STUDY OF CHROMATIN INSULATORS IN GENE REGULATION AND GENOME EVOLUTION IN *DROSOPHILA*

by

ZHIBO MA

(Under the Direction of Haini N. Cai)

ABSTRACT

Chromatin insulators are widely distributed in the genome and mediate formation of chromatin loops, but their roles in gene regulation remain poorly understood. The complex expression pattern of the *Drosophila* homeotic gene *Sex combs reduced (Scr)* is directed by an unusually long regulatory sequence harboring diverse cis elements and an intervening neighbor gene *fushi tarazu (ftz).* In this project, we identified the SF2 insulator between the *ftz* gene downstream enhancer and the *Scr* enhancers in the distal region. Both SF1 and SF2 locate precisely at the boundaries of polycomb-mediated silent chromatin domains and known insulator proteins bind to both SF1 and SF2 in 0-12 fly embryos as well as cell lines from different developmental origins. 3C experiments showed that interaction between SF1 and SF2 is developmentally regulated. By using a transgene assay, we showed that SF2 has strong enhancer blocking activity and pairing of SF1 and SF2 cancels enhancer blocking activity in an orientation dependent manner. We also demonstrated that both SF1 and SF2 insulators are functionally conserved in distantly diverged *Drosophila* species. CRISPR-Cas9 mediated knockout of SF1 or SF2 not only leads to the misregulation of *Scr* and *ftz* genes, but also affected stage- and tissue-

specific chromatin architecture in the Antennapedia complex. Flies with SF1 or SF2 deletion have significantly reduced overall-fitness during development.

SF1 and SF2 flank an evolutionarily inverted *ftz* gene unit in the genus of *Drosophila*. The breakpoints of this inversion map immediately to the inner sides of the loop anchor sites. We further demonstrated the precise matching of independent inversion breakpoints and insulator sites at the *Dfd* Hox gene locus. Both pairs of insulator sites form long-range chromatin loops in *Drosophila* embryos and their homologs from distantly related *Drosophila* species also show conserved insulator activity. Through genome wide analysis, we further demonstrated the precise association between evolutionary breakpoints and chromatin insulator sites in several *Drosophila* species that diverged from *Drosophila melanogaster* 15 to 60 million years ago. Together, we provide evidence that chromatin insulator formed loops may contribute both to the formation and fixation of evolutionary rearrangement.

INDEX WORDS: *Drosophila*, Chromatin insulator, Chromatin Boundary, Chromatin loop, Gene regulation, Genome rearrangement, Chromatin breakpoint, Genome evolution, Antennapedia complex, HOX gene, SF1, SF2

STUDY OF CHROMATIN INSULATORS IN GENE REGULATION AND GENOME EVOLUTION IN *DROSOPHILA*

by

ZHIBO MA

B.S., China Agricultural University, China, 2010

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

CHROMATIN INSULATORS REGULATE GENE EXPRESSION AND PROTECT GENOME DURING EVOLUTION

by

ZHIBO MA

Major Professor: Haini N. Cai

Committee: Douglas Menke
Edwards T. Kipreos
Haini N. Cai
Xiaoyu Zhang

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2018

DEDICATION

To my loving family

To my friends

To you who are reading it

To the three cats: Xiaotiao Ma, Xiaomeng Yue, and Miaomiao Mei

To all the delicious foods

To all the flies who helped me in this school

and kept me passionate about research

ACKNOWLEDGEMENTS

I would like to thank all the people who guided me, inspired me, and encouraged me in these seven years. Leaving all my family and friends to study abroad was a big decision for me. The past seven years turned out to be a journey that helped to learn how to think independently in research, to gain better ideas about who I am, and to grow into an independent man, I would first express my appreciation to my advisor and my committee members. Thank you to Dr. Cai for your mentorship, your wisdom, insights, criticism, and for encouraging me to think more rather than doing bench work. To Dr. Kipreos for your passion about science and critical advices. To Dr. Menke for your suggestions and comments to the project as well as your kindness and encouragement. To Dr. Zhang for your patience and critical comments to the project. I would also say thank you to Dr. Ping Shen, for your advices and support with expenses to the fly meetings. My special gratitude to my collaborator Burair Alsaihati, for your intelligence, patience, and help with developing the tools for the bioinformatics project.

My sincere thanks also go to my lab mates. Thank you to Matthew Romine for the help with everything and for keeping me accompanied late at night in lab. Thank you to Carly Duffy for helping me with the project and proof-reading the manuscript. Thank you to Nuria Waddington Negrao for your energy, humor, and the friendship. Thank you to Qing Zhao, for your help with the 3D DNA FISH work. Thank you to Joan Han, for your help with the fly and PCR work. Thank you to Jeffrey Thomas Gardner for your humor, intelligence and friendship.

I am especially grateful to the new friends I met in Athens. Thank you to Xiao Li, for your help since the first week I arrived in Athens. Thank you to Huan Yang, for your delicious

craw fishing boiling. Thank you to Peiwei Liu, for your friendship, encouragement, delicious food, and the help with proof-reading the manuscript. Thank you to Yiwen Zhang and Yuhan Pu for your friendship and help in the past. Thank you to Munisha Mumingjiang for the delicious Xinjiang food. Thank you to all the colleges in cell bio department!

Most importantly, I would like to thank my family. Thank you to my parents, for supporting me at your best and for letting me leave the family and chase my dream. Thank you to my sister for always sharing your happiness with me. Thank you to my fiancé, for your love, support, happiness! Love you!

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Nuclear organization and genome function

Chromosomes inside of a nucleus carry the genetic information of an organism. They meet the dual challenges of packaging long linear genomic DNA into the small nuclear space, while at the same time, enabling proper access to genetic information by spatiotemporal requirement. In the past two decades, benefiting from advance in technology, the organization of genomic DNA has increasingly been recognized as important for the function of the genome (Misteli, 2007; Bickmore and van Steensel, 2013; Belmont, 2014; Pombo and Dillon, 2015; Bonev and Cavalli, 2016). Mountains of evidence from both case studies and genome wide analysis indicate that chromatin organization interplays with essential biological functions of the genome, including transcription, replication, DNA repair, genome rearrangement and alternative splicing (Pope et al., 2014; Berthelot et al., 2015; Ruiz-Velasco et al., 2017; Fabre and Zimmer, 2018; Li et al., 2018). However, despite the quick progress in the chromatin organization field, we still understand little about the underlying mechanism of how chromatin organization is established, regulated and maintained. Also, very little is known about the mechanisms of how chromatin organization affects different aspects of the genome functions.

1.1.1 Nuclear organization at different scale

To fit an entire genome into the confined space of the nucleus, genomic DNA needs be compacted thousands of folds. In human, a diploid human genome DNA measures about 2 meters long, which needs to be packed into a nucleus of ~6μm in diameter. The model organism

Drosophila melanogaster's genome measures ~110cm long and must be packaged into a nucleus of ~5μm. DNA is known to assemble with histones and other proteins into higher order chromatin structures. The long-standing text-book model of chromatin organization is that the primary "beads-on-the-string" 11nm nucleosomes assemble into 30nm fibers that further fold into 120nm chromonema, 300 to 700nm chromatids, and ultimately mitotic chromosomes. Further extrapolating from this model, silent chromatin is generally in the status of 30nm and 120nm compact fiber which restricts the access of transcription factors (Pierce, 2006). However, recent evidence from both chromatin conformation capture, cryo-EM and X-ray scattering analysis, call into questions of 30nm chromatin fiber in interphase genome (Dekker, 2008; Eltsov et al., 2008; Ahmed et al., 2010; Fussner et al., 2012; Maeshima et al., 2014; Sanborn et al., 2015). Instead, it seems that chromatin is mainly configured as irregularly folded 10nm chromatin fibers in both open and closed domains, and nucleosomes are assembled in heterogeneous groups, called "clutches", in a cell-type dependent manner (Maeshima et al., 2014; Ricci et al., 2015). The canonical hierarchical model of chromatin was also challenged by recent microscopic observations. A comparative study with cryo-EM and conventional EM on a special region in human HL-60 cells found that the 30nm fiber in this region could be observed only following aldehyde fixation, but not in cryo-sections. This suggests that the 30nm fiber visualized by conventional EM could be an artifact (Eltsov et al., 2008). New observation of chromatin structure in the nucleus using the new method ChromEM, which combines electron microscopy tomography with a labeling method that selectively enhance the contrast of DNA also does not support the existence of 30nm or larger chromatin fibers (Ou et al., 2017). In contrast to the previous in vitro structures, ChromEM method showed that chromatin is a flexible and disordered 5-24nm diameter granular chain that is packed together at different concentration

densities (Ou et al., 2017). High resolution microscopic analysis and high-throughput chromosome conformation capture assay (HiC) in many species revealed that eukaryotes' genome is organized at different scales: chromosome territories, chromatin compartments, Topologically Associated Domains (TADs, also called physical domains), and chromatin loops (Ou et al., 2017; Serizay and Ahringer, 2018).

Chromosome territories and chromosome compartments

The concept of chromosome territories was first proposed by Carl Rabl in 1885, where he believes the DNA of each chromosome occupies a defined volume of the nucleus and only overlaps with its immediate neighbors. Early fluorescent in situ hybridization microscopic evidence confirmed the existence of chromatin territories in multiple eukaryotic species (Cremer and Cremer, 2010). Studies of genome organization using high-throughput chromosome conformation capture technique also supported the chromosome territory concept by showing chromatin interactions predominantly happen within each chromosome (Lieberman-Aiden et al., 2009; Stevens et al., 2017).

The first high-throughput HiC by Lieberman-Aiden *et al.* revealed that the entire human genome is portioned into two spatially separate compartments such that greater frequency of interaction occurs within each compartment rather than across compartments (Lieberman-Aiden et al., 2009). Compartment A is more closely associated with open, accessible, actively transcribed chromatin, and compartment B is associated with silent chromatin (Lieberman-Aiden et al., 2009; Rao et al., 2014). Recent higher resolution HiC suggested that the A/B compartments can be further subdivided into two A and four B subcompartments, including the polycomb-mediated facultative heterochromatin subcompartments (Rao et al., 2014). Stevens *et al.* conducted HiC on single-cell and provided a new view of the spatial localization of the A and

B compartments inside of nucleus. This study showed that the silent B compartment chromatin is located near the nuclear periphery and center nucleoli, while the active A compartment is organized in an inner ring-structure in between the peripheral and center layers of B compartment (Stevens et al., 2017). These are consistent with microscopic observations of the locations of active and silent chromatin within nucleus (Croft et al., 1999; Kind et al., 2013).

Chromatin compartments are highly associated with chromatin state domains defined by active or repressive histone marks, and they are dynamic during development. Dixon *et al*. found that about 36% of active A compartment and silent B compartment altered during human ES cell differentiation (Dixon et al., 2015). Artificially altering local chromatin state through targeting histone modifiers at "unlocked" loci can drive repositioning to different compartment (Wijchers et al., 2016). Chromatin compartmentalization is only observed during interphase and it disappears in the mitotic stage across the genome (Naumova et al., 2013). However, chromatin compartmentalization pattern could be inherited in the daughter cells, potentially due to chromatin bookmarking (Giorgetti et al., 2013). The recent ChromEM study also showed evidence inside of nuclei that the overall primary structure of chromatin polymer does not change but instead collapses into compact loops and interaction arrays in mitotic chromosomes, which also helps explain the rapid dynamic of chromatin condensation and the inheritance of chromatin organization through cell division (Ou et al., 2017). Depletion of CTCF and cohesin eliminates Topologically Associated Domains, but does not significantly affect chromatin compartment formation and linear chromatin histone modification domain pattern (Nora et al., 2017; Rao et al., 2017). These data suggest a model where the formation of chromosome compartments relies on linear chromatin domains instead of TADs (Carelli et al., 2017; Rao et al., 2017).

Topologically associated domains

Topologically associated domains (TADs), also called physical domains or contact domains, is one of the most interesting discoveries about the chromosomal secondary structure in recent high-throughput chromosome conformation capture studies. It was discovered that chromosomes are topologically segregated into sub-megabase sized domains, TADs, and regions within the same TAD interact with each other much more frequently than regions from different TADs (Lieberman-Aiden et al., 2009; Dixon et al., 2012; Sexton et al., 2012; Sexton and Cavalli, 2015). TADs are generally considered to be relatively conserved among different cell types and even across different species (Dixon et al., 2012; Rao et al., 2014; Vietri Rudan et al., 2015). However, to what extent of this conservation is uncertain. This uncertainty is partially due to the difference in sequencing depth of current data sets, different analysis algorism, and the nested feature of TADs where a large TAD could be further divided into smaller subTADs. Significant reorganization of the 3D genome at TADs level has been observed during cell lineage specification, animal development, and environmental stimulation (Andrey et al., 2013; Phillips-Cremins et al., 2013; Dixon et al., 2015; Li et al., 2015a; Beagan et al., 2016). TAD structures seems to also be variable from cell to cell as revealed by single cell HiC experiments (Stevens et al., 2017) and polymer modeling (Giorgetti et al., 2014). This suggests that TADs reflect an average of interactions at the single cell level where enhancer-promoter contacts emerge as probabilistic events. This is consistent with the previously observed cell to cell transcriptional heterogeneity (Boettiger and Levine, 2009; Vera et al., 2016; Mohammed et al., 2017).

Many evidences from TAD associated features support the view that TADs represent functional domains. The spatial segregation of the genome into TADs correlates very well with many genomic features such as histone modification domains, gene activity domain, lamina

association, and genome replication timing (Handoko et al., 2011; Sexton et al., 2012; Pope et al., 2014; Rao et al., 2014). In *Drosophila,* HiC achieves a higher resolution due to the ~17-fold smaller genome comparing to mammalian genome. Sexton *et al.* found that TADs are significantly associated with four major classes of epigenetic marks, as defined by cluster analysis of protein binding and histone modification pattern (Sexton et al., 2012). Out of the four classes, "null" domains are weakly enriched with the insulator protein Su(Hw) and are not enriched for any available marks. These null domains on average have normal levels of gene density, but lower transcriptional activity output. They span about half of the genome (Sexton et al., 2012). Transcriptionally active domains are associated well with H3K4me3, H3K36me3, and hyperacetylation, and they comprise 42% of the domains and 22% of the genome(Sexton et al., 2012). Enhancers and their target genes are typically restricted within the same active TAD domain (Symmons et al., 2014; Grubert et al., 2015). There are two classes of repressive domains, one is marked by polycomb-group proteins and associated with H3K27me3, and the other is bound by the heterochromatin proteins, HP1 and Su(var)3-9 , and associated with H3K9me2 (Sexton et al., 2012). Similar correlation between TADs and domains of epigenetic marks was also observed in mammals (Dixon et al., 2012; Rao et al., 2014). In addition, strong correlation between TADs and genome replication timing domains has been observed in mammals and yeast (Pope et al., 2014; Dileep et al., 2015; Eser et al., 2017).

However, it remains uncertain whether TADs constitute a structural blueprint for the regulatory landscapes, or if TADs are a result of transcriptional silencing and activation machineries. Pharmaceutical inhibition of transcription in *Drosophila* early stage embryos showed that TAD formation emerges during zygotic genome activation independent of transcription (Hug et al., 2017). However, Hug *et al.* used two inhibitors specific to transcription

elongation process, which made it the relationship between TADs and transcription initiation and pre-initiation process uncertain. In both *Drosophila* and *mammals*, poised interaction between distant enhancers and promoters has been detected before gene activation. Ablation of H3K27me3 in mouse ES cells by knockout of polycomb group gene *Eed* had no significant effect on the overall TAD structure at X-inactivation center (Nora et al., 2012). In some case studies, knockouts of CTCF sites led to the spreading of either H3K4me3 or H3K27ac marked active chromatin (Narendra et al., 2015; Willi et al., 2017) or H3K27me3 marked repressive chromatin (Luo et al., 2018). RNAi knock downs of insulator proteins in *Drosophila* caused spreading of 75% repressive H3K27me3 spreading, but not the actively transcribed domains (Schwartz et al., 2012). However, targeted degradation of CTCF in mammalian cells only caused a restricted local gain of the H3K27me3 signal at the initially bound CTCF site, but did not trigger spreading of H3K27me3, even though it caused extensive ectopic contacts across the initial TAD boundary (Nora et al., 2017). Acute depletion of cohesin eliminated almost all TADs. Surprisingly similar to CTCF depletion, chromatin compartments and histone modification domains were not significantly affected upon cohesin depletion (Rao et al., 2017). Restoration of CTCF and cohesin, respectively, led to the reformation of TADs pattern(Nora et al., 2017; Rao et al., 2017). Although current observations seem conflicting, more and more evidence support the view that linear chromatin domains may provide a primary level of 11nm chromatin organization and regulation upon which higher-level organization mechanisms act. The organization mechanism seems to also include organism-specific additions (Schwartz et al., 2012; Feng et al., 2014; Grob et al., 2014; Gabdank et al., 2016).

In addition, although the widespread TAD structures described in mammals and *Drosophila* are considered as functional units of the genome, they are not observed in

Caenorhabditis elegans (Crane et al., 2015; Gabdank et al., 2016) and *Arabidopsis thaliana* (Moissiard et al., 2012; Feng et al., 2014; Grob et al., 2014). Both *C. elegans* and *A. thaliana* chromosomes are demarcated by alternating chromatin domains of active and repressive histone marks (Feng et al., 2014; Evans et al., 2016), which is similar with what's been observed in mammals and *Drosophila* (Sexton et al., 2012; Rao et al., 2014). This difference may be due to technical or biological limitations, such as sequencing depth of HiC and compactness of genome. But it is also possible that TADs are not a general genomic structure in all organisms. Rowley et al. have proposed an CTCF-independent model of higher genome organization formation (Rowley et al., 2017). By using high resolution HiC map in *Drosophila* cell lines, Rowley et al found that the boundaries of many TADs are formed by smaller domains, which they termed compartment domains. A compartment domain is comprised of one or more genes of the same transcription status and it physically clusters with other compartment domains that share the same transcriptional status in the absence of CTCF. Simulation based on this model showed that transcriptional status of genes are good predictors of overall genome structure both in *Drosophila* and other species. Especially, simulation of *Arabidopsis thaliana* and *Caenorhabditis elegans* genome structures showed a consistent interaction map compared to the actual HiC map. A large domain observed in the actual HiC aligns very well with blocks of silenced regions separated by a small transcribed region. This domain is also captured in the computer simulation which is based on transcriptional status. Furthermore, according to this model, the rather constant transcriptional status pattern along chromatin is consistent with the actual HiC map. This study provides a novel insight into genome structure by showing that transcriptional status may play a critical role in higher order chromatin organization formation in all eukaryotes, with insulators playing an important role in domain formation in some organisms.

Chromatin loops

The idea that chromatin fiber is looped is one of the oldest in cell biology. A chromatin loop forms when two genomic sites are brought into closer physical proximity of each other than to intervening sequences. The earliest direct evidence for chromatin loops came from the observation of "lamp-brush" chromosomes by Walther Flemming in 1882 (Andraszek and Smalec, 2011). Chromatin loops are clearly seen in lamp-brush chromosomes during the prolonged meiotic prophase in oocytes of many species. These loops are sites of intense transcription as oocytes prepare to store huge amount of components needed for rapid cell division after fertilization (Andraszek and Smalec, 2011). Similar loops, called "puffs", have also been observed in *Drosophila* polytene chromosomes under stress condition (Zhao et al., 1995). Recent 3C-based experiments demonstrated that chromatin loops have been implicated in virtually all levels of chromatin organization. Chromatin loops include, but are not limited to, enhancer-gene loops, silencer-engaged loops, insulator engaged loops, enhancer-terminator loops, promoter-intron loops, and other uncharacterized loops. However, due to the limitation of resolution, it is difficult to study local chromatin loops using an unbiased genome wide approach. Therefore, the current understanding of local chromatin loops comes mainly from single-locus studies.

Local chromatin loops are critical in both positive and negative gene regulation. In mammals, the mouse and human β-globin loci have been extensively studied as models for long range developmentally regulated enhancer-gene loops. This gene cluster contains a distant locus control region (LCR) and a cluster of β-globin genes that are active at different developmental stages. LCR has been shown to specifically interact with only one β-globin gene promoter at a time in a developmental stage and tissue specific way(Palstra et al., 2003). Multiple CTCF

insulator protein binding sites have been found within LCR involved in mediating long-range enhancer-promoter interactions. CTCF-based interactions form a pre-structure that brings regulatory sequences located 130kb away and the *β*-globin cluster into proximity (Drissen et al., 2004). These pre-formed loops, also called poised enhancer-gene loops, provide physical proximity of genes and their regulatory elements, which is believed to facilitate timely regulation of transcription upon developmental or environmental stimuli. Additional tissue specific transcription factors such as EKLF and GATA1 mediate the final step of LCR to promoter interaction (Drissen et al., 2004; Vakoc et al., 2005). The mouse *TH2* locus revealed tissue specific poised loops formed by promoter-promoter interaction and promoter-enhancer interaction (Spilianakis and Flavell, 2004). The mouse *sonic hedgehog* (*shh*) gene requires an enhancer 800kb away in the zone of polarizing activity. *Shh* is expressed in all the cells in this zone and 3C experiments in the limb buds detected interactions between *shh* and this enhancer. However, looping between this distant enhancer and the *shh* gene is only observed in 18% of cells, which suggests that long range interactions are transient (Amano et al., 2009). This long-range enhancer-promoter loop at *shh* locus is also a pre-formed structure independent from the transcription activity as it was also found at *shh* inactive limb bud cells (Amano et al., 2009; Dixon et al., 2012). Mouse *Igf2/H19* locus demonstrated that chromatin loops may also contribute to gene silencing and the formation of loops can be regulated through the methylation of architectural protein binding sites. The accessibility of a distal enhancer to *H19* and *Igf2* genes is regulated by the methylation status of CTCF binding sites near the promoters and enhancer. On the maternal allele, the CTCF sites are unmethylated, which allows interactions and loops out *Igh2*. This looping-out of *H19* restricts the accessibility of the enhancer and keeps maternal *Igh2* allele silent. There is still much to learn about the contacts between enhancers, promoters, and

silencers, which can be highly flexible, diverse and dynamic rather than following a single pattern of multiple enhancers clustering around a promoter.

Local chromatin loops are also important in maintaining the individuality and specific gene-expression properties of neighboring genes. In *Drosophila*, insulator-mediated loops have been intensively studied in gene regulation. Insulators have been identified by their ability to impede the spreading of heterochromatin and communication between promoter and distal enhancers. It has been observed at numerous genomic loci that insulators form loops through physical interactions. The current model proposes that chromatin loops anchored by insulators ensure the fidelity of enhancers and their promoters. The specialized chromatin structure (scs) and scs' were the first pair of insulators found flanking the *Hsp70* "heat-shock puff" locus (Udvardy et al., 1985). Scs and scs', bound by Zw5 and BEAF respectively, were found to physically interact and insulate the flanked gene from the influence of surrounding chromatin (Blanton et al., 2003). Almost at the same time, studies from the *gypsy* retrotransposon revealed a new insulator called *gypsy* insulator, which is bound by the SuHw protein (Holdridge and Dorsett, 1991; Geyer and Corces, 1992). When the g*ypsy* insulator is located in between, it blocks the distal enhancer from activating the promoter in a directional manner. Pairing of two copies of *gypsy* insulators allows bypass of the distal enhancer (Cai and Shen, 2001; Muravyova et al., 2001). This bypass of insulator pairing is insulator-insulator specific and orientation dependent (Majumder and Cai, 2003; Kyrchanova et al., 2008a; Kyrchanova et al., 2013; Kyrchanova et al., 2016). A list of insulator elements was discovered in the *Drosophila* bithorax complex. The ~300kb region contains three HOX genes *Ubx, Abd-A, and Abd-b,* which determine the anterior-posterior body segments identity. Three genes are divided into nine functionally independent regulatory unit by insulator mediated interactions (Maeda and Karch,

2007; Kyrchanova et al., 2011). Interactions between insulator elements can also mediate vast long range intra-chromosomal loops and transvection between loci located on different chromosomes (Savitskaya et al., 2006; Schoborg et al., 2013; Fujioka et al., 2016). Studies from the *Drosophila eve* locus demonstrate that loops formed between two insulators, Homie and NHomie, not only insulate the eve locus from adjacent repressive chromatin spreading but also promote distal enhancer-promoter communication (Fujioka et al., 2009; Fujioka et al., 2013). Developmentally regulated insulator-insulator interaction was also reported in *Drosophila* to keep *ftz* gene insulated from neighboring HOX genes in early embryos (Li et al., 2015b). A novel stage specific protein complex, ELBA, has been identified responsible for Fab7 insulator mediated insulation in *Drosophila* embryos (Aoki et al., 2008; Aoki et al., 2012).

Long range interactions engaged by silencers are best represented by polycomb mediated repression in *Drosophila*. Silencers, such as polycomb response elements (PRE), repress promoters at a distance by nucleating complexes. In *Drosophila,* the Mcp PRE in the bithorax complex could control the expression of Abd-B gene over a distance of 60kb and the Fab7 insulator/PRE contributes to the silence of ANT-C over 10Mb (Bantignies et al., 2011). PREs pair/cluster into polycomb bodies. Upon heat shock, the 3D genomic organization widely reorganize. Polycomb group proteins bind to enhancers and gene promoters to form silent polycomb bodies (Li et al., 2015a). Silent chromatin loops also form through lamina associated domains (LADs) and matrix attachment regions (MARs). Genes may be brought close to the silent nuclear periphery or nuclear matrix region by LADs or MARs (van Steensel and Belmont, 2017).

Chromatin looping might be more prevalent and important for gene expression than what is commonly thought. Studies from both yeast and human found that gene-looping, which is also

called promoter-terminator looping, appears to be a general phenomenon of RNAPII transcribed genes (Tan-Wong et al., 2008; Laine et al., 2009; Tan-Wong et al., 2012; Bratkowski et al., 2018). It was found that gene looping could not only guide the proper orientation of RNAPII, but also facilitates transcription memory (Laine et al., 2009; Tan-Wong et al., 2012). In *Drosophila,* chromatin insulators located immediately downstream of reporter gene were found looping with their target promoters in transgenic lines, which supported the basal activity of the reporter gene (Erokhin et al., 2011). Ruiz-Valasco *et al.* recently demonstrated that CTCF-mediated chromatin loops between promoters and intragenic regions regulate alternative splicing. When exons are in physical proximity with their promoters, CTCF binding correlates with exon inclusion in spliced mRNA (Ruiz-Velasco et al., 2017). I speculate that future studies that focus more on the local scale chromatin loops will unravel novel general principles of chromatin loops, and its role in gene regulation.

Several approaches have been utilized in determining the formation of long range chromatin loops. Currently, the most widely used experimental approaches are 3D-FISH and chromosome conformation capture (3C) based methods. The FISH method directly measures the distance of two or more loci inside of nuclei, but is limited by its hypothesis-driven characteristic and is low throughput. 3C-based methods indirectly reflect physical distance in high resolution by measuring proximity-based ligation frequency at population average level. 3C derived HiC method enables mapping of chromosomal interactions at a genome wide scale, however, its effective resolution, which is contingent on sequencing depth and restriction fragments, limit the characterization of specific chromatin loops. Novel enhanced Hi-C derived methods, such as ChiA-PET (Fullwood and Ruan, 2009), Capture HiC (Jager et al., 2015) , and HiChIP (Mumbach et al., 2016), have been developed to either allow more efficient characterization of

chromatin loops at global scale, or allow protein factor-centric chromatin conformation investigation. The CRISPR affinity purification *in situ* of regulatory elements (CAPTURE) approach was recently developed to unbiasedly identify locus-specific chromatin-regulating protein/RNA complexes and single locus to whole genome long-range DNA interactions (Liu et al., 2017).

1.1.2 Factors involved in formation of higher order chromatin organization

An important remaining question in the field of chromatin is how the structural features of the chromatin are established, maintained and regulated. Progress achieved in the recent years by the development of 3C-based techniques and the availability of genome wide maps of chromatin binding proteins have made it possible to correlate protein occupancy with higher order chromatin structures, as well as point-to-point loop interactions. The factors and processes involved in the establishment, maintenance and regulation are largely unknown, but mediator complex, insulator binding proteins, polycomb group complex, and noncoding RNAs are emerging as likely candidates of essential architectural factors (Van Bortle and Corces, 2012).

Architectural factors for Insulator-mediated chromatin interactions

Insulators are DNA-protein complexes that are experimentally defined by their ability to impede the spreading of silent chromatin and/or their ability in blocking distal enhancers from activating the promoter. It has also been found enriched at TAD boundaries in both mammals and flies (Sexton et al., 2012; Rao et al., 2014). Different species seem to utilize a different set of components in order to establish higher order structures of chromatin. In fruit flies, several architectural proteins are enriched at loop anchors and TAD boundaries (Sexton et al., 2012; Van Bortle and Corces, 2012). This allows fruit flies to build different TAD boundaries or different chromatin loops by using different combinations of architectural proteins (Phillips-Cremins et

al., 2013; Van Bortle et al., 2014). However, in vertebrates, a partially different set of factors may fulfill the similar role (Ong and Corces, 2014).

-*CTCF*

CTCF is perhaps the factor that received the most attention in chromatin organization. CTCF is characterized as an insulator protein with enhancer blocking activity (Kurukuti et al., 2006). It is an evolutionarily conserved 11-zinc-finger DNA binding protein and is ubiquitously expressed (Klenova et al., 1993). Studies in *Drosophila* later identified the ortholog dCTCF with similar zinc-finger domains and insulator activity (Moon et al., 2005). Despite the conserved zinc-finger domains from fruit fly to human, it can bind to the conserved target motif with a variety of DNA sequence variations (Holohan et al., 2007). The variant sequences to which CTCF binds and the protein factors with which CTCF interacts have been suggested to contribute to the different roles CTCF plays (Ghirlando and Felsenfeld, 2016). Binding of CTCF to genomic DNA could be affected by the DNA methylation status (Kurukuti et al., 2006). Only some CTCF motifs contain a CpG island at the correct place to be used for regulation of CTCF binding (Flavahan et al., 2016). Between ~30-60% of CTCF binding sites are cell type specific across 38 human cell types (Chen et al., 2012).The neighboring binding sites for other factors may also modulate CTCF function (Weth and Renkawitz, 2011). For example, Smad associates with CTCF at the *Igf2/H19* locus (Bergstrom et al., 2010) as well as many other sites in fruit fly in a CTCF dependent manner, which indicates it has a potential role in BMP signaling response at certain genomic loci (Van Bortle et al., 2014). CTCF/cohesin complex can also recruit the core promoter factor TAF3 and form a TAF3 dependent loop with the promoter, which often brings distant enhancers into close physical proximity of its target promoter (Liu et al., 2011). It has been demonstrated that CTCF could also undergo poly(ADP) ribosylation, sumoylation and

phosphorylation to modify its activity (Klenova et al., 2001; MacPherson et al., 2009; Guastafierro et al., 2013).

In mammals, CTCF is highly associated with cohesin complex at the loop anchors and is especially enriched at TADs boundaries. Using a very stringent criteria, Rao *et al.* found CTCF and cohesin occupy about 86% of the identified loops in human and mouse cell lines. And the vast majority (92%) of the identified loops contain convergent CTCF sites at the anchors (Rao et al., 2014). A large fraction of identified loops (38%) coincide with interactions between adjacent TAD boundaries (Rao et al., 2014). However, many short-range CTCF-mediated loops and more weakly formed loops, which are often located within TADs, do not involve convergent CTCF sites (Tang et al., 2015). It has been revealed in human cell lines that boundary strength is associated with CTCF signal strength and super-enhancers are usually insulated by strong boundaries (Gong et al., 2018). This indicates that strong CTCF mediated loops demarcate the boundaries of contact domains. Additionally, deletion of CTCF sites at TAD boundaries result in the extension of TAD (Narendra et al., 2015; Willi et al., 2017; Luo et al., 2018). Inversion of CTCF sites with a flanked enhancer in the protocadherin locus rewire the enhancer-promoter interaction (Guo et al., 2015). This further supports the model that CTCF mediated loops may be directional.

However, *Drosophila* CTCF (dCTCF) seems to differ fundamentally in its function from the human homolog in TAD formation. In contrast to humans and mice, dCTCF sites are found in less than one third of the TAD boundaries, and they show no directional preference when interacting with other sites along the chromosome. CTCF enriched TAD boundaries mostly flank inactive chromatin (Ramirez et al., 2018). *Drosophila* cohesin complex could bind to transcriptionally active genes independently of dCTCF (Misulovin et al., 2008). The

colocalization between dCTCF and cohesin complex in *Drosophila* is still under debate. Bartkuhn *et al.* mapped the ancillary cohesin subunit, Scc3, using ChIP-chip and found no strong overlap of dCTCF with Scc3(Bartkuhn et al., 2009). However, Van Bortle et al. later showed the opposite result by mapping the binding profile of the essential cohesin subunit Rad21(Van Bortle et al., 2014). Nearly half of all high confidence dCTCF sites overlap with Rad21, similar to percentages originally identified in Hela cells (Wendt et al., 2008). Instead of cohesin, dCTCF is commonly thought to rely on a Centrosomal Protein 190 kDa (CP190), which contains BTP/POZ domains that are capable of forming stable multimers (Bonchuk et al., 2011; Van Bortle and Corces, 2012). 62% of dCTCF sites colocalize with CP190 (Schwartz et al., 2012), but dCTCF/CP190 is only enriched in a small subset of TADs boundaries that mainly flank repressive domains in the sub-kb resolution HiC map(Ramirez et al., 2018; Wang et al., 2018). Nevertheless, CTCF mediated insulator loops still carry conserved function between *Drosophila* and vertebrates. It has been reported that Nuclear Remodeling Factor (NURF) and dREAM complexes bind to CP190 and are required for enhancer blocking activity (Bohla et al., 2014). NURF is highly conserved and is found to interact with CTCF in mammals (Qiu et al., 2015).

-Drosophila specific insulator proteins

In *Drosophila,* many DNA-binding insulator proteins have been identified: dCTCF, Su(Hw), Boundary Element Factor 32 (BEAF-32), Zest-White 5 (zw5), GAGA factor (GAF), Elba, Pita, ZIPIC, Insulator binding factor 1 (Ibf1), Ibf2 (Bushey et al., 2009; Aoki et al., 2012; Cuartero et al., 2014; Maksimenko et al., 2015). Although each DNA-binding protein has its own preferred DNA binding motif sequence, many of them are found to be colocalized throughout the genome in different combinations (Bushey et al., 2009). Different subclasses of insulator protein combination have been found to be associated with different genomic locations, different gene

activities and TAD boundary strength (Bushey et al., 2009; Negre et al., 2010; Schwartz et al., 2012; Van Bortle et al., 2014; Ramirez et al., 2018). Different from vertebrates, CTCF enriched boundaries only represent a small fraction of boundaries flanking inactive chromatin (Schwartz et al., 2012; Sexton et al., 2012; Ramirez et al., 2018; Wang et al., 2018). The chromatin interaction network mediated by these insulator proteins have been probed more and more in depth, however, the mechanism of how these insulator mediated loops regulate gene expression is still uncertain. It has been pointed out that, unlike the prototype insulators, the majority of these insulator binding sites do not show robust enhancer blocking activity as assayed in transgenic flies and the insulator activity of a given site with certain combination of insulator proteins is often genomic context dependent (Schwartz et al., 2012).

Drosophila insulator proteins have been shown to anchor chromatin loops and are enriched at TAD boundaries (Sexton et al., 2012; Van Bortle et al., 2014; Ramirez et al., 2018; Wang et al., 2018). In contrast to vertebrate, *Drosophila* insulators do not rely on cohesin to establish loop interactions. Instead, the BTB/POZ domain proteins, CP190 and Mod(mdg4), play important role in forming insulator-mediated chromatin loop interactions in *Drosophila*. CP190 and Mod(mdg4) could directly interact with DNA-binding insulator proteins and form a homodimer or homomultimer, which bridges long range interactions (Cubenas-Potts and Corces, 2015; Ali et al., 2016). Additionally, other factors including DNA replication Related Element binding Factor (DREF), cohesin, condensin, Chromator, L3(mbt), DNA Topoisomerase II, and M1BP have also been identified to associate or colocalize with insulator proteins. Clustering of these factors are strongly associated with enhancer blocking activity and TAD boundary strength (Van Bortle et al., 2014; Ramirez et al., 2018). High occupancy insulator protein binding sites associate with robust formation of TAD boundaries and insulation strength. Low occupancy sites

appear to associate with gene-specific loops within TAD (Van Bortle et al., 2014). Except the developmental stage specific factor, Elba, most *Drosophila* insulator binding proteins are ubiquitously expressed. However, a subset of insulator protein binding sites are found to be cell type specific (Bushey et al., 2009), which suggest the existence of cell type specific regulation of insulator protein loading. Modifications of CP190 and Mod(mdg4) by sumoylation have been reported to affect their insulator activity in a partner-dependent manner (Capelson and Corces, 2006; Golovnin et al., 2012; Jox et al., 2017). Furthermore, different subclasses of insulator sites may recruit distinct factors to achieve spatiotemporal loading of the architectural proteins and the loop interactions.

-tDNA related factors

tRNA genes were first discovered as a barrier of silent chromatin at the *HMR* locus in *Saccharomyces cerevisiae* (Donze et al., 1999). It's insulator activity is subsequently found conserved throughout yeast, fruit fly, mouse, and human (Scott et al., 2006; Ebersole et al., 2011; Raab et al., 2012; Van Bortle et al., 2014). However, not all tDNAs are insulators. TFIIIC binding at tRNA promoters is essential for tDNA mediated insulator activity (Donze and Kamakaka, 2001; Ebersole et al., 2011; Raab et al., 2012). Furthermore, TFIIIC also binds to many loci throughout the genome that do not have RNAPIII, which is called Extra TFIIIC (ETC) and exist in yeast, fruit fly, mouse and human (Moqtaderi and Struhl, 2004; Moqtaderi et al., 2010; Ebersole et al., 2011; Raab et al., 2012; Van Bortle et al., 2014). TFIIIC sites have been shown to form long range chromatin interactions specifically with other TFIIIC sites, not with intervening RNAPII transcribed genes (Raab et al., 2012). A cohesin complex mutation in yeast disrupted tDNA mediated boundary activity (Donze et al., 1999). Several studies from yeast, mouse, fruit fly, and human subsequently suggest that TFIIIC/Cohesin association is highly

conserved (Ebersole et al., 2011; Van Bortle et al., 2014; Yuen et al., 2017). TFIIIC and CTCF has been found both enriched at TAD boundaries in *Drosophila*, mouse, and human(Van Bortle et al., 2014; Yuen et al., 2017). This suggests that TFIIIC may plan an important role in eukaryotic genome organization.

tDNA sites cluster in the nucleolus in a condensin-dependent manner, but loss of condensin mediated clustering of tDNA at the nucleolus do not significantly change the tDNA transcription (Haeusler et al., 2008). ETC loci in yeast localize to the nuclear periphery in a TFIIIC dependent manner, however, loss of ETC localization to periphery does not affect insulator function (Hiraga et al., 2012). Cohesin depletion in human cells eliminated almost all TADs but did not cause spreading of repressive chromatin domain (Rao et al., 2017). Which further questions the function of tDNA as barriers at chromatin domain boundaries. All these observations in yeast bring up the questions of how TFIIIC functions in genome organization, which remains unclear.

Architectural factors for enhancer-gene chromatin interactions

-Mediator complex

Enhancer-promoter interaction is important for driving high-level and spatiotemporally regulated gene expression. It accounts for a large proportion of chromatin loops. A mediator complex is a transcriptional coactivator in all eukaryotes that is found at both enhancers and promoters of actively transcribed genes. This complex is evolutionarily conserved with compositional variations among different species (Allen and Taatjes, 2015). It could interact with cohesin independently from CTCF to bring enhancers and promoters into physical contacts (Kagey et al., 2010). Knockdown of mediator components weakens the chromatin looping

(Kagey et al., 2010; Phillips-Cremins et al., 2013). Mediator complex has also been implicated in the formation of promoter-terminator looping in yeast genes (Mukundan and Ansari, 2013).

-*Yin and Yang 1*

Yin and Yang 1 (YY1) is a zinc-finger protein that binds hypomethylated DNA sequences. It is ubiquitously expressed and capable of forming homodimers, which could facilitate loop formation (Lopez-Perrote et al., 2014). Recent work from Weintraub et al. showed that YY1 preferentially binds to interacting enhancers and promoters and functions as a structural mediator of enhancer-promoter loop interactions (Weintraub et al., 2017). Deletion of YY1 motif in the promoters resulted in reduced YY1 binding and reduced contact frequency between the enhancers and the promoters. Acute YY1 protein depletion significantly reduced YY1-occupied enhancer-promoter interactions and expression of thousands of genes was changed (up and down), with the greatest changes in genes that originally were occupied by YY1 (Weintraub et al., 2017). Requirement of YY1 in formation of cell type specific enhancer-promoter interactions has been reported at several loci (Hwang et al., 2013; Kleiman et al., 2016; Beagan et al., 2017; Ou et al., 2018; Zaprazna et al., 2018). These evidences argue that YY1 is more of a general structural regulator of enhancer-promoter interactions for a large collection of genes, including cell type specific loops. Cell-type-specific loops are reflection of the cell-type-specific enhancers. YY1 is also found capable of recruiting polycomb complex to DNA target, which plays an important role in silencing genes (Zlotorynski, 2018).

Polycomb proteins

The polycomb group complex (PcG) is one of the prominent epigenetic silencing systems that play important role in cell differentiation and cell identity maintenance (Di Croce and Helin, 2013).PcG proteins are chromatin-associated factors that could locally modify chromatin through

histone modifying activities to regulate their target genes. The core components and silencing mechanism are generally conserved throughout fungi, plant, insects, and mammals (Del Prete et al., 2015). PcG is recruited to thousands of repression initiation sites by various recruitment factors, including DNA binding proteins and noncoding RNAs (Del Prete et al., 2015). However, the recruitment mechanism diverges significantly from fruit fly, to plants, to mammals. In *Drosophila,* PcG is recruited at a Polycomb Response Element (PRE), which are clustered binding sites for proteins that recruit PcG and TrxG complexes. However, the existence of PRE in plants and mammals is poorly defined. GAGA sites, CpG islands, and long noncoding RNA have been shown to play an important role in recruiting PcG in plants and mammals (Del Prete et al., 2015; Entrevan et al., 2016). PcG is crucial in establishment and maintenance of silent chromatin compaction in H3K27me3 marked domains and block the deposition of activating histone marks. Other than the regulatory role at the linear chromatin level, recent studies also revealed that PcG participates in chromatin looping and long-range interactions between TADs marked by H3K27me3 (Sexton et al., 2012; Eagen et al., 2015).

The first evidence for chromatin loops formed by physical interactions of PREs with each other and with target promoters was from *Drosophila* Biothorax complex (Lanzuolo et al., 2007). All the polycomb bound elements form a clustered configuration, while active promoter loops out from the cluster in corresponding cells. The PRC2 component, EZH2, was subsequently shown to mediate multi-loop organization formation and gene silencing of the GATA-4 locus in mammalian cells (Tiwari et al., 2008). In both *Drosophila* and mouse, polycomb mediated clustering could also bring together polycomb regulated regions that are separated over several Mb apart or even on different chromosomes to coordinate and maintain the repressive status (Bantignies et al., 2011; Denholtz et al., 2013). Interestingly, in the mouse *Meis2* locus,

disruption of RING1B mediated loop between a silencer and *Meis2* promoter is required to allow the enhancer to activate Meis2 gene in midbrain. RING1B is also required to bridge the interaction between the enhancer and promoter (Kondo et al., 2014). However, Li et al demonstrated in one *Drosophila* PRE, Mcp, that insulators, not RRE, are required for long-range interactions between polycomb targets(Li et al., 2011). He also showed subsequently that insulator proteins target both transgenes containing a PRE and endogenous PcG regulated genes to polycomb bodies when two genes are repressed (Li et al., 2013a). This indicates that PcG mediated loops are complex and may also involve other looping factors including insulator proteins.

In addition, PRC2 loss showed no effect on TAD boundaries or global contact pattern of the X chromosome in mouse ESCs (Nora et al., 2012). Targeted recruitment of PcG can reposition the original active chromatin compartment into the repressive chromatin compartment (Wijchers et al., 2016). This suggest that PCGs help maintain the repressive chromatin but do not facilitate TAD boundary establishment.

Non-coding RNAs

High throughput analysis of mammalian transcriptome revealed that ~75% of the human genome is transcribed, however, only less than 2% encodes proteins (Tomita et al., 2017). Although, whether all of these are functional is still under debate, it is evident that there are many functional noncoding RNAs. This includes thousands of long noncoding RNAs (lncRNA) of which the transcription is highly regulated during development in both *Drosophila* and mammals (Chen et al., 2016; Tomita et al., 2017). The major functional mechanism for ncRNAs discovered so far is to engage in RNA-DNA, RNA-protein, RNA-RNA interactions through their primary sequences and secondary structures (Guttman and Rinn, 2012). These physical

interactions help recruit transcriptional regulators to a specific locus and regulate target gene expression. Short noncoding RNAs (ncRNA) also play a very important role in heterochromatin formation and maintenance, and in protecting genome stability through suppressing transposon activities (Tomita et al., 2017). Besides its important role in epigenetic regulation in the linear dimension of the genome, ncRNA has started to emerge as an important player in organizing higher order chromatin structures.

Evidence from many organisms revealed that ncRNAs participate in nuclear organization, including establishing insulators and domain boundaries, mediating long-range chromatin loops, forming polycomb bodies and insulator bodies, and anchoring nuclear periphery/matrix associated silencing (Clemson, 1996; Ong and Corces, 2008; Cohen and Jia, 2014; Nwigwe et al., 2015). *Drosophila* RNAi machinery has been reported to influence nuclear organization and function of *gypsy* and dCTCF insulators (Lei and Corces, 2006; Moshkovich et al., 2011). Other than the small ncRNA, mRNA was also found to be a functional component of *Drosophila gypsy* insulator complex (Matzat et al., 2013). Some of the classic chromatin insulators, such as Fab8 and *Scr* DTE, are transcribed (Zhou et al., 1999; Calhoun and Levine, 2003). During X chromosome inactivation, the ncRNAs *Xist* not only targets the silencing histone modification and mediates the relocation of inactive X chromosome to the nuclear lamina, but also causes significant large-scale remodeling of the 3D architecture of the inactive X chromosome. Typical TAD structure disappears and two new 'mega-domains' form across the inactive X chromosome (Nora et al., 2012; Giorgetti et al., 2016). The architectural protein CTCF and mediator complex have been found to be able to directly bind RNA. The mediator complex member, MED12, was reported to bind specifically to ncRNAs called enhancer RNAs (eRNAs, also called activating ncRNA) (Lai et al., 2013). eRNA is a type of relatively short

ncRNA that is transcribed from the enhancer region. Knock down of these eRNAs causes decreased binding of mediator to ncRNA regulated genes and diminishes loop formation between the enhancer and its target gene (Lai et al., 2013). CTCF protein can also directly bind to many ncRNAs and have different effects on regulation. Some ncRNA binds to CTCF and enhance the insulator function by acting as scaffold to stabilize the interaction between CTCF and other factors that mediate looping (Yao et al., 2010). Conversely, some other ncRNAs directly binds to CTCF and interfere with the binding of CFCF to DNA (Sun et al., 2013). Architectural protein YY1 was also recently shown to bind RNA, which was suggested to reinforce transcription factor binding at the regulatory element (Sigova et al., 2015). Additionally, ncRNA transcription could be controlled by signaling factors, such as p53, estrogen receptor α, and NOTCH (Li et al., 2013b; Melo et al., 2013; Trimarchi et al., 2014). This indicates that ncRNAs may help establish and regulate cell type specific enhancer-promoter loops.

However, the current understanding of lncRNAs are mostly based on single lncRNA case studies. This is partially due to the considerable lag and difficulties in lncRNA annotation when compared to protein-coding genes. Although it seems very promising, it's difficult to tell for now whether establishing and maintaining chromatin structure is a general function of ncRNAs.

1.1.3 Mechanism of higher order chromatin organization formation

Loop extrusion with barrier model

The mechanism of higher order chromatin organization formation remains unknown. Most of the current evidence favors the loop-extrusion model. The loop extrusion model proposes that "loop extrusion factors" (LEFs) translocate along chromatin DNA, holding together progressively more distant genomic loci along chromosome and produce dynamically

expanding chromatin loop until they encounter a boundary element that prevent further translocation (Fudenberg et al., 2016). Cohesin and condensin may function as loop extrusion factors, and CTCF and other insulators function as boundaries. This model recapitulates many features of the interphase chromosome organization revealed by HiC and other methods. In addition, cohesin complex is found located at the inner side of loop relative to CTCF sites (Nagy et al., 2016). Since CTCF and cohesin are considered as the major players in the loop extrusion model. Perturbations of CTCF and cohesin would be crucial for testing this model. Recent evidence from CTCF depletion (Nora et al., 2017), cohesin depletion(Rao et al., 2017), and cohesin loading factor, Wapl, depletion (Haarhuis et al., 2017; Wutz et al., 2017) in human cells all strongly supported the role of loop extrusion in interphase chromatin organization. CTCF depletion disrupts CTCF-mediated chromatin looping and TAD insulation (Nora et al., 2017). Cohesin depletion eliminates almost all TADs and chromatin loops(Rao et al., 2017). Wapl depletion led to the loss of restriction of cohesin at CTCF boundary.

An important element of the loop-extrusion model is the active translocation of cohesin along chromatin rather than static interaction. Condensin has been demonstrated as a molecular motor capable of ATP-dependent translocation along DNA (Terakawa et al., 2017) and form chromatin loop by extrusion (Ganji et al., 2018). However, to date, there is no evidence to support the active translocation capability of cohesin as a driver. Transcription process has been shown to act as driving force of chromatin loop extrusion during formation of loops and TADs in interphase chromosome (Ocampo-Hafalla et al., 2016; Racko et al., 2018), but the cohesin-dependent feature in both active and inactive region argues against PolII as the primary driving force for loop extrusion. Additionally, not all the loop anchors associate with cohesin. It would be interesting to find out whether this model also applies to cohesin independent loop formation.

Another important question with this model is how the cohesin and condensin complex successfully slide through loop anchors that are not domain boundaries and, how they only stop at domain boundaries.

Phase separation model

The loop extrusion with barrier model does not explain all the features of the higher order genome organization. Upon cohesin or CTCF depletion, chromatin loops and TAD structures are disrupted, however, chromosomal A/B compartment organization is largely unaffected (Nora et al., 2017; Rao et al., 2017). This indicates that chromosomal compartment organization may use a different mechanism compared to chromatin loops and TAD structures.

A growing body of work has shown that liquid-liquid phase separation can drive the formation of non-membrane compartments inside of the nucleus and cytoplasm (Mitrea and Kriwacki, 2016). The phase separation model proposes that the cooperative binding of specific combination of factors to chromatin DNA leads to high local density of proteins and nucleotide concentration. Such high concentration of multi-molecule assemblies form gel-like phase-separated compartment, also called membrane-less organelles. This model is supported by numerous evidences of the liquid-feature of the nucleus (Mitrea and Kriwacki, 2016). Examples of such phase-separated compartments in the nucleus include nucleoli, which are the largest and best studied phase separated compartment for rapid ribosome biogenesis (Boisvert et al., 2007); nuclear speckles, which are strong compartments for mRNA splicing factors (Boisvert et al., 2007); and Cajal bodies, which are sites for small nuclear RNPs assembly (Morris, 2008). This model is consistent with most of the recently discovered features of the genome organization. Rao *et al* and Nora *et al* showed that depletion of CTCF or cohesin in mammals disrupts TADs, but do not significantly affect active/repressive chromatin domains and the active/repressive

compartments (Nora et al., 2017; Rao et al., 2017). Transcriptional status-based genome structure simulation generates genome contact maps that largely agree with the actual HiC map from many organisms, especially in *C. elegans* and *A. thaliana* which don't have TAD structures (Rowley et al., 2017). The interactions detected between the compartmental domains of the same transcriptional status (Rowley et al., 2017) and the evidence for enrichment of house-keeping genes at a fraction of TAD boundaries in mammals (Dixon et al., 2012) are consistent with the phase separation model.

Phase separation model has been proposed to explain the established and recently described features of transcriptional control via enhancers (Hnisz et al., 2017). These features include super-enhancer clustering, high vulnerability of super enhancers, transcriptional bursting patterns of enhancers, and the ability of an enhancer to produce simultaneous activation of multiple genes (Hnisz et al., 2017). Furthermore, two independent groups recently reported that HP1 bound heterochromatin show liquid-like properties and phase separate into liquid droplets both in *Drosophila* embryos and human cell lines (Larson et al., 2017; Strom et al., 2017). This is consistent with the observations that heterochromatin is assessible to some large protein complexes, such as DNA repair machinery, but is restricted to transcription factors (Watts, 2016). In addition, phosphorylation of HP1 was shown to promote the formation of phase-separated droplets, which suggests that post-translational modifications could be a way of regulating the access to such phase-separated compartment.

3D genome organization might also provide other possible mechanisms of creating novel patterns of gene expression. Local genomic rearrangement is often observed at TAD boundaries in cancer cells. Disruption of TADs usually led to fusion of neighboring TADs, which would bring together previously insulated regions (Guo et al., 2015; Lupianez et al., 2015; Sanborn et

al., 2015; Flavahan et al., 2016; Hnisz et al., 2016). Acemel et al. proposed a new model that processes such as gene duplication and neofunctionalization, which classically thought to occur in a stepwise manner, can actually occur simultaneously with the formation of neo-TADs (Acemel et al., 2017).

1.2 Genome organization and genome evolution

The transcriptional regulatory function of cis-regulatory elements on genes put evolutionary selection pressure on genome rearrangement that breaks the synteny between regulatory elements and their target gene. Considering that TADs might represent the functionally isolated units of the genome, they might also provide structural basis for genome evolution. Consistently, it has been showed that the synteny blocks of regulatory elements and gene tend to be conserved, and the boundaries of synteny blocks coincide with the boundaries of TADs (Berthelot et al., 2015). (Harmston et al., 2017). Most examples in the literature that reported TADs disruption are associated deleterious effect (Trimarchi et al., 2014; Guo et al., 2015; Lupianez et al., 2015; Flavahan et al., 2016; Hnisz et al., 2016; Luo et al., 2018). The relative conservation of TADs across different species indicates that genetic alterations that disrupt TADs are negatively selected during evolution. For instance, about 76% of mouse TAD boundaries are also boundaries in human, and about 54% of human boundaries are also boundaries in mouse (Dixon et al., 2012). These are all consistent with the proposed scenario that TADs, as a functionally isolated structural unit, might be shuffled in an evolutionarily neutral fashion during evolution (Berthelot et al., 2015). However, current understanding of the association of synteny block breakpoints and TAD boundaries is mainly based on these low-resolution studies. In Harmston's experiment, the synteny block borders and TAD boundaries matching was counted within a range of a few hundred kb to 1Mb. Only 29% of synteny blocks

overlap within 120kb of TADs in human (Harmston et al., 2017). In Dixson's analysis of TAD conservation, the comparison is based on low resolution HiC map (20kb) and relatively relaxed boundary size cutoff (size less than 400kb), which may neglect the changes in smaller scale. Rowley et al. showed in a high-resolution HiC map in *Drosophila* that TAD boundaries revealed in low resolution (10-40kb) are actually composed of smaller compartment domains which contains one or multiple genes of the same transcriptional status (Rowley et al., 2017). This makes it hard to understand the mechanistic relationship between genomic rearrangement breakpoints and TAD boundaries at such low resolution.

Although a genome rearrangement that disrupts a TAD is disfavored during evolution, analysis from the evolutionary perspective demonstrated that TADs may facilitate the appearance of new CREs by allowing the interactions of these novel CREs with their potential target genes that may be located a few hundreds of kilobases away but in the same TAD (Acemel et al., 2017). Changes in 3D structures at certain loci have been shown to be required for new regulatory mechanisms associated to morphological novelties during evolution. The best example is the Hox cluster. HoxA and HoxD clusters are essential for proper limb development. The proper regulation of each of these two clusters is closely linked to two neighboringTADs, with a developmentally shifting boundary dividing the cluster into two parts. This 3D configuration facilitates Hox genes on each side of the TAD boundary to preferentially contact regulatory elements located either on the anterior or posterior side of the Hox cluster (Andrey et al., 2013). The developmentally controlled shifting of TAD boundary between central Hox genes is critical for proper limb development (Andrey et al., 2013). Comparative genomic study from Acemel et al showed that the conserved synteny block of the Hox cluster in vertebrates includes both 5' and 3' regions, which contain the regulatory regions (Acemel et al., 2016). However, in

amphioxus, which is an invertebrate chordate, the Hox cluster synteny blocks only share with the vertebrate on the 3' side. Subsequent reconstructed 3D architecture revealed that amphioxus Hox cluster is only imbedded in a single TAD. This indicates that there must be a chromosomal rearrangement that brought a new genomic region to the 5' side of the Hox cluster between the last common ancestor of chordates and the last common ancestor of vertebrates. The new bisected TADs at this locus provided structural basis for the possibility of switching two separate sets of long range interactions, which is required for the limb development in vertebrates.

1.3 Three-dimensional genome organization and known human disease

Disruption of TAD boundaries causes pathogenic rewiring of gene-enhancer interactions in mammal's limb development (Lupianez et al., 2015). Distinct human limb malformation is associated with deletions, inversions, or duplications in the WNT6/IHH/EPHA4/PAX3 locus (Lupianez et al., 2015). A CRISPR-cas9 mediated mouse model that mimics the human disease alleles demonstrated that disruption of CTCF-associated boundary cause ectopic wiring of EPHA4 enhancers with nearby developmental genes and drive ectopic expression in the limb (Lupianez et al., 2015).

Evidence for disruption of chromatin organization in cancer has been reported as a result of either chromosomal rearrangements or compromised CTCF binding near oncogenes or tumor suppresser genes (Hnisz et al., 2016; Luo et al., 2018). Hox gene dysregulation is a common feature of acute myeloid leukemia. Disruption of CTCF sites in mouse HoxA cluster demonstrated its role in regulating Hox gene expression and maintaining the oncogenic transcription program for leukemic transformation (Luo et al., 2018). Hypermethylation of these binding sites caused CTCF insulator dysfunction which causes upregulation of the oncogene in IDH mutant gliomas (Flavahan et al., 2016). Chromosomal rearrangements, a hallmark in

cancers, often lead to the loss of insulation and activate oncogenes by exposing them to new enhancers. Microdeletions at boundary elements are often found in leukemia patient genomes (Hnisz et al., 2016). Genome regional mutation rate has been shown to be associated with DNA accessibility (Schuster-Bockler and Lehner, 2012). Indeed, CTCF/cohesin has been reported as a major mutational hotspot in multiple cancer types (Katainen et al., 2015). Recurrent genomic breakpoints of translocations in leukemia correlates with disruption of TAD boundaries (Valton and Dekker, 2016). In addition, topoisomerase II mediated the release of torsional strain at chromosomal loop anchors which generates DNA double-strand breaks that drive multiple oncogenic translocations (Canela et al., 2017). All these evidences suggest that higher order genome organization is crucial for maintaining proper gene regulation.

1.4 Manipulate chromatin organizations

Approaches that can efficiently manipulate linear and 3D organization of genome are key to determine the functional significance of genome organization. As the importance of chromatin organization emerging, manipulation of chromatin organization has been attempted to either reinstate the silent status of heterochromatin region or dissect the function of 3D genome organization in gene regulation. These manipulation methods include but are not limited to targeted DNA sequence manipulation (Guo et al., 2015; Lupianez et al., 2015; Narendra et al., 2015; Willi et al., 2017; Luo et al., 2018), architectural protein depletion (Van Bortle et al., 2012; Nora et al., 2017; Rao et al., 2017; Wutz et al., 2017), forced chromatin looping (Ahanger et al., 2014; Deng et al., 2014; Hao et al., 2017), targeted local chromatin remodeling (Casas-Delucchi et al., 2012; Wijchers et al., 2016), and the chemically inducible CRISPR-mediated chromosomal looping (Morgan et al., 2017).

With the prevalent use of homologous recombination-based genome editing and CRISPR-cas9 based genome editing techniques, targeted genome sequence manipulation is no longer a big challenge. Deletion, or inversion of boundary element containing CTCF sites led to mis-regulation of genes, spreading of active/repressive chromatin domains, and reorganization of local chromatin structure (Bantignies et al., 2011; Guo et al., 2015; Lupianez et al., 2015; Narendra et al., 2015; Willi et al., 2017; Luo et al., 2018). Inducible depletion of architectural proteins recently has been used to address the relationship between genome architecture and gene regulation (Nora et al., 2017; Rao et al., 2017; Wutz et al., 2017). The inducible degradation system of the plant hormone, auxin, enables acute conditional and reversible depletion of proteins in vivo (Zhang et al., 2015). The forced protein-protein interaction in combination of targeted protein binding to specific loci allow researchers to dissect the relationship 3D organization and regulatory genomic functions. Programable DNA binding zinc-finger proteins or CRISPR/dCas9(endonuclease defective cas9) have been used to target protein of interest to a specific locus in the genome. Fusing a forced protein-protein interaction system to the targeted DNA binding protein allows artificially forming chromatin loops inside of nucleus (Ahanger et al., 2014; Deng et al., 2014; Hao et al., 2017; Morgan et al., 2017). Adding a chemically inducible protein-protein component in the bivalent dCas9 system even allows precise temporal control of chromatin loop formation in vivo. However, the potential influence of the widely used chemical inducer, plant hormone auxin or ABA, on animal cells is under debate (Vildanova and Smirnova, 2016). The efficiency of chromatin loop formation in the bivalent dCas9 system still needs to be improved (Hao et al., 2017). Nevertheless, the forced (inducible) chromatin looping technique will significantly aid the understanding of endogenous loops and enabling creation of new regulatory connections. Targeted local chromatin remodeling

has also been demonstrated to be an effective method in manipulating higher order genome organization. Heterochromatin targeted MeCP2 reinstate the de-repressed heterochromatin in rett mutant cells (Casas-Delucchi et al., 2012). Wijchers used lacO/lacR system fused with NONOG, SUV39H1, or EZH2 demonstrated “locked” and “unlocked” genomic region in the context of 3D genome. Targeted modification of local chromatin status at the “unlocked” locus is capable of driving the switching between A/B compartments (Wijchers et al., 2016).

The fast development of these new tools that allows direct manipulation of 3D genome *in vivo* will significantly help researchers to dissect the detailed mechanisms of higher genome organization formation and the relationship between 3D chromosomal organization and genome function.

CHAPTER 2

CHROMATIN LOOP ANCHORED BY INSLATOR SF1 AND SF2 REGULATES PROPER *SCR/FTZ* GENE EXPRESSION DURING *DROSOPHILA* EMBRYONIC DEVELOPMENT

2.1 Introduction

Three-dimensional organization of the eukaryotic genome are highly complex, both in terms of its complex dynamic organization, and its relationship to genome function. Recent chromosomal conformation capture (3C)-based experiments, in junction with microscopy-based evidence, revealed that local scale chromatin loops consist the basic unit of the eukaryotic genome (Kantidze and Razin, 2009; Sexton et al., 2012; Maeshima et al., 2014; Rao et al., 2014; Ou et al., 2017). A major subset of these loops is mediated by a specialized DNA-protein complex called chromatin insulators, also called chromatin boundary elements (CBEs). Chromatin insulators have been well characterized by their ability to block distal regulatory elements (such as enhancers and silencers) from acting on the promoters and restricting the spread of repressive/active chromatin domains (Valenzuela and Kamakaka, 2006; Gurudatta and Corces, 2009; Van Bortle and Corces, 2012; Ali et al., 2016). In addition, evidence from the genome wide 3C based-experiments showed that chromatin insulators are enriched at Topologically Associated Domain (TAD) boundaries as well as local chromatin loop anchors, which are mostly located in the intergenic regions. These observations suggest that chromatin insulators play important role in insulating genes from neighboring genes as well as organizing the higher order chromatin organization. However, the mechanistic relationship between

insulator mediated chromatin loops and gene regulation at the endogenous genomic context is still unclear.

Current understanding of insulator functions mainly comes from enhancer blocking activity tests and the heterochromatin barrier activity test on transgene constructs. Genome-wide mapping of chromatin insulator sites revealed that insulator proteins localize to thousands of sites characterized by conserved motifs. About 50-80% of these sites are located in the intergenic regions in both *Drosophila* and mammals (Kim et al., 2007; Bushey et al., 2009; Jiang et al., 2009; Negre et al., 2010). Furthermore, chromatin insulators also localize to the borders of repressive/active chromatin domains throughout yeast (Donze et al., 1999; Scott et al., 2006), *Drosophila* (Bushey et al., 2009; Negre et al., 2010; Van Bortle et al., 2012), and mammals (Cuddapah et al., 2009). The genome distribution of the insulator sites is consistent with the chromatin insulator model learnt from experiments done in various artificial transgene systems. However, a recent study in *Drosophila* demonstrated that most of these insulator binding sites have little enhancer blocking activity and the enhancer blocking activity of a given insulator site is insertion site dependent (Schwartz et al., 2012). This suggest two possible scenarios of these many thousands of insulator sites: chromatin insulator may function in a genomic context dependent manner; Or unlike the prototype insulators, the majority of the insulator binding sites in the genome don't have robust enhancer blocking activity. Recent depletion of CTCF protein in mammalian cells supports the first scenario. Upon CTCF depletion, many genes are upregulated due to loss of insulation from their neighboring enhancers (Nora et al., 2017). However, the functional role of insulator sites at their endogenous genomic context remains largely unclear.

Drosophila has been the pioneering model organism used in chromatin insulator studies. In contrast to vertebrates, multiple insulator proteins including, dCTCF, BEAF-32, Su(Hw), zw5,

GAF, CP190, Mod(mdg4), Elba, Pita, ZIPIC, Ibf1, and Ibf2 have been identified in *Drosophila* (Bushey et al., 2009; Aoki et al., 2012; Cuartero et al., 2014; Maksimenko et al., 2015). Intensively studied chromatin insulator elements in *Drosophila* include but are not limited to Mcp, Fab7, and Fab8 elements in the Bithorax complex, the SF1 insulator in the Antennapedia complex, Homie and Nhomie insulators in the *eve* locus, scs and scs' insulators in the *hsp70* locus, and gypsy insulators. The *Drosophila* Hox cluster has been used as a model locus to study the function of insulators in regulating developmental genes. It has been postulated from transgenic studies that multiple insulators located within the regulatory sequences of *Abdominal B (Abd-B)* gene restrict the expression of *Abd-B* into segment-specific domains. However, studying of chromatin insulators at their endogenous genomic context has been very rare (Wolle et al., 2015).

We have previously identified a chromatin insulator, named SF1, within the large regulatory region of the homeotic gene *Sex comb reduced (Scr)* (Belozerov et al., 2003). The *Scr* locus in the *Drosophila* Antennapedia complex serves as a very good model to study all the proposed functions of chromatin insulators that we learnt from transgenic systems, which includes enhancer blocking, barrier, and loop formation. The *Scr* upstream regulatory region is interrupted by a nested non-homeotic gene, *fushi tarazu (ftz).* The ~14kb *ftz* gene unit splits the *Scr* regulatory region into proximal and distal cis-regulatory regions (Fig. 2.1B). The distal cis-regulatory region contains key cis-regulatory elements of *Scr*, including the parasegment 2 (PS2) and thoracic segment 1 (T1) enhancers which activate *Scr* in the PS2 and T1 tissue during embryogenesis and a polycomb response element (PRE) that maintains the *Scr* repression in late development stages (Fig. 2.1B) (Pattatucci et al., 1991; Gindhart et al., 1995). The *ftz* gene unit also contains multiple distal enhancers (Hiromi et al., 1985; Pick et al., 1990; Schier and

Gehring, 1993). Despite such complex cis-regulatory network and the juxtaposition of the *ftz* and *Scr*, these two neighboring genes are expressed in distinct spatial-temporal patterns. It has been suggested that SF1 may insulate the *Scr* gene from its neighboring cis-regulatory elements of the *ftz* gene (Belozerov et al., 2003). However, it remains unclear how the *ftz* gene is insulated from the *Scr* distal cis-regulatory elements and how *Scr* distal enhancers overcome the long genomic distance (~25kb) and the block of the SF1 insulator (Fig. 2.1B). Furthermore, *ftz* is active in tissues where *Scr* may be repressed. How is *ftz* protected from the polycomb silenced chromatin that maintains the repression of *Scr*? We hypothesize that the SF1 insulator may interact with a yet unknown boundary downstream of *ftz*, designated SF2, to form a chromatin loop. This putative loop would insulate *ftz* gene from influence by the *Scr* enhancers as well as insulate *ftz* enhancers from acting on *Scr* promoters. Both insulators will also function as barriers to the neighboring silent chromatin.

In this project, we identified the SF2 insulator located between the *ftz* gene downstream enhancer and the *Scr* enhancers in the distal region. Both SF1 and SF2 are bound by known insulator associated protein factors in 0-12 hours old embryos as well as in cell lines derived from different developmental origins. SF1 and SF2 locate precisely at the boundaries of polycomb-mediated silent chromatin domains. 3C experiments showed that interaction between SF1 and SF2 is developmentally regulated. By using a transgene assay, we showed that SF2 has strong enhancer blocking activity and pairing of SF1 and SF2 cancels enhancer blocking activity in an orientation dependent manner. We also demonstrated that both SF1 and SF2 insulators are functionally conserved in distantly diverged *Drosophila* species. CRISPR-Cas9 mediated knockout of SF1 or SF2 not only leads to the misregulation of *Scr* and *ftz* genes, but also affected

the chromatin architecture in the Antennapedia complex. Flies with SF1 or SF2 deletion have significantly reduced overall-fitness during development.

2.2 Results

2.2.1 Identification of SF2 in the *Scr* distal regulatory region.

To search for potential insulators that loop with SF1, we used 3C to search the entire *Scr* distal regulatory region for DNA sequences that capture SF1 in early (4-8hrs) and late stage (10-14hrs) fly embryos (Fig. 2.1C). Since genomic sites located within 100kb are known to capture each other at relatively high background frequencies, a distance-frequency background reference curve was used (dashed curve in Fig. 2.1C). Among the eleven EcoRI fragments (named R1 to R11 in Fig. 2.1C), R1, R5-R6, and R10 capture SF1 in the early stage at a frequency of 50% or higher above the expected distance-capture frequency background (Fig. 2.1C). In contrast, other regions capture SF1 at near or below background level (Fig. 2.1C). R1 and R5-R6 frag only capture SF1 in the early stage (4-8hrs), but not in the late stage (10-14hrs) embryos. These results indicate that SF1 interacts with multiple DNA elements in the *ftz* downstream *Scr* distal regulatory region to form developmentally regulated chromatin loops in fly embryo. It is worth noting that R1 is located in between *ftz* stripe 1/5 enhancer and the *Scr* distal regulatory region, where the hypothesized SF2 is located (Fig. 2.1B). Interaction between R1 and SF1 could loop out *ftz* gene, bringing the distal enhancers closer to the *Scr* promoter. Similarly, SF1-R6 fragment pairing could facilitate interactions between *Scr* promoter and the late polycomb response element (PRE). The relatively high capture of R4 to R7 in the early stage (4-8hrs) embryos and low capture frequency of R4 to R6 in the late stage (10-14hrs) embryos coincide very well with the expression pattern of a developmentally regulated long ncRNA (Fig. S2.1). Long ncRNA has been shown to interact with insulators in *cis* and paly important role in both *Drosophila* and

mammals (Clemson, 1996; Ong and Corces, 2008; Cohen and Jia, 2014; Nwigwe et al., 2015). However, whether this lncRNA is involved in the loop formation and *Scr/ftz* gene regulation in this region still needs to be further tested and is not the focus of the current study. R10 is located at the end of *Scr* distal regulatory region. This region is bound by multiple insulator proteins (Negre et al., 2010) and has been shown to also tether promoter region (Calhoun and Levine, 2003). An SF1-R10 loop could separate *Scr* enhancers from the neighboring *Antp* gene.

Although SF1 interacts with multiple DNA element in the *Scr* distal regulatory region, my project mainly focused on the R1 fragment, which I thought would be the best candidate for SF2 insulator. The timing of loop formation and the extent of SF1-R1 coincide precisely with the timing of *ftz* gene activity (Fig. 2.5A-F) and the limit of *ftz* gene domain (Fig. 2.1B, Fig. 2.3B). The strong capture of SF1 and R1 occurs at a time where both *Scr* and *ftz* genes are activated by their long-range enhancers. Ectoderm expression of the *ftz* gene stops after ~8hrs in fly embryos (Fig. 2.5E-F) (Kellerman et al., 1990), and SF1-R1 loop falls apart in the late stage embryos (10-14hrs) (Fig. 2.1C).

2.2.2 A conserved 2kb sub-fragment of R1 fragment contains insulator activity

Since the 3.9kb R1 fragment was arbitrarily generated by the restriction enzyme used in the 3C experiment, we further narrowed down the SF2 candidate sequence to a 2kb sub-fragment of R1 fragment by referring to the major known insulator proteins' binding profiles (Fig. 2.3A) (Negre et al., 2010). Known insulator proteins, such as CTCF, CP190 and Mod(mdg4), bind both at SF1 and the 2kb SF2 candidate in 0-12hrs old embryos. In addition to those classic insulator factors, newly discovered insulator protein, ZIPIC equally binds both SF1 and the 2kb SF2 candidate. Insulator binding factor (Ibf1) and Ibf2 preferably bind to the 2kb SF2 candidate over SF1 in S2 cells (Fig. S2.3) (Maksimenko et al., 2015). Su(Hw) and Mod(mdg4) only bind to SF2

in S2 cells (Fig. S2.3). These suggests that this 2kb sub-fragment may be a good candidate of the SF2 insulator. Furthermore, Ibf1/Ibf2 complex has been reported to be required for CP190 associated insulator and may function independently from CTCF. High affinity binding of Ibf1/2 complex and Su(Hw)-mod(mdg4) complex at SF2 suggest that SF2 may be different in protein complex components than SF1.

To test whether SF2 has enhancer-blocking activity, we tested this 2kb SF2 candidate for enhancer blocking activity by using an established insulator assay in transgenic fly embryos (Belozerov et al., 2003; Majumder and Cai, 2003). The assay transgene contains two tissue-specific enhancers, Neuroectoderm (NEE) and Hairy strip 1 (H1), between two divergently transcribed *lacZ* and *miniwhite* reporters (pBWNHZ, Fig. 2.2A). Previous studies have shown that insulators such as SF1 can block the distal NEE enhancer and reduce *lacZ* reporter gene expression in the horizontal stripes when inserted between NEE and H1 enhancers (Cai and Levine, 1997; Belozerov et al., 2003; Majumder and Cai, 2003). When the 2kb SF2 candidate sequence was inserted between NEE and H1 with either orientation, it significantly reduced the NEE-driven *lacZ* expression without affecting the anterior vertical stripe driven by the H1 enhancer (Fig. 2.2F-G,Q). Consistently, SF2 also blocked the H1 enhancer without affecting NEE function on the *miniwhite* reporter (Fig.2.2K). Work from our lab, as well as others, have previously shown that two *gypsy* insulators flanking an enhancer increases the blocking efficiency of the enclosed enhancer, possibly due to improved paring from the proximity of the two insulators (Cai and Shen, 2001; Kyrchanova et al., 2013). To test whether SF1 in proximity can augment the enhancer-blocking activity of SF2, we inserted a copy of SF1 distal to NEE in the downstream region of *miniwhite* reporter gene (~4.5kb away from NEE, Fig. 2.2A). Indeed,

increased blocking of NEE was observed in these transgenic embryos (Fig. 2.2H,Q). This result suggests that SF1 and SF2 can loop with each other to better block the NEE enhancer.

We've also noticed that the *ftz* gene domain is presented in an inverted orientation in some distantly diverged *Drosophila* species (Fig. 2.8). To see whether SF1 and SF2 are functionally conserved among *Drosophila* species. We first aligned SF1 and SF2 sequences to the other 11 *Drosophila* species whose genome has been sequenced. These 12 *Drosophila* species represent a large range of divergence time (up to 40 million years) during evolution. Both SF1 and SF2 are conserved at the DNA sequence level (Fig. S2.2A-B). We further tested the enhancer-blocking activity of the SF1 and SF2 orthologs cloned from *Drosophila willistoni*, which diverged from *Drosophila melanogaster* about 35-million years ago and has an inverted *ftz* gene unit in *Drosophila melanogaster* transgenic embryos. Both SF1 and SF2 cloned from *Drosophila willistoni* showed strong enhancer-blocking activity (Fig. 2.2I-K), that are similar to *Drosophila melanogaster* SF1 and SF2 (Fig. 2.2D,F,Q). These results suggest that SF1-SF2 formed chromatin loop plays a conserved role in regulating independent *Scr/ftz* gene expression among *Drosophila* species.

2.2.3 SF1-SF2 paring facilitates enhancer bypass in transgenic embryos

A unique behavior called "enhancer bypass" has been reported for certain chromatin insulators, in which the blocking of the distal enhancer is neutralized when a pair of such insulators are put together in cis (Cai and Shen, 2001; Muravyova et al., 2001; Melnikova et al., 2004; Kyrchanova et al., 2007; Kyrchanova et al., 2016). We have proposed in our hypothesis that SF1 and SF2, when paired *in vivo*, could allow the *Scr* distal enhancers to bypass both insulators to interact with the *Scr* promoter (Fig. 2.8). To test this hypothesis, we inserted both SF1 and SF2 in tandem between the NEE and H1 enhancers in the enhancer-blocking transgene.

Indeed, we observed a complete recovery of NEE-driven lacZ expression and H1-driven miniwhite expression in these transgenic embryos (Fig. 2.2M). It's worth noticing that SF1-SF2 pairing cancels enhancer-blocking activity in arrangement-orientation dependent manner (Fig. 2M-N). The SF1 and SF2 pair only cancel the enhancer blocking activity when they are arranged in the same orientation (based on the orientation presented in the genome) (Fig. 2.2M). Pairs arranged in opposite directions (Fig. 2.2N) not only do not allow bypassing of distal NEE enhancer, but also enhances the enhancer blocking activity of each insulator alone (Fig. 2.2Q). This is consistent with the orientation-dependent interactions observed in other insulators (Kyrchanova et al., 2007; Kyrchanova et al., 2013; Rao et al., 2014; Fujioka et al., 2016). The enhancer bypass activity of insulator pairing is considered as a strong evidence for insulator-insulator interaction *in vivo* (Cai and Shen, 2001; Muravyova et al., 2001; Melnikova et al., 2004; Kyrchanova et al., 2008b; Kyrchanova et al., 2016). To test if SF1-SF2 interaction observed in *Drosophila melanogaster* is conserved in other distantly diverged *Drosophila* species, we tested enhancer-bypass activity of SF1 and SF2 orthologs cloned from *Drosophila willistoni*. The result shows that *Drosophila willistoni* SF1 and SF2 showed surprisingly similar orientation-dependent enhancer bypass activity (Fig. 2.2O-Q). This indicates that the SF1-SF2 formed loop may be a functionally conserved feature that is required for proper regulation of *Scr* and *ftz* gene in *Drosophila* species. This is the first conserved enhancer bypass reported for two endogenous pairing insulators.

2.2.4 SF1-SF2 loop demarcates the active *ftz* chromatin domain in *Drosophila* embryos

During embryogenesis, the *ftz* gene is highly expressed in many parasegments/segments where *Scr* is repressed due to the assembly of polycomb mediated silent chromatin (Fig. 2.4A-E, Fig. 2.5A-E) (Mahaffey and Kaufman, 1987; Doe et al., 1988; Krause et al., 1988; Pattatucci and

Kaufman, 1991; Gindhart et al., 1995). Barrier activity has been known as one of the characteristic features of chromatin insulators (Valenzuela and Kamakaka, 2006; Gurudatta and Corces, 2009). To assess whether the SF1-SF2 formed loop plays any role in demarcating distinct chromatin domain boundaries, we examined the histone modification profile of *Drosophila* embryos in different developmental stages (ModENCODE data set: http://gbrowse.modencode.org/fgb2/gbrowse/fly/) (mod et al., 2010). As shown in fig. 2.3B, the SF1 and SF2 loop detected in the early stage (4-8hrs) embryos coincide precisely with a domain of low H3K9me3 and H3K27me3 repressive chromatin markers at this stage. The highly active transcription of the *ftz* gene in about half of the body segments during this stage is highly associated with the strong depletion of repressive marks in this domain (Fig. 2.3B). Furthermore, the disassociation of SF1-SF2 formed loop in the late stage (10-14hrs) embryos is accompanied by the spreading of silent chromatin into the *ftz* domain (12-16hr, Fig. 2.3B). The active/poised enhancer marker, H3K4me1, are enriched at the enhancer elements in both *ftz* and *Scr* domains during early development (Fig. S2.4), consistent with enhancer mediated activation of both genes (Hon et al., 2009; Boros, 2012; Calo and Wysocka, 2013; Koenecke et al., 2017). The difference of enhancer-driven transcription between *Scr* and *ftz* is also evident in their transcription starting time during fly embryogenesis (Hiromi et al., 1985; Martinez-Arias et al., 1987; Doe et al., 1988; Krause et al., 1988; LeMotte et al., 1989). Altogether, these evidences indicate that *ftz* and *Scr* genes are independently regulated in separate domains and the dynamically regulated SF1-SF2 loop could underlie the chromatin domains in this region.

2.2.5 Knock-out SF1 cause mis-regulation of *Scr* gene

To further evaluate the functional importance of SF1 and SF2 insulator in *Scr* and *ftz* gene regulation, we generated SF1 or SF2 knockout flies using CRISPR-Cas9 technique.

Knockout mutant lines were backcrossed multiple generations with w^{1118} wild type to minimize the variation from potential genomic background differences. RNA in situ hybridization against *Scr* and *ftz* gene was applied to analyze the transcriptional pattern in SF1-knockout or SF2-knockout mutant embryos. Deletion of SF1 or SF2 both caused enhancers-promoters cross-communication between *Scr* and *ftz* gene.

Scr gene is expressed in parasegment 2 in early stage embryos (Fig. 2.4A-E) (Mahaffey and Kaufman, 1987; Martinez-Arias et al., 1987; LeMotte et al., 1989), which is mainly driven by the PS2 enhancer located ~25kb away in the *Scr* distal regulation region (Fig. 2.1Y) (Gindhart et al., 1995). In SF1-knockout embryos, *Scr* gene is expressed in a 7-strip *ftz* gene pattern in all even-numbered parasegments during early embryonic stage (Fig. 2.4I-L). Not all the 7 strips are stained equally: the first (parasegment 2) and the second (parasegment 4) stripes stain the strongest, followed by the sixth strip, and the remaining four stripes show rather weak staining (Fib 2.4I-K). It has been reported that the relatively even-distributed *ftz* transcripts in 7 strips (Fig. 2.5A-D) are achieved through the cooperation of multiple *ftz* enhancers (Hiromi and Gehring, 1987; Schier and Gehring, 1993). The *ftz* distal enhancer cluster (red oval "D" in Fig. 2.4V) contains a separable PS4 enhancer sub fragment., which is located on the distal end from *ftz* gene and is closet to *Scr* promoter (Hiromi and Gehring, 1987). Strong parasegment 4 expression of *Scr* gene after SF1 knockout may be due to the close proximity of this *ftz* PS4 enhancer to the *Scr* promoter.

Knockout of SF1 also abolished the long-range *Scr* PS2 enhancer-driven *Scr* expression in parasegment 2 (Fig. 2.4J-L). In wild type, both *Scr* and *ftz* are expressed in parasegment 2 at the start of gastrulation (stage 5) and the *ftz* gene is more restricted to the anterior part of the parasegment as the embryos further develop until the germ-band is fully extended (Fig 2.5 B-E)

(Doe et al., 1988; Krause et al., 1988; Kellerman et al., 1990) The ectoderm 7-stripe expression of the *ftz* gene starts to fade out around stage 8 (Fig. 2.5D) and completely disappears around stage 10 (Fig. 2.5E). Stripe 1 of the *ftz* gene fades out relatively earlier comparing to the other stripes (Fig 2.5C-E) (Krause et al., 1988; Kellerman et al., 1990). The *Scr* gene is expressed in the full parasegment 2 throughout early embryonic stages (Fig 2.4A-E) (Mahaffey and Kaufman, 1987; Martinez-Arias et al., 1987). Starting around stage 8, the parasegment 2 expands in width and *Scr* expression in the full parasegment 2 becomes further intensified (Fig. 2.4D-E) and the mesoderm parasegment 3 expression starts to emerge (Fig 2.4E) (Mahaffey and Kaufman, 1987; Martinez-Arias et al., 1987; LeMotte et al., 1989; Paré et al., 2009). However, in SF1-knockout embryos, *Scr* transcripts in parasegment 2 shows very similar pattern to that of *ftz* gene during early embryonic stages (Fig 2.5A-E). Parasegment 2 expression of *Scr* is restricted to the anterior part of the parasegment 2 and it gradually fades out from stage 7 to stage 9 (Fig 2.4J-L). No *Scr* transcripts are detected in the parasegment 2 except the residual level remaining near the anterior border from stage 8 to stage 9 (Fig 2.4K-L), when the parasegments expand in width and normally the *Scr* transcripts further accumulates in parasegment 2 (Fig 2.4C-D). This indicates that the original long-range *Scr* PS2 enhancer-driven *Scr* expression in parasegment 2 seems to be abolished and *Scr* expression is mainly driven by the ectopic *ftz* distal enhancers (Fig. 2.4Y). After stage 10, *Scr* expression in the SF1-knockout embryos seem to be similar to the wild type till the very late stage of embryogenesis (Fig. 2.4E-F,M-N). This suggests that the enhancer-driven *Scr* expression switches to different enhancer(s) after mid-embryogenesis (Gindhart et al., 1995). At the very late stage of embryogenesis, stage 15-17, SF1-knockout embryos show weak ectopic *Scr* expression in a region between the labial and subesophageal ganglion of CNS (Fig.

2.4O-P). Salivary glands in SF1-knockout embryos also shows high frequency of abnormal positioning and *Scr* staining (Fig. S2.5C-D,E).

SF2-knockout embryos did not show obvious mis-regulation of the *Scr* gene (Fig. 2.4Q-X). In summary, deletion of the SF1 insulator not only leads to the loss of insulation between *ftz* distal enhancers and the *Scr* promoter, but also abolished the long-range communication between the *Scr* promoter and the *Scr* PS2 enhancer located 25kb away in the distal regulatory region (Fig 2.4Y).

2.2.6 Knockout of SF2 cause ectopic expression *ftz* gene in *Scr* pattern

In wild type, the *ftz* gene is expressed in a distinct 7-stripe pattern in the ectoderm of all odd-numbered parasegments during early embryonic stage (Fig. 2.4A-E) (Hiromi and Gehring, 1987; Krause et al., 1988; Schier and Gehring, 1993). The ectoderm expression of *ftz* completely fades out around mid-embryogenesis when the germ-band is fully extended (Fig. 2.5A-E). *ftz* is then transiently expressed a second time in the developing CNS (Fig. 2.5F-G) (Hiromi and Gehring, 1987; Doe et al., 1988; Krause et al., 1988). To see if deletion of SF1 or SF2 cause ectopic action of neighboring *Scr* regulatory elements on *ftz* gene, we analyzed the *ftz* transcript pattern in wildtype, SF1-knockout, and SF2-knockout fly embryos. SF1-knockout embryos have no obvious mis-regulation of the *ftz* gene (Fig. 2.5H-N). In SF2-knockout embryos, in addition to the normal 7-stripe expression in the ectoderm during early embryonic stage, *ftz* gene is ectopically expressed in ectoderm of parasegment 2 from stage 6 to late stage 10 (Fig. 2.5O-T). This ectopic expression in parasegment 2 coincides very well with the *Scr* PS2 enhancer activity pattern (Fig. 2.4A-E,I-M), which suggests that *Scr* PS2 enhancer located in the *Scr* distal regulatory region starts to act on the neighboring *ftz* gene promoter after the loss of the SF2 insulator (Fig. 2.5V).

2.2.7 SF1 and SF2 insulators facilitate tissue- and stage-specific 3D chromatin organization of Antennapedia complex

It has been reported that disruption of chromatin boundaries not only cause mis-wiring of long-range enhancers, but also reorganized the three-dimensional TAD structure (Guo et al., 2015; Lupianez et al., 2015). Unpublished data from our lab also suggests that SF1 function as an organizational hub of many chromatin loops in the ~350kb Antennapedia complex (Li and Ma, unpublished). To see if the deletion of SF1 or SF2 affects the local chromatin architecture, we performed 3D DNA FISH with probes targeting the proximal, middle, and distal end of the Antennapedia complex (Fig. 2.6A) at early and late embryonic developmental stages. Three different tissues (head, thoracic segments, and posterior segments) represent different gene activities of *labial (lab), Scr*, and *Antennapedia (Antp).*

In stage 5 embryos, deleting either SF1 or SF2 had little effect on the average three-dimensional distance in the head tissue, but caused opposite effects in thoracic segments and posterior segments tissues (Fig. 2.6B). In thoracic segments, deleting either SF1 or SF2 significantly decreased the average physical distance between the *Scr* and *lab* gene when compared to that in wild type (t-test, $p = 0.01$)(Fig. 2.6B). In contrast, deleting SF1 in the posterior segments caused significant increase in average physical distance between each pair of the three hox genes: *Scr-lab* (t-test, $p = 0.004$), *Scr-Antp* (t-test, $p = 0.003$), and *lab-Antp* (t-test, $p = 0.04$). All three genes, *lab, Scr,* and *Antp* genes are transcriptionally repressed in the posterior segments epidermis (Tomancak et al., 2007; Bantignies et al., 2011). This indicates that deleting SF1 leads to the overall decompaction of the Antennapedia complex in the posterior segments. This is consistent with the finding in our lab that SF1 serves as an interaction hub which selectively tethers multiple insulators across the Antennapedia complex, including an

insulator (LP2) near the *lab* gene and another insulator (AU1) near the Antp promoter (Li and Ma, data not shown, unpublished). Comparing to SF1 deletion, deleting SF2 only led to moderate increase of average physical distance between *Scr* and *Antp* in the posterior segments (Fig. 2.6B). In Li's unpublished 3C experiment, SF2 also strongly interacts with an insulator (AU1) near the *Antp* promoter region. Thus, knock-out of SF2 disrupts both SF1-SF2 formed loop and SF2-AU1 formed loop, which might contribute to the increased distance between *Scr* and *Antp* gene (Fig. 2.6B). *Scr* distal regulatory region also contains a major PRE, which plays an important role in polycomb mediated silencing of *Scr* gene (Pattatucci et al., 1991). It is worth noticing that in the SF1 knockout, disruption of SF1-SF2 loop not only led to the loss of insulation between *ftz* distal enhancers and *Scr* promoter, but also abolished the silencing effect from the PRE located in the distal regulatory region in the posterior segments epidermis (Fig. 2.4I-L,Y) (Pattatucci et al., 1991). However, in the SF2 knockout, the insulation between *ftz* distal enhancers and *Scr* promoter seems undisrupted (Fig. 2.4Q-X). This indicates that upon SF2 deletion, SF1 could still effectively insulate *ftz* distal enhancer from activating *Scr* promoter, possibly through looping with other insulators. The silencing effect from the PRE-element in *Scr* distal regulatory region was not significantly affected in SF2 knock-out (Fig 2.4Q-X), which suggests that the silencing effect on *Scr* promoter from this *Scr* distal PRE might depend on SF1, but does not require the SF1-SF2 loop. Instead, the SF1-R10 formed loop does not changed during embryonic development (Fig. 2.1C), which could bring this distal PRE close to the *Scr* promoter and silence the *Scr* gene in the posterior segments. Long-range interaction between the *Scr* proximal tethering element (PTE) near the *Scr* promoter and the R10 region is also known to facilitate communication between the *Scr* promoter and *Scr* distal regulatory elements (Calhoun and Levine, 2003).

In the late stage of the embryo (stage 17), the overall chromatin architecture is more compact than that in early stage embryos, which is consistent with the fact that Hox genes in Antennapedia complex are mostly silent in the epidermis in late embryonic stage (Tomancak et al., 2007). Deletion of SF1 or SF2 both led to a significant increase in the average physical distance between all three pairs of probes (Fig. 2.6C), which suggests that deletion of SF1 or SF2 both led to significant decompaction of the Antennapedia complex in the thoracic region. SF1 deletion also led to the overall decompaction of the Antennapedia complex in the posterior segments (Fig. 2.6C). SF2 deletion only caused an increase of the average physical distance between *Scr* and *Antp*, which suggests that the interaction between SF2 and *Antp* promoter in late stage embryos. The SF1-SF2 formed loop is known to fall apart around mid-embryogenesis when *ftz* gene expression fades out in the ectoderm tissue (Fig. 2.5E-F). However, deletion of SF1 or SF2 in the late stage of the embryo both led to global decompaction of the Antennapedia complex in thoracic and posterior segments (Fig. 2.6C). This suggests that SF1 and SF2 may still interact with other insulators in the Antennapedia complex to facilitate the 3D chromatin organization in the late embryonic stage.

In contrast to thoracic and posterior segments, SF1 or SF2 deletion had little effect on the overall compactness of Antennapedia complex in the head tissue (Fig. 2-6B,C). This indicates that SF1 and SF2 formed loops play little role in organizing 3D chromatin architecture in this tissue. It is possible that at different stages or tissues, a different mechanism plays a dominant role in organizing the local chromatin architecture.

We also noticed extensive variability in the 3D distance between the same pair of genomic loci among different cells in the same tissue and developmental stage. This is consistent with previous observations (Lanzuolo et al., 2007; Amano et al., 2009; Bantignies et al., 2011;

Giorgetti et al., 2014) and the model that 3D chromatin organization of a local region at the single cell level is probabilistic instead of being static (Stevens et al., 2017).

In summary, our data suggest that SF1 and SF2 insulators are participated in the dynamic tissue- and stage-specific chromatin organization of Antennapedia complex in *Drosophila.*

2.2.8 SF1 or SF2 knock-out flies have significantly lower overall-fitness and survival rate

Although homozygous SF1-knockout or SF2-knockout flies are both viable, we have noticed a significant degree of lethality at different stages of development, especially in stressful conditions. To quantitively describe the over-all fitness of the SF1 or SF2 knockout flies during development, we did a hatching and survival test in a Petri-dish chamber with apple juice plates and yeast paste. Wild type w^{1118} flies showed very consistently high hatching/survival rate throughout development with little variation among different repeats (Fig. 2.7B). In contrast, SF1 knock-out flies showed significantly lower hatching/survival rate at each developmental stage, and SF2 knock-out flies have an obvious drop in egg hatching rate (Fig. 2.7B). Further examination demonstrated that this egg hatching defect of SF2 knockout flies is due to egg fertilization failure and the percentage of fertilized eggs in SF2-knock-out flies gradually drops as they age (Fig. 2.7D). They become nearly infertile after two weeks. We further tested whether this egg fertilization defect is gender-specific by crossing the same aged male or female with wild type flies. The result showed that egg hatching rate of both w^{1118}♂x SF2KO♀ and SF2KO♂x w^{1118}♀ are intermediate compared to that of w^{1118} and homozygous SF2 knockout flies (Fig. 2.7C). SF2 knockout males seems to have a more severe defect in fertilizing eggs. This is coincided with the high expression of the *ftz* gene in adult testis (Chintapalli et al., 2007). It will be interesting to see whether the SF2 knockout affected *ftz* gene expression in the adult testis

and to figure out what caused the decrease of fertility in SF2 knockout flies. After hatching, the survival rate of SF2 knockout flies seems to be similar to that of the wild type flies.

In addition, both SF1 and SF2 knockout flies showed high variation of survival at later developmental stages (Fig. 2.7B), which are associated with more stressful living condition in the Petri-dish chamber. Based on SF1 and SF2's role in organizing the chromatin architecture in Antennapedia complex, it is likely that the drop in their fitness is due to mis-regulation of the spatial-temporal expression of genes in the Antennapedia complex, such as *Scr* and *ftz*.

2.3 Discussions

In the current study, we attempted to use *Drosophila Scr-ftz* gene locus in the Antennapedia complex as a model to demonstrate at the endogenous locus the functional importance of chromatin insulators in controlling gene activity and maintaining the regulatory independence between neighboring genes. We have identified a developmentally regulated chromatin loop anchored by two chromatin insulators, SF1 and SF2, flanking *ftz* gene unit in the *Drosophila* Antennapedia complex. We showed that SF1 and SF2 are bound by known insulator proteins and demarcate the boundary of active *ftz* gene domain in the early stage fly embryos. We found that the deletion of SF1 or SF2 caused genomic context-dependent mis-regulation of the *Scr* and *ftz* gene, which includes loss of insulation and abolishing long-range enhancer-promoter interactions. In addition, we provided evidence that deletion of SF1 or SF2 causes broader chromatin reorganization in the ~350kb Antennapedia complex. Flies with SF1 or SF2 deletion have significant lower overall fitness and survival rate.

2.3.1 Developmentally regulated chromatin loops anchored by two insulators

Although it has been reported both in *Drosophila* and mammals that chromatin insulators form chromatin loops through tethering in long-distance, developmentally regulated chromatin

loops anchored by insulators has been rare (Palstra et al., 2003). In the current study, together with unpublished data generated by a previous lab member, Mo Li, we identified a developmentally regulated chromatin loop in the in the *Scr-ftz* locus. Two insulators, SF1 and SF2, form long-range interactions during early embryonic stage when the flanked *ftz* gene is transcriptionally active. This loop isolates the *ftz* gene embedded in the *Scr* regulatory sequences. The SF1-SF2 loop corelates well with the timing of the *ftz* gene activity in the epidermis tissue. The *ftz* gene transcription is completely off when the germ-band starts to retract (around 7-8hrs of development) (Fig. 2.4A-F). This correlates well with the SF1-SF2 anchored chromatin loop, which is shown to fall apart at later stages (10-14hrs) when the *ftz* gene is completely off in the epidermis. However, known insulator protein factors, CTCF and CP190, seem to constantly bind to SF1 and SF2, as evidenced by the similar binding pattern in several *Drosophila* cell lines in which both *Scr* and *ftz* genes are silent (mod et al., 2010; Negre et al., 2010). Other developmentally regulated factors or post-transcriptional modifications that bridge the SF1-SF2 loop remain to be identified.

2.3.2 Regulatory role of SF1-SF2 formed chromatin loop in *Scr-ftz* region

By using CRISPR-Cas9 mediated deletion of SF1 or SF2 insulators in the current study, we provided evidence that chromatin loops tethered by SF1 and SF2 could address several challenges to proper gene regulation in the *Scr-ftz* gene region (Fig. 2.8).

i) Restriction of enhancer access. The *Scr* regulatory region contains a nested pair rule gene, *ftz,* and their enhancers are scattered on both sides of the *ftz* gene (Fig. 2.1A, Fig. 2.8). How are enhancer-promoter interactions specified for these two genes? Our lab has previously shown in a transgenic assay that SF1 insulator is capable of blocking *ftz* enhancers from activating *Scr* promoters (Majumder and Cai, 2003). However, it remains unclear how *ftz* is

insulated from the *Scr* regulatory elements located downstream of *ftz* and whether chromatin insulator at its genomic context works the same way as that demonstrated in a transgene assay. In the current work, we showed that SF1 and SF2 pair transiently to enclose the *ftz* gene domain, including all its enhancers. The timing and extent of this loop coincides with the timing of the *ftz* gene activity in the epidermis (Fig. 2.5A-F) and reduced access of *ftz* promoters to the outside *Scr* enhancers *in vivo* (Li et al., 2015b). In transgenic embryos, inserting SF1 5kb distal to NEE can augment the enhancer blocking activity by SF2 insulator (Fig. 2.2H,Q), supporting the notion that the SF1-SF2 formed loop restricts enhancer access. Pairing of SF1 and SF2 is orientation dependent to show enhancer bypass activity in our transgenic assay. This is consistent with the finding in *Drosophila* Fab8 insulator. dCTCF sites in *Drosophila* Fab8 is necessary for the enhancer blocking activity but not sufficient for enhancer bypass (Kyrchanova et al., 2016). Inversion of dCTCF sites only does not disrupt enhancer bypass activity of Fab8 at its endogenous locus (Kyrchanova et al., 2016). These observations indicate that the chromatin insulator is a complex element consisted of multiple functional units, including asymmetrically distributed components that account for its directional activities. We further examined the function of SF1 and SF2 in regulating local enhancer trafficking at their original genomic context. Knockout of SF1 leads to ectopic activation of *Scr* gene by *ftz* distal enhancers (Fig. 2.4I-L) and SF2 deletion causes ectopic expression of the *ftz* gene by the nearby *Scr* PS2 enhancer (Fig. 2.5P-T). Our study further indicates that the SF1-SF2 loop correlates with domains of enhancers access. Such delineation of enhancer domains by chromatin insulators is reminiscent of the Fab insulators subdividing the iab enhancers into parasegment-specific domains in the *Abd-B* regulatory region. In addition, our data together further suggests that a

chromatin insulator formed long-range loop could help establish physically separated chromatin domains at a sub-TAD scale.

ii) Facilitation of the *Scr* distal enhancers. The *Scr* regulatory sequences are interrupted by the *ftz* gene domain and multiple insulators, among which SF1 and SF2 contain strong and ubiquitous enhancer blocking activity (Fig. 2.2D-G,K,Q). These could pose impediments to the *Scr* distal enhancers communicating with *Scr* promoter. Previous studies have shown that pairing of chromatin insulators in cis may lead to reduction or cancellation of their enhancer-blocking activity due to changes in chromatin loop configurations (Cai and Shen, 2001; Muravyova et al., 2001; Melnikova et al., 2004; Savitskaya et al., 2006; Kyrchanova et al., 2007; Kyrchanova et al., 2008a; Kyrchanova et al., 2008b; Kyrchanova et al., 2016). Based on this, we have postulated that pairing between SF1 and SF2 would loop out the *ftz* gene domain, allow the *Scr* distal enhancers to bypass the block of both insulators, and facilitate the *Scr* distal enhancer by bringing the distal enhancers physically closer to the *Scr* promoter. In this study, we have shown that tandem arrangement of SF1 and SF2 indeed neutralizes the block of the distal enhancers in a transgenic setting (Fig. 2.2M). The enhancer-blocking cancellation effect by SF1 and SF2 pairing is orientation dependent. They only show enhancer-blocking cancellation activity when they are arranged together in the same relative orientations (SF1 forward + SF2 forward or SF2 forward + SF1 forward, Fig. 2.2M-N). This is consistent with the previous observation of orientation-dependent loop formation in both *Drosophila* and mammals (Kyrchanova et al., 2008a; Rao et al., 2014; Fujioka et al., 2016). In addition, knockout of the SF1 insulator not only caused the loss of insulation, but also abolished the activity of the PS2 enhancer located ~25kb away in the *Scr* distal regulatory region (Fig. 2.4J-L,Y). Surprisingly, deletion of SF2 did not abolish the activity of this long-range enhancer (Fig. 2.4Q-U). This might be due to the

redundancy from multiple chromatin loops in this *Scr-ftz* region. In addition to SF2, SF1 also strongly interacts with at least two chromatin insulators, R6 and R10, in the *Scr* distal regulatory region during early embryonic stages (Fig. 1C) (Li et al., 2015b). Upon SF2 deletion, SF1 may still interact with chromatin insulators in this region to facilitate the communication between *Scr* promoter and *Scr* distal enhancers. This is consistent with the changes of *Scr* transcription pattern in SF1 knockout and SF2 knockout embryos (Fig. 2.4-2.5). This hypothesis could be further validated through deletion of other SF1-thethering insulator elements in the *Scr* distal regulatory region in the future experiments. In addition to the SF1 anchored loops in the *Scr-ftz* region, the *Scr* proximal tethering element located right upstream of the *Scr* promoter has been reported to form long-range interaction with the *Scr* distal tethering element in the *Scr* distal regulatory region (overlaps with our R10 fragment, Fig. 2.1B-C) to facilitate the communication between the *Scr* promoter and the *Scr* distal regulatory elements (Calhoun et al., 2002; Calhoun and Levine, 2003).

iii) Separation of distinct chromatin domains. The *ftz* gene is expressed in many tissues where the *Scr* gene is repressed during early development. How does *ftz* remain active amid the polycomb-mediated repressive chromatin assembled in the surrounding *Scr* regions? Among the chromatin insulators identified in the *Scr-ftz* region, SF1 and SF2 precisely colocalize with the transient boundaries of repressive/active chromatin domains at the *ftz* gene locus during early embryonic stages (Fig. 2.3B). The active *ftz* gene domain is marked by low H3K9me3 and low H3K27me3 at the early embryonic stage, surrounded by repressive chromatin domains marked by high H3K9me3 and high H3K27me3 (Fig. 2.3B). Chromatin insulators are known to function as barriers to heterochromatin, so we propose that SF1 and SF2 may also function as barriers to restrict the spreading of polycomb-mediated repressive domain into the active *ftz* gene

domain. To address this question, another lab member has begun H3K27me3 ChIP-seq experiments in SF1-knockout and SF2-knockout embryos.

It has been reported in mammals that the disruption of the CTCF-binding site led to the spread of either H3K4me3 or H3K27ac marked active chromatin domains (Narendra et al., 2015; Willi et al., 2017), or H3K27me3 marked repressive chromatin domain (Luo et al., 2018). Interestingly, three recent studies showed that the acute protein factor depletion of CTCF (Nora et al., 2017), cohesin (Rao et al., 2017), or cohesin-loading complex member (WAPL) (Wutz et al., 2017) in mammalian cell lines disrupted chromatin insulator formed loops but did not trigger significant genome-wide H3K27me3 domains spreading. It has been previously shown in chicken *β-globin* locus that CTCF-dependent enhancer blocking activity and barrier activity are separable features of insulator elements (Recillas-Targa et al., 2002). The *Drosophila* Mcp insulator element also contains a separable insulator element and PRE-element that mediates long range interactions (Li et al., 2011). These observations suggest that barrier activity is not a universal feature of chromatin insulators. SF1 and SF2 might represent a class of composite insulators which consists of enhancer-blocking, loop-tethering, and heterochromatin-barrier function.

2.3.3 Broad influence on chromatin architecture in Antennapedia complex by SF1 and SF2

Previous 3C experiments in our lab revealed an organizational hub of chromatin loops anchored by SF1 and multiple other insulators located in the ~350kb *Drosophila* Antennapedia complex. In this study, we provided structural evidence that SF1 or SF2 deletion causes broad changes of chromatin architecture in the~350kb region of Antennapedia complex. The changes in chromatin organization upon SF1 or SF2 deletion differs in different tissues and developmental stages (Fig. 2.6B-C). In thoracic segments of early stage embryos, deleting either

SF1 or SF2 significantly decreased the average physical distance between the *Scr* and *lab* gene when compared to that in wild type. In contrast, deleting SF1 led to the overall decompaction of the Antennapedia complex in the posterior segments where the majority of the HOX genes in the Antennapedia complex are silent (Fig. 2.6B). Deleting SF2 only led to moderate increase of average physical distance between *Scr* and *Antp* in the posterior segments of early stage embryos (Fig. 2.6B). In late embryonic stage, the overall chromatin architecture is more compact than that in early stage embryos, which is consistent with the fact that Hox genes in the Antennapedia complex are mostly silent in the epidermis in late embryonic stage. Although the SF1-SF2 anchored chromatin loop falls apart in the late embryonic stages, deletion of SF1 or SF2 both led to different degrees of decompaction of the Antennapedia complex in the thoracic and posterior segments (Fig. 2.6C). This suggests that SF1 and SF2 may still interact with other insulators in the Antennapedia complex to facilitate the 3D chromatin organization. It will be interesting to see how SF1 or SF2 anchored loop interactions reprogram during development as well as how deletion of one insulator affects the "wiring" of the other insulators and broader gene activities in the Antennapedia complex the in the future experiments.

2.4 Material and methods

2.4.1 Enhancer-blocking assay with transgenic *Drosophila* embryos

Enhancer-blocking assay including spacer- and SF1- containing transgenic *Drosophila* lines were described previously (Cai and Levine, 1997; Cai and Shen, 2001; Belozerov et al., 2003). The original pWNHZ vector was modified by adding a phiC31 attB site at the NsiI site downstream of miniwhite reporter gene. The attB site was PCR cloned from the pBphi-yellow vector, which as a gift from Dr. Chris Rushlow. The Individual chromatin insulator elements were cloned by PCR (see table 2.1 for primers), TOPO-cloned, purified after further digestion,

and inserted into the AscI or Not1 sites between the NEE and H1 enhancers in pBWNHZ vector (see Fig. 2.2A). All the transgene constructs were inserted at the VK33 attP docking site via phiC-31 integrase mediated site-specific transformation. Microinjections were performed in the Cai lab. In situ hybridization with *lacZ* and *white* anti-sense RNA probes were performed as previously described (Cai et al., 2001). Whole mount in situ hybridization and visual assessment of reporter expression were performed according to an existing procedure (Cai and Levine, 1997; Cai et al., 2001; Belozerov et al., 2003). For each pBWNHZ transgene, about 100 to 150 embryos were scored double-blindly. Briefly, blastoderm stage embryos were scored for lacZ levels in the NEE domain (two horizontal strips in the lateral ventral ectoderm) against the H1 domain (vertical ring/strip in the head region) (Fig. 2.2B-P). Based on the intensity ratio of the lacZ in H1/NEE domains, embryos were ranked into five categories from complete block: NEE/H1 ≤ 2%, strong block: 2% < NEE/H1 ≤ 10%, medium block: 10% < NEE/H1 ≤ 40%, weak block: 40% < NEE/H1 ≤ 70%, no block: NEE/H1 > 70%.

2.4.2 3D DNA FISH

FISH analysis was performed following a published protocol (Dernburg, 2011). Loci of interest were cloned from genomic DNA by PCR into the PCRII-TOPO vector (Invitrogen, CA) with the following primers listed in table 2.1. The probes for *lab, Scr,* and *Antp* loci were made from two neighboring 5 kb DNA fragments that were cloned separately. For fluorescence labeling of probes, plasmid DNA were fragmented with six 4-base-recongnizing restriction enzymes to ~100 bp and labeled with Cy3-, Cy5- or Dig-UTP in terminal transferase reactions according to manufacturer's instructions (TdT labeling kit, Roche). For embryo FISH, stage 4 to stage 17 embryos were collected and fixed in glass vials containing 4ml of heptane and 4ml of fixative (4% freshly prepared paraformaldehyde in PBS). Fixation was carried out by shaking the

vial on an orbital shaker at 200rpm for 25 min. The hybridization and immunostaining steps were performed as described (Dernburg, 2011). To simultaneously label the nuclear envelop, a rabbit anti-lamin antibody (a gift from the Fisher Lab) was used in conjunction with Alexa Fluor® 647 goat anti-rabbit IgG. Confocal microscopy of the FISH and immunostaining experiments were done using a Zeiss LSM 710 Confocal Microscope. The three-dimensional distance between FISH signals were calculated using the distance measurement function of the Zeiss ZEN 2 (blue edition) software.

2.4.3 Knockout of SF1 and SF2 via CRISPR-Cas9

CRISPR-Cas9 mediated knockout was generated according to an established protocol by Fillip Port from CRISPR fly design (Port et al., 2014). Briefly, the protospacer sequences were introduced into the pCFD4 vector by PCR using roche Tgo high fidelity DNA polymerase (Roche, Cat. No. 0.186172001) (see primers sequences in table S2.1). Four different versions of SF1 gRNA pairs (SF1_gRNA1-1+SF1_gRNA2; SF1_gRNA1-1+SF1+300bp_gRNA2; SF1_gRNA1-2+SF1_gRNA2; SF1_gRNA1-2+SF1+300bp_gRNA2) and one version of SF2 gRNAs pair (SF2_gRNA1+SF2_gRNA2) were generated. pCFD4 vector was cut by BbsI and the larger backbone piece (6.4kb) was gel purified. The gel purified PCR insert and the BbsI-cut vector backbone were treated with T4 DNA polymerase for 30mins at room temp (22 ^{o}C), then heat inactivated at 65 ºC for 20min. The T4 DNA polymerase-treated insert and vector backbone were mixed, incubated, then transformed to DH5α competent cells using the ligase-independent cloning method (Li and Elledge, 2007). Recombinant plasmids were verified by Sanger sequencing by using the following sequencing primer: GACACAGCGCGTACGTCCTTCG. Transgenic flies carrying the pCFD4-gRNAs construct were made by microinjection in Cai lab. The pCFD4-gRNAs transgene was inserted at the attP2 site via phiC-31 mediated integration.

The phiC-31 integrase carrying X chromosome was replaced with y^1v^1 X chromosome. Homozygous actin5C-cas9 virgin females flies were mated with a homozygous transgenic fly that express gRNAs. F1 was crossed with a double balancer fly and germline carried knockout was screened in F2 (see screen primers in table S2.1). Each successful knock-out line was backcrossed to w^{1118} wild type for at least 8 generations, then kept as a homozygous line.

2.4.5 Hatching and survival assay

2-4 days old adult flies were fed with high protein yeast paste for three days in a collection bottle with apple juice plate (plastic food bottle with many holes for aeration). Each collection bottle contains about 200 female and 150 males. On day 4, eggs were collected on apple juice plates (2% agar with 20% Kroger concentrated 100% apple juice) with a thin layer of fresh yeast paste at 25 ºC. For egg hatching rate over aging test, SF2-knockout flies are fed on apple juice plate with yeast paste for various lengths of time before eggs were collected. Each collection is limited in one-hour windows to achieve maximal synchronization of the eggs. A total of 4 collections to be used for used for counting hatch/survival at 36hr (1^{st} instar), 60hr (2^{nd} instar), 96hr (3^{rd} instar), 168hr (pupae), and 264hr (adults) (Fig. 2.7A). Each plate is divided into 3 equal sized pieces as replicates and the egg number is counted until 100. Excessive eggs on each agar piece were removed. Then each piece of agar with 100 embryos were transferred onto a fresh apple juice plate and kept in a 100mm Petri-dish plate chamber in a 25 ºC ventilated incubator with natural day/night light cycles. 3 layers of 1cm x 2cm tissue paper with 100µl ddH_2O per day was add in the Petri-dish chamber to keep moister. Fresh 30% (w/v) yeast paste was added to the center of the apple juice plate on day 2 (25µl), day 3 (100µl), day 4 (200µl), day 5 (300µl), and day 6 (250µl). The hatching or survival rate at each stage was calculated by dividing the total number of properly developed larvae by the total egg number (100). Each of

this assay was repeated at least three times for each group. And the survival rate plot was generated by using GraphPad Prism 7 software.

2.4.6 Processing of ChIP-seq data

Raw ChIP-seq data in *Drosophila* S2 cells were obtained from the SRA database: Trl (GSE40646), BEAF-32 (GSE52962), CP190, CTCF, Mod(mdg4) and SuHw (GSE41354), Ibf1 and Ibf2 (GSE47559), ZIPIC and Pita (GSE54337). The sequencing reads were mapped to dm6 genome assembly with bowtie 2 with default settings, then the mapped bam files were normalized to 1x genome coverage using the bamCoverage tool in the deeptools tool set. The mapped bam files were visualized in EaSeq version 1.05. Chr3R:6842900-6898714 is visualized by using a filled tracks tool. The X-axis represents the 55778 bp surrounding the positions and was segmented in to 400 bins and smoothed for 1 bin.

Table 2.1 DNA primer sequences

Primer Name	**Sequences (5'->3')**
Asc1_SF1_F	AGGCGCGCCACGCTTGATTACCACGTGC
Asc1_SF1_R	AGGCGCGCCGGATTCCCCATCCTATACCC
NsiI_AttB_F	CCAATGCATGTCGACGATGTAGGTCACG
PstI_AttB_R	AACTGCAGTTCGGCTTGTCGACATGC
NotI_SF2_F	TGCGGCCGCCCATCCTCTTGTGAGGCTGG
NotI_SF2_R	TGCGGCCGCCTGATTGACGAATTGCGTGCG
D.wilistoni_SF1_F	TTGCGGCCGCAAAAACTCCTTTCGCTTTGGC
D.wilistoni_SF1_R	TTGCGGCCGCAATCGCAGCTCAAATTGAAGTC
D.wilistoni_SF2_F	TCAACAATTTTGGCATCTAGGCA
D.wilistoni_SF2_R	CTGATTGACGAATTGCGTGC
D.virilis_SF1_F	TTGCGGCCGCTGTTTGCCATTTCGTGAGCG
D.virilis_SF1_R	TTGCGGCCGCGCATATGGCTTTGGGCCTC
D.virilis_SF2_F	AGGCGCGCCGGTTATTTCAAAAATGAGAACGACATGC
D.virilis_SF2_R	AGGCGCGCCTGTCGTTCTGATTGACGAATTGC
D.mel_DS1_F	TTGCGGCCGCTGTAAAACCTGATCGTAATGTTTC
D.mel_DS1_R	TTGCGGCCGCGCTTAACCAGATATGTACGTATAAATG
D.mel_DP1_F	AGGCGCGCCCTCTTCCTTCTCCTCCTCCTC
D.mel_DP1_R	AGGCGCGCCGCAAACATCAATGTCAACTGTAACG
D.mel_DP1_ChIP-peak_F	AGGCGCGCCGGATTAACATAAACTGAATGCCTGGA
D.mel_DP1_ChIP-peak_R	AGGCGCGCCTTGTGCGGACATTGCAACG
D.virilis_DS1_F	TTGCGGCCGCCTCTCAGCTTTACTTGAGTGAGC
D.virilis_DS1_R	TTGCGGCCGCCCAAGGGCTGGCTATTTACAC
D.virilis_DP1_F	AGGCGCGCCGAGGGACTGGTACCATCTGC
D.virilis_DP1_R	AGGCGCGCCATTCCAGAAGCCAACAGTGACC
D.virilis_DP1-CTCF_F	AGGCGCGCCTGTAACTGCATTGGTAAGATCTCG
D.virilis_DP1-CTCF_R	AGGCGCGCCCCTTCAAGGTGTCCGTTCC
SF1_gRNA1-1	TATATAGGAAAGATATCCGGGTGAACTTCgctcttgaaatactgcgccaGTTTTAGAGCTAGAAATAGCAAG
SF1_gRNA1-2	TATATAGGAAAGATATCCGGGTGAACTTCGtgtccttctaacaggttctGTTTTAGAGCTAGAAATAGCAAG
SF1_gRNA2	ATTTTAACTTGCTATTTCTAGCTCTAAAACactccgaattcaactccaacGACGTTAAATTGAAAATAGGTC
SF1+300bp_gRNA2	ATTTTAACTTGCTATTTCTAGCTCTAAAACatgcttctatgcgacagctCGACGTTAAATTGAAAATAGGTC

SF2_gRNA1	TATATAGGAAAGATATCCGGGTGAACTTCGgaggatggtttg ggatcgcGTTTTAGAGCTAGAAATAGCAAG
SF2_gRNA2	ATTTTAACTTGCTATTTCTAGCTCTAAAACgatatttctgtttact tgcCGACGTTAAATTGAAAATAGGTC
SF1_del_screen_F	GTGGATTGTTGCGCATGTC
SF1_del_screen_R	CAACTGTTCAGTGATATTGTAGGAGTG
SF2_del_screen_F	ACAAATGTACACTTCCCAGTGTC
SF2_del_screen_R	GAGTTGCTCTCGCTGGATT
Lab_probe1st5k_F	GCAGTCGTATCGTGGTATCTACA
Lab_probe1st5k_R	TGGCCTTAGACTCTTTTGACCG
Lab_probe2nd5k_F	AAGGCTAGGCTCCTTGGGATT
Lab_probe2nd5k_R	GGACAATATGATGGACGTAAGCAGC
Scr_probe1st5k_F	GAGGACATCGCAAAACAGTCG
Scr_probe1st5k_R	GAACTTATGTAGCAATATTTGGTCTAAAGC
Scr_probe2nd5k_F	GCTTTAGACCAAATATTGCTACATAAGTTC
Scr_probe2nd5k_R	CTTGACATATGAATGAATCAGGTGG
Antp_probe1st5k_F	TGAAAAACAAGAGACCCGGCG
Antp_probe1st5k_R	AAACCGTTACGGCTGCCATTT
Antp_probe2nd5k_F	AATCTGTTTTAAATTATTTTCCCTTCGC
Antp_probe2nd5k_R	TGGCGACATGAGTCGGTTAATTA
Scr_sharedExonPprobe_F	CCATGGCAATCGAACCGAG
Scr_sharedExonProbe_R	CAGGTACTGTGAATGCCAATGG
*ftz*_RNAprobe_F	TTTGCTATATATGCAGGATCTG
*ftz*_RNAprobe_R	TGTGCAACTCATCGCGTCG

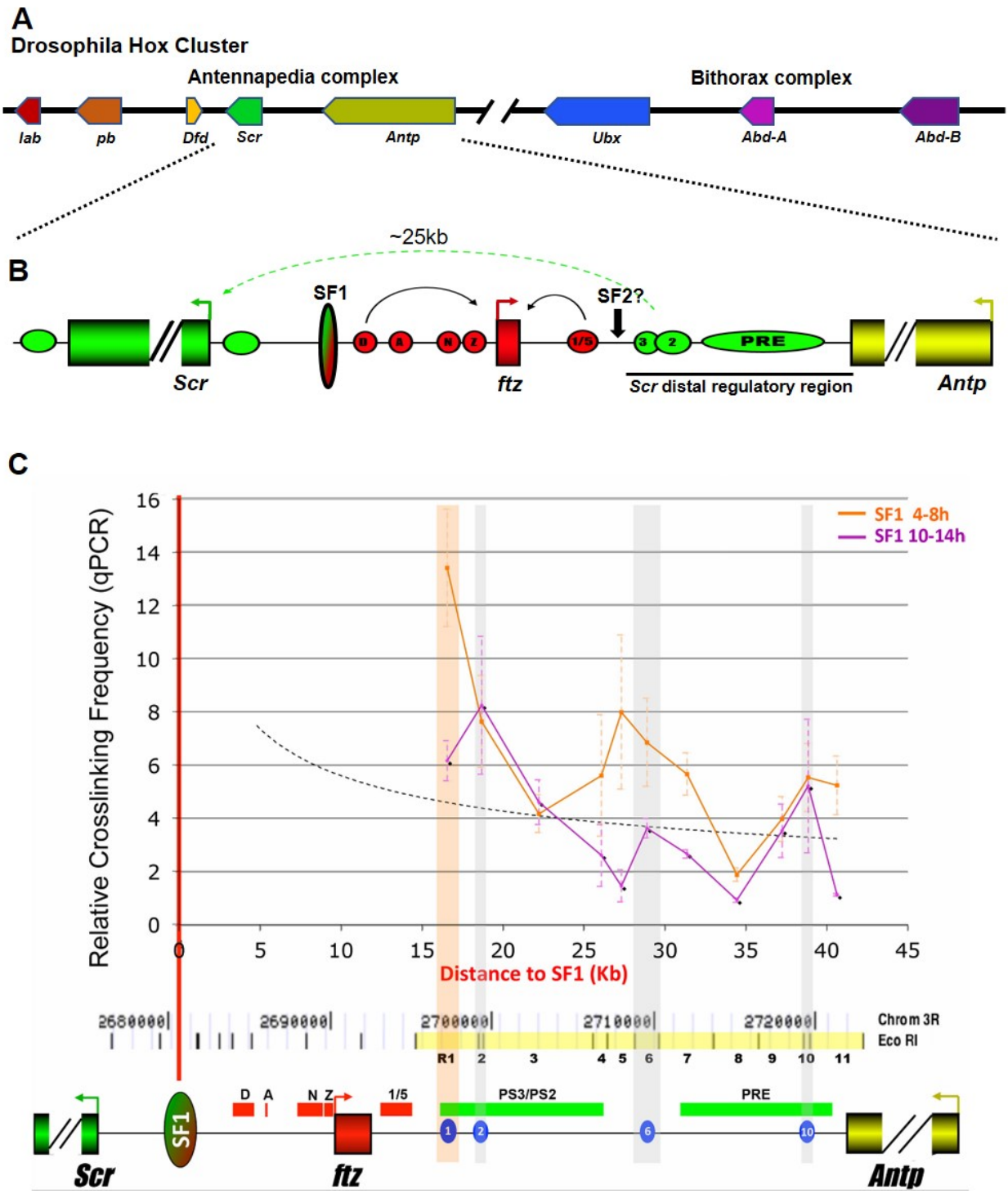
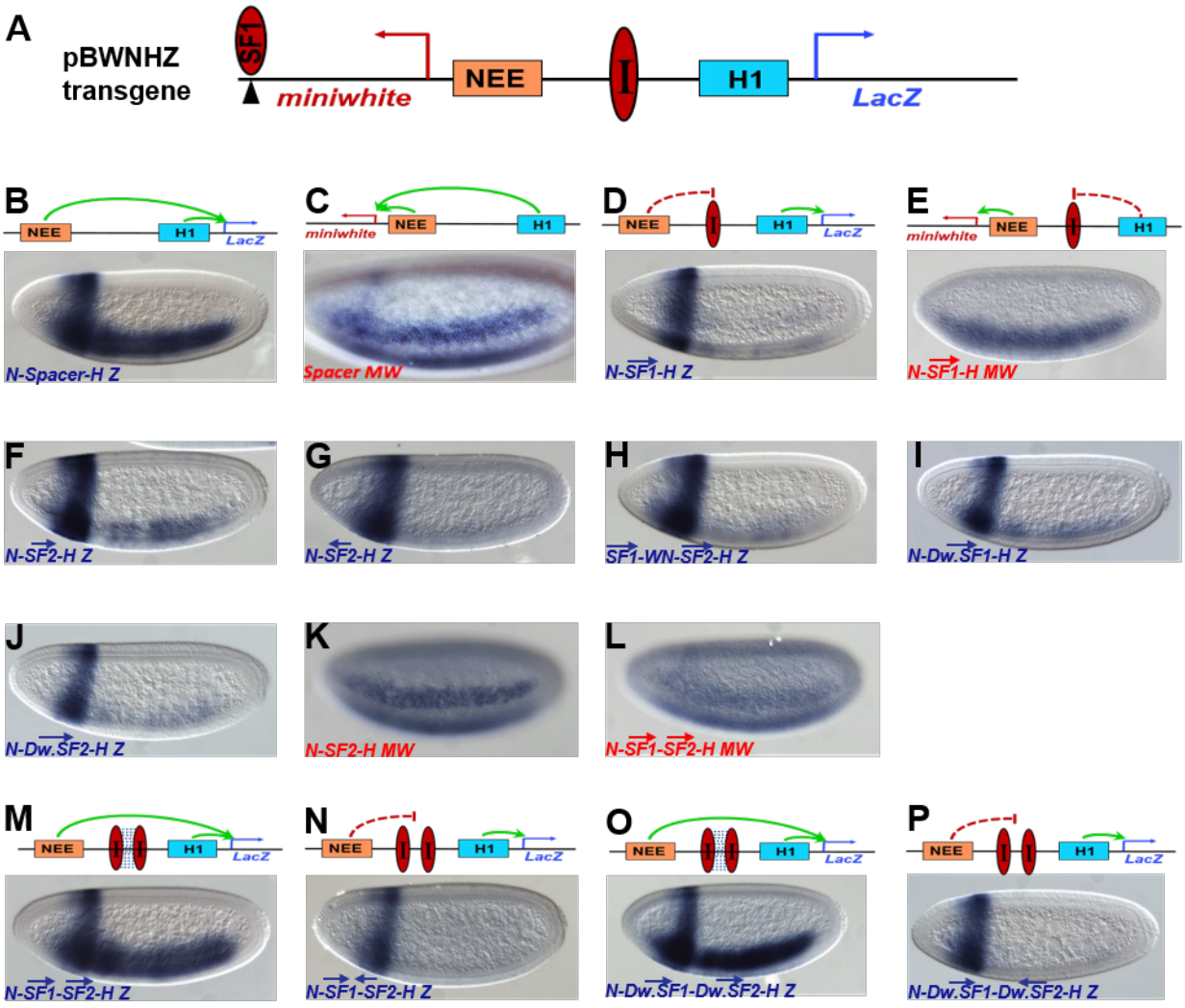

A
Drosophila Hox Cluster
Antennapedia complex
Bithorax complex
lab
pb
Dfd
Scr
Antp
Ubx
Abd-A
Abd-B
~25kb
B
SF1
SF2?
D
A
N
Z
1/5
3
2
PRE
Scr
ftz
Scr distal regulatory region
Antp
C
Relative Crosslinking Frequency (qPCR)
SF1 4-8h
SF1 10-14h
Distance to SF1 (Kb)
2680000
2690000
2700000
2710000
2720000
Chrom 3R
Eco RI
R1
PS3/PS2
PRE
SF1
Scr
ftz
Antp

Figure 2.1 Identification of SF1-tethering elements in the *Drosophila ftz-Antp* interval. (A) Schematic representation of *Drosophila* Hox cluster, which consists of the Antennapedia complex and the Bithorax complex that are about 10Mb apart. Hox genes are represented as filled pentagon arrows. (B) Schematic representation of zoomed-in *Scr-Antp* region of the Antennapedia complex. Filled boxes represents the *Scr* (green), *ftz* (red), and *Antp* (yellow) genes. Horizontal arrow heads and horizontal ovals represent promoters and enhancers, respectively. Red enhancers: D, *ftz* distal enhancer; A, *ftz* AE1 enhancer; N, *ftz* neurogenic enhancer; Z, *ftz* zebra enhancer; 1/5, *ftz* strip 1/5 enhancer. Labeled green enhancers: 3, *Scr* PS3 enhancer or *Scr* T1 enhancer; 2, *Scr* PS2 enhancer; PRE, *Scr* PRE. Vertical green/red hybrid oval represents SF1 insulator. Curved black arrow represents the *ftz* enhancer-promoter interaction. Dashed green arrow represents long-range *Scr* enhancer-promoter interaction. Solid black line under the green ovals marks the *Scr* distal regulatory region. The vertical black arrow marks the expected position of SF2 insulator. (C) SF1 insulator interacts with multiple DNA elements in the *ftz-Antp* interval. Chromosomal conformation capture (3C) frequencies between SF1 and R1 to R11 fragments were quantitated by qPCR and plotted over distance to SF1 insulator. Dashed curve, a distance-frequency power trend line (Li et al., 2015b). Capture frequencies from the early stage (4-8hrs) and late stage (10-14hrs) are represented in orange and purple color, respectively. Vertical shaded bars indicate regions of interest (orange, R1; gray, R2, R6, and R10). At the bottom, a genomic map of the Scr-Antp region, drawn to scale. EcoR1 sites are marked in yellow and labeled numerically from R1 to 11. SF1-tethering elements are shown as blue ovals.


A
pBWNHZ
transgene
SF1
miniwhite
NEE
I
H1
LacZ
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
N-Spacer-H Z
Spacer MW
N-SF1-H Z
N-SF1-H MW
N-SF2-H Z
N-SF2-H Z
SF1-WN-SF2-H Z
N-Dw.SF1-H Z
N-Dw.SF2-H Z
N-SF2-H MW
N-SF1-SF2-H MW
N-SF1-SF2-H Z
N-SF1-SF2-H Z
N-Dw.SF1-Dw.SF2-H Z
N-Dw.SF1-Dw.SF2-H Z

Q

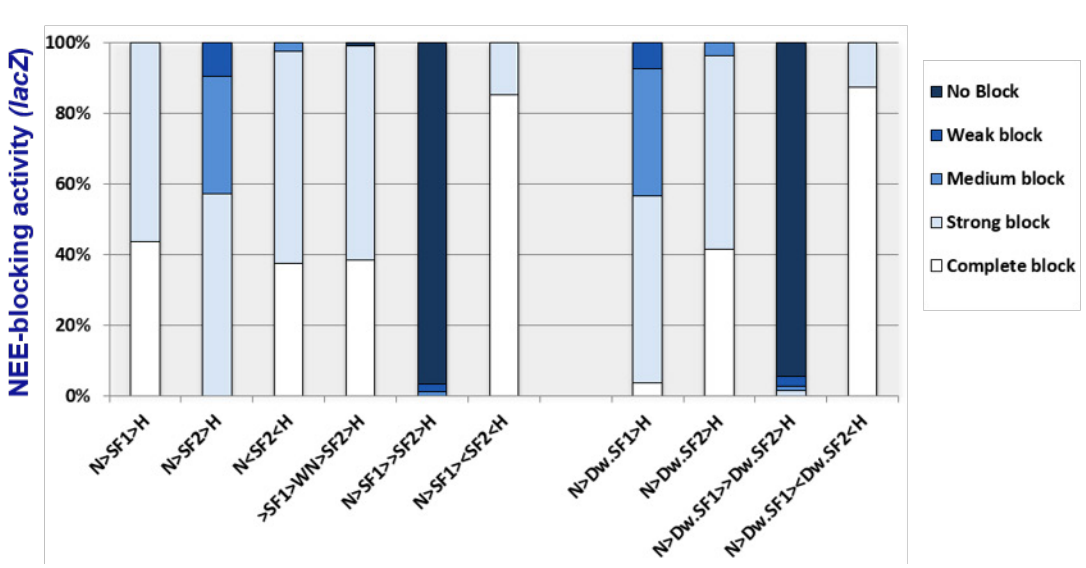
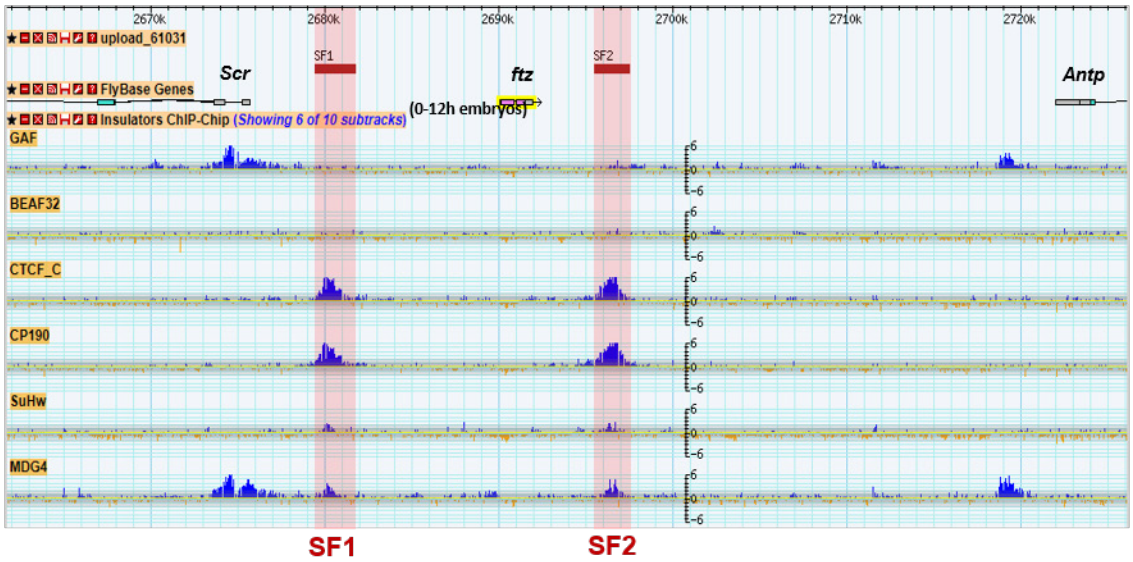

NEE-blocking activity (lacZ)
100%
80%
60%
40%
20%
0%
N>SF1>H
N>SF2>H
N<SF2<H
>SF1>WN>SF2>H
N>SF1>>SF2>H
N>SF1><SF2<H
N>Dw.SF1>H
N>Dw.SF2>H
N>Dw.SF1>>Dw.SF2>H
N>Dw.SF1><Dw.SF2<H
No Block
Weak block
Medium block
Strong block
Complete block

Figure 2.2 SF1 and SF2 show strong enhancer-blocking activity in transgenic *Drosophila*. (A) Diagram of the pBWNHZ transgene vector containing divergently transcribed *miniwhite* and *lacZ* reporters flanking the NEE and H1 enhancers. Red oval with label "I", insulator or spacer control between NEE and H1; Red oval with label "SF1", SF1 insulator inserted in the SF1-WN-SF2-H transgene (see materials and methods for details). (B-E, M-P) Representative enhancer-promoter communication of the tested reporter gene is represented with a schematic diagram. The green curve arrow represents active enhancer-promoter interaction, dashed red line represents restricted enhancer access. Enhancers, insulators, and reporter gene are represented the same as in (A). The blue dashed lines between the two insulator ovals of the diagram in (M) and (N) represent interaction between insulators). (B to P) Representative images of transgenic embryos after whole-mount *in situ* hybridization with the anti-*lacZ* (B, D, F-J, M-P) or the anti-*white* (C, E, K, L,) RNA probe. The transgene contained in these embryos are labeled at the bottom left of each photo, with probe indicated (Z, lacZ; MW, m*iniwhite*). Embryos are shown in sagittal views with anterior to the left and dorsal up. The arrow on top of the transgene label represents the orientation of the insulator elements under the arrow. The orientation is determined by comparing to the orientation of the insulator in the genome of *Drosophila melanogaster.* (Q) Quantitation of NEE-blocking in the whole neuroectoderm in transgenic embryos stained with the anti-*lacZ* probe (see Methods and Materials for details).

A


2670k
2680k
2690k
2700k
2710k
2720k
upload_61031
SF1
SF2
Scr
ftz
Antp
FlyBase Genes
(0-12h embryos)
Insulators ChIP-Chip (Showing 6 of 10 subtracks)
GAF
BEAF32
CTCF_C
CP190
SuHw
MDG4
SF1
SF2

B

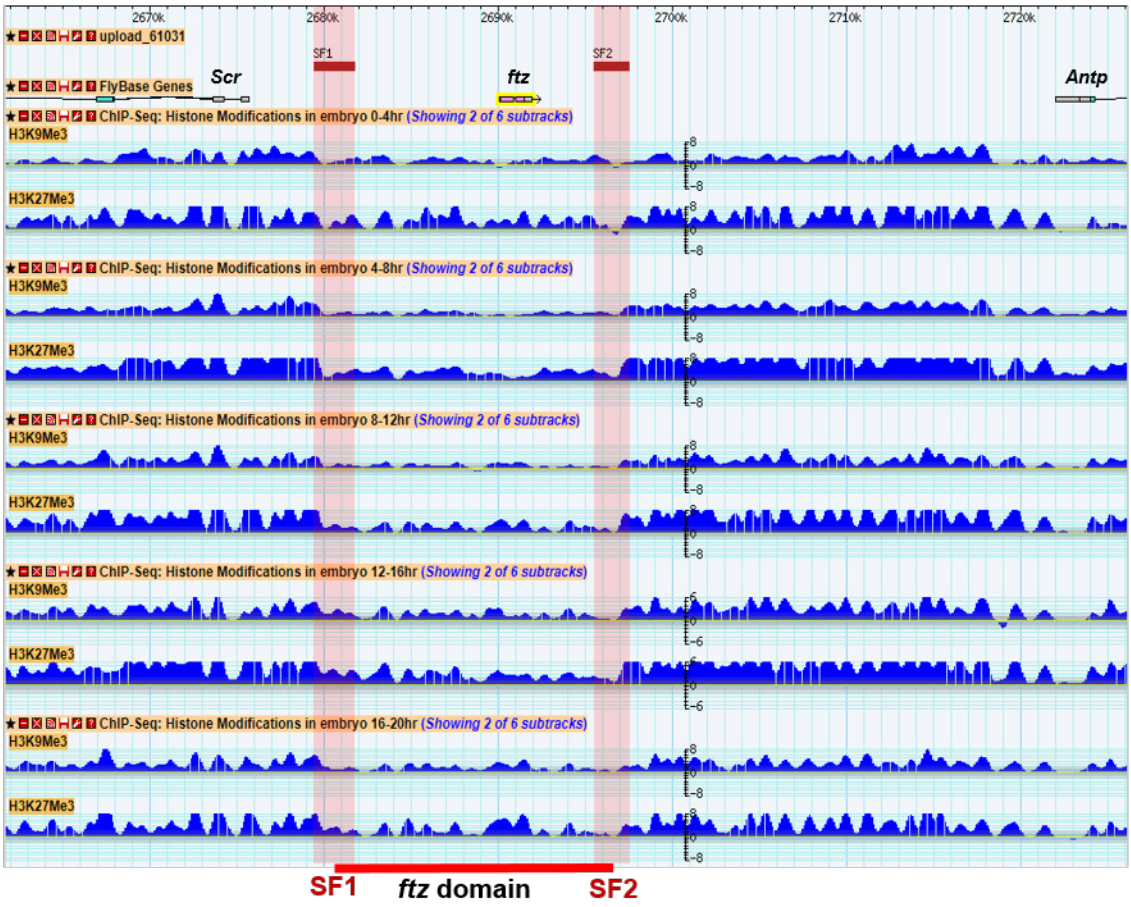

2670k
2680k
2690k
2700k
2710k
2720k
upload_61031
SF1
SF2
Scr
ftz
Antp
FlyBase Genes
ChIP-Seq: Histone Modifications in embryo 0-4hr (Showing 2 of 6 subtracks)
H3K9Me3
H3K27Me3
ChIP-Seq: Histone Modifications in embryo 4-8hr (Showing 2 of 6 subtracks)
H3K9Me3
H3K27Me3
ChIP-Seq: Histone Modifications in embryo 8-12hr (Showing 2 of 6 subtracks)
H3K9Me3
H3K27Me3
ChIP-Seq: Histone Modifications in embryo 12-16hr (Showing 2 of 6 subtracks)
H3K9Me3
H3K27Me3
ChIP-Seq: Histone Modifications in embryo 16-20hr (Showing 2 of 6 subtracks)
H3K9Me3
H3K27Me3
SF1
ftz domain
SF2

Figure 2.3 SF1 and SF2 insulators demarcate the boundaries active *ftz* domain. (A) Screenshot of the binding profile of the *Drosophila* classic insulator protein factors in *Scr-Antp* interval from modENCODE GBroswer (http://gbrowse.modencode.org/fgb2/gbrowse/fly/) (Negre et al., 2010). (B) Screenshot of the repressive histone modification ChIP-seq profile in *Scr-Antp* region throughout differnet developmental stages (http://gbrowse.modencode.org/fgb2/gbrowse/fly/). Vertical shaded red boxes in (A) and (B) mark the SF1 and SF2 insulators, rescpectively.

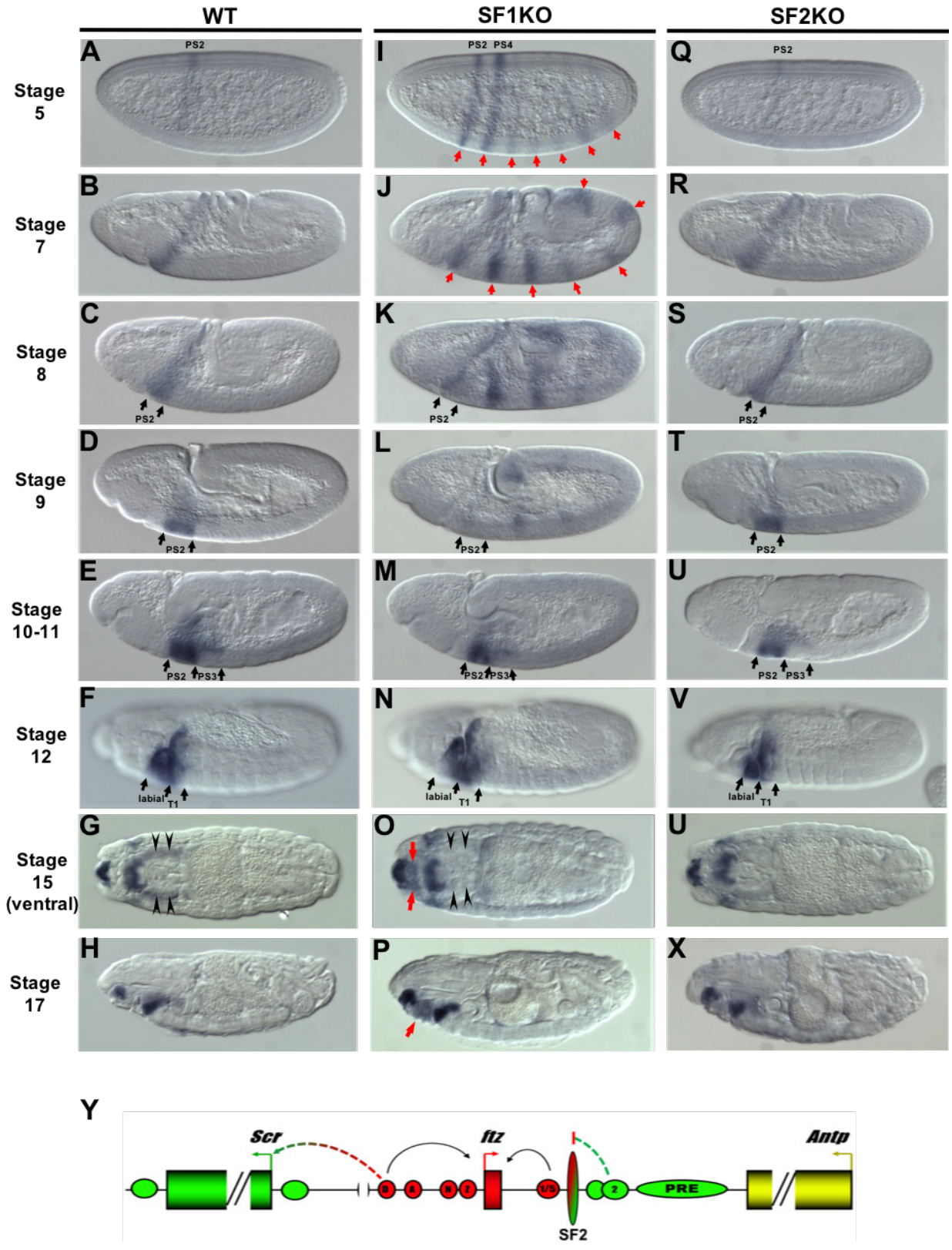

WT
SF1KO
SF2KO
Stage 5
Stage 7
Stage 8
Stage 9
Stage 10-11
Stage 12
Stage 15 (ventral)
Stage 17
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
U
X
PS2
PS4
PS3
labial
T1
Y
Scr
ftz
Antp
PRE
SF2

Figure 2.4 Knock-out SF1 causes misregulation of the *Scr* gene in *ftz* gene pattern. (A-X) Representative embryo images after RNA in situ hybrydization with the *Scr* probe. Black arrows mark the boundaries of parasegment 2 and 3. PS2 (parasegment 2) and PS3 (parasegment 3) are marked between arrows. Red arrows point to the ectopic expression of *Scr* gene. Black arrowheads in (G) and (O) mark the position of the salivary gland. Embryos are shown in saggital view with anterior to the left and dorsal up except in (G,O,U, which are show in transverse view, with ventral side facing the reader). (A-H) Representative wild type embryos in differnent stages (labeled on the left side). (I-P) Representative SF1-knockout embryos in different stages. (Q-X) Representative SF2-knockout embryos in different stages. (Y) Summarized model on *Scr* misregulatioin after SF1 deletion. Diagram is drawn similarly to Fig2.1B. The red-green hybrid-color curved arrow represents the ectopic action of *ftz* distal enhancers on *Scr* promoter. And the green dash curve with a red block represents the blocked communication between *Scr* PS2 enhancer and *Scr* promoter.

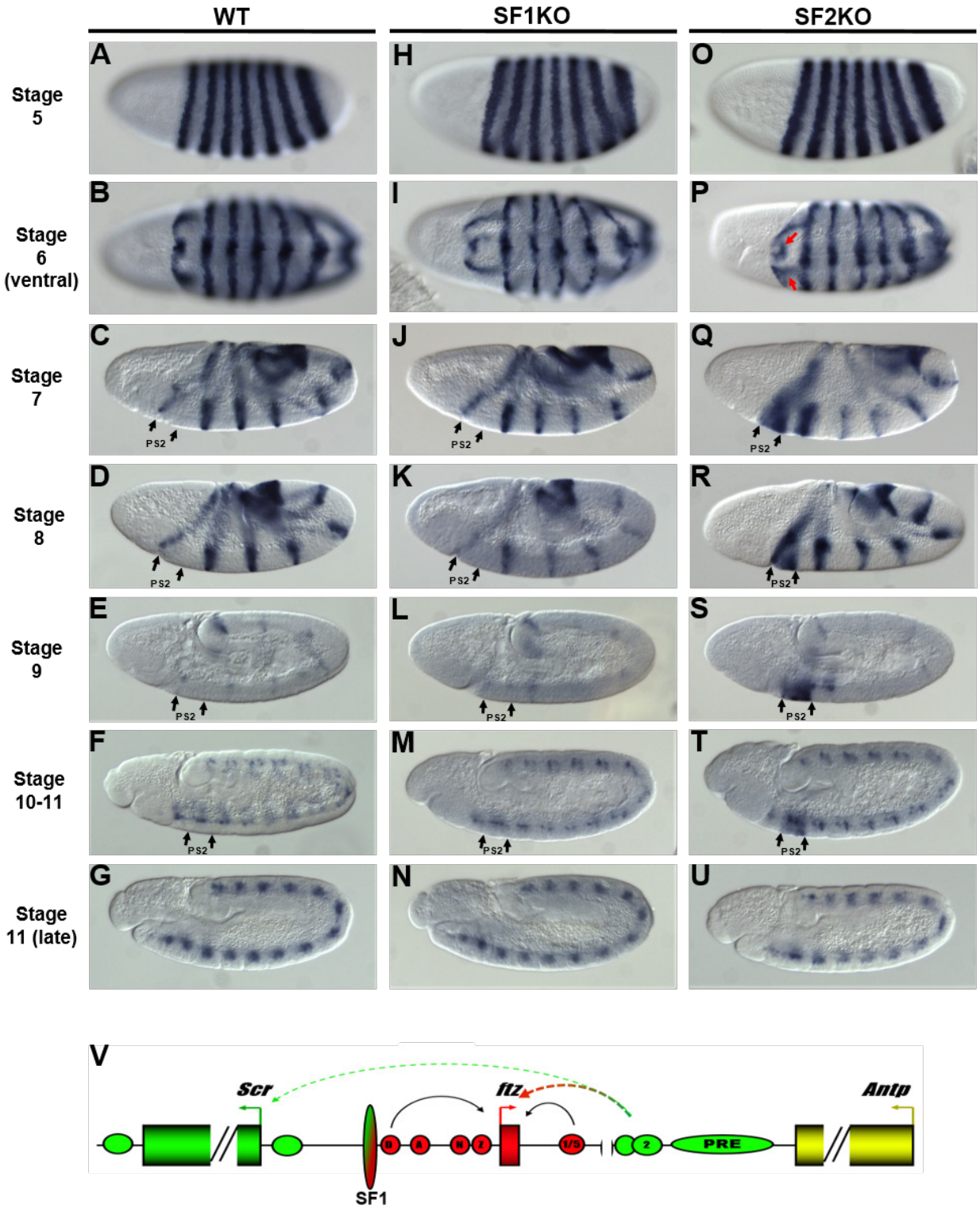

WT
SF1KO
SF2KO
Stage 5
Stage 6 (ventral)
Stage 7
Stage 8
Stage 9
Stage 10-11
Stage 11 (late)
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
PS2
V
Scr
ftz
Antp
SF1
PRE

Figure 2.5 Knock-out SF2 causes ectopic expression of *ftz* gene by neighboring *Scr* enhancer. (A-U) Representative embryo images after RNA in situ hybrydization with *ftz* probe. Black arrows mark the boundaries of parasegment 2 and 3. Parasegment 2 (PS2) is marked between black arrows. Red arrows point to the ectopic expression of the *Scr* gene. Embryos are shown in saggital view with anterior to the left and dorsal up except in (B, I, P, which are show in transverse view, with ventral side facing the reader). (A-G) Representative wild type embryos in differnent stages (labeled on the left side). (H-N) Representative SF1-knockout embryos in different stages. (O-U) Representative SF2-knockout embryos in different stages. (V) Summarized model on *ftz* gene misregulatioin after SF2 deletion. Diagram is drawn similarly to Fig2.1B. The green-red hybrid-color curved arrow represents the ectopic action of *Scr* PS2 enhancer on the neighoring *ftz* promoter.

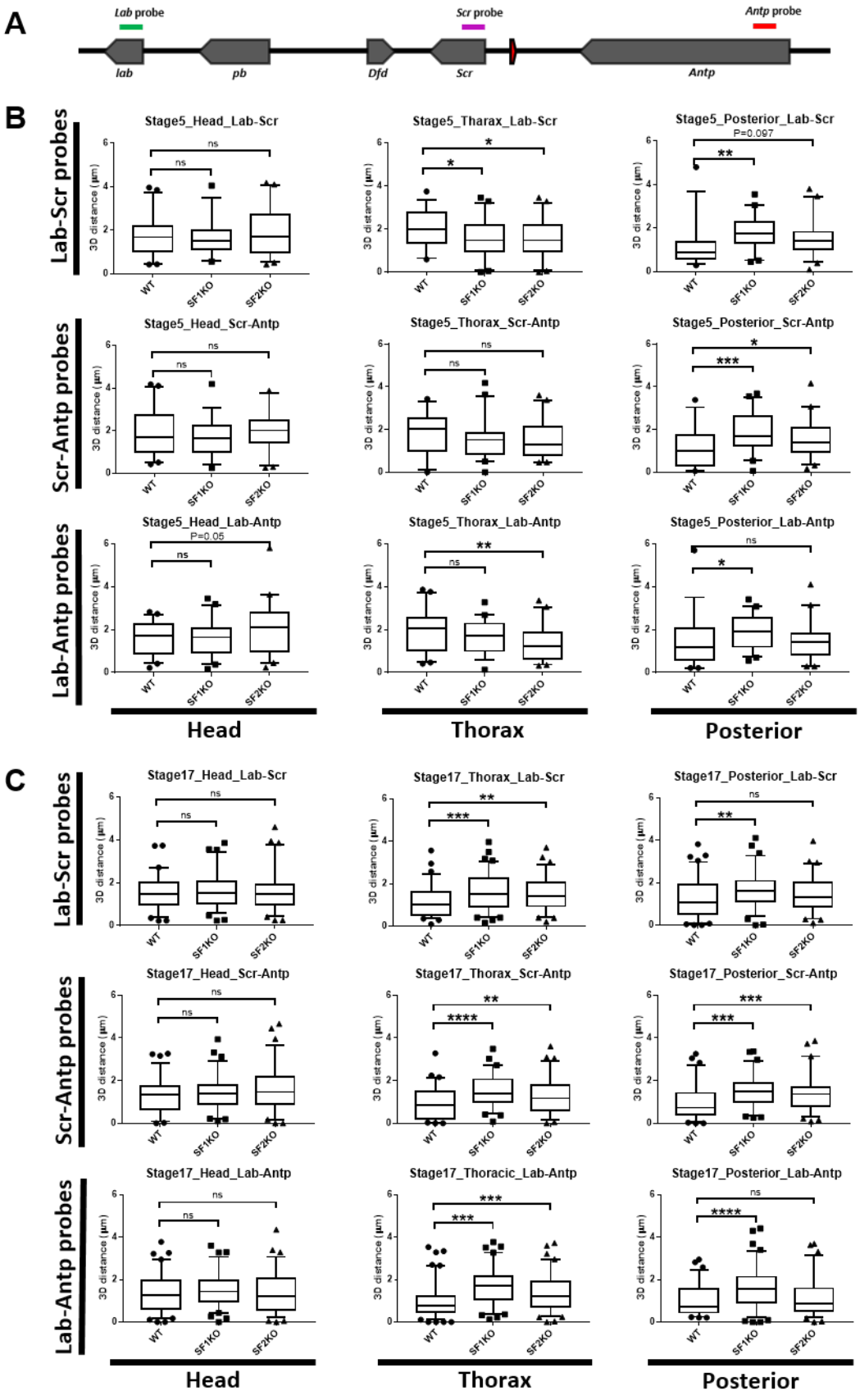

A
Lab probe
Scr probe
Antp probe
lab
pb
Dfd
Scr
Antp
B
Lab-Scr probes
Scr-Antp probes
Lab-Antp probes
Stage5_Head_Lab-Scr
Stage5_Tharax_Lab-Scr
Stage5_Posterior_Lab-Scr
P=0.097
Stage5_Head_Scr-Antp
Stage5_Thorax_Scr-Antp
Stage5_Posterior_Scr-Antp
Stage5_Head_Lab-Antp
P=0.05
Stage5_Thorax_Lab-Antp
Stage5_Posterior_Lab-Antp
3D distance (µm)
WT
SF1KO
SF2KO
Head
Thorax
Posterior
C
Stage17_Head_Lab-Scr
Stage17_Thorax_Lab-Scr
Stage17_Posterior_Lab-Scr
Stage17_Head_Scr-Antp
Stage17_Thorax_Scr-Antp
Stage17_Posterior_Scr-Antp
Stage17_Head_Lab-Antp
Stage17_Thoracic_Lab-Antp
Stage17_Posterior_Lab-Antp

Figure 2.6 Chromatin architectural change after SF1 or SF2 knockout as revealed by 3D DNA FISH. (A) A schematic diagram represents the *Drosophila* Antennapedia complex. Hox genes are represented as filled grey pentagon arrows. The *ftz* gene is represented by filled red pentagon arrow. Genomic target locations of *lab* probe (green), *Scr* probe (purple), and *Antp* probe (red) are marked with horizontal bars on top of the corresponding Hox gene. (B) Represent measurement from stage 5 embryos. (C) Represent measurement from stage 17 embryos. (B-C) Box plot of 3D distance between pairs of probes. Measurements are presented by tissues (marked by horizontal black bar) and probe-pair (marked by vertical black bar). P-values (ns, $p>0.05$; *, $p\leq0.05$; **, $p\leq0.01$; ***, $p\leq0.001$; ****, $p\leq0.0001$) of unpaired t-test between WT and knockout groups are indicated on top. Each group has at least 40 nuclei measured.

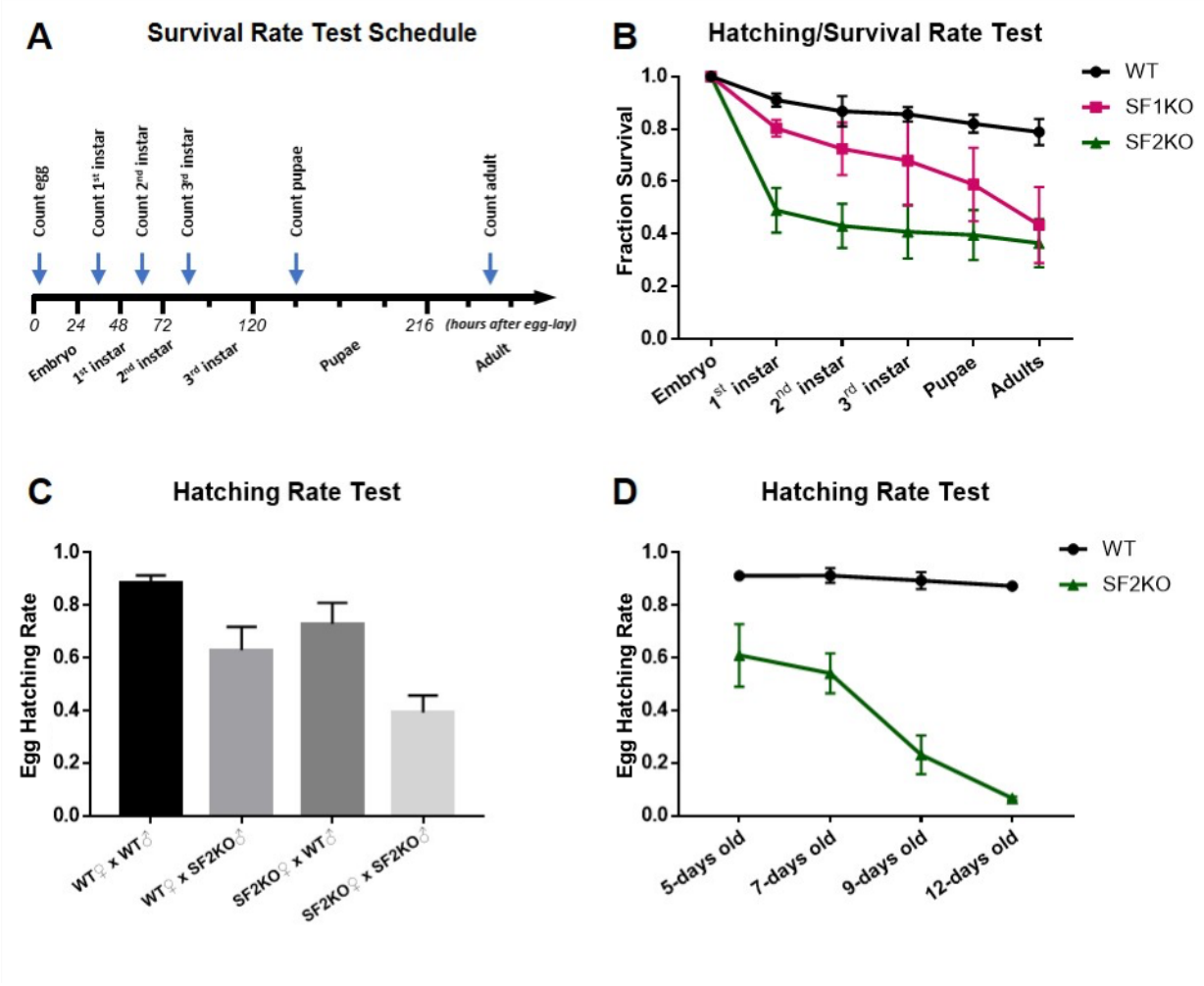

A
Survival Rate Test Schedule
Count egg
Count 1st instar
Count 2nd instar
Count 3rd instar
Count pupae
Count adult
0
24
48
72
120
216
(hours after egg-lay)
Embryo
1st instar
2nd instar
3rd instar
Pupae
Adult
B
Hatching/Survival Rate Test
Fraction Survival
1.0
0.8
0.6
0.4
0.2
0.0
WT
SF1KO
SF2KO
Embryo
1st instar
2nd instar
3rd instar
Pupae
Adults
C
Hatching Rate Test
Egg Hatching Rate
1.0
0.8
0.6
0.4
0.2
0.0
WT♀ x WT♂
WT♀ x SF2KO♂
SF2KO♀ x WT♂
SF2KO♀ x SF2KO♂
D
Hatching Rate Test
Egg Hatching Rate
1.0
0.8
0.6
0.4
0.2
0.0
WT
SF2KO
5-days old
7-days old
9-days old
12-days old

Figure 2.7 Knockout of SF1 or SF2 reduces overall fitness during *Drosophila* development. (A) Represents the hatching/survival test schedule. The horizontal bar with time represents the key transition point between stages during *Drosophila* development. Blue arrows point to the time for hatching or survival count at each stage. (B) Hatching/survival rate test in wild type (WT), SF1-knockout (SF1KO), and SF2-knockout (SF2KO) flies. (C) Egg hatching rate test in different male and virgin female crosses (See materials and methods for details.). (D) Egg hatching rate test over age of SF2-knockout flies and w^{1118} wild type. Error bars represent the standard deviation.

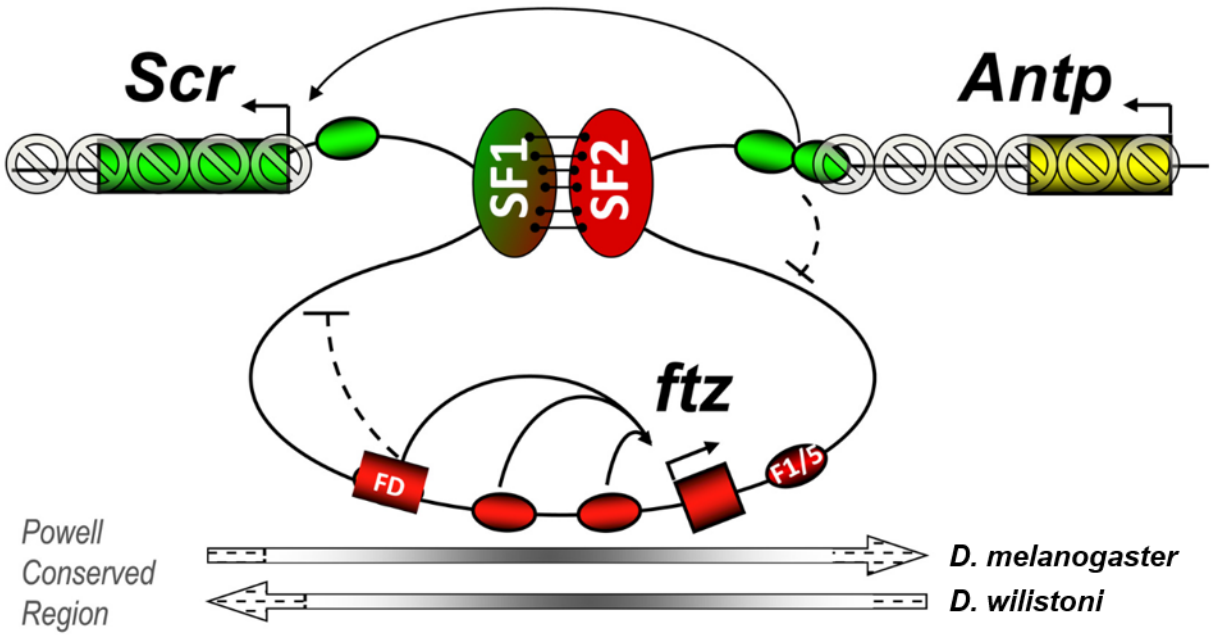
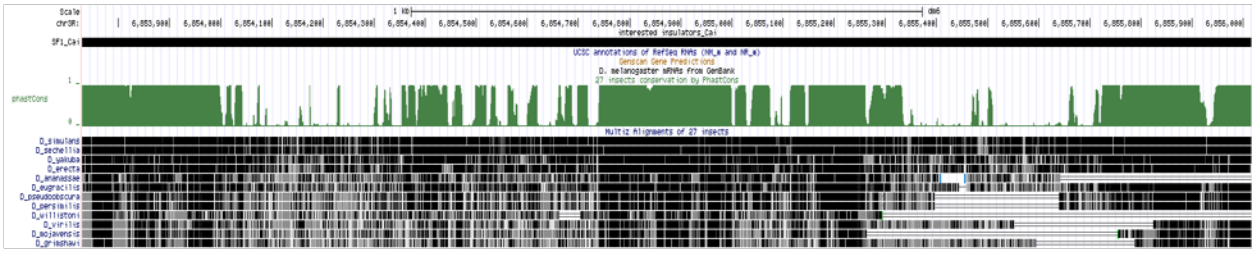

Scr
Antp
SF1
SF2
ftz
FD
F1/5
Powell
Conserved
Region
D. melanogaster
D. wilistoni

Figure 2.8 Model showing SF1-SF2 loop remodels enhancer traffic and modulate chromatin structure in *Scr-Antp* regioin. SF1-SF2 interaction forms a chromatin loop that includes that entire *ftz* gene unit. The loop restricts the access between enhancers and promoters of neighoring genes (blocked dash arrows). It facilitates the *Scr* distal enhancer PS2 to bypass both SF1 and SF2 and act on the *Scr* promoter over long-distance. This loop may also insulate *ftz* gene unit from the encroachment of polycomb mediated repressive chromatin assembled on *Scr* and *Antp* genes. Genes, promoters, enhancers, promoter-enhancer communication and insulators are represented similarly as in Fig. 2.1. The grey forbiden-sign represent the polycomb mediated silent chromatin. The two horizontal arrows below represent the evolutionarily conserved but flipped *ftz* gene domain in *Drosophila melanogaster* and *Drosophila willistoni* (Aka Powell Conserved Region)


6860k
6870k
6880k
6890k
6900k
Restriction Sites
EcoRI restriction site
EcoRI
R4 R5 R6
R9-10
Gene Span
lncRNA:CR45900
ftz
lncRNA:CR44931
lncRNA:CR45559
lncRNA:CR44932
Antp
0-2hr
2-4hr
4-6hr
6-8hr
8-10hr
10-12hr
12-14hr
14-16hr
16-18hr
18-20hr
20-22hr
22-24hr

Figure S2.1 RNA-seq profile during different embryonic stages in the *ftz-Antp* interval. This is a snapshot of flybase GBrowser with EcoR1 restriction site track, gene span track and RNA-seq track of whole fly embryos at different developmental stages (http://flybase.org/cgi-bin/gbrowse2/dmel/)(Graveley et al., 2011). The EcoR1 fragments, R4-R6 and R9-R10, that overlap with lncRNAs are marked in the restriction sites track. The bottom tack displays the $\log_2$ RNA expression data (citation).

A



B

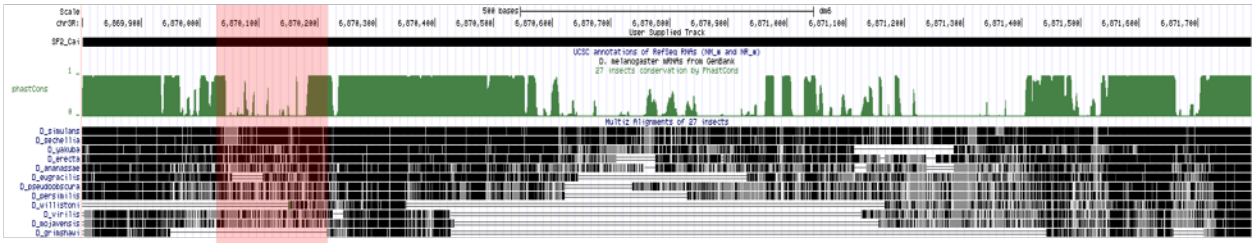


Figure S2.2. SF1 and SF2 contains conserved sequence blocks in *Drosophila* species. Both (A) and (B) are captured images of UCSC Genome Browser displaying user defined insulator track, phastCons 27 insects' conservation track (Green), and multiz Alignment track for 12 *Drosophila* species. Black blocks in the bottom multiz alignment represents conserved sequences (https://genome.ucsc.edu). Single or double line in the multiz alignment track represent alignment gaps due to insertion or deletion (Kent et al., 2002). (A) Represents the SF1 region. (B) Represent the SF2 region. Red shaded box marks the mapped rearrangement breakpoints region.

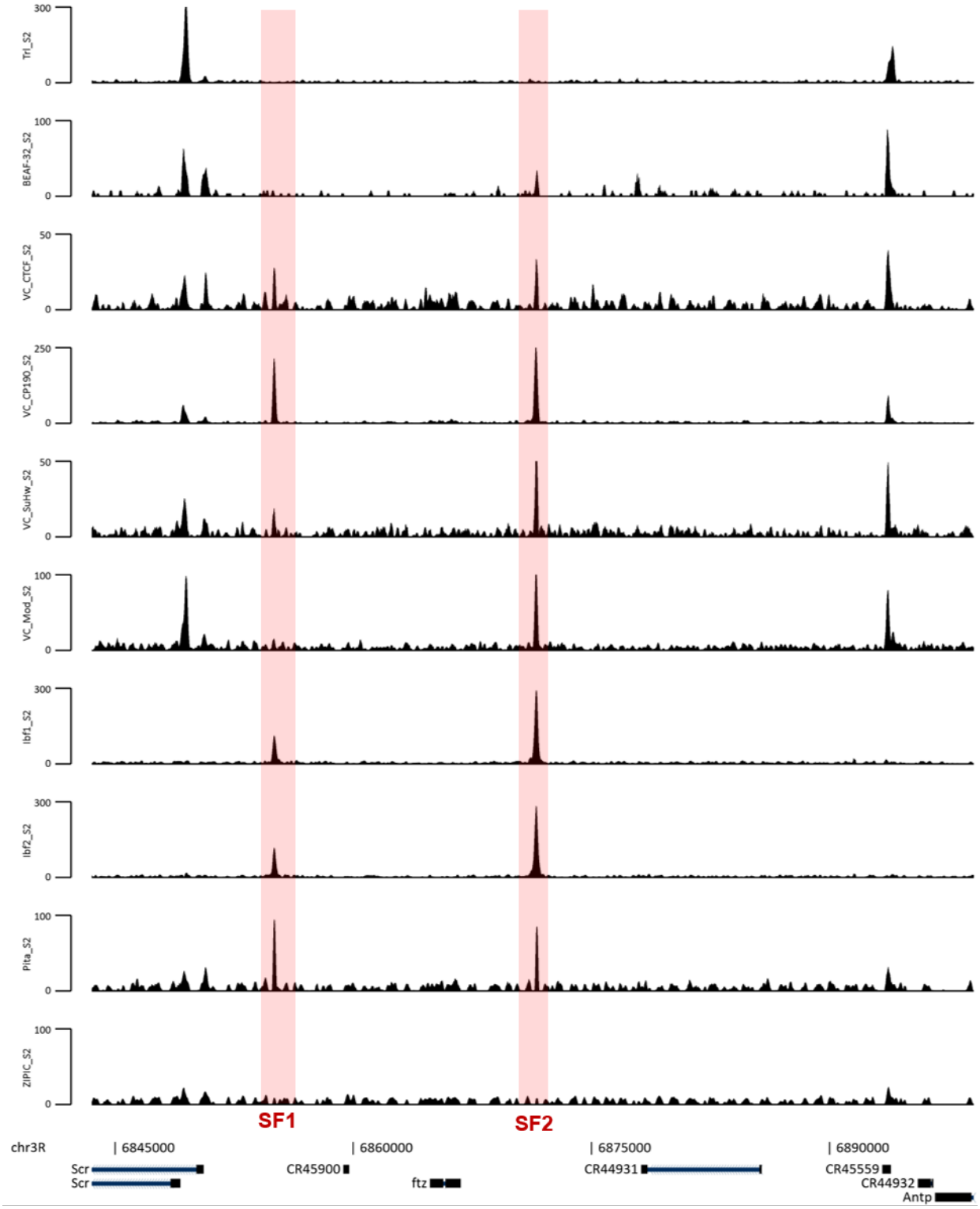

Trl_S2
BEAF-32_S2
VC_CTCF_S2
VC_CP190_S2
VC_SuHw_S2
VC_Mod_S2
Ibf1_S2
Ibf2_S2
Pita_S2
ZIPIC_S2
SF1
SF2
chr3R
6845000
6860000
6875000
6890000
Scr
Scr
CR45900
ftz
CR44931
CR45559
CR44932
Antp

Figure S2.3 Insulator protein factors bind to SF1 and SF2 in S2 cells. Tracks of Trl(GAF), BEAF-32, CTCF, CP190, SuHw, Mod(mdg4), Ibf1, Ibf2, Pita, and ZIPIC ChIP-seq profile displaying the *Scr-Antp* interval. Vertical red shaded box marks the position of SF1 and SF2 insulators. Bottom track displays the gene features in this region. The X-axis represents the chr3R:6842900-6898714 and was segmented in to 400 bins and smoothed for 1 bin.

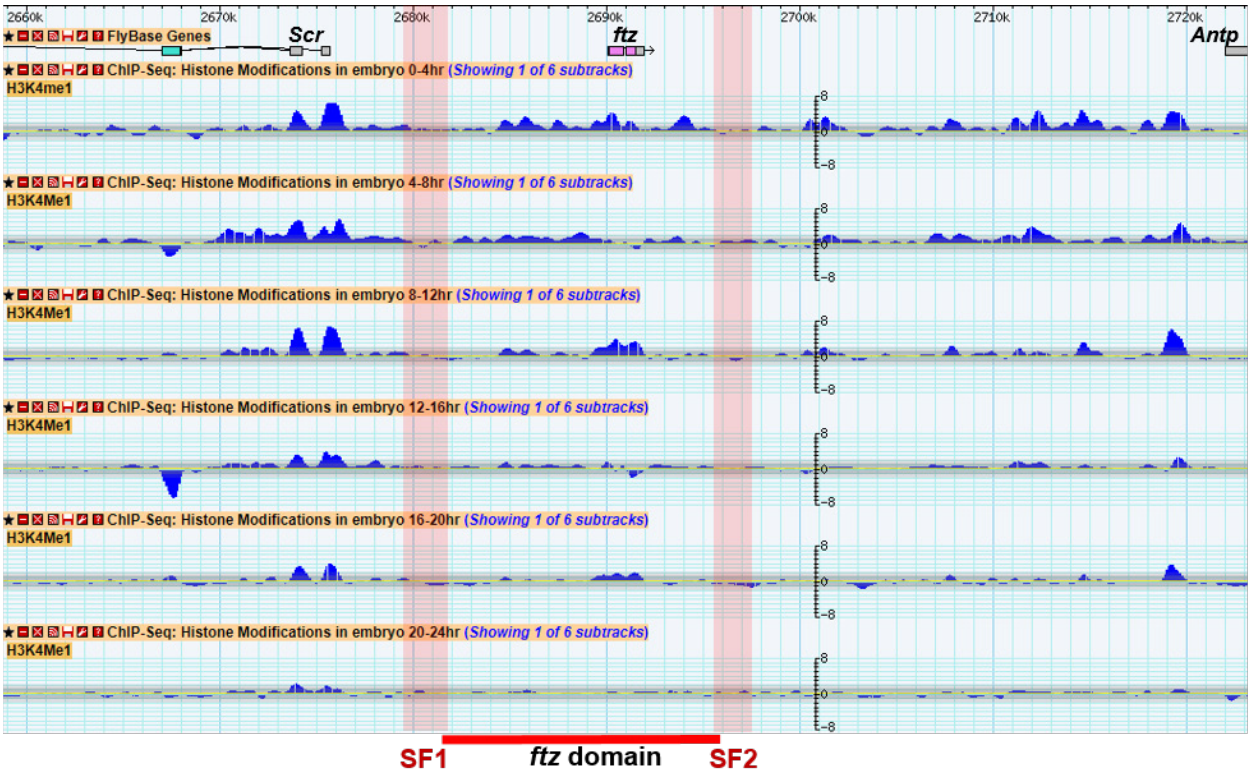
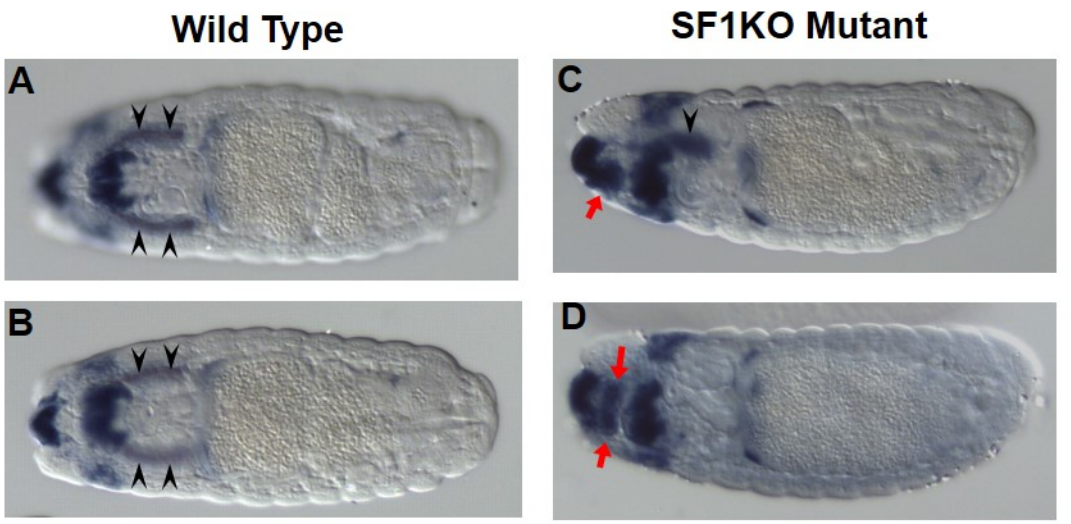

2660k
2670k
2680k
2690k
2700k
2710k
2720k
FlyBase Genes
Scr
ftz
Antp
ChIP-Seq: Histone Modifications in embryo 0-4hr (Showing 1 of 6 subtracks)
H3K4me1
ChIP-Seq: Histone Modifications in embryo 4-8hr (Showing 1 of 6 subtracks)
H3K4me1
ChIP-Seq: Histone Modifications in embryo 8-12hr (Showing 1 of 6 subtracks)
H3K4me1
ChIP-Seq: Histone Modifications in embryo 12-16hr (Showing 1 of 6 subtracks)
H3K4me1
ChIP-Seq: Histone Modifications in embryo 16-20hr (Showing 1 of 6 subtracks)
H3K4me1
ChIP-Seq: Histone Modifications in embryo 20-24hr (Showing 1 of 6 subtracks)
H3K4me1
8
0
-8
SF1
ftz domain
SF2

Figure S2.4 H3K4me1 profile during embryonic development in *Scr-Antp* interval. This is a snapshot of the modENCODE GBrowser displaying the H3K4me1 track of different embryonic stages (http://gbrowse.modencode.org/fgb2/gbrowse/fly/). Blue peaks represent the ChIP signal and the y-axis value means the times of standard deviation (SD) (mod et al., 2010). Light red-shaded boxes mark the position of SF1 and SF2 insulator. The *ftz* gene domain is marked with the red horizontal bar at the bottom.


Wild Type
SF1KO Mutant
A
B
C
D

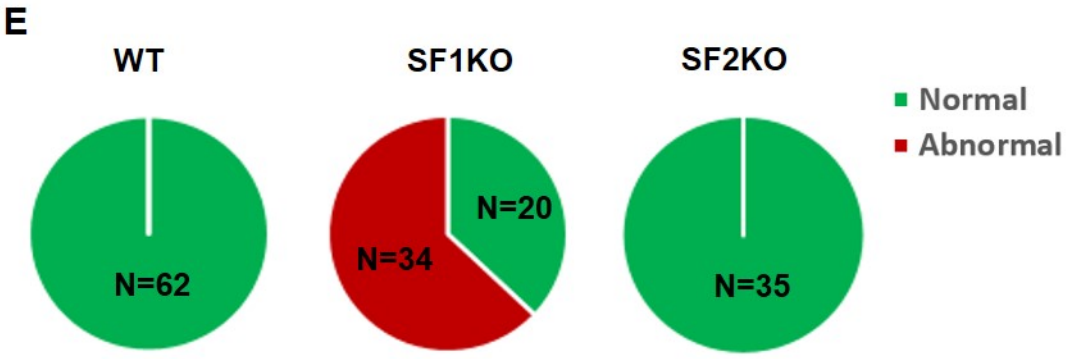

E
WT
SF1KO
SF2KO
Normal
Abnormal
N=62
N=34
N=20
N=35

Figure S2.5 Abnormality in salivary gland in late stage SF1 knockout embryos. (A-D) Representative embryo images of normal salivary glands positioning in wild type (A-B) and abnormal salivary glands positioning in SF1 knockout mutant embryos (C-D) in late embryonic stages. Embryos were relatively stained more heavily with RNA in situ hybridization probe of *Scr* gene. Transverse view with ventral side facing reader and anterior side to the left. The red arrows point to the ectopic expression of *Scr*. The solid black arrow heads points to the position of stained developed salivary glands. (C-D) Representatives of abnormal salivary gland staining with one side missing (C), or both sides missing (D). (E) Quantitative analysis of embryos from stage 15 to stage 17 on the positioning of stained salivary glands in wild type, SF1-knockout and SF2-knockout.

CHAPTER 3

CHROMATIN INSULATORS PROTECT GENOME FUNCTION FROM GENOMIC REARRANGEMENT DURING EVOLUTION

3.1 Introduction

Genomic rearrangements are a major source of evolutionary divergence in eukaryotic genomes, as well as a major cause of genetic diseases and many cancers. It has been shown both in cancer genomes and in evolution that breakpoints of genome rearrangement occur non-randomly (Hinsch and Hannenhalli, 2006; Gordon et al., 2009; Quinlan et al., 2010; Gostissa et al., 2011; Drier et al., 2013) for reasons that remain unclear. The observed non-random distribution of rearrangements breakpoints has been interpreted mainly from two non-exclusive perspectives: i) a direct reflection of rearrangements preferentially occurring at "fragile regions", or ii) a result of surviving rearrangements from the selection pressure on those that occur in functional regions where breakages are highly deleterious. Studies from cancer genomes and cell lines have revealed that certain chromosome regions are prone to higher frequency of breakage both naturally and under stress (Tsai and Lieber, 2010; Gostissa et al., 2011; Drier et al., 2013). These regions are statistically associated with high GC content, repeated sequences, gene densities and DNA hypomethylation (Gordon et al., 2009; Drier et al., 2013). On the other hand, certain functional units, such as gene, co-expressed gene clusters, and conserved genomic regulatory blocks (GRBs), have been reported to exert strong constraints on chromatin reorganization (Hinsch and Hannenhalli, 2006; Bhutkar et al., 2008; Lemaitre et al., 2009;

Berthelot et al., 2015; Harmston et al., 2017). However, the mechanism underlying the occurrence and evolutionary fixation remain largely unclear.

At the DNA level, chromosomal rearrangements are thought to mainly result from erroneous double-strand break repair when simultaneous breaks occur in close physical proximity in the nucleus (Quinlan et al., 2010; Zhang et al., 2012; Seeber and Gasser, 2017). Thus, mechanistic factors that could influence the occurrence of rearrangements may include the double-strand breakage frequency at breakpoints and the physical proximity of breakpoint loci inside of the nucleus. Eukaryotic chromosomes are known to occupy distinct nuclear territories (Cremer and Cremer, 2010) and are further organized into physically insulated domains by chromatin loops at different scales including active/repressive compartments, TADs, and long-range chromatin loops (Lieberman-Aiden et al., 2009; Sexton et al., 2012; Rao et al., 2014; Serizay and Ahringer, 2018). Chromatin insulators have been revealed as a major evolutionarily conserved player in long-range chromatin loop formation and 3D genome organization (Cubenas-Potts and Corces, 2015; Ghirlando and Felsenfeld, 2016; Acemel et al., 2017; Harmston et al., 2017; Serizay and Ahringer, 2018). In both *Drosophila* and mammals, chromatin insulator anchored loops are distributed throughout the genome, creating a network of long-range contacts spanning multiple scales, including not only loops between borders of strongly demarcated TADs but also local scale loops within these domains (Bushey et al., 2009; Dixon et al., 2012; Sexton et al., 2012; Phillips-Cremins et al., 2013; Rao et al., 2014). Such a highly organized feature of genomic DNA led to the question of whether such close spatial proximity of loop anchors promotes preferential rearrangement.

It has been previously reported in the mouse pro-B cell that RAG endonuclease mediated chromatin rearrangement at the antigen receptor locus is one important double-strand break

partner for recurrent translocations that activates various oncogenes in lymphoid malignance (Gostissa et al., 2011). Such translocations in leukemia were subsequently shown to be highly enriched with double-strand break loci that are in close spatial proximity (Zhang et al., 2012). However, this correlation is based on a very low resolution of HiC map (1Mb), which makes the link between chromatin loop anchors and genome rearrangement remain inclusive.

We have recently identified a developmentally regulated chromatin loop anchored by two chromatin insulators, SF1 and SF2, in the *Drosophila* Antennapedia complex (Fig. 2.1C). This SF1-SF2 anchored loop flanks a non-Hox *ftz* gene that is nested inside of the upstream cis-regulatory region of the Hox gene, *Scr.* Both *ftz* and *Scr* are independently regulated through gene specific long-range enhancers during early embryogenesis (Pick et al., 1990; Pattatucci et al., 1991; Gindhart et al., 1995; Calhoun and Levine, 2003). We have shown that interaction between SF1-SF2 not only loops out the complete *ftz* gene unit and insulates the cross-talk between enhancers and promoters of *Scr* and *ftz*, but it also facilitates the long-range interaction between *Scr* distal enhancers and the *Scr* promoter. Furthermore, the SF1 and SF2 flanked *ftz* gene unit has been found in an inverted orientation in several distantly diverged *Drosophila* species, including *Drosophila willistoni*. The breakpoints of this inversion were precisely mapped immediately to the inner sides of the loop anchor sites. To further investigate whether anchors of the loop formed by chromatin insulators could facilitate rearrangement near insulator sites and help fix the rearrangement during evolution, we first examined two inversion cases in the highly conserved Antennapedia complex in the genus of *Drosophila*. We demonstrated the conserved function of chromatin insulators in gene insulation in *Drosophila* species that are distantly related to *Drosophila melanogaster*. Then, we showed the precise association between evolutionary breakpoints and chromatin insulator sites in several *Drosophila* species that

diverged from *Drosophila melanogaster* 15 to 60 million years ago. Together, we provide evidence that chromatin insulator formed loops may contribute both to the formation and fixation of evolutionary rearrangement.

3.2 Results

3.2.1 Conserved SF1 and SF2 insulator sequences demarcate the evolutionary breakpoints of *ftz* gene inversion

In chapter 2, we demonstrated in *Drosophila melanogaster* that the interaction between SF1-SF2 insulators loop out the *ftz* gene unit and insulate the cross-talk of enhancers and promoters of *Scr* and *ftz* gene. This long-range interaction also facilitates the communication of the *Scr* PS2 enhancer that is located 25kb away on the other side of the *ftz* gene and *Scr* promoter, and it also potentially prevents the spread of the polycomb mediated silent chromatin domain (Fig 2.8). The complete *ftz* gene unit, which includes all the identified cis-regulatory sequences and gene sequences, is inverted in several *Drosophila* species that diverged from *Drosophila melanogaster* about 60 million years ago (Fig. 2.8, fig 3.2). Both breakpoints of the inversion are mapped immediately inside of SF1 and SF2 insulators. This led us to ask the question: did the SF1-SF2 mediated loop facilitate the formation or fixation of this inversion event? If the interaction between SF1 and SF2 observed in *Drosophila melanogaster* did play a positive role in this inversion event during evolution, we would expect SF1 and SF2 to be functionally conserved. To test the conservation, we first examined the conservation at DNA sequences level. The multiple-species DNA alignment showed that both SF1 and SF2 contains conserved sequence blocks (Fig S2.2). SF1 contains three conserved blocks. Previous work from our lab demonstrated that the full insulator activity of SF1 requires all three regions of sequences ((Belozerov et al., 2003), and unpublished data). SF2 contains a highly conserved sequence

block that is located immediately outside of the inversion breakpoint in all four species that have the inverted *ftz* (Fig S2.2). We further tested the enhancer blocking activity of the SF1 and SF2 homologs cloned from two of the most distantly diverged *Drosophila* species, *D. virilis* and *D. willistoni* in *D. melanogaster* transgenic embryos. SF1 and SF2 cloned from both *D. virilis* and *D. willistoni* showed strong enhancer blocking activity (Fig. 2.2H-J,Q, fig. 3.1B-C,P). Furthermore, pairing of the Dv.SF1 and Dv.SF2 also displayed orientation-dependent enhancer blocking-cancellation activity (Fig 2.2M-P,Q, fig. 3.1J-K,Q). This unique behavior observed in certain chromatin insulator pairs is also called "enhancer bypass", which indicates the direct interaction between paired insulators (Cai and Shen, 2001; Muravyova et al., 2001; Majumder and Cai, 2003; Melnikova et al., 2004; Kyrchanova et al., 2007; Kyrchanova et al., 2008a; Kyrchanova et al., 2008b; Kyrchanova et al., 2016). In summary, these results suggest that the insulator activity of SF1 and SF2 are functionally conserved in *Drosophila* species that diverged from *Drosophila melanogaster* about 60 million years ago. The enhancer-bypass activity of SF1 and SF2 cloned from *Drosophila virilis* and *Drosophila willistoni* in the transgene assay further indicates that the chromatin loop mediated by SF1 and SF2 interaction may also be functionally conserved among the *Drosophila* species.

3.2.2 Functionally conserved chromatin insulators mark the evolutionary breakpoints of independent *Dfd* Hox gene inversions in *Drosophila*

Hox genes are a set of highly conserved transcription factors that specify body segment identity. Unlike most of the genes that scatter about randomly inside of genomes, Hox genes are tightly clustered together with a conserved 3' to 5' gene order during evolution (Lemons and McGinnis, 2006; Negre and Ruiz, 2006; Noordermeer et al., 2011). The 3' to 5' order of Hox gene in the cluster reflects an anterior to posterior order of expression both in *Drosophila* and

vertebrates (Lemons and McGinnis, 2006; Negre and Ruiz, 2006). Within the cluster, due to the shared global control regions that activate Hox genes over long-range, the order and orientation of Hox genes play a critical role in the spatiotemporal activation of Hox genes in vertebrates (Negre and Ruiz, 2006; Narendra et al., 2015; Acemel et al., 2016). In *Drosophila*, Hox genes rely on gene-specific local enhancers rather than remote shared enhancer clusters (Acemel et al., 2016). All Hox genes in the sequenced *Drosophila* species, which spans up to ~60 million years of evolution (Singh et al., 2009; Gregg and Matthew, 2017), are still arranged in an evolutionarily conserved single orientation within the cluster, except the *Dfd* gene (Fig 3.2) (Negre and Ruiz, 2006). This reflects the functional constraint on the Hox gene order and orientation. However, the *Dfd* gene has at least independently inverted twice among the sequenced *Drosophila* species (Fig 3.2). In addition, it has been shown that chromatin insulators play an important role in demarcating insulated enhancer domains or gene domains between neighboring Hox genes in the *Drosophila* Bithorax complex (Moon et al., 2005; Aoki et al., 2008; Kyrchanova et al., 2008b; Kyrchanova et al., 2011; Wolle et al., 2015). Within the Antennapedia complex, we have identified SF1 and SF2, which keeps the independent regulation of *Scr* and *ftz* (Belozerov et al., 2003). To investigate whether interactions between chromatin insulators has played a role in the inversions of the *Dfd* locus in the genus of *Drosophila*. We examined the inversion breakpoints of the *Dfd* locus. The result shows that the inversion breakpoints from multiple *Drosophila* species all map to the same loci on both sides of the *Dfd* gene that co-localize with known insulator protein binding sites (Fig 3.2B, fig. S3.1). In addition, DS1 and DP1 insulators also precisely mark the boundaries of the polycomb mediated repressive chromatin domain throughout embryonic stages (Fig. S3.1). We cloned these two insulator elements (Fig 3.2B), at the *Dfd* inversion breakpoints and their homologs in the *Drosophila*

virilis which diverged from *Drosophila melanogaster* around 60 million years ago (Fig 3.2A) (Singh et al., 2009; Gregg and Matthew, 2017). Insulator activity was tested in an established enhancer-blocking activity transgene assay in *Drosophila melanogaster* embryos (Cai and Levine, 1997; Cai and Shen, 2001; Belozerov et al., 2003). The results show that both DS1 and DP1 from *D. melanogaster* and *D. virilis* have medium to strong enhancer blocking activity (Fig. 3.1E-I,Q). Pairing of DS1 and DP1 from both species show orientation-dependent "enhancer bypass" activity (Fig. 3.1L-O,Q). Enhancer bypass activity of insulator pairing has been shown to only form between selected insulator pairs that directly interact with each other (Cai and Shen, 2001; Muravyova et al., 2001; Calhoun and Levine, 2003; Majumder and Cai, 2003; Melnikova et al., 2004; Kyrchanova et al., 2007; Kyrchanova et al., 2008a; Kyrchanova et al., 2008b). Direct long-range interaction between DS1 and DP1 region is confirmed in the Recent high-resolution HiC maps generated from early stage fly embryos (Fig S3.2B) (Stadler et al., 2017). A HiC map in embryo-origin Kc cells also shows strong interaction between DS1 and DP1 loci (Juicebox archive, data not shown) (Durand et al., 2016; Cubenas-Potts et al., 2017).

Altogether, our result suggests, like the SF1 and SF2 interaction, that the interaction between DS1 and DP1 loops out *Dfd* locus and insulate it from the influence of nearby regions. This insulation effect from insulator mediated loops may reduce or prevent the potential mis-regulation of Hox genes in this region caused by *Dfd* inversion. Furthermore, the close physical proximity between DS1 and DP1 might also potentially contributed to the formation of rearrangement between these two anchors.

3.2.3 Genome-wide identification of evolutionary breakpoints in *Drosophila*

To analyze the correlation between evolutionary breakpoints and chromatin insulator sites genome wide, we first developed a tool to identify the evolutionary rearrangements from

whole genome pair-wise alignment. We selected 14 *Drosophila* species to compare to *Drosophila melanogaster* in this study because they have comparable high-quality genome assemblies and span at least 60 million years of evolution (Fig 3.2A) (Singh et al., 2009; Obbard et al., 2012; Gregg and Matthew, 2017). New high-quality genome assemblies of these 14 species have recently been generated by using the long-reads Oxford Nanopore sequencing technology (Miller et al., 2018). Benefiting from the long reads length, 5.9kb on average, these newly generated high-quality genome assemblies significantly reduce the gaps present in the current published reference genomes (Miller et al., 2018), which could minimize false positive breakpoints caused by large gaps in the reference genome assembly compared to the previously published short-reads genome assembly (data not shown). Whole genome pairwise alignment against *Drosophila melanogaster* was done using the improved whole genome pair-wise alignment pipeline (Kent et al., 2003; Harris, 2007). The alignment data consists of consecutive alignment blocks generated from the chaining and netting process (Kent et al., 2003). These consecutive alignments, called fills, are ordered hierarchically according to alignment, which contains information about genome rearrangements (Kent et al., 2003). We developed a tool that extracts the breakpoints of a rearranged fill by comparing it to the corresponding parental level gap (parent-child finder) and breakpoints of a rearranged fill by comparing to its surrounding sibling fills at the top level (between-siblings finder) (Fig S3.3, also see materials and methods for details). To overcome potential alignment artifacts and smaller local variations between genomes, we artificially applied a minimum rearranged fragment size filter of 10kb. Rearrangement events reported within the *Drosophila melanogaster* heterochromatin region are also removed due to the nature of high repeat sequences and low assembly quality (See table 3.1 for included non-heterochromatin regions).

We fist examined the distribution of the rearranged fragment size. The size of the rearranged fragment ranges from a few kilo base pairs to a million base pairs in length, with more small-sized fragments than large-sized fragments (Fig S3.4). This indicates that our tool could identify both small and large rearrangements and is consistent with the character of the alignment netting process (Kent et al., 2003). The breakpoints identified from all 14 *Drosophila* species are in a single base-pair resolution, and the majority of them are smaller than 1kb (Fig. S3.5). A very small proportion of breakpoints span a large distance (3-8% are larger than 2kb, Fig S3.5), which is mostly because of either the large-repeat sequences being removed from the alignment process or missing alignments. To get more accurate enrichment test of chromatin insulator protein factors at rearrangement breakpoints, we only kept breakpoints that are smaller than 2kb (Fig. S3.5). After applying the filters, we then examined the total number of identified rearrangements between *Drosophila melanogaster* and the other 14 *Drosophila* species. As expected, *Drosophila* species distantly related to *Drosophila melanogaster* have more rearrangement events identified compared to the species that are more closely related to *Drosophila melanogaster* (Fig 3.3A). This reflects the divergence time from *Drosophila melanogaster*. In addition, the inverted *ftz* locus and *Dfd* locus have both been precisely identified (Fig. 3.3B-C), which further confirms the sensitivity and accuracy of our tool.

In summary, we have identified about 200 to 1300 high-quality evolutionary breakpoints in those *Drosophila* species that diverged from *Drosophila melanogaster* 15- to 60-million years ago (Fig 3.3A). The number of identified rearrangements reflects the divergence time. This is a good sample size for the downstream protein factors enrichment analysis. Species that are very closely related to *D. melanogaster*, including *D. sim, D. mau, D. sec, D. ere, and D. yak,* have

too few rearrangements of which the quality needs more complex examination (see discussion). Thus, they are not included in the downstream protein factors enrichment analysis.

3.2.4 Chromatin insulators are enriched at evolutionary breakpoints

To investigate whether the coincidence observed between chromatin insulator sites and rearrangement breakpoints at SF1-SF2 flanked *ftz* locus and DS1-DP1 flanked *Dfd* locus is true genome wide, we examined the enrichment of chromatin insulator proteins around evolutionary breakpoints. ChIP-chip data of chromatin insulator proteins in 0-12 hour old wild type fly embryos was obtained from modENCODE database and was plot around the breakpoints pooled from all the species. Among the six classic chromatin insulator-associated factors, BEAF-32, dCTCF, CP190, Su(Hw), and Mod(mdg4), all show significant enrichment at evolutionary breakpoints (Fig3.4A-B). BEAF-32, dCTCF, and CP190 show higher enrichment than Su(Hw) and Mod(mdg4). GAF factor shows weak enrichment near breakpoints but not significantly different from the simulated intergenic breakpoints controls (Fig 3.4A-B). Although each individual insulator protein often binds to their target sequence independently, subclasses of different insulator proteins often found jointly bound at insulator sites (Negre et al., 2010; Schwartz et al., 2012). BEAF, dCTCF, and CP190 are often clustered together. Mod(mdg4) binds to insulator sites through the BTB domain mediated interaction with Su(Hw) protein. GAF mostly does not cluster with any of the other five insulator proteins (Negre et al., 2010). It has also been reported that clustering of insulator proteins also scales with TAD border strength (Van Bortle et al., 2014), which indicates importance of these jointly bound insulator sites in chromatin looping formation. These observed enrichment pattern at rearrangement breakpoints reflect the joint binding of BEAF-32, dCTCF and CP190 sub-class, as well as the association of Su(Hw) and Mod(mdg4) (Fig 3.4B).

Based on our previous study of SF1 and SF2 flanked *ftz* locus and DS1 and DP1 flanked *Dfd* locus, we have previously hypothesized that the close physical proximity between anchors of chromatin insulator loops might facilitate the rearrangement between anchor sites. The precise matching of chromatin insulator sites and evolutionary breakpoints (Fig 3.4A-B, Fig 3.5) seems to favor this hypothesis. However, it is also very likely that rearrangements occurred within the gene region have been lost during evolution due to deleterious effect. Thus, evolutionary breakpoints would enrich in the intergenic region. Genome-wide mapping of *Drosophila* chromatin insulator sites showed that chromatin insulators are mostly located in the intergenic regions suggesting their role in insulating neighboring genes (Negre et al., 2010). In addition, compared to mammalian genomes, *Drosophila* genome are known to be much more compact, with an average intergenic region of 5.5kb (Fig. S3.6). To investigate whether this enrichment of chromatin insulator proteins at evolutionary breakpoints is simply because chromatin insulators are enriched in the relatively small intergenic regions, we have simulated breakpoints located only in the intergenic region, with the same sample number, breakpoint size and chromosome distribution compared to the evolutionary breakpoints identified from each species. As shown in Fig. 3.4A, unlike the simulated random breakpoints in the genome, simulated intergenic breakpoints show enriched insulator protein ChIP signals around the breakpoints, but they are less centered and weaker on average (Fig 3.4A, fig. S3.7). This is consistent with the compactness of the fly genome. However, further clustering analysis of the ChIP signal demonstrated that, unlike evolutionary breakpoints, the majority of the simulated intergenic breakpoints do not precisely match the insulator ChIP signal peaks (Fig 3.4C). In contrast, insulator proteins are strongly enriched at least at half of the breakpoint centers within about a 1kb range (Fig. 3.4B). In addition to these strongly enriched sites, about another 25%

breakpoints show relatively weaker insulator protein enrichment at both the center of breakpoints and nearby regions (cluster 2 and 3 in Fig 3.4B). These results suggest that the enrichment of insulator proteins at evolutionary breakpoints are not simply due to the compactness of the fly genome and the enrichment of insulator sites in intergenic regions. Enrichment analysis in breakpoints identified from each species showed similar or better matching of insulator signals around the center of breakpoints (Fig S3.7, data not shown). Permutation test of breakpoints with $1x10^6$ simulations in the intergenic region show that the matching of insulator sites with breakpoints is significant ($p=0.0003$ for *D.eug* and $p<10^{-6}$ for all other species, Fig 3.6A). This further suggests the direct link between chromatin insulators and breakpoints of evolutionary rearrangement.

To further test our hypothesis, we also analyzed the matching of insulator sites and breakpoint pairs. Each rearrangement event will generate a pair of breakpoints. If the close physical proximity brought by anchors of chromatin insulator loops facilitate genome rearrangement at anchor sites, one would expect both breakpoints to match insulator sites in the common ancestral genome where the rearrangement occurred. Considering the conserved function of insulators within the genus of *Drosophila* and between *Drosophila* and mammals, we may speculate that most of these insulator sites are kept during evolution in the descendent *Drosophila melanogaster.* ChIP signal enrichment analysis at breakpoint pairs of insulator proteins show that about half of the pooled breakpoints show both ends match insulator sites. Breakpoints pairs with only one side matched to an insulator site, only represent about 30-40% (Fig 3.5B). The ratio of breakpoint pairs with both end match insulator site is far more than the expected ratio by chance (9-16%). Further analysis using breakpoints identified from each species showed the same pattern as the pooled breakpoints (data not shown). Permutation test in

each species shows that the matching of both breakpoints of a rearrangement with insulator site within a 1kb range from the center of the breakpoint is highly significant for all the tested species ($p<10^{-6}$ for all the species with 10^6 simulations, fig 3.6B-C)

3.3 Discussion and conclusions

Inspired by the precise matching of breakpoints of an evolutionary inversion at the *ftz* gene locus in *Drosophila*, we have hypothesized that the close proximity brought together by interactions between chromatin insulators might contribute to evolutionary rearrangement in at least two different aspects. Firstly, the close physical proximity between the chromatin insulator anchors may directly facilitate the double-strand breaks mediated by rearrangement between anchor sites. Secondly, the insulation effect from the enhancer-blocking activity and boundary activity of chromatin insulators could help prevent potential deleterious effects from rearranging the relative locations of enhancers and promoters near the rearrangement breakpoints.

3.3.1 Inversion of chromatin insulator loops in the highly conserved Hox cluster

To test our hypothesis described above, we first demonstrated that SF1 and SF2 insulator elements are functionally conserved in *Drosophila* species that diverged from *Drosophila melanogaster* about 60 million years ago. Homologs of SF1 and SF2 in *Drosophila willistoni*, and *Drosophila virilis* both show strong enhancer blocking activity and orientation dependent enhancer-bypass activity. In addition to the SF1 and SF2 flanked ftz gene locus inversion, one of the Hox genes, *Dfd*, is independently inverted at least twice during evolution (Fig 3.2A). Breakpoints of these independent inversion also map precisely to two functionally conserved chromatin insulators in *Drosophila melanogaster* (Fig. 3.1, figS3.1). In addition, a previously published high resolution map in both early fly embryo mixed tissue and embryonic origin Kc cells shows a strong long-range interaction between these two insulators (Fig S3.2, Juicebox

archive, data not shown) (Cubenas-Potts and Corces, 2015; Stadler et al., 2017). Results from these two highly conserved and developmentally regulated loci in *Drosophila* indicates that chromatin insulators may play an important role in chromatin rearrangement.

3.3.2 Quality of identified rearrangement breakpoints

The total number of breakpoints identified by our methods using the newly published genome assemblies of 14 *Drosophila* species well reflect the divergence time between *Drosophila melanogaster* and each of these species. Among these 14 species, 5 of them are very closely related to *Drosophila melanogaster* (diverged less than 5 million years ago). Very few (~15 to 50) rearrangements larger than 10kb were identified in these species. Upon close examination, most of these rearrangements are at the top level that represent assembled scaffolds or scaffolds broken by large repeated sequences (data not shown). Those 14 newly sequenced genomes are assembled into large scaffolds/contigs, which means DNA sequences are not assembled fully intro chromosomes. The breakpoints reported at the border of each contig may or may not be the true breakpoints. Considering that it only consists of a small proportion in those more distantly related *Drosophila* species, we did not remove them from the downstream analysis. Better assembled reference genomes in those species would reduce the false discovery rate of such events. The breakpoints are mapped mostly in a small range, about 90% within 2kb and about 80% within 1kb, which indicates precise mapping of rearrangement breakpoints.

Previous studies have reported that transposon mediated rearrangement and processed pseudogenes may result in short non-syntenic alignment fills (Kent et al., 2003; Mills et al., 2006). We applied a 10kb minimum size filter to avoid potential bias from local chromatin variation and transposon mediated re-arrangements. The goal of this study is to find evidence of correlation between evolutionary breakpoints and anchors of long-range loops mediated by

chromatin insulators. Analyzing rearrangements larger than an artificial 10kb threshold seems not biasing the result.

3.3.3 Active facilitation or negative selection?

The majority of these identified rearrangement are likely to be functional elements as evidenced by their conservation over long evolutionary periods, and they almost all contain at least one complete gene unit (data not shown). It is likely that rearrangements that disrupted these functional elements were selected against, and, therefore, are not present in the population. This is consistent with reports in the literature that genomic regulatory blocks (GRB) and genes are highly conserved during evolution (Hinsch and Hannenhalli, 2006; Bhutkar et al., 2008; Lemaitre et al., 2009; Berthelot et al., 2015; Harmston et al., 2017). Previous studies also showed that TADs are conserved ancient domains that coincide with synteny blocks (Berthelot et al., 2015; Dixon et al., 2015; Harmston et al., 2017). These observations suggest that 3D organization plays an important role in gene evolution. Insulation of TADs from the neighboring genome might protect rearrangements that contain a complete TAD from having deleterious effects. However, current understanding of the association of synteny block breakpoints and TAD boundaries are mainly based on low-resolution analysis. In Harmston's experiment, matching between synteny block borders and TAD boundaries was counted within a range of hundred kilobases to 1Mb, and only 29% of synteny blocks overlaps within 120kb of TADs in humans (Harmston et al., 2017). In Dixson's analysis of TAD conservation, the comparison is based on low resolution HiC map (20kb) and relatively relaxed boundary size cutoff (size less than 400kb). Rowley et al. recently showed in *Drosophila* that TAD boundaries viewed in low resolution (10-40kb) are actually composed of smaller compartment domains which contain one or multiple genes of the same transcriptional status (Rowley et al., 2017). It is hard to understand

the mechanistic relationship between genomic rearrangement breakpoints and TAD boundaries at such a low resolution.

The availability of 15 complete high-quality genomes of various species of the genus *Drosophila* provide a unique opportunity to analyze genome-scale rearrangements among a group of related species over a large range of divergence time up to ~60 million years (Fig 3.2A). Our study reported evolutionary breakpoints in single base pair resolution. Around 85% of the total breakpoints span less than 1kb and 60% are mapped within 500bp in the *Drosophila melanogaster* genome. Those breakpoints were identified through pair-wise genome alignments between two species that diverged 15-60 million years ago. The insulator protein factors are found precisely enriched at the center of breakpoints (Fig 3.4A-B, fig 3.5) within a range of ~1kb. About half of the rearrangements have both ends matching insulator sites within a 1kb distance from the center of the breakpoint (Fig 3.5B). This ratio is significantly higher than the possibility of both ends matching insulators by chance. Our results strongly support the direct relationship between chromatin insulator sites and evolutionary breakpoints. Our data does not exclude the possibility of negative selection, most of the breakpoints are located within the intergenic region and nearly half of the breakpoints do not match or precisely match chromatin insulator sites (Fig 3.4B). Results from simulated breakpoints in the intergenic region argues against the possibility that the enrichment of insulator sites at breakpoints are simply because of the compactness of the fly genome and the enrichment of insulator sites in the intergenic region (Fig 3.4C).

Topoisomerase II (TOP2) is known to cut both strands of DNA through an ATP dependent manner. TOP2B activity in mammals is critical to release torsional stress at sites of active transcription. TOPO2 inhibition induces increased genome rearrangement (Kantidze and

Razin, 2009; Canela et al., 2017). In vertebrates, DNA topoisomerase II beta (TOP2B) has recently been reported to physically interact with cohesin and CTCF at TAD boundaries (Uusküla-Reimand et al., 2016; Barutcu et al., 2017). Recent work showed that TOP2B is enriched at the loop anchors at the B-cell receptor locus, which may directly drive the rearrangement (Canela et al., 2017). A paper published earlier this year demonstrated that CTCF sites mediate accessibility of RAG endonuclease substrates to initiate antibody heavy chain variable region recombination (Jain et al., 2018). These studies indicate that chromatin insulators may directly interact with DNA nucleases. Nucleases such as topoisomerase II at the loop anchors may cause increased vulnerability to TOPII mediated double stranded breaks. *Drosophila* topoisomerase II has been shown to modulate the function of the Su(Hw) insulator (Ramos et al., 2011). These results suggest the direct interplay between chromatin insulators and DNA nucleases. Considering the well-known role of chromatin insulators in forming chromatin loops, our data suggest that, together with negative selection, the close physical proximity brought by the loops formed between chromatin insulators may have directly contributed to the formation of evolutionary rearrangement. The enhancer-blocking and boundary activity of insulator anchored chromatin loops may have contributed to the fixation of such rearrangements during evolution. Interaction between chromatin insulators with nucleases, such as topoisomerase, may directly contribute to the vulnerability of double strand breaks formation.

3.4 Materials & Methods

3.4.1 Enhancer-blocking assay with transgenic *Drosophila* embryos

Enhancer-blocking assay including spacer- and SF1- containing transgenic *Drosophila* lines were described previously (Cai and Levine, 1997; Cai and Shen, 2001; Belozerov et al., 2003). The original pWNHZ vector was modified by adding a phiC31 attB site at the NsiI site

downstream of miniwhite reporter gene. The attB site was PCR cloned from the pBphi-yellow vector, which as a gift from Dr. Chris Rushlow. The Individual chromatin insulator elements were cloned by PCR (see table 3.2 for primers), TOPO-cloned, purified after further digestion, and inserted into the AscI or Not1 sites between the NEE and H1 enhancers in pBWNHZ vector (see Fig. 3.1A). All the transgene constructs were inserted at the VK33 attP docking site via phiC-31 integrase mediated site-specific transformation. Microinjections were performed in the Cai lab. In situ hybridization with *lacZ* and *white* anti-sense RNA probes were performed as previously described (Cai et al., 2001). Whole mount in situ hybridization and visual assessment of reporter expression were performed according to an existing procedure (Cai and Levine, 1997; Cai et al., 2001; Belozerov et al., 2003). For each pBWNHZ transgene, about 100 to 150 embryos were scored double-blindly. Briefly, blastoderm stage embryos were scored for lacZ levels in the NEE domain (two horizontal strips in the lateral ventral ectoderm) against the H1 domain (vertical ring/strip in the head region) (Fig. 2.2B-P). Based on the intensity ratio of the lacZ in H1/NEE domains, embryos were ranked into five categories from complete block: NEE/H1 ≤ 2%, strong block: 2% < NEE/H1 ≤ 10%, medium block: 10% < NEE/H1 ≤ 40%, weak block: 40% < NEE/H1 ≤ 70%, no block: NEE/H1 > 70%.

3.4.2 Breakpoints deification

Newly published genome assemblies by Miller et al were obtained from https://github.com/danrdanny/*Drosophila*15GenomesProject. Then each genome was aligned to *Drosophila melanogaster* (dm6) genome using lastz. chrUn and chr_**_random were not included. The alignments are further processed into long consecutive blocks through chaining and netting process as describe previously (Kent et al., 2003). Then the net file was further

annotated by netSynteny and netClass tools developed by the UCSC Genome Browser team. Pre-compiled unix tools were obtained from http://hgdownload.soe.ucsc.edu/admin/exe/.

The net file was used to extract rearrangements between *Drosophila melanogaster* and the species of interest. Two different methods were used to identify breakpoints. The first method utilizes the hierarchy structure of alignment fills presented in the net files. By comparing the fill of a child level to the gap in the parent level, the child level fill is reported as a rearranged fragment with breakpoint coordinates defined between the edges of the gap and the edges of the fill (illustrated in Fig. S3.3). If a large gap contains multiple child level fills, then the breakpoint of a fill is narrowed down to edges of the qualified neighboring fills. The parent-child method searches through all levels in the net file. The second method complements the parent-child method by identifying the rearrangements presented only at the top level. Each fill at the top level is reported as a rearrangement event with breakpoints reported between the edges of this fill and its qualified neighboring fills. Rearrangements identified through this method might be affected by the assembly quality as each contif/scaffold is treated as a chromosome. Result from both methods includes the chromosome name, start and end of the first breakpoint, start and end of the second breakpoint, and the type of breakpoint.

Identified breakpoints were subjected to further filtering. Breakpoints located in the heterochromatin regions are removed (see table 3.1 for the included chromatin region). Breakpoints with either end larger than 2kb were removed.

3.4.3 Visualizing alignment data in UCSC Genome browser

In order to visualize the net files, a local version of UCSC Genome browser site was built following the “Installing a Genome Browser locally with the GBiC installer” user guide. All the 14 genomes of *Drosophila* species were loaded onto the local Genome Browser by following the

instructions at UCSC genomewiki site (http://genomewiki.ucsc.edu/index.php/Building_a_new_genome_database). Then the alignment chain and net files were loaded to the local genome browser database by using the hgLoadNet and hgLoadChain tools of the kent tree utilities (http://hgdownload.soe.ucsc.edu/admin/exe/). Then the loaded net tracks in the genome browser were used to visually evaluate the breakpoints.

3.4.4 Chromatin insulator proteins enrichment analysis

Protein enrichment analysis at breakpoints

For the enrichment analysis at breakpoint center, all the breakpoint pairs are pooled into one set to breakpoints in bed format with redundant breakpoints removed. Then this list of breakpoints was used as region set for the computeMatrix tools in the deeptools v.3.1.0 tool set. *Drosophila* 0-12h embryos ChIP-chip data of chromatin insulator proteins were obtained from modENCODE with accession ID: BEAF, #21; CP190, #22; CTCF-C, #769, CTCF-N, #770; Su(Hw), #27; Mod(mdg4), #24; GAF, #23. Wig files were lifted from dm3 to dm6 based coordinate by using UCSC liftover tool. Then the wig file is converted to bigwig format using the wigToBigWig tool developed by UCSC Genome browser team. Then these bigwig format ChIP-chip signal files were used as score files for the computeMatrix tools in the deeptools v.3.1.0 tool set. Heatmap was generated using the plotHeatmap tool. Breifly, all the regions are centered at the breakpoint center, and the ChIP-signals are aligned to the region and recoded as one line in the heatmap. The signals are reflected through the color. The clustering analysis was done using the -kmeans option of the plotHeatmap tool.

Protein enrichment analysis at both breakpoints of a rearrangement

For enrichment analysis at both rearrangement breakpoints, deeptool computeMatrix scale-region tool was used to generate the matrix of the heatmap, which is subsequently plotted

by using plotHeatmap tool. Breifly, two breakpoints of a pair were aligned at the centers of both breakpoints. Then the region in between was scaled to the same length. To preserve the true signal around the breakpoint, 4kb upstream and downstream of each breakpoint center were not scaled. The clustering analysis was done using the -kmeans option of the plotHeatmap tool.

Breakpoints simulations

Control breakpoints were simulated in two ways. The simulated random breakpoints were generated by randomly choosing breakpoint pairs in the genome, with the same breakpoint size, the same rearrangement fragment size, and the same distribution of chromosomes. The simulated intergenic breakpoints were generated by randomly choosing a single breakpoint in the intergenic region only. Simulated breakpoints of this type were not in pairs. Each simulated intergenic breakpoint was generated with the same breakpoint size, and same chromosome name. Breakpoint simulation from both methods excluded the heterochromatin area (see table 3.1 for the included chromatin region).

3.4.5 Permutation test of breakpoints and insulator sites matching

A simulation test was performed by simulating at least 10^6 samples and examining the matching rate of the breakpoint to the peak annotation of insulator sites. The annotated peaks of the insulator protein ChIP-chip experiment were obtained from modENCODE (Negre et al., 2010). The average matching rate from the simulations were taken and the p-value of the simulation test was recorded. For the simulation test of both ends match, matching rate of rearrangement with both ends match and with single end match were recoded separately. The average matching rates were also calculated separately, as well as the p-value of the simulation test.

Table 3.1 Coordinates of non-heterochromatin regions used in filtering breakpoints (dm6)

Chromosome Name	Start Coordinate	End Coordinate
chr2L	5370	22415564
chr2R	4130838	25257890
chr3L	21890	24550420
chr3R	4174280	32069400
chrX	122507	23020445

Table 3.2 DNA primer sequences

Dm_DP1_F	CTCTTCCTTCTCCTCCTCCTC
Dm_DP1_R	GCAAACATCAATGTCAACTGTAACG
Dvir_DS1_F	CTCTCAGCTTTACTTGAGTGAGC
Dvir_DS1_R	CCAAGGGCTGGCTATTTACAC
Dvir_DP1_F	TGTAACTGCATTGGTAAGATCTCG
Dvir_DP1_R	GAGGGACTGGTACCATCTGC
Dvir_SF1_F	TTGCGGCCGCCTCATTTGTGTTCATTATCGTCCTG
Dvir_SF1_R	TTGCGGCCGCTCGACATAGGTGTGTGTAAGC
Dvir_SF2_F	GTTTCATGCTGTGGCCACA
Dvir_SF2_R	TGTCGTTCTGATTGACGAATTGC

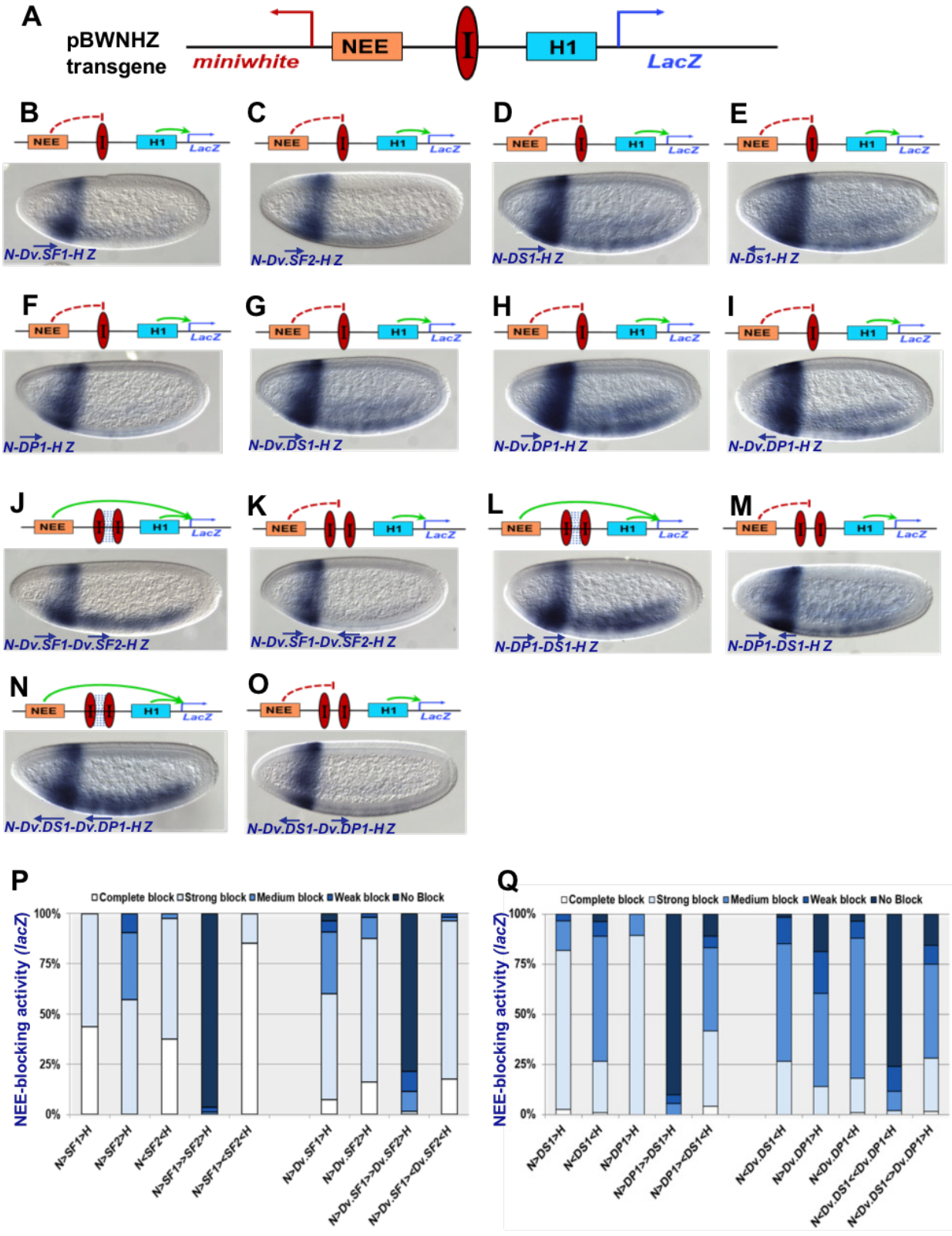

A
pBWNHZ
transgene
miniwhite
NEE
I
H1
LacZ
B
N-Dv.SF1-H Z
C
N-Dv.SF2-H Z
D
N-DS1-H Z
E
N-Ds1-H Z
F
N-DP1-H Z
G
N-Dv.DS1-H Z
H
N-Dv.DP1-H Z
I
N-Dv.DP1-H Z
J
N-Dv.SF1-Dv.SF2-H Z
K
N-Dv.SF1-Dv.SF2-H Z
L
N-DP1-DS1-H Z
M
N-DP1-DS1-H Z
N
N-Dv.DS1-Dv.DP1-H Z
O
N-Dv.DS1-Dv.DP1-H Z
P
Complete block
Strong block
Medium block
Weak block
No Block
NEE-blocking activity (lacZ)
100%
75%
50%
25%
0%
N>SF1>H
N>SF2>H
N<SF2<H
N>SF1>>SF2>H
N>SF1><SF2<H
N>Dv.SF1>H
N>Dv.SF2>H
N>Dv.SF1>>Dv.SF2>H
N>Dv.SF1><Dv.SF2<H
Q
Complete block
Strong block
Medium block
Weak block
No Block
NEE-blocking activity (lacZ)
100%
75%
50%
25%
0%
N>DS1>H
N<DS1<H
N>DP1>H
N>DP1>>DS1>H
N>DP1><DS1<H
N<Dv.DS1<H
N>Dv.DP1>H
N<Dv.DP1<H
N<Dv.DS1<<Dv.DP1<H
N<Dv.DS1<>Dv.DP1>H

Figure 3.1 Enhancer-blocking activity assay in transgenic *Drosophila*. (A) Diagram of the pBWNHZ transgene vector containing divergently transcribed miniwhite and lacZ reporters flanking the NEE and H1 enhancers. Red oval with label "I", insulator or spacer control between NEE and H1. (B-O) Representative enhancer-promoter communication of the tested reporter gene is represented with a schematic diagram on top of the embryo image. The green curve arrow represents active enhancer-promoter interaction, dashed red line represents restricted enhancer access. Enhancers, insulators, and reporter gene are represented the same as in (A). The blue dashed lines between the two insulator ovals of the diagram in (J, L, N) represent interaction between insulators. Representative images of transgenic embryos after whole-mount in situ hybridization with the anti-lacZ RNA probe. The transgene contained in these embryos are labeled at the bottom left of each photo, with probe indicated (Z, lacZ). Embryos are shown in sagittal views with anterior to the left and dorsal up. The arrow on top of the transgene label represents the orientation of the insulator elements under the arrow. The orientation is determined by comparing to the orientation of the insulator in the genome of *Drosophila melanogaster*. "Dv" the transgene name represents *Drosophila virilis*. (P,Q) Quantitation of NEE-blocking in the whole neuroectoderm in transgenic embryos stained with the anti-lacZ probe (see Methods and Materials for details).

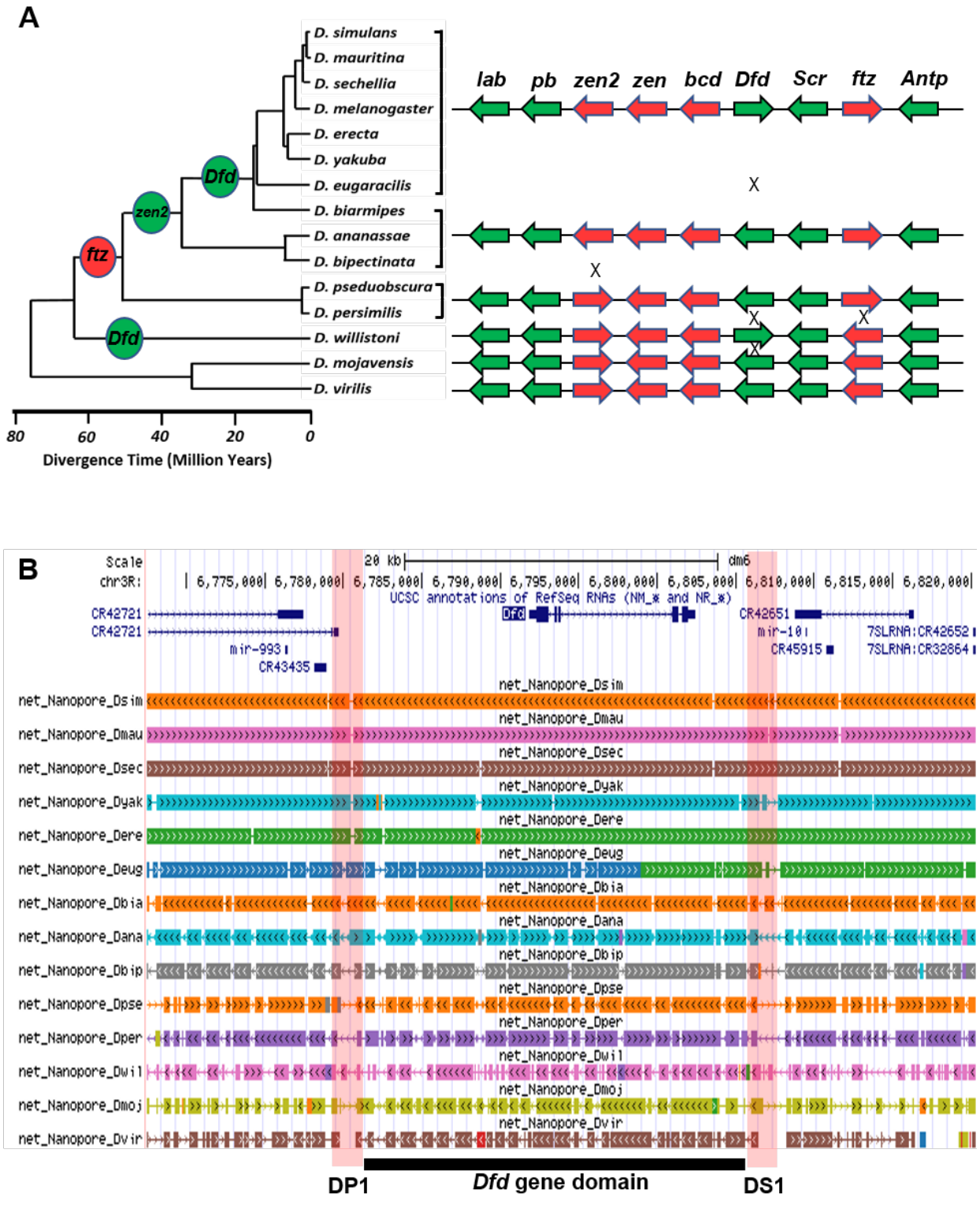
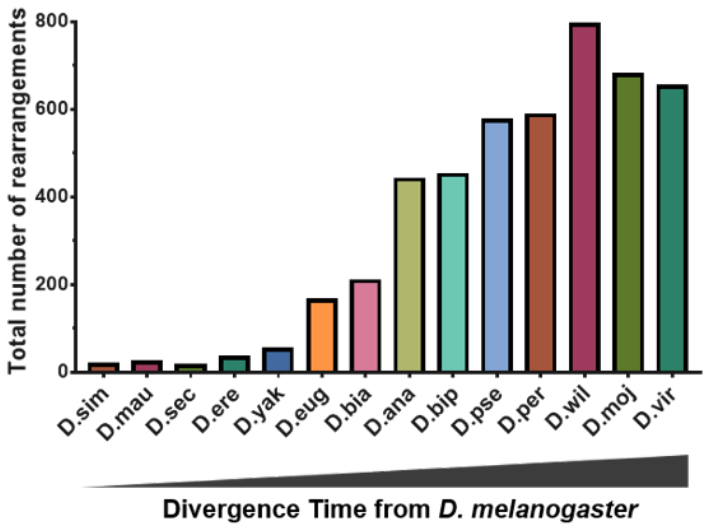
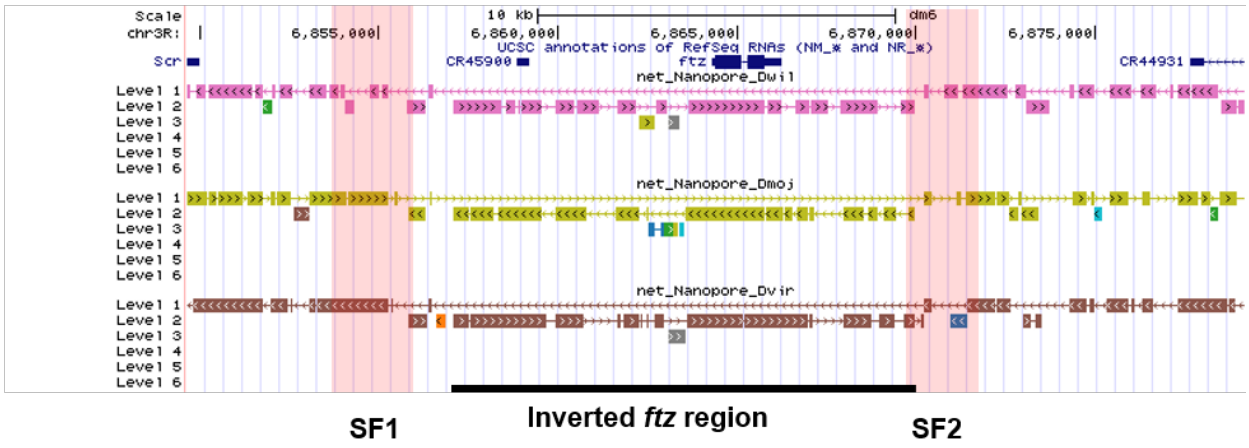

A
D. simulans
D. mauritina
D. sechellia
D. melanogaster
D. erecta
D. yakuba
D. eugaracilis
D. biarmipes
D. ananassae
D. bipectinata
D. pseudoobscura
D. persimilis
D. willistoni
D. mojavensis
D. virilis
Dfd
zen2
ftz
Dfd
80
60
40
20
0
Divergence Time (Million Years)
lab
pb
zen2
zen
bcd
Dfd
Scr
ftz
Antp
X
B
Scale
chr3R:
20 kb
dm6
6,775,000
6,780,000
6,785,000
6,790,000
6,795,000
6,800,000
6,805,000
6,810,000
6,815,000
6,820,000
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
CR42721
CR42721
mir-993
CR43435
Dfd
CR42651
mir-10
CR45915
7SLRNA:CR42652
7SLRNA:CR32864
net_Nanopore_Dsim
net_Nanopore_Dmau
net_Nanopore_Dsec
net_Nanopore_Dyak
net_Nanopore_Dere
net_Nanopore_Deug
net_Nanopore_Dbia
net_Nanopore_Dana
net_Nanopore_Dbip
net_Nanopore_Dpse
net_Nanopore_Dper
net_Nanopore_Dwil
net_Nanopore_Dmoj
net_Nanopore_Dvir
DP1
Dfd gene domain
DS1

Figure 3.2 Evolutionary inversions in the *Drosophila* Antennapedia complex. (A) A schematic representation of genes in the Antennapedia complex region among different species of the *Drosophila* genus. The phylogeny tree was modified from (Gregg and Matthew, 2017). Species that has been recently sequenced by using the long-reads Nanopore technology are included. The gene orientation was generated based on the comparison to *Drosophila melanogaster* by using pairwise-alignment net tracks and gene orientation in *Drosophila melanogaster*. Green arrows represent Hox genes and red arrows represent Hox-derived protein coding genes that already lost homeotic functions. The size of and distance between genes are not in scale. The back crosses between genes of different species represent evolutionary inversion. Inversions are mapped onto the phylogeny tree through comparative analysis. (B) Net tracks generated from pair-wised whole genome alignment between *Drosophila melanogaster* and other *Drosophila* species are displayed in the UCSC Genome browser. The colors in the net track were automatically generated to represent chromosomes. The color is not comparable between different tracks due to the nature of the unfinished scaffold genomes. Arrows in the colored boxes in each track represent the relative alignment direction to *Drosophila melanogaster*. Vertical red shaded boxes mark the cloned DP1 and DS1 locations. Black line at the bottom marks the complete *Dfd* gene domain.

A


Total number of rearrangements
800
600
400
200
0
D.sim
D.mau
D.sec
D.ere
D.yak
D.eug
D.bia
D.ana
D.bip
D.pse
D.per
D.wil
D.moj
D.vir
Divergence Time from D. melanogaster

B


Scale
chr3R:
10 kb
dm6
6,855,000
6,860,000
6,865,000
6,870,000
6,875,000
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
Scr
CR45900
ftz
CR44931
net_Nanopore_Dwil
net_Nanopore_Dmoj
net_Nanopore_Dvir
Level 1
Level 2
Level 3
Level 4
Level 5
Level 6
SF1
Inverted ftz region
SF2

C

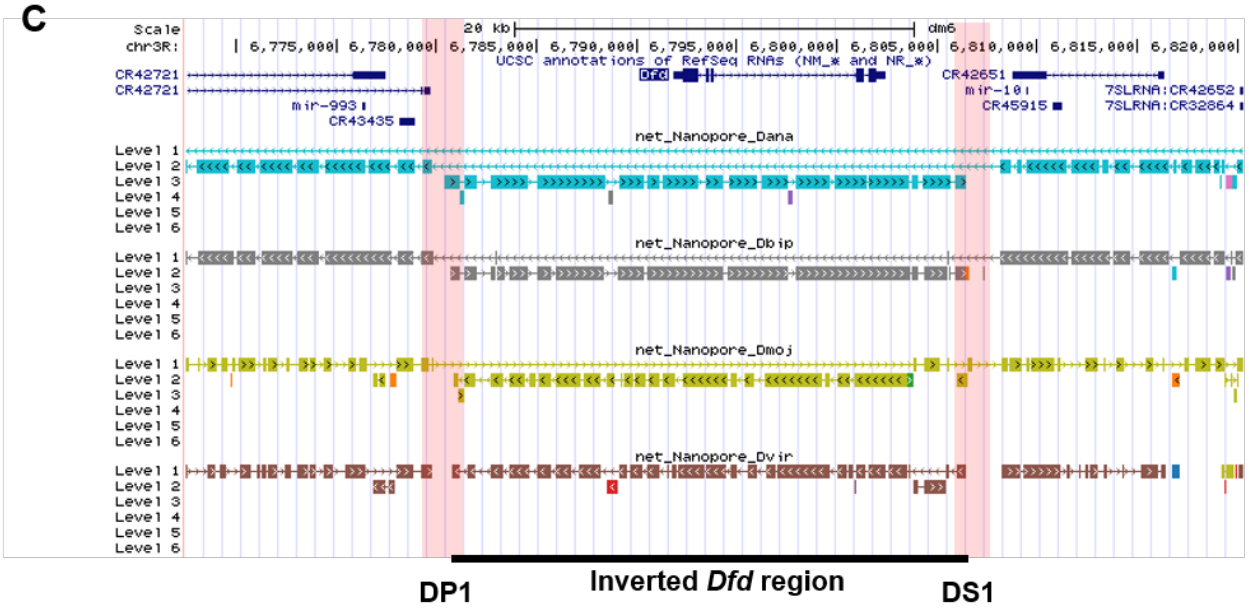

Scale
chr3R:
20 kb
dm6
6,775,000
6,780,000
6,785,000
6,790,000
6,795,000
6,800,000
6,805,000
6,810,000
6,815,000
6,820,000
UCSC annotations of RefSeq RNAs (NM_* and NR_*)
CR42721
CR42721
Dfd
CR42651
mir-993
CR43435
mir-10
CR45915
7SLRNA:CR42652
7SLRNA:CR32864
net_Nanopore_Dana
net_Nanopore_Dbip
net_Nanopore_Dmoj
net_Nanopore_Dvir
Level 1
Level 2
Level 3
Level 4
Level 5
Level 6
DP1
Inverted Dfd region
DS1

Figure 3.3 Total number of rearrangements in each species reflects evolutionary divergence time in genus of *Drosophila*. (A) total number of rearrangements of each species plotted in a column bar graph. Y axis represents the total count. X axis represents each species arranged with increasing divergence time from *Drosophila melanogaster*. (B) Representative inversion identified in the previously characterized *ftz* locus. The red shaded vertical boxes mark the location of cloned SF1 and SF2 insulators. The black line at the bottom marks the inverted *ftz* gene region. (C) Representative inversion identified in the previously characterized *Dfd* locus (B-C) The net tracks of pair-wise alignment are displayed in the UCSC Genome browser. The UCSC annotation of RefSeq track is displayed on top. Level 1 represents the top level, level 6 represents the lowest level in the hierarchy structure of net track. Chromosomes marked in different color are automatically assigned in each track. Colors between different genomes are not comparable. The direction of alignment are marks with arrows inside of the colored horizontal boxes (fills). The red shaded vertical boxes mark the location of cloned DP1 and DS1 insulators. The black line at the bottom marks the inverted *Dfd* gene region.

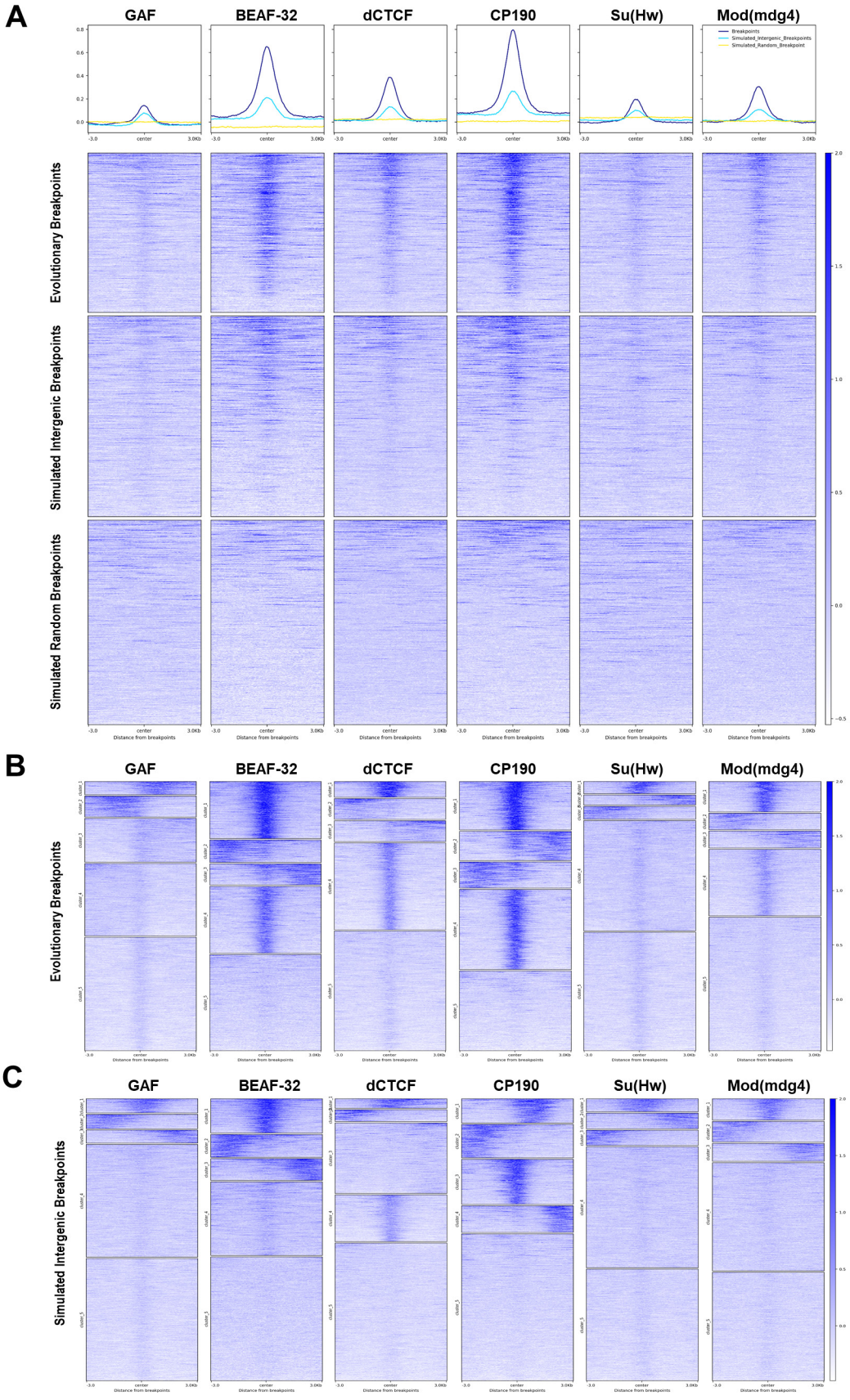
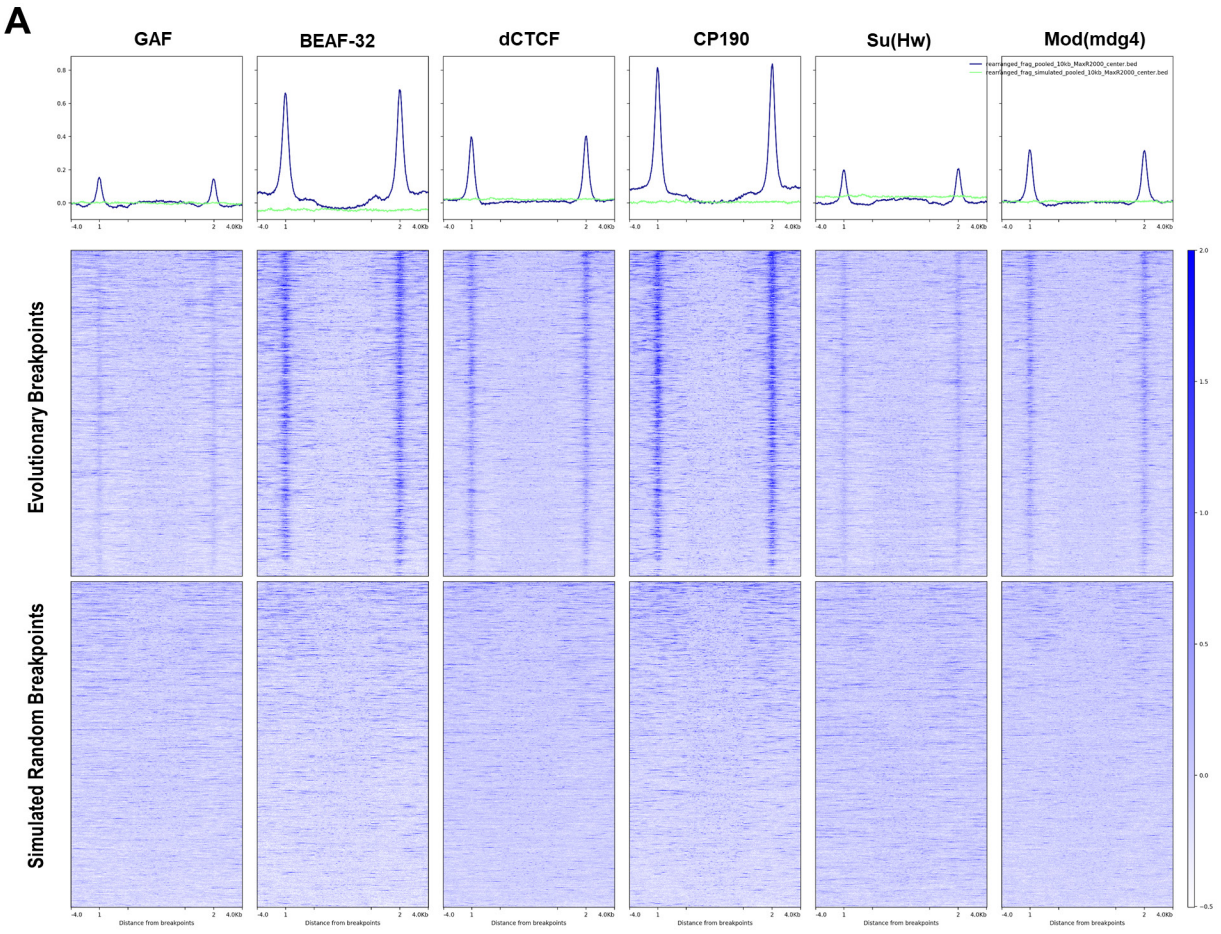

A
GAF
BEAF-32
dCTCF
CP190
Su(Hw)
Mod(mdg4)
Breakpoints
Simulated_Intergenic_Breakpoints
Simulated_Random_Breakpoint
Evolutionary Breakpoints
Simulated Intergenic Breakpoints
Simulated Random Breakpoints
-3.0
center
3.0Kb
Distance from breakpoints
B
GAF
BEAF-32
dCTCF
CP190
Su(Hw)
Mod(mdg4)
Evolutionary Breakpoints
C
GAF
BEAF-32
dCTCF
CP190
Su(Hw)
Mod(mdg4)
Simulated Intergenic Breakpoints

Figure 3.4 Chromatin insulators are enriched at evolutionary breakpoints. (A) Heatmap of chromatin insulator protein enrichment in a ±3kb region around the center of pooled breakpoints from all the species. Top profile plot represents the average signal around the center of all breakpoints. Lines in different colors represent different groups of breakpoints: evolutionary breakpoints, simulated intergenic breakpoints, and simulated random breakpoints (see materials and methods). Regions in each group are sorted on a descending order of the mean signal intensity. Signal intensity is represented by the darkness of the blue color. (B-C) Heatmap of chromatin insulator protein enrichment in a ±3kb region around the center of breakpoints by using kmeans clustering analysis with kmeans=5. (B) Represents the clustering analysis at evolutionary breakpoints. (C) Represents the clustering analysis of at the simulated intergenic breakpoints with kmeans=5. (A-C) The center of heatmap represent the center of breakpoints. The ±3kb distance from the center of breakpoints is marked at the bottom of each heatmap.


A
GAF
BEAF-32
dCTCF
CP190
Su(Hw)
Mod(mdg4)
rearranged_frag_pooled_10kb_MaxR2000_center.bed
rearranged_frag_simulated_pooled_10kb_MaxR2000_center.bed
Evolutionary Breakpoints
Simulated Random Breakpoints
Distance from breakpoints

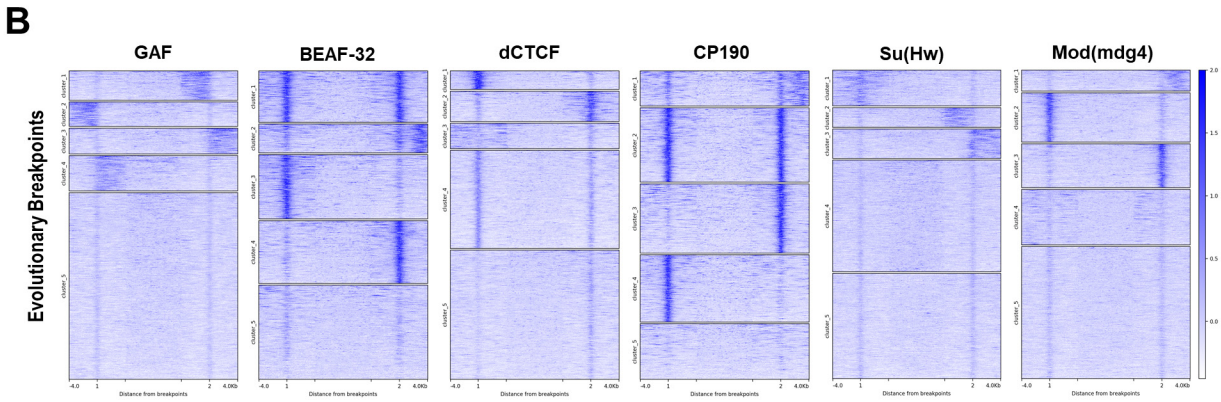
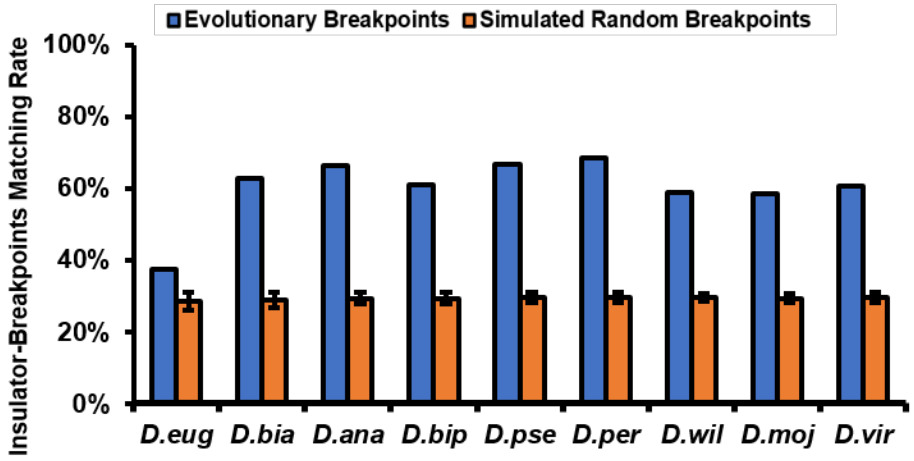
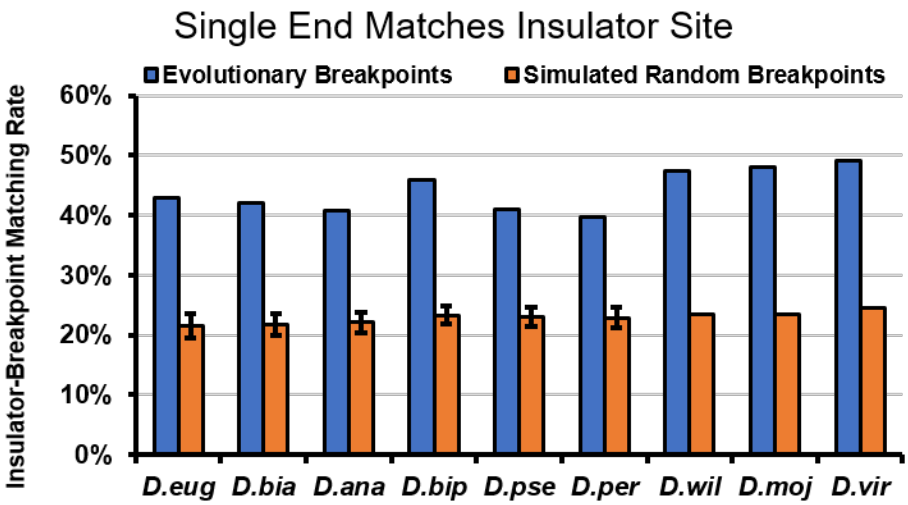

B
GAF
BEAF-32
dCTCF
CP190
Su(Hw)
Mod(mdg4)
Evolutionary Breakpoints
Distance from breakpoints

Figure 3.5 Chromatin insulators are enriched at both breakpoints of a rearrangement. (A) Heatmap of chromatin insulator protein enrichment at the centers of both breakpoints of a rearrangement, and the region in between. Pooled rearrangement from all the distantly related species are used (see methods and materials for details). 4kb regions both upstream and downstream of each breakpoint center not scaled to guarantee the true representation of signal around breakpoints. The rest of the region in between are scaled to the same length. Top profile plot represents the average signal. Lines in different color represent different groups of breakpoints: evolutionary breakpoints, and simulated random breakpoints (see materials and methods). Regions in each group are sorted on a descending order of the mean signal intensity. Signal intensity is represented by the darkness of the blue color. (B) Heatmap of chromatin insulator protein enrichment at the centers of both breakpoints of a rearrangement, and the region in between by using kmeans clustering analysis with kmeans=6. (A-B) 1 and 2 marks the center of breakpoint #1 and #2 of a rearrangement. Unscaled region around breakpoint centers are marked. The center regions between the inner two marks are scaled to the same length.

A


Evolutionary Breakpoints
Simulated Random Breakpoints
Insulator-Breakpoints Matching Rate
100%
80%
60%
40%
20%
0%
D.eug
D.bia
D.ana
D.bip
D.pse
D.per
D.wil
D.moj
D.vir

B


Single End Matches Insulator Site
Evolutionary Breakpoints
Simulated Random Breakpoints
Insulator-Breakpoint Matching Rate
60%
50%
40%
30%
20%
10%
0%
D.eug
D.bia
D.ana
D.bip
D.pse
D.per
D.wil
D.moj
D.vir

C

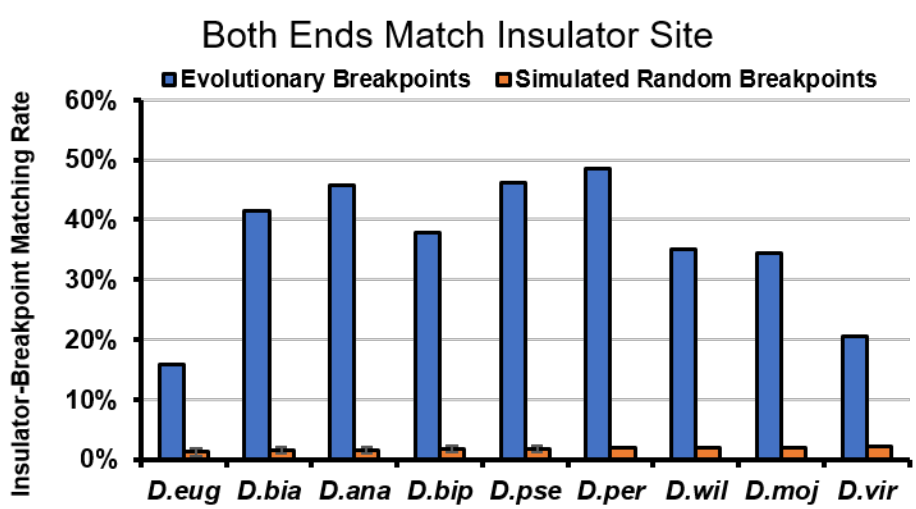

Both Ends Match Insulator Site
Evolutionary Breakpoints
Simulated Random Breakpoints
Insulator-Breakpoint Matching Rate
60%
50%
40%
30%
20%
10%
0%
D.eug
D.bia
D.ana
D.bip
D.pse
D.per
D.wil
D.moj
D.vir

Figure 3.6 Permutation test of matching between breakpoints and insulator sites. (A) Permutation test of breakpoints matching insulator sites using simulated intergenic breakpoints. The blue bars represent the matching rate of evolutionary breakpoints with insulator sites within 1kb range from the breakpoint center. The orange bars represent the average matching rate from the simulated intergenic breakpoints. (B) Permutation test of rearrangements with only single end matching insulator sites. Blue bars represent the matching rate of evolutionary breakpoints with insulators sites within a 1kb range from the breakpoint center. Orange bars represent the average matching rate from the randomly simulated breakpoints. (C) Permutation tests of rearrangements with both ends matching insulator sites. Blue bars represent the matching rate of evolutionary breakpoints. Orange bars represent the average matching rate of randomly simulated samples. (A-C) Error bars represent the standard deviation.

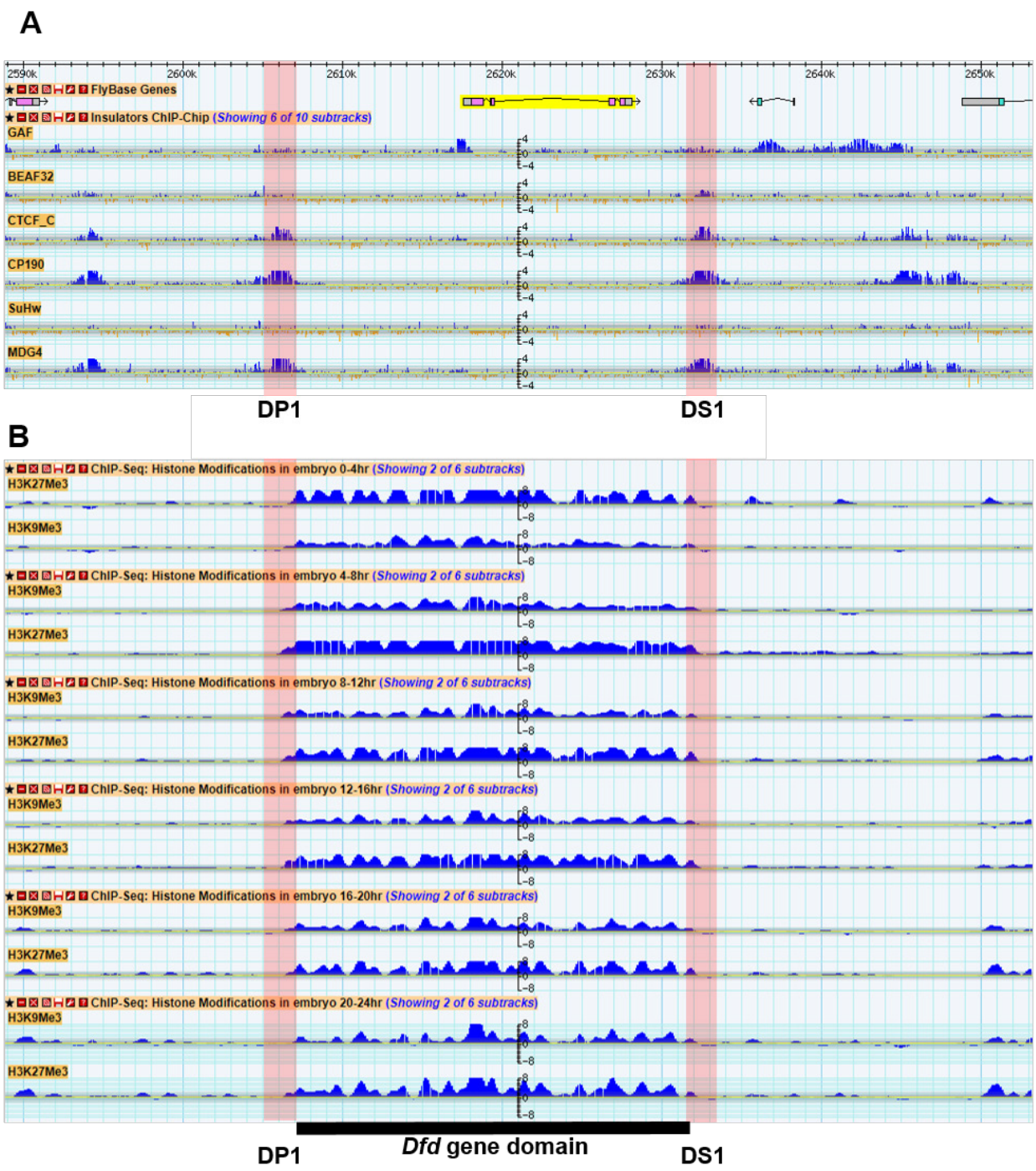

A
2590k
2600k
2610k
2620k
2630k
2640k
2650k
FlyBase Genes
Insulators ChIP-Chip (Showing 6 of 10 subtracks)
GAF
BEAF32
CTCF_C
CP190
SuHw
MDG4
DP1
DS1
B
ChIP-Seq: Histone Modifications in embryo 0-4hr (Showing 2 of 6 subtracks)
H3K27Me3
H3K9Me3
ChIP-Seq: Histone Modifications in embryo 4-8hr (Showing 2 of 6 subtracks)
H3K9Me3
H3K27Me3
ChIP-Seq: Histone Modifications in embryo 8-12hr (Showing 2 of 6 subtracks)
H3K9Me3
H3K27Me3
ChIP-Seq: Histone Modifications in embryo 12-16hr (Showing 2 of 6 subtracks)
H3K9Me3
H3K27Me3
ChIP-Seq: Histone Modifications in embryo 16-20hr (Showing 2 of 6 subtracks)
H3K9Me3
H3K27Me3
ChIP-Seq: Histone Modifications in embryo 20-24hr (Showing 2 of 6 subtracks)
H3K9Me3
H3K27Me3
DP1
Dfd gene domain
DS1

Figure S3.1 DP1 and DS1 insulators demarcate the boundaries Dfd gene domain. (A) Screenshot of the binding profile of the *Drosophila* classic insulator protein factors in the *Dfd* regioin from modENCODE GBroswer (http://gbrowse.modencode.org/fgb2/gbrowse/fly/) (Negre et al., 2010). (B) Screenshot of the repressive histone modification ChIP-seq profile in *Dfd* region throughout differnet developmental stages (http://gbrowse.modencode.org/fgb2/gbrowse/fly/). Vertical shaded red boxes in (A) and (B) mark the location of DP1 and DS1 insulators. Black line at the bottom marks the complete *Dfd* gene domain.

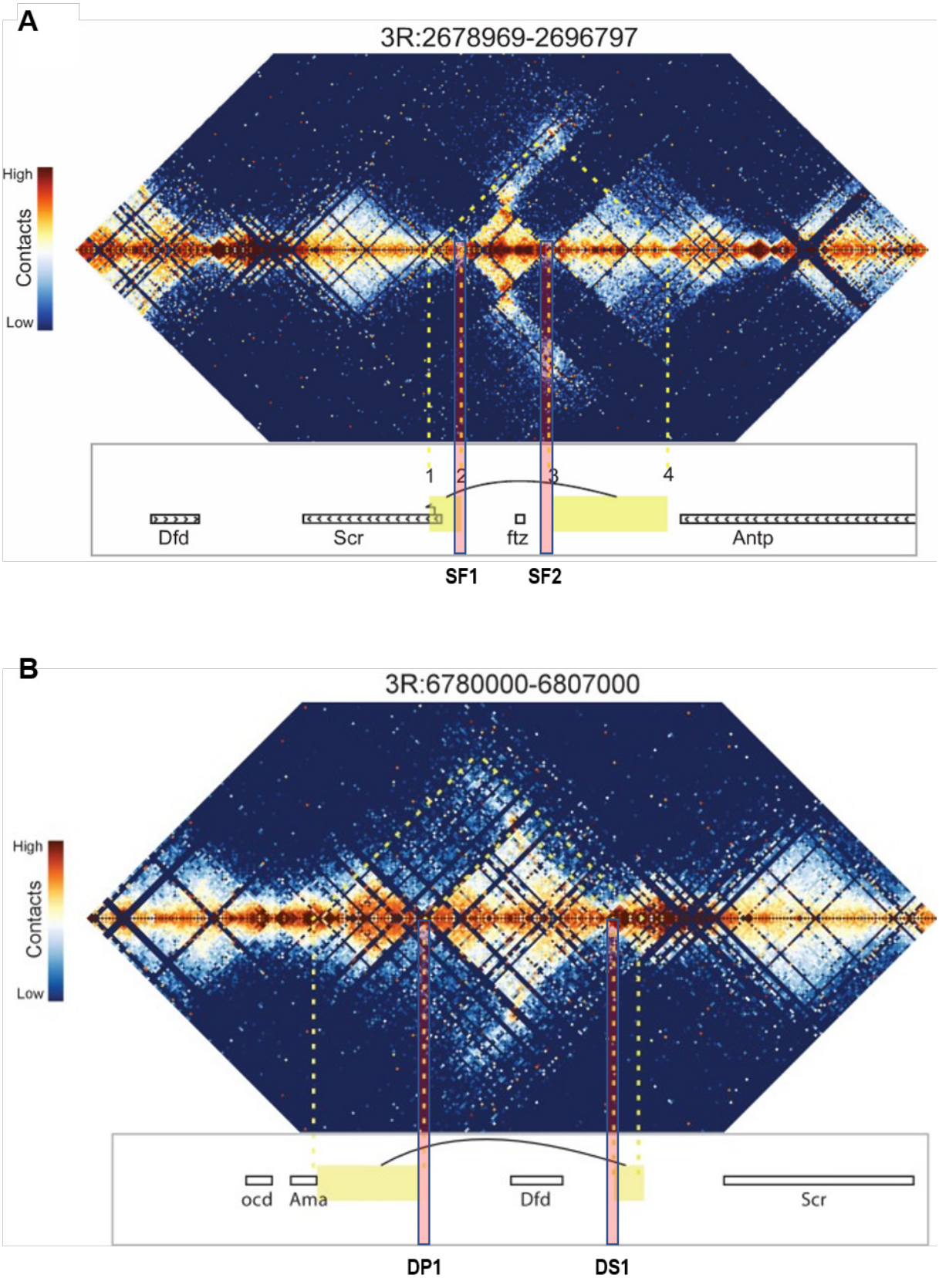
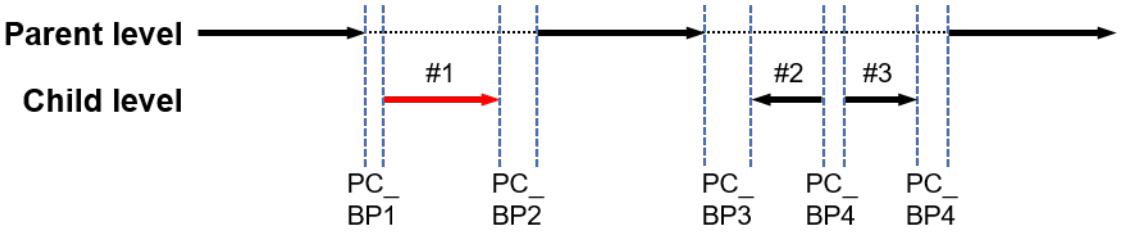

A
3R:2678969-2696797
High
Contacts
Low
1
2
3
4
Dfd
Scr
ftz
Antp
SF1
SF2
B
3R:6780000-6807000
High
Contacts
Low
ocd
Ama
Dfd
Scr
DP1
DS1

Figure S3.2. Long- range interactions observed in the *ftz* and *Dfd* regions in early stage embryo HiC. (A) HiC interaction map showing the Scr-Antp region. Strong long-range interaction forms between SF1 and SF2 insulators. 3' end of Antp region (our R9-10 region described in chapter2) also shows extensive interactions with SF1 and Scr promoter region. The number 1 to 4 marks the four anchors of strong crossing-TAD long-range loop interactions. Number 2 and 3 are SF1 and SF2, respectively. (B) HiC interaction map showing the *Dfd* locus. DP1 and DS1 form strong interactions at the borders of the *Dfd* gene domain. Vertical red shaded box marks the location of insulators on the HiC map. This HiC map is adopted from Stadler et al's publication on *elife* with a Creative Commons Attribution license (Stadler et al., 2017).


Parent level
Child level
#1
#2
#3
PC_ BP1
PC_ BP2
PC_ BP3
PC_ BP4
PC_ BP4

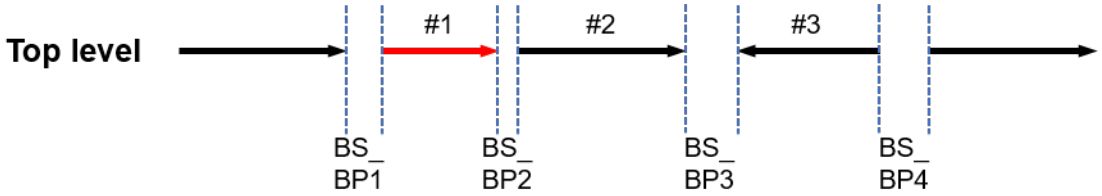
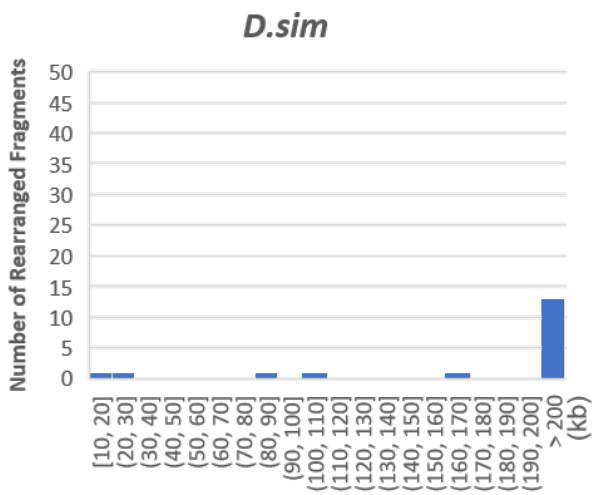
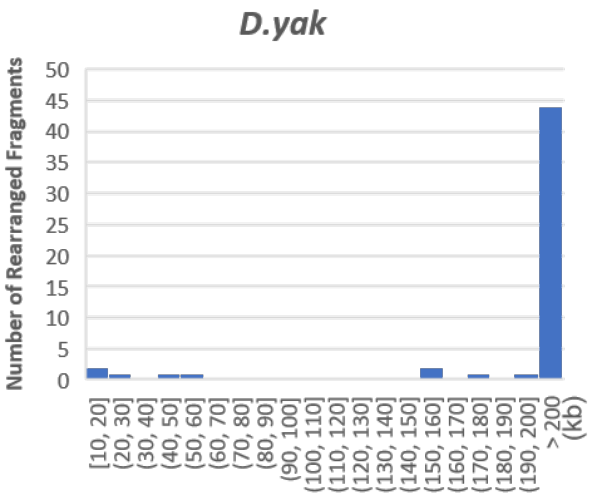
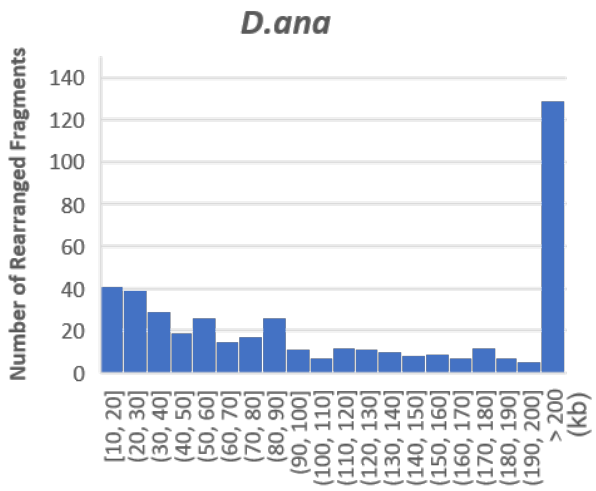
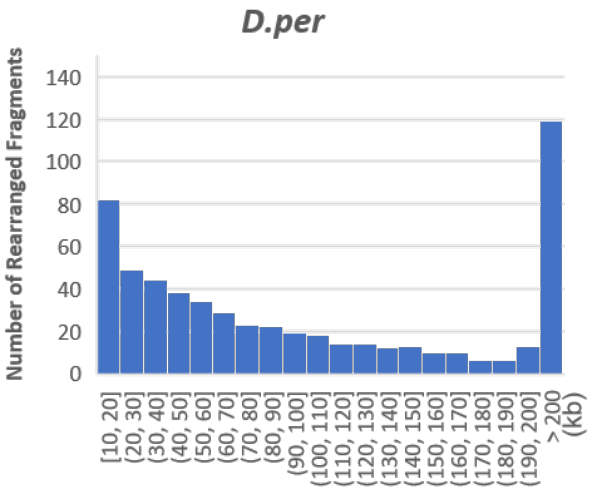

Top level
#1
#2
#3
BS_ BP1
BS_ BP2
BS_ BP3
BS_ BP4

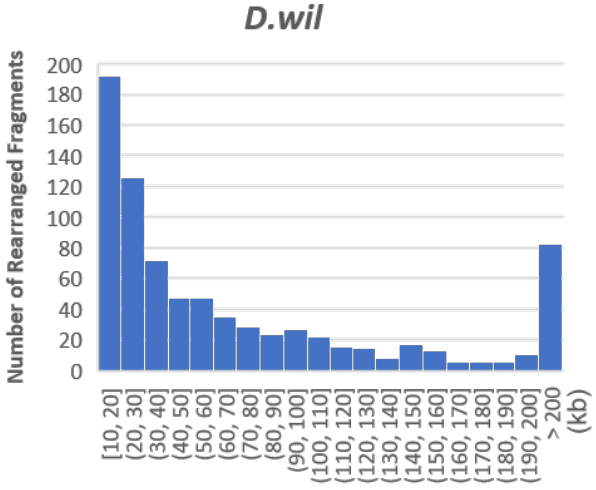
Figure S3.3 Illustration of breakpoints extraction strategy. (A) Diagram illustrating the parent-child breakpoint finding method. Consecutive alignment blocks (fills) are represented by horizontal arrows. Parent level means alignment fills with higher score. Gaps in the parent level fill are represented by the dashed horizontal line in between the black arrows. Arrows at the child level are alignment blocks that map to the gap of parent level fill in the target genome (*Drosophila melanogaster*). Regions between the gap edges and child level fill edges represent the breakpoints, which are marked with the blue vertical dashed lines. PC_BP represent breakpoints identified through the parent-child finding method. Different chromosomes (scaffolds) are marked with different colors. #1 to #3 represent three different types of rearrangements: #1 external translocation, #2 inversion or internal translocation, #3 internal translocation (See materials and methods for classification details). (B) Diagram illustrating the between-siblings breakpoint finding method. Fills at the top level usually represents the highest score alignment blocks. Gap regions between neighboring top-level fills is extracted as breakpoints. Gaps from two sides of a fill are reported as a pair. #1 to #3 represent three different types of rearrangements: #1, external translocation and #2, 3, internal translocations. (See materials and methods for classification details).


D.sim
Number of Rearranged Fragments
0
5
10
15
20
25
30
35
40
45
50
[10, 20]
(20, 30]
(30, 40]
(40, 50]
(50, 60]
(60, 70]
(70, 80]
(80, 90]
(90, 100]
(100, 110]
(110, 120]
(120, 130]
(130, 140]
(140, 150]
(150, 160]
(160, 170]
(170, 180]
(180, 190]
(190, 200]
> 200
(kb)


D.yak
Number of Rearranged Fragments
0
5
10
15
20
25
30
35
40
45
50
[10, 20]
(20, 30]
(30, 40]
(40, 50]
(50, 60]
(60, 70]
(70, 80]
(80, 90]
(90, 100]
(100, 110]
(110, 120]
(120, 130]
(130, 140]
(140, 150]
(150, 160]
(160, 170]
(170, 180]
(180, 190]
(190, 200]
> 200
(kb)


D.ana
Number of Rearranged Fragments
0
20
40
60
80
100
120
140
[10, 20]
(20, 30]
(30, 40]
(40, 50]
(50, 60]
(60, 70]
(70, 80]
(80, 90]
(90, 100]
(100, 110]
(110, 120]
(120, 130]
(130, 140]
(140, 150]
(150, 160]
(160, 170]
(170, 180]
(180, 190]
(190, 200]
> 200
(kb)


D.per
Number of Rearranged Fragments
0
20
40
60
80
100
120
140
[10, 20]
(20, 30]
(30, 40]
(40, 50]
(50, 60]
(60, 70]
(70, 80]
(80, 90]
(90, 100]
(100, 110]
(110, 120]
(120, 130]
(130, 140]
(140, 150]
(150, 160]
(160, 170]
(170, 180]
(180, 190]
(190, 200]
> 200
(kb)


D.wil
Number of Rearranged Fragments
0
20
40
60
80
100
120
140
160
180
200
[10, 20]
(20, 30]
(30, 40]
(40, 50]
(50, 60]
(60, 70]
(70, 80]
(80, 90]
(90, 100]
(100, 110]
(110, 120]
(120, 130]
(130, 140]
(140, 150]
(150, 160]
(160, 170]
(170, 180]
(180, 190]
(190, 200]
> 200
(kb)

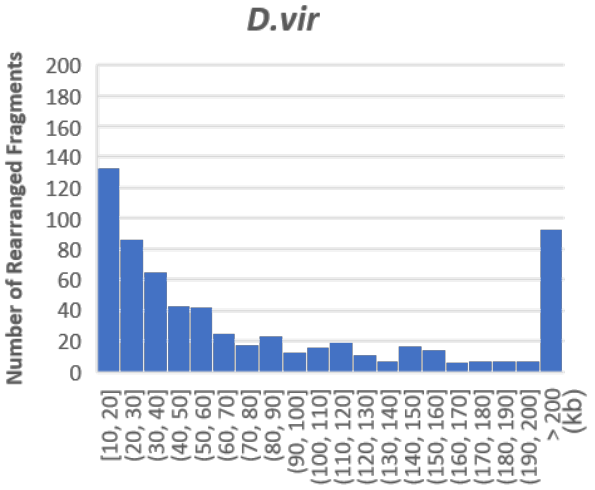

D.vir
Number of Rearranged Fragments
0
20
40
60
80
100
120
140
160
180
200
[10, 20]
(20, 30]
(30, 40]
(40, 50]
(50, 60]
(60, 70]
(70, 80]
(80, 90]
(90, 100]
(100, 110]
(110, 120]
(120, 130]
(130, 140]
(140, 150]
(150, 160]
(160, 170]
(170, 180]
(180, 190]
(190, 200]
> 200
(kb)

Figure S3.4 Size distribution of identified rearrangements in genus of *Drosophila*. Six species were selected to represent closely and distantly related *Drosophila* species to *Drosophila melanogaster* (refer to Fig 3.2). Total rearrangements (after the 10kb minimum size filter) identified between *Drosophila melanogaster* and the representative species (marked in the title section of each plot) were plot using histogram, with bin size of 20kb. Y axis represents the count of rearrangement. X axis represents the fragment size groups in kb.

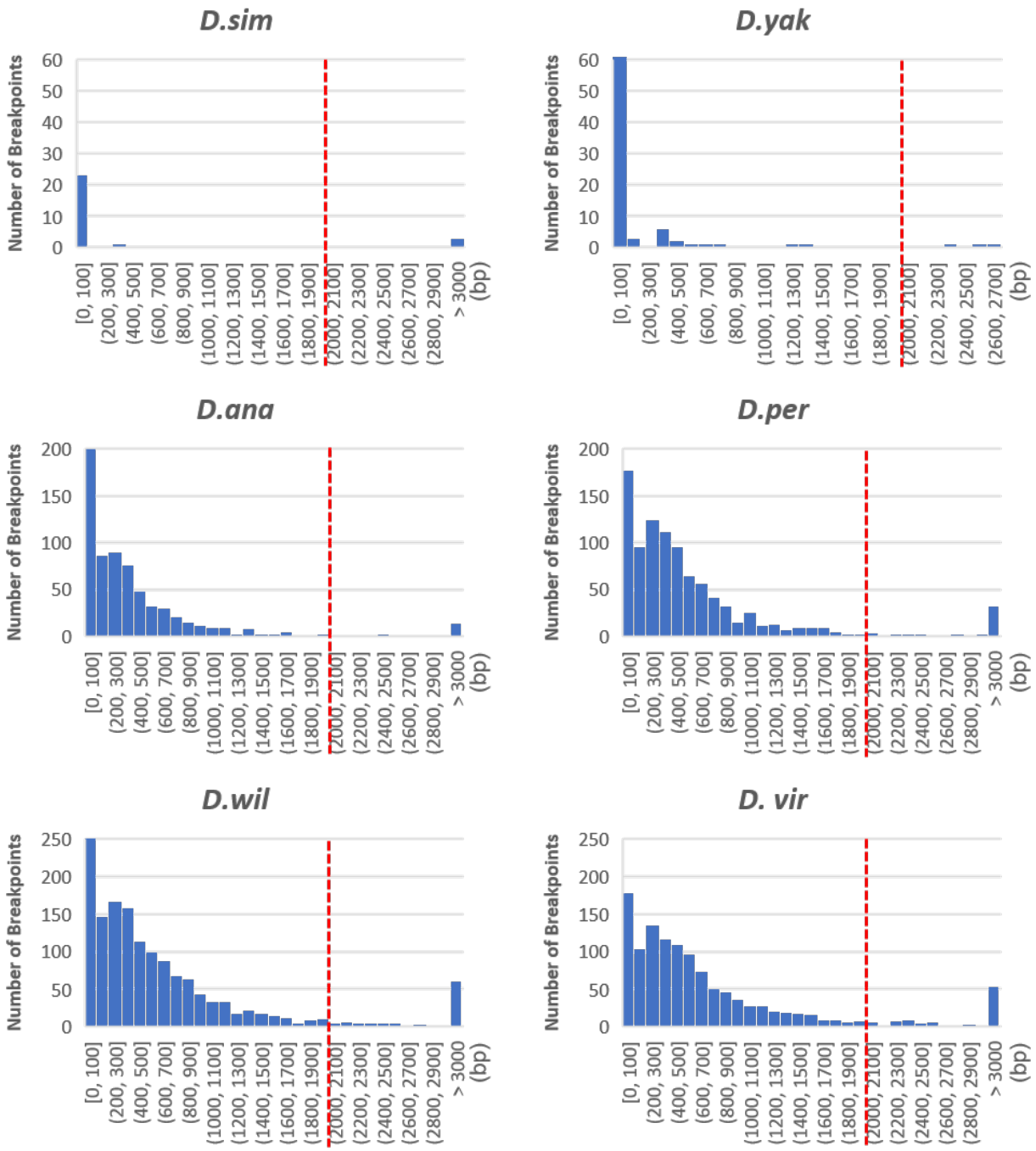

D.sim
Number of Breakpoints
60
50
40
30
20
10
0
[0, 100]
(200, 300]
(400, 500]
(600, 700]
(800, 900]
(1000, 1100]
(1200, 1300]
(1400, 1500]
(1600, 1700]
(1800, 1900]
(2000, 2100]
(2200, 2300]
(2400, 2500]
(2600, 2700]
(2800, 2900]
> 3000
(bp)
D.yak
Number of Breakpoints
60
50
40
30
20
10
0
[0, 100]
(200, 300]
(400, 500]
(600, 700]
(800, 900]
(1000, 1100]
(1200, 1300]
(1400, 1500]
(1600, 1700]
(1800, 1900]
(2000, 2100]
(2200, 2300]
(2400, 2500]
(2600, 2700]
(bp)
D.ana
Number of Breakpoints
200
150
100
50
0
[0, 100]
(200, 300]
(400, 500]
(600, 700]
(800, 900]
(1000, 1100]
(1200, 1300]
(1400, 1500]
(1600, 1700]
(1800, 1900]
(2000, 2100]
(2200, 2300]
(2400, 2500]
(2600, 2700]
(2800, 2900]
> 3000
(bp)
D.per
Number of Breakpoints
200
150
100
50
0
[0, 100]
(200, 300]
(400, 500]
(600, 700]
(800, 900]
(1000, 1100]
(1200, 1300]
(1400, 1500]
(1600, 1700]
(1800, 1900]
(2000, 2100]
(2200, 2300]
(2400, 2500]
(2600, 2700]
(2800, 2900]
> 3000
(bp)
D.wil
Number of Breakpoints
250
200
150
100
50
0
[0, 100]
(200, 300]
(400, 500]
(600, 700]
(800, 900]
(1000, 1100]
(1200, 1300]
(1400, 1500]
(1600, 1700]
(1800, 1900]
(2000, 2100]
(2200, 2300]
(2400, 2500]
(2600, 2700]
(2800, 2900]
> 3000
(bp)
D. vir
Number of Breakpoints
250
200
150
100
50
0
[0, 100]
(200, 300]
(400, 500]
(600, 700]
(800, 900]
(1000, 1100]
(1200, 1300]
(1400, 1500]
(1600, 1700]
(1800, 1900]
(2000, 2100]
(2200, 2300]
(2400, 2500]
(2600, 2700]
(2800, 2900]
> 3000
(bp)

Figure S3.5 Size distribution of identified breakpoints in genus of *Drosophila*. Six species were selected to represent closely and distantly related *Drosophila* species to *Drosophila melanogaster* (Refer to Fig 3.2). The breakpoints identified between *Drosophila melanogaster* and the representative species (marked in the title section of each plot) were plot using a histogram with bin size of 100bp. Y axis represent the count of breakpoint in each bin. X axis represent the breakpoints sizes in bp. The red vertical dashed line marks the 2kb maximum size cutoff applied before any downstream analysis.

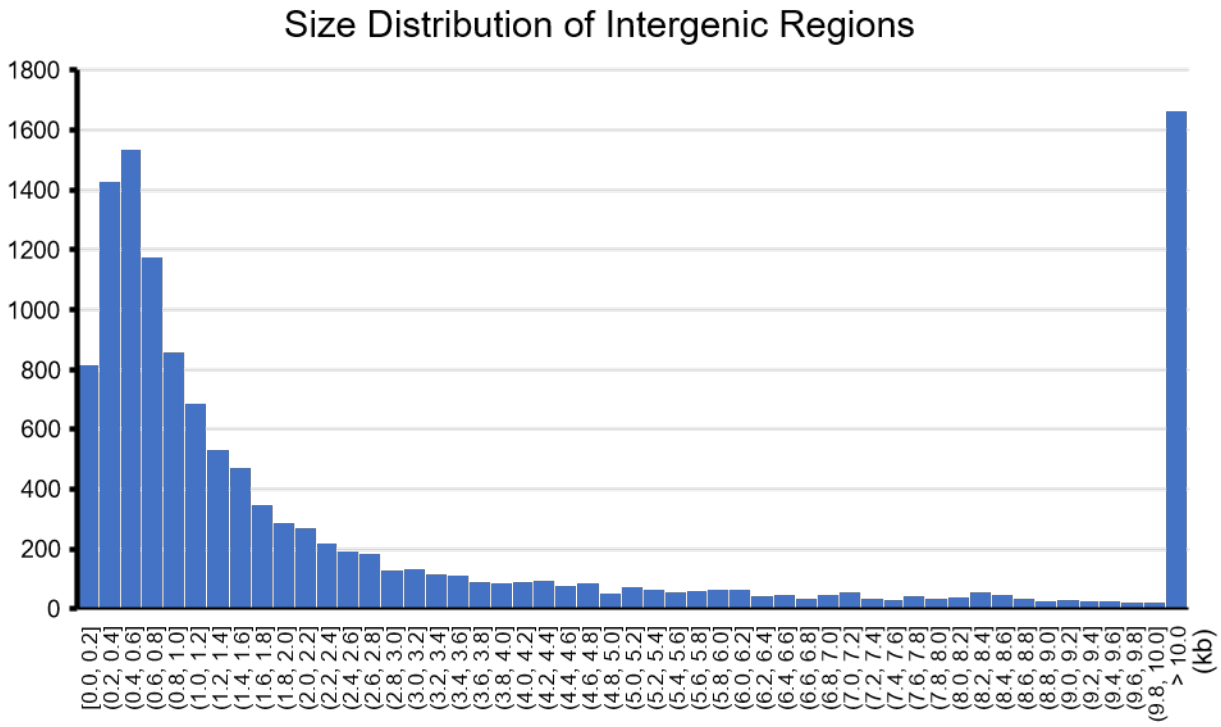

Size Distribution of Intergenic Regions
1800
1600
1400
1200
1000
800
600
400
200
0
[0.0, 0.2]
(0.2, 0.4]
(0.4, 0.6]
(0.6, 0.8]
(0.8, 1.0]
(1.0, 1.2]
(1.2, 1.4]
(1.4, 1.6]
(1.6, 1.8]
(1.8, 2.0]
(2.0, 2.2]
(2.2, 2.4]
(2.4, 2.6]
(2.6, 2.8]
(2.8, 3.0]
(3.0, 3.2]
(3.2, 3.4]
(3.4, 3.6]
(3.6, 3.8]
(3.8, 4.0]
(4.0, 4.2]
(4.2, 4.4]
(4.4, 4.6]
(4.6, 4.8]
(4.8, 5.0]
(5.0, 5.2]
(5.2, 5.4]
(5.4, 5.6]
(5.6, 5.8]
(5.8, 6.0]
(6.0, 6.2]
(6.2, 6.4]
(6.4, 6.6]
(6.6, 6.8]
(6.8, 7.0]
(7.0, 7.2]
(7.2, 7.4]
(7.4, 7.6]
(7.6, 7.8]
(7.8, 8.0]
(8.0, 8.2]
(8.2, 8.4]
(8.4, 8.6]
(8.6, 8.8]
(8.8, 9.0]
(9.0, 9.2]
(9.2, 9.4]
(9.4, 9.6]
(9.6, 9.8]
(9.8, 10.0]
> 10.0
(kb)

Figure S3.6 Size distribution of the intergenic regions in *Drosophila melanogaster*. The distribution of intergenic regions is plot in a histogram. Intergenic regions were extracted from the Ensemble genome annotation file using CDS (see materials and methods). Y-axis represents the counts. X-axis represents the size groups in unit of kb.

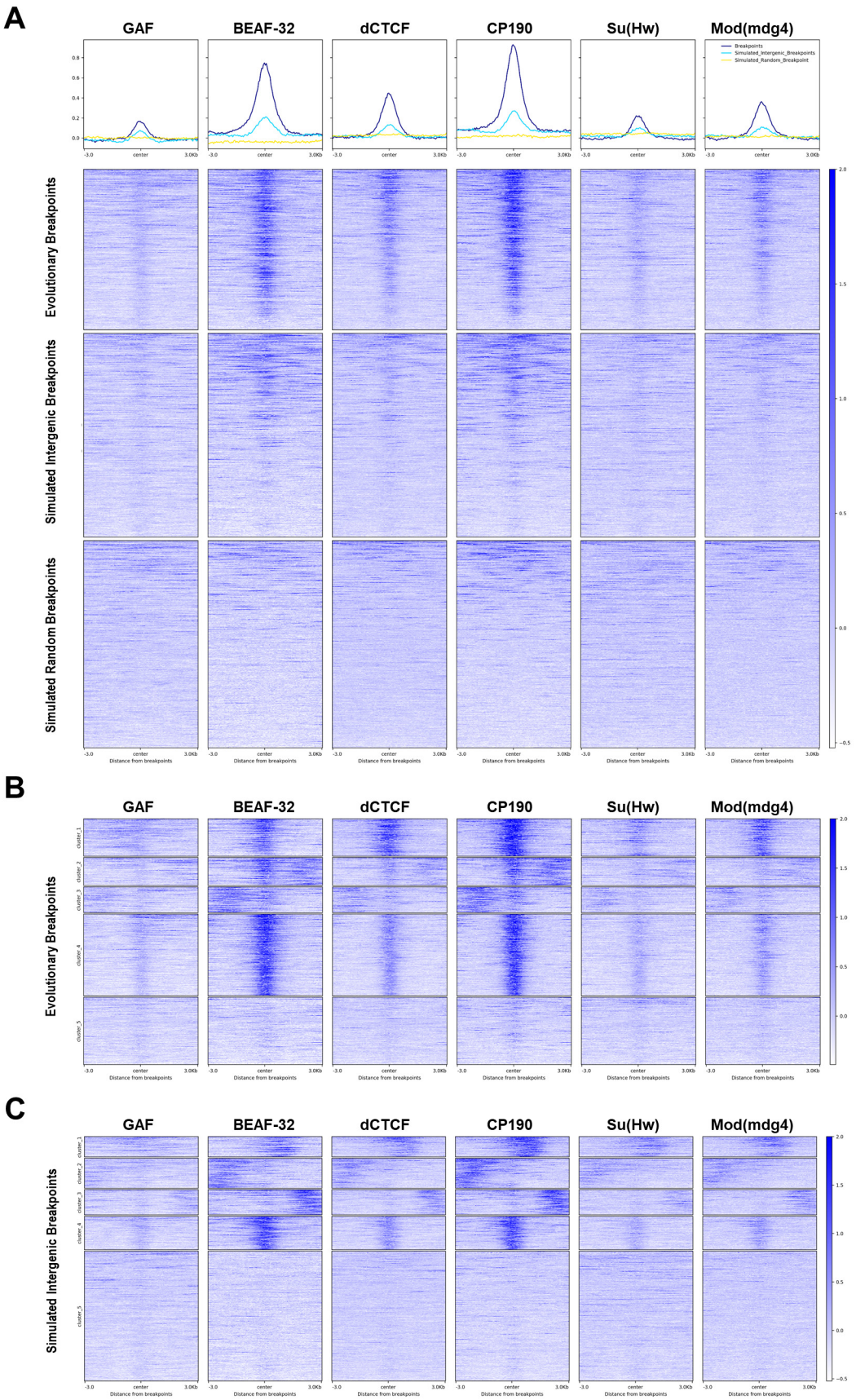

A
GAF
BEAF-32
dCTCF
CP190
Su(Hw)
Mod(mdg4)
Breakpoints
Simulated_Intergenic_Breakpoints
Simulated_Random_Breakpoint
Evolutionary Breakpoints
Simulated Intergenic Breakpoints
Simulated Random Breakpoints
-3.0
center
3.0Kb
Distance from breakpoints
B
Evolutionary Breakpoints
cluster_1
cluster_2
cluster_3
cluster_4
cluster_5
C
Simulated Intergenic Breakpoints

Figure S3.7 Chromatin insulators are enriched at evolutionary breakpoints in *Drosophila pseudoobscura*. (A) Heatmap of chromatin insulator protein enrichment around the center of breakpoints identified between *Drosophila melanogaster* and *Drosophila pseudoobscura*. Top profile plot represents the average signal around the center of all breakpoints. Lines in different colors represent different groups of breakpoints: evolutionary breakpoints, simulated intergenic breakpoints, and simulated random breakpoints (see materials and methods). Regions in each group is sorted on a descending order of the mean signal intensity. Signal intensity is represented by the darkness of the blue color. (B-C) Heatmap of chromatin insulator protein enrichment around the center of breakpoints by using kmeans clustering analysis with kmeans=5. (B) Represents the clustering analysis at evolutionary breakpoints. (C) Represents the clustering analysis at the simulated intergenic breakpoints with kmeans=5.

REFERENCES

Acemel, R.D., Maeso, I., and Gomez-Skarmeta, J.L. (2017). Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. Wiley Interdiscip Rev Dev Biol *6*.

Acemel, R.D., Tena, J.J., Irastorza-Azcarate, I., Marletaz, F., Gomez-Marin, C., de la Calle-Mustienes, E., Bertrand, S., Diaz, S.G., Aldea, D., Aury, J.M.*, et al.* (2016). A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. Nat Genet *48*, 336-341.

Ahanger, S.H., Gunther, K., Weth, O., Bartkuhn, M., Bhonde, R.R., Shouche, Y.S., and Renkawitz, R. (2014). Ectopically tethered CP190 induces large-scale chromatin decondensation. Sci Rep *4*, 3917.

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., and Bazett-Jones, D.P. (2010). Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PLoS One *5*, e10531.

Ali, T., Renkawitz, R., and Bartkuhn, M. (2016). Insulators and domains of gene expression. Curr Opin Genet Dev *37*, 17-26.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol *16*, 155-166.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. Dev Cell *16*, 47-57.

Andraszek, K., and Smalec, E. (2011). Structure and functions of lampbrush chromosomes. Termedia Publishing House.

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. (2013). A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science *340*, 1234167.

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

Aoki, T., Schweinsberg, S., Manasson, J., and Schedl, P. (2008). A stage-specific factor confers Fab-7 boundary activity during early embryogenesis in Drosophila. Mol Cell Biol *28*, 1047-1060.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell *144*, 214-226.

Bartkuhn, M., Straub, T., Herold, M., Herrmann, M., Rathke, C., Saumweber, H., Gilfillan, G.D., Becker, P.B., and Renkawitz, R. (2009). Active promoters and insulators are marked by the centrosomal protein 190. EMBO J *28*, 877-888.

Barutcu, A.R., Lian, J.B., Stein, J.L., Stein, G.S., and Imbalzano, A.N. (2017). The connection between BRG1, CTCF and topoisomerases at TAD boundaries. Nucleus *8*, 150-155.

Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C.V., Gillis, D.R., and Phillips-Cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res *27*, 1139-1152.

Beagan, J.A., Gilgenast, T.G., Kim, J., Plona, Z., Norton, H.K., Hu, G., Hsu, S.C., Shields, E.J., Lyu, X., Apostolou, E.*, et al.* (2016). Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. Cell Stem Cell *18*, 611-624.

Belmont, A.S. (2014). Large-scale chromatin organization: the good, the surprising, and the still perplexing. Curr Opin Cell Biol *26*, 69-78.

Belozerov, V.E., Majumder, P., Shen, P., and Cai, H.N. (2003). A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of Drosophila. EMBO J *22*, 3113-3121.

Bergstrom, R., Savary, K., Moren, A., Guibert, S., Heldin, C.H., Ohlsson, R., and Moustakas, A. (2010). Transforming growth factor beta promotes complexes between Smad proteins and the CCCTC-binding factor on the H19 imprinting control region chromatin. J Biol Chem *285*, 19727-19737.

Berthelot, C., Muffato, M., Abecassis, J., and Roest Crollius, H. (2015). The 3D organization of chromatin explains evolutionary fragile genomic regions. Cell Rep *10*, 1913-1924.

Bhutkar, A., Schaeffer, S.W., Russo, S.M., Xu, M., Smith, T.F., and Gelbart, W.M. (2008). Chromosomal Rearrangement Inferred From Comparisons of 12 Drosophila Genomes. Genetics *179*, 1657-1680.

Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. Cell *152*, 1270-1284.

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

Boettiger, A.N., and Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science *325*, 471-473.

Bohla, D., Herold, M., Panzer, I., Buxa, M.K., Ali, T., Demmers, J., Kruger, M., Scharfe, M., Jarek, M., Bartkuhn, M.*, et al.* (2014). A functional insulator screen identifies NURF and dREAM components to be required for enhancer-blocking. PLoS One *9*, e107765.

Boisvert, F.M., van Koningsbruggen, S., Navascues, J., and Lamond, A.I. (2007). The multifunctional nucleolus. Nat Rev Mol Cell Biol *8*, 574-585.

Bonchuk, A., Denisov, S., Georgiev, P., and Maksimenko, O. (2011). Drosophila BTB/POZ domains of "ttk group" can form multimers and selectively interact with each other. J Mol Biol *412*, 423-436.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat Rev Genet *17*, 661-678.

Boros, I.M. (2012). Histone modification in Drosophila. Brief Funct Genomics *11*, 319-331.

Bratkowski, M., Unarta, I.C., Zhu, L., Shubbar, M., Huang, X., and Liu, X. (2018). Structural dissection of an interaction between transcription initiation and termination factors implicated in promoter-terminator cross-talk. J Biol Chem *293*, 1651-1665.

Bushey, A.M., Ramos, E., and Corces, V.G. (2009). Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. Genes Dev *23*, 1338-1350.

Cai, H.N., and Levine, M. (1997). The gypsy insulator can function as a promoter-specific silencer in the Drosophila embryo. The EMBO Journal *16*, 1732-1741.

Cai, H.N., and Shen, P. (2001). Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. Science *291*, 493-495.

Cai, H.N., Zhang, Z., Adams, J.R., and Shen, P. (2001). Genomic context modulates insulator activity through promoter competition. Development (Cambridge, England) *128*, 4339-4347.

Calhoun, V.C., and Levine, M. (2003). 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.

Calhoun, V.C., Stathopoulos, A., and Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. Proceedings of the National Academy of Sciences *99*, 9243-9247.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.

Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., Kieffer-Kwon, K.R., Pekowska, A., Zhang, H., Rao, S.S.P.*, et al.* (2017). Genome Organization Drives Chromosome Fragility. Cell *170*, 507-521 e518.

Capelson, M., and Corces, V.G. (2006). SUMO conjugation attenuates the activity of the gypsy chromatin insulator. EMBO J *25*, 1906-1914.

Carelli, F.N., Sharma, G., and Ahringer, J. (2017). Broad Chromatin Domains: An Important Facet of Genome Regulation. Bioessays *39*.

Casas-Delucchi, C.S., Becker, A., Bolius, J.J., and Cardoso, M.C. (2012). Targeted manipulation of heterochromatin rescues MeCP2 Rett mutants and re-establishes higher order chromatin organization. Nucleic Acids Res *40*, e176.

Chen, B., Zhang, Y., Zhang, X., Jia, S., Chen, S., and Kang, L. (2016). Genome-wide identification and developmental expression profiling of long noncoding RNAs during Drosophila metamorphosis. Sci Rep *6*, 23330.

Chen, H., Tian, Y., Shu, W., Bo, X., and Wang, S. (2012). Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. PLoS One *7*, e41374.

Chintapalli, V.R., Wang, J., and Dow, J.A.T. (2007). Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nature Genetics *39*, 715.

Clemson, C.M. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. The Journal of Cell Biology *132*, 259-275.

Cohen, A.L., and Jia, S. (2014). Noncoding RNAs and the borders of heterochromatin. Wiley Interdiscip Rev RNA *5*, 835-847.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature *523*, 240-244.

Cremer, T., and Cremer, M. (2010). Chromosome territories. Cold Spring Harb Perspect Biol *2*, a003889.

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. J Cell Biol *145*, 1119-1131.

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

Cubenas-Potts, C., and Corces, V.G. (2015). Architectural proteins, transcription, and the three-dimensional organization of the genome. FEBS Lett *589*, 2923-2930.

Cubenas-Potts, C., Rowley, M.J., Lyu, X., Li, G., Lei, E.P., and Corces, V.G. (2017). Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. Nucleic Acids Res *45*, 1714-1730.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res *19*, 24-32.

Dekker, J. (2008). Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. J Biol Chem *283*, 34532-34540.

Del Prete, S., Mikulski, P., Schubert, D., and Gaudin, V. (2015). One, Two, Three: Polycomb Proteins Hit All Dimensions of Gene Regulation. Genes (Basel) *6*, 520-542.

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A.*, et al.* (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell *158*, 849-860.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell *13*, 602-616.

Dernburg, A.F. (2011). In Situ Hybridization to Somatic Chromosomes in Drosophila. Cold Spring Harbor Protocols *2011*.

Di Croce, L., and Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. Nat Struct Mol Biol *20*, 1147-1155.

Dileep, V., Ay, F., Sima, J., Vera, D.L., Noble, W.S., and Gilbert, D.M. (2015). Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program. Genome Res *25*, 1104-1113.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W.*, et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331-336.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Doe, C.Q., Hiromi, Y., Gehring, W.J., and Goodman, C.S. (1988). Expression and function of the segmentation gene fushi tarazu during Drosophila neurogenesis. Science *239*, 170-175.

Donze, D., Adams, C.R., Rine, J., and Kamakaka, R.T. (1999). The boundaries of the silenced HMR domain in Saccharomyces cerevisiae. Genes Dev *13*, 698-708.

Donze, D., and Kamakaka, R.T. (2001). RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. EMBO J *20*, 520-531.

Drier, Y., Lawrence, M.S., Carter, S.L., Stewart, C., Gabriel, S.B., Lander, E.S., Meyerson, M., Beroukhim, R., and Getz, G. (2013). Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. Genome Research *23*, 228-235.

Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev *18*, 2485-2490.

Durand, Neva C., Robinson, James T., Shamim, Muhammad S., Machol, I., Mesirov, Jill P., Lander, Eric S., and Aiden, Erez L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Systems *3*, 99-101.

Eagen, K.P., Hartl, T.A., and Kornberg, R.D. (2015). Stable Chromosome Condensation Revealed by Chromosome Conformation Capture. Cell *163*, 934-946.

Ebersole, T., Kim, J.H., Samoshkin, A., Kouprina, N., Pavlicek, A., White, R.J., and Larionov, V. (2011). tRNA genes protect a reporter gene from epigenetic silencing in mouse cells. Cell Cycle *10*, 2779-2791.

Eltsov, M., Maclellan, K.M., Maeshima, K., Frangakis, A.S., and Dubochet, J. (2008). Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. Proc Natl Acad Sci U S A *105*, 19732-19737.

Entrevan, M., Schuettengruber, B., and Cavalli, G. (2016). Regulation of Genome Architecture and Function by Polycomb Proteins. Trends Cell Biol *26*, 511-525.

Erokhin, M., Davydova, A., Kyrchanova, O., Parshikov, A., Georgiev, P., and Chetverina, D. (2011). Insulators form gene loops by interacting with promoters in Drosophila. Development *138*, 4097-4106.

Eser, U., Chandler-Brown, D., Ay, F., Straight, A.F., Duan, Z., Noble, W.S., and Skotheim, J.M. (2017). Form and function of topologically associating genomic domains in budding yeast. Proc Natl Acad Sci U S A *114*, E3061-E3070.

Evans, K.J., Huang, N., Stempor, P., Chesney, M.A., Down, T.A., and Ahringer, J. (2016). Stable Caenorhabditis elegans chromatin domains separate broadly expressed and developmentally regulated genes. Proc Natl Acad Sci U S A *113*.

Fabre, E., and Zimmer, C. (2018). From dynamic chromatin architecture to DNA damage repair and back. Nucleus *9*, 161-170.

Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E. (2014). Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. Mol Cell *55*, 694-707.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suva, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110-114.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep *15*, 2038-2049.

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

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

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

Fullwood, M.J., and Ruan, Y. (2009). ChIP-based methods for the identification of long-range chromatin interactions. J Cell Biochem *107*, 30-39.

Fussner, E., Strauss, M., Djuric, U., Li, R., Ahmed, K., Hart, M., Ellis, J., and Bazett-Jones, D.P. (2012). Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. EMBO Rep *13*, 992-996.

Gabdank, I., Ramakrishnan, S., Villeneuve, A.M., and Fire, A.Z. (2016). A streamlined tethered chromosome conformation capture protocol. BMC Genomics *17*, 274.

Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. Science *360*, 102-105.

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

Ghirlando, R., and Felsenfeld, G. (2016). CTCF: making the right connections. Genes Dev *30*, 881-891.

Gindhart, J.G., Jr., King, A.N., and Kaufman, T.C. (1995). Characterization of the cis-regulatory region of the Drosophila homeotic gene Sex combs reduced. Genetics *139*, 781-795.

Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell *157*, 950-963.

Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E.*, et al.* (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*, 575-579.

Giorgetti, L., Servant, N., and Heard, E. (2013). Changes in the organization of the genome during the mammalian cell cycle. Genome Biol *14*, 142.

Golovnin, A., Volkov, I., and Georgiev, P. (2012). SUMO conjugation is required for the assembly of Drosophila Su(Hw) and Mod(mdg4) into insulator bodies that facilitate insulator complex formation. J Cell Sci *125*, 2064-2074.

Gong, Y., Lazaris, C., Sakellaropoulos, T., Lozano, A., Kambadur, P., Ntziachristos, P., Aifantis, I., and Tsirigos, A. (2018). Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. Nat Commun *9*, 542.

Gordon, J.L., Byrne, K.P., and Wolfe, K.H. (2009). Additions, Losses, and Rearrangements on the Evolutionary Route from a Reconstructed Ancestor to the Modern Saccharomyces cerevisiae Genome. PLoS Genetics *5*.

Gostissa, M., Alt, F.W., and Chiarle, R. (2011). Mechanisms that Promote and Suppress Chromosomal Translocations in Lymphocytes. Immunology *29*, 319-350.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W.*, et al.* (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473.

Gregg, T., and Matthew, H. (2017). Drosophila 25 species phylogeny.

Grob, S., Schmid, M.W., and Grossniklaus, U. (2014). Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Mol Cell *55*, 678-693.

Grubert, F., Zaugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A.*, et al.* (2015). Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. Cell *162*, 1051-1065.

Guastafierro, T., Catizone, A., Calabrese, R., Zampieri, M., Martella, O., Bacalini, M.G., Reale, A., Di Girolamo, M., Miccheli, M., Farrar, D.*, et al.* (2013). ADP-ribose polymer depletion leads to nuclear Ctcf re-localization and chromatin rearrangement(1). Biochem J *449*, 623-630.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y.*, et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900-910.

Gurudatta, B.V., and Corces, V.G. (2009). Chromatin insulators: lessons from the fly. Brief Funct Genomic Proteomic *8*, 276-282.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339-346.

Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yanez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B.*, et al.* (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell *169*, 693-707 e614.

Haeusler, R.A., Pratt-Hyatt, M., Good, P.D., Gipson, T.A., and Engelke, D.R. (2008). Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes. Genes Dev *22*, 2204-2214.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F.*, et al.* (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet *43*, 630-638.

Hao, N., Shearwin, K.E., and Dodd, I.B. (2017). Programmable DNA looping using engineered bivalent dCas9 complexes. Nat Commun *8*, 1628.

Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merkenschlager, M., and Lenhard, B. (2017). Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. Nat Commun *8*, 441.

Harris, R.S. (2007). Improved pairwise alignment of genomic dna (Pennsylvania State University), pp. 84.

Hinsch, H., and Hannenhalli, S. (2006). Recurring genomic breaks in independent lineages support genomic fragility. BMC Evolutionary Biology *6*, 1-12.

Hiraga, S., Botsios, S., Donze, D., and Donaldson, A.D. (2012). TFIIIC localizes budding yeast ETC sites to the nuclear periphery. Mol Biol Cell *23*, 2741-2754.

Hiromi, Y., and Gehring, W.J. (1987). Regulation and function of the Drosophila segmentation gene fushi tarazu. Cell *50*, 963-974.

Hiromi, Y., Kuroiwa, A., and Gehring, W.J. (1985). Control elements of the Drosophila segmentation gene fushi tarazu. Cell *43*, 603-613.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13-23.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A.*, et al.* (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science *351*, 1454-1458.

Holdridge, C., and Dorsett, D. (1991). Repression of hsp70 heat shock gene transcription by the suppressor of hairy-wing protein of Drosophila melanogaster. Mol Cell Biol *11*, 1894-1900.

Holohan, E.E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S., and White, R. (2007). CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. PLoS Genet *3*, e112.

Hon, G.C., Hawkins, R.D., and Ren, B. (2009). Predictive chromatin signatures in the mammalian genome. Hum Mol Genet *18*, R195-201.

Hug, C.B., Grimaldi, A.G., Kruse, K., and Vaquerizas, J.M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell *169*, 216-228 e219.

Hwang, S.S., Kim, Y.U., Lee, S., Jang, S.W., Kim, M.K., Koh, B.H., Lee, W., Kim, J., Souabni, A., Busslinger, M.*, et al.* (2013). Transcription factor YY1 is essential for regulation of the Th2 cytokine locus and for Th2 cell differentiation. Proc Natl Acad Sci U S A *110*, 276-281.

Jager, R., Migliorini, G., Henrion, M., Kandaswamy, R., Speedy, H.E., Heindl, A., Whiffin, N., Carnicer, M.J., Broome, L., Dryden, N.*, et al.* (2015). Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. Nat Commun *6*, 6178.

Jain, S., Ba, Z., Zhang, Y., Dai, H.-Q., and Alt, F.W. (2018). CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning. Cell.

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

Jox, T., Buxa, M.K., Bohla, D., Ullah, I., Macinkovic, I., Brehm, A., Bartkuhn, M., and Renkawitz, R. (2017). Drosophila CP190- and dCTCF-mediated enhancer blocking is augmented by SUMOylation. Epigenetics Chromatin *10*, 32.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kantidze, O.L., and Razin, S.V. (2009). Chromatin loops, illegitimate recombination, and genome evolution. Bioessays *31*, 278-286.

Katainen, R., Dave, K., Pitkanen, E., Palin, K., Kivioja, T., Valimaki, N., Gylfe, A.E., Ristolainen, H., Hanninen, U.A., Cajuso, T.*, et al.* (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. Nat Genet *47*, 818-821.

Kellerman, K.A., Mattson, D.M., and Duncan, I. (1990). Mutations affecting the stability of the fushi tarazu protein of Drosophila. Genes Dev *4*, 1936-1950.

Kent, J.W., Baertsch, R., Hinrichs, A., Miller, W., and Haussler, D. (2003). Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. Proceedings of the National Academy of Sciences *100*, 11484-11489.

Kent, J.W., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and and Haussler, D. (2002). The Human Genome Browser at UCSC. Genome Research *12*, 996-1006.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231-1245.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A., and van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. Cell *153*, 178-192.

Kleiman, E., Jia, H., Loguercio, S., Su, A.I., and Feeney, A.J. (2016). YY1 plays an essential role at all stages of B-cell differentiation. Proc Natl Acad Sci U S A *113*, E3911-3920.

Klenova, E.M., Chernukhin, I.V., El-Kady, A., Lee, R.E., Pugacheva, E.M., Loukinov, D.I., Goodwin, G.H., Delgado, D., Filippova, G.N., Leon, J.*, et al.* (2001). Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. Mol Cell Biol *21*, 2221-2234.

Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E., and Lobanenkov, V.V. (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Molecular and Cellular Biology *13*, 7612-7624.

Koenecke, N., Johnston, J., He, Q., Meier, S., and Zeitlinger, J. (2017). Drosophila poised enhancers are generated during tissue patterning with the help of repression. Genome Res *27*, 64-74.

Kondo, T., Isono, K., Kondo, K., Endo, T.A., Itohara, S., Vidal, M., and Koseki, H. (2014). Polycomb potentiates meis2 activation in midbrain by mediating interaction of the promoter with a tissue-specific enhancer. Dev Cell *28*, 94-101.

Krause, H.M., Klemenz, R., and Gehring, W.J. (1988). Expression, modification, and localization of the fushi tarazu protein in Drosophila embryos. Genes & Development *2*, 1021-1036.

Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci U S A *103*, 10684-10689.

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

Kyrchanova, O., Ivlieva, T., Toshchakov, S., Parshikov, A., Maksimenko, O., and Georgiev, P. (2011). Selective interactions of boundaries with upstream region of Abd-B promoter in Drosophila bithorax complex and role of dCTCF in this process. Nucleic Acids Res *39*, 3042-3052.

Kyrchanova, O., Maksimenko, O., Stakhov, V., Ivlieva, T., Parshikov, A., Studitsky, V.M., and Georgiev, P. (2013). Effective blocking of the white enhancer requires cooperation between two main mechanisms suggested for the insulator function. PLoS Genet *9*, e1003606.

Kyrchanova, O., Mogila, V., Wolle, D., Deshpande, G., Parshikov, A., Cleard, F., Karch, F., Schedl, P., and Georgiev, P. (2016). Functional Dissection of the Blocking and Bypass Activities of the Fab-8 Boundary in the Drosophila Bithorax Complex. PLoS Genet *12*, e1006188.

Kyrchanova, O., Toshchakov, S., Parshikov, A., and Georgiev, P. (2007). Study of the functional interaction between Mcp insulators from the Drosophila bithorax complex: effects of insulator pairing on enhancer-promoter communication. Mol Cell Biol *27*, 3035-3043.

Kyrchanova, O., Toshchakov, S., Podstreshnaya, Y., Parshikov, A., and Georgiev, P. (2008b). Functional interaction between the Fab-7 and Fab-8 boundaries and the upstream promoter region in the Drosophila Abd-B gene. Molecular and cellular biology *28*, 4188-4195.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497-501.

Laine, J.P., Singh, B.N., Krishnamurthy, S., and Hampsey, M. (2009). A physiological role for gene loops in yeast. Genes Dev *23*, 2604-2609.

Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat Cell Biol *9*, 1167-1174.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature *547*, 236-240.

Lei, E.P., and Corces, V.G. (2006). RNA interference machinery influences the nuclear organization of a chromatin insulator. Nat Genet *38*, 936-941.

Lemaitre, C., Zaghloul, L., Sagot, M.-F., Gautier, C., Arneodo, A., Tannier, E., and Audit, B. (2009). Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. BMC Genomics *10*, 1-12.

Lemons, D., and McGinnis, W. (2006). Genomic evolution of Hox gene clusters. Science (New York, NY) *313*, 1918-1922.

LeMotte, P.K., Kuroiwa, A., Fessler, L.I., and Gehring, W.J. (1989). The homeotic gene Sex Combs Reduced of Drosophila: gene structure and embryonic expression. The EMBO journal *8*, 219-227.

Li, H.B., Muller, M., Bahechar, I.A., Kyrchanova, O., Ohno, K., Georgiev, P., and Pirrotta, V. (2011). Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in Drosophila melanogaster. Mol Cell Biol *31*, 616-625.

Li, H.B., Ohno, K., Gui, H., and Pirrotta, V. (2013a). Insulators target active genes to transcription factories and polycomb-repressed genes to polycomb bodies. PLoS Genet *9*, e1003436.

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H.Q., Ong, C.T., Cubenas-Potts, C., Hu, M., Lei, E.P., Bosco, G.*, et al.* (2015a). Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. Mol Cell *58*, 216-231.

Li, M., Ma, Z., Liu, J.K., Roy, S., Patel, S.K., Lane, D.C., and Cai, H.N. (2015b). An Organizational Hub of Developmentally Regulated Chromatin Loops in the Drosophila Antennapedia Complex. Mol Cell Biol *35*, 4018-4029.

Li, M.Z., and Elledge, S.J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nature Methods *4*.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., *et al.* (2013b). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516-520.

Li, Y., Hu, M., and Shen, Y. (2018). Gene Regulation in the 3D Genome. Hum Mol Genet.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., Cao, H., Ni, M., Liu, Y., Gu, Z., *et al.* (2017). In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell *170*, 1028-1043 e1019.

Liu, Z., Scannell, D.R., Eisen, M.B., and Tjian, R. (2011). Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. Cell *146*, 720-731.

Lopez-Perrote, A., Alatwi, H.E., Torreira, E., Ismail, A., Ayora, S., Downs, J.A., and Llorca, O. (2014). Structure of Yin Yang 1 oligomers that cooperate with RuvBL1-RuvBL2 ATPases. J Biol Chem *289*, 22614-22629.

Luo, H., Wang, F., Zha, J., Li, H., Yan, B., Du, Q., Yang, F., Sobh, A., Vulpe, C., Drusbosky, L., *et al.* (2018). CTCF boundary remodels chromatin domain and drives aberrant HOX gene transcription in acute myeloid leukemia. Blood.

Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., *et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.

MacPherson, M.J., Beatty, L.G., Zhou, W., Du, M., and Sadowski, P.D. (2009). The CTCF insulator protein is posttranslationally modified by SUMO. Mol Cell Biol *29*, 714-725.

Maeda, R.K., and Karch, F. (2007). Making connections: boundaries and insulators in Drosophila. Curr Opin Genet Dev *17*, 394-399.

Maeshima, K., Imai, R., Tamura, S., and Nozaki, T. (2014). Chromatin as dynamic 10-nm fibers. Chromosoma *123*, 225-237.

Mahaffey, J.W., and Kaufman, T.C. (1987). Distribution of the Sex combs reduced gene products in Drosophila melanogaster. Genetics *117*, 51-60.

Majumder, P., and Cai, H.N. (2003). The functional analysis of insulator interactions in the Drosophila embryo. Proc Natl Acad Sci U S A *100*, 5223-5228.

Maksimenko, O., Bartkuhn, M., Stakhov, V., Herold, M., Zolotarev, N., Jox, T., Buxa, M.K., Kirsch, R., Bonchuk, A., Fedotova, A.*, et al.* (2015). Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. Genome Res *25*, 89-99.

Martinez-Arias, A., Ingham, P.W., Scott, M.P., and Akam, M.E. (1987). The spatial and temporal deployment of Dfd and Scr transcripts throughout development of Drosophila. Development (Cambridge, England) *100*, 673-683.

Matzat, L.H., Dale, R.K., and Lei, E.P. (2013). Messenger RNA is a functional component of a chromatin insulator complex. EMBO Rep *14*, 916-922.

Melnikova, L., Juge, F., Gruzdeva, N., Mazur, A., Cavalli, G., and Georgiev, P. (2004). Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. Proc Natl Acad Sci U S A *101*, 14806-14811.

Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A., Elkon, R., Melo, S.A., Leveille, N., Kalluri, R.*, et al.* (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol Cell *49*, 524-535.

Miller, D.E., Staber, C., Zeitlinger, J., and Hawley, S.R. (2018). High-quality genome assemblies of 15 Drosophila species generated using Nanopore sequencing. bioRxiv, 267393.

Mills, R.E., Bennett, E.A., Iskow, R.C., Luttig, C.T., Tsui, C., Pittard, W.S., and Devine, S.E. (2006). Recently mobilized transposons in the human and chimpanzee genomes. American journal of human genetics *78*, 671-679.

Misteli, T. (2007). Beyond the sequence: cellular organization of genome function. Cell *128*, 787-800.

Misulovin, Z., Schwartz, Y.B., Li, X.Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D.*, et al.* (2008). Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. Chromosoma *117*, 89-102.

Mitrea, D.M., and Kriwacki, R.W. (2016). Phase separation in biology; functional organization of a higher order. Cell Commun Signal *14*, 1.

mod, E.C., Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L.*, et al.* (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science *330*, 1787-1797.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J.*, et al.* (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep *20*, 1215-1228.

Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P.*, et al.* (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science *336*, 1448-1451.

Moon, H., Filippova, G., Loukinov, D., Pugacheva, E., Chen, Q., Smith, S.T., Munhall, A., Grewe, B., Bartkuhn, M., Arnold, R.*, et al.* (2005). CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. EMBO Rep *6*, 165-170.

Moqtaderi, Z., and Struhl, K. (2004). Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes. Mol Cell Biol *24*, 4118-4127.

Moqtaderi, Z., Wang, J., Raha, D., White, R.J., Snyder, M., Weng, Z., and Struhl, K. (2010). Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. Nat Struct Mol Biol *17*, 635-640.

Morgan, S.L., Mariano, N.C., Bermudez, A., Arruda, N.L., Wu, F., Luo, Y., Shankar, G., Jia, L., Chen, H., Hu, J.F.*, et al.* (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. Nat Commun *8*, 15993.

Morris, G.E. (2008). The Cajal body. Biochim Biophys Acta *1783*, 2108-2115.

Moshkovich, N., Nisha, P., Boyle, P.J., Thompson, B.A., Dale, R.K., and Lei, E.P. (2011). RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. Genes Dev *25*, 1686-1701.

Mukundan, B., and Ansari, A. (2013). Srb5/Med18-mediated termination of transcription is dependent on gene looping. J Biol Chem *288*, 11384-11394.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods *13*, 919-922.

Muravyova, E., Golovnin, A., Gracheva, E., Parshikov, A., Belenkaya, T., Pirrotta, V., and Georgiev, P. (2001). Loss of insulator activity by paired Su(Hw) chromatin insulators. Science *291*, 495-498.

Nagy, G., Czipa, E., Steiner, L., Nagy, T., Pongor, S., Nagy, L., and Barta, E. (2016). Motif oriented high-resolution analysis of ChIP-seq data reveals the topological order of CTCF and cohesin proteins on DNA. BMC Genomics *17*, 637.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017-1021.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science *342*, 948-953.

Negre, B., and Ruiz, A. (2006). HOM-C evolution in Drosophila: is there a need for Hox gene clustering? Trends in genetics : TIG *23*, 55-59.

Negre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A.*, et al.* (2010). A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet *6*, e1000814.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., Laat, W., and Duboule, D. (2011). The dynamic architecture of Hox gene clusters. Science (New York, NY) *334*, 222-225.

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944 e922.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.*, et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.

Nwigwe, I.J., Kim, Y.J., Wacker, D.A., and Kim, T.H. (2015). Boundary Associated Long Noncoding RNA Mediates Long-Range Chromosomal Interactions. PLoS One *10*, e0136104.

Obbard, D.J., Maclennan, J., Kim, K.-W., Rambaut, A., O'Grady, P.M., and Jiggins, F.M. (2012). Estimating Divergence Dates and Substitution Rates in the Drosophila Phylogeny. Molecular Biology and Evolution *29*, 3459-3473.

Ocampo-Hafalla, M., Munoz, S., Samora, C.P., and Uhlmann, F. (2016). Evidence for cohesin sliding along budding yeast chromosomes. Open Biol *6*, 150178.

Ong, C.T., and Corces, V.G. (2008). Modulation of CTCF insulator function by transcription of a noncoding RNA. Dev Cell *15*, 489-490.

Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet *15*, 234-246.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O'Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science *357*.

Ou, X., Huo, J., Huang, Y., Li, Y.F., Xu, S., and Lam, K.P. (2018). Transcription factor YY1 is essential for iNKT cell development. Cell Mol Immunol, 1-10.

Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet *35*, 190-194.

Paré, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y., and McGinnis, W. (2009). Visualization of Individual Scr mRNAs during Drosophila Embryogenesis Yields Evidence for Transcriptional Bursting. Current Biology *19*, 2037-2042.

Pattatucci, A.M., and Kaufman, T.C. (1991). The homeotic gene Sex combs reduced of Drosophila melanogaster is differentially regulated in the embryonic and imaginal stages of development. Genetics *129*.

Pattatucci, A.M., Otteson, D.C., and Kaufman, T.C. (1991). A functional and structural analysis of the Sex combs reduced locus of Drosophila melanogaster. Genetics *129*, 423-441.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y.*, et al.* (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281-1295.

Pick, L., Schier, A., Affolter, M., Schmidt-Glenewinkel, T., and Gehring, W.J. (1990). Analysis of the ftz upstream element: germ layer-specific enhancers are independently autoregulated. Genes Dev *4*, 1224-1239.

Pierce, B.A. (2006). Genetics : a conceptual approach, 2nd ed. Book News, Inc.

Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. Nat Rev Mol Cell Biol *16*, 245-257.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K.*, et al.* (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402-405.

Port, F., Chen, H.-M., Lee, T., and Bullock, S.L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proceedings of the National Academy of Sciences *111*.

Qiu, Z., Song, C., Malakouti, N., Murray, D., Hariz, A., Zimmerman, M., Gygax, D., Alhazmi, A., and Landry, J.W. (2015). Functional interactions between NURF and Ctcf regulate gene expression. Mol Cell Biol *35*, 224-237.

Quinlan, A.R., Clark, R.A., Sokolova, S., Leibowitz, M.L., Zhang, Y., Hurles, M.E., Mell, J.C., and Hall, I.M. (2010). Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Research *20*, 623-635.

Raab, J.R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P.A., Haussler, D., and Kamakaka, R.T. (2012). Human tRNA genes function as chromatin insulators. EMBO J *31*, 330-350.

Racko, D., Benedetti, F., Dorier, J., and Stasiak, A. (2018). Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes. Nucleic Acids Res *46*, 1648-1660.

Ramirez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Gruning, B.A., Villaveces, J., Habermann, B., Akhtar, A., and Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun *9*, 189.

Ramos, E., Torre, E.A., Bushey, A.M., Gurudatta, B.V., and Corces, V.G. (2011). DNA Topoisomerase II Modulates Insulator Function in Drosophila. PLoS ONE *6*.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S.*, et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

Rao, S.S.P., Huang, S.-C., Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., and Bochkov, I.D. (2017). Cohesin loss eliminates all loop domains. Cell *171*, 305.

Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. Proc Natl Acad Sci U S A *99*, 6883-6888.

Ricci, M.A., Manzo, C., Garcia-Parajo, M.F., Lakadamyali, M., and Cosma, M.P. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. Cell *160*, 1145-1158.

Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y., and Corces, V.G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol Cell *67*, 837-852 e837.

Ruiz-Velasco, M., Kumar, M., Lai, M.C., Bhat, P., Solis-Pinson, A.B., Reyes, A., Kleinsorg, S., Noh, K.M., Gibson, T.J., and Zaugg, J.B. (2017). CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. Cell Syst *5*, 628-637 e626.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J.*, et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A *112*, E6456-6465.

Savitskaya, E., Melnikova, L., Kostuchenko, M., Kravchenko, E., Pomerantseva, E., Boikova, T., Chetverina, D., Parshikov, A., Zobacheva, P., Gracheva, E.*, et al.* (2006). Study of long-distance functional interactions between Su(Hw) insulators that can regulate enhancer-promoter communication in Drosophila melanogaster. Mol Cell Biol *26*, 754-761.

Schier, A.F., and Gehring, W.J. (1993). Analysis of a fushi tarazu autoregulatory element: multiple sequence elements contribute to enhancer activity. EMBO J *12*, 1111-1119.

Schoborg, T., Kuruganti, S., Rickels, R., and Labrador, M. (2013). The Drosophila gypsy insulator supports transvection in the presence of the vestigial enhancer. PLoS One *8*, e81331.

Schuster-Bockler, B., and Lehner, B. (2012). Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature *488*, 504-507.

Schwartz, Y.B., Linder-Basso, D., Kharchenko, P.V., Tolstorukov, M.Y., Kim, M., Li, H.B., Gorchakov, A.A., Minoda, A., Shanower, G., Alekseyenko, A.A.*, et al.* (2012). Nature and function of insulator protein binding sites in the Drosophila genome. Genome Res *22*, 2188-2198.

Scott, K.C., Merrett, S.L., and Willard, H.F. (2006). A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. Curr Biol *16*, 119-129.

Seeber, A., and Gasser, S.M. (2017). Chromatin organization and dynamics in double-strand break repair. Current Opinion in Genetics & Development *43*, 9-16.

Serizay, J., and Ahringer, J. (2018). Genome organization at different scales: nature, formation and function. Curr Opin Cell Biol *52*, 145-153.

Sexton, T., and Cavalli, G. (2015). The role of chromosome domains in shaping the functional genome. Cell *160*, 1049-1059.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978-981.

Singh, N.D., Larracuente, A.M., Sackton, T.B., and Clark, A.G. (2009). Comparative Genomics on the Drosophila Phylogenetic Tree. Annual Review of Ecology, Evolution, and Systematics *40*, 459-480.

Spilianakis, C.G., and Flavell, R.A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat Immunol *5*, 1017-1027.

Stadler, M.R., Haines, J.E., and Eisen, M.B. (2017). Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early Drosophila melanogaster embryo. eLife *6*.

Stevens, T.J., Lando, D., Basu, S., Atkinson, L.P., Cao, Y., Lee, S.F., Leeb, M., Wohlfahrt, K.J., Boucher, W., O'Shaughnessy-Kirwan, A.*, et al.* (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature *544*, 59-64.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241-245.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013). Jpx RNA activates Xist by evicting CTCF. Cell *153*, 1537-1551.

Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. Genome Res *24*, 390-400.

Tan-Wong, S.M., French, J.D., Proudfoot, N.J., and Brown, M.A. (2008). Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. Proc Natl Acad Sci U S A *105*, 5160-5165.

Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M., and Proudfoot, N.J. (2012). Gene loops enhance transcriptional directionality. Science *338*, 671-675.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B.*, et al.* (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611-1627.

Terakawa, T., Bisht, S., Eeftens, J.M., Dekker, C., Haering, C.H., and Greene, E.C. (2017). The condensin complex is a mechanochemical motor that translocates along DNA. Science *358*, 672-676.

Tiwari, V.K., McGarvey, K.M., Licchesi, J.D., Ohm, J.E., Herman, J.G., Schubeler, D., and Baylin, S.B. (2008). PcG proteins, DNA methylation, and gene repression by chromatin looping. PLoS Biol *6*, 2911-2927.

Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., and Rubin, G.M. (2007). Global analysis of patterns of gene expression during Drosophilaembryogenesis. Genome Biology *8*, 1-24.

Tomita, S., Abdalla, M.O., Fujiwara, S., Yamamoto, T., Iwase, H., Nakao, M., and Saitoh, N. (2017). Roles of long noncoding RNAs in chromosome domains. Wiley Interdiscip Rev RNA *8*.

Trimarchi, T., Bilal, E., Ntziachristos, P., Fabbri, G., Dalla-Favera, R., Tsirigos, A., and Aifantis, I. (2014). Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. Cell *158*, 593-606.

Tsai, A.G., and Lieber, M.R. (2010). Mechanisms of chromosomal rearrangement in the human genome. BMC Genomics *11*, 1-9.

Udvardy, A., Maine, E., and Schedl, P. (1985). The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. J Mol Biol *185*, 341-358.

Uusküla-Reimand, L., Hou, H., Samavarchi-Tehrani, P., Rudan, M., Liang, M., Medina-Rivera, A., Mohammed, H., Schmidt, D., Schwalie, P., Young, E.J.*, et al.* (2016). Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. Genome Biology *17*, 182.

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell *17*, 453-462.

Valenzuela, L., and Kamakaka, R.T. (2006). Chromatin insulators. Annu Rev Genet *40*, 107-138.

Valton, A.-L., and Dekker, J. (2016). TAD disruption as oncogenic driver. Current Opinion in Genetics & Development *36*, 34-40.

Van Bortle, K., and Corces, V.G. (2012). Nuclear organization and genome function. Annu Rev Cell Dev Biol *28*, 163-187.

Van Bortle, K., Nichols, M.H., Li, L., Ong, C.T., Takenaka, N., Qin, Z.S., and Corces, V.G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol *15*, R82.

Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J.E., and Corces, V.G. (2012). Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. Genome Res *22*, 2176-2187.

van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell *169*, 780-791.

Vera, M., Biswas, J., Senecal, A., Singer, R.H., and Park, H.Y. (2016). Single-Cell and Single-Molecule Analysis of Gene Expression Regulation. Annu Rev Genet *50*, 267-291.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep *10*, 1297-1309.

Vildanova, M.S., and Smirnova, E.A. (2016). [Effects of Different Classes of Plant Hormones on Mammalian Cells]. Tsitologiia *58*, 5-15.

Wang, Q., Sun, Q., Czajkowsky, D.M., and Shao, Z. (2018). Sub-kb Hi-C in D. melanogaster reveals conserved characteristics of TADs between insect and mammalian cells. Nature communications *9*, 188.

Watts, F.Z. (2016). Repair of DNA Double-Strand Breaks in Heterochromatin. Biomolecules *6*.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L.*, et al.* (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573-1588 e1528.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T.*, et al.* (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796-801.

Weth, O., and Renkawitz, R. (2011). CTCF function is modulated by neighboring DNA binding factors. Biochem Cell Biol *89*, 459-468.

Wijchers, P.J., Krijger, P.H.L., Geeven, G., Zhu, Y., Denker, A., Verstegen, M., Valdes-Quezada, C., Vermeulen, C., Janssen, M., Teunissen, H.*, et al.* (2016). Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. Mol Cell *61*, 461-473.

Willi, M., Yoo, K.H., Reinisch, F., Kuhns, T.M., Lee, H.K., Wang, C., and Hennighausen, L. (2017). Facultative CTCF sites moderate mammary super-enhancer activity and regulate juxtaposed gene in non-mammary cells. Nat Commun *8*, 16069.

Wolle, D., Cleard, F., Aoki, T., Deshpande, G., Schedl, P., and Karch, F. (2015). Functional Requirements for Fab-7 Boundary Activity in the Bithorax Complex. Mol Cell Biol *35*, 3739-3752.

Wutz, G., Varnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J.*, et al.* (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J *36*, 3573-3599.

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D., and Felsenfeld, G. (2010). Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. Genes Dev *24*, 2543-2555.

Yuen, K.C., Slaughter, B.D., and Gerton, J.L. (2017). Condensin II is anchored by TFIIIC and H3K4me3 in the mammalian genome and supports the expression of active dense gene clusters. Sci Adv *3*, e1700191.

Zaprazna, K., Basu, A., Tom, N., Jha, V., Hodawadekar, S., Radova, L., Malcikova, J., Tichy, B., Pospisilova, S., and Atchison, M.L. (2018). Transcription factor YY1 can control AID-mediated mutagenesis in mice. Eur J Immunol *48*, 273-282.

Zhang, L., Ward, J.D., Cheng, Z., and Dernburg, A.F. (2015). The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. Development *142*, 4374-4384.

Zhang, Y., McCord, R., Ho, Y.-J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial Organization of the Mouse Genome and Its Role in Recurrent Chromosomal Translocations. Cell *148*, 908-921.

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

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

Zlotorynski, E. (2018). Gene expression: The yin and yang of enhancer-promoter interactions. Nat Rev Mol Cell Biol *19*, 75.