Structure and function of an ectopic Polycomb chromatin domain

Sandip De, Yuzhong Cheng, Ming-an Sun, Natalie D. Gehred, Judith A. Kassis*

Polycomb group proteins (PcGs) drive target gene repression and form large chromatin domains. In *Drosophila*, DNA elements known as Polycomb group response elements (PREs) recruit PcGs to the DNA. We have shown that, within the *inverted-engrailed (inv-en)* Polycomb domain, strong, constitutive PREs are dispensable for Polycomb domain structure and function. We suggest that the endogenous chromosomal location imparts stability to this Polycomb domain. To test this possibility, a 79-kb *en* transgene was inserted into other chromosomal locations. This transgene is functional and forms a Polycomb domain. The spreading of the H3K27me3 repressive mark, characteristic of PcG domains, varies depending on the chromatin context of the transgene. Unlike at the endogenous locus, deletion of the strong, constitutive PREs from the transgene leads to both loss- and gain-of function phenotypes, demonstrating the important role of these regulatory elements. Our data show that chromatin context plays an important role in Polycomb domain structure and function.

INTRODUCTION

Polycomb group proteins (PcGs) are critical for organismal development and stem cell maintenance (1). PcGs were first found in *Drosophila* as repressors of homeotic genes, and PcG repression is one of the earliest epigenetic regulatory mechanisms to be identified. In *Drosophila,* nearly all PcG proteins are subunits of one of four principal protein complexes: Polycomb repressive complexes 1 and 2 (PRC1 and PRC2), Pho repressive complex (PhoRC), and Polycomb repressive deubiquitinase (PR-DUB) (2). PcG protein complexes bind to DNA elements known as Polycomb group response elements (PREs), deposit the repressive chromatin modification mark H3K27me3, and drive chromatin compaction leading to gene repression.

Genes repressed by PcG are covered with H3K27me3 and are thought to form their own topologically associating domains (TADs) (3, 4). In *Drosophila,* TADs vary from a few kilobases to several hundred kilobases in size. Regulatory DNAs present within TADs preferentially interact with genes located within the same domain, with limited contacts outside of the TAD boundaries. In *Drosophila,* genome-wide data suggest that mini-domains formed by actively transcribed regions form the boundaries of some TADs (5), while other TAD boundaries are demarcated by insulator elements (6, 7). In mammals, the insulator protein CTCF colocalizes with a subset of TAD boundaries (4). The current understanding in chromatin biology hypothesizes that the folding of chromatin into domains assists in the packaging of long stretches of DNA inside the eukaryotic nucleus. Further, it is the organization of these domains that facilitates spatial and temporal regulation of genes within them. Thus, understanding how the domains are formed is of prime importance.

How large PcG domains/TADs are formed is a central question. To date, researchers have extensively studied how PcG proteins are recruited to specific DNA sequences and which proteins are present in PcG complexes. PREs are required for the recruitment of PcG proteins in *Drosophila* and are thought to initiate the formation of a Polycomb domain. We have been studying the 113-kb *PcG* domain that encompasses the *inverted (inv)* and *engrailed (en)* genes of *Drosophila.* Unexpectedly, deletion of strong, constitutive PREs from the endogenous *inv-en* domain had little effect on *inv-en* PcG domain organization (8). Weak PREs present in the *inv-en* domain were sufficient to establish and maintain the overall domain organization (8), and some weak PREs overlap with enhancers present within the *inv-en* domain (9, 10). Similarly, deletion of the *bxd* PRE had a mild effect on *Ubx* expression, whereas deletion of the *iab7* PRE resulted in misexpression of *Abd-B* in very specific parasegments (11, 12). A recent report showed that deletion of two PREs from the *dac* locus causes prominent structural and functional changes in the locus (13).

*Functional studies of PREs in transgenes have shown that PRE activity is highly influenced by the chromosomal insertion sites (2). Some studies have shown that a PRE can act in cooperation with other flanking regulatory DNA as a positive or negative regulatory element depending on the chromatin context (14), which emphasizes the fact that chromatin context contributes to PcG recruitment and function. The effect of chromatin context has been intensely investigated in *Drosophila* with respect to heterochromatin formation. Genes that are juxtaposed to heterochromatin, either by rearrangement or by transposition, show a variegated phenotype. This is a result of the gene being silenced in some of the cells in which it is normally active. Chromatin context can also influence an enhancer’s activity (15, 16). Recently, Wijchers et al. (17), by artificially recruiting chromatin regulators to specific loci, showed that chromatin context can heavily influence genomic contacts. How chromatin context can influence Polycomb domain formation has not yet been investigated.

Here, we report the effect of different chromatin contexts on the formation and function of a Polycomb domain in *Drosophila.* We inserted a 79-kb transgene containing the *en* gene in three different locations of the *Drosophila* genome. The *en* gene encodes a homeodomain-containing protein that is a key developmental regulator in *Drosophila,* necessary for embryonic segmentation and development of the posterior compartment of imaginal discs. The *en* gene resides in an ~113-kb Polycomb domain alongside another homeodomain gene, *inv.* Unlike *en,* the *inv* gene is not required for development (18). We used chromatin immunoprecipitation sequencing (ChIP-seq) and circular chromosome conformation capture sequencing (4C-seq) techniques to characterize the Polycomb domain at the ectopic sites. We found that the PcG domain was formed efficiently at two of these sites and that there was differential spreading of the repressive H3K27me3 mark to the region flanking the insertion

SCIENCE ADVANCES | RESEARCH ARTICLE

S1, A and B). Examination of the H3K27me3 ChIP-seq profile in some 3R, 1 kb before the 3′ CG2781 end of [3R:3808570, R5] (fig. ′ end of the 3′ domain, 4.7 kb before chromosome has broken within the 3′ end of the 3′ inv-en domain, from the 3′ end of tou. In contrast, H3K36me3, a mark of actively transcribed genes, is strong over E(Pc) and tou but undetectable over the inv and en transcription units (Fig. 1A).

What stops the spreading of the H3K27me3 mark? Genome-wide ChIP-chip and ChIP-seq data in many different cell types fail to detect high levels of insulator proteins at the ends of the inv-en domain and other PcG domains (5). Instead, we and others suggest that it is the presence of the actively transcribed genes, in this case the ubiquitously transcribed E(Pc) and tou genes, that stops further spreading of H3K27me3 (Fig. 1A). As an initial test of this hypothesis, we examined the distribution of the H3K27me3 over the mut-en region (T(2;3)en80F) (19). In this x-ray–induced mutation, the second chromosome has broken within the inv-en domain, 4.7 kb before the 3′ end of the tou gene (2R:7461660, R5), and joined to chromosome 3R, 1 kb before the 3′ end of CG2781 [3R:3808570, R5] (fig. S1, A and B). Examination of the H3K27me3 ChIP-seq profile in T(2;3)en80F/Df(2R)enX31 (which contains a deletion of the entire inv-en region and flanking DNA) revealed that H3K27me3 spreads for about 16 kb over CG2781 but falls off at CG14463, an actively transcribed gene covered with H3K36me3 (fig. S1, B and C). These data are consistent with the model that PRC2 is recruited to PREs and methylates histone H3 in flanking nucleosomes until it encounters actively transcribed genes.

A 79-kb HA-en transgene rescues inv-en double mutants PREs can be identified in the genome via ChIP-seq as PcG protein binding DNA fragments within H3K27me3 domains. Within the inv-en domain, there are four strong, constitutive PREs found in all cell lines, all tissues, and all stages of development. These are identified by red stars in Fig. 1A. In addition, there are numerous smaller peaks—some stage specific—that we have shown to be functional PREs (Fig. 1A, green stars) (8). 4C-seq analysis reveals that these small peaks interact with the en transcription unit (8). In the following set of experiments, we sought to answer three questions: (i) Can a large en transgene set up a functional en gene with a typical H3K27me3 domain at an ectopic site in the genome? (ii) Would the structure of the ectopic H3K27me3 domain, assayed by PcG binding and 4C interactions, look like the endogenous locus? (iii) Would the H3K27me3 mark spread into flanking DNA? To answer these questions, we used the phi-C31 transgene system to insert a 79-kb HA-en transgene into the genome with chromosomal attP landing sites (10).

The 79-kb HA-en transgene, extending from within an inv intron to the 3′ end of the tou gene (Fig. 1B), rescues inv-en double mutants into viable, fertile adults (10). inv and en are coexpressed and functionally redundant. The inv gene is dispensable in the laboratory (18). Inserting this transgene into a number of different chromosomal insertion sites allows us to determine the effect of flanking chromatin on the formation and function of the H3K27me3 domain. We reasoned that inserting an active HA-en gene might not allow for the recovery of transgenic lines in which HA-En is misexpressed. In addition, we might not recover lines where the regulatory DNA present in the En transgene causes misexpression of the PcG genes. For this reason, we chose three different attP landing sites: attP3, attP40, and VK00033 (fig. S2). The attP3 and attP40 landing sites have been shown to support different levels of gene expression in third instar larvae in a tissue-specific manner, with attP3 being generally repressive except in the nervous system (20). Thus, these insertion sites might lead to differing levels of En expression. We included VK00033 as a landing site because a 21-kb en transgene including an En–EGFP (enhanced green fluorescent protein) fusion protein has been successfully inserted into this site (21). The 21-kb construct contains regulatory DNA for early En embryonic stripes but lacks enhancers for most other aspects of En expression, including larval imaginal disc enhancers (21).

We were successful in obtaining Haen79 transgenic lines with insertions into attP3 and attP40. However, despite numerous attempts, we were not able to obtain insertion into the VK00033 site. To test whether this might be due to misexpression HA-En protein, we introduced a stop codon into the Haen79 gene so that it produced a nonfunctional protein (Haen79stop). We obtained a transgenic line with Haen79stop at VK00033. Haen79stop was expressed outside of the canonical En-expressing region in Haen79stop/VK00033 embryos and imaginal discs (fig. S3, A and B). We hypothesize that misexpression of an active En protein in these tissues would lead to dominant lethality and the inability to recover transgenic lines.

Haen79 transgenes form H3K27me3 domains We wanted to compare the H3K27me3 domain and three-dimensional (3D) structure of the 79-kb HA-en transgenes with that of a similar domain at the endogenous location. For this purpose, we used CRISPR-Cas9 to delete 33 kb of the inv region, creating a 80-kb endogenous en locus (en80), similar to the 79-kb HA-en transgene. We left a 1-kb region at the 3′ end of E(Pc) intact so as to not interfere with the transcript termination of E(Pc). en80 flies are homozygous viable and fertile and express En correctly (fig. S4A). en80 flies hold their wings out, a phenotype common to other mutants that lack inv DNA (8). ChIP-seq experiments on en80 third instar larval brains and imaginal discs showed that H3K27me3 accumulates from the 3′ end of E(Pc) to the 3′ end of tou, similar to the wild-type (WT) inv-en domain (Fig. 1A). Furthermore, the PcG proteins Pho and Ph bind in two large peaks upstream of the en transcription units and in several smaller peaks within the domain (Fig. 1A). The ChIP-seq profiles of H3K27me3, Pho, and Ph in en80 look nearly identical to
those of the intact inv-en domain. We also tested the H3K36me3 level in en80 over the gene E(Pc) and did not observe any significant difference in the level of the histone mark in comparison to the WT (fig. S4B). This shows that the 80 kb present in en80 has all the DNAs necessary to set up a normal H3K27me3 domain and suggests that our HAen79 transgene should be able to do the same.

**Fig. 1. Characterization of en80.** (A) ChIP-seq distribution of Pho, Ph, and H3K27me3 at the inv-en domain in WT (rows 1 to 3), en80 (rows 4 to 6), HAen79@attP40 en3110 (rows 7 to 9), and HAen79@attP3; en3110 (rows 10 to 12) is shown. Inv and en strong, constitutive PREs are indicated by red stars, and weak PcG peaks are indicated by green stars. The region deleted in en80 is highlighted with a crossed box. ChIP-seq signals from transgenes are within the red dashed box. Positions of the genes (navy blue) are shown at the bottom. The arrows indicate the direction of the transcription. The genomic coordinates are shown at the top (FlyBase R5). (B) The extent of genomic DNA present in the HAen79 transgene is shown (2R:7386838–7466000). (C) Zoomed-in view of the en upstream region. The ChIP-seq scale has been readjusted to highlight the weak peaks; H3K27me3 data are not included. At one of the small peaks, Pho and Ph are strongly reduced in size in HAen79@attP40 and HAen79@attP3 in comparison with en80 (red dotted box). All transgenes and en deletions are homozygous.

H3K27me3, Pho, and Ph binding was examined in HAen79@attP40 en\textsuperscript{Δ1.5} and HAen79@attP3; en\textsuperscript{Δ10} larval brains and discs. en\textsuperscript{Δ10} flies contain a 110-kb deletion of the entire inv-en domain, excluding 1 kb at the 3' ends of E(Pc) and tou (fig. S3C). Except for the 1-kb at the 3' ends of tou and E(Pc), the ChIP-seq profiles of the en region seen in Fig. 1A are derived exclusively from the 79-kb HA-en transgenes.

The ChIP-seq results demonstrate that an H3K27me3 domain forms over the transgene at both chromosomal insertion sites (Fig. 1A). Further, Pho and Ph binding in the transgenic regions looks similar to that of the sites (Fig. 1A). Further, Pho and Ph binding in the transgene at both chromosomal insertion regions seen in Fig. 1A are derived exclusively from the 79-kb E(Pc) and en tou E(Pc) domain, excluding inv-en flies contain a 110-kb deletion of the entire E(Pc) domain, excluding inv-en genes and either E(Pc) or tou, as these two genes do not flank the transgenes at the ectopic locations (fig. S5B).

We next examined the 3D structure of these H3K27me3 domains via 4C-seq. As bait, we used the DNA within the strong, constitutive PREs just upstream of the en transcription unit. Because PREs are known to interact with each other, these interactions may contribute to the structure of the H3K27me3 domain. Comparison of the 4C interaction data between the WT inv-en domain and the en80 locus shows nearly identical structures; most interactions occur within the H3K27me3 domain, but weak interactions extend into the flanking genes (fig. S5A). The 4C-seq data from the HAen79 transgenes inserted at attP3 and attP40 show similar interactions within the transgene, although the levels of some interaction peaks appear diminished (fig. S5B, red arrowheads indicate weaker interactions). As expected, there are no interactions between the transgenes and either E(Pc) or tou, as these two genes do not flank the transgenes at the ectopic locations (fig. S5B).

We next examined the spreading of H3K27me3 and the 4C interactions of the 79-kb HA-en transgenes with DNA flanking the attP40 and attP3 insertion sites. Neither attP40 nor attP3 is covered by H3K27me3 in WT larvae (fig. S2; also see en80 data in Fig. 2). The attP40 insertion site is in a gene-rich region that contains a number of actively transcribed genes with Pho and Ph weakly bound at their promoters (note that scale is the same as that in Fig. 1C). At attP40, the H3K27me3 mark spreads to flanking genes in a manner largely associated with exons (Fig. 2A). This is prominent over the large exon of the Msp300 gene, located almost 15 kb away from the attP40 site. In Kc167 cells, H3K27me3 is spread over the Msp300 in an almost identical manner (http://chorogenome.ie-freiburg.mpmp.de:5001/#browser/2L:5,059,845-5,157,052) (7). Thus, while H3K27me3 spreads over Msp300 in some cell types, it is not spread over Msp300 in WT larval brains and discs (our samples). The presence of the HAen79 transgene at attP40 causes H3K27me3 to form over Msp300. The attP3 landing site is in a nontranscribed region (“null” chromatin) in WT tissues and Kc167 cells (http://chorogenome.ie-freiburg.mpmp.de:5001/#browser/X:20,120,880-20,282,321) (7). In this case, a low level of H3K27me3 spreads outward from the HAen79 transgene insertion site over about a 100-kb region (Fig. 2B).

We also examined the interactions of the flanking genomic DNA with the enPREs present in the transgene via 4C-seq. At attP40, strong interactions are seen in an ~20-kb region flanking the insertion site, but the strength of the interactions drops at genes covered with H3K36me3 (Fig. 3A). The presence of both H3K27me3 and H3K36me3 over Msp300 suggests that this gene is expressed in some cells and repressed in others in our larval samples. Similar to that at attP40, with HAen79 inserted at the attP3 site, the 4C-seq interaction profile with enPREs is greatly reduced at flanking H3K36me3 domains; because it is in a nontranscribed region, the interactions extend over an ~130-kb region (Fig. 3B). We also looked at the mini-white (w) and mini-yellow (y) marker genes, present in the transgene and at the attP sites, respectively, and observed the presence of H3K27me3 and 4C-seq signal over these genes (fig. S6). Further, we performed 4C-seq in the breakpoint mutant T(2;3)en\textsuperscript{ES}/en\textsuperscript{X31} mentioned previously. As expected, the inv-en domain was observed to interact with the CG2781 gene (fig. S7A), while no interactions were observed over the tou gene in T(2;3)en\textsuperscript{ES}/en\textsuperscript{X31} (fig. S7B), which no longer abuts the en DNA. Similar to the transgenes, we also observed that interactions between the enPREs and genes near the breakpoint decrease at transcribed genes (fig. S7A). Together, these data suggest that H3K36me3 reduces interactions between PREs and flanking DNA.

The endogenous en80 domain is resilient to loss of strong PREs

We have previously shown that deletion of the strong, constitutive PREs from the endogenous en locus does not disrupt the formation of the H3K27me3 domain or the regulation of En expression (8). Is this true for the endogenous 80-kb en domain (en80) or the 79-kb transgenes? Using CRISPR-Cas9, we created a chromosome that contained a 1.5-kb deletion of the strong en PREs within en80 (en80\textsuperscript{Δ1.5}). en80\textsuperscript{Δ1.5} contains the same 1.5-kb deletion as that present on the inv\textsuperscript{ES}en80\textsuperscript{Δ1.5} chromosome (8) and also deletes all inv PREs. The en\textsuperscript{Δ1.5} deletion removes both strong PREs upstream of the en transcription unit. Transgenic assays show that both PREs are contained within this 1.5-kb fragment (22). ChIP-seq assays show that some PcG proteins are still associated with the en promoter in flies with the en\textsuperscript{Δ1.5} deletion (8). We do not know whether the promoter-associated PcG proteins are directly recruited to the en promoter or are associated with the promoter via cross-linking with distant PREs. As seen in inv\textsuperscript{ES}en\textsuperscript{Δ1.5} larvae, en80\textsuperscript{Δ1.5} larvae accumulate H3K27me3 over the inv-en domain (fig. S8). Further, these flies are homozygous viable and fertile and have the same cuticular fusion defect in the abdominal midline seen in inv\textsuperscript{ES}en\textsuperscript{Δ1.5} flies (8). In contrast, deletion of the same 1.5-kb fragment from the HAen79 transgene had effects on both the silencing and expression of En, as detailed below. Our data show that the endogenous location of the 80-kb en domain imparts stability to the locus not seen in the 79-kb transgene.

The strong PREs are important for both silencing and activation of the HA-en transgene

The HAen79\textsuperscript{Δ1.5} transgenic construct was created by deleting the 1.5-kb fragment containing the en strong, constitutive PREs from HAen79. We obtained a transgenic line with HAen79\textsuperscript{Δ1.5} inserted into attP40. After numerous injections, we also obtained five females with the putative genotype HA-en79\textsuperscript{Δ1.5}attP3, but these flies were infertile and we were unable to confirm their genotype. Thus, we proceeded with the analysis of flies with HAen79\textsuperscript{Δ1.5}attP40. These flies have a dominant abdominal phenotype; the pigmentation and bristle pattern of the abdomen are severely disrupted in flies with one or two copies of HAen79\textsuperscript{Δ1.5} (Fig. 4) in the presence of WT inv and en. This abdominal phenotype is similar to the phenotype of T(2;3)en\textsuperscript{ES} flies that has been shown to be due to misexpression of En during abdominal formation at the pupal stage (23). Further, a similar phenotype was observed in flies that have one copy of
HAen79@attP40 in combination with a strong, viable mutation in polyhomeotic (ph-p410), a component of PRC1 (Fig. 4A, bottom). This phenotype was not present in ph-p410; attP40 vector males or ph-p410/+; HAen79@attP40 females, as they have a WT copy of ph. We therefore suggest that deletion of the PREs in HAen79∆1.5 leads to misexpression of En in the abdomen during pupal development.

Furthermore, we posit that this is due to loss of PcG regulation in this tissue.

Somatic chromosomes are paired in Drosophila. In a phenomenon known as “transvection,” regulatory DNA on one chromosome can affect the expression of a gene on the homologous chromosome in trans. We reasoned that if the abdominal phenotype of HAen79∆1.5.
flies is due to a loss of Polycomb silencing, we should be able to correct this phenotype by adding a PRE to the homologous chromosome. When inserted at attP40, both the en PREs and PRE_D (from the Ubx gene) could largely correct the abdominal phenotype (Fig. 4B) when put in trans to the HAen79\textsuperscript{4A,5} transgene. In HAen79\textsuperscript{4A,5}/enPREs and PRE_D flies, the abdominal pigmentation and bristle patterns are restored. However, there remains a defect in fusion of the cuticle at the midline in abdominal segments seen most dramatically in the HAen79\textsuperscript{4A,5}/PRE_D flies. In contrast, neither vector alone@attP40 nor the enPREs@attP3 could correct the abdominal pigmentation or bristle pattern defects (Fig. 4, B and C). These data strongly support the hypothesis that the abdominal phenotype is caused by loss of PcG silencing in the abdomen of HAen79\textsuperscript{4A,5} flies.

We were surprised to find that HAen79\textsuperscript{4A,5} could not rescue inv-en double mutants. Therefore, we investigated the expression pattern
of this transgene in embryos. While HA-en from HAen79 is expressed like En throughout embryogenesis, HA-en expression from HAen79∆1.5 is reduced in late embryos, most evident in the nervous system (Fig. 5A). Thus, the 1.5-kb fragment that includes the enPREs, although dispensable at the endogenous locus, is required for the ability of the HAen79 transgene to rescue en-in double mutants.

Although there is reduced HA-En expression (Fig. 5A), no misexpression of HA-En is detected in HAen79∆1.5 embryos or larvae, indicating that the abdominal misexpression is developmental stage and tissue specific. We examined the H3K27me3 level in HAen79∆1.5 larval brains and discs and compared it to the level in HAen79. Because HAen79∆1.5 cannot survive in the absence of endogenous inv-en DNA, we could not assay H3K27me3 over the inv-en region. Instead, we assayed the levels over the y marker gene that is present at the attP40 site and two sites flanking the insertion site. These data show that the level of H3K27me3 accumulated over the HAen79∆1.5 is lower than that of HAen79 (Fig. 5B).

Enhancer interaction with flanking genes at ectopic sites

inv and en share regulatory DNA. Enhancers for inv-en are located up to 50 kb upstream of the en promoter—about 90 kb downstream of the inv promoter (10). Nevertheless, these enhancers do not activate the expression of genes flanking the inv-en domain at the endogenous locus. Some of the inv-en enhancers exhibit promoter specificity and are thus unable to activate flanking promoters, while others are more promiscuous (10, 24). Without the activity of the E(Pc) and tou genes acting as boundaries for the inv-en domain, we wondered whether the genes flanking the HAen79 would be expressed like En. The mini-w gene is present as a reporter gene within the HAen79 construct, located adjacent to the 5′ end of the HAen79 DNA. The mini-y gene is present at the attP40 and attP3 sites, downstream of the mini-w gene. At attP40, both mini-w and mini-y are expressed in subsets of en stripes in embryos (Fig. 6B). Both mini-w and mini-y transcripts become evident at stage 11, much later than endogenous en stripes that first appear at stage 6. Curiously, mini-y
transcripts are predominantly in the thoracic segments, and mini-w is expressed in the abdominal segments. By stage 13, mini-y is hardly detectable but mini-w is evident predominantly in the abdominal segments. Only a subset of en enhancers is able to activate these two reporter genes. We also examined the expression of Msp300 and pGANT-5; neither was expressed like En at any time during embryogenesis (fig. S9A).

In HAen79@attP3 embryos, mini-w expression is nearly identical to that in HAen79@attP40 embryos; however, mini-y is not expressed in stripes at any stage (Fig. 6D). In contrast, CG1504 present just downstream of the HA-en transgene was expressed like En in a transient manner in stage 10 embryos (Fig. 6D). These observed differences in expression of flanking genes during embryonic development might be due to either promoter specificity or local chromatin conformation affecting enhancer availability. Another annotated gene, CG1631, present within the first exon of CG1504, did not show expression like En (fig. S9B). Our experiments suggest that when PcG-regulated genes are transcribed, flankingly ubiquitously expressed genes are acting as boundaries to the “transcriptionally ON” PcG domain.

**DISCUSSION**

PcG repression establishes an epigenetic memory that permits the heritable propagation and maintenance of this repression throughout development, so identifying the components regulating PcG domain structure and function is very important. Our primary aim in this work was to study the effect of “chromatin context” on a PcG domain; the following main conclusions can be drawn from this study. First, the en PcG domain at ectopic genomic sites rescues the null mutants of inv-en. This shows that the information to form the chromatin domain is primarily present within the domain itself. Second, spreading of the chromatin state at different sites of the genome is very much dependent on the local “context” itself. Third, a repressive chromatin domain can interact with null chromatin (has no chromatin mark) but stays segregated from the active chromatin. Fourth, a fragment of DNA containing the strong, constitutive PREs is required at an ectopic location to ensure PcG silencing in all tissues. Strikingly, the endogenous locus is resilient to loss of the same DNA fragment, indicating the importance of context on PcG chromatin domain formation and function. Fifth, ubiquitously expressed flanking genes may act as boundaries for enhancers within PcG domains. Our data provide experimental evidence for the hypothesis that chromosomal neighborhood plays an important role in regulating gene expression.

**The endogenous PcG domain is more robust**

Transgenic assays to test “functional PREs” have highlighted the effect of chromatin context on reporter gene expression. For example, the strong PREs upstream of en, present in the 1.5-kb fragment studied here, only act to repress reporter transgene expression in about 50% of chromosomal insertion sites (14). Previously, we have shown that these PREs are dispensable from the endogenous inv-en locus (8). The normal phenotype of en80NULL flies also supports these data. However, here we show that the HAen79@attP40 transgene has a higher level of H3K27me3 accumulation than HAen79pGANT-5@attP40. This reduced level of H3K27me3 is sufficient to maintain correct en expression in embryos and imaginal discs. Notably, the absence of strong PREs in HAen79pGANT-5@attP40 caused misexpression of en during adult abdomen development. Our data emphasize the importance of context itself on the stability and resiliency of the PcG domain toward modification (mutation or deletion) of regulatory sequences present within the domain.

Our data also show that deletion of the 1.5-kb DNA fragment that includes the strong en PREs rendered the HAen79 transgene unable to rescue inv-en mutants. Strikingly, HA-en expression from HAen79pGANT-5 was very low in the embryonic nervous system, a result not seen with HAen79 or in en80NULL flies. These data suggest that the embryonic nervous system enhancers are not able to interact well with the HA-en promoter in the absence of the 1.5-kb DNA fragment. This same deletion, when present at the endogenous locus, does not cause a loss of en expression in the nervous system. This suggests that the structure of the endogenous locus is resilient to the loss of this DNA. We suggest that the chromosomal location of the endogenous locus aids in the proper folding of en to facilitate enhancer-promoter communication in the embryonic nervous system. Another explanation for this observation is that the 1.5-kb PRE could be acting as a Trithorax response element (TRE) and bind to Trithorax group proteins for proper en expression. In any case, the essential role of 1.5-kb PRE/TRE at the ectopic locus is alleviated by the “local context” effect at the endogenous locus.

**Spreading of the repressive mark requires sequence-specific recruitment of the PcG proteins**

It has been proposed that establishment and inheritance of H3K27me3 are dependent on two factors: (i) sequence-specific recruitment of the chromatin modifiers and (ii) the ability of H3K27me3 to act as a template for PRC2 to bind and modify other nucleosomes present in the vicinity (25–27). In our experiments, we show that the spreading of H3K27me3 to the flanking chromatin surrounding the insertion site is not very efficient, although no active chromatin mark was present in the vicinity (particularly at attP3). However, our 4C-Seq analysis showed that the ectopic en PcG domain interacted with flanking chromatin significantly until it encountered the active domain. These observations indicate that mere interactions between the PcG domain and flanking chromatin were not able to spread the repressive mark efficiently. Our observation highlights the importance of PcG recruitment via PREs for PcG domain formation.

One of the surprising observations in this study was the deposition of the repressive mark over the exons of Msp300 in larvae that contain HAen79@attP40. As discussed in the Results, this accumulation is also observed in Kc167 cells but not in larval samples that did not have HAen79 inserted at attP40. Our assumption is that some regions in the genome are more susceptible to the PcG regulation and that the weak binding peaks of PcGs near the attP40 site might be acting as PREs to facilitate spreading of the mark. We posit that, when a PcG domain is inserted into attP40, it renders the adjacent chromatin more likely to form an H3K27me3 domain. We note that the Msp300 gene is covered by both H3K27me3 and H3K36me3. This is likely due to the mixed cell population in our larval brains and discs. What about the spreading of the H3K27me3 mark over the exons? This is unusual; however, pre-mRNA and cotranscriptional activity have been linked to the local chromatin structure (28). In a genomic study of many histone modifications in human and Caenorhabditis elegans DNA, the H3K27me3 mark was enriched over exons (28). In addition, physical interaction between mammalian splicing factors (U2snRNP and Sf3b1) and PcG proteins (Zfp144 and Rfn2) was reported to be required for proper repression of Hox genes (29). We speculate that cotranscriptional recruitment...
of PcG proteins over the Msp300 gene in the transgenic line might have established the exon-specific deposition of the H3K27me3 mark. We propose that the exon-specific H3K27me3 accumulation can act as an intermediate step to repress transcriptionally active target genes and establish PcG domains during development.

Active and repressed domain segregation appears to be a basic principle of chromatin organization

Two interesting questions for chromatin biologists are the following: How does a meter-long genome fit into a nucleus and how does this folding influence genome function? High-resolution Hi-C experiments have given structural insights into interphase chromosomes in eukaryotic nuclei. According to the Hi-C data, chromatin is organized into TADs or “contact domains” (3, 4, 30), and the TADs form compartments A (enriched with active domains) and B (enriched with inactive domains) (31, 32). While the Drosophila genome has “compartments,” the existence of TADs in Drosophila is disputed (5), and despite the evidence that TADs and compartments are important for chromatin organization and function, basic information about how these structures are formed and maintained has been
incomplete. In our data, we show that it is the intrinsic property of chromatin to segregate based on histone modifications and gene activity. This property of chromatin is not locus specific; in different chromatin contexts, repressed chromatin tends to segregate from active domains. Biochemical and molecular evidence on the antagonistic behavior between the H3K27me3 and H3K36me3 modifications also support our claim (33–35) and provide evidence toward H3K36me3 as a chromatin component that restricts the PcG-mediated spread of H3K27me3.

How are large PcG domains formed and maintained? The establishment of the PcG domain and the spreading of H3K27me3 in the PPREs present in the PcG domain during cell cycle 14 (36). Repressive loops between PPREs within PcG domains are also formed during cell cycle 14 (13). These loops are proposed to play a synergistic role in establishing PcG domains (13). Surprisingly, deletion of some strong PPREs in situ resulted in weak phenotypes (8, 11, 12), suggesting redundancy of PcG recruitment. Here, we provide evidence that apart from the minor PPREs present in the large PcG domains, chromatin context itself is a critical factor that determines robustness and function of PcG domains.

MATERIALS AND METHODS

Fly strains

Construction of HAen79 and HAen79stop and generation of their transgenic flies used the same methods as previously described (10). In HAen79, the same DNA fragment deleted in en (8) was deleted from the HAen79 construct [2R:7,353,743..7,386,877 (R5)], with an AT inserted at the deletion junction. The enD insertion in situ deletion was generated by P element imprecise excision and has 32 base pairs of unknown origin present at the deletion site (8). In HAen79stop, the DNA sequence that encodes En amino acids 479 to 499 was changed from CAGATCAAG to TAGTAATGA. In the enPREs@ attP40 line, a polymerase chain reaction (PCR) fragment that contains the sequence deleted in enD was cloned into modified attB–P[acman]-ApR vector [the eye and testis enhancer from the gypsy promoter (10)] by Asc I and Pac I restriction sites (PCR primers are GATGGCGCGC-GCGGTCAGAAGCTTGTGCTCGGAG and GCCGACGGCAACAGCGGA). In the PRED@attP40 line, PRED was also cloned into the modified vector (PCR primers are GGGG-GGCGGCTCTCAATCTTGTGGCGGA and GGGGTTAACGACATCGAGATCGGTGTGGCGGA). The deletion of 110 kb (en) and 33 kb (enD) of the inv- domain was achieved through CRISPR (http://flycrispr.molbio.wisc.edu). The guide RNA (gRNA) target sequences were cloned into pU6-*BsrI-chiRNA, and the plasmids were injected into y w; M[en-] flies (37) by Rainbow Transgenic Flies Inc. Desired deletions were screened by PCR and DNA sequencing. The coordinates of the deleted sequence in enD was 2R:7,353,743..7,463,977 (R5), with 3 nucleotides CCC inserted at the junction. The coordinates of the deleted sequence in en80 was 2R:7,353,743..7,386,877 (R5). To generate the en80, a y w; M[en-] strain was made, and gRNA plasmids were injected into this fly strain. Detailed information about the experimental protocol is available upon request. The molecular basis in T(2;2)enD was determined by sequencing PCR product encompassing the break/junction. PCR primers are CAC-GATGATATCATGCTCGACA and TCCCTACAATAAAAA GCCAAT.

Immunostaining and RNA in situ experiments

The procedure for antibody staining of imaginal discs has been described previously (8). Rabbit anti-En (1:500; SC 28640, Santa Cruz Biotechnology Inc.) and mouse anti-HA (1:100; PRB-101C, Abcam) were used for the immunostaining experiments.

For RNA in situ, we used pBS II SK(+) plasmid for cloning PCR fragments amplified by the following primers: yellow, GGGG-TACCCGAGCTAATTCTGATCCA and GGGGATCTCTCTC GTCTCCTGTATCATC; Msp300, GGGGTACCACAATAGCCCAACGGGAACAG and GGGGATCCGCTCCTCTCTCTCATCACG; CG1504, GGGGATCCGAAACCTGTGACACTCTCG and GGGGTACCAGCTGACAGTTGTAAGG; and CG1631, GGGG- GTACCTGCAATTCCAGAGACCCAG and GGGGATCCGCGTACTCAAAGAGCTCCA. Digoxigenin-labeled RNA antisense probe synthesis and whole-mount in situ hybridization were carried out as previously described (38).

ChIP and ChIP–quantitative PCR

Protocol for carrying out ChIP with anti-Pho and anti-Ph antibodies and with anti-H3K27me3 in larval tissue has been described previously (8). A 1:200 dilution of anti-H3K36me3 (ABE435, Millipore) antibody was used for ChIP.

ChIP-seq and 4C-seq

Following purification of immunoprecipitated DNA, Illumina libraries were prepared using the TrueSeq DNA Sample Prep Kit V2 as described (http://ethanomics.wordpress.com/chip-seq-library-construction-using-the-illumina-truseq-adapters/). Chromosome conformation capture (3C) and 4C-seq protocol was followed as described previously (8, 39). We used Dpn II and Nla III for making the libraries.

Bioinformatic analysis

ChIP-seq reads were aligned to reference genome [dm3, BDGP (Berkeley Drosophila Genome Project) Release 5] using Bowtie v1.1.2 (40) with the following parameters: -n 2 -l 28 -k 1 --best. The PCR duplicates for each dataset were removed using the rndmp function of SAMtools (41). Then, the resulted bam files were converted to tdf format by using igvtools for visualization in IGV (Integrative Genomics Viewer) (42).

Reads from 4C-seq were aligned to a reduced genome consisting of all unique sequences adjacent to Dpn II sites with bowtie2 (40), with -n 0, --5 23. The counts for each fragment were then obtained from the mapped SAM files according to the procedure provided in 4C-ker (43). For visualization, the counts for each sample were linearly scaled; thus, their sum is 1 M. To detect the interacting regions near the bait, we performed near-bait analysis using 4C-ker (43), with k = 5. However, the trans-analysis from 4C-ker does not work for our data; thus, we modeled the trans–chromosome read counts as negative binomial distribution, which have been widely used to model RNA sequencing—, ChIP-seq—, and Hi-C–related data, to detect significant trans–chromosome interactions. Briefly, we moved a sliding window with five digested fragments along all trans chromosomes. The number of reads resided within each window was determined, and then the two parameters of negative binomial distribution (size and mu) were estimated using the maximum likelihood approach. The P value for each window was determined as the probability that the observed count is higher than expected according to negative binomial distribution. Last, the P values were adjusted to get the false discovery rate (FDR). The same procedure was
applied to all replicate datasets, and only regions determined as significant (FDR < 0.05) in all replicates were called as significant interaction regions.

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/1/eaa9739/DC1


Acknowledgments: We thank H. Smith for carrying out all Illumina NGS. We thank E. Lei, K. Pfeifer, and A. Perkins for valuable input on the paper. We also want to thank M. Fujioka (mini-w plasmid) and K. G. Ten Hagen (Pgan5 plasmid) for providing the plasmids to make the RNA probes. This study used the computational resources of the NIH High-Performance Computing Biowulf cluster (https://hpc.nih.gov). Funding: This work was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH. Author contributions: S.D. and J.A.K. conceived and designed the experiments. S.D., Y.C., M.-a.S., N.D.G., and J.A.K. performed the experiments, analyzed the data, and wrote the paper. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Sequencing data generated in this study are deposited in the NCBI BioProject database (PRJNA494709). Additional data related to this paper may be requested from the authors.

Submitted 1 August 2018
Accepted 28 November 2018
Published 9 January 2019
10.1126/sciadv.aau9739

Structure and function of an ectopic Polycomb chromatin domain
Sandip De, Yuzhong Cheng, Ming-an Sun, Natalie D. Gehred and Judith A. Kassis

DOI: 10.1126/sciadv.aau9739

ARTICLE TOOLS  http://advances.sciencemag.org/content/5/1/eaau9739
SUPPLEMENTARY MATERIALS  http://advances.sciencemag.org/content/suppl/2019/01/07/5.1.eaau9739.DC1
REFERENCES  This article cites 43 articles, 18 of which you can access for free  http://advances.sciencemag.org/content/5/1/eaau9739#BIBL
PERMISSIONS  http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service