



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2022

Critical role of CTCF and Cp190 in Drosophila genome organization and gene regulation

Anjali Anjali

Anjali Anjali, 2022, Critical role of CTCF and Cp190 in Drosophila genome organization and gene regulation

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB_CFE13B78295C6

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Center for Integrative Genomics

**Critical role of CTCF and Cp190 in *Drosophila* genome
organization and gene regulation**

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Anjali Anjali

Master de l'Université Maharaja Sayajirao de Baroda

Jury

Prof. Stephan Gruber, Président
Prof. Maria Cristina Gambetta, Directrice de thèse
Prof. David Gatfield, Co-directeur de thèse
Prof. Bart Deplancke, Expert
Prof. Christian Fankhauser, Expert

Lausanne 2022



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Stephan	Gruber
Directeur·trice de thèse	Madame	Prof.	Maria Cristina	Gambetta
Co-directeur·trice	Monsieur	Prof.	David	Gatfield
Expert·e·s	Monsieur	Prof.	Bart	Deplancke
	Monsieur	Prof.	Christian	Fankhauser

le Conseil de Faculté autorise l'impression de la thèse de

Anjali Anjali

Master of Science in biotechnology, The Maharaja Sayajirao University of Baroda, Inde

intitulée

**Critical role of CTCF and Cp190 in *Drosophila*
genome organization and gene regulation**

Lausanne, le 17 juin 2022

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Stephan Gruber

Acknowledgments

I would like to express my deepest gratitude to my PhD advisor Prof. Maria Cristina Gambetta, whose insight, knowledge, and patience steered me through this research. I want to acknowledge her for allowing me to be a part of her new lab and guiding me throughout my Ph.D.

This project would not have been possible without the support of many people. I would like to thank all the collaborators for their crucial help in my PhD. I am also thankful to my thesis committee members for their valuable inputs.

I am fortunate to have pleasant and supportive colleagues. I thank Anastasiia Semenova, Giriram Kumar, Bihan Wang, Marion Mouginot, and Pascal Cousin for making my stay in Gambetta lab a nice and wonderful memory. They are the great bunch of people in and out of the lab.

I also thank Marion, Pascal, Jon and Anastasiia in helping me translate some parts of my thesis in French. I would also like to thank my Ph.D. mentor Prof Liliane Michalik for her generous and kind support.

I would like to acknowledge the support from the University of Lausanne for the courses and financial support.

I like to thank my friends Pragya, Chetan, Sol Santillán, and my group quarantine sufferers for all the support and fun.

I want to express my gratitude to my parents, brother, and Paramvir. Their tremendous understanding and encouragement have helped me a lot.

Abstract

The eukaryotic genome is partitioned into contact domains (sometimes referred to as Topologically Associating Domains, or TADs). These domains are chromatin regions that interact in 3D more frequently within themselves than with nearby domains. A particular class of proteins known as insulator binding proteins (IBPs) bind at the boundaries of these domains. In mammals, CTCF - the only known IBP to date – forms contact domain boundaries by stalling loop-extruding cohesin. This mechanism explains why CTCF and cohesin binding is enriched at contact domain boundaries in mammals. Even though *Drosophila* also has CTCF, the role of CTCF in forming contact domain boundaries was debated because CTCF binding is not clearly enriched at fly contact domain boundaries.

In contrast to CTCF, another IBP called Cp190 was recently reported to bind to all contact domain boundaries. We examined the effects of complete absence of CTCF and Cp190, separately and in combination, on genome organization and gene regulation. We found that CTCF and Cp190 are both essential for fly viability, but Cp190 loss leads to much earlier lethality (embryonic lethality) than loss of CTCF (late pupal lethality). We found that CTCF binds to and is required to form a small subset of contact domain boundaries. Cp190 recruitment to CTCF-occupied boundaries was strictly CTCF-dependent at some boundaries, and at least partially CTCF-independent at other boundaries. CTCF and Cp190 coregulated the expression of a subset of genes present near CTCF and Cp190 co-occupied boundaries. In contrast to CTCF, we found that Cp190 is the major architectural protein in *Drosophila* as 25% of contact domain boundaries were affected by Cp190 loss (compared to only 10% of boundaries affected by CTCF loss). Although Cp190 binds to most contact domain boundaries, Cp190 was only required to form boundaries that are distal to (i.e. more than 200 bp from) a transcriptional start site (TSS) whereas Cp190 was dispensable to form the remaining 75% of boundaries that are all TSS-proximal. In addition to Cp190's role in contact domain boundary formation, Cp190 also blocked regulatory crosstalk of tested developmental genes with nearby enhancers. However, Cp190 was dispensable for long-range enhancer-promoter communication at tested loci. This demonstrated that, in contrast to preconceived notions, Cp190 is not merely a “looping factor” organizing

the fly genome into a series of loops that either connect or segregate gene promoters and enhancers.

Taken together, our results demonstrated that CTCF and Cp190 are critical factors required for genome organization and gene regulation. This revealed for the first time that diverse mechanisms exist to partition animal genomes into physical and regulatory domains. This advanced our fundamental understanding of the relevance of genome folding for gene regulation in *Drosophila*.

Résumé

Le génome eucaryote est divisé en domaines de contact (parfois appelés domaines d'association topologique ou TAD). Ces domaines sont des régions de chromatine qui interagissent en 3D plus fréquemment en eux-mêmes qu'avec des domaines voisins. Une classe particulière de protéines, connues sous le nom de protéines de liaison isolantes (IBP) se lie aux limites de ces domaines. Chez les mammifères, CTCF – l'unique IBP connu à ce jour - forme les limites des domaines de contacts, en bloquant la cohésine durant l'extrusion de boucle d'ADN. Ce mécanisme explique pourquoi la liaison de CTCF et de la cohésine est enrichie aux limites des domaines de contacts, chez les mammifères. La drosophile possède également CTCF, mais son rôle dans la formation des limites des domaines de contacts est débattu. En effet, CTCF n'y est pas clairement enrichi.

Contrairement à CTCF, une autre IBP appelée Cp190 a récemment été signalé comme se liant à toutes les frontières des domaines de contacts. Nous avons examiné les effets de l'absence complète de CTCF et Cp190, séparément et en combinaison, sur l'organisation du génome et la régulation des gènes. Nous avons constaté que CTCF et Cp190 sont tous deux essentiels à la viabilité des drosophiles, mais que la perte de Cp190 entraîne une létalité beaucoup plus précoce (létalité embryonnaire) que la perte de CTCF (létalité pupale tardive). Nous avons constaté que CTCF est nécessaire pour former un petit sous-ensemble de frontières de domaines de contacts et qu'il se lie uniquement à ces frontières-là.

Le recrutement de CTCF aux frontières des domaines de contacts occupées par Cp190 est strictement dépendant du type de frontière : CTCF est requis à certaines frontières, et il est partiellement requis à d'autres frontières.

CTCF et Cp190 co-régule l'expression d'un sous-ensemble de gènes présents près des frontières des domaines de contacts qu'ils occupent ensemble. Contrairement à CTCF, nous avons constaté que Cp190 est la principale protéine architecturale chez la drosophile : En effet, 25 % des limites des domaines de contacts sont affectés par la perte de Cp190 alors que seulement 10 % des limites sont affectés par la perte de

CTCF. Bien que Cp190 se lie à la plupart des limites des domaines de contacts, Cp190 est nécessaire pour former des limites distales (c'est-à-dire à plus de 200 pb) d'un site de début de transcription (TSS) mais n'est pas essentiel pour former les 75% restants des limites des domaines de contacts qui sont toutes proche des TSS. En plus du rôle de Cp190 dans la formation des limites des domaines de contacts, Cp190 est également nécessaire pour bloquer les interactions indésirées entre des gènes de développement testés et des enhancers/amplificateur présent à proximité. Cependant, Cp190 est primordiale pour la communication entre amplificateur et promoteur séparé par de longue distance, aux loci testés. Démontrant que, contrairement aux idées préconçues, Cp190 n'est pas simplement un « facteur de boucle » organisant le génome de la drosophile en une série de boucles qui connectent ou séparent les promoteurs et les amplificateurs de gènes.

Pris ensemble, nos résultats démontre que CTCF et Cp190 sont des facteurs critiques nécessaires à l'organisation du génome et à la régulation des gènes. Cela met en lumière que pour la première fois, divers mécanismes existent pour partitionner les génomes animaux en différent domaines physiques très régulés. Nos découvertes mettent en lumière la pertinence de l'étude du repliement du génome pour la régulation des gènes chez la drosophile.

List of abbreviations

BEAF-32 - Boundary element-associated factor of 32

BED - BEAF-32 and DREF

bHLH - Basic helix-loop-helix

BTB - Broad-Complex, Tramtrack, and Bric-a-brac

Cp190 - Centrosomal protein 190 kDa

CTCF - CCCTC-binding factor

DNA - Deoxyribonucleic acid

Ibf - Insulator binding factor

IBP - Insulator binding proteins

Mod(mdg4) - Modifier of mdg4

NIPBL - Nipped B-like protein

PCR - Polymerase chain reaction

PREs - Polycomb group response elements

RE - Regulatory Element

RNA - Ribonucleic acid

SDS - Sodium dodecyl sulfate

SHH - Sonic Hedgehog

Su(Hw) - Suppressor of Hairy wing

TADs - Topologically associated domain

ZAD - Zinc finger associated domain

ZIPIC - Zinc-finger protein interacting with Cp190

ZnF - Zinc finger

List of figures

Figure 1. Mutation in REs affects gene expression.

Figure 2. Overview of the gene expression pathway.

Figure 3. Early concepts of genome organization in interphase nuclei.

Figure 4. Schematic image representing different aspects of chromatin organization.

Figure 5. Overview of a Hi-C experiment.

Figure 6. Hi-C revealed new levels of chromosomal organizations.

Figure 7. Chromatin loop extrusion by cohesin happens in three steps.

Figure 8. Clinical examples of chromosomal rearrangements that alter TADs.

Figure 9. Insulator binding proteins.

Figure 10. Properties of insulators.

Figure 11. Differential enrichment of DNA motifs at fly TAD boundaries.

Table of contents

ACKNOWLEDGMENTS	I
ABSTRACT	II
RÉSUMÉ	IV
LIST OF ABBREVIATIONS	VI
LIST OF FIGURES	VII
TABLE OF CONTENTS	VIII
INTRODUCTION	1
1.1 LINEAR VIEW OF THE GENOME	1
1.2 3D ORGANIZATION OF THE GENOME	3
1.3 GENOME ORGANIZATION IN MAMMALS	7
1.4 CONNECTION BETWEEN TOPOLOGICALLY ASSOCIATED DOMAINS AND GENE REGULATION	9
1.5 DROSOPHILA AS A MODEL TO STUDY THE FUNDAMENTALS OF GENOME ORGANIZATION	11
1.6 DROSOPHILA INSULATOR BINDING PROTEINS (IBPs)	12
1.7 THE ROLE OF Cp190 IN INSULATION AND GENOME ORGANIZATION	15
1.8 DO TADS MERELY ARISE FROM TRANSCRIPTION-RELATED PROCESSES IN FLIES?	17
2 PROJECT AIMS	19
3 FIRST PUBLICATION	20
3.1 SHORT SUMMARY OF THE FIRST PUBLICATION	21
4 SECOND PUBLICATION	21
4.1 SHORT SUMMARY	24
5 DISCUSSION	25
NOTE: ALL THE FIGURES REFERRED IN THIS SECTION ARE FROM THE PUBLISHED ARTICLE ATTACHED SECTION IN 11.2	25
5.1 SUMMARY OF FINDINGS	25
5.2 DIVERSE WAYS TO FORM BOUNDARIES IN THE FLY GENOME	26
5.3 Cp190 PREVENTS PROMISCUOUS REGULATION OF SEVERAL TESTED GENES	29
5.4 Cp190 IS NEITHER REQUIRED FOR HOX GENE NOR HOMIE-MEDIATED DISTAL ACTIVATION	30
6 CONCLUSION	34
7 METHODS	35
7.1 CHROMATIN PREPARATION FROM FLY EMBRYOS	35
7.2 CHIP-SEQ	35
7.3 CHIP-SEQ DATA ANALYSIS	37
7.4 LIBRARY PREPARATION USING NEBNext ULTRA II DNA LIBRARY PREP KIT FOR ILLUMINA	38
7.5 HI-C	40
7.6 HI-C DATA ANALYSIS	44
7.7 INSULATOR REPORTER ASSAY	45
7.8 RNA-FISH	46
7.9 RECOMBINANT PROTEIN PULL-DOWNS	49
7.10 WESTERN BLOT ANALYSIS	51
DROSOPHILA VIABILITY TEST	52
9 MATERIALS	53
10 REFERENCES	62

11. ANNEXED ARTICLES	72
11.1 CTCF LOSS HAS LIMITED EFFECTS ON GLOBAL GENOME ARCHITECTURE IN DROSOPHILA DESPITE CRITICAL REGULATORY FUNCTION	
11.2 ESSENTIAL ROLE OF Cp190 IN PHYSICAL AND REGULATORY BOUNDARY FORMATION	

Introduction

1.1 Linear view of the genome

The complete genetic information in an organism is known as the genome. It consists of nucleotide sequences of DNA that transfer genetic information via sexual or asexual reproduction in all organisms known so far. Genetic information exists in two forms of DNA. One is in the form of a genes and the other is in the form of non-protein-coding DNA, which was initially considered to be the junk part of the genome. However, several elements of non-coding DNA have now been identified as regulatory elements (REs) that control the expression of genes. REs contain binding sites for transcription factors or non-coding RNA molecules that regulate gene expression by activating, sustaining, or suppressing transcription. Enhancers, silencers, and insulators are the main types of REs and are defined as follows:

- Enhancers contain short DNA motifs that recruit transcriptional factors to activate transcription of their target genes, somewhat independently of the orientation and positioning of enhancers relative to their targets (Bulger and Groudine 2011). Transcriptional factors that bind to enhancers are known as transcriptional activators.
- Silencers are the repressive counterparts of enhancers. They are DNA elements capable of repressing transcription by recruiting transcriptional repressors (Segert et al. 2021).
- Insulators are DNA elements that recruit Insulator binding proteins (IBPs) to block the influence of other regulatory elements' effects on gene promoters to avoid unwanted gene regulation (discussed in section 1.7) (reviewed in (Özdemir and Gambetta 2019)). To block regulatory element-promoter communication, insulators must be present in between them.

Any alteration in a gene or RE can have a severe effect on gene expression level and may affect the viability of an organism. Pioneering work from Ed Lewis showed that mutations in regulatory sequences of the *Drosophila Ultrabithorax* gene alter its expression, and this causes dramatic development phenotypes (Fig-1) (Lewis 1978).

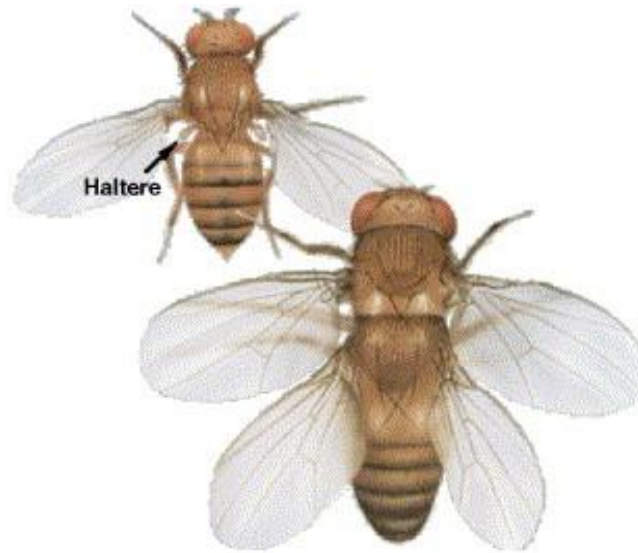


Figure 1. Mutation in REs affects gene expression. Mutation in a regulatory sequence in the Bithorax complex results in an extra pair of wings in *Drosophila*. A wildtype fly has a pair of wings and a pair of halteres (black arrow) required for balance in flight. Ed Lewis identified mutations in regulatory sequences of the *Ultrabithorax* gene that convert halteres into another pair of wings. Figure source: Akbari et al. 2006

Therefore, every process influencing gene expression is tightly regulated. These processes can be roughly grouped into three categories: (1) genomic, (2) transcriptomic, and (3) proteomic. These categories control gene expression via various processes (Fig-2). Gene regulation can be considered the most significant one as any perturbation at this level will ultimately affect the downstream levels.

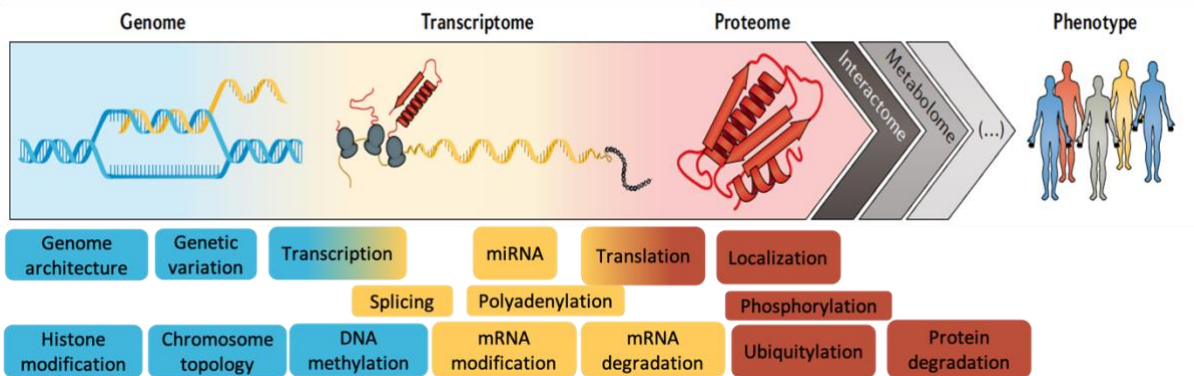


Figure 2. Overview of the gene expression pathway. A general schematic of three levels that control gene expression and connect genotype to phenotype. Several processes that regulate these categories are shown below. Figure source: Buccitelli and Selbach 2020

1.2 3D organization of the genome

The genome is not randomly organized inside the nucleus, and its three-dimensional positioning inside the nucleus plays a critical role in proper genome functioning. Detailed studies have described severe effects of changes in the linear sequences of the chromatin. Even without any alteration in the linear sequence of the chromatin, changes in the 3D positioning can also severely affect gene expression.

The concept of the 3D organization of chromosomes is not recent. In 1885, Carl Rabl observed *Salamandra maculate* and *Proteus* cell nuclei under the microscope and proposed a theory on the internal structure of the interphase nucleus. It states, in modern terminology, that “*Chromosomes maintain their individuality as well as their anaphase-telophase orientation throughout interphase and occupy distinct territories in the interphase nucleus*” (Cremer et al. 1982) (Fig-3). Using *Ascaris megalocephala*, the cytologist Theodor Boveri also showed that chromosomes occupy distinct territories inside the nucleus. He proposed that any change in their organization might lead to pathogenicity (Boveri 1887)

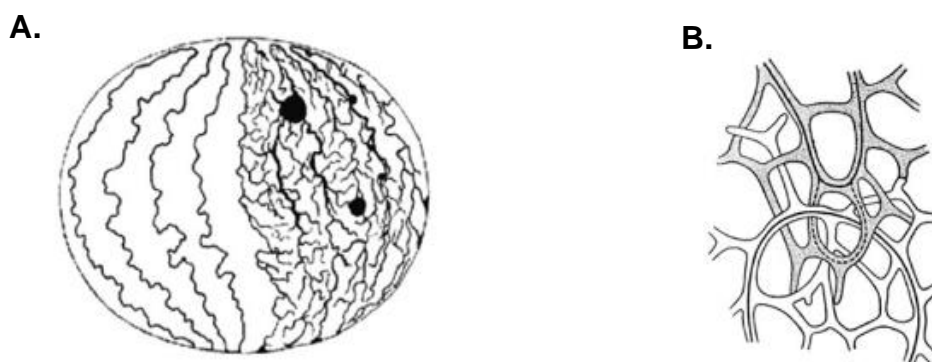


Figure 3. Early concepts of genome organization in interphase nuclei. **A.** Carl Rabl's view on chromosomal arrangements in *Proteus* and *Salamandra* interphase nucleus. **A.** Side view of an interphase nucleus, chromatin threads are segregated in the nucleus. **B.** Boveri's sketch of two neighboring chromosomal territories in sponge like structures because of chromatin bundles. The continuous line is shown as a primary chromatin thread. Figure source: Cremer and Cremer 2010

The concept of DNA organization is not limited to the higher chromosomal level (Fig-4). The smallest unit of genome folding is the nucleosome, consisting of ~147 bp of DNA wrapped around an octamer of core histones (Kornberg 1974). Nucleosomes are

further compacted into chromatin. Chromatin exists in two forms: (a) Euchromatin is an easy to transcribe, less condensed, "beads on a string" structure; (b) Heterochromatin is usually the non-transcribed, highly condensed part of chromatin. Further higher levels of chromatin organization had remained inaccessible for decades. Using electron microscopy combined with ChromEM, Ou et al. revealed the ultra-structure and 3D organization of chromatin fibers. They showed that chromatin forms various shapes that vary from diameter between 5 and 24 nm to achieve structural compaction. This variability highlights that chromatin can have many different ways of arranging itself to achieve structural compaction and high packaging density (Ou et al. 2017).

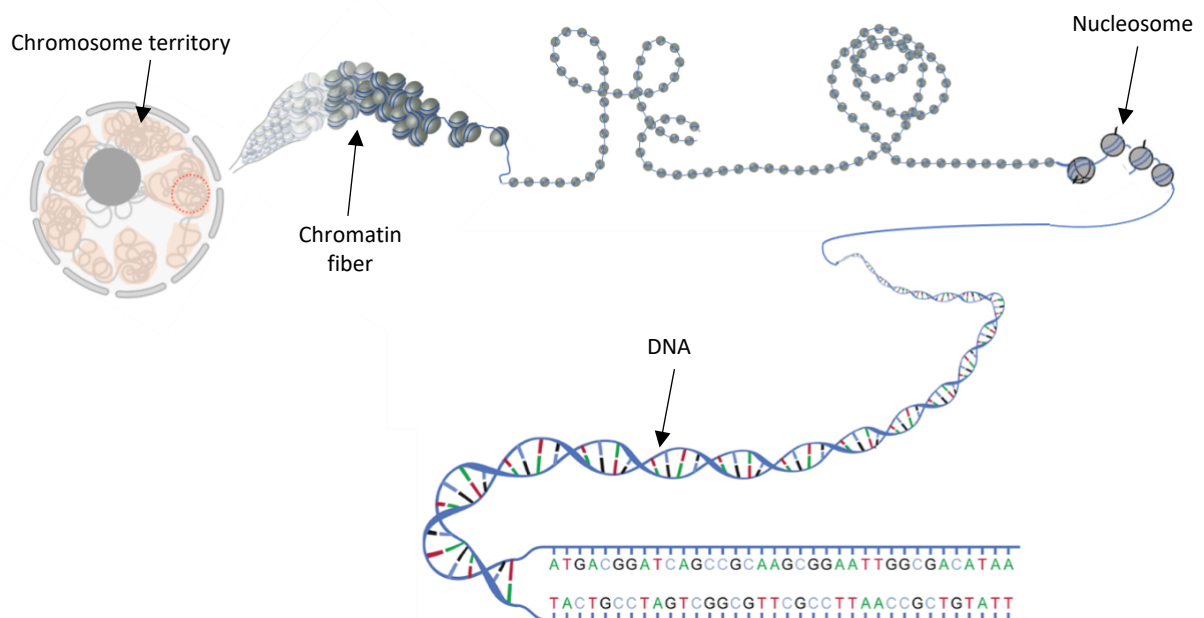


Figure 4. Schematic image representing different aspects of chromatin organization. Chromosome territories within the nucleus are shaded in orange and are composed of chromatin fibers. Chromatin fibers are the coiled and condensed assemblies of nucleosomes. In eukaryotic cells, nucleosomes are the fundamental unit of chromatin consisting of ~147 bp of DNA wrapped around a histone octamer.

In the last two decades, genome editing, High-throughput Chromosomal Conformation Capture (Hi-C) techniques, single-cell genomics, super-resolution imaging, and computational modeling have been valuable tools to further our understanding of genome organization and gene regulation. These technologies not only reestablished the existence of a century-old concept of chromosome territories but also revealed several new layers of genome organization (Cremer and Cremer 2001; Simonis et al. 2006; Cremer et al. 2008; Lieberman-Aiden et al. 2009; Stevens et al. 2017). Hi-C has

been the most widely used strategy (Lieberman-Aiden et al. 2009). In a Hi-C experiment (Fig-5), chromatin is briefly cross-linked with neighboring loci with formaldehyde. The cross-linked DNA is fragmented by restriction digestion, but neighboring loci remain attached to each other due to cross-linking. The 5' overhangs of the restriction digested DNA fragments are filled with nucleotides, one of which is biotinylated. This results in blunt-ended DNA fragments. These DNA fragments are then ligated with neighboring cross-linked DNA. This ligation results in a library of chimeric DNA that are further fragmented and pulled down using streptavidin beads. Chimeric DNA results from ligation of DNA loci that are close in 3D space though they are normally far apart on the linear genome. These DNA sequences are identified by paired-end sequencing. The frequency of chimeric reads is plotted on a Hi-C heatmap to infer the three-dimensional proximity between any two genomic loci (Fig-5).

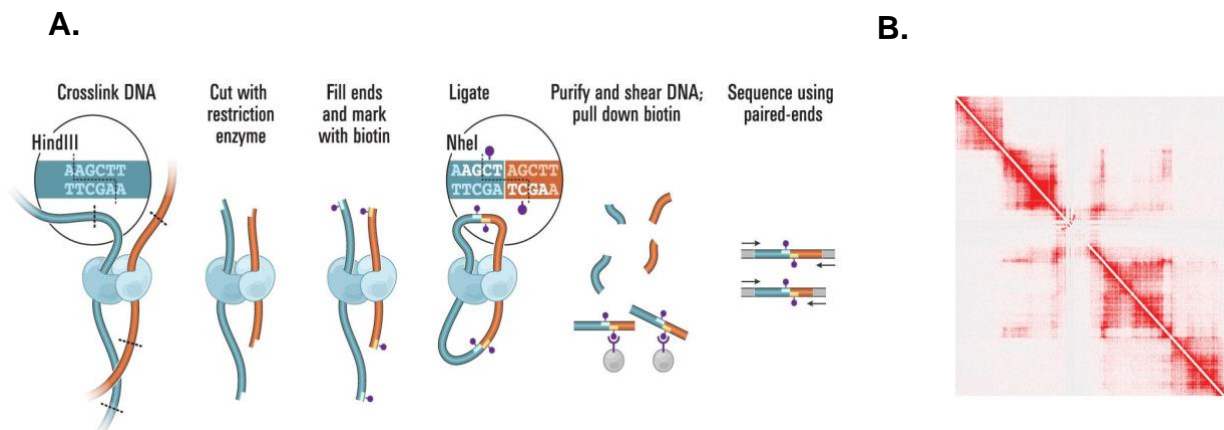


Figure 5. Overview of a Hi-C experiment. **A.** Chromatin fibres are crosslinked with their neighboring loci. DNA is digested using restriction enzymes resulting in sticky ends. Overhangs are filled with nucleotides, one of which is biotinylated. This leads to blunting of DNA fragments. These DNA fragments are ligated with their neighboring DNA fragment by blunt-end ligation. Ligated DNA is purified and sheared. These DNA fragments are enriched using streptavidin beads and subjected to paired-end sequencing. **B.** A representative Hi-C heatmap shows a portion of a genome-wide contact matrix. Figure source: Lieberman-Aiden et al., 2009

Hi-C technology provides a quantifiable metric to measure contact probabilities between pairs of loci genome-wide (Lieberman-Aiden et al. 2009). Hi-C maps confirmed the presence of chromosome territories and further identified two types of spatial compartments in mammalian nuclei: a primarily open chromatin compartment A and a primarily closed chromatin compartment B (Lieberman-Aiden et al. 2009; Zhang et al. 2012). Compartment A consists of transcriptionally active euchromatin,

and compartment B consists of transcriptionally silent heterochromatin regions, generally located at the peripheries of the nucleus or nucleolus. Hi-C also revealed chromosomes are organized into “Topologically Associated Domains” (TADs) (Dixon et al. 2012). TADs are sub-megabase scale genomic regions that self-interact within their domain more frequently than outside their domain (Dixon et al. 2012; Nora et al. 2012; Sexton et al. 2012). TADs are loop-like structures, and their loop anchors correspond to TAD boundaries (Fig-6).

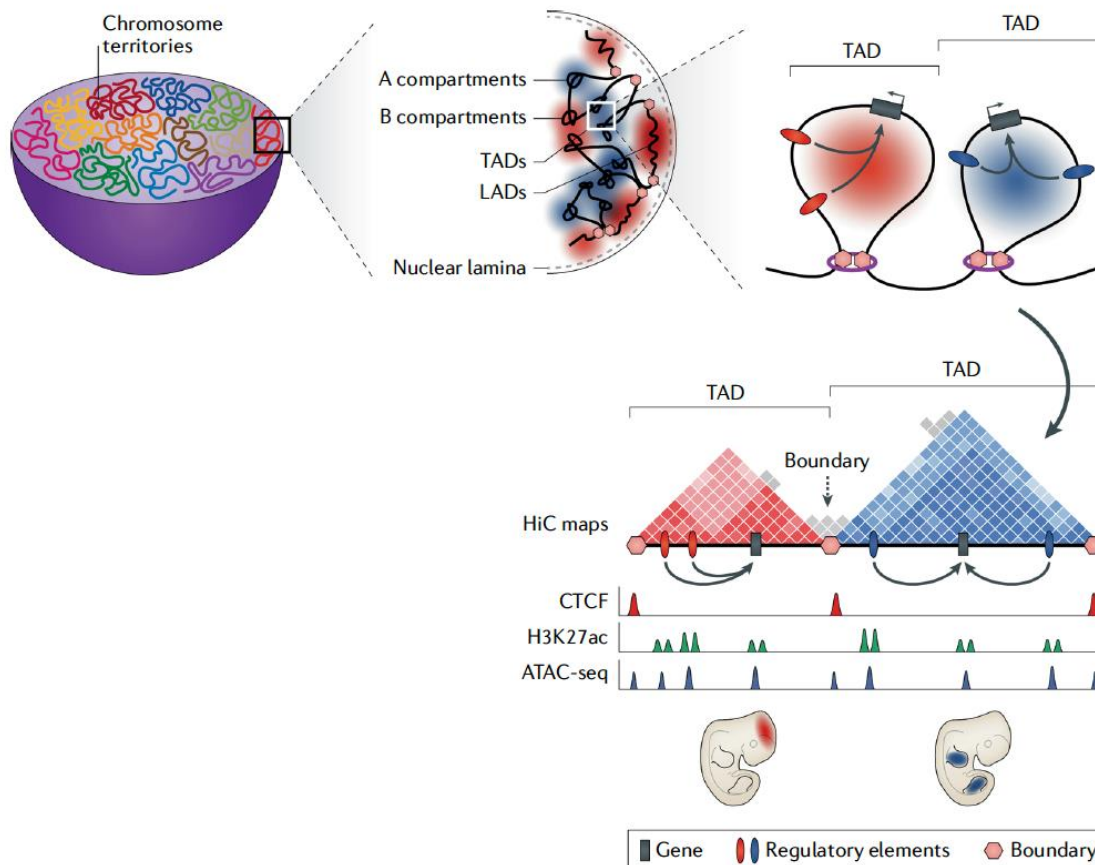


Figure 6. Hi-C revealed new levels of chromosomal organizations. Chromosomes are segregated into A and B compartments. Compartment A is primarily composed of transcriptionally active euchromatin and compartment B with compacted and transcriptionally silent heterochromatin. Topologically associated domains (TADs) are identified in Hi-C heat maps as pyramid-like structures with higher interaction frequency within themselves. Intra-TAD interactions between regulatory elements and gene promoters are higher than inter-TAD interactions. These interactions lead to the tissue-specific expression pattern of genes. Histone modifications and assay for transposase-accessible chromatin using sequencing (ATAC-seq) can help identify regulatory elements. Figure source: Spielmann et al. 2018

Chromatin packaging within a TAD also brings correct REs close to their target genes. In a non-linear genome, a rearrangement at the 3D level may change the gene

expression profile. For example, the interaction between an enhancer and a gene promoter can be disrupted by changing their 3D interaction profiles (discussed in section 1.4).

1.3 Genome organization in mammals

Non-random genome organization is supported by the polymeric nature of chromosomes, which limits their mobility. The movement of each chromatin locus would require moving its neighboring loci, which might also be restricted due to nearby topological interactions or steric hindrance. Therefore, the probability of random interactions between two loci 100 kb or more apart is estimated to be very low during a mammalian cell cycle (Dekker and Mirny 2016).

Most genome organization studies have been performed in mammalian cells. Several molecular players that influence mammalian genome organization have been identified. In vertebrates, TAD boundaries are bound by CCCTC-binding factor (CTCF), an 82 kDa protein with highly conserved zinc finger motifs initially identified for its enhancer-blocking activity (Bell et al. 1999; Rao et al. 2014; Tang et al. 2015). CTCF anchored boundaries are also co-occupied by cohesin, a multiprotein ring-like complex (Bell et al. 1999; Rao et al. 2014; Tang et al. 2015). This dual occupancy mostly has an inward orientation of CTCF binding motifs with 5' and 3' motifs facing each other (Bell et al. 1999; Parelho et al. 2008; Wendt et al. 2008; Rao et al. 2014; Tang et al. 2015). Because CTCF and cohesin co-localize at TAD boundaries, cohesin has been hypothesized to extrude chromatin loops bidirectionally until it gets stalled at TAD boundaries due to interaction with CTCF (Fig-7) (Sanborn et al. 2015; Fudenberg et al. 2016)(Li et al. 2020) Using biochemical reconstitution and single-molecule imaging, Davidson et al. (2019) and Kim et al. (2019) provided the first evidence of cohesin- nipped B-like protein (NIPBL) mediated loop extrusion (Davidson et al. 2019; Kim et al. 2019).

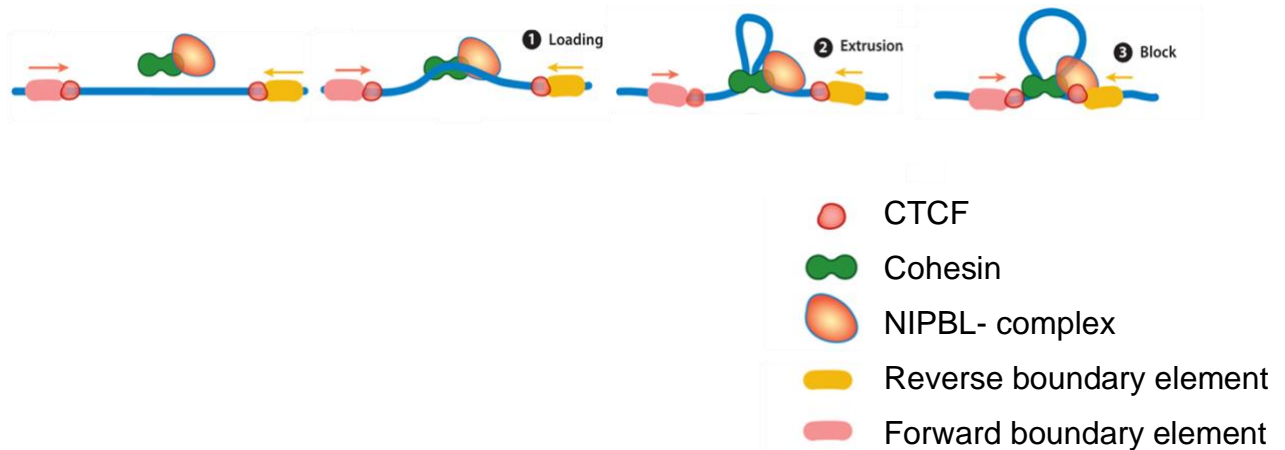


Figure 7. Chromatin loop extrusion by cohesin happens in three steps. 1) Cohesin is loaded onto chromatin by the nipped B-like (NIPBL) complex. 2) NIPBL stabilizes the loop extrusion process by staying associated with cohesin. 3) A chromatin loop is extruded through cohesin until convergently oriented CTCF acts as a barrier at loop anchors. Figure source: Ghosh and Meyer 2021

In addition to its boundary function, CTCF-mediated loops are also thought to exert an insulator activity by preventing physical and regulatory cross-talk between chromatin regions within a loop with those that are outside of the loop. An illustration that TADs and gene regulation are linked is the finding that disrupting CTCF binding sites in *Hox* gene clusters disrupts TAD boundaries and results in the expansion of active chromatin into inactive *Hox* loci (Narendra et al. 2015). Thus, CTCF binding sites insulate active and repressed *Hox* gene regions from each other (Narendra et al. 2015). Moreover, deleting CTCF binding sites at TAD boundaries in *HoxA* and *HoxC* clusters also results in *Hox* gene misregulation and developmental defects (Narendra et al. 2016). Studies in human pluripotent stem cells also showed that some CTCF-anchored sites influence the expression of pluripotency genes (Ji et al. 2016).

CTCF is essential for mouse embryonic development as *ctcf* null mouse embryos fail to implant (Moore et al. 2012). CTCF is vital for cell survival as conditional depletion of CTCF in the limb mesenchyme induces extensive apoptosis during limb development (Soshnikova et al. 2010). Likewise, deletion of CTCF in developing mouse brains also leads to apoptosis in dividing neuroprogenitor cells of the forebrain (Watson et al. 2014). However, we still do not know which events are triggered upon CTCF loss that lead to such severe defects and lethality.

Auxin-inducible acute depletion of endogenous CTCF in mouse embryonic stem cells resulted in an absolute and dose-dependent requirement of CTCF for loops and TAD formation (Nora et al. 2017). Upon degradation of cohesin, loop domains are lost genome-wide. After cohesin recovery, loop domains are formed within minutes. However, such genome-wide effects unexpectedly have a modest effect on global transcription (Nora et al. 2017; Rao et al. 2017). Similar findings were observed upon depletion of the cohesin-loading factor called nipped B-like protein (NIPBL) in mouse liver, which also leads to the global loss of TADs (Fig-6) (Schwarzer et al. 2017). In short, these observations show that CTCF and cohesin are essential for mammalian genome organization, though how important TADs are for gene regulation was debated between studies.

1.4 Connection between topologically associated domains and gene regulation

The impact of genome architecture disruption on gene regulation is still ambiguous because of various contradicting findings discussed below. Support for a role of genome architecture in gene regulation comes from various pathological structural variations that disturb TAD integrity and cause misexpression of several genes. Specific examples of such clinical lines of evidence are discussed below (Fig-8):

- Neo-TAD formation: Duplicating a locus containing the *KCNJ2* gene and REs associated with the *Sox9* gene causes neo-TAD formation in the duplicated locus. Within the neo-TAD, *Sox9* REs ectopically activate the *KCNJ2* gene in mouse limbs causing Cooks syndrome – a malformation of apical structures of digits. This duplication has also been identified in humans (Fig-8 B) (Kurth et al. 2009; Franke et al. 2016).
- TAD fusion: Deleting a TAD boundary at the *lamin B1* gene locus causes forebrain directed enhancer adoption by *lamin B1*, leading to adult-onset demyelinating leukodystrophy (Fig-8 C) (Giorgio et al. 2015).
- TAD shuffling: An inversion at the *WNT6* and *EPHA4* locus shuffles TADs at *WNT6* and *EPHA4* loci. This brings the *Epha4* enhancer closer to the *Wnt6* gene, and drives ectopic expression of *Wnt6* in distal limb tissues. This

inversion causes F-syndrome in mice and humans, syndactyly of finger and thumb (Fig-8 D) (Lupiáñez et al. 2015).

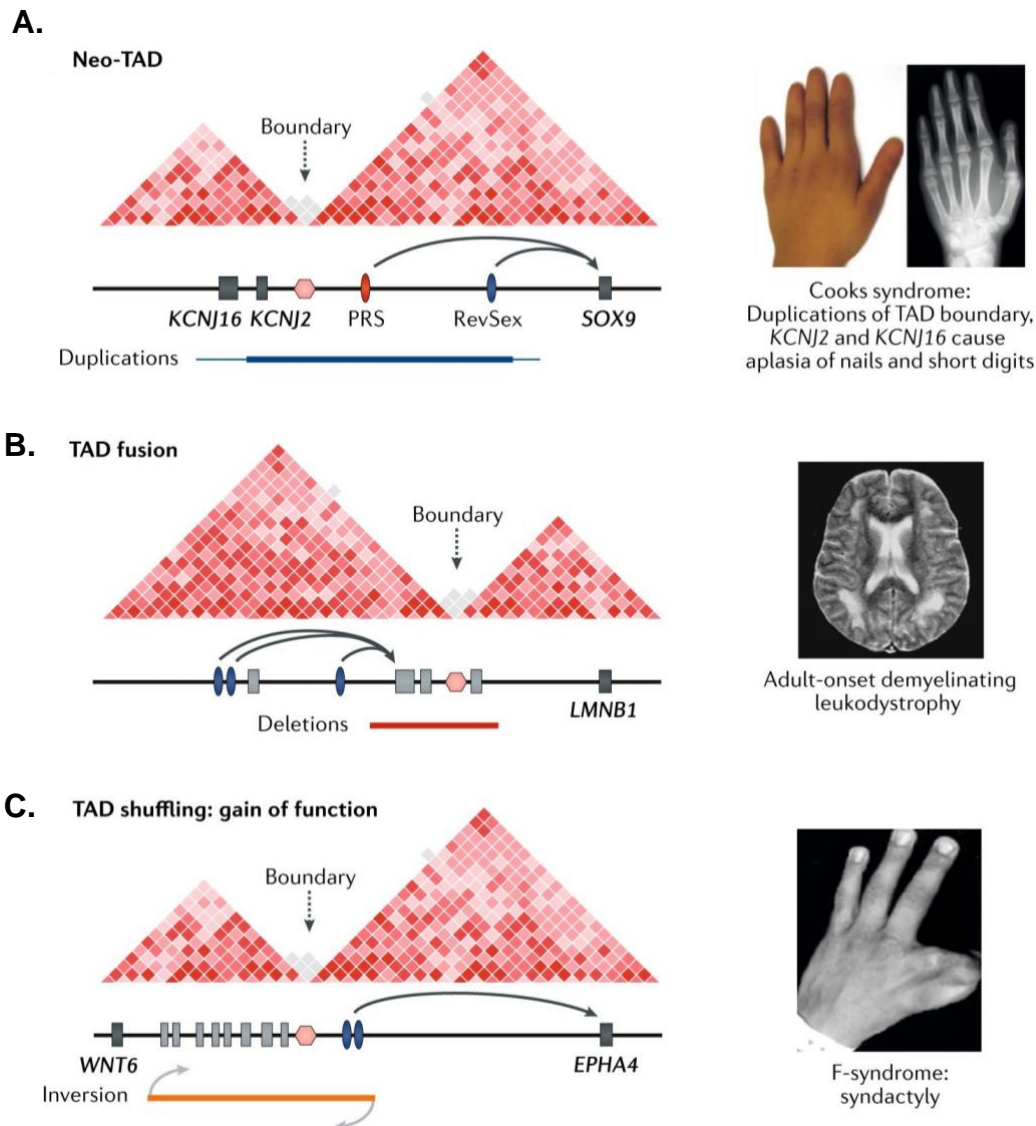


Figure 8. Clinical examples of chromosomal rearrangements that alter TADs. Pathologies are thought to result from ectopic gene activation in inappropriate tissues arising from different scenarios shown in A-C. **A.** TAD boundary duplication at the *SOX9* locus leads to a neo-TAD and causes Cooks syndrome. **B.** TAD boundary deletion causes enhancer adoption by *LMNB1* and causes adult-onset demyelinating leukodystrophy. **C.** Inversion of the *EPHA4* locus causes enhancer adoption and misregulation of *WNT6*, associated with F-syndrome. Figure source: Spielmann et al. 2018

On the other hand, some other studies contradict the coupling of chromatin topology and gene regulation. Therefore, topological control of gene expression may be highly locus-specific. Some lines of evidence that contradict the connection between genome architecture and gene regulation are listed below:

- Acute depletion of CTCF in mouse embryonic stem cells altered TAD organization, but altered the expression of very few genes (Nora et al. 2017).
- Loss of cohesin or cohesin loading factor caused TADs and Hi-C peaks to vanish globally, but there were no global transcriptional changes (Nora et al. 2017; Rao et al. 2017).
- *Sonic Hedgehog (SHH)* gene controls growth and patterning during embryonic development. Substantial alteration of the *SHH* TAD does not affect *SHH* expression (Williamson et al. 2019).
- Single nucleosome resolution of folding of the mammalian genome using Micro-C showed that fine-scale chromatin organization features are shaped by transcription, but higher-order structure remains intact upon inhibition of promoter melting and transcriptional elongation (Hsieh et al. 2020).

1.5 *Drosophila* as a model to study the fundamentals of genome organization

Drosophila provides novel insights into the principles of genome organization as folding principles appear different from mammals. Like mammals, the *Drosophila* genome is also organized into chromosomal territories, compartments, and TADs. However, these domains may be formed by different molecular mechanisms than mammals. *Drosophila* also has dCTCF – an ortholog of mammalian CTCF. CTCF is essential for embryonic development in mammals, but *dCTCF* null mutant flies lacking maternal and zygotic CTCF survive until the pupal stage (Gambetta & Furlong, 2018). CTCF and cohesin are required for mammalian TAD formation, but it was not clear whether *Drosophila* TADs arise from loop extrusion because most TAD boundaries are not bound by dCTCF (Uljanov et al. 2016; Wang et al. 2018a). Moreover, dCTCF and cohesin localization in *Drosophila* is still debatable; one study reported that 75% of dCTCF binding sites do not overlap with a cohesin subunit (stromalin) (Ramírez et al. 2018); whereas another study reported more extensive co-localization between dCTCF and the cohesin subunit Rad21 where nearly half of dCTCF sites colocalized with cohesin Rad21 (Bortle et al. 2014). Also, CTCF motifs at CTCF-occupied boundaries are not convergently oriented at fly TAD boundaries, in contrast to mammals (Rowley et al. 2017). Lack of evidence for a dCTCF mediated loop extrusion mechanism questioned the idea of a conserved mechanism of TAD formation even in

animals with CTCF. *dCTCF* knockdown-based experiments in *Drosophila* tissue culture cells did not reveal an apparent effect of dCTCF on transcription (Bartkuhn et al. 2009; Bortle et al. 2012; Schwartz et al. 2012). Although some cell line-based studies showed CTCF binding at TAD boundaries (Ramírez et al. 2018; Wang et al. 2018a), it remained unknown to what extent dCTCF defines genome architecture in flies. Therefore, assessing the biological role of CTCF in TAD formation and gene regulation was necessary to understand the fundamentals of genome organization.

1.6 *Drosophila* insulator binding proteins (IBPs)

Unlike mammals, where CTCF is the only known IBP, several IBPs have been identified in *Drosophila* including dCTCF, Cp190, Su(Hw), Ibf1, Ibf2, Pita, ZIPIC, Mod(mdg4), BEAF-32, and others. These IBPs were identified based on their ability to bind to insulators or facilitate insulator function mostly in transgenic reporters (Geyer and Corces 1992; Gerasimova et al. 1995; Zhao et al. 1995; Pai et al. 2004; Sultana et al. 2011). Some IBPs were identified as proteins co-purifying with a well-known IBP called Centrosomal protein 190 kDa (Cp190) in Cp190 purifications (Cuartero et al. 2014; Maksimenko et al. 2015; Zolotarev et al. 2017). Cp190 is also enriched at *Drosophila* TAD boundaries. Some known fly IBPs are schematized below (Fig-9).

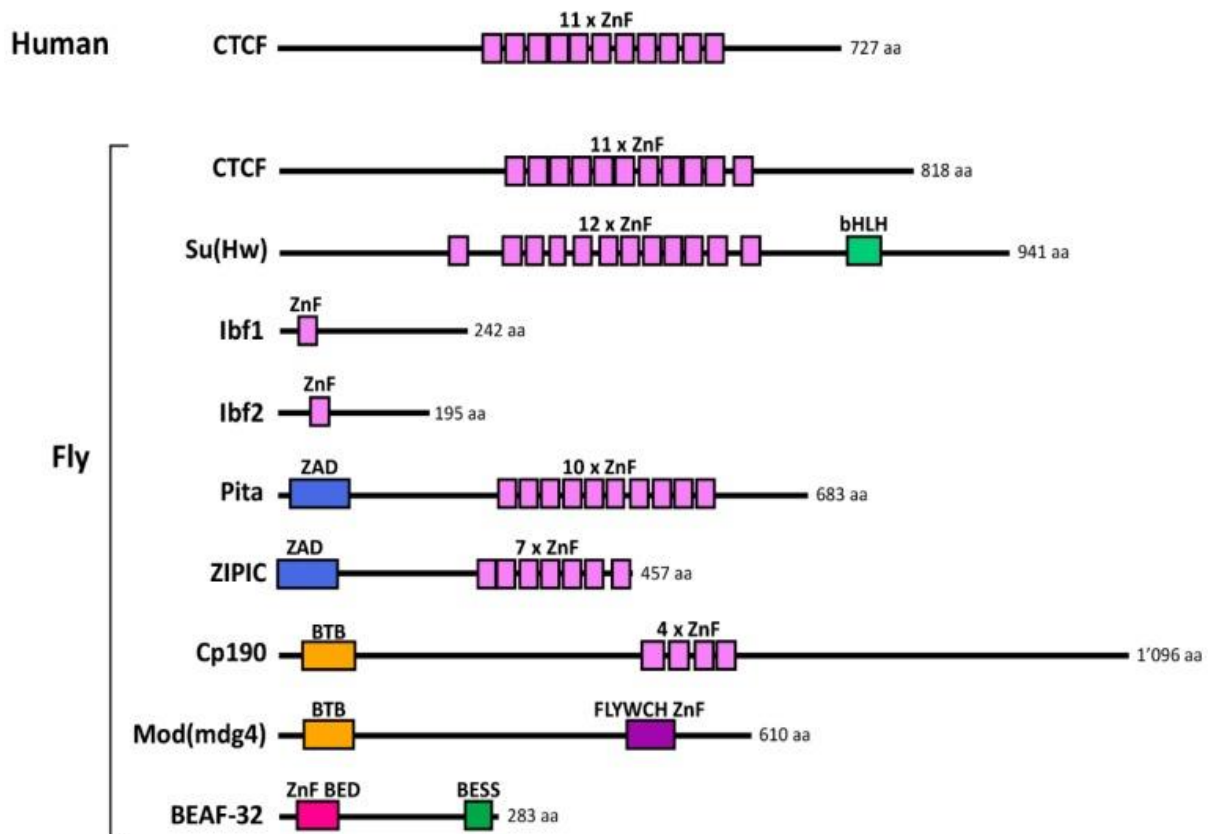


Figure 9. Insulator binding proteins. Humans only have one Insulator binding protein (IBP) - CTCF - known so far. Well-known *Drosophila* IBPs are schematized. Abbreviations are CTCF: CCCTC-binding factor; Su(Hw): Suppressor of Hairy wing; Ibf: Insulator binding factor; ZIPIC: Zinc-finger protein interacting with Cp190; Cp190: Centrosomal protein 190 kDa; Mod(mdg4): modifier of mdg4; BEAF-32: Boundary element-associated factor of 32 kDa; ZnF: zinc finger; ZAD: zinc finger associated domain; BTB: Broad-Complex, Tramtrack, and Bric-a-brac; bHLH: basic helix-loop-helix; BED: BEAF-32 and DREF; BESS: BEAF-32, Suvar(3)7 and Stonewall. Figure source: Özdemir and Gambetta 2019

IBPs are thought to exert various properties of insulators, briefly described below:

- *Blocking enhancer/silencer - promoter communication:* Insulators do not directly impact the functionalities of enhancers and silencers, but can block their influence on gene promoters when placed in between, even though enhancers and silencers regulate transcription through different mechanisms (Fig-10) (Kellum and Schedl 1991, 1992; Cai and Levine 1995; Sigrist and Pirrotta 1997; Mallin et al. 1998)
- *Blocking the spread of histone marks:* Some insulators are present between two different chromatin states with unique histone modifications (Fig-10). For example, studies of *Hox* gene clusters and the *even-skipped* (*eve*) gene locus showed that

insulators act as a barrier preventing histone modification spreading (Kahn et al. 2006; Fujioka et al. 2013; Bowman et al. 2014).

- *Facilitating long-distance enhancer-promoter communication:* The enhancer blocking and enhancer facilitating functions of insulators seem paradoxical. Nevertheless, insulators bridge developmental gene promoters to distal enhancers in flies and mammals to facilitate their communication (Fig-10) (Furlong and Levine 2018). For example, ‘Homing insulator of eve’, also known as the Homie insulator, was identified downstream of the *eve* gene. The name Homie insulator comes from “Homing pigeon” because transgenes containing Homie tend to integrate in the vicinity of the endogenous Homie insulator locus when remobilized by P element transposition (Fujioka et al. 2009). The Homie insulator was shown to mediate long-range enhancer-promoter communication between endogenous *eve* enhancers and reporter genes in a Homie-containing transgene inserted up to 140 kb away (Fujioka et al. 2009).
- *Insulator bypassing to avoid insulation activity:* There are also examples of fly insulators canceling each other out when they are present next to each other in a cis arrangement (Fig-10) (Cai and Shen 2001; Muravyova et al. 2001). When a pair of insulators are present between a RE and a gene promoter, the paired insulators no longer block RE-promoter communication (Kyrchanova et al. 2008).
- *Facilitating trans-interactions:* Pairing of homologous chromosomes can lead to transvection – an ability of regulatory elements on one chromosome to regulate the transcription of a gene on the other homologous chromosome (Fig-10) (Lewis 1954). Insulators have been shown to support trans interactions between polycomb group response elements (PREs) present on different chromosomes (Sigrist and Pirrotta 1997). Pairing between insulators has also facilitated enhancer communication in trans (Kravchenko et al. 2005). Insulators have also been reported to increase the stability of allele-allele pairing, possibly explaining how insulators allow enhancers to activate gene promoters in *trans* (Lim et al. 2018).
- *Influencing genome architecture:* Hi-C studies have shown that *Drosophila* TAD boundaries are demarcated by various insulator proteins complexes (Fig-10) (Ramírez et al. 2018; Wang et al. 2018a). However, their role in TAD boundary formation and maintenance remained speculative and was not directly addressed.

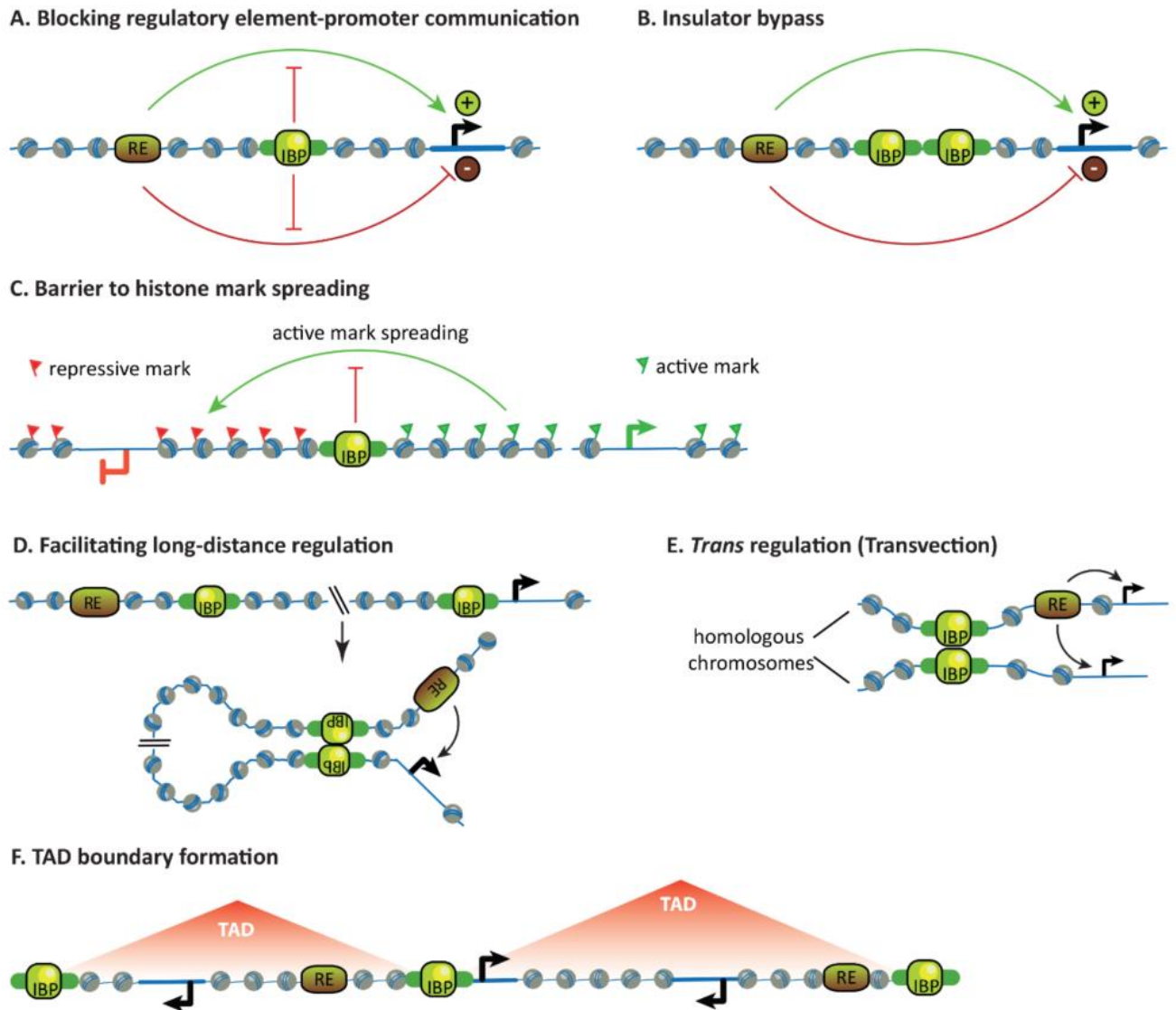


Figure 10. Properties of insulators. Insulators block the communication between regulatory elements and gene promoters (**A**), support bypassing of insulators when paired (**B**), act as a barrier to stop histone mark spreading (**C**), facilitate long-distance communication between regulatory elements and promoters (**D**), facilitate trans interaction of homologous chromosomes via transvection (**E**), and are enriched at TAD boundaries (**F**). Figure source: Özdemir and Gambetta 2019.

1.7 The role of Cp190 in insulation and genome organization

Other than CTCF, Cp190 appears to be the most critical IBP in the *Drosophila* genome because it seems to be a unifying cofactor, co-localizing at many sites bound by all other IBPs (Bartkuhn et al. 2009; Bushey et al. 2009; Nègre et al. 2010; Bortle et al. 2012; Schwartz et al. 2012). Cp190 was initially identified as a factor associated with centrosomes and microtubules, but later studies showed that it also binds to specific sites on polytene chromosomes (Whitfield et al. 1988, 1995). Cp190 was identified in

a genetic screen as an interacting partner of two IBPs Su(Hw) and Mod(mdg4)2.2 that bind to the *gypsy* insulator – a 350 bp DNA sequence of the *gypsy* transposon containing 12 copies of binding site for Su(Hw) (Pai et al. 2004). Cp190 interacting proteins like BEAF-32, Ibf1, Ibf2, Pita, ZIPIC and CTCF exert insulator activity (Cuartero et al. 2014) (Moshkovich et al. 2011; Maksimenko et al. 2015). Cp190 is not thought to bind DNA directly, and was instead shown to be recruited to chromosomes by some of these DNA binding IBPs (Gerasimova et al. 2007; Mohan et al. 2007; Schwartz et al. 2012; Soshnev et al. 2012). All these results suggested that Cp190 is important for insulator function, though the mechanism remained unclear. Several studies suggested that Cp190 bridges distal sites and thereby organizes the fly genome into chromosomal loops (Liang et al. 2014). For example, *in vitro* studies reported that Cp190 interacts with BEAF-32 via Cp190's C-terminal domain and homodimerizes via Cp190's N-terminal BTB/POZ domain (Vogelmann et al. 2014). This led to the popular hypothesis that DNA binding IBPs provide DNA specificity and recruit Cp190 to function as a “glue” for long range contacts (Liang et al. 2014; Vogelmann et al. 2014; Mourad and Cuvier 2016).

On the other hand, Cp190 not only binds to insulators but is also present at active transcriptional start sites (Bartkuhn et al. 2009). Studies in cell lines showed that Cp190 is present at nearly all TAD boundaries suggesting it may have a possible role in TAD boundary formation (Hou et al. 2012; Sexton et al. 2012; Ali et al. 2017; Bag et al. 2021). A ubiquitously expressed transcription factor, M1BP, was shown to interact with Cp190 and activate transcription at TAD borders (Bag et al. 2021) raising the question of whether Cp190 has a dual role in gene insulation and transcriptional activation depending on the TAD boundary.

Despite so many studies implying that Cp190 may be critical for TAD formation and gene insulation, our understanding of Cp190's role in these functions remained limited because of the lack of an appropriate *in vivo* model. To precisely understand the role of Cp190 in boundary formation and insulation, the impact of complete loss of Cp190 on endogenous gene transcription and global genome architecture needed to be determined.

1.8 Do TADs merely arise from transcription-related processes in flies?

Apart from IBPs, features of active transcription are associated with fly TAD boundaries. Most *Drosophila* TAD boundaries are enriched in active chromatin marks and divergently transcribed gene promoters (El-Sharnouby et al. 2016; Luzhin et al. 2019) (Chathoth and Zabet 2019). Ramírez et al. classified TAD boundaries as being in majority (75%) promoter promoters and the rest (25%) being non-promoter boundaries. By performing hierarchical clustering of various IBP DNA motifs present at boundaries, they found that promoter and non-promoter boundaries are differentially enriched in specific DNA motifs (Fig-12)(Ramírez et al. 2018). Single-cell Hi-C results in *Drosophila* suggested lower cell to cell variability in TAD profiles than mammalian TADs, possibly because many TAD boundaries correspond to active transcription (Ulianov et al. 2021). This makes sense as a population of clonal cells transcribe similar genes. Inhibiting transcription led to decreased Hi-C contacts inside TADs (Rowley et al. 2017). A Hi-C study on early developing embryos suggested that *Drosophila* TADs emerge during zygotic genome activation (Hug et al. 2017). Together, these studies suggested that transcription itself may be required to form fly TADs. On the other hand, other evidence suggested that TADs emerge independently of transcription. In early *Drosophila* embryos, tissue-specific gene expression patterns arise despite tissue-invariant chromatin conformations (Hug et al. 2017; Ing-Simmons et al. 2021). Another study on *Drosophila* early embryos showed that regulatory element-gene promoter looping appears before the emergence of TADs and zygotic genome activation (Espinola et al. 2021). Moreover, extensive chromosomal rearrangements present in balancer chromosomes that disrupt TADs, long-range loops, and promoter interactions do not strongly affect the expression for most genes (Ghavi-Helm et al. 2019). In short, conflicting observations from these studies reveal that the connection between gene transcription and TAD formation is complex, but transcription itself is likely insufficient to determine genome folding.

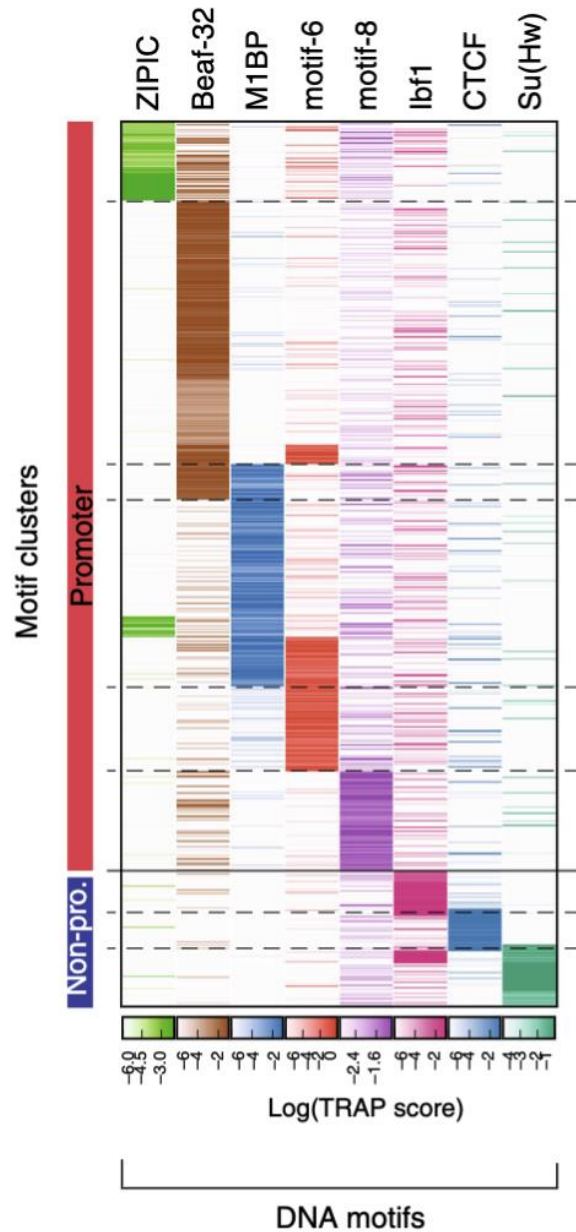


Figure 11. Differential enrichment of DNA motifs at fly TAD boundaries. Promoter and non-promoter boundaries are clustered based on motif binding affinities. Higher scores represent higher confidence IBP binding and dashed lines delineate the motif clusters. Figure source: Ramírez 2018

2 Project aims

The work reported in this thesis aimed to investigate how two essential IBPs, CTCF and Cp190 define genome architecture and affect gene regulation in *Drosophila*. For this, we used flies completely lacking (both maternal and zygotic) CTCF (*CTCF⁰*), Cp190 (*Cp190⁰*), or both factors (*double⁰*) combined. Hi-C, ChIP-seq, RNA-seq and RNA in situ hybridization techniques were deployed to characterize the molecular and developmental phenotypes of our mutants.

Specific project aims were:

- to describe the impact of CTCF and/or Cp190 loss on genome architecture;
- to identify the interacting partners of CTCF and Cp190 in *Drosophila* embryonic nuclear extracts;
- to test various CTCF and Cp190 bound genomic loci for insulator activity in a plasmid-based reporter assay;
- to identify genes that are misexpressed upon loss of CTCF or Cp190;
- to compare gene misexpression arising from Cp190 loss or deletion of Cp190 occupied insulators at a developmental gene locus;
- to test Cp190's role in insulation and long-range insulator pairing using a reporter system in transgenic flies.

3 First Publication

CTCF loss has limited effects on global genome architecture in *Drosophila* despite critical regulatory functions

Anjali Kaushal*, Giriram Mohana*, Julien Dorier*, Isa Özdemir, Arina Omer, Pascal Cousin, Anastasiia Semenova, Michael Taschner, Oleksandr Dergai, Flavia Marzetta, Christian Iseli, Yossi Eliaz, David Weisz, Muhammad Saad Shamim, Nicolas Guex, Erez Lieberman Aiden† & **Maria Cristina Gambetta†**

* Equal authors

† Corresponding authors

Status: Published in Nature Communications, 12 February 2021

DOI: <https://doi.org/10.1038/s41467-021-21366-2>

My contribution to the publication

I performed the western blotting mentioned in Supplementary figure 2a. I optimized and constructed the reporter plasmid for insulator screen shown in figure 4a. I prepared samples for Cp190 ChIP-seq (Figure 5-a-f). I performed RNA-FISH on embryos of various genotypes and obtained their images using confocal microscopy (Figure-6c and e). I also performed RNA-seq on *Cp190^{KO}* samples shown in figure-6. I performed pull-down of the negative control for mass-spectrometry analysis (Supplementary Figure 5a).

Note: Published article is attached in the annexed section 11.1.

3.1 Short Summary of the first publication

CTCF is an essential to form the majority of contact domain boundaries in mammals, by stalling chromatin loop extruding cohesin. In *Drosophila*, whether CTCF contributes to genome organization was debated, mostly because only a fraction of fly contact domain boundaries are occupied by CTCF. To understand the relevance of CTCF in fly genome organization and its impact on gene regulation, we studied *CTCF⁰* mutants completely lacking maternal and zygotic CTCF. We found that CTCF is required for fly viability and its expression in neurons is critical for pupal hatching. We showed that loss of CTCF affects only about 10% of contact domain boundaries and these affected boundaries can be categorized into two types: CTCF dependent (completely lost in *CTCF⁰* mutants) and partially CTCF dependent (retained but weakened relative to wildtype) boundaries. This showed that CTCF independent mechanisms form majority of fly boundaries.

Despite CTCF's limited role in contact domain boundary formation, CTCF was important for the correct expression of a subset of genes at CTCF-occupied boundaries. RNA-sequencing on wildtype and *CTCF⁰* mutant larval central nervous systems (CNSs) revealed that some genes not normally expressed in the nervous system were inappropriately expressed in *CTCF⁰* mutant nervous systems. Some other genes had reduced expression in *CTCF⁰* mutants. Investigating spatial misexpression patterns of selected genes by RNA fluorescent in situ hybridization (RNA-FISH) showed that different genes were misexpressed in various patterns in mutant CNSs. Our results demonstrated that CTCF regulates the expression of a subset of genes near CTCF-occupied boundaries in the CNS.

We also found that the boundary associated factor Cp190 directly binds to CTCF. We showed that CTCF recruits Cp190 to the many CTCF-occupied boundaries and coregulates a subset of genes together with CTCF.

In short, the work presented in this chapter revealed that – in stark contrast to mammals - CTCF plays a limited role in genome organization in *Drosophila*. CTCF nevertheless critically maintains correct gene expression in the *Drosophila* CNS. CTCF recruits Cp190 to CTCF-occupied contact domain boundaries and coregulates

a subset of genes together with Cp190. This demonstrated that CTCF can associate with a regulatory cofactor. This work also set the foundation for the next step in our research focused on understanding whether Cp190 was also required to form physical and regulatory contact domain boundaries.

4 Second publication

Essential role of Cp190 in physical and regulatory boundary formation

Anjali Kaushal*, Julien Dorier, Bihan Wang, Giriram Mohana, Michael Taschner, Pascal Cousin, Patrice Waridel, Christian Iseli, Anastasiia Semenova, Simon Restrepo, Nicolas Guex, Erez Lieberman Aiden, Maria Cristina Gambetta†

* Lead author

† Corresponding author

Status: Accepted for publication in Science Advances

My contribution to the publication

I was involved in the discussion for designing experiments. I optimized and performed CTCF and Cp190 ChIP-seq (Figure-1,2, and 3) and prepared samples for Hi-C (Figure-1,2, and 3). I also performed Cp190 pull-down for mass spectrometry analysis (Figure-4a). I generated CRISPR Knockout cassette for *SF1^{KO}* and *SF2B^{KO}* mutants (Figure-5). I performed RNA-FISH on embryos of various genotypes and obtained their images using confocal microscopy (Figure-5c, 6c, 7a and b). I gave input for manuscript writing.

Note: Published article is attached in the annexed section 11.2.

4.1 Short Summary

Here, we assessed the importance of Cp190 in *Drosophila* genome organization and gene regulation using mutant embryos completely lacking Cp190 (*Cp190⁰*), CTCF (*CTCF⁰*) or both factors combined (*double⁰*).

We found that Cp190 is the major factor required for contact domain boundary formation known to date, as one quarter of boundaries were lost in the absence of Cp190. Cp190 was specifically required to form boundaries that are distal to (i.e. more than 200 bp away from) a transcriptional start site (TSS). TSS-proximal sites are also occupied by Cp190 but were not affected by Cp190 loss. This suggests that Cp190 independent mechanisms form boundaries at TSS-proximal sites. Importantly, we found that Cp190 was critical for the formation of CTCF occupied boundaries. In other words, CTCF binding to DNA alone is not sufficient for boundary formation unlike what is thought to be the case in mammals. We showed that Cp190 assembles into several different multiprotein complexes in *Drosophila* nuclear extracts. Many of these Cp190 interacting proteins are known to bind to different classes of contact domain boundaries. We showed that various classes of Cp190 occupied boundaries (each bound by different Cp190-containing complexes) all show comparable enhancer-blocking activity in a plasmid-based insulator reporter assay. We further validated that Cp190 is essential for enhancer-blocking at developmental gene loci that become misexpressed in ectopic tissues due to regulatory crosstalk with nearby enhancers, in *Cp190⁰* mutant embryos. Inconsistent with preconceived notions in the field, however, Cp190 was only required for the enhancer-blocking activity of developmental gene contact domain boundaries we tested, but Cp190 was dispensable for the abilities of the contact domain boundaries we tested to mediate long-range enhancer-promoter communication. This demonstrated that the enhancer-blocking and -facilitating functions previously attributed to some fly contact domain boundaries are separable.

Collectively, our observations demonstrated that Cp190 is essential factor for *Drosophila* genome organization and to prevent regulatory crosstalk at critical developmental genes. Our work revealed that diverse mechanisms evolved to form physical and regulatory boundaries in animal genomes.

5 Discussion

Note: All the figures referred in this section are from the published article attached section in 11.2

5.1 Summary of findings

The following points summarize the overall conclusions from chapter 4 (article is in the annexed section) :

- Cp190 is essential for fly viability as *Cp190⁰* animals die during embryogenesis, a more severe phenotype than *CTCF⁰* flies (fig-1A).
- Cp190 associates into several multiprotein complexes with well-known IBPs in *Drosophila* nuclear extracts (Fig-4A).
- Genomic loci bound by Cp190 in diverse multiprotein complexes show similar insulator activity in a plasmid-based insulator reporter assay (Fig-4B).
- Cp190 is the major architectural protein described to date in flies as it is required to form one quarter of all contact domain boundaries (Fig-1B).
- Even though Cp190 binds to most contact domain boundaries, it is only required to form TSS-proximal boundaries but is dispensable to form TSS-distal boundaries (Fig-1B).
- Some boundaries are weakened but not lost in *Cp190⁰* mutants. This suggests that Cp190-independent mechanisms can also form boundaries (Fig-1B, E).
- Cp190-dependent and Cp190-independent boundaries are differentially bound by different DNA-binding IBPs (Fig-1B,D).
- Cp190 is present at CTCF-occupied boundaries and is required for their formation (Fig-2).

- Cp190 is recruited to some CTCF-occupied boundaries strictly CTCF-dependently, and is recruited to other CTCF-occupied boundaries at least partially CTCF-independently (Fig-3).
- Cp190 blocks *Scr* gene from inappropriate activation by nearby *ftz* enhancers in early embryos. In older embryos, Cp190 blocks crosstalk between an enhancer located 30 kb downstream of *Scr* gene and *Scr* promoter. Cp190 also prevents ectopic expression of other developmental genes, but the degree of misexpression caused by Cp190 loss varies from gene to gene (Fig-5, 6).
- Although insulators were proposed to mediate long-range enhancer-promoter communication (Fujioka et al. 2016), Cp190 is not essential for the regulation of the developmental genes we tested by their known long-range enhancers (Fig-6, S6-S7).
- Cp190 is critical for the enhancer-blocking activity of the Homie insulator but not for Homie-Homie insulator pairing that facilitates long-range enhancer-promoter communication in a transgenic reporter (Fig-7).

5.2 Diverse ways to form boundaries in the fly genome

CTCF forms a significant fraction of contact domain boundaries in mammals. Mammalian CTCF forms directional boundaries by stalling chromatin loop extruding cohesin when CTCF is bound in a specific orientation. We found that CTCF co-purified with cohesin subunits in *Drosophila* nuclear extracts (Kaushal et al. 2021) (Fig-4). We do not yet know, however, whether CTCF forms contact domain boundaries in flies by stalling cohesin. Cohesin-mediated loops extrusion in flies has yet to be demonstrated because of the following discrepancies:

- In mammals, CTCF sites are present at the majority of contact domain boundaries (Rao et al. 2014). However, CTCF is not the major boundary

associated factor in *Drosophila* (Wang et al. 2018b; Kaushal et al. 2021; Ulianov et al. 2021).

- CTCF and cohesin colocalize genome-wide in mammals (Wendt et al. 2008; Pugacheva et al. 2020). However, CTCF and cohesin do not specifically colocalize in flies (Cubebñas-Potts et al. 2017; Ramírez et al. 2018).
- CTCF-dependent boundaries of TADs have convergently orientated CTCF binding motifs in mammals, but this directionality is not present in flies (Rao et al. 2014; Rowley et al. 2017).

Our results support the notion that genome architecture in flies is more complex and requires additional factors like Cp190 to establish contact domain boundaries. Cp190 is critical for forming one quarter of contact domain boundaries as these boundaries are lost or weakened upon loss of Cp190 (fig-1B). The fact that some Cp190 occupied-boundaries are weakened but retained upon Cp190 loss suggests that Cp190-independent mechanisms also form boundaries at these sites.

We find two classes of contact domain boundaries: (1) TSS-distal boundaries that rely on Cp190, several of which are severely affected upon loss of Cp190; (2) TSS-proximal boundaries that are present near active TSS (within ± 200 bp) and are also occupied by Cp190 but not affected by Cp190 loss. Our DNA motif enrichment analysis for various DNA binding IBPs showed that Cp190 occupied TSS-distal boundaries to be enriched in CTCF, Ibf1, and Su(Hw) motifs. Cp190 occupied TSS-proximal boundaries are more enriched in BEAF-32, M1BP, and ZIPIC motifs. These observations further support our findings that Cp190 is recruited to some TSS-distal boundaries by CTCF and Su(Hw) (Fig-1B, D, S4).

How are TSS-proximal boundaries formed? It is still unclear what factors drive the formation of boundaries at TSS-proximal sites. The most common speculated factors are active transcription, chromatin modifiers, RNA polymerase II, promoter-associated factors or other insulator binding proteins like BEAF-32 (Li et al. 2015; Bonev et al. 2017; Hug et al. 2017; Rowley et al. 2017). Our results show that despite having Cp190 at TSS-proximal sites, Cp190 is dispensable for their boundary formation. So the

question arises, what role is Cp190 playing at these sites if it is not involved in boundary formation? The results from the plasmid-based assay show that all tested DNA fragments representing loci bound by diverse Cp190-containing complexes all show insulation activity. Therefore, it is possible that at TSS-proximal loci, instead of boundary formation, Cp190 might be required for insulation to prevent regulatory crosstalk between active gene promoters and nearby enhancers.

In wildtype embryos, CTCF ChIP peaks with high and low CTCF occupancy are present at contact domain boundaries. These CTCF occupied boundaries also show Cp190 co-occupancy (Fig-2). In *CTCF⁰* mutants, boundary retention at previously CTCF-occupied boundaries correlates with Cp190 presence (Fig-3). This suggests that there are also CTCF-occupied boundaries at which CTCF-independent mechanisms recruit Cp190. Most CTCF-occupied boundaries are lost or weakened in *Cp190⁰* mutants. This suggests that Cp190 plays an important role in boundary formation at CTCF-occupied boundaries.

Unlike CTCF high and low occupancy ChIP peaks, we observed that CTCF intermediate peaks did not colocalize with contact domain boundaries or Cp190 in wildtype embryos. We do not know why CTCF standalone peaks are not present at contact domain boundaries. These standalone sites may exert different activity than CTCF and Cp190 co-bound sites in the genome. As previously described, CTCF standalone peaks were frequently present in introns (Fig-S2 C) (Schwartz et al. 2012). Interestingly, a few of these CTCF standalone loci have also been shown to possess no insulation or repression activity in a reporter based study (Schwartz et al. 2012).

Collectively, these results show that Cp190 plays a decisive role in boundary formation in flies as more than one quarter of contact domain boundaries are lost or weakened upon loss of Cp190. Several independent mechanisms recruit Cp190 to different contact domain boundaries.

5.3 Cp190 prevents promiscuous regulation of several tested genes

We had previously shown that loss of CTCF or Cp190 resulted in misexpression of some genes at CTCF-dependent contact domain boundaries (Kaushal et al. 2021). It was difficult to hypothesize if misexpression was due to loss of insulation because we could not determine the regulatory elements driving these misexpressions. Therefore, to systematically understand the cause of misexpression, we focused on the well-characterized HOX gene locus. HOX genes are critical developmental genes that are essential to specify the anterior-posterior axis of flies. The ANT-C is one of the two HOX gene complexes. *ftz* is a non-HOX gene present between *Scr* and *Antp* HOX genes. *ftz* is a pair-rule gene whose early embryonic expression is driven by stripe enhancers contained within the *ftz* contact domain. *Scr* was misexpressed in dynamic spatio-temporal patterns in *Cp190⁰* mutants. *ftz* stripe enhancers caused ectopic expression of *Scr* gene in a *ftz-like* pattern in early *Cp190⁰* embryos despite the fact that a *ftz* contact domain boundary was retained between *Scr* and *ftz* genes (Fig-5A, C). *Scr* misexpression was more severe in *SF1^{KO}* compared to *Cp190⁰* mutants, indicating that boundary deletion has a more severe effect than loss of Cp190 (Fig-5C). This further suggests that additional factors are present on SF1 DNA that also insulate *Scr* promoter from *ftz* enhancers. In older *Cp190⁰* embryos, *Scr* was misexpressed in hindgut and anal plate tissues (Fig-6C). Molecular analysis of regulatory elements around *Scr* had previously identified an enhancer located 30 kb downstream of *Scr* promoter that drives an *Scr-LacZ* fusion gene in a non *Scr*-like patterns in the hindgut and anal plate (Gindhart et al. 1995). Cp190 binds between *Scr* gene and these enhancers in wildtype embryos (Fig-6B), likely to prevent regulatory crosstalk between them. We also observed *Dfd* misexpression in the nervous system of *Cp190⁰* mutants in addition to its wildtype expression. However, it was challenging to identify the causative enhancer due to the presence of several neuronal enhancers around *Dfd* gene. In contrast, we found no effect of loss of Cp190 on *Antp* gene as it was normally expressed in *Cp190⁰* mutants. The diverse effects of loss of Cp190 at the ANT-C locus reveal that Cp190 is critical for preventing ectopic activation of *Scr* and *Dfd* genes but it is dispensable for correct expression of the *Antp* gene in embryos. We speculate that other factors may block ectopic activation of the *Antp* gene, or alternatively *Antp* promoter may not be susceptible to activation by inappropriate local enhancers.

The BX-C HOX locus defines the identity of the posterior thorax and abdominal segments of the fly. The tissue-specific expression of BX-C HOX genes is established by segment-specific regulatory domains containing their respective enhancers and silencers. Regulatory domains act independently, and boundaries between them overlap with genetically identified insulators. Deleting these boundaries results in regulatory domain fusion that leads to misexpression of BX-C genes (Mateo et al. 2019) (Maeda & Karch 2015). Cp190 binds to these boundaries along with various combinations of other IBPs, including CTCF. Surprisingly, we observed very mild effects of Cp190 loss on BX-C gene expression and contact domain boundaries (Fig-S7). There was subtle misexpression of *Ubx* and *abd-A*, whereas *Abd-B* was expressed normally in *Cp190⁰* mutants (Fig-S7B). However, *Abd-B* is misexpressed in *CTCF⁰* mutants (Gambetta and Furlong 2018). Our observations at the BX-C locus suggest that additional factors present on BX-C boundaries exert insulation independently of Cp190 to ensure proper BX-C gene expression.

Importantly, we did not observe any loss of endogenous expression of all tested genes. This indicates that Cp190 plays a role in preventing inappropriate regulatory crosstalk but not in facilitating endogenous expression of these tested genes. However, due to the limited number of tested genes, Cp190's role in controlling endogenous expression and insulation remains to be explored at other genes.

In short, Cp190 loss mildly affects contact domain boundaries at HOX loci but nevertheless resulted in ectopic activation of some HOX genes by neighboring enhancers. However, Cp190 might not be the only factor required for gene insulation. The variable effects of Cp190 loss on HOX genes indicate that additional factors are also able to block regulatory crosstalk of HOX genes independently of Cp190.

5.4 Cp190 is neither required for HOX gene nor Homie-mediated distal activation

Insulators have been thought to perform two opposite functions - to block crosstalk between genes and regulatory elements, and facilitate long-distance enhancer promoter communication. These two features of insulators pose a paradox: how do

insulators prevent regulatory crosstalk but at the same time jump over intervening loci to ensure specific long distance regulation? The unified molecular mechanism that would explain these two features of insulators is unclear. The “looping model” was proposed to account for these opposite activities of insulators, whereby two insulators pair with each other to form a chromatin loop that not only brings distal enhancers close to their target genes but also blocks regulatory crosstalk with non target enhancers present on another loop (Fujioka et al. 2016). Insulator replacement experiments at the BX-C locus have shown that the boundaries between BX-C regulatory domains can block crosstalk between adjacent regulatory domains and also facilitate long range activation of HOX gene promoters. This has been shown in the context of the BX-C *Mcp* insulator which acts as a boundary between regulatory domains that control the expression of *abd-A* and *Abd-B* genes. Replacement of *Mcp* with *Fab-7* or *Fab-8* insulators (that are normally found between *Abd-B* regulatory domains) caused inappropriate activation of the *Abd-B* promoter by the distal *abd-A* regulatory domain (Postika et al. 2018).

Yet we did not observe any impairment of long-range enhancer mediated activation of any of the genes we tested upon loss of Cp190. In the case of *Scr*, we only observed loss of endogenous *Scr* expression in early embryos upon SF1 boundary deletion, but not upon loss of Cp190. Others have also reported loss of *Scr* expression upon SF1 boundary deletion (Yokoshi et al. 2020). This phenotype was interpreted as a consequence of the loss of communication between the *Scr* promoter and a hypothetical distal enhancer located 35 kb upstream of *Scr* gene beyond the *ftz* contact domain (Fig-x). Yokoshi et al. proposed that SF1 and SF2b pairing is critical to loop the *ftz* contact domain out to avoid *Scr-ftz* regulatory crosstalk and facilitate *Scr* promoter interaction with its distal enhancer (Yokoshi et al. 2020). Recently, work from Batut et al. proposed that additional DNA sequences known as tethering elements foster physical interactions between *Scr* promoter and its distal enhancer (Batut et al. 2022). Tethering elements are functionally different from insulators, contact domain boundaries and enhancers.

Our interpretation of these results is different. We find that the SF1 boundary overlaps an enhancer that drives an *Scr-like* expression pattern in early embryos (Fig-S5B)(Kvon et al. 2014). Moreover, *Scr* expression was normal in an SF2b deletion

that deletes the other *ftz* contact domain boundary (Fig-5C). Therefore, we find that both Cp190 and an eventual pairing of *ftz* contact domain boundaries are dispensable to drive endogenous *Scr* expression in early embryos. The relevance of *Scr*'s putative distal enhancer in *Scr* transcription activation remains unclear because neither deletion of the tethering element nor *Scr* distal enhancer deletion abolished *Scr* expression in early embryos (Batut et al. 2022). As discussed previously, *Scr* is expressed in a *ftz*-like pattern upon loss of Cp190, which shows that Cp190's role is to prevent local regulatory cross-talk rather than bridge *Scr* to a distal enhancer (Fig-5 and 6).

In addition, BX-C gene expression driven by long-range enhancers was unaffected in *Cp190⁰* mutants. This further shows that Cp190 is not critical for long-range enhancer-promoter communication at these loci. Kyrchanova et al. showed that replacement of the boundary present between two *Abd-B* regulatory domains (*iab-6* and *iab-7*) by different insulators that recruit known IBPs like Pita, Su(Hw), and CTCF, blocked regulatory crosstalk between the regulatory domains but failed to support long-range regulation by *iab-6* of its target gene *Abd-B*. However, upon replacement of the boundary by a modified version that does not recruit IBPs but recruits another multiprotein complex known as the Late Boundary Complex, long-range regulation by *iab-6* of its target gene *Abd-B* was restored. This suggests that the ability of HOX boundaries to support long-range enhancer-promoter communication relies on factors other than insulator binding proteins (Kyrchanova et al. 2019).

We further showed complementary observations using the Homie insulator in a transgenic fly. In this transgenic fly, the Homie insulator placed in between divergently transcribed reporter genes (*GFP* and *LacZ*) was integrated ~140 kb upstream of the *even skipped* (*eve*) gene locus with local endogenous *eve* enhancers. The endogenous Homie insulator is present downstream of the *eve* gene (Fig-7A). Previous studies described that Homie insulators physically pair in a head-to-head orientation and bring *eve* enhancers close to the transgenic reporter genes to activate their expression (Fujioka et al. 2016)(Chen et al. 2018) Since Cp190 binds to the Homie insulator, we further showed that Cp190 is critical for Homie mediated enhancer-blocking activity but not for Homie mediated activation of the transgenic reporter genes by distal *eve* enhancers (Fig-7B). However, these are limited

examples, and we cannot rule out that Cp190 facilitates long-range enhancer-promoter communication at other loci.

6 Conclusion

Whether and how the three-dimensional organization of the genome ensures the correct spatial-temporal expression of genes are fundamental questions that have been extensively studied in the last two decades. Insulators and their binding proteins were hypothesized to play a critical role in the 3D organization of the fly genome, but previous studies mostly studied the properties of a few transgenic insulators. On the other hand, previous genome-wide studies had partially knocked-down IBPs in *Drosophila* cell lines and failed to reveal biologically interpretable transcriptional changes. Thus, our understanding of the roles of insulators in shaping the genome and affecting gene expression remained limited.

The knowledge derived from my studies highlight the essential role of Cp190 in forming physical and regulatory contact domain boundaries. Our work lays a solid foundation to further our understanding of fundamentals of genome organization. Several questions remain to be answered. For example, the role of Cp190 at TSS-proximal boundaries is still unclear. Profiling genome-wide transcriptional defects in *Cp190⁰* mutants for example by single cell RNA sequencing could reveal how gene expression is affected around both TSS-distal and TSS-proximal boundaries.

The presence of Cp190 at both TSS-distal and TSS-proximal boundaries raises interesting questions. What makes Cp190 critical for boundary formation at TSS-distal boundaries? Are there different Cp190 mediated molecular mechanisms that govern boundary formation and enhancer-blocking activity? Future studies addressing these questions would further our understanding of how these boundaries form differently.

Our work highlights the essential role that Cp190 plays in contact domain boundary formation and maintaining tissue specific expression pattern. This improves our current understanding of *Drosophila* genome. More broadly, our finding that diverse mechanisms exist to partition genomes into independent domains (beyond CTCF-dependent mechanisms previously described in mammals) highlight that it will be exciting to explore whether analogous mechanisms also exist in other species.

7 Methods

7.1 Chromatin preparation from fly embryos

- 0-12 hours old embryos were dechorionated in 50% bleach solution for 2 minutes at room temperature and washed thoroughly under running tapwater until the smell of bleach was completely gone.
- These dechorionated embryos were snap frozen and stored in -80°C until we reached the desired volume.
- Approximately 100 μl volume of embryos per biological replicate (three biological replicates per genotype) were collected.
- After collecting the desired volume of embryos, these embryos were dounce homogenized with 15 strokes in an ice cold 15 ml dounce homogenizer in 5 ml crosslinking solution and transferred to a 15 ml falcon tube and rotated at room temperature until 12 minutes.
- Nuclei were pelleted by centrifuging the falcon tubes for 2 minutes at 2000 g (room temperature) and supernatant was discarded.
- Nuclei were washed 10 minutes in 1 ml stop solution by rotation in a 1 ml tube.
- Nuclei were again pelleted by centrifuging 2 minutes at 2000g at room temperature and stop solution was discarded.
- The same steps of nuclei washing and pelleting by centrifugation were repeated with solution A and then with solution B.
- After these thorough washes, nuclei pellet was sonicated in 100 μl RIPA buffer in AFA microtubes in a Covaris S220 sonicator for 5 minutes with a peak incident power of 140 W, a duty cycle of 5% and 200 cycles per burst.
- Sonicated chromatin was centrifuged to pellet insoluble material and supernatant was snap-frozen and stored in -80°C .

7.2 ChIP-seq

- ChIP was performed by incubating 40 μl chromatin with 2 μl of rabbit polyclonal antibody against CTCF¹⁻²⁹³ or Cp190¹⁻¹⁰⁹⁶ (Kaushal et al. 2021) overnight at 4°C .

- To this, 25 μ l of pre-mixed Protein A and G Dynabeads (Thermo Fisher 100-01D and 100-03D) were added for 3 hours at 4°C.
- These beads were washed for 10 minutes with RIPA with 140 mM NaCl using a magnetic rack. The same washes were done four times with RIPA with 500 mM NaCl, once with LiCl buffer and twice with TE buffer.
- Beads were resuspended in 100 μ l of TE for further steps.
- RNA from these samples and from 10 μ l of the chromatin as input (90 μ l of TE was added to make up 100 μ l volume) was removed by incubating 2 μ l of RNase (10ng/ μ l) for 30 minutes at 37°C shaking at 750 rpm.
- These samples were treated with proteinase K digestion by adding 3 μ l of 20% SDS (0.5% final) and 12 μ l of 5 μ g/ μ l proteinase K at 37°C and reverse crosslinked for 6 hours at 65°C shaking at 750 rpm in a programmable thermoblock.
- These samples were placed on a magnetic rack and supernatant was transferred into a new tube.
- DNA was purified using QIAGEN Minelute polymerase chain reaction (PCR) purification kit and following steps were performed:
 - Eluates from the previous step was mixed with the 600 μ l of PB buffer and 10 μ l of 3M sodium acetate pH 5.2.
 - This was loaded on minelute column stored at 4°C, centrifuged maximum speed for 1 minute at room temperature and flow through was discarded.
 - The column was incubated 5 minutes with 750 μ l of PE buffer and centrifuged maximum speed for 1 minute at room temperature. This step was repeated one more time.
 - Minelute column was further dried by centrifuging for 1 minute at room temperature to avoid any PE buffer carryover.
 - Purified DNA from column was eluted by incubating pre-warmed 30 μ l of 10 mM Tris-HCL pH 8.5 for 1 minute at room temperature and centrifuging maximum speed.
 - This step was repeated by reloading the eluted sample on the same column to further increase the DNA concentration.

- After elution, samples were stored at 4°C.
- ChIP-seq libraries were prepared using the NEBNext Ultra II DNA Library Prep kit for Illumina (method discussed in other section).
- An 10 nM equimolar pool of multiplexed ChIP-seq libraries at 4 nM was sequenced on 2 lanes of an Illumina HiSeq4000 150 bp paired-end.

7.3 ChIP-seq data analysis

- Paired-end ChIP-seq reads were demultiplexed and mapped to dm6 genome using Micmap-derivative of fetchGWI versatile tool for rapidly searching multiple genomes that index database or query set (Iseli et al. 2007).
- Some samples were sequenced twice and their reads were merged using Samtools v1.10 that manipulates alignment in SAM, BAM, and CRAM formats. This tool performs sorting, merging and indexing, and can retrieve reads in any region swiftly.
- Only sequencing data from chromosome 2, 3, 4, and X were used. ChIP-seq peaks were called using R package csaw v1.16.1 using window width of 20 bp and spacing of 10 bp. Duplicate reads and blacklisted regions were ignored (Amemiya et al. 2019). The removal of the ENCODE blacklist is an essential quality measure when analyzing functional genomics data. Filtering of these regions is crucial before applying any threshold, normalization, or peak calling as this can significantly bias the results (Carroll et al. 2014).
- Background enrichment of reads was calculated as the median for overall samples in comparison of average reads per 2 kb bins and windows with less than 2 fold enrichment over background were filtered out.
- Data were normalized using TMM method in csaw, a Bioconductor package for differential binding analysis of ChIP-seq data (Lun and Smyth 2016).
- Results obtained were clustered by combining adjacent windows and combined p-values were calculated using csaw and Benjamini-Hochberg method, a tool that decreases the false discovery rate.
- Regions that had false discovery rate (FDR) < 0.01 and |best.logFC| > 1 were identified as differentially bound regions. Genuine Cp190 peaks were called by

differential analysis of ChIP-seq signals as follows: WT versus *Cp190⁰* for Cp190 peaks in WT ; *CTCF⁰* versus *Cp190⁰* for Cp190 peaks in *CTCF⁰*. Similarly, genuine peaks CTCF were identified in WT and *Cp190⁰*.

- One of the replicate for CTCF ChIP-seq in *Cp190⁰* failed therefore analysis were performed with two replicates.
- ChIP occupancy was defined as the best.log2FC obtained from csaw. Overlapping ChIP-seq peaks were defined as peaks that share at least 1 bp and likewise overlapping DNA motifs were also defined as those share at least 1 bp with DNA motif.
- CTCF or Cp190 bound contact boundaries were defined as those having a ChIP peak within ± 2 kb of the boundary.
- Promoter proximal ChIP peaks were defined as those present within ± 200 bp of a transcriptional start site and peaks away from ± 200 bp of any transcriptional start site were defined as promoter distal peaks.
- CTCF peaks within an intron were identified using gene annotation from FlyBase release FB2020_06.

7.4 Library preparation using NEBNext Ultra II DNA Library Prep kit for Illumina

End repair and A tailing

- Following components were mixed by pipetting up and down ten times to the samples:

	Volume for 30 μ l (≤ 100 ng input DNA)
sonicated DNA	25 μ l
nuclease-free H ₂ O	0 μ l
NEBNext Ultra II end prep reaction buffer	3.5 μ l
NEBNext Ultra II end prep enzyme mix	1.5 μ l

- This reaction mixture was mixed by pipetting up and down 10 times and given a quick spin.
- The reaction mixture was placed on thermocycler: heat lid to 75°C → 20°C 30 min → 65°C 30 min → hold 4°C

Adaptor ligation

Appropriate indexed adapters were added by using following components:

	vol for 46.75 μ l (\leq 100 ng input DNA)
end prep reaction	30 μ l
NEBNext Ultra II ligation master mix	15 μ l
NEBNext ligation enhancer	0.5 μ l
annealed adapters	1.25 μ l of 15 μ M (400 nM final)

- The reaction mixture was mixed thoroughly by pipetting up and down, given a quick spin, and incubated at 20°C for 15 minutes.

Clean-up adapter-ligated DNA

- In the adapter ligated DNA 42 μ l (0.9 volume) of SPRIselect beads were added, vortexed briefly and incubated for 10 minutes at room temperature.
- Samples were quickly spun, incubated on magnetic rack for 3 minutes and supernatant was discarded.
- Beads were washed twice with 200 μ l of 80% ethanol and air-dried for 5 minutes to avoid any carryover of ethanol in next steps.
- Beads were incubated for 2 minutes with 102 μ l (for Hi-C) or 18 μ l (for ChIP-seq) of 10mM Tris pH 8 to elute at room temperature.
- Samples were given a quick spin and placed on magnetic for 5 minutes.
- 100 μ l (for Hi-C samples) or 18 μ l (for ChIP-seq samples) of supernatant was transferred to a new 0.5 ml DNA LoBind tube.

PCR amplification

Following components were added to the purified adapter ligated DNA to perform PCR amplification:

	[stock]	[final]	dilution	vol for 50 μ l / sample
KAPA HiFi Fidelity 5 MM	2 x	1 x	5 x	10 μ l
Primer PE 1.0 and 2.0	10 μ M	1 μ M	10x	10 μ l
adapter-ligated DNA	10 μ M	1 μ M	10x	27.5 μ l

- The reaction mixture was pipetted up and down 10 times, given a quick spin, and placed on thermocycler with parameters as heat lid 110°C → 98°C 45 sec → 8-12x (98°C 15 sec → 58°C 30 sec → 72°C 30 sec) → 72°C 1 min → 10°C hold
- PCR clean-up was performed using SPRI select beads exactly as described in the previous step and amplified DNA was eluted in 20 μ l of 10 mM Tris pH7.5 in a 0.5 ml LoBind tube.

7.5 Hi-C

Sample preparation

- Approximately 100 2-6 hours old embryos (four replicates per genotype) were dechorionated in 50% bleach for 2 minutes at room temperature and thoroughly washed under running tap water until smell of bleach was gone.
- These embryos were transferred to an eppendorf tube and washed twice with 1 ml RPMI supplement with 10% fetal bovine serum (FBS).
- Embryos were thoroughly crushed using a micro-pestle until the solution became cloudy and volume was increased to 500 μ l.
- The 500 μ l cell was transferred in a 15 ml polypropylene falcon tube and leftover from eppendorf tube was also transferred to the same 15 ml falcon tube by washing the eppendorf twice with 1 ml RPMI+ 10% FBS
- 7.5 ml of RPMI+10% FBS was added to the 15 ml falcon tube to make up the volume to 10 ml.
- Cells were palleted by centrifuging at 300 g for 6 minutes.
- 667 μ l of freshly opened 16% paraformaldehyde was added to fix the cells and incubated for exactly 10 minutes at room temperature.
- To quench the unreacted paraformaldehyde, 930 μ l of 2.5M glycine was added and incubated for 5 minutes at room temperature.
- Cells were palleted by centrifuging at 300 g for 6 minutes at 4°C.

- Supernatant was discarded and pallet was carefully washed with ice cold 10 ml PBS.
- After washing, cells were again palleted down by centrifuging at 300 g for 6 minutes at 4°C and PBS was removed without disturbing the pallet.
- Pellet in falcon was stored in -80°C.

Simultaneous restriction fill-in ligation

- Samples stored at -80°C were thawed on ice for 1 minute/
- 1 ml of ice cold lysis buffer was added on the samples, mixed by flicking, and incubated at 4°C for 30 minutes.
- After incubation, samples were centrifuged 3 minutes at 500 g at 4°C and supernatant was discarded leaving roughly 20 µl to avoid any pallet loss. 700 µl of ice cold lysis buffer was added and pallet was detached by strictly pipetting up and down only three times..
- The pallet was transfer to a 1 ml eppendorf tube and left over cells in the falcon tube were further transferred to eppendorf by adding 300 µl of lysis buffer back into the falcon tube.
- Lysis buffer was removed by pelleting the cells down for 3 minutes at 500 g at 4°C.
- Pellet was incubated for 10 minutes at 62°C with 50 µl of 0.5 % SDS solution to permeabilized the cells.
- To quench the remaining SDS, cells were further incubated 15 minutes at 37°C with 160 µl of water and 30 µl of 10% Triton X-100.
- 28 µl of of 10xT4 DNA ligase buffer was mixed by flicking and cells were pelleted by centrifuging at 500 g at 4°C and supernatant was removed carefully without disturbing the pellet.
- Following components were added by preparing a master mix and 65 µl of master mix was added to the pellet to perform simultaneous restriction, fill in, and ligation:

	[stock]	[final]	dilution	vol for 70 µl final
H ₂ O				29.5 µl
PEG4000				
T4 DNA ligase buffer	10 x	1 x	10 x	7 µl

Triton X-100	10%	0.5%	20 x	3.5 μ l
biotin-11-dUTP	1 mM	114 μM	8.75 x	8 μ l
dATP+dCTP+dGTP premix	1 mM each	57 μM	17.5 x	4 μ l
DNA polymerase I, large (Klenow) fragment	5 U / μ l		10 U	2 μ l
MseI	50 U / μ l	50 U	50 U	1 μ l
Csp6I	50 U / μ l	500 U	250 U	5 μ l
(Hi-) T4 DNA ligase				5 μ l
(nuclei pellet				5 μ l)

Reaction was set up in a thermocycler with following parameters: heat lid 65°C → hold 37°C → 70 x [37°C 2min → 16°C 5 min]

DNA purification and library preparation

- 150 μ l of 10 mM Tris pH 8 was added make the reaction mixture less viscous and centrifuged at 500 g at 4°C for 5 minutes. Supernatant was carefully removed without disturbing the pellet.
- Following components were added to do the proteinaseK digestion:

	[stock]	[final]	dilution	vol for 100 μ l final
(nuclei pellet				10 μ l)
Tris pH 8	10 mM	4 mM	2.3 x	79 μ l
SDS	10%	0.3%	30 x	3 μ l
NaCl	5 M	150 mM	30 x	3 μ l
(proteinase K	0.8 U / μ l			5 μ l)

This reaction mixture was incubated at 37°C for first half of the night and then set for reverse crosslinking by incubation at 65°C for 6 hours.

- After decrosslinking, samples were sonicated in AFA microtubes with following settings to aim for 300-500 bp:

Covaris	S220	E220
peak power	70	175
duty factor	20	10
cycles per burst	500	200

duration	110 sec	140 sec
temperature range	18-22°C	18-22°C
vol	120 µl	100 µl

DNA clean-up

- Sonicated DNA was transferred to a 1.5 ml eppendorf and 100 µl of SPRI select beads were added, vortexed briefly and incubated for 10 minutes at room temperature.
- Samples were quickly spun, incubated on magnetic rack for 3 minutes and supernatant was discarded.
- Beads were washed twice with 200 µl of 80% ethanol and air-dried for 5 minutes to avoid any carryover of ethanol in next steps.
- Beads were incubated for 2 minutes with 28 µl of 10mM Tris pH 8 to elute at room temperature.
- Samples were given a quick spin and placed on magnetic for 5 minutes.
- 26 µl (for ChIP-seq samples) of supernatant was transferred to a new 0.5 ml DNA LoBind tube.

Library preparation for next generation sequencing

Libraries were prepared using NEBNext Ultra II DNA Library Prep kit for Illumina kit and similar protocol was followed as mentioned in previous section but an additional step was added which is mentioned below.

Biotin pull down was performed after adapter-ligation purification step

- 50 µl (per sample) of Dynabeads MyOne Streptavidin T1 beads were aliquoted in 1.5 ml eppendorf tubes and storage solution was removed by placing the eppendorf on a magnetic rack.
- T1 beads were resuspended in 1 ml 1x Tween Binding and Washing Buffer (1xTWB) and incubated at 55°C for 10 minutes at 650 rpm.
- 1x TWB was removed from the beads by putting them on a magnetic rack for 1 minute.
- Beads were washed twice with 500 µl of 1x TWB and finally resuspended in 100 µl of 1x TWB.

- These beads resuspended in 1x TWB were added to the samples and incubated at 55°C for 15 minutes at 650 rpm to allow binding of biotinylated DNA to streptavidin beads.
- Supernatant was discarded by putting beads on a magnetic rack for 1 minute and beads were washed with 500 μ l 1xTWB at 55°C and then with 500 μ l of 10 mM Tris pH 8.
- On-bead PCR amplification was performed to amplify the DNA as mentioned in the library preparation protocol.

After library preparation, 4 nM equimolar pools of multiplexed Hi-C libraries were subjected to 150 bp paired-end sequencing on HiSeq4000 instrument.

7.6 Hi-C data analysis

- A table containing the position of all restriction sites used for Hi-C were pre-identified in the dm6 genome.
- The sequencing reads were analysed with a Perl script from Micmap package to identify the fusion sites. Micmap allows mapping of short reads in fastq format onto a reference genome.
- Reads were trimmed and mapped to dm6 genome using Micmap. Following criteria was considered to discard some reads that : mapped to a non-unique position in the reference; had indels and more than 2 mismatches; had fusion of oppositely oriented reads within 2 kb; were PCR duplicates.
- Only sequencing data from chromosome 2, 3, 4, and X were used and chromosome arms were considered as separate chromosome.
- Samples were down sampled to 13 million reads to correlate the biological replicates.
- Hi-C contact matrices were created at 10 kb resolution and these matrices were then normalized. Low coverage regions were removed before normalization.
- Pearson correlation coefficients were determined for normalized matrices and the calculated value was ≥ 0.936 for all replicates, meaning that Hi-C matrices for WT and mutants were globally similar.
- Quadruplicated replicates per genotype were pooled together and were down sampled to 79 million contacts per genotype and raw Hi-C matrices were

obtained by binning at 2 kb resolution and then normalized. Low coverage regions were filtered out before normalization and Hi-C heat maps were visualized in R.

- Contact domain boundaries in Hi-C matrix were called based on TopDom method that identifies contact boundaries along with a set of statistical methods for evaluating their boundaries based on bin signals (Shin et al. 2016).
- A and B compartment calling was performed using method proposed by Lieberman Aiden et al. (Lieberman-Aiden et al. 2009).

7.7 Insulator reporter assay

- We used the same insulator reporter plasmid as reported in our previous publication (Kaushal et al. 2021). This reporter plasmid comprised of an enhancer present in the middle of two reporter genes *EGFP* and *mCherry*.
- Selected genomic loci that showed Cp190 occupancy with various other IBPs were PCR-amplified from genomic DNA and cloned in the reporter system between *EGFP* and the enhancer.
- Control reporter plasmids had a fragment of the bacterial Kanamycin resistance gene (as a negative control) or the gypsy insulator (as a positive control) in between the enhancer and *EGFP*.
- Some of the genomic fragments with BEAF-32 motifs were mutagenized by PCR to mutate 1 bp in a BEAF-32 motif (TATCGATW to TAGCGATW).
- All plasmids with various genomic loci integrated in their reporter system were transfected in parallel with negative and positive control reporter plasmids into S2 cells in duplicates in a 96-well plate using 60 ng of reporter plasmid per replicate and Effectene (QIAGEN) following the manufacturer's instructions.
- After 48 hours, *EGFP* and *mCherry* fluorescence were measured for each samples on a NovoCyte Flow Cytometer. The distribution of *mCherry/EGFP* fluorescence was measured for every transfected cell for each experiment and their median *mCherry/EGFP* ration was calculated per experiment.
- The average *mCherry/EGFP* log₂ ratio was calculated for the negative control and substracted from all *mCherry/EGFP* log₂ rations obtained from each experiment.

- The average mCherry/EGFP log2 ratio was calculated for the the gypsy insulator reporter (positive control) and was set to 100% insulator strength. Relative insulator strengths of each tested fragments were measured with respect to gypsy insulator reporter as % of gypsy insulator strength.

7.8 RNA-FISH

Probe preparation

- To generate the RNA probes for *Scr*, *Dfd*, *Antp*, *Ubx*, *abd-A*, and *Abd-B* full length complementary DNA clones were used to PCR amplify the cDNA. For *ftz*, dm6 coordinates chr3R:6864324-6865765 (as published by Yokoshi et al, 2020 (Yokoshi et al. 2020) were cloned and then PCR amplified.
- After PCR amplification, in vitro transcription and RNA labelling was performed. The reaction mixture composition is described in the following table:

	[stock]	[final]	dilution	vol for 10 μ l
Transcription buffer (-Roche or NEB)	10 x	1 x	10 x	1
RNA labeling mix (Dig, Fluo, Bio) (-Roche)	10 x	1 x	10 x	1
SUPERase In RNase inhibitor	20 U / μ L	1 U / μ l	20 x	0.5
RNA polymerase (T7, T3, SP6) (-Roche or NEB)	20 U / μ L	20 U		0.5
DNA	\geq 70 ng / μ L	500 ng		\leq 7

nuclease-free ddH ₂ O				
----------------------------------	--	--	--	--

- The reaction mixture was assembled at room temperature to avoid precipitation of any components. These reactions were incubated for 3 hours at 37°C.
- After DNase I digestion for 20 minutes at 37°C, 15 µl of nuclease free water was added in each reaction mixture to increase the volume and probes were fragmented using 25 µl of carbonation buffer by incubating 20 minutes at 65°C.
- Fragmented probes were precipitated by adding following components

	[stock]	[final]	dilution	vol (in uL)
Fragmented RNA				50
NaOAc pH 5.2	3 M	300 mM	10 x	5
LiCl precipitation solution pH 8	7.5 M LiCl 50 mM EDTA	1.25 M LiCl 8 mM EDTA	6 x	8.5
tRNA	50 mg / mL			2
EtOH	100%			300

- This reaction mixture was incubated at 20°C atleast for 20 minutes and later centrifuged 30 minutes at maximum speed at 4°C to pallet the RNA.
- Supernatant was discarded and pallet was washed by adding 500 µl of 70% ethanol to remove any residual salt.
- Pallet in 70% ethanol were centrifuged maximum speed for 5 minutes at 4°C.
- Supernatant was carefully removed without touching the pallet.
- Pallet was airdried for 5 minutes.
- Once pallet was dried, it was dissolved in 50 µl of Hybridization buffer A by incubating at 37°C for 10 minutes.
- After this, probes were stored at -20°C for future use.

Fixing embryos

- Overnight and 2-6 hours collections of embryos from various genotypes were dechorionated by submerging them in 50% sodium hypochlorite solution for 2 minutes.
- Embryos were thoroughly washed under tap water until the smell of bleach was completely gone.
- They were transferred into a glass vial containing 2 ml of fixing solution and 4 ml of heptane. Embryos were fixed for 30 minutes on a shaking platform with maximum speed.
- After this, fixing solution was removed first and then heptane was removed carefully without disturbing the embryos.
- Embryos were rinsed with 2 ml heptane to remove the remnants of the fixing solution.
- Embryos were devitellinized by shaking vigorously 5 minutes in 1 ml heptane and 2 ml methanol.
- After this, embryos were washed in 1 ml methanol twice and stored in 1 ml methanol at -20°C.

In Situ hybridization

Rehydration and post-fixation steps

- To rehydrate the embryo, they were first incubated in 1 ml 50% methanol/50% PBS with 0.1% Tween-20.
- Embryos were washed 3 times 10 minutes in 1 ml of PBS with 0.1% Tween-20.
- They were post fixed in 4% paraformaldehyde solution for 20 minutes at room temperature.
- Post fixed embryos were washed 3 times quickly in 1 ml PBS with 0.1% Tween-20 to remove any traces of paraformaldehyde.
- After this, these embryos were washed 3 times 10 minutes in 1 ml of PBS with 0.1% Tween-20.
- After removing PBS with 0.1% Tween-20 of the last wash, embryos were first nutated in 1 ml of 50% PBTween/50% Hybridization buffer B for 5 minutes and then in 100% Hybridization buffer B for 5 minutes.

Blocking and probe addition

- Embryos were incubated at least for 3 hours (maximum overnight) in Hybridization buffer A at 65°C in a thermoblock at 800 rpm agitation.
- 5 µl of each probes were added into 250 µl of Hybridization buffer A solution and heated at 80°C for 10 minutes and then immediately incubated on ice for 5 minutes. Probes were stored on ice until used.
- Hybridization buffer A solution was removed from the embryos after the incubation and 250 µl of diluted probes were added to the samples overnight shaking at 65°C.

Antibody incubation

- Samples were incubated twice 30 minutes with 1 ml 10x diluted Roche Western Blocking Reagent in freshly prepared PBT with 0.1% TritonX-100 (PBTriton).
- Samples were incubated overnight at 4°C in anti-Dig peroxidase diluted 1:2000 in PBS, 0.1% TritonX-100 and 1x Western blocking reagent.
- Samples were washed 6 times 10 minutes in PBTween at room temperature.

TSA reaction

- Samples were labelled Cyanine 3 tyramide in TSA Plus kit (Perkin Elmer NEL753001KT) for 3 minutes at room temperature and immediately washed 4 times quickly in PBTween.
- After this step, samples were washed 6 times 10 minutes in 1 ml PBTween and 1 µl of 100xDAPI was added in the last step to stain DNA.
- These samples were finally mounted with VECTASHIELD mounting medium to capture the images of these embryos on a Zeiss LSM 880 microscope with a 20x objective.
- These images were visualized using Fiji software v2.1.0/1.53c.

7.9 Recombinant protein pull-downs

***Drosophila* nuclear extract preparation**

- 30 grams of 0-14 hrs old wildtype (OregonR) embryos were dechorionated by submerging them in 50% sodium hypochlorite solution for 2 minutes.

- Embryos were thoroughly washed under tap water until the smell of bleach was completely gone.
- Dechorionated embryos were collected in a dounce-homogenized in containing 30 ml of NU1 buffer and incubated for 10 minutes to allow cooling down and hypo-osmotic swelling.
- Then embryonic cells were disrupted by giving 30 strokes using a rotating device, at 4°C.
- Nuclei were filtered through a double layer miracloth and pelleted at 9000'rpm for 15 min at 4°C.
- Nuclei pellet was resuspended and lysed in 30 ml of high-salt buffer by rotating for 20 minutes at 4°C.
- Lysed nuclei were ultracentrifuged for 1 hour with SW40 rotor at 38000 rpm at 4°C.
- By removing the lipid layer, soluble nuclear extract was dialyzed using a dialysis buffer.
- Soluble nuclear extract was aliquoted into 1 ml microfuge tubes and snap-frozen in liquid nitrogen, and stored at -80°C.

Purification of interacting partners of Cp190 and negative control

- 250 µl of wildtype embryonic nuclear extract (0-14hrs) was incubated for 2 hours at 4°C with each 50 µl of GFP tagged recombinantly purified Cp190 or GFP tagged Nsel (negative control).
- 20 µl of GFP binder beads were added to this to pull down the baits and its interacting partner proteins by incubating for another 2 hours at 4°C.
- These GFP binder beads were washed twice with IP buffer and later with IP minimum buffer. Each wash was performed for 10 minutes at 4°C.
- In the last wash, supernatant was removed and beads were resuspended in 25 µl of minimal IP buffer.
- Since recombinant baits were tagged with GFP which may also interact with nuclear extract proteins, a 3C proteolytic site was introduced between GFP tag and the bait. 1ul of 3C protease was added to each samples and incubated at 4°C for 2 hours with constant flicking.

- Supernatant was transferred carefully (without any carryover beads) to a new tube and 8 µl of 4xSDS loading buffer was added to each tube.
- These beads were washed with minimal IP buffer once.
- To dissociate GFP and its partner proteins from the beads, beads were incubated with 25 µl of citrate buffer for 3 minutes on ice (constant flicking each minute).
- The elute from this step was transferred to a new tube and immediately neutralized by adding 2.5 µl (1/10th vol) of 1.5 M Tris pH 8.8.
- Cp190 and Nsel pull-downs were analysed using pre-casted SDS-PAGE gel (Tris-Glycine, 4-12%) and stained by Coomassie.
- After SDS PAGE analysis, samples were given for mass spectroscopy analysis.

7.10 Western Blot analysis

Preparation of embryo extracts:

- 6-10 hour old embryos were dechorionated and homogenized in 1 ml of 1x SDS loading buffer and heated at 80° C for 5 minute.
- These homogenized samples were sonicated in a Bioruptor on high intensity setting for 10 minutes (30 seconds on, 30 seconds off).
- Cell debris were removed by centrifuging for 3 minute.
- Supernatant from each sample was collected and aliquoted into two fresh microfuge tubes and snap frozen in liquid nitrogen before storing it in -80 .

Steps followed in western blot analysis:

- The samples and pre-stained ladder were loaded in a pre-casted acrylamide gel and ran between 150-200 volt for 1 hour.
- Pre-cool transfer buffer was prepared in advance while samples were running.
- After running the samples for appropriate time, acrylamide gel was blotted on a 6cm x 8 cm nitrocellulose membrane (pre-equilibrated in dH₂O) immersed in a 1x western transfer buffer.
- The membrane was washed 3 times with PBTween and then blocked with 5% skimmed milk for 1 hour at room temperature.
- It was incubated overnight with following primary antibodies in blocking reagent:

Primary antibody	concentration
------------------	---------------

Rabbit anti CP190-FL	1:2000
Affinity purified rabbit anti CTCF-C	1:2000
Mouse anti alpha-tubulin	1:10000

- After primary antibody incubation, membrane was washed three times with PBSTween for 5 minute each.
- It was incubated with the following concentration of secondary antibody in blocking reagent for at least 1-2 hour at 4°C.

Secondary antibody	concentration
Anti-rabbit	1:5000
Anti-mouse	1:5000

- Membrane was developed using ECL reagents and imaged using Fiji software.

Drosophila viability test

- 3 sets of 60-100 embryos of desired collected in a fly culture vial.
- Vials were incubated at 25° and the number of unfertilized eggs and hatched larvae were counted 2 days later.
- Once pupae started to hatch, number of adult flies that completely emerged from pupal case were also counted.
- These tests were performed in triplicate experiments for each genotype.

9 Materials

1x western transfer buffer

	Final	Volume for 1.4 lt
10X SDS buffer	1X	140 ml
100% EtOH	20%	280 ml
ddH ₂ O		Complete up to 1.4 lt

2x Laemmli buffer

Reagent	Molecular weight	Concentration (M or %)	2x	Add for 5 ml of 2X
Tris base	121.14	0.0625 M	0.125 M	0.0747 g
SDS	288.37	0.07 M (2%)	0.14 M (4%)	0.2 g
glycerol	92.09	10%	20%	1 ml
Bromphenol blue	691.94	-	-	10 mg
2-mercapto-ethanol	78.13	5%	10%	0.5 ml

PBTween

	[stock]	[final]	dilution	vol for 50 ml
nuclease-free H ₂ O				45 ml
PBS	10 x	1 x	10 x	5 mL
Tween20	20%	0.1%	200 x	250 µL

Prepare freshly before use.

NU1 buffer :

	[stock]	[final]	dilution	vol for 4 mL
HEPES pH 7.6 (8)	1 M	15 mM	66.6 x	3.89 mL
KCl	3 M	10 mM	300 x	
MgCl ₂	1 M	5 mM	200 x	
EDTA pH 8	0.5 M	0.1 mM	5'000 x	

EGTA pH 8	100 mM	0.5 mM	200 x	
Sucrose	342.3 g/mol	350 mM		
DTT	1 M	1 mM	1'000 x	4 µL
Complete	50 x	1 x	50 x	80 µL
PMSF	200 mM	1 mM	200 x	20 µL

Add DTT, Complete and PMSF just before use.

Low-salt buffer :

	[stock]	[final]	dilution	vol for 6 mL
HEPES pH 7.6 (8)	1 M	15 mM	66.6 x	5.84 mL
glycerol	100%	20%	5 x	
MgCl ₂	1 M	1.5 mM	666.6 x	
KCl	3 M	20 mM	150 x	
EDTA pH 8	0.5 M	0.2 mM	2'500 x	
DTT	1 M	1 mM	1'000 x	6 µL
Complete	50 x	1 x	50 x	120 µL
PMSF	200 mM	1 mM	50 x	30 µL

Add DTT, Complete and PMSF just before use.

High-salt buffer :

	[stock]	[final]	dilution	vol for 4 mL
HEPES pH 7.6 (8)	1 M	15 mM	66.6 x	3.87 mL
glycerol	100%	20%	5 x	
MgCl ₂	1 M	1.5 mM	666.6 x	
KCl	3 M	800 mM	150 x	
EDTA pH 8	0.5 M	0.2 mM	2'500 x	
DTT	1 M	1 mM	1'000 x	4 µL
Complete	50 x	1 x	50 x	80 µL
PMSF	100 mM	1 mM	100 x	40 µL

Add DTT, Complete and PMSF just before use.

IP buffer

	[stock]	[final]	dilution	vol for 500 ml
Tris pH 7.5	1 M	50 mM	20 x	25 ml
potassium acetate	8 M	150 mM	50 x	10 ml
MgCl ₂	1 M	2 mM	500 x	1000 µL
glycerol	100%	10%	10 x	50 ml
DTT	1 M	1 mM	1'000 x	500 µL
MilliQ H ₂ O				up to 500 ml

Add DTT just before use.

Dialysis buffer

	[stock]	[final]	dil.	vol for 1 L
Tris pH 7.5	1 M	50 mM	20 x	50 ml
potassium acetate	98.15 g / mol	150 mM		14.7 g
MgCl ₂	1 M	2 mM	500 x	2 ml
EDTA pH 8	500 mM	0.2 mM	2'500 x	400 µL
glycerol	100%	20%	5 x	200 ml
MilliQ				
DTT	1 M	0.5 mM	2'000 x	500 ul

Hybridization buffer A

	[stock]	[final]	dilution	vol for 50 ml
Formamide	100%	50%	2 x	25 ml
SSC pH 5	20 x	5 x (75 mM sodium citrate, 750 mM NaCl)	4 x	12.5 ml
salmon sperm ssDNA	10 mg / mL	100 ug / ml	100 x	500 µL
Heparin	50 mg / ml	50 ug / ml	1'000 x	50 µL
Tween20	20%	0.1%	200 x	250 µL

nuclease-free ddH ₂ O				11.7 mL
-------------------------------------	--	--	--	---------

Store in a Falcon @ -20C.

Hybridization buffer B

	[stock]	[final]	dilution	vol for 50 ml
Formamide	100%	50%	2 x	25 ml
SSC pH 5	20 x	5 x (75 mM sodium citrate, 750 mM NaCl)	4 x	12.5 ml
nuclease-free ddH ₂ O				12.5 ml

Fixing solution:

	[stock]	[final]	dilution	vol for 2 mL
PBS	1 x	1 x		1.5 mL
paraformaldehyde	16%	4%	4 x	0.5 mL

Prepare fresh before use.

Cross linking solution

	[stock]	[final]	dilution	vol for 500 ml
Hepes pH 7.9	1 M	50 mM	20x	25 ml
EDTA pH 8	500 mM	1 mM	500x	1 ml
EGTA pH 8	250 mM	0.5 mM	500x	1 ml
NaCl	5 M	100 mM	50x	10 ml
MilliQ ddH ₂ O				463 ml

Tris contains reactive amine which causes cross-linking of formaldehyde to Tris so hampering fixation.

Filter 0.22 µm to sterilize.

Store at 4°C.

STOP solution

	[stock]	[final]	dilution	vol for 500 ml
glycine	75.07 g/mol	125 mM		4.7 g
Triton X-100	20%	0.01%	2'000x	250 µl
PBS	1x	1x		up to 500 ml

Filter 0.22 µm to sterilize.

Store at 4°C.

Solution A

	[stock]	[final]	dilution	vol for 500 ml
Hepes pH 7.9	1 M	10 mM	100x	5 ml
EDTA pH 8	500 mM	10 mM	50x	10 ml
EGTA pH 8	250 mM	0.5 mM	500x	1 ml
Triton X-100	20%	0.25%	80x	6.25 ml
MilliQ ddH ₂ O				477.75 ml

Filter 0.22 µm to sterilize. First, you can filter solution B and then solution A with the same filter.

Store at 4°C.

Solution B

	[stock]	[final]	dilution	vol for 500 ml
Hepes pH 7.9	1 M	10 mM	100x	10 ml
EDTA pH 8	500 mM	1 mM	500x	2 ml
EGTA pH 8	250 mM	0.5 mM	500x	2 ml
Triton X-100	20%	0.01%	2000x	500 µl
NaCl	5 M	200 mM	25x	40 ml
MilliQ ddH ₂ O				954.5 ml

Filter 0.22 µm to sterilize. First, you can filter solution B and then solution A with the same filter.

Store at 4°C.

RIPA 140 mM NaCl

	[stock]	[final]	dilution	vol for 1 L total
Tris-HCl pH 8	1 M	10 mM	100 x	10 mL
NaCl	1 M	140 mM	~7.14 x	140 mL
EDTA pH 8	0.5 M	1 mM	500 x	2 mL
Triton X-100	100%	1%	100 x	10 mL
SDS	10%	0.1%	100 x	10 mL
sodium deoxycholate	10%	0.1%	100 x	10 mL
complete	50 x	1 x	50 x	
MilliQ ddH ₂ O				818 mL

Filter 0.22 um to sterilize.

Store at 4C.

Add complete to an aliquot of RIPA 140 mM that will be used to adjust the chromatin to 1 mL for o/n incubation with ChIP antibody.

Don't add protease inhibitors to the wash buffers (not necessary and may interfere with proteinase K digestion).

RIPA 140 mM NaCl

	stoc k	final	dilutio n	vol for 1 L total
Tris-HCl pH 8	1 M	10 mM	100 x	10 mL
NaCl	1 M	140 mM	~7.14 x	140 mL
EDTA pH 8	0.5 M	1 mM	500 x	2 mL
Triton X-100	100 %	1%	100 x	10 mL
SDS	10%	0.1%	100 x	10 mL
sodium deoxycholate	10%	0.1%	100 x	10 mL
complete	50 x	1 x	50 x	
ddH ₂ O				818 mL

Filter 0.22 um to sterilize.

Store at 4C.

Add complete to an aliquot of RIPA 140 mM that will be used to adjust the chromatin to 1 mL for o/n incubation with CHIP antibody.

Don't add protease inhibitors to the wash buffers (not necessary and may interfere with proteinase K digestion).

RIPA 500 mM NaCl

	stock	final	dilution	vol for 1 L
Tris-HCl pH 8	1 M	10 mM	100 x	10 mL
NaCl	58.44 g / mol	500 mM		29.22 g
EDTA pH 8	0.5 M	1 mM	500 x	2 mL
Triton X-100	100%	1%	100 x	10 mL
SDS	10%	0.1%	100 x	10 mL
sodium deoxycholate	10%	0.1%	100 x	10 mL
ddH ₂ O				up to 1 L

Filter 0.22 um to sterilize.

Store at 4C.

LiCl buffer

	stock	final	dilution	vol for 500 mL
Tris-HCl pH 8	1 M	10 mM	100x	5 mL
LiCl	42.39 g/mol	250 mM		5.3 g
EDTA pH 8	0.5 M	1 mM	500x	1 mL
Igepal CA-630	100%	0.5%	200x	2.5 mL
sodium deoxycholate	10%	0.5%	20x	25 mL
ddH ₂ O				up to 500 mL

Filter 0.22 um to sterilize.

Store at 4C.

TE buffer

	stock	final	dilution	vol for 1 L
Tris-HCl pH 8	1 M	10 mM	100x	10 mL
EDTA pH 8	0.5 M	1 mM	500x	2 mL
ddH ₂ O				988 mL

Filter 0.22 um to sterilize.

Store at 4C.

20% Triton X-100

10 ml Triton X-100 100% in 40 ml milliQ H₂O in a Falcon.

Rotate as long as necessary until thoroughly mixed.

Filter to sterilize.

Store at 4°C.

10% (w/v) sodium deoxycholate

5 g in 50 ml milliQ H₂O in a Falcon.

Rotate as long as necessary until thoroughly mixed.

Filter to sterilize.

Store at 4°C protected from light.

1% (w/v) sodium deoxycholate

Make 50 ml from 10% stock.

Store at 4°C protected from light.

1% SDS

Make 50 ml from 20% stock.

Store at RT (crystallizes at 4°C).

1.4 M NaCl

4.1 g NaCl and milliQ up to 50 ml final volume in a Falcon.

Filter to sterilize.

Store at 4°C.

3M NaOAc

61.5 g sodium acetate anhydrous in 200 ml milliQ H₂O.

Adjust pH to 5.2 with glacial acetic acid under the hood.

Add milliQ H₂O to 250 ml final.

Filter to sterilize.

Store at RT

10 References

- Akbari OS, Bousum A, Bae E, Drewell RA (2006) Unraveling cis-regulatory mechanisms at the abdominal-A and Abdominal-B genes in the *Drosophila bithorax* complex. *Dev Biol* 293:294–304. <https://doi.org/10.1016/j.ydbio.2006.02.015>
- Ali T, Krüger M, Bhujji S, et al (2017) Chromatin binding of Gcn5 in *Drosophila* is largely mediated by CP190. *Nucleic Acids Res* 45:2384–2395. <https://doi.org/10.1093/nar/gkw1178>
- Amemiya HM, Kundaje A, Boyle AP (2019) The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep-uk* 9:9354. <https://doi.org/10.1038/s41598-019-45839-z>
- Bag I, Chen S, Rosin LF, et al (2021) M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity. *Nat Commun* 12:4170. <https://doi.org/10.1038/s41467-021-24407-y>
- Bartkuhn M, Straub T, Herold M, et al (2009) Active promoters and insulators are marked by the centrosomal protein 190. *Embo J* 28:877–888. <https://doi.org/10.1038/emboj.2009.34>
- Batut PJ, Bing XY, Sisco Z, et al (2022) Genome organization controls transcriptional dynamics during development. *Sci New York N Y* 375:566–570. <https://doi.org/10.1126/science.abi7178>
- Bell AC, West AG, Felsenfeld G (1999) The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. *Cell* 98:387–396. [https://doi.org/10.1016/s0092-8674\(00\)81967-4](https://doi.org/10.1016/s0092-8674(00)81967-4)
- Bonev B, Cohen NM, Szabo Q, et al (2017) Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171:557-572.e24. <https://doi.org/10.1016/j.cell.2017.09.043>
- Bortle KV, Nichols MH, Li L, et al (2014) Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol* 15:R82. <https://doi.org/10.1186/gb-2014-15-5-r82>
- Bortle KV, Ramos E, Takenaka N, et al (2012) *Drosophila* CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res* 22:2176–2187. <https://doi.org/10.1101/gr.136788.111>
- Boveri T (1887) *Zellen-Studien*. <https://doi.org/10.5962/bhl.title.6755>
- Bowman SK, Deaton AM, Domingues H, et al (2014) H3K27 modifications define segmental regulatory domains in the *Drosophila bithorax* complex. *Elife* 3:e02833. <https://doi.org/10.7554/elife.02833>

- Buccitelli C, Selbach M (2020) mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* 21:630–644. <https://doi.org/10.1038/s41576-020-0258-4>
- Bulger M, Groudine M (2011) Functional and Mechanistic Diversity of Distal Transcription Enhancers. *Cell* 144:825. <https://doi.org/10.1016/j.cell.2011.02.026>
- Bushey AM, Ramos E, Corces VG (2009) Three subclasses of a *Drosophila* insulator show distinct and cell type-specific genomic distributions. *Gene Dev* 23:1338–1350. <https://doi.org/10.1101/gad.1798209>
- Cai H, Levine M (1995) Modulation of enhancer–promoter interactions by insulators in the *Drosophila* embryo. *Nature* 376:533–536. <https://doi.org/10.1038/376533a0>
- Cai HN, Shen P (2001) Effects of cis Arrangement of Chromatin Insulators on Enhancer-Blocking Activity. *Science* 291:493–495. <https://doi.org/10.1126/science.291.5503.493>
- Carroll TS, Liang Z, Salama R, et al (2014) Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Frontiers Genetics* 5:75. <https://doi.org/10.3389/fgene.2014.00075>
- Chathoth KT, Zabet NR (2019) Chromatin architecture reorganization during neuronal cell differentiation in *Drosophila* genome. *Genome Res* 29:613–625. <https://doi.org/10.1101/gr.246710.118>
- Chen H, Levo M, Barinov L, et al (2018) Dynamic interplay between enhancer-promoter topology and gene activity. *Nat Genet* 50:1296–1303. <https://doi.org/10.1038/s41588-018-0175-z>
- Cremer M, Grasser F, Lanctôt C, et al (2008) Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol Biology Clifton N J* 463:205–39. https://doi.org/10.1007/978-1-59745-406-3_15
- Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2:292–301. <https://doi.org/10.1038/35066075>
- Cremer T, Cremer C, Baumann H, et al (1982) Rabl’s model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Hum Genet* 60:46–56
- Cremer T, Cremer M (2010) Chromosome Territories. *Csh Perspect Biol* 2:a003889. <https://doi.org/10.1101/cshperspect.a003889>
- Cuartero S, Fresán U, Reina O, et al (2014) Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. *Embo J* 33:637–647. <https://doi.org/10.1002/embj.201386001>
- Cubeñas-Potts C, Rowley MJ, Lyu X, et al (2017) Different enhancer classes in *Drosophila* bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res* 45:1714–1730. <https://doi.org/10.1093/nar/gkw1114>

- Davidson IF, Bauer B, Goetz D, et al (2019) DNA loop extrusion by human cohesin. *Science* 366:1338–1345. <https://doi.org/10.1126/science.aaz3418>
- Dekker J, Mirny L (2016) The 3D Genome as Moderator of Chromosomal Communication. *Cell* 164:1110–1121. <https://doi.org/10.1016/j.cell.2016.02.007>
- Dixon JR, Selvaraj S, Yue F, et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–380. <https://doi.org/10.1038/nature11082>
- El-Sharnouby S, Fischer B, Magbanua JP, et al (2016) Regions of very low H3K27me3 partition the Drosophila genome into topological domains. *Biorxiv* 072900. <https://doi.org/10.1101/072900>
- Espinola SM, Götz M, Bellec M, et al (2021) 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. <https://doi.org/10.1038/s41588-021-00816-z>
- Franke M, Ibrahim DM, Andrey G, et al (2016) Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 538:265–269. <https://doi.org/10.1038/nature19800>
- Fudenberg G, Imakaev M, Lu C, et al (2016) Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* 15:2038–2049. <https://doi.org/10.1016/j.celrep.2016.04.085>
- Fujioka M, Mistry H, Schedl P, Jaynes JB (2016) Determinants of Chromosome Architecture: Insulator Pairing in cis and in trans. *Plos Genet* 12:e1005889. <https://doi.org/10.1371/journal.pgen.1005889>
- Fujioka M, Sun G, Jaynes JB (2013) The Drosophila eve Insulator Homie Promotes eve Expression and Protects the Adjacent Gene from Repression by Polycomb Spreading. *Plos Genet* 9:e1003883. <https://doi.org/10.1371/journal.pgen.1003883>
- Fujioka M, Wu X, Jaynes JB (2009) A chromatin insulator mediates transgene homing and very long-range enhancer-promoter communication. *Development* 136:3077–3087. <https://doi.org/10.1242/dev.036467>
- Furlong EEM, Levine M (2018) Developmental enhancers and chromosome topology. *Science* 361:1341–1345. <https://doi.org/10.1126/science.aau0320>
- Gambetta MC, Furlong EEM (2018) The Insulator Protein CTCF Is Required for Correct Hox Gene Expression, but Not for Embryonic Development in Drosophila. *Genetics* 210:129–136. <https://doi.org/10.1534/genetics.118.301350>
- Gerasimova TI, Gdula DA, Gerasimov DV, et al (1995) A drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell* 82:587–597. [https://doi.org/10.1016/0092-8674\(95\)90031-4](https://doi.org/10.1016/0092-8674(95)90031-4)

- Gerasimova TI, Lei EP, Bushey AM, Corces VG (2007) Coordinated Control of dCTCF and gypsy Chromatin Insulators in *Drosophila*. *Mol Cell* 28:761–772. <https://doi.org/10.1016/j.molcel.2007.09.024>
- Geyer PK, Corces VG (1992) DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. *Gene Dev* 6:1865–1873. <https://doi.org/10.1101/gad.6.10.1865>
- Ghavi-Helm Y, Jankowski A, Meiers S, et al (2019) Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet* 51:1272–1282. <https://doi.org/10.1038/s41588-019-0462-3>
- Ghosh RP, Meyer BJ (2021) Spatial Organization of Chromatin: Emergence of Chromatin Structure During Development. *Annu Rev Cell Dev Bi* 37:1–34. <https://doi.org/10.1146/annurev-cellbio-032321-035734>
- Gindhart JG, King AN, Kaufman TC (1995) Characterization of the cis-regulatory region of the *Drosophila* homeotic gene *Sex combs reduced*. *Genetics* 139:781–95. <https://doi.org/10.1093/genetics/139.2.781>
- Giorgio E, Robyr D, Spielmann M, et al (2015) A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Hum Mol Genet* 24:3143–3154. <https://doi.org/10.1093/hmg/ddv065>
- Hou C, Li L, Qin ZS, Corces VG (2012) Gene Density, Transcription, and Insulators Contribute to the Partition of the *Drosophila* Genome into Physical Domains. *Mol Cell* 48:471–484. <https://doi.org/10.1016/j.molcel.2012.08.031>
- Hsieh T-HS, Cattoglio C, Slobodyanyuk E, et al (2020) Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell* 78:539–553.e8. <https://doi.org/10.1016/j.molcel.2020.03.002>
- Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM (2017) Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* 169:216–228.e19. <https://doi.org/10.1016/j.cell.2017.03.024>
- Ing-Simmons E, Vaid R, Bing XY, et al (2021) Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning. *Nat Genet* 53:487–499. <https://doi.org/10.1038/s41588-021-00799-x>
- Iseli C, Ambrosini G, Bucher P, Jongeneel CV (2007) Indexing Strategies for Rapid Searches of Short Words in Genome Sequences. *Plos One* 2:e579. <https://doi.org/10.1371/journal.pone.0000579>
- Ji X, Dadon DB, Powell BE, et al (2016) 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* 18:262–275. <https://doi.org/10.1016/j.stem.2015.11.007>

- Kahn TG, Schwartz YB, Dellino GI, Pirrotta V (2006) Polycomb Complexes and the Propagation of the Methylation Mark at the *Drosophila* Ubx Gene*. *J Biol Chem* 281:29064–29075. <https://doi.org/10.1074/jbc.m605430200>
- Kaushal A, Mohana G, Dorier J, et al (2021) CTCF loss has limited effects on global genome architecture in *Drosophila* despite critical regulatory functions. *Nat Commun* 12:1011. <https://doi.org/10.1038/s41467-021-21366-2>
- Kellum R, Schedl P (1992) A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol* 12:2424–2431. <https://doi.org/10.1128/mcb.12.5.2424-2431.1992>
- Kellum R, Schedl P (1991) A position-effect assay for boundaries of higher order chromosomal domains. *Cell* 64:941–950. [https://doi.org/10.1016/0092-8674\(91\)90318-s](https://doi.org/10.1016/0092-8674(91)90318-s)
- Kim Y, Shi Z, Zhang H, et al (2019) Human cohesin compacts DNA by loop extrusion. *Science* 366:1345–1349. <https://doi.org/10.1126/science.aaz4475>
- Kornberg RD (1974) Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* 184:868–871. <https://doi.org/10.1126/science.184.4139.868>
- Kravchenko E, Savitskaya E, Kravchuk O, et al (2005) Pairing between gypsy Insulators Facilitates the Enhancer Action in trans throughout the *Drosophila* Genome. *Mol Cell Biol* 25:9283–9291. <https://doi.org/10.1128/mcb.25.21.9283-9291.2005>
- Kurth I, Klopocki E, Stricker S, et al (2009) Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia. *Nat Genet* 41:862–863. <https://doi.org/10.1038/ng0809-862>
- Kvon EZ, Kazmar T, Stampfel G, et al (2014) Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* 512:91–95. <https://doi.org/10.1038/nature13395>
- Kyrchanova O, Sabirov M, Mogila V, et al (2019) Complete reconstitution of bypass and blocking functions in a minimal artificial Fab-7 insulator from *Drosophila* bithorax complex. *Proc National Acad Sci* 116:13462–13467. <https://doi.org/10.1073/pnas.1907190116>
- Kyrchanova O, Toshchakov S, Podstreshnaya Y, et al (2008) Functional Interaction between the Fab-7 and Fab-8 Boundaries and the Upstream Promoter Region in the *Drosophila* Abd-B Gene. *Mol Cell Biol* 28:4188–4195. <https://doi.org/10.1128/mcb.00229-08>
- Lewis EB (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* 276:565–570. <https://doi.org/10.1038/276565a0>
- Lewis EB (1954) The Theory and Application of a New Method of Detecting Chromosomal Rearrangements in *Drosophila melanogaster*. *Am Nat* 88:225–239. <https://doi.org/10.1086/281833>

- Li L, Lyu X, Hou C, et al (2015) Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing. *Mol Cell* 58:216–231. <https://doi.org/10.1016/j.molcel.2015.02.023>
- Li Y, Haarhuis JHI, Cacciatore ÁS, et al (2020) The structural basis for cohesin–CTCF-anchored loops. *Nature* 578:472–476. <https://doi.org/10.1038/s41586-019-1910-z>
- Liang J, Lacroix L, Gamot A, et al (2014) Chromatin Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator Proteins and Pol II Pausing. *Mol Cell* 53:672–681. <https://doi.org/10.1016/j.molcel.2013.12.029>
- Lieberman-Aiden E, Berkum NL van, Williams L, et al (2009) Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326:289–293. <https://doi.org/10.1126/science.1181369>
- Lim B, Heist T, Levine M, Fukaya T (2018) Visualization of Transvection in Living *Drosophila* Embryos. *Mol Cell* 70:287–296.e6. <https://doi.org/10.1016/j.molcel.2018.02.029>
- Lun ATL, Smyth GK (2016) csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res* 44:e45–e45. <https://doi.org/10.1093/nar/gkv1191>
- Lupiáñez DG, Kraft K, Heinrich V, et al (2015) Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 161:1012–1025. <https://doi.org/10.1016/j.cell.2015.04.004>
- Luzhin AV, Flyamer IM, Khrameeva EE, et al (2019) Quantitative differences in TAD border strength underly the TAD hierarchy in *Drosophila* chromosomes. *J Cell Biochem* 120:4494–4503. <https://doi.org/10.1002/jcb.27737>
- Maksimenko O, Bartkuhn M, Stakhov V, et al (2015) Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome Res* 25:89–99. <https://doi.org/10.1101/gr.174169.114>
- Mallin DR, Myung JS, Patton JS, Geyer PK (1998) Polycomb group repression is blocked by the *Drosophila* suppressor of Hairy-wing [su(Hw)] insulator. *Genetics* 148:331–9. <https://doi.org/10.1093/genetics/148.1.331>
- Mateo LJ, Murphy SE, Hafner A, et al (2019) Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 568:49–54. <https://doi.org/10.1038/s41586-019-1035-4>
- Mohan M, Bartkuhn M, Herold M, et al (2007) The *Drosophila* insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *Embo J* 26:4203–4214. <https://doi.org/10.1038/sj.emboj.7601851>
- Moore JM, Rabaia NA, Smith LE, et al (2012) Loss of Maternal CTCF Is Associated with Peri-Implantation Lethality of Ctfc Null Embryos. *Plos One* 7:e34915. <https://doi.org/10.1371/journal.pone.0034915>

- Moshkovich N, Nisha P, Boyle PJ, et al (2011) RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. *Gene Dev* 25:1686–1701.
<https://doi.org/10.1101/gad.16651211>
- Mourad R, Cuvier O (2016) Computational Identification of Genomic Features That Influence 3D Chromatin Domain Formation. *Plos Comput Biol* 12:e1004908.
<https://doi.org/10.1371/journal.pcbi.1004908>
- Muravyova E, Golovnin A, Gracheva E, et al (2001) Loss of Insulator Activity by Paired Su(Hw) Chromatin Insulators. *Science* 291:495–498.
<https://doi.org/10.1126/science.291.5503.495>
- Narendra V, Bulajić M, Dekker J, et al (2016) CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Gene Dev* 30:2657–2662.
<https://doi.org/10.1101/gad.288324.116>
- Narendra V, Rocha PP, An D, et al (2015) CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347:1017–1021.
<https://doi.org/10.1126/science.1262088>
- Nègre N, Brown CD, Shah PK, et al (2010) A Comprehensive Map of Insulator Elements for the Drosophila Genome. *Plos Genet* 6:e1000814.
<https://doi.org/10.1371/journal.pgen.1000814>
- Nora EP, Goloborodko A, Valton A-L, et al (2017) Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169:930-944.e22. <https://doi.org/10.1016/j.cell.2017.05.004>
- Nora EP, Lajoie BR, Schulz EG, et al (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485:381–385. <https://doi.org/10.1038/nature11049>
- Ou HD, Phan S, Deerinck TJ, et al (2017) ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Sci New York N Y* 357:eaag0025.
<https://doi.org/10.1126/science.aag0025>
- Özdemir I, Gambetta MC (2019) The Role of Insulation in Patterning Gene Expression. *Genes-basel* 10:767. <https://doi.org/10.3390/genes10100767>
- Pai C-Y, Lei EP, Ghosh D, Corces VG (2004) The Centrosomal Protein CP190 Is a Component of the gypsy Chromatin Insulator. *Mol Cell* 16:737–748.
<https://doi.org/10.1016/j.molcel.2004.11.004>
- Parelho V, Hadjur S, Spivakov M, et al (2008) Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell* 132:422–433.
<https://doi.org/10.1016/j.cell.2008.01.011>
- Postika N, Metzler M, Affolter M, et al (2018) Boundaries mediate long-distance interactions between enhancers and promoters in the Drosophila Bithorax complex. *Plos Genet* 14:e1007702. <https://doi.org/10.1371/journal.pgen.1007702>

- Pugacheva EM, Kubo N, Loukinov D, et al (2020) CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *P Natl Acad Sci Usa* 117:2020–2031. <https://doi.org/10.1073/pnas.1911708117>
- Ramírez F, Bhardwaj V, Arrigoni L, et al (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* 9:189. <https://doi.org/10.1038/s41467-017-02525-w>
- Rao SSP, Huang S-C, Hilaire BGS, et al (2017) Cohesin Loss Eliminates All Loop Domains. *Cell* 171:305–320.e24. <https://doi.org/10.1016/j.cell.2017.09.026>
- Rao SSP, Huntley MH, Durand NC, et al (2014) A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159:1665–1680. <https://doi.org/10.1016/j.cell.2014.11.021>
- Rowley MJ, Nichols MH, Lyu X, et al (2017) Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell* 67:837–852.e7. <https://doi.org/10.1016/j.molcel.2017.07.022>
- Sanborn AL, Rao SSP, Huang S-C, et al (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc National Acad Sci* 112:E6456–E6465. <https://doi.org/10.1073/pnas.1518552112>
- Schwartz YB, Linder-Basso D, Kharchenko PV, et al (2012) Nature and function of insulator protein binding sites in the *Drosophila* genome. *Genome Res* 22:2188–2198. <https://doi.org/10.1101/gr.138156.112>
- Schwarzer W, Abdennur N, Goloborodko A, et al (2017) Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551:51–56. <https://doi.org/10.1038/nature24281>
- Segert JA, Gisselbrecht SS, Bulyk ML (2021) Transcriptional Silencers: Driving Gene Expression with the Brakes On. *Trends Genet* 37:514–527. <https://doi.org/10.1016/j.tig.2021.02.002>
- Sexton T, Yaffe E, Kenigsberg E, et al (2012) Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome. *Cell* 148:458–472. <https://doi.org/10.1016/j.cell.2012.01.010>
- Shin H, Shi Y, Dai C, et al (2016) TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res* 44:e70–e70. <https://doi.org/10.1093/nar/gkv1505>
- Sigrist CJA, Pirrotta V (1997) Chromatin Insulator Elements Block the Silencing of a Target Gene by the *Drosophila* Polycomb Response Element (PRE) but Allow trans Interactions Between PREs on Different Chromosomes. *Genetics* 147:209–221. <https://doi.org/10.1093/genetics/147.1.209>











- Simonis M, Klous P, Splinter E, et al (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat Genet* 38:1348–1354. <https://doi.org/10.1038/ng1896>
- Soshnev AA, He B, Baxley RM, et al (2012) Genome-wide studies of the multi-zinc finger *Drosophila* Suppressor of Hairy-wing protein in the ovary. *Nucleic Acids Res* 40:5415–5431. <https://doi.org/10.1093/nar/gks225>
- Soshnikova N, Montavon T, Leleu M, et al (2010) Functional Analysis of CTCF During Mammalian Limb Development. *Dev Cell* 19:819–830. <https://doi.org/10.1016/j.devcel.2010.11.009>
- Spielmann M, Lupiáñez DG, Mundlos S (2018) Structural variation in the 3D genome. *Nat Rev Genet* 19:453–467. <https://doi.org/10.1038/s41576-018-0007-0>
- Stevens TJ, Lando D, Basu S, et al (2017) 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544:59–64. <https://doi.org/10.1038/nature21429>
- Sultana H, Verma S, Mishra RK (2011) A BEAF dependent chromatin domain boundary separates myoglianin and eyeless genes of *Drosophila melanogaster*. *Nucleic Acids Res* 39:3543–3557. <https://doi.org/10.1093/nar/gkq1297>
- Tang Z, Luo OJ, Li X, et al (2015) CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* 163:1611–1627. <https://doi.org/10.1016/j.cell.2015.11.024>
- Ulianov SV, Khrameeva EE, Gavrilov AA, et al (2016) Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res* 26:70–84. <https://doi.org/10.1101/gr.196006.115>
- Ulianov SV, Zakharova VV, Galitsyna AA, et al (2021) Order and stochasticity in the folding of individual *Drosophila* genomes. *Nat Commun* 12:41. <https://doi.org/10.1038/s41467-020-20292-z>
- Vogelmann J, Gall AL, Dejardin S, et al (2014) Chromatin Insulator Factors Involved in Long-Range DNA Interactions and Their Role in the Folding of the *Drosophila* Genome. *Plos Genet* 10:e1004544. <https://doi.org/10.1371/journal.pgen.1004544>
- Wang Q, Sun Q, Czajkowsky DM, Shao Z (2018a) Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nat Commun* 9:188. <https://doi.org/10.1038/s41467-017-02526-9>
- Wang Q, Sun Q, Czajkowsky DM, Shao Z (2018b) Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nat Commun* 9:188. <https://doi.org/10.1038/s41467-017-02526-9>
- Watson LA, Wang X, Elbert A, et al (2014) Dual Effect of CTCF Loss on Neuroprogenitor Differentiation and Survival. *J Neurosci* 34:2860–2870. <https://doi.org/10.1523/jneurosci.3769-13.2014>

- Wendt KS, Yoshida K, Itoh T, et al (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451:796–801. <https://doi.org/10.1038/nature06634>
- Whitfield WG, Chaplin MA, Oegema K, et al (1995) The 190 kDa centrosome-associated protein of *Drosophila melanogaster* contains four zinc finger motifs and binds to specific sites on polytene chromosomes. *J Cell Sci* 108 (Pt 11):3377–87. <https://doi.org/10.1242/jcs.108.11.3377>
- Whitfield WG, Millar SE, Saumweber H, et al (1988) Cloning of a gene encoding an antigen associated with the centrosome in *Drosophila*. *J Cell Sci* 89 (Pt 4):467–80. <https://doi.org/10.1242/jcs.89.4.467>
- Williamson I, Kane L, Devenney PS, et al (2019) Developmentally regulated Shh expression is robust to TAD perturbations. *Development* 146:dev179523. <https://doi.org/10.1242/dev.179523>
- Yokoshi M, Segawa K, Fukaya T (2020) Visualizing the Role of Boundary Elements in Enhancer-Promoter Communication. *Mol Cell* 78:224-235.e5. <https://doi.org/10.1016/j.molcel.2020.02.007>
- Zhang Y, McCord RP, Ho Y-J, et al (2012) Spatial Organization of the Mouse Genome and Its Role in Recurrent Chromosomal Translocations. *Cell* 148:908–921. <https://doi.org/10.1016/j.cell.2012.02.002>
- Zhao K, Hart CM, Laemmli UK (1995) Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell* 81:879–889. [https://doi.org/10.1016/0092-8674\(95\)90008-x](https://doi.org/10.1016/0092-8674(95)90008-x)
- Zolotarev N, Maksimenko O, Kyrchanova O, et al (2017) Opbp is a new architectural/insulator protein required for ribosomal gene expression. *Nucleic Acids Res* 45:gkx840. <https://doi.org/10.1093/nar/gkx840>

11. Annexed Articles

11.1 CTCF loss has limited effects on global genome architecture in *Drosophila* despite critical regulatory functions

CTCF loss has limited effects on global genome architecture in *Drosophila* despite critical regulatory functions

Anjali Kaushal ^{1,14}, Giriram Mohana ^{1,14}, Julien Dorier ^{2,14}, Isa Özdemir ¹, Arina Omer³, Pascal Cousin¹, Anastasiia Semenova¹, Michael Taschner ⁴, Oleksandr Dergai¹, Flavia Marzetta^{2,5}, Christian Iseli ², Yossi Eliaz^{3,6,7}, David Weisz ³, Muhammad Saad Shamim ^{3,8,9}, Nicolas Guex ², Erez Lieberman Aiden^{3,7,10,11,12,13}✉ & Maria Cristina Gambetta ¹✉

Vertebrate genomes are partitioned into contact domains defined by enhanced internal contact frequency and formed by two principal mechanisms: compartmentalization of transcriptionally active and inactive domains, and stalling of chromosomal loop-extruding cohesin by CTCF bound at domain boundaries. While *Drosophila* has widespread contact domains and CTCF, it is currently unclear whether CTCF-dependent domains exist in flies. We genetically ablate *CTCF* in *Drosophila* and examine impacts on genome folding and transcriptional regulation in the central nervous system. We find that CTCF is required to form a small fraction of all domain boundaries, while critically controlling expression patterns of certain genes and supporting nervous system function. We also find that CTCF recruits the pervasive boundary-associated factor Cp190 to CTCF-occupied boundaries and co-regulates a subset of genes near boundaries together with Cp190. These results highlight a profound difference in CTCF-requirement for genome folding in flies and vertebrates, in which a large fraction of boundaries are CTCF-dependent and suggest that CTCF has played mutable roles in genome architecture and direct gene expression control during metazoan evolution.

¹Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland. ²Bioinformatics Competence Center, University of Lausanne, Lausanne, Switzerland. ³The Center for Genome Architecture, Baylor College of Medicine, Houston, TX, USA. ⁴Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland. ⁵Vital-IT, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland. ⁶Department of Physics, University of Houston, Houston, TX, USA. ⁷Center for Theoretical Biological Physics, Rice University, Houston, TX, USA. ⁸Department of Bioengineering, Rice University, Houston, TX, USA. ⁹Medical Scientist Training Program, Baylor College of Medicine, Houston, TX, USA. ¹⁰Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. ¹¹Departments of Computer Science and Computational and Applied Mathematics, Rice University, Houston, TX, USA. ¹²Broad Institute of MIT and Harvard, Cambridge, USA. ¹³Shanghai Institute for Advanced Immunochemical Studies, Shanghai Tech University, Shanghai, China. ¹⁴These authors contributed equally: Anjali Kaushal, Giriram Mohana, Julien Dorier. ✉email: erez@erez.com; mariacristina.gambetta@unil.ch

A wide range of animal genomes are partitioned into a series of contact domains (CDs) that exhibit increased physical proximity among loci within them. An evolutionarily conserved mechanism of such genome folding is thought to be compartmentalization, reflecting the segregation of chromosomal domains based on their transcriptional and epigenetic states^{1–3}. In vertebrates, chromosomal loops are additionally extruded on underlying compartmental domains through a process involving DNA-bound CTCF molecules which stall loop-extruding cohesin complexes at domain boundaries^{1,4–10}. CTCF-dependent extrusion-based boundaries either reinforce or counteract compartmental domain boundaries, depending on the locus. Overall, a large fraction of boundaries in the vertebrate genome are CTCF-dependent^{9,11}.

Intriguingly, although *Drosophila* has widespread CDs and CTCF, it is currently unclear whether CTCF-dependent domains exist in *Drosophila*. High-resolution genome-wide Hi-C maps of formaldehyde-crosslinking frequencies between pairs of DNA fragments (as a measurement of their proximity in 3D-space) were recently generated in *Drosophila* tissue culture cells^{2,12–15}. These studies highlighted the lack of hallmarks of CTCF-mediated domains observed in vertebrate cells. Rather, evidence suggests that CDs in flies are formed by CTCF-independent compartmentalization and other transcription-related processes, as most boundaries lie between domains with different histone modifications or at promoters of highly transcribed genes^{2,12,16–18}.

Crucially, the functional importance of genome folding into CTCF-dependent domains is not fully understood in any organism. CTCF is essential for the viability of mammalian cells^{11,19,20}, whereas it is dispensable for early development in *Drosophila*²¹. Assessing whether or not CTCF-mediated domains exist in *Drosophila* is important for understanding their relevance for genome function. Recent studies have perturbed specific CDs in flies to address their biological roles without knowing whether they are CTCF-mediated or compartmental^{22–24}, yet different types of CDs may have different functions.

CTCF-dependent domains in mammals generally comprise regulatory elements and their target promoters^{25–27}. This suggested that CTCF somehow limits regulatory crosstalk between CDs, and fosters regulatory interactions within them. This model is, however, difficult to test in mammals because global perturbation of CTCF leads to cell death. Acute depletion of CTCF protein in mouse embryonic stem cells followed by transcriptional profiling did not reveal widespread transcriptional changes¹¹. Alternatively, deletion of CTCF binding sites near developmental genes in cultured cells and mice identified some sites where CTCF appears to critically prevent developmental defects and disease^{28–30}, and many CTCF sites that did not appear functional^{31–33}. These diverse results paint an opaque picture of how CTCF impacts gene expression. Previous studies that partially knocked-down CTCF in *Drosophila* cell lines also did not reveal clear effects on transcription^{34–36}. Analysis of the homeotic phenotype of CTCF⁰ mutants completely lacking both maternal and zygotic CTCF suggested that CTCF blocks regulatory crosstalk between elements on either side of some CTCF binding sites²¹. A fundamental question arising from comparative studies in flies and humans is how CTCF impacts transcription, and how this relates to its uncertain architectural function in flies. Whether CTCF stably associates with partner proteins to effect its functions also remains unclear.

Here, we show using CTCF⁰ mutant *Drosophila* that CTCF is critically required in neurons for fly viability. We examine the effects of CTCF loss on genome folding and transcriptional regulation in the central nervous system (CNS) and investigate the molecular basis of CTCF function.

Results

CTCF expression in neural stem cells (NSCs) or neurons is essential for fly viability. To identify a biologically relevant tissue in which to study CTCF function in *Drosophila*, we used previously described CTCF knock-out (CTCF^{KO}) mutants and CTCF⁰ mutants that additionally lack maternally inherited CTCF²¹. Some CTCF^{KO} mutants (60%) hatch into adults with spasmodic movements suggesting a neurological phenotype that might be the cause of their short lifespan (Figs. 1a, 1b, Supplementary Movie 1). We tested the relevance of CTCF expression in the nervous system by performing tissue-specific knock-out and rescue experiments. Specifically, we used Gal4 drivers active in NSCs, mature neurons or muscles to drive conditional excision of a CTCF rescue transgene (knock-out) or UAS-CTCF expression (rescue) in CTCF mutant genetic backgrounds. Loss of CTCF expression in NSCs or neurons compromised the ability of flies to hatch to a comparable extent as loss of all zygotic CTCF expression (Fig. 1a) and severely shortened the life span of flies that did hatch (Fig. 1b, Supplementary Movie 2). On the other hand, loss of CTCF in muscle only slightly impaired adult hatching and life span (Figs. 1a, b).

In contrast to CTCF^{KO}, CTCF⁰ mutants never hatch from the pupal case (Fig. 1c). Conditional expression of CTCF in NSCs or neurons of CTCF⁰ mutants strongly rescued hatching (Fig. 1c) and adults were capable of coordinated movements and survived for several days (Fig. 1d, Supplementary Movie 3). On the other hand, expressing CTCF in muscles of CTCF⁰ mutants barely rescued hatching (Fig. 1c, d).

Together, these results show that CTCF expression is critically required in neurons for pupal hatching and adult viability. Consistently, CTCF is more highly expressed in the nervous system than in other tissues^{37,38}. Analyses of molecular phenotypes of CTCF⁰ mutants described hereafter were therefore performed in dissected CNSs of third instar larvae, a developmental stage at which CTCF⁰ mutants are fully viable.

Physical insulation defects in CTCF⁰ mutants. To address whether CTCF is required to form CD boundaries in flies, Hi-C was performed on CNSs dissected from wildtype (WT) and CTCF⁰ larvae in biological triplicate using two 4-cutter restriction enzymes for enhanced resolution. Hi-C maps consisting of 200 million reads per genotype were obtained by combining the correlated biological replicates (see Methods, Supplementary Table 1). Hi-C maps from whole bodies of single flies of the same genotypes were also generated. In parallel, CTCF binding sites were mapped in larval CNSs by chromatin immunoprecipitation sequencing (ChIP-seq) with a polyclonal antibody specifically recognizing CTCF (Supplementary Fig. 2a) in WT and in CTCF⁰ animals as control. Only 740 CTCF peaks were defined as enriched in WT relative to CTCF⁰ CNSs, of which 77% overlapped a CTCF consensus motif (Supplementary Fig. 2b, Supplementary Data 1).

To assess the relation between CTCF peaks and CD boundaries genome-wide in WT CNS Hi-C maps, boundaries were identified at 2 kb resolution with TopDom (see “Methods”, Supplementary Table 2, Supplementary Data 2 and 3). Very few (<1%) boundaries defined in this study potentially correspond to small CDs defined in even higher resolution Hi-C studies (see “Methods”). Domain boundaries were enriched within ± 1 kb of several (36%) CTCF peaks (Fig. 2a). Conversely, a CTCF peak was located within ± 1 kb of only 8% of all boundaries (Fig. 2b). This indicates that while CTCF peaks are frequently at domain boundaries, CTCF is only present at a small fraction of all boundaries in flies.

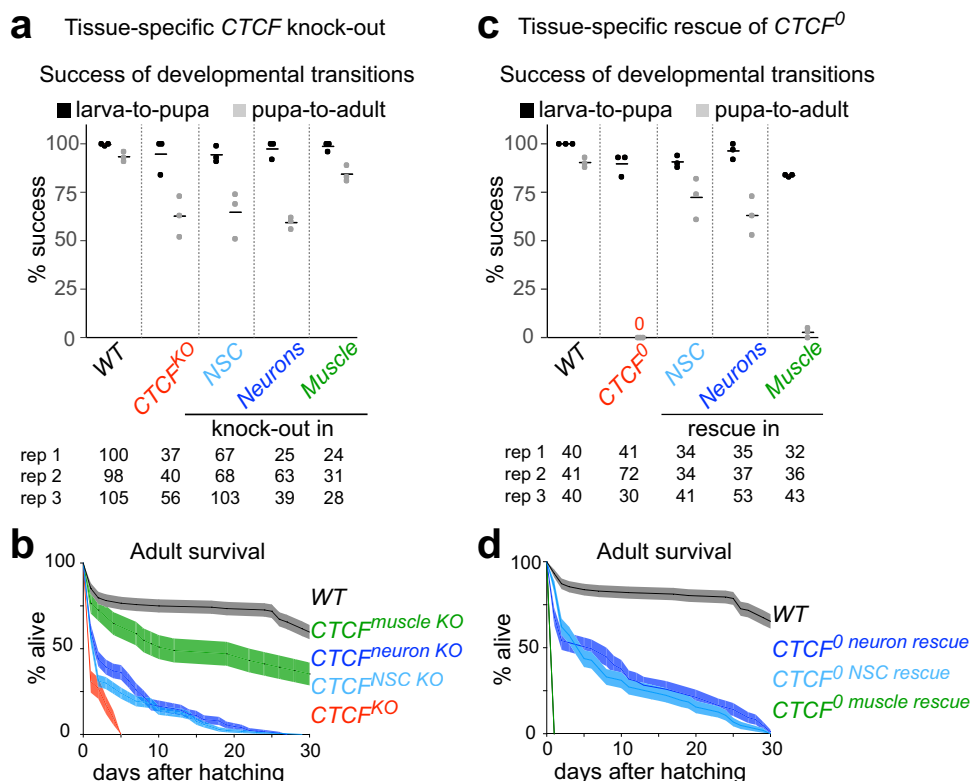


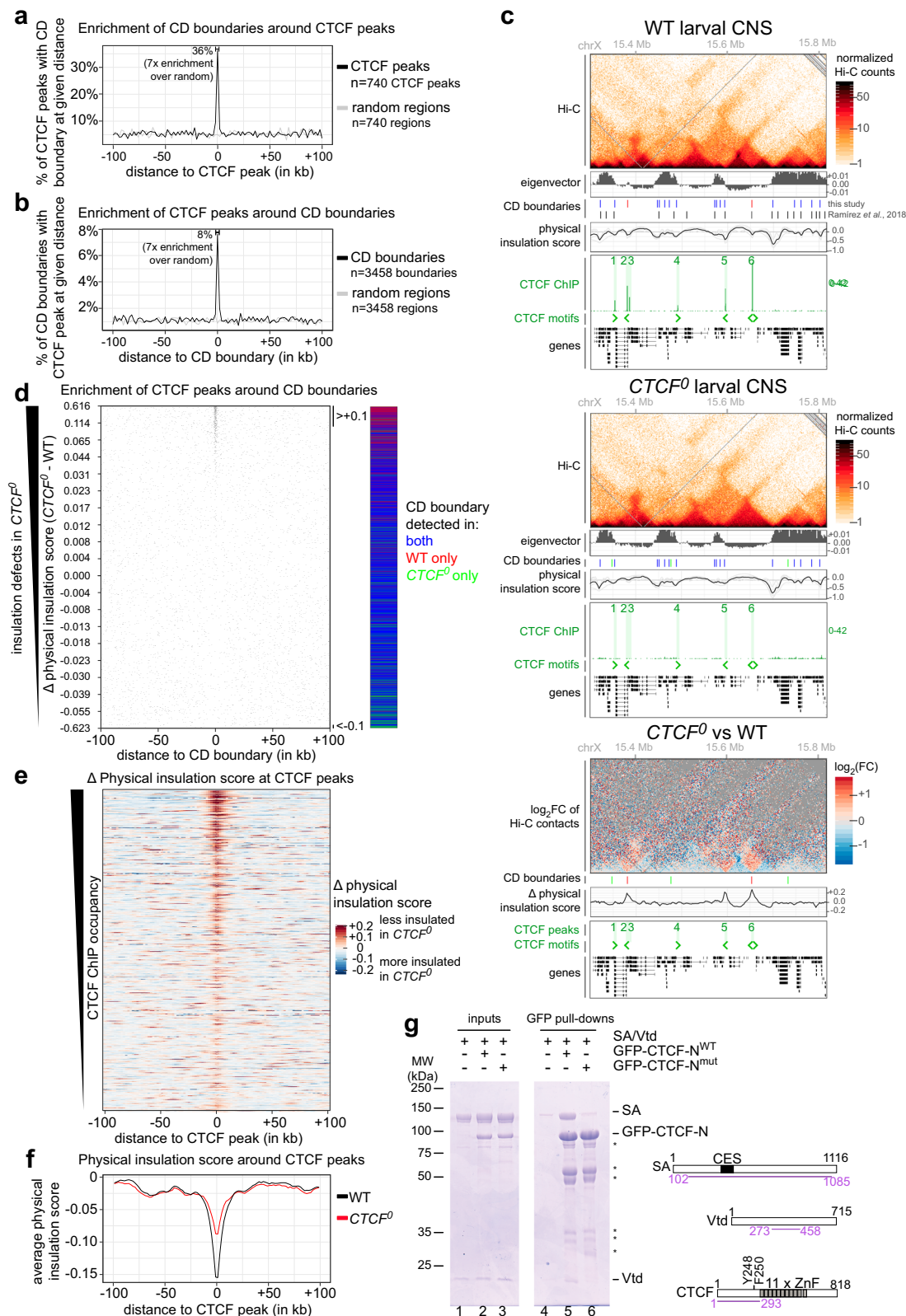
Fig. 1 CTCF expression in neural stem cells or neurons is essential for *Drosophila* viability. **a** Percentages (in y) of wildtype (WT), *CTCF*^{KO} and animals lacking CTCF in neural stem cells (NSCs), neurons or muscle that successfully transition from third instar larva to pupa (black) and from pupa to adult (gray) in *n* = 3 biological replicates (rep 1-3), each containing the indicated number of animals per genotype. Horizontal lines show means. **(b)** Percentage of live adults of the same genotypes up to 30 days after pupal hatching. 40 animals of each genotype were analyzed in triplicate; dark lines show mean and shading shows ±standard deviation. **c** Same as **(a)** but for WT, *CTCF*⁰ and animals with CTCF expression restricted to NSCs, neurons or muscle. **d** Same as **(b)** but for genotypes indicated in **(c)**.

WT and *CTCF*⁰ Hi-C maps were globally similar, and most (84%) domain boundaries were detected in both WT and *CTCF*⁰ mutants. Nevertheless, specific CDs were visibly less physically insulated from the neighboring domain in *CTCF*⁰ mutants (Fig. 2c, Supplementary Fig. 2c, Supplementary Table 3). Clearly disrupted domain boundaries in *CTCF*⁰ mutants frequently occurred at former CTCF peaks (Fig. 2d). Of 135 strongly affected domain boundaries that were lost in *CTCF*⁰ mutants, 89 (66%) were at former CTCF peaks (Supplementary Table 2). To determine how generally physical insulation defects are observed at former CTCF peaks in the absence of CTCF (irrespective of their localization at CD boundaries identified by TopDom), physical insulation score differences between WT and *CTCF*⁰ mutants were measured across all 740 CTCF peaks. Boundary defects in *CTCF*⁰ mutants were observed at most former CTCF peaks, with more prominent defects visible at CTCF peaks that are highly occupied in WT (Fig. 2e, Supplementary Fig. 2d). CTCF-dependent boundaries were variably positioned relative to neighboring genes (see examples in Fig. 2c: CTCF peaks 2, 3, 5 and 6 are respectively in an intron, at the end of a gene, within 1 kb of a gene promoter or intergenic). Many CTCF-dependent boundaries were similarly affected in Hi-C maps from whole-body flies of the same genotypes, indicating that CTCF is required to form physical boundaries in most cell types (Supplementary Fig. 2e). Together, these results strongly suggest that CTCF mediates the formation of physical boundaries.

Whereas domain boundaries were abolished at several former CTCF peaks in *CTCF*⁰ mutants, they were partially retained at other peaks that are similarly occupied by CTCF in WT (Supplementary Fig. 2c, compare boundary defects at CTCF

peaks 5 and 6). Of 343 WT CD boundaries bound by CTCF, only 125 (36%) were fully lost in *CTCF*⁰ mutants (Supplementary Table 2). This resulted in a lower average physical insulation score at former CTCF peaks in *CTCF*⁰ mutant CNS Hi-C maps (Fig. 2f). These observations are not due to the presence of contaminating CTCF, as CTCF RNA and protein are undetectable by RNA-seq and ChIP-seq (Fig. 2c and next section). As *CTCF*⁰ mutants lack CTCF from the beginning of development, residual boundaries can also not be explained by a role of CTCF in the establishment but not maintenance of boundaries. Rather, this observation suggests that at some sites, CTCF reinforces boundaries redundantly established by other mechanisms, a scenario also observed in mammalian cells^{1,2}. We define CTCF-occupied CD boundaries present only in WT as strictly CTCF-dependent, and those that are present in *CTCF*⁰ (generally weaker than in WT) as partially CTCF-dependent. These two types of CTCF-dependent boundaries are contrasted later in the “Results” section.

A region in the N-terminus of human CTCF directly interacts with cohesin and stabilizes cohesin on DNA^{10,39}, partly explaining how human CTCF forms CD boundaries. Vertebrate and fly CTCF N-termini are highly diverged, yet a 10 amino acid residue stretch in CTCF’s N-terminus that binds to cohesin in human cells is present at a similar distance from the zinc finger domain in fly CTCF¹⁰ (boxed in Supplementary Fig. 2f). We therefore tested whether two residues critical for cohesin interaction in human CTCF (Y226 F228, homologous to Y248 F250 in fly CTCF) mediate direct interaction of fly CTCF with the SA-Vtd (homologous to human SA2-SCC1) complex. For this, GFP-tagged recombinant WT and Y248A F250A point mutant



CTCF N-termini were mixed with an untagged SA-Vtd subcomplex and purified on GFP binder beads. WT but not mutant CTCF versions retained SA-Vtd (Fig. 2g). Therefore, despite profound divergence, the fly CTCF N-terminus interacts directly with cohesin *in vitro*. This interaction was suggested to impart directionality to CTCF-dependent boundaries in mammalian cells^{10,39}, but we find that CTCF has at best a very weak

preference to establish directional boundaries (Supplementary Fig. 2g) consistent with a previous study².

We conclude that *Drosophila* CTCF is required to form physical boundaries with strengths generally proportional to its occupancy on DNA. Other mechanisms reinforce CTCF-dependent boundaries at some sites and explain the formation of most boundaries in flies.

Fig. 2 Physical insulation defects in *CTCF⁰* mutants. **a** Percentage of $n = 740$ CTCF peaks with at least one contact domain (CD) boundary at a given distance (per 2 kb bins) around the CTCF peak. Enrichment of CD boundaries around the same number of random positions (gray) is shown as control. **b** Percentage of $n = 3458$ CD boundaries with at least one CTCF peak at a given distance (per 2 kb bins) around the CD boundary. Enrichment of CTCF peaks around the same number of random positions (gray) is shown as a control. **c** Example locus (dm6 coordinates) Hi-C maps, eigenvector values (positive for A compartment, negative for B compartment), CD boundaries from this study (color-coded as in Fig. 2d) and a Hi-C study in cultured cells¹⁷, physical insulation score (calculated with different window sizes in gray, average in black), CTCF ChIP-seq (CTCF peaks highlighted and numbered), CTCF motif orientations in DNA, and gene tracks in WT (top) and *CTCF⁰* (middle) larval CNSs. (Below) Differential (*CTCF⁰* minus WT) Hi-C map and physical insulation score. **d** Position of CTCF peaks around all CD boundaries defined in any genotype ($n = 3970$ boundaries) ranked by physical insulation score differences measured in *CTCF⁰* minus WT Hi-C maps. Visibly weaker boundaries in *CTCF⁰* (score $> +0.1$) or in WT (score < -0.1) are bracketed. Boundaries are color-coded in all figures as present in both WT and *CTCF⁰* (blue), only in WT (red) or only in *CTCF⁰* (green). **e** Physical insulation score differences measured in *CTCF⁰* minus WT Hi-C maps around CTCF peaks, ranked by CTCF ChIP occupancy in WT. **f** Average physical insulation scores around CTCF peaks in WT (black) and *CTCF⁰* (red). **g** GFP pull-down of tagged CTCF N-terminus (residues 1–123) that is WT (GFP-CTCF-N^{WT}) or Y248A F250A point mutant (GFP-CTCF-N^{mut}) mixed with untagged recombinant cohesin subcomplex (residues 102–1085 of SA and 273–458 of Vtd). Specific retention of cohesin by CTCF (lane 5) is higher than the background binding of SA-Vtd to beads (lanes 4, 6). Asterisks mark GFP-CTCF-N degradation. CES conserved essential surface, ZnF zinc finger.

CTCF impacts expression patterns of genes near CTCF peaks.

To understand how CTCF impacts transcription, we performed RNA sequencing (RNA-seq) on cDNA libraries from mRNA purified from WT and *CTCF⁰* larval CNSs in triplicate. This confirmed the absence of *CTCF* mRNA in *CTCF⁰* samples (Supplementary Fig. 3a). 392 (~3% of all) genes were significantly differentially expressed (DE) in *CTCF⁰* mutants (with adjusted p -value < 0.05 and $|\text{fold-change}| > 1.5$) (Fig. 3a, Supplementary Data 4). *CTCF⁰* mutants therefore do not show widespread transcriptional defects, though changes occurring in subsets of cells in the CNS such as CTCF's previously validated target gene *Abdominal-B* elude our analysis²¹.

Some DE genes had decreased expression in *CTCF⁰* mutant CNSs compared to WT (Fig. 3b). Several DE genes with increased expression in *CTCF⁰* CNSs are normally not expressed in the CNS but rather restricted to other specialized tissues like testes (*Intraflagellar transport 52*), tendons (*Thrombospondin*), and the peripheral nervous system (*Odorant receptor 67d*) (Figs. 3c, 3d, Supplementary Fig. 3b). Some ectopic transcripts lacked annotated start and termination sites suggesting that they are cryptic (Supplementary Fig. 3b). RNA fluorescent in situ hybridization (RNA-FISH) analysis showed that genes with increased expression in *CTCF⁰* CNSs were misexpressed in various patterns, possibly driven by locus-specific enhancers (Fig. 3e).

Indirect transcriptional changes are expected in *CTCF⁰* mutants, which lack CTCF since the beginning of development, and we asked whether CTCF regulates genes in the vicinity of its binding sites. 10% of DE genes had a CTCF peak within ± 1 kb of their transcriptional start site (TSS) (ninefold enrichment over randomly sampled matched non-DE genes) (Fig. 3f), a result that was not very different for genes with increased versus decreased expression in *CTCF⁰* mutants (Supplementary Fig. 3c). Conversely, 5% of CTCF peaks were located within ± 1 kb of a DE gene TSS (9-fold enrichment over randomly sampled matched non-DE genes) (Fig. 3g). These results suggest that, depending on the locus, CTCF may directly repress or activate the transcription of nearby genes, or alternatively CTCF may shield promoters from inappropriate enhancers or silencers as observed at *Hox* gene loci^{21,40}.

Could the structural defects observed in *CTCF⁰* Hi-C maps be secondary consequences of gene misregulation in the vicinity of former CTCF peaks? Some CTCF-dependent domain boundaries were located far from genes (Fig. 2c, CTCF peak 6 is 9 kb away from the closest gene) and are thus unlikely to be impacted by transcription. Others were located near genes whose expression increased (Supplementary Fig. 2c, peak 3), decreased (Supplementary Fig. 2c, peak 6) and in most cases remained unchanged (Supplementary Fig. 2c, peak 7). Few (8%) DE genes were located

in different A/B compartments in *CTCF⁰* mutants relative to WT, indicating that differential gene expression mostly occurred without large changes in higher-order spatial chromatin configuration (Supplementary Fig. 3d, Supplementary Data 5). Together, these results indicate that the pervasive weakening of physical boundaries observed at former CTCF peaks in *CTCF⁰* mutants (Fig. 2e–f) is not a mere consequence of altered transcription.

CTCF occupancy scales with enhancer-blocker activity in a reporter assay.

Previous studies of the functionality of CTCF binding sites stably integrated into the fly genome suggested that most of them lack insulator activity (i.e., the ability to block regulatory crosstalk)³⁶, at least in single copies⁴⁰. Here, we tested CTCF peaks in a quantitative reporter assay. The reporter comprises an enhancer positioned between two fluorescent reporter genes (EGFP and mCherry) driven by minimal *Heat-shock-protein-70* (*Hsp70*) promoters (Fig. 4a). Test fragments were cloned in between EGFP and the enhancer, maintaining the enhancer at a similar distance from both reporter genes. Reporter plasmids were then transiently transfected into *Drosophila* S2 cells, and relative EGFP and mCherry intensities were measured in thousands of single cells with a cell analyzer (Supplementary Fig. 4a). An insulator should reduce EGFP expression while mCherry expression should remain high. Control experiments with a neutral spacer or the well-characterized *gypsy* insulator⁴¹ validated the assay (Fig. 4b, lanes 1 and 2). Two CTCF peaks near genes whose expression decreased (peak G from Fig. 3b) or increased (peak N from Fig. 3e) in *CTCF⁰* mutants had similar effects as *gypsy* (Fig. 4b, lanes 3 and 4). EGFP levels in the presence of *gypsy* or CTCF peaks were not strongly reduced below basal levels measured in enhancer-less control reporters (Supplementary Fig. 4b), suggesting that these tested sequences mostly impaired enhancer-mediated EGFP expression. Additional CTCF peak regions (Supplementary Fig. 4c, average size 360 bp) were tested and their relative insulator strengths were estimated from the median ratio of mCherry-over-EGFP fluorescence measured in single cells. Eleven out of 14 tested CTCF peaks selectively reduced EGFP intensities to various degrees that globally scaled with CTCF ChIP-seq occupancy measured in S2 cells⁴² (Fig. 4c) and that appeared independent of the endogenous locations of CTCF peaks relative to their nearest genes (Supplementary Fig. 4c) and of combinatorial co-binding with other fly insulator-binding proteins on the cloned fragments (Supplementary Fig. 4d). Mutating two base pairs of a CTCF motif in one of these fragments abolished its activity (Fig. 4c, fragment N mut); thus, the reporter specifically reveals the activity of a single CTCF

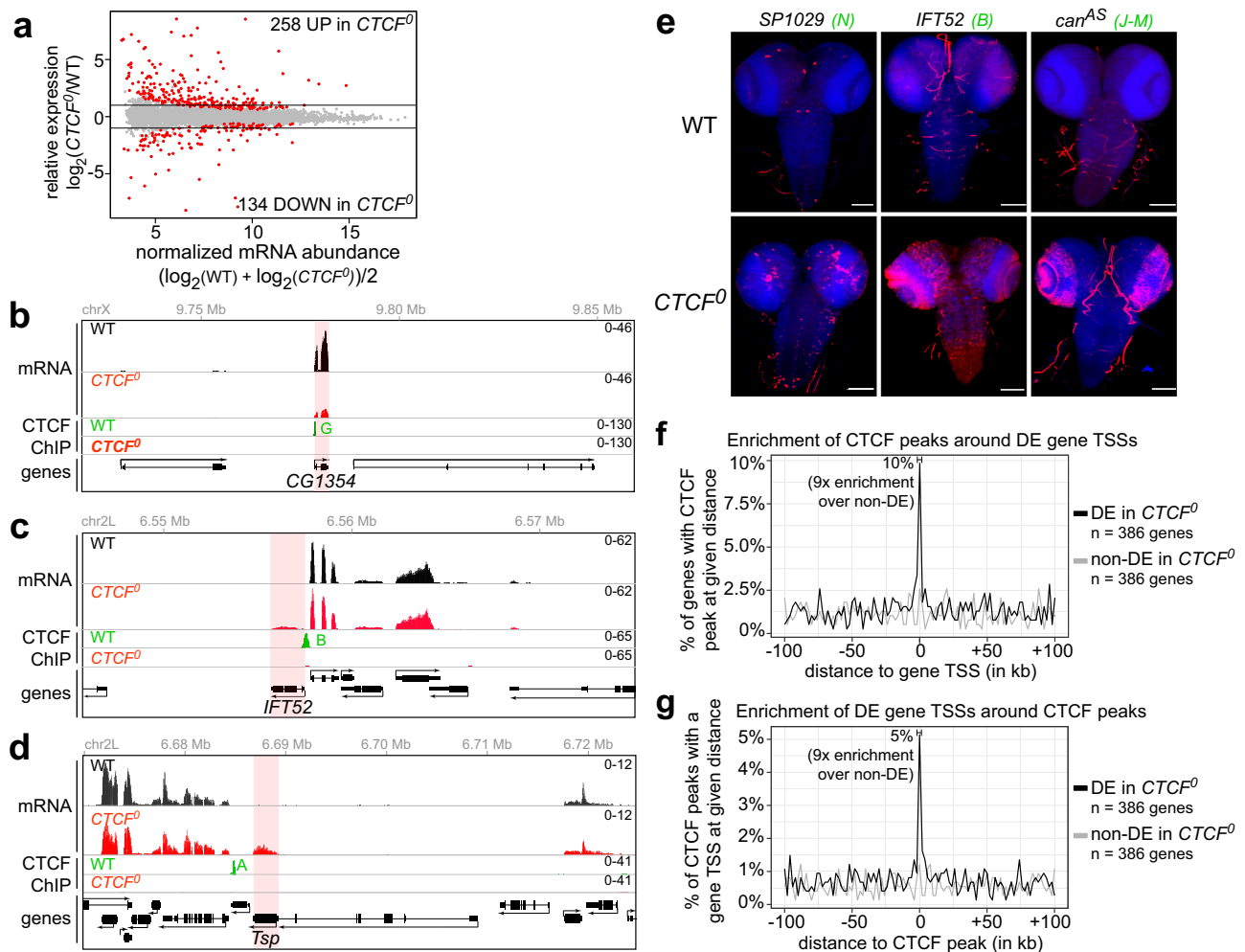


Fig. 3 CTCF impacts expression patterns of genes near CTCF peaks. **a** RNA-seq MA plot of $CTCF^0$ versus WT CNSs with mean abundance (in x) plotted as a function of enrichment (in y). Differentially expressed (DE) genes ($p_{adj} < 0.05$ and $|\text{fold change}| > 1.5$) are red. **b–d** RNA-seq signals in WT (black) and $CTCF^0$ (red) larval CNSs, and CTCF ChIP-seq signals in WT (green) and $CTCF^0$ (red) larval CNSs at *CG1354* (**b**), *IFT52* (**c**) and *Tsp* (**d**) loci. Differentially transcribed regions are shaded in red. Scales in tracks of all figures indicate reads per million. In all figures, CTCF peaks labeled by capital letters were tested in Fig. 4c. **e** RNA-FISH with antisense probes (red) against indicated transcripts in CNSs of wildtype and $CTCF^0$ larvae stained by DAPI (blue) (scale bars 100 μm). mRNAs of *SP1029*, *IFT52* and an antisense transcript overlapping *can* (shown in Supplementary Fig. 3b) are normally not expressed in wildtype CNSs (background signal is sometimes visible in trachea) and are misexpressed in different patterns in $CTCF^0$ mutants. All animals showed similar misexpression patterns for a given transcript. **f** Percentage (in y) of $n = 386$ DE genes in $CTCF^0$ larval CNSs (black) or $n = 386$ randomly sampled expression-level-matched non-DE genes (gray) with at least one of 740 CTCF peaks at a given distance (per 2 kb bins) around the gene TSS, measured in the direction of transcription (in x). Ten percent of DE genes have at least one CTCF peak within ± 1 kb of their TSS, which is ninefold higher than the average enrichment at the sampled non-DE genes. **g** Percentage (in y) of CTCF peaks with at least one of $n = 386$ DE gene TSSs (black) or $n = 386$ randomly sampled expression-level-matched non-DE gene TSSs (gray) at a given distance (per 2 kb bins) around CTCF peaks, measured in the direction of transcription (in x). Five percent of CTCF peaks have at least one DE gene TSS within ± 1 kb, which is 9-fold higher than at the sampled non-DE TSSs.

binding site. Taken together, these observations indicate that CTCF sites in the reporter do not strongly directly repress or activate transcription but rather insulate a promoter from an enhancer.

CTCF recruits Cp190 to a subset of Cp190-bound domain boundaries. To further understand how CTCF functions, we asked whether it stably associates with partner proteins that contribute to its activity. Unbiased identification of CTCF partners from *Drosophila* embryonic nuclear extracts in biological duplicates by mass spectrometry reproducibly identified known insulator-binding proteins Centrosomal protein 190 kDa (Cp190) and Insulator binding factors 1 and 2 (Ibf1 and Ibf2) as enriched CTCF interactors relative to negative control (Supplementary

Fig. 5a). Reciprocal Cp190 purifications published by others also identified Ibf1, Ibf2 and CTCF among other proteins⁴³. Traces of the cohesin complex also co-purified with CTCF (Supplementary Fig. 5a) reminiscent of transient interactions between cohesin and CTCF seen in mammalian cells⁴⁴.

CTCF was previously shown to directly interact with Cp190⁴⁵, yet the relevance of this interaction remained unclear. No common target genes are known⁴⁶ and a mutant version of CTCF reported to no longer interact with Cp190 was largely functional in vivo⁴⁵. We performed pull-downs of GFP-tagged CTCF fragments co-expressed in bacteria with Cp190's BTB (Broad-Complex, Tramtrack and Bric-a-brac) domain and found that amino acids 698–771 in CTCF C-terminus directly interact with Cp190 BTB (Supplementary Fig. 5b). Importantly, this stretch in

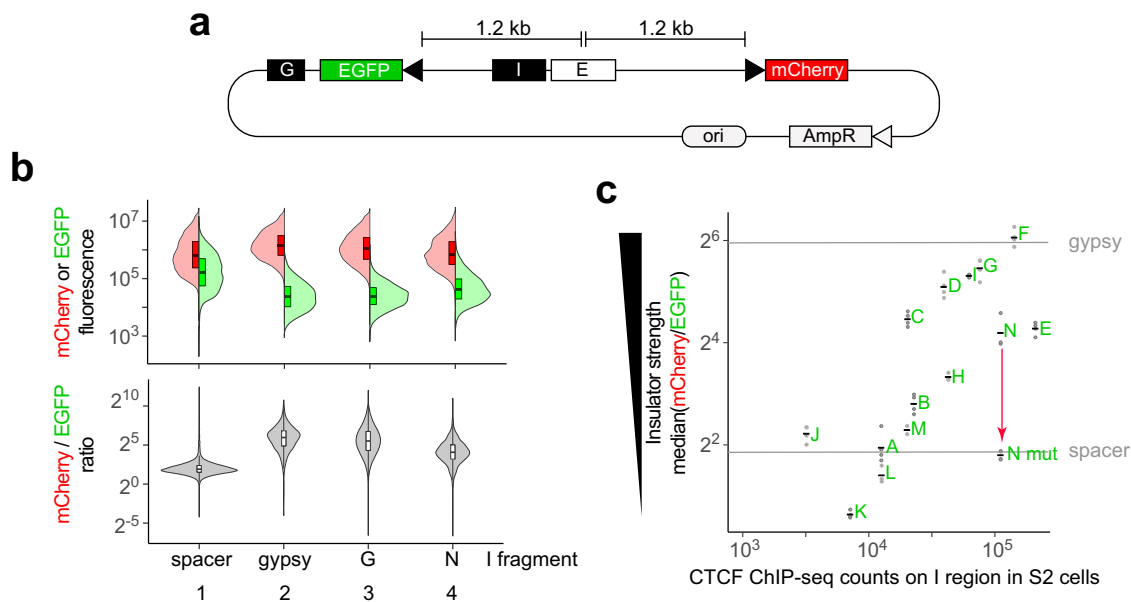


Fig. 4 CTCF occupancy scales with enhancer-blocker activity in a reporter assay. **a** In the reporter plasmid, a test insulator I is cloned in between an enhancer E and EGFP, and mCherry serves as a reference (elements are drawn to scale, arrowheads represent *Hsp70* minimal promoters). A *gypsy* insulator G is present downstream of EGFP to block EGFP activation by the enhancer (which in a circular plasmid molecule is both upstream and downstream of EGFP) from the left. **b** Split violin plots (thick lines mark medians, boxes mark interquartile ranges) show distributions of mCherry (left) and EGFP (right) fluorescence intensities (\log_{10} values in y) measured in thousands of single S2 cells transiently transfected with reporters with indicated I fragments (in x). mCherry-to-EGFP ratios (\log_2 values in y) in single cells are shown below. For each reporter, merged biological triplicates are plotted. **c** Median mCherry-to-EGFP ratios in single transfected S2 cells (\log_2 values in y) relative to CTCF ChIP-seq counts in S2 cells⁴² (\log_{10} values in x) on selected CTCF peaks (labeled A–N, Supplementary Fig. 4c) cloned as I fragments. $n = 2$ (M), 3 (H, I, L, N) or 4 (A–G, J, K, N mut) biological replicates (dots) and mean values (horizontal lines) are shown. As a reference, mean values obtained with the *gypsy* insulator or a neutral spacer are indicated as horizontal lines. As a control, a CTCF motif in fragment N was mutated, leading to fragment N mut for which CTCF ChIP occupancy was not determined.

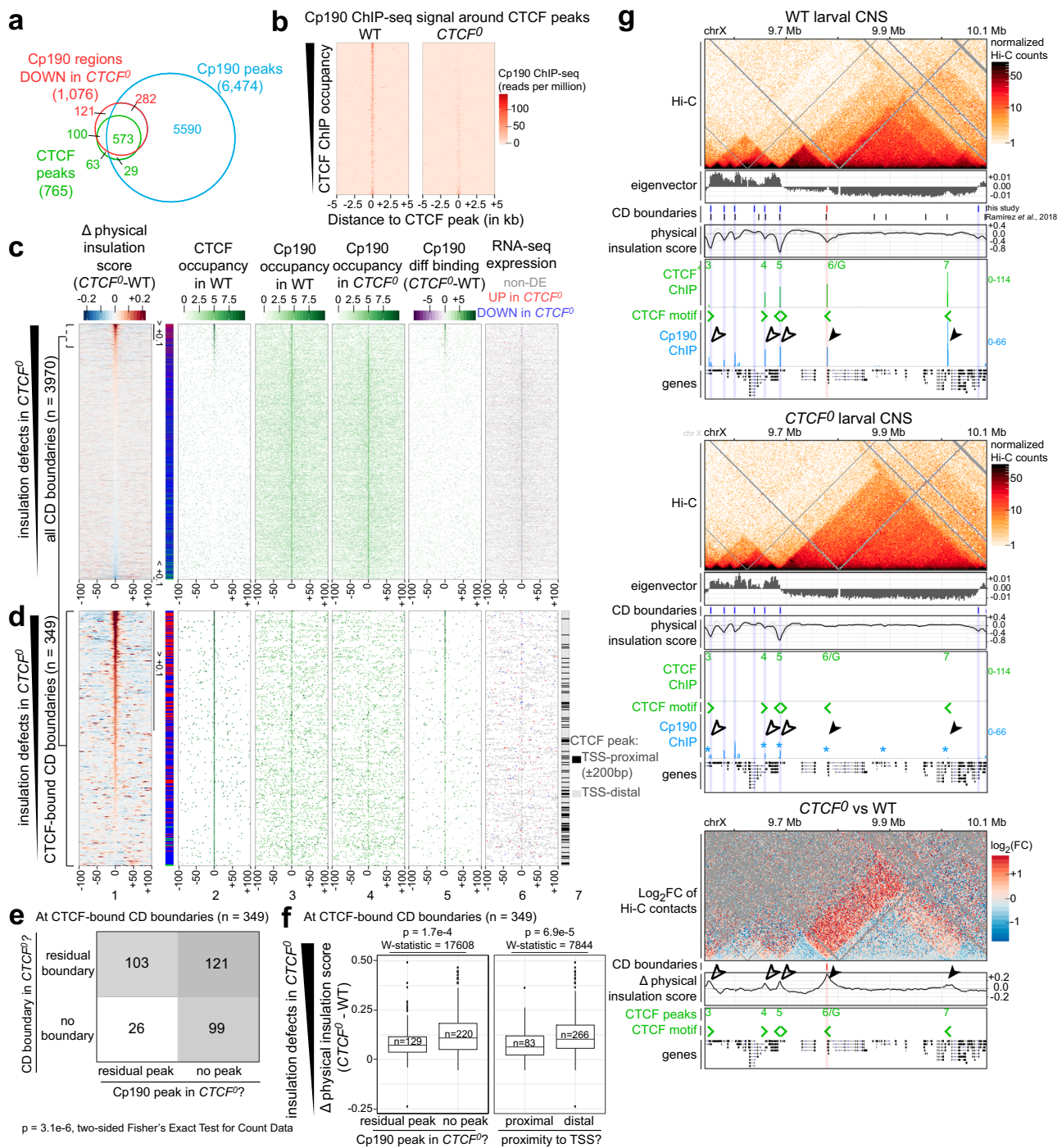
CTCF does not overlap the previously deleted region (amino acid residues 774–818) that was used to conclude that CTCF's interaction with Cp190 was unimportant in vivo.

To assess the genome-wide overlap between CTCF and Cp190 binding sites in larval CNSs, specific Cp190 peaks were identified by ChIP-seq with a polyclonal anti-Cp190 antibody in WT and in *Cp190^{KO}* animals with a CRISPR-Cas9 mediated deletion of the *Cp190* open reading frame as control (Supplementary Fig. 5c). 6,473 Cp190 peaks were enriched in WT relative to *Cp190^{KO}* CNSs (Fig. 5a, Supplementary Data 6). Cp190 colocalized with CTCF at most (79%) CTCF peaks and was additionally present at many other sites (Fig. 5a), consistent with other studies^{35,36,47}. We profiled Cp190 binding sites in WT and *CTCF⁰* larval CNSs and found that Cp190 was normally recruited to most Cp190 peaks in *CTCF⁰* mutants with the exception of former CTCF peaks, at which Cp190 was globally reduced (Figs. 5a, 5b, Supplementary Data 7 and 8). In *CTCF⁰* mutants, Cp190 was lost from former higher-occupancy CTCF peaks but only reduced at former lower-occupancy CTCF peaks (Fig. 5b). We therefore distinguish between strictly CTCF-dependent Cp190 peaks (lacking a detectable Cp190 peak when comparing *CTCF⁰* and *Cp190^{KO}* mutants) and partially CTCF-dependent Cp190 peaks (with a detectable Cp190 peak in *CTCF⁰* relative to *Cp190^{KO}* mutants, generally weaker in *CTCF⁰* than in WT).

Unlike CTCF, Cp190 binding was enriched at CD boundaries genome-wide (Fig. 5c lane 3, Supplementary Figs. 5d, e)^{2,15,17}. Outside of CTCF peaks, Cp190-occupied domain boundaries were often proximal to transcribed TSSs (Fig. 5c, lane 6). In *CTCF⁰* mutants, residual Cp190 binding at former CTCF-occupied boundaries was significantly associated with boundary retention (Figs. 5d–f). Seventy-five percent of strictly CTCF-dependent boundaries lacked a residual Cp190 peak, and 80% of

residual Cp190 peaks were associated with a residual boundary in *CTCF⁰* mutants (Fig. 5e). CD boundary defects in *CTCF⁰* mutants were also less severe at former TSS-proximal CTCF peaks (within 200 bp of a gene TSS) than at former TSS-distal CTCF peaks (Fig. 5f). This suggests that either Cp190 itself, its associated factors, or transcription at Cp190-bound TSSs may redundantly contribute to the formation of physical boundaries independently of CTCF and may synergize with CTCF at partially CTCF-dependent Cp190 peaks (see examples in Fig. 5g).

CTCF and Cp190 co-regulate a subset of target genes. To assess whether loss of Cp190 results in transcriptional changes shared with *CTCF⁰* mutants, RNA-seq was performed on *Cp190^{KO}* larval CNSs in biological triplicate. Overall, 440 DE genes were observed in *Cp190^{KO}* mutant CNSs compared to WT, of which 192 went up and 248 went down relative to WT (with adjusted p -value < 0.05 and |fold-change| > 1.5) (Supplementary Fig. 6a, Supplementary Data 9). Since Cp190 is bound to many more sites than CTCF (Fig. 5a), we did not expect that many transcriptional changes in *Cp190^{KO}* mutants would be shared in *CTCF⁰* mutants. Surprisingly, however, a considerable fraction of DE genes in *CTCF⁰* and *Cp190^{KO}* mutants were common (31% of all DE genes in *CTCF⁰* and 26% of all DE genes in *Cp190^{KO}*) and concordantly changed in similar directions and to similar degrees relative to WT (Fig. 6a). This is exemplified at the *SP1029* (Fig. 6b–c) and *CG15478* (Fig. 6d–e) genes that are proximal to a CTCF and Cp190 co-bound peak (peak 1/N in Fig. 6b, peak 2 in Fig. 6d). In the absence of CTCF, Cp190 is additionally lost from these peaks (Figs. 6b and d, middle), a CD boundary is disrupted (Supplementary Figs. 6b and c), and the gene is expressed at increased (*SP1029* in Fig. 6b, middle) or decreased (*CG15478* in Fig. 6d, middle) levels relative to



WT. In the absence of Cp190, CTCF remains bound at *SP1029* (Fig. 6b, bottom) and *CG15478* (Fig. 6d, bottom) which are nevertheless also similarly misexpressed relative to WT (Figs. 6b and d, bottom). This suggests that Cp190 is required for CTCF function independently of CTCF binding to DNA. To more stringently compare *SP1029* and *CG15478* misexpression in the absence of CTCF or Cp190, we visualized their mRNAs in embryos completely lacking maternal and zygotic CTCF (*CTCF⁰*) or Cp190 (*Cp190⁰*). Already at 11 h of development, *CTCF⁰* and *Cp190⁰* embryos ectopically expressed *SP1029* in the same cells (in the nervous system and additional cell types) (Fig. 6c) and failed to express WT levels of *CG15478* in the nervous system (Fig. 6e). We conclude that Cp190 is a critical partner of CTCF for regulating a subset of common genes (see summary model in Fig. 6f).

Discussion

CTCF-dependent CDs have been proposed to regulate the communication between genes and their regulatory elements. Here, we analyzed *Drosophila* that developed in the complete absence of CTCF and reached the following conclusions: (1) CTCF is most critically required in neuronal cells for adult viability (Fig. 1). (2) Domain boundary defects in *CTCF⁰* mutants are overwhelmingly associated with CTCF-bound sites, consistent with a mechanism in which CTCF can form boundaries (Fig. 2). At the same time, the vast majority of boundaries are CTCF-independent. (3) CTCF prevents ectopic activation and silencing of certain genes in its vicinity (Fig. 3). (4) Sites bound by CTCF do not directly repress or activate transcription, but rather functionally insulate promoters and enhancers in a reporter assay in S2 cells (Fig. 4). (5)

Fig. 5 CTCF recruits Cp190 to a subset of Cp190-bound domain boundaries. **a** Overlap between CTCF (green) and Cp190 (blue) peaks in WT, and regions with reduced Cp190 binding in *CTCF⁰* relative to WT (red). Some peaks were split for three-way comparisons (see “Methods”). **b** Cp190 ChIP-seq signal in WT or *CTCF⁰* around CTCF peaks, ranked by CTCF occupancy in WT. **c** Distribution of indicated datasets around CD boundaries defined in any genotype ($n = 3970$ boundaries) ranked by insulation defects in *CTCF⁰*. (1) Insulation score differences in *CTCF⁰* minus WT. Visibly weaker boundaries in *CTCF⁰* (score $> +0.1$) or in WT (score < -0.1) are bracketed. On the right, boundaries are classified as in Fig. 2d. (2–4) ChIP occupancy of CTCF peaks in WT, Cp190 peaks in WT or Cp190 peaks in *CTCF⁰*. (5) Differential Cp190 ChIP occupancy in *CTCF⁰* minus WT. (6) Expressed TSSs in WT and *CTCF⁰* with similar (gray), increased (red) or decreased (blue) expression in *CTCF⁰* relative to WT. **d** As above for CD boundaries with a CTCF peak within ± 2 kb ($n = 349$ boundaries) centered on the closest CTCF peak classified as TSS-proximal (within ± 200 bp of a TSS) or distal (lane 7). **e** Numbers of CD boundaries bound by CTCF in WT ($n = 349$ boundaries) that are present or absent in *CTCF⁰* mutants, and whose associated CTCF peak overlaps or not a residual Cp190 peak in *CTCF⁰* mutants (p -val = $3.1e-6$, two-sided Fisher’s Exact Test for Count Data). **f** Physical insulation score differences in *CTCF⁰* minus WT at CTCF-bound CD boundaries ($n = 349$ boundaries) are higher when the associated CTCF peak does not overlap a residual Cp190 peak in *CTCF⁰* mutants, or a TSS within 200 bp (indicated p values and W -statistics from two-sided Wilcoxon rank-sum test with continuity correction). Box plot: center line, median; box limits, upper and lower quartiles; whiskers, $1.5\times$ interquartile ranges; points, outliers; $n =$ CTCF peaks of each category (in x). **g** Example locus like Fig. 2c also displaying Cp190 ChIP-seq signal in WT and *CTCF⁰* mutant larval CNSs. Asterisks mark Cp190 peaks in *CTCF⁰* mutants with reduced occupancy relative to WT revealed by differential analysis. Solid arrowheads mark strictly CTCF-dependent boundaries (the second boundary was not called by TopDom), empty arrowheads mark partially CTCF-dependent boundaries.

Cp190 directly binds to the C-terminus of CTCF and is recruited to CTCF peaks in a strictly or partially CTCF-dependent manner (Fig. 5). Residual Cp190 binding at former CTCF peaks coincides with residual boundary retention in *CTCF⁰* mutants (Fig. 5). (6) CTCF binding to DNA alone is not sufficient for correct expression patterns of a subset of genes that also rely on Cp190. Below we discuss how this work furthers our understanding of genome folding in *Drosophila*, CTCF’s role in transcriptional regulation and the molecular basis thereof.

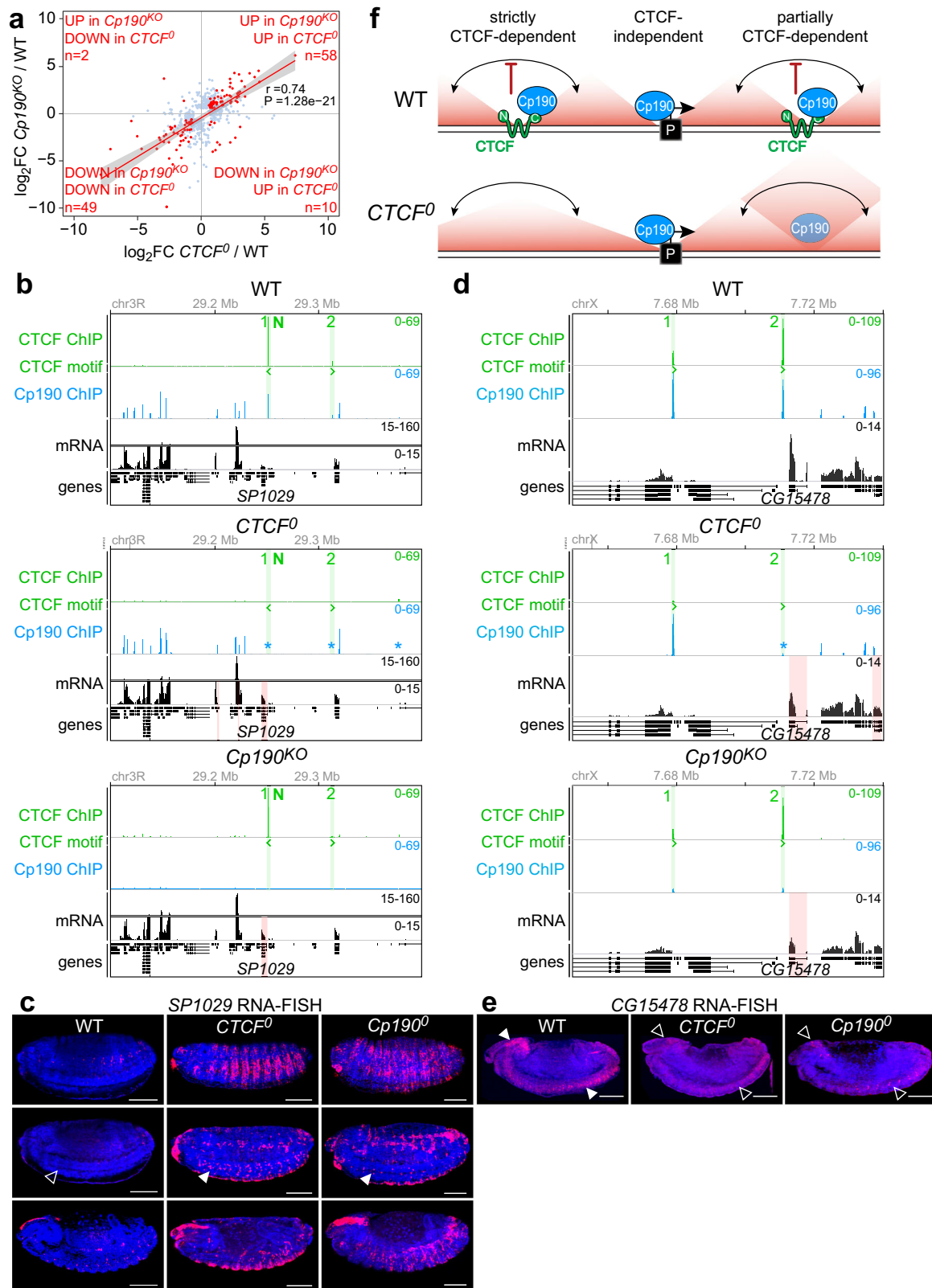
Relaxed requirement of CTCF for *Drosophila* genome architecture. In comparison to vertebrates, the principles of genome folding into CDs in *Drosophila* are less clear. On the one hand, the majority of fly CDs were proposed to form by compartmentalization of domains with different transcriptional states or because actively transcribed genes cluster, with little contribution from architectural proteins acting independently of transcription^{2,48}. On the other hand, analyses of enriched transcription factor motifs at domain boundaries defined at high-resolution revealed that 77% were enriched in core promoter motifs (and called promoter boundaries) and the remaining 23% were enriched in motifs of insulator-binding proteins like CTCF, su(Hw) and Ibf1 (and called non-promoter boundaries)¹⁷. This suggested that architectural proteins may form some domain boundaries. By completely ablating CTCF *in vivo*, we definitively show that CTCF contributes to the formation of a small fraction (below 10%) of domain boundaries in *Drosophila* (Fig. 2). This strongly contrasts with the mammalian genome where extrusion-based mechanisms are responsible for the formation of a large fraction of boundaries. This demonstrates that although domain formation is ubiquitous in different species, the contributions of different mechanisms can vary widely. The limited role that CTCF plays in global genome architecture in flies is nevertheless consistent with our finding that CTCF binding sites are an order of magnitude less frequent in flies (~ 800 peaks in 130 Mb genome) than in humans ($\sim 80,000$ peaks in 3 billion bp genome)⁴⁹, and the fact that alternative boundary-forming mechanisms exist in flies.

At strictly CTCF-dependent boundaries, CTCF can form boundaries independently of the presence/absence of a nearby TSS and of detectable transcriptional changes in nearby genes (Figs. 2c and 5d). At partially CTCF-dependent boundaries, defects in *CTCF⁰* mutants are limited by redundant boundary-forming mechanisms often associated with CTCF-independent recruitment of Cp190, Cp190-associated factors or the presence of Cp190-bound transcribed gene TSSs (Figs. 5c–g and 6f). Cp190 marks both promoter and non-promoter boundaries (Fig. 5c)^{15,17},

and it remains to be clarified whether Cp190 or its associated factors directly contribute to domain boundary formation (through similar or unrelated mechanisms as CTCF) or whether boundary formation is governed by transcription of Cp190-bound TSSs. Pervasive transcriptional perturbation globally affects Hi-C contact maps^{2,16,48}, indicating that transcription itself or the transcription machinery at least reinforces CDs. Finally, we note that apart from CTCF, the transcription factor Zelda has also been shown to affect CD boundaries in flies: Zelda depletion in early *Drosophila* embryos led to partial disruption of former Zelda-occupied domain boundaries, and to concurrent loss of RNA polymerase II recruitment which may account for the observed boundary defects¹⁶.

Whether *Drosophila* CTCF, like its mammalian counterpart, forms CD boundaries in concert with loop-extruding cohesin remains unclear because of discrepancies between flies and mammals. (1) In mammalian Hi-C maps, CTCF sites at both anchors of an extruded loop often engage in high-frequency contacts⁴ not seen in *Drosophila*² (Fig. 2c, Supplementary Fig. 2c). (2) CTCF and cohesin colocalize genome-wide in mammals^{49,50}, but cohesin does not colocalize specifically with CTCF in *Drosophila*^{13,17}. Fly CTCF may therefore not have a robust or unique ability to stall or stabilize loop-extruding cohesin complexes, despite their ability to interact *in vitro* (Fig. 2g). (3) CTCF-dependent boundaries are directional in mammals^{4,5,51} but lack clear directionality in flies (Supplementary Fig. 2g)². All these discrepancies could nevertheless be expected given the probable differences in how fly CTCF interacts with extruding cohesin (Supplementary Fig. 2f). Indeed, previous *in silico* simulations⁶ and experiments affecting loop-extrusion processivity across CTCF-dependent boundaries in human cells^{7,9,10} described CDs with weaker corner interactions more similar to domains observed in flies. The N-terminus of DNA-bound mammalian CTCF may stall or stabilize cohesin by directly interacting with cohesin subunits and regulators^{10,39,52,53} via binding interfaces that are not all conserved in fly orthologs (Supplementary Fig. 2f). Our results suggest that direct interaction of fly CTCF N-terminus with cohesin is insufficient to form directional chromosomal loops.

Impact of CTCF on transcriptional regulation. Functional studies of how CTCF impacts expression are challenging in mammalian cells. Recent studies that manipulated CTCF binding sites at specific loci have moderated our view of how critical CTCF is for patterned gene expression, but a limitation is that effects can be masked by unperturbed CTCF sites nearby that function redundantly^{31–33}.



Our transcriptional analyses of *Drosophila* *CTCF*⁰ CNSs showed that CTCF is required for patterned expression of selected genes in the CNS while at the same time being dispensable for orchestrating other complex gene expression programs. Gene misexpression may result from defective gene insulation from local regulatory elements, as supported by the

binding of CTCF between certain neuronal and non-neuronal genes in vivo (Figs. 3c, d), the increased expression of these genes in *CTCF*⁰ larval CNSs (Figs. 3c–e) and the enhancer-blocking activity of CTCF peaks in S2 cells (Fig. 4b–c). Our reporter assay is independent of chromatin environment, allowing quantitative measurements of insulator activity that reveal a direct relation to

Fig. 6 CTCF and Cp190 co-regulate a subset of target genes. **a** DE genes (with $\text{padj} < 0.05$ and $|\text{fold change}| > 1.5$) in $CTCF^0$ and/or $Cp190^{KO}$ mutant larval CNSs relative to WT with detectable expression in both differential RNA-seq analyses (omitting 55 DE genes in $CTCF^0$ and 54 DE genes in $Cp190^{KO}$ that had low counts in the other differential analysis) are plotted in light blue and red. DE genes common in $CTCF^0$ and $Cp190^{KO}$ mutants are highlighted in red and counted in each quadrant. Pearson's correlation coefficient and p-value show correlated changes of common DE genes. The red line shows linear regression and gray shadowing the corresponding 95% confidence interval. **b** Extended *SP1029* gene locus displaying CTCF ChIP-seq (peaks numbered and highlighted in green), CTCF motif orientations in DNA, Cp190 ChIP-seq, and mRNA-seq tracks (DE genes highlighted in red) in WT (top), $CTCF^0$ (middle) and $Cp190^{KO}$ CNSs (bottom). Asterisks mark Cp190 peaks in $CTCF^0$ mutants with reduced occupancy relative to WT revealed by differential analysis. **c** Lateral views of 11 h old embryos of labeled genotypes (columns) in 3 confocal sections (rows) subjected to *SP1029* RNA-FISH (scale bars 100 μm). Arrowheads mark *SP1029* misexpression in the nerve chord of $CTCF^0$ and $Cp190^0$ mutants (filled arrowheads), not occurring in WT embryos (empty arrowhead). **d** As Fig. 6b for the extended *CG15478* gene locus. Residual Cp190 ChIP signal in $Cp190^{KO}$ mutants could be maternally deposited Cp190 or non-specific ChIP signal. **e** As Fig. 6c for *CG15478* RNA-FISH. Arrowheads mark *CG15478* expression in the brain and nerve chord of WT embryos (filled arrowheads), strongly reduced in $CTCF^0$ and $Cp190^0$ mutants (empty arrowheads). **f** Wildtype *Drosophila* contact domain boundaries are strictly CTCF-dependent, partially CTCF-dependent, or not bound by CTCF. CTCF recruits Cp190 to CTCF-dependent boundaries, and Cp190 is recruited independently to additional boundaries many of which are close to transcribed gene promoters. In $CTCF^0$ mutants, Cp190 is lost from strictly CTCF-dependent boundaries, while at other former CTCF peaks residual Cp190 binding is associated with partial boundary retention. CTCF-dependent boundaries can prevent regulatory crosstalk (double-sided arrows) between genes and regulatory elements positioned on either side, and Cp190 co-regulates a subset of genes together with CTCF.

the efficiency of CTCF recruitment. These findings are consistent with our previous characterization of *Hox* gene misexpression in $CTCF^0$ mutants, which phenocopies deletions of insulator boundaries that maintain the independence of some *Hox* regulatory domains²¹. Our ability to detect gene misregulation in $CTCF^0$ larval CNSs likely depends on genomic context, notably the presence of regulatory elements active in this organ in a sufficiently large number of cells to detectably alter transcription.

Why aren't gene misexpression defects in $CTCF^0$ mutants more widespread? Recent studies have emphasized that specific communication between regulatory elements and gene promoters is controlled at many levels, of which CTCF provides one. In particular, enhancer-promoter compatibility⁵⁴ and regulation of the chromatin properties of regulatory elements themselves⁵⁵ also determine whether or not regulatory elements and promoters functionally communicate. CTCF may also function redundantly with other insulator-binding proteins in *Drosophila* to limit regulatory crosstalk in this compact genome. Unlike what is known in mammals, flies have a family of insulator-binding proteins, many of which have DNA binding domains with which they target specific loci⁵⁶.

Molecular basis of how CTCF impacts gene regulation. Whether CTCF's ability to form physical boundaries explains its conserved genetic insulator activity remains an open question^{1,57}. An ideal scenario to address this would be to separate boundary formation from gene insulator function. Human CTCF with mutated critical cohesin-interacting residues was largely functional, but CD boundaries were only partially disrupted¹⁰. We observed that some DE genes in $CTCF^0$ mutants are close to partially CTCF-dependent boundaries (Fig. 5d, lane 6). Gene misregulation in the absence of CTCF may therefore occur despite significant retention of a physical boundary, but we did not definitively confirm that these DE genes are direct CTCF targets.

We found that CTCF functionally cooperates with a stably bound regulatory cofactor, expanding the view of how CTCF may impact gene regulation. The relevance of the CTCF-Cp190 interaction has been debated. On the one hand, Cp190 was assumed to be required for CTCF's insulator function based on the observations (1) that the enhancer-blocking activity of a *Hox* gene insulator in transgenic reporter assays depended on both CTCF and Cp190, and (2) that CTCF failed to be recruited to many sites on polytene chromosomes in $Cp190$ mutants^{58,59}. The latter observation was, however, not reproduced in genome-wide ChIP experiments in Cp190 knock-down cells³⁶. On the other

hand, no common CTCF and Cp190 target genes were known⁴⁶, and the interaction between CTCF and Cp190 was recently concluded to be dispensable in vivo⁴⁵. The latter conclusion was based on deleting residues in CTCF that did not interact with Cp190 in our pull-down experiments (Supplementary Fig. 5b). We identified genes with concordant transcriptional changes upon loss of either CTCF and Cp190 that are potentially directly regulated by both proteins.

Is this interaction conserved in vertebrates? Around 40 Cp190-like proteins comprising an N-terminal BTB domain and zinc fingers exist in humans⁶⁰, but Cp190 does not have a direct ortholog. The C-terminus of human CTCF is capable of interacting with the BTB domain of a Cp190-like protein called KAISO in yeast two-hybrid experiments⁶¹, reminiscent of the interaction between fly CTCF C-terminus and the BTB domain of Cp190 (Supplementary Fig. 5b). Whether CTCF transiently interacts with a BTB domain-containing protein in human cells or whether this interaction has not been maintained in vivo remains to be clarified.

How do Cp190 and CTCF collaborate? Incomplete overlap of DE genes in $CTCF^0$ and $Cp190^{KO}$ mutants suggests that CTCF requires Cp190 at some loci but not others (Fig. 6a). Alternatively, additional common targets may be masked by other transcriptional changes in $Cp190^{KO}$ mutants or by maternal Cp190 rescuing early defects in these mutants. How Cp190 functions is not known, but it may contribute to CTCF's insulator activity similarly to how Cp190 contributes to the activities of *gypsy* and some *Hox* gene boundary insulators^{46,62}. Cp190 may help CTCF form CD boundaries, or Cp190 may function independently of boundary formation through unknown mechanisms that could uncover paradigms for controlling the communication between genes and regulatory elements.

Methods

Tissue-specific CTCF loss-of-function. ($CTCF^{KO}$, *UAS-FLP*)/*TM6B* heterozygotes were crossed to $CTCF^{KO}/TM6B$ heterozygotes for an independently isolated $CTCF^{KO}$ allele that also carried an FRT-flanked genomic *CTCF* rescue transgene and one of various Gal4 drivers: expressed in neuroblasts [*wormiu-Gal4* (Bloomington stock 56553)], mature neurons [*elav-Gal4* (Bloomington stock 25750)], or muscles [*Mef2-Gal4* (Bloomington stock 25756)]. Resulting non-*TM6B* animals were transheterozygous for $CTCF^{KO}$ alleles, derived from a WT maternal germline, and expressed *UAS-FLP* under the control of a Gal4 driver leading to tissue-specific excision of the *CTCF* rescue transgene. *w¹¹¹⁸* (wildtype) and $CTCF^{KO}$ transheterozygous animals were used as controls.

Tissue-specific rescue of $CTCF^0$ mutants. Females trans-heterozygous for two independently isolated $CTCF^{KO}$ alleles were rescued with an FRT-flanked genomic *CTCF* rescue transgene that was excised in their germline by expressing FLP

recombinase under the control of *nanos* regulatory sequences. These females were crossed to *CTCF^{KO}/TM6B* males carrying a *UAS-CTCF-3xHA* transgene (FlyORF stock F000619) and a *Gal4* driver mentioned above or no *Gal4* driver as control. Resulting non-TM6B animals were transheterozygous for *CTCF^{KO}* alleles, derived from a maternal germline devoid of *CTCF* (*CTCF⁰* mutant background) and expressed *UAS-CTCF* under the control of a *Gal4* driver. *w¹¹¹⁸* animals were used as WT control.

Drosophila viability tests. Three sets of 30–40 third instar larvae of desired genotypes were transferred into separate vials and the number of pupae and fully hatched adults was recorded. The average percentage and standard deviation of animals alive at each developmental stage and over a 30-day period after hatching were scored and plotted in Kaplan-Meier survival plots with 5% confidence intervals from the triplicate experiments.

Antibodies. For this study, polyclonal rabbit antibodies were raised against *CTCF¹⁻²⁹³* and *Cp190¹⁻¹⁰⁹⁶*. Proteins were recombinantly purified in *E. coli* by tandem affinity purification using N-terminal GFP- and C-terminal His-tags. Tags were cleaved off by 3C protease and used for immunization.

Western blotting. Forty third-instar larval CNSs per biological replicate were dissected in ice-cold PBS. Samples were sonicated in 100 μ l of 20 mM Tris pH 7.5, 500 mM NaCl, 0.1% Triton X-100, 1 \times complete protease inhibitors (Roche) in a Bioruptor (settings on high, 5 min, 4 $^{\circ}$ C). Extracts were centrifuged for 5 min at maximum speed and total protein was quantified by Qubit protein assay (ThermoFisher). Calibrated amounts of extract from WT, *CTCF⁰* and *CTCF^{OE}* animals were loaded on a 4–12% acrylamide gel and probed with rabbit anti-*CTCF¹⁻²⁹³* crude serum (diluted 1:1000) and mouse anti-tubulin clone B-5-1-2 (Sigma T5168, diluted 1:10,000). *CTCF^{OE}* animals expressed a *CTCF* cDNA under the control of upstream activating sequences (UAS) driven by a ubiquitous *tubulin-Gal4* driver, and served as control. Chemiluminescence pictures of nitrocellulose membranes were imaged in Fiji v2.1.0/1.53c.

Chromatin preparation from larval CNSs. 60 third-instar larval cuticles per biological replicate (two biological replicates per sample except *CTCF* ChIP-seq in WT performed in biological triplicates) were dissected in ice-cold PBS, then cross-linked 15 min at room temperature in 1.8% (v/v) paraformaldehyde, 50 mM HEPES pH 8, 100 mM NaCl, 1 mM EDTA, 1 mM EGTA. Crosslinking was stopped by washing for 10 min in 1 ml PBS, 0.01% Triton-X100, 125 mM glycine, then cuticles were washed for 10 min in 10 mM HEPES pH 7.6, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100. CNSs were dissected from the cuticles in 10 mM HEPES pH 7.6, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01% Triton X-100, then sonicated in 120 μ l of RIPA buffer (10 mM Tris-HCl pH 8, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, protease inhibitor cocktail) in AFA microtubes in a Covaris S220 sonicator for 5 min with a peak incident power of 140 W, a duty cycle of 5% and 200 cycles per burst. Sonicated chromatin was centrifuged to pellet insoluble material and snap-frozen.

ChIP-seq. ChIP was performed with 2 μ l of rabbit polyclonal antibody crude sera against *CTCF¹⁻²⁹³* or *Cp190¹⁻¹⁰⁹⁶*, each incubated with half of the chromatin prepared from a biological replicate overnight at 4 $^{\circ}$ C. Twenty-five microliters of pre-mixed Protein A and G Dynabeads (Thermo Fisher 100-01D and 100-03D) were added for 3 h at 4 $^{\circ}$ C, then washed for 10 min each once with RIPA, four times with RIPA with 500 mM NaCl, once in LiCl buffer (10 mM Tris-HCl pH 8, 250 mM LiCl, 1 mM EDTA, 0.5% Igepal CA-630, 0.5% sodium deoxycholate) and twice in TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA). DNA was purified by RNase digestion, proteinase K digestion, reversal of crosslinks at 65 $^{\circ}$ C for 6 h, and elution from a QIAGEN Minelute PCR purification column. ChIP-seq libraries were prepared using the NEBNext Ultra II DNA Library Prep kit for Illumina. An equimolar pool of multiplexed ChIP-seq libraries at 4 nM was sequenced on the Illumina HiSeq4000 (150 bp paired-end).

ChIP-seq analysis. Paired-end ChIP-seq reads were demultiplexed and mapped to the dm6 genome using Micmap, a derivative of the fetchGWI tool⁶³. Only chromosomes 2, 3, 4, and X were used. ChIP-seq peaks were called using the R package *csaw*⁶⁴ v1.16.1 using a window width of 20 bp and spacing of 10 bp, ignoring duplicate reads. A background enrichment was evaluated as the median over all samples in the comparison of the average number of reads per 2 kb bins. Windows with less than threefold enrichment over background were filtered out. Data were normalized using the TMM method⁶⁵ implemented in *csaw*. Differential binding analysis in *csaw* is based on the quasi-likelihood framework implemented in the *edgeR* package⁶⁶. Results obtained on different windows were combined into regions by clustering adjacent windows. Combined *p*-values were evaluated for each region using *csaw* and Benjamini & Hochberg method was applied to control the false discovery rate. Regions with false discovery rate (FDR) < 0.01 and |fold change| > 2 were considered as differential binding regions and are reported in Supplementary Data files 1, 6, 7, and 8. Genuine *CTCF* peaks were identified by differential analysis of ChIP-seq signals in WT versus *CTCF⁰* as being lower in the

mutant samples relative to WT. Genuine *Cp190* peaks were similarly identified by differential analysis of ChIP-seq signals in WT versus *Cp190^{KO}* (*Cp190* peaks in WT) or in *CTCF⁰* versus *Cp190^{KO}* (*Cp190* peaks in *CTCF⁰*). Additional differential analyses were performed for *Cp190* ChIP-seq signal in WT versus *CTCF⁰* (for Fig. 5a). We defined ChIP occupancy as the best.log2FC obtained from *csaw* in the respective differential analysis. We defined peak positions as the best.pso obtained from *csaw*. To count overlaps between *CTCF* and *Cp190* peaks in three-way comparisons shown in Fig. 5a, some *CTCF* and *Cp190* peaks were split into 2 or 3 sub-regions. Specifically, 740 WT *CTCF* peaks were split into 765 peaks, 6473 WT *Cp190* peaks were split into 6474 peaks, and 1045 differentially bound *Cp190* regions with lower occupancy in *CTCF⁰* relative to WT were split into 1076 peaks. Accompanying the *CTCF* ChIP-seq, matches to the *Drosophila* *CTCF* motif MA0531.1 downloaded from the JASPAR website were indicated in all figures.

Hi-C library preparation. 60 third-instar larval CNSs (~600,000 cells) per biological replicate were dissected in ice-cold PBS. CNSs or a single whole-bodied female fly were crushed in RPMI supplemented with 10% fetal bovine serum using a micro-pestle. Cells were fixed in 1% (v/v) paraformaldehyde for 10 min at room temperature. The Hi-C libraries were prepared using *MboI* and *MseI* as restriction enzymes. Restricted ends were marked with biotin, then ligated. Fragmented DNA was enriched for pairwise DNA junctions by biotin pull-down using Dynabeads MyOne Streptavidin T1 beads following the manufacturer's instructions. Illumina sequencing libraries were prepared with standard protocols. 4 nM equimolar pools of multiplexed Hi-C libraries were subjected to paired-end sequencing on Illumina HiSeqX Ten and HiSeq4000 instruments.

Hi-C data processing. We pre-computed a table containing the positions of all restriction sites used for Hi-C present in the dm6 genome. The FASTQ read pairs were analyzed with a Perl script available for download in the Micmap⁶³ package (see Code Availability) to locate and separate fusion sites using the patterns /GATCGATC/, /TTATAA/, /GATCTAA/ and /TTAGATC/. The maximal length of each read was trimmed at 60 nucleotides, then reads were mapped to the dm6 genome using Micmap and matched to their closest pre-computed genomic restriction site. Read pairs were discarded if they (1) mapped to non-unique positions in the reference genome, (2) had indels or >2 mismatches per read, (3) represented fusion of 2 oppositely oriented reads within 2 kb of each other, which may have not resulted from ligation of 2 digested fragments (these fragments were used to estimate local copy number status of the underpinning genomic region), (4) were likely additional copies of a given read pair, i.e., likely PCR duplicates. Only chromosomes 2, 3, 4, and X were considered.

To assess the correlation of biological replicates, samples were downsampled to 45 million contacts per replicate. Raw Hi-C contact matrices were created by binning Hi-C pairs at 10 kb resolution. These matrices were then normalized with the ICE normalization implemented in *iced* v0.5.2⁶⁷. Low coverage regions (bins with no contacts and those with the 2% smallest total number of contacts among bins) were filtered out. Pearson correlation coefficients were determined for every pair of normalized matrices by flattening each matrix and evaluating the Pearson correlation coefficient for the resulting vector, using only pairs of bins at a genomic distance below 1 Mb. The limitation on the distance was introduced to compare contacts at a scale relevant to the analyses performed in this manuscript which were at the level of CDs. Resulting Pearson correlation coefficients were ≥ 0.949 for all replicates, showing that they were well correlated and that WT and *CTCF⁰* Hi-C matrices were globally similar. For the analyses presented in the main figures, pooled replicates of the same genotype were downsampled to 200 million contacts per genotype. Raw Hi-C contact matrices obtained by binning Hi-C pairs at 2 kb resolution were then normalized with the ICE normalization implemented in *iced* v0.5.2⁶⁷. Low coverage regions (bins with no contacts and those with the 2% smallest total number of contacts among bins) were filtered out before normalization (these regions are marked by gray lines in Hi-C maps shown in the figures).

For each normalized Hi-C contact matrix, CD boundaries were called using TopDom⁶⁸. Given a window size *w*, a physical insulation score was defined for each bin *i* as:

$$\log_2 \frac{\text{binSignal}_i}{\sum_{i-w/2 < j < i+w/2} \text{binSignal}_j} \quad (1)$$

where *binSignal_i* is the average normalized Hi-C contact frequency between *w* bins upstream of bin *i* and *w* bins downstream of bin *i* determined by TopDom. The strength of a boundary at bin *i* was thus estimated as the log₂ of the *binSignal* value at bin *i* normalized by its local average on a window of size *w*. With this definition, lower insulation scores indicate stronger boundaries. We extracted CD boundaries and physical insulation scores for Hi-C matrices at 2 kb resolution using window sizes 20, 40, 80, and 160 kb. CD boundaries found with all window sizes were merged, and the average insulation score obtained with all window sizes was retained. To facilitate comparisons of CD boundaries found in WT and *CTCF⁰* genotypes and avoid mismatches due to small fluctuations of CD boundary positions obtained with different window sizes or genotypes, groups of consecutive boundaries (i.e., within 2 kb of each other) were merged. Groups of consecutive boundaries were replaced by the boundary with the lowest insulation score (average of both genotypes for boundaries common to WT and *CTCF⁰*).

Hi-C maps were visualized in R and Juicebox⁶⁹ (see Supplementary Table 3 for links to interactive maps for browsing).

A/B compartment calling. A/B compartment calling was performed following the method proposed in Lieberman Aiden et al.⁷⁰ Each individual chromosome arm (chr2L, chr2R, chr3L, chr3R, chr4, chrX) was analyzed separately. Normalized Hi-C contact matrices at 2 kb resolution were considered after discarding invalid bins (low coverage regions) and bins around centromeres (chosen for exclusion as dm6 coordinates >22,170,000 for chr2L, <5,650,000 for chr2R, >22,900,000 for chr3L, <4,200,000 for chr3R). Observed-over-expected matrices were generated by dividing the normalized Hi-C contact matrices by the average number of normalized Hi-C contacts at the corresponding genomic distance. For each chromosome arm, the first eigenvector of the correlation matrix was obtained by principal component analysis of the observed-over-expected matrix. Each eigenvector was then centered around zero by subtracting its mean value, then multiplied by the sign of the Pearson correlation between the eigenvector and the number of expressed gene TSSs per 2 kb bin. 2 kb bins with positive eigenvector values were assigned to compartment A, those with negative eigenvector values were assigned to compartment B. chr4 eigenvectors appeared to reflect a large-scale structure that separated the chromosome into two halves, and were thus excluded from Supplementary Fig. 3d.

Comparison with CD boundaries from other Hi-C studies. To assess whether CD boundaries called in our study could correspond to small CDs resolved in higher resolution Hi-C contact maps (analyzed at 500 bp resolution instead of 2 kb used here), we compared our CD boundary calls to CD coordinates published by Eagen et al.¹⁴, and Ramirez et al.¹⁷ (converted from dm3 to dm6 genome coordinates using the liftOver tool <http://genome.ucsc.edu/cgi-bin/hgLiftOver>) in Kc167 tissue culture cells. We counted how many small (≤ 4 kb) CDs identified in those published studies were close (within 2 kb) to one of our CD boundaries. We could have potentially mis-called such small domains as a domain boundary. The result is that Eagen et al. did not report CDs smaller than 6 kb. Only 31 of our domain boundaries were within 2 kb of a ≤ 4 kb CD identified by Ramirez et al. Thus, very few (31/3970, or <1%) of our domain boundaries may correspond to a small domain defined by Ramirez et al. We next asked: How many domain boundaries that disappear in *CTCF*⁰ mutants could correspond to small domains? The result is that very few (4/567, or <1%) of our domain boundaries identified only in WT were within 2 kb of a ≤ 4 kb CD identified by Ramirez et al. Domain boundaries identified by Ramirez et al. are displayed together with domain boundaries identified in this study in all Hi-C screenshots throughout the manuscript for comparison.

RNA-seq on larval CNSs. WT, *CTCF*⁰ and *Cp190*^{KO} mutant third instar larval brains were dissected in ice-cold PBS. For RNA isolation, triplicates of 60 larval brains each were homogenized in TRIzol LS (ThermoFisher) with pestles (VWR) on ice. RNA was extracted following the manufacturer's instructions, remaining DNA digested with DNase I (Roche), and RNA was purified using RNAClean XP beads (Beckman Coulter). Strand-specific mRNA-seq libraries were prepared from 1 μ g of total RNA after mRNA selection with NEBNext Oligo d(T)25 beads, using the NEBNext Ultra directional RNA library prep kit for Illumina following the manufacturer's instructions. Multiplexed libraries were sequenced on one lane of a HiSeq2500 (100 bp paired-end for *CTCF*⁰ and WT control) or a HiSeq4000 (150 bp single-end for *Cp190*^{KO} and WT control).

Differential RNA-seq analysis. RNA-seq reads were mapped both to the dm6 *Drosophila melanogaster* reference genome and to Flybase gene models and transcripts (dmel-all-r6.26.gtf.gz) using Micmap⁶³. The results of both mappings were combined into spliced alignments in BAM file format. Then, htseq-count (v0.9.1) was used to produce read counts per gene⁷¹. Statistical analysis was performed in R (v3.5.1). Genes with <1 count per million in at least three replicate samples were filtered out using EdgeR (v3.22.5)⁶⁶. Normalization and differential expression analysis were performed in DESeq2 (v1.22.1)⁷² individually for both WT versus *CTCF*⁰ and WT versus *Cp190*^{KO} samples. Statistical significance was tested by Wald test and the Benjamini-Hochberg method was used for multiple testing adjustment. A significance threshold of [fold change] > 1.5 and *p*-adjusted < 0.05 was used to identify DE genes. The R package ggplot2 (v3.2.1) was used for data visualization.

RNA-FISH. Labeled RNA probes were generated by in vitro transcription with Dig-UTP labeling mix (Roche 11277073910) and T7 RNA polymerase (Roche 10881767001) antisense to full-length complementary DNA clones of *SP1029* (FI20034) and *IFT52* (MIP14443), genomic DNA amplified from dm6 coordinates chr3L: 10263888–10266244, or cDNAs amplified using gene-specific primers from a cDNA library prepared from *Drosophila* embryos (see Supplementary Data 10 for primer sequences). After DNase I digestion for 20 min at 37 °C, probes were fragmented by incubating 20 min at 65 °C in 60 mM Na₂CO₃, 40 mM NaHCO₃ pH 10.2, precipitated in 300 mM sodium acetate pH 5.2, 1.25 M LiCl, 50 mg/ml tRNA and 80% EtOH, resuspended in 50% formamide, 75 mM sodium citrate pH 5, 750 mM NaCl, 100 μ g/ml salmon sperm DNA, 50 μ g/ml heparin and 0.1% Tween20, and stored at –20 °C. Embryos or third instar larval cuticles were fixed in

4% paraformaldehyde for 30 min at room temperature, washed, and then stored in 100% MeOH at –20 °C for at least overnight. Samples were rehydrated in PBS with 0.1% Tween20, post-fixed in 4% paraformaldehyde for 20 min at room temperature, progressively equilibrated to hybridization buffer (50% formamide, 75 mM sodium citrate pH 5, 750 mM NaCl) and heated to 65 °C. RNA probes were diluted 1:50 in hybridization buffer, denatured at 80 °C for 10 min then placed on ice, and added to the samples overnight shaking at 65 °C. Samples were washed 6 times 10 min in hybridization buffer at 65 °C, then progressively equilibrated to PBS with 0.1% Triton X-100. Samples were incubated overnight at 4 °C in anti-dig peroxidase (Roche 11207733910) diluted 1:2000 in PBS, 0.1% Triton X-100, 1 \times Western blocking reagent (Sigma 1921673). Samples were washed six times 10 min in PBS with 0.1% Tween20, labeled with Cyanine 3 tyramide in the TSA Plus kit (Perkin Elmer NEL753001KT) for 3 min at room temperature, washed 6 times 10 min in PBS with 0.1% Tween20, and finally mounted with DAPI to stain DNA. Images were acquired on a Zeiss LSM 880 microscope with a $\times 20$ objective and visualized with Fiji software v2.1.0/1.53c.

Insulator reporter. An insulator reporter (Fig. 4a) was designed with an enhancer (*OpIE2*) equidistant from EGFP and mCherry fluorescent reporters with basal *Hsp70* promoters. A *gypsy* insulator is present in the reporter plasmid, downstream of the EGFP transcription unit. Selected CTCF binding sites (Supplementary Fig. 4c) were PCR-amplified from genomic DNA and cloned in between the enhancer and EGFP. Control reporters had a neutral spacer (a fragment of the bacterial *Kanamycin* resistance gene) or the *gypsy* insulator in between the enhancer and EGFP. In addition, one CTCF binding site (fragment N) was mutagenized by PCR to mutate 2 bp in a CTCF motif (ATGTCAGAGGGCGCT converted to ATGTCAGACAGCGCT). All plasmids were transfected in parallel into S2 cells (originally purchased from ATCC, reference number CRL-1963) in triplicates in a 96-well plate using 100 ng of reporter plasmid and Effectene (QIAGEN) following the manufacturer's instructions. After 48 h, fluorescence was measured on a NovoCyte Flow Cytometer (ACEA) using FITC and PE-TexasRed detection settings. Recordings were gated to discard measurements of untransfected cells (Supplementary Fig. 4a). Distributions of mCherry/EGFP fluorescence ratios in thousands of single transfected cells were plotted and the median mCherry/EGFP ratio was extracted for each experiment. The average of these median values obtained for each replicate is plotted in Fig. 4c as a function of the total CTCF ChIP-seq read counts in S2 cells on the cloned fragment tested in the insulator reporter—extracted using bedtools multicov⁷³ applied to CTCF ChIP-seq data in S2 cells⁴² (GEO accession GSM1015410).

Recombinant protein pull-downs

Purification of N-terminal CTCF constructs. The sequence encoding WT or Y248A F250A mutant versions of the dmCTCF N-terminus (residues 1–293) were cloned into a pET-based vector with an N-terminal GFP-tag and a C-terminal His₆ tag. The constructs were transformed into an *E. coli* expression strain (Rosetta), and 1 liter cultures were grown in TB-medium to an OD(600) of 1.0 at 37 °C. The culture temperature was then reduced to 18 °C and IPTG was added to a final concentration of 0.5 mM. Cells were harvested after overnight incubation at 18 °C by centrifugation, and the cell pellet was resuspended in 2 volumes of Lysis Buffer (50 mM Tris pH 7.5, 300 mM NaCl, 5 % glycerol, 25 mM Imidazole). Cells were opened by sonication, and the lysate was clarified by centrifugation at 50,000 \times g at 4 °C. The supernatant was loaded onto a 5 ml HisTrap column (GE Healthcare), washed extensively with Lysis Buffer, and the bound material was eluted with Lysis Buffer supplemented with 400 mM Imidazole. The eluate was then diluted 10-fold with buffer QA (20 mM Tris pH 7.5, 100 mM NaCl, 5% glycerol), and the resulting solution was loaded onto a 5 ml HiTrap-Q column (GE Healthcare). After washing the column with 5 column volumes (cV) of QA buffer, the bound material was eluted with a 5 cV gradient from QA to QB (20 mM Tris pH 7.5, 1000 mM NaCl, 5% glycerol). Fractions containing the CTCF protein at sufficient purity were identified by SDS-PAGE followed by Coomassie staining. Proteins aliquots were snap-frozen in liquid nitrogen and stored at –80 °C.

Purification of SA-Vtd complex. The sequences encoding dmSA (residues 102–1085) and Vtd (Rad21) (residues 273–458) were cloned into a pET-based vector with an N-terminal His₁₀-TwinStrep-3C tag on SA. The complex was expressed in 1 liter of *E. coli* (Rosetta) grown in TB. Growth, induction of expression, and cell harvesting and lysis were carried out as described for CTCF constructs. Clarified lysates were loaded onto a 5 ml StrepTrap column (GE Healthcare), washed with 5 cV of Lysis buffer, and bound material was eluted with 8 cV of elution buffer (20 mM Tris pH 7.5, 100 mM NaCl, 5 % glycerol, 2.5 mM des-thiobiotin). The eluate was loaded on a 5 ml HiTrap-Q column (GE Healthcare), and after washing the column with 5 column volumes (cV) of QA buffer, the bound material was eluted with a 5 cV gradient from QA to QB (20 mM Tris pH 7.5, 1000 mM NaCl, 5% glycerol). Fractions containing the purified SA-Vtd complex were identified by SDS-PAGE and Coomassie staining, pooled, aliquoted, snap-frozen in liquid nitrogen, and stored at –80 °C.

Pull-downs between CTCF and SA-Vtd. Proteins were diluted to a final concentration of 2.5 μ M in 500 μ l of binding buffer (20 mM Tris pH 7.5, 150 mM

potassium acetate, 10 % glycerol) and allowed to bind to each other at 4 °C for 2 h. Twenty microliters of this solution was removed as 'input' sample and boiled in SDS-PAGE loading buffer. GFP-binder beads (Agarose beads covalently bound to GFP-nanobody; 20 µl per reaction) were washed in binding buffer and added to the binding reactions for 30 min at 4 °C on a rotating wheel to bind to the GFP-tagged CTCF construct. Beads were harvested by centrifugation (1 min, 700 × g) and washed twice with 1 ml of binding buffer. The final immobilized material was eluted by boiling in 50 µl of SDS-PAGE loading buffer. Inputs and pull-downs were loaded onto a 12% SDS-PAGE gel, and the proteins were visualized by staining with Coomassie.

Pull-downs between C-terminal CTCF constructs and Cp190 BTB domain. Expression plasmids encoding GFP-His-tagged constructs of the C-terminal domain of CTCF (all with Ampicillin resistance) were co-transformed with an expression plasmid carrying a His-tagged Cp190 BTB-domain (with Kanamycin resistance) into the *E.coli* Rosetta strain. Colonies were inoculated in 10 ml TB cultures and grown at 37 °C to an OD(600) of 1. The culture temperature was then reduced to 18 °C, and 0.5 mM IPTG was added to induce protein expression. Cells were harvested after overnight incubation at 18 °C, and the pellets were resuspended in 2 volumes of lysis buffer (50 mM Tris pH 7.5, 200 mM NaCl, 5% glycerol, 25 mM Imidazole). Cells were lysed by sonication and the lysate was clarified by centrifugation at 16000 g for 10 min at 4 °C. The lysates were split into two halves, which were incubated for 1 h at 4 °C with either 20 µl of GFP-binder resin or 20 µl of Ni(2+)-NTA resin, to pull down only CTCF-constructs or both CTCF and Cp190-BTB, respectively. The beads were then washed three times with 1 ml of Lysis buffer to remove non-specifically bound proteins. The bound material was eluted either by boiling in SDS-loading buffer (for GFP pull-downs) or by incubation with Lysis buffer supplemented with 500 mM Imidazole (for Ni(2+)-NTA pull-downs), and analysed by SDS-PAGE followed by Coomassie staining.

Co-purification of CTCF interactors from embryo nuclear extracts. Soluble nuclear protein extracts were prepared from WT (OregonR) 0–14 h embryos. Thirty grams of embryos were dechorionated, taken up in 30 ml of NU1 buffer (15 mM HEPES pH 7.6, 10 mM KCl, 5 mM MgCl₂, 0.1 mM EDTA pH 8, 0.5 mM EGTA pH 8, 350 mM sucrose, 2 mM DTT, 0.2 mM PMSF), and dounce-homogenized. The lysate was filtered through a double layer of miracloth, then centrifuged 15 min at 9000 rpm at 4 °C. The nuclei pellet was resuspended and lysed in 30 ml of high-salt buffer (15 mM HEPES pH 7.9, 400 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 20% glycerol, 1 mM DTT, protease inhibitor cocktail) rotating for 20 min at 4 °C, and ultracentrifuged 1 h with a SW40 rotor at 38000 rpm at 4 °C. The lipid layer was removed by suction and the soluble nuclear extract was dialyzed into 15 mM HEPES pH 7.9, 200 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA pH 7.9, 20% glycerol, 1 mM DTT with a 6–8 kDa molecular weight cut-off membrane. Soluble nuclear extract was snap-frozen in liquid nitrogen, and stored at –80 °C. *Drosophila* CTCF^{1–293} fused to an N-terminal GFP-3C tag and a 3C-His₆ C-terminal tag was purified from bacterial lysates by Ni-NTA affinity then ion-exchange chromatography as described above. Purified GFP-3C-CTCF^{1–293}-3C-His₆ was immobilized on GFP binder beads, of which 30 µl bead volume were then incubated with 6 mg of *Drosophila* embryo nuclear extract in a total volume of 10 ml of IP buffer (50 mM Tris-Cl pH 7.5, 150 mM potassium acetate, 2 mM MgCl₂, 10% glycerol, 0.1 mM DTT, 0.2% Igepal, 1× complete protease inhibitor cocktail) rotating for 3 h at 4 °C. Beads were washed three times with IP buffer, rotating for 10 min at 4 °C for each wash. Proteins were eluted with 3 C protease, adjusted to 1× SDS-loading buffer and loaded on an SDS-PAGE gel. A duplicate experiment was similarly performed with nuclear protein extracts prepared from another biological replicate embryo sample. Peptides covering the entire CTCF full-length protein were recovered, indicating that pull-downs with CTCF N-terminus recovered interactors of full-length CTCF.

Mass spectrometry analysis. Protein samples were separated by SDS-PAGE and stained by Coomassie. Gel lanes between 15–300 kDa were excised into five pieces and digested with sequencing-grade trypsin. Extracted tryptic peptides were dried and resuspended in 0.05% trifluoroacetic acid, 2% (v/v) acetonitrile. Tryptic peptide mixtures were injected on a Dionex RSLC 3000 nanoHPLC system (Dionex, Sunnyvale, CA, USA) interfaced via a nanospray source to a high-resolution mass spectrometer LTQ-Orbitrap Velos Pro. Peptides were loaded onto a trapping microcolumn Acclaim PepMap100 C18 (20 mm × 100 µm ID, 5 µm, Dionex) before separation on a C18 reversed-phase custom-packed column using a gradient from 4 to 76% acetonitrile in 0.1 % formic acid. In data-dependent acquisition controlled by Xcalibur software (Thermo Fisher), the 10 most intense multiply charged precursor ions detected with a full MS survey scan in the Orbitrap were selected for collision-induced dissociation (CID, normalized collision energy NCE = 35%) and analysis in the ion trap. The window for precursor isolation was of 4.0 *m/z* units around the precursor and selected fragments were excluded for 60 s from further analysis. Data files were analyzed with MaxQuant 1.6.3.4 incorporating the Andromeda search engine^{74,75} for protein identification and quantification based on IBAQ intensities⁷⁶. The following variable modifications were specified: cysteine carbamidomethylation (fixed) and methionine oxidation and protein N-terminal acetylation (variable). The sequence databases used for searching were *Drosophila*

melanogaster and *Escherichia coli* reference proteomes based on the UniProt database (www.uniprot.org, versions of 31 January 2019, containing 21,939 and 4915 sequences respectively), and a contaminant database containing the most usual environmental contaminants and the enzymes used for digestion (keratins, trypsin, etc). Mass tolerance was 4.5 ppm on precursors (after recalibration) and 0.5 Da on CID fragments. Both peptide and protein identifications were filtered at 1% FDR relative to hits against a decoy database built by reversing protein sequences. The MaxQuant output table proteinGroups.txt was processed with Perseus software⁷⁷ to remove proteins matched to the contaminants database as well as proteins identified only by modified peptides or reverse database hits. Next, the table was filtered to retain only proteins identified by a minimum of two peptides, the IBAQ quantitative values were log-2 transformed and missing values imputed with a constant value of 9.

Generation of Cp190^{KO} animals. We cloned ~1.5 kb homology arms (dm6 coordinates chr3R:152761111–15274519 and chr3R:15271056–15269404) into the pHD-DsRed-attP vector⁷⁸. Guide RNAs close to the start and stop codons of the Cp190 open reading frame were cloned into pCFD3 vector⁷⁹. Plasmids were co-injected into *nanos-Cas9* embryos⁷⁹. Experiments were performed in animals transheterozygous for two independent knockout alleles.

Generation of Cp190⁰ animals. Cp190^{KO} mutants were rescued into viable and fertile adults with an FRT-flanked 7 kb Cp190 genomic rescue transgene (dm6 coordinates chr3R:15269425–15276409) amplified by PCR. The Cp190 rescue cassette was excised from male and female germlines through *nanos-Gal4:VP16* (NGVP16)-driven expression of UAS-FLP. Cp190⁰ animals were collected from crosses between such males and females.

Statistics and reproducibility. All described replicate experiments are biological (not technical) replicates. For all box plots: center line, median; box limits, upper and lower quartiles; upper whisker extends to the largest value no further than 1.5× interquartile range from the upper hinge; lower whisker extends to the smallest value no further than 1.5× interquartile range from the lower hinge; points, outliers. Figure 2g: This experiment was repeated twice from independently grown bacterial cultures, with similar results. Figure 3e and Supplementary Fig. 1a–b: *n* = 10 independent third instar larvae per genotype were examined over two independent experiments each. All animals showed similar expression patterns for a given gene, that was characteristic of each genotype. RNA-FISH probes for additional genes were tested on larval nervous systems but discarded because they showed an inconsistent pattern (variable, asymmetric signal in the optic lobes in all genotypes) that we concluded was non-specific background. Figure 6c, e: *n* = 50 independent embryos per genotype were examined over two independent RNA-FISH experiments each. All animals showed similar expression patterns for a given gene, that was characteristic of each genotype. Supplementary Fig. 2a: The experiment was repeated twice with independently prepared extracts, with similar results. Supplementary Fig. 5b: The pull-down experiments were repeated twice from independently grown bacterial cultures, with similar results.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All sequencing data (Hi-C, ChIP-seq, RNA-seq) that support the findings of this study were deposited in Gene Expression Omnibus with accession code GSE146752. Hi-C maps are browsable on Juicebox (links in Supplementary Table 3). Mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD019487. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request. Additional information is provided in Supplementary Data files 1–10 and a reporting summary for this Article is available as a Supplementary Information file. Source data are provided with this paper.

Code availability

All software used as described in the Methods to map, visualize and analyze data is published open source and freely available for download in the following links: “Micmap v2.20200223 [<https://github.com/sib-swiss/micmap>]”; “DESeq2 v1.22.2 [<https://bioconductor.org/packages/release/bioc/html/DESeq2.html>]”; “HTSeq v0.9.1 [<https://github.com/anders-anders/htseq>]”; “iced v0.5.2 [<https://github.com/hiclib/iced>]”; “TopDom v0.0.2 [<https://github.com/jasminzhoulab/TopDom>]”; “R v3.5.1 [<https://www.R-project.org/>]” with packages “csaw v1.16.1 [<https://bioconductor.org/packages/release/bioc/html/csaw.html>]”, “edgeR v3.22.5 [<https://bioconductor.org/packages/release/bioc/html/edgeR.html>]”, “Eulerr v6.0.0 [<https://cran.r-project.org/package=eulerr>]” and “ggplot2 v3.1.0 [<https://ggplot2.tidyverse.org/>]”; “bedtools multicov v2.29.2 [<https://bedtools.readthedocs.io/en/latest/>]”; “Juicebox v1.5.1 [[aidenlab.org/juicebox](https://github.com/gambettalab/juicebox)]”. Custom scripts are provided in “link [<https://github.com/gambettalab/kaushal2020/>]”.

Received: 18 May 2020; Accepted: 21 January 2021;

Published online: 12 February 2021

References

- Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–309. e24 (2017).
- Rowley, M. J. et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* **67**, 837–852.e7 (2017).
- Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 1–800 (2018).
- Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
- Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).
- Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
- Haarhuis, J. H. I. et al. The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693–707.e14 (2017).
- Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
- Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
- Li, Y. et al. The structural basis for cohesin–CTCF-anchored loops. *Nature* **578**, 1–9 (2020).
- Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
- Chathoth, K. T. & Zabet, N. R. Chromatin architecture reorganization during neuronal cell differentiation in *Drosophila* genome. *Genome Res* **29**, 613–625 (2019).
- Cubeñas-Potts, C. et al. Different enhancer classes in *Drosophila* bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* **45**, 1714–1730 (2017).
- 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).
- Wang, Q., Sun, Q., Czajkowsky, D. M. & Shao, Z. Sub-kb Hi-C in *D. melanogaster* reveals conserved characteristics of TADs between insect and mammalian cells. *Nat. Commun.* **9**, 188 (2018).
- 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.e19 (2017).
- Ramírez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
- Ulianov, S. V. et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* **26**, 70–84 (2016).
- Moore, J. M. et al. Loss of maternal CTCF is associated with peri-implantation lethality of Ctf null embryos. *PLoS ONE* **7**, e34915 (2012).
- Soshnikova, N., Montavon, T., Leleu, M., Galjart, N. & Duboule, D. Functional analysis of CTCF during mammalian limb development. *Dev. Cell* **19**, 819–830 (2010).
- Gambetta, M. C. & Furlong, E. E. M. The insulator protein CTCF is required for correct Hox gene expression, but not for embryonic development in *Drosophila*. *Genetics* **210**, 129–136 (2018).
- Arzate-Mejía, R. G., Cerecedo-Castillo, A. J., Guerrero, G., Furlan-Magaril, M. & Recillas-Targa, F. In situ dissection of domain boundaries affect genome topology and gene transcription in *Drosophila*. *Nat. Commun.* **11**, 894 (2020).
- Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).
- Yokoshi, M., Segawa, K. & Fukaya, T. Visualizing the role of boundary elements in enhancer-promoter communication. *Mol. Cell* **78**, 224–235.e5 (2020).
- Andrey, G. et al. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**, 1234167–1234167 (2013).
- Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
- Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
- Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
- Narendra, V. et al. Transcription. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021 (2015).
- Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E. O. & Reinberg, D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.* **30**, 2657–2662 (2016).
- 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).
- Paliou, C. et al. Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proc. Natl Acad. Sci. USA* **116**, 12390–12399 (2019).
- Rodríguez-Carballo, E. et al. The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–2281 (2017).
- Bartkuhn, M. et al. Active promoters and insulators are marked by the centrosomal protein 190. *EMBO J.* **28**, 877–888 (2009).
- Bortle, K. V. et al. *Drosophila* CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res.* **22**, 2176–2187 (2012).
- Schwartz, Y. B. et al. Nature and function of insulator protein binding sites in the *Drosophila* genome. *Genome Res.* **22**, 2188–2198 (2012).
- Tomancak, P. et al. Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **8**, R145 (2007).
- Brown, J. B. et al. Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**, 393–399 (2014).
- Nora, E. P. et al. Molecular basis of CTCF binding polarity in genome folding. *Nat. Commun.* **11**, 5612 (2020).
- Kyrchanova, O. et al. The insulator functions of the *Drosophila* polyductyl C2H2 zinc finger protein CTCF: necessity versus sufficiency. *Sci. Adv.* **6**, eaz3152 (2020).
- Geyer, P. K. & Corces, V. G. DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. *Genes Dev.* **6**, 1865–1873 (1992).
- Ong, C.-T., Van Bortle, K., Ramos, E. & Corces, V. G. Poly(ADP-ribosylation) regulates insulator function and intrachromosomal interactions in *Drosophila*. *Cell* **155**, 148–159 (2013).
- Cuartero, S., Fresán, U., Reina, O., Planet, E. & Espinàs, M. L. Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. *EMBO J.* **33**, 637–647 (2014).
- Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6**, 2848 (2017).
- Bonchuk, A. et al. Functional role of dimerization and CP190 interacting domains of CTCF protein in *Drosophila melanogaster*. *BMC Biol.* **13**, 63 (2015).
- Savitsky, M., Kim, M., Kravchuk, O. & Schwartz, Y. B. Distinct roles of chromatin insulator proteins in control of the *Drosophila* Bithorax complex. *Genetics* **115**, 179309 (2016).
- Nègre, N. et al. A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet.* **6**, e1000814 (2010).
- Rowley, M. J. et al. Condensin II counteracts cohesin and RNA polymerase II in the establishment of 3D chromatin organization. *Cell Rep.* **26**, 2890–2903.e3 (2019).
- Pugacheva, E. M. et al. CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc. Natl Acad. Sci. USA* **201911708** (2020).
- Wendt, K. S. et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
- Tang, Z. et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
- Wutz, G. et al. ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin-STAG1 from WAPL. *eLife* **9**, e52091 (2020).
- Hansen, A. S. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus* **11**, 132–148 (2020).
- Haberle, V. et al. Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* **570**, 801 (2019).
- Kraft, K. et al. Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat. Cell Biol.* **21**, 305–310 (2019).
- Özdemir, I. & Gambetta, M. C. The role of insulation in patterning gene expression. *Genes* **10**, 767 (2019).
- Yatskevich, S., Rhodes, J. & Nasmyth, K. Organization of chromosomal DNA by SMC complexes. *Annu Rev. Genet.* **53**, 445–482 (2019).
- Gerasimova, T. I., Lei, E. P., Bushey, A. M. & Corces, V. G. Coordinated control of dCTCF and gypsy chromatin insulators in *Drosophila*. *Mol. Cell* **28**, 761–772 (2007).
- Mohan, M. et al. The *Drosophila* insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *EMBO J.* **26**, 4203–4214 (2007).

60. Letunic, I., Doerks, T. & Bork, P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* **43**, D257–D260 (2015).
61. Defossez, P.-A. et al. The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *J. Biol. Chem.* **280**, 43017–43023 (2005).
62. Pai, C.-Y., Lei, E. P., Ghosh, D. & Corces, V. G. The centrosomal protein CP190 is a component of the gypsy chromatin insulator. *Mol. Cell* **16**, 737–748 (2004).
63. Iseli, C., Ambrosini, G., Bucher, P. & Jongeneel, C. V. Indexing strategies for rapid searches of short words in genome sequences. *PLoS ONE* **2**, e579 (2007).
64. Lun, A. T. L. & Smyth, G. K. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* **44**, e45 (2015).
65. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
66. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinform. Oxf. Engl.* **26**, 139–140 (2009).
67. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
68. Shin, H. et al. TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res.* **44**, e70 (2015).
69. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
70. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
71. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
72. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550–521 (2014).
73. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
74. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
75. Cox, J. et al. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
76. Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
77. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
78. Gratz, S. J. et al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* **196**, 961–971 (2014).
79. Port, F., Chen, H.-M., Lee, T. & Bullock, S. L. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc. Natl Acad. Sci. USA* **111**, E2967–E2976 (2014).

Acknowledgements

We thank Winship Herr, Richard Benton, Jean-Yves Roignant and Naoko Mizuno for critical comments on the manuscript. We thank Patrice Waridel and Manfredo Quadroni

for mass spectrometry analyses. We thank René Dreos for advice on statistical analyses. MCG thanks Eileen Furlong for support during early phases preceding this work. Deep sequencing was performed at the Genomic Technologies Facility (GTF), mass spectrometry was performed at the Protein Analysis Facility (PAF) and imaging was performed at the Cellular Imaging Facility (CIF) at the Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, Switzerland. This work was supported by the Swiss National Science Foundation (SNSF #184715 to M.C.G.) and the University of Lausanne.

Author contributions

M.C.G. conceived the project and designed experiments. E.L.A. and A.O. conceived and designed Hi-C experiments. A.K., G.M., I.O., A.O., M.T., P.C., A.S., and M.C.G. performed the experiments. J.D., P.C., A.S., O.D., F.M., C.I., Y.E., D.W., M.S.S., and N.G. analyzed data. Y.E. created links interactive browsing of Hi-C and ChIP-seq data in Juicebox. M.C.G. prepared the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21366-2>.

Correspondence and requests for materials should be addressed to E.L.A. or M.C.G.

Peer review information *Nature Communications* thanks Elphège Nora, Sergey Razin, Félix Recillas-Targa and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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

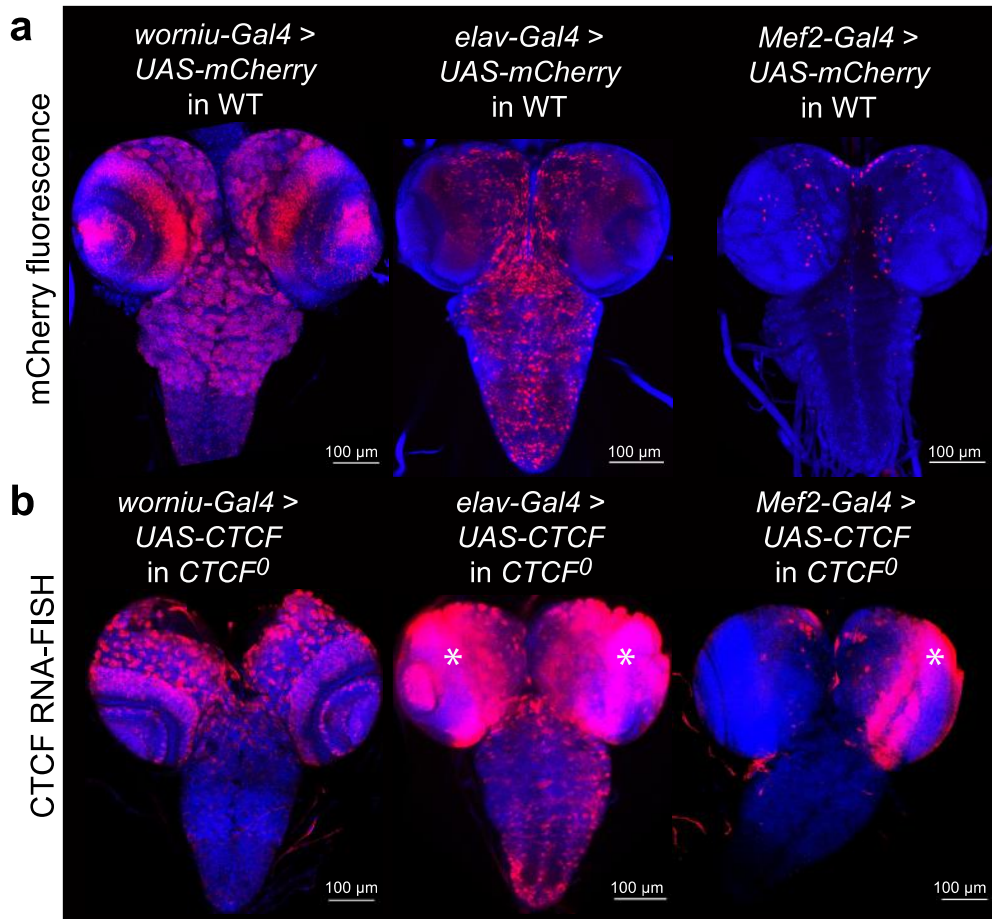


Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

CTCF loss has limited effects on global genome architecture in *Drosophila* despite critical regulatory functions

SUPPLEMENTARY INFORMATION

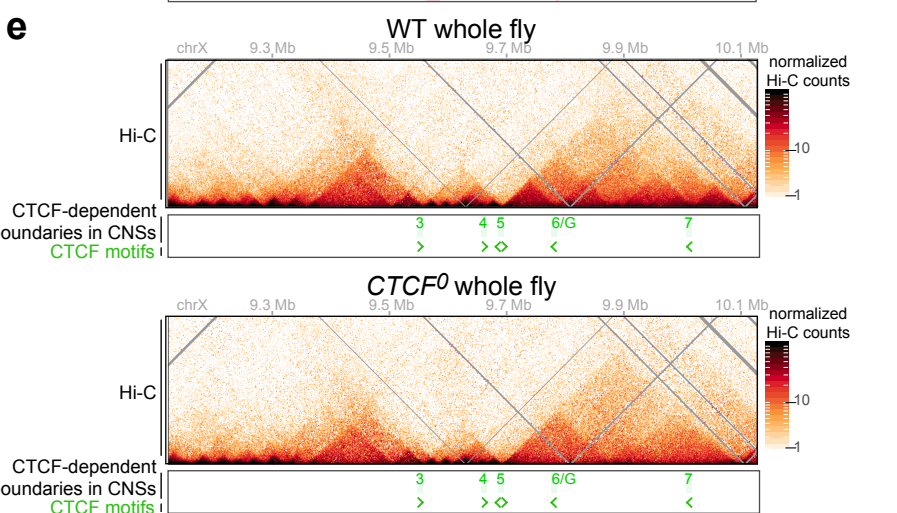
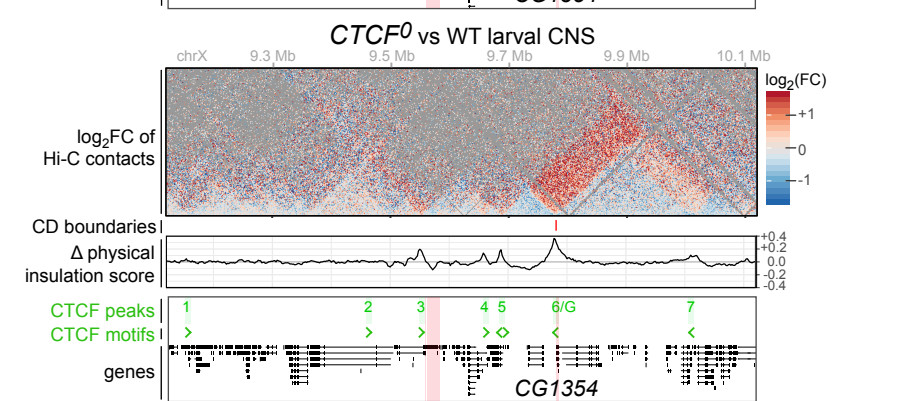
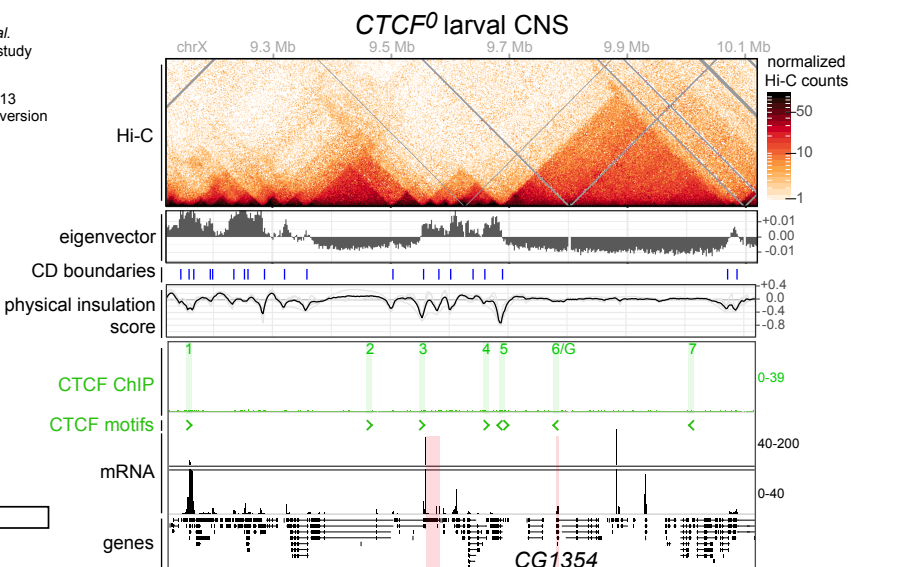
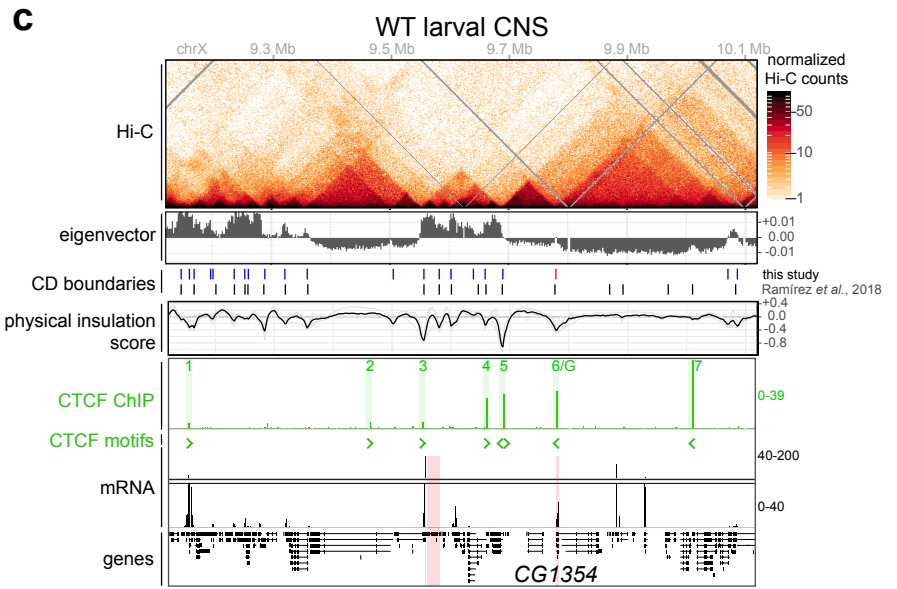
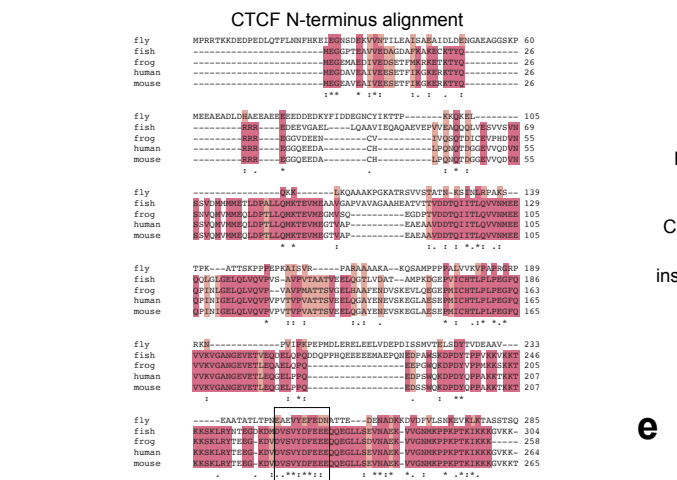
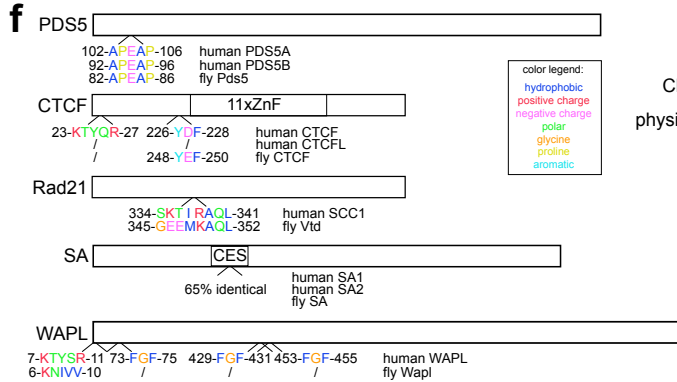
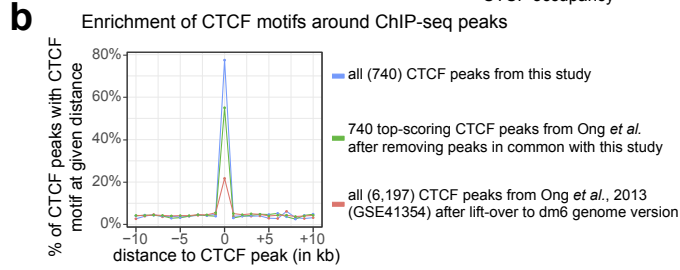
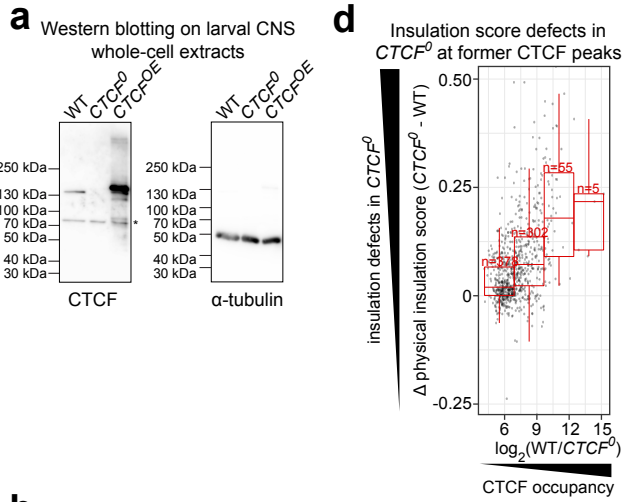


Supplementary Figure 1: Tissue-specific rescue of *CTCF⁰* mutants.

(a) Wildtype third instar larval central nervous systems of animals expressing *UAS-mCherry* (red) under the control of Gal4 drivers used in Fig.1 that are active in neural stem cells (*worniu-Gal4*), mature neurons (*elav-Gal4*) or muscle (*Mef2-Gal4*). Gal4-expressing cells are marked by mCherry direct immunofluorescence. Scale bars 100 μ m.

(b) Third instar larval central nervous systems of *CTCF⁰* mutants analyzed in Fig. 1c-d, in which *UAS-CTCF* was expressed in restricted cells under the control of Gal4 drivers. Samples were labeled by RNA-FISH with an antisense probe to *CTCF* mRNA. White asterisks mark variable non-specific signal visible in the optic lobes of some *CTCF⁰* mutants, also those without a *UAS-CTCF* transgene. Scale bars 100 μ m.

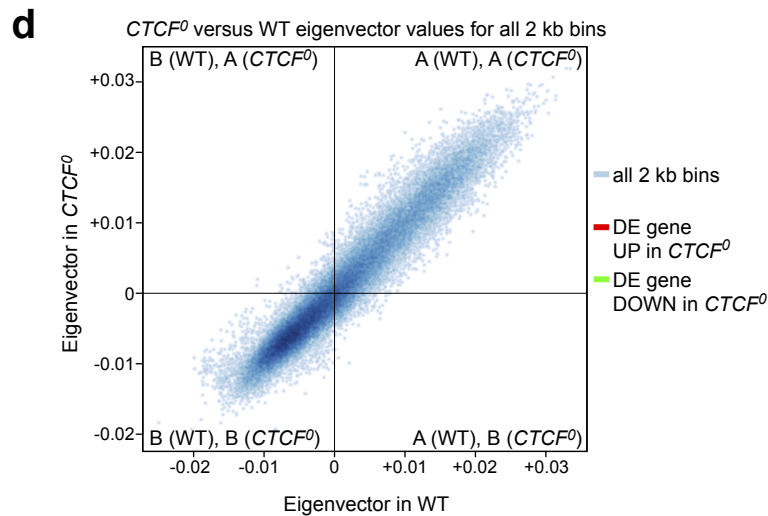
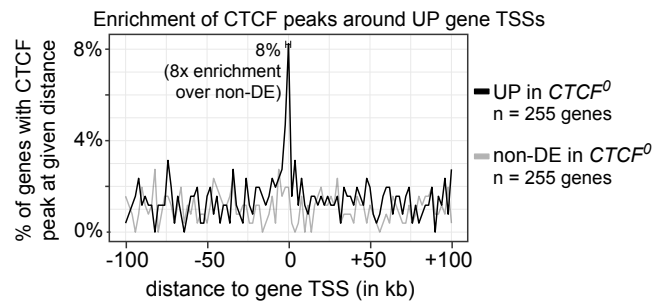
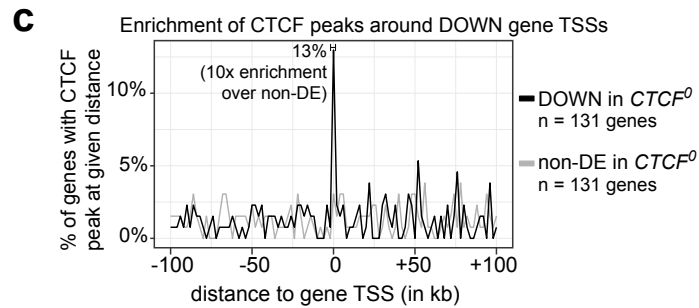
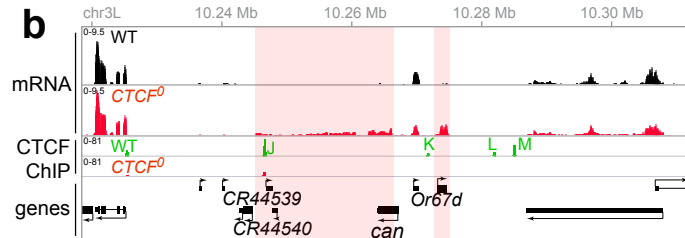
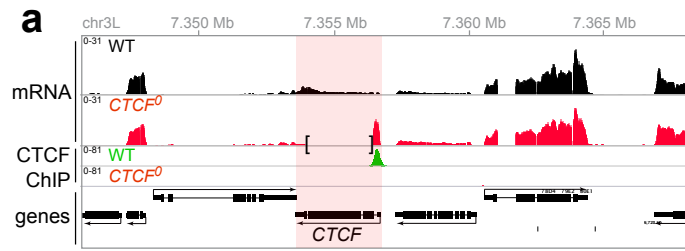
Supplementary Figure 2



Supplementary Figure 2: Characterization of CTCF antibody and CTCF-dependent contact domain boundaries.

- (a) Western blotting of whole-cell extracts from WT, *CTCF⁰* and CTCF-overexpressing (*CTCF^{OE}*) larval CNSs probed with anti-CTCF (asterisk marks cross-reacting band), then with anti-alpha-tubulin to verify equal loading of each extract.
- (b) Percentage of CTCF peaks from this or a published study¹ with at least one CTCF motif (JASPAR motif MA0531.1) at a given distance.
- (c) Example locus like Fig. 2c additionally displaying mRNA-seq tracks (genes differentially expressed in *CTCF⁰* highlighted). CTCF-dependent boundaries are observed near genes with unchanged (peak 7), decreased (*CG1354* near peak 6/G) or increased (peak 3) expression in *CTCF⁰* CNSs.
- (d) Physical insulation score differences measured in *CTCF⁰* minus WT Hi-C maps as a function of CTCF occupancy measured by ChIP-seq [$\log_2(\text{WT}/\text{CTCF}^0)$] for each CTCF peak (dots). Box plots of indicated n CTCF peaks binned by ChIP occupancy are overlaid. Box plots in d and g: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile ranges; points, outliers.
- (e) Hi-C maps like (c) generated from single whole-bodied WT or *CTCF⁰* flies. CTCF peaks 3-7 at which boundary defects were detected in *CTCF⁰* larval CNS Hi-C maps are marked.
- (f) Fly and human CTCF and cohesin subunits and regulators implicated in TAD boundary formation. DNA-bound CTCF zinc-fingers (ZnF) form a semi-permeable barrier to loop-extruding cohesin in mammals². Human CTCF YDF and fly CTCF YEF bind to the conserved essential surface in cohesin³ (Fig. 2g). Mammalian CTCF KTYQR (similar to WAPL KTYSR) binds to PDS5A APEAP²; but fly CTCF lacks this motif and does not co-purify with Pds5 in vivo (Supplementary Fig. 5a). Human CTCF binds to cohesin competitively with WAPL (possibly via WAPL FGF) in vitro³. (Bottom) Clustal Omega alignment of fly and vertebrate (fish, frog, human, mouse) CTCF N-termini.
- (g) Box plots of \log_2 ratios (n indicated in the figure) of right-over-left Hi-C interactions in WT established by forward (red) or reverse (blue) pointing CTCF motifs, at increasing distances from the CTCF peak (in 16 kb bins). Colored lines connect means. \log_2 ratios between forward and reverse pointing CTCF motifs were significantly different (indicated adjusted p-values <0.05) between 16 and 192 kb (two-sided Wilcoxon rank-sum test with Benjamini-Hochberg multiple testing correction).

Supplementary Figure 3



Supplementary Figure 3: Transcriptional defects in *CTCF*⁰ mutant CNSs.

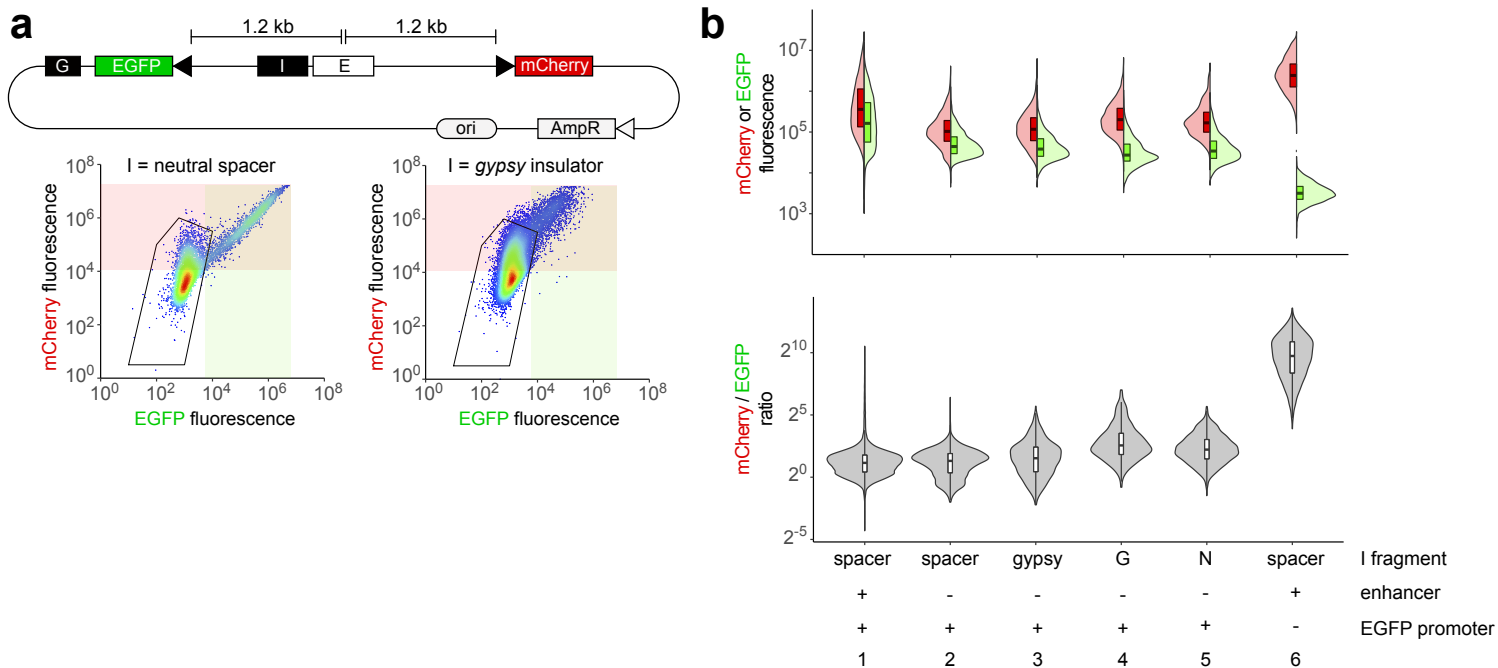
(a) Like Figs. 3b-d for the *CTCF* locus. The deleted region in *CTCF*⁰ mutants is bracketed. Increased transcription of *CTCF* 5'UTR in *CTCF*⁰ mutants could be due to potential CTCF-autoregulation or driven by 3xP3 regulatory sequences of the knocked-in *DsRed* selection marker ^{4,5}.

(b) As above for a locus harboring a cryptic transcript (transcribed from left-to-right) in *CTCF*⁰ CNSs, and *Or67d*.

(c) (Top) Percentage (in y) of n = 131 DE genes with decreased expression in *CTCF*⁰ larval CNSs (black) or n = 131 randomly sampled expression-level-matched non-DE genes (grey) with at least one of 740 CTCF peaks at a given distance (per 2 kb bins) around the gene TSS, measured in the direction of transcription (in x). 13% of DOWN genes have at least one CTCF peak within ± 1 kb of their TSS, which is 10-fold higher than the average enrichment at the sampled non-DE genes. (Bottom) Percentage (in y) of n = 255 DE genes with increased expression in *CTCF*⁰ larval CNSs (black) or n = 255 randomly sampled expression-level-matched non-DE genes (grey) with at least one of 740 CTCF peaks at a given distance (per 2 kb bins) around the gene TSS, measured in the direction of transcription (in x). 8% of UP genes have at least one CTCF peak within ± 1 kb of their TSS, which is 8-fold higher than the average enrichment at the sampled non-DE genes. 6 out of 392 DE genes were omitted from these analyses because they overlapped blacklisted regions in the CTCF ChIP-seq analysis ⁶.

(d) Eigenvector values in *CTCF*⁰ mutants (in y) versus WT (in x) for every 2 kb bin of chromosomes 2, 3 and X (density plot in blue), with bins overlapping DE gene TSSs highlighted in red (for genes with increased expression in *CTCF*⁰) or green (for genes with decreased expression in *CTCF*⁰). Bins with positive eigenvector values are in the A (active) compartment, those with negative eigenvector values are in the B (inactive) compartment. Bins in the top left and bottom right quadrants are considered to be located in opposite compartments in *CTCF*⁰ mutants relative to WT.

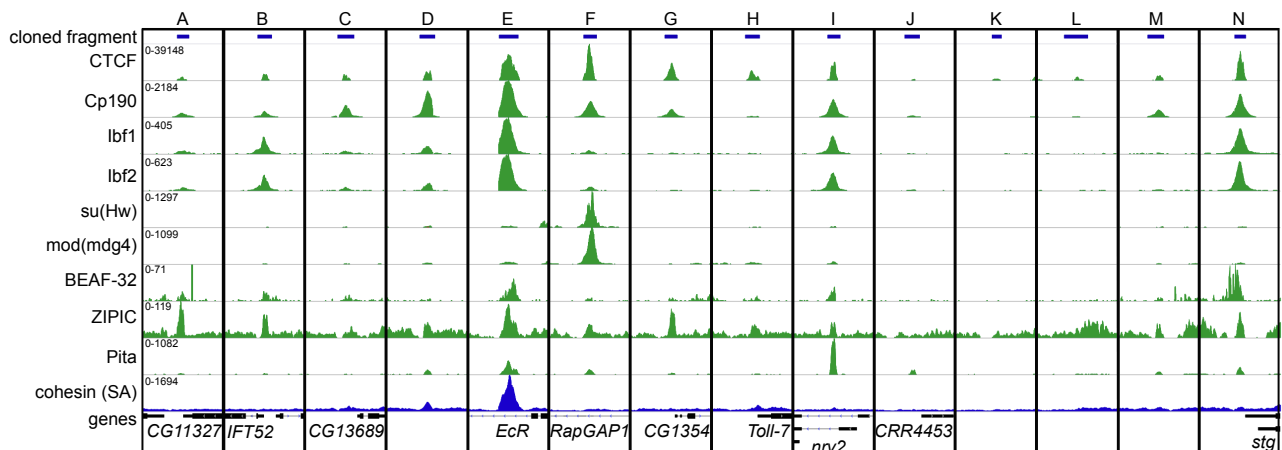
Supplementary Figure 4



c

label	dm6 coordinates	size	close to differentially expressed gene in <i>CTCF</i> ⁰	position relative to closest gene	distance from fragment center to closest annotated TSS	# matches to CTCF motif	CTCF-dependent boundary in larval CNS?
A	chr2L:6684687-6684980	294 bp	<i>Tsp</i>	overlaps TTS	467 bp	1	partially CTCF-dependent
B	chr2L:6557352-6557697	346 bp	<i>IFT52</i>	overlaps TSS	22 bp	1	partially CTCF-dependent
C	chr2L:602317-602729	411 bp	<i>CG13689</i>	upstream of TSS	291 bp	3	strictly CTCF-dependent
D	chr2L:17784631-17785007	377 bp	none	downstream of TTS	2,151 bp	2	strictly CTCF-dependent
E	chr2R:6151039-6151512	474 bp	none	intron	5,853 bp	7	partially CTCF-dependent
F	chr2L:7534948-7535274	327 bp	none	intron	4,512 bp	3	strictly CTCF-dependent
G	chrX:9778493-9778804	312 bp	<i>CG1354</i>	upstream of TSS	96 bp	1	strictly CTCF-dependent
H	chr2R:19826269-19826627	359 bp	cryptic transcript	overlaps TSS	130 bp	0	strictly CTCF-dependent
I	chr2L:6797475-6797786	312 bp	none	intron	563 bp	2	strictly CTCF-dependent
J	chr3L:10246350-10246920	571 bp	<i>can</i>	upstream of TSS	1,843 bp	2	partially CTCF-dependent
K	chr3L:10271587-10271968	382 bp	<i>can</i>	upstream of TSS	1,412 bp	1	no boundary
L	chr3L:10281692-10282362	671 bp	<i>can</i>	downstream of TTS	7,318 bp	1	no boundary
M	chr3L:10284788-10285203	579 bp	<i>can</i>	downstream of TTS	10,000 bp	1	no boundary
N	chr3R:29251563-29251840	278 bp	<i>SP1029</i>	overlaps TTS	3,623 bp	2	strictly CTCF-dependent

d



Supplementary Figure 4: Insulator reporter assay.

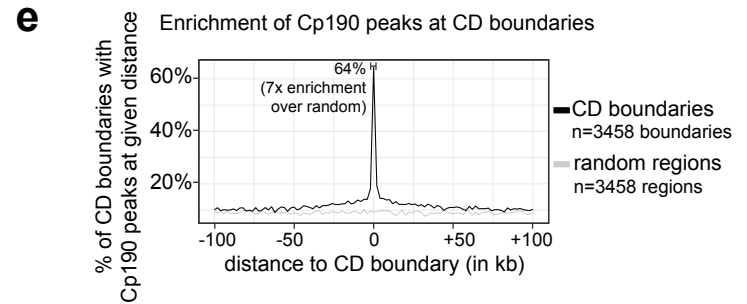
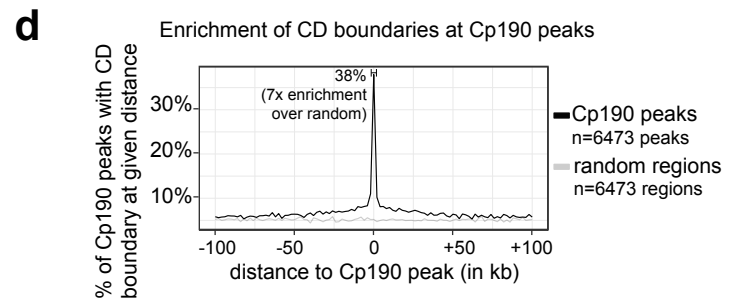
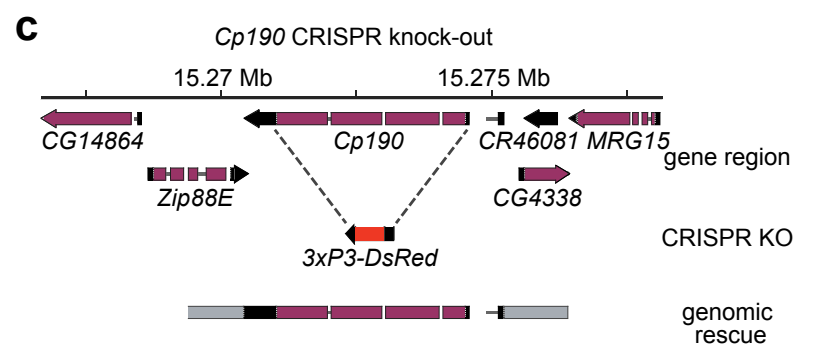
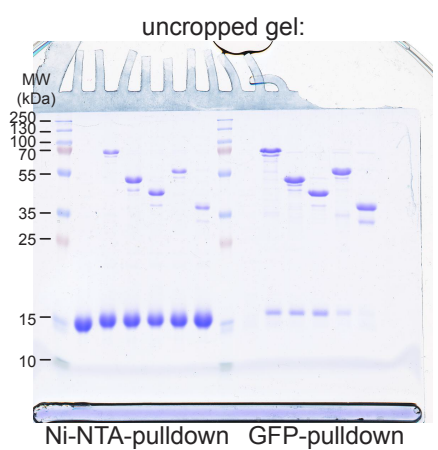
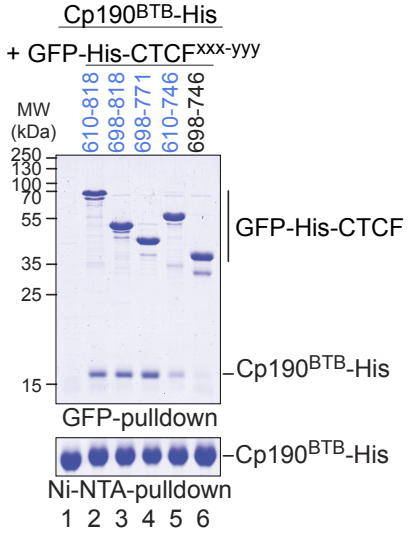
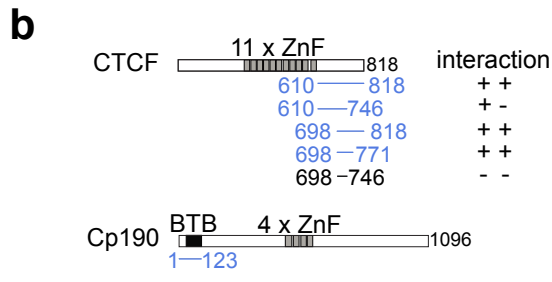
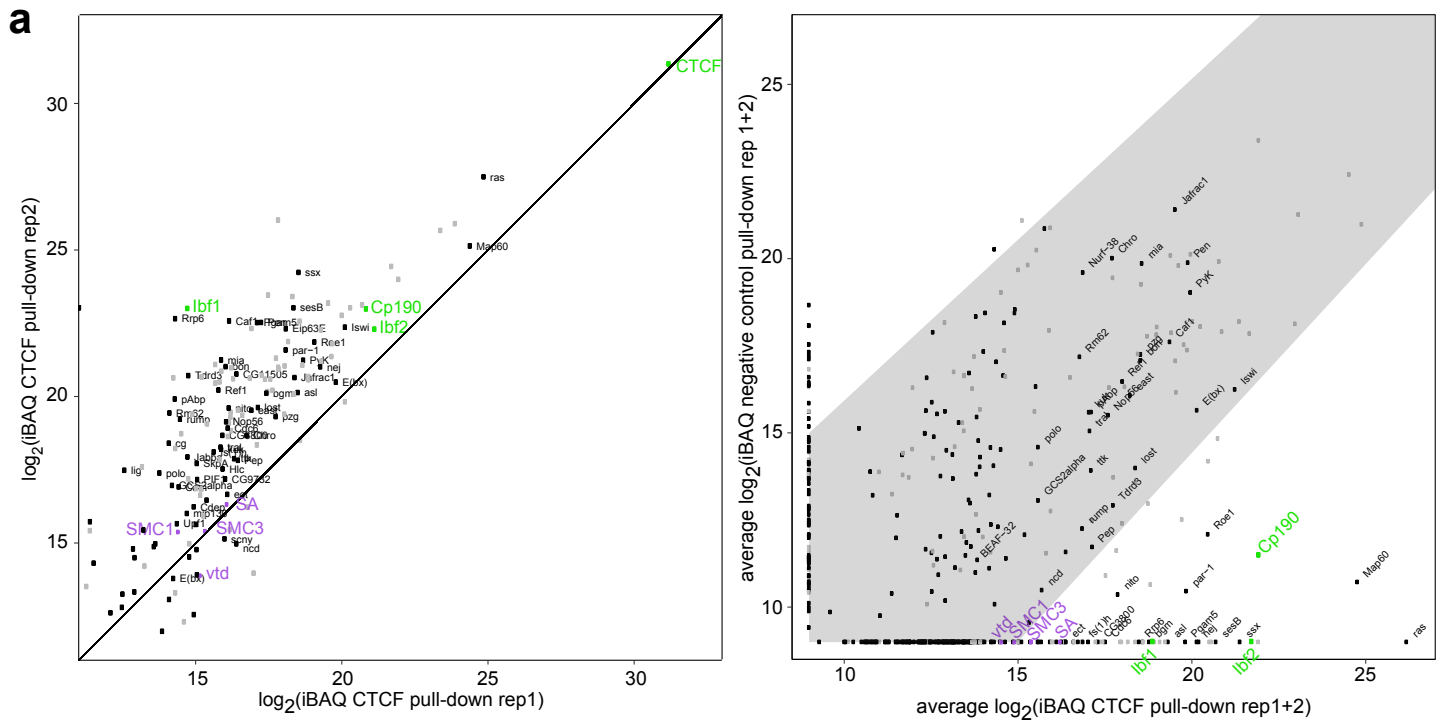
(a) Scatter plots of EGFP and mCherry fluorescence intensities in S2 cells transiently transfected with reporters with no insulator (left) or *gypsy* (right) cloned as I fragments. Gated untransfected cells (polygon) were excluded in Figs. 4b-c and in Supplementary Fig. 4b. Ranges of EGFP and mCherry fluorescence values considered to be above background (when outside of the black polygon) are shaded in green and red.

(b) Split violin plots (thick lines mark medians, boxes mark interquartile ranges) of mCherry (left) and EGFP (right) fluorescence intensities measured in thousands of single S2 cells (merged biological duplicates) transiently transfected with reporters with or without EGFP promoters or enhancers, and with the indicated I fragments. mCherry-to-EGFP ratios (\log_2 values) in single cells are shown below. Reporters in lanes 1 (with promoter and enhancer), 2 (with promoter but no enhancer) and 6 (no promoter and no enhancer) reveal enhancer-activated, basal and background EGFP fluorescence levels. Reporters shown in lanes 2-5 lack an enhancer and differ by the cloned I fragment.

(c) For each cloned CTCF peak (labeled A-N in all figures) tested in Fig. 4c: dm6 coordinates; size; name of the nearby gene if this gene was differentially expressed in *CTCF⁰* mutant CNSs; position of the CTCF peak relative to the closest gene (irrespective of whether the gene is differentially expressed in *CTCF⁰* mutants or not); distance from the center of the cloned CTCF peak to the closest annotated TSS; number of matches to the Jaspar insect CTCF motif (MA0531.1); and whether this CTCF peak overlapped a CTCF-dependent boundary in larval CNSs. (TSS: transcription start site, TTS: transcription termination site)

(d) Published ChIP-seq profiles in S2 cells of indicated insulator-binding proteins and of cohesin SA on fragments A-N. The following datasets were re-mapped and visualized: CTCF, mod(mdg4) and su(Hw) from GSE41354¹; Cp190, Ibf1 and Ibf2 from GSE47559⁷; Pita and ZIPIC from GSE54337⁸; BEAF-32 from GSE52962⁹ and dSA from GSE85191¹⁰. Scales show total counts.

Supplementary Figure 5



Supplementary Figure 5: CTCF stably binds to the pervasive boundary-associated factor Cp190.

(a) Proteins identified by mass spectrometry with indicated \log_2 intensity-based absolute quantification (iBAQ) values. Known insulator-binding proteins (green) and cohesin subunits (purple) are indicated; abundant likely common contaminants are marked by grey dots (see Source Data). (Left) Proteins with iBAQ values >11 reproducibly co-purified with recombinant GFP-tagged CTCF N-terminus mixed with *Drosophila* embryonic nuclear extracts in biological duplicates. CTCF bait was added in excess to extracts and is disproportionately abundant in the pull-down. (Right) Average \log_2 iBAQ values from biological duplicates of CTCF pull-downs or negative control pull-downs using recombinant GFP. Proteins not enriched by more than 64-fold ($6 \log_2$ iBAQ units) in the CTCF pull-down relative to the negative control are shaded in grey.

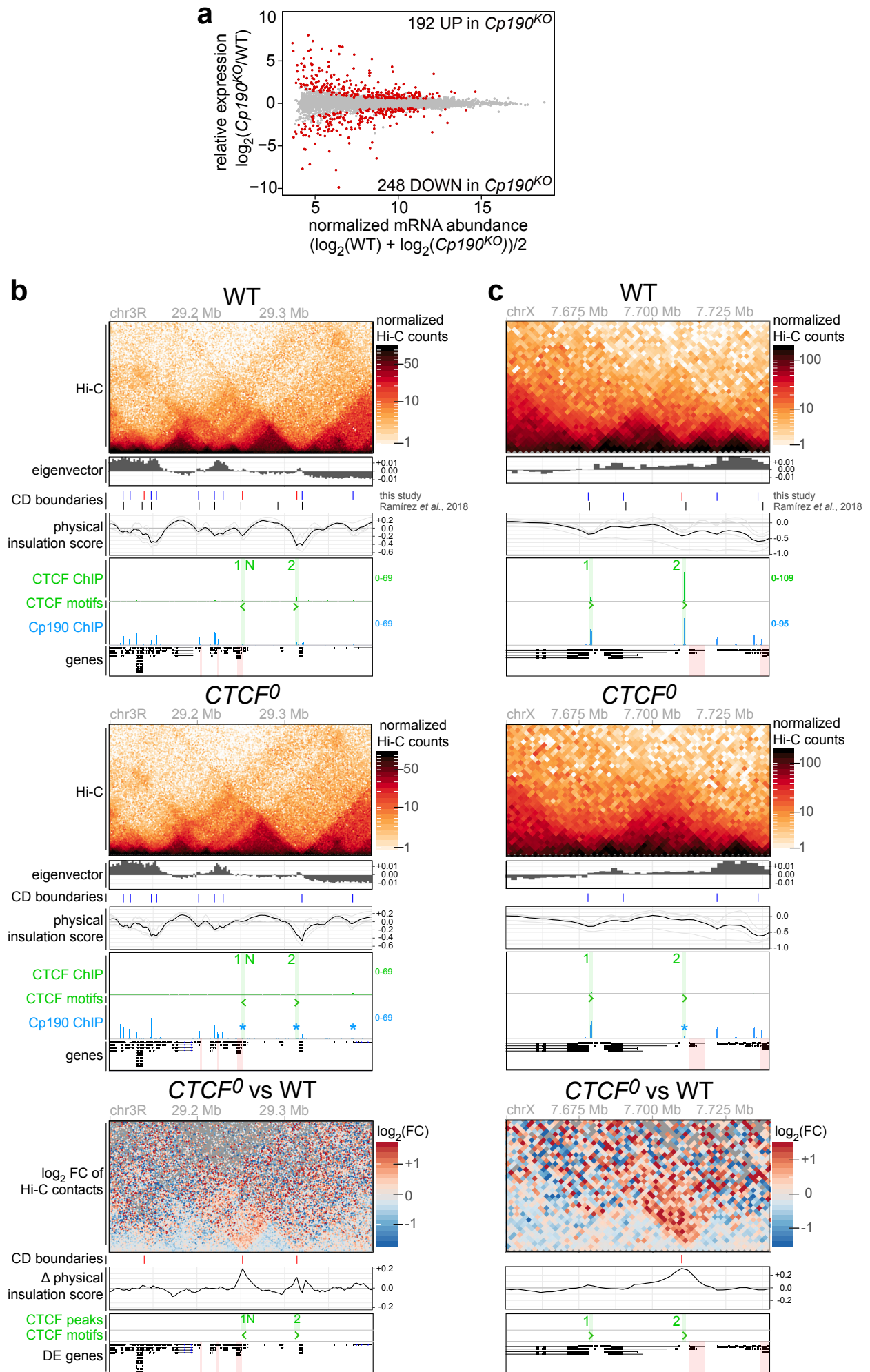
(b) GFP-His-tagged CTCF constructs were co-expressed with His-tagged Cp190^{BTB} in bacteria (blue fragments interacted, black did not). Extracts were split in half and subjected to GFP or Ni-NTA pull-downs (to control for similar amounts of Cp190^{BTB} in each extract). CTCF⁶⁹⁸⁻⁷⁷¹ was the smallest fragment that retained Cp190^{BTB} with similar efficiency as CTCF's entire C-terminal domain. Further C-terminally truncated CTCF⁶¹⁰⁻⁷⁴⁶ retained Cp190^{BTB} more weakly, and further N-terminally truncated CTCF⁶⁹⁸⁻⁷⁴⁶ bound even more weakly. The uncropped gel is below.

(c) *Cp190* extended gene region with coding (purple) and noncoding (black) exons and introns (lines). In *Cp190*^{KO} mutants, a *DsRed* selection marker replaces *Cp190* open reading frame. *Cp190*⁰ mutants were generated by excising (with FLP recombinase) the indicated FRT-flanked genomic rescue fragment from germlines of conditionally rescued *Cp190*^{KO} mothers and fathers.

(d) Percentage of $n=6473$ Cp190 peaks in WT with at least one CD boundary in WT at a given distance (per 2 kb bins). Enrichment of CD boundaries around the same number of random positions (grey) is shown as control.

(e) Percentage of $n=3458$ CD boundaries in WT with at least one Cp190 peak in WT at a given distance (per 2 kb bins). Enrichment of Cp190 peaks around the same number of random positions (grey) is shown as control.

Supplementary Figure 6



Supplementary Figure 6: Transcriptional misregulation in *Cp190*^{KO} CNSs and Hi-C maps of loci from Fig. 6.

(a) RNA-seq MA plot of *Cp190*^{KO} versus WT larval CNSs with mean abundance (in x) plotted as a function of enrichment (in y). DE genes ($p_{adj} < 0.05$ and $|\text{fold change}| > 1.5$) are red.

(b) *SP1029* gene region shown in Fig. 6b with Hi-C maps, eigenvector values (positive for A compartment, negative for B compartment), CD boundaries from this study and a Hi-C study in cultured cells¹¹, physical insulation score (calculated with different window sizes in grey, average in black), CTCF ChIP-seq (CTCF peaks highlighted and numbered), CTCF motif orientations in DNA, Cp190 ChIP-seq (asterisks mark Cp190 peaks in *CTCF*⁰ mutants with reduced occupancy relative to WT revealed by differential analysis), and gene tracks (differentially expressed genes in *CTCF*⁰ relative to WT shaded in red) in WT (top) and *CTCF*⁰ (middle) larval CNSs. (Below) Differential (*CTCF*⁰ minus WT) Hi-C maps and physical insulation score.

(c) Same as (b) but for the *CG15478* extended gene region shown in Fig. 6d.

Supplementary Table 1: Quality metrics of Hi-C reads.

Hi-C sample	total reads	interchr	intrachr	intrachr <20 kb	intrachr >20 kb
WT larval brain (combined replicates)	200,000,000 (downsampled)	24,618,605	175,381,395	50,788,347	124,593,048
<i>CTCF⁰</i> larval brain (combined replicates)	200,000,000 (downsampled)	27,076,150	172,923,850	47,787,104	125,136,746
WT whole fly (single replicate)	11,0261,440	22201387	88060053	26894987	124,593,048
<i>CTCF⁰</i> whole fly (single replicate)	83,302,985	19,068,061	64,234,924	16,155,686	48,079,238

Supplementary Table 2: Contact domain (CD) boundary counts.

(Column 1) Total CD boundaries called in both WT and *CTCF⁰* mutants (common), or only in WT, or only in *CTCF⁰* mutant larval CNS Hi-C maps. CD boundaries were split into those with (columns 2 and 3) or without (columns 4 and 5) a CTCF peak within ± 2 kb (the resolution at which CD boundaries were called). Strongly affected CD boundaries (columns 3 and 5) were defined as having a physical insulation score difference between *CTCF⁰* minus WT Hi-C maps > 0.1 (weaker boundary in *CTCF⁰* relative to WT) for boundaries only called in WT, or < -0.1 (stronger boundary in *CTCF⁰* relative to WT) for boundaries only called in *CTCF⁰*, or an absolute value > 0.1 for common boundaries.

	1	2	3	4	5
	All CD boundaries	CD boundaries with CTCF peak ± 2 kb		CD boundaries without CTCF peak ± 2 kb	
		all	strongly affected	all	strongly affected
Common	2891	218	75	2673	51
Only in WT	567	125	89	442	46
Only in <i>CTCF⁰</i>	512	6	1	506	22
Total	3970	349	165	3621	119

Supplementary Table 3: Interactive links to browse Hi-C and CHIP-seq data on Juicebox.

Differential Hi-C maps of WT (A) and *CTCF⁰* (B) larval CNSs are displayed for the indicated genomic intervals (1 Mb each). Additional tracks are: CTCF ChIP-seq peaks in WT (track 1), CTCF peaks intersected with underlying motifs in forward (green) or reverse (red) orientations (track 2), Cp190 ChIP-seq peaks in WT (track 3), and Cp190 ChIP-seq signal that is differentially enriched in *CTCF⁰* mutants relative to WT (track 4).

dm6 genomic coordinates	url
NT_033779.5:1-1,000,000	http://bit.ly/2TM42hd
NT_033779.5:1,000,000-2,000,000	http://bit.ly/2VSG0DV
NT_033779.5:5,000,000-6,000,000	http://bit.ly/2xgSijn
NT_033779.5:6,000,000-7,000,000	http://bit.ly/2PTGFRO
NT_033779.5:7,000,000-8,000,000	http://bit.ly/2PQQ3FK
NT_033779.5:14,000,000-15,000,000	http://bit.ly/3avWeHl
NT_033778.4:8,000,000-9,000,000	http://bit.ly/2TnejBz
NT_037436.4:3,000,000-4,000,000	http://bit.ly/2PU06tU
NT_037436.4:6,000,000-7,000,000	http://bit.ly/3awLuZ7
NT_037436.4:14,000,000-15,000,000	http://bit.ly/3aEI5I5
NT_033777.3:4,800,000-5,800,000	http://bit.ly/2VQiLud
NT_033777.3:7,000,000-8,000,000	http://bit.ly/3aydD1Z
NT_033777.3:14,000,000-15,000,000	http://bit.ly/3cAfmpr
NT_033777.3:16,000,000-17,000,000	http://bit.ly/2PUXaNH
NT_033777.3:28,500,000-29,500,000	http://bit.ly/2VMq1aJ
NC_004354.4:9,500,000-10,500,000	http://bit.ly/2vAMyN6
NC_004354.4:15,000,000-16,000,000	http://bit.ly/2TIN2sl
NC_004354.4:17,000,000-18,000,000	http://bit.ly/2vltGvm
NC_004354.4:20,500,000-21,500,000	http://bit.ly/2VSxBke

Supplementary References

1. Ong, C.-T., Van Bortle, K., Ramos, E. & Corces, V. G. Poly(ADP-ribosyl)ation Regulates Insulator Function and Intrachromosomal Interactions in *Drosophila*. *Cell* **155**, 148–159 (2013).
2. Nora, E. P. *et al.* Molecular basis of CTCF binding polarity in genome folding. *Nat Commun* **11**, 5612 (2020).
3. Li, Y. *et al.* The structural basis for cohesin–CTCF-anchored loops. *Nature* **578**, 1–9 (2020).
4. Gratz, S. J., Harrison, M. M., Wildonger, J. & O'Connor-Giles, K. M. Precise Genome Editing of *Drosophila* with CRISPR RNA-Guided Cas9. *Methods in molecular biology (Clifton, N.J.)* **1311**, 335–348 (2015).
5. Gambetta, M. C. & Furlong, E. E. M. The Insulator Protein CTCF Is Required for Correct Hox Gene Expression, but Not for Embryonic Development in *Drosophila*. *Genetics* **210**, 129–136 (2018).
6. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep-uk* **9**, 9354 (2019).
7. Cuartero, S., Fresán, U., Reina, O., Planet, E. & Espinàs, M. L. Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. *The EMBO Journal* **33**, 637–647 (2014).
8. Maksimenko, O. *et al.* Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome Research* **25**, 89–99 (2015).
9. Liang, J. *et al.* Chromatin Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator Proteins and Pol II Pausing. *Molecular Cell* **53**, 672–681 (2014).
10. Henriques, T. *et al.* Widespread transcriptional pausing and elongation control at enhancers. *Genes & development* **32**, 26–41 (2018).
11. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature Communications* **9**, 189 (2018).

11.2 Essential role of Cp190 in physical and regulatory boundary formation

GENETICS

Essential role of Cp190 in physical and regulatory boundary formation

Anjali Kaushal¹, Julien Dorier², Bihan Wang¹, Giriram Mohana¹, Michael Taschner³, Pascal Cousin¹, Patrice Waridel⁴, Christian Iseli², Anastasiia Semenova¹, Simon Restrepo⁵, Nicolas Guex², Erez Lieberman Aiden^{6,7,8,9}, Maria Cristina Gambetta^{1*}

Boundaries in animal genomes delimit contact domains with enhanced internal contact frequencies and have debated functions in limiting regulatory cross-talk between domains and guiding enhancers to target promoters. Most mammalian boundaries form by stalling of chromosomal loop-extruding cohesin by CTCF, but most *Drosophila* boundaries form CTCF independently. However, how CTCF-independent boundaries form and function remains largely unexplored. Here, we assess genome folding and developmental gene expression in fly embryos lacking the ubiquitous boundary-associated factor Cp190. We find that sequence-specific DNA binding proteins such as CTCF and Su(Hw) directly interact with and recruit Cp190 to form most promoter-distal boundaries. Cp190 is essential for early development and prevents regulatory cross-talk between specific gene loci that pattern the embryo. Cp190 was, in contrast, dispensable for long-range enhancer-promoter communication at tested loci. Cp190 is thus currently the major player in fly boundary formation and function, revealing that diverse mechanisms evolved to partition genomes into independent regulatory domains.

INTRODUCTION

Chromosomal contact domains are ubiquitous in different species, but there is evidence that they form through diverse mechanisms. Two fundamental questions are as follows: How are contact domains formed, and what is their function? Contact domains are known to form through compartmentalization of transcriptionally active and inactive domains or extrusion of chromosomal loops by cohesin until cohesin is stalled by DNA-bound CTCF at domain boundaries (1–4). Contact domains arising from these different mechanisms have respectively been dubbed “compartmental domains” or “topologically associating domains” (TADs) (1, 5, 6). CTCF is required to form a large proportion of mammalian contact domain boundaries but less than 10% of all boundaries in *Drosophila* (7). In *Drosophila*, three-quarters of contact domain boundaries are located at promoters of highly and ubiquitously expressed genes and hence called promoter boundaries, while the remaining one-quarter are nonpromoter boundaries occupied by different DNA binding proteins such as CTCF or Su(Hw) (suppressor of Hairy-Wing) (8). Promoter and nonpromoter boundaries in flies all share a common feature: They are bound by Centrosomal protein 190 kDa (Cp190) (7, 8). The diversity of boundary-associated factors in flies compared to mammals raises the possibility that additional proteins other than CTCF form physical boundaries in chromosomes through yet unknown mechanisms.

How genome folding into contact domains affects gene regulation has been intensely investigated by studying mammalian

CTCF. A major challenge is that CTCF is essential for mammalian cell survival, and acute CTCF depletion results in few transcriptional effects (2). Locus-specific CTCF-dependent boundary perturbations led to different conclusions on their relevance for gene regulation (9–13). Contact domains generally contain co-regulated genes and their cognate regulatory elements (8, 14, 15) and, in some cases, foster efficient activation of promoters by enhancers within the same domain (16–18). Conversely, contact domain boundaries can exert genetic insulator activity by strongly dampening communication between regulatory elements and gene promoters in flanking domains, in exceptional cases preventing developmental defects and human disease (18–20). An emerging model in mammals is that cohesin-mediated loop extrusion brings enhancers into functional contact with compatible promoters all the way until cohesin is stalled by CTCF (21, 22).

In contrast to mammals, CTCF has a limited role at selected boundaries in *Drosophila* (7). Despite the presence of Cp190 at nearly all boundaries, studies have not yet addressed whether Cp190 is critical for gene regulation specificity. It can seem puzzling that Cp190 associates with very different types of boundaries (promoter and nonpromoter), and it remains unclear whether it exerts different activities at these sites (23, 24).

Cp190 was identified in a genetic screen as essential for the activity of the well-characterized *gypsy* insulator and was later shown to be required at additional insulators (25–27). It remains, however, unclear how relevant Cp190 is for natural gene expression (7, 28–30). Cp190 copurifies with diverse proteins, indicating assembly into complexes whose exact compositions remain unclear because it is challenging to assemble them recombinantly. Several Cp190 interactors exert insulator activity in transgenic reporter assays, suggesting that Cp190 is an essential insulator cofactor (31, 32). Cp190 is recruited to chromosomes by sequence-specific DNA binding insulator proteins (7, 26, 27, 30). For example, CTCF recruits Cp190 to CTCF-dependent boundaries (7). In *CTCF*⁰ mutants, residual Cp190 binding was observed at some former CTCF peaks, and this importantly correlated with boundary retention (7). This had raised

¹Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland.

²Bioinformatics Competence Center, University of Lausanne, 1015 Lausanne, Switzerland.

³Department of Fundamental Microbiology, University of Lausanne, 1015 Lausanne, Switzerland.

⁴Protein Analysis Facility, University of Lausanne, 1015 Lausanne, Switzerland.

⁵arcoris bio AG, Lüssirainstrasse 52, 6300 Zug, Switzerland.

⁶The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA.

⁷National Institute of Genetics, 1111 Yaya, Mishima, Shizuoka 411-8540, Japan.

⁸UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA 6009, Australia.

⁹Shanghai Institute for Advanced Immunological Studies, ShanghaiTech, Pudong 20120, China.

*Corresponding author. Email: mariacristina.gambetta@unil.ch

Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

Downloaded from https://www.science.org on June 20, 2022

the possibility that Cp190 synergizes with CTCF to form contact domain boundaries.

Apart from limiting regulatory cross-talk, some Cp190-bound insulators were shown to physically pair and thereby bring linked promoters and distal regulatory elements into close proximity to enable long-range regulation (33–35). In extreme examples, physical pairing of insulators enabled regulation of a promoter by an enhancer 140 kb away or by an enhancer located on the homologous chromosome (36–38). Observations from seminal studies led to the influential model in which DNA-bound insulator proteins pair through Cp190 acting as a universal glue, for example by dimerizing through its BTB (Broad-Complex, Tramtrack and Bric a brac) domain (29, 33, 39, 40).

Here, we directly address the biological relevance of the major fly boundary-associated factor Cp190. We performed Hi-C, Capture-C, and chromatin immunoprecipitation sequencing (ChIP-seq) in *Cp190⁰* mutants completely lacking Cp190 to uncover that Cp190 is required to form most nonpromoter boundaries. Promoter boundaries are insensitive to Cp190 loss and are thus formed through separate mechanisms. By optimizing our genetic strategy to generate not only *Cp190⁰* and *CTCF⁰* single mutants but also *double⁰* mutants lacking both CTCF and Cp190 products combined, we demonstrate that Cp190 is required for DNA-bound CTCF to form a robust boundary. We show that Cp190 associates with various insulator proteins in the context of core complexes with shared subunits. We quantify the relative enhancer-blocking activities of these complexes in a reporter assay to assess whether promoter and nonpromoter boundaries have different insulator activities. Last, by exploring how gene regulation is affected at well-characterized developmental loci, we were able to clearly assess the relevance of widespread contact domain boundary impairment for both inhibition and facilitation of enhancer-promoter communication at these loci during embryogenesis. In *Cp190⁰* mutants, all tested developmental genes are expressed in their endogenous and, in some cases, additional ectopic patterns in a manner consistent with characterized enhancers in flanking contact domains. We reveal that Cp190 is critical for the ability of classical developmental gene boundaries that we tested to block enhancers but not to mediate long-range gene activation by distal enhancers located in another contact domain. These findings demonstrate that diverse mechanisms exist to fold genomes into independent gene regulatory domains beyond what is currently known in mammals and refine our understanding of how fly contact domain boundaries affect gene regulation.

RESULTS

Cp190 is required to form nonpromoter boundaries in fly embryos

To address whether Cp190 is required to form contact domain boundaries, we performed Hi-C on 2- to 6-hour-old wild-type (WT) and *Cp190⁰* embryos lacking both maternal and zygotic *Cp190* gene products (fig. S1A). This early developmental stage was chosen to avoid confounding effects of lethality of *Cp190⁰* mutants, as half of *Cp190⁰* mutants arrest development at late embryogenesis with the remaining animals dying as young larvae (Fig. 1A). Two four-cutter restriction enzymes were combined for enhanced resolution, and Hi-C maps consisting of 80 million reads per genotype were obtained by combining the four biological replicates (table S1). In parallel, Cp190 binding sites were mapped in embryos by anti-Cp190 ChIP-seq performed in biological triplicates. A total of

2791 Cp190 peaks were defined as enriched in WT relative to *Cp190⁰* mutants (data S1). To assess the relation between Cp190 peaks and contact domain boundaries in 2- to 6-hour-old embryo Hi-C maps, boundaries were identified at 2-kb resolution with TopDom (data S2) (41). Boundaries within ± 2 kb of a Cp190 ChIP peak were defined as Cp190-occupied boundaries. Physical insulation scores were also measured for every 2-kb bin in the genome to determine how strongly Hi-C contacts are depleted across a region of interest such as a boundary or a ChIP peak (small physical insulation scores indicate strong physical insulation) (data S3). The difference of physical insulation scores in mutant versus WT embryos was calculated to assess boundary defects in mutants.

In Fig. 1B (top), all 2334 contact domain boundaries called in WT embryos are ranked by strongest (top) to weakest (bottom) boundary defects in *Cp190⁰* embryos. WT and *Cp190⁰* Hi-C maps were globally similar, and compartmental interactions were unaffected (fig. S1, B and C, and data S4). However, 22% of all contact domain boundaries detected in WT were lost in *Cp190⁰* mutants (Fig. 1B, lanes 1 and 2). Additional boundaries were retained but weaker in *Cp190⁰* than in WT, and overall, 26% of all WT boundaries were either lost or strongly weakened in *Cp190⁰* (Fig. 1B, lanes 1 to 5). The remaining three-quarters of boundaries were unaffected by Cp190 loss (Fig. 1B, lanes 1 to 5). Many of these unaffected boundaries were proximal [within ± 200 base pairs (bp)] to an active transcription start site (TSS) [Fig. 1B (lane 7) and fig. S1D]. Consistently, boundary defects were significantly higher at promoter-distal boundaries than at promoter-proximal boundaries in *Cp190⁰* mutants, although Cp190-occupied promoter-distal and promoter-proximal boundaries had similar strengths in WT embryos (Fig. 1C and fig. S1, E and F).

A previous analysis of boundaries defined in a high-resolution Hi-C study revealed that promoter and nonpromoter boundaries are differentially enriched in DNA motifs and differentially bound by the corresponding transcription factors (8). Consistently, we found that the most common motifs enriched at nonpromoter boundaries [CTCF, Ibf1 (insulator binding factor 1), and Su(Hw)] were visibly enriched at Cp190-dependent boundaries, while motifs enriched at promoter boundaries [such as BEAF-32 (boundary element-associated factor of 32 kD), M1BP (motif 1 binding protein), core motif-6, and ZIPIC (zinc-finger protein interacting with CP190)] were visibly enriched at Cp190-independent boundaries [Fig. 1B (lane 8) and fig. S1D]. Conversely, boundary defects measured in *Cp190⁰* mutants at boundaries occupied by Cp190 in WT were significantly higher when Cp190 peaks overlapped CTCF, Ibf1, or Su(Hw) motifs (Fig. 1D and fig. S1G). In contrast, boundary defects in *Cp190⁰* were not higher at boundaries occupied by Cp190 in WT that overlapped BEAF-32, ZIPIC, M1BP, or core motif-6 than at boundaries not overlapping these motifs (Fig. 1D and fig. S1G).

To determine how generally physical insulation defects are observed at former Cp190 peaks in *Cp190⁰* irrespective of their localization at contact domain boundaries defined by TopDom, physical insulation score differences between WT and *Cp190⁰* mutants were measured across all 2791 Cp190 peaks [ranked by ChIP occupancy in Fig. 1B (bottom)]. Domain boundaries were enriched within ± 2 kb of many (49%) Cp190 peaks (Fig. 1C, lane 1). Boundary defects in *Cp190⁰* mutants were only observed at a subset of former Cp190 peaks, with more prominent defects visible at Cp190 peaks that are highly occupied in WT [Fig. 1C (lane 5) and fig. S1H].

These results are illustrated at the *Antennapedia* complex (ANT-C) comprising five HOX genes that determine the identity of anterior

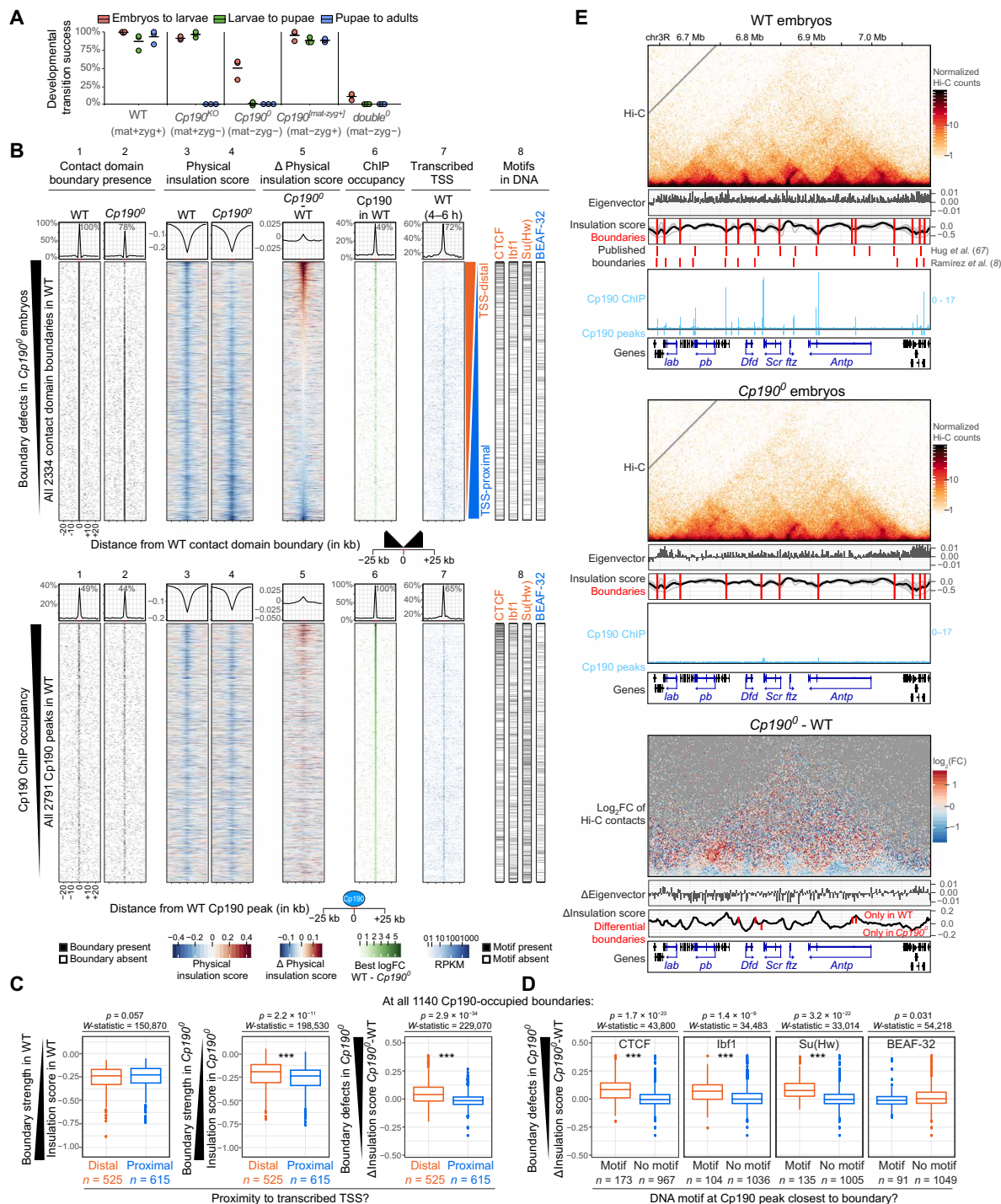


Fig. 1. Cp190 is required to form nonpromoter boundaries in fly embryos. (A) Percentages of indicated genotypes (with/without maternal/zygotic protein) that completed indicated developmental transitions in three biological replicates. Horizontal lines show means. (B) Top: Distribution of indicated datasets in 2-kb bins ± 25 kb around all WT contact domain boundaries ranked by physical insulation defects in $Cp190^0$. Lane 8 shows the presence of indicated DNA motifs in the central 2-kb bin. Summarized values (average physical insulation score or percentage of WT boundaries with boundary/ChIP peak/transcribed TSS present) across 2-kb bins are shown. Enrichments ± 2 kb around the central boundary are indicated. Bottom: Same but for all WT Cp190 ChIP peaks ranked by ChIP occupancy. (C) Physical insulation scores or differences measured at all 1140 Cp190-occupied boundaries whose most boundary-proximal Cp190 peak is distal or proximal to a transcribed TSS in WT embryos (75). P values and W -statistics from two-sided Wilcoxon rank sum test with continuity correction are indicated. (D) Similar to (C) but at all 1140 Cp190-occupied boundaries whose most boundary-proximal Cp190 peak overlaps or does not overlap the indicated DNA motif. (E) Example locus (dm6 coordinates) Hi-C maps (2-kb resolution), eigenvector values (2-kb resolution, positive/negative for A/B compartments), physical insulation score (calculated with different window sizes in gray, average in black), and contact domain boundaries (red lines) from this (above) and published (below) Hi-C studies (8, 67) and Cp190 ChIP-seq (reads per million). Homeobox genes are blue. Differential ($Cp190^0$ minus WT) values are shown below.

body segments (Fig. 1E). Cp190 was bound at most contact domain boundaries in WT embryos but was undetectable in *Cp190⁰* (Fig. 1E). Some boundaries were lost in *Cp190⁰*, and others were weaker but retained (Fig. 1E). The consequences of these blurred contact domain boundaries on HOX gene regulation are described later in the “Cp190 prevents regulatory cross-talk between early patterning gene loci” section.

We conclude that Cp190 is required to form one-quarter of all fly domain boundaries and is thus the major architectural protein required for fly domain boundary formation described to date. Although Cp190 is widely associated with domain boundaries, it only mediates formation of nonpromoter boundaries. Some boundaries are weakened but persist in *Cp190⁰* mutants, suggesting that Cp190 synergizes with other boundary-forming mechanisms at these sites.

Cp190 is required for boundary formation at CTCF peaks

Our finding that a subset of Cp190-dependent boundaries is enriched for CTCF motifs (Fig. 1B, lane 8) and the fact that CTCF recruits Cp190 to CTCF binding sites (7, 26, 27) led us to hypothesize that Cp190 is an essential cofactor required for the ability of CTCF to form robust boundaries. This hypothesis predicts that only those CTCF peaks that are cobound by Cp190 would be physical boundaries. To test this, we identified 1477 CTCF peaks defined as enriched in WT relative to *CTCF⁰* mutants by anti-CTCF ChIP-seq (data S5) and assessed the location of boundaries around CTCF ChIP peaks. Thirteen percent of all contact domain boundaries in WT embryos were located within ± 2 kb of a CTCF peak (fig. S2A, lane 11), consistent with previous reports that CTCF is only enriched at a subset of boundaries (7, 8). When all CTCF peaks were ranked from highest (top) to lowest (bottom) ChIP occupancy (Fig. 2A, lane 11), CTCF peaks with high or low CTCF occupancy were associated with contact domain boundaries but an abundant class of CTCF peaks with intermediate ChIP occupancy were clearly not [Fig. 2A (lane 1) and fig. S2B]. CTCF peaks without associated boundaries did not colocalize with Cp190, whereas higher and lower occupancy CTCF peaks did (Fig. 2A, lane 13). CTCF+Cp190 colocalization was significantly positively associated with localization at a boundary (fig. S2C). By assessing their genomic locations, we realized that CTCF peaks not colocalizing with Cp190 correspond to previously described CTCF standalone peaks that are frequently located in introns in contrast to CTCF+Cp190 cobound peaks (fig. S2D) (30, 42). We had not noticed that CTCF standalone peaks are not physical boundaries in our previous analysis of *CTCF⁰* mutant larval central nervous systems (7) because many CTCF peaks with intermediate occupancy in WT embryos are low occupancy peaks in WT larval nervous systems (fig. S2, E and F). We had thus previously assumed that domain boundaries were difficult to detect at weakly occupied CTCF peaks. These new results in *CTCF⁰* embryos, however, clearly reveal that physical boundaries are only present at CTCF sites that are co-occupied by Cp190. Cp190 might therefore be required for boundary formation at CTCF peaks, or alternatively, boundaries may not be established when CTCF binds at specific genomic locations (such as introns).

If Cp190 is a CTCF cofactor required for robust boundary formation, a second prediction is that CTCF-dependent boundaries should also depend on Cp190. To test this, we performed Hi-C on *CTCF⁰* and *double⁰* (completely lacking both CTCF and Cp190; fig. S1A) 2- to 6-hour-old embryos, in parallel to the WT and *Cp190⁰* embryos described above (table S1). Cp190 ChIP peaks were also mapped in *CTCF⁰*, and CTCF ChIP peaks were mapped in *Cp190⁰*.

Fewer *double⁰* mutants completed embryogenesis compared to *Cp190⁰*, indicating that additional loss of CTCF aggravated the embryonic lethality of *Cp190⁰* (Fig. 1A). In *CTCF⁰* mutants, physical insulation defects were observed at many former CTCF+Cp190 cobound peaks [Fig. 2A (lanes 2 and 5) and data S2 and S3]. Cp190 binding was reduced at most of these sites [Fig. 2A (lanes 13, 14, and 16) and data S6 and S7], and we therefore could not say whether CTCF acts alone or together with Cp190 to form these boundaries. In *Cp190⁰* mutants, however, CTCF binding was largely unaffected [Fig. 2A (lanes 11, 12, and 15), and data S8 and S9], yet boundaries were defective at formerly Cp190 cobound CTCF peaks relative to WT (Fig. 2A, lane 6). An example of a domain boundary located at a CTCF+Cp190 cobound peak that relies on both CTCF and Cp190 is shown in Fig. 2B. This demonstrates that Cp190 is required for DNA-bound CTCF to form a robust boundary.

The fact that only a subset of CTCF peaks colocalize with Cp190 suggests that CTCF also exerts Cp190-independent functions. To test this, we introduced transgenes expressing truncated *CTCF* versions completely lacking CTCF N (*CTCF^{ΔN}*) or C termini (*CTCF^{ΔC}*) into *CTCF⁰* animals. *CTCF^{ΔC}* mutants lack the Cp190-interacting domain, and *CTCF^{ΔN}* lack the cohesin-interacting domain (7). Both truncated CTCF proteins were expressed in vivo (fig. S2G) and enabled about one-third of *CTCF⁰* mutants, which never hatch from the pupal case, to hatch into very short-lived adults (fig. S2H). The fact that CTCF lacking its C terminus retains some limited function indeed suggests that CTCF exerts some Cp190-independent functions.

Boundary defects in *CTCF⁰* correlate with Cp190 retention

If Cp190 is required for robust boundary formation, then a third prediction is that boundaries co-occupied by CTCF+Cp190 will be retained in *CTCF⁰* mutants if Cp190 is retained at the boundary despite CTCF loss. To test this, we focused on CTCF-occupied boundaries, defined as those within ± 2 kb of a CTCF peak in WT (Fig. 3A, lanes 1 and 11). All 312 CTCF-occupied boundaries were ranked from strongest (top) to weakest (bottom) physical insulation defects in *CTCF⁰* relative to WT (Fig. 3A, lane 5). Only 28% of these boundaries were lost in *CTCF⁰* [Figs. 3, A (lane 2) and B]. Several boundaries were even unexpectedly reinforced in *CTCF⁰* relative to WT (Fig. 3A, lane 5). Boundaries that remained intact in *CTCF⁰* retained a residual Cp190 peak (Fig. 3A, lanes 5 and 14), revealing that Cp190 is recruited there at least partially CTCF independently. Boundaries at which Cp190 was retained in *CTCF⁰* mutants were often promoter-proximal (Fig. 3A, lanes 14 and 18), suggesting that Cp190 may be redundantly recruited by promoter-associated factors. Conversely, boundaries that retained a residual Cp190 peak were significantly less weakened in *CTCF⁰* than those that lost Cp190 [Fig. 3C and fig. S3A (top)]. This effect was not seen in *Cp190⁰* [Fig. 3C and fig. S3A (middle)] or *double⁰* [Fig. 3C and fig. S3A (bottom)], indicating that boundary retention in *CTCF⁰* correlates with Cp190 presence.

More than half of CTCF-occupied boundaries that were intact in *CTCF⁰* were lost or weakened in *Cp190⁰* (Fig. 3B). In addition, the average physical insulation score defect measured at CTCF-occupied boundaries was larger in *Cp190⁰* than in *CTCF⁰* (Fig. 3A, top of lanes 5 and 6). Cp190 is therefore also required to form boundaries occupied by CTCF but unaffected or, in some cases, unexpectedly reinforced in *CTCF⁰*. Nevertheless, some boundaries were more strongly affected in *CTCF⁰* than in *Cp190⁰* (Fig. 3A, top of lane 10), indicating that CTCF retains some ability to form boundaries without Cp190 at several sites.

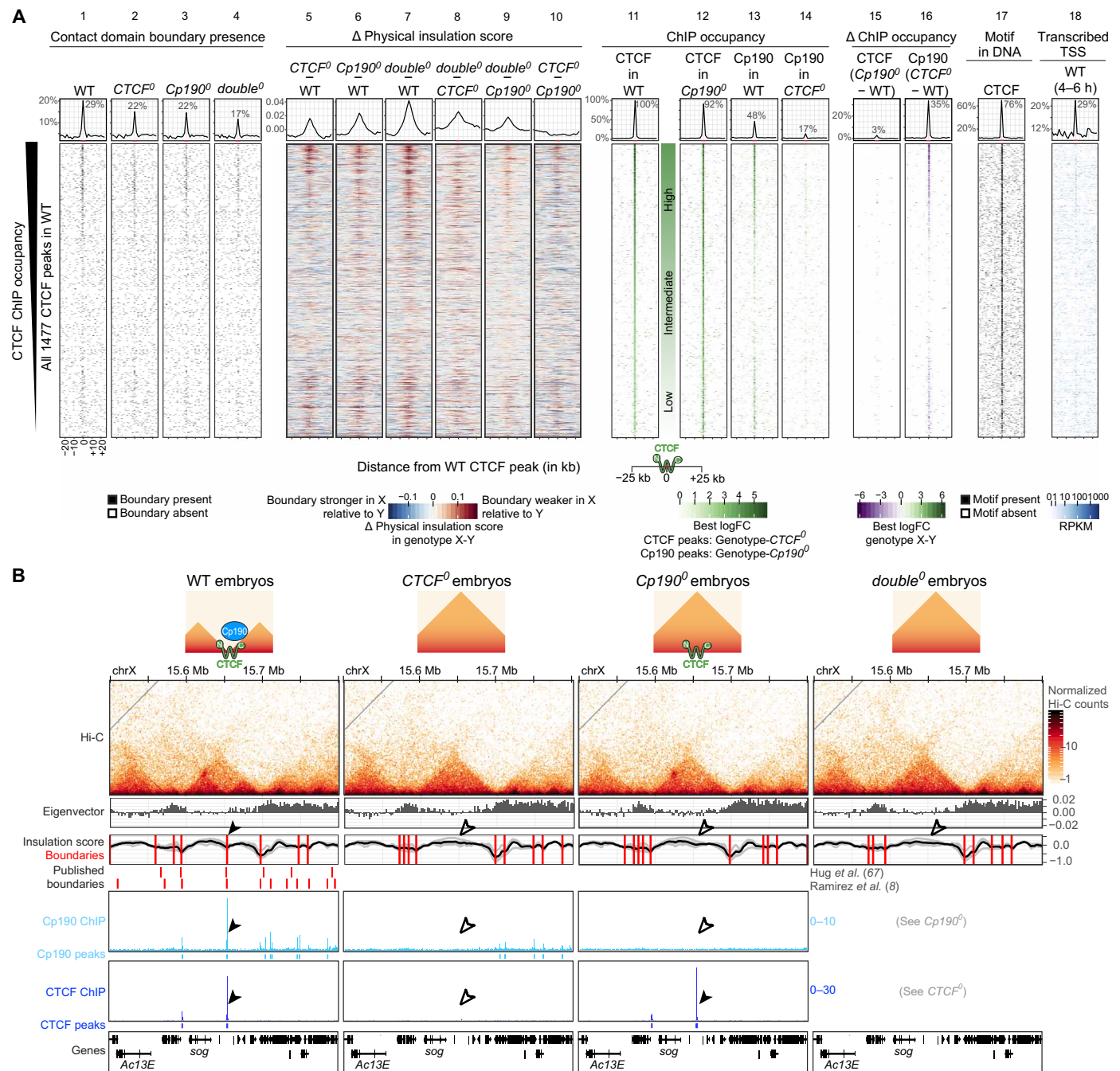


Fig. 2. Cp190 is required for boundary formation at CTCF peaks. (A) Distribution of indicated datasets in ± 25 -kb windows centered around all 1477 CTCF peaks identified in WT ranked by ChIP occupancy. (Lanes 1 to 4) Presence of contact domain boundaries called in each genotype by TopDom in 2-kb bins around the peak center. (Lanes 5 to 10) Physical insulation score differences measured in genotype X (top) minus genotype Y (bottom) by Hi-C. (Lanes 11 and 12) CTCF or (Lanes 13 and 14) Cp190 ChIP occupancy in indicated genotypes. (Lanes 15 and 16) Differential CTCF and Cp190 ChIP occupancy in genotype X (top) minus genotype Y (bottom). (Lane 17) CTCF motif presence in DNA. (Lane 18) Expressed TSSs in WT 4- to 6-hour-old embryos (75). Summarized values (average physical insulation score or percentage of WT CTCF peaks with boundary/ChIP peak/differentially bound region/CTCF motif/transcribed TSS present) across 2-kb bins are shown above with indicated enrichments ± 2 kb around the central CTCF peak (highlighted in red on the x axis). **(B)** Example locus (dm6 coordinates) Hi-C maps and ChIP-seq tracks presented as in Fig. 1E. Arrowheads point to a CTCF and Cp190 cobound peak located at a domain boundary in WT; empty arrowheads indicate their absence in the mutants. ChIP-seq scale is reads per million. Cartoons summarize CTCF and Cp190 binding status and boundary presence/absence. RPKM, reads per kilobase per million reads.

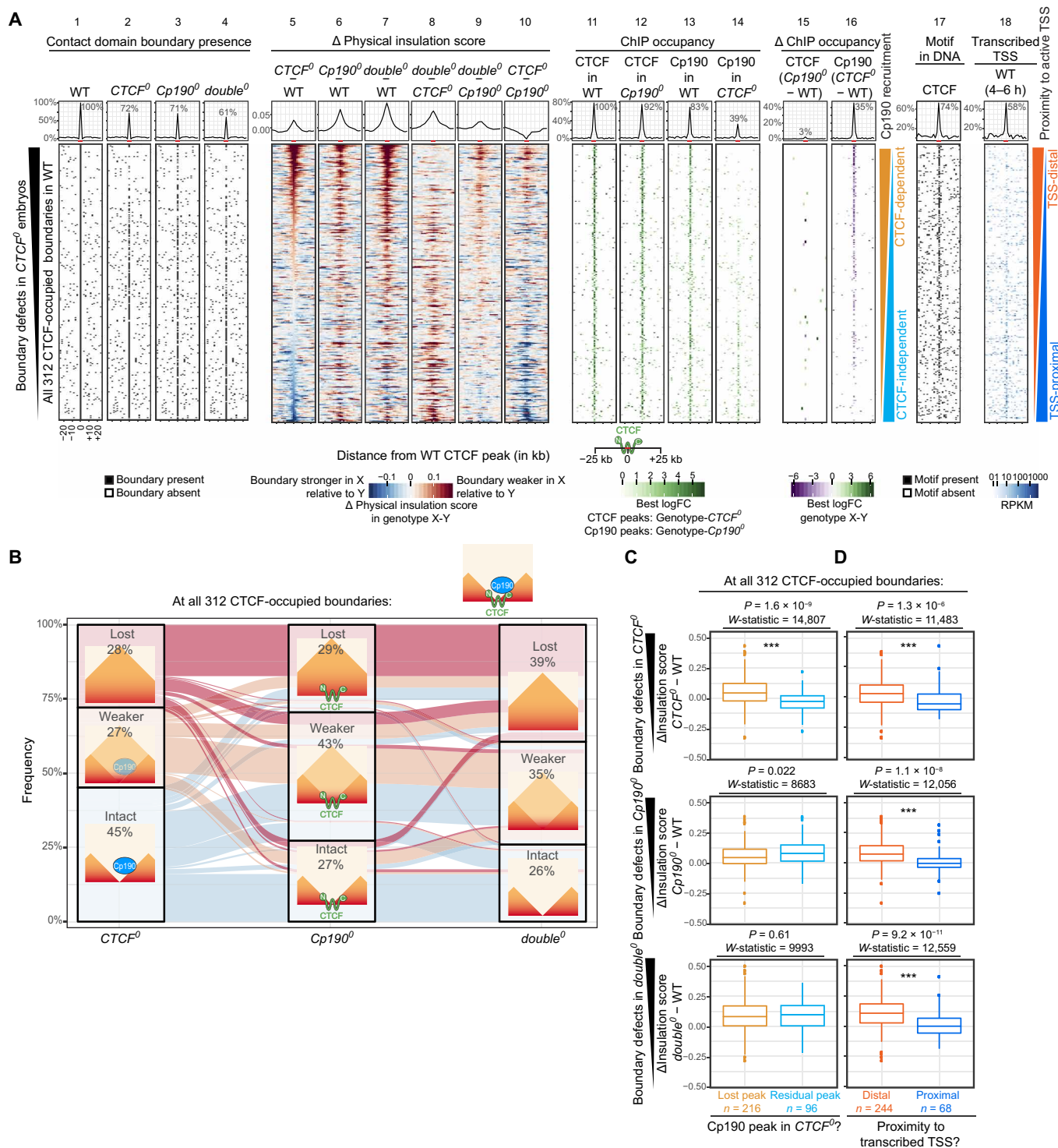


Fig. 3. CTCF-occupied boundaries are differentially sensitive to CTCF or Cp190 loss. (A) Same as Fig. 2A but for all 312 CTCF-occupied boundaries ranked by physical insulation defects in $CTCF^0$ relative to WT. **(B)** Alluvial plot of the same data as in (A). Lost boundaries are absent in the mutant but present in WT. Weaker and intact boundaries are retained in the mutant but have a delta physical insulation score in the mutant minus WT ≥ 0.01 or < 0.01 , respectively. Cartoons show typical CTCF and Cp190 ChIP occupancy in each scenario. **(C)** Box plots of data shown in (A). Physical insulation score differences in indicated mutants minus WT at all 312 CTCF-occupied boundaries at which Cp190 recruitment to the nearest CTCF peak to the boundary is strictly CTCF dependent (lost Cp190 peak in $CTCF^0$) or at least partially CTCF independent (residual Cp190 peak at the former CTCF peak in $CTCF^0$). P values and W -statistics from two-sided Wilcoxon rank sum test with continuity correction are indicated. Box plot: Center line, median; box limits, upper and lower quartiles; whiskers, 1.5 \times interquartile ranges; points, outliers; n = CTCF-occupied boundaries of indicated categories. **(D)** Same as (C) but for CTCF-occupied boundaries at which the nearest CTCF peak to the boundary is distal (further than ± 200 bp) or proximal (within 200 bp) to an actively transcribed TSS (RPKM > 0) in WT embryos (75). (C and D) *** indicate high statistical significance ($P \leq 0.0001$). The precise P values are shown in the figures.

Seventy-two to 74% of CTCF-occupied boundaries were lost or at least measurably weaker in *Cp190⁰* or *double⁰* mutants (Fig. 3B), indicating that Cp190 is required to form robust boundaries at many, but not all, CTCF-occupied boundaries. Boundary defects in all genotypes were significantly weaker at CTCF-occupied boundaries that were TSS-proximal than those that were TSS-distal [Fig. 3, A (lanes 5 to 7 and 18) and D, and fig. S3B], again suggesting that transcription-dependent mechanisms may redundantly form boundaries, as we observed for all Cp190-occupied boundaries (Fig. 1C).

An example locus illustrating these results is shown in fig. S3C. We conclude that Cp190 reinforces several CTCF-occupied boundaries independently of CTCF. CTCF-occupied boundaries to which Cp190 is recruited by additional factors other than CTCF are more sensitive to Cp190 loss because CTCF loss is compensated by redundant Cp190 recruitment.

Su(Hw) recruits Cp190 to a distinct subset of Cp190-dependent boundaries

Of boundaries occupied by Cp190 in WT and lost in *Cp190⁰*, only 40% are co-occupied by CTCF. Su(Hw) occupies a distinct subset of nonpromoter boundaries than CTCF (8) and directly recruits Cp190 to some of its binding sites (30). Some Cp190-dependent boundaries are enriched for Su(Hw) motifs (Fig. 1B), suggesting that Su(Hw) recruits Cp190 to these sites to form boundaries. To test this, we identified Su(Hw)-dependent Cp190 peaks. Cp190 ChIP-seq could not directly be performed on *su(Hw)⁰* mutant embryos lacking maternal and zygotic Su(Hw) because Su(Hw) is essential for female germline development (43). Instead, we performed Cp190 ChIP-seq in larval central nervous systems of *su(Hw)^{KO}* mutants with a deletion of the *su(Hw)* open reading frame and diluted maternal Su(Hw) and in WT and *Cp190^{KO}* mutants as controls (fig. S4A and data S10 to S12) and subsequently intersected Cp190 peaks identified in larval central nervous systems with Cp190-occupied boundaries in WT embryos. Of 1140 Cp190-occupied boundaries in WT embryos, 1125 (99%) overlapped a Cp190 peak in WT larval central nervous systems. Among these, 88 of 1125 (8%) did not overlap a Cp190 peak in *su(Hw)^{KO}* larval central nervous systems. Physical insulation defects in *Cp190⁰* embryos were significantly larger at boundaries overlapping Su(Hw)-dependent Cp190 peaks than at boundaries overlapping Su(Hw)-independent Cp190 peaks (fig. S4B). Together, these results suggest that Cp190 is recruited to independent sites by CTCF and Su(Hw) to form boundaries (fig. S4C).

Diverse Cp190 complexes exert similar enhancer-blocking activity in a reporter assay

Our finding that Cp190 associates with both promoter and non-promoter boundaries but is only required to form the latter (Fig. 1B) raises the question of whether Cp190 exerts different activities at different boundaries, possibly in the context of distinct multiprotein complexes. To test this, we first clarified the compositions of distinct Cp190-containing complexes since these complexes were previously purified from different sources using different protocols, precluding their direct comparison. We pulled down CTCF (7), Su(Hw), Chro, and Cp190 from the same batches of *Drosophila* embryonic nuclear extracts (data S13). Cp190 copurified with all expected insulator-binding proteins such as Ibf1, Ibf2, mod(mdg4), pita, CTCF, Su(Hw), BEAF-32, and ZIPIC (Fig. 4A) (25, 27, 31, 32). Cp190, CTCF, Su(Hw), and Chro pull-downs identified partially overlapping sets

of copurifying proteins, and all contained Cp190, Cp60 [Cp190's partner protein at centrosomes (44)], and CG1737 [which previously copurified with HP1a (heterochromatin protein 1) (45)] (Fig. 4A). Recombinant Cp190-Cp60 complexes directly interacted with CTCF C terminus, Su(Hw) N or C terminus, or full-length Ibf1 or 2, and CTCF directly interacted with Cp60 in addition to its previously known direct interaction with Cp190 (fig. S4, D and E) (7). Cohesin subunits (SMC1, SMC3, SA, and vtd) were specifically enriched in CTCF and Cp190 pull-downs (Fig. 4A). Proteins copurifying with CTCF, Su(Hw), or Chro generally colocalized with these proteins in published ChIP-seq experiments (Fig. 4B) (31, 33, 40, 46).

We then tested whether different Cp190 complexes assembled at separate loci exert enhancer-blocking activity in a quantitative insulator reporter assay (7). Test fragments (345 to 888 bp long, average 496 bp) were cloned in between an enhanced green fluorescent protein (EGFP) reporter and an enhancer, while an mCherry reporter present at a similar distance from the enhancer serves as a reference. Relative EGFP and mCherry intensities were measured in thousands of single transfected *Drosophila* S2 (Schneider's *Drosophila* Line 2) cells with a cell analyzer. Sites bound by CTCF, Su(Hw), or Chro+Pz+BEAF-32 specifically reduced EGFP fluorescence to varying degrees relative to the well-characterized *gypsy* insulator (Fig. 4C) (7, 47). We mutagenized two boundaries each containing two pairs of overlapping BEAF-32 motifs (48). Single point mutations in both BEAF-32 motif pairs had a stronger effect than mutating a single pair, indicating that each pair of overlapping BEAF-32 motifs contributes independently to insulator activity (Fig. 4C). Prior transgenic insulator reporter assays in flies concluded that only a subset of insulator protein-bound sites are insulators or that sites must be multimerized to reveal insulator function (30, 49–51). Insulator activity depends on chromatin context (30), and the robustness of our transiently transfectable reporter suggests that it is chromatin context independent. We conclude that Cp190 assembles into diverse multisubunit protein complexes bound at distinct genomic loci that exert similar enhancer-blocking activities in a reporter assay.

Cp190 prevents regulatory cross-talk between early patterning gene loci

Given its critical boundary function, we then investigated how Cp190 affects the expression of well-studied developmental genes. We first focused on the ANT-C locus comprising essential developmental genes and harboring several contact domain boundaries that were defective in *Cp190⁰* (Figs. 1E and 5, A and B, and fig. S5A). More specifically, we focused on the extended *Sex combs reduced* (*Scr*) locus because the spatial and temporal activity patterns of enhancers present in a 70-kb region around *Scr* have been systematically characterized in WT embryos (52, 53), enabling us to interpret gene misexpression phenotypes in *Cp190⁰* mutants with respect to local enhancers. *Scr* is a HOX gene conferring segmental identity to specific anterior body segments, and its neighboring gene *fushi tarazu* (*ftz*) is a pair-rule homeodomain gene required to help segment the very early embryo (54). *Scr* and *ftz* are expressed in independent spatiotemporal patterns in early embryos. *ftz* is expressed in seven equally spaced stripes beginning after zygotic genome activation, while *Scr* is expressed later in early gastrulae in a band of cells that partially overlaps the first *ftz* stripe (Fig. 5C). *ftz* expression is driven by stripe enhancers contained within the *ftz* contact domain (numbered 2 to 5 in Fig. 5B) (52, 55, 56). This domain interrupts a larger contact domain containing *Scr* and its enhancers, including a

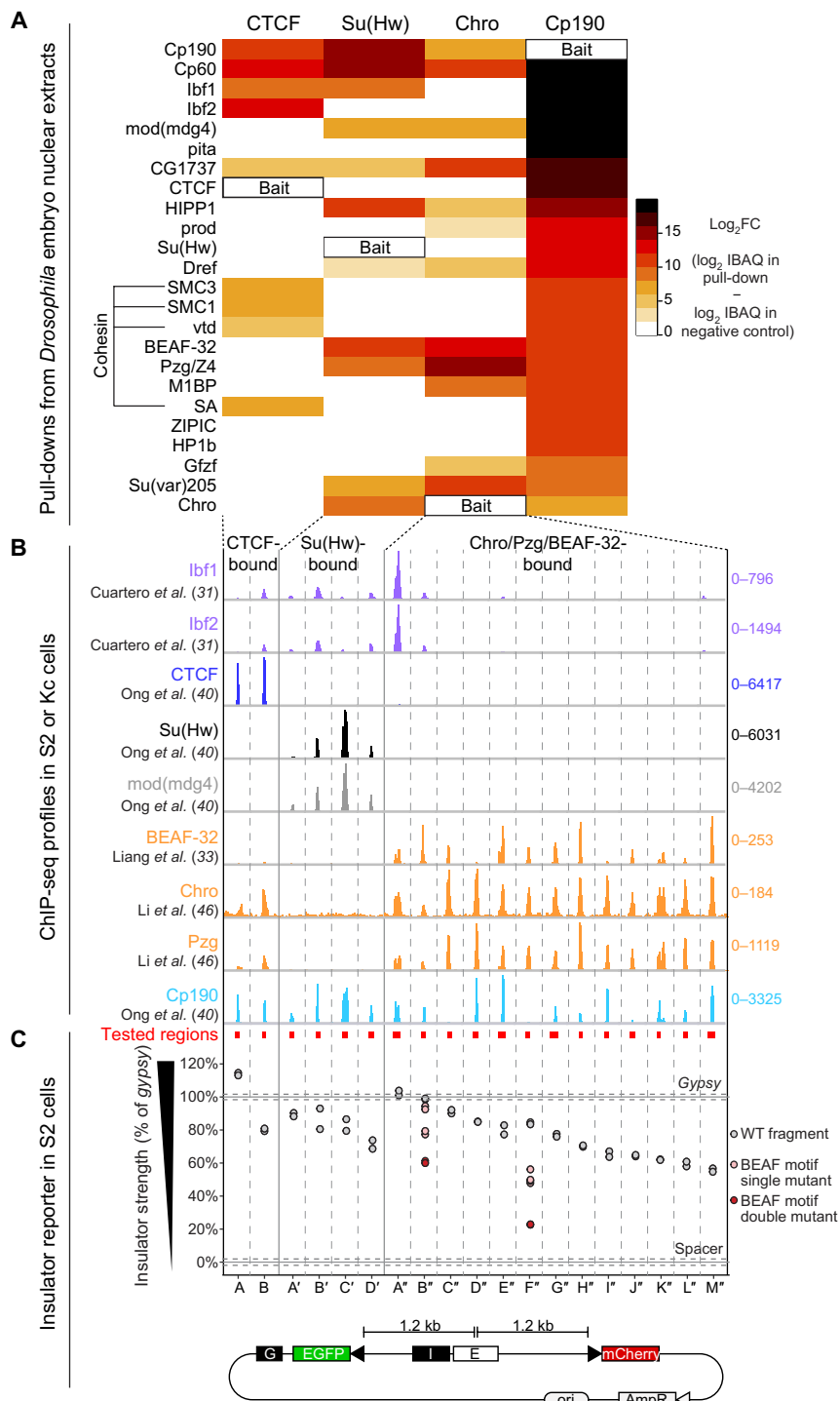


Fig. 4. Cp190 complexes exert similar enhancer-blocking activity in a quantitative reporter assay. (A) Enrichments of indicated proteins (rows) in pull-downs with indicated GFP-tagged baits (columns, Su(Hw)[1-219], CTCF[1-293], Chro[613-926], and full-length Cp190) from the same batches of embryo nuclear extracts, analyzed by mass spectrometry (MS) and ranked by their specific enrichment in the Cp190 pull-down. Scale bar indicates log₂ fold change (FC) of average intensity based absolute quantification (iBAQ) values of biological duplicate pull-downs over negative control pull-downs with unrelated GFP-tagged bait. CTCF pull-downs were previously described (7). **(B)** Published ChIP-seq profiles in S2 or Kc cell lines of indicated insulator proteins ±1 kb around the cloned genomic fragments (345 to 888 bp long indicated by red boxes, loci separated by vertical lines). Scales show total counts. **(C)** Insulator strengths of cloned genomic fragments measured in S2 cells transiently transfected with reporters with indicated I (insulator) test fragments, expressed as percentage of *gypsy* insulator strength (set to 100%). Insulators block EGFP activation by the enhancer (E). A *gypsy* insulator (“G”) blocks EGFP activation by the enhancer from the left. Fragments were tested in biological duplicates (dots). Horizontal lines show average values obtained with *gypsy* or a neutral spacer (*n* = 8 biological replicates); dotted lines show SDs. Tested fragments were bound by CTCF (A and B), Su(Hw) (A’ to D’), or Chro+Pzg+BEAF-32 (A” to M”). Single (pink dots) or two (red dots) pairs of non-overlapping BEAF motifs in B” and F” were mutated.

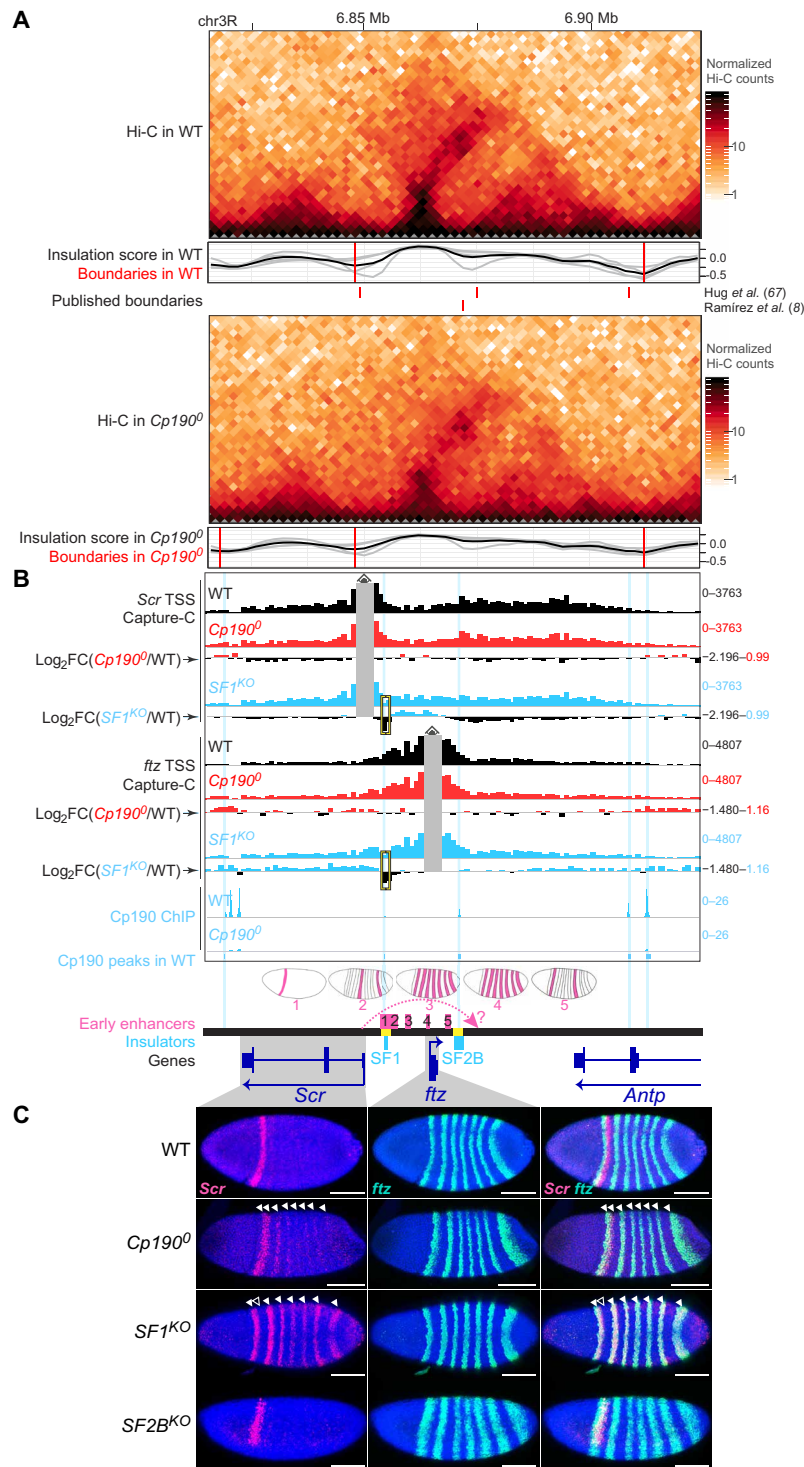


Fig. 5. Cp190 prevents regulatory cross-talk between early patterning gene loci. (A) Extended *ftz* locus (dm6 coordinates) Hi-C maps presented as in Fig. 1E. **(B)** NG Capture-C profiles (1-kb resolution) around *Scr* or *ftz* TSS viewpoints showing average normalized reads (in reads per million) of biological triplicates excluding bins ± 2 kb around the viewpoint (gray). Differential Capture-C profiles in mutant versus WT are shown as log₂ fold change profiles obtained from diffHic. Yellow brackets mark the deleted boundary in *SF1^{KO}*. WT Cp190 ChIP peaks are highlighted in blue (two prominent Cp190 ChIP signals in WT overlap a blacklisted region and were not called peaks). Early stripe enhancers with schematized expression patterns are numbered 1 to 5 (references in Materials and Methods). Dotted arrow shows *Scr* regulation by a hypothetical distal enhancer (question mark). *SF1^{KO}* and *SF2B^{KO}* deletions are yellow. **(C)** RNA-fluorescence in situ hybridization (FISH) with cohybridized antisense probes against *Scr* (red) and *ftz* (green) mRNAs in 4',6-diamidino-2-phenylindole (DAPI)-stained early gastrula embryos (anterior left; posterior right; scale bars, 100 μ m; merged images on the right). In *Cp190⁰*, *Scr* is expressed in its WT stripe and in ectopic *ftz* stripes (filled arrowheads). In *SF1^{KO}* embryos, *Scr* is lost in its WT stripe (empty arrowhead) and only expressed in *ftz* stripes (filled arrowheads). In *SF2B^{KO}* embryos, *Scr* and *ftz* expression seem normal (embryo rotation reveals a normal ventral gap in *Scr* expression).

putative distal *Scr* enhancer located downstream of *ftz* (question mark in Fig. 5B) (52). Two characterized insulators named *Scr-ftz* (SF) boundaries SF1 and SF2B (57, 58) overlap *ftz* contact domain boundaries (Fig. 5B). According to a published model, SF1 and SF2B pair and thereby “loop” *ftz* out of the *Scr* domain to prevent *Scr-ftz* regulatory cross-talk and enable the *Scr* promoter to skip the intervening *ftz* domain and reach its putative distal enhancer (shown as a dotted arrow in Fig. 5B) (58, 59). To clarify how SF boundaries function and understand how Cp190 contributes to boundary function, we analyzed boundary defects at higher resolution and examined *Scr* and *ftz* expression in *Cp190⁰* mutants and in embryos carrying deletions of SF1 (*SF1^{KO}*) or SF2B (*SF2B^{KO}*).

Cp190 binds to *ftz* contact domain boundaries in WT, and *ftz* boundaries were slightly weakened in *Cp190⁰* Hi-C maps (Fig. 5A and fig. S5A). Interdomain contacts between *Scr* and *ftz* contact domains were not significantly increased in higher-resolution next-generation (NG) Capture-C (60) experiments on 2- to 6-hour-old WT and *Cp190⁰* embryos with viewpoints in *Scr* and *ftz* TSSs (Fig. 5B and data S14). In *Cp190⁰* mutants, *Scr* was expressed in its endogenous stripe and in seven stripes overlapping *ftz* expression (Fig. 5C). This suggests that *ftz* enhancers are able to ectopically activate *Scr* transcription upon Cp190 loss despite retention of a physical boundary.

Deletion of SF1 boundary DNA led to stronger interdomain contacts between *Scr* TSS and *ftz* contact domains than Cp190 loss, as revealed by simultaneous NG Capture-C on 2- to 6-hour-old *SF1^{KO}* embryos (Fig. 5B and data S15). In contrast to *Cp190⁰*, *SF1^{KO}* embryos lost *Scr* expression in its endogenous anterior stripe, and *Scr* expression was completely replaced by the *ftz* pattern (Fig. 5C) as recently described (59). This result had previously been interpreted to support the model in which SF1-SF2B pairing is required to bridge *Scr* to its putative distal enhancer downstream of *ftz* (58, 59). Inconsistent with this model, however, *Scr* expression was normal in *SF2B^{KO}* mutant embryos (Fig. 5C). Instead, we found that *SF1^{KO}* embryos likely lose endogenous *Scr* expression because SF1 deletion concomitantly deletes an enhancer that we noticed was active in an early *Scr*-like stripe (labeled 1 in Fig. 5B and fig. S5B) (53). We conclude that both Cp190 protein and SF1 DNA critically form a regulatory boundary, ensuring independent regulation of *Scr* and *ftz*, but neither Cp190 nor SF1-SF2B pairing is required for endogenous *Scr* expression. Stronger *Scr* misexpression in *ftz* stripes observed in *SF1^{KO}* than in *Cp190⁰* (Fig. 5C) correlates with stronger *Scr* TSS-*ftz* interdomain contacts observed upon SF1 deletion than upon Cp190 loss (Fig. 5B). We note that we do not know how Cp190 is recruited to *ftz* boundaries, as they are not CTCF- or Su(Hw)-dependent Cp190 peaks and *CTCF⁰* mutants did not show any contact domain boundary or gene misexpression defects at this locus (figs. S5, A and C).

Cp190 is dispensable for HOX gene activation by long-range enhancers

ftz stripe enhancers are only active in early embryos, and *Cp190⁰* and *SF1^{KO}* older embryos no longer ectopically expressed *Scr* in stripes (fig. S5D). Instead, older *Cp190⁰* embryos misexpressed *Scr* in the hindgut and anal plate (Fig. 6, A to C). Near-complete characterization of embryonic enhancers in the extended *Scr* locus previously identified a single hindgut and anal plate enhancer 30 kb downstream of the *Scr* promoter that could activate transcription from the *Scr* promoter in a transgene (labeled 7 in Fig. 6B) (52, 53). Cp190 normally binds to a contact domain boundary separating

this enhancer from the *Scr* promoter, and both Hi-C (Fig. 6A) and NG Capture-C (Fig. 6B) revealed qualitatively weakly increased contacts in broad contiguous regions across former Cp190 peaks in *Cp190⁰* mutants. This strongly suggests that in *Cp190⁰* mutants, the *Scr* promoter is ectopically activated by a long-range enhancer from which it was formerly insulated. Cp190 is therefore required to insulate *Scr* from noncognate enhancers but not to bridge *Scr* to distal enhancers (summarized in Fig. 6D).

We also examined the expression of additional ANT-C HOX genes other than *Scr* in WT and *Cp190⁰* embryos (fig. S6A). Expression of *Antennapedia* (*Antp*) was normal in *Cp190⁰* embryos (fig. S6B). In contrast, *Deformed* (*Dfd*) was strongly ectopically expressed in the nervous system of *Cp190⁰* mutants in addition to being expressed in its endogenous pattern (fig. S6B). Several neuronal enhancers have been annotated both within and flanking the *Dfd* contact domain, and we are not able to hypothesize which of these may be the culprit enhancer driving *Dfd* misexpression in *Cp190⁰* mutants. Nevertheless, as in the case of *Scr*, ectopic *Dfd* expression is overlaid onto its endogenous expression pattern in *Cp190⁰* mutants. Abdominal HOX genes of the bithorax complex (BX-C) are expressed from a separate locus than ANT-C (fig. S7A). These genes are controlled by body segment-specific enhancers delimited by boundaries that maintain the independence of these enhancer domains (61, 62). Expression of *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*) was mostly normal in *Cp190⁰* mutants (fig. S7B). Therefore, Cp190 is not essential for abdominal HOX gene activation by their long-range enhancers (over more than 50 kb in the case of the *iab-5* enhancer domain driving *Abd-B* transcription in parasegment 10). Graded expression of *Ubx* and *abd-A* was, however, somewhat altered, suggesting that enhancer domains were inadequately insulated from each other (fig. S7B). These effects were subtle compared to the more severe phenotypes of BX-C boundary deletions (63, 64), revealing that redundant mechanisms maintain the independence of BX-C regulatory domains. Together, we conclude that in all cases examined (*Scr*, *ftz*, *Antp*, *Dfd*, *Ubx*, *abd-A*, and *Abd-B*), developmental regulator genes were expressed in their endogenous patterns and, in some cases, in additional cells upon Cp190 loss.

Cp190 is required for enhancer-blocking but not long-range pairing by the Homie insulator

If Cp190 is indeed required for enhancer-blocking but not distal enhancer-facilitating functions of insulators, these two functions should be differentially sensitive to Cp190 loss. We tested this hypothesis using the classical Homie insulator known to support one of the longest-range enhancer-promoter contacts described in flies (36, 37). Homie overlaps a Cp190-occupied contact domain boundary downstream of *eve*, but Cp190 was not visibly required for formation of this boundary (fig. S8).

Published Homie transgenes contain divergently transcribed reporter genes (*GFP* and *LacZ*) and are integrated 142 kb from *even-skipped* (*eve*), another pair-rule homeodomain gene similar to *ftz* (37). Local *hebe* gene enhancers close to the transgene integration site activate reporter gene expression in neurons of midstage WT embryos, except when Homie is present in between and specifically shields *LacZ* from *hebe* enhancers (Fig. 7A). In WT animals, transgenic Homie physically pairs with endogenous Homie in a head-to-head orientation and supports long-range reporter gene activation by *eve* enhancers active in anal plate, cardiac mesoderm, and specific neurons of mid-stage embryos (36, 37) (Fig. 7A). When

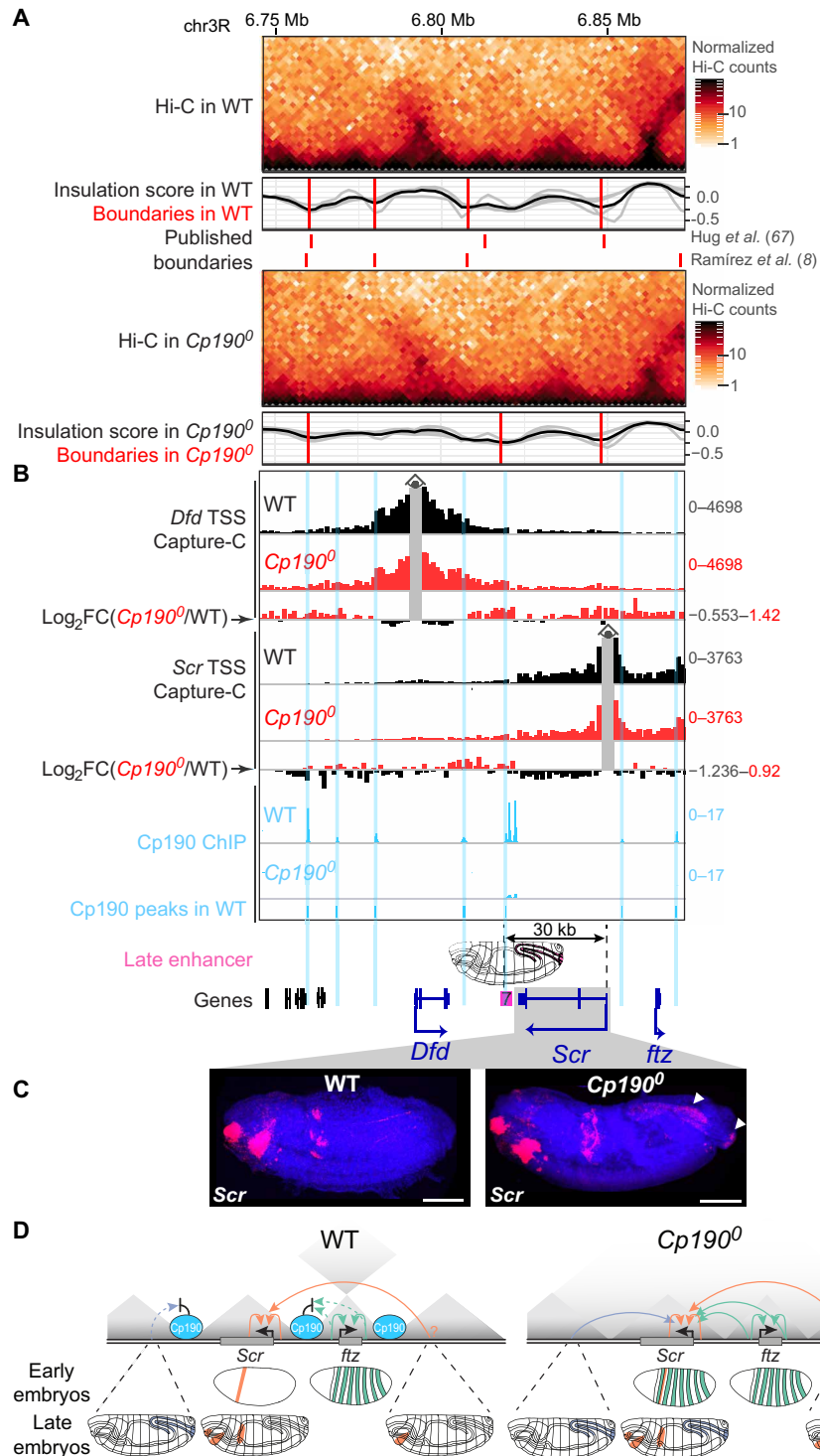


Fig. 6. Cp190 is dispensable for ectopic *Scr* activation by a long-range enhancer. (A) Similar to Fig. 5A but showing contact domains downstream of *Scr*. **(B)** NG Capture-C profiles presented as in Fig. 5B but around *Dfd* or *Scr* TSS viewpoints in indicated genotypes. Enhancer 7 drives schematized reporter gene expression in the hindgut and anal plate of older embryos in transgene assays (52) and is separated from the *Scr* promoter (30 kb away) by Cp190 ChIP peaks. **(C)** RNA-FISH with antisense probes (red) against *Scr* mRNA in late-stage (stage 16) DAPI-stained embryos (anterior left; posterior right; scale bars, 100 μ m). *Scr* is normally expressed in labial and prothoracic segments and the anterior midgut and is additionally expressed in the hindgut and anal plate (left and right arrowheads) of *Cp190⁰* mutants. **(D)** Summarized *Scr* misexpression phenotypes in *Cp190⁰* early and late embryos. Effective (solid arrows) or blocked (dotted arrows) transcriptional activation of promoters by indicated enhancers is shown (hindgut and anal plate enhancer in blue; *Scr* enhancers in orange including a putative distal enhancer marked by a question mark; *ftz* enhancers in green). In *Cp190⁰* embryos, *Scr* is activated by its endogenous enhancers and additionally by formerly insulated enhancers, resulting in cumulated expression patterns.

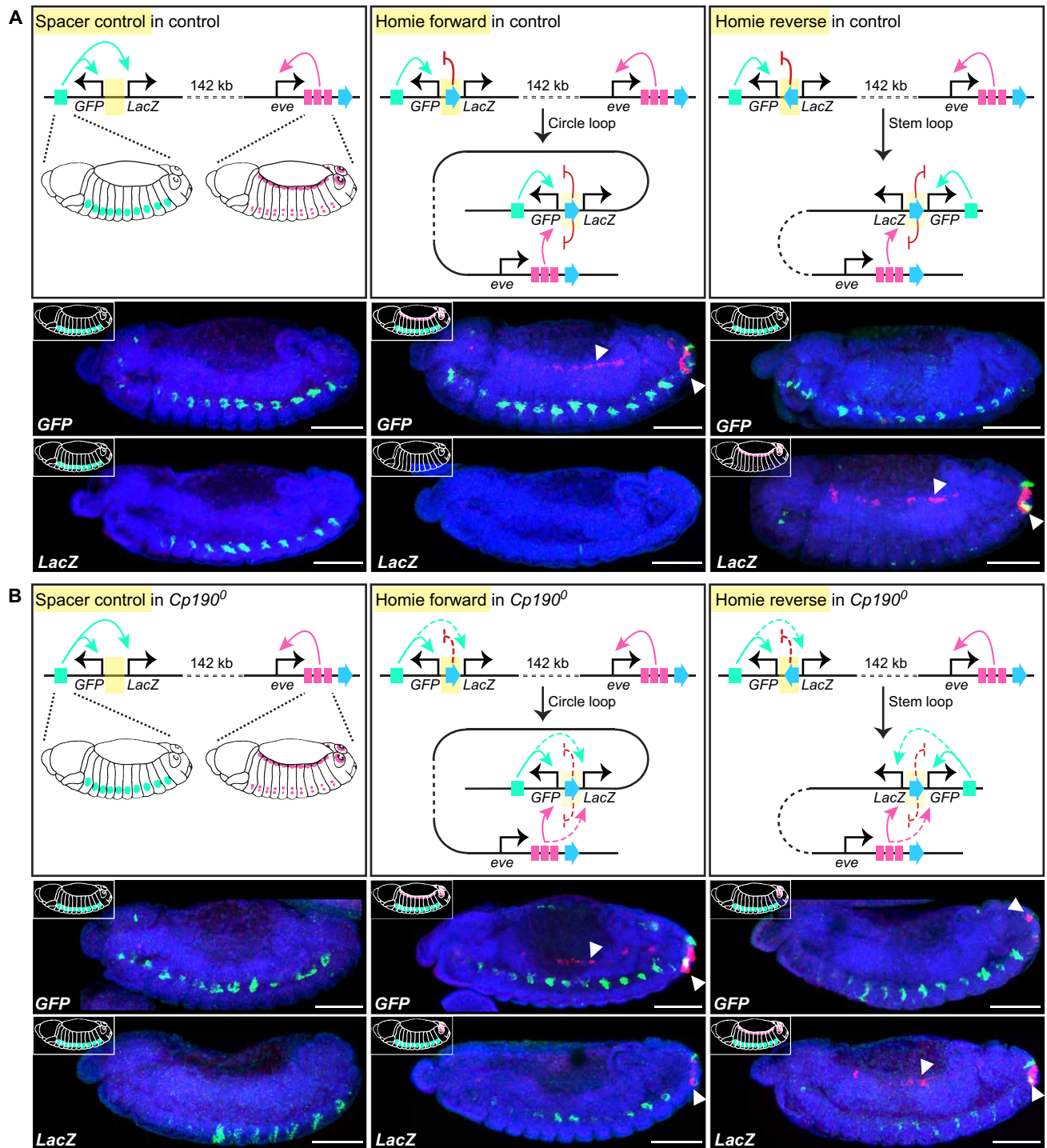


Fig. 7. Cp190 supports enhancer-blocking but not long-range pairing by Homie. (A) WT expression of Fujioka *et al.* (37) transgenes with divergently transcribed *GFP* and *LacZ* reporter genes, integrated 142 kb upstream of *eve* in the vicinity of local *hebe* enhancers (green). *eve* endogenous enhancers (pink) are respectively active in anal plate, cardiac mesoderm, and specific neurons. When Homie insulator is between *GFP* and *LacZ*, it pairs in a head-to-head orientation with endogenous Homie downstream of *eve*, leading to schematized *GFP* and *LacZ* expression patterns. Below, RNA-FISH with antisense probes against *GFP* (top) or *LacZ* (bottom) mRNAs in midstage (stage 13) DAPI-stained control embryos with Cp190 (anterior left; posterior right; scale bars, 100 μ m). RNA-FISH signal was false-colored green or pink when it was respectively detected in a deep (showing *hebe* enhancer-driven neuronal expression) or surface (showing cardiac mesoderm and anal plate expression marked by arrowheads) confocal slice. Note that neuronal expression driven by *hebe* enhancers masks that driven by *eve* enhancers, and anal plate signal is visible in all confocal slices. (B) Expression of same transgenes in *Cp190⁰*. Homie still supports long-distance reporter gene activation by distal *eve* enhancers (arrowheads) but is a weaker enhancer-blocker (as seen both by *LacZ* activation by *hebe* enhancer and activation of *LacZ/GFP* in the Homie forward/reverse transgenes, respectively, by the *eve* anal plate enhancer, although Homie is still able to block the *eve* cardiac enhancer).

transgenic Homie is cloned in the same orientation as endogenous Homie (called forward orientation), it forms a circle loop enabling *GFP* activation by both local *hebe* and distal *eve* enhancers while preventing *LacZ* activation (37) (Fig. 7A). When transgenic Homie is cloned in the opposite reverse orientation, it forms a stem loop enabling *LacZ* activation by distal *eve* enhancers while ensuring that *GFP* is only activated by its nearby *hebe* enhancers (Fig. 7A) (37).

When the same Homie and control transgenes were introduced into *Cp190⁰* mutants, reporter genes were still activated by *eve* long-range enhancers with similar efficiencies as in WT (Fig. 7B), revealing that Cp190 is not essential for Homie pairing. *GFP* and *LacZ* reporters were, however, expressed in partially overlapping patterns in *Cp190⁰*, indicating that Homie's enhancer-blocking activity partially relies on Cp190. Concretely, *LacZ* was activated (albeit more weakly than *GFP*) by *hebe* enhancers from which it was formerly insulated (Fig. 7B). In addition, Homie pairing enabled both *GFP* and *LacZ* activation by the *eve* anal plate enhancer in *Cp190⁰* mutants, although Homie still blocked activation of *LacZ* (in Homie forward) and *GFP* (in Homie reverse) by the *eve* cardiac mesoderm enhancer (Fig. 7B). The anal plate enhancer is known to activate Homie transgenes located at much larger distances from *eve* (up to 2 Mb away) than the cardiac mesoderm enhancer, potentially suggesting that the anal plate enhancer is stronger and thus requires fully functional Homie to be blocked (37).

These results demonstrate that enhancer-blocking and enhancer-pairing functions traditionally ascribed to the classical Homie insulator are separable and reveal that Cp190 is only required for efficient enhancer blocking. Notably, these results were obtained in embryos heterozygous for Homie transgene, arguing that transvection could not influence the result.

DISCUSSION

Cp190 was hypothesized 17 years ago to organize the genome into chromosomal loops and thereby ensure gene regulation specificity (25). Here, we tested this model by analyzing *Drosophila* completely lacking Cp190, CTCF, and both factors. We reached the following conclusions: (i) Cp190 is critical for early development (Fig. 1A). (ii) Cp190 is required to form most promoter-distal boundaries but is dispensable to form promoter-proximal boundaries (Fig. 1). (iii) Cp190 is recruited to CTCF-dependent boundaries and is required for their formation (Fig. 2). (iv) While Cp190 is strictly recruited by CTCF to some of these boundaries, it reinforces other CTCF-occupied boundaries CTCF-independently (Fig. 3). (v) Cp190 assembles into diverse multisubunit complexes that share similar enhancer-blocking activity in a quantitative insulator reporter assay in transfected cells (Fig. 4). (vi) Cp190 critically insulates the HOX gene *Scr* from inappropriate enhancers located up to 30 kb away from the *Scr* promoter (Figs. 5 and 6). (vii) In contrast, Cp190 was largely dispensable for activation of HOX genes by distal enhancers (Fig. 6 and figs. S6 and S7). (viii) Cp190 is similarly only critical for the enhancer-blocking activity of the classical Homie insulator but not for Homie pairing. Below, we discuss how this work advances our understanding of how contact domain boundaries are formed and affect transcriptional regulation.

Diversity of boundary-forming mechanisms

Drosophila contact domains frequently align with active/inactive compartmental domains, raising the question of whether fly contact

domains are formed directly by architectural proteins assembled at boundaries or indirectly by transcription-related processes (1, 7, 65). Our studies of *Drosophila* completely lacking Cp190 or CTCF demonstrate that these proteins form a subset of domain boundaries that are distal to sites of transcription (Figs. 2 and 3) (7). At least two distinct mechanisms of boundary formation therefore exist, one relying on architectural proteins and the other correlating with transcribed promoters (1, 66).

Part of this study focused on CTCF peaks and revealed three lines of evidence that Cp190 promotes boundary formation at CTCF peaks: (i) Only CTCF peaks colocalizing with Cp190 are present at domain boundaries (Fig. 2A). (ii) Most CTCF-occupied boundaries are lost or weakened in *Cp190⁰*, although CTCF remains bound (Figs. 2 and 3). (iii) Residual Cp190 binding at former CTCF-occupied boundaries coincides with boundary retention in *CTCF⁰* mutants (Fig. 3) (7). We do not know why some CTCF-occupied boundaries were unexpectedly reinforced in *CTCF⁰* mutants relative to WT (Fig. 3, A and C). Even boundaries not bound by CTCF were stronger in *CTCF⁰* relative to WT (fig. S2A). A hypothetical explanation is that boundary strength is somehow redistributed to the remaining Cp190-dependent boundaries in the absence of CTCF.

To summarize, whereas CTCF forms a large fraction of mammalian contact domains by directly blocking extruding cohesin, our results support the notion that flies use Cp190 as an adaptor protein recruited DNA sequence-specifically by proteins such as CTCF and Su(Hw) to form robust physical boundaries at these sites. Cp190 is therefore more widely required for boundary formation than individual DNA binding proteins such as CTCF.

Seventy-eight percent of contact domain boundaries are retained in *Cp190⁰* mutants (Fig. 1B). How are these boundaries formed? (i) At promoter boundaries, it is still unclear what drives boundary formation: transcription itself, active chromatin modifications, RNA polymerase II, promoter-associated factors, or insulator proteins at promoter boundaries such as BEAF-32 (1, 46, 66, 67). Our results do not support a model in which Cp190 drives promoter boundary formation (24). (ii) At nonpromoter boundaries such as SF1 that was partially retained in *Cp190⁰*, deletion of SF1 boundary DNA more strongly increased contacts between flanking contact domains than Cp190 loss (Fig. 5B), revealing that Cp190 is less important than other SF1-associated factors to form this boundary.

Cp190 prevents promiscuous gene regulation at tested loci

We previously reported that CTCF and Cp190 co-regulate a subset of genes near CTCF-dependent boundaries (7). We did not know which regulatory elements were driving gene misexpression and hence could not say whether misregulation arose from regulatory cross-talk between formerly insulated loci. By assessing how Cp190 loss affects gene expression in the best-characterized *Drosophila* developmental loci, we found that *Scr* was ectopically expressed in patterns that could be predicted on the basis of annotated enhancers located up to 30 kb away from its promoter (Figs. 5 and 6). *Scr* misexpression patterns in *Cp190⁰* evolved dynamically during embryogenesis, reflecting changing enhancer activities. *Scr* was ectopically activated by *ftz* enhancers despite retention of *ftz* contact domain boundaries.

These results, together with our findings that all tested Cp190 binding sites exert insulator activity in a reporter assay (Fig. 4), and that Cp190 is required for efficient Homie enhancer-blocking activity

(Fig. 7), all consolidate the original notion that Cp190 is critical for gene insulation. A well-understood function of genetic insulators is the regulation of abdominal HOX gene expression along the anterior-posterior body axis by ensuring that segment-specific regulatory domains containing HOX gene enhancers and silencers act independently (61, 68). These HOX insulators coincide with contact domain boundaries between regulatory domains, and boundary deletion results in contact domain fusion (65). We were therefore surprised that expression of abdominal HOX genes and contact domain boundaries themselves were mildly affected in *Cp190⁰* mutants compared to such boundary deletions (fig. S7) (63, 64). Moreover, *Abd-B* is misexpressed in *CTCF⁰* but not *Cp190⁰* embryos (fig. S7B) (69). This indicates that other factors are able to exert genetic insulation independently of Cp190 at abdominal HOX gene boundaries. It also remains to be determined how widely Cp190 protects other genes from inappropriate regulation.

Cp190 is not essential for HOX gene and Homie-mediated distal activation

Drosophila insulators are traditionally thought to form chromosomal loops exerting seemingly contradictory effects, both blocking and facilitating regulatory element-promoter communication by respectively segregating or connecting these elements. SF1 and SF2B boundary pairing was thus proposed to shield *Scr* from *ftz* enhancers and bridge *Scr* to its putative distal enhancer (58, 59). We instead found that both Cp190 and SF1-SF2B boundary pairing are dispensable for *Scr* embryonic expression (Fig. 5C). The putative distal *Scr* enhancer is located 25 kb upstream of the *Scr* promoter (fig. S5B), and its relevance for *Scr* transcriptional activation remains uncertain, but we find that an enhancer located even further away (30 kb downstream of the *Scr* promoter) is able to activate *Scr* transcription in *Cp190⁰* mutants (Fig. 6). Our finding that abdominal HOX genes are also expressed in patterns normally driven by their long-range enhancers in *Cp190⁰* mutants further suggests that Cp190 is not essential for long-distance enhancer-promoter pairing at these loci. Consistently, the abilities of abdominal HOX boundaries to support long-distance activation of HOX promoters by their distal enhancers was recently suggested to rely on uncharacterized factors other than insulator proteins (70).

We further demonstrate that Cp190 is dispensable for Homie's ability to mediate transcriptional activation by distal enhancers 142 kb away (Fig. 7). We suggest that Cp190 is not a "looping factor" critical for distal enhancer-promoter pairing. However, we have not assessed the relevance of Cp190 for fostering enhancer-promoter communication at other loci and thus cannot exclude that Cp190 supports long-range regulation of other genes, for example, within a contact domain by bringing promoters and their cognate enhancers into enhanced three-dimensional proximity.

Molecular basis of Cp190 function

CTCF, Su(Hw), and BEAF-32 colocalize with Cp190 at only a subset of their respective binding sites (Fig. 2A) (30, 42). Whether Cp190 colocalization with these proteins is regulated or instead dictated by the underlying DNA sequence is debated (29, 30, 40, 42, 71). We did not detect differentially enriched DNA motifs in our set of embryonic CTCF standalone versus CTCF+Cp190 cobound sites, and we do not know why CTCF standalone sites have intermediate ChIP occupancy in embryos (Fig. 2A). Standalone sites may exert different activities than sites cobound by Cp190. By testing a few

Su(Hw) sites in a transgenic insulator assay, Su(Hw) standalone sites were proposed to be repressors unlike Su(Hw)+Cp190 cobound sites (30). Similarly, CTCF standalone sites were proposed to lack insulator activity unlike some CTCF+Cp190 cobound sites (30). We now show that Cp190 imparts physical boundary activity to sites to which it is recruited by CTCF or Su(Hw) (Fig. 2A and fig. S4, B and C). This activity may underlie co-regulation of some genes near CTCF-dependent boundaries by both CTCF and Cp190 (7).

How does Cp190 form boundaries? Two main models were proposed to explain how insulator proteins fold chromosomes: (i) by pairwise looping between contact domain boundaries or (ii) by stalling loop-extruding cohesin at contact domain boundaries. The first model originally proposed that Cp190 interacts with DNA-bound insulator proteins through its C-terminal domain and dimerizes with distal Cp190-bound sites via its BTB domain (33, 39). It later became clear, however, that Cp190 BTB interacts directly with, for example, CTCF and Su(Hw) (7, 72). On the other hand, whether insulator proteins stall loop-extruding cohesin in flies is still debated (7), and cohesin has not yet been shown to play a major role in fly contact domain formation (1, 46). We find that both CTCF and Cp190 copurify with cohesin, but we do not know whether Cp190 interacts with cohesin independently of CTCF (7).

Not all CTCF binding sites in mammalian cells are associated with physical boundaries, and the contribution of non-CTCF proteins to boundary reinforcement has recently been explored (73). Our finding that Cp190 is recruited to DNA-bound CTCF to reinforce boundaries in *Drosophila* highlights that it will be interesting to further investigate whether analogous mechanisms are deployed across species and locus-specifically within a species.

MATERIALS AND METHODS

Drosophila melanogaster crosses

Using the same genetic strategy used to generate *CTCF⁰* (69) and *Cp190⁰* (7) mutants, *double⁰* mutants were generated for this study by recombining knockout mutations of the entire open reading frames of *CTCF* and *Cp190* and rescuing the double knockout animals by excisable FRT (flippase recognition target)-flanked genomic *CTCF* and *Cp190* rescue fragments. These rescue fragments were excised from the germ lines of conditionally rescued mothers and fathers expressing FLP (Flippase) under the control of *nanos* regulatory sequences. WT embryos with a matched genetic background were used as control in all Hi-C, NG Capture-C, and ChIP-seq experiments. Similar to *double⁰* embryos, WT embryos were generated by excising the same FRT-flanked genomic *CTCF* and *Cp190* rescue fragments from the germ lines of mothers and fathers expressing FLP enzyme under the control of *nanos* regulatory sequences, but these flies were WT for *CTCF* and *Cp190*.

su(Hw)^{KO}, *SF1^{KO}*, and *SF2B^{KO}* mutants were generated by CRISPR-Cas9-mediated genome editing using two single guide RNAs flanking the regions chosen for deletion: 3157 bp of the entire *su(Hw)* open reading frame (dm6 coordinates chr3R:14307954-14304798) for *su(Hw)^{KO}*, 2041 bp (dm6 coordinates chr3R:6853644-6855684) for *SF1^{KO}*, or 2122 bp (dm6 coordinates chr3R:6869630-6871751) for *SF2B^{KO}*. Guide RNAs were cloned into pCFD3 (Addgene, 49410). One-kilobase left and right homology arms were cloned into pHD-DsRed-attP vector (Addgene, 51019) for homology-directed repair leading to the integration of a *DsRed* fluorescent selection marker in each knockout allele. Primers used for cloning guide RNAs and

homology arms of the donor plasmid are provided in table S2. Both guide RNA plasmids and the homology repair plasmid were injected into flies expressing Cas9 in their germ line (*nanos-Cas9*).

Homie and control transgenes inserted 142 kb upstream of *eve* originally described in figure 3 of Fujioka *et al.* (37) were introduced into the *Cp190⁰* mutant background by recombining them onto the same second chromosome also harboring the FLP transgene.

Drosophila viability tests

Three sets of between 60 and 90 embryos of desired genotypes were aligned on a glass coverslip and vertically inserted into a fly culture vial. Vials were placed at 25°C, and unfertilized eggs and hatched larvae were counted 2 days later. The vials were later scored for the numbers of pupae and adult flies that completely emerged from the pupal case. The numbers of hatched embryos, pupae, and adults were counted in the triplicate experiments for each genotype.

Western blotting

For Western blotting presented in fig. S1A, 6- to 10-hour embryos were dechorionated, homogenized in SDS sample buffer, sonicated for 10 cycles (30 s on and 30 s off) in a Bioruptor on high-intensity settings, and centrifuged. The supernatants were loaded on a 4 to 12% acrylamide gel and probed with rabbit anti-CTCF-N diluted 1:2000, rabbit anti-full-length Cp190 diluted 1:2000 (7), and mouse anti-tubulin clone B-5-1-2 (Sigma-Aldrich, T5168) diluted 1:10,000. Chemiluminescence pictures of nitrocellulose membranes were imaged in Fiji v2.1.0/1.53c.

For Western blotting presented in fig. S2G, 40 third-instar larval central nervous systems per biological replicate were dissected in ice-cold phosphate-buffered saline (PBS). Samples were sonicated in 100 μ l of 20 mM tris (pH 7.5), 500 mM NaCl, 0.1% Triton X-100, and 1 \times cOmplete protease inhibitors (Roche) in a Bioruptor on high-intensity settings for 5 min at 4°C. Extracts were centrifuged for 5 min at maximum speed, and total protein was quantified by Qubit protein assay (Thermo Fisher Scientific). Calibrated amounts of extract from *CTCF⁰* animals rescued by TAP (Tandem Affinity Purification)-tagged transgenic versions *CTCF^{WT}*, *CTCF^{ΔN}*, or *CTCF^{ΔC}* were loaded on a 4 to 12% acrylamide gel and probed with rabbit peroxidase anti-peroxidase antibody complex (Sigma-Aldrich, P1291) diluted 1:2000 and mouse anti-tubulin clone B-5-1-2 (Sigma-Aldrich, T5168) diluted 1:10,000.

Chromatin preparation from fly embryos

Approximately 400 0- to 14-hour-old embryos per biological replicate (three biological replicates prepared per genotype) were dechorionated in bleach diluted 1:1 in water for 2 min at room temperature, extensively rinsed with water, transferred to an Eppendorf, flash-frozen, and stored at -80°C. Embryos were homogenized in a glass 15-ml Dounce in 5 ml of cross-linking solution [50 mM Hepes (pH 7.9), 1 mM EDTA (pH 8), 0.5 mM EGTA (pH 8), 100 mM NaCl, and 1.8% formaldehyde] with 15 strokes, transferred to a 15-ml Falcon tube, and rotated at room temperature. Cross-linking was stopped after 15 min by pelleting nuclei for 2 min at 2000g and rotating for 10 min in stop solution (1 \times PBS, 125 mM glycine, and 0.01% Triton X-100). Nuclei were washed for 10 min in solution A [10 mM Hepes (pH 7.9), 10 mM EDTA (pH 8), 0.5 mM EGTA (pH 8), and 0.25% Triton X-100] and then for 10 min in solution B [10 mM Hepes (pH 7.9), 1 mM EDTA (pH 8), 0.5 mM EGTA (pH 8), 0.01% Triton X-100, and 200 mM NaCl]. Nuclei were sonicated in 100 μ l of

radioimmunoprecipitation assay (RIPA) buffer [10 mM tris-HCl (pH 8), 140 mM NaCl, 1 mM EDTA (pH 8), 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, and 1 \times cOmplete protease inhibitor cocktail] in AFA microtubes in a Covaris S220 sonicator for 5 min with a peak incident power of 140 W, a duty cycle of 5%, and 200 cycles per burst. Sonicated chromatin was centrifuged to pellet insoluble material and snap-frozen.

Chromatin preparation from larval central nervous systems

Thirty third-instar larval cuticles per biological replicate (two biological replicates per sample) were dissected in ice-cold PBS and then cross-linked for 15 min at room temperature in 1.8% (v/v) paraformaldehyde, 50 mM Hepes (pH 8), 100 mM NaCl, 1 mM EDTA, and 1 mM EGTA. Cross-linking was stopped by washing for 10 min in 1 ml of PBS, 0.01% Triton X-100, and 125 mM glycine. Then, cuticles were washed for 10 min in 10 mM Hepes (pH 7.6), 10 mM EDTA, 0.5 mM EGTA, and 0.25% Triton X-100. Central nervous systems were dissected from the cuticles in 10 mM Hepes (pH 7.6), 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and 0.01% Triton X-100 and then sonicated in 100 μ l of RIPA buffer [10 mM tris-HCl (pH 8), 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, and protease inhibitor cocktail] in AFA microtubes in a Covaris S220 sonicator for 5 min with a peak incident power of 140 W, a duty cycle of 5%, and 200 cycles per burst. Sonicated chromatin was centrifuged to pellet insoluble material and snap-frozen.

ChIP-seq

ChIP was performed with 2 μ l of rabbit polyclonal antibody crude sera against CTCF¹⁻²⁹³ or Cp190¹⁻¹⁰⁹⁶ (7) each incubated with half of the chromatin prepared from a biological replicate overnight at 4°C. Premixed Protein A and G Dynabeads (25 μ l; Thermo Fisher Scientific, 100-01D and 100-03D) were added for 3 hours at 4°C and then washed for 10 min each once with RIPA, four times with RIPA with 500 mM NaCl, once in LiCl buffer [10 mM tris-HCl (pH 8), 250 mM LiCl, 1 mM EDTA, 0.5% IGEPAL CA-630, and 0.5% sodium deoxycholate], and twice in TE buffer [10 mM tris-HCl (pH 8) and 1 mM EDTA]. DNA was purified by ribonuclease digestion, proteinase K digestion, reversal of cross-links at 65°C for 6 hours, and elution from a QIAGEN MinElute PCR (polymerase chain reaction) purification column. ChIP-seq libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina. An equimolar pool of multiplexed ChIP-seq libraries at 4 nM was sequenced on two lanes of an Illumina HiSeq 4000 150-bp paired-end.

ChIP-seq analysis

Paired-end ChIP-seq reads were demultiplexed and mapped to the dm6 genome using micmap v2.20200223 (<https://github.com/sib-swiss/micmap>), a derivative of the fetchGWI tool. For samples that were sequenced twice, reads were merged with Samtools v1.10 (<http://www.htslib.org/>). Only chromosomes 2, 3, 4, and X were used. ChIP-seq peaks were called using the R package csaw v1.16.1 (<https://bioconductor.org/packages/release/bioc/html/csaw.html>) using a window width of 20 bp and spacing of 10 bp, ignoring duplicate reads, and blacklisted regions by ENCODE. A background enrichment was evaluated as the median over all samples in the comparison of the average number of reads per 2-kb bins. Windows with less than twofold (for ChIP-seq in embryos) or threefold

(for ChIP-seq in larval central nervous systems, which gives better signal-to-noise ratio) enrichment over background were filtered out. Data were normalized using the TMM (trimmed mean of M-values) method implemented in *csaw*. Differential binding analysis in *csaw* is based on the quasi-likelihood framework implemented in the *edgeR* package v3.22.5 (<https://bioconductor.org/packages/release/bioc/html/edgeR.html>). Results obtained on different windows were combined into regions by clustering adjacent windows. Combined *P* values were evaluated for each region using *csaw*, and the Benjamini-Hochberg method was applied to control the false discovery rate. Regions with false discovery rate <0.01 and $|\text{best.logFC}| > 1$ were considered as differentially bound regions. Genuine Cp190 peaks were identified by differential analysis of ChIP-seq signals in WT versus *Cp190⁰* embryos (Cp190 peaks in WT embryos; data S1), in *CTCF⁰* versus *Cp190⁰* embryos (Cp190 peaks in *CTCF⁰* embryos; data S6), in WT versus *Cp190^{KO}* larval central nervous systems (Cp190 peaks in WT larval central nervous system; data S10), or in *su(Hw)^{KO}* versus *Cp190^{KO}* larval central nervous systems [Cp190 peaks in *su(Hw)^{KO}* larval central nervous system; data S11] as being lower in *Cp190* mutants relative to WT, *CTCF⁰*, or *su(Hw)^{KO}*, respectively. Genuine CTCF peaks were similarly identified by differential analysis of ChIP-seq signals in WT versus *CTCF⁰* embryos (CTCF peaks in WT embryos; data S5) or in *Cp190⁰* versus *CTCF⁰* embryos (CTCF peaks in *Cp190⁰* embryos; data S8) as being lower in *CTCF⁰* relative to WT or *Cp190⁰*, respectively. One replicate of CTCF ChIP in *Cp190⁰* embryos failed at the library preparation step; hence, differential analysis was performed with the two remaining replicates. Additional differential analyses were performed for Cp190 ChIP-seq in WT versus *CTCF⁰* embryos (data S7), for CTCF ChIP-seq in WT versus *Cp190⁰* embryos (data S9), and for Cp190 ChIP-seq in WT versus *su(Hw)^{KO}* larvae (data S12). Throughout the manuscript, the following conventions are used when comparing ChIP data to other data. ChIP occupancy was defined as the best.log2FC obtained from *csaw* in the respective differential analysis. ChIP peak (and differentially bound region) positions were defined as the best.pos obtained from *csaw*, and regions were defined as the $[\text{start}, \text{end}]$ interval obtained from *csaw*. Overlapping ChIP peaks (and differentially bound regions) were defined as those with peak regions sharing at least 1 bp. Similarly, ChIP peaks overlapping a DNA motif were defined as those with peak regions sharing at least 1 bp with the motif. CTCF- or Cp190-occupied boundaries were defined as those with a ChIP peak position within ± 2 kb of the boundary. Promoter-proximal and promoter-distal ChIP peaks were defined as those with peak positions within ± 200 bp or further away from the closest TSS, respectively. In fig. S2D, CTCF peaks were defined as present in an intron when the peak position was inside an intron but not in an exon using gene annotations from FlyBase release FB2020_06.

Hi-C

About 100 2- to 6-hour-old embryos per biological replicate (four biological replicates per genotype) were dechorionated in bleach diluted 1:1 in water for 2 min at room temperature, extensively rinsed with water, transferred to an Eppendorf, and crushed in RPMI supplemented with 10% fetal bovine serum using a micropestle. Nuclei were fixed in 1% (v/v) formaldehyde for 10 min at room temperature. Cross-linking was stopped by adding 200 mM glycine; then, nuclei were washed in PBS and snap-frozen for -80°C storage. Nuclei were restricted with *Mse* I and *Csp* 6I; restricted ends were marked with biotin and then ligated. DNA was purified

by proteinase K digestion and reverse cross-linking at 65°C for 6 hours, then sonicated in AFA microtubes in a Covaris S220 sonicator, and purified on SPRIselect beads (Beckman Coulter). DNA was end-repaired, A-tailed, and ligated to barcoded adapters using the NEBNext Ultra II DNA Library Prep Kit for Illumina and then enriched for pairwise DNA junctions by biotin pull-down using Dynabeads MyOne Streptavidin T1 beads following the manufacturer's instructions. Libraries were amplified using KAPA HiFi HotStart Ready Mix and purified on SPRIselect beads. Four nanomolar equimolar pools of multiplexed Hi-C libraries were subjected to 150-bp paired-end sequencing on HiSeq 4000 instruments.

Hi-C analysis

We precomputed a table containing the positions of all restriction sites used for Hi-C present in the dm6 genome. The FASTQ read pairs were analyzed with a Perl script available for download in the micmap package v2.20200223 (<https://github.com/sib-swiss/micmap>) to locate and separate fusion sites using the patterns */GTATAC/, /TTATAA/, /GTATAA/, and /TTATAC/*. The maximal length of each read was trimmed at 60 nucleotides (nt); then, reads were mapped to the dm6 genome using micmap and matched to their closest precomputed genomic restriction site. Read pairs were discarded if they (i) mapped to non-unique positions in the reference genome; (ii) had indels or more than two mismatches per read; (iii) represented fusion of two oppositely oriented reads within 2 kb of each other, which may have not resulted from ligation of two digested fragments; and (iv) were likely PCR duplicates. Only chromosomes 2, 3, 4, and X were considered, and chromosome arms were treated as separate chromosomes.

To assess correlation of biological replicates, samples were downsampled to 13 million contacts per replicate. Raw Hi-C contact matrices were created by binning Hi-C pairs at 10-kb resolution. These matrices were then normalized with the ICE (iterative correction and eigenvector decomposition) normalization implemented in *iced* v0.5.2 (<https://github.com/hiclib/iced>). Low-coverage regions (bins with no contacts and those with the 5% smallest total number of contacts among bins) were filtered out before normalization. Pearson correlation coefficients were determined for every pair of normalized matrices by flattening each matrix and evaluating the Pearson correlation coefficient for the resulting vector using only pairs of bins at a genomic distance below 1 Mb. The limitation on the distance was introduced to compare contacts at a scale relevant to the analyses performed in this manuscript, which were at the level of contact domains. Resulting Pearson correlation coefficients were ≥ 0.936 for all replicates, showing that they were well correlated and that WT and mutant Hi-C matrices were globally similar. For the analyses presented in the main figures, pooled quadruplicate replicates of the same genotype were downsampled to 79 million contacts per genotype. Raw Hi-C contact matrices obtained by binning Hi-C pairs at 2-kb resolution were then normalized with the ICE normalization implemented in *iced* v0.5.2. Low-coverage regions (bins with no contacts and 5% of bins with the smallest total number of contacts but at least one contact) were filtered out before normalization (these regions are marked by gray lines in Hi-C maps shown in the figures). Hi-C maps were visualized in R.

For each normalized Hi-C contact matrix, contact domain boundaries were called using TopDom v0.0.2 (<https://github.com/jasminzhoulab/TopDom>) (41). Given a window size w_i , a physical insulation score was defined for each bin i as

$$\log_2 \frac{\text{binSignal}_i}{\sum_{i-w/2 < j < i+w/2} \text{binSignal}_j} \quad (1)$$

where binSignal_i is the average normalized Hi-C contact frequency between w bins upstream of bin i and w bins downstream of bin i determined by TopDom. The strength of a boundary at bin i was thus estimated as the \log_2 of the binSignal value at bin i normalized by its local average on a window of size w . With this definition, lower physical insulation scores indicate stronger boundaries. We extracted contact domain boundaries and physical insulation scores for Hi-C matrices at 2-kb resolution using window sizes 20, 40, 80, and 160 kb. Contact domain boundaries found with all window sizes were merged (boundaries with an insulation score >-0.1 were ignored), and the average insulation score obtained with all window sizes was retained. Boundaries with an average insulation score >-0.1 were filtered out. To facilitate comparisons of contact domain boundaries between genotypes and avoid mismatches due to small fluctuations of domain boundary positions obtained with different window sizes or genotypes, groups of consecutive boundaries (i.e., within 2 kb of each other) found in any of the four genotypes (WT, *Cp190*⁰, *CTCF*⁰, and *double*⁰) were replaced by the boundary with the lowest global insulation score (sum of insulation scores over all genotypes having this boundary), with insulation score taken from the corresponding genotype at the new boundary position. Boundaries in blacklisted regions were also filtered out to have a set of boundaries comparable to the set of ChIP peaks.

A/B compartment calling

A/B compartment calling was performed following the method proposed by Lieberman-Aiden *et al.* (74). Each individual chromosome arm (chr2L, chr2R, chr3L, chr3R, chr4, and chrX) was analyzed separately. In addition, to avoid having eigenvectors dominated by a chromosomal rearrangement in chr2L present in all four genotypes (WT, *Cp190*⁰, *CTCF*⁰, and *double*⁰), chr2L was split into three subregions analyzed independently (region 1 < 9471500 , $9471500 < \text{region 2} < 13657500$, and region 3 > 13657500). Normalized Hi-C contact matrices at 2-kb resolution were considered after discarding invalid bins (low-coverage regions filtered before ICE normalization) and bins around centromeres (chosen for exclusion as dm6 coordinates >22170000 for chr2L, <5650000 for chr2R, >22900000 for chr3L, and <4200000 for chr3R). Observed-over-expected matrices were generated by dividing the normalized Hi-C contact matrices by the average number of normalized Hi-C contacts at the corresponding genomic distance and clipped to the 99.9th percentile to avoid instabilities due to very large values. For each chromosome arm or chr2L region, the first eigenvector of the correlation matrix was obtained by principal components analysis of the observed-over-expected matrix. Each eigenvector was then multiplied by the sign of the Spearman correlation between the eigenvector and the number of gene TSSs per 2-kb bin (gene list taken from FlyBase release FB2020_06; ftp://ftp.flybase.net/genomes/Drosophila_melanogaster_dmel_r6.37_FB2020_06/gtf/dmel-all-r6.37.gtf.gz) and then centered around zero by subtracting its mean value. For chr2L, centering around zero was performed after merging eigenvectors from all regions. Two-kilobase bins with positive eigenvector values were assigned to compartment A; those with negative eigenvector values were assigned to compartment B. chr4 first eigenvector failed to capture the relevant A/B compartment structure and was thus excluded from fig. S1C.

Capture-C

The NG Capture-C (60) protocol was adapted as follows. Hi-C was performed exactly as described above on 2- to 6-hour-old WT, *Cp190*⁰, and *SFI*^{KO} embryos in biological triplicates. Five hundred nanograms of each of the nine Hi-C libraries (4.5 μg total DNA) was multiplexed and hybridized with 2.5 μg of Cot DNA, 2.5 μg of salmon sperm, 2 nmol of blocking oligos, and 16 fmol of xGen Lockdown probe pool [Integrated DNA Technologies (IDT)] consisting of 5'-biotinylated 120-nt single-stranded DNA capture probes listed in table S3 complementary to both ends of each of the 10 restriction fragments selected as viewpoints. Hybridization (24 hours at 65°C) and washes were performed with the xGen Hybridization and Wash Kit (IDT, 1080577) according to the manufacturer's protocol. The first capture was PCR-amplified with KAPA HiFi HotStart ReadyMix in a 50- μl reaction with 14 cycles to obtain 1 μg of postcapture DNA. This DNA was subsequently subjected to a second capture identical to the first, after which only nine PCR cycles were necessary to obtain sequencing-ready DNA. Capture-C libraries were subjected to 150-bp paired-end sequencing on a NovaSeq 6000.

Capture-C analysis

As for Hi-C analysis, we precomputed a table containing the positions of all restriction sites used for Hi-C present in the dm6 genome. To extract the read pairs corresponding to each captured region (called viewpoint), all capture probe sequences were split into consecutive 25-mers, and a Perl script was used to scan all read pairs in the raw FASTQ files and generate a specific pair of FASTQ files per viewpoint in which either read of the pair had an exact match to one of the 25-mers. The FASTQ read pairs were mapped to the dm6 genome using STAR (v2.7.7a; <https://github.com/alexdobin/STAR>) with parameters tuned to map the expected chimeric read pairs generated by Hi-C (--chimOutType WithinBAM, --chimSegmentMin 10, --outFilterMultimapNmax 1, --outFilterMismatchNoverLmax 0.04, --scoreGapNoncan 0, --scoreGapGCAG 0, --scoreGapATAC 0, --alignIntronMax 1, --chimScoreJunctionNonGTAG 0). Each matching pair was assigned to their closest precomputed genomic restriction site. Only chimeric read pairs (as defined by STAR) were retained.

For each viewpoint per sample, only informative read pairs were considered, i.e., only unique read pairs (after discarding probable PCR duplicates) with at least one read mapping to the viewpoint restriction fragment or one of its two neighboring restriction fragments. A vector of raw counts between the viewpoint and other regions in the genome was obtained by partitioning the genome into 1-kb bins (rounded to the nearest restriction site) and evaluating, for each bin, the number of read pairs with one end associated with the viewpoint and the other end associated with a restriction fragment overlapping the bin. Bins located <2 kb or >100 kb from the viewpoint restriction fragment were discarded.

For each comparison (*Cp190*⁰ versus WT in data S14 and *SFI*^{KO} versus WT in data S15), a differential analysis was done with diffHic v1.14.0 (<https://www.bioconductor.org/packages/release/bioc/html/diffHic.html>) using vectors of raw counts for all viewpoints. Data were normalized by library size. The Benjamini-Hochberg method was applied to control the false discovery rate. Capture-C read pair quality metrics are shown in table S4.

Analysis of published datasets

RNA-seq data in 4- to 6-hour-old WT embryos [mE_mRNA_em4-6hr_(FBIC0000088)] (75) were downloaded from FlyBase release

FB2020_06 (http://ftp.flybase.org/releases/FB2020_06/precomputed_files/genes/gene_rpkm_matrix_fb_2020_06.tsv.gz). Reads per kilobase per million reads (RPKM) values were calculated only for the unique exonic regions of the gene (excluding segments that overlap other genes), except for genes derived from di- or polycistronic transcripts, in which case, all exons were used in the RPKM expression calculation. For visualizing TSSs mapped by PRO-seq (Precision Run-On Sequencing) in 3- to 4-hour-old WT embryos and CAGE (cap analysis of gene expression) in 2- to 4-hour-old WT embryos (76) in fig. S1D, these published datasets were lifted over from dm3 to dm6 coordinates using the CrossMap tool (v0.6.0; <https://crossmap.readthedocs.io/en/latest/>). These datasets contain signed counts of 5' ends of reads per base pair (with positive/negative values corresponding to reads on the positive/negative strand), and only the absolute values of the counts were plotted in fig. S1D.

Published ChIP-seq profiles in S2 or in Kc cells when not available in S2 cells were remapped to genome version dm6 using the same pipeline described above for ChIP-seq analysis and visualized in Fig. 4B: Ibf1 (SRR837792) and Ibf2 (SRR837793) in S2 cells from GSE47559 (31); CTCF (SRR580343), Su(Hw) (SRR580339), mod(mdg4) (SRR580341), and Cp190 (SRR580337) in S2 cells from GSE41354 (40); BEAF-32 (SRR1042411) in S2 cells from GSE52962 (33); ZIPIC (SRR1141009) in S2 cells from GSM1313421 (32); and Pzq (SRR1636808) and Chro (SRR1636762) in Kc cells from GSE63518 (46).

To assess whether contact domain boundaries called in our study overlap those of previously published Hi-C contact maps, we compared our boundaries to those called by Hug *et al.* (67) using 3- to 4-hour-old WT embryo Hi-C maps binned at 2-kb resolution and by Ramírez *et al.* (8) using Kc167 tissue culture cell Hi-C maps binned at Dpn II restriction fragment resolution. Boundaries from Ramírez *et al.* (8) were converted from dm3 to dm6 genome coordinates using the LiftOver tool (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>). Published boundaries are displayed together with those from this study in all Hi-C screenshots in the manuscript for comparison.

To visualize and assess boundary defects around DNA motifs, we used known motifs [JASPAR motifs MA0531.1 for CTCF, MA0533.1 for Su(Hw), MA0529.2 for BEAF-32, and MA1459.1 for M1BP or published core motif-6 from Ohler *et al.* (77)]. The Ibf1 and ZIPIC motifs were rediscovered from the published ChIP-seq datasets mentioned above. For each ChIP-seq dataset, peaks were called with MACS2, the 500 peaks with the highest scores were selected, and sequences ± 100 bp around the ChIP peak summits were extracted and submitted to MEME (<https://meme-suite.org/meme/tools/meme>). The Ibf2 motif was similarly also rediscovered and was almost identical to that of Ibf1 as expected (31); hence, we only used the Ibf1 motif in Fig. 1 and fig. S1. The genome was then scanned for occurrences of each considered motif using PWMScan (<https://cgg.epfl.ch/pwmtools/pwmscan.php>) with default parameters. For visualizing deoxyribonuclease (DNase) hypersensitive sites mapped by DNase sequencing in 4- to 6-hour-old whole WT embryos (78) in fig. S1D, this published dataset was lifted over from dm3 to dm6 coordinates using the LiftOver tool.

RNA-fluorescence in situ hybridization

For single RNA-fluorescence in situ hybridization (FISH), labeled RNA probes were generated by in vitro transcription with Dig-UTP (digoxigenin-11-uridine triphosphate) labeling mix (Roche, 11277073910) and T7 RNA polymerase (Roche, 10881767001) antisense to full-length

cDNA clones of *Scr*, *Dfd*, *Antp* (LD33666), *Ubx* (RE43738), *abd-A* (RE04174), *Abd-B* (RE47096), and *ftz* [dm6 coordinates chr-3R:6864324-6865765 as originally published by Yokoshi *et al.* (59)] (see table S2 for primer sequences). After DNase I digestion for 20 min at 37°C, probes were fragmented by incubating for 20 min at 65°C in 60 mM Na₂CO₃ and 40 mM NaHCO₃ (pH 10.2); precipitated in 300 mM sodium acetate (pH 5.2), 1.25 M LiCl, tRNA (50 mg/ml), and 80% ethanol; resuspended in 50% formamide, 75 mM sodium citrate (pH 5), 750 mM NaCl, salmon sperm DNA (100 μ g/ml), heparin (50 μ g/ml), and 0.1% Tween 20; and stored at -20° C. Embryos were fixed in 4% paraformaldehyde for 30 min at room temperature, washed, and then stored in 100% methanol at -20° C. Samples were rehydrated in PBS with 0.1% Tween 20, postfixed in 4% paraformaldehyde for 20 min at room temperature, progressively equilibrated to hybridization buffer [50% formamide, 75 mM sodium citrate (pH 5), and 750 mM NaCl], and heated to 65°C. RNA probes were diluted 1:50 in hybridization buffer, denatured at 80°C for 10 min, then placed on ice, and added to the samples for overnight shaking at 65°C. Samples were washed six times for 10 min in hybridization buffer at 65°C and then progressively equilibrated to PBS with 0.1% Triton X-100. Samples were incubated overnight at 4°C in anti-Dig peroxidase (Roche, 11207733910) diluted 1:2000 in PBS, 0.1% Triton X-100, and 1 \times Western blocking reagent (Sigma-Aldrich, 1921673). Samples were washed six times for 10 min in PBS with 0.1% Tween 20, labeled with Cyanine 3 Tyramide in the TSA Plus kit (PerkinElmer, NEL753001KT) for 3 min at room temperature, washed six times for 10 min in PBS with 0.1% Tween 20, and lastly mounted with 4',6-diamidino-2-phenylindole (DAPI) to stain DNA. Images were acquired on a Zeiss LSM 880 microscope with a $\times 20$ objective and visualized with Fiji software v2.1.0/1.53c.

For double RNA-FISH, we used MUSE technology (arcobis bio). Twenty probes with around 30 bases antisense to the mRNA of interest (*Scr* or *ftz*; complementary sequences in table S2) were ordered with MUSE overhangs from IDT. Equimolar pools containing 100 nM each probe (100 \times stock) were made in 2 \times SSC (pH 7; 300 mM NaCl and 30 mM sodium citrate) and stored at -20° C. Embryos were fixed, stored, and rehydrated as for single RNA-FISH. Probe pools were diluted 100 \times and incubated with embryos at 40°C overnight. Samples were washed six times for 10 min in hybridization buffer at 40°C and then progressively equilibrated to and washed in 2 \times SSCT [2 \times SSC (pH 7) and 0.1% Tween 20]. MUSE signal was then detected by hybridizing nanoamplifiers and ATTO-488 and ATTO-550 readout probes according to the manufacturer's instructions. Embryos were DAPI-stained and mounted on microscope slides in 2 \times SSC for immediate imaging on a Zeiss LSM 880 microscope with a $\times 20$ objective. Images were processed with Fiji software v2.1.0/1.53c.

Insulator reporter

An insulator reporter plasmid (Fig. 4C) (7) comprises an enhancer (*OpIE2*) equidistant from EGFP and mCherry fluorescent reporters with basal *Hsp70* promoters. A *gypsy* insulator is present in the reporter plasmid, downstream of the EGFP transcription unit. Selected genomic loci were PCR-amplified using primer sequences in table S2 from genomic DNA and cloned in between the enhancer and EGFP. Control reporters had a neutral spacer (a fragment of the bacterial *Kanamycin* resistance gene) or the *gypsy* insulator in between the enhancer and EGFP. In addition, genomic fragments with BEAF-32 motifs were mutagenized by PCR to mutate 1 bp in a

BEAF-32 motif (TATCGATW to TAGCGATW). All plasmids of a given class [CTCF-bound, Su(Hw)-bound, or Chro+Pzgf+BEAF-32-bound] were transfected in parallel with spacer and *gypsy* control reporters into S2 cells in duplicates in a 96-well plate using 60 ng of reporter plasmid per replicate and Effectene (QIAGEN) following the manufacturer's instructions. S2 cells were originally purchased from the American Type Culture Collection (reference number CRL-1963) in 2006. After 48 hours, fluorescence was measured on a NovoCyte Flow Cytometer (ACEA) using fluorescein isothiocyanate and phycoerythrin–Texas Red detection settings. Distributions of mCherry/EGFP fluorescence ratios in thousands of single transfected cells were plotted, and the median mCherry/EGFP ratio was extracted for each experiment. The average mCherry/EGFP log₂ ratio obtained with the neutral spacer control was subtracted from all mCherry/EGFP log₂ ratios obtained. Then, the average mCherry/EGFP log₂ ratio obtained with the *gypsy* controls was set to 100% insulator strength. Relative insulator strengths of the tested fragments are expressed as percentage of *gypsy* insulator strength in Fig. 4C.

Copurification of insulator protein interactors from embryo nuclear extracts

Soluble nuclear protein extracts were prepared from WT (OregonR) 0- to 14-hour embryos. Thirty grams of embryos was dechorionated, taken up in 30 ml of NU1 buffer [15 mM Hepes (pH 7.6), 10 mM KCl, 5 mM MgCl₂, 0.1 mM EDTA (pH 8), 0.5 mM EGTA (pH 8), 350 mM sucrose, 2 mM dithiothreitol (DTT), and 0.2 mM phenylmethylsulfonyl fluoride], and Dounce-homogenized. The lysate was filtered through a double layer of Miracloth and then centrifuged for 15 min at 9000 rpm at 4°C. The nuclei pellet was resuspended and lysed in 30 ml of high-salt buffer [15 mM Hepes (pH 7.9), 400 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 20% glycerol, 1 mM DTT, and protease inhibitor cocktail] rotating for 20 min at 4°C and ultracentrifuged for 1 hour with an SW 40 rotor at 38,000 rpm at 4°C. The lipid layer was removed by suction, and the soluble nuclear extract was dialyzed into 15 mM Hepes (pH 7.9), 200 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA (pH 7.9), 20% glycerol, and 1 mM DTT with a 6- to 8-kDa molecular weight cutoff membrane. Soluble nuclear extract was snap-frozen in liquid nitrogen and stored at –80°C.

Baits with an N-terminal GFP-3C tag and a 3C-His₆ C-terminal tag were purified from bacterial lysates by Ni–nitrilotriacetic acid affinity. Baits were full length when soluble or spanned soluble portions of baits when the full-length protein was not soluble. The following baits were used: Su(Hw)[1-219] (N terminus), CTCF[1-293] [N terminus; results reproduced from Kaushal *et al.* (7)], Cp190[1-1096] (full length), Chro[613-926] (C terminus), or *Xenopus* Nse I (full length) as an unrelated bait for negative control. Numbering is based on UniProt accession numbers P08970 [Su(Hw)], Q9VS55 (CTCF), Q24478 (Cp190), and Q86BS3 (Chro). Purified GFP-3C-bait-3C-His₆ was immobilized on GFP binder beads, of which 30- μ l bead volume was then incubated with 6 mg of *Drosophila* embryo nuclear extract in a total volume of 10 ml of IP buffer [50 mM Tris-Cl (pH 7.5), 150 mM potassium acetate, 2 mM MgCl₂, 10% glycerol, 0.1 mM DTT, 0.2% IGEPAL, and 1 \times cOmplete protease inhibitor cocktail] rotating for 3 hours at 4°C. Beads were washed three times with IP buffer, rotating for 10 min at 4°C for each wash. Proteins were eluted with 3C protease, adjusted to 1 \times SDS loading buffer, and loaded on an SDS–polyacrylamide gel electrophoresis (SDS–PAGE) gel. A duplicate experiment was similarly performed with nuclear protein extracts prepared from another biological replicate embryo sample.

Although CTCF and Su(Hw) pull-downs were only performed with N-terminal portions, peptides spanning full-length CTCF and Su(Hw) were respectively recovered, indicating that interactors of full-length proteins could be recovered. Pull-downs that we also performed with CTCF and Su(Hw) C-terminal portions did not recover additional interactors.

Mass spectrometry analysis

For CTCF [previously described in (7)], Cp190, Chro, and their respective Nse I negative control pull-downs, protein samples were separated by SDS–PAGE and stained by Coomassie. Gel lanes between 10 and 300 kDa were excised into 5 to 10 pieces and digested with sequencing-grade trypsin. Extracted tryptic peptides were dried and resuspended in 0.05% trifluoroacetic acid (TFA) and 2% (v/v) acetonitrile. Tryptic peptide mixtures were injected on a Dionex RSLC 3000 nanoHPLC system (Dionex, Sunnyvale, CA, USA) interfaced via a nanospray source to a high-resolution mass spectrometer LTQ-Orbitrap Velos Pro (Thermo Fisher Scientific, Bremen, Germany) or timsTOF Pro (Bruker, Bremen, Germany). Peptides were loaded onto a trapping microcolumn Acclaim PepMap 100 C18 [20 mm by 100 μ m; inside diameter (ID), 5 μ m; Dionex] before separation on a C18 reversed-phase custom-packed column (75 μ m ID by 40 cm; 1.8- μ m particles; ReproSil-Pur, Dr. Maisch) using a gradient from 4 to 76% acetonitrile in 0.1% formic acid.

In the LTQ-Orbitrap Velos instrument, the 10 most intense multiply charged precursor ions detected with a full mass spectrometry (MS) survey scan in the Orbitrap [resolution of 60,000 at mass/charge ratio (*m/z*) of 400] were selected for collision-induced dissociation (normalized collision energy, 35%) and analysis in the ion trap. The window for precursor isolation was of 4.0 *m/z* units around the precursor, and selected fragments were excluded for 60 s from further analysis.

In the timsTOF instrument, data-dependent acquisition was carried out using a standard TIMS PASEF method with ion accumulation for 100 ms for both the survey MS1 scan and the TIMS-coupled MS2 scans. Duty cycle was kept at 100%. Up to 10 precursors were targeted per TIMS scan. Precursor isolation was done with 2 or 3 *m/z* windows below or above *m/z* of 800, respectively. The minimum threshold intensity for precursor selection was 2500. If the inclusion list allowed it, then precursors were targeted more than once to reach a minimum target total intensity of 20,000. Collision energy was ramped linearly only on the basis of the 1/*k* values from 20 (at 1/*k* = 0.6) to 59 eV (at 1/*k* = 1.6). Total duration of a scan cycle including one survey and 10 MS2 TIMS scans was 1.16 s. Precursors could be targeted again in subsequent cycles if their signal increased by a factor 4.0 or more. After selection in 1 cycle, precursors were excluded from further selection for 60 s. Mass resolution in all MS measurements was approximately 35,000.

For Su(Hw) and its respective Nse I negative control pull-down, samples were digested following a modified version of the iST method. After dilution 1:1 (v/v) with 1% sodium deoxycholate, 100 mM Tris (pH 8.6), and 10 mM DTT buffer, reduced disulfides were alkylated by adding ¼ vol of 160 mM chloroacetamide (final, 32 mM) and incubating at 25°C for 45 min in the dark. Samples were adjusted to 3 mM EDTA and digested with 1 μ g of trypsin/LysC mix (Promega, #V5073) under gentle shaking for 1 hour at 37°C, followed by a second 1-hour digestion with a second identical aliquot of proteases. To remove sodium deoxycholate, two sample

volumes of isopropanol containing 1% TFA were added to the digests, and the samples were desalted on a cation exchange plate (Oasis MCX microelution plate; Waters Corp., Milford, MA, prod.#186001830BA) by centrifugation. After washing with isopropanol/1% TFA, peptides were eluted in 250 μ l of 80% acetonitrile (MeCN), 19% water, and 1% (v/v) ammonia. Eluates after steric exclusion chromatography desalting were dried and resuspended in 2% MeCN and 0.1% TFA before injection on a timsTOF mass spectrometer using a 110-min gradient from 4 to 76% acetonitrile in 0.1% formic acid.

Data files were analyzed with MaxQuant 1.6.3.4 incorporating the Andromeda search engine for protein identification and quantification based on iBAQ intensities. The following variable modifications were specified: cysteine carbamidomethylation (fixed) and methionine oxidation and protein N-terminal acetylation (variable). The sequence databases used for searching were *D. melanogaster* and *Escherichia coli* (strain K12) reference proteomes based on the UniProt database (www.uniprot.org; versions of 24 August 2020, containing 22,039 and 4391 sequences, respectively) and a contaminant database containing the most usual environmental contaminants and the enzymes used for digestion (keratins, trypsin, etc.). Both peptide and protein identifications were filtered at 1% false discovery rate relative to hits against a decoy database built by reversing protein sequences. The MaxQuant output tables proteinGroups.txt were processed with Perseus software to remove proteins matched to the contaminants database and proteins identified only by modified peptides or reverse database hits. Next, the tables were filtered to retain only proteins identified by a minimum of two peptides; the iBAQ quantitative values were \log_2 -transformed, and *E. coli* proteins were removed. Missing values were imputed with the lowest value measured. Results are represented in Fig. 4A as the \log_2 average fold enrichment of proteins in biological duplicate pull-down experiments relative to that of a negative control pull-down (with GFP-tagged *Xenopus* Nse I as unrelated bait) done in parallel. Enrichments of baits themselves are not plotted because baits were exogenously added in large excess relative to endogenous proteins present in the nuclear extracts.

Recombinant protein pull-downs

Expression plasmids encoding untagged full-length Cp190[1-1096] and/or Cp60[1-440] (with kanamycin resistance) were cotransformed with a plasmid expressing GFP-tagged CTCF[1-293] (N terminus), CTCF[610-818] (C terminus), Su(Hw)[1-219] (N terminus), Su(Hw)[724-941] (C terminus), Ibf1[1-242] (full length), or Ibf2[1-195] (full length) (with ampicillin resistance) into the *E. coli* Rosetta strain. Numbering is based on UniProt accession numbers Q24478 (Cp190), Q7K180 (Cp60), Q9VS55 (CTCF), P08970 [Su(Hw)], Q9VHG5 (Ibf1), and Q9VHG6 (Ibf2). Colonies were inoculated in 10 ml of TB cultures and grown at 37°C to an OD₆₀₀ (optical density at 600 nm) of 1. The culture temperature was then reduced to 18°C, and 0.5 mM isopropyl- β -D-thiogalactopyranoside was added to induce protein expression. Cells were harvested after overnight incubation at 18°C, and the pellets were resuspended in 2 volumes of lysis buffer [50 mM tris (pH 7.5), 200 mM NaCl, 5% glycerol, and 25 mM imidazole]. Cells were lysed by sonication, and the lysate was clarified by centrifugation at 16,000g for 10 min at 4°C. The lysates were incubated for 1 hour at 4°C with 20 μ l of GFP-binder resin. The beads were washed three times with 1 ml of lysis buffer and then boiled in SDS loading buffer to elute purified proteins, which were then visualized by SDS-PAGE and Coomassie staining.

Annotation of embryonic enhancers at *Scr-ftz-Antp* locus

Figure 5B shows stripe enhancers active in early embryos. Enhancer 1 is VT37564 active in a stripe posterior to the cephalic furrow, thus overlapping *Scr* expression (53). Enhancer 2 is VT37565 active in *ftz* stripes 3 and 6 (53). Enhancer 3 is the *ftz* upstream element active in all seven *ftz* stripes (56). Enhancer 4 is the *ftz* zebra enhancer active in all seven stripes (79). Enhancer 5 is *ftz*DE active in *ftz* stripes 1 and 5 (55).

Figures S5B and 6B show selected enhancers active in later embryos and relevant to the discussed phenotypes. Enhancer 6 is VT37574 active in an anterior segment overlapping *Scr* expression (53). VT37574 overlaps the 3.7-kb Hind III and 7-kb Eco RI fragments (52) and also the T1 enhancer (80) described to be active in labial or prothoracic segments overlapping *Scr* expression. Enhancer 7 is the 6.8-kb Xba I fragment driving expression in hindgut and anal plate (52). This enhancer overlaps two inactive enhancers [VT37547 (53) and the 6.7-kb Bam HI fragment (52)] such that the probable active enhancer is entirely separated from the *Scr* promoter by an annotated Cp190 peak in WT (see Fig. 6B).

Statistics and reproducibility

All described replicate experiments are biological (not technical) replicates. For all box plots, the center line denotes the median, box limits are upper and lower quartiles, the upper whisker extends to the largest value no further than 1.5 \times interquartile range from the upper hinge, the lower whisker extends to the smallest value no further than 1.5 \times interquartile range from the lower hinge, and points indicate outliers. Contingency tables in fig. S2 are colored by $\log_{10}(n_{\text{observed}}/n_{\text{expected}})$, where n_{expected} is the expected value assuming independence of rows and columns. This value was obtained for each cell as (row sum) \times (column sum)/(table sum).

Animals were not separated by sex. Samples were grouped according to genotype (WT or various mutants). The investigators were not blinded during data collection, as the biological groups (genotypes) were well defined and handled in parallel. Computational analysis was performed by data scientists different from the researchers who collected the data. No data were excluded from the analyses.

For *Drosophila* viability tests, at least 100 animals were analyzed per genotype because clear differences between genotypes were visible already at this scale. For RNA-FISH experiments, approximately 20 embryos were examined per genotype over two independent experiments and only representative phenotypes that were observed in all animals are shown. These numbers were chosen because they revealed that phenotypes were reproducibly detected in all animals and because sample collection beyond this scale was rate limiting. For ChIP-seq, Hi-C, and Capture-seq experiments, at least 100 embryos were collected or 30 third instar larval brains were dissected per replicate because these numbers allowed sufficient material to be amplified for NG sequencing library preparation with a limited number of PCR cycles to avoid overamplification. This number was sufficient because all biological replicates were well correlated.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abl8834>

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

- M. J. Rowley, M. H. Nichols, X. Lyu, M. Ando-Kuri, I. S. M. Rivera, K. Hermetz, P. Wang, Y. Ruan, V. G. Corces, Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* **67**, 837–852.e7 (2017).

2. E. P. Nora, A. Goloborodko, A.-L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, B. G. Bruneau, Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
3. S. S. P. Rao, S.-C. Huang, B. G. S. Hilaire, J. M. Engreitz, E. M. Perez, K.-R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, E. L. Aiden, Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
4. F. Zenk, Y. Zhan, P. Kos, E. Löser, N. Atinbayeva, M. Schächtle, G. Tiana, L. Giorgetti, N. Iovino, HP1 drives de novo 3D genome reorganization in early *Drosophila* embryos. *Nature* **593**, 289–293 (2021).
5. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
6. E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, E. Heard, Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature* **485**, 381–385 (2012).
7. A. Kaushal, G. Mohana, J. Dorier, I. Özdemir, A. Omer, P. Cousin, A. Semenova, M. Taschner, O. Dergai, F. Marzetta, C. Iseli, Y. Eliaz, D. Weisz, M. S. Shamim, N. Guex, E. L. Aiden, M. C. Gambetta, CTCF loss has limited effects on global genome architecture in *Drosophila* despite critical regulatory functions. *Nat. Commun.* **12**, 1011 (2021).
8. F. Ramirez, V. Bhardwaj, L. Arrigoni, K. C. Lam, B. A. Grüning, J. Villaveces, B. Haberman, A. Akhtar, T. Manke, High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
9. V. Narendra, P. P. Rocha, D. An, R. Raviram, J. A. Skok, E. O. Mazzone, D. Reinberg, CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021 (2015).
10. J. M. Downen, Z. P. Fan, D. Hnisz, G. Ren, B. J. Abraham, L. N. Zhang, A. S. Weintraub, J. Schuijers, T. I. Lee, K. Zhao, R. A. Young, Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
11. W. A. Flavahan, Y. Drier, B. B. Liaw, S. M. Gillespie, A. S. Venteicher, A. O. Stemmer-Rachamimov, M. L. Suvà, B. E. Bernstein, Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
12. A. Despang, R. Schöpflin, M. Franke, S. Ali, I. Jerković, C. Paliou, W.-L. Chan, B. Timmermann, L. Wittler, M. Vingron, S. Mundlos, D. M. Ibrahim, Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* **51**, 1263–1271 (2019).
13. E. Rodríguez-Carballo, L. Lopez-Delisle, Y. Zhan, P. J. Fabre, L. Beccari, I. El-Idrissi, T. A. Huynh, H. Ozadam, J. Dekker, D. Duboule, The *HoxD* cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–2281 (2017).
14. O. Delaneau, M. Zazhytska, C. Borel, G. Giannuzzi, G. Rey, C. Howald, S. Kumar, H. Ongen, K. Popadin, D. Marbach, G. Ambrosini, D. Bielser, D. Hacker, L. Romano, P. Ribaux, M. Wiederkehr, E. Falconnet, P. Bucher, S. Bergmann, S. E. Antonarakis, A. Reymond, E. T. Dermitzakis, Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* **364**, eaat8266 (2019).
15. O. Symmons, V. V. Uslu, T. Tsujimura, S. Ruf, S. Nassari, W. Schwarzer, L. Ettwiller, F. Spitz, Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
16. J. Zuin, G. Roth, Y. Zhan, J. Cramard, J. Redolfi, E. Piskadlo, P. Mach, M. Kryzhanovska, G. Tihanyi, H. Kohler, P. Meister, S. Smallwood, L. Giorgetti, Nonlinear control of transcription through enhancer-promoter interactions. *bioRxiv* 10.1101/2021.04.22.440891 [Preprint]. 2021.
17. C. Paliou, G. Guckelberger, R. Schöpflin, V. Heinrich, A. Esposito, A. M. Chiariello, S. Bianco, C. Annunziatella, J. Helmuth, S. Haas, I. Jerković, N. Brieske, L. Wittler, B. Timmermann, M. Nicodemi, M. Vingron, S. Mundlos, G. Andrey, Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12390–12399 (2019).
18. O. Symmons, L. Pan, S. Remeseiro, T. Aktas, F. Klein, W. Huber, F. Spitz, The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev. Cell* **39**, 529–543 (2016).
19. D. G. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, S. Mundlos, Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
20. M. Franke, D. M. Ibrahim, G. Andrey, W. Schwarzer, V. Heinrich, R. Schöpflin, K. Kraft, R. Kempfer, I. Jerković, W.-L. Chan, M. Spielmann, B. Timmermann, L. Wittler, I. Kurth, P. Cambiaso, O. Zuffardi, G. Houge, L. Lambie, F. Brancati, A. Pombo, M. Vingron, F. Spitz, S. Mundlos, Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
21. E. S. M. Vos, C. Valdes-Quezada, Y. Huang, A. Allahyar, M. J. A. M. Versteegen, A.-K. Felder, F. van der Vegt, E. C. H. Uijtewaal, P. H. L. Krijger, W. de Laat, Interplay between CTCF boundaries and a super enhancer controls cohesin extrusion trajectories and gene expression. *Mol. Cell* **81**, 3082–3095.e6 (2021).
22. A.-L. Valton, S. V. Venev, B. Mair, E. Khokhar, A. H. Y. Tong, M. Usaj, K. S. K. Chan, A. A. Pai, J. Moffat, J. Dekker, A cohesin traffic pattern genetically linked to gene regulation. *bioRxiv* 10.1101/2021.07.29.454218 [Preprint]. 2021.
23. T. Ali, M. Krüger, S. Bhujji, M. Jarek, M. Bartkuhn, R. Renkawitz, Chromatin binding of Gcn5 in *Drosophila* is largely mediated by CP190. *Nucleic Acids Res.* **45**, 2384–2395 (2017).
24. I. Bag, S. Chen, L. F. Rosin, Y. Chen, C.-Y. Liu, G.-Y. Yu, E. P. Lei, M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity. *Nat. Commun.* **12**, 4170 (2021).
25. C.-Y. Pai, E. P. Lei, D. Ghosh, V. G. Corces, The centrosomal protein CP190 is a component of the gypsy chromatin insulator. *Mol. Cell* **16**, 737–748 (2004).
26. M. Mohan, M. Bartkuhn, M. Herold, A. Philippen, N. Heini, I. Bardenhagen, J. Leers, R. A. White, R. Renkawitz-Pohl, H. Saumweber, R. Renkawitz, The *Drosophila* insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *EMBO J.* **26**, 4203–4214 (2007).
27. T. I. Gerasimova, E. P. Lei, A. M. Bushey, V. G. Corces, Coordinated control of dCTCF and gypsy chromatin insulators in *Drosophila*. *Mol. Cell* **28**, 761–772 (2007).
28. M. Savitsky, M. Kim, O. Kravchuk, Y. B. Schwartz, Distinct roles of chromatin insulator proteins in control of the *Drosophila* Bithorax complex. *Genetics* **202**, 601–617 (2016).
29. A. M. Wood, K. Van Bortle, E. Ramos, N. Takenaka, M. Rohrbach, B. C. Jones, K. C. Jones, V. G. Corces, Regulation of chromatin organization and inducible gene expression by a *Drosophila* insulator. *Mol. Cell* **44**, 29–38 (2011).
30. Y. B. Schwartz, D. Linder-Basso, P. V. Kharchenko, M. Y. Tolstokrov, M. Kim, H.-B. Li, A. A. Gorchakov, A. Minoda, G. Shanover, A. A. Alekseyenko, N. C. Riddle, Y. L. Jung, T. Gu, A. Plachetka, S. C. R. Elgin, M. I. Kuroda, P. J. Park, M. Savitsky, G. H. Karpen, V. Pirrotta, Nature and function of insulator protein binding sites in the *Drosophila* genome. *Genome Res.* **22**, 2188–2198 (2012).
31. S. Cuartero, U. Fresán, O. Reina, E. Planet, M. L. Espinàs, Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. *EMBO J.* **33**, 637–647 (2014).
32. O. Maksimenko, M. Bartkuhn, V. Stakhov, M. Herold, N. Zolotarev, T. Jox, M. K. Buxa, R. Kirsch, A. Bonchuk, A. Fedotova, O. Kyrchanova, R. Renkawitz, P. Georgiev, Two new insulator proteins, Pita and ZlPIC, target CP190 to chromatin. *Genome Res.* **25**, 89–99 (2015).
33. J. Liang, L. Lacroix, A. Gamot, S. Cuddapah, S. Queille, P. Lhoumaud, P. Lepetit, P. G. P. Martin, J. Vogelmann, F. Court, M. Hennion, G. Micas, S. Urbach, O. Bouchez, M. Nölmann, K. Zhao, E. Emberly, O. Cuvier, Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and Pol II pausing. *Mol. Cell* **53**, 672–681 (2014).
34. I. Comet, B. Schuettengruber, T. Sexton, G. Cavalli, A chromatin insulator driving three-dimensional Polycomb response element (PRE) complexes and Polycomb association with the chromatin fiber. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2294–2299 (2011).
35. O. Kyrchanova, D. Chetverina, O. Maksimenko, A. Kullyev, P. Georgiev, Orientation-dependent interaction between *Drosophila* insulators is a property of this class of regulatory elements. *Nucleic Acids Res.* **36**, 7019–7028 (2008).
36. H. Chen, M. Levo, L. Barinov, M. Fujioka, J. B. Jaynes, T. Gregor, Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
37. M. Fujioka, H. Mistry, P. Schedl, J. B. Jaynes, Determinants of chromosome architecture: Insulator pairing in cis and in trans. *PLoS Genet.* **12**, e1005889 (2016).
38. B. Lim, T. Heist, M. Levine, T. Fukaya, Visualization of transvection in living *Drosophila* embryos. *Mol. Cell* **70**, 287–296.e6 (2018).
39. J. Vogelmann, A. L. Gall, S. Dejardin, F. Allemand, A. Gamot, G. Labesse, O. Cuvier, N. Nègre, M. Cohen-Gonsaud, E. Margeat, M. Nölmann, Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the *Drosophila* genome. *PLoS Genet.* **10**, e1004544 (2014).
40. C.-T. Ong, K. Van Bortle, E. Ramos, V. G. Corces, Poly(ADP-ribosyl)ation regulates insulator function and intrachromosomal interactions in *Drosophila*. *Cell* **155**, 148–159 (2013).
41. H. Shin, Y. Shi, C. Dai, H. Tjong, K. Gong, F. Alber, X. J. Zhou, TopDom: An efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res.* **44**, e70 (2015).
42. A. M. Bushey, E. Ramos, V. G. Corces, Three subclasses of a *Drosophila* insulator show distinct and cell type-specific genomic distributions. *Genes Dev.* **23**, 1338–1350 (2009).
43. R. M. Baxley, A. A. Soshnev, D. E. Koryakov, I. F. Zhimulev, P. K. Geyer, The role of the Suppressor of Hairy-wing insulator protein in *Drosophila* oogenesis. *Dev. Biol.* **356**, 398–410 (2011).
44. D. R. Kellogg, B. M. Alberts, Purification of a multiprotein complex containing centrosomal proteins from the *Drosophila* embryo by chromatography with low-affinity polyclonal antibodies. *Mol. Biol. Cell* **3**, 1–11 (1992).

45. A. A. Alekseyenko, A. A. Gorchakov, B. M. Zee, S. M. Fuchs, P. V. Kharchenko, M. I. Kuroda, Heterochromatin-associated interactions of Drosophila HP1a with dADD1, HIPP1, and repetitive RNAs. *Genes Dev.* **28**, 1445–1460 (2014).
46. L. Li, X. Lyu, C. Hou, N. Takenaka, H. Q. Nguyen, C.-T. Ong, C. Cubeñas-Potts, M. Hu, E. P. Lei, G. Bosco, Z. S. Qin, V. G. Corces, Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol. Cell* **58**, 216–231 (2015).
47. P. K. Geyer, V. G. Corces, DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes Dev.* **6**, 1865–1873 (1992).
48. E. Emberly, R. Blattes, B. Schuettengruber, M. Hennion, N. Jiang, C. M. Hart, E. Käs, O. Cuvier, BEAF regulates cell-cycle genes through the controlled deposition of H3K9 methylation marks into its conserved dual-core binding sites. *PLoS Biol.* **6**, 2896–2910 (2008).
49. N. Nègre, C. D. Brown, L. Ma, C. A. Bristow, S. W. Miller, U. Wagner, P. Kheradpour, M. L. Eaton, P. Loriaux, R. Sealton, Z. Li, H. Ishii, R. F. Spokony, J. Chen, L. Hwang, C. Cheng, R. P. Auburn, M. B. Davis, M. Domanus, P. K. Shah, C. A. Morrison, J. Zieba, S. Suchy, L. Senderowicz, A. Victorsen, N. A. Bild, A. J. Grundstad, D. Hanley, D. M. MacAlpine, M. Mannervik, K. Venken, H. Bellen, R. White, M. Gerstein, S. Russell, R. L. Grossman, B. Ren, J. W. Posakony, M. Kellis, K. P. White, A cis-regulatory map of the *Drosophila* genome. *Nature* **471**, 527–531 (2011).
50. K. Van Bortle, M. H. Nichols, L. Li, C.-T. Ong, N. Takenaka, Z. S. Qin, V. G. Corces, Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol.* **15**, R82 (2014).
51. K. C. Scott, A. D. Taubman, P. K. Geyer, Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength. *Genetics* **153**, 787–798 (1999).
52. J. G. Gindhart Jr., A. N. King, T. C. Kaufman, Characterization of the cis-regulatory region of the Drosophila homeotic gene *Sex combs reduced*. *Genetics* **139**, 781–795 (1995).
53. E. Z. Kvon, T. Kazmar, G. Stampfel, J. O. Yáñez-Cuna, M. Pagani, K. Schernhuber, B. J. Dickson, A. Stark, Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91–95 (2014).
54. B. T. Wakimoto, F. R. Turner, T. C. Kaufman, Defects in embryogenesis in mutants associated with the antennapedia gene complex of *Drosophila melanogaster*. *Dev. Biol.* **102**, 147–172 (1984).
55. V. C. Calhoun, M. Levine, Long-range enhancer–promoter interactions in the Scr-Antp interval of the Drosophila Antennapedia complex. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9878–9883 (2003).
56. Y. Hiromi, W. J. Gehring, Regulation and function of the Drosophila segmentation gene *fushi tarazu*. *Cell* **50**, 963–974 (1987).
57. V. E. Belozerop, P. Majumder, P. Shen, H. N. Cai, A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of Drosophila. *EMBO J.* **22**, 3113–3121 (2003).
58. M. Li, Z. Ma, J. K. Liu, S. Roy, S. K. Patel, D. C. Lane, H. N. Cai, An organizational hub of developmentally regulated chromatin loops in the Drosophila antennapedia complex. *Mol. Cell Biol.* **35**, 4018–4029 (2015).
59. M. Yokoshi, K. Segawa, T. Fukaya, Visualizing the role of boundary elements in enhancer-promoter communication. *Mol. Cell* **78**, 224–235.e5 (2020).
60. J. O. J. Davies, J. M. Telenius, S. J. McGowan, N. A. Roberts, S. Taylor, D. R. Higgs, J. R. Hughes, Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* **13**, 74–80 (2016).
61. R. K. Maeda, F. Karch, The open for business model of the bithorax complex in Drosophila. *Chromosoma* **124**, 293–307 (2015).
62. J. Mihalý, S. Barges, L. Sipos, R. Maeda, F. Cléard, I. Hogga, W. Bender, H. Gyurkovics, F. Karch, Dissecting the regulatory landscape of the Abd-B gene of the bithorax complex. *Development* **133**, 2983–2993 (2006).
63. H. Gyurkovics, J. Gausz, J. Kummer, F. Karch, A new homeotic mutation in the Drosophila bithorax complex removes a boundary separating two domains of regulation. *EMBO J.* **9**, 2579–2585 (1990).
64. W. Bender, M. Lucas, The border between the ultrabithorax and abdominal-A regulatory domains in the Drosophila bithorax complex. *Genetics* **193**, 1135–1147 (2013).
65. L. J. Mateo, S. E. Murphy, A. Hafner, I. S. Cinquini, C. A. Walker, A. N. Boettiger, Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
66. B. Bonev, N. M. Cohen, Q. Szabo, L. Fritsch, G. L. Papadopoulos, Y. Lubling, X. Xu, X. Lv, J.-P. Hugnot, A. Tanay, G. Cavalli, Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017).
67. C. B. Hug, A. G. Grimaldi, K. Kruse, J. M. Vaquerizas, Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**, 216–228.e19 (2017).
68. I. Özdemir, M. C. Gambetta, The role of insulation in patterning gene expression. *Genes* **10**, 767 (2019).
69. M. C. Gambetta, E. E. M. Furlong, The insulator protein CTCF is required for correct *Hox* gene expression, but not for embryonic development in *Drosophila*. *Genetics* **210**, 129–136 (2018).
70. O. Kyrchanova, M. Sabirov, V. Mogila, A. Kurbidaeva, N. Postika, O. Maksimenko, P. Schedl, P. Georgiev, Complete reconstitution of bypass and blocking functions in a minimal artificial *Fab-7* insulator from *Drosophila bithorax* complex. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 13462–13467 (2019).
71. R. M. Baxley, J. D. Bullard, M. W. Klein, A. G. Fell, J. A. Morales-Rosado, T. Duan, P. K. Geyer, Deciphering the DNA code for the function of the Drosophila polydactyl zinc finger protein Suppressor of Hairy-wing. *Nucleic Acids Res.* **45**, 4463–4478 (2017).
72. L. Melnikova, M. Kostyuchenko, V. Molodina, A. Parshikov, P. Georgiev, A. Golovnin, Interactions between BTB domain of CP190 and two adjacent regions in Su(Hw) are required for the insulator complex formation. *Chromosoma* **127**, 59–71 (2018).
73. T. Xiao, X. Li, G. Felsenfeld, The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023127118 (2021).
74. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
75. B. R. Graveley, A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin, L. Yang, C. G. Artieri, M. J. van Baren, N. Boley, B. W. Booth, J. B. Brown, L. Cherbas, C. A. Davis, A. Dobin, R. Li, W. Lin, J. H. Malone, N. R. Mattiuzzo, D. Miller, D. Sturgill, B. B. Tuch, C. Zaleski, D. Zhang, M. Blanchette, S. Dudoit, B. Eads, R. E. Green, A. Hammonds, L. Jiang, P. Kapranov, L. Langton, N. Perrimon, J. E. Sandler, K. H. Wan, A. Willingham, Y. Zhang, Y. Zou, J. Andrews, P. J. Bickel, S. E. Brenner, M. R. Brent, P. Cherbas, T. R. Gingeras, R. A. Hoskins, T. C. Kaufman, B. Oliver, S. E. Celniker, The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479 (2011).
76. O. Mikhaylichenko, V. Bondarenko, D. Harnett, I. E. Schor, M. Maleš, R. R. Viales, E. E. M. Furlong, The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev.* **32**, 42–57 (2018).
77. U. Ohler, G. Liao, H. Niemann, G. M. Rubin, Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol.* **3**, research0087.1 (2002).
78. J. P. Reddington, D. A. Garfield, O. M. Sigalova, A. K. Calviello, R. Marco-Ferreres, C. Girardot, R. R. Viales, J. F. Degner, U. Ohler, E. E. M. Furlong, Lineage-resolved enhancer and promoter usage during a time course of embryogenesis. *Dev. Cell* **55**, 648–664.e9 (2020).
79. Y. Hiromi, A. Kuroiwa, W. J. Gehring, Control elements of the Drosophila segmentation gene *fushi tarazu*. *Cell* **43**, 603–613 (1985).
80. V. C. Calhoun, A. Stathopoulos, M. Levine, Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9243–9247 (2002).

Acknowledgments: We are grateful to A. Omer for help with Hi-C. Homie transgenes were provided by M. Fujioka and J. B. Jaynes. We thank F. Karch and W. Herr for helpful discussions. Deep sequencing was performed at the Genomic Technologies Facility (GTF); MS was performed at the Protein Analysis Facility (PAF), and imaging was performed at the Cellular Imaging Facility (CIF) at the Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, Switzerland. **Funding:** This work was supported by the Swiss National Science Foundation grant no. 184715 (M.C.G.) and the University of Lausanne (M.C.G.). **Author contributions:** A.K., B.W., M.T., G.M., A.S., P.C., and M.C.G. performed the experiments. P.W. performed MS analyses. S.R. provided MUSE experiment protocols and reagents. M.C.G. performed Hi-C using a method adapted from E.L.A. J.D., P.C., C.I., and N.G. analyzed the data and generated the plots. J.D., P.W., and C.I. wrote their respective sections in Materials and Methods. M.C.G. wrote the manuscript with input from all authors. **Competing interests:** S.R. developed MUSE technology marketed by arcoris bio AG and used in Fig. 5C. S.R. has shares and sits on the board of arcoris bio AG. E.L.A. discloses that he received payments from the Broad Institute for his expert witness testimony and is part of the Scientific Advisory Board for Colossal Laboratories and Biosciences. The authors declare no other competing interests. **Data and materials availability:** All sequencing data (Hi-C, NG Capture-C, and ChIP-seq) that support the findings of this study were deposited in Gene Expression Omnibus with accession code GSE180376. MS proteomics data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD027795. Reused data from Kaushal *et al.* (7) were previously deposited with the identifier PXD019487. All other relevant data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 18 August 2021
Accepted 1 April 2022
Published 13 May 2022
10.1126/sciadv.abl8834

Essential role of Cp190 in physical and regulatory boundary formation

Anjali KaushalJulien DorierBihan WangGiriram MohanaMichael TaschnerPascal CousinPatrice WaridelChristian IseliAnastasiia SemenovaSimon RestrepoNicolas GuexErez Lieberman AidenMaria Cristina Gambetta

Sci. Adv., 8 (19), eabl8834. • DOI: 10.1126/sciadv.abl8834

View the article online

<https://www.science.org/doi/10.1126/sciadv.abl8834>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science Advances (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

Supplementary Materials for

Essential role of Cp190 in physical and regulatory boundary formation

Anjali Kaushal, Julien Dorier, Bihan Wang, Giriram Mohana, Michael Taschner, Pascal Cousin, Patrice Waridel, Christian Iseli, Anastasiia Semenova, Simon Restrepo, Nicolas Guex, Erez Lieberman Aiden, Maria Cristina Gambetta*

*Corresponding author. Email: mariacristina.gambetta@unil.ch

Published 13 May 2022, *Sci. Adv.* **8**, eabl8834 (2022)
DOI: 10.1126/sciadv.abl8834

The PDF file includes:

Figs. S1 to S8
Tables S1 to S4
Legends for data S1 to S15

Other Supplementary Material for this manuscript includes the following:

Data S1 to S15

Figure S1

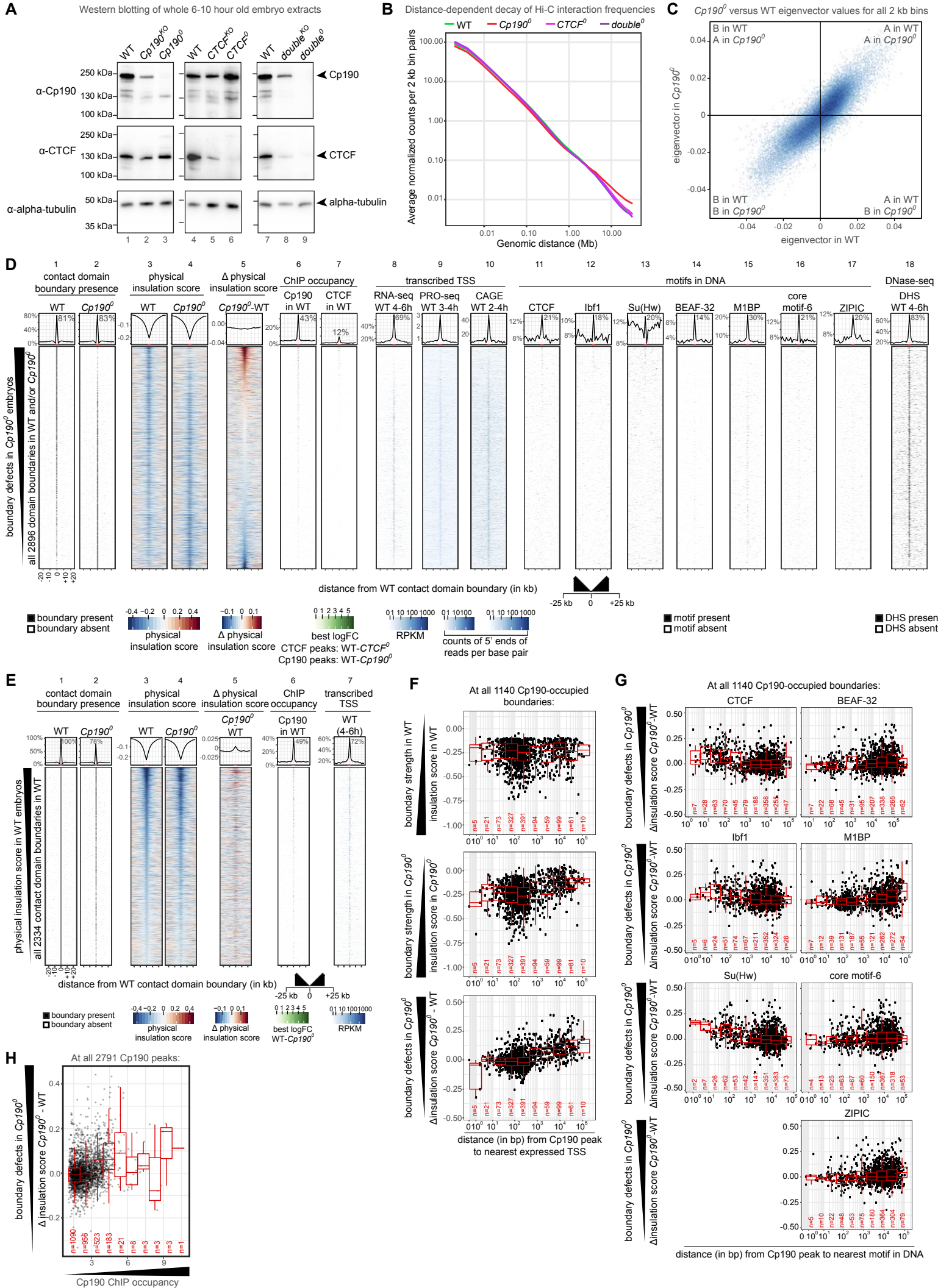


Fig. S1. Cp190 is required to form non-promoter boundaries in fly embryos. Related to Figure 1.

- A. Western blotting of whole cell extracts from 6-10 hour old embryos of indicated genotypes. Blotting membranes were cut at the 70 kDa marker height. The upper part was probed first with anti-CTCF and then with anti-Cp190, while the lower part was probed with anti-alpha-tubulin to verify similar loading of each extract. Each picture delimited by a black border is of the same membrane. No Cp190 and/or CTCF signal is detected in *Cp190⁰* (lane 3), *CTCF⁰* (lane 6), and *double⁰* (lane 9) embryo extracts. The reduced Cp190 and/or CTCF signals in *Cp190^{KO}* (lane 2), *CTCF^{KO}* (lane 5), and *double^{KO}* (lane 8) extracts represent maternally deposited proteins. The intense signals in wildtype extracts (lanes 1, 4, 7) represent summed maternally deposited and zygotically expressed CTCF and Cp190 proteins. CTCF levels seem normal in *Cp190⁰* (lane 3), and Cp190 levels seem normal in *CTCF⁰* (lane 6).
- B. Average normalized Hi-C counts per pair of 2 kb bins (in y) as a function of genomic distance (using exponentially increasing bins in x) per genotype (merged biological quadruplicates downsampled to 79 million contacts shown for separately colored genotypes). This shows that all genotypes show comparable distance-dependent decays of Hi-C interaction frequencies up to over 1 Mb.
- C. First eigenvector values in *Cp190⁰* (in y) versus WT (in x) for every 2 kb bin of chromosomes 2, 3 and X (density plot in blue, one point per 2 kb bin). Bins with positive eigenvector values are in the A (active) compartment, those with negative eigenvector values are in the B (inactive) compartment. Bins in the top left and bottom right quadrants are considered to be located in opposite compartments in *Cp190⁰* mutants relative to WT. This shows that compartments are not strongly affected in *Cp190⁰* mutant embryos.
- D. Like top of Fig. 1B, but for all 2896 contact domain boundaries identified in 2-6 hour old WT and/or *Cp190⁰* embryos ranked by strongest (top) to weakest (bottom) physical insulation defects in *Cp190⁰* relative to WT. (1-2) Presence of boundaries called in each genotype by TopDom in 2 kb bins around the boundary center. (3-4) Physical insulation scores measured in each genotype. (5) Physical insulation score differences measured in *Cp190⁰* minus WT by Hi-C. (6-7) Cp190 and CTCF ChIP occupancy in WT. (8-10) Expressed transcription start sites (TSSs) detected by RNA-seq in WT 4-6 hour embryos [Graveley et al. (75)], by PRO-seq in WT 3-4 hour embryos or by CAGE in 2-4 hour embryos [Mikhaylichenko et al. (76)]. (11-17) Indicated motifs in DNA. (18) DNase-hypersensitive sites (DHSs) mapped by DNA-seq in WT 4-6 hour embryos [Reddington et al. (78)]. Summarized values (average physical insulation score/counts of 5' ends measured by PRO-seq or CAGE, or percentage of WT boundaries with boundary/ChIP peak/transcribed TSS/DNA motif/DHS present) across 2 kb bins are plotted on top, and enrichment ± 2 kb around the central boundary (in red on x axis) is indicated. Color-coded value ranges are shown at the bottom. This shows that Cp190-dependent boundaries are generally TSS-distal, depleted for motifs enriched in promoter boundaries (BEAF-32,

M1BP, core motif-6, ZIPIC), and instead enriched in motifs found in non-promoter boundaries (CTCF, Ibf1, Su(Hw)). In contrast, boundaries only found in *Cp190⁰* but absent in WT are not visibly associated with any of the analyzed features in WT embryos, and their origins remain unclear.

- E. Like top of Fig. 1B, but all 2334 WT contact domain boundaries are ranked by strongest (top) to weakest (bottom) physical insulation scores measured in WT. (1-2) Presence of boundaries called in each genotype by TopDom in 2 kb bins around the boundary center. (3-4) Physical insulation scores measured in each genotype. (5) Physical insulation score differences measured in *Cp190⁰* minus WT by Hi-C, showing that Cp190 loss affects boundaries of variable strengths. (6) Cp190 ChIP occupancy in WT. (7) Expressed transcription start sites (TSSs) in WT 4-6 hour embryos [Graveley et al. (75)].
- F. Scatter plot of physical insulation scores measured by Hi-C in WT (top) or in *Cp190⁰* (middle) embryos, or of physical insulation score differences in *Cp190⁰* minus WT (bottom) (in y) at all 1140 Cp190-occupied boundaries (i.e. all boundaries within ± 2 kb of a Cp190 ChIP peak in WT, points) versus distance (in bp, in $\log_{(x+1)}$ transformed x axis) from the position of the most boundary-proximal Cp190 peak to the nearest transcribed TSS (RPKM>0) in WT 4-6 hour old embryos [expression data from Graveley et al. (75)]. Box plots of indicated n Cp190 peaks binned by distance to expressed TSS are overlaid. Box plot center line is median; box limits are upper and lower quartiles; whiskers are 1.5x interquartile ranges. This shows that Cp190-occupied boundaries that are close to or far from expressed TSSs have similar strengths in WT, but boundaries far from expressed TSSs are more severely affected by Cp190 loss. These scatter plots accompany the box plots shown in Fig. 1C.
- G. Scatter plot of physical insulation score differences measured in *Cp190⁰* minus WT Hi-C maps (in y) at all 1140 Cp190-occupied boundaries (points) versus distance (in bp, in $\log_{(x+1)}$ transformed x axis) from the position of the most boundary-proximal Cp190 peak to the center of the nearest indicated motif. Motifs enriched in non-promoter boundaries are in the first column; motifs enriched in promoter boundaries are in the second column. Box plots of indicated n Cp190 peaks binned by distance to motif are overlaid. This shows that insulation defects in *Cp190⁰* are higher closer to motifs enriched in non-promoter boundaries (CTCF, Ibf1, Su(Hw)) than to motifs enriched at promoter boundaries (BEAF-32, M1BP, core motif-6, ZIPIC). These scatter plots accompany the box plots shown in Fig. 1D.
- H. Scatter plot of physical insulation score differences measured in *Cp190⁰* minus WT Hi-C maps (in y) versus Cp190 ChIP occupancy (in x) for each WT Cp190 ChIP peak (points). Insulation scores are measured at the nearest bin boundary to the Cp190 peak position. Box plots of indicated n Cp190 peaks binned by ChIP occupancy are overlaid. This shows that insulation defects in *Cp190⁰* scale with Cp190 ChIP occupancy, except at rare very high occupancy Cp190 ChIP peaks.

Figure S2

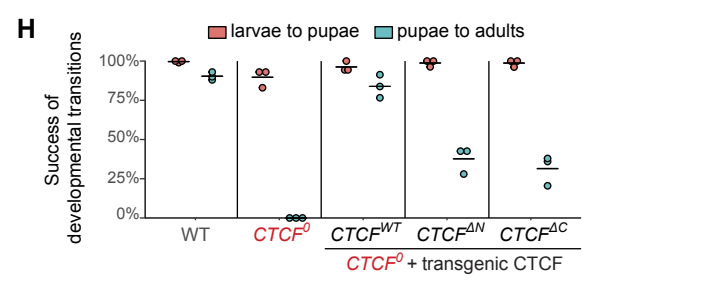
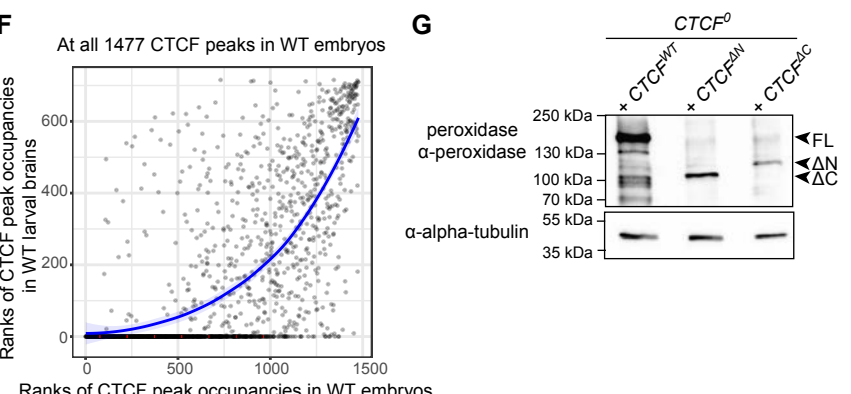
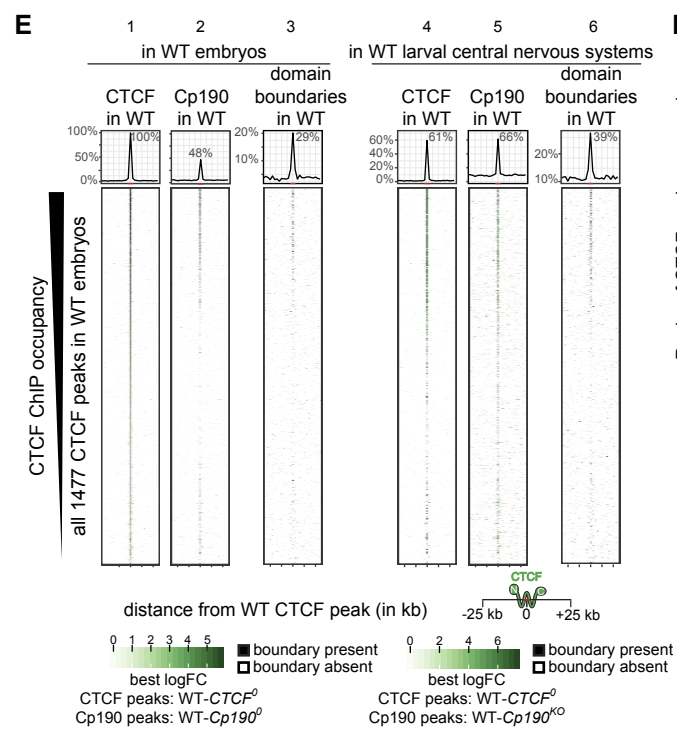
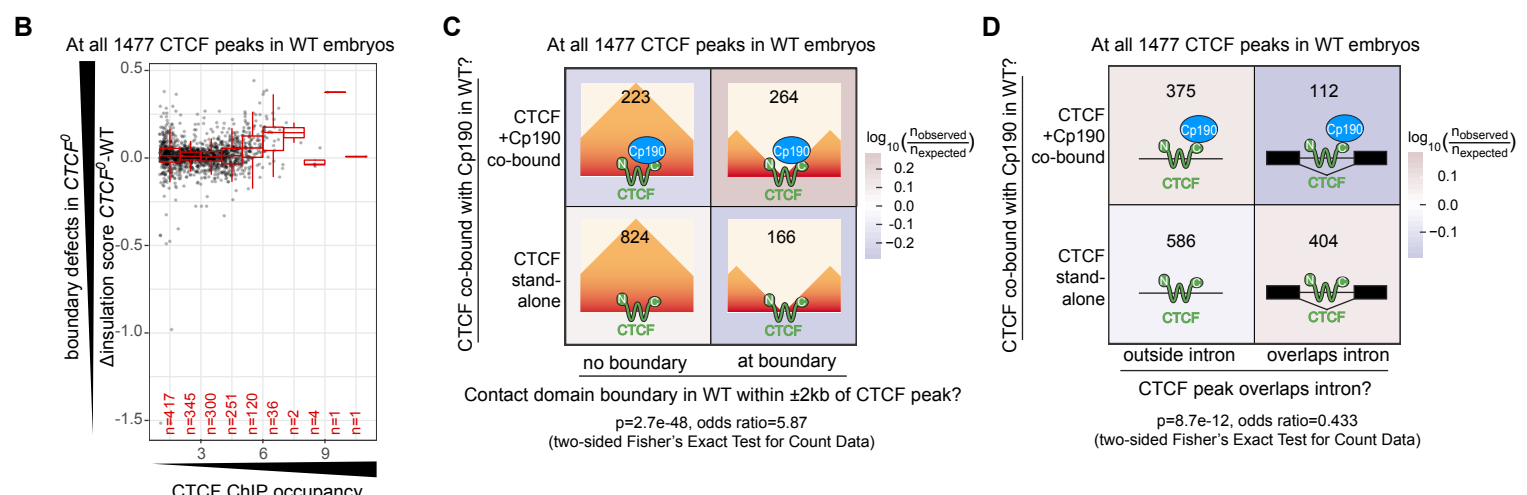
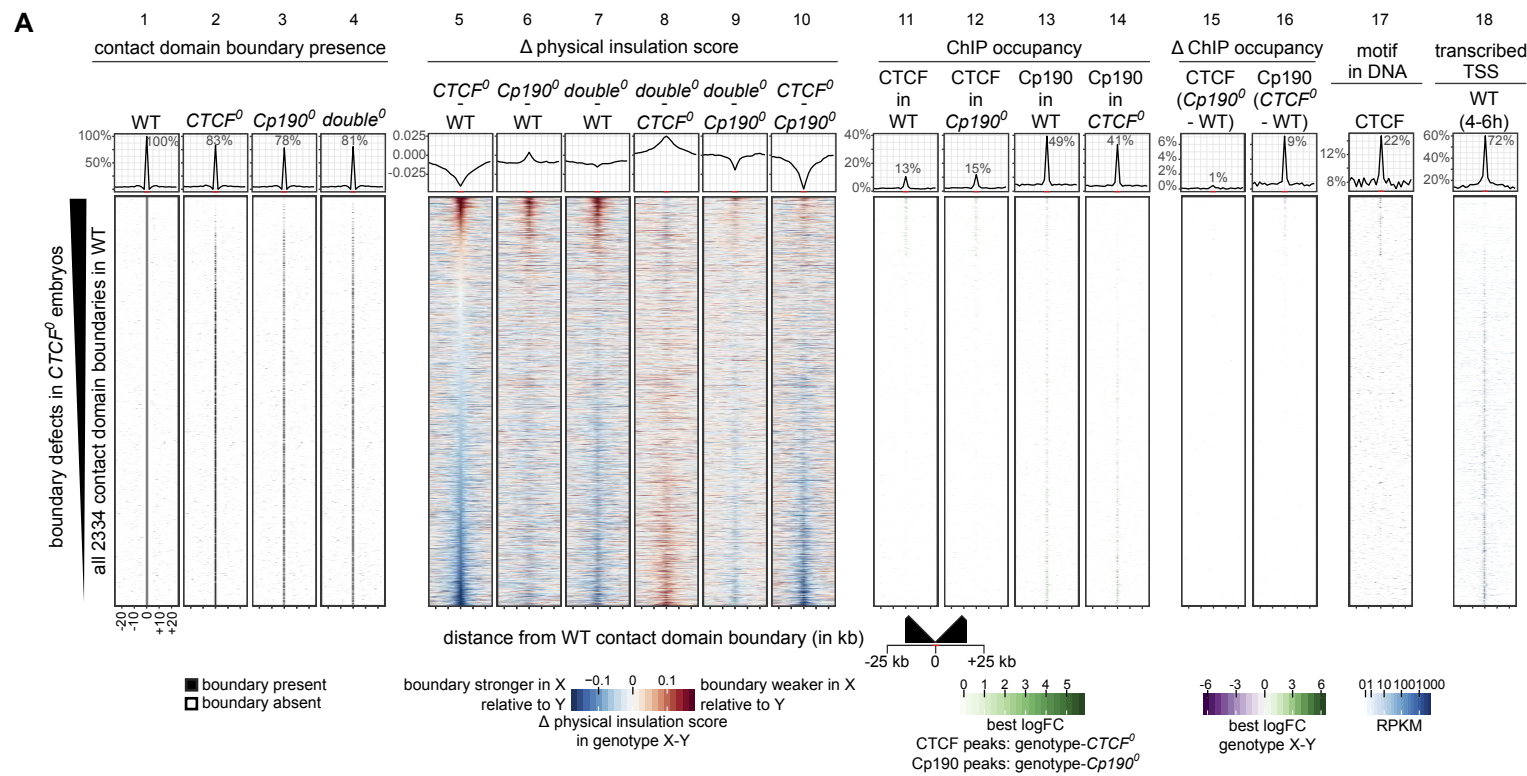


Fig. S2. Cp190 is required for boundary formation at CTCF peaks. Related to Figure 2.

- A. Distribution of indicated datasets in ± 25 kb windows centered around all 2334 contact domain boundaries identified in WT embryos ranked by strongest (top) to weakest (bottom) boundary defects measured in *CTCF⁰* mutant embryos. (1-4) Presence of contact domain boundaries called in each genotype by TopDom in 2 kb bins around the boundary center. (5-10) Physical insulation score differences measured in genotype X (top) minus genotype Y (bottom) by Hi-C. (11-12) CTCF or (13-14) Cp190 ChIP occupancy in indicated genotypes. (15-16) Differential CTCF and Cp190 ChIP occupancy in genotype X (top) minus genotype Y (bottom). (17) CTCF motif presence in DNA. (18) Expressed TSSs in WT 4-6 hour old embryos [Graveley et al. (75)]. Summarized values (average physical insulation score or percentage of WT boundaries with boundary/ChIP peak/differentially bound region/motif/transcribed TSS present) across 2 kb bins are shown above with indicated enrichments ± 2 kb around the central boundary. Color-coded value ranges are shown at the bottom. This shows that CTCF is present ± 2 kb from 13% of all WT boundaries (lane 11), and that boundary defects in *CTCF⁰* mutants (lanes 2, 5) are specifically observed at boundaries occupied by CTCF in WT (lane 11).
- B. Scatter plot of physical insulation score differences measured in *CTCF⁰* minus WT Hi-C maps (in y) versus CTCF ChIP occupancy (in x) for each CTCF peak (points). Insulation scores are measured at the nearest bin boundary to the CTCF peak position. Box plots of indicated n CTCF peaks binned by ChIP occupancy are overlaid. Box plot center line is median; box limits are upper and lower quartiles; whiskers are 1.5x interquartile ranges. This shows that insulation defects in *CTCF⁰* are observed at low and high occupancy CTCF ChIP peaks, but not at intermediate occupancy CTCF ChIP peaks (which are not at boundaries – see Fig. 2A).
- C. Numbers of CTCF peaks in WT (n=1477 peaks) that overlap a Cp190 peak in WT or not (rows), and have a contact domain boundary in WT within ± 2 kb of the peak position or not (columns). Cells are colored by $\log_{10}(\text{observed}/n_{\text{expected}})$, where n_{expected} is the expected value assuming independence of rows and columns (see Methods). CTCF+Cp190 co-localization is significantly positively associated with localization at a boundary (odds ratio and p-value are from two-sided Fisher's Exact Test for Count Data).
- D. Numbers of CTCF peaks in WT (n=1477 peaks) that overlap a Cp190 peak in WT or not (rows), and whose peak position is inside an intron or not (columns). Cells are colored by $\log_{10}(\text{observed}/n_{\text{expected}})$, where n_{expected} is the expected value assuming independence of rows and columns (see Methods). CTCF co-localization with Cp190 is significantly negatively associated with localization in an intron (odds ratio and p-value are from two-sided Fisher's Exact Test for Count Data).
- E. Occupancy of CTCF ChIP peaks in WT embryos (this study) compared to occupancy of CTCF peaks previously mapped in WT larval central nervous systems [from Kaushal et al.

(7)]. All 1477 WT CTCF peaks in embryos are ranked from highest (top) to lowest (bottom) ChIP occupancy (lane 1). Intermediate occupancy CTCF peaks frequently do not colocalize with Cp190 (lane 2) or with contact domain boundaries in WT (lane 3) (these panels were reproduced from Fig. 2A). Intermediate CTCF peaks in WT embryos are often low or zero occupancy CTCF peaks in larval central nervous systems (lane 4), and also overlap less frequently a Cp190 peak (lane 5) and a contact domain boundary (lane 6) in WT larval central nervous systems.

- F. Scatter plot of ranks of CTCF peak occupancies in WT larval central nervous systems (in y) versus ranks of CTCF peak occupancies in WT embryos (in x). Ranks are assigned by ordering non-zero peak occupancies by increasing order and assigning increasing ranks from 1 to the number of non-zero peak occupancies (see Methods). Ties are replaced by average rank, while rank 0 is assigned to unoccupied peaks. Each point is a CTCF peak in WT embryos, with CTCF peaks in WT larval central nervous systems [data from Kaushal et al. (7)] set to rank 0 when no CTCF peaks in WT larval central nervous systems overlap the corresponding CTCF peak in WT embryos. Loess fit (blue line) with 95% confidence interval (blue shade) is added. This illustrates that intermediate occupancy CTCF peaks in WT embryos are generally low occupancy CTCF peaks in larval central nervous systems.
- G. Western blotting of whole-cell extracts prepared from dissected larval central nervous systems of *CTCF⁰* animals rescued with transgenes expressing wildtype CTCF (*CTCF^{WT}*) or CTCF completely lacking N- (*CTCF^{ΔN}*) or C- (*CTCF^{ΔC}*) termini, under the control of *CTCF* regulatory sequences. *CTCF^{ΔN}* lacks amino acids 1-293, *CTCF^{ΔC}* lacks amino acids 610-818 (numbering according to Uniprot accession Q9VS55). *CTCF^{WT}* and *CTCF^{ΔN}* were expressed as C-terminal tandem affinity purification (TAP) tag fusions, *CTCF^{ΔC}* was expressed as an N-terminal TAP tag fusion. Extracts were probed with peroxidase anti-peroxidase antibody complex that binds to the TAP tag (top), then with anti-alpha-tubulin to verify equal loading of each extract. Observed sizes of CTCF versions are marked by arrowheads on the right.
- H. Percentages (in y) of wildtype (WT), *CTCF⁰*, and *CTCF⁰* animals rescued with *CTCF^{WT}*, *CTCF^{ΔN}*, or *CTCF^{ΔC}* transgenically expressed proteins that successfully transitioned from third instar larva to pupa (pink) and from pupa to adult (blue) in 3 biological replicates (each starting with between 83 and 100 larvae). Horizontal lines show means.

Figure S3

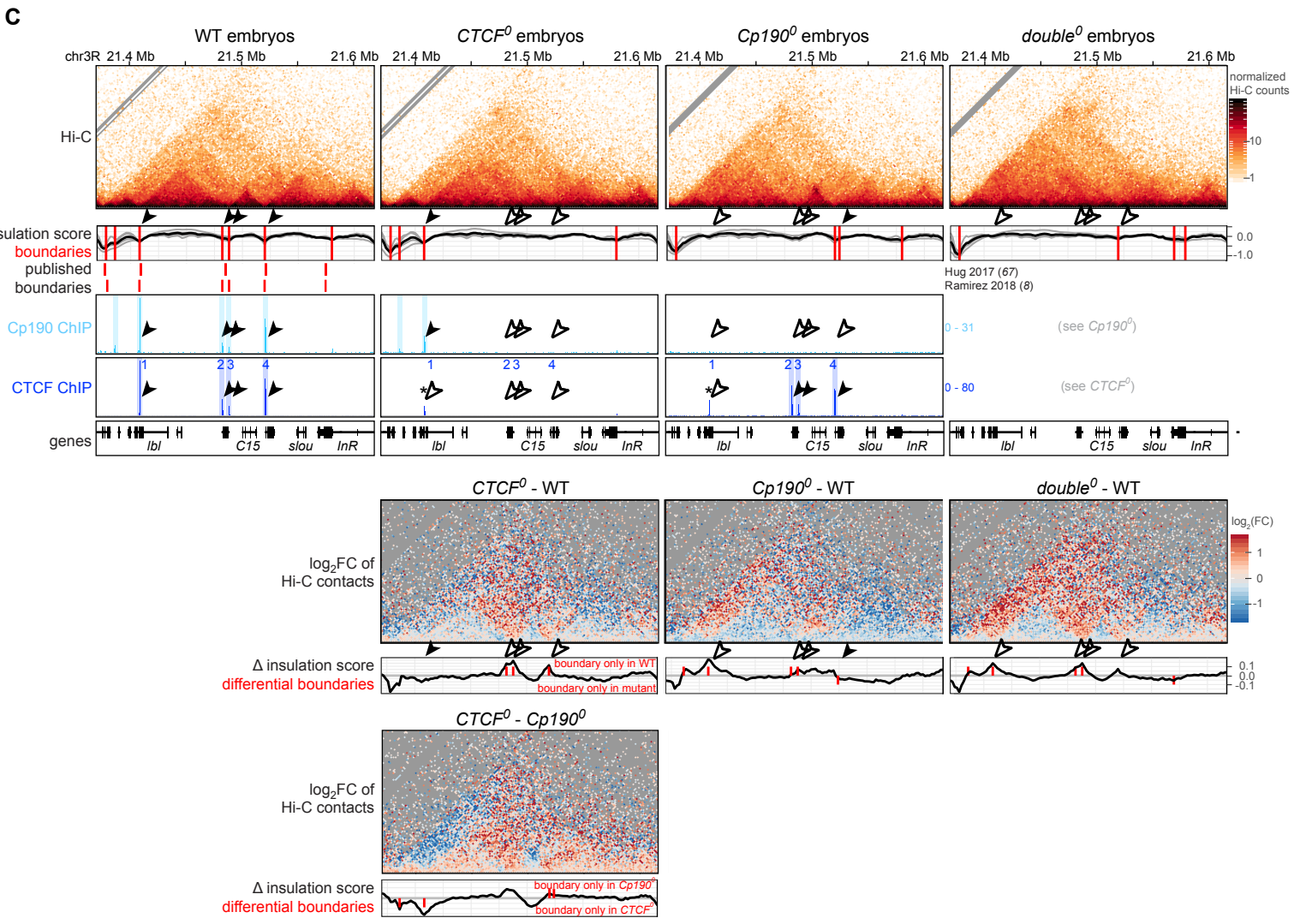
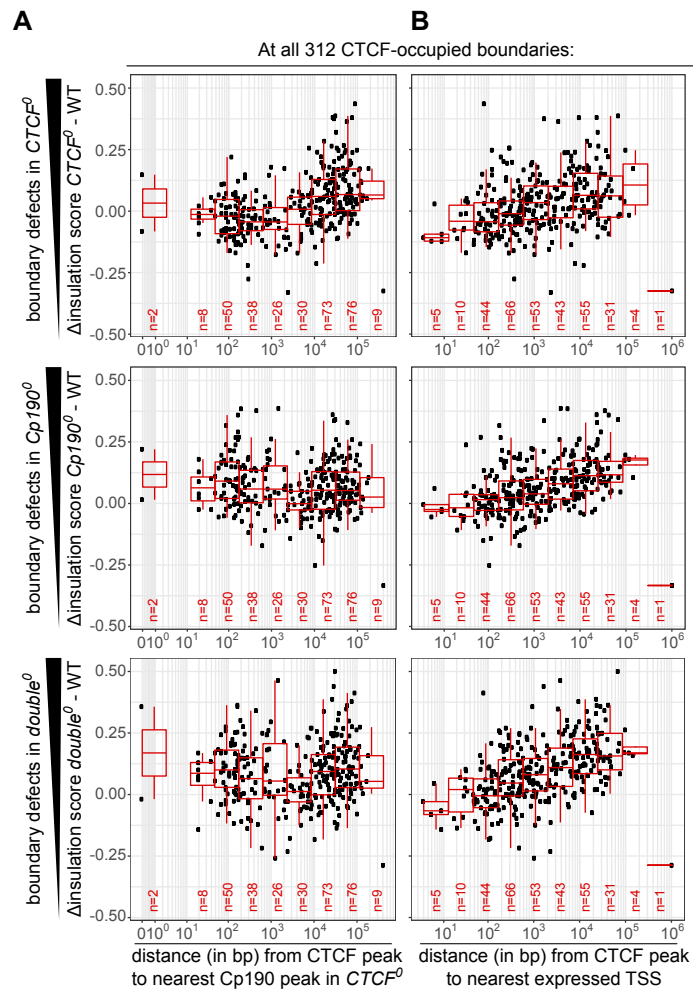


Fig. S3. CTCF-occupied boundaries are differentially sensitive to loss of CTCF or Cp190. Related to Figure 3.

- A. Scatter plot of physical insulation score differences measured by Hi-C in *CTCF⁰* (top), *Cp190⁰* (middle) or *double⁰* (bottom) minus WT (in y) at all 312 boundaries occupied by CTCF in WT (i.e. all boundaries within ± 2 kb of a CTCF ChIP peak in WT, points) versus distance (in bp, in $\log_{(x+1)}$ transformed x axis) from the position of the most boundary-proximal CTCF peak to the nearest Cp190 peak in *CTCF⁰*. Box plots of indicated n CTCF peaks binned by distance to the Cp190 peak are overlaid. Box plot center line is median; box limits are upper and lower quartiles; whiskers are 1.5x interquartile ranges. This shows that boundary defects in *CTCF⁰* are smaller when the boundary is close to a residual Cp190 peak, which is not the case in *Cp190⁰* or *double⁰*. These scatter plots accompany the box plots shown in Fig. 3C.
- B. Same as A but versus distance from the position of the most boundary-proximal CTCF peak to the nearest transcribed TSS (RPKM>0) in WT 4-6 hour old embryos [expression data from Graveley et al. (75)]. This shows that insulation defects in all genotypes are smaller at boundaries close to transcribed TSSs. These scatter plots accompany the box plots shown in Fig. 3D.
- C. Example locus (dm6 coordinates) Hi-C maps, physical insulation score (calculated with different window sizes in gray, average in black) and contact domain boundaries (vertical red lines) from this study (above) and published Hi-C studies in embryos [Hug et al. (67)] and tissue culture cells [Ramírez et al. (8)] (below), Cp190 ChIP-seq with Cp190 peaks defined in the respective genotype relative to *Cp190⁰* highlighted in light blue, CTCF ChIP-seq with CTCF peaks defined in the respective genotype relative to *CTCF⁰* highlighted in dark blue and numbered 1 to 4, and gene tracks (only longest isoform of each protein-coding gene shown) in embryos of the indicated genotypes. ChIP-seq scale is reads per million. Differential Hi-C maps (mutants minus WT in row 2, and *CTCF⁰* minus *Cp190⁰* in row 3), physical insulation score and contact domain boundaries are shown below. A non-specific CTCF ChIP-seq signal detected in *CTCF⁰* is marked by a black asterisk (the signal was higher in WT and thus called a CTCF peak, but it was not sufficiently high in *Cp190⁰* to be called a CTCF peak in the differential analysis). Arrowheads point to CTCF+Cp190 co-bound peaks located at domain boundaries in WT, while empty arrowheads indicate sites where these peaks are absent in the indicated genotypes. Cp190 co-localizing with CTCF peak 1 is partially CTCF-dependent because its peak height is reduced but the peak is still detected in *CTCF⁰*. Cp190 co-localizing with CTCF peaks 2-4 are strictly CTCF-dependent because these peaks are lost in *CTCF⁰*. This locus illustrates that contact domain boundaries at partially CTCF-dependent Cp190 peaks (peak 1) are more strongly affected in *Cp190⁰* than in *CTCF⁰*, whereas the strictly CTCF-dependent Cp190 peaks are more strongly affected in *CTCF⁰* than in *Cp190⁰*.

Figure S4

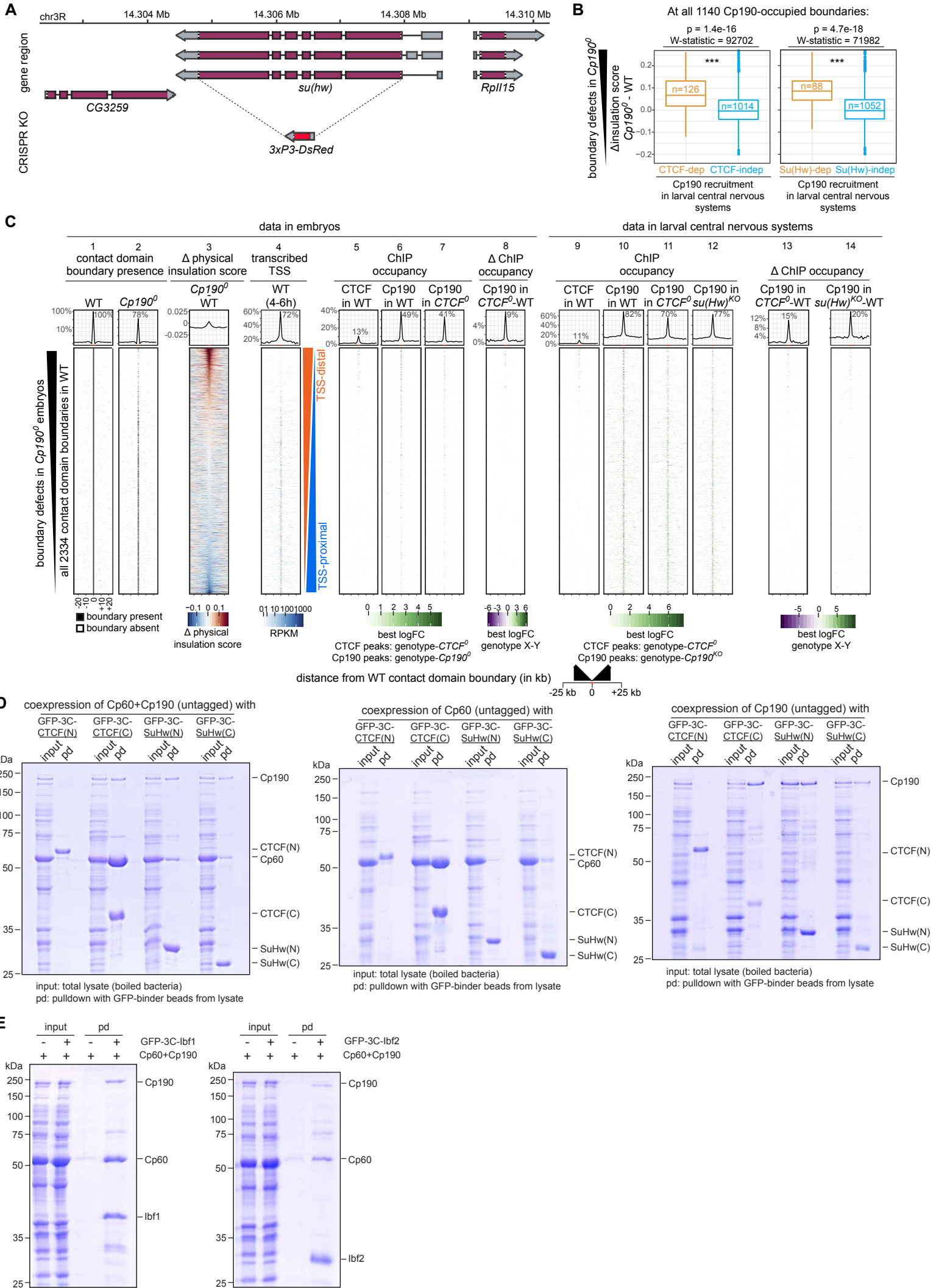


Fig. S4. Cp190 assembly into diverse multisubunit complexes. Related to Figure 4.

- A. *su(Hw)* extended gene region with coding (purple) and noncoding (grey) exons and introns (lines). In *su(Hw)^{KO}* mutants, a *DsRed* selection marker replaces the *su(Hw)* open reading frame.
- B. Boxplots of physical insulation score differences in *Cp190⁰* minus WT embryos at all 1140 contact domain boundaries occupied by Cp190 in WT (i.e. all boundaries within ± 2 kb of a Cp190 ChIP peak in WT) at which the Cp190 peak in embryos overlaps a CTCF-dependent Cp190 peak in WT larval central nervous systems [left, data from Kaushal et al. (7)] or a Su(Hw)-dependent Cp190 peak in WT larval central nervous systems (right). (Note that Su(Hw)-dependent Cp190 peaks had to be mapped in larvae instead of embryos because *su(Hw)⁰* mutants cannot be generated.) Indicated p value and W-statistic from two-sided Wilcoxon rank-sum test with continuity correction. Box plot center line is median; box limits are upper and lower quartiles; whiskers are $1.5\times$ interquartile ranges; points are outliers; n = Cp190-occupied boundaries of indicated categories.
- C. Distribution of indicated datasets in ± 25 kb windows centered around all 2334 contact domain boundaries identified in WT ranked by strongest (top) to weakest (bottom) physical insulation defects in *Cp190⁰* relative to WT. (1-2) Presence of contact domain boundaries called in each genotype by TopDom in 2 kb bins around the boundary center. (3) Physical insulation score differences measured in *Cp190⁰* minus WT by Hi-C. (4) Expressed TSSs in WT 4-6 hour old embryos [expression data from Graveley et al. (75)]. (5-7) CTCF and Cp190 ChIP occupancy in embryos of indicated genotypes. (8) Differential Cp190 ChIP occupancy in *CTCF⁰* minus WT embryos. (9-12) CTCF and Cp190 ChIP occupancy in WT, *CTCF⁰* or *su(Hw)^{KO}* larval central nervous systems [data in lanes 9 and 11 is from Kaushal et al. (7)]. (13-14) Differential Cp190 ChIP occupancy in *CTCF⁰* or *su(Hw)^{KO}* mutants minus WT larval central nervous systems. Summarized values (average physical insulation score or percentage of WT boundaries with boundary/ChIP peak/differentially bound region/transcribed TSS present) across 2 kb bins are shown above with indicated enrichments ± 2 kb around the central boundary (highlighted in red on x axis). Color-coded value ranges are shown at the bottom. Lanes 1-4 and 6 were reproduced from Fig. 1B. This shows that Cp190 is recruited to Cp190-dependent boundaries by CTCF or Su(Hw).
- D. GFP pull-down of tagged CTCF N-terminus (residues 1-123), CTCF C-terminus (residues 610-818), Su(Hw) N-terminus (residues 1-219), or Su(Hw) C-terminus (residues 724-941), each co-expressed with untagged Cp190 and Cp60 (left), Cp60 alone (middle) or Cp190 alone (right). Positions of molecular weight ladder bands are marked on the left, those of co-expressed proteins are marked on the right. Amino acid numbering is based on Uniprot entries Q9VS55 (CTCF) and P08970 (Su(Hw)).

E. GFP pull-down of tagged full-length Ibf1 (left) or Ibf2 (right), each co-expressed with untagged Cp190 and Cp60. GFP pull-down in the absence of Ibf1 or Ibf2 is shown as negative control. Positions of molecular weight ladder bands are marked on the left, those of co-expressed proteins are marked on the right.

Figure S5

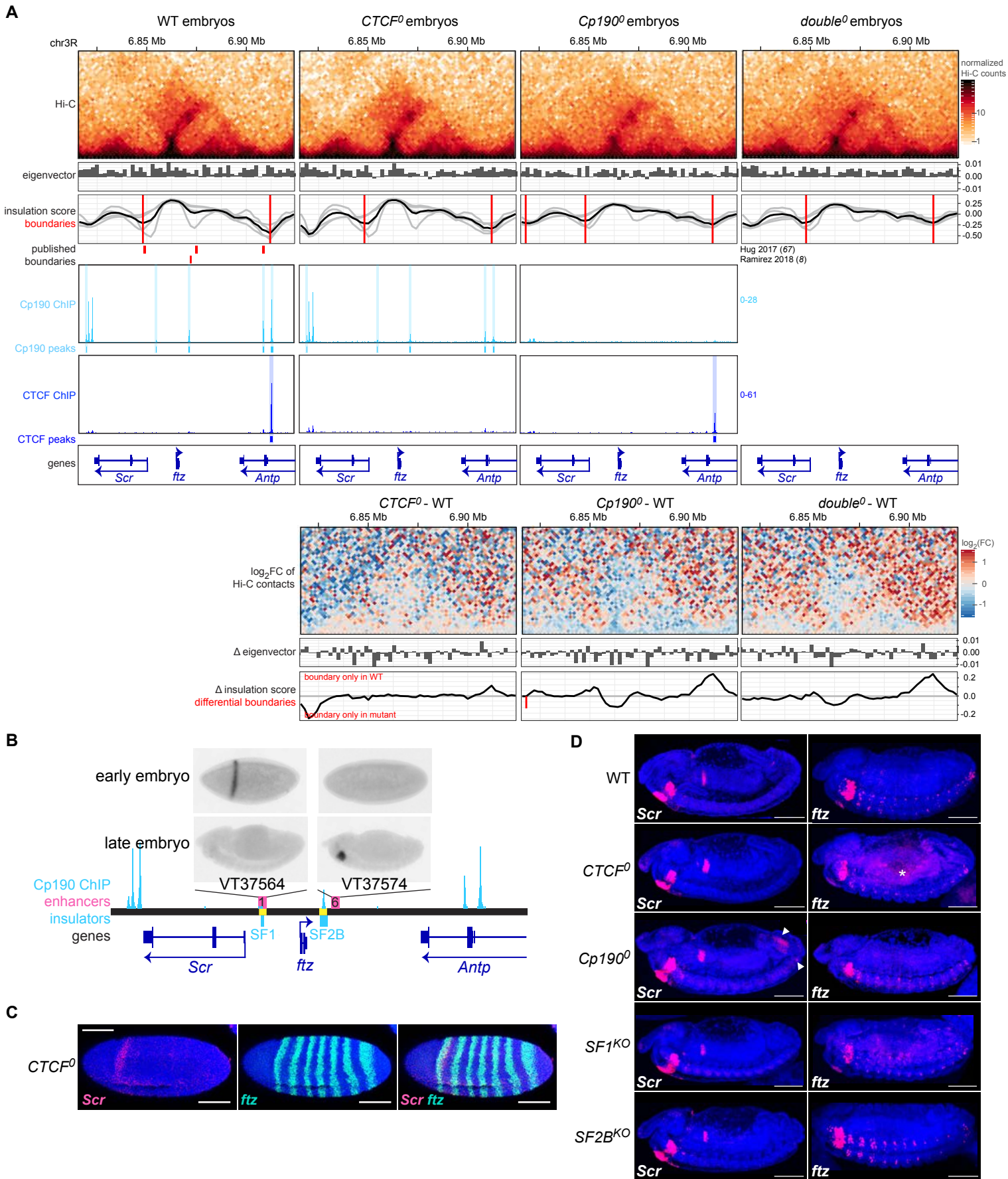


Fig. S5. Cp190 prevents regulatory crosstalk between early patterning gene loci. Related to Figure 5.

- A. Same locus shown in Fig. 5A additionally showing eigenvector values (2 kb resolution, positive for A compartment, negative for B compartment), and CTCF ChIP-seq with CTCF peaks defined in the respective genotype relative to *CTCF*⁰ in dark blue, in embryos of the indicated genotypes. Differential Hi-C maps, physical insulation scores and contact domain boundaries in the respective mutants minus WT are shown below. Note limitations in TopDom boundary calls at this locus: the *ftz* upstream boundary was called but is shifted to the left of its visible position in the Hi-C map, and the *ftz* downstream boundary was not called robustly by TopDom and was hence subsequently filtered out in our analysis (see Methods). *ftz* boundaries are therefore visible by Hi-C but were not called because they challenge the classical definition of boundaries due to the unusually strong inter-domain interactions occurring across them at this exceptional locus.
- B. Same locus shown in A. Enhancer 1 (also shown in Fig. 5B) is VT37564 [Kvon et al. (53)] and drives expression in early gastrulae overlapping *Scr* expression (i.e. it is expressed in a stripe just posterior to the cephalic furrow like *Scr*), as shown on top. Enhancer 1 is silent in older embryos (shown below), and *SFI*^{KO} older embryos consistently express *Scr* normally [as shown in panel D row 1 and previously reported by Yokoshi et al. (59)]. Enhancer 6 is VT37574 [Kvon et al. (53)] and drives expression in the labial segment overlapping *Scr* expression (compare to *Scr* expression at the same embryonic stage shown in panel D row 1), consistent with the possibility that it may be a distal *Scr* enhancer in older embryos. Enhancer 6 is silent in young embryos, further suggesting that it is unlikely to be responsible for early *Scr* expression. Pictures of reporter gene expression driven by these enhancers were obtained from <https://enhancers.starklab.org/>.
- C. Like Fig. 5C. RNA-FISH with co-hybridized antisense probes against *Scr* (red) and *ftz* (green) mRNAs in early gastrula *CTCF*⁰ embryos stained with DAPI (blue) and imaged from the side (anterior left, posterior right, scale bars 100 μm). Single *Scr* and *ftz* images are shown on the left, and merged images of the same embryo is shown on the right. This shows that *Scr* and *ftz* appear normally expressed in *CTCF*⁰ embryos.
- D. RNA-FISH with antisense probes (red) against *Scr* (left) or *ftz* mRNAs (right) in stage 14 (mid-embryogenesis) embryos stained with DAPI (anterior left, posterior right, scale bars 100 μm). In WT embryos, *Scr* is expressed in anterior (labial and prothoracic) segments and the anterior midgut. *Scr* is expressed normally in all genotypes (rows) except *Cp190*⁰ mutants which misexpress *Scr* in the hindgut and anal plate (arrowheads). Misexpression in anal plate becomes stronger in older embryos (see Fig. 6C). In WT embryos, *ftz* is expressed in a subset of cells of the ventral nerve cord. *ftz* is expressed normally in all genotypes (white asterisk marks background staining).

Figure S6

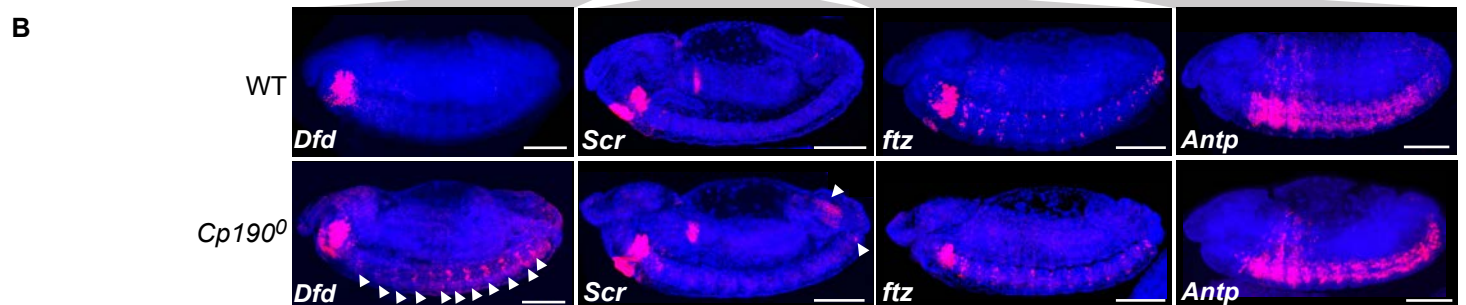
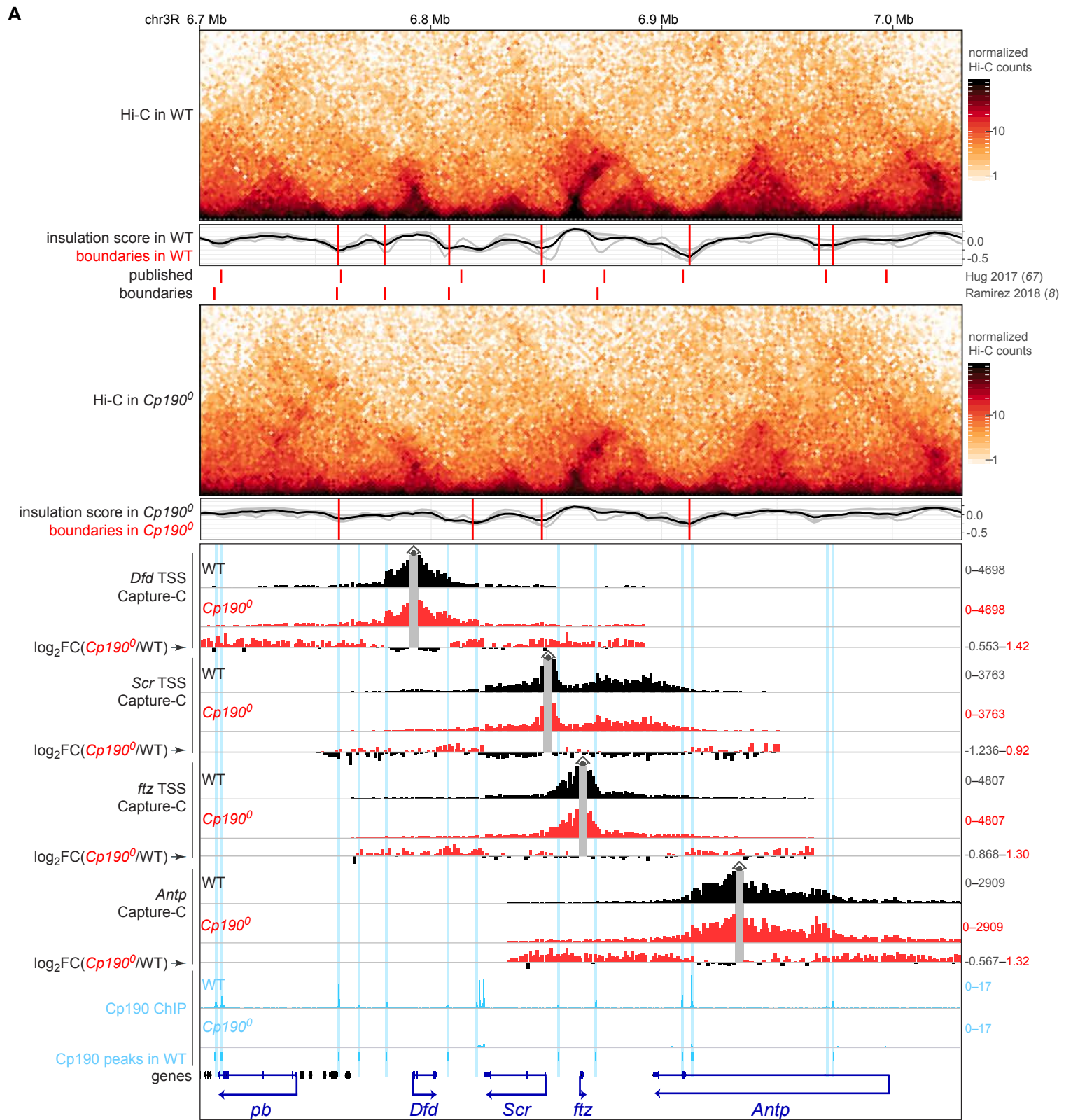
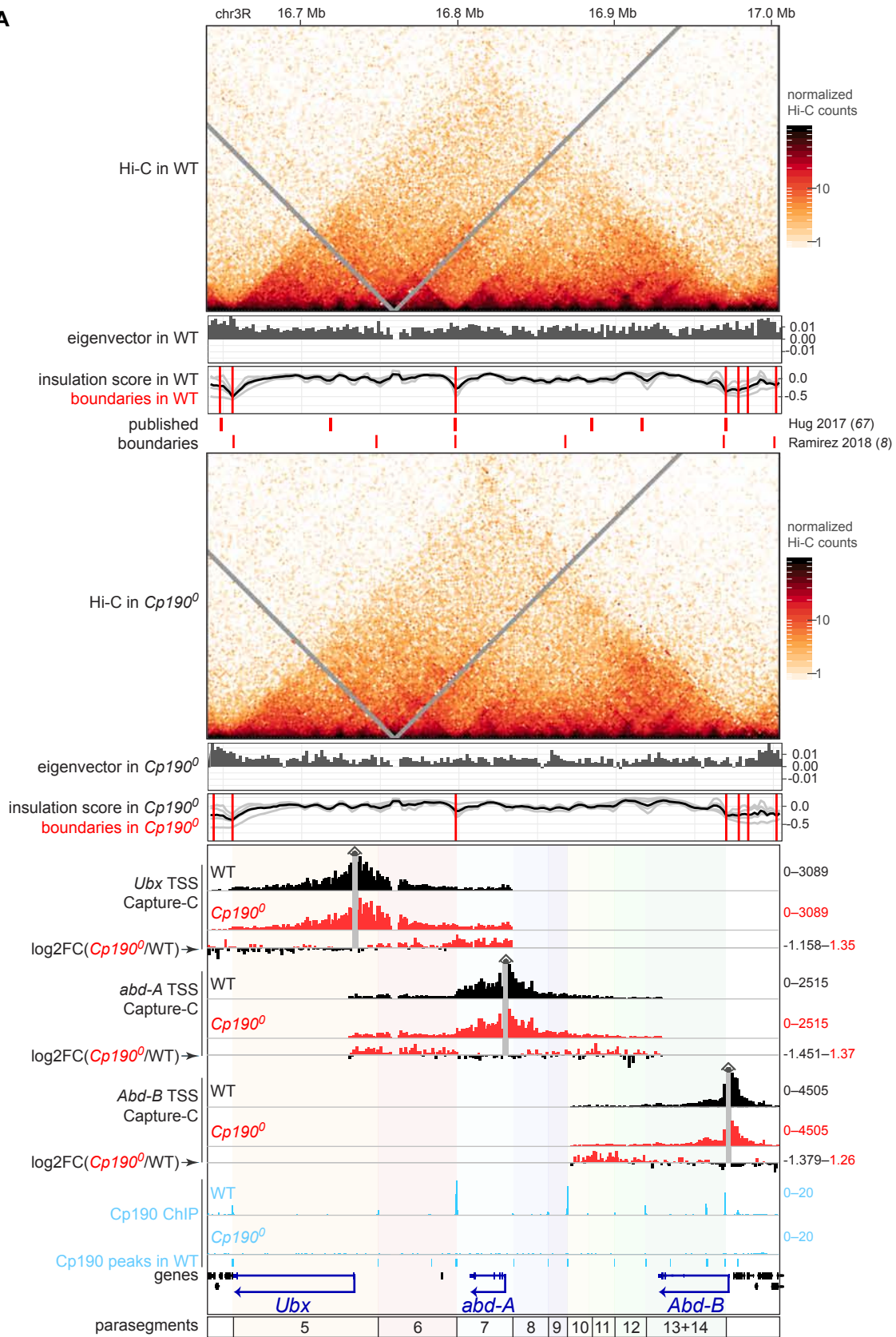


Fig. S6. ANT-C HOX gene expression in *Cp190⁰* mutants. Related to Figure 6.

- A. Portion of the ANT-C locus (dm6 coordinates) Hi-C maps (2 kb resolution), physical insulation score (calculated with different window sizes in gray, average in black) and contact domain boundaries (vertical red lines) from this study (above) and published Hi-C studies in WT 3-4 hour old embryos [Hug et al. (67)] and tissue culture cells [Ramírez et al. (8)] (below), NG Capture-C profiles binned at 1 kb resolution around *Dfd*, *Scr*, *ftz* and *Antp* viewpoints in WT (top) and *Cp190⁰* (bottom) embryos. Average normalized Capture-C reads (in reads per million) obtained from biological triplicates of each genotype are shown, excluding bins ± 2 kb around the viewpoint (grey rectangles below eye symbols). Differences in Capture-C profiles in *Cp190⁰* versus WT are shown as \log_2 fold-change profiles obtained from diffHic, with ratios >0 indicating increased contacts in *Cp190⁰*, and values <0 indicating decreased contacts in *Cp190⁰*. Below Capture-C profiles, Cp190 ChIP-seq in indicated genotypes (in reads per million), Cp190 peaks defined as enriched in WT relative to *Cp190⁰* in blue, and gene tracks (only longest isoform of each protein coding gene shown, homeobox genes are blue). This shows that for each viewpoint in *Cp190⁰* mutants, Capture-C contacts are qualitatively weakly increased in broad contiguous regions beyond former Cp190 peaks at the expense of intra-domain contacts.
- B. RNA-FISH with antisense probes (red) against *Dfd*, *Scr*, *ftz* or *Antp* mRNAs in stage 14 (mid-embryogenesis) embryos stained with DAPI (blue) and imaged from the side (anterior left, posterior right, scale bars 100 μm). *Scr* and *ftz* images are reproduced from Fig. S5D for completeness. In WT embryos, *Dfd* is expressed in mandibular and maxillary segments. In *Cp190⁰* embryos, *Dfd* is additionally misexpressed in the nervous system (arrowheads). In WT embryos, *Antp* is expressed in anterior segments and in the nerve cord. In *Cp190⁰* embryos, *Antp* appears normally expressed.

Figure S7

A



B

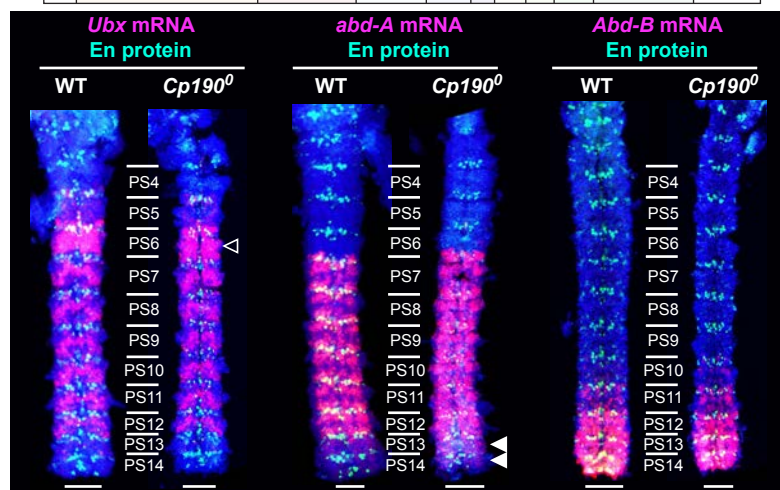


Fig. S7. Cp190 is dispensable for abdominal HOX gene activation by long-range enhancers. Related to Figure 6.

- A. Bithorax-complex locus (dm6 coordinates) Hi-C maps (2 kb resolution), eigenvector values (2 kb resolution, positive for A compartment, negative for B compartment), physical insulation score (calculated with different window sizes in gray, average in black) and contact domain boundaries (vertical red lines) from this study (above) and published Hi-C studies in WT 3-4 hour old embryos [Hug et al. (67)] and tissue culture cells [Ramírez et al. (8)] (below), NG Capture-C profiles binned at 1 kb resolution around *Ubx*, *abd-A* and *Abd-B* TSS viewpoints in WT (top) and *Cp190⁰* (bottom) embryos. Average normalized Capture-C reads (in reads per million) obtained from biological triplicates of each genotype are shown, excluding bins ± 2 kb around the viewpoint (grey rectangles below eye symbols). Differences in Capture-C profiles in *Cp190⁰* versus WT are shown as \log_2 fold-change profiles obtained from diffHic, with ratios >0 indicating increased contacts in *Cp190⁰*, and values <0 indicating decreased contacts in *Cp190⁰*. Below Capture-C profiles, Cp190 ChIP-seq in indicated genotypes (in reads per million), Cp190 peaks defined as enriched in WT relative to *Cp190⁰* in blue, and gene tracks (only longest isoform of each protein coding gene shown, homeobox genes are blue). Enhancer domains driving expression of *Ubx* (shades of orange), *abd-A* (shades of blue) and *Abd-B* (shades of green) in the indicated parasegments are colored. This shows that in *Cp190⁰* mutants, Capture-C contacts are qualitatively weakly increased in broad contiguous regions beyond former Cp190 peaks at the expense of intra-domain contacts.
- B. Ventral nerve cords dissected from stage 15 embryos (at mid-embryogenesis) (oriented with anterior up) of the indicated genotypes subjected to RNA-FISH with probes against the indicated abdominal HOX gene mRNAs (red), followed by immunostaining with anti-Engrailed (green) to mark parasegment boundaries and DAPI-labeling (blue) of DNA. Parasegments (PS) are labeled. Phenotypes consistently seen in *Cp190⁰* mutants are marked by arrowheads: the empty arrowhead shows lower *Ubx* mRNA levels in PS6 in *Cp190⁰* than in WT, and solid arrowheads show higher *abd-A* mRNA levels in PS13 and PS14 in *Cp190⁰* than in WT. Scale bars below each nerve cord show 50 μm .

Figure S8

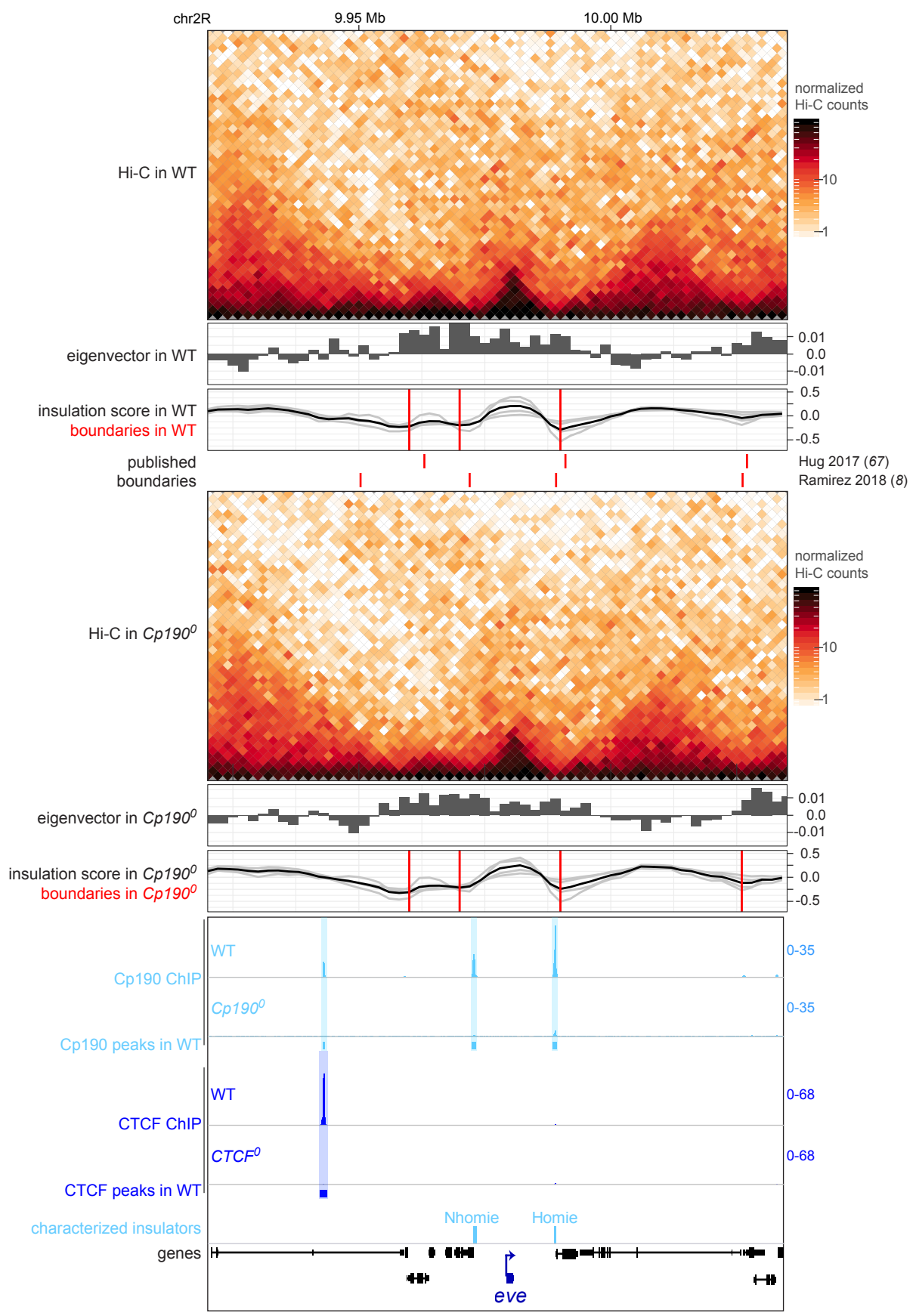


Fig. S8. Homie overlaps a Cp190-occupied boundary. Related to Figure 7.

eve locus (dm6 coordinates) Hi-C maps (2 kb resolution), eigenvector values (2 kb resolution, positive for A compartment, negative for B compartment), physical insulation score (calculated with different window sizes in gray, average in black) and contact domain boundaries (vertical red lines) from this study (above) and published Hi-C studies in WT 3-4 hour old embryos [Hug et al. (67)] and tissue culture cells [Ramírez et al. (8)] (below), Cp190 ChIP-seq (in reads per million), Cp190 peaks defined as enriched in WT relative to *Cp190⁰* in light blue, CTCF ChIP-seq, CTCF peaks defined as enriched in WT relative to *CTCF⁰* in dark blue, characterized Nhomie and Homie insulators [Fujioka et al. (37)], and gene tracks (only longest isoform of each protein-coding gene shown, homeobox genes are blue). This shows that Homie overlaps a Cp190-occupied boundary in WT that is not visibly affected in *Cp190⁰* mutants.

Hi-C sample	total reads	interchr	intrachr	intrachr (<20 kb)	intrachr (≥20 kb)
WT_embryo_ rep1234_79M	79000000	14918886	64081114	24847640	39233474
Cp190 ⁰ _embryo_ rep1234_79M	79000000	20871310	58128690	22155105	35973585
CTCF ⁰ _embryo_ rep1234_79M	79000000	15880776	63119224	26388098	36731126
double ⁰ _embryo_ rep1234_79M	79000000	13048469	65951531	27630124	38321407

Table S1. Quality metrics of Hi-C reads.

Primer name	Primer sequence (5'-3')	Purpose
dmscr_1	CGAATTTATGACTGTCCATCA GCGCCAG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_2	GCACACGCGCAATTTGTTAC AGATCACG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_3	AGCGACGGATTTGGACGGCT TTTGGTAT	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_4	TTGCGTCCTACCTGCAGGCTT ATGTTTTG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_5	GCGGAAATCAACGTGCACAA TTACGAAGG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_6	GCACTGAACCACGATTTTCA CTTGTGCTGAG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_7	GCGTTCACCTCCGTTCAAGT GAGTTTGTTT	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_8	ATTTTAAAGCCAGGGGTCGT TGTCGTGG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_9	GAAGTGGTACGAGGACATCG CAAACAG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_10	TCGTTGGCGTACTTGCAACTG ATGTTTCG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_11	TTCTTTCCATTGCCGCTGTTT TGGCTACTG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_12	TTGTCGTTTCGTCTCGCCATT GGCATT	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_13	AGTCTCTAACCAGTACCCGA AAAGTGCC	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_14	GAGATCAGAACTCCTGCGGA TACTTGATG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_15	AGTTTTAGGTCCCGTCTCTGA TTCCGAT	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_16	TTCGACAATGTCCGCCTTGA ACTCCAGTT	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_17	TTAAAGAGTGACGTCGATTG CTCGTGGTTCC	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_18	CCCTCGAGCATTTCGCATAGA AAGGTTTAG	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_19	CTGCTCATCGTGCAGCTTACG TGCTAAAA	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmscr_20	GTTGGAGATTGGGCGATACA AACGAAGAC	MUSE-RNA-FISH probe against <i>Scr</i> mRNA
dmftz_1	CTTCTCTAACTCTGCGATGTG CACGCAACG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA

dmftz_2	TTGCTGCCTGAATTATCGTAG TAGGTGG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_3	GCCCTGATAATTGGAGGTGT TCTGATAG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_4	AGTAGCAGCTCTCCGAGTAA CTCTCCTG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_5	AAGCAGCATCATCTTCGGCC TTGCGCTT	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_6	CTTCAGCTTCTTCACGGGATT GGTGAGC	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_7	CTGCTCGACGGTTGTGTAGA AATAGTCGG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_8	ATTCGATGATTGATCTCCTGG CTGACAG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_9	CTCCTCGATGTGCGACCAATT GAAATCG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_10	GCGTTTCGAGTCTTTGCAATC TGATGCCAAAG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_11	TCTTGATCTGCCTTTCGCTCA GGCTCAG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_12	CTCTGGGGAAGAGAGTAACT GAGCATCG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_13	AATGGTTCGAGAGAAGTGCGC TTCGGTTTCGT	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_14	CTTCGTTCTCGGCTGTGTCAT TTGCGTG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_15	TCTTTTGCCTCTGCCTTCTGC ACTTGCG	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_16	AGCAAGGCTCCTTTTCTGTTT GCGCTGC	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_17	AGCTAATCGATCGCTGAGAA CCCATC	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
dmftz_18	AGCTGAGTGTTTTGGGCTTGT GTTTGGC	MUSE-RNA-FISH probe against <i>ftz</i> mRNA
CTCF A fw	CTCCGGAATATTAGGTCTCAT ACTTTTCATTTCAATTTG CGG	CTCF-bound insulator cloning (Fig. 4b)
CTCF A rv	GGCTCAAGCAGTGGGTCTCC CATTTTTTGGTTCATATGAAG CGC	CTCF-bound insulator cloning (Fig. 4b)
CTCF B fw	CTCCGGAATATTAGGTCTCAT ACTGCATCATTTTGTAGTTGT CC	CTCF-bound insulator cloning (Fig. 4b)

CTCF B rv	GGCTCAAGCAGTGGGTCTCC CATTAAATTGGGAAATAAACT CTAGC	CTCF-bound insulator cloning (Fig. 4b)
Su(Hw) A' fw	CTCCGGAATATTAGGTCTCAT ACTAGCCCTGCAACTCAATG G	Su(Hw)-bound insulator cloning (Fig. 4b)
Su(Hw) A' rv	GGCTCAAGCAGTGGGTCTCC CATTATGCCAATTGACTGCAT GG	Su(Hw)-bound insulator cloning (Fig. 4b)
Su(Hw) B' fw	CTCCGGAATATTAGGTCTCAT ACTTTGGGGATTGTTTGTAAT GTTG	Su(Hw)-bound insulator cloning (Fig. 4b)
Su(Hw) B' rv	GGCTCAAGCAGTGGGTCTCC CATTGCTGGCAACATTTTAG TGG	Su(Hw)-bound insulator cloning (Fig. 4b)
Su(Hw) C' fw	CTCCGGAATATTAGGTCTCAT ACTGGTTTTACATTTGACTTC TTCG	Su(Hw)-bound insulator cloning (Fig. 4b)
Su(Hw) C' rv	GGCTCAAGCAGTGGGTCTCC CATTGATCGGCTAAGTGAA GTGG	Su(Hw)-bound insulator cloning (Fig. 4b)
Su(Hw) D' fw	CTCCGGAATATTAGGTCTCAT ACTCGCACTTCCCCACCAAC C	Su(Hw)-bound insulator cloning (Fig. 4b)
Su(Hw) D' rv	GGCTCAAGCAGTGGGTCTCC CATTGCTGGCGGCATACAAA TAG	Su(Hw)-bound insulator cloning (Fig. 4b)
Chro/Pzg/BEAF A" fw	CTCCGGAATATTAGGTCTCAT ACTGTCACAGGATTGCTGGT GG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF A" rv	GGCTCAAGCAGTGGGTCTCC CATTAAATTGTGTTTCACTGCT TTGC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF B" fw	CTCCGGAATATTAGGTCTCAT ACTAACAACAAGCGCATATG TTGG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF B" rv	GGCTCAAGCAGTGGGTCTCC CATTAGCGAATAATTTAATTC ACACC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF C" fw	CTCCGGAATATTAGGTCTCAT ACTTCCGAAAATTTCTTCTCT GACC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)

Chro/Pzg/BEAF C" rv	GGCTCAAGCAGTGGGTCTCC CATTCAAGAAAGTATATAAA ATGCCG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF D" fw	CTCCGGAATATTAGGTCTCAT ACTGCTGTTGTTGTTTCGAAAT GTTG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF D" rv	GGCTCAAGCAGTGGGTCTCC CATTCACTTGCCTTAGAC GCGGC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF E" fw	CTCCGGAATATTAGGTCTCAT ACTGACAAAAGTCCACAAA ATC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF E" rv	GGCTCAAGCAGTGGGTCTCC CATTGAAATGTAATGCATTC GACTC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF F" fw	CTCCGGAATATTAGGTCTCAT ACTAACGTTTTGCCGGCTAC GC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF F" rv	GGCTCAAGCAGTGGGTCTCC CATTGGCCATGTTAAAGTCT GTG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF G" fw	CTCCGGAATATTAGGTCTCAT ACTTGTCTATCCTCTTCCTT GG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF G" rv	GGCTCAAGCAGTGGGTCTCC CATTGCCCAACTCGACGCTA GC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF H" fw	CTCCGGAATATTAGGTCTCAT ACTCATGTTTAACCAACTTAC CTAG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF H" rv	GGCTCAAGCAGTGGGTCTCC CATTGGCACAAAACAATGGC ATTTG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF I" fw	CTCCGGAATATTAGGTCTCAT ACTACTAGCAACAAGACGCA C	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF I" rv	GGCTCAAGCAGTGGGTCTCC CATTATGGCTTTCGCTTGTTG ATTCC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF J" fw	CTCCGGAATATTAGGTCTCAT ACTGCTCCAAAATTTCCCGCT CG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)

Chro/Pzg/BEAF J" rv	GGCTCAAGCAGTGGGTCTCC CATTAACAAATTTCCCTGCA TAG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF K" fw	CTCCGGAATATTAGGTCTCAT ACTCCGCTGCTTATACGCACC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF K" rv	GGCTCAAGCAGTGGGTCTCC CATTTACCTTCAGTACTTAAT GTTCC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF L" fw	CTCCGGAATATTAGGTCTCAT ACTTCTGCAGTGCTCGAAAG TGTC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF L" rv	GGCTCAAGCAGTGGGTCTCC CATTGGCGAATTTGCGTGAA ATTG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF M" fw	CTCCGGAATATTAGGTCTCAT ACTGAATGCACACTTGCAAT TTGCG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF M" rv	GGCTCAAGCAGTGGGTCTCC CATTGCGGCTGACGCCTTAT AAAC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF B" mut1 fw	CGATAAGCCAAACATATGC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF B" mut1 rv	CTAGTCTGGAGCTATTGCAG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF B" mut2 fw	CGATTGTTTTTGTGAAAAGG	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF B" mut2 rv	CTATTACCTCACCGCCTAAA AC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF F" mut1 fw	CGATATTTAATTGCGTAGCC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF F" mut1 rv	CTAACACCACTGTCATACC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF F" mut2 fw	CGATTGTTTGTACATGGATT G	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
Chro/Pzg/BEAF F" mut2 rv	CTAGACCGCTGCTGCCCCAC	Chro/Pzg/BEAF-3 insulator cloning (Fig. 4b)
su(Hw) L sgRNA fw	GTCGCCTTGGAGGCACTCAT GT	su(Hw) sgRNA left
su(Hw) L sgRNA rv	AAACACATGAGTGCCTCCAA GG	su(Hw) sgRNA left
su(Hw) R sgRNA fw	GTCGGTACGCATGCATCAAG GGC	su(Hw) sgRNA right

su(Hw) R sgRNA rv	AAACGCCCTTGATGCATGCG TAC	su(Hw) sgRNA right
su(Hw) LHA fw	GTCGACGAATTCTCCGCCTCC TGTTTGTAATC	su(Hw) homology arm left
su(Hw) LHA rv	GTCGACgcgggcgcTGTTGGTGA TACCAGCctg	su(Hw) homology arm left
su(Hw) RHA fw	GTCGACACGCGTCTTGATGC ATGCGTACAATTACC	su(Hw) homology arm right
su(Hw) RHA rv	gagaGCTCTTCgGACAGCCAGT CTATGTCGCCAGT	su(Hw) homology arm right
SF1 L sgRNA fw	gtcgGTCCTTCTAACAGGTTCT	SF1 sgRNA left
SF1 L sgRNA rv	AAACAGAACCTGTTAGAAGG AC	SF1 sgRNA left
SF1 R sgRNA fw	GTCGCTGAAATCAGCATAAA AGG	SF1 sgRNA right
SF1 R sgRNA rv	AAACCCTTTTATGCTGATTC AG	SF1 sgRNA right
SF2b L sgRNA fw	GTCGGTTTGGCTCGGAGATTT CA	SF2b sgRNA left
SF2b L sgRNA rv	AAACTGAAATCTCCGAGCCA AAC	SF2b sgRNA left
SF2b R sgRNA fw	GTCGGCAAGTAAACAGAAAT ATC	SF2b sgRNA right
SF2b R sgRNA rv	AAACGATATTTCTGTTTACTT GC	SF2b sgRNA right
SF1 LHA fw	GAGACACCTGCGAGATCGCG GCCTTGGAGGTCATTAGG	SF1 homology arm left
SF1 LHA rv	GAGACACCTGCGAGACTACA CCTGTTAGAAGGACATTCAG	SF1 homology arm left
SF1 RHA fw	ccccagttggggcactacGCTCTTCGT ATAGGTGGGATTGCAGACAG C	SF1 homology arm right
SF1 RHA rv	GTCGCCCTTGAACCTCGATTGC TCTTCGGACAAATGCAATAA CTGGGAGTGG	SF1 homology arm right
SF2b LHA fw	GAGACACCTGCGAGATCGCC CGTGCCGTTTTTATAGCC	SF2b homology arm left
SF2b LHA rv	GAGACACCTGCGAGACTACT CATGGAAAGTACAAGAAAC	SF2b homology arm left
SF2b RHA fw	ccccagttggggcactacGCTCTTCGT ATATCTGGAAGACACAATTA TAC	SF2b homology arm right

SF2b RHA rv	GTCGCCCTTGAAC TCGATTGC TCTTCGGACTGCATGTGCGAT CTGAAGTT	SF2b homology arm right
Scr fw	ATGGATCCCGACTGTTTTGCG	Scr RNA-FISH probe
T7 Scr rv	AATACGACTCACTATAGGTG CGCTTAGGTGCGCGAACTGC	Scr RNA-FISH probe
ftz fw	ATTCGCAAAC TCAACCAGCGT	ftz RNA-FISH probe
T7 ftz rv	AATACGACTCACTATAGGGA TGGTAGAGGTCCTGTGG	ftz RNA-FISH probe
Dfd fw	ATGAGCTCTTTTCTCATGGG	Dfd RNA-FISH probe
T7 Dfd rv	GAAATTAATACGACTCACTA TAGGCAGGGCCGTCAGATCG TAG	Dfd RNA-FISH probe
abd-A fw	TTCGCCGTGTTTATTGTTCC	abd-A RNA-FISH probe
T7 abd-A rv	GAAATTAATACGACTCACTA TAGGAACAAGGCAAAAGGTT GTCG	abd-A RNA-FISH probe

Table S2. Primer list.

List of primers used.

Taf1_TSS_L	GTACAAAACGCAATTGACCTTCCCAAGCTCCCAACCCCTTGTC AAATCTAATCGCGGATTTTCTGGTGCCAGTGGGGAGCCATGAG GCATCTTTGCTTTGTCCATCGACGGTCACACTTA
Taf1_TSS_R	GTACTIONACTTGGAAAGGGAACCAATGTTTTCCCGGAGCTCCGC ATCGAAACCGGTGCCCCCGCGCCCCTCTCCGTCGTCATCTTGCA GCAGTCTGCCCTCGGAGTCGATGTTGCCGAAAA
Dfd_TSS_L	GTACGAGGCGAAGGGTGAGAAAAGGTGAACAGACATGGCTGC TACATATGAATGGCTTTATTTGTGCGCGTGTGTGGGCCGTCTT CGATTGCTTCGGCAGTGAAAACCCGTTTTGATGG
Dfd_TSS_R	TTAAAAGCCGTGGGGCCTTGAATTCTGAGAACTTGATTATG GTGTTATTTCCGGTTATAAAAAACGTTTTTACAACGAATTTCA ATCCACACAGGGAGAAAAGTCGGACGTTTCGG
Scr_TSS_L	GTACGAATACGAGAACGTGTATCTGTATCTTTTTATGTTTGCAG AAGACCGAGCGCGCTGCTTGGCGACTGAAGTGCGCCACGTTCA ATTCACGTTTGAGCCATAAAAACCGACAAAACG
Scr_TSS_R	TTAAGCTTCAGCAGGAGACTTTGCATTTTGCCAGGCCAATAAA AATGGAAATGAACGCGGAGCGGAACGTTTATAAAAATAAAT GCGAAATAAATAATGTGGCGGTGATTTAGCGGAACA
ftz_TSS_L	GTACGTGACTGTGCCACGCCAGCGCCTCCGAGGATGTCGAC TACTTGACGTCTACTCGCCCCAGTCGCAGACGCAGAAGCTGA AGAATGGCGACTTTGCCACCCCTCCGCCAACCAC
ftz_TSS_R	TTAATCGTGTGTGATGCCTACCTGATGCCAAAGTCTCCTCGATG TGCGACCAATTGAAATCGCCGGCTCCATTCGGGGCTGTCAAA TTCGATGATTGATCTCCTGGCTGACAGCTGACG
Antp_TSS_L	TTAAGAACTGGGAATGAACCGAACGGAATAAACGCGTGAAA CGACGGCAAACCGAAAGCAAAGTGCGAATCGAGCGGCGGCG GCACGTCTATTTGGCACGTCCATTCATTCACAACCTG
Antp_TSS_R	TTAACGAGCCACTAGTTACCTATCCAGCACTGTTTGCGTTCTCC AGCCGGCGAGAGTGTGTGTGTGTTTGCACAAAAACAATGCCGC CATTTATCATTTTCTATGGTGGCCGTGTTGCTG
ArgRS-m_TSS_L	TTAAAAACAGAACTTGCGCGAGAGCCACGCTCAGATTCTTAT TTTCTTGCCCTAGCCAATCGATATCTATTTTTCATGACAGCCCC GTGGTATGCAATACTTTTTAGCCGCTTCGCGCT
ArgRS-m_TSS_R	TTAAATTTGCAGACCTAACCTTCTTGCCAAAAATTTCCCAAGT ATCTAAAAGCTTGTTTTGCCTATTGGTAACGCAGTTTCGTGCAC CCACCGACGCCCATGCAAACAGATGTTTCAGTG
stck_TSS_L	TTAAAAAGAACTCATTTTCATAGTGATTGTGGGTTTTCATAGCA AGTTTACAAGCTCTTGATTCTCTTATCAATACATATTTTCTAAT GATAAGTATCAGCGAAACTGTTATGACATTTGT
stck_TSS_R	GTACTIONACTTGTAAACAAGACCGGTTTCGCATGCCAATTTAG GCTGCAGGCAAGAAATCCCTTTAGCATTGAATTTGATTGATTTT CTTTTTGCTATGTTACACGCCAATTGGGGAAA
Ubx_TSS_L	TTAACGATTTGAACGATTATTAGCCATAGTGCTGACCGAACGG GCGCGAATGCGTCTGCTGACGGATTTCTCGAATCTGGACGCC AATCTAGTGGAAGCGACTTTTCCGCGCGGCATT
Ubx_TSS_R	TTAACTGAACGAACACTCAAGAGAGAGCGCAAGAGCGCTCA AAAACAATCTGGTTTTGAGCGTTTCGCTGGCTCTCTGTTTCTGT TTTCCACTCGTTTTTLAGGCCGAGTCGAGTGAGTT

abd-A_TSS_L	TTAAAAGAGTATAAAAATTTTCGTGTAATTGGTAATTCTTGCTGC CAGCGCGGGCAGCGGCGTTCGACAGAAACGGCGGCAGAGCGGC AGCGACTGAGGCGCTTTGAGTCGTTGGAGACTTT
abd-A_TSS_R	TTAAAAGCATAGCACTCAAAGCGGGGCTCCAATAGTTTCCAT TCTCACGTTATAAAGAGTGAAATTAGGAGAGGCACTCAAATCG GGAAATTACTCACTCAAAGCAGCGAGGCGGCCAT
Abd-B_TSS_L	TTAACACTTTTCGAGCAAGAGCGCGCCGGCGGTGGCCAAGTGTT AGAGCCAGTTTTTTGGTTTTGGAGTCCGAGGTGCGGAGCGTCAT CGAGTGAAGTGAATGGTGTGCGAGTTCTCCGCTG
Abd-B_TSS_R	TTAAGGGTGCGGGTGCGAGTCGAAAAAAAAAGATATATGCCCGT CATATATCTAGGCGGTTCCCCATGCCGCTTTATTTATTTGTTTAT CTACTATTTATGGTCCCAAGAGCCGAAGCCGA
Blocking oligo 1	CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCTGCT GAACCGCTCTCCGATCT
Blocking oligo 2	AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACG ACGCTCTTCCGATCT

Table S3. Capture-C probe list and blocking oligos.

List of probes and blocking oligos used for Capture-C.

Capture-C sample	total read pairs	interchr	intrachr	intrachr (<20 kb)	intrachr (≥20 kb)
WT_embryo_rep1	222071	35016	187055	65749	121306
WT_embryo_rep2	218013	37884	180129	58335	121794
WT_embryo_rep3	372540	81031	291509	107001	184508
Cp190 ⁰ _embryo_rep1	263835	35939	227896	76655	151241
Cp190 ⁰ _embryo_rep2	255505	43570	211935	63822	148113
Cp190 ⁰ _embryo_rep3	352172	25730	326442	143762	182680
SF1 ^{KO} _embryo_rep1	342434	55638	286796	99181	187615
SF1 ^{KO} _embryo_rep2	281072	44945	236127	81849	154278
SF1 ^{KO} _embryo_rep3	389964	68287	321677	107680	213997

Table S4. Quality metrics of Capture-C read pairs.

Only informative read pairs were considered, i.e. only unique read pairs (after discarding probable PCR duplicates) with at least one read mapping to a viewpoint restriction fragment or one of its 2 neighboring restriction fragments.

Data S1. ChIPseq_Cp190_WT0-Cp1900_embryo.xlsx

Differential Cp190 ChIP-seq analysis in wildtype versus *Cp190⁰* embryos (Cp190 peaks in WT embryos).

Data S2. Hi-C_CD_boundaries.xlsx

Contact domain boundaries found in WT (column exist.WT_0), *double⁰* (column exist.CTCF_0_Cp190_0), *CTCF⁰* (column exist.CTCF_0) and *Cp190⁰* (column exist.Cp190_0), with physical insulation scores measured at each boundaries in each genotype (columns score.).

Data S3. Hi-C_insulation_scores.xlsx

Physical insulation scores measured for all bins in the genome by Hi-C in WT (score.WT_0), *double⁰* (column score.CTCF_0_Cp190_0), *CTCF⁰* (column score.CTCF_0) and *Cp190⁰* (column score.Cp190_0) embryos.

Data S4. Hi-C_eigenvectors.xlsx

First eigenvector and A/B compartments.

Data S5. ChIPseq_CTCF_WT0-CTCF0_embryo.xlsx

Differential CTCF ChIP-seq analysis in wildtype versus *CTCF⁰* embryos (CTCF peaks in WT embryos). Contains only peaks in up direction on chr2,3,4 and X.

Data S6. ChIPseq_Cp190_CTCF0-Cp1900_embryo.xlsx

Differential Cp190 ChIP-seq analysis in *CTCF⁰* versus *Cp190⁰* embryos (Cp190 peaks in *CTCF⁰* embryos). Contains only peaks in up direction on chr2,3,4 and X.

Data S7. ChIPseq_Cp190_CTCF0-WT0_embryo.xlsx

Differential Cp190 ChIP-seq analysis in *CTCF⁰* versus WT embryos (differential Cp190-bound regions in *CTCF⁰* versus WT embryos). Contains only differentially bound regions on chr2,3,4 and X.

Data S8. ChIPseq_CTCF_Cp1900-CTCF0_embryo.xlsx

Differential CTCF ChIP-seq analysis in *Cp190⁰* versus *CTCF⁰* embryos (CTCF peaks in *Cp190⁰* embryos). Contains only peaks in up direction on chr2,3,4 and X.

Data S9. ChIPseq_CTCF_Cp1900-WT0_embryo.xlsx

Differential CTCF ChIP-seq analysis in *Cp190⁰* versus WT embryos (differential CTCF-bound regions in *Cp190⁰* versus WT embryos). Contains only differentially bound regions on chr2,3,4 and X.

Data S10. ChIPseq_Cp190_WT-Cp190KO_larva.xlsx

Differential Cp190 ChIP-seq analysis in wildtype versus *Cp190^{KO}* larval central nervous systems (Cp190 peaks in WT larval central nervous systems). Contains only peaks in up direction on chr2,3,4 and X.

Data S11. ChIPseq_Cp190_SuHwKO-Cp190KO_larva.xlsx

Differential Cp190 ChIP-seq analysis in *su(Hw)^{KO}* versus *Cp190^{KO}* larval central nervous systems (Cp190 peaks in *su(Hw)^{KO}* larval central nervous systems). Contains only peaks in up direction on chr2,3,4 and X.

Data S12. ChIPseq_Cp190_SuHwKO-WT_larva.xlsx

Differential Cp190 ChIP-seq analysis in *su(Hw)^{KO}* versus WT larval central nervous systems (differential Cp190-bound regions in *su(Hw)^{KO}* versus WT larval central nervous systems). Contains only differentially bound regions on chr2,3,4 and X.

Data S13. IBAQ.xlsx

List of intensity-based absolute quantification (IBAQ) values of proteins detected by mass spectrometry in pull-down experiments with baits shown in Fig. 4A and the respective negative control pull-downs performed in parallel.

Data S14. CaptureC_Cp190-WT0_embryo.xlsx

Differential Capture-C analysis in *Cp190⁰* versus WT embryos.

Data S15. CaptureC_SF1KO-WT0_embryo.xlsx

Differential Capture-C analysis in *SF1^{KO}* versus WT embryos.