartments that enable homologous TAD association and pairing (**Fig. 2L**). As gene activity is also a feature of buttons, the mechanisms that promote specific enhancer-promoter interactions on the same chromosome may also act between chromosomes to pair active regions together (**Fig. 2L**). Additional small DNA elements may also facilitate pairing (*14-17*). Complementary work from Erceg, AlHaj Abed, & Goloborodko, *et al.* (*27*) and AlHaj Abed, Erceg, & Goloborodko, *et al.* (*28*) using Hi-C also reveals variable levels of pairing across the genome, with implications for genome function.

Our data indicate that pairing and transvection are mechanistically separable: TADs and gene activity facilitate pairing, while an insulator element facilitates transvection to the endogenous *spineless* locus. Consistent with our findings using endogenous alleles, an insulator is required for

transvection but not pairing between transgenes containing the *snail* enhancer and the *eve* promoter (*29*).

We find that the *ss* locus drives pairing and performs transvection in the eye but not in the antenna. Our results support a model in which different buttons drive pairing in different cell types. In this model, local buttoning or unbuttoning at a specific gene determines its transvection efficiency in a given cell type. Variation in levels of pairing or transvection across cell types has been observed for a number of loci (*12, 13, 21*), suggesting that differences in pairing between cell types may be a general mechanism regulating gene expression.

The mechanisms driving chromosome pairing and transvection have remained a mystery of fly genetics since their initial discoveries by Nettie Stevens and Ed Lewis (*9, 10*). Our results provide strong support for the button model of pairing initiation and offer the first evidence of a general feature, specialized TADs, that drives homologous chromosomes together. Furthermore, we find that pairing is necessary but not sufficient for transvection and that distinct elements are required for these processes. Both pairing and transvection are cell-type-specific, suggesting that tighter pairing in a given cell type enables more efficient transvection in that cell type. Our findings suggest a general mechanism that drives homologous chromosome pairing and interchromosomal gene regulation across organisms to facilitate processes including X-inactivation and imprinting.

Materials and methods can be found in the supplementary materials.

Figure Legends

Figure 1: "Button" loci are interspersed along chromosomes.
A. Zipper versus button models of homologous chromosome pairing. Yellow squares: button loci (high pairing affinity).
B. Two-step process of pairing and transvection.
C. Two-color DNA FISH strategy: If a transgene drove pairing, red and green FISH punctae would be close together in the nucleus. If a transgene did not drive pairing, red and green FISH punctae would be far apart in the nucleus, similar to a control.
D-F. Control, pairer, and non-pairer examples. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.
G. ~1 Mb region of chromosome 3R used for pairing screen. Orange boxes indicate locations of the major developmental genes *spineless* (*ss*), *ultrabithorax* (*ubx*), *abdominal-a* (*abd-a*), and *abdominal-b* (*abd-b*).
H. Quantifications for all transgenes from the initial screen. Black: control, blue: pairers, gray: non-pairers. **T:** contains a TAD. Control data are the same as in **Fig. 2F** (2L-3R control), **S10J**, and **S12Q** (site 1 control). ****=p<0.0001, ***=p<0.001, **=p<0.005, *=p<0.05, ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test.

Figure 2: Specialized TADs contribute to button activity and drive pairing.
A. Representative Hi-C heat map and directionality index (NCBI GSE38468) showing TADs in the region covered by *Transgenes E* and *F*. Dotted lines: TAD boundaries. Black bars: TADs. See **Fig. S4A** for TAD assessment.

B. One-color DNA FISH strategy: If a transgene drove pairing, the two red FISH punctae would be close together in the nucleus and indistinguishable as separate dots. If a transgene did not drive pairing, the two red FISH punctae would be far apart in the nucleus, similar to a control.
C-E. Control, pairer, and non-pairer examples. *Transgenes Y* and *Z* were taken from Chromosome 3R and inserted at site 1. Control assessed the distance between sites on Chromosomes 2L and 3R with no transgenes present. Scale bars=1 μm. White: Lamin B, red: probes against endogenous sequence and transgene.
F. Quantifications for additional transgenes. **T:** contains a TAD. Black: controls, blue: pairers, gray: non-pairers. ***=p<0.001, **=p<0.005, *=p<0.05, ns=p>0.05, one-way ANOVA on ranks with Dunn's multiple comparisons test (for *Transgenes U-W*, *Y-Z*, *AA-BB*) or Wilcoxon rank-sum test (for *Transgene X*). 3L-X control data are the same as in **Fig. S1E**. 2L-3R control data are the same as in **Fig. 1H, S10J,** and **S12Q** (site 1 control).
G. Percentage of pairers versus non-pairers spanning TADs. Blue: pairers, gray: non-pairers.
H-K. Quantifications for pairers and non-pairers tested in **Fig. 1, 2,** and **S1**. Blue: pairers, gray: non-pairers. *=p<0.05, ns=p>0.05, Wilcoxon rank-sum test (**H-I, K**) or unpaired t-test with Welch's correction (**J**).
H. Insulator ChIP peaks.
I. Transgene length
J. H3K4me3 ChIP peaks.
K. Percentage of genes active.
L. Model for button-driven pairing.

Figure 3: Pairing is necessary but not sufficient for transvection.
A. Representative directionality index (NCBI GSE38468) showing the TAD that defines the *ss* button. Black bar: TAD. See **Fig. S4A** for TAD assessment.
B. Spineless (Ss) activates Rh4 and represses Rh3.
C. Ss is expressed in ~70% of R7s. Green: Ss, red: Prospero (R7 marker), white circles: Ss-expressing R7s.
D. Rh3 (blue) and Rh4 (red) expression in wild type R7s.
E. *ss* alleles and transgenes. ins: insulator, sil 1: silencer 1, enh: enhancer, sil 2: silencer 2. Smaller black arrows: transcription start sites. Gray rectangles: exons. Dotted gray lines: region required for transvection.
F. Strategy used to assess pairing and transvection from site 1 in **Fig. 3G-L**. Gray arrow with "?" indicates that *Transgenes S* and *T* were tested for transvection.
G, I, K, O, Q, S. Scale bars= 1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.
H, J, L, P, R, T. Red: Rh4, blue: Rh3.
G, I, K. Pairing assay images of *site 1 control*, *Transgene S site 1,* and *Transgene T site 1*.
H, J, L. Rh3 and Rh4 expression in *wild type control* (Ss(Rh4)=70%), *Transgene S site 1* (Ss(Rh4)=57%), and *Transgene T site 1* (Ss(Rh4)=55%). The slight decrease in Rh4 frequency for *Transgene S site 1* and *Transgene T site 1* is likely due to background genetic effects.
M. Natural chromosome looping forces transgenes into close proximity with endogenous *ss*, mimicking pairing and facilitating transvection. Gray arrow with "?" indicates that *Transgenes S* and *T* were tested for transvection.
N. Strategy used to assess pairing and transvection from site 2 in **Fig. 3O-T**.
O, Q, S. Pairing assay images of *site 2 control*, *Transgene S site 2*, and *Transgene T site 2*.

P, R, T. Rh3 and Rh4 expression in *wild type control* (Ss(Rh4)=70%), *Transgene S site 2* (Ss(Rh4)=98%) and *Transgene T site 2* (Ss(Rh4)=78%).

Figure 4: *ss* pairing and transvection are cell-type-specific.
A. Constitutive vs. cell-type-specific pairing models.
B. Third instar larval eye-antennal disc. The *ss* button drove pairing in the larval eye but not the larval antenna. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.
C, E, G, I. Genotypes tested for transvection. Gray rectangles: exons. Smaller black arrows: transcription start sites. Red X indicates an uncharacterized mutation in the $ss^{arista\ 1}$ sequence. Red X over gray arrow indicates an absence of transvection between alleles in the arista.
D, F, H, J. Arista phenotype. Scale bars=50 μm. White arrows indicate arista.

References

1. T. Cremer, M. Cremer, Chromosome territories. *Cold Spring Harb Perspect Biol* **2**, a003889 (2010).
2. K. P. Eagen, Principles of Chromosome Architecture Revealed by Hi-C. *Trends Biochem Sci*, (2018).
3. T. Sexton *et al.*, Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458-472 (2012).
4. Y. Guo *et al.*, CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900-910 (2015).
5. D. G. Lupianez *et al.*, Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).
6. E. J. Clowney *et al.*, Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell* **151**, 724-737 (2012).
7. K. L. Reddy, J. M. Zullo, E. Bertolino, H. Singh, Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**, 243-247 (2008).
8. E. F. Joyce, J. Erceg, C. T. Wu, Pairing and anti-pairing: a balancing act in the diploid genome. *Curr Opin Genet Dev* **37**, 119-128 (2016).
9. N. M. Stevens, A Study of the Germ Cells of Certain Diptera, With Reference to the Heterochromosomes and the Phenomena of Synapsis. *J Exper Zool* **5**, 359-374 (1906).
10. E. B. Lewis, The Theory and Application of a New Method of Detecting Chromosomal Rearrangements in *Drosophila melanogaster*. *Am Nat* **88**, 225-239 (1954).
11. Y. Hiraoka *et al.*, The onset of homologous chromosome pairing during Drosophila melanogaster embryogenesis. *J Cell Biol* **120**, 591-600 (1993).
12. J. C. Fung, W. F. Marshall, A. Dernburg, D. A. Agard, J. W. Sedat, Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations. *J Cell Biol* **141**, 5-20 (1998).
13. M. J. Gemkow, P. J. Verveer, D. J. Arndt-Jovin, Homologous association of the Bithorax-Complex during embryogenesis: consequences for transvection in Drosophila melanogaster. *Development (Cambridge, England)* **125**, 4541-4552 (1998).
14. M. Fujioka, H. Mistry, P. Schedl, J. B. Jaynes, Determinants of Chromosome Architecture: Insulator Pairing in cis and in trans. *PLoS Genet* **12**, e1005889 (2016).
15. F. Bantignies, C. Grimaud, S. Lavrov, M. Gabut, G. Cavalli, Inheritance of Polycomb-dependent chromosomal interactions in Drosophila. *Genes Dev* **17**, 2406-2420 (2003).
16. J. Vazquez, M. Muller, V. Pirrotta, J. W. Sedat, The Mcp element mediates stable long-range chromosome-chromosome interactions in Drosophila. *Mol Biol Cell* **17**, 2158-2165 (2006).
17. M. Ronshaugen, M. Levine, Visualization of trans-homolog enhancer-promoter interactions at the Abd-B Hox locus in the Drosophila embryo. *Dev Cell* **7**, 925-932 (2004).
18. I. W. Duncan, Transvection effects in Drosophila. *Annu Rev Genet* **36**, 521-556 (2002).
19. D. J. Mellert, J. W. Truman, Transvection is common throughout the Drosophila genome. *Genetics* **191**, 1129-1141 (2012).
20. J. R. Bateman, J. E. Johnson, M. N. Locke, Comparing enhancer action in cis and in trans. *Genetics* **191**, 1143-1155 (2012).
21. A. J. Blick *et al.*, The Capacity to Act in Trans Varies Among Drosophila Enhancers. *Genetics* **203**, 203-218 (2016).
22. B. J. Beliveau *et al.*, Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc Natl Acad Sci U S A* **109**, 21301-21306 (2012).

23. J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).
24. M. F. Wernet *et al.*, Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440**, 174-180 (2006).
25. R. J. Johnston, Jr., C. Desplan, Interchromosomal communication coordinates intrinsically stochastic expression between alleles. *Science* **343**, 661-665 (2014).
26. G. Morata, P. A. Lawrence, Development of the eye-antenna imaginal disc of Drosophila. *Dev Biol* **70**, 355-371 (1979).
27. J. Erceg, J. AlHaj Abed, A. Goloborodko, B.R. Lajoie, G. Fudenberg, N. Abdennur, M. Imakaev, R.B. McCole, S.C. Nguyen, W. Saylor, E.F. Joyce, T.N. Senaratne, M.A. Hannan, G. Nir, J. Dekker, L.A. Mirny, C.-ting Wu, Structured homolog pairing during zygotic genome activation. *In prep*.
28. J. AlHaj Abed, J. Erceg, A. Goloborodko, S. C. Nguyen, R. B. McCole, W. Saylor, G. Fudenberg, B. R. Lajoie, J. Dekker, L. A. Mirny, C.-ting Wu, Global *trans*-homolog pairing underlies a genome-wide functional level of organization in the *Drosophila* genome. *In prep*.
29. B. Lim, T. Heist, M. Levine, T. Fukaya, Visualization of Transvection in Living Drosophila Embryos. *Mol Cell* **70**, 287-296 e286 (2018).

Materials and Methods

Drosophila lines

Flies were raised on standard cornmeal-molasses-agar medium and grown at 25° C.

Fly line	Full genotype	Source	Figures
wild type control	*yw; +; +* or *yw; pm181>Gal4, UAS>mcd8GFP/CyO; +* or *yw; sp/CyO; +*	*(1)*	1D, H; 2C, F; 3C-D, G-H, O-P; 4C-D; S1C-E; S3B, D; S8C, E; S10B, F-G, J; S12B, H, N, Q; S14B, D, F, H
Transgene B site 1	*yw; pBac{CH321-38G20}VK00037; +*	*(2)**	1E, G-H; 2G-K; S4A; S7A; S9A-F
Transgene D site 1	*yw; pBac{CH321-25M02}VK00037; +*	*(2)**	1F-H; 2G-K; S4A; S7B; S9A-F
Transgene A site 1	*yw; pBac{CH321-94A21}VK00037; +*	*(2)**	1G-H; 2G-K; S1A; S4A; S7A; S9A-F
Transgene C site 1	*yw; pBac{CH321-86F17}VK00037; +*	*(2)**	1G-H; 2G-K; S1A; S4A; S7A; S9A-F
Transgene E site 1	*yw; pBac{CH321-28L15}VK00037; +*	*(2, 3)*	1G-H; 2A, G-K; 3E; S1A; S4A; S7A; S9A-F; S14A, C-D
Transgene F site 1	*yw; pBac{CH321-23C04}VK00037; +*	*(2)**	1G-H; 2A, G-K; S1B; S4A; S7B; S9A-F
Transgene G site 1	*yw; pBAC{CH321-02A24}VK00037; +*	*(2)**	1G-H; 2G-K; S1B; S4A; S7B; S9A-F
Transgene H site 1	*yw; pBAC{CH321-92J22}VK00037; +*	*(2)**	1G-H; 2G-K; S1B; S4A; S7B; S8A; S9A-F
Transgene I site 1	*yw; pBAC{CH321-95F12}VK00037; +*	*(2)**	1G-H; 2G-K; S1B; S4A; S7B; S8A; S9A-F
Transgene J site 1	*yw; pBAC{CH321-71G17}VK00037; +*	*(2)**	1G-H; 2G-K; S1B; S4A; S7B; S8A; S9A-F
Transgene K site 1	*yw; pBAC{CH321-50E16}VK00037; +*	*(2)**	1G-H; 2G-K; S1B; S4A; S7B; S8A; S9A-F
Transgene L site 1	*yw; pBAC{CH321-60D22}VK00037; +*	*(2)**	1G-H; 2G-K; S1B; S4A; S7B; S9A-F
Transgene M site 1	*yw; pBAC{CH321-58G7}VK00037; +*	*(2)**	1G-H; 2G-K; S1B; S4A; S7B; S9A-F
Transgene N site 1	*yw; pBAC{CH321-96A10}VK00037/+; +*	*(2)**	1G-H; 2G-K; S1B; S4A; S7B; S9A-F
Transgene O site 1	*yw; pBAC{CH321-45F07}VK00037; +*	*(2)**	1G-H; 2G-K; S1A; S4A; S7A; S9A-F
Transgene P site 1	*yw; pBAC{CH321-58J11}VK00037; +*	*(2)**	1G-H; 2G-K; S1B; S4A; S7B; S9A-F
Transgene Q site 1	*yw; pBAC{CH321-52D18}VK00037/+; +*	*(2)**	1G-H; 2G-K; S1B; S4A; S7B; S9A-F
Transgene Y site 1	*yw; pBAC{CH321-47D18}VK00037; +*	*(2)**	2D, F-K; S5E; S7A; S9A-F
Transgene Z site 1	*w^{118}; PBac{y[+mDint2]w[+mc]=pros-GFP.FPTB}VK00037; +*	Bloomington*, (2)*	2E-K; S2B; S6A; S7B; S9A-F
Transgene U site 3	*w^{118}; +; Dp(1;3)DC212, PBac{y[+mDint2]w[+mC]=DC212} VK00033*	Bloomington*, (4)*	2F-K; S1C; S5A; S7A; S9A-F
Transgene V site 3	*w^{118}; +; Dp(1;3)DC550, PBac{y[+mDint2]w[+mC]=DC550} VK00033*	Bloomington*, (4)*	2F-K; S1C; S5B; S7A; S9A-F

Transgene W site 3	*w^{118}; +; Dp(1;3)DC305, PBac{y[+mDint2]w[+mC]=DC305} VK00033*	Bloomington, *(4)*	2F-K; S1C; S5C; S7A; S9A-F
Transgene X site 3	*yw; +; pBAC{CH321-43H12}VK00033*	*(2)**	2F-K; S1C; S5D; S7A; S9A-F
Transgene AA site 3	*yw; +; pBAC{CH321-16A21}VK00033*	*(2)**	2F-K; S1C; S6B; S7B; S9A-F
Transgene BB site 3	*yw; +; pBAC{CH321-68L02}VK00033*	*(2)**	2F-K; S1C; S6C; S7B; S9A-F
Transgene CC site 3	*w^{118}; +; Dp(1;3)DC129, PBac{y[+mDint2]w[+mC]=DC129} VK00033*	Bloomington, *(4)*	2G-K; S1D-E; S6D; S7B; S9A-F
Transgene DD site 3	*w^{118}; +; Dp(1;3)DC372, PBac{y[+mDint2]w[+mC]=DC372} VK00033*	Bloomington, *(4)*	2G-K; S1D-E; S6E; S7B; S9A-F
Transgene S site 1	*yw; pBac{pBJ250}VK00037; +*	*(3)*$^{+}$	3E, I-J; S11B; S12Q
Transgene T site 1	*yw; pBAC{pBJ205}VK00037; +*	*(3)*$^{+}$	3E, K-L; S11G; S12Q
duplication	*yw; Dp(3;2)P10/CyO; +*	Bloomington, *(5)*	3E; S4A; S8A-B, D-E
$ss^{inversion}$/ +	*yw; +/CyO; In(3R)P/+*	Bloomington, *(6)*	3E; S10A-C
$ss^{high\ freq\ null}$/ $ss^{inversion}$	*yw; +; ss^{52}/In(3R)P*	Bloomington, *(3, 6)*	3E; S10D
Transgene S-ss^{def}/ $ss^{inversion}$	*yw; +; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330/In(3R)P*	Bloomington, *(3, 6, 7)*	3E; S10E
Transgene S site 2	*yw; +; pBAC{pBJ250}VK00027*	*(3)*$^{+}$	3E, Q-R; S11C; S12Q
Transgene T site 2 for third instar larvae	*yw; pm181>Gal4, UAS>mcd8GFP/CyO; pBAC{pBJ205}VK00027/TM6B*	*(1, 3)*$^{+}$	3E, S; S12Q
Transgene T site 2 for pupae and adults	*yw; +; pBAC{pBJ205}VK00027*	*(3)*$^{+}$	3E, T; S11H
Transgene E site 3	*yw; +; pBac{CH321-28L15}VK00033*	*(2)**	3E; S3A, C-D; S14E, G-H
Transgene S site 3	*yw; +; pBac{pBJ250}VK00033*	*(3)*$^{+}$	3E; S11D; S12C-D, Q
Transgene S site 4 for pupae and adults	*yw; +; pBAC{pBJ250}ZH-86Fb*	*(3)*$^{+}$	3E; S11E; S12J
Transgene S site 5	*yw; +; pBAC{pBJ250}VK00028*	*(3)*$^{+}$	3E; S11F; S12O-Q
Transgene T site 3	*yw; +; pBAC{pBJ205}VK00033*	*(3)*$^{+}$	3E; S11I; S12E-F, Q
Transgene T site 4	*yw; +; pBAC{pBJ205}ZH-86Fb*	*(3)*	3E; S11J; S12K-L, Q
Transgene S site 4 for third instar larvae	*yw; pm181>Gal4, UAS>mcd8GFP/CyO; pBAC{pBJ250}ZH-86Fb/TM6B*	*(1, 3)*	3E; S12I, Q
Transgene E+ $ss^{protein\ null}$	*yw; pBAC{CH321-28L15}VK0037/+; $ss^{d115.7}$/Df(3R)Exel7330*	Bloomington, *(2, 7-9)*	3E; S13A-B
Transgene E+ $ss^{high\ freq\ null}$	*yw; pBAC{CH321-28L15}VK0037/+; ss^{52}/Df(3R)Exel7330*	Bloomington, *(2, 3, 7, 8)*	3E; S13A, C
Transgene E+ Transgene S-ss^{def}/ $ss^{protein\ null}$	*yw; pBAC{CH321-28L15}VK00037/CyO; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330/ $ss^{d115.7}$*	Bloomington, *(2, 3, 7, 8)*	3E; S13D-E
Transgene S-ss^{def}/$ss^{upstream\ del}$	*yw; +; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330/$ss^{upstream\ deletion}$*	Bloomington, *(3, 7)*	3E; S17D-F
Transgene S-ss^{def}/$ss^{arista\ 1}$	*yw; +; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330/ss^{a}*	Bloomington, *(3, 7, 10, 11)*	3E; S17G-I
Transgene S-ss^{def}/$ss^{arista\ 2}$	*yw; +; pBAC{pBJ250}ZH-86Fb, Df(3R)Exel7330/ss^{a40a}*	Bloomington, *(3, 7, 11)*	3E; S17J-L
$ss^{protein\ null}$/ss^{def}	*yw; +; $ss^{d115.7}$/Df(3R)Exel6269*	Bloomington, *(7, 8)*	4E-F
$ss^{arista\ 1}$/ss^{def}	*yw; +; ss^{a}/Df(3R)Exel6269*	Bloomington, *(7, 10, 11)*	4G-H; S15A-C
$ss^{arista\ 1}$/$ss^{protein\ null}$	*yw; +; ss^{a}/$ss^{d115.7}$*	Bloomington, *(8, 10, 11)*	4I-J; S15D-F
rearrangements	*yw; +; TM2/TM6B*	N/A	S10H-J

$ss^{arista\ 2}/ss^{def}$	*yw; +; ss^{a40a}/Df(3R)Exel6269*	Bloomington, *(7, 11)*	S15G-I
$ss^{arista\ 2}/ss^{protein\ null}$	*yw; +; $ss^{a40a}/ss^{d115.7}$*	Bloomington, *(8, 11)*	S15J-L
$ss^{enh\ del}/ss^{def}$	*yw; +; $ss^{enhancer\ deletion}$/Df(3R)Exel6269*	Bloomington, *(7)*	S16A-B
$ss^{enh\ del}/ss^{protein\ null}$	*yw; +; $ss^{enhancer\ deletion}/ss^{d115.7}$*	*(8)*	S16C-D
$ss^{upstream\ del}/ss^{def}$	*yw; +; $ss^{upstream\ deletion}$/ Df(3R)Exel6269*	Bloomington, *(7)*	S17A-C

*Constructs were purchased from the CHORI *Drosophila melanogaster* BAC library collection (*2*) and sent to BestGene Inc. (Chino Hills, CA) or Rainbow Transgenic Flies, Inc. (Camarillo, CA) for injection.
+Constructs were generated in (*3*) and sent to BestGene Inc. (Chino Hills, CA) or Rainbow Transgenic Flies, Inc. (Camarillo, CA) for injection.

Constructs were inserted via PhiC31 integration at the following landing sites:

Landing site	Cytological coordinates	Genome coordinates
site 1 (VK00037)	22A3	2L: 1,582,820
site 2 (VK00027)	89E11	3R: 17,052,863
site 3 (VK00033)	65B2	3L: 6,442,676
site 4 (ZH-86Fb)	86F8	3R: 11,808,607
site 5 (VK00028)	92F1	3R: 20,549,650

Oligopaints probe libraries

Probe set	Oligopaints library name	Genome coordinates targeted	Conjugated fluorophore	Figures
site 1 neighboring sequence	right of 2L>22A3 transgene insertion site	2L: 1,582,821-1,642,821	Cy5	1D-F; 2C; 3G, I, K; 4B; S1A-B; S2B; S14B
ss	old *ss* 90K library	3R: 16,374,660-16,430,430	Cy3	1D; 2C; 3G; S10A, G, I; S14B-C
Transgene B neighboring endogenous sequence	downstream of 38G20	3R: 16,263,284-16,313,284	Cy3	1E
Transgene D neighboring endogenous sequence	downstream of 25M02	3R: 16,381,436-16,431,436	Cy3	1F
Transgene Y	*bicoid* 25-kb left extension+*bicoid* DNA+*bicoid* 25-kb right extension	3R: 6,729,194-6,787,593	Cy3	2D
Transgene Z	*prospero* DNA	3R: 11,246,862-11,407,918	Cy3	2E; S2B
upstream of *Transgenes S* and *T*	upstream of pBJ250 and pBJ205	3R: 16,340,760-16,390,760	Cy3	3I-K, O-Q
site 2 neighboring sequence	pBJ250>3R(89E11) insertion site	3R: 16,992,863-17,052,863	Cy5	3O, Q, S
upstream of *ss*	*spineless* 50-kb extension (left)	3R: 16,320,533-16,370,533	Cy3	3S; S8C-D; S12H-I, K, N-O
downstream of *ss*	*spineless* 50-kb extension (right)	3R: 16,435,681-16,485,681	Cy3	4B; S1A; S3B-C; S12B-C, E; S14F-G

Transgene A neighboring endogenous sequence	downstream of 94A21	3R: 16,240,324-16,290,324	Cy3	S1A
Transgene C neighboring endogenous sequence	downstream of 86F17	3R: 16,324,960-16,374,960	Cy3	S1A
Transgene O neighboring endogenous sequence	downstream of 45F07	3R: 17,026,709-17,076,709	Cy3	S1A
Transgene F neighboring endogenous sequence	downstream of 23C04	3R: 16,455,152-16,505,152	Cy3	S1B
Transgene G neighboring endogenous sequence	upstream of 02A24	3R: 16,350,218-16,400,218	Cy3	S1B
Transgene H neighboring endogenous sequence	upstream of 92J22	3R: 16,390,309-16,440,309	Cy3	S1B
Transgene I neighboring endogenous sequence	upstream of 95F12	3R: 16,459,720-16,509,720	Cy3	S1B
Transgene J neighboring endogenous sequence	upstream of 71G17	3R: 16,511,320-16,561,320	Cy3	S1B
Transgene K neighboring endogenous sequence	downstream of 50E16	3R:16,691,689-16,741,689	Cy3	S1B
Transgene L neighboring endogenous sequence	downstream of 60D22	3R: 16,844,756-16,894,756	Cy3	S1B
Transgene M neighboring endogenous sequence	upstream of 58G07	3R: 16,739,235-16,789,235	Cy3	S1B
Transgene N neighboring endogenous sequence	upstream of 96A10	3R: 16,844,621-16,894,621	Cy3	S1B
Transgene P neighboring endogenous sequence	upstream of 58J11	3R: 16,967,427-17,017,427	Cy3	S1B
Transgene Q neighboring endogenous sequence	upstream of 52D18	3R: 17,043,366-17,093,366	Cy3	S1B
site 3 neighboring sequence	pBJ250>3L(65B2) insertion site	3L: 6,442,676-6,502,676	Cy5	S1C-D; S3B-C; S12B-C, E; S14F-G
3L-2R control probe	*egfr* DNA	2R: 21,520,393-21,560,246	Cy3	S1C
Transgene U	*sp1* DNA	X: 9,697,559-9,778,741	Cy3	S1C-D
Transgene V	*merlin* 25-kb left extension+*merlin* DNA+*merlin* 25-kb right extension	X: 19,663,948-19,718,977	Cy3	S1C
Transgene W	*scalloped* 25-kb left extension+*scalloped* DNA	X: 15,778,880-15,827,986	Cy3	S1C
Transgene X	*yki* 25-kb left extension+*yki* DNA+*yki* 25-kb right extension	2R: 24,040,405-24,093,757	Cy3	S1C
Transgene AA	upstream of *clamp* DNA	2L: 22,115,720-22,165,720	Cy3	S1C
Transgene BB	downstream of *smo* DNA	2L: 282,167-332,167	Cy3	S1C
Transgene CC	*CG15930* 25-kb left extension+*CG15930* DNA+*CG15930* 25-kb right extension	X: 5,288,125-5,342,409	Cy3	S1D
Transgene DD	*phf7* 25-kb left extension+*phf7* DNA+*phf7* 25-kb right extension	X: 20,134,872-20,191,696	Cy3	S1D

neighboring duplication breakpoint	*spineless* duplication onto chromosome 2	2R: 14,522,912-14,582,912	Cy5	S8C-D
site 4 neighboring sequence	pBJ250>J36 insertion site	3R: 11,748,607-11,808,607	Cy5	S12H-I, K
site 5 neighboring sequence	downstream of 92F1 insertion site	3R: 20,549,650-20,599,650	Cy5	S12N-O
secondary sequence 1	sec 1	N/A	Cy3	Targets all Cy3-conjugated probes
secondary sequence 2	sec 2	N/A	Cy5	Targets all Cy5-conjugated probes

Antibodies

Antibodies and dilutions were as follows: mouse anti-Lamin B (DSHB ADL67.10 and ADL84.12), 1:100; rabbit anti-GFP (Invitrogen), 1:500; rabbit anti-Rh4 (gift from C. Zuker, Columbia University), 1:50; mouse anti-Rh3 (gift from S. Britt, University of Texas at Austin), 1:50; mouse anti-Prospero (DSHB MR1A), 1:10; rat anti-Elav (DSHB 7E8A10), 1:50; guinea pig anti-Ss (gift from Y.N. Jan, University of California, San Francisco), 1:500. All secondary antibodies (Molecular Probes) were Alexa Fluor-conjugated and used at a dilution of 1:400.

Antibody staining (pupal and adult eyes)

Dissections were performed as described in references (*9, 12-14*). Eyes were dissected and fixed at room temperature for 15 minutes in 4% formaldehyde diluted in 1X PBX (PBS+0.3% Triton-X), then washed three times in 1X PBX. Eyes were incubated overnight at room temperature in primary antibody diluted in 1X PBX, then washed three times in 1X PBX and incubated in PBX at room temperature for ≥3 hours. Secondary antibody diluted in 1X PBX was added and incubated overnight at room temperature. Eyes were then washed three times in 1X PBX and incubated in PBX at room temperature for ≥2 hours. Adult eyes were mounted in SlowFade Gold (Invitrogen), and pupal eyes were mounted in Vectashield (Vector Laboratories, Inc.). Images were acquired on a Zeiss LSM700 confocal microscope.

The adult eye dissection protocol was used for **Fig. 3H, J, L, P, R, T; Fig. S10C-E; Fig. S12D, F, J, L, P; Fig. S13B-C, E; Fig. S15B, E, H, K; Fig. S16B, D;** and **Fig. S17B, E, H, K**. The pupal dissection protocol was used for **Fig. 3C** and **Fig. S11B-J**.

Oligopaints probe design

Probes for DNA FISH were designed using the Oligopaints technique (*15, 16*). Target sequences were run through the bioinformatics pipeline available at http://genetics.med.harvard.edu/oligopaints/ to identify sets of 42-bp (for old ss 90K probes) or 50-bp (for all other probes) optimized probe sequences (i.e. “libraries”) tiled across the DNA sequence of interest. Five 19-bp barcoding primers, gene F and R; universal (univ) F and R, and either sublibrary (sub) F or random (rando) R, were appended to the 5’ and 3’ ends of each probe sequence (**Fig. S18A-B**). To ensure that all probes were the same length, an additional 8-bp random sequence was added to the 3’ end of the old ss 90K probes. The gene F and R primers allowed PCR amplification of a probe library of interest out of the total oligo pool, and the univ F and R primers allowed conjugation of fluorophores, generation of single-stranded DNA probes, and PCR addition of secondary sequences to amplify probe signal. The *ss* 50-kb left and right extension libraries had a sub F primer between the gene and universal forward primers to allow PCR amplification of probes targeting a

specific sub-region of the locus of interest (**Fig. S18A**). All other probe libraries had a rando R primer appended at the 3' end to maintain a constant sequence length between all probes (**Fig. S18B**).

Barcoding primer sequences were taken from a set of 240,000 randomly generated, orthogonal 25-bp sequences (*17*) and run through a custom script to select 19-bp sequences with ≤15-bp homology to the *Drosophila* genome. Primers were appended to probe sequences using the orderFile.py script available at http://genetics.med.harvard.edu/oligopaints/. Completed probe libraries were synthesized as custom oligo pools by Custom Array, Inc. (Bothell, WA), and fluorescent FISH probes were generated as described in references (*15, 16*).

DNA FISH

DNA FISH was performed using modified versions of the protocols described in references (*15, 16*). 20-50 eye-antennal discs attached to mouth hooks from third instar larvae were collected on ice and fixed in 129 μL ultrapure water, 20 μL 10X PBS, 1 μL Tergitol NP-40, 600 μL heptane, and 50 μL fresh 16% formaldehyde. Tubes containing the fixative and eye discs were shaken vigorously by hand, then fixed for 10 minutes at room temperature with nutation. Eye discs were then given three quick washes in 1X PBX, followed by three five-minute washes in PBX at room temperature with nutation. Eye discs were then removed from the mouth hooks and blocked for 1 hour in 1X PBX+1% BSA at room temperature with nutation. They were then incubated in primary antibody diluted in 1X PBX overnight at 4°C with nutation. Next, eye discs were washed three times in 1X PBX for 20 minutes and incubated in secondary antibody diluted in 1X PBX for two hours at room temperature with nutation. Eye discs were then washed two times for 20 minutes in 1X PBX, followed by a 20-minute wash in 1X PBS. Next, discs were given one 10-minute wash in 20% formamide+2X SSCT (2X SSC+.001% Tween-20), one 10-minute wash in 40% formamide+2X SSCT, and two 10-minute washes in 50% formamide+2X SSCT. Discs were then predenatured by incubating for four hours at 37°C, three minutes at 92°C, and 20 minutes at 60°C. Primary probes were added in 45 μL hybridization buffer consisting of 50% formamide+2X SSCT+2% dextran sulfate (w/v), + 1 μL RNAse A. All probes were added at a concentration of ≥5 pmol fluorophore/μL. For FISH experiments in which a single probe was used, 4 μL of probe was added. For FISH experiments in which two probes were used, 2 μL of each probe was added. After addition of probes, eye discs were incubated at 91°C for three minutes and at 37°C for 16-20 hours with shaking. Eye discs were then washed for 1 hour at 37°C with shaking in 50% formamide+2X SSCT. 1 μL of each secondary probe was added at a concentration of 100 pmol/μL in 50 μL of 50% formamide+2X SSCT. Secondary probes were hybridized for 1 hour at 37°C with shaking. Eye discs were then washed twice for 30 minutes in 50% formamide+2X SSCT at 37°C with shaking, followed by three 10-minute washes at room temperature in 20% formamide+2X SSCT, 2X SSCT, and 2X SSC with nutation. Discs were mounted in SlowFade Gold immediately after the final 2X SSC wash, and imaged using a Zeiss LSM700 confocal microscope.

Generation of CRISPR lines

CRISPR lines were generated as described in references (*18-21*). For both $ss^{enh\ del}$ and $ss^{upstream\ del}$, sense and antisense DNA oligos for the forward and reverse strands of four gRNAs were designed to generate BbsI restriction site overhangs. The oligos were annealed and cloned into the pCFD3 cloning vector (Addgene, Cambridge, MA). A single-stranded DNA homology bridge was generated with 60-bp homologous regions flanking each side of the predicted cleavage site and an EcoRI (for $ss^{enh\ del}$) or NaeI (for $ss^{upstream\ del}$) restriction site to aid in genotyping. The gRNA constructs

(125 ng/μl) and homologous bridge oligo (100 ng/μl) were injected into *Drosophila* embryos (BestGene, Inc., Chino Hills, CA). Single males were crossed with a balancer stock (*yw; +; TM2/TM6B*), and F1 female progeny were screened for the insertion via PCR, restriction digest, and sequencing. Single F1 males whose siblings were positive for the deletion were crossed to the balancer stock (*yw; +; TM2/TM6B*), and the F2 progeny were screened for the deletion via PCR, restriction digest, and sequencing. Deletion-positive flies from multiple founders were used to establish independent stable stocks.

The following oligos were used for the $ss^{enh\ del}$ CRISPR:

Oligo name	Sequence
Homologous bridge	CAATTTAATTGAGCTCCCAAGTGCTGGGAAGCAGCTGCCCTTTGAATTGGGCTTCTCACCGAATTC TGGCCTGGCTTTGGAGCTCCTTTTGGTGAGAGACCAAAAGAGATTCCGCTGCGCGAATCG
gRNA 1F	GTCGTAATATTCGCTAGGACCTA
gRNA 1R	AAACTAGGTCCTAGCGAATATTAC
gRNA 2F	GTCGAATTGGGCTTCTCACCCCT
gRNA 2R	AAACAGGGGTGAGAAGCCCAATTC
gRNA 3F	GTCGCCAGGCCATGTGGGCATTT
gRNA 3R	AAACAAATGCCCACATGGCCTGGC
gRNA 4F	GTCGCTCCAAAGCCAGGCCATGT
gRNA 4R	AAACACATGGCCTGGCTTTGGAGC
genotype F	CTTAGCTTCAAGCGGCTCCG
genotype R	GAATAACGTCAACTGTGCCA

The following oligos were used for the $ss^{upstream\ del}$ CRISPR:

Oligo name	Sequence
Homologous bridge	TGAGTTGATTGAAGGCTGTAAGAGCAGATTACAGTGGGGCGGAGGCCCAAGTCTGGATCT GCCGGCCTCTGGGTATTCATTTTTTTCGACTTGGCAATTGCAAATGCAAAACCATTTCATTTGCCG
gRNA 1F	GTCGTCGTCTAGCCTAGAAGCGTT
gRNA 1R	AAACAACGCTTCTAGGCTAGACGA
gRNA 2F	GTCGGGCCCAAGTCTGGATCTCCC
gRNA 2R	AAACGGGAGATCCAGACTTGGGCC
gRNA 3F	GTCGCAAAACAATATGAGGTCTAA
gRNA 3R	AAACTTAGACCTCATATTGTTTTGC
gRNA 4F	GTCGAAGTGGCCTGGGCTTATCTC
gRNA 4R	AAACGAGATAAGCCCAGGCCACTT
genotype F	GACCATTTAAGCGGCTACAAA
genotype R	GGTGGTCAGTCGGCAAATGAA

Scanning electron microscopy

Adult *Drosophila* heads were removed and immediately mounted on a pin stub without fixation or sputtering. Heads were imaged at high vacuum at a voltage of 1.5 kV. All SEM was performed on a FEI Quanta ESEM 200 scanning electron microscope. SEM was used for **Fig. 4D, F, H, J; Fig. S15C, F, I, L;** and **Fig. S17C, F, I, L.**

Pairing quantifications

All quantifications were performed in 3D on z-stacks with a slice thickness of 0.2 μm. Quantifications were performed manually using Fiji (*22, 23*). To chart the z position of each FISH dot, a line was drawn through the dot and the Plot Profile tool was used to assess the stack in which the

dot was brightest. To determine the x-y distance between the two FISH dots, a line was drawn from the center of one dot to the center of the other dot and the length of the line was measured with the Plot Profile tool. The distance between the FISH dots was then calculated in 3D. A total of 50 nuclei from three eye discs were quantified for each genotype (i.e. N=3, n=50).

For experiments in which the transgene and endogenous site were both labeled with red fluorescent probes, FISH punctae ≤0.4 μm apart could not be distinguished as separate and were assigned a distance of 0.4 μm apart. For all controls in **Fig. 2F**, green probes labeling the transgene insertion site were pseudocolored red and data were quantified in the same way as experiments in which the transgene and endogenous site were both labeled with red probes. Thus, 3L-X control data in **Fig. 2F** are the same as in **Fig. S1E**, but the data were re-quantified with the green probes pseudocolored red. Similarly, 2L-3R control data in **Fig. 2F** are the same as in **Fig. 1H, S10J,** and **S12Q** (site 1 control), but the data were re-quantified with the green probes pseudocolored red.

Adult eye quantifications

The frequencies of Rh4- and Rh3-expressing R7s were scored manually for at least eight eyes per genotype. R7s co-expressing Rh3 and Rh4 were scored as Rh4-positive. 100 or more R7s were scored for each eye. For **Fig. S17E, H,** and **K**, only the ventral half of each eye was scored.

Hi-C mapping and TAD calling

Directionality index scores were calculated across 15-kb windows, stepping every 5 kb, by finding the log2 transform of the difference in the ratios of downstream versus upstream summed observed over expected interactions ranging from 15 kb to 100 kb in size. The expected value of a bin was defined as the sum of the product of fragment corrections for each valid fragment pair with both interaction fragments falling within the bin.

Directionality indices were generated using 14 published Hi-C datasets (*24-27*):

Dataset	NCBI Accession Number
1	GSE38468
2	GSE38468
3	GSE61471
4	GSE61471
5	GSE61471
6	GSE63515
7	GSE63515
8	GSM2679637
9	GSM2679640
10	GSM2679641
11	GSM2679642
12	GSM2679643
13	GSM2679644
14	GSM2679645

TADs were read from the beginning of a positive directionality index peak to the end of a negative directionality index peak. Parameters for calling a TAD were as follows: **1)** The positive peak must have a signal of ≥0.8; **2)** The negative peak must have a signal of ≤-0.8; and **3)** The TAD must be

present in at least two datasets. Any transgene covering ≥95% of a TAD was considered to span a TAD.

mRNA sequencing and analysis

RNA-seq was performed on three biological replicates, each consisting of 30 third instar larval eye discs. Eye discs were dissected in 1X PBS, separated from the mouth hooks and antennal discs, and placed directly into 300 μL of Trizol. RNA was purified using a Zymo Direct-zol RNA MicroPrep kit (catalog number R2062). mRNA libraries were prepared using an Illumina TruSeq Stranded mRNA LT Sample Prep Kit (catalog number RS-122-2101). Sequencing was performed using an Illumina NextSeq 500 (75 bp, paired end). Sequencing returned an average of 23,048,349 reads per replicate.

The following pipeline was used for mRNA-sequencing analysis: 1) FASTQ sequencing datasets were assessed for quality using FastQC; 2) Pseudoalignment with the *Drosophila* dm6 transcriptome and read quantifications were performed using kallisto (*28*); 3) Transcript abundance files generated by kallisto were joined to a file containing the genomic coordinates of all *Drosophila* mRNA transcripts (dmel-all-r6.20.gtf, available from Flybase); 4) The joined transcript coordinate file was compared to a file containing the coordinates of all tested transgenes using the bedtools intersect tool (http://bedtools.readthedocs.io/en/latest/content/tools/intersect.html). The output file contained a list of all of the active genes per transgene.

Assessment of chromatin marks and ncRNA, Polycomb Group Complex, and insulator density

ncRNA content of transgenes was assessed manually using the GBrowse tool on FlyBase. tRNAs, miRNAs, snoRNAs, and lncRNAs were included in the analysis of ncRNA content.

Transgenes were evaluated for insulator binding sites, Polycomb Group Complex binding sites, and the presence of chromatin marks using publicly available modENCODE ChIP-seq datasets. The following ChIP-seq datasets were used for this analysis:

Protein/chromatin mark	**modENCODE dataset ID(s)**
BEAF-32	21
Su(Hw)	27, 901, 4104, 4105
CTCF	769, 770, 908, 2638, 2639
Cp190	22
Mod(Mdg4)	24, 4094
GAF	23, 2568, 3238, 3245, 3397, 3814, 3830, 5028
Pcl	3237, 3813, 3816, 3960
Pc	325, 326, 816, 948, 3791, 3957, 5064
dRING	927, 928, 3750, 5071, 5255
Pho	3894
H3K27me3	346, 767, 869, 919
H3K4me3	392, 397, 967

For each protein or chromatin mark, .bed files containing the genomic coordinates of all ChIP peaks in each ModENCODE dataset were downloaded and merged into one file using the bedtools merge tool (http://bedtools.readthedocs.io/en/latest/content/tools/merge.html). The merged file was compared to a .bed file containing the genomic coordinates of all transgenes using the bedtools intersect tool (http://bedtools.readthedocs.io/en/latest/content/tools/intersect.html). This pipeline output the number of protein or chromatin mark ChIP peaks contained in each transgene. The

number of ChIP peaks for BEAF-32, Su(Hw), CTCF, Cp190, Mod(Mdg4), and GAF were added together to calculate the total number of insulator binding sites per transgene in **Fig. 2H**.

Statistical analysis

All datasets were tested for a Gaussian distribution using a D'Agostino and Pearson omnibus normality test and a Shapiro-Wilk normality test. If either test indicated a non-Gaussian distribution, datasets were tested for statistical significance using a Wilcoxon rank-sum test (for single comparisons) or a one-way ANOVA on ranks with Dunn's multiple comparisons test (for multiple comparisons). If both the D'Agostino and Pearson and the Shapiro-Wilk tests indicated a Gaussian distribution, datasets were tested for statistical significance using an unpaired t-test with Welch's correction (for single comparisons) or an ordinary one-way ANOVA with Dunnett's multiple comparisons test (for multiple comparisons).

Supplemental Figure Legends

Supplemental Figure 1: A subset of transgenes interspersed across the genome drive pairing.
A-B. Pairer and non-pairer transgenes from the initial screen of a ~1 Mb region of chromosome 3R. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.
C. Additional transgenes taken from chromosomes X (*Transgenes U-W*), 2R (*Transgene X*), and 2L (*Transgenes AA and BB*) and inserted at site 3.
D. Additional transgenes taken from chromosome X and inserted at site 3. Pairing was assessed with a two-color FISH strategy. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.
E. Quantifications for transgenes in **Fig. S1D**. Neither transgene contained a TAD. Black: control, gray: non-pairers. ns=p>0.05, one way ANOVA on ranks with Dunn's multiple comparisons test. Control data are the same as in **Fig. 2F** (3L-X control), but were quantified with the probes labeling the transgene insertion site in green and the probes labeling the endogenous site in red.

Supplemental Figure 2: Probes neighboring paired sequences give offset probe signals.
A. Schematic of FISH strategy used to label transgene and endogenous sequences and the region directly neighboring the transgene insertion site.
B. Probes directly neighboring the transgene insertion site could be distinguished from probes labeling the transgene sequence itself, despite being immediately downstream on the DNA.
C. When a transgene drives pairing with its endogenous site, the two copies of the transgene are paired with each other and the two copies of the endogenous site are paired with each other due to homologous chromosome pairing. Therefore, one green FISH puncta (neighboring the transgene insertion site) and one red FISH puncta (neighboring the endogenous site) are observed.
The experiment in **Fig. S2B** showed that two sets of probes targeting neighboring regions on the DNA could be distinguished from each other. As the red and green probes in **Fig. S2C** were neighboring the paired sites, not directly targeting the paired sites, it was therefore feasible that their signals did not completely overlap, despite being close in 3D space.

Supplemental Figure 3: Buttons drive pairing in a position-independent manner.
A. DNA FISH strategy used to assess pairing of *Transgene E* inserted at site 3 with its endogenous locus.
B-C. Control and *Transgene E* at site 3. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.
D. Quantifications for **Fig. S3B-C**. Black: control, blue: pairer. *=p<0.05, Wilcoxon rank-sum test. Data for control are the same as in **Fig. S12Q** (site 3 control).

Supplemental Figure 4: TAD calls across 14 Hi-C datasets for the region on chromosome 3R used for the initial pairing screen.
A. Red lines indicate a directionality index signal of 0.8 or -0.8. Black bars indicate TAD calls.

Supplemental Figure 5: TAD calls across 14 Hi-C datasets for Transgenes U-Y.
A-E. Red lines indicate a directionality index signal of 0.8 or -0.8. Black bars indicate TAD calls.

Supplemental Figure 6: TAD calls across 14 Hi-C datasets for Transgenes Z-DD.
A-E. Red lines indicate a directionality index signal of 0.8 or -0.8. Black bars indicate TAD calls.

Supplemental Figure 7: A higher percentage of pairers encompass entire TADs than non-pairers
A. Representative directionality indices showing the percentage of a TAD covered by each pairing transgene. Black bars indicate final TAD calls generated from analysis of 14 Hi-C datasets (*24-27*). Representative directionality indices are from NCBI accession numbers GSE38468 (*Transgenes C, U)*, GSE61471 (*Transgenes A, B, E, O, W, X*), GSM2679637 (*Transgene Y*), and GSE63515 (*Transgene V*).
B. Representative directionality indices showing the percentage of a TAD covered by each non-pairing transgene. Black bars indicate final TAD calls generated from analysis of 14 Hi-C datasets(*24-27*). Representative directionality indices are from NCBI accession numbers GSE38468 (*Transgenes G, H, N, P*), GSE61471 (*Transgenes D, F, I, J, K, L, M, Z, BB, CC*), GSE63515 (*Transgene DD*), GSM2679644 (*Transgene Q*), and GSM2679637 (*Transgene AA*).

Supplemental Figure 8: A 460-kb duplication drives pairing.
A. Representative HiC heat map and directionality index (NCBI GSE38468) showing large TAD covered by the duplication. Dotted lines: TAD boundaries. Black bar: TAD. See **Fig. S4A** for TAD assessment.
B. DNA FISH strategy used to assess pairing of the duplication with its endogenous locus.
C-D. Control and duplication. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous locus, green: probes neighboring duplication breakpoint.
E. Quantifications for **Fig. S3C-D**. Black: control, blue: pairer. *=p<0.05, Wilcoxon rank-sum test.

Supplemental Figure 9: **Pairing is not correlated with Polycomb Group Complex binding sites, repressive chromatin marks, or ncRNAs.**
A-F. Graphs showing the number of Polycomb Group Complex or H3K27me3 ChIP peaks or the number of ncRNAs per transgene for all pairers and non-pairers tested in **Fig. 1**, **2**, and **S1**. Blue: pairers, gray: non-pairers. ns=p>0.05, Wilcoxon rank-sum test.

Supplemental Figure 10: The *ss* button drives pairing and transvection despite chromosome rearrangements.
A. $ss^{inversion}$ / + example. Scale bar=1 μm. White: Lamin B, red: probes against endogenous *ss*.
B. Quantification of **Fig. S10A**. ns=p>0.05, Wilcoxon rank-sum test. Data for wild type are the same as in **Fig. S10J**. Blue: pairers.
C. Schematic and representative image of $ss^{inversion}$ / + adult R7s. Ss(Rh4)=63%. $ss^{inversion}$ / + had no effect on the normal Rh3:Rh4 ratio. ins: insulator, sil 1: silencer 1, enh: enhancer, sil 2: silencer 2. Smaller black arrows: transcription start sites. Red: Rh4, blue: Rh3.
D. Schematic and representative image of $ss^{inversion}$ / $ss^{high\ freq\ null}$ adult R7s. Ss(Rh4)=80%. $ss^{high\ freq\ null}$ produced no functional Ss protein, but it performed transvection to increase the expression frequency of *ss* on other chromosomes (*3*). $ss^{high\ freq\ null}$ upregulated expression frequency from $ss^{inversion}$, indicating that $ss^{inversion}$ performed transvection. Black X indicates that there is a mutation in the second silencer of *ss* that disrupts the protein coding sequence of $ss^{high\ freq\ null}$. Smaller black arrows: transcription start sites. Red: Rh4, blue: Rh3.

E. Schematic and representative image of *Transgene S-ss*^{def} / *ss*^{inversion} adult R7s. Ss(Rh4)=99%. *Transgene S* was recombined onto a chromosome with a *ss* deficiency to examine *Transgene S* transvection with mutant *ss* alleles. *Transgene S* performed transvection to upregulate expression of wild type, endogenous *ss* (*3*). *Transgene S* upregulated expression of *ss* on the *ss*^{inversion} allele, indicating that *ss*^{inversion} performed transvection. Red: Rh4, blue: Rh3.
F&H. DNA FISH strategies used to assess endogenous *ss* pairing in wild type and chromosome rearrangement backgrounds.
G&I. Wild type and chromosome rearrangement examples. Scale bars=1 μm. White: pm181 (R7 marker)>GFP in **B**, Lamin B in **D**. Red: probes against endogenous *ss.*
J. Quantifications for **Fig. S10G, I**. ****=p<0.0001, one-way ANOVA on ranks with Dunn's multiple comparisons test. Negative control data are the same as in **Fig. 1H**, **2F** (2L-3R control), and **S12Q** (site 1 control). Data for wild type are the same as in **Fig. S10B**. Black: control, blue: pairers.

Supplemental Figure 11: *Transgenes S* and *T* are expressed in 100% of R7 photoreceptors at all insertion sites.
A. Schematic of *Transgenes S* and *T* with GFP tags. ins: insulator, sil1: silencer, enh: enhancer. Black arrows: transcription start sites.
B-F. Representative images of *Transgene S*>GFP expression in mid-pupal R7 photoreceptors for all insertion sites. Red: Elav (photoreceptors), blue: Prospero (R7s), green: *Transgene S*>GFP. White circles indicate representative R7s.
G-J. Representative images of *Transgene T*>GFP expression in mid-pupal R7 photoreceptors for all insertion sites. Red: Elav (photoreceptors), blue: Prospero (R7s), green: *Transgene T*>GFP. White circles indicate representative R7s.

Supplemental Figure 12: Pairing is necessary but not sufficient for *ss* transvection.
A, G, M. DNA FISH strategies used to test pairing and transvection of *Transgenes S* and *T* at each insertion site. Gray arrow with "?" indicates that *Transgenes S* and *T* were tested for transvection.
B-C, E, H-I, K, N-O. DNA FISH examples for control, *Transgene S*, and *Transgene T* at each insertion site. Scale bars=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.
D, F, J, L, P. Representative images of adult eyes for *Transgene S* and *Transgene T* at each insertion site. Red: Rh4, blue: Rh3.
D. Ss(Rh4)=71%
F. Ss(Rh4)=76%
J. Ss(Rh4)=98%
L. Ss(Rh4)=74%
P. Ss(Rh4)=98%

Supplemental Figure 13: *Transgene E* does not perform transvection.
A. Transvection assay to test whether mutant *ss* alleles alter expression from *Transgene E.* Red X indicates *ss* mutant allele, which is *ss*$^{protein\ null}$ from **Fig. S13B** or *ss*$^{high\ freq\ null}$ from **Fig. S13C**. Gray arrow with "?" indicates that *Transgene E* was tested for transvection.
B-C. Rh3 and Rh4 expression in *Transgene E* + *ss*$^{protein\ null}$ / *ss*^{def} (Ss(Rh4)=52%) and *Transgene E* + *ss*$^{high\ freq\ null}$ / *ss*^{def} (Ss(Rh4)=51%).
D. Schematic of *Transgene E* + *Transgene S-ss*^{def} / *ss*$^{protein\ null}$ genotype. *Transgene S* was recombined onto a chromosome with a *ss* deficiency to examine *Transgene S* transvection with

mutant *ss* alleles. *Transgene S* performed transvection to upregulate expression of wild type, endogenous *ss* (*3*). In the *Transgene E* + *Transgene S-ss*^{def} / *ss*$^{protein\ null}$ genotype, the endogenous *ss* locus was hemizygous for a protein coding null allele of *ss*, which produced no functional Ss protein. Therefore, any functional Ss protein in this genotype was produced by *Transgene E*, and an increase in Ss (Rh4) expression frequency indicated that *Transgene S* was performing transvection to upregulate Ss expression from *Transgene E*. Red X over the *ss* locus indicates that the *ss* allele is a protein coding null. Gray arrow with "?" indicates that *Transgene E* was tested for transvection.
E. Adult eye for the *Transgene E* + *Transgene S-ss*^{def} / *ss*$^{protein\ null}$ genotype. *ss* expression frequency was not upregulated (Ss(Rh4)=53%), indicating that *Transgene S* did not perform transvection with *Transgene E*. Red: Rh4, blue: Rh3.

Supplemental Figure 14: Pairing driven by the *ss* button is cell-type specific.
A. DNA FISH strategy used to assess *Transgene E* pairing at site 1 in the larval antenna.
B-C. *Site 1 control* and *Transgene E site 1* in the larval antenna. Scale bars=1 μm. White: Lamin B, red: probes against endogenous *ss* and *Transgene E*.
D. Quantifications for **Fig. S14B-C**. Black: control, gray: non-pairer. ns=p>0.05, unpaired t-test with Welch's correction.
E. DNA FISH strategy used to assess pairing of *Transgene E* at site 3 with its endogenous locus in the larval antenna.
F-G. *Site 3 control* and *Transgene E site 3* examples in the larval antenna. Scale bar=1 μm. White: Lamin B, red: probes neighboring endogenous sequence, green: probes neighboring transgene insertion site.
H. Quantifications for **Fig. S14F-G**. ns=p>0.05, unpaired t-test with Welch's correction.

Supplemental Figure 15: *ss* mutant alleles with arista-specific phenotypes do not perform transvection in the arista.
A, D, G, J. Genotypes tested for transvection. Gray rectangles: exons. Smaller black arrows: transcription start sites. Red X indicates an uncharacterized mutation in the *ss*$^{arista\ 1}$ or *ss*$^{arista\ 2}$ sequence. Red X over gray arrow indicates an absence of transvection between alleles in the arista.
B. *ss*$^{arista\ 1}$ / *ss*^{def} adult eye. *ss*$^{arista\ 1}$ / *ss*^{def} had no effect on eye development. Red: Rh4, blue: Rh3. Ss(Rh4)=53%.
C. *ss*$^{arista\ 1}$ / *ss*^{def} arista. *ss*$^{arista\ 1}$ / *ss*^{def} caused aristapedia. Scale bar= 50 μm. White arrow indicates arista.
E. *ss*$^{arista\ 1}$ / *ss*$^{protein\ null}$ adult eye. *ss*$^{arista\ 1}$ / *ss*$^{protein\ null}$ had no effect on eye development. Red: Rh4, blue: Rh3. Ss(Rh4)=53%.
F. *ss*$^{arista\ 1}$ / *ss*$^{protein\ null}$ arista. *ss*$^{arista\ 1}$ / *ss*$^{protein\ null}$ had aristapedia, indicating that transvection did not occur to rescue the mutant *ss* phenotype. Scale bar= 50 μm. White arrow indicates arista.
H. *ss*$^{arista\ 2}$ / *ss*^{def} adult eye. *ss*$^{arista\ 2}$ / *ss*^{def} had no effect on eye development. Red: Rh4, blue: Rh3. Ss(Rh4)=60%.
I. *ss*$^{arista\ 2}$ / *ss*^{def} arista. *ss*$^{arista\ 2}$ / *ss*^{def} caused aristapedia. Scale bar= 50 μm. White arrow indicates arista.
K. *ss*$^{arista\ 2}$ / *ss*$^{protein\ null}$ adult eye. *ss*$^{arista\ 2}$ / *ss*$^{protein\ null}$ had no effect on eye development. Red: Rh4, blue: Rh3. Ss(Rh4)=62%.
L. *ss*$^{arista\ 2}$ / *ss*$^{protein\ null}$ arista. *ss*$^{arista\ 2}$ / *ss*$^{protein\ null}$ had aristapedia, indicating that transvection did not occur to rescue the mutant *ss* phenotype. Scale bar= 50 μm. White arrow indicates arista.

Supplemental Figure 16: *ss*$^{protein\ null}$ performs transvection in the eye.
A. Schematic of *ss* enhancer deletion (*ss*$^{enh\ del}$) allele over a *ss* deficiency (*ss*^{def}). Gray rectangles: exons. Smaller black arrow: transcription start site.
B. *ss*$^{enh\ del}$ / *ss*^{def} adult eye. *ss*$^{enh\ del}$ / *ss*^{def} caused a near complete loss of Ss/Rh4 expression. Red: Rh4, blue: Rh3. Ss(Rh4)=0.1%.
C. Schematic of *ss*$^{enh\ del}$ over *ss*$^{protein\ null}$. Through transvection, the functional enhancer of *ss*$^{protein\ null}$ acted on the functional protein coding region of *ss*$^{enh\ del}$ to rescue Ss expression (green arrow). Enh: enhancer, gray rectangles: exons. Smaller black arrows: transcription start sites.
D. *ss*$^{enh\ del}$ / *ss*$^{protein\ null}$ adult eye. *ss*$^{protein\ null}$ rescued Ss expression from the *ss*$^{enh\ del}$ allele. Red: Rh4, blue: Rh3. Ss(Rh4)=59%.

Supplemental Figure 17: Transgene S performs transvection in the eye but not in the arista.
A. Schematic of the *ss*$^{upstream\ del}$ allele over the *ss*^{def} allele. *ss*$^{upstream\ del}$ is a CRISPR allele in which 12.7 kb of the upstream regulatory regions of *ss* are deleted. Red X indicates the deletion of regulatory regions directly upstream of the *ss* locus.
B. *ss*$^{upstream\ del}$ / *ss*^{def} adult eye. *ss*$^{upstream\ del}$ displayed Ss(Rh4) expression in 85% of R7s. Red: Rh4, blue: Rh3.
C. *ss*$^{upstream\ del}$ / *ss*^{def} arista. *ss*$^{upstream\ del}$ caused aristapedia. Scale bar= 50 μm. White arrow indicates arista.
D. Schematic of the *Transgene S-ss*^{def} allele over the *ss*$^{upstream\ del}$ allele. *Transgene S* was recombined onto a chromosome with a *ss* deficiency to examine *Transgene S* transvection with mutant *ss* alleles. Red X indicates the deletion of regulatory regions directly upstream of the *ss* locus. Gray arrow with a "?" indicates that *Transgene S* was tested for transvection with the *ss*$^{upstream\ del}$ allele.
E. *Transgene S-ss*^{def} / *ss*$^{upstream\ del}$ adult eye. *Transgene S-ss*^{def} upregulated Ss(Rh4) expression from the *ss*$^{upstream\ del}$ allele into 100% of R7s, indicating that *Transgene S* performed transvection with *ss* $^{upstream\ del}$ in the eye. Red: Rh4, blue: Rh3.
F. *Transgene S-ss*^{def} / *ss*$^{upstream\ del}$ arista. *Transgene S-ss*^{def} / *ss*$^{upstream\ del}$ had aristapedia, indicating that *Transgene S* did not perform transvection to rescue *ss*$^{upstream\ del}$ expression in the arista. Scale bar= 50 μm. White arrow indicates arista.
G. Schematic of the *Transgene S-ss*^{def} allele over the *ss*$^{arista\ 1}$ allele. Red X indicates an uncharacterized mutation in the *ss*$^{arista\ 1}$ allele. Gray arrow with "?" indicates that *Transgene S* was tested for transvection with the *ss*$^{arista\ 1}$ allele.
H. *Transgene S-ss*^{def} / *ss*$^{arista\ 1}$ adult eye. *Transgene S-ss*^{def} upregulated Ss(Rh4) expression from the *ss*$^{arista\ 1}$ allele into 99% of R7s, indicating that *Transgene S* performed transvection with *ss*$^{arista\ 1}$ in the eye. Red: Rh4, blue: Rh3.
I. *Transgene S-ss*^{def} / *ss*$^{arista\ 1}$ arista. *Transgene S-ss*^{def} / *ss*$^{arista\ 1}$ had aristapedia, indicating that *Transgene S* did not perform transvection to rescue *ss*$^{arista\ 1}$ expression in the arista. Scale bar= 50 μm. White arrow indicates arista.
J. Schematic of the *Transgene S-ss*^{def} allele over the *ss*$^{arista\ 2}$ allele. Red X indicates an uncharacterized mutation in the *ss*$^{arista\ 2}$ allele. Gray arrow with "?" indicates that *Transgene S* was tested for transvection with the *ss*$^{arista\ 2}$ allele.
K. *Transgene S-ss*^{def} / *ss*$^{arista\ 2}$ adult eye. *Transgene S-ss*^{def} upregulated Ss(Rh4) expression from the *ss*$^{arista\ 2}$ allele into 100% of R7s, indicating that *Transgene S* performed transvection with *ss*$^{arista\ 2}$ in the eye. Red: Rh4, blue: Rh3.

L. *Transgene S-ss*^{def} / *ss*$^{arista\ 2}$ arista. *Transgene S-ss*^{def} / *ss*$^{arista\ 2}$ had aristapedia, indicating that *Transgene S* did not perform transvection to rescue *ss*$^{arista\ 2}$ expression in the arista. Scale bar= 50 μm. White arrow indicates arista.

Supplemental Figure 18: Barcoding primer scheme for DNA Oligopaints FISH probes.
A. Schematic of barcoding primer scheme for Oligopaints probe libraries containing sublibraries. univ: universal primer, sub: sublibrary primer.
B. Schematic of barcoding primer scheme for Oligopaints probe libraries without sublibraries. univ: universal primer, rando: random primer.

Supplemental References

1. C. H. Lee, T. Herman, T. R. Clandinin, R. Lee, S. L. Zipursky, N-cadherin regulates target specificity in the Drosophila visual system. *Neuron* **30**, 437-450 (2001).
2. K. J. Venken *et al.*, Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. *Nat Methods* **6**, 431-434 (2009).
3. R. J. Johnston, Jr., C. Desplan, Interchromosomal communication coordinates intrinsically stochastic expression between alleles. *Science* **343**, 661-665 (2014).
4. K. J. Venken *et al.*, A molecularly defined duplication set for the X chromosome of Drosophila melanogaster. *Genetics* **186**, 1111-1125 (2010).
5. E. B. Lewis, A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-570 (1978).
6. F. Payne, An experiment to test the nature of the variations on which selection acts. *Indiana Univ Studies* **5**, 1-45 (1918).
7. A. L. Parks *et al.*, Systematic generation of high-resolution deletion coverage of the Drosophila melanogaster genome. *Nat Genet* **36**, 288-292 (2004).
8. D. M. Duncan, E. A. Burgess, I. Duncan, Control of distal antennal identity and tarsal development in Drosophila by spineless-aristapedia, a homolog of the mammalian dioxin receptor. *Genes Dev* **12**, 1290-1303 (1998).
9. R. J. Johnston, Jr. *et al.*, Interlocked feedforward loops control cell-type-specific Rhodopsin expression in the Drosophila eye. *Cell* **145**, 956-968 (2011).
10. G. Morata, P. A. Lawrence, Development of the eye-antenna imaginal disc of Drosophila. *Dev Biol* **70**, 355-371 (1979).
11. C. H. Waddington, A note on some alleles of aristapedia. *J. Genet.* **51**, 123-129 (1952).
12. S. U. Thanawala *et al.*, Regional modulation of a stochastically expressed factor determines photoreceptor subtypes in the Drosophila retina. *Dev Cell* **25**, 93-105 (2013).
13. D. Jukam *et al.*, The insulator protein BEAF-32 is required for Hippo pathway activity in the terminal differentiation of neuronal subtypes. *Development (Cambridge, England)* **143**, 2389-2397 (2016).
14. H. Y. Hsiao *et al.*, Dissection and immunohistochemistry of larval, pupal and adult Drosophila retinas. *J Vis Exp*, 4347 (2012).
15. B. J. Beliveau *et al.*, Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat Commun* **6**, 7147 (2015).
16. B. J. Beliveau *et al.*, Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc Natl Acad Sci U S A* **109**, 21301-21306 (2012).
17. M. R. S. Qikai Xu, Gregory J. Hannon, Stephen J. Elledge, Design of 240,000 orthogonal 25mer DNA barcode probes. *PNAS* **106**, 2289-2294 (2008).
18. C. Anderson *et al.*, Natural variation in stochastic photoreceptor specification and color preference in Drosophila. *Elife* **6**, (2017).
19. J. Yan *et al.*, Regulatory logic driving stable levels of defective proventriculus expression during terminal photoreceptor specification in flies. *Development (Cambridge, England)* **144**, 844-855 (2017).
20. S. J. Gratz, J. Wildonger, M. M. Harrison, K. M. O'Connor-Giles, CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. *Fly (Austin)* **7**, 249-255 (2013).

21. F. Port, H. M. Chen, T. Lee, S. L. Bullock, Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc Natl Acad Sci U S A* **111**, E2967-2976 (2014).
22. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).
23. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012).
24. C. Hou, L. Li, Z. S. Qin, V. G. Corces, Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell* **48**, 471-484 (2012).
25. L. Li *et al.*, Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol Cell* **58**, 216-231 (2015).
26. B. Schuettengruber *et al.*, Cooperativity, specificity, and evolutionary stability of Polycomb targeting in Drosophila. *Cell Rep* **9**, 219-233 (2014).
27. M. R. Stadler, J. E. Haines, M. B. Eisen, Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early Drosophila melanogaster embryo. *Elife* **6**, (2017).
28. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527 (2016).

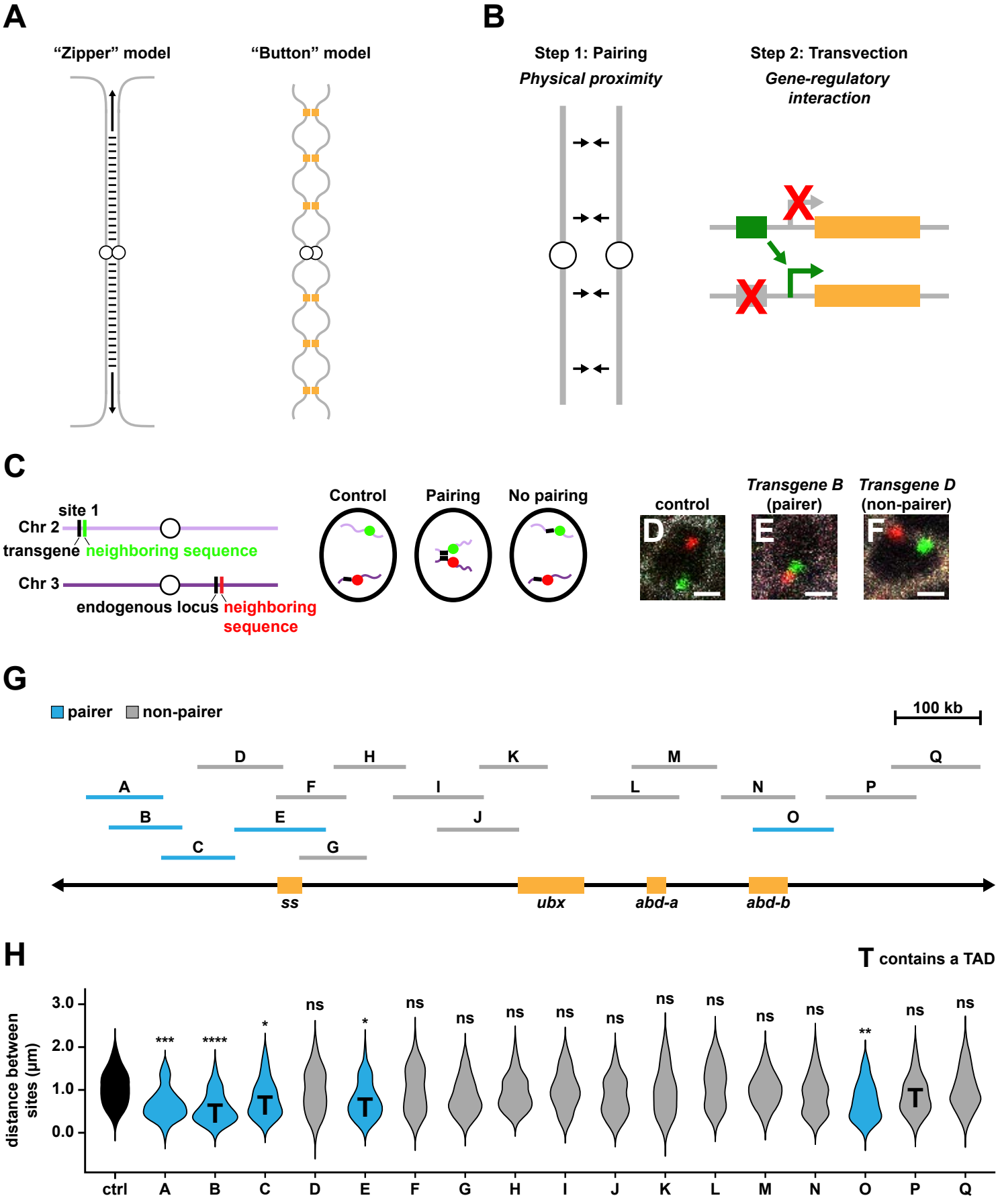

A
"Zipper" model
"Button" model
B
Step 1: Pairing
Physical proximity
Step 2: Transvection
Gene-regulatory interaction
C
site 1
Chr 2
transgene
neighboring sequence
Chr 3
endogenous locus
neighboring sequence
Control
Pairing
No pairing
control
Transgene B (pairer)
Transgene D (non-pairer)
D
E
F
G
pairer
non-pairer
100 kb
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
ss
ubx
abd-a
abd-b
H
T contains a TAD
distance between sites (µm)
3.0
2.0
1.0
0.0
***
****
*
ns
*
ns
ns
ns
ns
ns
ns
ns
ns
ns
**
ns
ns
T
T
T
T
ctrl
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q

Figure 1

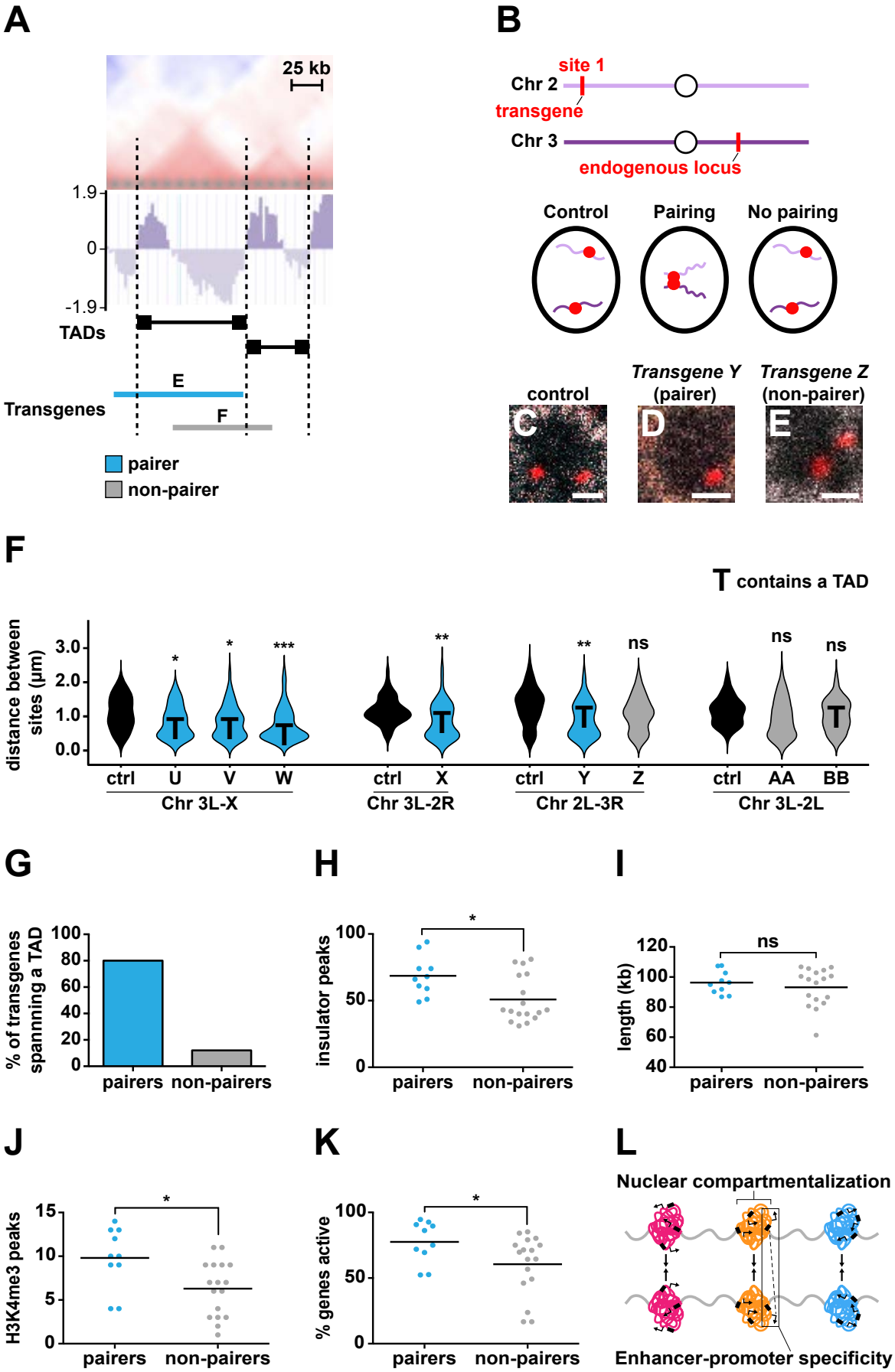

A
25 kb
1.9
0
-1.9
TADs
Transgenes
E
F
pairer
non-pairer
B
site 1
Chr 2
transgene
Chr 3
endogenous locus
Control
Pairing
No pairing
control
Transgene Y (pairer)
Transgene Z (non-pairer)
C
D
E
F
T contains a TAD
distance between sites (µm)
3.0
2.0
1.0
0.0
ctrl
U
V
W
Chr 3L-X
X
Chr 3L-2R
Y
Z
Chr 2L-3R
AA
BB
Chr 3L-2L
ns
G
% of transgenes spannning a TAD
100
80
60
40
20
0
pairers
non-pairers
H
insulator peaks
100
50
0
I
length (kb)
120
100
80
60
40
J
H3K4me3 peaks
15
10
5
0
K
% genes active
L
Nuclear compartmentalization
Enhancer-promoter specificity

Figure 2

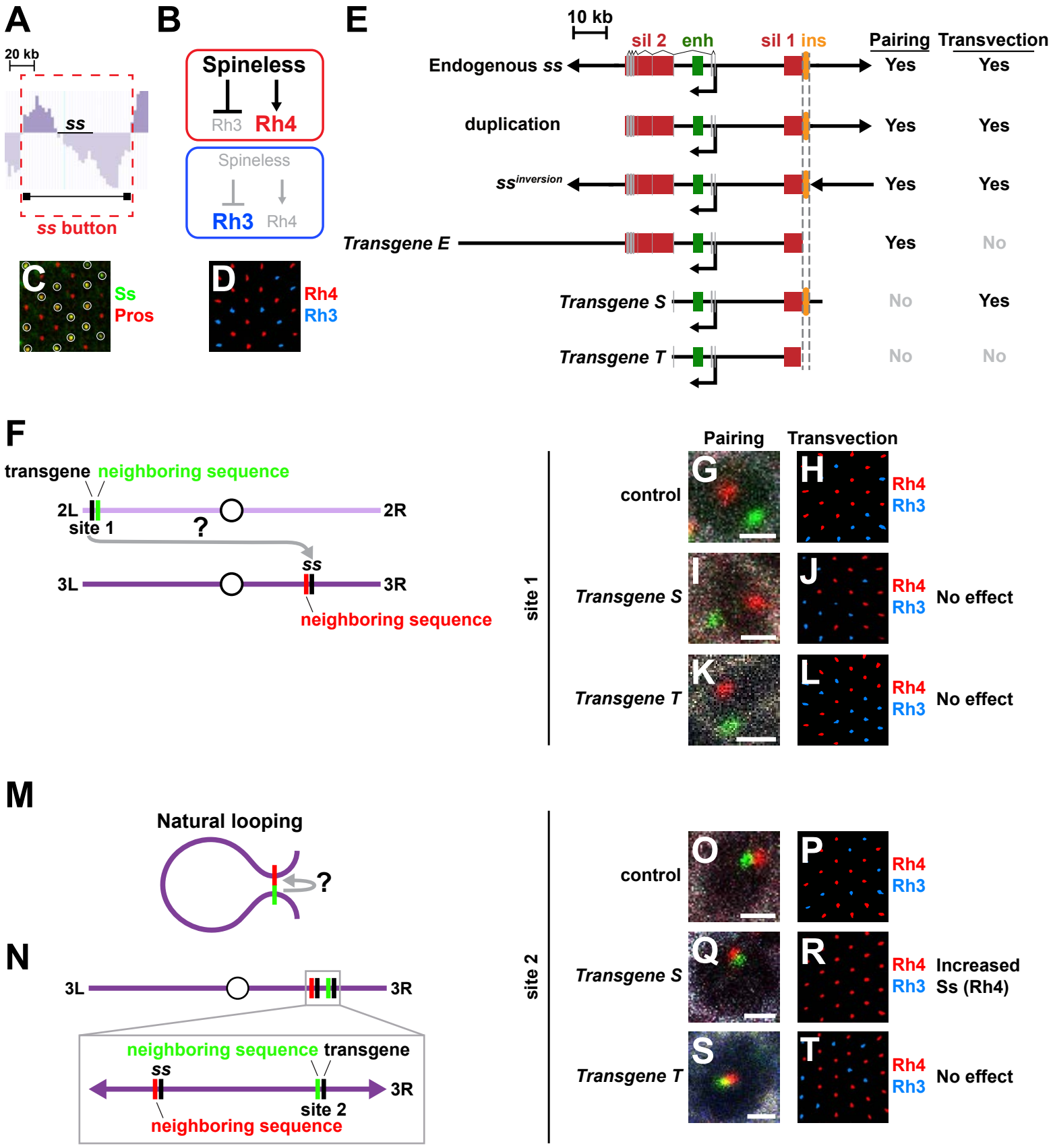

A
20 kb
ss
ss button
B
Spineless
Rh3
Rh4
Spineless
Rh3
Rh4
C
Ss
Pros
D
Rh4
Rh3
E
10 kb
sil 2
enh
sil 1
ins
Pairing
Transvection
Endogenous ss
Yes
Yes
duplication
Yes
Yes
ssinversion
Yes
Yes
Transgene E
Yes
No
Transgene S
No
Yes
Transgene T
No
No
F
transgene
neighboring sequence
2L
site 1
?
2R
3L
ss
3R
neighboring sequence
Pairing
Transvection
control
G
H
Rh4
Rh3
site 1
Transgene S
I
J
Rh4
Rh3
No effect
Transgene T
K
L
Rh4
Rh3
No effect
M
Natural looping
?
N
3L
3R
neighboring sequence
transgene
ss
3R
site 2
neighboring sequence
site 2
control
O
P
Rh4
Rh3
Transgene S
Q
R
Rh4
Rh3
Increased
Ss (Rh4)
Transgene T
S
T
Rh4
Rh3
No effect

Figure 3

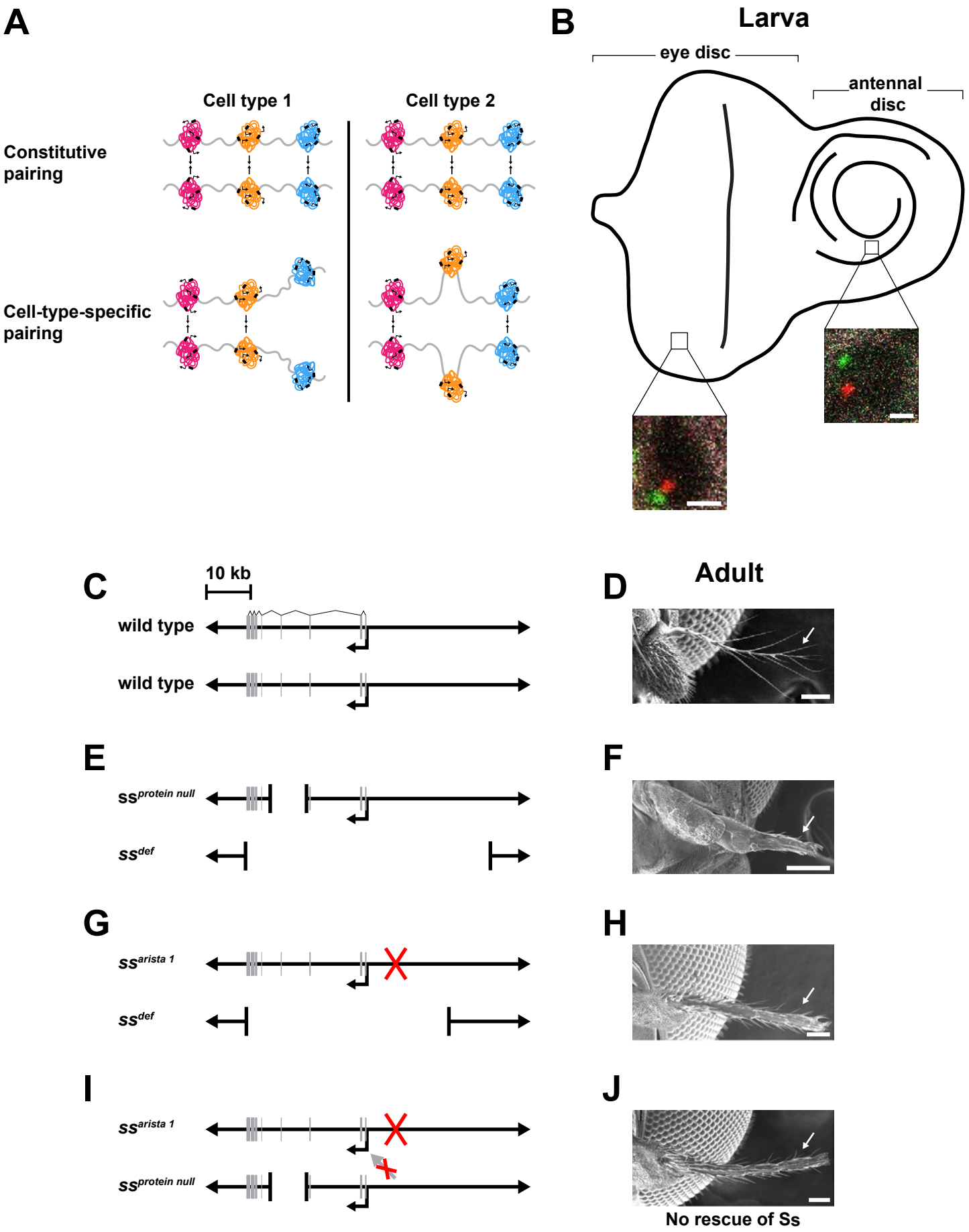
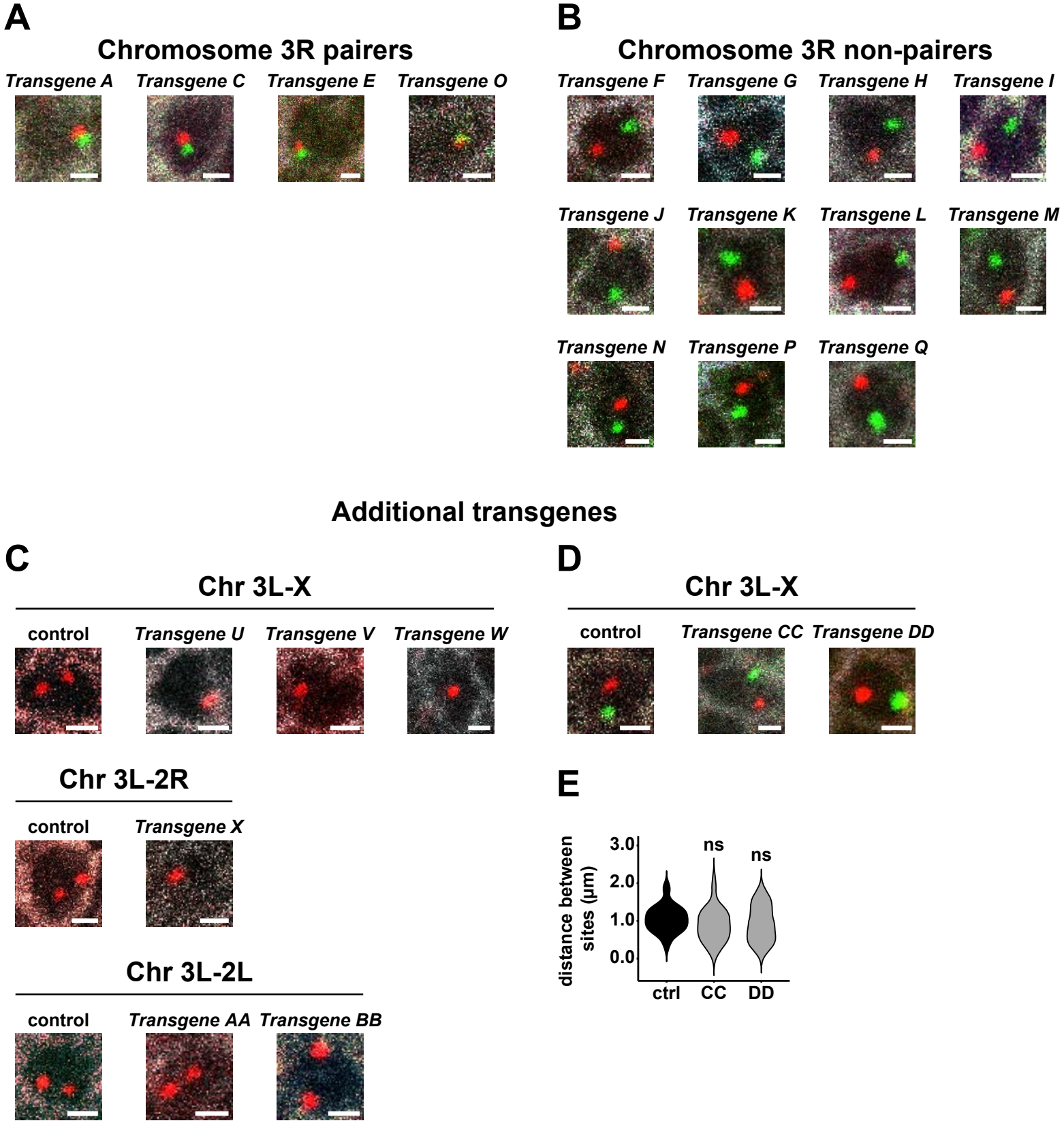
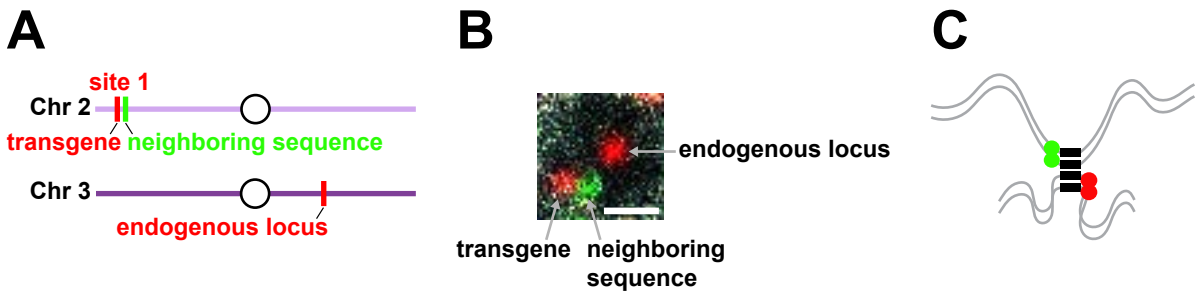

A
Cell type 1
Cell type 2
Constitutive pairing
Cell-type-specific pairing
B
Larva
eye disc
antennal disc
C
10 kb
wild type
wild type
D
Adult
E
ss protein null
ss def
F
G
ss arista 1
ss def
H
I
ss arista 1
ss protein null
J
No rescue of Ss

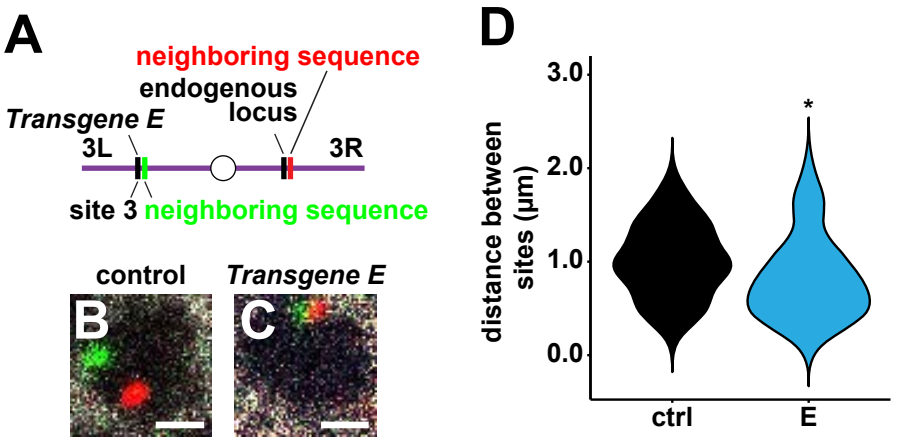
Figure 4



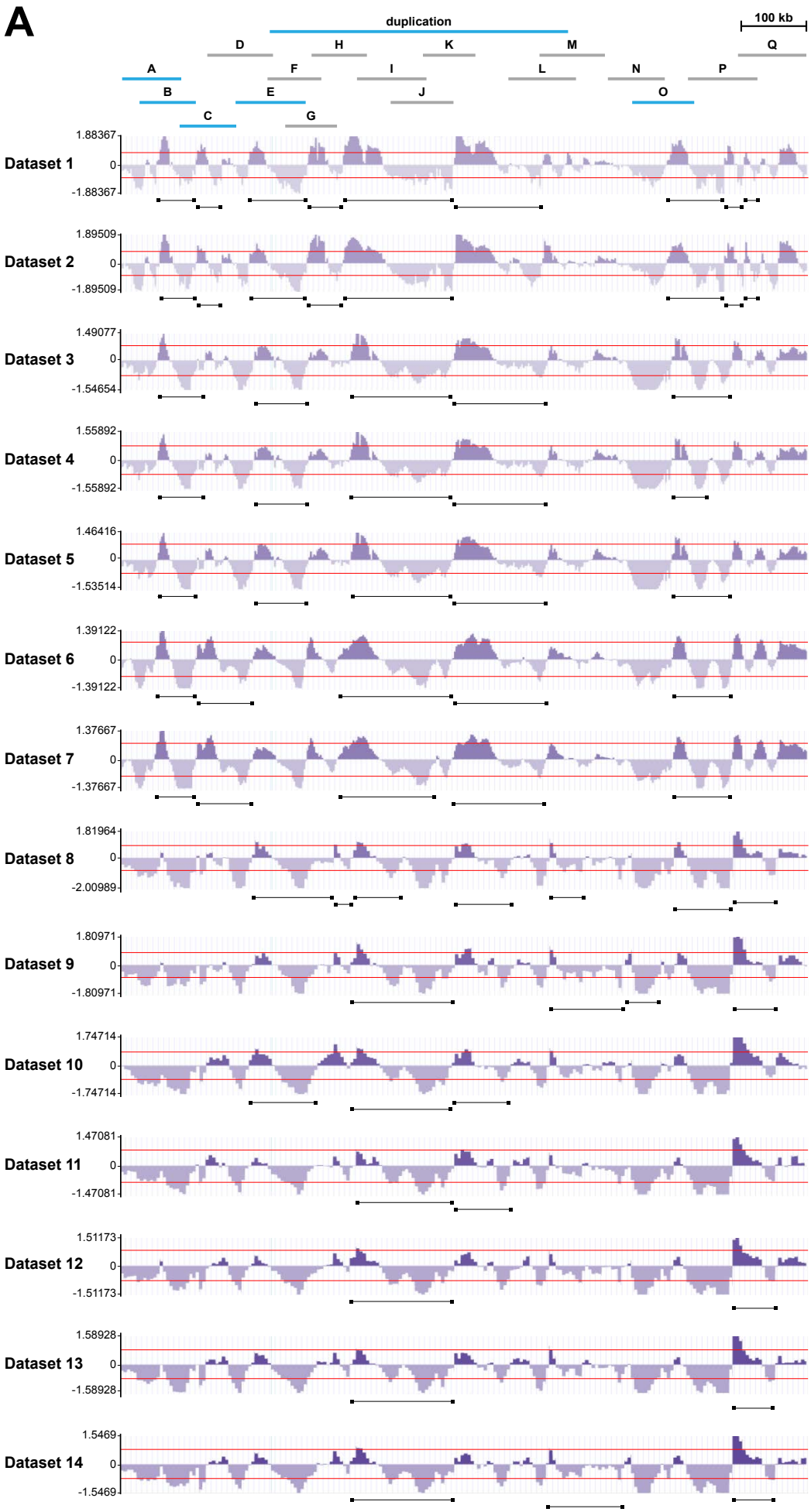
Supplemental Fig. 1



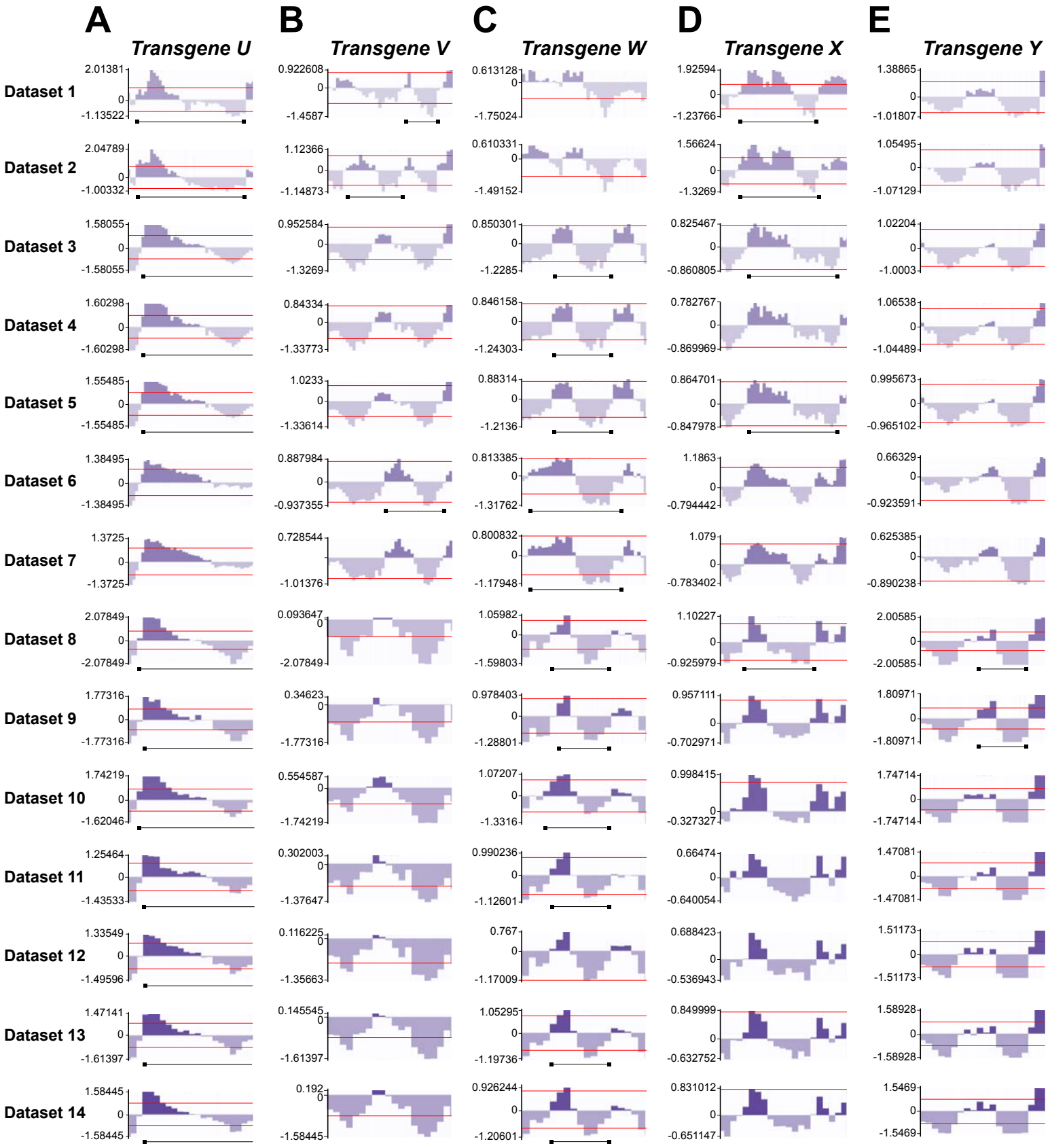
Supplemental Fig. 2



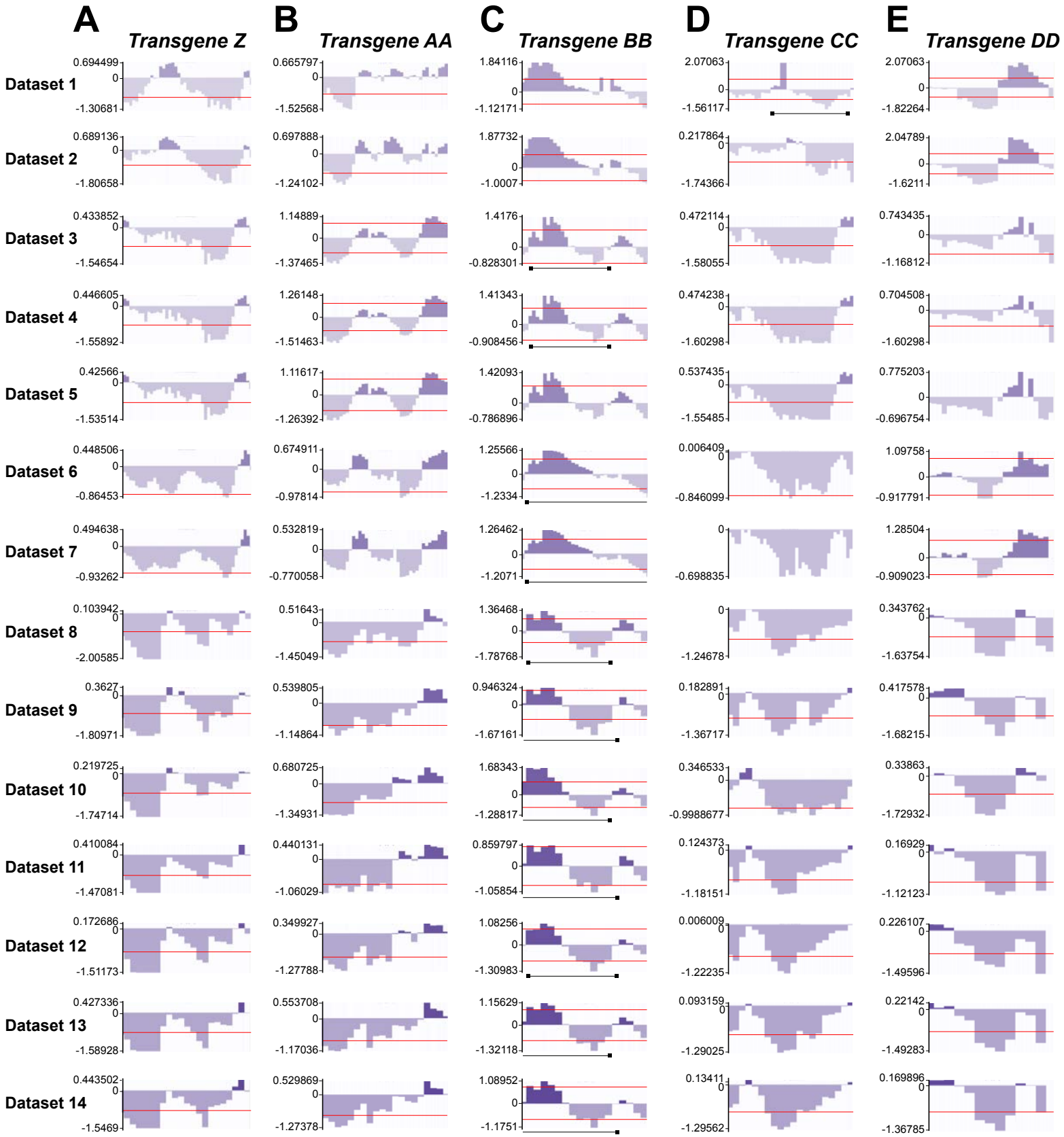
Supplemental Fig. 3



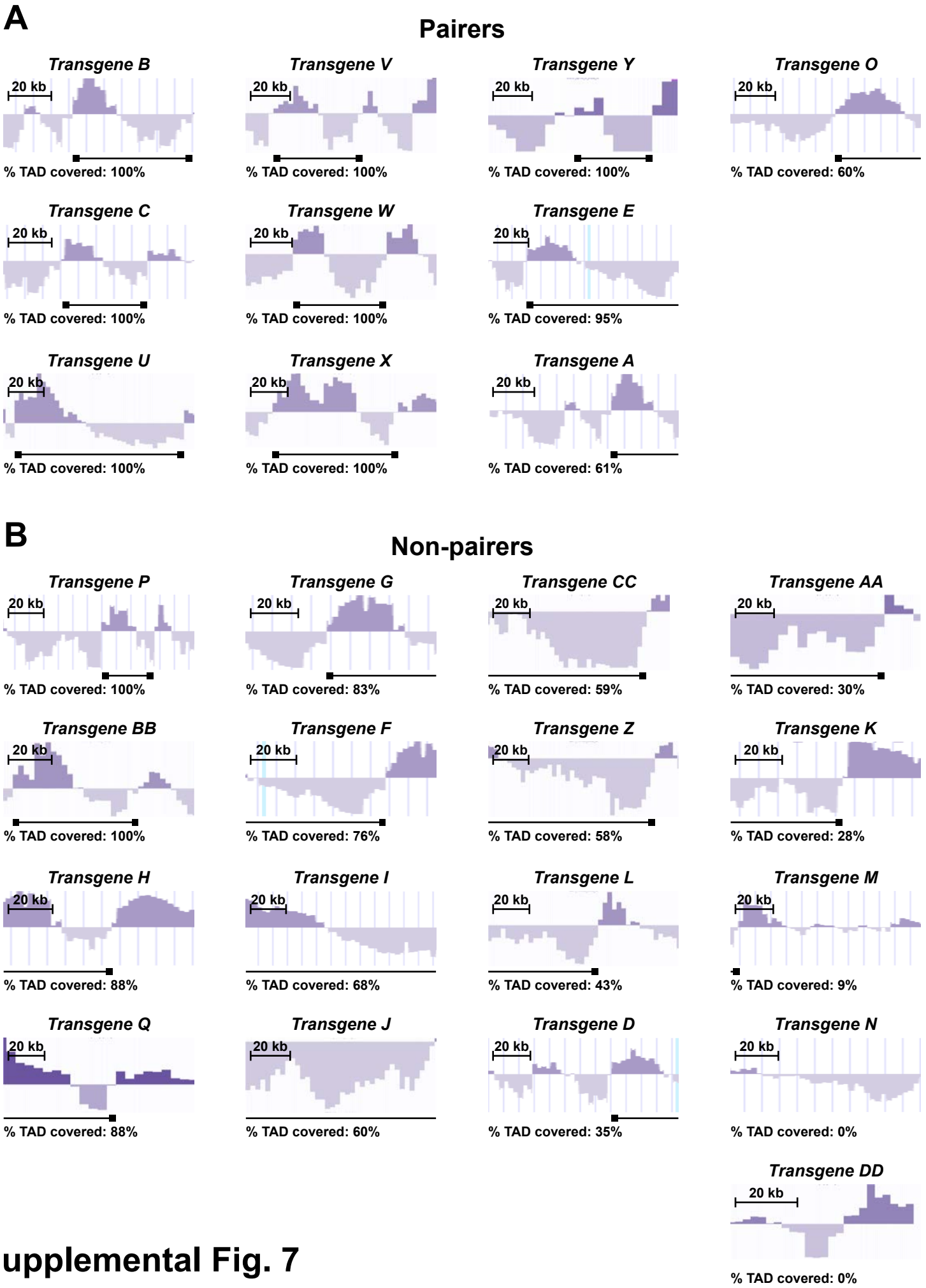
Supplemental Fig. 4



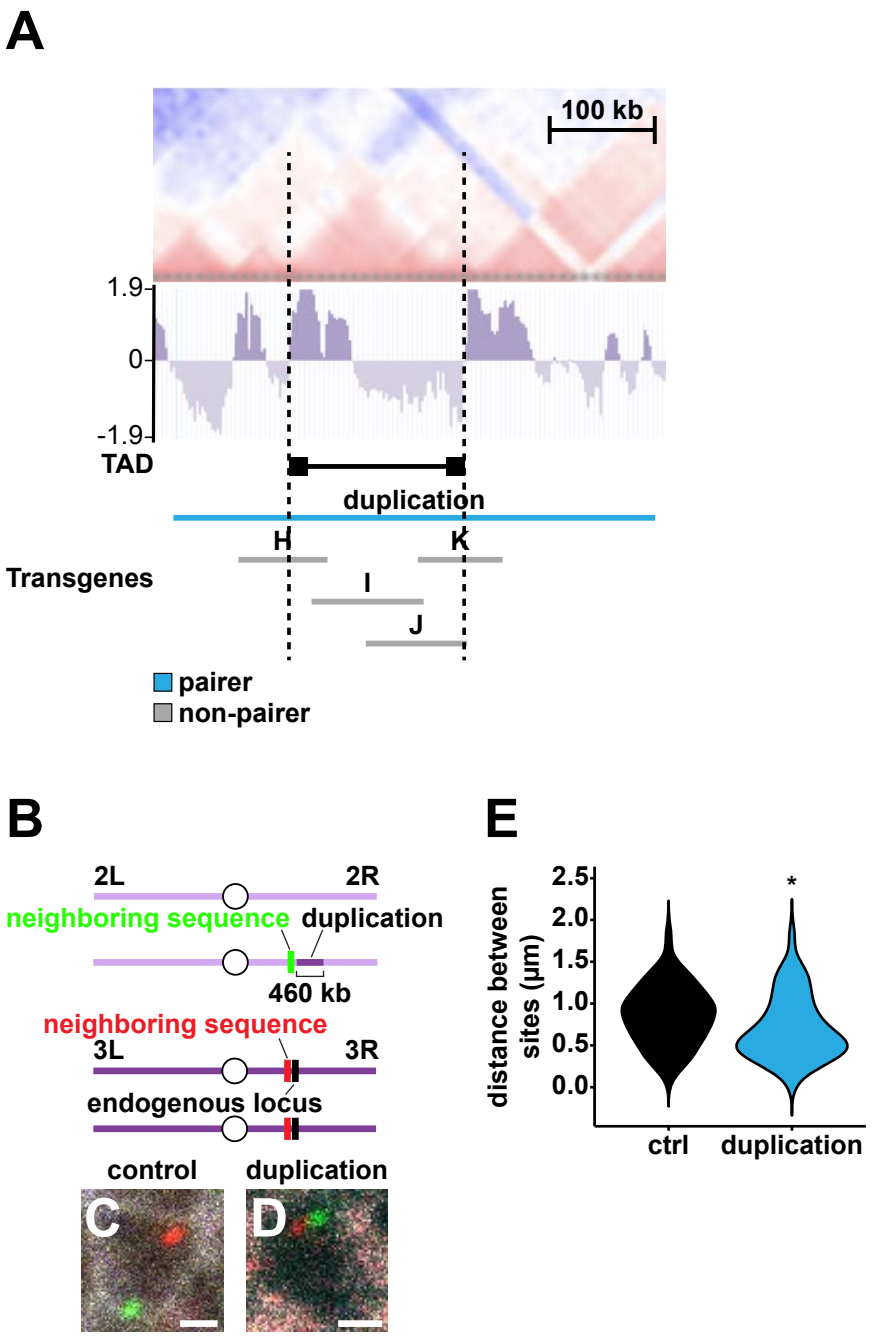
Supplemental Fig. 5



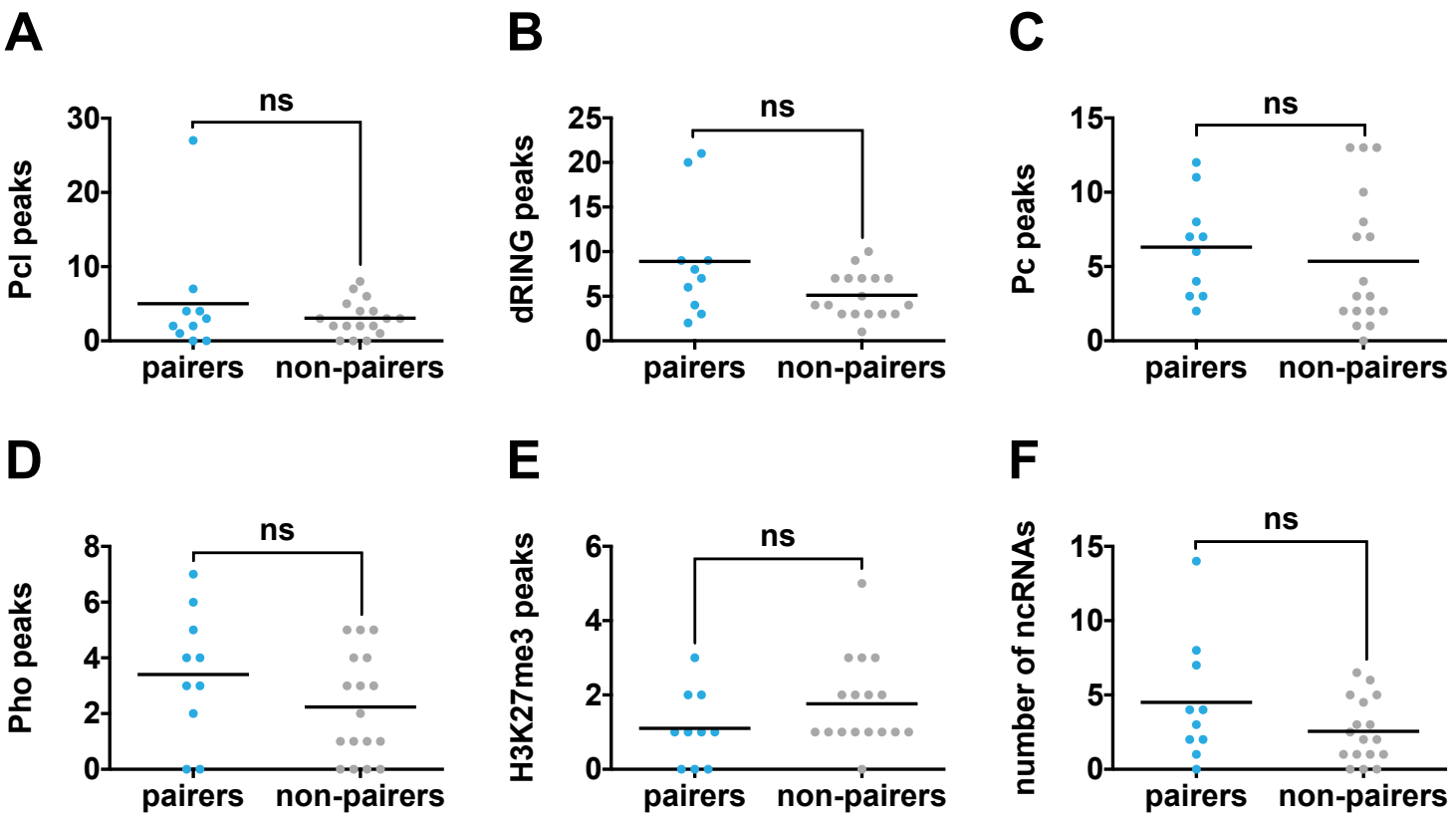
Supplemental Fig. 6



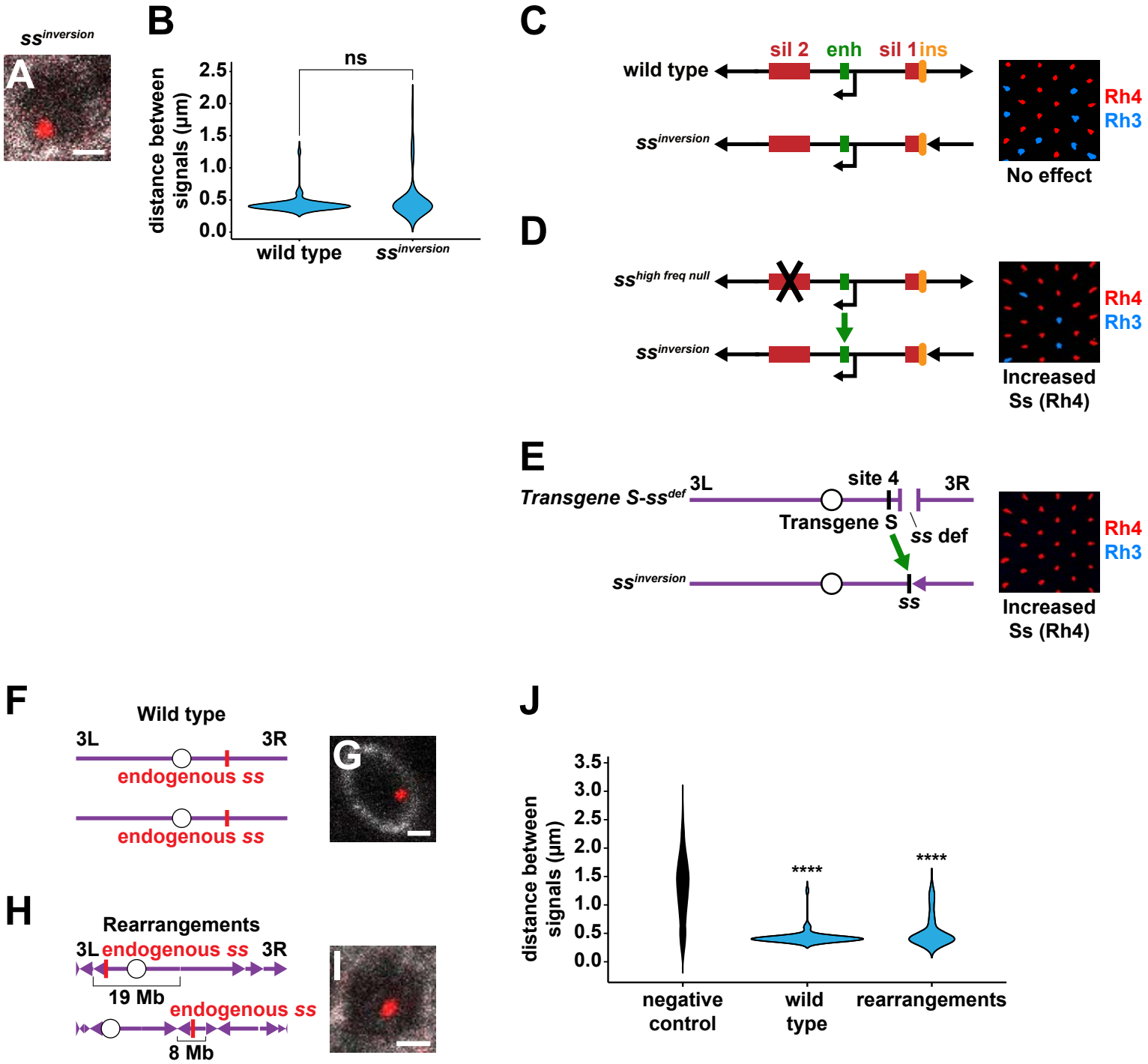
Supplemental Fig. 7



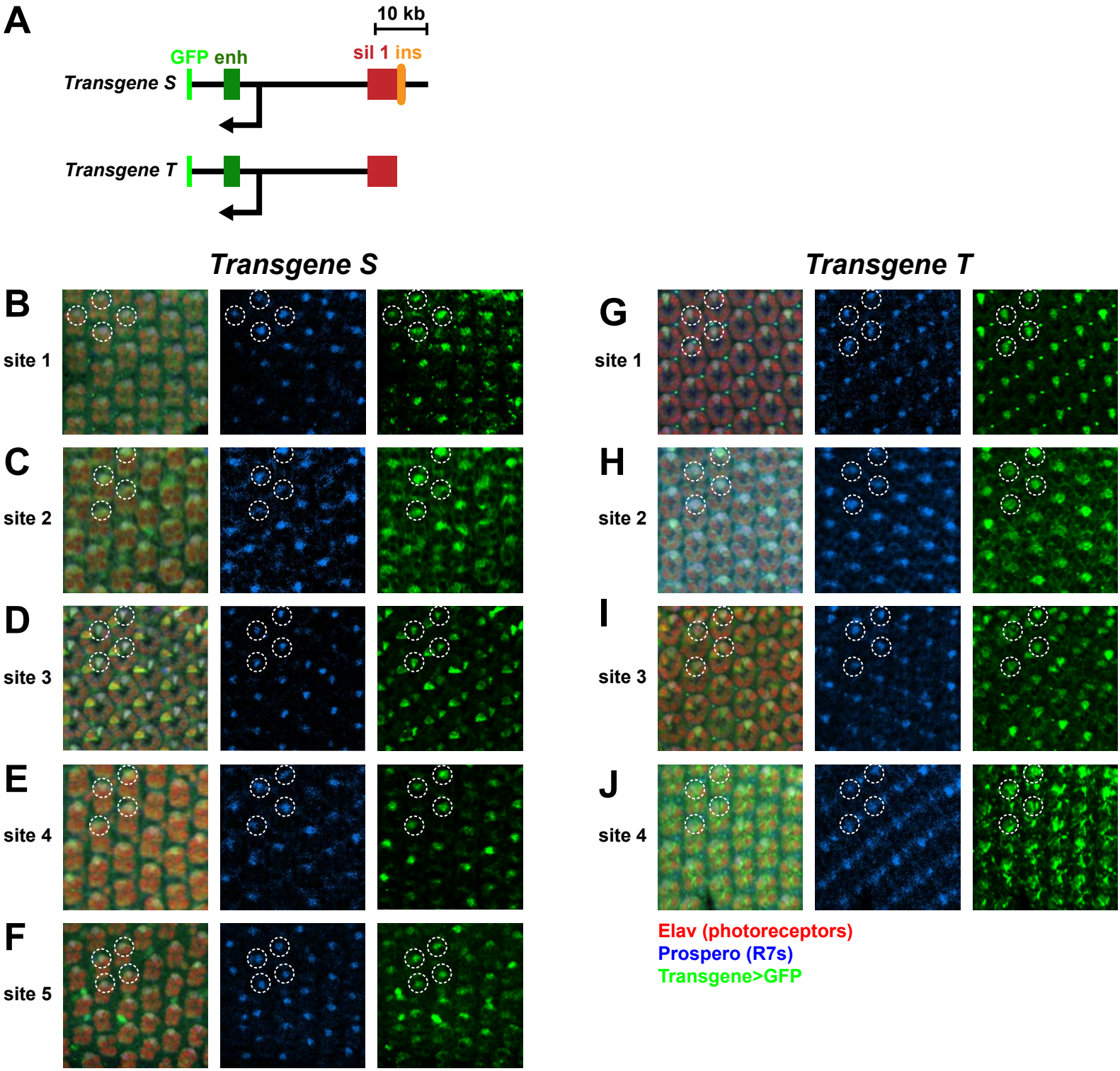
Supplemental Fig. 8



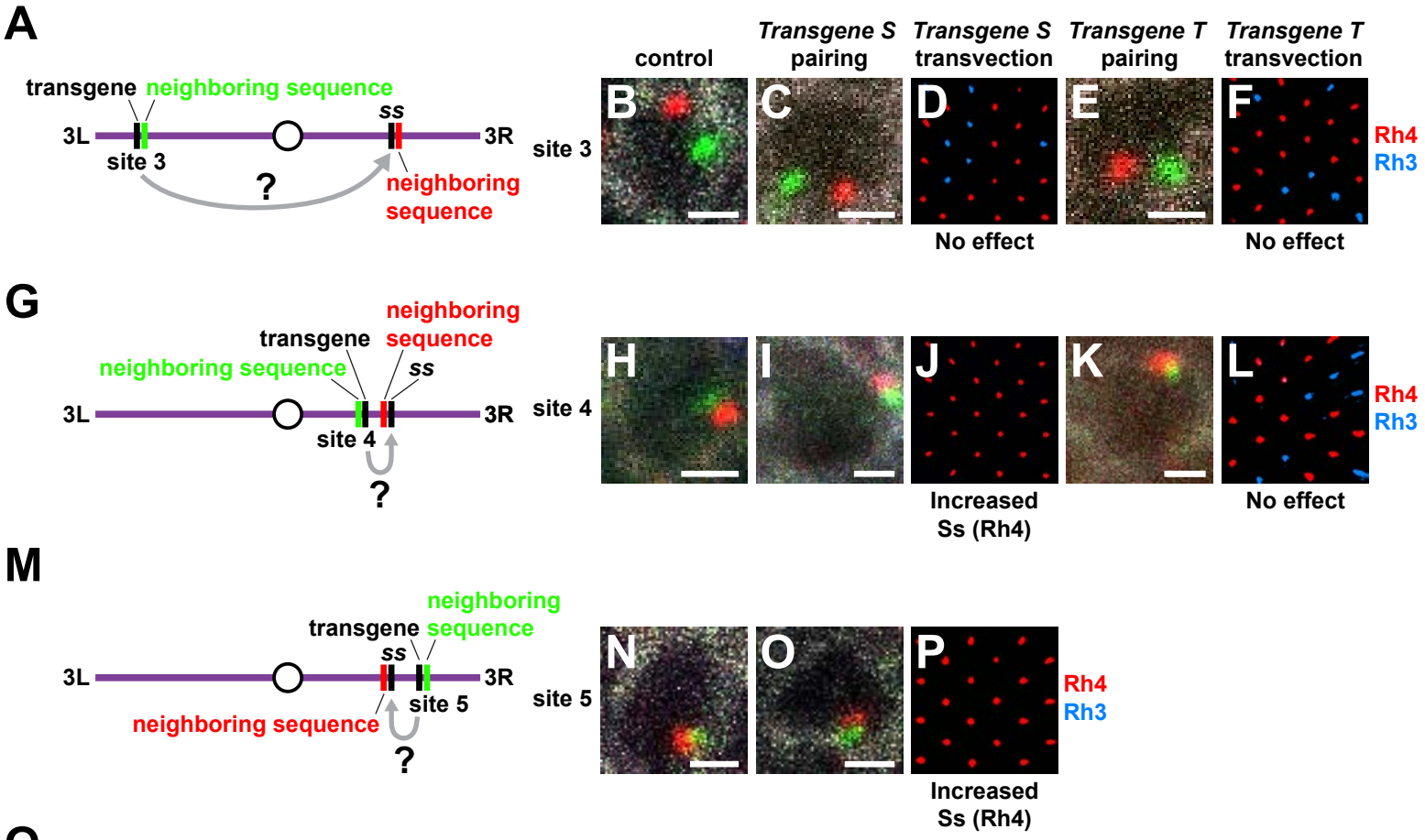
Supplemental Fig. 9

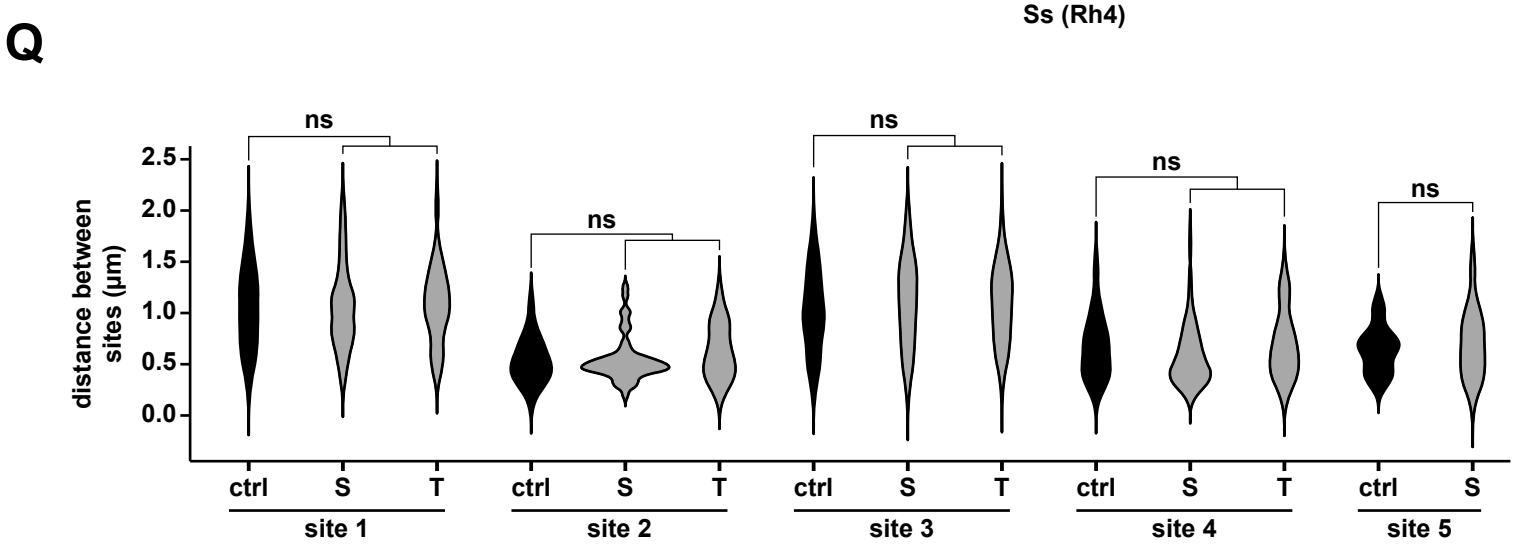


Supplemental Fig. 10

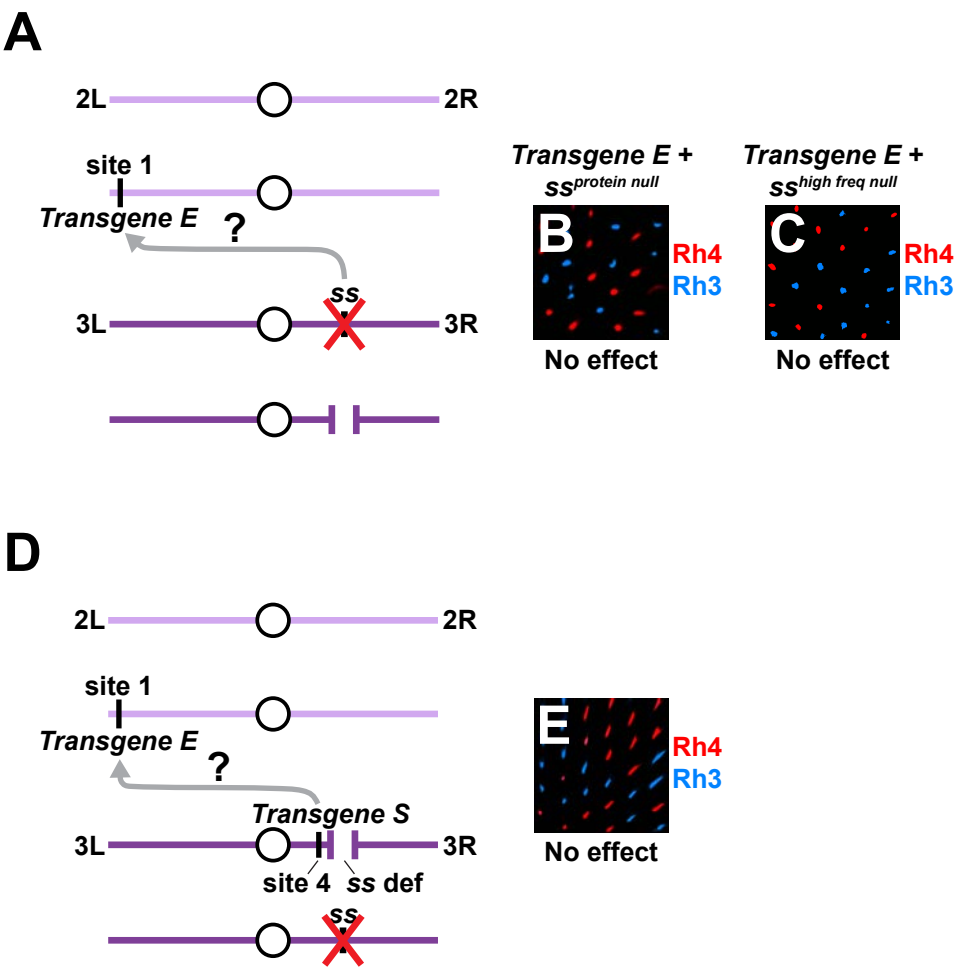


Supplemental Fig. 11

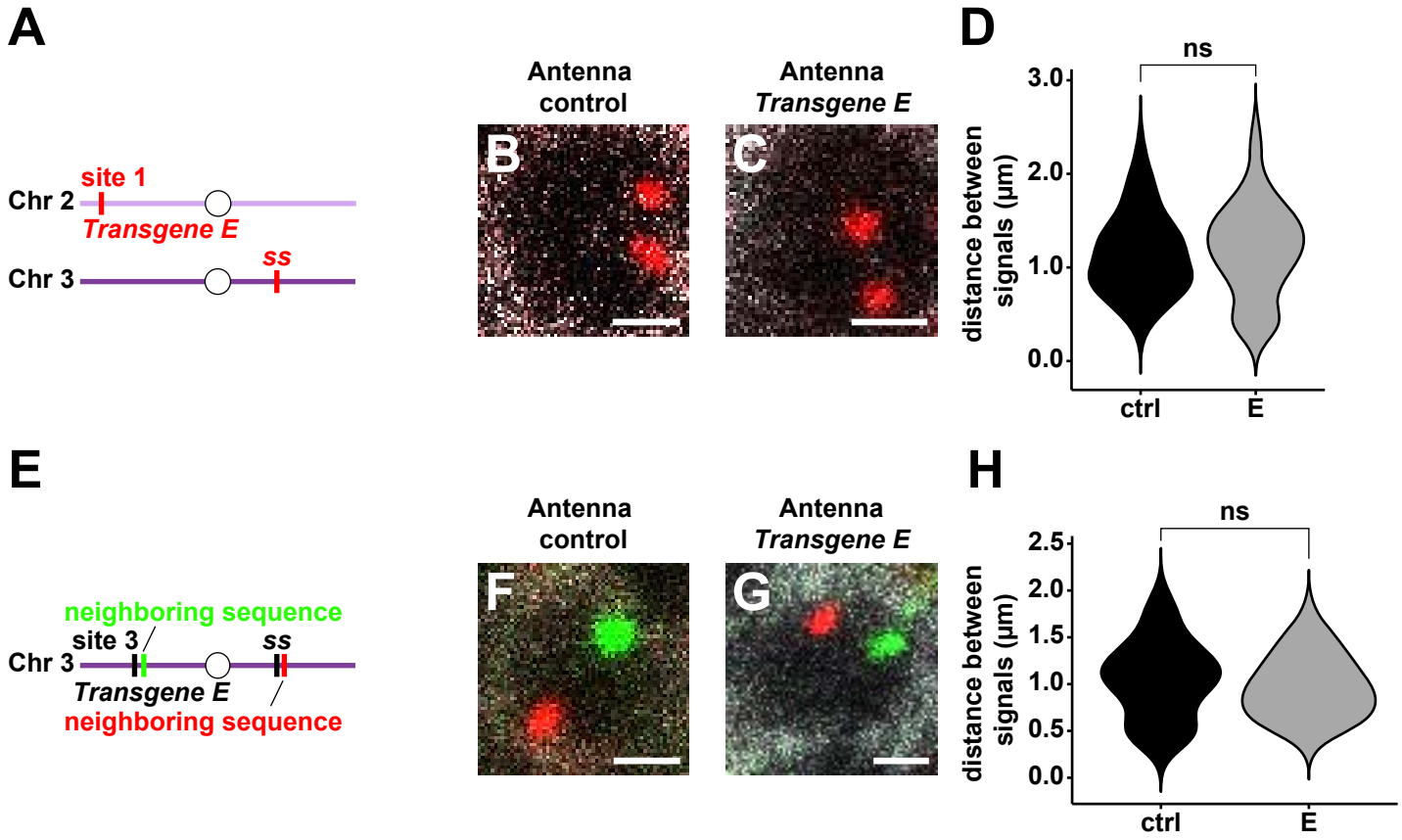




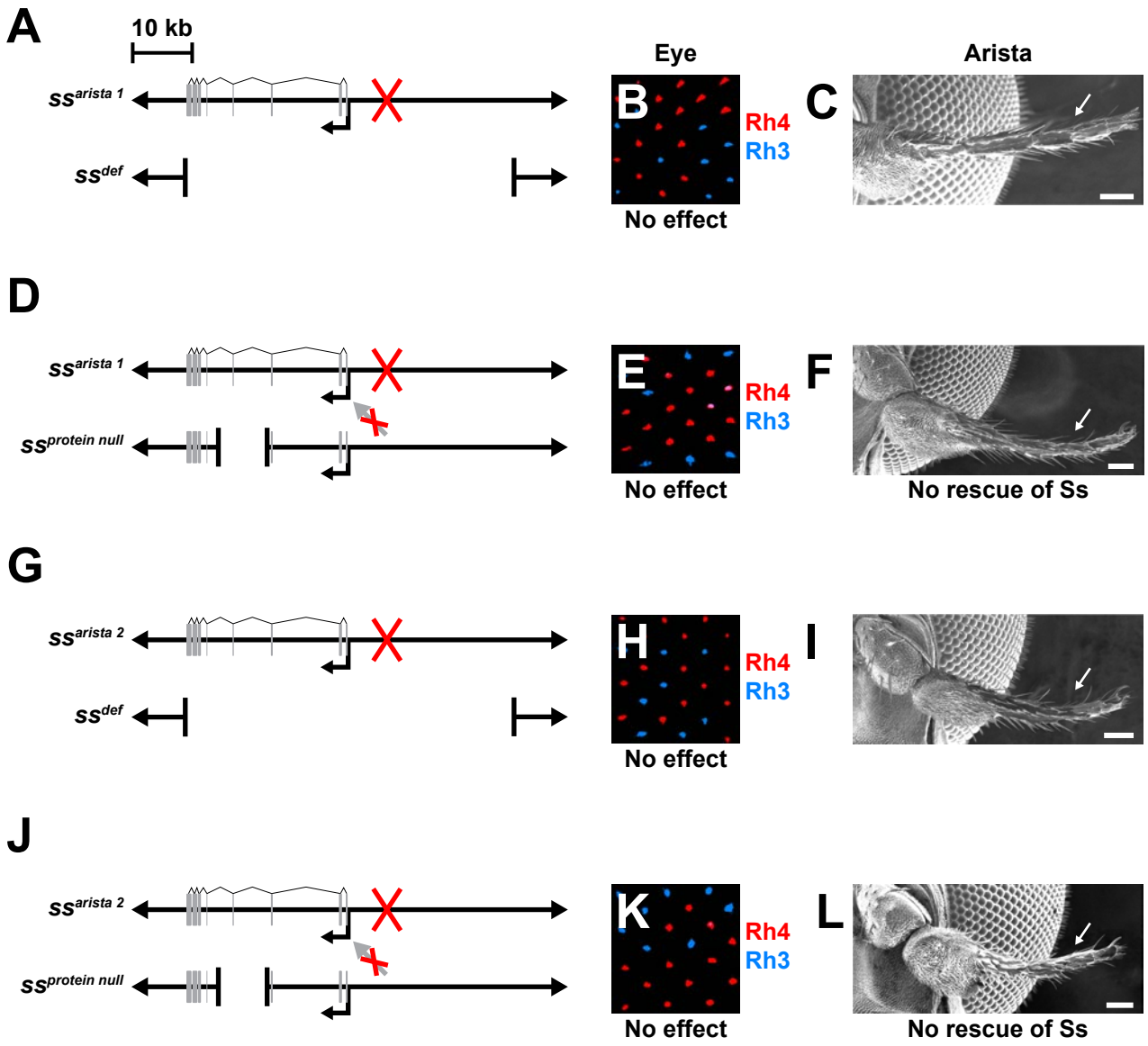
Supplemental Fig. 12



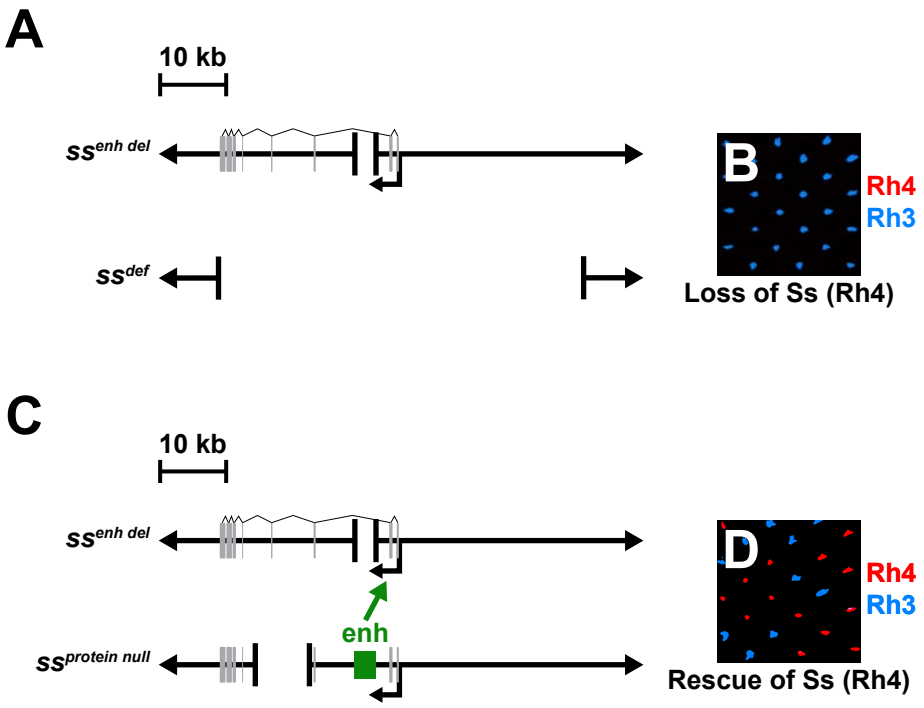
Supplemental Fig. 13



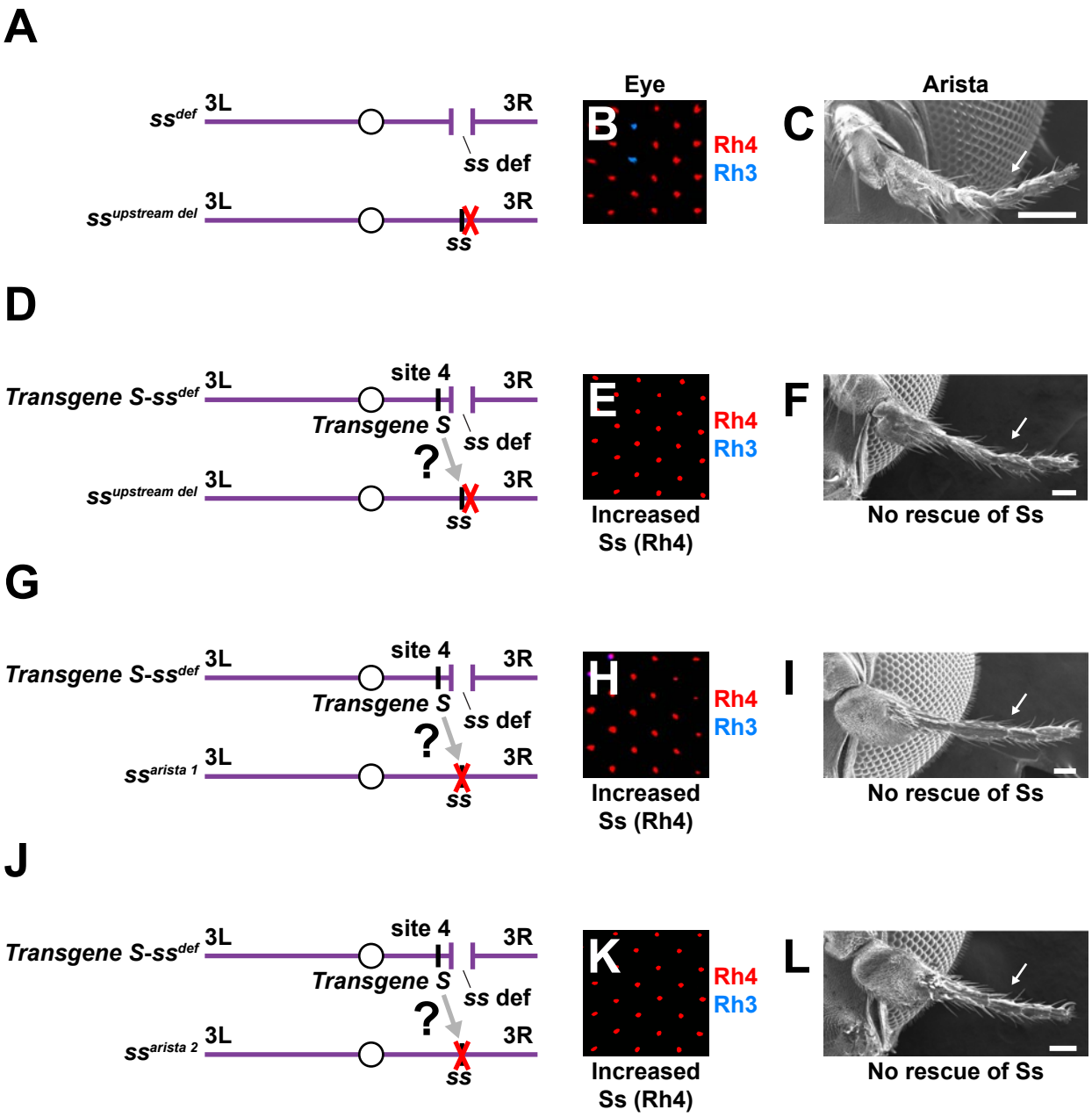
Supplemental Fig. 14



Supplemental Fig. 15

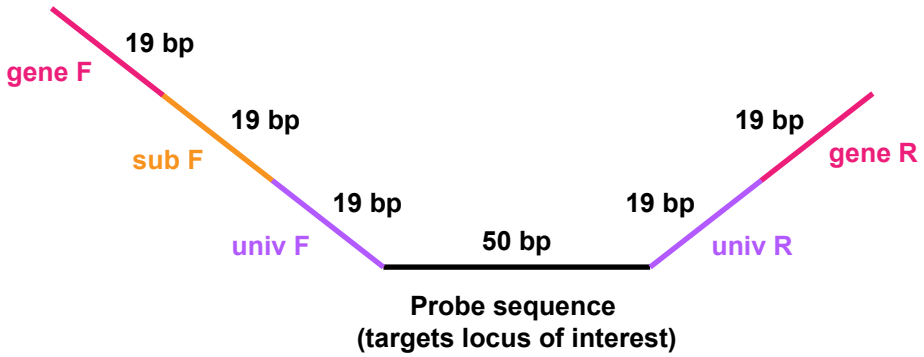


Supplemental Fig. 16

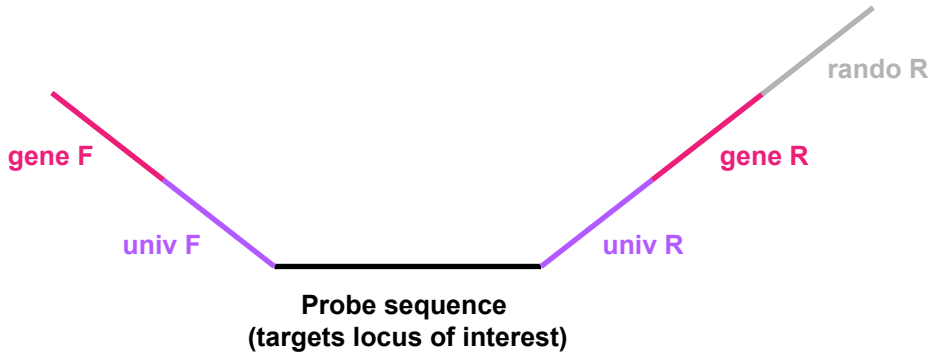


Supplemental Fig. 17

A



B



Supplemental Fig. 18