nature structural & molecular biology

Article

https://doi.org/10.1038/s41594-024-01375-7

A PRE loop at the *dac* locus acts as a topological chromatin structure that restricts and specifies enhancer–promoter communication

Received: 16 November 2023

Accepted: 22 July 2024

Published online: 16 August 2024

Check for updates

Sandrine Denaud[1], Marion Bardou[2,4], Giorgio-Lucio Papadopoulos[1,4], Stefan Grob[3], Marco Di Stefano[1], Gonzalo Sabarís[1], Marcelo Nollmann[2], Bernd Schuettengruber[1] & Giacomo Cavalli[1]

Three-dimensional (3D) genome folding has a fundamental role in the regulation of developmental genes by facilitating or constraining chromatin interactions between *cis*-regulatory elements (CREs). Polycomb response elements (PREs) are a specific kind of CRE involved in the memory of transcriptional states in *Drosophila melanogaster*. PREs act as nucleation sites for Polycomb group (PcG) proteins, which deposit the repressive histone mark H3K27me3, leading to the formation of a class of topologically associating domain (TAD) called a Polycomb domain. PREs can establish looping contacts that stabilize the gene repression of key developmental genes during development. However, the mechanism by which PRE loops fine-tune gene expression is unknown. Using clustered regularly interspaced short palindromic repeats and Cas9 genome engineering, we specifically perturbed PRE contacts or enhancer function and used complementary approaches including 4C-seq, Hi-C and Hi-M to analyze how chromatin architecture perturbation affects gene expression. Our results suggest that the PRE loop at the *dac* gene locus acts as a constitutive 3D chromatin scaffold during *Drosophila* development that forms independently of gene expression states and has a versatile function; it restricts enhancer–promoter communication and contributes to enhancer specificity.

Eukaryotic genomes are highly organized within the three-dimensional (3D) nuclear space. The development of chromatin conformation capture (3C)-based methods and advanced microscopy approaches highlighted the importance of the 3D chromatin topology of *cis*-regulatory elements (CREs) in gene regulation[1]. Regulatory interactions of CREs define, maintain and change the expression pattern of key developmental genes to ensure proper development. Loss of these control mechanisms is a frequent feature of cancer and disease[2,3].

In interphase nuclei, chromosomes are hierarchically organized into topologically associating domains (TADs)[4–6] that modulate gene regulation, although their exact function is debated. Extensive rearrangement of chromosomes and TADs in the *Drosophila* genome does not correlate with changes in gene expression[7] and removal of proteins involved in TAD border formation have no dramatic effect on gene expression[8]. On the other hand, genomic rearrangements of TADs can cause gene misexpression and diseases[9–13]. TADs facilitate local

[1]Institute of Human Genetics, UMR9002 CNRS, University of Montpellier, Montpellier, France. [2]Centre de Biologie Structurale, IUMR5048 CNRS, INSERM U1054, University of Montpellier, Montpellier, France. [3]Department of Plant and Microbial Biology and Zurich-Basel Plant Science Center, University of Zurich, Zurich, Switzerland. [4]These authors contributed equally: Marion Bardou, Giorgio-Lucio Papadopoulos. e-mail: bernd.schuttengruber@igh.cnrs.fr; giacomo.cavalli@igh.cnrs.fr

https://doi.org/10.1038/s41594-024-01375-7

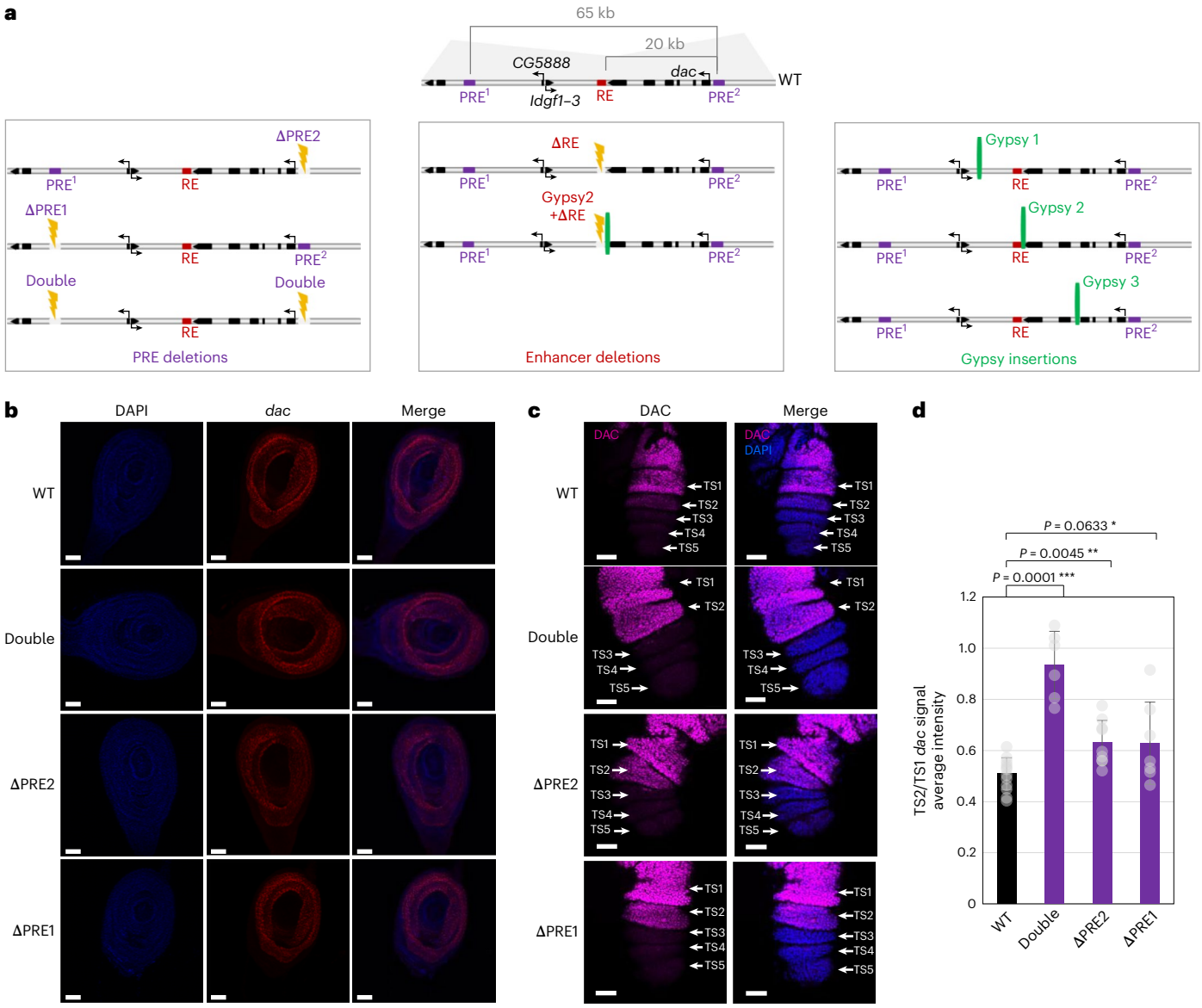


Fig. 1 | Expression of *dac* gene during *Drosophila* development upon loss of PRE function and/or PRE looping. a, Top, schematic representation of the *dac* TAD (WT). The gray shadow represents the H3K27me3 domain. Bottom, CRISPR–Cas9 mutant fly lines used in this study. Left, PRE deletion lines; middle, enhancer deletion lines; right, gypsy insertion lines. The orange lightning indicates a mutated PRE deletion line. The gypsy 1 line was described in a previous study[22]. Enhancer deletion lines and gypsy 2 and gypsy 3 lines were generated in this study. **b**, DAC immunostaining analysis of third-instar larval imaginal leg discs of WT and PRE deletion lines. White bars indicate 30 µm. **c**, DAC immunostaining analysis in early pupal imaginal leg discs (4–5 h after pupation). TS1–TS5 are indicated. White bars indicate 30 µm. At this pupal stage, the imaginal disc is composed of a single cell layer that connects in a 'tubal' shape. The images correspond to one side of the pupal leg disc imaged by confocal microscopy. **d**, Quantification of DAC immunostaining signals. The average signal ratio between TS2 and TS1 segments is plotted for WT versus mutant lines. A minimum of six pupal discs (WT, $n = 13$; double, $n = 6$; ΔPRE2, $n = 9$; ΔPRE1, $n = 7$) were scored. Error bars indicate the s.d. ***$P < 0.001$, **$P < 0.01$ and *$P < 0.1$ (two-sided unpaired t-test).

promoter–enhancer interactions and prevent inappropriate interactions between different TADs[4,9,14]. Genome organization can shape transcription dynamics by two complementary mechanisms; 'tethering elements' (TEs) within TADs foster interactions between CREs, whereas insulators and/or TAD boundaries prevent inappropriate interactions between enhancers and promoters[15].

In *Drosophila*, a specific class of TADs is characterized by the presence of the H3K27me3 Polycomb mark, which can cover several hundreds of kilobases, including key developmental regulatory genes[16,17]. Within these Polycomb domains, epigenetic regulatory sequences named Polycomb response elements (PREs) act as nucleation sites for the recruitment of the Polycomb repressive complexes 2 and 1 (PRC2 and PRC1), which are responsible for the deposition and spreading of H3K27me3 and H2AK118ub, respectively, as reviewed in the literature[18]. Many Polycomb domains contain multiple PREs and genes, which tend to be coregulated or involved in related developmental pathways[16,17]. PREs can participate in the maintenance of both active and repressed gene expression states throughout development, as reviewed in the literature[19], and a subset of PREs can engage in specific chromatin contacts within Polycomb domains, forming chromatin loops (PRE loops)[20–22]. PRC1 is likely to have an important role in PRE looping because PRC1 is critical for chromatin condensation of Polycomb domains and the establishment of their long-range interactions in mammals, as reviewed in the literature[18]. Looping interactions might be mediated by oligomerization of the sterile alpha motif domain of Polyhomeotic (PH), a subunit of PRC1, which is crucial for the condensation

of individual Polycomb domains[23] and for mediating long-range Polycomb domain interactions[24,25].

Using the Polycomb domain associated with the leg patterning gene *dachshund* (*dac*) as a paradigm, we previously showed that the loss of PRE contacts induced a very specific gain-of-function phenotype in the adult fly leg, without affecting gene expression during early development[22]. While this suggests that PRE loops have repressive functions and contribute to stabilize gene silencing during development, another PRE at the homeobox gene locus that coincides with a loop anchor of TEs is involved in gene activation at the embryonic stage[15]. This suggests that chromatin loops involving PREs can mediate both gene activation and repression but many questions concerning PRE loop formation and their functions remain unsolved. It has not been defined whether PRE loops are spatially or temporally regulated and the effects of the loss of PRE contacts on the global 3D chromatin architecture of Polycomb TADs are unknown. Furthermore, the nature, timing and underlying molecular mechanisms of gene misexpression upon loss of a PRE loop remain to be studied.

To address these questions, we exploited the *dac* Polycomb domain by generating mutant fly lines in which we specifically interfered with PRE contacts and/or enhancer function and analyzed the effects on TAD architecture and gene expression during *Drosophila* development. Intriguingly, both PREs at the *dac* locus were identified as TEs in early embryogenesis[15]. Here, we show that the PRE loop constitutes a 3D chromatin scaffold of the Polycomb domain that forms independently of gene expression states and is present at all stages during fly development. Insertion of an insulator at various positions generates local insulation within Polycomb domains without creating a new TAD boundary. Intriguingly, the insulator blocks PRE looping, whereas an essential leg enhancer can bypass the insulator to activate the *dac* gene. Loss of PRE looping does not lead to widespread gene activation but correlates with the activation of the *dac* gene at a precise developmental stage and tissue. Importantly, the modulation of physical proximity between the leg enhancer and promoter is not a major determinant for gene activation. Lastly, we show that reduced PRE looping results in a loss of enhancer specificity, leading to ectopic activation of a neighboring gene. We propose that PRE loops form a topological scaffold structure within Polycomb domains that have a dual function; they restrict promoter–enhancer communication, while they also contribute to enhancer–promoter specificity.

Results

dac expression during development upon loss of PRE looping

To analyze the importance of PRE looping for the 3D organization of Polycomb domains and gene expression, we created three classes of clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 mutant flies associated with the *dac* gene locus (Fig. 1a). The first class comprised PRE deletion lines that affect both Polycomb group (PcG) recruitment and PRE looping[22]. The second category corresponded to the deletion of an enhancer driving *dac* expression in the leg. The third class consisted of inserting a gypsy insulator sequence between two PREs at various positions upstream or downstream of the leg enhancer.

We previously showed that the disruption of PRE looping does not change the embryonic expression pattern of the *dac* gene. However, a very specific gain-of-function phenotype was observed in the adult fly: the appearance of extra sex comb (ESC) bristles on the second tarsal segment (TS2) of male flies on the first legs[22]. To determine the nature and timing of *dac* misexpression inducing this phenotype, we analyzed *dac* expression patterns during leg development in PRE deletion fly lines by performing immunostaining of third-instar larval imaginal leg discs and early pupal imaginal leg discs, when leg segmentation takes place. In larval leg discs, *dac* is expressed in a ring-like shape corresponding to the medial leg structures (trochanter, femur, tibia and TS1) and no notable changes in *dac* expression pattern can be observed at the larval stage upon mutation of PRE sequences (Fig. 1b).

At the early pupal stage, *dac* is normally expressed in the proximal leg part and TS1, whereas expression in TS2 is low and it is completely absent in the more distal segments (TS3–TS5) (Fig. 1c). Importantly, the expression of *dac* is significantly increased, specifically in TS2 of the developing leg in PRE mutant fly lines, whereas *dac* remains repressed in TS3–TS5 (Fig. 1c,d). No significant difference in global mRNA levels was observed by reverse transcription (RT)–qPCR analysis in pupal and larval imaginal leg discs (Extended Data Fig. 1a). This shows that PRE deletion leads to derepression of *dac* specifically in cells of the TS2 segment of pupal leg discs. Increased *dac* expression at this precise developmental stage and tissue is likely involved in transforming TS2 into TS1 identity, consistent with the induction of ectopic sex combs (normally found exclusively on the more proximal segment TS1) on the more distal segment TS2 (Extended Data Fig. 1b).

The PRE loop acts as a constitutive 3D chromatin scaffold

We hypothesized that changes in DAC expression, leading to the transformation of leg segment identity, may involve alterations in 3D chromatin architecture in mutant flies. To test this, we performed Hi-C experiments of third-instar larval imaginal leg discs and early pupal leg discs and compared them to Hi-C data in embryos (Fig. 2a). Interestingly, *dac* intra-TAD interactions decrease during development (Fig. 2b). This might reflect global decondensation of the domain correlating with an increase in *dac*-expressing cells during the larval and pupal stages (approximately 20% *dac*-expressing cells in late embryos versus 50% *dac*-expressing cells in leg discs). Importantly, the local contact enrichment of the PRE loop is present at similar levels from embryos to pupae (Fig. 2c). Furthermore, no additional chromatin loop or major change in intra-TAD structure or TAD borders was observed during these stages of development. This indicates that the PRE loop is a prominent chromatin feature of the *dac* TAD that persists at all stages of *Drosophila* development. Because the proportion of *dac*-expressing cells greatly increases during development, we hypothesized that the PRE loop within the *dac* TAD may not be the key element determining cell-type-specific expression.

To test this model, we determined the conformation of the *dac* TAD in the different leg segments expressing or not expressing *dac*. For this, we turned to Hi-M, a multiplexed DNA fluorescence in situ hybridization (FISH) technology that captures chromatin conformations in single cells[26]. We designed and amplified an oligopaint library tiling most of the *dac* TAD with different barcodes, reaching a mean resolution of 4 kb (Supplementary Table 1). This library was hybridized and each barcode was sequentially imaged in early pupal leg discs (Methods). Ensemble Hi-M maps obtained from pooled nuclei comprising TS1–TS4 confirmed the presence of a prominent long-range chromatin interaction between barcodes including and adjacent to the two PRE sequences (Fig. 2d). Hi-M maps obtained from different replicates were highly correlated (Extended Data Fig. 2b). As in Hi-C maps, the PRE loop includes additional neighboring chromatin regions, indicating that it involves an extensive set of chromatin interactions in the whole regions surrounding the PREs. Next, we analyzed the chromatin organization of the *dac* TAD in the different TSs, where the *dac* gene is highly expressed (TS1), weakly expressed (TS2) or repressed (TS3 and TS4). Interestingly, PRE interactions can be observed with similar frequency in all these segments (Fig. 2e). Taken together, these results show that the *dac* PRE loop forms with similar frequencies at different stages of development and within leg segments displaying different levels of *dac* expression. Thus, we conclude that the regulation of *dac* expression does not require changes in PRE looping.

A ring enhancer (RE) is necessary and sufficient for *dac* expression

Because the loss of PRE function induces activation of the *dac* gene exclusively during leg development, we searched for putative regulatory regions driving expression of the *dac* gene specifically in the leg.

https://doi.org/10.1038/s41594-024-01375-7

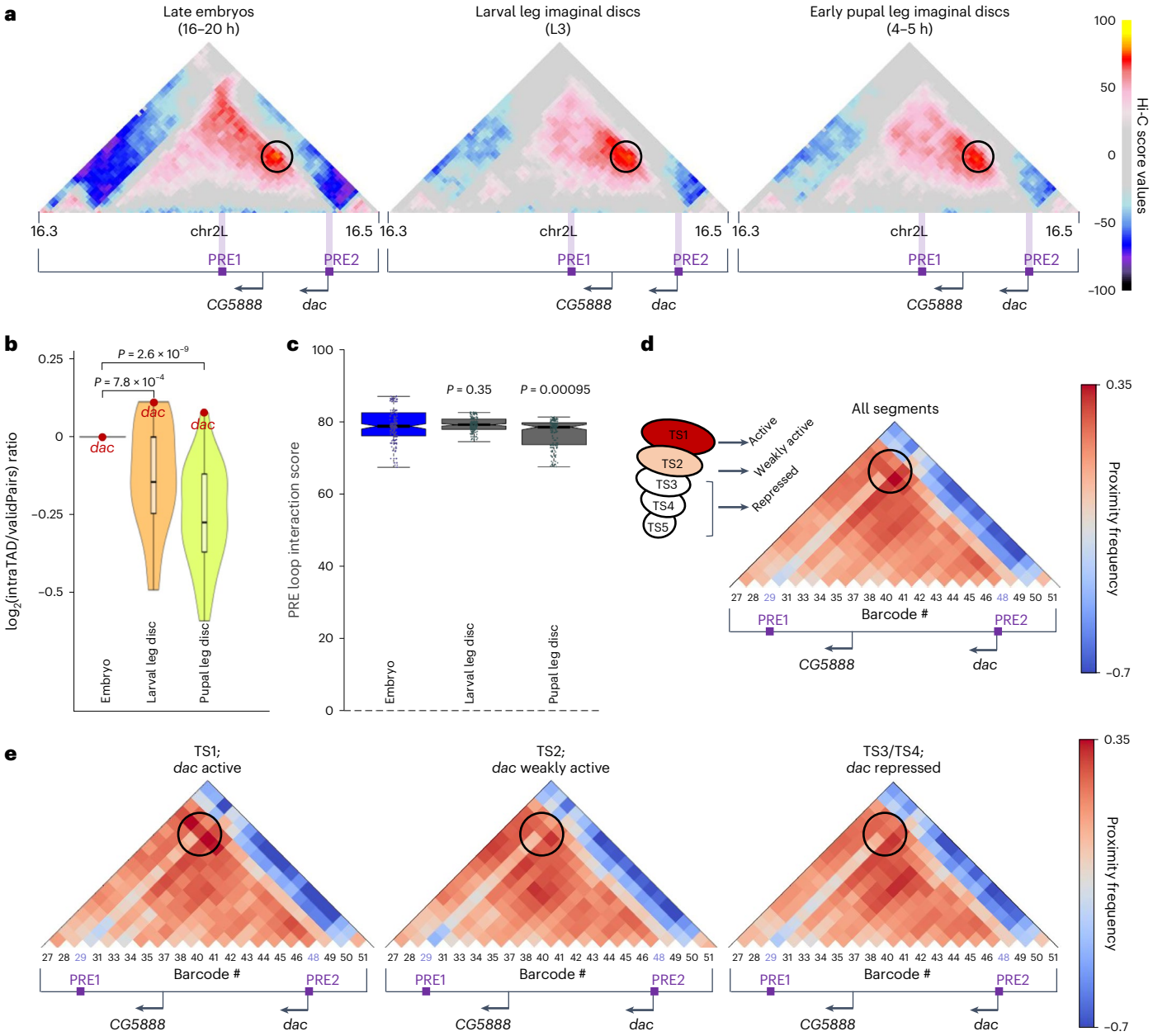


Fig. 2 | PRE loop during *Drosophila* development and *dac*-expressing or *dac*-repressed cells. a, Hi-C score maps (Methods) of a 200-kb region at 3-kb resolution on chr2L including the *dac* gene locus in whole late embryos (left), third-instar leg imaginal discs (middle) and early pupal stage (4–5 h after pupation; right). The black circle indicates the position of the *dac* PRE loop. Violet bars indicate the position of PREs. Black arrows indicate gene promoters of the *dac* and the CG5888 genes. **b**, Distributions of the $\log_2$ ratios of the frequencies of observed Hi-C contacts within Polycomb-associated TADs in chr2L in embryos over the larval and pupal leg discs ($n = 25$ values in each distribution). The contact frequency for each condition is computed against the total number of valid pairs of the corresponding condition. The ratio is computed as the contact frequency in embryos over the equivalent quantity in larvae or pupae. Hence, a positive value of the $\log_2$ ratio for the *dac* domain indicates a decrease in contact frequency during the analyzed developmental stages. Box plots show median (central line), the 25th and 75th percentiles (box limits) and 1.5 × IQR (whiskers). Outliers are not shown. The unpaired two-sided Wilcoxon statistical test was used to estimate the reported *P* values. **c**, Quantification of Hi-C scores the *dac* PRE loop in embryos ($n = 195$), larval leg discs ($n = 172$) and early pupal leg discs ($n = 183$). Reported *P* values result from comparing the embryos with the larval and the pupal leg disc distributions, respectively. The number of points per distribution is reported (Methods). Box plots show the median (central line), the 75th and 25th percentiles (box limits) and 1.5 × IQR (whiskers). The unpaired two-sided Wilcoxon statistical test was used to estimate the reported *P* values. **d**, Ensemble Hi-M proximity matrices generated from TS1–TS5 for early pupal leg discs (4–5 h after pupation). On the scale bar, red and blue represent, respectively, high and low proximity frequencies. The proximity frequency was normalized by the genomic distance. The black circle indicates the position of the PRE1–PRE2 loop within the *dac* TAD. Barcodes 29 (PRE1) and 48 (PRE2) are highlighted in purple. **e**, Ensemble Hi-M proximity maps in the early pupal stage: TS1 (*dac* active), TS2 (*dac* weakly active) and TS3 and TS4 (*dac* repressed). The proximity frequency was normalized by the genomic distance. Matrices were generated from 6,535 traces for TS1, 7,363 traces for TS2 and 9,993 traces for TS3 and TS4. The black circle indicates the position of the PRE1–PRE2 barcodes. Barcodes 29 (PRE1) and 48 (PRE2) are highlighted in purple.

https://doi.org/10.1038/s41594-024-01375-7

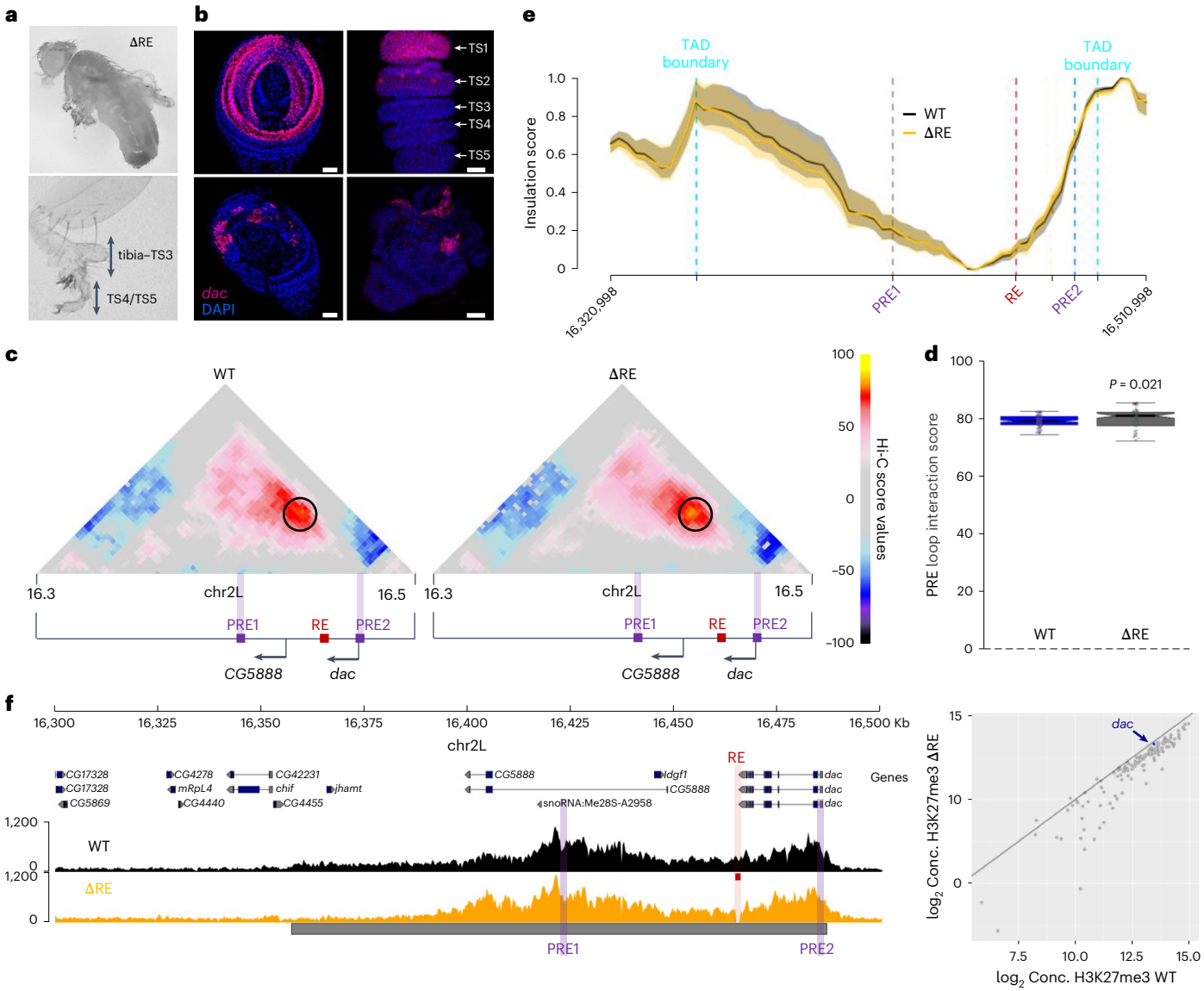


Fig. 3 | Phenotypic, transcriptional and 3D chromatin conformational consequences of the RE deletion. a, Representative picture of homozygous adult flies and legs homozygous mutant for the RE. The tibia and TSs are fused. **b**, DAC immunostaining analysis in WT and ΔRE larval third-instar imaginal leg disc (left) or early pupal imaginal leg discs (4–5 h after pupation) (right). White bars indicate 30 µm. **c**, Hi-C score (Methods) maps of a 200-kb region at 3-kb resolution on chr2L at *dac* gene locus in third-instar imaginal leg disc in WT or ΔRE flies. A black circle indicates the position of the *dac* PRE loop. Violet bars indicate the position of PREs. Black arrows indicate gene promoters of the *dac* and *CG5888* genes. The red square indicates the RE. **d**, Quantification of the *dac* PRE loop interaction scores. Hi-C interaction score in WT larval leg discs ($n = 172$) and ΔRE discs ($n = 73$). Reported *P* values result from comparing the WT and ΔRE distributions. The number of points per distribution is reported (Methods). Box plots show the median (central line), the 25th and 75th percentiles (box limits) and 1.5 × IQR (whiskers). The unpaired two-sided Wilcoxon statistical test was used to estimate the reported *P* values. **e**, Insulation profile shown at 3-kb resolution along the *dac* 200-kb region in larval WT and ΔRE leg discs is shown as the mean value (line) ± the s.d. (shaded area) over the ISs computed using five different values of the window parameter (100, 150, 200, 250 and 300 kb; Methods). **f**, CUT&RUN profile for H3K27me3 mark at the *dac* domain performed in third-instar imaginal leg disc in WT or ΔRE flies. The red bar indicates the position of the RE. Violet bars indicate the positions of PREs. The gray bar below H3K27me3 tracks demarcates the PcG TAD[22]. The scatter plot on the right shows the H3K27me3 enrichment in the 131 *Drosophila* Polycomb domains in ΔRE as a function of WT flies ($\log_2$ scale). The *dac* Polycomb domain is highlighted in blue.

Interestingly, a well-conserved 567-bp sequence was previously shown to recapitulate *dac* expression in a ring-like shape in larval leg discs in transgenic reporter gene assays[27]. Therefore, this regulatory region 20 kb downstream of the *dac* promoter (Fig. 1a) was called an RE.

Knowing that the RE is sufficient to recapitulate *dac* expression in reporter assays, we asked whether the RE is essential for regulating the correct *dac* expression pattern at the endogenous *dac* TAD. Therefore, we created a mutant fly line carrying a 1-kb deletion encompassing the RE sequence (ΔRE) (Fig. 1a). Heterozygous flies for the RE deletion were viable and did not show any morphological phenotype. In contrast, homozygous flies displayed short and deformed legs with fused segments, characteristic of *dac* loss of function (Fig. 3a). Strong loss of *dac* expression was observed by immunostaining experiments in imaginal leg discs and early pupal discs (Fig. 3b). Although *dac* expression was absent in the large majority of cells, a few random spots of cells expressing *dac* could be observed and morphological perturbations of the imaginal disc appeared at the early pupal stage during metamorphosis (Fig. 3b).

Together, these results show that the RE is the major CRE driving the expression of the *dac* gene in larval and early pupal leg discs and is both necessary and sufficient for *dac* activation in the leg disc.

https://doi.org/10.1038/s41594-024-01375-7

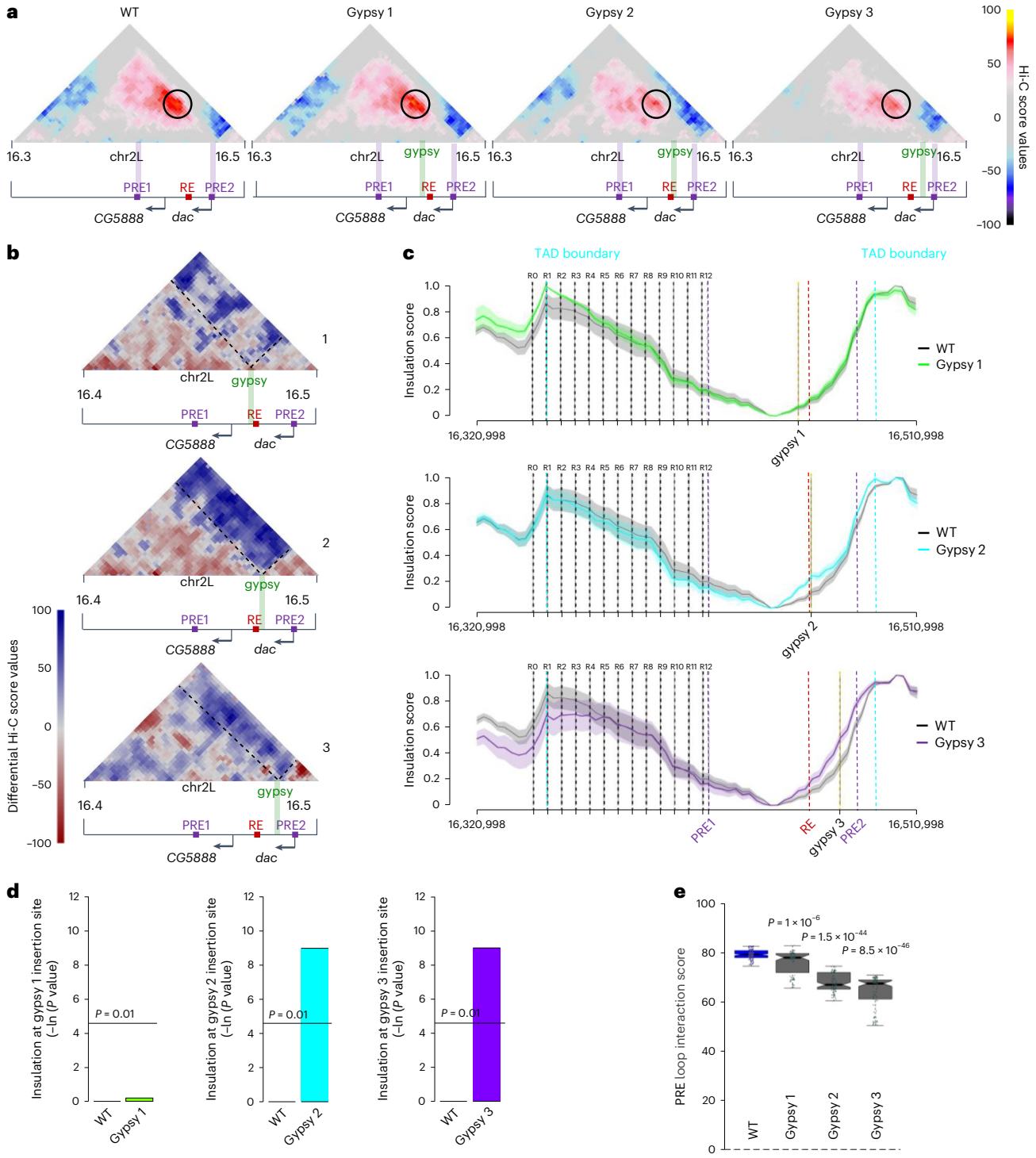


Fig. 4 | Consequences of gypsy insertions for the *dac* Polycomb domain architecture analyzed by Hi-C. a, Hi-C score (Methods) maps of a 200-kb region at 3-kb resolution on chr2L including the *dac* gene locus in third-instar imaginal leg disc carrying the indicated gypsy insertions. The black circle indicates the position of the *dac* PRE loop. Violet bars indicate the position of PREs. Black arrows indicate the gene promoters of the *dac* and *CG5888* genes. The red square indicates the RE. The green bar indicates the position of the gypsy insulator. **b**, Differential Hi-C scores maps (WT versus gypsy 1 (top), gypsy 2 (middle) or gypsy 3 (bottom) mutants) of a 150-kb region of the *dac* gene in third-instar imaginal leg disc (Methods). Black dashed lines within differential Hi-C maps indicate the position of local insulation. Black arrows indicate the gene promoters of the *dac* and *CG5888* genes. The red square indicates the RE. The green bar indicates the position of the gypsy insulator. **c**, Insulation profile shown at 3-kb resolution along the *dac* 200-kb region in larval WT and gypsy 1, gypsy 2 and gypsy 3 mutant leg discs is shown as the mean value (line) ± the s.d. (shaded area) over the ISs computed using five different values of the window parameter (100, 150, 200, 250 and 300 kb; Methods). Gypsy insulator insertions (gypsy 1, gypsy 2 and gypsy 3) are indicated. **d**, *P* values from the comparisons of ISs at gypsy 1, gypsy 2 and gypsy 3 insertion sites between the WT and the corresponding fly line (Methods). The *P* values resulted from a two-sided Welch *t*-test between the WT condition and each of the gypsy mutant lines at the corresponding locus. **e**, Quantification of the *dac* PRE loop Hi-C interaction score in WT ($n = 172$), gypsy 1 ($n = 106$), gypsy 2 ($n = 106$) and gypsy 3 ($n = 111$) mutant flies. Reported *P* values result from comparing the WT and gypsy 1, gypsy 2 and gypsy 3 distributions. The number of points per box plot is reported (Methods). Box plots show the median (central line), the 25th and 75th percentiles (box limits) and 1.5 × IQR (whiskers). The unpaired two-sided Wilcoxon statistical test was used to estimate the reported *P* values.

Loss of *dac* expression and PRE looping

RE deletion with a concomitant loss of *dac* expression affects the formation and topology of the repressive H3K27me3 domain (PRE looping in particular).

Therefore, we first performed Hi-C experiments in the wild-type (WT) or ΔRE larval imaginal leg discs to determine the consequences of RE deletion on the PRE loop and global TAD architecture (Fig. 3c). Analysis of the Hi-C data and quantification of the interaction frequency of the PRE loop revealed that overall TAD structure, boundaries and PRE looping were not significantly affected upon deletion of the RE (Fig. 3d,e).

Next, we performed cleavage under targets and release using nuclease (CUT&RUN) experiments using H3K27me3 antibodies in WT or ΔRE leg imaginal discs (Fig. 3f). We did not observe any major changes in the distribution of the repressive H3K27me3 mark across the *dac* locus and quantitative analysis of H3K27me3 levels confirmed that the deposition of H3K27me3 across the TAD was not notably changed upon deletion of the RE. This indicates that the RE does not have a role in PRE function and the deposition of H3K27me3. Overall, these experiments show that the PRE loop is formed independently of *dac* expression states and the RE has no effect on PRE function and global TAD architecture.

Gypsy insertion induces insulation and reduces PRE looping

The identification and characterization of the RE as an essential CRE driving *dac* expression in the leg strongly suggest that derepression of *dac* in TS2 upon loss of PRE function is mediated by the RE. To disentangle the functional relationship of PRE looping, RE–promoter communication and *dac* expression, we generated a panel of mutant fly lines, where we inserted a gypsy insulator between the RE and the *dac* promoter and/or between the two PREs (Fig. 1a). The gypsy insulator element has the ability to reduce enhancer–promoter or PRE–PRE interactions when placed between these elements[28,29]. In a previously characterized fly line (gypsy 1), the insulator was inserted upstream of the RE and was not located between the RE and *dac* promoter (gypsy 1 line, Fig. 1a). This insulator is not expected to interfere with the interaction of the RE and the *dac* promoter. We, therefore, created two lines where we inserted the gypsy insulator at different positions between the RE and the *dac* transcription start site (TSS) (gypsy 2 and gypsy 3 lines, Fig. 1a). We hypothesized that, in these lines, the gypsy element should interfere with both PRE looping and RE–promoter communication.

Quantitative chromatin immunoprecipitation (qChIP) experiments confirmed that all three gypsy insertions generated an ectopic Su(Hw)-binding site, indicating that the insulator sequence is functional (gypsy 2 and gypsy 3 lines, Extended Data Fig. 3a; Su(Hw) binding to gypsy 1 was analyzed in a previous study[22]). Additionally, CUT&RUN experiments in larval imaginal leg discs showed that PRE-mediated deposition of H3K27me3 was not notably changed upon insertion of the gypsy insulator in all three gypsy lines, confirming that the recruitment of PcG complexes to PREs is not affected (Extended Data Fig. 3b,c).

To analyze the physical insulation activity of the gypsy element and its impact on PRE looping, we performed Hi-C experiments in larval imaginal leg discs of the three gypsy insertion lines (Fig. 4). Visual inspection of Hi-C contact maps revealed that no new TAD borders were formed around the gypsy insertion sites (Fig. 4a). A moderate insulation activity of the gypsy sequences could be visualized by plotting the maps of differential score enrichments of interactions in the WT line versus each of the gypsy mutant lines (Fig. 4b). Gypsy 1 insertion had the weakest effect on local insulation, whereas, in comparison to gypsy 2 and gypsy 3 lines, the WT had stronger contact enrichment in the region bypassing the gypsy insertion sites, indicating an insulation activity of gypsy on the surrounding chromatin. Accordingly, quantification of the insulation scores (ISs) at the gypsy insertion sites showed a significant increase in insulation at the gypsy 2 and gypsy 3 insertion sites, whereas no significant differences between the insulation profiles of the different conditions were observed in the region upstream of the PRE1 site (Extended Data Fig. 4a,b). Although gypsy 1 insertion showed no significant increase in local insulation in the Hi-C approach (Fig. 4c,d), we observed an insulation of the surrounding chromatin at the gypsy 1 insertion site insulation activities using 4C-seq experiments with viewpoints upstream or downstream of the gypsy insertion (Extended Data Fig. 5a,b). This suggests that, albeit with different strength, all three gypsy elements have insulation activity.

Next, we analyzed how the physical insulation activity in the different gypsy lines affects the PRE loop. Quantification of the looping interactions between the two PRE regions revealed a significant reduction in PRE contacts upon gypsy insertions in each of the three mutant lines (Fig. 4e). Quantitative differences were still observed between the gypsy lines, with gypsy 1 insertion weakly reducing PRE contacts, whereas gypsy 2 and gypsy 3 insertions had stronger effects on PRE looping. Reduced PRE looping of the gypsy 1 line was further confirmed by 4C-seq experiments using the *dac* promoter as a viewpoint (Extended Data Fig. 5b).

Lastly, we analyzed changes in gypsy insulation activity and its effects on PRE looping during *Drosophila* development. For this, we exclusively used the gypsy 2 line and performed 4C-seq experiments at different developmental stages (embryos, larval leg discs and pupal leg discs). Using the *dac* promoter (PRE2) (Extended Data Fig. 6a) and the RE (Extended Data Fig. 6b) as viewpoints, we observed that gypsy insertion resulted in a similar physical insulation activity at all developmental stages and PRE looping was reduced at all developmental stages investigated (Extended Data Fig. 6a,b).

Altogether, these experiments indicated that gypsy insertion interferes with PRE looping during all stages of fly development investigated, although the position of insertion of the gypsy insulator affects the strength of the insulation effect.

Gypsy insulator does not block RE function

Next, we asked whether gypsy insertions affect enhancer–promoter interactions and change *dac* expression. We predicted that insertion of the gypsy insulator upstream of the RE (gypsy 1), which interferes with PRE looping, might induce *dac* expression in TS2, as is the case in the PRE deletion lines (Fig. 1). On the other hand, gypsy 2 and gypsy 3 insulators inserted between the RE and the *dac* promoter might block enhancer–promoter communication.

Intriguingly, insertion of the gypsy insulators between the RE and the *dac* promoter (gypsy 2 and gypsy 3) resulted in a significant increase in *dac* expression in TS2 of pupal leg discs (Fig. 5a,b), whereas *dac* activation upon gypsy insertion downstream of the RE (gypsy 1) was much weaker. Moreover, activation of *dac* in TS2 in the gypsy 2 and gypsy 3 lines induced the *dac* gain-of-function phenotype (ESC) on adult male fly legs (Fig. 5c). The penetrance of ESC in the gypsy 2 and gypsy 3 lines (about 25% and 50% of all male flies, respectively) was significantly stronger than the penetrance observed in the gypsy 1 line. Notably, there was a good correlation between the penetrance of ESC and the levels of overexpression of *dac* in TS2 (Fig. 5b,c), which in turn inversely correlated with the reduction in PRE contacts (Fig. 4e), suggesting that loss of PRE looping and *dac* activation in TS2 are functionally linked.

The absence of enhancer-blocking activity when the insulator was inserted between the RE and promoter is surprising, given the previously reported enhancer-blocking function of the gypsy element[30]. One possible explanation could be that gypsy blocks the RE but, in this case, a shadow enhancer might take over and begin to induce *dac* expression.

To test whether the overexpression of *dac* upon gypsy insertions is induced by the RE, we deleted the RE in the presence of the gypsy 2 insertion (gypsy 2 + ΔRE line; Fig. 1a). Indeed, if the gypsy insertion induces *dac* overexpression independently of the RE, we would expect at least a partial rescue of the loss-of-function phenotype upon deletion of the RE (Fig. 3a). Instead, we observed that homozygous flies carrying gypsy 2 + ΔRE displayed the same strong crippled leg phenotype

https://doi.org/10.1038/s41594-024-01375-7

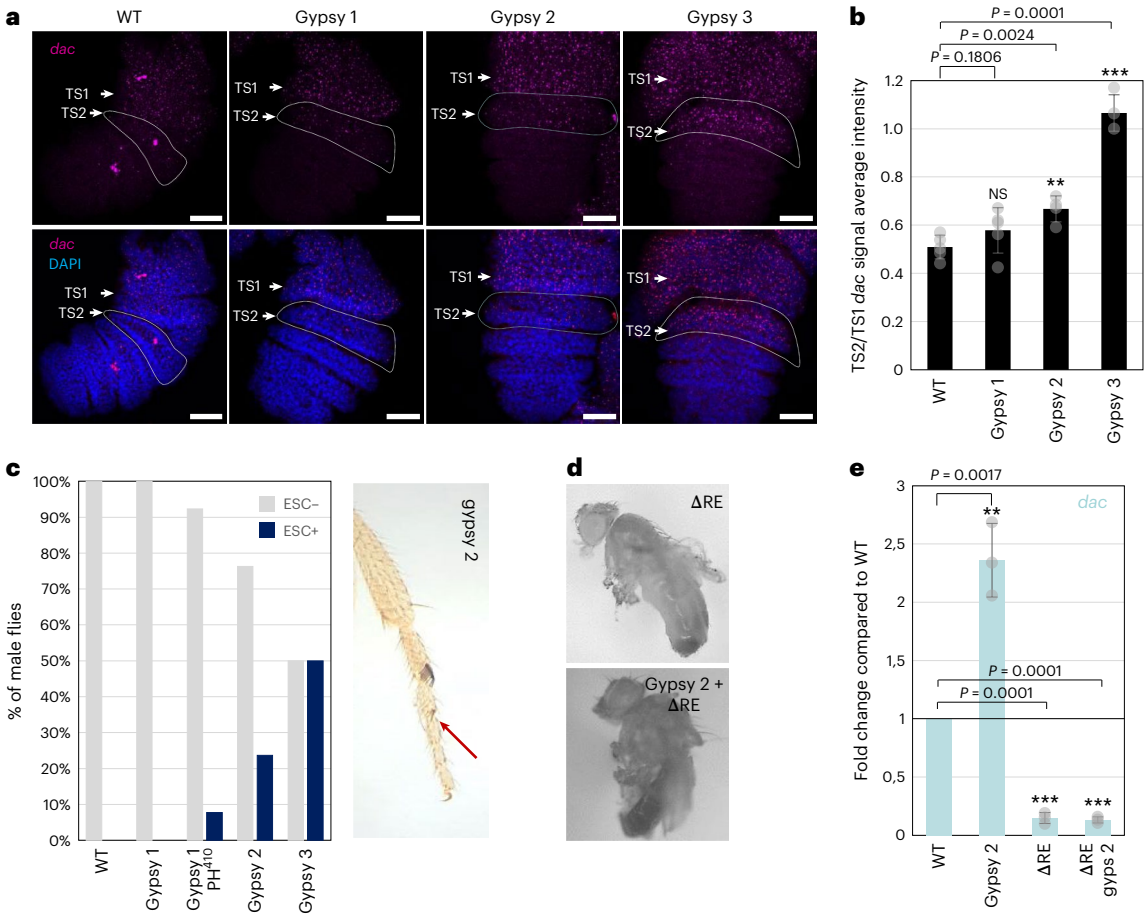


Fig. 5 | Transcriptional and phenotypic consequences of gypsy insulator insertions at the *dac* Polycomb domain. a, RNA FISH images of WT and gypsy insertions in early pupal imaginal leg discs (4–5 h after pupation) (*dac* gene, violet). White bars indicate 30 µm. TS2 is encircled by a dashed line. These images are *z* projections of one side of the pupal leg disc representing a single cell layer. **b**, Quantification of FISH signals of the *dac* gene in TS2 versus TS1. For each segment of the pupal leg disc, the number of *dac* FISH spots was quantified and normalized by the surface of the segment. The *y* axis represents the average intensity ratio of TS2:TS1 values of four imaginal discs ($n = 4$). Data are presented as the mean values ± s.d. (error bars) of replicates. $^{**}P < 0.01$ and $^{***}P < 0.001$ (two-sided unpaired *t*-test). NS, not significant. **c**, Quantification of the ESC phenotype in the indicated fly lines grown at 25 °C. A minimum of 50 male flies were scored. The ESC phenotype in gypsy 1 flies can only be observed in a sensitized mutant background, where levels of the PcG protein PH are reduced (PH^{410} mutant background)[22]. **d**, Representative example of adult flies carrying a deletion of the RE and/or an insertion of the gypsy insulator (gypsy 2). **e**, RT–qPCR analysis of third-instar imaginal leg discs in the indicated fly lines. The fold change between the indicated mutant and WT (control) line is shown. Three independent experiments were performed. Data are presented as the mean values ± s.d. (error bars) of replicates. $^{***}P < 0.001$ and $^{**}P < 0.01$ (two-sided unpaired *t*-test).

compared to the deletion of the RE alone, with a complete penetrance (Fig. 5d). Furthermore, RT–qPCR analysis showed that *dac* expression was reduced to the same extent in the gypsy 2 + ΔRE line compared to the ΔRE alone (Fig. 5e).

Altogether, these results indicate that the RE can bypass the gypsy insulator to activate *dac* expression irrespective of the genomic location of the gypsy insulator. Moreover, activation of the *dac* gene upon gypsy insertion and reduced PRE looping are strictly dependent on the presence of RE.

Unchanged enhancer–promoter proximity upon *dac* gene activation

Next, we tested whether the bypass of the insulator and activation of the *dac* gene involves changes in the physical proximity between enhancer–promoter sequences. For this, we performed Hi-M experiments in pupal leg discs in the gypsy 2 mutant, which allowed us to analyze RE–promoter distances specifically in the TS where *dac* is activated. Comparing the interaction profiles of gypsy 2 mutant cells to WT cells confirmed that the PRE loop is reduced in all TSs (TS1, TS2 or TS3 and TS4) independent of the *dac* expression status (Extended Data Fig. 7a). Furthermore, we observed increased short-range interactions upstream or downstream of the gypsy insulator, whereas long-range interactions between sequences upstream and downstream of the insulator insertion were reduced, consistent with the insulator activity of the gypsy sequence (Extended Data Fig. 7a,b).

Next, we compared the distances between the RE and the *dac* promoter in the TSs where *dac* is active (TS1), moderately expressed (TS2) or inactive (TS3 and TS4) in WT or gypsy 2 mutant flies. Therefore, we created virtual 4C plots derived from Hi-M experiments using the *dac* promoter as a viewpoint (Fig. 6a). Remarkably, despite the upregulation of *dac* in the gypsy mutant, the RE–*dac* promoter distance was not decreased but displayed a very modest increase between WT and the gypsy mutant in TS2 (≤10 nm). A similar weak increase in RE–promoter distances was observed in TSs where *dac* remains repressed (TS3 and TS4), indicating that this increase in RE–promoter distance was not sufficient for gene activation. We note that the RE–*dac* promoter distance changes were statistically significant for these segments (Wilcoxon two-sided rank test, Fig. 6a) but were, in all cases, comparable to the measurement error in the median distance as estimated by bootstrapping analysis (~8–25 nm; Methods). We, therefore, concluded that

https://doi.org/10.1038/s41594-024-01375-7

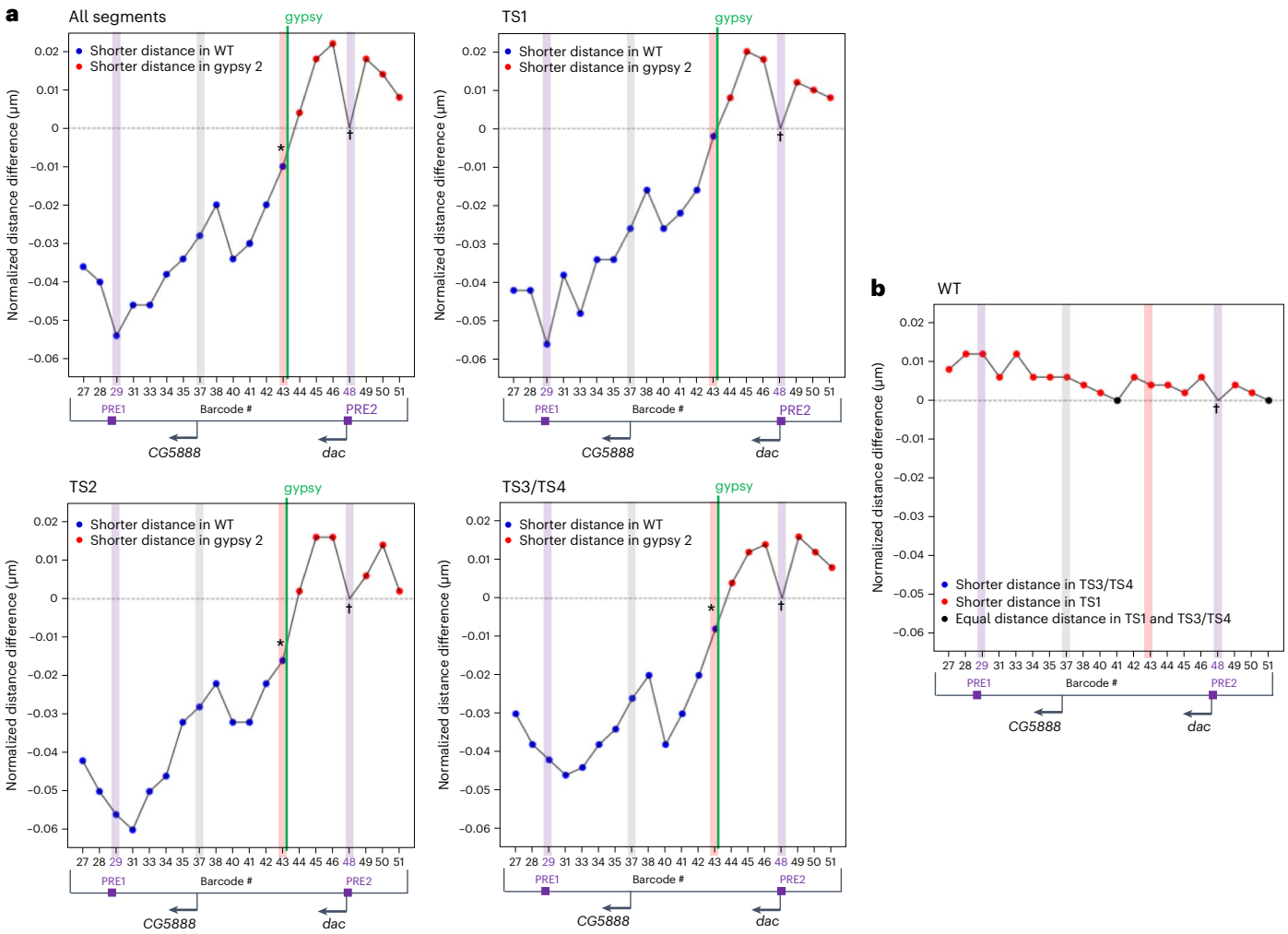


Fig. 6 | RE–*dac* promoter distances in TSs of early pupal leg discs in WT versus gypsy 2 flies determined by Hi-M. a, Virtual 4C plots derived from Hi-M PWD difference matrices of WT and gypsy 2 mutant conditions of pupal leg discs in the indicated TSs. Barcode 48 (*dac* promoter) was used as a viewpoint (black cross). Blue and red dots represent shorter distances for the WT condition and gypsy 2 mutant condition, respectively. Barcode 43 (salmon bar) indicates the position of the RE. Violet bars highlight the positions of PRE1 (barcode 29) and PRE2 (barcode 48). The green bar indicates the position of the gypsy 2 insulator. The gray bar indicates the position of the CG5888 promoter (barcode 37). Asterisks represent *P* values obtained by a Wilcoxon two-sided rank test (*$P < 0.05$). **b**, Virtual 4C plots derived from Hi-M PWD difference matrices of TS3 + TS4 and TS1 in WT pupal leg discs. Blue and red dots represent shorter distances for TS3 + TS4 and TS1, respectively. Barcode 43 (salmon bar) indicates the position of the RE. Violet bars (barcodes 29 and 48) highlight the positions of PRE1 and PRE2, respectively. The green bar indicates the position of the gypsy insulator. The gray bar indicates the position of the CG5888 promoter (barcode 37).

activation of *dac* expression in TS2 upon loss of PRE looping is not the consequence of large-scale changes in RE–promoter distances. Likewise, RE–*dac* promoter distances exhibited no significant changes when we compared TS1 (where *dac* is active) to TS3 and TS4 (where *dac* is completely repressed) in WT pupal leg discs (Fig. 6b). Importantly, the distance distributions across individual cells between PREs or enhancer–promoter elements showed no evidence for bimodality in all TSs including TS2, which comprises a mixed population of *dac* expressed and repressed cells (Extended Data Fig. 8a,b). This argues against the existence of different chromatin conformations within the examined TSs.

Together, these results indicate that increased physical proximity between the RE and *dac* promoter does not appear to be a major mechanism involved in *dac* gene activation.

The PRE loop contributes to enhancer–promoter specificity

Because the major structural effect of gypsy insertion is to reduce PRE looping at the *dac* domain, we asked whether other genes within the domain are affected by the loss of looping interactions. Therefore, we first performed RT–qPCR analysis at different developmental stages in mutant fly lines, where we interfered with either PRE function (double, ΔPRE2), PRE looping (gypsy 2) or enhancer function (ΔRE) (Fig. 7a). We found that the *CG5888* gene was induced already at the larval stage upon loss of PRE contacts. The *Idgf1* gene was also activated, whereas two other genes, *Idgf2* and *Idgf3*, remained repressed at all developmental stages analyzed (Extended Data Fig. 9a). RNA FISH analysis confirmed the transcriptional activation of *CG5888* upon the reduction in PRE looping in both gypsy 2 and gypsy 3 larval leg discs (Fig. 7b and Extended Data Fig. 9b,c). Interestingly, *CG5888* was mainly overexpressed in cells that also expressed *dac* and where the RE was active (Fig. 7c), resulting in a similar ring-shaped expression. Importantly, expression of the *CG5888* gene was not significantly changed upon deletion of the RE alone (Fig. 7b,d,e), indicating that *CG5888* expression is not controlled by the RE in WT conditions. However, the ectopic expression of the *CG5888* gene upon PRE looping reduction depends on the presence of the RE because *CG5888* activation was lost when we deleted the RE in the presence of the gypsy 2 insertion (ΔRE + gypsy 2) (Fig. 7b,d,e). Importantly, the perturbation of the PRE loop did not

https://doi.org/10.1038/s41594-024-01375-7

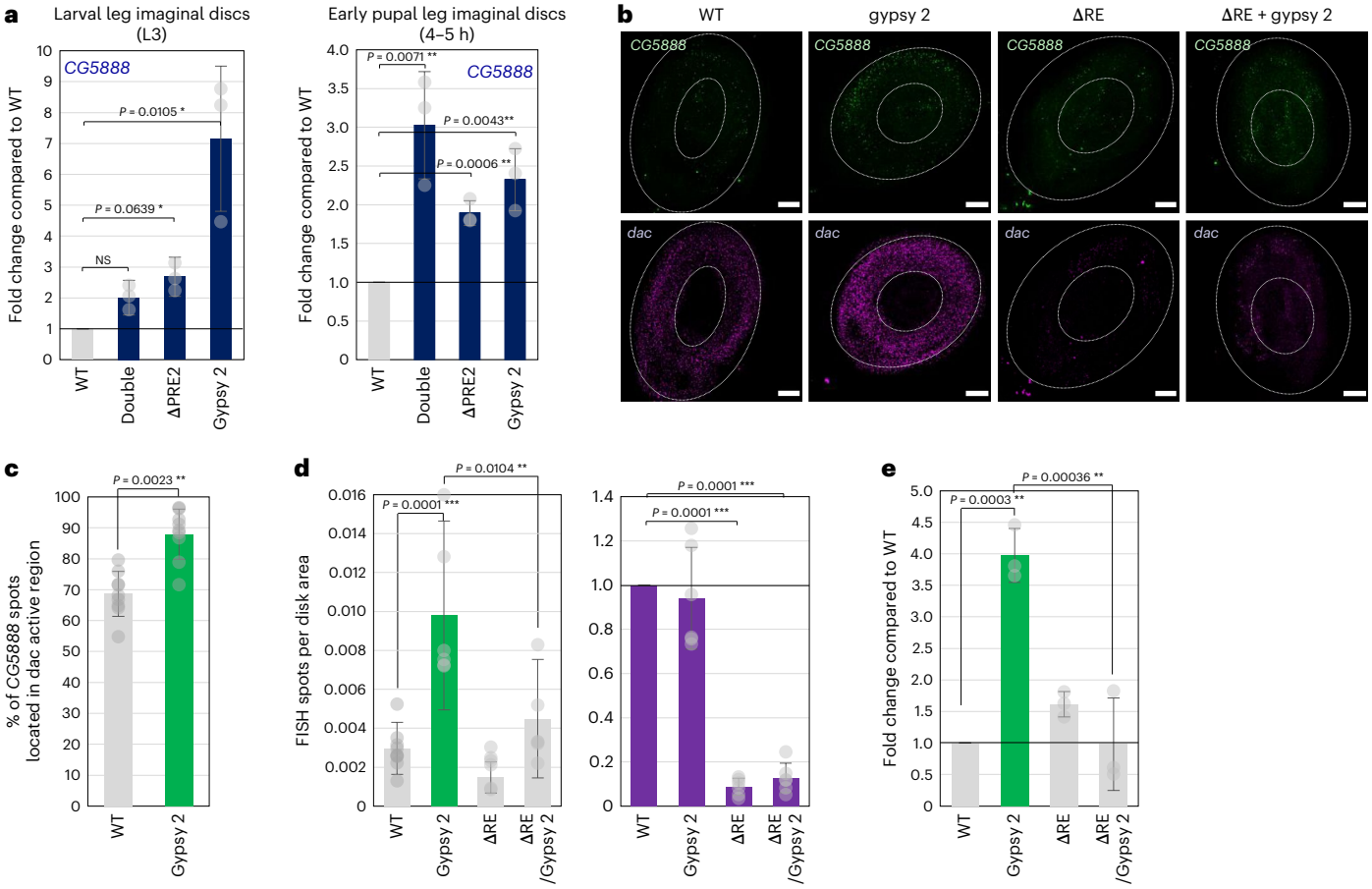


Fig. 7 | Consequences of gypsy insertions and/or RE deletion on *CG5888* and *dac* gene expression. **a**, RT–qPCR analysis at the indicated developmental stages using primers specific for *CG5888* gene. The fold change between the indicated mutant and WT (control) line is shown. Three independent experiments were performed. Data are presented as the mean values ± s.d. (error bars) of replicates. $^{*}P < 0.1$ and $^{**}P < 0.01$ (two-sided unpaired *t*-test). **b**, RNA FISH images of WT, gypsy 2 insertion, ΔRE and ΔRE + gypsy 2 in third-instar imaginal leg discs (*dac* gene, violet; *CG5888* gene, green). White bars indicate 30 µm. **c**, Quantification of *CG5888* RNA FISH signals observed in *dac*-expressing cells in WT and gypsy 2 third-instar imaginal leg discs. Nine pupal discs ($n = 9$) were scored. Data are presented as the mean values ± s.d. (error bars) of replicates. $^{**}P < 0.01$ (two-sided unpaired *t*-test). **d**, Quantification of *CG5888* (left) or *dac* (right) RNA FISH signals in the indicated fly lines. The number of FISH spots was quantified and normalized by the total disc area (*CG5888*) or by the number of FISH spots in the WT condition (*dac*). The *y* axis represents the average of six imaginal disc values. Error bars indicate the s.d. $^{***}P < 0.001$, $^{**}P < 0.01$ and $^{*}P < 0.1$ (two-sided unpaired *t*-test). **e**, RT–qPCR analysis of WT, gypsy 2 insertion, ΔRE and ΔRE + gypsy 2 in third-instar imaginal leg discs using primers specific for the *CG5888* gene. The fold change between the indicated mutant and WT (control) line is shown. Three independent experiments were performed. Data are presented as the mean values ± s.d. (error bars) of replicates. $^{**}P < 0.01$ (two-sided unpaired *t*-test).

lead to the activation of all genes within the *dac* TAD (Extended Data Fig. 9a) and was restricted to the region where the RE was active during leg development. This indicates that the loss of the PRE loop does not create a global permissive environment facilitating transcription per se but rather regulates RE specificity toward permissive promoters such as the *CG5888* gene. Altogether, these results suggest that, in addition to TADs, which can restrict enhancer–promoter communication across TAD borders, PRE loops within TADs constitute an additional gene-regulatory layer by contributing to intra-TAD enhancer specificity (Extended Data Fig. 10a).

Discussion

We showed that the *dac* PRE loop constitutes a topological chromatin structure that has a versatile function. On one hand, it can regulate enhancer–promoter communication in a manner specific to the developmental stage and tissue. On the other hand, the PRE loop contributes to enhancer–promoter specificity by restricting the enhancer activity to its specific target promoter (the *dac* gene), as shown by the RE-dependent illegitimate activation of the *CG5888* gene upon insertion of a gypsy insulator between the two PREs.

Gypsy insulator reduces PRE looping but not enhancer function

Insulator elements are DNA sequences that act as chromatin boundaries and regulate interactions between genomic regulatory elements. The gypsy element is one of the best-characterized insulators. It contains three core components, Su(Hw), CP190 and Mod(mdg4), and the zinc finger protein CLAMP that promotes gypsy enhancer-blocking activity[31].

At the *dac* TAD, the gypsy insulator sequence interferes with the interaction of the two PREs when inserted between them. However, gypsy insertion between the RE and the *dac* promoter does not block enhancer–promoter communication. Although surprising, these results are consistent with previous work showing that the Su(Hw) protein binds at thousands of sites throughout the genome yet is not systematically associated with the physical boundaries of gene units[5]. Although more than 20 enhancers have been shown to be blocked by the gypsy insulator (as discussed in a previous study[30] and references therein), this element does not necessarily establish an impermeable chromatin barrier[32,33]. In all the enhancer–promoter pairs that were analyzed in these studies, each element of the pair was located close to the gypsy element and insulation might perhaps depend on the ability

of the gypsy element to reduce short-range chromatin contacts. Furthermore, a recent genome-wide analysis suggested that gypsy-binding proteins do not generally act by blocking loop formation but rather by inducing local insulation at their binding sites[34].

In transgenic reporter assays, the gypsy insulator behaves as a chromatin border that is able to block spreading of H3K27me3 and prohibit contacts between a PRE and a distal promoter[29,35]. In contrast, at the *dac* TAD, insertion of the same gypsy insulator between the two PREs does not interfere with the formation of the repressive H3K27me3 domain. This is in agreement with the presence of endogenous Su(Hw)-binding sites within the *dac* TAD that, at this genomic locus, do not act as classical chromatin domain borders and do not interfere with the deposition of H3K27me3. This indicates that the function and effect of the gypsy insulator on the physical interaction between CREs or the partitioning of chromatin domains are context dependent.

PRE contacts constitute a specific form of TE loops

A recent paper used the Micro-C technique to demonstrate that the genome is organized by insulator elements and so-called TEs in early *Drosophila* embryos[15]. These TEs correspond to 'organizational elements' that form chromatin loops to facilitate specific enhancer–promoter contacts for rapid gene activation.

Intriguingly TEs and PREs show several similarities. First, both PRE loop and TE loop disruptions have little effect on the overall structure of TADs, whereas they have an impact on enhancer–promoter communication. Secondly, both elements are frequently bound by the GAGA factor, which has been proposed to mediate chromatin loops[36]. Lastly and most compellingly, we observed that 48% (109 of 225) of PREs identified in embryos coincided with TEs (Extended Data Fig. 10b), notably including the two PREs of the *dac* gene locus. We, therefore, propose that PRE loops actually constitute a specific form of TE loops.

In contrast to the *dac* PRE loop, which restricts enhancer–promoter communication in pupal leg discs, a chromatin loop between TEs (that also corresponds to PREs) at the *scr* locus was shown to promote enhancer–promoter contacts[15]. Another example of a TE loop that corresponds to a PRE loop involves the regulation of the *cut* (*ct*) gene. Importantly, a natural insertion of a gypsy retrotransposon between the two PREs and TEs of the *ct* gene locus leads to *ct* downregulation and loss of enhancer–promoter communication[28]. Thus, chromatin loops involving PREs and TEs can clearly mediate both gene activation and repression. Intriguingly deletion of both looping anchor points of the *dac* PRE loop resulted in reduced *dac* gene expression in embryos[22]. Although this reduced expression does not result in developmental defects, this suggests that even the same PRE or TE loop can have different functions in gene activation and repression depending on the developmental stage or tissue. These antipodal functions of PRE and TE loops might be mediated by the presence of developmental or stage-specific factors. Interestingly, a recent study analyzed chromatin loops of paralogous gene pairs, proposing an additional function of chromatin loops in the fine-tuning of coordinated expression levels of genes with related function[37]. As is the case for the *dac* PRE loop, the same chromatin loop can have different functions in the regulation of gene expression[37].

Together, these data suggest that PRE and TE loops constitute a chromatin scaffold structure that is used to regulate enhancer–promoter communication positively or negatively, depending on the gene locus, the developmental stage and the specific tissue.

PRE looping as a regulator of enhancer–promoter communication

Loss of PcG binding to the PREs is unlikely to be the major mechanism leading to *dac* gene activation upon loss of PRE looping. We previously observed that deletion of both PREs, with concomitant loss of PcG binding, was not sufficient to globally activate *dac* expression but only resulted in the same tissue-specific activation of the *dac* gene as the insertion of the gypsy insulator. Moreover, PcG proteins were shown to remain associated with PREs when their target genes are active[38,39]. In contrast, activation of the *dac* gene in TS2 upon loss of PRE looping is strictly dependent on the RE, indicating that gene activation involves changes in enhancer–promoter communication.

A popular model of enhancer–promoter communication proposes chromatin looping as a mechanism to bring the two regulatory regions in close proximity[40] and experimentally forced enhancer–promoter looping contributes to gene activation[41]. By applying Hi-M to detect chromatin 3D organization in single nuclei[42], including loops at distances similar to that separating the RE from the *dac* TSS, we were unable to detect increased contact frequency between the RE and the *dac* promoter compared to surrounding regions. In addition, no significant differences in contact frequencies of the *dac* promoter and RE were observed in repressed versus active cells, or upon loss of PRE looping, suggesting that no stable contact is needed for *dac* transcription. This agrees with a previous report showing no differences between enhancer–promoter pairs in active or inactive transcriptional states during early *Drosophila* development[42]. In addition, live-cell imaging analyses challenge the idea of stable loops as a general mechanism for all enhancer–promoter communication[43,44]. Lastly, increased enhancer–promoter distance has been reported to accompany *Shh* gene activation during neural differentiation[45], supporting evidence that enhancers can act at a distance.

Several 'action-at-a-distance' models have been proposed to explain the functional interaction of distant enhancers with their promoters[46], including liquid–liquid phase separation, where enhancers function as binding surfaces that concentrate transcriptional activators in 'transcriptional condensates', allowing them to induce gene expression even when promoters are not in absolute proximity. Alternatively, in the transcription factor activity gradient model, enhancer-bound coactivators can activate target promoters within a 'permissive range'. It is conceivable that the PRE loop might act as a crucial scaffold to create a particular chromatin environment or a 'permissive range', thereby regulating activation of the *dac* gene and enhancer specificity.

It is worth mentioning that the RE at the *dac* locus is not associated with the previously well-established molecular signatures of active enhancers (such as H3K4me1 and H3K27Ac enrichment) in embryos or larval disc tissues[47]. This surprising lack of histone marks at an essential enhancer might indicate that other such enhancers with an important function but undetectable epigenomic signature exist in the genome. In the future, it will be interesting to study whether chromatin looping or contact-independent mechanisms are differently involved in gene activation for different classes of enhancer elements.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-024-01375-7.

References

1. Jerkovic, I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. *Nat. Rev. Mol. Cell Biol.* **22**, 511–528 (2021).
2. Peng, A. et al. Regulation of 3D organization and its role in cancer biology. *Front. Cell Dev. Biol.* **10**, 879465 (2022).
3. Deng, S., Feng, Y. & Pauklin, S. 3D chromatin architecture and transcription regulation in cancer. *J. Hematol. Oncol.* **15**, 49 (2022).
4. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
5. Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).

6. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
7. Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).
8. Despang, A. et al. Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* **51**, 1263–1271 (2019).
9. Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene–enhancer interactions. *Cell* **161**, 1012–1025 (2015).
10. Franke, M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
11. Narendra, V. et al. CTCF establishes discrete functional chromatin domains at the *Hox* clusters during differentiation. *Science* **347**, 1017–1021 (2015).
12. Guo, Y. et al. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
13. Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in *IDH* mutant gliomas. *Nature* **529**, 110–114 (2016).
14. Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
15. Batut, P. J. et al. Genome organization controls transcriptional dynamics during development. *Science* **375**, 566–570 (2022).
16. Schwartz, Y. B. et al. Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* **38**, 700–705 (2006).
17. Schuettengruber, B. et al. Functional anatomy of Polycomb and Trithorax chromatin landscapes in *Drosophila* embryos. *PLoS Biol.* **7**, e13 (2009).
18. Schuettengruber, B., Bourbon, H. M., Di Croce, L. & Cavalli, G. Genome regulation by Polycomb and Trithorax: 70 years and counting. *Cell* **171**, 34–57 (2017).
19. Schuettengruber, B. & Cavalli, G. Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* **136**, 3531–3542 (2009).
20. Eagen, K. P., Aiden, E. L. & Kornberg, R. D. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc. Natl Acad. Sci. USA* **114**, 8764–8769 (2017).
21. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**, 216–228 (2017).
22. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J. M. & Cavalli, G. Polycomb-dependent chromatin looping contributes to gene silencing during *Drosophila* development. *Mol. Cell* **71**, 73–88 (2018).
23. Kundu, S. et al. Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation. *Mol. Cell* **65**, 432–446 (2017).
24. Isono, K. et al. SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* **26**, 565–577 (2013).
25. Wani, A. H. et al. Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat. Commun.* **7**, 10291 (2016).
26. Cardozo Gizzi, A. M. et al. Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. *Mol. Cell* **74**, 212–222 (2019).
27. Giorgianni, M. W. & Mann, R. S. Establishment of medial fates along the proximodistal axis of the *Drosophila* leg through direct activation of *dachshund* by Distalless. *Dev. Cell* **20**, 455–468 (2011).
28. Jack, J. & DeLotto, Y. Structure and regulation of a complex locus: the cut gene of *Drosophila*. *Genetics* **139**, 1689–1700 (1995).
29. Comet, I., Schuettengruber, B., Sexton, T. & Cavalli, G. A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *Proc. Natl Acad. Sci. USA* **108**, 2294–2299 (2011).
30. Kuhn, E. J., Viering, M. M., Rhodes, K. M. & Geyer, P. K. A test of insulator interactions in *Drosophila*. *EMBO J.* **22**, 2463–2471 (2003).
31. Bag, I., Dale, R. K., Palmer, C. & Lei, E. P. The zinc-finger protein CLAMP promotes gypsy chromatin insulator function in *Drosophila*. *J. Cell Sci.* **132**, jcs226092 (2019).
32. Morris, J. R., Chen, J. L., Geyer, P. K. & Wu, C. T. Two modes of transvection: enhancer action in trans and bypass of a chromatin insulator in *cis*. *Proc. Natl Acad. Sci. USA* **95**, 10740–10745 (1998).
33. Scott, K. C., Taubman, A. D. & Geyer, P. K. Enhancer blocking by the *Drosophila* gypsy insulator depends upon insulator anatomy and enhancer strength. *Genetics* **153**, 787–798 (1999).
34. Kahn, T. G. et al. Topological screen identifies hundreds of Cp190- and CTCF-dependent *Drosophila* chromatin insulator elements. *Sci. Adv.* **9**, eade0090 (2023).
35. Comet, I. et al. PRE-mediated bypass of two Su(Hw) insulators targets PcG proteins to a downstream promoter. *Dev. Cell* **11**, 117–124 (2006).
36. Li, X. et al. GAGA-associated factor fosters loop formation in the *Drosophila* genome. *Mol. Cell* **83**, 1519–1526 (2023).
37. Pollex, T. et al. Chromatin gene-gene loops support the cross-regulation of genes with related function. *Mol. Cell* **84**, 822–838 (2024).
38. Papp, B. & Muller, J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* **20**, 2041–2054 (2006).
39. Bowman, S. K. et al. H3K27 modifications define segmental regulatory domains in the *Drosophila bithorax* complex. *eLife* **3**, e02833 (2014).
40. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
41. Petrascheck, M. et al. DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Res.* **33**, 3743–3750 (2005).
42. Espinola, S. M. et al. *cis*-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early *Drosophila* development. *Nat. Genet.* **53**, 477–486 (2021).
43. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166**, 358–368 (2016).
44. Alexander, J. M. et al. Live-cell imaging reveals enhancer-dependent *Sox2* transcription in the absence of enhancer proximity. *eLife* **8**, e41769 (2019).
45. Benabdallah, N. S. et al. Decreased enhancer–promoter proximity accompanying enhancer activation. *Mol. Cell* **76**, 473–484 (2019).
46. Yang, J. H. & Hansen, A. S. Enhancer selectivity in space and time: from enhancer–promoter interactions to promoter activation. *Nat. Rev. Mol. Cell Biol.* **25**, 574–591 (2024).
47. Loubiere, V., Papadopoulos, G. L., Szabo, Q., Martinez, A. M. & Cavalli, G. Widespread activation of developmental gene expression characterized by PRC1-dependent chromatin looping. *Sci. Adv.* **6**, eaax4001 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the

Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024

Methods

Fly work and generation of mutant flies by CRISPR–Cas9 genome engineering

All flies were raised on standard corn meal yeast extract medium at 25 °C. CRISPR–Cas9 mutant fly lines double, ΔPRE1, ΔPRE2 and gypsy 1 were described in a previous study[22]. Sequences of guide RNAs (gRNAs) used to create fly lines gypsy 2, gypsy 3, ΔRE and gypsy 2 + ΔRE are described in Supplementary Table 2. Sense and antisense oligonucleotides were annealed and phosphorylated by the T4 polynucleotide kinase (New England Biolabs, M0201S) before being inserted inside a pCFD3 plasmid (Addgene, 49410) previously digested by BbsI (NEB, R0539S). To create the pHD-dsRED donor plasmid (Addgene) containing a removable (floxed) 3XP3–dsRED construct flanked by loxP sites and DNA fragments having homology to the target regions (homology arms) serving as a template for homology-directed repair, 1.5-kb genomic DNA fragments were amplified by PCR (Supplementary Table 2) and inserted into the pHD-dsRED plasmid using Gibson assembly (kit NEBuilder; New England Biolabs, E2621S).

The gypsy insulator was amplified from the plasmid (Gy)w(Gy) described in a previous study[48] and introduced into the donor plasmid cut by SpeI and BglII using Gibson cloning (Supplementary Table 2). To generate mutant fly lines, gRNA-containing pCFD3 and pHD-dsRED donor plasmids were injected into flies expressing Cas9 in the germline (vas-Cas9(X) RFP^{-}; Bloomington, 55821). Injections and dsRED screening were performed by BestGene (https://www.thebestgene.com/). To remove the dsRED reporter construct, mutant flies were crossed with a fly line expressing CRE recombinase (Bloomington, 34516). To generate the gypsy 2 + ΔRE mutant line, gRNAs targeting the RE and corresponding donor plasmid were injected into gypsy 2 mutant lines previously generated and expressing Cas9 (vas-Cas9(III)). Coordinates and sequences of deleted regions can be found in Supplementary Table 3. Genotypes of mutant fly lines were confirmed by PCR genotyping and sequencing analysis of the mutated region.

Immunostaining experiments

For immunostaining, third-instar imaginal leg discs were dissected at room temperature in sterile Schneider medium. Pupae were selected at the very beginning of pupation, where pupae can be recognized by their white color (a pupal stage that lasts 1 h), and were dissected 3.5–4 h later. The discs were then fixed for 20 min in 4% formaldehyde and were permeabilized for 1 h in PBS + 0.5% Triton (for larval leg discs) or 0.8% Triton (for pupal leg discs). The samples were then incubated for 1 h in 3% BSA PBTr (1× PBS + 0.5% Triton X-100). DAC primary antibody was diluted 1:400 (DSHB mAbdac1-1) in 1% BSA PBTr and incubated overnight at 4 °C on a rotating wheel. The leg discs were washed in PBTr before adding the secondary antibody at 1:1,000 dilution (Thermo Fisher Scientific, A-31571) and incubated for 2 h at room temperature on a rotating wheel. Finally, the discs were extensively washed in PBTr. The proximal segments of the leg discs were removed by dissection to only keep the TSs, which were subsequently mounted on microscope slides using ProLong Gold reagent (Invitrogen, P36930). The different images were acquired on a Zeiss axioimager Z2 Apoptome Leica SP8 confocal microscope using the same settings for all mutant lines and analyzed using Fiji software.

RNA FISH experiments

RNA FISH probes were prepared with an RNA FISH probe kit (Thermo Fisher Scientific, F32956) from DNA probes amplified with the primers described in Supplementary Table 2. Third-instar imaginal leg discs were quickly dissected in Schneider medium. Pupae were selected at the very beginning of pupation, where pupae can be recognized by their white color (a pupal stage that lasts 1 h), and were dissected 3.5–4 h later. The discs were then fixed with 4% formaldehyde before being permeabilized with PBTr for 4 h. Subsequently, discs were incubated for 10 min with 50% PBT (PBS + 1% Triton), 50% hybridization solution (50% formamide, 5× saline–sodium citrate (SSC), 100 g ml^{-1} fragmented salmon testes DNA, 50 g ml^{-1} heparin and 0.2% Tween-20) at room temperature. The samples were incubated for 45 min and then 1 h in hybridization solution at 55 °C. In parallel, a previously tested optimal concentration of labeled probe was diluted in 50 µl of hybridization solution, heated for 2 min at 85 °C and chilled on ice to denature RNA secondary structures. The discs were then incubated overnight with 50 µl of probe solution at 55 °C. The day after, the samples were washed three times at 55 °C with hybridization solution and twice with PBT. The proximal segments of the leg discs are removed by dissection to only keep the TSs, which were mounted on microscope slides using ProLong Gold reagent (Invitrogen, P36930). Images were acquired on a Zeiss axioimager Z2 APopoteme Leca SP8 confocal microscope using the same settings for all mutant lines and analyzed using Fiji software.

Hi-C experiments

Hi-C experiments were performed using the EpiTect Hi-C Kit (Qiagen, 59971). All Hi-C experiments were performed in two or three independent experiments using 50 third-instar imaginal leg discs or early pupal discs. Briefly, discs were homogenized and fixed in activated buffer T and 2% formaldehyde using tissue masher tubes (Biomasher II (EOG-sterilized), 320103). Tissue was digested by adding 25 µl of collagenase I and II (40 mg ml^{-1}) for 1 h at 37 °C. Samples were centrifuged and supernatant was carefully aspirated, leaving ~250 µl of solution in the tube. Then, 250 µl of QIAseq beads equilibrated to room temperature were added to bind nuclei to the beads and all subsequent reactions were performed on the beads according to the manufacturer's protocol. Libraries were sequenced at BGI (https://www.bgi.com/) with 150-bp paired-end reads (approximately 400 million reads per replicate).

Hi-C analysis

Raw data from Hi-C sequencing were processed using the 'scHiC2' pipeline. Sequencing statistics are summarized in Supplementary Table 4. Valid interactions were stored in a database using the 'misha' R package (https://github.com/msauria/misha-package). Extracting the valid interactions from the *misha* database, the 'shaman' R package (https://bitbucket.org/tanaylab/shaman) was used for computing the Hi-C expected models, Hi-C scores with parameters $k = 250$ and $k_exp = 500$ (Figs. 2a, 3c and 4a) and differential Hi-C interaction scores with parameters $k = 250$ and $k_exp = 250$, with the compared datasets downsampled to have the same number of valid pairs in chromosome 2L (chr2L) for each comparison (Fig. 4b). Specifically, Hi-C scores quantify the contact enrichment (positive values) or depletion (negative values) of each bin of the map with respect to a statistical model used to evaluate the expected number of counts. To generate this expected model, we randomized the observed Hi-C contacts using a Markov chain Monte Carlo-like approach per chromosome[49]. Shuffling was conducted such that the marginal coverage and decay of the number of observed contacts with the genomic distance were preserved but any features of genome organization (for example, TADs or loops) were not. These expected maps were generated for each biological replicate separately and contained twice the number of observed *cis* contacts. Next, the score for each contact in the observed contact matrix was calculated using the k nearest neighbors (kNN) strategy[49]. In brief, the distributions of two-dimensional Euclidean distances between the observed contact and its nearest k_exp neighbors in the pooled observed and pooled expected (per cell type) data were compared, using Kolmogorov–Smirnov D statistics to visualize positive (higher density in observed data) and negative (lower density in observed data) enrichments. These D scores were then used for visualization (using a scale from −100 to +100) and are referred to as Hi-C scores in the text. Accordingly, the color scale of the Hi-C scores comprises both positive and negative values. When computing the differential Hi-C scores maps of Fig. 4b, the reference dataset was used as the expected model.

For each condition, the Hi-C interaction quantifications at the *dac* PRE loop (Figs. 2c, 3d and 4e) were performed by considering the Hi-C scores between two regions of 6 kb, chr2L:16419514–16425515 and chr2L:16482929–16488930), including the PRE1 and PRE2, respectively (Supplementary Table 3). The distributions of Hi-C scores (Figs. 2c, 3d and 4e) are represented as box plots showing the median (central line), the 75th and 25th percentiles (box limits) and 1.5 × the interquartile range (IQR; whiskers). Each of the comparisons of the Hi-C interaction quantifications at the *dac* PRE loop was performed between a reference condition (embryo in Fig. 2c and WT larvae in Figs. 3d and 4e) and each of the other conditions present in the same figure. An unpaired two-sided Wilcoxon statistical test (with the null hypothesis that the true median shift is equal to zero and assuming that the two variables are not normally distributed) was used to estimate the reported *P* values. The annotation of the Polycomb-associated TADs in chr2L from a previous study[5] was used to compute the number of Hi-C interactions within the PcG TAD, which were then normalized by the total number of valid pairs at the corresponding developmental stage (embryo, larvae or pupae). The distributions of these interaction frequencies are shown in the violin and box plots of Fig. 2b as the $\log_2$ ratios of embryo over the larval and pupal leg discs. The box plots show the median (central line), the 75th and 25th percentiles (box limits) and 1.5 × IQR (whiskers). An unpaired two-sided Wilcoxon statistical test (with the null hypothesis that the true median shift is equal to zero and assuming that the two variables are not normally distributed) was used to estimate the reported *P* values. The ISs[50] were computed on the observed Hi-C datasets binned at 2-kb resolution with windows of 100, 150, 200, 250 and 300 kb, resulting in five values per bin, and were stored in the misha database using an in-house R script. The mean and s.d. for each of the 2-kb bins were computed for the plots in Figs. 3e and 4c. The quantification of the ISs at gypsy insertions and R0–R12 regions was performed by applying a pairwise statistical comparison of the five IS quantifications per 2-kb bin. The *P* values in Fig. 4d and Extended Data Fig. 4b resulted from a Welch *t*-test (with the null hypothesis that the true difference in means is equal to zero and assuming that the variances of the samples are not equal) between the WT condition and each of the gypsy mutants at the corresponding locus. All plots of Hi-C maps (Figs. 2a, 3c and 4a,b), Hi-C interaction score comparisons (Figs. 2c, 3d and 4e), IS profiles (Figs. 3e and 4c) and *P* values of IS comparisons (Fig. 4d) were obtained with in-house R scripts.

Hi-M library preparation

The oligopaint library covering the *dac* region consists of 52-mer sequences with genome homology ordered from CustomArray. These sequences were obtained from the oligopaint public database (http://genetics.med.harvard.edu/oligopaints). From the initial design of the library, we selected 20-mers with an average probe density of 9–17 probes per kb. Each barcode contained 45 probes covering 3.8 kb on average (Supplementary Table 1). Each oligo was composed of five different regions: (1) a 21-nt forward universal priming region for library amplification; (2) two 20-nt readout regions separated by an A for barcoding; (3) a 42-nt genome homology region; (4) a duplication of one 20-nt readout region; and (5) a 21-nt reverse universal priming region.

The procedure for oligopaint library amplification was previously described[26,42,51,52]. It consists of seven steps: (1) an emulsion PCR (emPCR) to extract the *dac* library from the oligonucleotide pool using universal primers; (2) a limited-cycle PCR performed on the emPCR to identify the most efficient amplification cycle; (3) a large-scale PCR with a T7 promoter on the reverse primer; (4) an in vitro T7 transcription; (5) RT to transform RNAs into single-stranded DNA (ssDNA); (6) an alkaline hydrolysis for the removal of the intermediate RNA; and (7) ssDNA purification and concentration.

Each barcode is unique and specific to an adaptor oligo. The adaptor oligo serves as a bridge between the readout region and an Alexa Fluor 647-labeled secondary oligonucleotide. The fluorescently labeled part of the secondary probe is attached by a disulfide leakage that can be cleaved (chemical bleaching) during the sequential imaging of FISH probes[51]. For the fiducial, we used an adaptor oligo complementary to the reverse primer of the library and specific to a secondary probe bound to a noncleavable rhodamine red fluorophore. Adaptors and fluorescently labeled secondary probes were synthesized and purchased from Integrated DNA Technologies.

Hi-M library hybridization

Pupae were collected at the beginning of pupation (white pupae) and dissected 3.5–4 h later. The dissected leg disc were fixed with 4% formaldehyde before being permeabilized with PBTr for 4 h. The discs were then progressively washed in four different concentrations of Triton and pHM (2× SSC and 0.1 M NaH_2PO_4 pH 7; 20%, 50%, 80% and 100%) for 20 min in each buffer at room temperature on a rotating wheel. Then, the discs were incubated overnight in 225 pmol of the library diluted in 30 µl of FISH hybridization buffer (FHB; 50% formamide, 2× SSC, 0.5 mg ml^{-1} salmon sperm DNA and 10% dextran sulfate). The probes and the discs in pHM were heated at 80 °C. The incubation of the leg discs in the FHB + probe buffer was performed in a PCR machine from 80 °C to 37 °C with a temperature decrease of 1 °C every 10 min. The next day, discs were washed two times with 50% formamide, 2× SSC and 0.3% CHAPS and sequentially washed with four different concentrations of formamide and PBT (40%, 30%, 20% and 10% formamide) for 20 min per buffer on a rotating wheel. Finally, the discs were washed with PBT, fixed with 4% formaldehyde in PBS, washed with PBS and stored at 4 °C.

Hi-M imaging system

Hi-M experiments were performed with a homemade wide-field and epifluorescence microscope. This setup includes a rapid automated modular microscope (Applied Scientific Instrumentation) coupled with a microfluidic device as previously described[26,42]. The microscope and fluidics system were controlled using Qudi-HiM (our homemade hardware control package)[53]. The fluidics system permitted the automated and sequential hybridizations of the probes. The solutions were delivered to the sample by a combination of three eight-way valves (HVXM 8-5, Hamilton), a negative pressure pump (MFCS-EZ, Fluigent) and an FCS2 flow chamber (Bioptechs). The excitation was performed by three different lasers: 405 nm (Obis 405, 100 mW; Coherent), 561 nm (Sapphire 561 LP, 150 mW; Coherent) and 642 nm (VFL-0-1000-642-OEM1, 1 W; MPB communications). The fluorescence was collected through a Nikon APO ×60 (1.2 numerical aperture) water-immersion objective lens mounted on a closed-loop piezoelectric stage (Nano-F100, Mad City Labs). Images were acquired using a scientific complementary metal–oxide–semiconductor camera (ORCA Flash 4.0 V3, Hamamatsu) with an effective optical pixel size of 106 nm. To correct axial drift in real time, we used a homemade autofocus system composed of a 785-nm laser (OBIS 785, 100 mW; Coherent) and an infrared-sensitive camera (DCC1545M, Thorlabs).

Acquisition of Hi-M datasets

The proximal part of the pupal leg discs was removed by dissection to only keep the TSs. About 15–20 TSs were aligned on a 2% agar–PBS pad and then attached onto a 40-mm round coverslip previously functionalized with trimethoxysilane and 10% poly(L-lysine). The slide was then mounted onto the flow chamber. Pupal leg discs were first incubated with the fiducial adaptor (25 nM of the adaptor specific to the reverse primer, 2× SSC and 40% v/v formamide) for 20 min and then washed with a washing buffer solution (2× SSC and 40% v/v formamide) for 10 min. To complete the hybridization of the fiducial, we did a second round of incubation with the appropriate secondary oligo (25 nM of rhodamine-red-labeled probe, 2× SSC and 40% v/v formamide) for 20 min and washed again for 10 min with the washing buffer solution. After a 5-min wash with 2× SSC, we proceeded with nuclear staining with 0.5 µg ml^{-1} of DAPI in PBS for 20 min. After another 5-min wash

with 2× SSC, the imaging buffer (1× PBS, 5% glucose, 0.5 mg ml^{-1} glucose oxidase and 0.05 mg ml^{-1} catalase) was injected to limit fiducial photobleaching during the acquisition. An image stack (200 µm × 200 µm region of interest) was acquired for each of the 10–15 pupal leg discs. The DAPI and the fiducial were sequentially imaged (using 405-nm and 561-nm lasers) with a *z* step size of 250 nm for a total range of 17.5 µm.

Next, adaptor oligos and the secondary probe were sequentially hybridized, acquired and photobleached to image the whole *dac* oligopaint library. The following steps were performed for each of the 22 barcodes: (1) adaptor (40 nM of adaptor oligonucleotide, 2× SSC and 40% v/v formamide) injection and incubation for 10 min; (2) imaging probe (40 nM secondary probe, 2× SSC and 40% v/v formamide) injection and incubation for 10 min; (3) 10-min wash with washing buffer solution; (4) 5-min wash with 2× SSC; (5) imaging buffer injection and sequential acquisition of fiducial and barcode with 561-nm and 642-nm lasers; (6) chemical bleaching (2× SCC and 50 mM TCEP) of the imaging probe; and (7) 5-min wash with 2× SSC before a new cycle of hybridization.

Image processing and Hi-M analysis

Raw TIFF images were deconvolved using Huygens Professional 21.04 (Scientific Volume Imaging, https://svi.nl). Hi-M analysis was performed using pyHiM, a homemade analysis pipeline (https://pyhim.readthedocs.io/en/latest/)[54], as previously described[55]. First, images were *z*-projected by applying either the sum for DAPI channels or the maximum intensity for the barcodes and fiducial. For each cycle of hybridization, fiducial images were used to register the corresponding barcode image using global and local registration methods. Barcodes and fiducials were segmented in 3D using a neural network, followed by 3D localization of the center of each barcode mask[54]. The fiducial oligo bound to the universal priming regions, thus labeling the entire *dac* locus. Therefore, we built chromatin traces by combining the DNA FISH spots colocalizing within single fiducial masks. DAPI images were used to manually segment the different TSs (TS1, TS2, TS3, TS4 or TS5). Pairwise distance (PWD) matrices were calculated for each single chromatin trace. From a list of PWD maps, we calculated the proximity frequencies as the number of chromatin traces in which PWDs were within 250 nm, normalized by the number of chromatin traces containing both barcodes. Hi-M maps of the WT condition were generated from 51,622 total traces from 48 pupal leg discs from two independent biological replicates. Hi-M maps of the Gypsy 2 mutant were produced from 63,458 total traces of 51 pupal leg discs from two independent biological replicates. Hi-M matrices were generated for all the TSs combined (TS1, TS2, TS3, TS4 and TS5), as well as for TS1, TS2 and TS3 + TS4. Each trace contained at least 12% of the barcodes. Virtual 4C figures were obtained by plotting the PWDs between the anchored barcode or viewpoint with the remainder of the barcodes of an Hi-M matrix.

Wilcoxon two-sided rank tests between the PWD distributions of the barcodes containing the RE and the *dac* promoter were performed to test the hypothesis that two independent samples (for example, WT and gypsy mutant) were drawn from the same distribution. P values < 0.05 were considered significant to reject the hypothesis (that is, a 5% significance level).

We estimated the error in the measurement of the median RE–*dac* promoter distance by performing bootstrapping analysis. For this, we performed 1,000 bootstrapping cycles drawn from the experimental distribution of PWDs to estimate the s.d. in the determination of the median distance. The errors were between 8 and 25 nm for the WT condition.

4C-seq experiments

For 4C, either about 3,000 embryos were collected or 300 third-instar imaginal leg discs were dissected, homogenized and fixed in 2% formaldehyde diluted in nuclear permeabilization (NP) buffer (15 mM HEPES pH 7.6, 60 mM KCl, 15 mM NaCl, 4 mM $MgCl_2$, 0.1% Triton X-100, 0.5 mM DTT and 1× protease inhibitors (complete EDTA-free tablets; Roche, 11 873 580 001)) for 10 min at room temperature. Fixation was stopped by adding 2 M glycine for 5 min. The samples were then washed once in NP buffer and twice in 1.25× NEB3 buffer and the pellet of fixed cells was frozen in liquid nitrogen and conserved at −80 °C.

The chromatin pellet was then layered with 500 µl of 1.25× DpnII buffer without resuspension and centrifuged. The pellet was resuspended in 250 µl of 1.25× DpnII buffer. Then, 10 µl of 10% SDS was added and incubated for 20 min at 65 °C and 40 min at 37 °C. Chromatin was then split into 250-µl 1.25× DpnII buffer aliquots of 5–6 × 10^6 cells and incubated for 1 h at 37 °C with 3.3% Triton X (final concentration). Samples were digested with 500 units of DpnII overnight. The day after, DpnII enzyme was inactivated by heating the samples at 65 °C for 20 min. The fragments were then ligated for 5 h at 16 °C with T4 ligase (2,000 units per µl) and digested overnight with proteinase K at 65 °C. The day after, RNA was degraded by RNAse A solution for 1 h at 37 °C. DNA was purified with Ampure beads without size selection and digested overnight with NlaIII enzyme. The next day, the DNA fragments were circularized by overnight ligation with T4 ligase (2,000 units per µl) in a large buffer volume. Next, circularized DNA was purified by Ampure beads without size selection. Lastly, 4C PCR was performed with the primers described in Supplementary Table 2. The amplified DNA was purified with Ampure beads. The sequencing libraries were produced with an Illumina kit (Illumina, 20015964). Sequencing (paired-end sequencing with 150-bp reads; approximately 4 Gb per sample) was performed by Novogene (https://en.novogene.com/).

4C-seq processing and analysis

Using a custom-made python script, FASTQ sequencing files were split using 4C primer sequences to obtain individual FASTQ files only containing reads from a single viewpoint per genotype and tissue type. Accordingly, the reads were trimmed to remove viewpoint sequences up to the restriction sites. Subsequently, the trimmed reads were aligned against the DM6 reference assembly using Bowtie[56] with the parameters '-a -v 0 -m 1' (no mismatches and no multiple alignments allowed). The number of successfully aligned reads can be found in Supplementary Table 5. The aligned reads were mapped to restriction fragments and genomic bins of 1 kb using HiCdat[57] to obtain tabular files describing the number of reads (that is, contact frequencies) for a given fragment or genomic bin. All subsequent analysis steps were conducted using R. Depending on the 4C samples genotypes and viewpoints, contact frequencies arising from the viewpoint (±4 bins) and contact frequencies mapping to genotype-specific deletions were masked by setting them to zero (Supplementary Table 3). Then, data from individual samples were normalized for differing overall library size (counts per million).

To analyze differences between different genotypes, *t*-tests using triplicate data per genotype were performed for each 1-kb genomic bin along the region of interest (chr2L:1630000–16600000). No multiple-testing correction was performed. Subsequently, the differences of the average of triplicates were plotted and genomic bins that exhibited P values < 0.1 were highlighted.

qRT–PCR experiments

Embryos were collected in a 16–20-h developmental time window. Third-instar imaginal leg discs or early pupal leg discs (3.5–4 hs after pupation) were quickly dissected (<30 min) in Schneider medium and transferred into Trizol. RNA was extracted using Trizol reagent and purified using an RNA clean and concentrator kit (Zymo Research, R1015) following the instructions and using DNAse I (Qiagen, 79254). Then, 250 ng of purified RNA was used for RT using the Maxima first-strand complementary DNA synthesis kit for RT–qPCR with dsDNase (Thermo Fisher Scientific, K1671) following the manufacturer's recommendations. Finally, quantification of the RT product was performed on a LightCycler 480 (Roche) with the primers listed in Supplementary

Table 2. Data analysis was performed on LightCycler software. Expression levels were normalized to the housekeeping gene *RP49*.

qCHIP experiments

qChIP experiments were performed as described in a previous study[58] with minor modifications. Chromatin was sonicated using a Bioruptor Pico (Diagenode) for 7 min (30 s on, 30 s off). Su(HW) antibody was diluted 1:100 for the immunoprecipitation. After decrosslinking, DNA was purified using MicroChIP DiaPure columns from Diagenode. Enrichment of DNA fragments was analyzed using a real-time PCR LightCycler 480 (Roche). The primers used are indicated in Supplementary Table 2.

CUT&RUN experiments

CUT&RUN experiments were performed as described in the literature[59] with minor modifications. A total of 50 eye discs were dissected in Schneider medium, centrifuged for 3 min at 700*g* and washed twice with Wash+ buffer before the addition of concanavalin A-coated beads. MNase digestion (pAG-MNase Enzyme from Cell Signaling) was performed for 30 min on ice. After proteinase K digestion, DNA was recovered using SPRIselect beads and eluted in 50 µl of Tris-EDTA. DNA libraries for sequencing were prepared using the NEBNext Ultra II DNA library prep kit for Illumina. Sequencing (paired-end sequencing with 150-bp reads; approximately 2 Gb per sample) was performed by Novogene (https://en.novogene.com/). H3K27me3 antibody (Active Motif, 39155) was diluted 1:100. IgG antibody (1:100, Cell Signaling Technology, 2729S) was used as a control.

CUT&RUN analysis

The quality of the reads was assessed using FastQC. FASTQ files were aligned to the *D. melanogaster* reference genome dm6 using Bowtie 2 (version 2.4.2)[60] with the following parameters: '--local --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -I10 -X700'. SAM files were compressed into BAM files using SAMtools (version 1.16.1) and reads with low mapping quality (Phred score < 30) were discarded. Duplicate reads were removed using Sambamba markdup (version 1.0.0)[61] with the following parameters: '-r --hash-table-size 500000 --overflow-list-size 500000'. For visualization, replicates were merged using SAMtools 'merge' with default parameters and reads per kilobase per million mapped reads (RPKM)-normalized bigWig binary files were generated using the bamCoverage (version 3.5.5) function from deepTools2 (ref. 62) with the following parameters: '--normalizeUsing RPKM --ignoreDuplicates -e 0 -bs 10'. Genome browser plots were generated using the pyGenomeTracks package (version 3.8)[63]. The 131 *Drosophila* Polycomb domains[22] were used for differential enrichment analysis using the DESeq2 method from the 'DiffBind' R package (version 3.12.0). Differential quantification results of H3K27me3 levels within Polycomb domains are summarized in Supplementary Table 6.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequencing data reported in this paper were deposited to the Gene Expression Omnibus under accession number GSE247377. Source data are provided with this paper.

Code availability

All original code was deposited on GitHub (https://github.com/cavallifly/Denaud_et_al_NatStructMolBiol_2024) and is publicly available as of the date of publication. Any additional information required to reanalyze the data reported in this paper is available from the corresponding author upon request.

References

48. Silicheva, M. et al. *Drosophila* mini-white model system: new insights into positive position effects and the role of transcriptional terminators and gypsy insulator in transgene shielding. *Nucleic Acids Res.* **38**, 39–47 (2010).
49. Olivares-Chauvet, P. et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296–300 (2016).
50. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
51. Cardozo Gizzi, A. M. et al. Direct and simultaneous observation of transcription and chromosome architecture in single cells with Hi-M. *Nat. Protoc.* **15**, 840–876 (2020).
52. Messina, O. et al. 3D chromatin interactions involving *Drosophila* insulators are infrequent but preferential and arise before TADs and transcription. *Nat. Commun.* **14**, 6678 (2023).
53. Barho, F. et al. Qudi-HiM: an open-source acquisition software package for highly multiplexed sequential and combinatorial optical imaging. *Open Res. Eur.* **2**, 46 (2022).
54. Devos, X. et al. pyHiM, a new open-source, multi-platform software package for spatial genomics based on multiplexed DNA-FISH imaging. *Genome Biol.* **25**, 47 (2024).
55. Gotz, M., Messina, O., Espinola, S., Fiche, J. B. & Nollmann, M. Multiple parameters shape the 3D chromatin structure of single nuclei at the *doc* locus in *Drosophila*. *Nat. Commun.* **13**, 5375 (2022).
56. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
57. Schmid, M. W., Grob, S. & Grossniklaus, U. HiCdat: a fast and easy-to-use Hi-C data analysis tool. *BMC Bioinformatics* **16**, 277 (2015).
58. Schuettengruber, B. et al. Cooperativity, specificity, and evolutionary stability of Polycomb targeting in *Drosophila*. *Cell Rep.* **9**, 219–233 (2014).
59. Ahmad, K. CUT&RUN with *Drosophila* tissues V.1. *protocols.io* https://doi.org/10.17504/protocols.io.umfeu3n (2018).
60. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
61. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
62. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
63. Lopez-Delisle, L. et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).

Acknowledgements

We are grateful to V. Corces (Department of Human Genetics, Emory University School of Medicine, Atlanta) for the anti-Su(Hw) antibody. We would like to thank the MRI and *Drosophila* facilities (BioCampus Montpellier), CNRS, INSERM and the University of Montpellier. This work was supported by grants from the European Research Council (Advanced Grant 3DEpi, to G.C.), the European CHROMDESIGN ITN project (Marie Skłodowska-Curie grant agreement No 813327, to G.C.), the European E-RARE NEURO DISEASES grant 'IMPACT' from the Agence Nationale de la Recherche (PLASMADIFF3D, grant no. ANR-18-CE15-0010, and LIVCHROM, grant no. ANR-21-CE45-0011, to G.C.), the Fondation ARC (EpiMM3D, to G.C.), the Fondation pour la Recherche Médicale (EQU202303016280, to G.C.), the MSD Avenir Foundation (Project GENE-IGH) and the French National Cancer Institute (INCa, PIT-MM grant no. INCA-PLBIO18-362, to G.C.). This project was also funded by the European Union's Horizon 2020

Research and Innovation Program (grant no. 724429, to M.N.), the Bettencourt-Schueller Foundation (prize 'Coup d'élan pour la recherche Française', to M.N.), France-BioImaging (grant no. ANR-10-INBS-04, 'Investments for the Future', to M.N.) and the *Drosophila* facility (BioCampus Montpellier, CNRS, INSERM and the University of Montpellier). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

G.C. and B.S. conceptualized the study. S.D., M.B. and B.S. performed the experiments. S.G. performed the bioinformatic analysis of 4C experiments. M.N. and M.B. conceptualized and designed the Hi-M experiments. M.B. performed the image acquisition and data analysis. G.P. and M.D.S. performed the bioinformatic analysis of Hi-C experiments. G.S. performed the analysis of CUT&RUN data. B.S. and G.C. wrote the manuscript. B.S., C.G. and M.N. edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41594-024-01375-7.

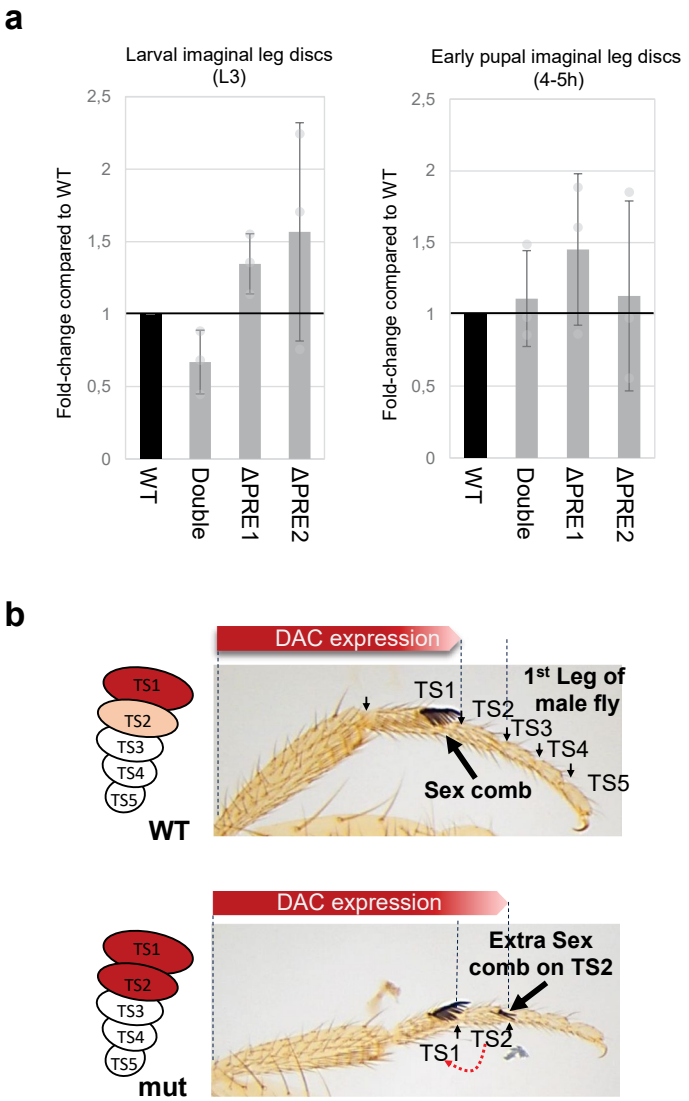
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-024-01375-7.

Correspondence and requests for materials should be addressed to Bernd Schuettengruber or Giacomo Cavalli.

Peer review information *Nature Structural & Molecular Biology* thanks Robert Illingworth, Son Nguyen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editors: Carolina Perdigoto, Jean Nakhle and Dimitris Typas, in collaboration with the *Nature Structural & Molecular Biology* team.

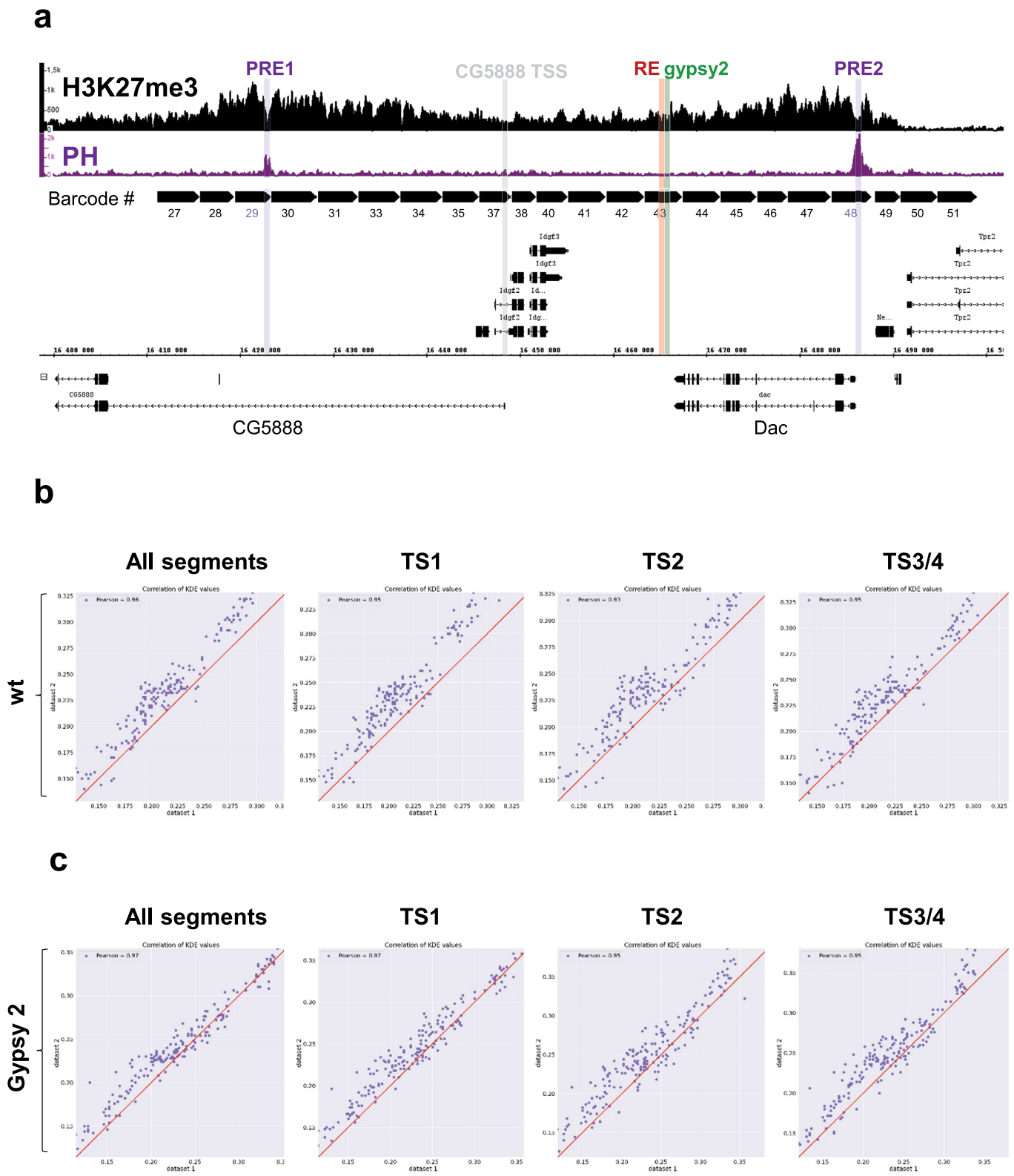
Reprints and permissions information is available at www.nature.com/reprints.

https://doi.org/10.1038/s41594-024-01375-7



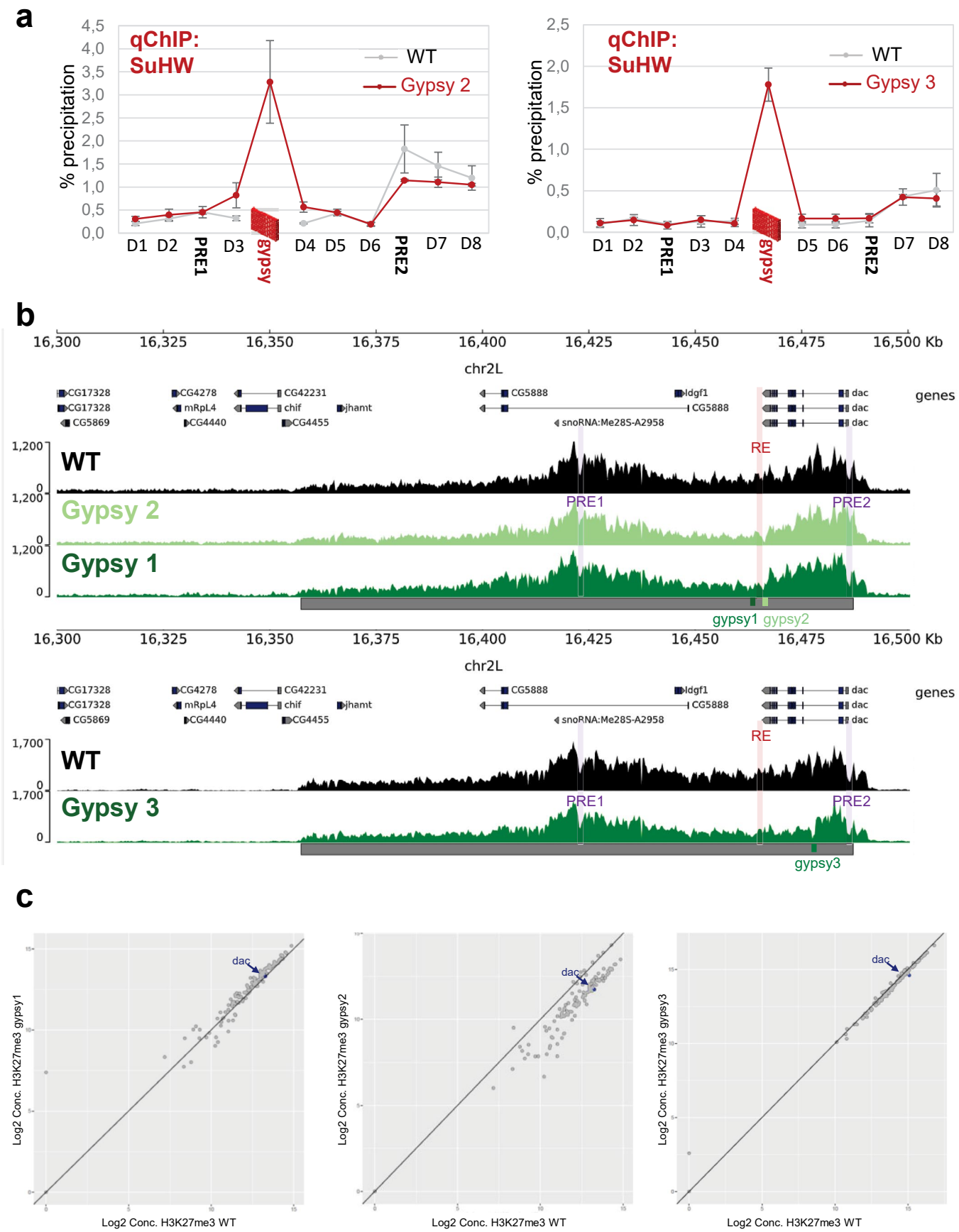
Extended Data Fig. 1 | *Dac* Gain of function phenotype upon loss of PRE function. (**a**) RT-qPCR analysis in the indicated fly lines and developmental stages using primers specific for *dac* gene. The fold change between the indicated mutant and wild type (control) line is shown. Three independent experiments have been performed. Data are presented as the mean values ± s.d (error bars) of replicates. (**b**) Representative examples of WT or PRE mutant 1st leg male flies. The *dac* expression domain is indicated by a red arrow. The black arrow indicates the extra sex comb bristles on tarsal segment 2 (TS2). Overexpression of *dac* in the TS2 segment partially transforms this more distal segment TS2 into the more proximal segment TS1 induces the ectopic sex comb teethes.

https://doi.org/10.1038/s41594-024-01375-7



Extended Data Fig. 2 | Hi-M library and Pearson correlations of individual replicates. (**a**) Schematic representation of the genomic positions of the barcodes of the *dac* gene locus. Violet bars indicate the positions of PRE1 (barcode 29) and PRE2 (barcode 48). Salmon bar indicates the position of the leg enhancer (RE), green bar indicates the position of the gypsy 2 insertion and gray bar indicates the position of the CG5888 promoter. (**b**) Scatter plots and Pearson correlations of two biological replicates of wild type pupal leg discs for all tarsal segments and specific tarsal segments (TS1, TS2 and TS3/4). (**c**) Scatter plots and Pearson correlations of two biological replicates of gypsy 2 mutant pupal leg discs for all tarsal segments and specific tarsal segments (TS1, TS2 and TS3/4).

https://doi.org/10.1038/s41594-024-01375-7



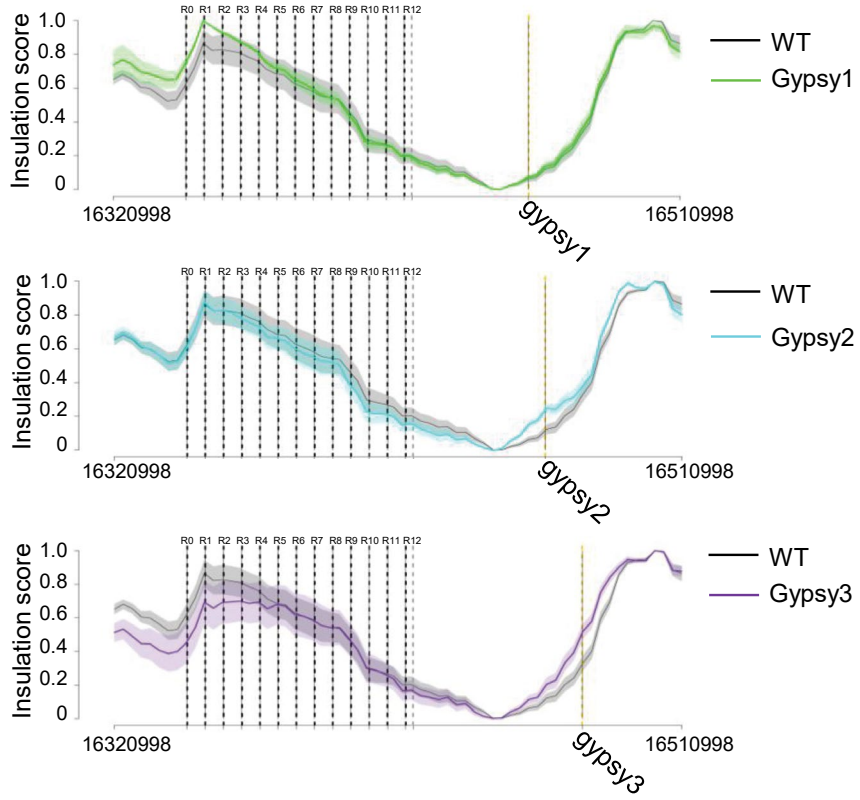
Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41594-024-01375-7

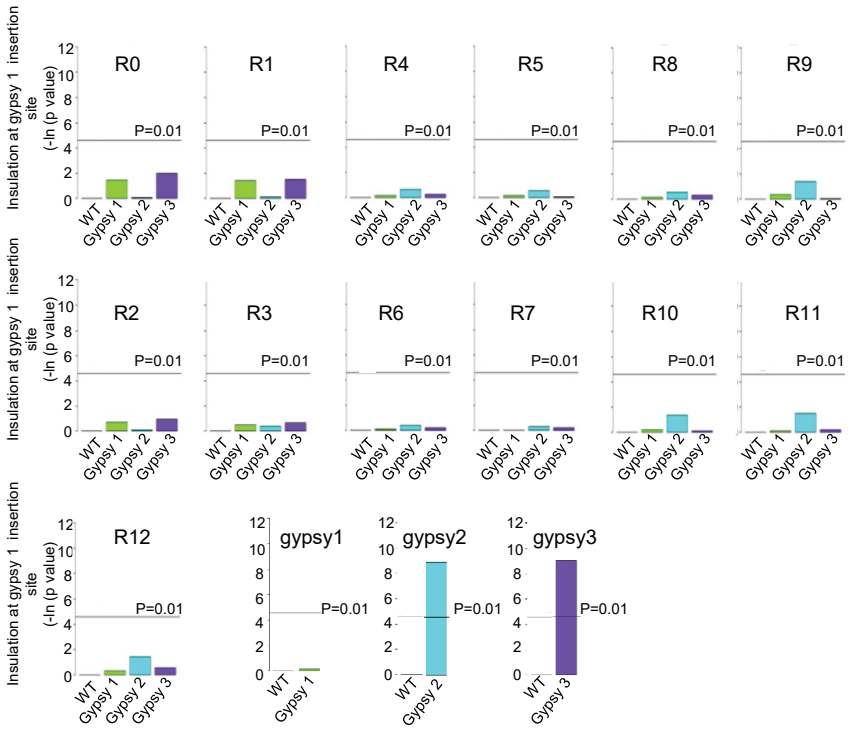
Extended Data Fig. 3 | Insulation activity of gypsy sequences at the *dac* TAD. (**a**) qChIP experiments using Su(HW) antibodies of WT embryos or embryos carrying the gypsy 2 or gypsy 3 insertion. D1-D8, PRE1, 2 and gypsy indicate PCR amplicons used for qChIP experiments along the *dac* gene locus (see Supplementary Table 2). The housekeeping gene Rp49 was used as negative control. Two independent replicates have been performed. Data are presented as the mean values ± s.d (error bars) of replicates. (**b**) CUT&RUN profiles for H3K27me3 at the *dac* gene locus, performed in 3rd instar imaginal leg discs of the indicated fly lines. Red bar indicates position of the RE. Violet bars indicate positions of PREs. Green bar position of the gypsy insulator. Grey bar below H3K27me3 tracks demarcates the PcG TAD[22]. (**c**) The scatter plots show the H3K27me3 enrichment in the 131 *Drosophila* Polycomb domains in gypsy 1 (left), gypsy 2 (middle) and gypsy 3 (right) as a function of WT flies (log2 scale). The *dac* Polycomb domain is shown in blue.

https://doi.org/10.1038/s41594-024-01375-7

a



b

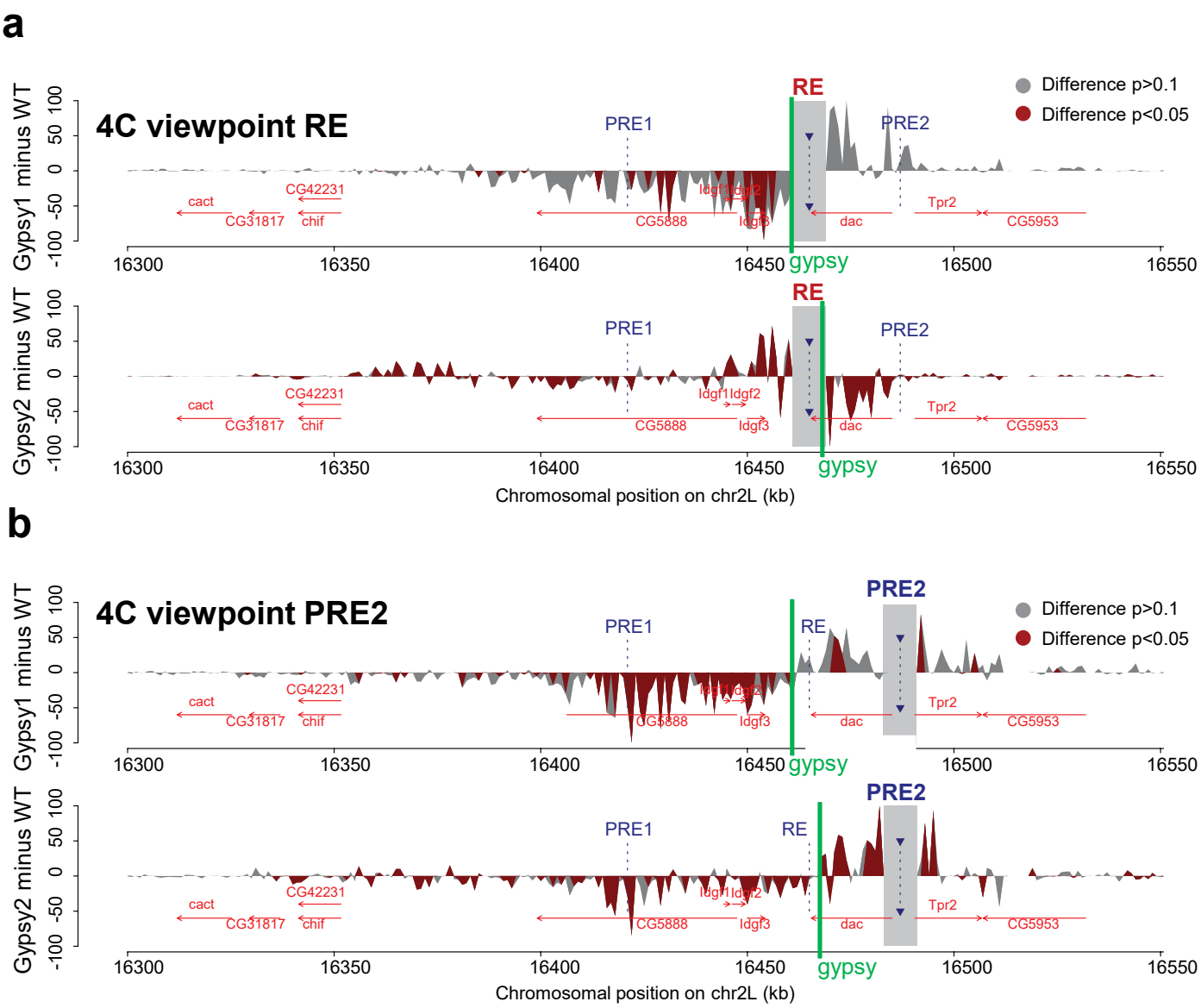


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41594-024-01375-7

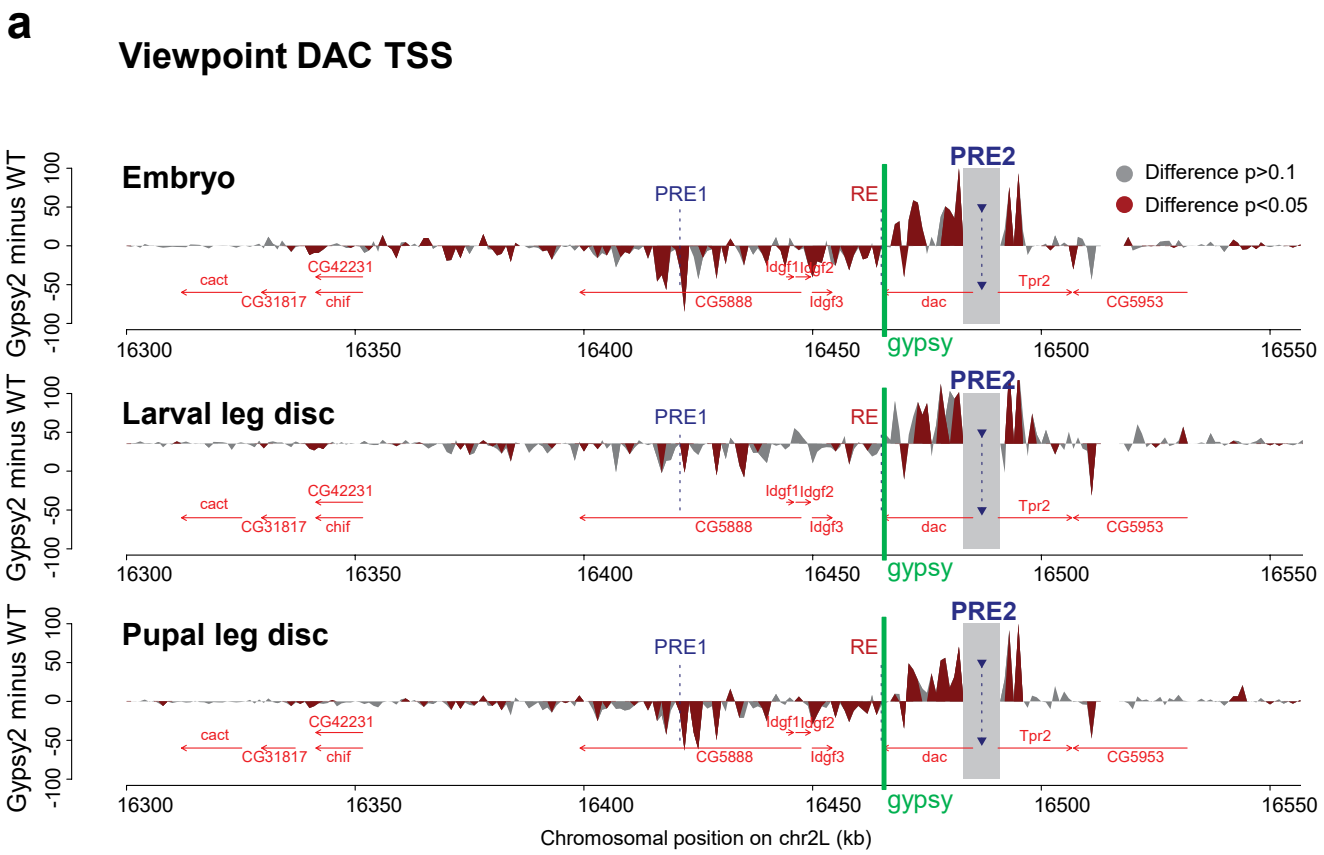
Extended Data Fig. 4 | Statistical analysis of the insulation changes at the *dac* domain in the gypsy lines. (**a**) The insulation profile per each condition is shown at 3kb resolution as the mean value (line) +/− the standard deviation (shaded area) over the insulation scores computed using 5 different values of the window parameter (w=100, 150, 200, 250, and 300 kb). Each panel show the WT together with one gypsy line at the time: Gypsy1, Gypsy2, and Gypsy3 respectively. These panels also show the locations of the 13 2kb-regions (R0-R12) spaced by 4 kb, spanning the domain from the TAD boundary to the PRE1 (chr2L:16,348,000-16,422,000) on which the differential statistical analysis has been applied. The *p-values* resulted from a Welch t-test between the WT condition and each of the gypsy mutant at the corresponding locus. (**b**) Bar-plots of the statistical analysis applied to the R0-R12 regions and the gypsy insertion sites. We notice that the only statistically significant increases of insulation are quantified at the Gypsy 2 and Gypsy 3 insertion sites in the corresponding lines.

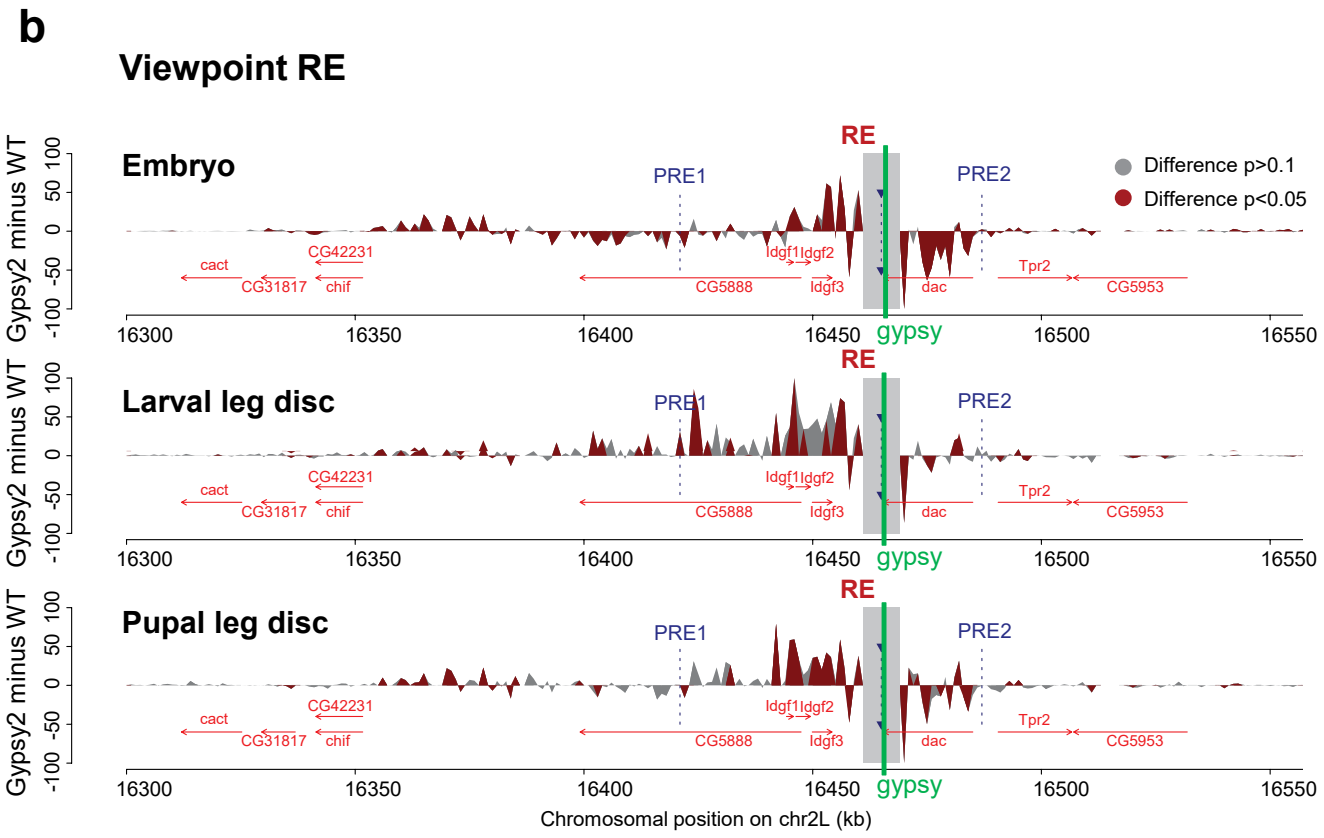
https://doi.org/10.1038/s41594-024-01375-7



Extended Data Fig. 5 | Gypsy 1 insertion has insulation activity and reduces PRE looping. (**a**) Differential 4C-seq profiles at the *dac* gene locus of the indicated fly lines using the leg enhancer (RE) as a viewpoint (gray bar). (**b**) Differential 4C-seq profiles at the *dac* gene locus of the indicated fly lines using the PRE2 as a viewpoint (gray bar). Reads of 3 replicates were aligned and pooled into 1 kb bins. Mean values of the indicated mutant line are subtracted from the control line and the calculated counts per million are plotted. Green bar indicates position of the gypsy insulator. Position of PREs is indicated by blue dotted lines.

https://doi.org/10.1038/s41594-024-01375-7



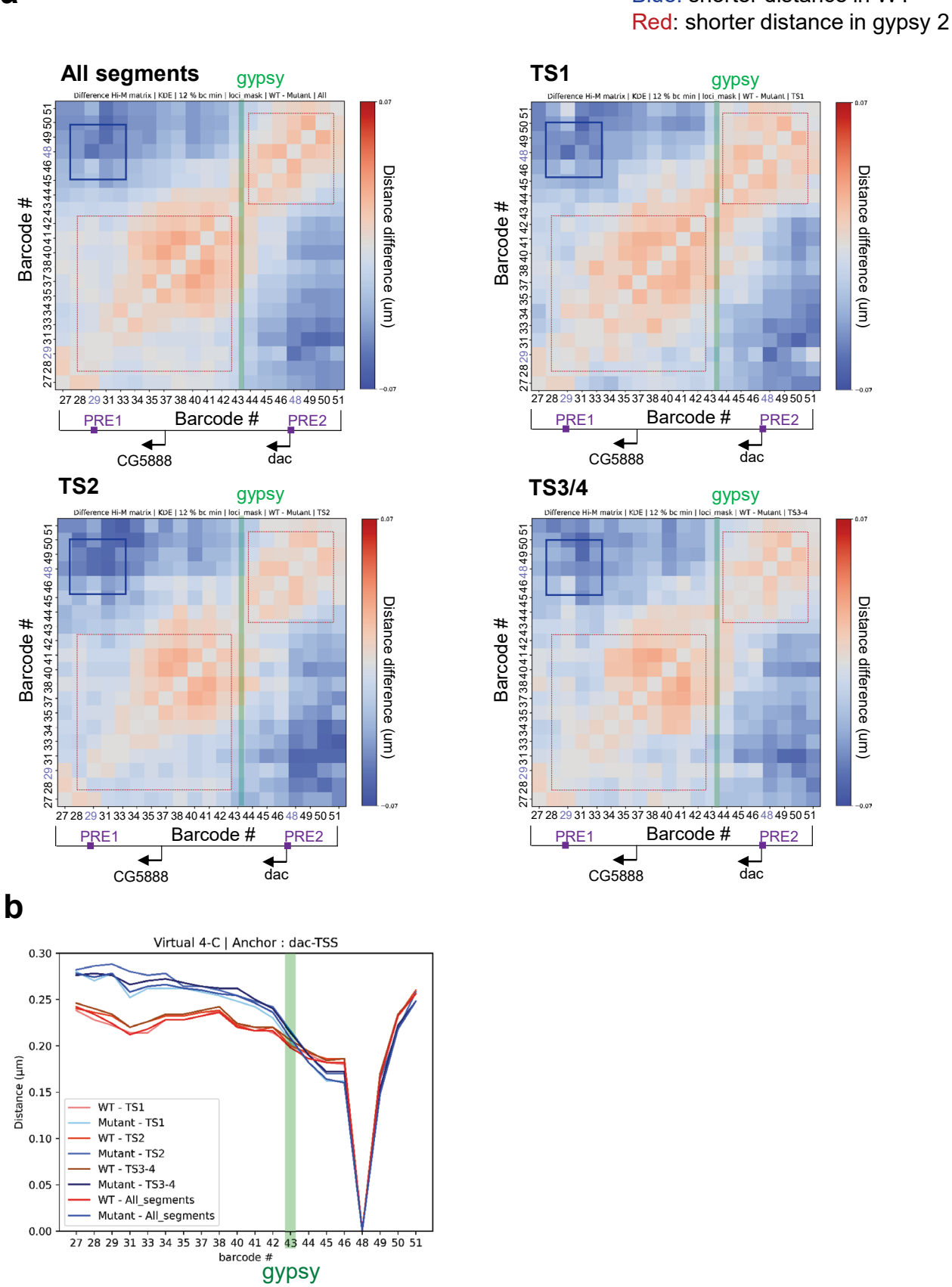


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41594-024-01375-7

Extended Data Fig. 6 | Developmental stage specific consequences of gypsy 2 insertion for the *dac* Polycomb domain architecture analyzed by 4C-seq. (**a**) Differential 4C-seq profiles at the *dac* gene locus of gypsy 2 versus WT fly lines using the dac promoter (PRE2) as a viewpoint (gray bar) at the indicated developmental stages. (**b**) Differential 4C-seq profiles at the *dac* gene locus of gypsy 2 versus WT fly lines using the leg enhancer (RE) as a viewpoint (gray bar) at the indicated developmental stages. Reads of 3 replicates were aligned and pooled into 1 kb bins. Mean values of the indicated mutant line are subtracted from the control line and the calculated counts per million are plotted. Green bar indicates position of the gypsy insulator. Position of PREs is indicated by blue dotted lines.

https://doi.org/10.1038/s41594-024-01375-7

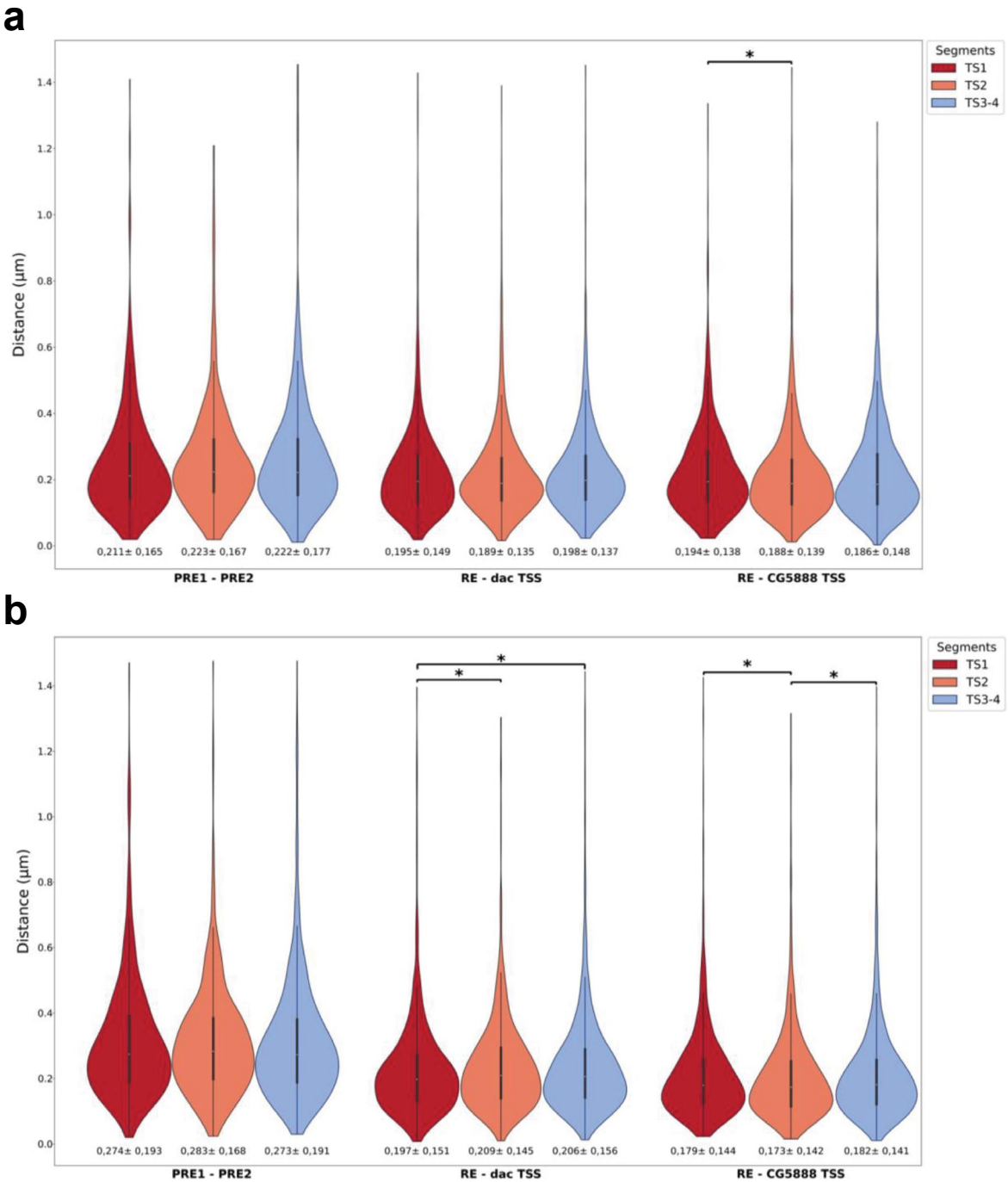


Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s41594-024-01375-7

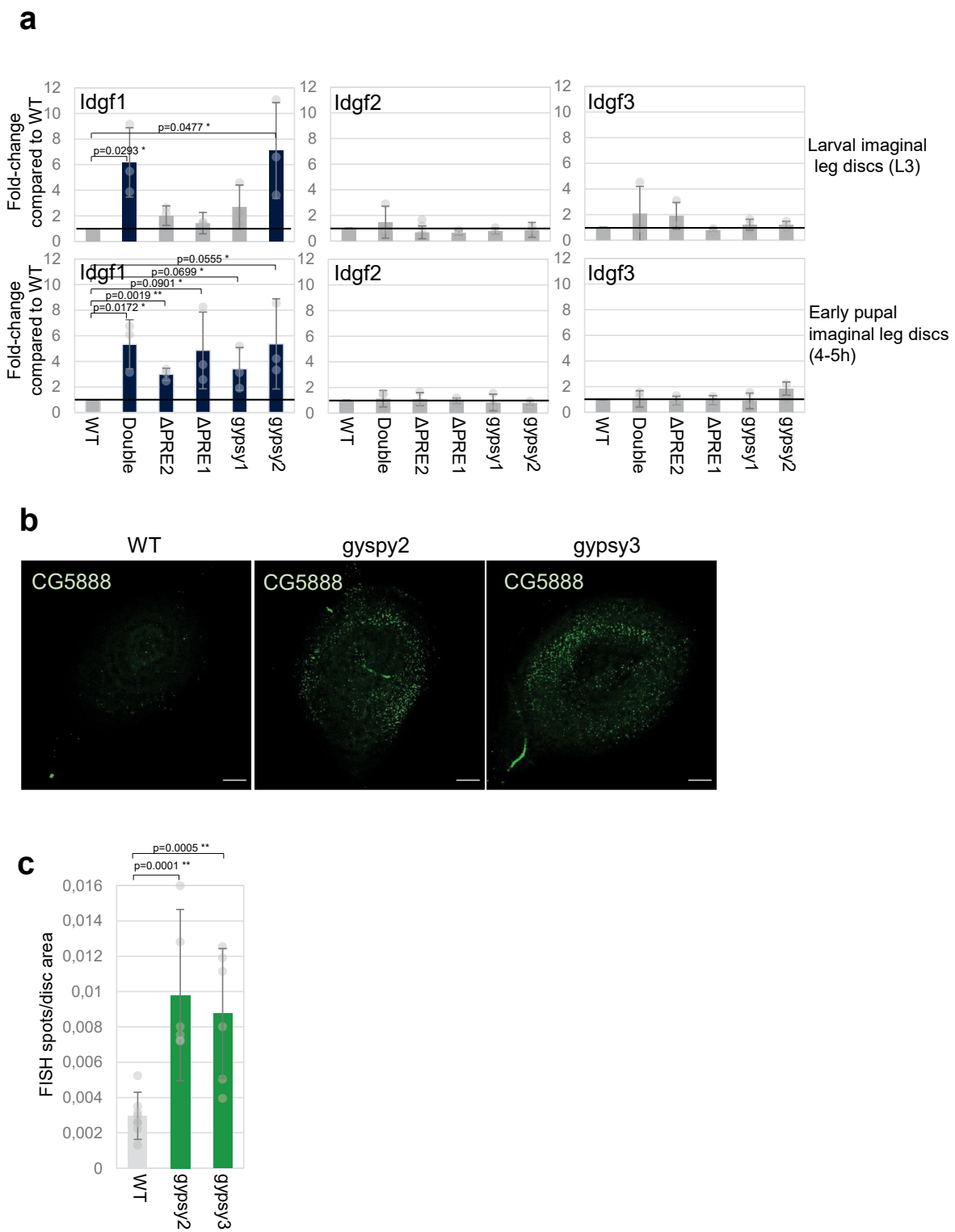
Extended Data Fig. 7 | Consequences of gypsy 2 insertion on chromatin topology observed by Hi-M experiments. (**a**) Ensemble Hi-M PWD difference matrix of WT and gypsy 2 mutant for all tarsal segments, TS1, TS2 and TS3/4 of pupal leg discs. Blue indicates shorter distances in WT and red in gypsy2 mutant condition. Green bar indicates the position of the gypsy insulator. Blue square indicates loss of long-range interactions around the position of the PRE1-PRE2 loop upon gypsy insertion. Red dashed square indicates increased short-range interactions up- and downstream the gypsy insulator insertion. (**b**) Virtual 4C plots derived from Hi-M PWD matrices using barcode 48 (*dac* TSS) as viewpoint. Green bar indicates the gypsy insulator insertion. Red tracks correspond to WT and blue tracks to gypsy 2 mutant condition.

https://doi.org/10.1038/s41594-024-01375-7



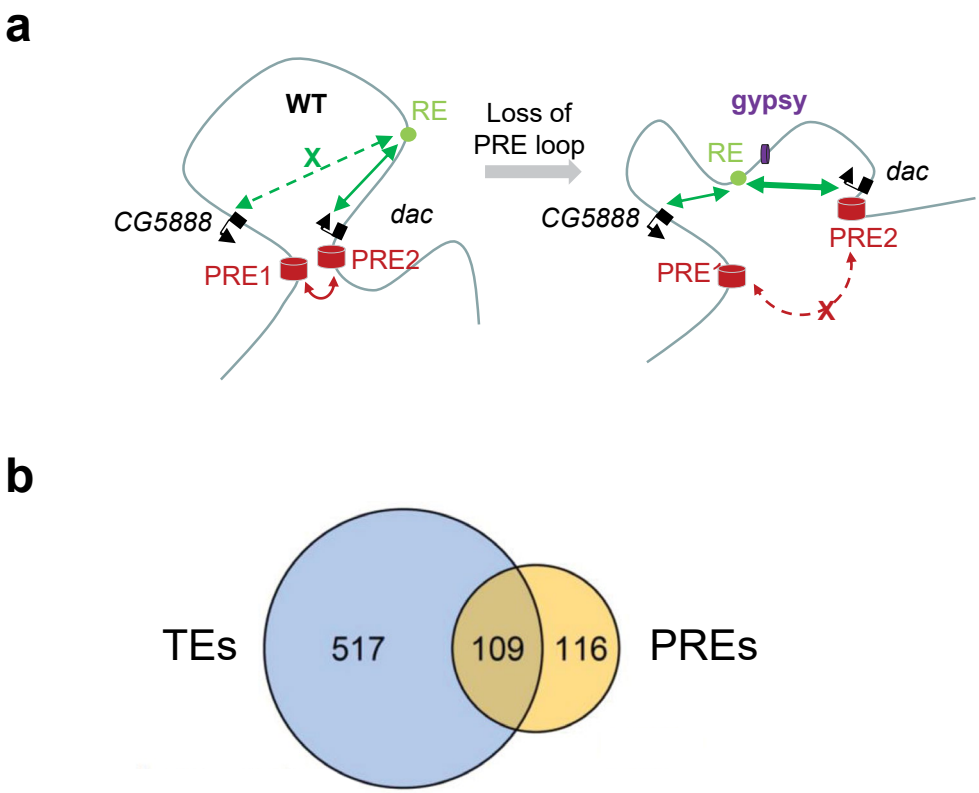
Extended Data Fig. 8 | Cell to cell variability of PRE-PRE or E-P distances measured by Hi-M. (a) Distances distributions between PRE1 and PRE2, RE and *dac* TSS and RE and *CG5888* TSS. The distances distributions were extracted for TS1 (red), TS2 (orange) and TS3/4 (blue) segments of the WT condition. Asterisk represent the p-value obtained with the Wilcoxon two-sided rank test: p-value [*] < 0.05 (rejected null hypothesis). **(b)** Distances distributions between PRE1 and PRE2, RE and *dac* TSS and RE and *CG5888* TSS. The distances distributions were extracted for TS1 (red), TS2 (orange) and TS3/4 (blue) segments of the gypsy 2 condition. Asterisk represent the p-value obtained with the Wilcoxon two-sided rank test: p-value [*] < 0.05 (rejected null hypothesis). White box represents the median, the thick black line represents the first and third quartiles, and the red/orange/blue areas represent the kernel density estimation of the distribution.

https://doi.org/10.1038/s41594-024-01375-7



Extended Data Fig. 9 | Importance of the RE for the activation of the CG5888 gene upon loss of the PRE loop. (**a**) RT-qPCR analysis at the indicated developmental stages using primers specific for *Idgf1, 2* and *3* genes. The fold change between the indicated mutant and wild type (control) line is shown. Three independent experiments have been performed. Data are presented as the mean values ± s.d (error bars) of replicates. * indicates p-value < 0.1, ** indicates p-value < 0.01 (two-sided unpaired t-test). (**b**) RNA FISH images of WT, gypsy2 and gypsy3 insertion, in 3rd instar imaginal leg discs (*CG5888* gene: green). White bar indicates 30 micrometers. (**c**) Quantification of CG5888 RNA FISH signals in WT and gypsy2 and gypsy3 imaginal leg discs. The number of CG5888 FISH signals has been quantified and normalized by the surface of the total disc area. Y axis represents the average of 6 imaginal discs values. Error bars indicate s.d. ** indicates p-value < 0.01 (two-sided unpaired t-test).

https://doi.org/10.1038/s41594-024-01375-7



Extended Data Fig. 10 | Function of PRE loop on gene expression. (**a**) A Model illustrating the architectural consequences of gypsy insulator insertion on PRE looping and RE communication with gene promoters. (**b**) Overlap of Tethering Elements (TEs) identified[15] in and embryonic PREs from[47].

nature portfolio

Corresponding author(s): Bernd Schuettengruber and Giacomo Cavalli

Last updated by author(s): Jun 21, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

All immunostaining and RNA FISH images were taken using a Leica SP8 confocal microscope.
Hi-M experiments were performed on Drosophila Tissue with a homemade widefield and epifluorescence microscope.
Data were collected using qudi-HiM, our open-source package for data acquisition available at https://github.com/NollmannLab/qudi-HiM
HiC was performed on Drosophila Tissue using the EpiTect Hi-C Kit, sequencing was performed by BGI.
4C-seq was performed on Drosophila Tissue using an in house protocol
RT-qPCR were performed using LightCycler480.

Data analysis

Immunostaining and RNA FISH were analysed using ImageJ2 win 64/Fiji.

RT-qPCR were analysed using LightCycler software.

Hi-M data were acquired using qudi-HiM, available at: https://github.com/NollmannLab/qudi-HiM, and at https://zenodo.org/record/6379944 (DOI: 10.5281/zenodo.6379944)
Hi-M data were analyzed using pyHiM release 0.6, available at (https://pyhim.readthedocs.io/en/latest/) (https://www.biorxiv.org/content/10.1101/2023.09.19.558412v1).
Barcodes and fiducials were segmented in 3D using a neural network, followed by 3D localisation of the center of each barcode mask (https://www.biorxiv.org/content/10.1101/2023.09.19.558412v1).

Hi-C raw data from Hi-C sequencing were processed by using the "shHiC2" pipeline.
Valid interactions were stored in a database using the “misha” R package (https://github.com/msauria/misha-package).
Extracting the valid interactions from the misha database, the "shaman" R package [https://bitbucket.org/tanaylab/shaman] has been used

for computing the Hi-C expected models, Hi-C scores with parameters k=250 and k_exp=500,and differential Hi-C interaction scores with parameters k=250 and k_exp=250 and per each comparison down-sampling the compared datasets to have the same number of valid-pairs in chr2L. The insulation score was computed on the observed Hi-C datasets binned at 2kb resolution with windows of 100 kb and stored in the misha database using an in-house R script. All plots of Hi-C maps, Hi-C interaction scores comparisons, insulation score (IS) profiles with window 100kb, p-values of IS comparisons (Fig. 4d) were obtained with in-house R scripts

4C-seq:
Using a custom-made python script, fastq sequencing files were split using 4C primer sequences to obtain individual fastq files only containing reads from a single viewpoint per genotype and tissue type. Thereby, the reads were trimmed to remove viewpoint sequences up to the restriction sites. Subsequently, the trimmed reads were aligned against the DM6 reference assembly using bowtie with the parameters -a -v 0 -m 1 (no mismatches and no multiple alignments allowed). The aligned reads were mapped to restriction fragments and genomic bins of 1kb size using HiCdat to obtain tabular files describing the number of reads (i.e. contact frequencies) for a given fragment or genomic bin, respectively. All subsequent analysis steps have been conducted using R.

Cut&Run:
Fastq files were aligned to the D. melanogaster reference genome dm6 using Bowtie 2 (v 2.4.2). Duplicate reads were removed using sambamba markdup (v 1.0.0); For visualization, replicates were merged using samtools merge with default parameters and reads per kilo base per million mapped reads (RPKM)-normalized bigWig binary files were generated using the bamCoverage (v 3.5.5) function from deepTools2 Full custGenome browser plots were generated using the pyGenomeTracks package (v 3.8)om code will be made available upon publication on a dedicated, public (Github) repository. The 131 Drosophila Polycomb domains 22 were used for the differential enrichment analysis using the DESeq2 method from the "DiffBind" R package (v 3.12.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All genomic data produced for this manuscript have been deposited on GEO repository under the public accession number GSE247377.
Oligo Paint sequences were obtained from the oligopaint public database (http://genetics.med.harvard.edu/oligopaints).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was defined in compliance with the gold standards of the field, such that relevant statistical parameters (mean, median...) would get stabilized. Cut&Run, 4C-seq were performed in duplicates, following encode's standards (https://www.encodeproject.org/chip-seq/

transcription_factor/#standards). qRT-PCR analysis have been performed in Triplicates. For sample size of Immunostaining experiments, RNA-FISH experiments see coresponding Figure Legends.

Data exclusions	Hi-C experiments were performed genome wide. No data were excluded and all are available on GEO. As our locus of interest for this paper - the dac locus - is located on chromosomee 2L, the analysis of Hi-C was done specifically on reads mapping to the 2L chromosome.
Replication	All experiments were performed several times on different days, and only consistent observations were reported (see replicates section for further details regarding biological replicates). qRT-PCR analysis have been performed in Triplicates. RNA FISH and Immunostaining analysis was performed in Duplicates. For sample size of Immunostaining experiments, RNA-FISH experiments see coresponding Figure Legends. qChIp was performed in Duplicates. Hi-M Cut&Run and 4C-seq experiments were performed in duplicates.
Randomization	For all experiments, flies or fly tissues were randomly collected by experimenters who did not know what phenotype or result was corresponding to the collected fly at the time of collection.
Blinding	Blinding is not applicable here, since data anonymization is not compatible with quality controls and the identity of control samples must be known in order to perform genomic data analysis. However, reported data are based on unbiased analysis avoiding confirmation bias and/or subjective assumption of qualitative data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☐	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Antibodies are described in the Methods section of the manuscript.

The following primary antibodies were used (Antibody, dilution, Provider, Catalogue number) for immunostaining:
-Dachshund (DAC), 1:400, (DSHB clone number: mAbdac1-1)

The following secondary antibodies were used (Antibody, dilution, Provider, Catalogue number) for immunostaining:
- Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647 1:1000, (Invitrogen, A-31571)

The following antibodies were used (Antibody, dilution, Provider, Catalogue number) for qChIP:
-Su(Hw), (1:100) gift from Victor Corces, Antibody is not commercially available

The following antibodies were used (Antibody, dilution, Provider, Catalogue number) for CUT&RUN:
- H3K27me3, 1:100, Active motif, 39155
- IgG, 1:100, Cell Signaling Technology, 2729S

Validation

Commercial antibodies are validated for the use of immunofluorescence and CUT&RUN. Data are available on the manufacturer's website.

Su(Hw) antibodies have been validated by previous paper(Antibody, Validation and references):
Gerasimova et al., 1995

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Drosophila strains were used for the study, as described in the Methods section. Fly lines created by CRISPR/Cas9 used in this study are: Double, ΔPRE1, ΔPRE2 and gypsy 1, gypsy2, gypsy3, deltaRE, gypsy2/deltaRE. Age of larvae were 9 to 11 days, Age of pupae were 11-13 days, Adult flies used to score the ESC were 3 days old.

Wild animals	The study did not involve wild animals
Reporting on sex	For PRE mutant phenotype, only male flies were scored, since female flies show no detectable phenotype.
Field-collected samples	The study did not involve samples collected from the field
Ethics oversight	This study was perfomed under the ethical approval N. n6906C2 of the MINISTERE DE L'ENSEIGNEMENT SUPERIEUR DE LA RECHERCHE ET DE L'INNOVATION, issued on April 8, 2020. Drosophila strains are reported in the manuscript and the methods section.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A