Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by

Gabriel Rodrigues Cavalheiro, M.Sc.
born in Rio de Janeiro, Brazil

Oral examination: 15th March 2021

A genetic dissection of TAD formation during embryogenesis

Referees: Dr. Kyung-Min Noh

Prof. Dr. Karsten Rippe

Inaugural-Dissertation

zur

Erlangung der Doktorwürde

der

Naturwissenschaftlich-Mathematischen Gesamtfakultät

der

Ruprecht-Karls-Universität

Heidelberg

vorgelegt von

Gabriel Rodrigues Cavalheiro, M.Sc.

aus Rio de Janeiro, Brasilien

Tag der mündlichen Prüfung: der 15. März 2021

Eine genetische Analyse der TAD-Bildung während der Embryogenese

Gutachter: Dr. Kyung-Min Noh

Prof. Dr. Karsten Rippe

I. Acknowledgments

I'm extremely grateful to Eileen Furlong, for providing me with the opportunity and all the means to pursuit a PhD. Her support and care as a mentor have been essential for my scientific development.

I'd like to thank the people that have worked most closely to me, and to whom I'm much in debt for the friendship and scientific help over the years. These are specially Tim Pollex, Rebecca Rodriguez-Viales, Charles Girardot and Martina Varisco.

In addition, I'd like to thank the many friends I made in the Furlong lab, for day-to-day great interactions, being always available to give me feedback and answer questions, and for all the fun times. Thank you Yacine, Raquel, Songjie, Katharina, Matteo, Federica, Alessandro, Mattia, Stefano, Elena, Adam, Alek, Anna, James, Olga S., Baekgyu, Olga M., Marijn, Nezha, Miguel, Maria, Lucia, Matthias, Marek, Imtiaz, Blanca. Thank you very much Marijn for revising my thesis and giving very useful comments.

I'm also fortunate to have very supportive TAC members that provided valuable advice during my PhD: Dr. Kyung-Min Noh, Dr. Karsten Rippe and Dr. Justin Crocker. Thank you Dr. Frank Lyko for joining in the defense committee.

I had luck to meet amazing people over the last 4 years in Heidelberg. I'd like to thank especially my good friends, Sergio Triana, Sergio Lembo, Karen, Mariana, Carol, Erica, Veronika, Piotrek, Ksenia, Philipp, Dimi, Ana, Lina and Gabe for all the nice memories.

I would like to thank Kristina for introducing me to the world of saxophone and jazz, and all the happy moments, support and love.

Thank you Sinja and Antonio and all macumba members. I had much fun playing drums with you. Will never forget playing during the carnival in Heidelberg!

I'm very fortunate to have done my PhD at EMBL, and having the chance to receive unparalleled support and guidance for my experiments, from collaborators and core facilities. Thank you ALMF, Gene Core, Flow cytometry facility, Ephrussi lab members and Alessandra Reversi for all the assistance, as well as all other EMBL staff that help to run the multiple great services in this institute.

Thank you Inés, for your constant support and all beautiful moments we've had together. You are a great source of joy and love in my life. I learn with you every day and couldn't ask for a better partner.

Gostaria por último de agradecer a minha família, em especial a minha mãe, Ana Lucia, e aos meus avós, Maria de Lourdes e Antônio, e ao Chico. Durante toda a minha vida vocês foram essenciais em todos os aspectos do meu desenvolvimento pessoal, e sempre fizeram de tudo para que eu tivesse as melhores oportunidades. Eu sou eternamente grato por tudo que fizeram por mim. Sem vocês nada seria possível.

II. Summary

Regulation of gene expression is a complex process that requires the action of DNA elements such as promoters, enhancers and insulators. In eukaryotic genomes, these elements are organized in self-interacting units, the Topologically-Associating Domains (TADs). The role of TADs in regulating gene expression is increasingly appreciated, though still poorly understood. Genetic dissection of TADs and their boundaries have shown that TADs can restrict enhancer-promoter interactions, but also promote enhancer-mediated activation. It is not clear by which mechanisms TADs can form, although insulator protein binding at TAD boundaries has been implicated. TADs are largely conserved between different cell types and an intriguing question is when they form during embryogenesis and whether they can help drive developmental gene expression programs during.

Using *Drosophila melanogaster* as a model organism, I decided to first investigate the timing of chromatin topology establishment in the early embryo. I observed that 3D spatial proximity across different loci appeared simultaneously with the onset of zygotic transcription, consistent with TADs being established during that stage. To interrogate the mechanisms regulating TAD establishment, I depleted individual insulator proteins (BEAF-32, CTCF and CP190) during the first steps of embryogenesis. Most TADs could form in the absence of those proteins. However, a mild loss of TAD insulation throughout the genome was detected. Specific loci appeared more severely affected upon loss of CP190, although it is not clear what distinguishes those loci from unaffected loci. Insulator depletion also caused gene expression defects that are potentially linked to topological alterations. Deletion of specific DNA elements within a TAD boundary revealed that transcription and insulator protein binding may both be required for the full insulation potential of certain boundaries.

To systematically investigate which factors determine TAD boundary function, I inserted several TAD boundaries into two TADs containing distinct properties. Boundary insertions disrupted local chromatin topology in nearly half of the cases, revealing a strong context-specificity for boundary function. While some boundaries altered chromatin topology in a locus- and orientation-specific manner, other boundaries affected both TADs. Some insertions caused developmental lethality, indicating effects on gene expression associated with topological alterations.

Overall, the findings presented here indicate a remarkable context-specificity of mechanisms and factors shaping the genome topology in *Drosophila*, possibly being required for the precise regulation of highly complex gene expression programs during embryonic development.

III. Zusammenfassung

Die Regulation der Genexpression ist ein komplexer Prozess, der das Zusammenspiel von DNA-Elementen wie Promotoren, Enhancer und Insulatoren benötigt. In eukaryotischen Genomen sind diese Elemente in selbstinteragierenden Einheiten organisiert, welche als topologisch-assoziierte Domänen (TADs) bezeichnet werden. Die Rolle von TADs in der Genregulation wird immer mehr anerkannt, ist aber dennoch schlecht verstanden. Die genetische Analyse von TADs und deren Grenzen hat gezeigt, dass TADs Enhancer-Promotor-Interaktionen einschränken können, aber auch Enhancer-vermittelte Aktivierung fördern. Es ist unklar durch welche Mechanismen TADs entstehen, jedoch wird das Binden von Insulatorproteinen an TAD- Grenzen damit in Zusammenhang gebracht. TADs sind zwischen verschiedenen Zelltypen weitgehend konserviert. Eine interessante Frage ist, wann in der Embyonalentwicklung sie gebildet werden und ob sie dabei helfen können Genexpressionsprogramme zu steuern.

Unter Verwendung von *Drosophila melanogaster* als Modellorganismus beschloss ich, zunächst den Zeitpunkt der Etablierung der Chromatintopologie im frühen Embryo zu untersuchen. Ich beobachtete, dass die räumliche Nähe zwischen verschiedenen Loci gleichzeitig mit dem Einsetzen der zygotischen Transkription auftrat, was mit der Etablierung von TADs in diesem Stadium zusammenfällt. Zur Untersuchung der Mechanismen, die die TAD-Etablierung regulieren, habe ich einzelne Insulatorproteine (BEAF-32, CTCF und CP190) während der ersten Schritte der Embryogenese verringert. Die meisten TADs konnten sich in Abwesenheit dieser Proteine bilden. Jedoch wurde ein leichter Verlust der TAD-Isolierung im gesamten Genom festgestellt. Spezifische Loci schienen nach dem Verlust von CP190 stärker betroffen zu sein, obwohl nicht klar ist, was diese Loci von nicht betroffenen Loci unterscheidet. Das verringerte Vorhandensein des Insulatorproteins verursachte auch Genexpressionsdefekte, die möglicherweise mit topologischen Veränderungen einhergehen. Das Entfernen spezifischer DNA-Elemente innerhalb einer TAD-Grenze zeigte, dass sowohl die Transkription als auch die Bindung des Insulatorproteins für das volle Isolationspotential bestimmter Grenzen erforderlich sein könnten.

Um systematisch zu untersuchen, welche Faktoren die TAD-Grenzfunktion bestimmen, habe ich mehrere TAD-Grenzen in zwei TADs mit unterschiedlichen Eigenschaften eingefügt. Grenzinsertionen störten in fast der Hälfte der Fälle die lokale Chromatintopologie und zeigten dabei eine starke Kontextspezifität für die Grenzfunktion. Während einige Grenzen die Chromatintopologie orts- und orientierungsspezifisch veränderten, wirkten sich andere Grenzen auf beide TADs aus. Einige Insertionen verursachten Letalität während der Embryonalentwicklung, was auf Auswirkungen auf die Genexpression verbunden mit topologischen Veränderungen hinweist.

Insgesamt weisen die hier vorgestellten Ergebnisse auf eine bemerkenswerte Kontextspezifität von Mechanismen und Faktoren hin, welche die Genomtopologie in *Drosophila* prägen und möglicherweise für die präzise Regulation hochkomplexer Genexpressionsprogramme während der Embryonalentwicklung erforderlich sind.

1 Table of Contents

Table of Figures

1 Introduction

1.1 *Cis*-regulatory elements and the regulation of gene expression

In the beginning of the 21st century, the first near-complete sequences of animal genomes (*Drosophila melanogaster* and human) became available (Adams et al., 2000; Venter et al., 2001). Those sequences revealed that genes correspond to only a small fraction of the whole genome, showing that the majority of animal genomes does not encode RNA or proteins.

Decades of work had previously shown that non-coding regulatory DNA elements play a crucial role in regulating the expression of genes (Blackwood & Kadonaga, 1998; Butler & Kadonaga, 2002; Gerasimova & Corces, 2001). With sequenced genomes and the advent of next-generation sequencing, it became possible to map the precise location of regulatory elements and therefore to dramatically improve the annotation of genomes.

This greatly facilitated studies on the function of such elements, in particular their role during embryonic development. The most well-studied genomic elements that regulate gene expression are promoters, enhancers and insulators.

1.1.1 Promoters

The promoter is an essential DNA element regulating the activation of gene transcription. Promoters are located at the 5' end of a gene, where RNA polymerase is recruited and binds to chromatin, initiating transcription (Haberle & Stark, 2018; Lenhard, Sandelin, & Carninci, 2012). These elements can lead to some basal level of gene expression. The length and DNA sequence of a promoter can vary between genes and species, and this difference allows regulation of transcriptional initiation. For example, in "broad" promoters transcription initiates at multiple locations in the genomic region (thus containing many Transcription Start Sites "TSSs") and tend to occur in housekeeping genes. "Narrow" promoters have a very focused region of transcriptional initiation (sometimes only at one base) and are enriched at genes that are precisely regulated (e.g. genes with tissue-specific expression). Moreover, promoter features can also influence the level of transcription of a gene (Haberle & Stark, 2018; Lenhard et al., 2012). While promoters regulate transcription initiation through a multitude of

mechanisms, enhancers are key DNA elements that modulate the basal transcriptional machinery, controlling transcription in a precise and dynamic manner.

1.1.2 Enhancers

To regulate a gene's expression in both space and time within an organism, and to dynamically respond to both internal and external cues, a gene needs regulatory elements beyond promoters. Enhancers are DNA sequences that drive transcription of a gene in a specific biological condition (e.g. in a tissue of a developing embryo, in cells receiving a molecular signal, etc.) (Figure 1A) (Field & Adelman, 2020; Furlong & Levine, 2018; Schoenfelder & Fraser, 2019).

Due to its modular nature, enhancers can activate their target gene irrespective of their own orientation, and their position in relation to that gene (Fig. 1B,C) (Field & Adelman, 2020; Furlong & Levine, 2018). Another important property of enhancers is their ability to activate genes from a distance (Furlong & Levine, 2018; Schoenfelder & Fraser, 2019). Even though enhancers can activate genes as far as ~1 Megabase (Mb) away, as shown for a mammalian *Sonic hedgehog* [*Shh*] enhancer (Lettice et al., 2003)), increasing their distance to a target gene may lead to less efficient activation (Furlong & Levine, 2018; Schoenfelder & Fraser, 2019).

The DNA sequence of enhancers and promoters determines which proteins are recruited to these elements. While proteins that recognize DNA sequences motifs, such as transcription factors, can bind to specific enhancers and promoters, other proteins can be recruited indirectly through protein-protein interactions (Field & Adelman, 2020). It has been proposed that enhancers could only efficiently activate promoters if the enhancer-bound proteins can interact with the promoter-bound proteins, leading to the idea that enhancers and promoters need to be compatible (Field & Adelman, 2020). This could explain why in some cases enhancers are not able to activate their closest gene in the chromosome, but rather "jump" over it to activate more distant genes (Fig. 1D). However, it is not clear what proportion of enhancers and promoters require this compatibility throughout the genome.

Multiple enhancers can regulate a single gene's expression, either acting additively or cooperatively to activate the gene in a given biological context (Fig. 1E) (Furlong & Levine, 2018; Spitz & Furlong, 2012). Alternatively, different enhancers can each drive the expression of the target gene in specific tissues or times (Fig. 1F). There are also some examples where a

single enhancer can also activate more than one gene simultaneously (Fig. 1G) (Furlong & Levine, 2018; Spitz & Furlong, 2012).

Besides driving dynamic gene expression, enhancers can also provide robustness. For example, a gene can be regulated by multiple enhancers with partially redundant activity that provide similar activation instructions (Spitz & Furlong, 2012). Under physiological conditions, mutations affecting one of these 'redundant' enhancers may not affect gene expression. However during stress conditions the functional role of 'redundant' enhancers becomes clear - loss of a 'redundant' enhancer can impact gene expression e.g. in elevated temperatures (Tsai, Galupa, & Crocker, 2020).

Although enhancers are the key regulators of gene expression in space and time, additional DNA elements, such as insulators, are needed to modulate their function and orchestrate complex gene expression programs.

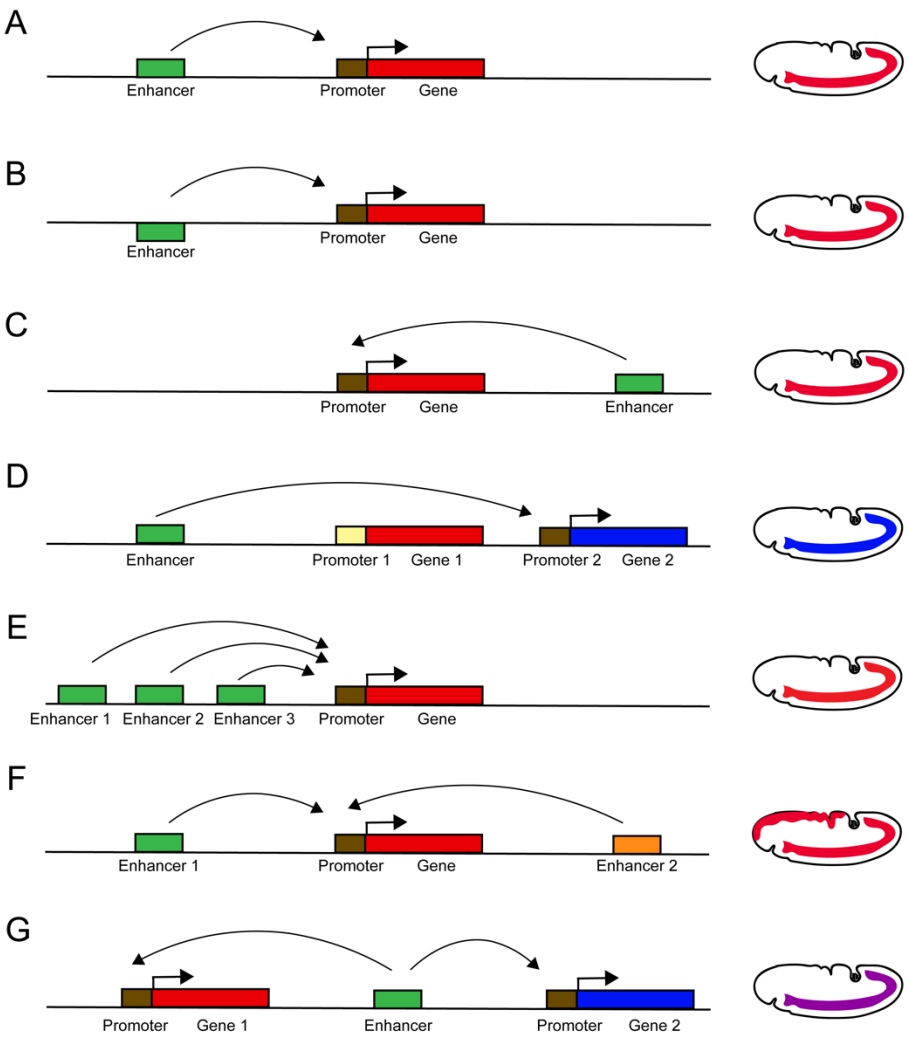

Figure 1: **Enhancer-promoter communication.**
(A) Enhancers can stimulate tissue-specific transcription from a distance.
(B) The orientation of the enhancer does not affect its ability to activate the target promoter.
(C) The relative position of the enhancer does not affect its ability to activate the target promoter.
(D) Enhancers can activate a target promoter by 'jumping' over a gene.
(E) Multiple enhancers can activate the same promoter.
(F) A promoter can be activated by different enhancers, sometimes leading to expression in different tissues/stages.
(G) An enhancer can activate two promoters at the same stage/tissue.
An embryo scheme is presented on the right, to illustrate the expression pattern of a hypothetic gene in the different conditions.

1.1.3 Insulators

Insulators were discovered in *Drosophila*, as DNA elements capable of regulating the action of enhancers on promoters, either by blocking their communication or shielding them from the influence of other elements in the genome (Ozdemir & Gambetta, 2019).

Insulators were first discovered as DNA regions that flank the heat-shock locus and hypothesized to restrict heat-shock-induced DNA decompaction. Those regions were named *specialized chromosome structure* (*scs*) and *scs'* (Udvardy, Maine, & Schedl, 1985). When *scs* and *scs'* elements are placed flanking a reporter gene, they are able to shield the reporter gene from the influence of nearby regulatory elements (Fig. 2A) (Kellum & Schedl, 1991). Moreover, when *scs* elements are placed between an enhancer and a promoter they can block enhancer-mediated activation (Kellum & Schedl, 1992) (Fig. 2A). Evidence for the existence of more insulator elements came from studies with *gypsy* transposable elements in flies (Geyer & Corces, 1992; Jack, Dorsett, Delotto, & Liu, 1991; Peifer & Bender, 1986). These elements can insert in different genomic locations, and if the insertion occurs between an enhancer and a promoter, it can block the enhancer-mediated activation (Fig. 2B) (Geyer & Corces, 1992; Jack et al., 1991; Peifer & Bender, 1986). In parallel, different studies attempted to understand the logic of *cis*-regulation of *Hox* genes' expression, also in flies (Gyurkovics, Gausz, Kummer, & Karch, 1990). Enhancers that regulate the expression of the *Hox* gene *Abdominal-B* (*Abd-B*) had been previously characterized. DNA deletions of regions between the enhancers resulted in activation of *Abd-B* in ectopic body segments, thus indicating a role of these intervening sequences in inhibiting the cross-talk between enhancers (Gyurkovics et al., 1990). These early locus-specific functional studies hypothesized that insulators would be widespread in the genome, helping to shape regulatory domains of gene expression, and avoid mis-regulation of genes by the wrong enhancers.

The main function assigned to insulators is to block enhancer-promoter communication (Fig. 2C), even though the exact mechanism is still unclear. Insulators are hypothesized to communicate with other insulators to exert their function, an idea that gained strength with the discovery of "insulator bypass" (Muravyova et al., 2001). When two copies of an insulator are positioned in tandem, in between an enhancer and a promoter, they cancel each other's enhancer-blocking function (Fig. 2D). Therefore, insulators can promote transcription by bringing activating enhancers close to their target promoters (Fig. 2D). Insulators have also been hypothesized to directly activate or repress transcription, even though this function has not been thoroughly studied (Cai & Levine, 1997; Wei & Brennan, 2001) (Fig. 2E). The *Hox* locus shows how insulators can shape gene expression by performing multiple functions. While insulators are required for blocking activity of *Hox* enhancers, they are also required for long-distance enhancer-promoter communication (Kyrchanova et al., 2019; Postika et al., 2018).

Similar to enhancers, insulators can act across long chromosomal distances, and even between chromosomes. The phenomenon of transvection occurs when regulatory elements in one chromosome are able to interact with regulatory elements in a homologous chromosome (Fig. 2F) (Duncan, 2002). Transvection was discovered in *Drosophila*, and shown to rely on insulator elements (Duncan, 2002; Fujioka, Mistry, Schedl, & Jaynes, 2016). Trans-chromosomal enhancer-promoter communication occurs in vertebrate genomes (Noordermeer et al., 2011), albeit it is rare. The higher frequency of transvection in *Drosophila* is likely facilitated by homologous chromosome pairing during *Drosophila* embryogenesis (Fujioka et al., 2016). Another function of insulators is to block the spread of heterochromatin (Fig. 2G), a phenomenon also discovered in *Drosophila*, and the basis for Position-Effect Variegation (PEV) (Gaszner & Felsenfeld, 2006). Altogether, the discovery of the multiple regulatory roles for insulator elements propelled the hypothesis that they could shape the formation of regulatory domains that allow sets of genes to be coordinately transcribed (Fig. 2H) (Bushey, Dorman, & Corces, 2008; J. Yang & Corces, 2012).

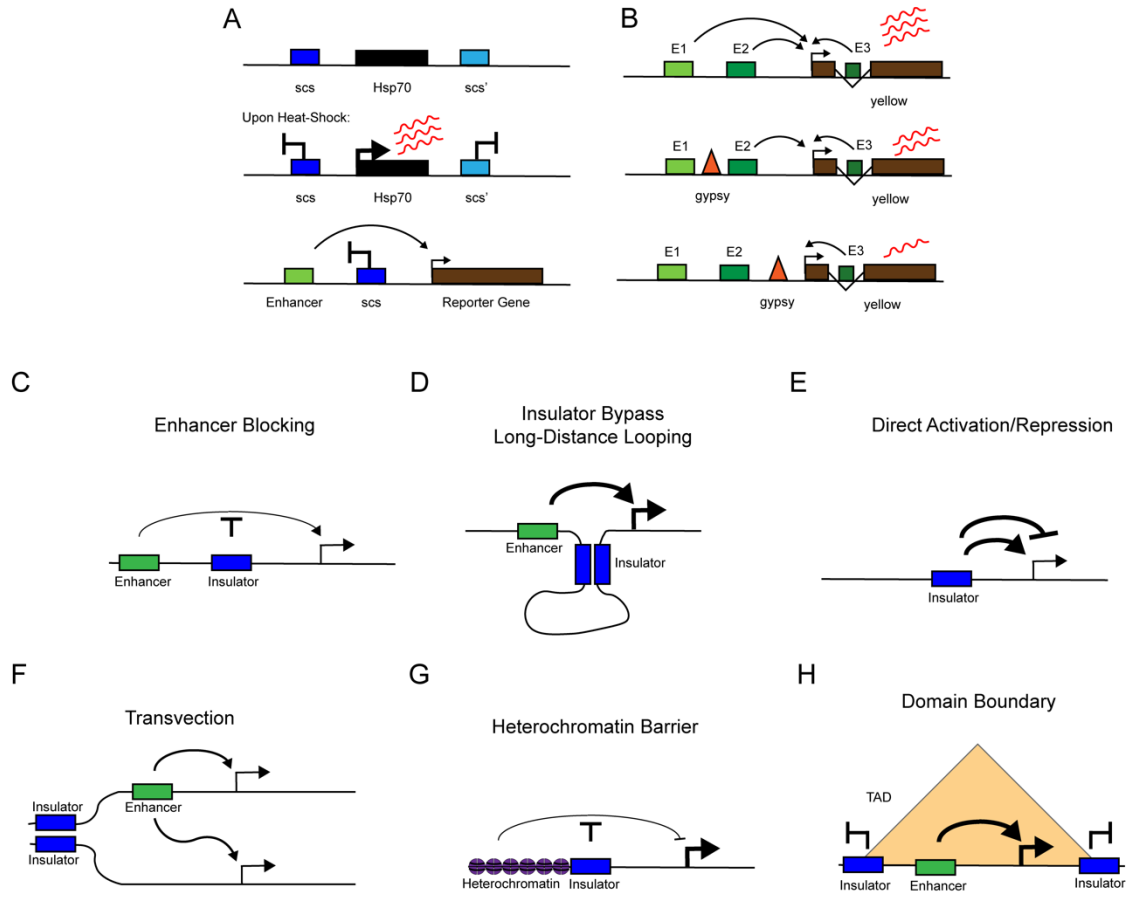

Figure 2: **Insulator elements**
(A) Scheme of the first discovered insulators in *Drosophila*, the *scs* and *scs*' elements, which flank the *Hsp70* gene, and allow its activation upon heat shock. The *scs* elements were also shown to block enhancer-mediated activation.
(B) Scheme of the *gypsy* insulator and its ability to block enhancer-mediated activation, demonstrated in the *yellow* locus, whenever placed in between a *yellow* enhancer and the promoter.
(C-H) Functions proposed for insulator elements.

Proteins that bind to insulators are called "insulator proteins". One of the first discovered insulator proteins was Suppressor of Hairy wing (Su(Hw)), which binds to the *gypsy* transposon in *Drosophila* (Geyer & Corces, 1992). In vertebrates, the first and only protein reported to have an insulator function was the CCCTC-binding factor (CTCF) (Bell, West, & Felsenfeld, 1999). *Drosophila* also has a CTCF homolog as well as several additionally characterized insulator proteins. *Drosophila* insulator proteins can be divided in 2 groups: (1) proteins that bind to DNA generally through C2H2 zinc-finger domains, recognizing a specific motif (CTCF, BEAF-32, Su(Hw), Zw5, Pita, ZIPIC, Ibf1/2 and Opbp) and (2) proteins that

interact with the 1st group, but do not bind to DNA directly (CP190 and Mod(Mdg4)) (Fig. 3) (Ozdemir & Gambetta, 2019). Insulator proteins have been shown to form homo-dimers, hetero-dimers, homo-trimers and multimers *in vitro*. Specifically, CTCF, Zw5, ZIPIC, Pita, CP190 and Mod(Mdg4) were shown to homodimerize. Mod(Mdg4) has several protein isoforms, which have the capacity to heterodimerize with some proteins with the BTB/POZ domain. BEAF-32 forms a homotrimer. CP190 and Mod(Mdg4) are able to form multimers (Kyrchanova & Georgiev, 2014).

Insulators depend on precise combinations of insulator proteins to function, as the loss of binding of one or more proteins compromises their function (Gerasimova & Corces, 1996). Moreover, replacing an insulator by a different insulator may not lead to enhancer blocking to the same extent (Hogga, Mihaly, Barges, & Karch, 2001; Iampietro, Cléard, Gyurkovics, Maeda, & Karch, 2008; Kyrchanova et al., 2016). Therefore, the combinatorial binding of insulator proteins and their genomic context are crucial for insulator function, suggesting that insulators may need to communicate with other "compatible" insulators in their genomic neighborhood to function. However, the molecular details of this, and the relative sequence dependencies, are not understood, especially in *Drosophila*, and is a question that my thesis aims to address.

An important open question is how genetic elements such as promoters, enhancers and insulators influence, and are influenced by, the 3D spatial organization of chromosomes. The close relationship between nuclear organization and the regulation of gene expression, and its relevance for processes such as embryonic development, is becoming increasingly clear.

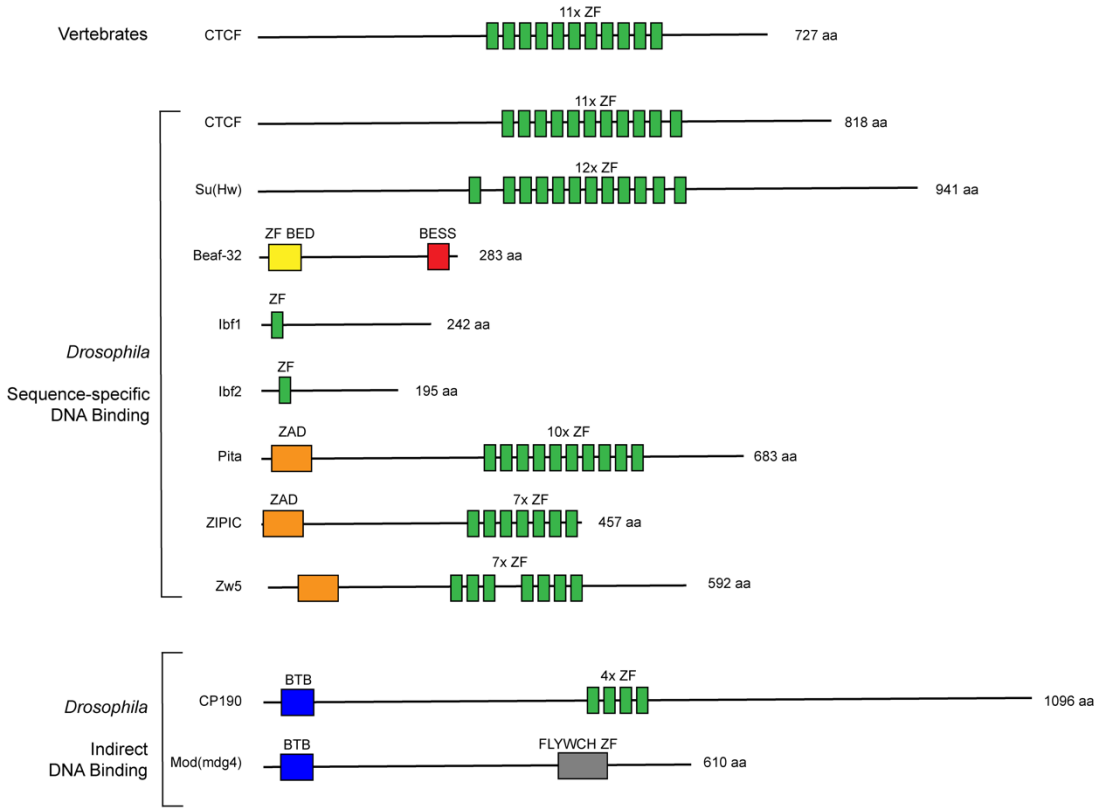

Figure 3: **Insulator proteins in *Drosophila* and Vertebrates.**
Figure adapted from (Ozdemir & Gambetta, 2019)

1.2 Chromatin organization in vertebrates and invertebrates – the genome is organized in three dimensional topologies of different scales

Since the initial observation of the cell nucleus by Antoni van Leeuwenhoek in the early 18th century, much has been learned about multiple aspects of chromatin organization within this organelle. Microscopy-based studies, and more recent DNA-sequencing based technologies, combined with the development of genetic tools across several model and non-model organisms are bringing to light: (1) how chromosomes are organized in the nuclear space, (2) the mechanisms regulating this organization and (3) the functional relevance for processes such as DNA replication, repair and transcription.

1.2.1 Euchromatin and heterochromatin

Early studies on the composition of the eukaryotic nucleus found that chromatin is not homogeneously organized, but displays regions with high or less chromatin density. The

chromatin-dense regions were named “heterochromatin” and low-density regions “euchromatin” (Heitz, 1928). It was noted that heterochromatin is preferentially positioned at the nuclear periphery and around a region with very low chromatin density, the nucleolus (Cremer & Cremer, 2001). Those observations led to the hypothesis that spatial chromatin organization in the nucleus is relevant for regulating DNA function.

1.2.2 Chromosome territories, Nuclear bodies and Hi-C Compartments

Chromosomes occupy discrete 3D volumes in the nuclear space during interphase, called chromosome territories (Fig. 4A) (Cremer & Cremer, 2001). In general, chromosomal surfaces are less condensed, resembling euchromatin, while their interior is more compact, composed of heterochromatin. There is limited mixing between the surface of different chromosomes, where highly transcribed genes are found, indicating that such regions are permissive for transcription (Cremer & Cremer, 2001; Pombo & Dillon, 2015). Active genes form spatial clusters within the nucleus, which were first observed through microscopy experiments, and named “transcription factories” (Pombo & Dillon, 2015). Transcription factories help to stimulate transcription by trapping RNA polymerase II molecules in a confined nanometer space, locally increasing their concentration (Fig. 4B).

Staining of chromatin by different methods reveals chromatin clusters frequently associated with specific proteins. These clusters are collectively denominated “nuclear bodies” or “nuclear speckles” and have various functions (Y. Chen & Belmont, 2019). Each cluster is enriched with a set of proteins that are associated to DNA/RNA metabolism roles, such as gene activation, repression, DNA repair, RNA post-transcriptional processing and export. Mutations that affect components of nuclear bodies, and disrupt their formation, are associated with human disease, further indicating a functional role for these structures in regulating gene expression (Y. Chen & Belmont, 2019).

The advent of Chromosome-Conformation-Capture (3C) and derived techniques allowed the topology of chromosomes to be assessed using molecular biology methods such as PCR or DNA sequencing (Dekker, Rippe, Dekker, & Kleckner, 2002; Denker & de Laat, 2016). In such methods, biological samples are fixed, the DNA is digested with a restriction enzyme, and ligated. Regions in close spatial proximity will be ligated together, irrespectively of their linear distance in the chromosome. 3C allows the detection of ligation events between 2 target regions by PCR, while the variations 4C and Hi-C by DNA sequencing. 4C probes

contacts between a single region with the rest of the genome and Hi-C detects ligation events between all regions of the genome simultaneously (Denker & de Laat, 2016).

Hi-C experiments provided an important view of long-range chromosomal interactions using an imaging-independent methodology (Denker & de Laat, 2016; Lieberman-Aiden et al., 2009). When analyzed on a chromosomal level, Hi-C maps display a checkerboard pattern, indicating that discrete genomic segments over an entire chromosome show an enriched interaction frequency in a cell population (Fig. 4B) (Lieberman-Aiden et al., 2009). Principal-component Analysis (PCA) of Hi-C interactions over chromosomes divides the genome into 2 main compartments, the A and B compartments (Fig. 4B). In general, chromosomal regions belonging to the A compartment contain chromatin in an "active" state, i.e. with transcribed genes and epigenetic features that are associated with transcription (Fig. 4B). The B compartment contains chromatin in an "inactive" state, rich in heterochromatin features (Fig. 4B) (Lieberman-Aiden et al., 2009). During the course of differentiation, or upon certain stimuli, whole regions may transition between A and B compartments, reflecting their activity state (Dixon et al., 2015; Phillips-Cremins et al., 2013). Whether A/B compartments and Nuclear Bodies represent the same physical entity in a living cell is under debate, but most of the evidence to date suggests that those represent the same phenomenon by completely orthogonal methods.

Chromosomal organization ultimately provides a scaffold for the regulation of gene transcription. While loci can display long-range physical associations highly linked to their transcriptional activity, the study of chromatin topology beyond chromosome territories, nuclear bodies and A/B compartments is crucial to understand the mechanisms linking chromatin spatial organization and transcriptional regulation.

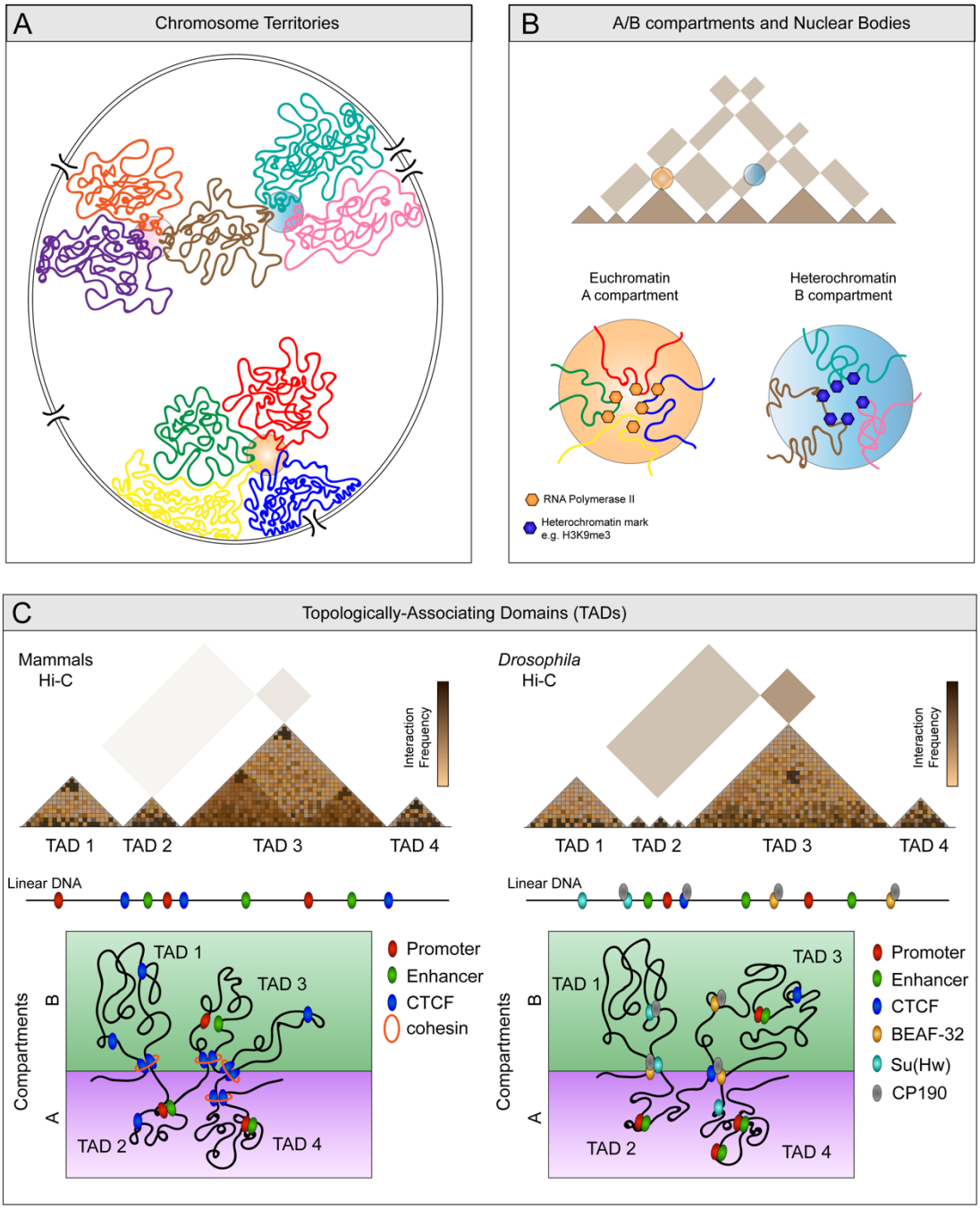

Figure 4**: Overview of the topological organization of the genome across different length scales.**
(A) Chromosome occupy their own territories, with limited intermingling.
(B) Chromosomes organize into A/B compartments (as observed in Hi-C experiments) and Nuclear Bodies (as observed in microscopy experiments).
(C) In the range of dozens of kilobases to few Megabases, chromosomes display TADs. TADs exhibit distinct features in mammals and Drosophila, such as the lack of looping interactions between TAD boundaries in the latter. Compartmental interactions (between A/B compartments) can also be more prominent in a local scale in *Drosophila* TADs. Drosophila TAD boundaries are bound by multiple combinations of insulator proteins (e.g. BEAF-32, CTCF, CP190, etc), while mammals only possess CTCF.

1.2.3 Topologically-Associating Domains (TADs)

Early studies suggested an organization of the genome in domains, in the range of hundreds of Kilobases (Kb) to a few Megabases (Mb). Imaging-based approaches such as labelling nascent DNA with short-pulses of fluorescent nucleotide analogues could detect chromatin domains that are coordinately replicated (Jackson, Hassan, Errington, & Cook, 1993). Such domains were hypothesized to be relevant for molecular processes such as DNA replication, transcription and DNA repair, but due to technical limitations, it was not possible to determine which genomic locations formed such domains, or even if the domain organization was a genome-wide feature or only restricted to a few special loci.

The advent of Hi-C helped to overcome limitations in resolution from microscopy techniques, although the caveat is that Hi-C is based on averaging over hundreds of thousands (sometimes millions) of cells. Hi-C revealed that the genome is organized as domains in the Kilobase to Megabase range. The first reports for the existence of such structures in a genome-wide scale named these "Topologically-Associating Domains" (TADs) (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Nora et al., 2012; Sexton et al., 2012). TADs, as detected by Hi-C, correspond to segments of the genome that self-interact with a higher frequency than with neighbouring regions (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). The organization of the genome into TADs has been reported across eukaryotes of diverse species, and even prokaryote genomes exhibit a TAD-like organization (Dekker & Heard, 2015). These domains can vary in size across species: domains in human and mouse cells often span several hundred Kb, up to 1-2Mb, while in flies they range from ~40-300 Kb (Dekker & Heard, 2015). TADs have been hypothesized to play a regulatory role in gene expression, as genes within TADs are often co-expressed in certain biological conditions (e.g. in a developing tissue) (Nora et al., 2012; Symmons et al., 2014).

TADs are separated, or insulated, by boundaries or inter-TAD regions. Visual inspection of TAD boundaries in Hi-C matrices reveals that some boundaries can be "sharp", while other boundaries extend to several Kb, rather resembling unstructured regions. High-resolution Hi-C studies showed that some unstructured regions actually consist of very small domains, sometimes encompassing a single gene (Rowley et al., 2019). Whether these smaller domains can also be considered TADs is currently under debate (Beagan & Phillips-Cremins, 2020).

One important distinction between mammalian and *Drosophila* TADs lies in interactions between TAD boundaries. In mammals, TAD boundaries typically exhibit a high

frequency focal interaction, as observed in Hi-C matrices (S. S. Rao et al., 2014) (Fig. 4C). In flies, this high-frequency interaction is rather uncommon, and it usually occurs within TADs rather than between their boundaries (Eagen, Aiden, & Kornberg, 2017; Ogiyama, Schuettengruber, Papadopoulos, Chang, & Cavalli, 2018)(Fig. 4C).

High-resolution Hi-C maps showed that TADs are nested structures, and weaker domains can be found within bigger TADs identified in lower-resolution Hi-C matrices (Hsieh et al., 2020; Krietenstein et al., 2020; S. S. Rao et al., 2014). These are often referred to as "subTADs", and are more dynamic than TADs across tissues and developmental stages (Dixon et al., 2015; Phillips-Cremins et al., 2013). On the other hand, TADs appear largely invariable between cell types in bulk experiments, although there is clear heterogeneity at the single-cell level (discussed below), and time-course Hi-C experiments showed that TADs remain constant during differentiation (Dixon et al., 2015; Phillips-Cremins et al., 2013). However, one caveat of Hi-C experiments is the failure to capture how dynamic chromatin topology in fact is, for example within a cell cycle or between cells in a homogeneous population.

The increase in resolution and use of Hi-C has led to the identification of different structures in Hi-C maps, such as subTADs, loop domains, compartmental domains, gene domains, etc. A unified nomenclature is still lacking and has been discussed in a number of recent publications (Beagan & Phillips-Cremins, 2020; de Wit, 2020). For simplicity, here we will refer to TADs as in the original definition, as regions of the genome with a higher self-interaction frequency than to its neighbors.

The mapping of TADs on the genome has been aided by algorithms that use different strategies to call TAD boundaries. Even though the precise position of many TAD boundaries is consistently identified by different algorithms, there is obvious divergence in calling the position of some boundaries (Zufferey, Tavernari, Oricchio, & Ciriello, 2018). This occurs due to (1) technical aspects of how Hi-C experiments affect the resolution of Hi-C matrices, but also due to (2) biological dynamics of TADs. An important defining property of a TAD boundary is its ability to insulate two TADs, and this is usually measured by an "insulation score" in Hi-C experiments. The insulation score measures a change in the directionality of chromatin contacts, by quantifying the interactions passing across each genomic bin. It defines boundaries by identifying a local minima (Lajoie, Dekker, & Kaplan, 2015).

Hi-C experiments are typically performed with millions of cells as starting material, and the structures observed in Hi-C maps are an average of chromatin interactions across the cell population. Meanwhile, microscopy-based experiments, such as DNA FISH reveal an extensive level of plasticity in intra and inter-TAD chromatin interactions across single cells

(Fraser, Williamson, Bickmore, & Dostie, 2015). Questions such as whether TADs exist in single cells, or are an entity derived from averaging of interactions across a cell population, cannot be addressed by conventional Hi-C or DNA FISH experiments, thus requiring new approaches, such as 3C-based techniques in single cells, 3C on native chromatin (without fixation), and super-resolution microscopy coupled to chromatin tracing (McCord, Kaplan, & Giorgetti, 2020). Rather than being static structures, recent studies indicate that TADs represent an ensemble of different chromatin conformation configurations across multiple cells (Bintu et al., 2018). While domains can be observed in single cells by both super-resolution microscopy and single-cell Hi-C, these domains do not map exactly to the same positions as TADs in a Hi-C map. However, the position of those domain's boundaries is statistically more favorable to be located on the Hi-C-detected boundaries.

Overall it is clear that the genome is organized as dynamic domains in the Kb to Mb scale, and this organization is flexible among single cells. However, TADs may not all be formed by the same mechanisms, and may have distinct roles in gene expression regulation.

1.3 Putative mechanisms of TAD formation

1.3.1 Focus on TAD-Boundaries: Loop-extrusion Model

The loop extrusion model is currently the most well-accepted model to explain TAD formation in mammals. This model was initially formed by integrating polymer modelling with experimental evidence from studies with two factors: the insulator protein CTCF and the multi-protein complex cohesin (Fudenberg et al., 2016; Sanborn et al., 2015). Up to 86% of all TAD boundaries are bound by the insulator protein CTCF in mammals (S. S. Rao et al., 2014). CTCF binds to an asymmetrical 11 base pairs (bp) motif in DNA. In a typical mammalian TAD, CTCF motifs at the boundaries are arranged in a convergent orientation (Fig. 5A) (de Wit et al., 2015; S. S. Rao et al., 2014). Meanwhile, the cohesin complex does not bind to DNA through a specific motif, but is rather continuously loaded and unloaded from chromatin by accessory proteins (Peters, Tedeschi, & Schmitz, 2008). The loop extrusion model proposes that once the cohesin complex is loaded, it entraps the chromatin polymer, extrudes a loop, until it gets stalled (Fig. 5A) (Fudenberg et al., 2016; Sanborn et al., 2015). The stalling occurs preferentially at sites bound by CTCF in a convergent orientation, thus at TAD boundaries

(Fig. 5A). This retention of CTCF at convergent CTCF sites leads to an increased interaction time between the TAD boundaries, which can be visualized in Hi-C maps as a focal signal between the corners of TADs (Fig. 5A) (Fudenberg et al., 2016; Sanborn et al., 2015). Experimental evidence for the requirement of CTCF and cohesin for TAD formation in mammals comes from studies where CTCF or components of the cohesin complex were removed (Fig. 5A) (Nora et al., 2017; S. S. P. Rao et al., 2017; Schwarzer et al., 2017). Removal of either strongly affects TAD structure, as TADs containing CTCF at their boundaries become undetectable in Hi-C maps (Fig. 5A). In addition to depletion of CTCF in *trans*, genetic deletions of CTCF binding sites at TAD boundaries were also able to disrupt TAD boundaries in some loci (de Wit et al., 2015; Sanborn et al., 2015).

As in mammals, *Drosophila* TAD boundaries are bound by insulator proteins (Cubenas-Potts et al., 2017; Ramirez et al., 2018; Sexton et al., 2012), but with some important differences. In flies, CTCF only binds to a subset of TAD boundaries and the CTCF motif does not exhibit the same convergent orientation (Ramirez et al., 2018; Sexton et al., 2012). In fact, in a given TAD, CTCF may even bind to only one boundary while the other boundary is occupied by different proteins. Different combinations of insulator proteins bind to TAD boundaries (Ramirez et al., 2018; Sexton et al., 2012), and it is not known if a motif orientation code similar to that of CTCF in mammals exists for any of the *Drosophila* insulator proteins. A schematic of *Drosophila* TADs is illustrated in Fig. 5B, along binding of various insulator proteins and their respective motifs.

There is no direct functional evidence to date that insulator proteins regulate genome-wide TAD formation in *Drosophila*. One study tried to address the role of a key insulator protein in flies, BEAF-32. Removal of BEAF-32 via RNAi in a *Drosophila* cell line did not cause major defects in chromosomal conformation as assessed by Hi-C (Ramirez et al., 2018). Combined depletion of BEAF-32 and the transcription factor M1BP caused most TADs to disappear. However, the global loss of TADs observed in that system is likely due to cells arresting in metaphase after M1BP depletion (Ramirez et al., 2018).

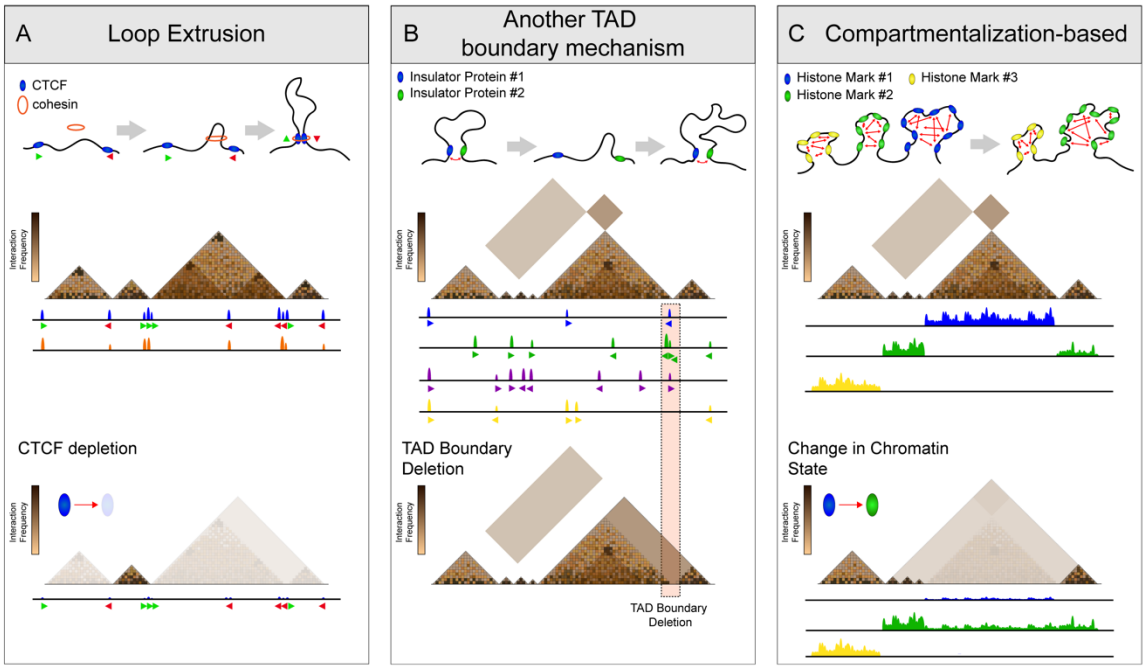

Figure 5 **Mechanisms of TAD formation.**
(A) The loop extrusion model involving CTCF and cohesin can explain the formation of the majority of mammalian TADs. CTCF and cohesin often bind to TAD boundaries, although not limited to these. Depletion of CTCF leads to disappearance of most TADs in population-level Hi-C.
(B) Other mechanisms relying on TAD boundary function might be into play in *Drosophila* TADs and a subset of mammalian TADs. Other proteins bound to TAD boundaries besides CTCF may be important for TAD boundary function. In such a mechanism, precise deletion of TAD boundaries could lead to TAD fusion.
(C) A compartmentalization-based mechanism may play a role in *Drosophila* TADs and a subset of mammalian TADs. Such mechanism would rely on the binding of proteins, or post-translational modification of histones, throughout a whole region, and their alternation along the genome. Changes in binding or erasure of modifications could alter TAD composition.

1.3.2 Focus on TAD-Boundaries: other mechanisms independent of CTCF and cohesin?

Even though a CTCF- and cohesin-based loop-extrusion model offers important insight into a potential mechanism of TAD formation in mammals, it cannot be the only mechanism behind the formation of those structures for the following reasons:

(1) In mammals not all TAD boundaries are bound by CTCF and cohesin.

(2) A subset of mammalian TADs and the majority of *Drosophila* TADs do not exhibit the high-frequency boundary interactions in Hi-C experiments, as the loop extrusion model would predict.

(3) Organisms that do not possess a CTCF homolog are still able to form TADs.

Even though a loop extrusion mechanism orchestrated by CTCF may not be required for TAD formation in *Drosophila*, it is possible that loop extrusion operates with distinct proteins in flies, since many insulator proteins exist in flies. However, most *Drosophila* TADs lack the strong focal Hi-C signal at TAD corners in Hi-C maps, indicating that fly TADs may be formed by a different mechanism (Rowley et al., 2017). Such a mechanism may still be TAD-boundary-centered (Fig. 5B). If a TAD boundary-based mechanism operates to form TADs in *Drosophila*, deletions encompassing boundaries would lead to TAD fusion (Fig. 5B). So far, this was indeed observed in at least two loci in flies (Arzate-Mejia, Josue Cerecedo-Castillo, Guerrero, Furlan-Magaril, & Recillas-Targa, 2020; Mateo et al., 2019).

Insulator proteins in both flies and mammals not only bind at TAD boundaries but also within TADs and at unstructured inter-TAD regions (Ramirez et al., 2018; Sexton et al., 2012). Multiple experimental evidence indicates that insulators need to be "compatible" for their communication and function (See for example "insulator bypass" in section 1.1.3). While that compatibility depends on the combinatorial binding of insulator proteins to an insulator, whether such a requirement exists at the level of TAD boundaries is an important open question. If such compatibility exists, one implication is that specific insulator proteins may be required only in a subset of TADs, and not throughout the genome, like CTCF in mammals.

1.3.3 Transcription and TAD boundaries

As both mammalian and *Drosophila* TAD boundaries can overlap with active promoters, transcription has been proposed as a possible driver of TAD boundary formation (Dixon et al., 2012; Sexton et al., 2012).

In mammals, several TAD boundaries overlap with RNA polymerase II and active transcription (Dixon et al., 2012; S. S. Rao et al., 2014). A subset of TAD boundaries colocalize with *tRNA* promoters (and therefore with the binding of RNA polymerase III) (Yuen, Slaughter, & Gerton, 2017). During early mammalian development, transposable retroelements are transiently transcriptionally active. Such active elements co-localize with early-embryo-specific TAD boundaries (Zhang et al., 2019). During differentiation, cell-type specific TAD boundaries often correlate with transcriptional activation and/or transcription-factor binding (Bonev et al., 2017; Stadhouders et al., 2018). However, artificially inducing transcriptional

activation (with a Cas9-based system) to an intra-TAD location during neuronal differentiation did not induce an ectopic boundary (Bonev et al., 2017).

Experimental and modelling approaches in *Drosophila* suggested that the division of the genome into domains could be partially explained by the alternation of transcriptionally active and inactive regions throughout the genome (Rowley et al., 2017; Ulianov et al., 2016). The factors associated with transcriptional activation and repression would help the partitioning of TADs, with active regions enriched in TAD boundaries and inter-TADs, while repressed or inert regions would be enriched within TADs.

In *Drosophila*, TAD boundaries co-localize with promoter elements very often (up to 80%), especially promoters of housekeeping genes (Ramirez et al., 2018; Sexton et al., 2012). The binding of a subset of insulator proteins, such as BEAF-32 and CP190, is enriched at transcriptionally active boundaries, and these proteins have been proposed to act as direct transcriptional activators (Bartkuhn et al., 2009; Jiang, Emberly, Cuvier, & Hart, 2009; Ramirez et al., 2018). Even though other insulator proteins such as CTCF and Su(Hw) also co-localize with TAD boundaries, those boundaries are not enriched at promoters (Ramirez et al., 2018). However, inhibition of transcriptional elongation by RNA polymerase II in *Drosophila* embryos or cell lines had only a subtle effect on genome-wide insulation at TAD boundaries, implying that transcription is not the crucial factor for chromatin topology at the level of TADs (Hug, Grimaldi, Kruse, & Vaquerizas, 2017; Rowley et al., 2017). Importantly, the drugs used in these studies do not abolish RNA polymerase II recruitment, and have only a slight effect on transcriptional initiation. It is possible that formation of the transcription preinitiation complex (PIC) and transcriptional initiation per se are more relevant for TAD boundary function than transcriptional elongation.

The fact that TAD boundaries overlap promoters suggested that transcription may contribute to TAD boundary formation in a subset of mammalian TADs, and in the majority of *Drosophila* TADs. However, even though transcription contributes to some level to insulation in transcriptionally-active boundaries, so far there is no strong evidence for it being essential for boundary function.

1.3.4 Compartmentalization-based mechanisms

Modelling studies and descriptive evidence suggested that a compartmentalization-based mechanism could drive TAD assembly, associated to features such as histone posttranslational modifications and chromatin-bound proteins (Fig. 5C) (Rowley et al., 2017;

Ulianov et al., 2016). Whether such compartmentalization would be driven by liquid-liquid phase-separation or other type of droplet formation is not clear.

Initial Hi-C studies in flies observed that TADs have a dominant type or a combination of epigenetic marks, correlated to their transcriptional status (Sexton et al., 2012). Interestingly, the contact frequency in TADs containing active histone marks and transcribed genes have a faster decay as a function of linear distance in comparison to other TADs (Sexton et al., 2012). This could indicate that chromatin in active TADs is less compact, or more mobile, than in inactive TADs. Conversely, heterochromatin and Polycomb-repressed chromatin are more compact and can form droplets *in vivo* (Boettiger et al., 2016; Erdel et al., 2020). Modelling studies have shown that chromosome structure at the level of TADs can be largely predicted based on data from histone marks and transcriptional status (Rowley et al., 2017). However, chromatin and transcriptional state have not been functionally demonstrated to drive chromatin topology at the level of TADs.

In mammals, further evidence for a compartmentalization mechanism operating at the level of TADs comes from cohesin depletion experiments. In the absence of cohesin, A/B compartments and sub-TAD structures become reinforced, while TADs vanish (S. S. P. Rao et al., 2017; Schwarzer et al., 2017). The remaining structures highly overlap with blocks of histone marks. This suggests that a cohesin-driven loop-extrusion mechanism counteract the chromatin-state compartmentalization of mammalian genomes.

Overall, there is an undergoing debate about which mechanisms regulate TAD formation across species. It is possible that different TADs are formed by different mechanisms, and this could reflect different roles in gene regulation. Genetic manipulations that affect TAD formation are needed to understand which mechanisms drive formation of different TADs.

1.4 TAD dynamics during embryonic development

Following the development of Hi-C, multiple studies characterized chromatin topology in diverse cell types, and during differentiation trajectories (Dixon et al., 2015; Phillips-Cremins et al., 2013; S. S. Rao et al., 2014; Schmitt et al., 2016). The surprising finding from those studies is that TADs are largely conserved in very different cell types throughout development. In *Drosophila*, studies in cell lines originated from different cell types, and in

embryos and larvae at different developmental timepoints agreed with those findings, showing largely invariant TADs (Cubenas-Potts et al., 2017; Eagen et al., 2017; Hug et al., 2017; Sexton et al., 2012).

Therefore, one important question is whether chromatin topology is established at the very first stages of embryonic life (zygote), or at some key stage during early embryonic development. Below, I introduce the early stages of embryonic development in *Drosophila*, with a special focus on gene regulation and chromatin topology during that period.

1.4.1 The Zygotic Genome Activation (ZGA) in Drosophila

In animal species, the beginning of embryonic development occurs in the absence of transcription from the zygotic genome (Vastenhouw, Cao, & Lipshitz, 2019). Instead, all initial developmental processes are regulated by mRNAs and proteins deposited in the oocyte in the female germline (Vastenhouw et al., 2019). The time scale of the initial development varies widely between species, and the start of transcription from the zygotic genome (Zygotic Genome Activation – ZGA) is triggered at distinct relative times in different species during early embryogenesis (Vastenhouw et al., 2019).

In *Drosophila*, following fertilization the zygotic nucleus divides several times before the onset of ZGA (O'Farrell, Stumpff, & Su, 2004). Those divisions occur in the absence of cytokinesis, and therefore are named “Nuclear Cycles” (NC). NCs consist of successions of back-to-back S and M cell cycle phases, without G phases, and are among the fastest animal cell cycles: the initial 10 divisions take around 10 minutes each to complete (O'Farrell et al., 2004). The nuclei collectively move towards the surface of the developing embryo, reaching the embryo’s surface around NC9. The first genes transcribed from the zygotic genome are detectable around NC8 (Schulz & Harrison, 2019). This is considered to be the first wave of genome activation, and lasts until NC14 (Schulz & Harrison, 2019). This wave is named the “minor” wave of ZGA, because only about one hundred genes are transcribed, all of which are very short and usually intron-less (K. Chen et al., 2013; De Renzis, Elemento, Tavazoie, & Wieschaus, 2007). This is likely due to the fact that the interphase during the nuclear cycles is very fast, and long genes would not be fully transcribed and properly spliced before the onset of the next mitosis. Additionally, most genes transcribed in the minor ZGA are targets of the pioneer transcription factor Zelda, which is maternally-deposited (Liang et al., 2008). During the minor ZGA, the nuclear cycles become progressively longer, from ~10 minutes at NC9 to ~20 minutes at NC13. At NC13, poised RNA polymerase II is detectable at thousands of gene

promoters (Blythe & Wieschaus, 2015). At NC14, the cell cycle is arrested, cell membranes are formed between the nuclei ("cellularization") and thousands of transcripts start to be produced from the zygotic genome (Schulz & Harrison, 2019). At the same time, most maternally-deposited mRNAs are degraded, and a proportion (~35%) is replaced by their zygotic versions. Certain mRNAs that were not maternally deposited begin to be transcribed from the zygotic genome. This succession of events transferring the control of embryonic development from maternal to zygotic genes is called the Maternal-to-Zygotic Transition or Mid-Blastula Transition (MZT or MBT), and is illustrated in (Fig. 6). The ZGA is a key event during the MZT.

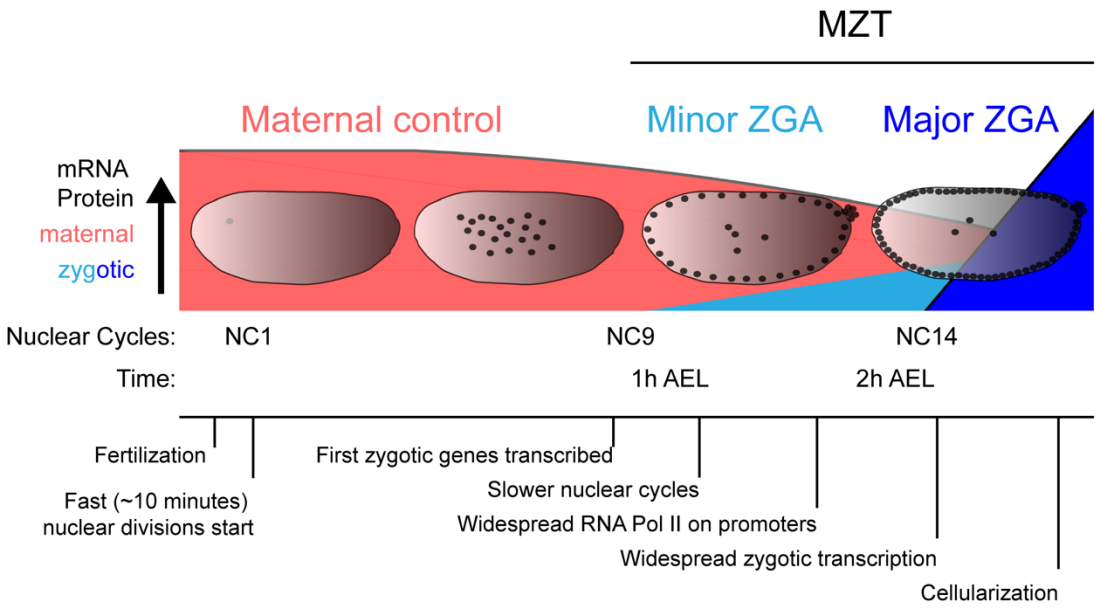

Figure 6: **The Maternal-to-Zygotic Transition (MZT) and Zygotic Genome Activation (ZGA) in *Drosophila.*** Figure adapted from (Vastenhouw et al., 2019).

1.4.2 Genome topology establishment in early embryos

A number of studies have described the state of genome topology during early embryonic development across different species (Vallot & Tachibana, 2020). Microscopy studies initially observed that chromatin organization in the nucleus changes dramatically during ZGA (FOE, 1993). Heterochromatin is detectable at the ZGA in *Drosophila*, along with foci displaying histone marks associated with heterochromatin (Rudolph et al., 2007).

In absolute time, mammalian embryos have their ZGA later than Zebrafish and *Drosophila*, however at a much earlier relative developmental time. In mice, the first transcribed genes are detected already at the late one-cell stage, and major ZGA occurs at late two-cell stage (48 hours after fertilization) (Schulz & Harrison, 2019). This early occurrence of ZGA hampers the study of genome topology in mammals, due to the scarcity of material for Hi-C experiments. Low-input Hi-C performed in mouse embryos across early development revealed that insulation at certain TAD boundaries is detectable at very early stages, although TADs are not readily detectable until after the ZGA (Du et al., 2017). TADs are strengthened gradually with each cell division, as measured by increases in the insulation score between TADs, and compartments follow a similar trend (Du et al., 2017). In Zebrafish, Hi-C analyses detected different dynamics for TAD formation. TADs can be detected shortly before ZGA, but then become undetectable during ZGA and only appear again at a later stage (Kaaij, van der Weide, Ketting, & de Wit, 2018).

A couple of studies have addressed the establishment of genome topology in *Drosophila*. Hug et al., isolated embryos from specific nuclear cycles before and after ZGA, while Ogiyama et al. performed Hi-C in embryos collected from different timepoints (NC9-13, early NC14, late NC14, and later time points) (Hug et al., 2017; Ogiyama et al., 2018). Both studies observe a progressive formation of TADs during the process of MZT (Fig. 7A-D). The earliest discernible genomic structures detected by Hi-C are at genomic regions containing genes activated during the minor ZGA, mostly Zelda targets. These genes form transient TAD boundaries in embryos pre-ZGA (Fig. 7E) and long-range loops with each other. At NC14, with the onset of the major ZGA, most TADs can be detected by Hi-C, and A/B compartments start to be visible (Fig. 7D). The TAD boundaries detected then largely match to boundaries detectable in other development stages, or cell lines (Fig. 7F)

Even though genome-wide studies showed that the major ZGA is a key event in the establishment of chromatin topology, a number of important questions remain unanswered, which I will address in this thesis:

1) Do individual loci have different dynamics? Can this be captured with sensitive locus-specific techniques (e.g. DNA FISH)?

2) Even though TADs are detectable following ZGA, it is not clear how their physical properties change through time. For example, how does the 3D proximity between regions within a TAD change during early and late development?

3) Which trans-acting factors regulate the establishment of TADs during ZGA?

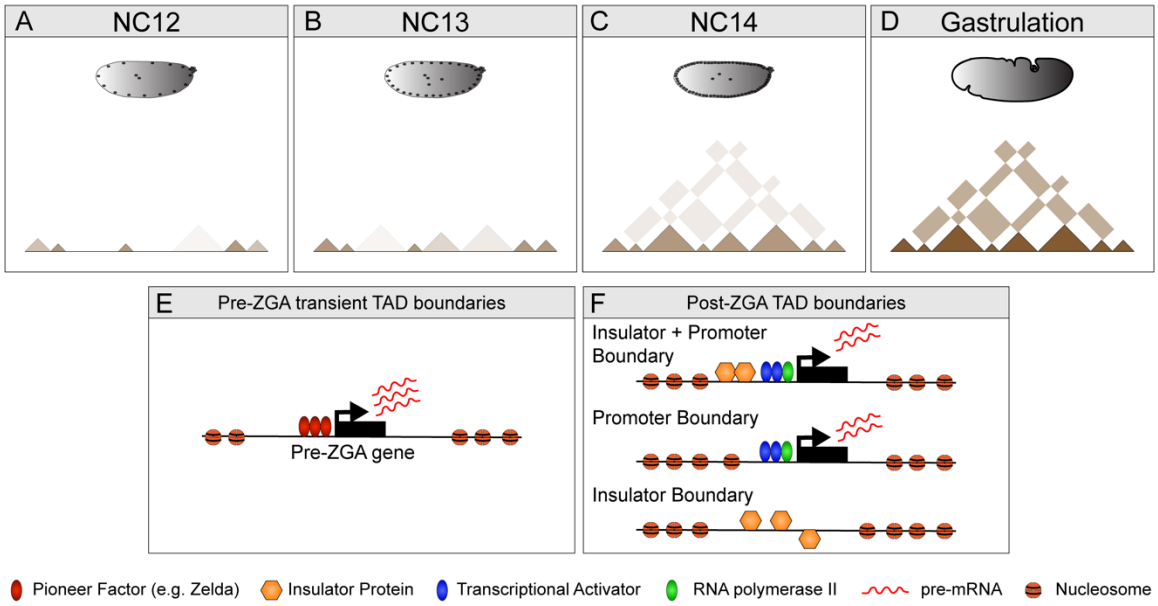

Figure 7 **Chromatin topology in the early Drosophila embryo.**
(A-D) Scheme of the changes in chromatin topology during early *Drosophila* development as assessed by Hi-C, from the minor ZGA (A) up to gastrulation (D).
(E-F) Differences in TAD boundaries detected prior to the major ZGA and after the major ZGA.

1.5 The functional role of TADs in gene regulation

1.5.1 TADs as gene expression regulatory domains

Insertion of reporter genes across random genomic sites can lead to extremely varied expression patterns (e.g. in a developing embryo) (Ruf et al., 2011; Symmons et al., 2014). However, if such insertions occur within the same TAD, the reporter is often expressed in a reproducible pattern (Symmons et al., 2014). In addition, TADs often span genes that are co-regulated in space and time during embryogenesis (Nora et al., 2012; Symmons et al., 2014). Those findings suggest that TADs represent gene expression regulatory domains. Functional studies with manipulated TAD structure *in trans* or *in cis* are revealing more about how TADs regulate gene expression.

1.5.2 Genetic manipulations in trans and their effects in gene expression

The effects of disturbing TADs genome-wide were tested by depletion of a key regulators of TAD formation, CTCF, in mammalian cells. Removing CTCF from embryonic

stem cells (ESCs) (Nora et al., 2012) or a cancer cell line (Hyle et al., 2019) caused a strong effect on topology, eliminating virtually all TADs displaying CTCF binding. This caused the mis-regulation of several hundred genes (Hyle et al., 2019; Nora et al., 2012). Those transcriptional defects did not affect embryonic stem cells differentiation into astrocytes in the absence of CTCF, although their viability was compromised after some days (Nora et al., 2012). Transdifferentiation of B cells into macrophages was also possible in the absence of CTCF, even though macrophages could not elicit a transcriptional response following LPS stimulation (Stadhouders et al., 2018).

In *Drosophila*, CTCF is not required for completion of embryogenesis, as flies without maternal and zygotic CTCF survive until late pupal stages, and a proportion hatches and dies early in adulthood. However, CTCF is required for correct *Hox* gene expression in the correct spatial domain in a developing fly (Gambetta & Furlong, 2018).

Therefore, even though CTCF is strongly required for TAD formation in mammals, and has important insulator roles in flies, it is not required for transcriptional changes associated to multiple cell differentiation events during development. CTCF seems rather required for proper transcriptional responses in specific contexts. However, it is important to note that mice knockout for CTCF die very early during embryogenesis, and therefore CTCF's role in an *in vivo* context might be more crucial than in *in vitro* differentiation systems (Wan et al., 2008).

1.5.3 Genomic rearrangements at TAD boundaries and their effects in gene expression

Genomic rearrangements (deletions, duplications, inversions and translocations) have the potential to alter genome topology and gene expression. For instance, genomic rearrangements that span TAD boundaries can lead to gene mis-expression during embryogenesis, and are associated with human developmental defects (Lupianez et al., 2015; Northcott et al., 2014). These rearrangements can allow the communication between non-cognate enhancer-promoter pairs (enhancer hijacking), by either removing barriers (e.g. TAD boundaries) or directly bringing enhancers close to promoters.

However, in many cases genomic rearrangements do not result in gene mis-expression (Akdemir et al., 2020; Ghavi-Helm et al., 2019). Often, genomic rearrangements that have an effect in gene expression require a perturbation in more than one aspect of genome topology (Fig. 8). This may occur due to a number of different possibilities:

1) Genomic rearrangements that don't disrupt a TAD boundary may not affect the required contact between an enhancer and promoter within the same TAD (Fig. 8B);

2) There is a distance-decay in the frequency of interactions between chromatin regions based on their linear distance, irrespectively of TADs. Genomic rearrangements that disrupt TAD boundaries may need to also bring a distant enhancer close to a promoter to induce strong gene mis-expression (e.g. in large inversions) (Fig. 8C,D) (Despang et al., 2019);

3) Intra-TAD interactions that do not rely on elements at the TAD boundary may be enough to sustain gene expression even after the TAD boundary is disrupted (Despang et al., 2019; Paliou et al., 2019; Williamson et al., 2019).

4) The compatibility of enhancers to their target promoter (e.g. via recruitment of specific trans-acting factors) could sustain gene activation after TAD topology is disrupted.

5) Other insulator binding sites may overtake TAD boundary function after the original boundary is disrupted. Deletions of insulator binding sites could just move the TAD boundary to the next binding site (Despang et al., 2019).

6) The robustness of boundaries can also vary. In the mammalian *HoxD* locus for example, a functional boundary spans several kilobases, and deleting CTCF sites within that boundary does not perturb its function (Rodríguez-Carballo et al., 2017). Only a larger deletion, that removes all CTCF sites, is able to abolish boundary function (Rodríguez-Carballo et al., 2017).

7) Gene expression in TADs that do not rely on a boundary-centered mechanism to be formed (e.g. compartmentalization instead of loop-extrusion) may not be affected by boundary disruptions.

8) Finally, even after disrupting one or more aspects of genome topology in a given locus, some genes may require an even more dramatic gene expression change to cause observable phenotypes (Paliou et al., 2019; Williamson et al., 2019).

Such observations and hypotheses propel important open questions: (1) Which TAD boundaries are functionally important to restrict enhancer-promoter mis-communication? (2) What can be considered a minimal functional TAD boundary? (3) How do TAD boundaries contribute to the robustness of gene expression?

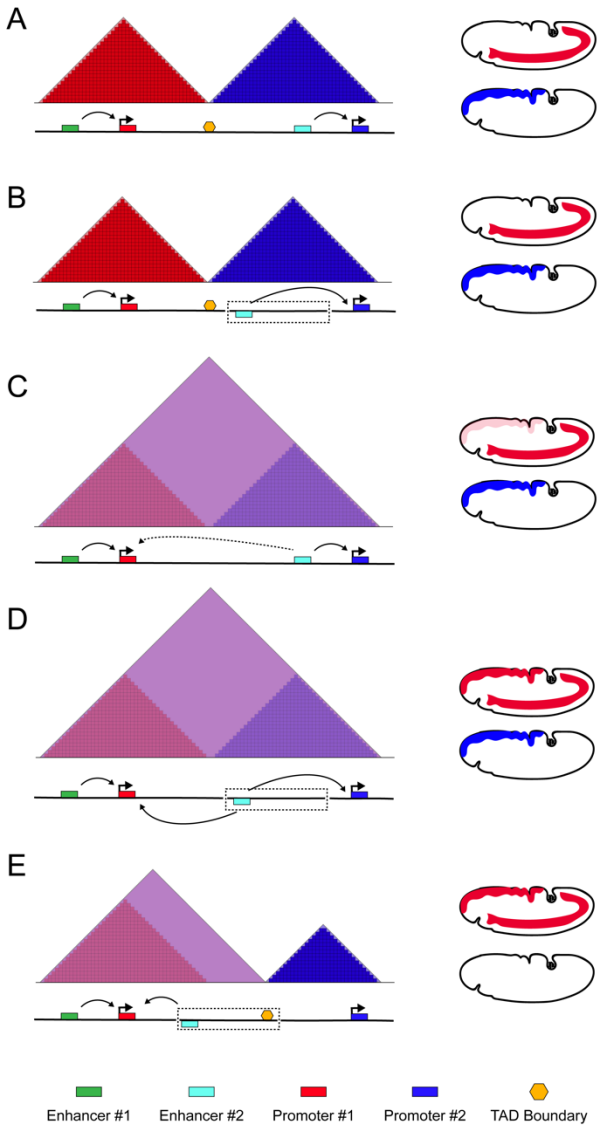

Figure 8: **Effects of genomic rearrangements on TAD structures and gene expression.**
(A) In the WT situation, enhancers are restricted to their target promoters within their TADs, and a boundary separates the red and blue TADs.
(B) Changing the position of an enhancer within the blue TAD does not affect its ability to find its target promoter.
(C) Deleting the TAD boundary leads to TAD fusion and may allow the blue enhancer to find the red promoter. However, the long distance may still impair strong activation of the new target.
(D) A combined loss of boundary and re-positioning of the blue enhancer may cause TAD fusion and allow the blue enhancer to activate the red promoter, while simultaneously activating the blue promoter.
(E) An inversion that spans the boundary, simultaneously bringing the blue enhancer to the red TAD and maintaining a boundary between the two TADs, can lead to simultaneous gain of ectopic expression of the red gene and loss of the blue gene's expression.
Figure adapted from (Ozdemir & Gambetta, 2019)

The field of chromatin topology and gene regulation has seen tremendous advances over the last decade. Hi-C maps with increasing resolution and advanced imaging techniques allow us to probe with increasing detail the dynamics of chromatin topology in multiple biological contexts. However, we still understand little about how topology actually impacts transcription (and vice-versa). After the consolidation of Hi-C and imaging-based techniques for descriptive studies of chromatin topology across multiple species and cell types, the field of gene regulation is increasingly moving towards mechanistical studies, using classic and new genetic tools to question the function of topology in regulating gene expression. In this thesis, I have used genetic manipulations in *cis* and *trans* in combination with imaging- and genomics-based techniques, to dissect mechanisms and function of chromatin topology in developing embryos.

2 Aims

The aims of my PhD are:

1. To evaluate the temporal dynamics of TADs formation during early embryonic development;

2. To perturb different trans-acting factors and assess their impact on TAD establishment and transcription during embryogenesis;

3. To characterize the genomic context-specificity and the essential features for TAD boundary function, and evaluate if ectopic boundaries can lead to gene expression defects and phenotypes.

3 Materials and Methods

3.1 *Drosophila* genetics and husbandry

A list of all *Drosophila melanogaster* transgenic lines used in this thesis is shown in Supplementary Table 1.

3.1.1 RNA interference (RNAi)

Drosophila stocks carrying shRNAs against specific targets, or carrying Gal4 transgenes were obtained from the Bloomington *Drosophila* Stock Center (BDSC) or the Vienna *Drosophila* Resource Center (VDRC). Crosses to obtain BEAF-32- or CP190-depleted embryos were carried out as described in (Staller et al., 2013). In order to obtain embryos efficiently depleted of BEAF-32 or CP190, I compared different Gal4 drivers and placed the flies in different temperatures. Those tests are detailed below.

3.1.1.1 **Selecting Gal4 drivers and temperatures for each target**

Virgin females from the lines carrying either BEAF-32 or CP190 shRNA were crossed to males from either the MTD-Gal4 or Mat-tub-Gal4 lines. Those Gal4 lines drive expression either during all stages of oogenesis (MTD-Gal4) or only during late stages (Mat-tub-Gal4) (Staller et al., 2013). The crosses were kept either at 25°C or 29°C, to compare Gal4 efficiency at different temperatures. In the F1, virgin females that carried one copy of either the CP190 or BEAF-32 shRNAs, and one copy of the Gal4 transgene(s) were crossed to either *yw* males or males carrying the CP190 shRNA or BEAF-32 shRNA (Fig. 9). F2 embryos were allowed to develop at 25°C or 29°C and checked for viability. The results of the test are shown in Table 1.

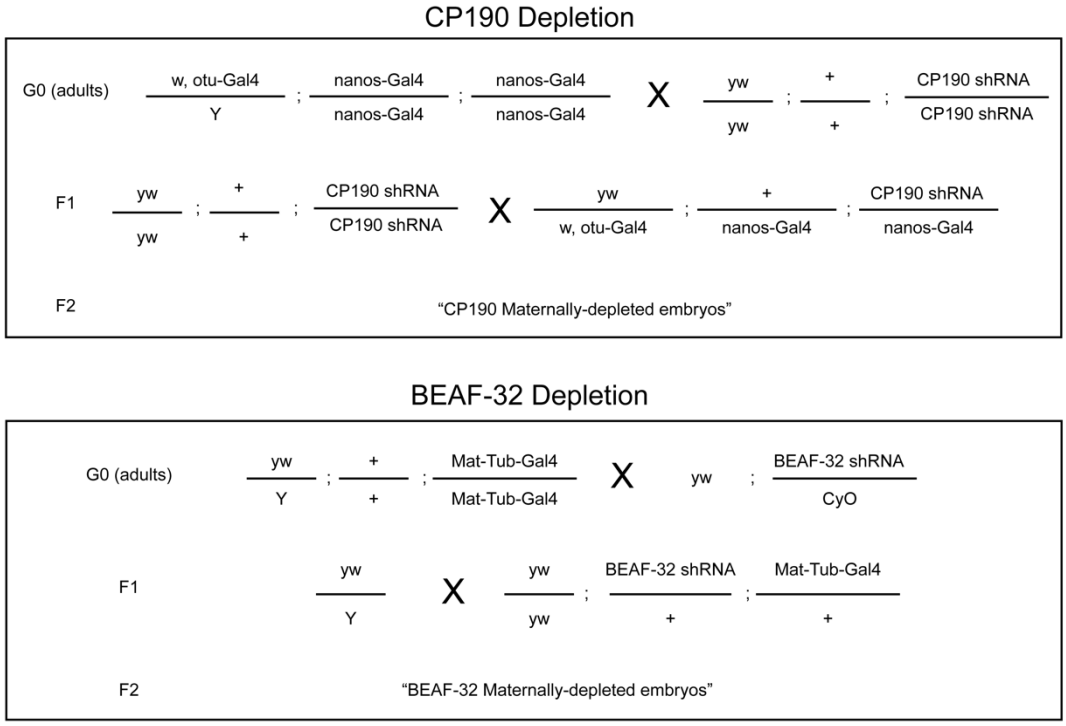

Figure 9: **Crossing scheme of the RNAi strategy.**

Table 1: Viability tests with different Gal4 drivers to induce shRNA expression against BEAF-32 or CP190 in the female germline.

Target	shRNA Line	Driver	25°C	29°C
BEAF-32	#330274	Mat-Tub-Gal4	partial lethality	reduced fertility
BEAF-32	#330274	MTD-Gal4	sterile	sterile
BEAF-32	#29734	Mat-Tub-Gal4	-	-
BEAF-32	#29734	MTD-Gal4	viable	viable
BEAF-32	#35642	Mat-Tub-Gal4	viable	viable
BEAF-32	#35642	MTD-Gal4	viable	viable
CP190	#33903	Mat-Tub-Gal4	-	partial lethality
CP190	#33903	MTD-Gal4	partial lethality	partial lethality
CP190	#33944	Mat-Tub-Gal4	-	-
CP190	#33944	MTD-Gal4	partial lethality	partial lethality
CP190	#35077	Mat-Tub-Gal4	viable	viable
CP190	#35077	MTD-Gal4	viable	viable

Knockout of BEAF-32 and CP190 from the zygote (without affecting the maternal contribution) was shown to impact embryo/larval viability (Butcher et al., 2004; Roy, Gilbert,

& Hart, 2007). Therefore, based on their partial lethal phenotypes and normal fertility, we selected the following lines to use for all experiments described in this thesis:

1) For BEAF-32 depletion, the shRNA line (#330274) was crossed to the Mat-Tub-Gal4 line at 25°C;
2) For CP190 depletion, the shRNA line (#33903) was crossed to the MTD-Gal4 line at 29°C.

3.1.2 Maternal-zygotic nulls and germline clones

The fly lines used to generate CTCF-depleted (maternal/zygotic null) embryos were previously generated by a previous postdoc in the lab (Dr. Maria Gambetta) (Gambetta & Furlong, 2018) and the crossing scheme is detailed in (Fig. 10). In brief, CTCF knockout flies were rescued with an FRT-flanked 5 kb CTCF genomic rescue transgene and developed into viable and fertile adults. The CTCF rescue cassette was excised from male and female germlines through nanos-Gal4-driven expression of UAS-FLP, as described in (Gambetta & Müller, 2014). CTCF-depleted embryos were collected from crosses between those males and females.

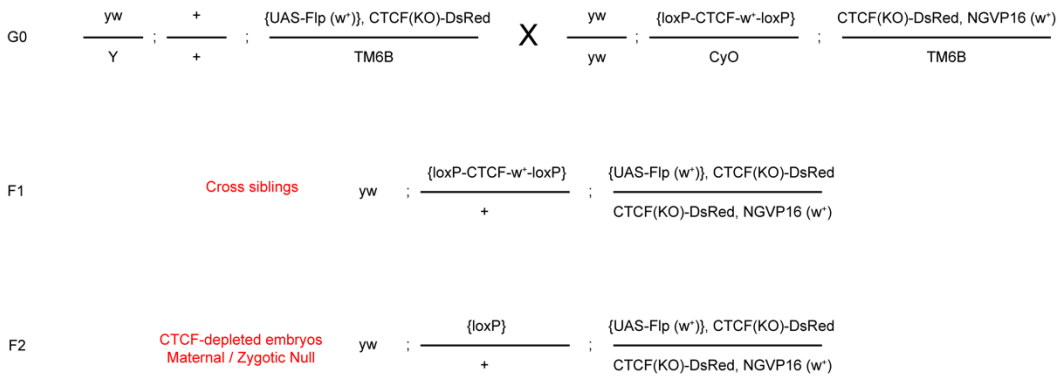

Figure 10: **Crossing scheme of the CTCF maternal/zygotic knockout strategy.**

The fly lines used to generate CP190- and Zld-depleted embryos via germline clones were previously generated. Their generation and crossing strategy to obtain germline mutant clones is detailed in (Chodagam, Royou, Whitfield, Karess, & Raff, 2005; Liang et al., 2008).

3.1.3 CRISPR Deletions

The transgenic fly generation for the CRISPR deletions in the *btsz* locus was performed by an internship student, Songjie Feng, under my supervision.

To generate flies with CRISPR deletions at the Btsz locus, the plasmids obtained as described in section 3.2.2 (containing gRNAs and a donor DsRed marker) were injected into *vasa-Cas9* embryos. These embryos have Cas9 expression in the germline, controlled by regulatory sequences of *vasa.* All injections were performed by Dr. Alessandra Reversi from EMBL's *Drosophila* Injection Facility. Hatching adults were crossed to *yw* females, and the progeny was screened for the DsRed-positive marker. Adults carrying the marker were crossed to flies carrying a GFP-marked 3rd chromosome balancer, and the progeny was assessed for viability. Flies carrying a disruption of the *btsz* promoter or the whole boundary were not able to be maintained as homozygous stocks. Conversely, deletion of the insulator binding site did not affect viability. The DsRed marker was then removed by crosses with a line expressing the Cre-recombinase. Deletions were confirmed by genotyping, as described in section 3.3.1.

3.1.4 TAD Boundary Insertions

In order to generate flies carrying TAD boundary insertions via RCME into MiMIC landing sites (Venken et al., 2011), we first obtained fly lines carrying the MiMIC acceptor cassettes in the selected TADs, from BDSC. We then crossed those with flies carrying the phiC31 integrase and Flp recombinase, to generate the final acceptor lines TAD A and TAD B that were used for both insertion strategies.

3.1.4.1 **TAD boundary Insertions via crosses**

Plasmids containing cloned boundaries along a helper P-element plasmid (see section 6.1.2) were injected into *yw* embryos for P-element transformation. Injections were performed by BestGene Inc. We obtained multiple random insertions into chromosomes 2 and 3. Those lines were used for crosses with flies carrying the acceptor cassette, to achieve RMCE, and the crossing scheme is shown in Fig. 11. The larvae were heat-shocked to induce Flp expression and the excision the donor cassette. If occurring in the germline, the donor cassette could then be integrated into the acceptor cassette through RMCE catalyzed by the phiC31 integrase. In

the F2 progeny, loss of the *yellow* marker indicated successful exchange between the donor and acceptor cassettes. Heterozygous flies could then be crossed to homozygosity.

This strategy displayed very low efficiency of RMCE events, given that after setting up a pilot experiment with multiple crosses (more than 50 independent crosses, for 10 boundaries into one acceptor site), we only obtained two desired insertions. We then decided to focus on a strategy based on injecting the donor plasmids directly into flies carrying the acceptor sites, as described below.

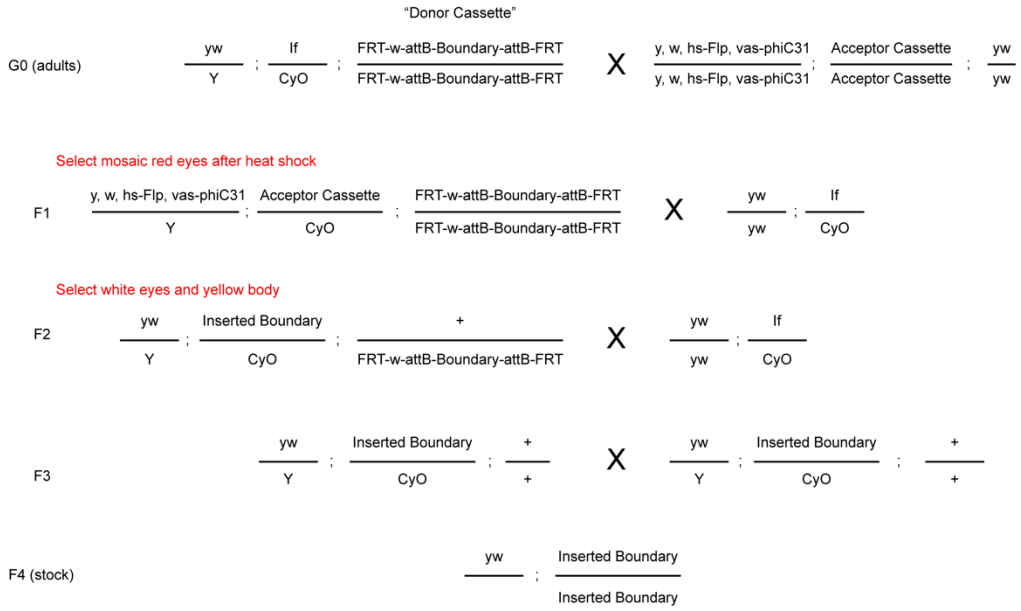

Figure 11: **Crossing scheme of the TAD boundary insertions 'via crosses' strategy.**

3.1.4.2 **TAD boundary insertions via injections**

Plasmids containing cloned boundaries were injected into embryos from the final acceptor lines TAD A and TAD B, which were allowed to develop. All injections were performed by Dr. Alessandra Reversi from EMBL's *Drosophila* Injection Facility. The crossing scheme used is illustrated in Fig. 12. In brief, male adults derived from injected embryos were crossed to *yw* females. F1 males were screened for DsRed-positive eyes, selected and crossed with virgin females carrying a GFP-marked balancer chromosome. F2 DsRed-positive siblings were crossed and the F3 progeny was screened for loss of the GFP-marked

balancer chromosome. When homozygous viable, stocks were kept in homozygosity, otherwise the insertions were kept over the GFP-marked balancer chromosome.

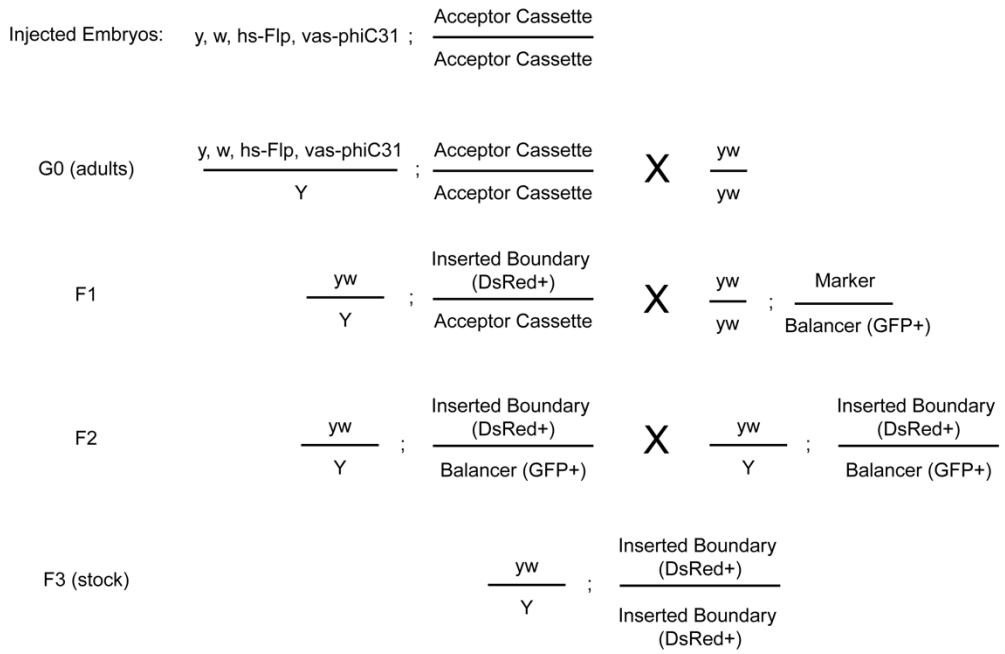

Figure 12: **Crossing scheme of the TAD boundary insertions 'via injection' strategy.**

3.2 Plasmid design and cloning

3.2.1 DNA FISH probe cloning

To guide DNA FISH probe design, Hi-C data was visualized in the Hi-C browser Juicebox (Durand et al., 2016) to search for coordinates of topological landmarks, such as TAD boundaries and anchors of high-frequency looping interactions. DNA FISH probes were then designed to target genome regions of interest, spanning around 7.0 kb in size. The DNA sequence of selected regions was retrieved using the UCSC genome browser (https://genome-euro.ucsc.edu/), and primers targeting selected regions were designed using Primer3 (within the Primer Blast tool from NCBI) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). All the primers used for DNA FISH probe cloning are listed in Supplementary Table 2. PCR was performed using as a template either *Drosophila* genomic DNA, or BACs from the BACPAC Resource Center (https://bacpacresources.org/). Following PCR and gel band DNA extraction using a gel extraction kit (#740609, Macherey-Nagel), the purified DNA was A-tailed and ligated into the linearized pGEMT-Easy vector (#A1360, Promega), following manufacturer's

instructions. The pGEMT-Easy vector with the inserted probe was then used for the fluorescent labelling step to generate the final probes, as described in section 3.6.1. Successful cloning was confirmed by gel electrophoresis and Sanger sequencing with either T7 or Sp6 standard primers.

Alternatively, BACs obtained from the BACPAC Resource Center (https://bacpacresources.org/) were directly used as probes for some of the DNA FISH experiments. Those were fluorescently labelled following the same procedure as for the PCR-derived probes, using the Nick-Translation kit, as described in section 3.6.1.

3.2.2 CRISPR cloning

The cloning for CRISPR deletions was performed by an internship student, Songjie Feng, under my supervision.

To generate deletions in the *btsz* TAD boundary by homology-directed repair, CRISPR donor and gRNA plasmids were constructed, following the strategy for "gene replacement with pHD-DsRed-attP" described in (Gratz, Rubinstein, Harrison, Wildonger, & O'Connor-Giles, 2015). The pHD-DsRed-attP vector, contains a screenable DsRed marker flanked on either side by loxP sites, allowing Cre-mediated removal of the cassette following the successful modification of the target locus. Homology arms can be inserted in that vector, to induce homology-directed repair and integration of the DsRed marker where Cas9 induces DNA breaks.

Four gRNAs were designed targeting the *btsz* TAD boundary. The two first ones to guide Cas9 to the 5' UTR of btsz and to a location upstream of the proximal promoter. The third and fourth ones targeted regions flanking the insulator binding site. This way, by using pairs of gRNAs we could produce three distinct deletions: by combining gRNAs #1 and #2 we could delete the promoter, with gRNAs #3 and #4 we could delete the insulator binding site and with gRNAs #1 and #4 we could delete the whole boundary (promoter + insulator binding site). All gRNAs were checked for potential off-target effects in the *Drosophila* genome using the flyCRISPR Target Finder (tools.flycrispr.molbio.wisc.edu/targetFinder/). gRNA sequences were generated by annealed oligo cloning and inserted into the BbsI site of the pU6-BbsI-gRNA vector. To generate the homology arms, we PCR amplified regions between 2-3kb starting from the directly upstream or downstream of cutting site of each gRNA, and inserted those into the SapI and AarI sites of the pHD-DsRed-attP vector, by using the In-Fusion cloning kit (#639650, Takara Bio USA, Inc.). All primers were designed using Primer3 (within the

Primer Blast tool from NCBI) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/), and are listed in Supplementary Table 2. All insertions were confirmed by Sanger sequencing.

3.2.3 TAD Boundary cloning

Selected TAD boundaries were amplified from *Drosophila* genomic DNA with primers designed using Primer3 (within the Primer Blast tool from NCBI). The 5' end of the primers was modified to be compatible with In-Fusion Cloning into the PstI site in the pBS-KS-attB1-2-DsRed-loxP or pW35-attb12 vectors.

3.2.3.1 **Injection strategy**

For the injection strategy, a plasmid previously assembled in the lab by Dr. Tim Pollex (pBS-KS-attB1-2-DsRed-loxP) was used (Fig. 13A). This plasmid contains attB sites in an inverted orientation. In between the attB sites, there is a reporter gene (DsRed) flanked by loxP sites, and a PstI restriction enzyme site which was used to insert the selected TAD boundaries using the In-Fusion cloning.

3.2.3.2 **P-element strategy**

For the crossing strategy, we used a plasmid previously assembled in the lab, pW35-attb1-2 (Fig. 13B). This plasmid contains attB sites in an inverted orientation. In between those sites, there is a PstI site that we used for inserting the TAD boundaries, using In-Fusion cloning. Adjacent to one of the attB sites, there is a mini-White marker. The attB sites and mini-White marker are flanked by FRT sites. All final plasmids were verified by gel electrophoresis and Sanger sequencing.

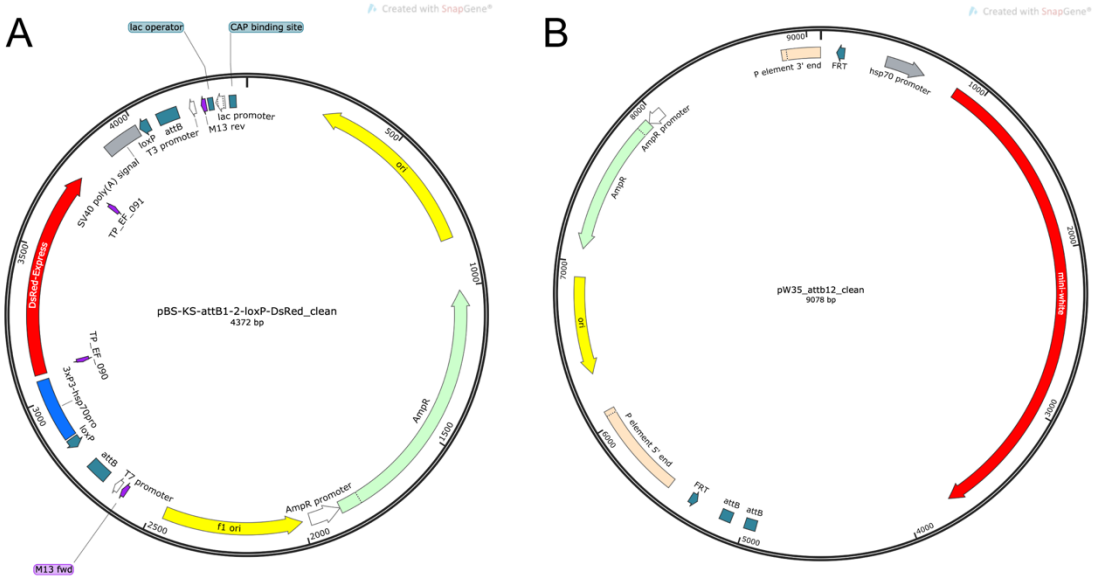

Figure 13: **Vectors used for TAD boundary cloning.**

3.3 Genotyping

3.3.1 Genotyping CRISPR deletions

Genotyping to confirm the CRISPR deletions was performed by an internship student, Songjie Feng, under my supervision, and followed recommendations presented in (Gratz et al., 2015).

Upon obtaining flies carrying the deletions, as described in section 3.1.3, the DNA of single flies was extracted to be used in a PCR reaction. A single PCR reaction was performed to confirm each deletion, by using primers flanking the deletion. The amplicon was then Sanger sequenced to confirm the deletions.

3.3.2 Genotyping the orientation of TAD boundary insertions

I performed PCR to detect the orientation of boundary insertions, based on the original strategy described in the MiMIC resource publication (Venken et al., 2011). Upon obtaining flies carrying the DsRed fluorescent marker, we extracted DNA of single flies, and performed PCR.

I carried out two PCR reactions to confirm the orientation of each insertion (Fig. 14A). Both reactions use a primer annealing to an internal sequence in the donor fragment, which points towards either the 5' or 3' ends of the insertion site, depending on the orientation. Reaction #1 uses a second primer on the left side of the insertion site, while reaction #2 uses a second primer on the right side. Consequently, flies carrying a single insertion can only have a PCR amplicon of ~1.4kb in either reaction #1 ("Orientation 1", or "Or1"), or reaction #2 ("Orientation 2", or "Or2"). Fig. 14B shows the corresponding PCR amplicons from reactions 1 and 2 in single flies which had the boundary 9.1 inserted into TAD B in either "Orientation 1" or "Orientation 2".

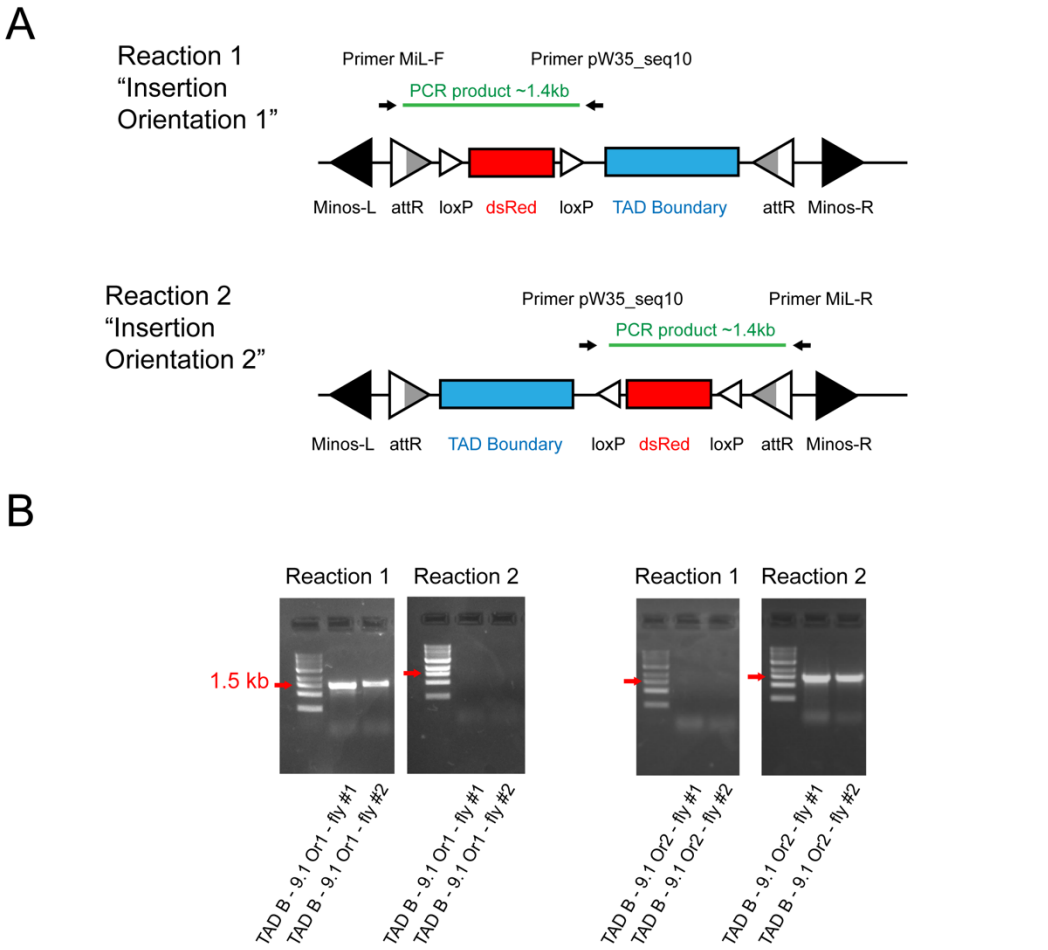

Figure 14: **Scheme of genotyping strategy to verify the orientation of TAD boundary insertions.**
(A) PCR products expected for each reaction designed to detect a specific orientation.
(B) Representative pictures of agarose gels showing the results of PCRs from (A) in distinct single flies.

3.3.3 Genotyping to confirm the identity and location of inserted TAD boundaries

I then performed another PCR reaction followed by Sanger sequencing, to confirm that the insertion occurred in the correct landing site, as well as to confirm the sequence of the inserted boundary. For this reaction, we used primers annealing to sequences flanking the original MiMIC acceptor site, either in TAD A or TAD B (Fig. 15A). This resulted in amplicons ranging from 3.4 to 4.4 kb, depending on the size of the inserted boundary (which ranged from 2 to 3kb). An example of the outcome of such a PCR is illustrated in (Fig. 15B). The DNA from those amplicons were extracted, and sequenced by Sanger sequencing to confirm the identity of the inserted boundary.

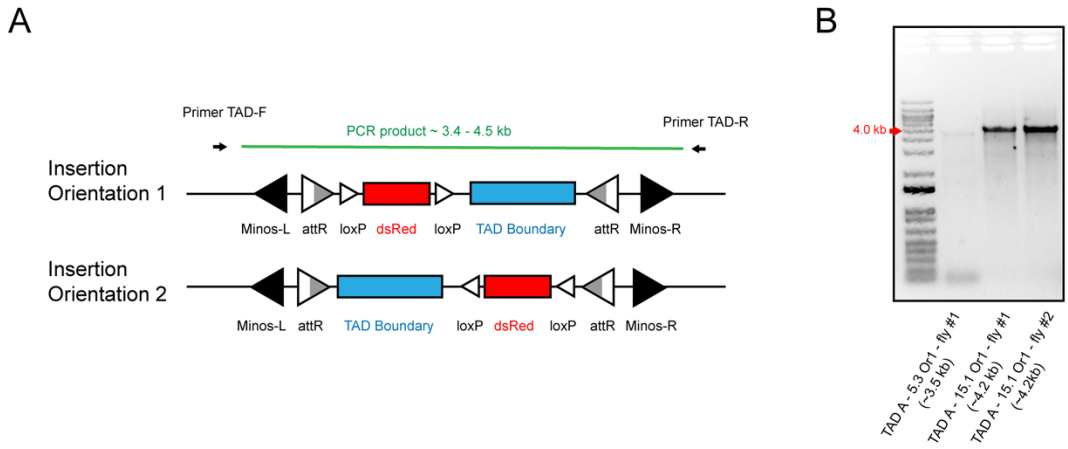

Figure 15: **Scheme of genotyping strategy to verify the insertion of correct TAD boundary into the correct landing site.**
(A) PCR product expected for the single reaction.
(B) Representative picture of an agarose gel showing the results of PCRs from (A) in distinct single flies.

3.4 Embryo Collection

The beginning of the embryo collection procedure was the same for all experiments. Adult flies (males and females) were placed in cages containing apple juice plates with yeast paste. Three consecutive one hour pre-lays were performed to discard embryos derived from eggs that were retained by females. Then, new plates were placed during the desired collection time. If needed, plates were removed and left aging at 25°C up to the desired developmental stage.

Embryos were washed from the plates with dH2O into a container with a sieve. Embryos were dechorionated for 2 minutes by incubation with 50% bleach, washed in dH2O

and PBT 0.1% (PBS 1x + Triton 0.1%), and processed according to the downstream experiment.

3.4.1 Embryo collection for DNA FISH

Dechorionated embryos were placed in a 1:1 mixture of fixative solution (formaldehyde 4% in PBS 1x) and heptane. The tube was vigorously shaken for 20 minutes. The lower phase (fixative) was removed, and the same amount of methanol was added. Embryos were shaken by hand for 1 minute for vitelline membrane removal. Embryos were then washed three times with Methanol 100% and stored in it at -20°C.

3.4.2 Embryo collection for Western Blot and RNA-seq

Dechorionated embryos were transferred to a tube containing ice-cold PBT 0.1%. Embryos were placed in a small petri dish with ice-cold PBT 0.1% (which was kept cold by being placed within a bigger petri dish containing ice) and analyzed under a stereoscope. Embryos at the NC14 stage were identified based on morphological features, as described in (Campos-Ortega & Hartenstein, 2013) and separated with the help of a needle. Embryos in the correct stage were then pipetted into a separate 2.0 ml tube with ice-cold PBT 0.1% and processed according to the experiment.

3.4.3 Embryo collection for ChIP-seq and Hi-C

Dechorionated embryos were placed in a 1:3 mixture of fixative solution (formaldehyde 1.8% in PBS 1x) and heptane. The tube was vigorously shaken for 15 minutes. Of note, specifically with embryos destined to CP190 ChIP-seq, the fixation conditions were different: 3.0% formaldehyde was used, and embryos were shaken during 30min. This was needed, since CP190 does not bind directly do chromatin, and could not be immunoprecipitated under the "standard" fixation. Fixation was stopped by replacing fixative and heptane with fixation stopping solution (PBT + 125mM Glycine), and the tube was shaken by hand for 1 minute. Embryos were spun down at 500 g for 1 minute, and then washed with PBT 0.1% once. Embryos were placed to dry on a Nitex membrane (20 μm) (#03-20/14, SEFAR) on top of absorbing paper. Once dried, embryos were transferred to a new tube, weighted, snap-frozen on liquid nitrogen, and stored at -80°C

3.5 Western Blot

3.5.1 Protein extraction

Manually-selected *Drosophila* embryos at the NC14 stage were kept in a 2.0 ml tube with ice-cold PBT 0.1% (100 embryos were counted and collected per replicate). PBT 0.1% was then replaced by 30 µl of loading buffer. Embryos were lysed manually, with a plastic pestle (#12-141-363, Thermo Fisher Scientific) couple to a protein low-binding tube (#90410, Thermo Fisher Scientific). Then, 70µl of loading buffer were added, for a total of ~100 µl (~1 µl per embryo). The embryo extract was incubated at 95°C for 10 minutes. This was followed by a centrifugation at maximum speed (20,000 g) for 3 minutes. The supernatant was transferred to a new protein low-binding tube, which was then stored at -80°C.

3.5.2 Western blot procedure

All Western blots were performed with biological triplicate samples. Protein samples were heated at 90°C for 10 minutes, and 25 µl were loaded on a pre-cast 4-20% gradient acrylamide gel (Mini-PROTEAN® TGX Stain-Free) (#4568095, Bio-Rad) and ran on 1x laemmli running buffer in an vertical electrophoresis system (#1658004, Bio-Rad). Proteins were ran at 180V until the loading dye reached the end of the gel (approximately 45 min). Since all samples contained the same amount of embryos, we did not measure protein concentration prior to loading, but rather loaded the same volume in all samples. The membrane was then washed in 1x transfer buffer. Proteins were transferred to a Nitrocellulose membrane (#GE10600002, Amersham) at 80V for 1h15 at 4°C, using the Mini Trans-Blot® Cell system (#1703930, Bio-Rad), following manufacturer's recommendations. Protein transfer was assessed with Ponceau staining, followed by five membrane washes with dH_2O, and three washes with 1x PBS + 0.1% Tween. The membrane was then blocked with 5% milk for 1h and incubated with primary antibodies, according to Table 2.

Table 2: Primary antibodies used for Western Blot.

Primary Antibody	Species	Source	Dilution	Incubation
anti-BEAF-32	Mouse	DSHB	1:100 in 5% milk	Overnight at 4°C
anti-CTCF	Rabbit	Rainer Renkawitz's Lab	1:3000 in 5% milk	Overnight at 4°C

anti-CP190	Rat	Pavel Georgiev's Lab	1:5000 in 5% milk	Overnight at 4°C
anti-alpha-Tubulin	Mouse	Abcam #ab7291	1:5000 in 5% milk	2 hours at room temperature

The membrane was then washed three times with PBS + 0.1% Tween and incubated with a secondary antibody diluted 1:10000 in 5% milk for 2 hours at room temperature (Table 3). This was followed by three washes with PBS + 0.1% Tween. Membranes were then developed using SuperSignal West Pico Chemiluminescent Substrate (#34079, Thermo Fisher Scientific) in a ChemiDoc MP Imaging System (#17001402, Bio-Rad).

Table 3: Secondary antibodies used for Western Blot.

Primary Antibody	Species	Source	Dilution	Incubation
anti-Mouse (HRP)	Goat	Cell Signalling #7076	1:10000 in 5% milk	2 hours at room temperature
anti-Rabbit (HRP)	Goat	Cell Signalling #7074	1:10000 in 5% milk	2 hours at room temperature
anti-Rat (HRP)	Goat	Cell Signalling #7077	1:10000 in 5% milk	2 hours at room temperature

3.5.3 Quantification of protein depletion

Western blots were quantified by using ImageJ, via its distribution FIJI (Schindelin et al., 2012). Experiments to evaluate BEAF-32, CP190 and CTCF depletion were performed independently, with depleted-embryos and control (*yw*) embryos independently collected for each experiment. Briefly, grayscale images of the western blots were opened in FIJI, and rectangles covering the bands of the expected sizes were drawn. Histograms displaying the signal intensity across the rectangles were generated, and the signal under the peak corresponding to the correct size band was calculated for each band separately. In each experiment, the mean value of three alpha-Tubulin bands was calculated and used as normalizer. Afterwards, the signal intensity values for insulator proteins was obtained in the same way, and the signal of each band was divided by the normalizer. Values were corrected for background intensity. A t-test was performed to determine statistical significance.

3.6 DNA Fluorescence *In Situ* Hybridization (DNA FISH)

3.6.1 DNA FISH probe preparation

DNA FISH was performed with either PCR-derived probes (~7.0 kb) or BACs (~20kb or 80kb, depending on the experiment). For probe labelling, the Nick Translation Kit (#7J0001, Abbott Bioscience) was used following the manufacturer's instructions, using fluorescently-labelled dUTP nucleotides (#NU-803-550-S, Jena Bioscience; #A32763, #C11397 Thermo Scientific). Probes were stored at -20°C

During DNA FISH we used, per embryo sample, 5 µl of fluorescently labelled probe mixed with 2 µl salmon sperm DNA (#D7656, Sigma Aldrich), 0.1 volume 3 M NaOAc (pH 5.2) and 3 volumes ethanol, and then vortexed. The mixture was precipitated by centrifugation at 17000g and 4°C for 30 minutes. Following a wash with 70% Ethanol, the probe mixture was centrifuged again, for 5 minutes, and the supernatant was discarded. The pellet was left to dry, and x µl of formamide were added according to the following equation: $x = N \times 6.25\ \mu l \div n$ (N = number of embryo samples, n = number of probes per sample). The probe mixture was left shaking at 37°C until prior to hybridization. Just before hybridization, probes were denatured at 75°C for 10 minutes, placed on ice, and mixed with 2x hybridization buffer.

3.6.2 DNA FISH

Embryos stored in 100% methanol were washed sequentially in 75%, 50%, 25% methanol (diluted in 2x SSCT) at room temperature for 5 minutes, each step on a mixer. Embryos were then washed three times for 5 minutes in 2x SSCT at room temperature on a mixer. This was followed by a wash in 20% formamide and a wash in 50% formamide, both for 10 minutes at room temperature on a mixer. Then, embryos were washed two times in 50% formamide at 37°C in an incubator for one hour each, while under rotation. The 50% formamide was removed, and embryos were denatured at 80°C for 15 minutes in a water bath. Embryos were then placed on ice and mixed with hybridization mix containing probes, by flicking the tube. Samples were overlaid with 1 drop mineral oil and allowed to hybridize overnight at 37°C, without shaking. The mineral oil was removed, and embryos were washed two times in 50% formamide (heated to 37°C), while rotating in an incubator at 37°C. Then, embryos were washed with 20% formamide at room temperature in a mixer, followed by three washes in 2x SSCT for 10 minutes each, at room temperature in a mixer. The supernatant was removed, and embryos were mixing with mounting medium containing DAPI, and mounted onto a slide.

3.6.3 Combined DNA/RNA FISH

Combined DNA/RNA FISH experiments were performed in embryos carrying TAD boundary insertions in TAD B. Some of the TAD B lines were homozygous lethal, therefore necessitating to be kept as a stock over a GFP-marked balancer chromosome. The RNA FISH served the purpose of detecting GFP, which marks embryos heterozygous for the boundary insertion. By selecting GFP-negative embryos, we could identify embryos homozygous for the inserted boundaries. For combined DNA/RNA FISH experiments, I followed the procedure described above for DNA FISH, with the only exception that the hybridization mix did not contain RNase, but a RNase inhibitor VRC (2 μM) (#R3380, Sigma Aldrich). After the last wash with 2x SSCT, instead of mounting, I continued with a RNA detection protocol by using the kit HCR v3.0 from Molecular Instruments, following the manufacturer's instructions. Samples were washed thrice with PBT 0.1% , post-fixed in 4% formaldehyde in PBT for 20 min on mixer and then washed five times with PBT 0.1%. Samples were then pre-hybridized in 30% probe hybridization buffer for 30 min at 37°C (without agitation*)*. Afterwards, samples were incubated overnight with probes diluted in 30% probe hybridization buffer at 37°C. Excess probes were removed by washing four times 15 min with 30% probe wash buffer at 37°C. Samples were washed thrice with 5x SSCT at room temperature for 5 min. This was followed by incubation in amplification buffer for 30 minutes at room temperature, and incubation with activated fluorescently-labelled hairpins overnight. Excess hairpins were washed away five times with 5x SSCT at room temperature. The supernatant was removed, and embryos were mixed with mounting medium containing DAPI, and mounted onto a slide.

3.6.4 Image Acquisition and embryo stage identification

Confocal images were acquired at the EMBL Advanced Light Microscopy Facility (ALMF), as z-stacks using a Leica SP8 confocal microscope equipped with a 100x objective (HC PL APO CS2 100x / NA 1.4 / Oil), a 405 nm laser and white-light laser (470-670 nm). The z-stack typically covered a single layer of nuclei, spanning the entire nuclei. Step size for z stacks was set to 200 nm. Leica HyD detectors were used to allow for efficient detection of DNA FISH spots from the small 7 kb DNA FISH probes. A single window in at least 3 embryos was acquired per genotype. To precisely determine the NC stage of early embryos, the amount of nuclei in a 50 μm^2 window was counted, according to (Blythe & Wieschaus, 2015), as shown in Fig. 16.

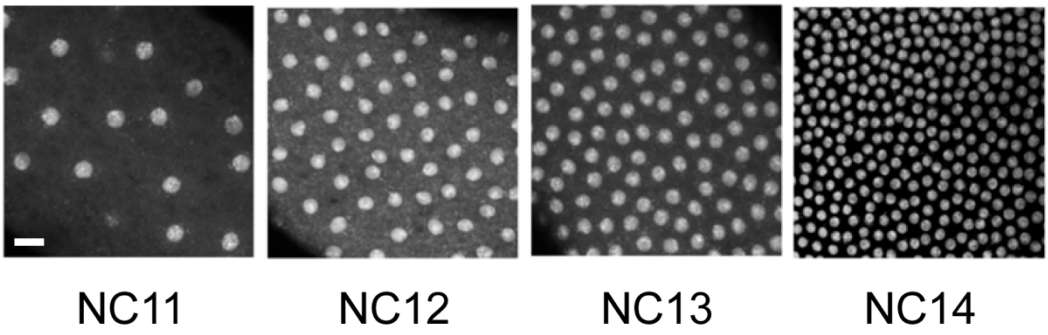

Figure 16: **Staging early *Drosophila* embryos.**
Drosophila embryos were staged according to the density of nuclei at their surface. The precise nuclear cycle could be determined by counting the amount of nuclei in given window. Scale bar = 10 µm.

3.6.5 Quantification of 3D distances in DNA FISH images

Images were first deconvolved by using the Huygens Professional software (SVI), using default parameters. For the quantification of distances between DNA FISH probes in the confocal images, I used a custom plugin developed by Christian Tischer in ALMF ("*Analyze FISH spots*") for use in FIJI. The plugin has two mains functions: (1) detect spots across different channels and (2) calculate the 3D distances between the spots.

1) The x, y and z coordinates of FISH spots are determined based on a manually provided value for signal intensity and background in each channel. Images were visually inspected in FIJI, and intensity and background values were selected. Following their detection, the plugin displays the spots in the image. The multi-point tool was then used to manually select "clusters" of spots in the different DNA FISH channels (two or three channels, depending on the experimental design).

2) After all points are manually marked, the FIJI plugin finds the nearest cluster of spots, and calculates the pair-wise distances between the spots in all channels. More specifically, it calculates two 3D distances between the Center-of-Mass (CoM) of each spot, or uses Difference-of-Gaussians (DoG) to find the centers of spots and calculates their 3D distances. Those distances were used in the DNA FISH plots used throughout this study. For results in chapter I, CoM distance measurements are plotted as DNA FISH spot distances. For chapter II, DoG distances are plotted, as a later analysis revealed that such measurements result in distance distributions more reproducible when quantified by different experimenters, even though the CoM and DoG measurements in general agreed. Multiple images were

independently quantified by Martina Varisco and I, obtaining largely similar findings. A Kolmogorov-Smirnov test was performed to compare the distribution of distances between a given condition and the WT.

3.7 Chromatin Immunoprecipitation (ChIP-seq)

All ChIP-seq experiments were performed in collaboration with Dr. Rebecca Rodriguez-Viales in the lab.

3.7.1 Chromatin extraction

Embryos were placed in ice-cold PBT with protease inhibitors, dounced on ice, and centrifuged at 400 g for one minute. The supernatant was transferred to a fresh tube, and centrifuged again at 1100 g for 10 minutes at 4°C. The supernatant was discarded and the pellet was resuspended in cell lysis buffer containing protease inhibitors. The pellet was dounced, transferred to fresh tubes and centrifuged at 2000 g for 4 minutes at 4°C. The nuclei were resuspended in 1 ml of ice-cold nuclear lysis buffer with protease inhibitors, and incubated on ice for 20 minutes. Nuclei were then sonicated using Bioruptor (Diagenode) to generate 250 bp chromatin fragments. The chromatin was centrifuged at 16000 g and the supernatant containing pure chromatin was then aliquoted into fresh tubes, and used directly for ChIP, or stored at -80C. The quality of chromatin was determined by agarose gel electrophoresis to observe chromatin fragment size distribution.

3.7.2 ChIP procedure

The obtained chromatin was used to perform ChIP as described previously (Bonn et al., 2012). The antibodies' concentrations listed below (Table 4) and IP conditions were optimized to ensure optimal balance between enrichment of positive, antibody bound fragments over negative regions and obtaining sufficient material to generate libraries for sequencing. These steps were controlled by performing quantitative real-time PCR (qPCR) and by measuring DNA concentration with Qubit fluorometer, respectively.

Table 4: Antibodies used for ChIP experiments.

Antibody	Species	Source	Amount	Incubation
anti-BEAF-32	Mouse	DSHB	100 µl	5 µg of chromatin
anti-CTCF	Rabbit	Rainer Renkawitz's Lab	1 µl	5 µg of chromatin
anti-CP190	Rat	Pavel Georgiev's Lab	1 µl	30 µg of chromatin*
anti-Su(Hw)	Mouse	Abcam #ab7291	3 µl	3 µg of chromatin

*embryo fixation is different, see embryo collection section.

The antibodies anti-CTCF, anti-CP190 and anti-Su(Hw) were incubated overnight with chromatin in 900 µl of IP dilution buffer, and 25 µl of magnetic protein A/G beads (#88802, Thermo Fischer Scientific) were added the next day for an additional 3 hour incubation on the rotating wheel.

For the BEAF-32 ChIP, 25 µl of protein G beads were combined with 100ul of the antibody and 300 µl RIPA buffer for 2 hrs. This was followed by two washes with RIPA and resuspension in 100 µl of RIPA, which was added to the purified chromatin, and followed as above.

3.7.3 ChIP-seq library preparation

After ChIP, chromatin was RNase (#10109142001, Roche) treated, reverse cross-linked overnight, and precipitated to obtain pure DNA. Libraries were generated using 10 ng of the starting material. Library preparation was performed using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB #E7645S), following the manufacturer's instructions. The quality of the libraries was assessed on Bioanalyzer (Agilent), and libraries displayed a peak around 350-500 bp. For each ChIP two independent biological replicates were obtained. ChIP-seq libraries were paired-end sequenced with 150 bp paired-end reads using Illumina NextSeq 500 platform at EMBL Genomics Core Facility.

3.7.4 ChIP-seq computational analyses

ChIP-seq computational analyses were performed by Dr. Charles Girardot from Genome Biology Computational Service (GBCS) at EMBL, largely using a local installation of a Galaxy platform maintained by (GBCS), unless stated otherwise.

3.7.4.1 **ChIP-seq data processing**

Illumina sequence files were demultiplexed and converted into FASTQ format. Sequencing quality was assessed using FastQC as provided in Galaxy (Afgan et al., 2016), and ChIP quality was estimated by cross-correlation using the "SPP" tool as suggested by ENCODE ChIP-seq guidelines (Landt et al., 2012). Reads were aligned against the *D. melanogaster* dm6 genome using Bowtie2 (Langmead & Salzberg, 2012) with default parameters. Reads mapping to multiple location (as identified by the XS tag set by bowtie2) and PCR duplicates were removed from Sequence Alignment/Map (SAM) output files using custom code and Picard's MarkDuplicates (http://broadinstitute.github.io/picard), respectively. ChIP and merged input bam files were converted into bigwig format using BamCoverage tool from deepTools library (Ramírez, Dündar, Diehl, Grüning, & Manke, 2014) with the following options: normalization to 1x genome coverage (effective genome size = 121xE6 nt), bin size = 25 nt, length of the average fragment = 150 nt. Input subtracted bigwig tracks were used along this thesis.

Reproducibility of ChIP replicates and final peak selection was achieved using the IDR pipeline presented in (Landt et al., 2012). MACS2 (Zhang et al., 2008) was used as the underlying peak caller with recommended options (--nomodel --pvalue=0.01 --to-large --extsize 150). To evaluate correlation between ChIP-seq samples, the plotCorrelation tool from deepTools library (Ramírez et al., 2014) was used in Galaxy to assess the Spearman correlation between replicates from all conditions.

3.7.4.2 **ChIP-seq Downstream Analyses**

The tool MEME-chip (Ma, Noble, & Bailey, 2014) was used to identify motifs based on the MACS2 peaks, using default parameters. To assess combinatorial binding, custom scripts were developed and utilized IDR-reproducible ChIP peaks obtained with MACS2. To calculate the overlap of ChIP peaks with genomic features, the tool ChIPseeker (Yu, Wang, & He, 2015) was used within the galaxy platform available in the Genome Biology Computational Service (GBCS). To calculate overlap of ChIP peaks specifically with TAD boundaries, a custom script was developed, which utilized IDR-reproducible ChIP peaks obtained with MACS2, and TAD calls as described in section 3.9.2.4.

3.8 RNA-seq

3.8.1 RNA Isolation

For RNA isolation, approximately 100 embryos at NC14 stage were manually-selected (see section 3.4.2) and homogenized in TRIzol LS (#10296028, Thermo Fisher Scientific) with a Cordless Motor for Pellet Mix and pestles (#47747-370, VWR) on ice. RNA was extracted according to the manufacturer's instructions, and the remaining DNA was digested with RNase-free DNase I (#4716728001, Roche) for 30 min. The RNA solution was purified a second time using Agencourt RNAClean XP beads (#A63987, Beckman Coulter).

3.8.2 RNA-seq library preparation

Strand-specific RNA-Seq was performed from 1 μg of total RNA using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (#E7420, NEB) according to the manufacturers' instructions. RNA-Seq was performed in three biological replicates, representing three independent embryo collections. The samples were multiplexed and sequenced on an Illumina NextSeq 500 (75-bp paired end) platform at the EMBL Genomics Core Facility

3.8.3 RNA-seq computational analyses

RNA-seq data processing was performed by Dr. Adam Rabinowitz in the lab.

3.8.3.1 **RNA-seq data processing**

Illumina sequence files were demultiplexed and converted into FASTQ format. Sequencing quality was assessed using FastQC as provided in Galaxy (Afgan et al., 2016). Reads were aligned against the *D. melanogaster* dm6 genome and transcriptome using STAR (Dobin et al., 2013) with default parameters. Reads mapping to multiple locations and PCR duplicates were removed from Sequence Alignment/Map (SAM) output files using Picard's MarkDuplicates (http://broadinstitute.github.io/picard). Read counts per gene were generated using the tool RSEM (B. Li & Dewey, 2011), and the counts were used with DESeq2 (B. Li & Dewey, 2011) using default parameters to compare gene expression between conditions. Genes were considered differentially expressed with a minimum Log2FoldChange of 0.7 and FDR <

0.05 in comparison to WT embryos. Maternally-deposited genes were removed from downstream analysis by using a custom script and a list of maternally-deposited genes generated in (Ghavi-Helm et al., 2019).

3.8.3.2 **RNA-seq downstream analyses**

For all downstream analyses, differentially expressed genes continued to be defined as having a minimum Log2FoldChange of 0.7 and FDR < 0.05 in comparison to WT embryos.

A custom script was used to compare the amount of differentially-expressed genes that is either unique or common between genotypes. Another custom script was used to obtain the position of the TSS of differential genes and calculate their distance to the nearest ChIP peak for a given insulator protein and to TAD boundaries. Only ChIP peaks reproducible between conditions using IDR were used, as described in the section 3.7.4.1. The TAD boundaries were called in Hi-C experiments with WT 2-3h AEL embryos, as described in the section 3.9.2.4.

3.9 Hi-C

All Hi-C experiments were performed in collaboration with Dr. Rebecca Rodriguez-Viales in the lab.

3.9.1.1 **Nuclei extraction**

Embryo collection was performed as described in the section 3.4.3. Nuclei isolation was performed as described in (Bonn et al., 2012) but without any antibody staining. Briefly, staged embryos were homogenized using glass douncers, dissociated through needles and filtered through Nitex membrane (20μm) (#03-20/14, SEFAR) to extract nuclei. The isolated nuclei were stained with DAPI. Nuclei were counted using the BD LSRFortessa™ X-20 Flow Cytometer at the EMBL Flow Cytometry Facility, with assistance of Malte Paulsen and Diana Ordonez. Aliquots of nuclei were prepared in individual 1.5ml tubes, the supernatant was removed, and nuclei pellet were stored at -80°C until use.

3.9.1.2 **Hi-C procedure**

We used a "Bridge-Adaptor in situ Hi-C protocol", corresponding to a in situ Hi-C (S. S. Rao et al., 2014) utilizing biotinylated bridge oligo adaptors to allow for multiplexing between samples, as described in (Ramani et al., 2017). The protocol was adapted and optimized in the lab by Dr. Rebecca Rodriguez-Viales.

First, 15 million frozen nuclei derived from single biological replicates were used as starting material for each replicate. Nuclei were resuspended in fresh ice-cold lysis buffer by pipetting up and down. After 30 min incubation on ice, the nuclei were spun down, the supernatant was discarded and the pellet resuspended in 0.5% SDS by pipetting up and down. The nuclei were permeabilized by incubating at 65°C for 10 min. The SDS was then quenched by adding Triton X-100, and incubating at 37°C for 15 min. This was followed by addition of DpnII (#R0543, NEB) and overnight incubation at 37°C. More DpnII was added in the morning followed by one hour incubation. The nuclei were spun down, the supernatant was discarded and the pellet resuspended in dH_2O and mixed well. This was followed by ligation of the biotinylated bridge oligos to the fragmented DNA, by using T4 DNA ligase HC (#EL0013, Thermo Fisher Scientific) the following annealed biotinylated oligos (*MKVII_F_1*: GATCGAGCTCGAGAA/iBiodT/T and *MKVII_R_1*: CTCGAGCTC) (90 μM). The reaction was incubated at 16°C overnight, and EDTA was then added to stop the ligation. Nuclei were spun down and washed four times with a solution containing SDS and BSA. At this point, the biotinylated bridge adaptors oligos were ligated to the DNA fragments, but not ligated to each other due to the lack of a phosphate group at their ends. Therefore, nuclei were treated with PNK for 1 hour at 37°C, and a new ligation was performed, at room temperature for 4 hours. Then, DNA was extracted, treated with Proteinase K and de-crosslinked. DNA was purified, treated with RNase (#10109142001, Roche) , and sheared to ~200-400 bp with Bioruptor Pico. The resultant biotinylated 3C library was size-selected with AMPure XP DNA Purification Beads (#A63881, Beckman Coulter), resuspended in 100μL of 10mM Tris-Cl pH 8.0 and quantified with Qubit, displaying a concentration in the range of 1-6 ng/μl. Biotinylated DNA fragments were then captured with 30 μl of Dynabeads Streptavidin M-280 (#11205D, Thermo Fisher Scientific), followed by washes and resuspension in 40 μl of 10 mM Tris-Cl pH 8.0. The Hi-C sequencing library was prepared with Accel-NGS 2S Plus DNA Library Kit for Illumina (#21024, SWIFT Biosciences), following manufacturer's instructions, with 12 PCR cycles. Libraries were quantified with Qubit, and their quality was assessed on Bioanalyzer,

displaying a peak around 350-500 bp. For each condition two independent biological replicates were performed. Hi-C libraries were paired-end sequenced with 150 bp paired-end reads using a Illumina NextSeq 500 platform at EMBL Genomics Core Facility.

3.9.2 Hi-C computational analyses

Hi-C data computational analysis was performed by Dr. Charles Girardot largely using a local installation of a Galaxy platform maintained by GBCS, and a HiCExplorer suite (Ramirez et al., 2018) installed in Galaxy, unless stated otherwise.

3.9.2.1 **Hi-C data processing**

Illumina sequencing files were demultiplexed and converted into FASTQ format. Sequencing quality was assessed using FastQC as provided in Galaxy (Afgan et al., 2016). Reads were aligned against the *D. melanogaster* dm6 genome using BWA-MEM (H. Li & Durbin, 2009) with default parameters. Reads mapping to multiple location (as identified by the XS tag set by bowtie2) and PCR duplicates were removed from Sequence Alignment/Map (SAM) output files using custom code and Picard's MarkDuplicates (http://broadinstitute.github.io/picard), respectively. Computational generation of Hi-C matrices was performed using the HiCExplorer tool, similarly to described in (Ghavi-Helm et al., 2019; Ramirez et al., 2018). BAM files were filtered and processed using hicBuildMatrix with options –restrictionSequence GATC --danglingSequence GATC --binSize 1000 --skipDuplicationCheck. The resulting contact matrices were summed across two biological replicates and normalized by iterative correction using hicCorrectMatrix and taking only chromosomes 2, 3, 4, X and Y.

3.9.2.2 **Hi-C quality control**

Basic metrics such as number of duplicated read pairs, dangling ends, and same fragment ligations were extracted from the Hi-C matrices using the MultiQC tool (Ewels, Magnusson, Lundin, & Käller, 2016). The proportion of short-range, long-range and trans-chromosomal contacts was also obtained with that tool. To further verify the quality of the Hi-C experiments, the tools hicPlotDistVsCounts and hicCorrelate from the HiCExplorer suite were used. In brief, hicPlotDistVsCounts compares the chromatin contact enrichment at different genomic ranges / distances up to whole chromosomes, between multiple Hi-C matrices. hicCorrelate computes pairwise correlations between different Hi-C matrices. The

correlation is computed taking the values from each pair of matrices and discarding values that are zero in both matrices. To further evaluate the correlation between Hi-C matrices with additional tools, we used GenomeDISCO (Ursu et al., 2018), HiCRep (T. Yang et al., 2017), HiC-Spector (Yan, Yardimci, Yan, Noble, & Gerstein, 2017) and QuASAR (Yardımcı et al., 2019).

3.9.2.3 **Insulation Score calculation and comparisons**

Insulation scores for each Hi-C individual and merged replicates were obtained with the hicFindTADs tool of the HiCExplorer suite (Ramirez et al., 2018). The insulation score corresponds to a TAD-separation score that was measured using the z-score of the Hi-C matrix and it was defined as the mean z-score of all the matrix contacts between the left and right regions. Hi-C data at 5kb resolution and a FDR < 0.05 for multiple testing correction were used to generate the insulation scores.

3.9.2.4 **TAD calling**

To call TADs from Hi-C data, we used an approach based on the TAD separation score, using hicFindTADs from HiCExplorer with default options (Ramirez et al., 2018). We used hi-C matrices at 5kb resolution and an FDR < 0.05 for TAD calling. We obtained the TAD separation score profiles as described above, for several different window sizes, ranging from 50–195 kb. Those different TAD separation scores were then averaged into an aggregate insulation score profile. The local minima of those scores were used to call TADs.

4 Results I

Investigating the temporal dynamics and trans-acting factors regulating chromatin topology establishment in the early *Drosophila* embryo

4.1 Chromatin architecture at the level of TADs is established at the onset of the major Zygotic Genome Activation (ZGA)

To understand when chromatin topology features such as TADs and long-range looping interactions are formed during early embryonic development in *Drosophila*, I performed 3D DNA FISH in tightly staged wild-type embryos spanning the first developmental stages (0-3h After Egg Laying (AEL)). Such staged collections contain embryos before and after the major ZGA, spanning all the nuclear cycle (NC) stages from fertilization (NC1) up to NC14, when the major ZGA takes places (see 1.4.1). I could confidently pinpoint the precise stage of single embryos, within a 10 minutes window, by quantifying the amount of nuclei in a given spatial window at the surface of the embryo (see 3.6.4).

I designed short (~7kb) DNA FISH probes designed to target sites of interest (e.g. TAD boundaries, loop anchors, intra-TAD regions, etc). To guide the DNA FISH probe design, I used very high resolution Hi-C data from WT 2-4h AEL embryos previously generated in the lab (unpublished). This Hi-C dataset comprises a mixture of embryos undergoing the major ZGA and gastrulation, and displays many features of chromatin topology (e.g. TADs, loops and compartments) that are also detectable later in development and in *Drosophila* cell lines (Ramirez et al., 2018; Sexton et al., 2012). After performing DNA FISH, I acquired confocal microscopy images and used a FIJI plugin developed at the ALMF in EMBL to obtain the position of the Center-of-Mass (CoM) of the DNA FISH probes (see 3.6.5). Then the plugin calculated the 3D distances between the probes CoMs across multiple cells, and I compared their distributions among WT embryos of different stages.

4.1.1 TADs containing high-frequency looping interactions

I initially selected two TADs containing genes important for embryonic development:

one TAD contains the paralogous genes *scylla* (*scyl*) and *charybde* (*chrb*), while the other TAD contains the paralogous genes *teashirt* (*tsh*) and *tiptop* (*tio*).

The *scyl-chrb* TAD contains a three-way high-frequency looping interaction that can be detected in Hi-C experiments with 2-4h AEL embryos, spanning the blastoderm and gastrulation stages. I compared the 3D distances between 3 probes targeting linearly equidistant genomic regions in the *scyl-chrb* locus (Fig. 17A). Two probes overlap the anchors of a high-frequency looping interaction spanning 235kb within the TAD, detectable in Hi-C experiments from 2-4h *Drosophila* embryos. The other probe targets a region outside of the TAD, 235kb from one of the loop anchors, and was used as a chromosomal equidistant control. The long-range loop connects the *scyl* and *chrb* genes, but it is not known whether this loop is involved in transcriptional regulation between those genes. In NC11-NC12 embryos (approx. 1-2h after egg laying [AEL]), the distribution of 3D distances between the probe pair targeting the loop anchors is not significantly different from the distribution of distances between the *chrb* anchor and the equidistant control region (Fig. 17B,C). However, in NC14 embryos (approx. 2-3h AEL) the median distance between *scyl-chrb* probes is significantly smaller in comparison to the distance across the TAD boundary (Fig. 17B,C), with 15.7% of nuclei closer than 250 μm, compared to 6.3% of the control probe. Interestingly, in later stage embryos (6-8h AEL) the median distance between *scyl* and *chrb* is even smaller than in NC14 embryos (Fig. 17 B), with 46.5% being within 250 μm or closer. This suggests a substantial reduction in cell-to-cell heterogeneity in TAD structure at mid-embryogenesis

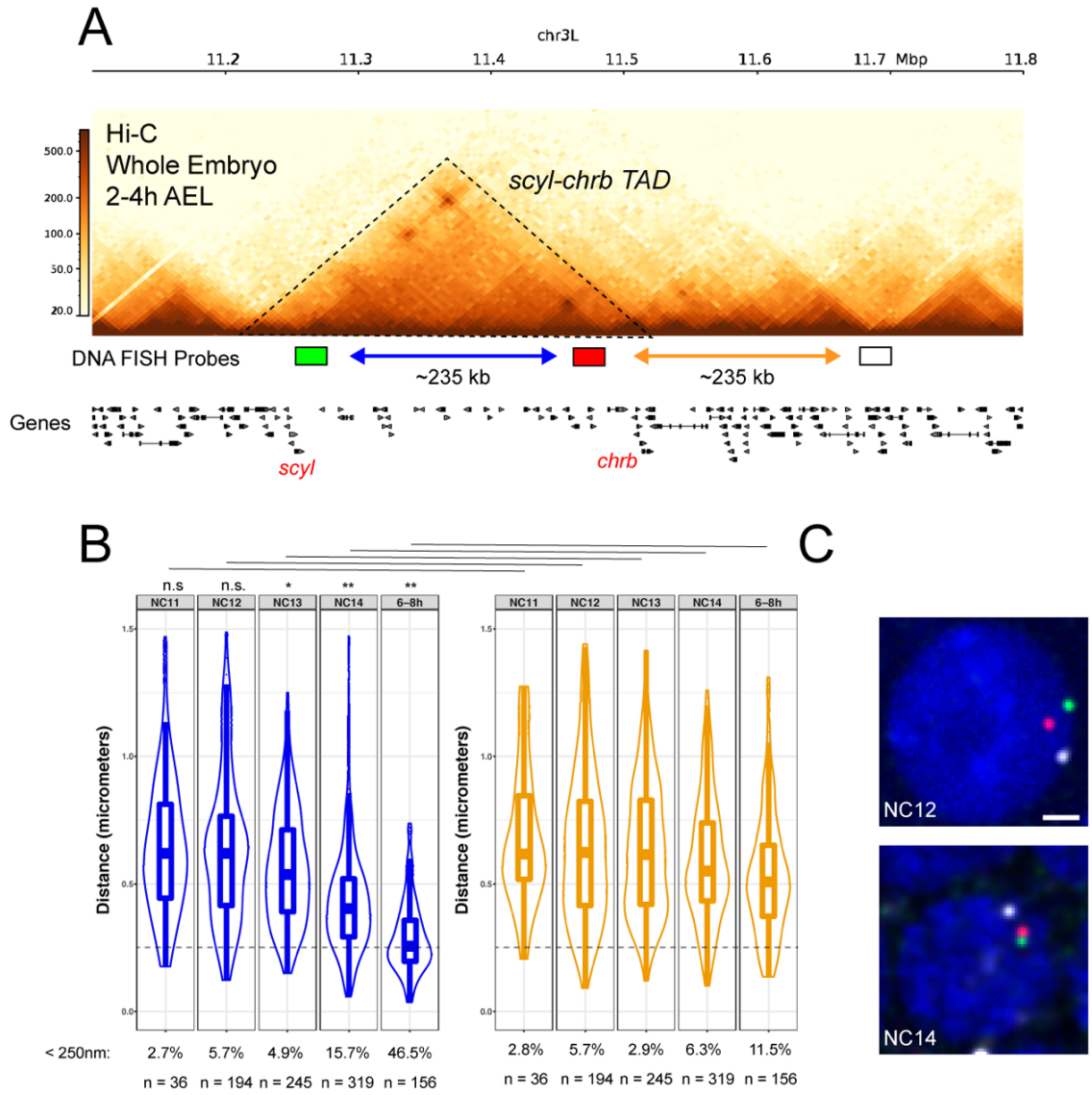

Figure 17: **DNA FISH to evaluate chromatin topology at the *scyl-chrb* locus during early embryogenesis.**
(A) Hi-C from 2-4h AEL embryos previously generated in the lab (unpublished). The position of the DNA FISH probes is shown below the Hi-C matrix. Blue and orange arrows indicate intra-TAD and inter-TAD comparisons, respectively.
(B) 3D distance distributions between intra-TAD probes (blue) and inter-TAD probes (orange) across the different indicated stages. n = number of alleles measured in at least 3 embryos.
(C) Representative single confocal slice showing the DNA FISH probes from (A) labelling a single region. Scale bar = 1 µm.
A Kolmogorov-Smirnov test was applied to evaluate statistical significance in (B): n.s. = non-significant, *p < 0.01, **p < 0.001.

The *tsh-tio* TAD also contains a three-way looping interaction, detectable in Hi-C experiments with 2-4h AEL embryos. I designed DNA FISH probes in a similar way, with two

probes targeting the anchors of the intra-TAD loop (~190 Kb), and a third probe equidistant outside of the TAD (Fig. 18A), and performed the same temporal analysis comparing the 3D distances between pairs of probes. As observed in the *scyl-chrb* locus, the proximity between the *tsh* and *tio* anchors, in comparison to the control probe, increased only at NC14 (Fig. 18B,C), indicating that the TAD is formed concomitantly to the major ZGA. Again, the median distance and the spread of the distance distribution between the loop anchors becomes even smaller at 6-8h AEL (Fig. 18B,C).

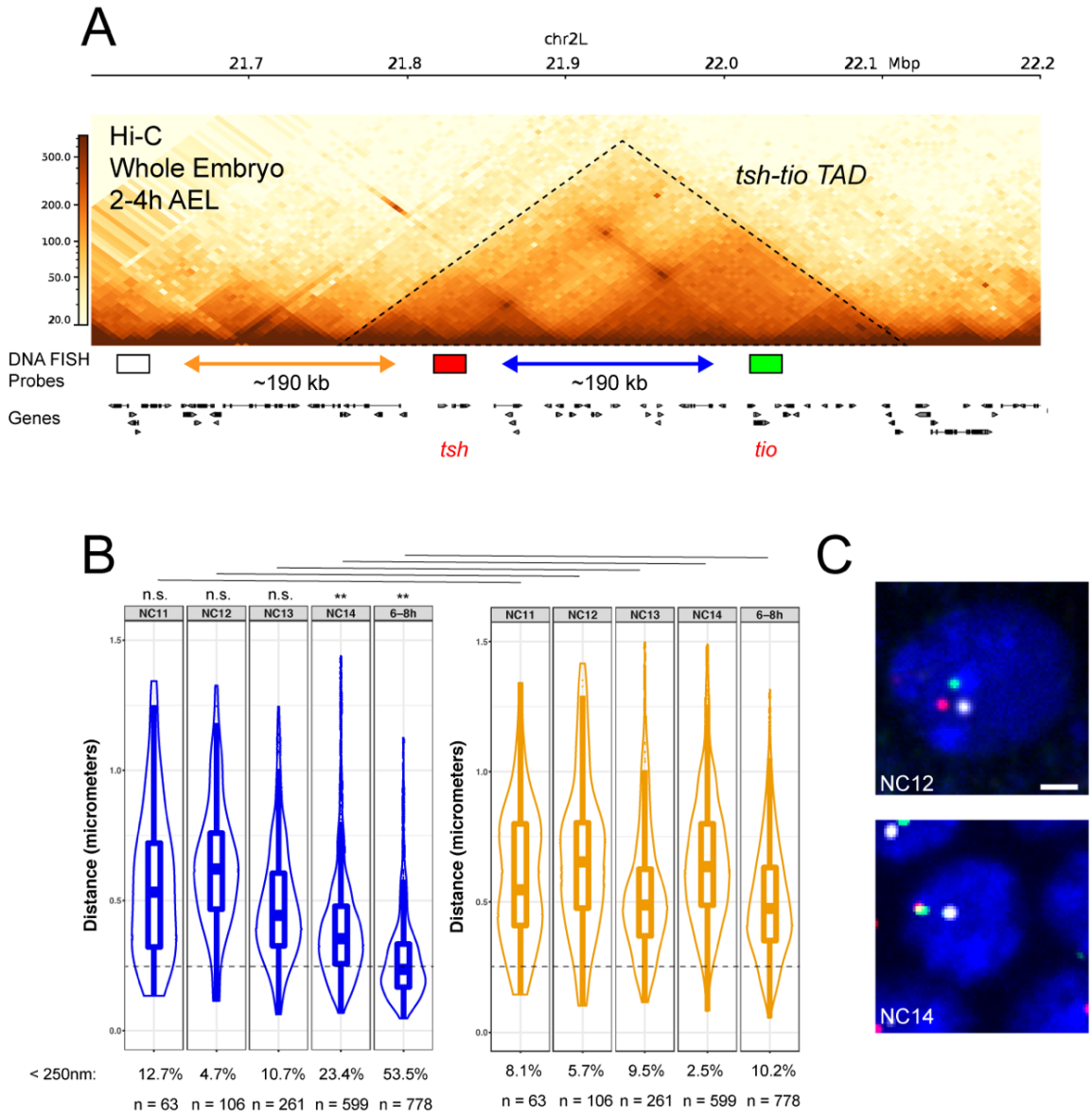

Figure 18: **DNA FISH to evaluate chromatin topology at the *tsh-tio* locus during early embryogenesis.**
(A) Hi-C from 2-4h AEL embryos previously generated in the lab (unpublished). The position of the DNA FISH probes is shown below the Hi-C matrix. Blue and orange arrows indicate intra-TAD and inter-TAD comparisons, respectively.

(B) 3D distance distributions between intra-TAD probes (blue) and inter-TAD probes (orange) across the different indicated stages. n = number of alleles measured in at least 3 embryos.
(C) Representative single confocal slice showing the DNA FISH probes from (A) labelling a single region. Scale bar = 1 μm.
A Kolmogorov-Smirnov test was applied to evaluate statistical significance in (B): n.s. = non-significant, *p < 0.01, **p < 0.001.

4.1.2 TADs that do not contain high-frequency looping interactions

The majority of TADs in *Drosophila* do not contain a high-frequency looping interaction, as present in the *scyl-chrb* and *tsh-tio* loci, at least with the resolution of Hi-C data currently available. Given that, I decided to evaluate the timing of chromatin topology establishment in a locus that does not contain such loops. I selected a large TAD (~350 Kb), whose left boundary overlaps with the gene *Lamin B receptor* (*LBR*). Moreover, while the previous TADs contained developmentally regulated genes, this TAD is gene poor, and largely transcriptionally inactive during embryogenesis.

I used the same experimental strategy as for the TADs with long-range loops. Two DNA FISH probes were designed targeting regions inside the *LBR* TAD and a third probe outside of the TAD in an equidistant chromosomal distance (control) (Fig. 19A). Again, I detected a smaller distribution of 3D distances between the 2 intra-TAD probes in comparison to the control probe across the TAD boundary in NC14 embryos (Fig. 19B,C). Both intra-TAD and inter-TAD distance distributions are not significantly different prior to MZT (NC12) (Fig. 19 B,C).

Taken together, these findings suggest that the *Drosophila* genome is in a relatively unstructured state prior to the ZGA, reaching a more organized state containing TADs and high frequency long-range loops, at the onset of ZGA. This data agrees with recently published findings showing the appearance of TADs (Hug et al., 2017) and some looping interactions (Ogyiama et al., 2018) concomitantly to the major ZGA in *Drosophila*.

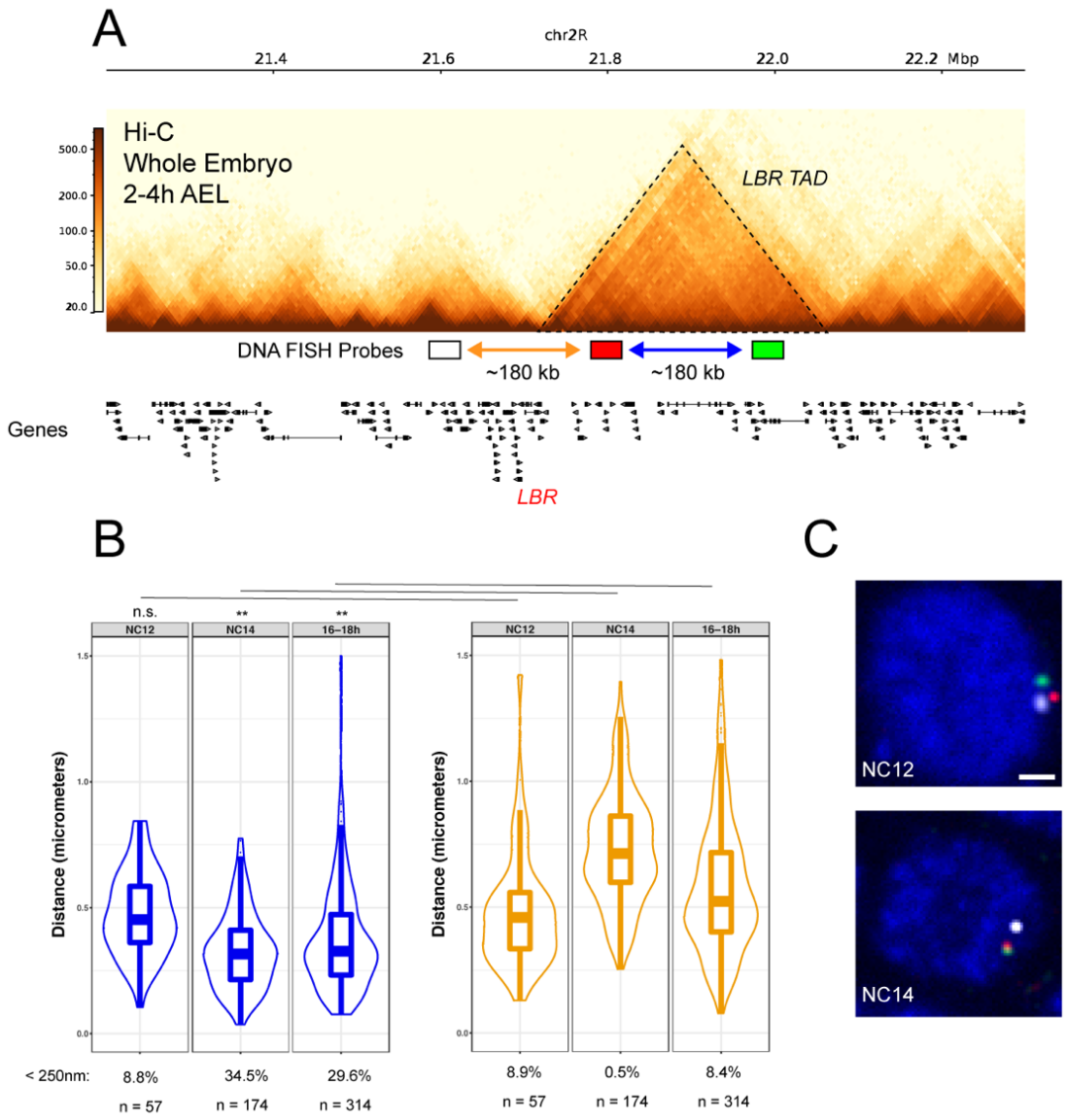

Figure 19: **DNA FISH to evaluate chromatin topology at the *LBR* locus during early embryogenesis.**
(A) Hi-C from 2-4h AEL embryos previously generated in the lab (unpublished). The position of the DNA FISH probes is shown below the Hi-C matrix. Blue and orange arrows indicate intra-TAD and inter-TAD comparisons, respectively.
(B) 3D distance distributions between intra-TAD probes (blue) and inter-TAD probes (orange) across the different indicated stages. n = number of alleles measured in at least 3 embryos.
(C) Representative single confocal slice showing the DNA FISH probes from (A) labelling a single region. Scale bar = 1 µm.
A Kolmogorov-Smirnov test was applied to evaluate statistical significance in (B): n.s. = non-significant, *p < 0.01, **p < 0.001.

4.2 Genome-wide profiling of insulator protein binding in the early embryo

The insulator protein CTCF is required for the formation of the majority of TADs and high-frequency looping interactions in vertebrates. In *Drosophila*, insulator proteins also regulate many different features of chromatin topology (see 1.1.3). Even though more than a dozen proteins with insulator function were found in flies to date, only a small number have been more well characterized, and therefore I chose to focus on BEAF-32, CP190 and CTCF. In order to study their roles during early embryonic establishment of chromatin topology, we first characterized their chromatin binding at the onset of ZGA. If these proteins have a role in setting up chromatin topology, they should be bound to chromatin during ZGA, and probably occupy TAD boundaries and/or loop anchors. Therefore, we performed ChIP-seq experiments for BEAF-32, CTCF and CP190 in wild-type embryos at NC14, in biological replicates. ChIP-seq experiments were done in collaboration with Dr. Rebecca Rodriguez-Viales in the Furlong lab. The analyses of this data, as described below, was done in collaboration with a bioinformatician in the Furlong lab Charles Girardot, and a master's student, Perrine Lacour.

4.2.1 ChIP-seq quality control

To assess how our embryonic NC14 ChIP-seq data compares to previously published ChIP-chip datasets from *Drosophila* embryos, we performed a Pearson correlation analyses. Our ChIP-seq samples generally show a very good correlation between replicates (Fig. 20A). Visual inspection of the ChIP-seq signal for the different insulators indicated that the peaks have strong enrichment over background, and present partial co-localization, as expected (Fig. 20B).

To confidentially identify peaks across biological replicates, we applied an IDR approach (1% cut-off). CP190 has the most peaks (4,396), followed by BEAF-32 (2,847) and CTCF (1,494). The numbers of CP190 and BEAF-32 peaks are in line with the number of peaks found at other biological stages or conditions, but the number of CTCF peaks at this stage is particularly low (Nègre et al., 2010; Van Bortle et al., 2012). Given that we observed a strong specific CTCF signal compared to background, the low number of peaks may reflect the biological role of CTCF at this early stage, and suggests differences in the requirements of CTCF in earlier and late stages of embryogenesis.

To evaluate which DNA-binding motifs are enriched within the ChIP-seq peaks for each of the insulator proteins, we performed motif analyses on the retrieved peaks, using the

MEME suite. For BEAF-32 and CTCF, the top DNA-binding motifs correspond to the canonical motifs for these proteins reported in the JASPAR database (Fig. 20C). Even though CP190 contains DNA-binding domains (C2H2 Zn-fingers), it has not been reported to bind to DNA directly, or at least through a clear motif (Kyrchanova & Georgiev, 2014), and is rather thought to be recruited to chromatin indirectly through the binding of other insulator proteins. The top-motif enriched under CP190 peaks is BEAF-32, corroborating to previous findings which show direct protein-protein interaction between those factors (Kyrchanova & Georgiev, 2014). Even though CTCF and CP190 were also previously shown to interact, the CTCF motif is not enriched at CP190 peaks. Other sequence motifs associated to insulator proteins (e.g. Su(Hw)), and transcription factors (e.g. Dorsal, Nau) appear enriched at CP190 peaks, perhaps indicating that CP190 binds to multiple regulatory elements. CP190 can interact with Su(Hw), but it is unclear if it can interact with those other proteins.

Overall, these analyses confirm the quality of our ChIP-seq data, and demonstrate that the three studied insulator proteins are bound to chromatin already during the major ZGA and suggest putative new partners of CP190 binding to DNA.

Figure 20: **ChIP-seq analysis of BEAF-32, CTCF and CP190 binding to chromatin in NC14 WT embryos.**
(A) Pearson correlation between all replicates performed for BEAF-32, CTCF and CP190 ChIPs and published 2-4h AEL ChIP-chip data for the same factors.
(B) Snapshot of tracks of merged ChIP-seq replicates.
(C) Top motifs found under ChIP-seq peaks using the MEME-ChIP tool (left), and the reported canonical motifs found at the Jaspar database (right).
(D) Upset plot of combinatorial binding of BEAF-32, CTCF and CP190 throughout the genome.
(E) Heatmaps of ChIP-seq signal ranked according to strongest CP190 peaks (top), BEAF-32 peaks (middle) and CTCF peaks (bottom).
(F) Pie charts of the distribution of ChIP peaks at different genomic features, obtained with *chipseeker*.
For all analysis, only ChIP peaks reproducible between replicates were used (IDR < 0.01).

4.2.2 Insulator proteins can co-occupy chromatin sites at ZGA

CP190, BEAF-32 and CTCF, can be found binding "alone" in ~2,200, ~1,300 and ~600 peaks, respectively (Fig. 20D). CP190/BEAF-32 combination occurs at ~1,300 sites, while CP190/CTCF occurs in ~600 sites and BEAF-32/CTCF combined binding is negligible (32 sites). Triple binding occurs in 232 sites (Fig. 20D). Therefore, BEAF-32 or CTCF can be found "alone" in as many sites as co-bound with CP190. This implies that when BEAF-32 or CTCF bind to chromatin mutually exclusively, there's a 50% chance of CP190 recruitment. However, when BEAF-32 and CTCF co-bind to a given site, CP190 is very often recruited (~90% of BEAF-32/CTCF co-bound sites have CP190 binding). Even though we can detect more than 200 sites which are co-occupied by CTCF, BEAF-32 and CP190, these sites are still a minority of insulator-bound sites. Nevertheless, these may have important biological implications, as it was previously shown that CTCF preferentially co-binds with other insulator proteins at the boundaries of H3K27me3-marked domains in cell lines (Van Bortle et al., 2012).

To analyze how binding strength relates to combinatorial binding, we rank-ordered either CP190, BEAF-32 or CTCF peaks, and plotted the quantitative ChIP signal for the other two insulator proteins across those peaks. This showed that the strongest BEAF-32 or CTCF peaks correlate with stronger CP190 signal (Fig. 20E). Interestingly, the strongest CP190 peaks are more associated with stronger CTCF compared to BEAF-32 binding (Fig. 20E). This might be due to the long-residence time of CTCF on chromatin (in the order of minutes in mammalian cells) (Hansen, Pustova, Cattoglio, Tjian, & Darzacq, 2017), while this time is not known for BEAF-32.

4.2.3 Which genomic sites are bound by insulator proteins?

To analyze which genomic features overlap insulator protein binding in NC14 embryos, we used *chipseeker* to quantify insulator binding at annotated genomic features such as promoters, exons and introns. CP190 and BEAF-32 are highly enriched at promoters (~70 and ~80% of peaks within 3 kb of an annotated promoter) (Fig. 20F). CTCF is also enriched, but to a lesser extent (~45% of peaks). Conversely, CTCF binds to a higher proportion of intergenic sites (14%) in comparison to CP190 and BEAF-32 (6% and 5%).

4.2.4 Insulator protein binding at TAD boundaries

Insulator binding has extensive overlap with *Drosophila* TAD boundaries at later developmental stages and in cell lines (Ramirez et al., 2018; Sexton et al., 2012). Accessible chromatin regions occupied by multiple insulator proteins are associated to domain boundaries in cell lines (Van Bortle et al., 2012). Here, we investigated whether (1) insulator proteins are already localized at TAD boundaries during ZGA and (2) whether specific combinations of insulator protein binding are enriched at TAD boundaries. We evaluated the overlap of BEAF-32/CTCF/CP190 ChIP-peaks at NC14 with TAD boundary from Hi-C data generated in embryos at the same stage. This revealed that TAD boundaries differ in terms of insulator protein occupancy at NC14: 84% of TAD boundaries at NC14 are occupied by at least one of the studied insulator proteins (668/819 of high confidence TAD boundaries). The most frequent combination is BEAF-32 and CP190, present at 30% of all TAD Boundaries, followed by CP190 binding "alone" (20%) (Fig. 21). Other combinations, such as CTCF/CP190 or CTCF/BEAF-32/CP190 occur less frequently (12% and 15%) (Fig. 21). BEAF-32 or CTCF "alone" occur only at a minor fraction of boundaries (4 and 1% respectively) (Fig. 21). Overall, the insulator binding combination most predictive of TAD boundaries is CP190/BEAF-32/CTCF, with 41% of all triple-bound sites overlapping a NC14 TAD boundary. The BEAF-32/CP190 and CTCF/CP190 combinations follow, with 17% of their respective ChIP peaks overlapping a TAD boundary.

Insulator protein binding by itself clearly cannot predict the position of TAD boundaries, as the majority of peaks are outside of TAD boundaries, both in our early stage dataset and in previously published datasets. Additional features (e.g. local context of combinatorial binding, binding at promoters, etc) are likely relevant additional factors for the establishment of TADs. For example, CP190 and BEAF-32 are associated with promoters and boundaries that overlap with promoters, while CTCF to a lesser extent (Ramirez et al., 2018). We are currently investigating what additional features could be involved in further diversifying TAD boundaries.

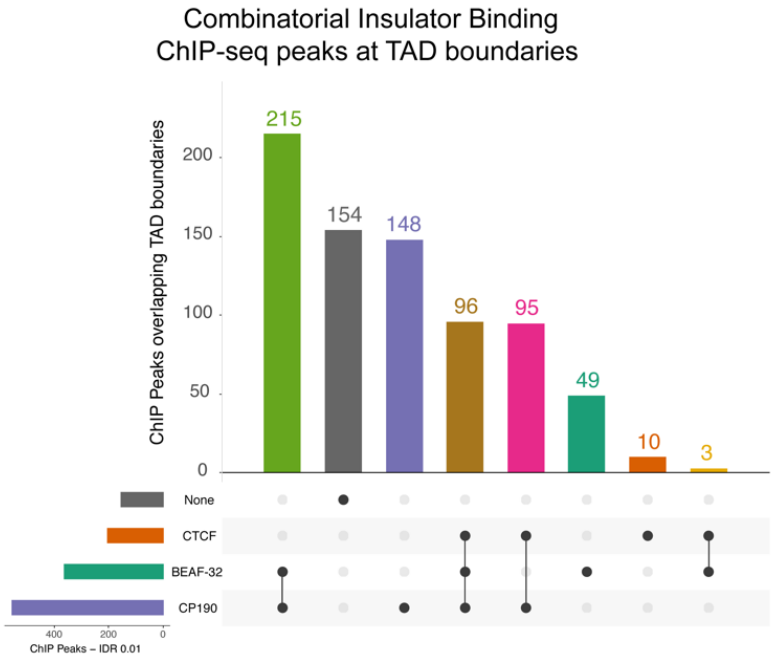

Figure 21: **Combinatorial binding of BEAF-32, CTCF and CP190 at TAD boundaries.**
Upset plot of combinatorial binding of BEAF-32, CTCF and CP190 at TAD boundaries obtained from Hi-C with WT 2-3h AEL embryos. Only ChIP peaks reproducible between replicates were used (IDR < 0.01).

4.3 Depleting candidate regulators of chromatin topology establishment in the early embryo

Since the establishment of chromatin topology occurs concomitantly to the onset of the major ZGA, we reasoned that trans-acting factors present in the embryo prior to the major ZGA may play a role in driving the formation of TADs and loops. These proteins would probably be maternally-loaded in the developing oocyte, and be present at least until the onset of the major ZGA. The insulator proteins BEAF-32, CTCF and CP190 are all maternally-deposited. To study the contributions of those candidates in the establishment of chromosome topology and gene expression at the onset of the major ZGA, I removed the maternal contribution of each factors. Two main strategies were used: RNA interference and genetic knockout in the germline.

4.3.1 RNA interference (RNAi)

This approach consists of targeting maternally-deposited mRNA with a shRNA. Since the candidate proteins are deposited maternally in the egg, RNAi against their mRNA must occur in the female germline. The chosen strategy for RNAi consisted of crossing flies expressing the transcriptional activator Gal4 specifically in the maternal germline, with flies

carrying an UAS-shRNA sequence integrated in the genome, activatable by the Gal4 protein. The advantage of this approach is that *Drosophila* lines carrying inducible shRNAs against hundreds of specific genes are available, making it easier and fast to generate embryos depleted of a given factor. However, this approach may not lead to total depletion of the targeted mRNAs, so the level of protein depletion needs to be accurately measured.

To induce shRNA expression in the germline, we tested two lines carrying distinct Gal4 drivers: one in which Gal4 is expressed throughout the whole oocyte maturation process (MTD-Gal4), and one in which Gal4 is expressed only late in oogenesis (Mat-Tub-Gal4). This distinction is useful particularly when studying genes that have an essential role in early oogenesis, as their premature depletion can arrest oogenesis and lead to sterility, precluding the development of eggs (Staller et al., 2013). The crossing scheme for the RNAi strategy is shown in (Fig. 9).

We therefore tested the efficiency of both the MTD-Gal4 or Mat-Tub-Gal4 drivers to deplete the maternal contribution of BEAF-32 or CP190 mRNA. Depleting BEAF-32 with the MTD-Gal4 driver led to arrested oogenesis and sterility, while the Mat-Tub-Gal4 allowed oogenesis to proceed. For CP190, we were able to obtain embryos depleted of its maternal contribution by using the MTD-Gal4 driver. Therefore, we chose to use the Mat-Tub-Gal4 driver to obtain BEAF-32 depleted embryos, and the MTD-Gal4 driver to obtain CP190-depleted embryos.

4.3.2 Germline knockout clones / maternal-zygotic nulls

This approach consists in genetic deletion of the actual gene from both the female germline and the embryo, removing all of its contribution to the early embryo. For this approach, we either used the "germline clone" strategy (Chou & Perrimon, 1996), or a strategy based on Flp-mediated excision of Frt-flanked alleles in the female germline (Gambetta & Müller, 2014). The advantage of those approaches is that the gene is completely removed, eliminating all mRNA and protein production. However, as some genes are required for germline formation, their complete removal can arrest oogenesis and lead to sterility, and therefore alternative approaches must be applied. The crossing scheme for the knockout strategy is shown in Fig. 10.

CTCF knockout in the female germline was previously developed by a former post-doc in the Furlong lab (Gambetta et al., 2018), and we used this strategy to remove CTCF. As BEAF-32 is essential for oogenesis (Gambetta & Furlong, 2018), we could not use this

approach in the female germline. CP190 germline clones have been previously generated (Chodagam et al., 2005), but in our hands, this approach led to a dramatic reduction in fertility: the amount of embryos we obtained was incompatible with follow-up experiments. We additionally used the germline clone approach to obtain embryos depleted of the transcription factor Zld, which was shown to be important for activation of hundreds of genes during the ZGA ((Liang et al., 2008).

4.3.3 Evaluating efficiency of protein depletion

The depletion of insulator proteins via either RNAi or genetic knockout needs to cause strong protein depletion at the time of ZGA, since we want to test their roles in setting up chromatin topology at that time point. To evaluate protein depletion efficiency, we performed Western blots for BEAF-32, CTCF and CP190 in embryos manually-selected at the NC14 stage, based on morphological features.

Depletion of BEAF-32 with the Mat-Tub-Gal4 driver and of CP190 with the MTD-Gal4 driver led to strongly reduced BEAF-32 or CP190 protein levels at ZGA (Fig. 22A,C). This reduction was of approximately 230- and 115-fold, respectively (Fig. 22B,D). As expected, depletion of CTCF using the knockout approach completely removed the CTCF protein from embryos at ZGA, resulting in a ~500-fold difference in protein levels (Fig. 22E,F). These results show that the RNAi and genetic knockout approaches are suitable to study the role of insulator proteins in chromatin topology and gene expression in the early embryo.

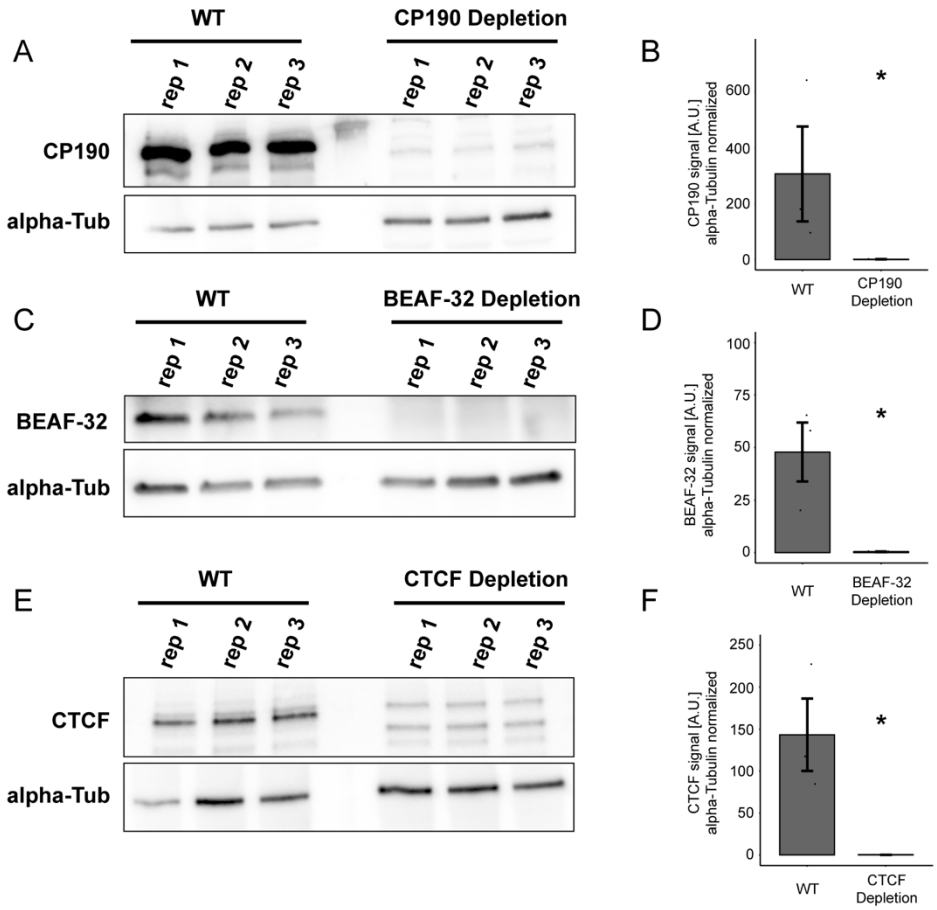

Figure 22: **Evaluation of depletion of insulator proteins at NC14.**
(A) Western Blot to detect the proteins CP190 (top) and alpha-Tubulin (bottom) in WT and CP190-depleted embryos manually selected at NC14.
(B) Quantification of the CP190 western blot.
(C) Western Blot to detect the proteins BEAF-32 (top) and alpha-Tubulin (bottom) in WT and BEAF-32-depleted embryos manually selected at NC14.
(D) Quantification of the BEAF-32 western blot.
(E) Western Blot to detect the proteins CTCF (top) and alpha-Tubulin (bottom) in WT and CTCF-depleted embryos manually selected at NC14.
(F) Quantification of the CTCF western blot.
In all experiments, alpha-Tubulin was used as the normalizer. T-test was used to assess statistical significance. *$p < 0.05$

4.4 Depletion of insulator proteins leads to transcriptional defects during ZGA

To evaluate what are the genome-wide transcriptional consequences of depleting insulator proteins, we performed RNA-seq in embryos manually-selected at the NC14 stage,

in biological triplicates. The computational analyses of the RNA-seq data was done in collaboration with a bioinformatician in the lab (Adam Rabinowitz).

4.4.1 RNA-seq quality control

To evaluate the quality of the datasets, we analyzed the amount of uniquely mapping reads among the replicates of all samples. We obtained from 20 to 60 million sequencing reads per replicate, and from those, 95-97% uniquely mapping reads (Fig. 23A). Then, we plotted the distribution of reads along gene bodies. As expected, our RNA-seq reads distribute homogeneously along gene bodies, in all genotypes (Fig. 23B). To assess the reproducibility between replicates, we performed PCA analyses. This showed that for all three insulator-depleted conditions, all three replicates separated well from the wild-type replicates, along the PC1 component (Fig. 23C).

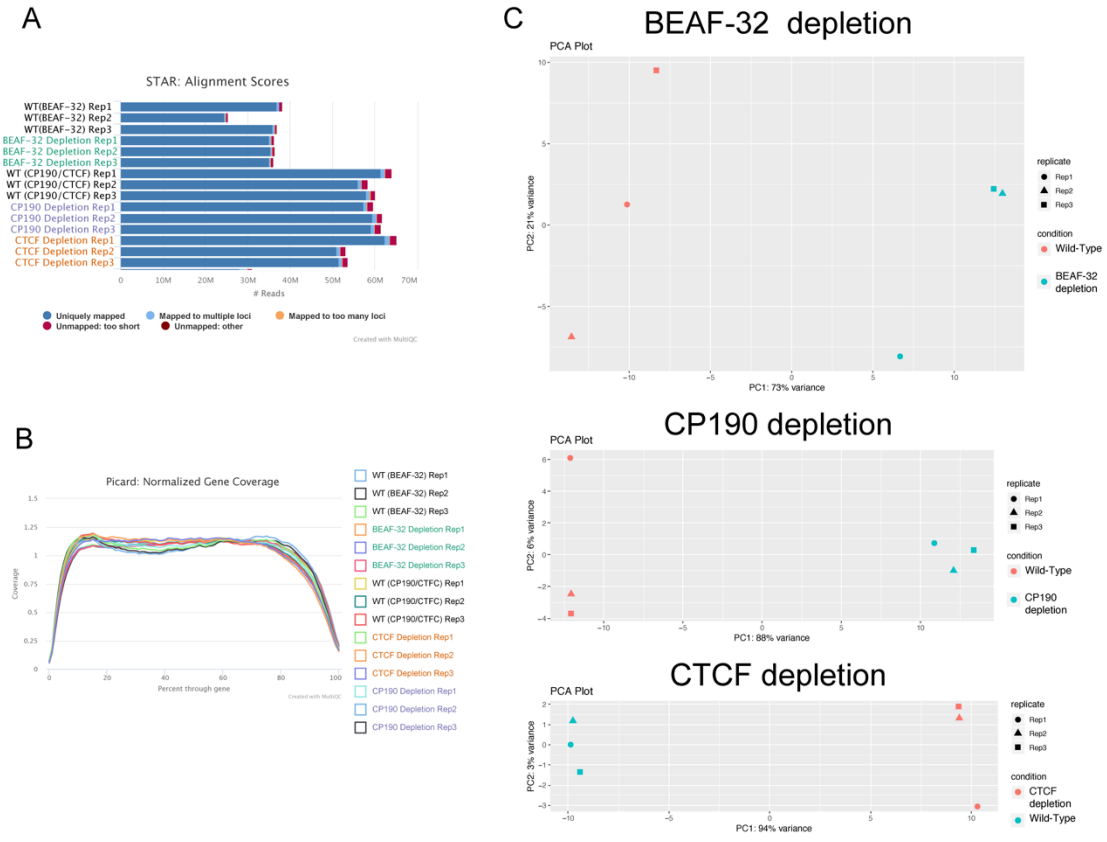

Figure 23: **Quality control of RNA-seq experiments.**
(A) Read alignment statistics over all replicates obtained with STAR.

(B) Aggregate view of coverage over gene bodies across all replicates generated with Picard.
(C) PCA plots generated by DESeq2, showing separation between replicates of BEAF-32 (top), CP190 (middle), CTCF (bottom) and their respective controls (WT).

4.4.2 Insulator protein depletion leads to the misregulation of hundreds of zygotically-expressed genes

During ZGA, the embryo contains a mixture of maternally-deposited and newly zygotically expressed mRNAs. In order to assess the more direct impact of insulator proteins on transcription, and avoid confounding indirect effects, we limited our analyses to genes that are zygotically-expressed but not maternally-deposited. We used a previously generated RNA-seq dataset from unfertilized eggs (Ghavi-Helm et al., 2019), to identify which genes are maternally deposited, and filtered those out of our analyses. Therefore, we focused all our analyses on genes that are strictly zygotically expressed. We used the cut-off of log2 fold-change of 0.7 and FDR < 0.05 to define Differentially-Expressed Genes (DEGs) in a given genotype compared to WT, at NC14. Using these stringent criteria, we found 323, 374 and 595 zygotic DEGs in embryos depleted of BEAF-32, CTCF or CP190 respectively (Fig. 24A-C). In BEAF-32 and CP190-depleted embryos there are 50/50% down- and up-regulated genes, while there are 38/62% down-/up-regulated in CTCF-depleted embryos.

4.4.3 Insulator proteins co-regulate gene expression

To understand how insulator proteins could cooperatively regulate transcription, we calculated the overlap between mis-regulated genes among the different genotypes. Even though combinatorial insulator binding is widespread, most DEGs are specific to a single insulator depletion (Fig. 24D). Embryos depleted of CP190 or CTCF share more mis-regulated genes than embryos depleted of CP190 or BEAF-32 (125 vs. 62 DEGs), while only 36 genes are mis-regulated in common after BEAF-32 or CTCF depletion (Fig. 24D). Moreover, the three genotypes share the mis-regulation of 65 genes (Fig. 24D). We noticed that even though CP190 and BEAF-32 have more extensive chromatin co-binding than CP190 and CTCF (both genome-wide and at promoters) (Fig. 20D-F), there are less DEGs shared between CP190 and BEAF-32 depletion than CP190 and CTCF depletion. Additionally, as mentioned above, we observed a correlation between the strongest CP190 and CTCF ChIP-seq peaks (Fig. 20E). These findings suggest that CP190 and CTCF cooperation may be more important for the

correct transcription of strictly zygotic genes than CP190 and BEAF-32, even though a subset of genes is sensitive to the loss of BEAF-32 and CP190.

4.4.4 DEGs are a mixture of direct targets of insulator proteins, and indirect effects

The binding of all three insulator proteins is enriched at promoters (Fig. 20F), and BEAF-32 and CP190 have been previously speculated to act as direct transcriptional activators, while some reports attribute a transcriptional repressor role for CTCF (Filippova et al., 1996). Therefore, we asked whether the DEGs could be direct transcriptional targets of insulator proteins. If that is the case, insulator protein binding should occur at the promoters of DEGs, or at associated regulatory elements. We analyzed the distances between the closest insulator protein ChIP-seq peaks to the TSS of DEGs in each genotype, and compared those distances to the background of not differentially-expressed genes. We found that BEAF-32- or CP190-depleted embryos have slightly higher than expected binding of the corresponding insulator at the DEG promoters, compared to the background (Fig. 24E,G). For CTCF DEGs, there is no enrichment for CTCF binding at their promoters of DEGs compared to background (Fig. 24F).

These findings indicate that, even though a subset of DEGs could be explained by loss of insulator protein binding directly at the promoter (or associated regulatory elements: e.g. enhancers in intronic sites), the DEGs in all three genotypes are a mixture of potential direct effects on promoter-proximal regulation and more indirect (cis long-range and trans) effects. This prompted us to investigate whether chromatin topology is affected in the absence of insulator proteins at the major ZGA, and if that could explain some of the transcriptional defects we observe.

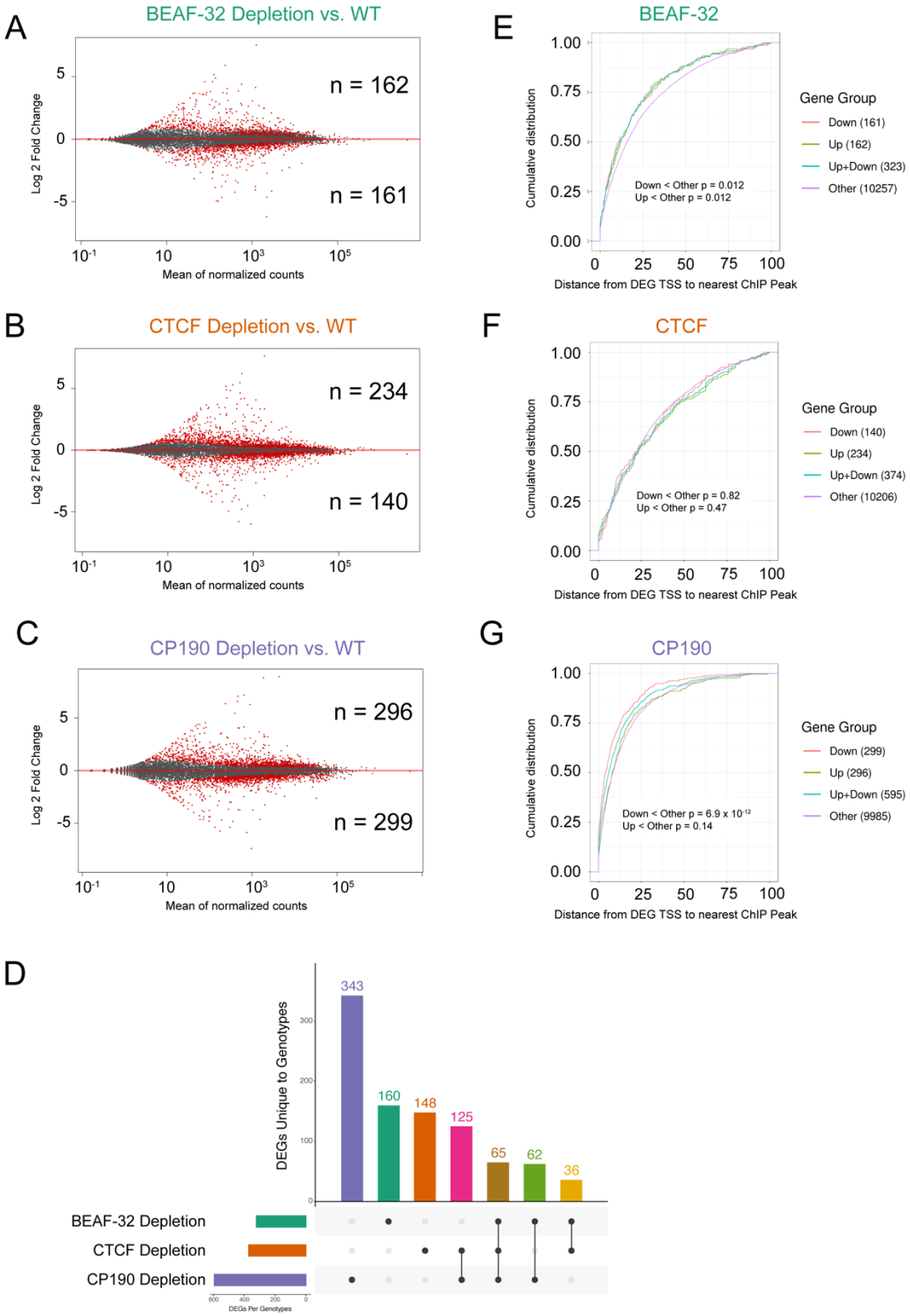

Figure 24: **Gene expression changes measured by RNA-seq in NC14 embryos following insulator protein depletion.**
(A-C) M-A plots of Log2 Fold-Change versus mean of normalized counts from RNA-seq experiments. In red are shown DEGs and in black all other genes. n = number of DEGs either up- or down-regulated (top and bottom numbers, respectively) in each condition.

(D) Upset plot showing the overlap of DEGs between conditions.
(E-G) Cumulative distribution plots showing the distance between the TSS of DEGs and the nearest ChIP peak. In each plot, the DEGs in a single genotype are shown relative to the distance to ChIP peaks of that same factor. Wilcoxon tests were performed for statistical analyses in (E-G).

4.5 Characterizing the effects of insulator protein depletion on global chromatin topology at ZGA

In order to evaluate how depleting insulator proteins affects genome-wide chromatin topology during ZGA, we performed Hi-C in embryos depleted of each of the three insulator proteins. Hi-C was performed with 2-3h AEL embryos, which corresponds to Bownes' Stage 5 (predominantly NC14 embryos), in biological duplicates. Hi-C experiments were done in collaboration with Dr. Rebecca Rodriguez-Viales in the Furlong lab. The computational analyses of this data, as described below, was done in collaboration with a bioinformatician in the Furlong lab, Charles Girardot, and a master's student, Perrine Lacour.

4.5.1 Hi-C quality control

Hi-C datasets from each replicate of WT and insulator-depleted embryos contained an equivalent number of reads, from 50 to 80 million usable read pairs per replicate (Fig. 25A). The proportion of short-range, long-range and trans-chromosomal contacts was compatible with reported experiments in the literature and from our lab (Ghavi-Helm et al., 2019; S. S. Rao et al., 2014; Sexton et al., 2012), although the trans-chromosomal contacts were increased for WT samples (36-42%) in comparison to insulator-depleted samples (22-32%) (Fig. 25B). The proportion of relative read orientation within read pairs followed an expected distribution, with an approximate equal number of inward, outward, left and right pairs in all conditions (Fig. 25C). The overall Spearman correlation of contact frequencies between all replicates and conditions was high (ranging from 0.81 to 0.87), indicating high reproducibility between replicates, and no striking differences between conditions (Fig. 25D). Similarly, calculating the correlation between all replicates and conditions using 4 different tools (GenomeDISCO, HiCRep, HiCSpector and QuASAR) revealed a high correlation between replicates of the same genotype (ranging from 0.65 to 0.95, depending on the tool) (Fig. 25D). We also noticed that in a few cases, replicates would correlate equally or even better with replicates from a different genotype, indicating that the global contact pattern is very similar between all conditions.

Evaluation of the contact frequency across different genomic scales revealed a similar distribution for all replicates in WT and insulator-depleted embryos (Fig. 25D).

Overall, these analyses indicate that the generated Hi-C datasets have high quality and reproducibility, with a relatively mild coverage, and suggest that depletion of any of the three insulators does not strikingly impact global chromatin conformation at the level of TADs or compartments (local or long-range interactions).

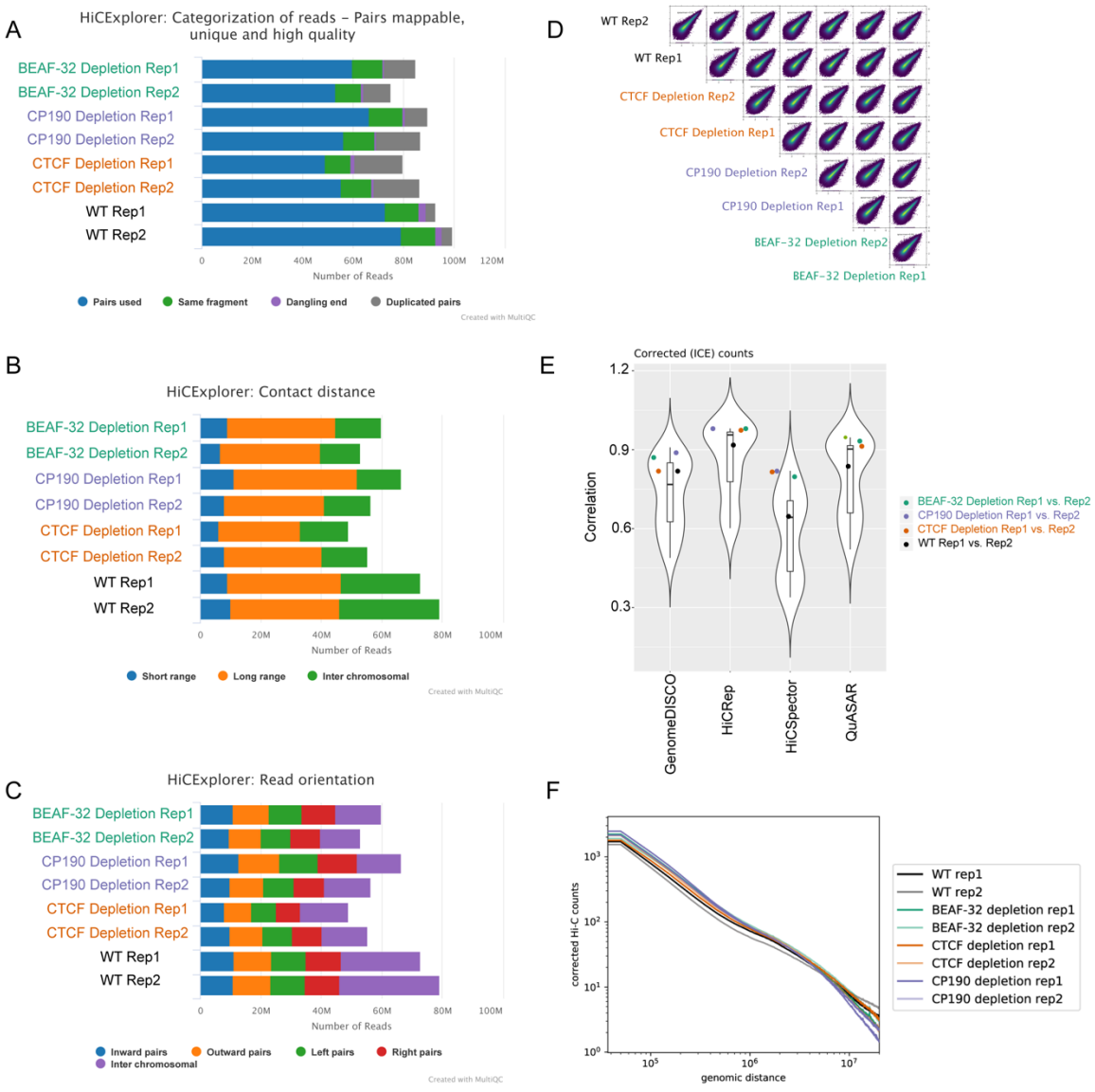

Figure 25: **Quality control of Hi-C experiments.**
(A) Read pair alignment statistics over all replicates.
(B) Proportion of ligation events (each read in a read pair) occurring between different chromosomal distances. Short range correspond to distances < 25kb and long range > 25kb.
(C) Proportion of different types of ligation events (given four possible orientations of ligation events) across the Hi-C library.
(D) Spearman correlation between all Hi-C replicates.

(E) Violin plots showing the correlation between replicates obtained with different methods. Points corresponding to correlation between replicates of the same genotype are coloured.
(F) Distribution of Hi-C chromatin interactions along different chromosomal distances, for each replicate.

4.5.2 Most TADs are still formed after BEAF-32, CTCF or CP190 depletion

Visual inspection of Hi-C maps showed that most TADs are able to form during ZGA upon depletion of BEAF-32, CP190 or CTCF (Fig 26A). To identify regions with differential contacts in insulator-depleted samples, we generated differential Hi-C maps by calculating the log2 ratio of contact frequencies between a given genotype and the wild-type. Such differential Hi-C maps support the observation that most TADs are not strongly affected by insulator depletion (Fig. 26B). Computational identification of TADs using the Hi-C Explorer suite (see 3.9.2.4) on Hi-C maps at 5kb resolution and using a FDR < 0.05, indicated a slight variation in the number of TADs found genome-wide across genotypes (Fig. 26C). The proportion of TADs per chromosome is not affected (Fig. 26D).

Interestingly, we could visually detect increased interactions across a handful of TAD boundaries, especially in CP190-depleted embryos, pointing to the existence of regions more susceptible to topological defects upon insulator protein loss. One example of affected region is shown in (Fig. 26E,F), in which two neighbouring TADs exhibit a higher interaction frequency across the TAD boundary, upon loss of CP190 (note the increased red signal across the marked TAD boundary).

We decided then to systematically characterize finer-scale changes in chromatin topology by performing more sensitive analyses, such as quantitative assessment of insulation at TAD boundaries, and single-cell DNA FISH targeting specific loci NC14 embryos.

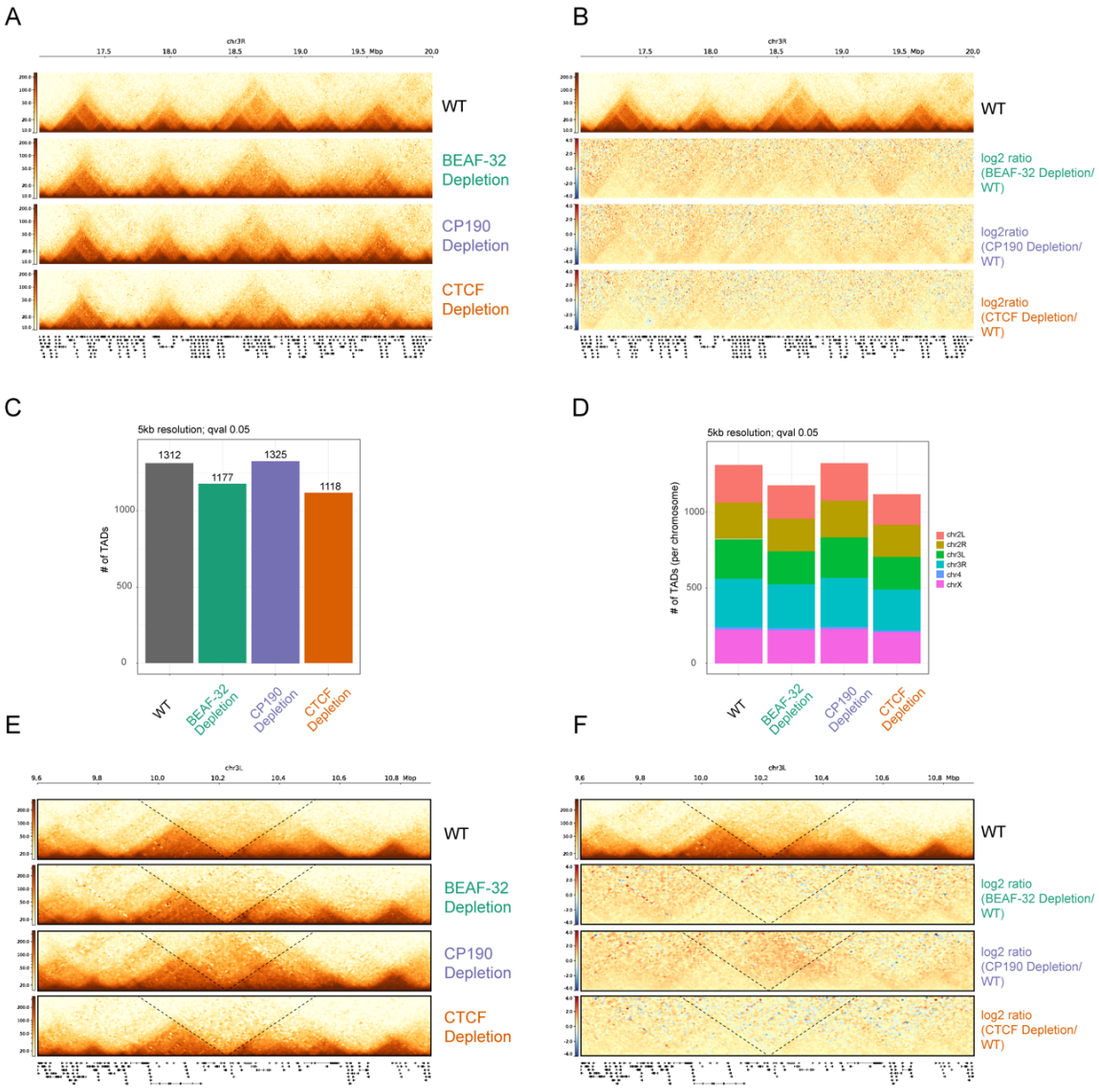

Figure 26: **TAD structure assessed by Hi-C in embryos with insulator protein depletion.**
(A) Hi-C snapshot of a 3 Mb region across different genotypes. Each Hi-C map is corresponds to merged replicates.
(B) Differential Hi-C map generated by calculating the log2 ratio between Hi-C contacts in a given genotype versus WT.
(C) Bar plots showing the number of TADs obtained for each genotype, using Hi-C data with 5 kb resolution and a qvalue < 0.05.
(D) Proportion of TADs per chromosome, using the same parameters as in (C).
(E) Hi-C maps showing a TAD boundary affected by CP190 depletion. Dashed lines indicate the position of the TAD boundary.
(F) Differential Hi-C maps (log2 ratio) of the same region shown in (E).

4.5.3 Identifying changes in fine-scale chromatin topology upon insulator protein loss

TAD boundaries can be defined as regions separating TADs, and the separation (insulation) level can be expressed as an insulation score (Lajoie et al., 2015)(see 3.9.2.3). This metric is useful to evaluate if the contact frequency across a boundary changes between biological conditions. Our ChIP-seq experiments showed that TAD boundaries are occupied by distinct combinations of insulator proteins at NC14 (Fig. 20). Does insulator protein depletion affect insulation at these different categories of TAD boundaries? To evaluate that, we first obtained the coordinates of insulator ChIP-seq peaks and TAD boundaries (called in WT Hi-C matrices). Then, we derived and compared the insulation scores at those coordinates in Hi-C matrices from WT- and insulator-depleted samples. We found that average global insulation at TAD boundaries is diminished (i.e. higher insulation score) in all three insulator-depleted conditions (Fig. 27A). To better understand this global effect, we divided all insulator binding sites or all TAD boundaries into groups based on combinatorial insulator protein occupancy, and then analyzed insulation changes separately in those groups.

In WT embryos, there's virtually no insulation at genomic sites lacking insulator binding or occupied solely by CTCF (Fig. 27B). Sites occupied by either BEAF-32 or CP190 display weak to modest insulation, to similar levels as sites co-occupied by CP190 and CTCF (Fig. 27B). Insulation increases in CP190/BEAF-32 co-occupied sites, and reaches the strongest levels in sites bound by CP190, BEAF-32 and CTCF (Fig. 27B).

At TAD boundaries in WT embryos, the strongest insulation is observed in boundaries occupied by the three proteins (Fig. 27C). Boundaries overlapping with BEAF-32 peaks, either alone or in combination with CP190, are slightly less insulated. Boundaries bound by CTCF and CP190, and not bound by any of the three factors, display a lower level of insulation (Fig. 27C).

We then analyzed changes in the insulation score across those different categories of ChIP peaks and TAD boundaries following insulator depletion.

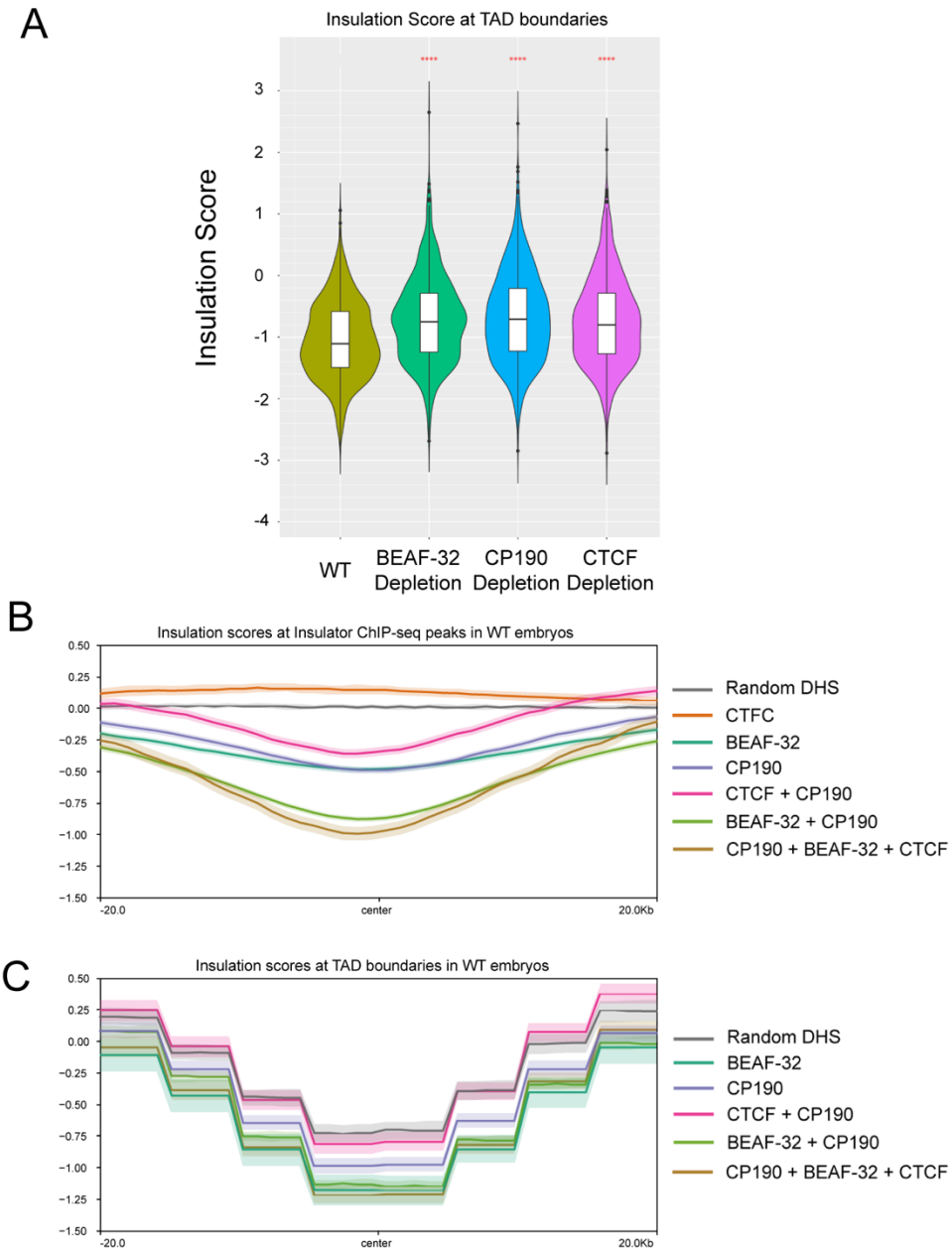

Figure 27: **Hi-C insulation scores vary according to combinatorial insulator protein binding, and are reduced following insulator protein depletion.**
(A) Violin plots of insulation scores derived from Hi-C matrices of the indicated genotypes.
(B) Averaged insulation scores in WT embryos across all ChIP-seq peaks overlapping with binding of the indicated insulator proteins at NC14. The line shades represent 95% confidence intervals.
(C) Averaged insulation scores in WT embryos across all TAD boundaries overlapping with binding of the indicated insulator proteins at NC14. The line shades represent 95% confidence intervals. For the calculation of insulation scores at TAD boundaries, bins of 5 kb were used, thus the lines are not as smooth as in the plots showing insulation at ChIP-peaks (B).

4.5.4 Insulator binding sites

Surprisingly, depletion of BEAF-32, CTCF or CP190 leads to a decrease in average insulation at sites bound by almost any combination of the three insulator proteins. For example, insulation at CP190 or BEAF-32 uniquely-bound sites is reduced in embryos depleted of any of the insulator proteins (Fig. 28). Similarly, at sites co-occupied by BEAF-32/CP190, insulation is reduced after depletion of either BEAF-32, CP190 or CTCF (Fig. 28). This suggests that insulation at these sites is sensitive not only to loss of CP190 or BEAF-32 at the bound site, but also to a global loss of any of the three insulator proteins. Triple-bound sites also have decreased insulation in any of the three genotypes (Fig. 28). In general, however, the biggest drop in insulation is seen upon depletion of the proteins that bind at those sites, indicating that the insulation loss due to an "indirect effect" is not as strong. Interestingly, the exception to these observations are sites co-occupied by CP190 and CTCF, which have decreased insulation only after depletion of CP190 or CTCF, but not BEAF-32 (Fig. 28). This indicates that CP190/CTCF co-occupied sites may be less prone to indirect effects on insulation caused by insulator protein depletion.

4.5.5 TAD Boundaries

TAD boundaries not occupied by any of the three factors have reduced insulation after depletion of BEAF-32, CTCF or CP190 (Fig. 28). Boundaries occupied by one insulator protein (BEAF-32 or CP190), two (BEAF-32/CP190 or CTCF/CP190) or three (BEAF-32/CTCF/CP190) insulator proteins have reduced insulation in any of the three genotypes (Fig. 28). However, boundaries occupied by the combination CTCF/CP190 lose insulation specifically after loss of either CTCF or CP190, but not BEAF-32. Boundaries occupied by only CTCF or the CTCF/BEAF-32 combination are too scarce, not allowing proper quantification.

Overall, our findings indicate that depletion of any of the three insulator proteins leads to a global decrease in insulation at insulator binding sites and TAD boundaries, even though most TADs can still be detected. This effect can often be observed even at TAD boundaries that are not occupied by the insulator protein being depleted. This indicates that either boundary insulation is sensitive to a global loss of insulator protein binding, or can be influenced by elements acting from a distance. A specific set of TAD boundaries, co-occupied by CP190 and CTCF, is affected specifically by CP190 and CTCF depletion, while other boundaries are affected by any insulator depletion.

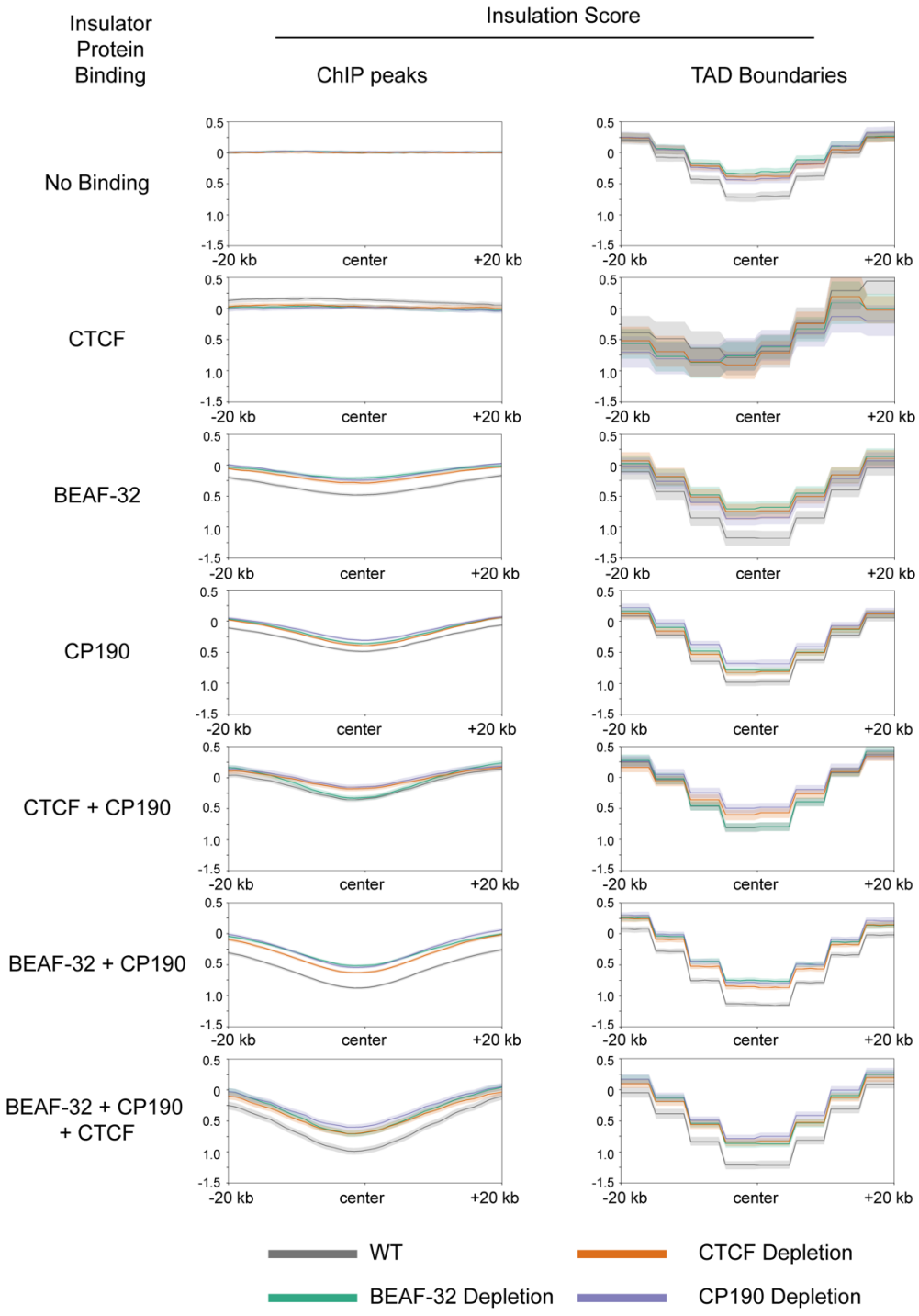

Figure 28: **Hi-C insulation scores across various categories of ChIP-seq peaks and TAD boundaries are reduced following insulator protein depletion.**
(Left) Comparison of averaged insulation scores across all ChIP-seq peaks overlapping with binding of the indicated insulator proteins at NC14. The line shades represent 95% confidence intervals. Each plot shows lines representing averaged insulation scores in different genotypes, as indicated.
(Right) Averaged insulation scores across all TAD boundaries overlapping with binding of the indicated insulator proteins at NC14. The line shades represent 95% confidence intervals. Each plot shows lines representing averaged insulation scores in different genotypes, as indicated For the calculation of insulation scores at TAD

boundaries, bins of 5 kb were used, thus the lines are not as smooth as in the plots showing insulation at ChIP-peaks (Left).

4.6 DNA FISH to assess changes in chromatin topology at specific loci with single-cell resolution

To evaluate if insulator protein depletion could lead to alterations in the establishment of chromatin topology at NC14 at the single cell level, we performed DNA FISH experiments targeting loci in which we observed chromatin topology establishment at NC14 (Fig. 17-20): the *tsh-tio*, *scyl-chrb*, and *LBR* TADs. We also evaluated if transcriptional changes at these loci following insulator protein depletion were associated with topological changes.

4.6.1 TADs containing high-frequency looping interactions

In the *tsh-tio* locus, the two flanking TAD boundaries are bound by BEAF-32 at NC14, but not CTCF (Fig. 29A). The left boundary is also bound by CP190. Additionally, there is a single intra-TAD site occupied by CTCF/CP190. This TAD contains three loop anchors which are bound by neither of the three insulator proteins at NC14. The anchors form a high-frequency looping interaction between the *tsh* and *tio* genes that can be detected in the high resolution 2-4h AEL Hi-C dataset (Fig. 18). Depletion of BEAF-32, but not CTCF, reduces transcription of *tsh* or *tio*, as measured by RNA-seq (Fig. 29B). The expression of *tsh* might be altered upon depletion of CP190 as well, even though it is not statistically significant (Fig. 29B).

We used the same DNA FISH probes targeting intra-TAD and inter-TAD regions as described above, to evaluate topological changes upon depletion of CP190 or CTCF (Fig. 18). Our DNA FISH measurements revealed that depletion of neither CP190 nor CTCF affects the formation of this TAD at NC14 (Fig. 29C). The difference between intra-TAD and inter-TAD distances is similar between WT embryos and insulator-depleted embryos. However, depletion of CP190 led to a subtle increase in contacts across the TAD boundary (Fig. 29C). These DNA FISH findings agree with the Hi-C data, which shows that the TAD is still formed upon depletion of any of the three insulator proteins (Fig. 29C,D). However, the increase in contacts is not observed in the Hi-C data, possibly due to lack of resolution or sensitivity.

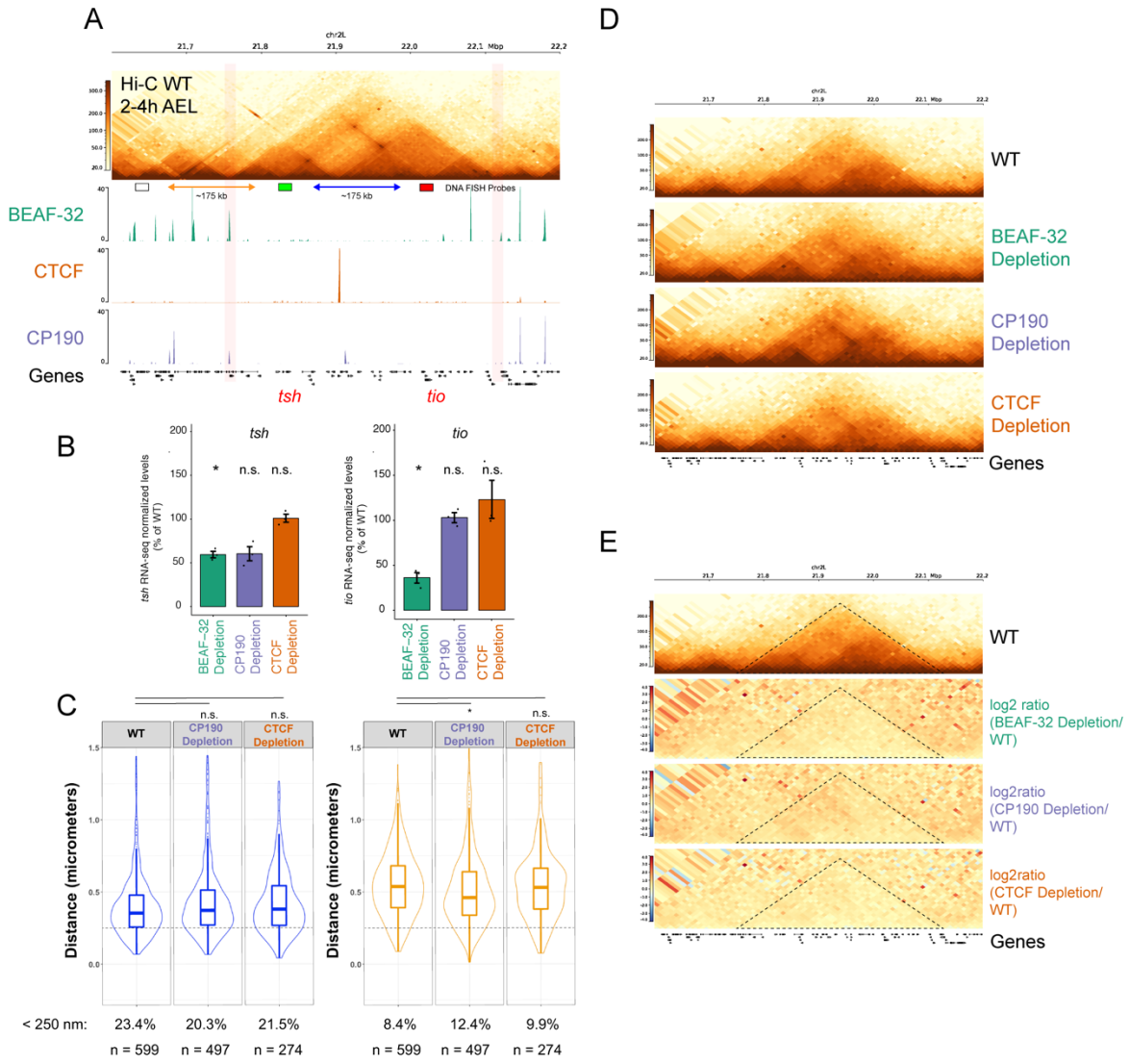

Figure 29: **DNA FISH and Hi-C to evaluate chromatin topology at the *tsh-tio* locus following insulator protein depletion.**

(A) Hi-C from 2-4h AEL WT embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins in WT NC14 embryos. The position of the DNA FISH probes is shown below the Hi-C matrix. Blue and orange arrows indicate intra-TAD and inter-TAD comparisons, respectively. The boundaries of the *tsh-tio* TAD are highlighted.

(B) Bar plots of RNA-seq normalized counts for *tsh* (left) and *tio* (right). Different genotypes are shown relative to WT. Significance was assessed according to DESeq2 standard procedure, with a correction for multiple testing. n.s. = non significant, *padj < 0.01

(C) 3D distance distributions between intra-TAD probes (blue) and inter-TAD probes (orange) across different genotypes. n = number of alleles measured in at least 3 embryos. A Kolmogorov-Smirnov test was applied to evaluate statistical significance: n.s. = non-significant, *p < 0.01, **p < 0.001

(D) Hi-C maps of the *tsh-tio* locus across different genotypes (5kb resolution).

(E) Differential Hi-C maps (log2 ratio of contacts between a given genotype and WT) of the *tsh-tio* locus across different genotypes. The dashed triangle indicates the *tsh-tio* TAD.

In the *scyl-chrb* locus, the left TAD boundary overlaps with BEAF-32 and CP190 binding, while the right TAD boundary overlaps with CP190. As in *tsh-tio*, this locus presents an intra-TAD three-way looping interaction. The leftmost of the loop anchors is bound by CTCF, while the other two anchors are not bound by any of the three insulator proteins (Fig. 30A). The three insulator proteins bind to intra-TAD sites in between the *scyl* and *chrb* loop anchors. The expression of *scyl* and *chrb* are affected by depletion of BEAF-32 or CTCF. While BEAF-32 loss leads to a slight up-regulation of *scyl* and down-regulation of *chrb*, CTCF depletion reduces expression of both genes (Fig. 30B).

We used the same DNA FISH probes described above to label intra-TAD and inter-TAD regions (Fig. 17). Our DNA FISH measurements revealed that depletion of either BEAF-32, CTCF or CP190 does not affect the formation of this TAD at NC14, as intra-TAD and inter-TAD measurements are similar between WT and insulator-depleted embryos (Fig. 30C). However, depletion of either BEAF-32 or CP190 led to a minor increase in proximity between the loop anchors, while CTCF depletion led to a minor decrease. None of the depletions affected the control probe distances, suggesting that the right TAD boundary remains fully functional. The TAD formation and increased intra-TAD proximity in the absence of BEAF-32 or CP190 are supported by Hi-C experiments (Fig. 30D,E).

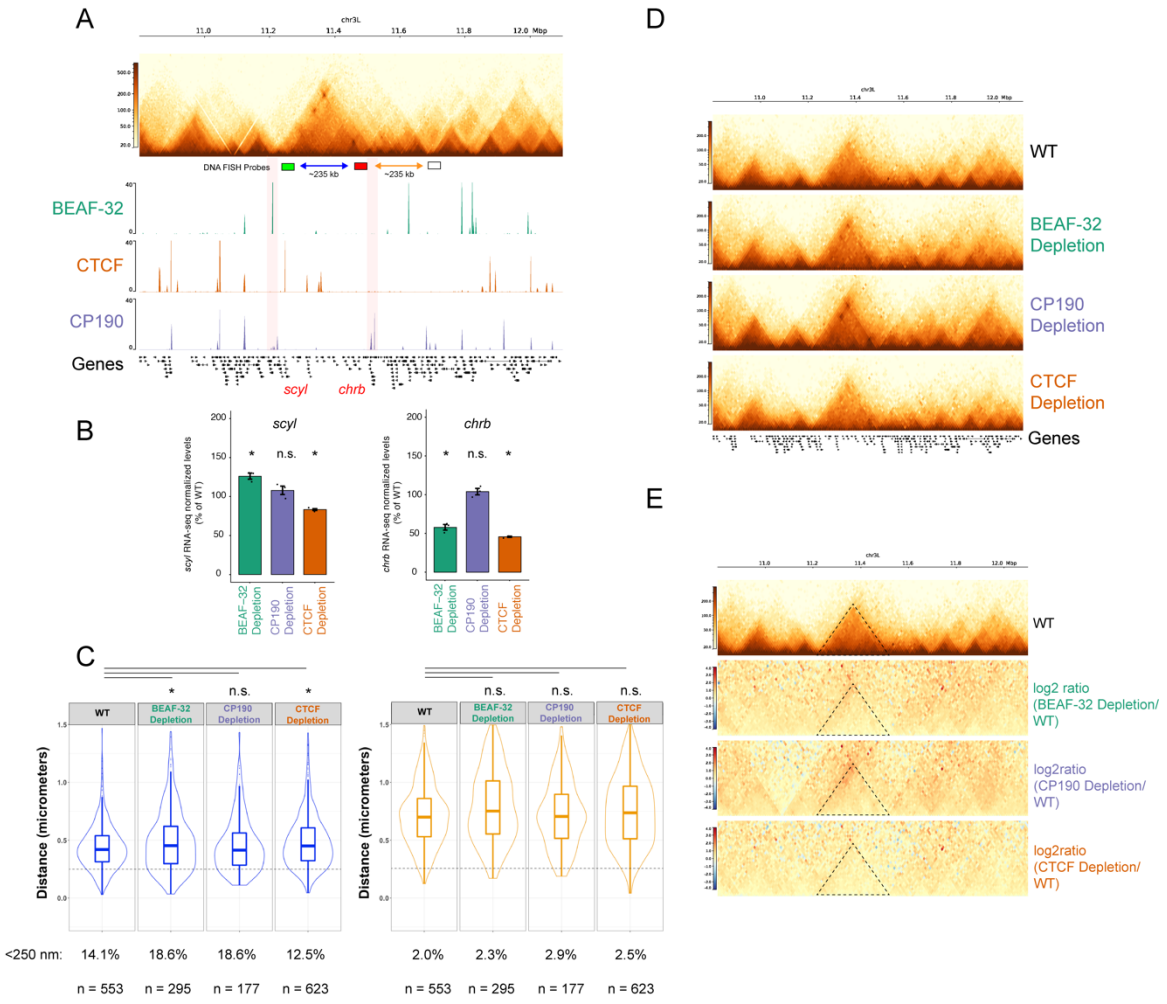

Figure 30: **DNA FISH and Hi-C to evaluate chromatin topology at the *scyl-chrb* locus following insulator protein depletion.**

(A) Hi-C from 2-4h AEL WT embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins in WT NC14 embryos. The position of the DNA FISH probes is shown below the Hi-C matrix. Blue and orange arrows indicate intra-TAD and inter-TAD comparisons, respectively. The boundaries of the *scyl-chrb* TAD are highlighted.

(B) Bar plots of RNA-seq normalized counts for *scyl* (left) and *chrb* (right). Different genotypes are shown relative to WT. Significance was assessed according to DESeq2 standard procedure, with a correction for multiple testing. n.s. = non significant, *padj < 0.01

(C) 3D distance distributions between intra-TAD probes (blue) and inter-TAD probes (orange) across different genotypes. n = number of alleles measured in at least 3 embryos. A Kolmogorov-Smirnov test was applied to evaluate statistical significance: n.s. = non-significant, *p < 0.01, **p < 0.001

(D) Hi-C maps of the *scyl-chrb* locus across different genotypes (5kb resolution).

(E) Differential Hi-C maps (log2 ratio of contacts between a given genotype and WT) of the *scyl-chrb* locus across different genotypes. The dashed triangle indicates the *scyl-chrb* TAD.

Our DNA FISH and Hi-C results indicated that the proximity between loop anchors within the *scyl-chrb* TAD may be slightly affected upon loss of insulator proteins. The current resolution of our Hi-C experiments, and the relative low frequency of these interactions at early stages (Fig. 17,18) (Ogiyama et al., 2018) does not allow more quantitative assessments of whether such loops are affected by the loss of insulator proteins.

4.6.2 TADs that do not contain high-frequency looping interactions

In the *LBR* TAD, the three insulator proteins bind at the right boundary, but do not bind at the left boundary, although BEAF-32 and CP190 occupy a site close to that boundary (Fig. 31A). The expression of *LBR* is slightly up-regulated upon depletion of any of the three insulator proteins (Fig. 31B). Intra-TAD distances measured by DNA FISH are similar between WT and CP190-depleted embryos, indicating that the TAD is able to form in the absence of this protein (Fig. 31C). However, we did detect a very subtle increase in the proximity between inter-TAD probes upon loss of CP190, suggesting that CP190 partially helps to insulate the left TAD boundary. Hi-C experiments showed that the LBR TAD can be detected in the absence of any of the three insulator proteins at NC14 (Fig. 31D). Corroborating our DNA FISH measurements, the Hi-C data indicates a subtle gain of contacts across the left TAD boundary, in all three genotypes in comparison to WT embryos (Fig. 31E).

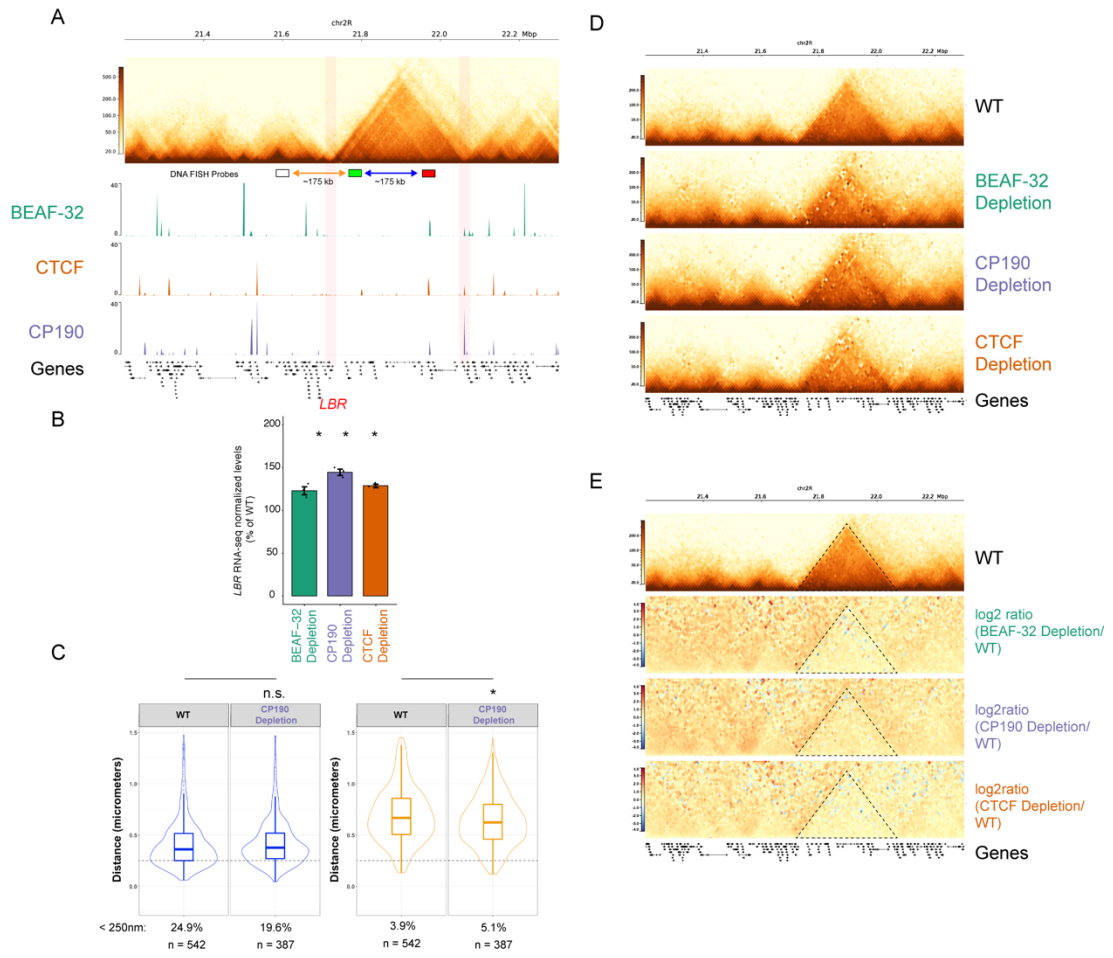

Figure 31: **DNA FISH and Hi-C to evaluate chromatin topology at the *LBR* locus following insulator protein depletion.**
(A) Hi-C from 2-4h AEL WT embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins in WT NC14 embryos. The position of the DNA FISH probes is shown below the Hi-C matrix. Blue and orange arrows indicate intra-TAD and inter-TAD comparisons, respectively. The boundaries of the *LBR* TAD are highlighted.
(B) Bar plot of RNA-seq normalized counts for *LBR*. Different genotypes are shown relative to WT. Significance was assessed according to DESeq2 standard procedure, with a correction for multiple testing. n.s. = non significant, *padj < 0.01
(C) 3D distance distributions between intra-TAD probes (blue) and inter-TAD probes (orange) in WT and CP190 depleted embryos. n = number of alleles measured in at least 3 embryos. A Kolmogorov-Smirnov test was applied to evaluate statistical significance: n.s. = non-significant, *p < 0.01, **p < 0.001
(D) Hi-C maps of the *LBR* locus across different genotypes (5kb resolution).
(E) Differential Hi-C maps (log2 ratio of contacts between a given genotype and WT) of the *LBR* locus across different genotypes. The dashed triangle indicates the *LBR* TAD.

In the three studied loci that establish their chromatin topology at NC14, TADs are able to form in the absence of BEAF-32, CP190 or CTCF, as shown by DNA FISH and Hi-C experiments. In some cases, there is a small loss of insulation at their boundaries or slight change in intra-TAD contacts after depletion of the insulator proteins, but this is not sufficient

to abolish TAD formation. We also observed slight changes in intra-TAD proximity following insulator protein depletion in the *scyl-chrb* locus. These subtle defects might be involved in gene expression changes observed at that locus, although this evidence is correlative.

We then decided to investigate the relationship between topological changes and transcriptional alterations in the absence of insulator proteins.

4.7 Evaluating the contribution of insulator proteins and transcription for the establishment of TAD boundaries

4.7.1 Disruption of transcription at TAD boundaries can be associated with defects in boundary function

We reasoned that since TAD boundaries often overlap with promoters and insulator binding sites, depletion of insulator proteins could simultaneously impact promoter transcriptional output and TAD boundary function. Therefore, we used our Hi-C and RNA-seq data from insulator-depleted embryos to evaluate if topological changes are associated to transcriptional defects. First, we asked whether DEGs tend to occur at TAD boundaries. The cumulative distance distribution of DEGs' promoters to TAD boundaries revealed that these promoters are not enriched at TAD boundaries above background in any insulator-depleted condition (Fig. 32). Nonetheless, we found examples of DEGs whose promoters overlap with TAD boundaries. For instance, the promoter of *scramb1* is occupied by CP190 and overlaps with a TAD boundary (Fig. 33A,B). CP190 depletion leads simultaneously to downregulation of *scramb1* (Fig. 33C) and increased contacts across that boundary (Fig. 33D,E).

In this case, as the expression of the gene is decreased, it does not reflect inappropriate activation of the gene by an enhancer across the weakened boundary (commonly referred to as "enhancer high-jacking" (Northcott et al., 2014). It could suggest a close relationship between transcription (or promoter activation) and TAD boundaries. However, it is difficult to distinguish between cause and consequence at such loci: does transcription influence TAD boundary formation, or is the boundary topology required for transcriptional activation?

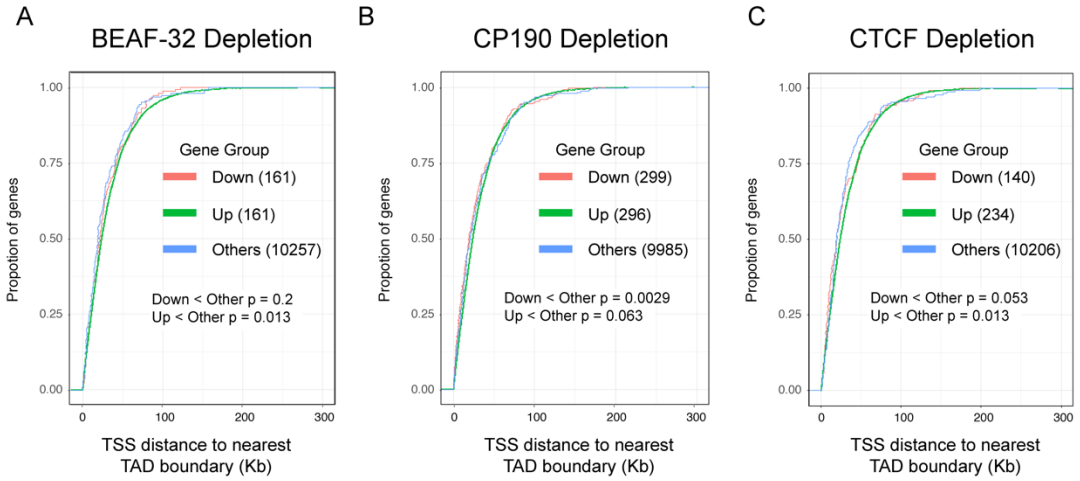

Figure 32: **DEGs can be found, but are not enriched, at TAD boundaries.**
(A-C) Cumulative distribution plots showing the distance between the TSS of DEGs in (A) BEAF-32-, (B) CP190 and (C) CTCF-depleted embryos and the nearest TAD Boundary. TAD boundaries were obtained from WT 2-3h AEL Hi-C data. Wilcoxon tests were performed for statistical analyses in (E-G).

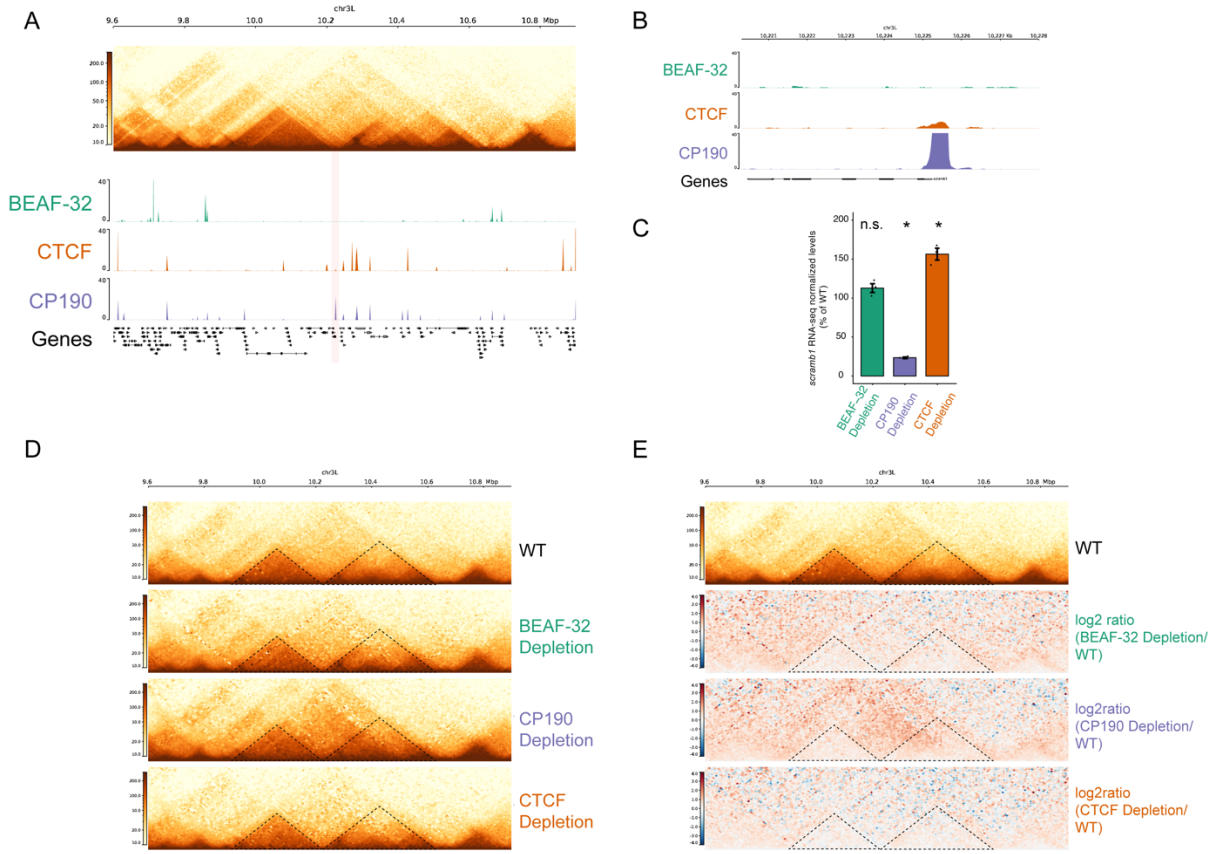

Figure 33: **DEGs can be found at disrupted TAD boundaries.**

(A) Hi-C from 2-4h AEL WT embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins in WT NC14 embryos. A TAD boundary overlapping with *scramb1* is highlighted.
(B) ChIP-seq tracks of insulator proteins in WT NC14 embryos at the *scramb1* TAD boundary.
(C) Bar plot of RNA-seq normalized counts for *scramb1*. Different genotypes are shown relative to WT. Significance was assessed according to DESeq2 standard procedure, with a correction for multiple testing. n.s. = non significant, *padj < 0.01
(D) Hi-C maps of the *scramb1* locus across different genotypes (5kb resolution). The dashed triangles indicates the TADs separated by the *scramb1* TAD boundary.
(E) Differential Hi-C maps (log2 ratio of contacts between a given genotype and WT) of the *scramb1* locus across different genotypes. The dashed triangles indicates the TADs separated by the *scramb1* TAD boundary.

4.7.2 Depletion of trans-acting factors reveals a combined role for insulator protein binding and transcription in TAD boundary function

We also found examples of TAD boundaries that lose insulation, but overlap with promoters that are not transcriptionally affected by insulator protein depletion. Such loci offer examples in which insulator protein binding could be directly responsible for TAD boundary function. We reasoned that dissection of individual, simpler, loci would help to tease apart different contributions of insulator proteins and transcription in setting up chromatin topology. We focused on a specific locus, which contains a TAD boundary that overlaps simultaneously with an insulator binding site and a promoter. The promoter regulates transcription of an isoform of *btsz*, a gene transcribed prior to NC14, and target of the pioneer transcription factor Zld (Fig. 34A). The insulator binding site is slightly distal to the promoter, and is occupied by CP190, but not BEAF-32 or CTCF (Fig. 34A).

First, we perturbed promoter function in *trans* by removing the maternal and zygotic contribution of *Zld* via a germline clone strategy, as Zld inactivation was previously shown to abolish *Btsz* transcription (Hug et al., 2017; Liang et al., 2008). Alternatively, we perturbed insulator protein binding at this boundary in trans by using embryos depleted of CP190. Importantly, depletion of CP190 (or BEAF-32 or CTCF) does not affect Btsz transcription, as measured by RNA-seq (Fig. 34B).

In order to assess boundary function, we measured by DNA FISH the 3D proximity between probes targeting regions within the two TADs that are separated by the Btsz TAD boundary. In WT embryos, the median 3D distance between the FISH probes is 447 nm (Fig. 34C). The strongest effect was observed after depletion of Zld, bringing the median distance down to 244 nm (Fig. 34C). Depletion of CP190 also affected boundary function, even though displaying a milder effect (median distance of 370 nm) (Fig. 34C). Hi-C experiments corroborate the DNA FISH measurements, also revealing higher interaction frequencies across

the Btsz boundaries upon loss of CP190 (Fig. 34D,E). These results suggest that both transcription and insulator protein binding are required for full TAD boundary function at the Btsz locus.

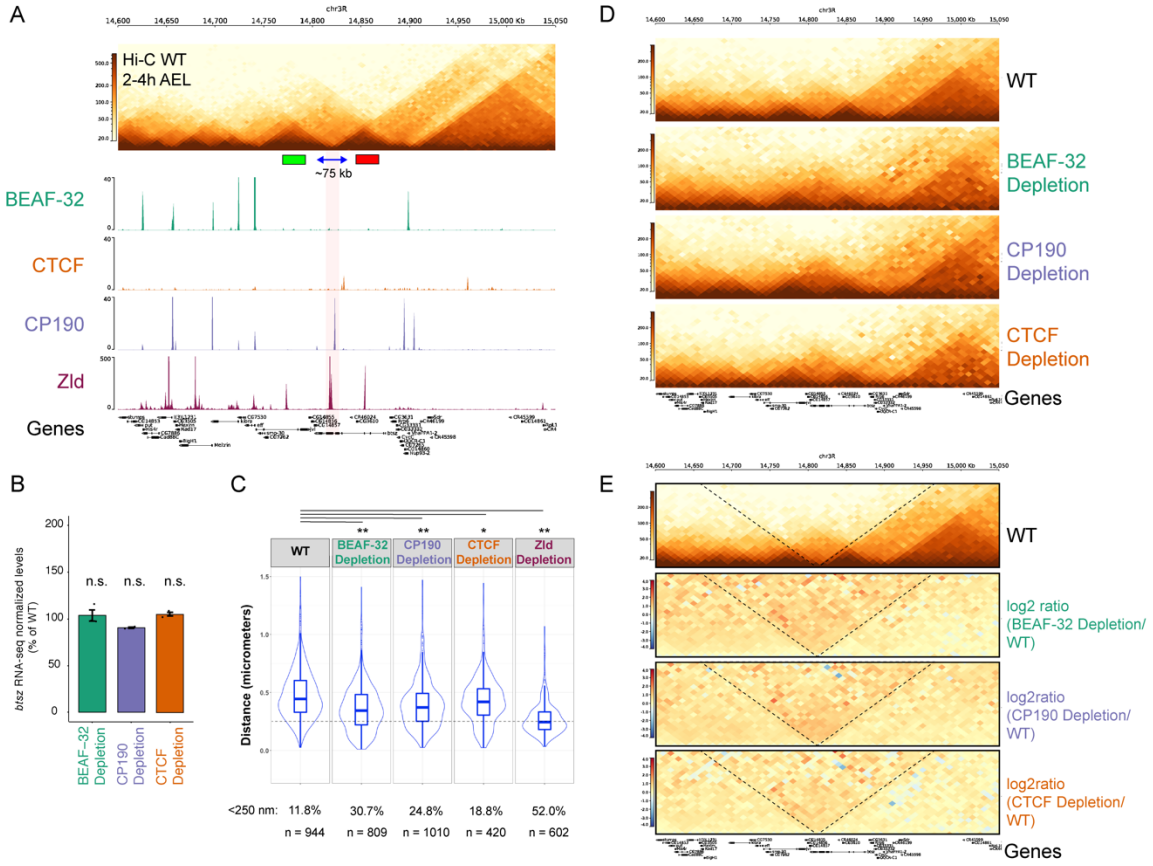

Figure 34: **DNA FISH and Hi-C to evaluate chromatin topology at the *btsz* locus following insulator protein and Zld depletion.**

(A) Hi-C from 2-4h AEL WT embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins and Zld in WT NC14 embryos. The position of the DNA FISH probes is shown below the Hi-C matrix.

(B) Bar plot of RNA-seq normalized counts for *btsz*. Different genotypes are shown relative to WT. Significance was assessed according to DESeq2 standard procedure, with a correction for multiple testing. n.s. = non significant, *padj < 0.01

(C) 3D distance distributions between DNA FISH probes indicated in (A) across different genotypes. n = number of alleles measured in at least 3 embryos. A Kolmogorov-Smirnov test was applied to evaluate statistical significance: n.s. = non-significant, *p < 0.01, **p < 0.001

(D) Hi-C maps of the *btsz* locus across different genotypes (5kb resolution).

(E) Differential Hi-C maps (log2 ratio of contacts between a given genotype and WT) of the *btsz* locus across different genotypes. The dashed lines indicates the *btsz* TAD boundary.

4.7.3 TAD boundary function may be partially regulated by elements acting from a distance

We next asked whether long-range effects could impact boundary function. We noticed that, while CP190 binds at the boundary, it is recruited there at a site without BEAF-32 or CTCF. Instead, ChIP-seq data from later timepoints shows co-occupancy of the CP190 site with another DNA-binding insulator protein (Su(Hw)) (not shown), which could be involved in CP190's recruitment. However, BEAF-32 binds to the adjacent TAD boundaries in the left and right directions at the *btsz* locus, while CTCF binds to a site distal from the boundary (~7 Kb) (Fig. 34A). Interestingly, both DNA FISH and Hi-C experiments revealed that the Btsz boundary is affected upon loss of BEAF-32 (to the same extent as CP190 loss), and CTCF (to a minor extent) (Fig. 34C-E). This suggests that TAD boundary function does not depend solely on trans-acting factors binding in *cis*, but also from a distance in *cis*. Furthermore, this might explain the widespread decrease in insulation at TAD boundaries detected after depletion of any of the three insulators (Fig. 28).

4.7.4 CRISPR genetic deletions pinpoint distinct elements within a boundary that mediate its function

To avoid potential confounding secondary effects from protein depletions in *trans*, and to more precisely discern the role of transcription and insulator protein binding in conferring boundary function, we decided to dissect the *btsz* TAD boundary by performing CRISPR deletions. The cloning and generation of flies carrying CRISPR deletions were performed by an internship student, Songjie Feng, under my supervision. We designed three distinct deletions: (1) promoter-proximal, encompassing the promoter and Zld binding sites, (2) insulator binding site distal to the promoter and (3) whole boundary deletion encompassing (1 + 2) (Fig. 35A,B). As expected given the requirement of *btsz* for embryo viability, deletion of the promoter or the whole boundary (1 and 3) are homozygous lethal. Deletion of the insulator binding site (2) is homozygous viable, corroborating the RNA-seq results that show unperturbed Btsz expression in CP190-depleted embryos. To assess the effects of the CRISPR deletions on boundary function, we again employed DNA FISH with probes across the boundary (Fig. 35A). All three deletions led to compromised boundary function, with different strengths. The deletions with stronger effect were the promoter and whole boundary deletions, while the insulator binding site had a mild, albeit detectable, effect (Fig. 35C). This indicates that insulator binding and transcription (or promoter function) are mechanisms that jointly

contribute to full boundary function, in the case of the *btsz* TAD boundary, which overlaps both features.

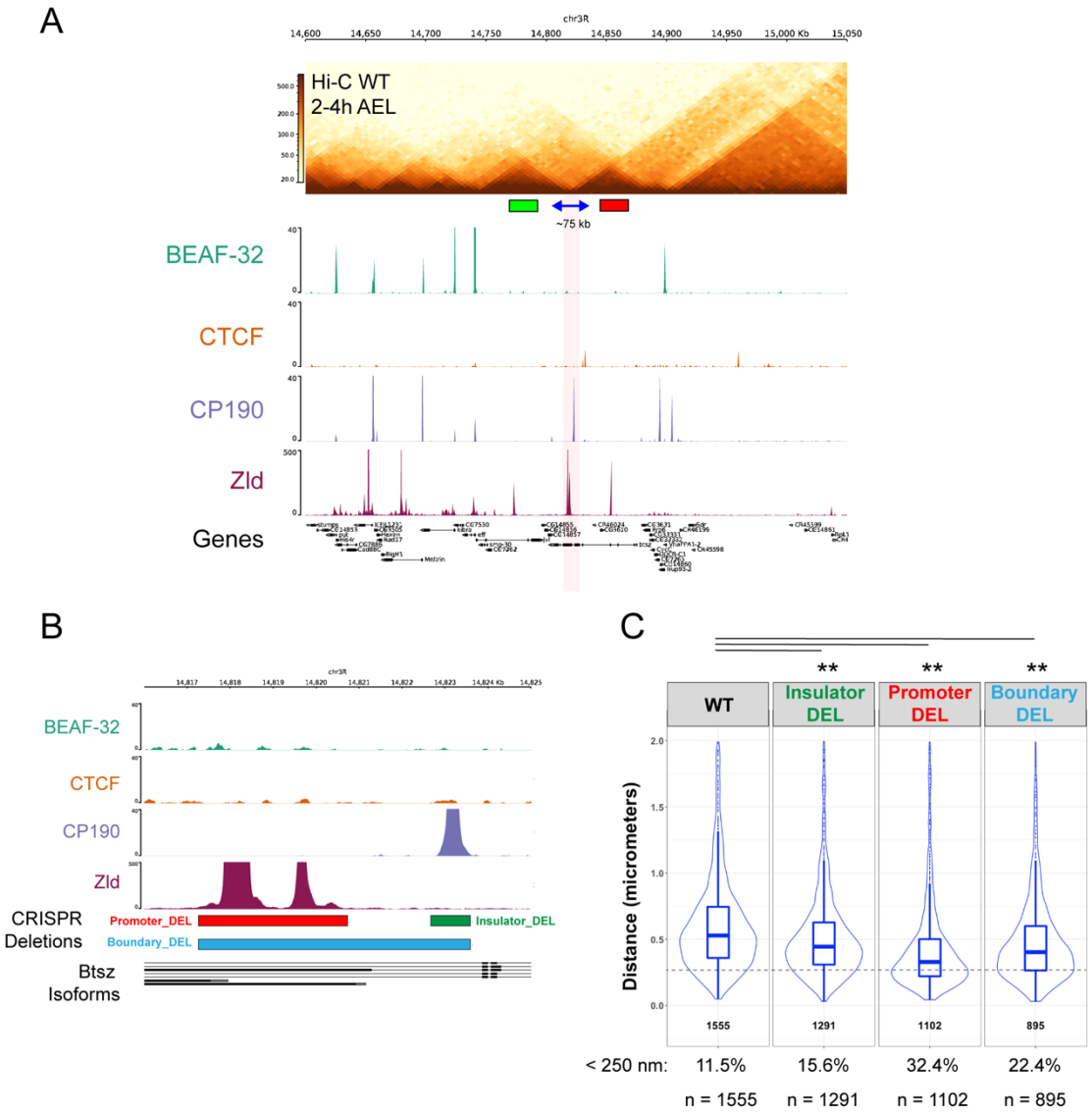

Figure 35: **Assessing the consequences of CRISPR deletions in the btsz TAD boundary by DNA FISH.**
(A) Hi-C from 2-4h AEL WT embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins and Zld in WT NC14 embryos. The position of the DNA FISH probes is shown below the Hi-C matrix.
(B) ChIP-seq tracks of insulator proteins and Zld in WT NC14 embryos at the *btsz* TAD boundary
(C) 3D distance distributions between DNA FISH probes indicated in (A) across different genotypes. n = number of alleles measured in at least 3 embryos. A Kolmogorov-Smirnov test was applied to evaluate statistical significance: n.s. = non-significant, *p < 0.01, **p < 0.001

5 Discussion and Perspectives I

5.1 What topological features do measurements by DNA FISH capture?

Recent studies suggested that TADs form throughout the genome during the onset of the major ZGA in *Drosophila* (Hug et al., 2017; Ogiyama et al., 2018). We performed DNA FISH experiments to carefully evaluate the dynamics of chromatin topology establishment in individual cells, at specific loci. By using DNA FISH with three probes targeting the same locus, we could measure the relative pair-wise 3D proximity between target sites. We designed our probes to compare intra-TAD and inter-TAD 3D proximity across different stages in three TADs; two TADs displaying high-frequency looping interactions by Hi-C and one TAD that does not. Our measurements for the three TADs are compatible with the establishment of TADs at NC14 (Figs. 17-19). However, there is still large cell-to-cell heterogeneity at this stage, with only 15% to 23% of nuclei having the loop anchors within 250 μm of each other. It is also difficult to distinguish if the high-frequency looping interactions at the *scyl-chrb* and *tsh-tio* TADs are already formed at NC14, or if the proximity results from TAD formation. Quantifying these distances at later embryonic stages showed that both the 3D proximity between loop anchors is higher, and there is much less cell-to-cell heterogeneity at 6-8h AEL in comparison to NC14: ~50% of nuclei have distances < 250 μm at 6-8h AEL (Figs. 17,18). Our Hi-C experiments with WT embryos at 2-3h AEL (mostly NC14 embryos) (Fig. 17,18) showed that the looping interactions are barely detectable at this stage, corroborating my DNA FISH findings. Interestingly, in the TAD not exhibiting looping interactions, the intra-TAD distance does not change from NC14 to a later timepoint (16-18h AEL) in WT embryos (Fig. 19).

We conclude that our DNA FISH measurements are compatible with TADs being established at the onset of the major ZGA. Although high-frequency loops may start being formed at that stage, they become more consistent with less cell-to-cell heterogeneity at later developmental stages. This is also compatible with published Hi-C data, indicating that loops become stronger with developmental time (Ogiyama et al., 2018).

5.2 Depletion efficiency

The depletion of BEAF-32, CTCF and CP190 was very efficient during the major ZGA; the stage when we assessed the topological and transcriptional consequences of those depletions (Fig. 22). Efficient reduction of CTCF was expected since we are using a knockout strategy; however, because BEAF32 and CP190 knockout flies are sterile, an RNAi strategy was needed. Even though BEAF-32 and CP190 RNAi was very efficient, the use of RNAi may have masked stronger phenotypic effects due to the presence of a small amount of proteins bound to chromatin. However, our western blot quantification indicates that this is not an issue, as a band was nearly not detectable even after long exposures (Fig. 22). Furthermore, we detected stronger topological and transcriptional effects after removing CP190 in comparison to CTCF (Fig. 24,26,33). This indicates that RNAi-mediated depletion of CP190 was sufficient to study its function. To more precisely quantify protein depletion levels, we could use a more quantitative strategy than regular Western blot. In fact, we are planning to quantify the amount of protein still bound to chromatin, by performing quantitative ChIP-seq experiments in insulator-depleted embryos.

5.3 TAD establishment is robust to depletion of CP190, BEAF-32 or CTCF

Our Hi-C experiments showed that most TADs were able to form upon depletion of BEAF-32, CTCF or CP190 (Fig. 26). This contrasts with the dramatic effects on chromatin topology observed upon CTCF depletion from mammalian cells (Nora et al., 2017). Three main possible explanations follow:

5.3.1 Achieved depletion levels may not be sufficient to disrupt topology

Studies of CTCF regulation of chromatin topology indicated that a strong depletion of this protein is required for detectable effects of chromatin topology. Reducing CTCF to 15% of its normal level is enough to sustain TAD insulation (Nora et al., 2017). Using our strategy, we completely removed CTCF protein by using a knockout system. Clearly this complete removal did not have a strong topological effect as observed for CTCF removal in mammalian cells. This indicates dramatic differential requirements for CTCF in chromatin topology across different animal species. However, for BEAF-32 and CP190, a non-complete depletion due to the use of RNAi might be compatible with TAD structure maintenance, even though the total level of those proteins following depletion is less than 1% of WT levels (Fig. 22).

The auxin-degron system was used to strongly deplete CTCF in mammalian cells, reducing the levels of chromatin-bound CTCF to less than 1% of WT (Nora et al., 2017). However, this system has not been adapted yet for use directly in fly embryos, possibly due to the difficulty in delivering drugs through the embryos' chorionic and vitelline membranes. One alternative is to use an optogenetic system to remove tagged proteins from the nucleus in a fast and controllable manner. Such a system has been recently adapted and enhanced for use in fly embryos in our lab (unpublished data from a previous PhD student, Dr. Anna Kögler), and is especially suitable for studies during early embryogenesis. It will be interesting to compare the efficiency of RNAi and optogenetic strategies in reducing chromatin binding for a target protein.

5.3.2 TAD boundaries may not rely entirely on insulator proteins to function

As described in the section 1.3.2, it is not clear by which mechanisms TADs form in *Drosophila*. Although flies have the required molecular components for a CTCF-cohesin loop extrusion mechanism, our results with CTCF-depleted embryos suggest that CTCF-mediated loop extrusion is not responsible for the formation of *Drosophila* TADs. Therefore, either loop extrusion operates with other proteins (or a functionally redundant set of proteins), or different mechanisms (e.g. compartmentalization-based) might be in place for TAD formation in flies.

Upon loss of insulator proteins, we did observe TAD boundary disruption at some loci (Figs. 26,33,34) and a global decrease in TAD boundary insulation (Fig. 27,28), underscoring that these proteins have a local and genome-wide influence in TAD formation.

At least at some loci, the combination of insulator protein binding and transcription may be important for TAD boundary function (Fig. 33,34). This hypothesis is supported by CRISPR deletions on the *btsz* locus, which has a TAD boundary simultaneously overlapping with a promoter and an insulator binding site (Fig. 35). Deleting the insulator binding site at that locus caused a decrease in TAD boundary insulation (Fig. 35). An even stronger decrease in insulation was observed upon deletion of the promoter (Fig. 35). However, deletion of the whole boundary (encompassing both the promoter and insulator binding site) led to an intermediate effect, relative to the deletions of promoter and insulator binding site (Fig. 35). As this is counter-intuitive, this observation requires further investigation.

5.3.3 Insulator proteins may be functionally redundant and compensate for the loss of single components

We removed three insulator proteins with very different properties (See insulator protein structures in Fig. 3). We explicitly focused on BEAF-32, CTCF and CP190 and excluded other insulator proteins for the following reasons:

(1) BEAF-32, CTCF and CP190 have been functionally characterized in depth, and bind to thousands of sites in the genome (Kyrchanova & Georgiev, 2014; Ozdemir & Gambetta, 2019);
(2) Some other insulator proteins bind to a more restricted set of genomic sites (e.g. ZIPIC binds to ~ 600 sites in fly chromatin (Zolotarev et al., 2016), Opbp binds to only ~ 40 sites (Zolotarev et al., 2017);
(3) Depletion of Su(Hw) leads to sterility, and we could therefore not examine its role in early embryogenesis
(4) *Mod(Mdg4)* has dozens of mRNA isoforms, making it very difficult to find shRNAs that target all isoforms.

It is possible that the loss of BEAF-32, CTCF and CP190 is compensated by other insulator proteins with similar characteristics. According to this hypothesis, CP190 would be less prone to compensation, since it acts as a general insulator co-factor, being recruited to chromatin by multiple insulator proteins. Consistent with this idea, my results show that loss of CP190 caused the strongest topological and transcriptional phenotypes (Fig. 24,26,33).

To address potential insulator protein redundancy, I am currently crossing *Drosophila* stocks to obtain flies carrying shRNAs against combinations of two insulators (BEAF-32/CTCF, CTCF/CP190 and BEAF-32/CP190), and will evaluate the topological and transcriptional consequences of removing those pairs. Nevertheless, as insulator proteins display multiple protein-protein interaction possibilities (e.g. homotypic and heterotypic, between DNA-binding proteins and/or their co-factors), it is very challenging to genetically disrupt all possible combinations of insulator interactions in *Drosophila*.

5.3.4 Insulator protein depletion leads to a global decrease in TAD insulation

Analysis of insulation at TAD boundaries revealed a surprising effect upon loss of BEAF-32, CTCF or CP190: TAD boundaries that are not occupied by those proteins displayed reduced insulation in the depleted embryos (Fig. 28). These findings are supported by DNA FISH measurements in the *btsz* locus (Fig. 34). TAD boundary insulation at the *btsz* locus is compromised upon depletion of BEAF-32, even though this insulator protein only binds to the next TAD boundaries (both upstream and downstream). We envisage two possible explanations for this 'indirect' loss of insulation:

(1) In *Drosophila*, the two boundaries of a TAD are often bound by distinct insulator proteins, many of which can bind as both homo and hetero-dimers, and higher order complexes. Supposing that interactions between TAD boundaries are required for TAD formation and insulation, perhaps loss of proteins at one boundary will affect the insulation at the other boundary as well.
(2) Loss of insulator proteins may cause a direct, specific decrease in insulation at the TAD boundaries they occupy. This in turn, may cause a global indirect effect on the conformation of chromosomes, impacting other TAD boundaries as well.

These possibilities are difficult to distinguish, and may require additional locus-specific genetic manipulations in *cis* to be disentangled.

5.4 Correlation between topological and transcriptional defects

An important open question is whether topological changes caused by the depletion of insulator proteins correlates with transcriptional mis-regulation measured by RNA-seq. Analyses to address this were hampered in our case by the difficulty in automatically detecting regions with altered topology in the Hi-C data, as the topology phenotypes are mild. We used different software to automatically detect differential interactions between Hi-C maps in insulator-depleted embryos in comparison to WT (Djekidel, Chen, & Zhang, 2018; Lun & Smyth, 2015; Ramirez et al., 2018). However, those analyses failed to capture differential interactions. We suspect that such algorithms are more suited to detect punctual changes in interactions (such as changes in high-frequency loops) which are more commonly found in mammalian TADs (e.g. focal interactions between CTCF occupied boundaries). The most visible changes we observed were a gain of interactions between two or more whole TADs (see examples in Fig. 26 and Fig. 33), rather than changes in focal interactions.

Two recently developed computational tools to detect differential interactions in Hi-C maps may help. Both use computer vision and image analysis algorithms to find altered patterns in Hi-C contact matrices between biological conditions (Galan et al., 2020; Matthey-Doret et al., 2020). That is an innovative approach, which shifts the quantification of differential interactions from measuring Hi-C interaction frequencies to an image-analysis-based comparison. Preliminary analyses suggest that the most obvious Hi-C changes in our datasets are captured by those methods.

Once we systematically detected differences between Hi-C datasets, in an unbiased way, we will move to the important task of correlating those changes with transcriptional defects.

5.5 Conclusions

Taken together, my results underscore the robustness of the *Drosophila* genome to the loss of individual insulator proteins. Moreover, they highlight the context-specificity of chromatin organization, given that loci are differentially affected by loss of insulator proteins. CP190 depletion led to overall stronger transcriptional and topological disruptions than BEAF-32 or CTCF depletion, consistent with its role as an insulator protein that can be recruited to chromatin by multiple insulator proteins. Experiments challenging chromatin organization through combined depletion of insulator proteins will likely have more severe affects and may shed light on requirements of topology for specific loci. Moreover, my results indicate that both insulator protein binding and transcription can be involved in insulation of a TAD boundary. New methods to systematically perturb and detect changes in chromatin topology across conditions may help to connect topological changes to transcriptional defects, and reveal the relationship between topology and gene expression.

6 Results II

Understanding the molecular determinants of TAD boundary function and their implications for gene regulation

In vertebrates, CTCF and cohesin have been recently demonstrated to play an essential role in TAD formation, via the binding of CTCF at convergent sites at TAD boundaries (de Wit et al., 2015; Nora et al., 2017; S. S. Rao et al., 2014; Sanborn et al., 2015). Their depletion leads to disruption of most TADs in population-based assays, such as Hi-C (Nora et al., 2017; S. S. P. Rao et al., 2017; Schwarzer et al., 2017). Though most boundaries depend on CTCF/cohesin to form in vertebrates, the extent of their functional diversity remains unexplored. It is unclear whether all boundaries are equivalent, i.e. would any CTCF-occupied boundary be able to replace a mutated CTCF-boundary in such a system?

In *Drosophila*, decades of work revealed that insulator elements are highly diverse (Kyrchanova & Georgiev, 2014; Ozdemir & Gambetta, 2019). Classic studies have shown that insulators can either block or stimulate enhancer-promoter communication, depending on their arrangement, and the local genomic context (Geyer & Corces, 1992; Muravyova et al., 2001). Other studies have shown that discrete regions within a single insulator element can be responsible for context-specific functions, such as blocking or promoting enhancer-promoter communication (Iampietro et al., 2008; Kyrchanova et al., 2016).

Studies on *Drosophila* insulators tend to focus on a handful of elements that have been thoroughly characterized. Though each of the thousands of genomic sites occupied by insulator proteins is a potential functional insulator element, we still don't know (1) whether these sites are relevant for gene regulation, (2) whether they are active at different developmental stages and tissues, or (3) how they act. Similar to vertebrates, a proportion of insulator protein binding sites in *Drosophila* overlaps with TAD boundaries (Ramirez et al., 2018; Sexton et al., 2012).

During my thesis, I have explored what determines if an insulator binding site forms a functional TAD boundary, and what is the relevance of TAD boundaries for gene regulation. More specifically:

1) How does the genomic context influence TAD boundary function?

2) What are the required and sufficient components for TAD boundary function in a given locus?

3) Can the reorganization of the chromatin topology in a given locus/TAD affect local transcription?

We therefore decide to insert endogenous *Drosophila* TAD boundaries into ectopic locations in the middle of TADs, to evaluate whether such boundaries remain functional, affect the topology of the targeted locus and are able to locally disrupt gene expression. The approach is outlined in (Fig. 36). In brief, we cloned selected TAD boundaries from *Drosophila*, and inserted them into other TADs (Fig. 36A). These TADs may contain high-frequency looping interactions or not (Fig. 36B). We observed different outcomes from such insertions. Some of these insertions do not cause detectable topological disruption, and were therefore not able to function as a boundary in an ectopic location. However, approximately half of the insertions we generated produce an effect, which varies depending on the locus (Fig. 36C).

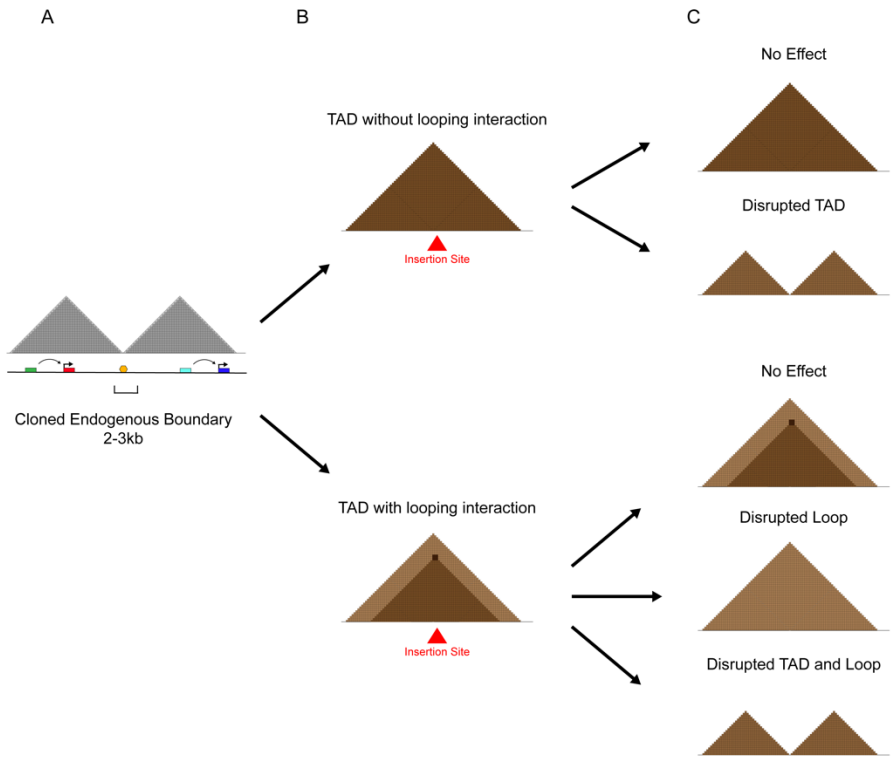

Figure 36: **Outline of the general strategy for TAD boundary insertions.**
(A) TAD boundaries were selected and fragments spanning 2-3 kb around them were cloned.
(B) The cloned TAD boundaries were inserted in the middle of TADs containing different properties.

(C) Different chromatin topology outcomes are expected: acceptor TADs may not be affected; intra TAD proximity may be disrupted due to creation of a new TAD boundary; if present, high-frequency loops may be disrupted without affecting TAD structure.

The work described in this chapter is the result of a collaboration with multiple colleagues, being led by a collaborating Ph.D. student Martina Varisco and I. Dr. Rebecca Rodriguez Viales assisted with ChIP-seq and Capture-Hi-C experiments. Dr. Charles Girardot provided support with computational analyses. Dr. Raquel Marco-Ferreres and Dr. Rebecca Rodriguez-Viales aided with fly work.

6.1 A recombination-mediated cassette exchange strategy to insert TAD boundaries into ectopic locations

6.1.1 The MiMIC system

To generate boundary insertions, we used a strategy based on the *Minos-mediated integration cassette* (MiMIC) resource (Venken et al., 2011), which consists of performing Recombination-Mediated Cassette Exchange (RMCE) in the *Drosophila* genome. In brief, three components are necessary: flies that carry an acceptor DNA cassette flanked by attP sites integrated into a single genomic location, a donor DNA cassette flanked by attB sites that is exchangeable with the acceptor cassette, and the phiC31 integrase enzyme (Fig. 37A). In this system, the insertions can happen in either orientation, randomly (Fig. 37B). As there are no other acceptor sites in the *Drosophila* genome, another key advantage of the system is that the donor DNA cassette can only insert at this defined location. There are no off-target effects, and the site of integration is molecularly mapped, and can be simply followed by eye/body color.

Flies carrying integrated acceptor cassettes at hundreds of genomic locations were generated as part of the MiMIC resource (Venken et al., 2011). Each fly line carries an acceptor cassette in an unique insertion site. The acceptor cassette is flanked by Minos transposon sequences, and attP sites in an inverted orientation (Fig. 37A). The site also contains a splice acceptor motif and an EGFP transgene that can be used as a gene expression reporter when the acceptor cassette is integrated within an intronic sequence. Additionally, there is a yellow+ marker, whose loss can be used to detect the integration event.

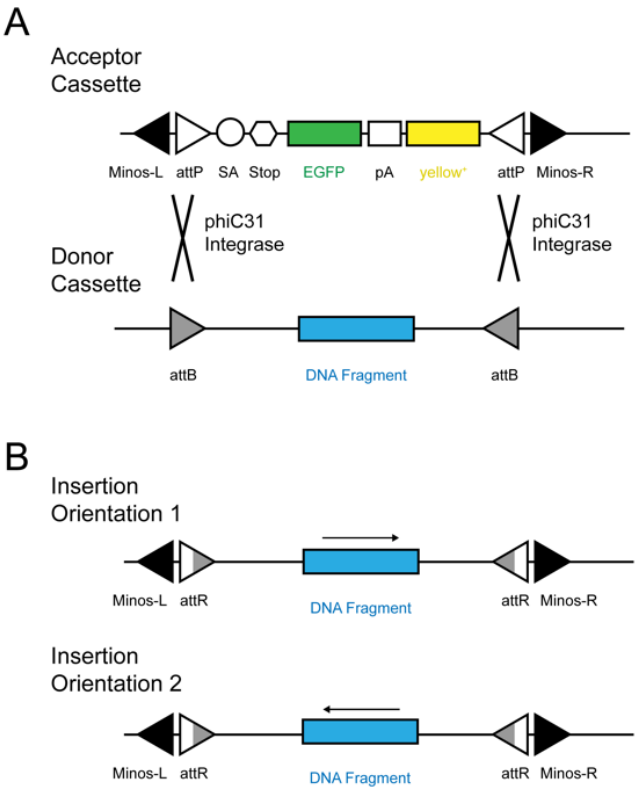

Figure 37: **Recombinase-Mediated Cassette Exchange (RMCE) is accomplished with the MiMIC system.**
(A) Scheme representing RCME. An acceptor cassette, flanked by Minos transposon sequences and attP sites in an inverted orientation, can be swapped with a donor cassette flanked by attB sites, in a reaction catalysed by the phiC31 integrase.
(B) In this strategy, insertion of the donor cassette can occur in either orientation, randomly.
SA = Splicing acceptor site, STOP = stop codon, pA = polyadenylation site

We pursued two main ways to deliver the donor DNA for this RMCE strategy: (1) by injecting the flies with plasmid DNA containing the donor cassette or (2) by generating transgenic flies that carry an excisable copy of the donor cassette integrated into the genome, followed by crosses with flies carrying the acceptor cassette.

6.1.2 Outline of the injection-based strategy

First, we generated plasmids containing a donor DNA cassette, which consisted in an endogenous TAD boundary (~2-3kb, see below) and an excisable *dsRed* marker, both flanked by attB sites in an inverted orientation (see Methods) (Fig. 38A). Those plasmids were injected into *Drosophila* embryos which carried an acceptor cassette at a location of interest, and the

phiC31 integrase. In the presence of the donor and acceptor DNA cassettes, the phiC31 integrase can recombine the attB and attP sites (Fig. 38A). This leads to the exchange between the donor and acceptor DNA fragments, and thus integration of the donor DNA into the acceptor site of interest, which can be tracked with the dsRed marker. Importantly, the insertion can occur in either orientation randomly (Fig. 38B).

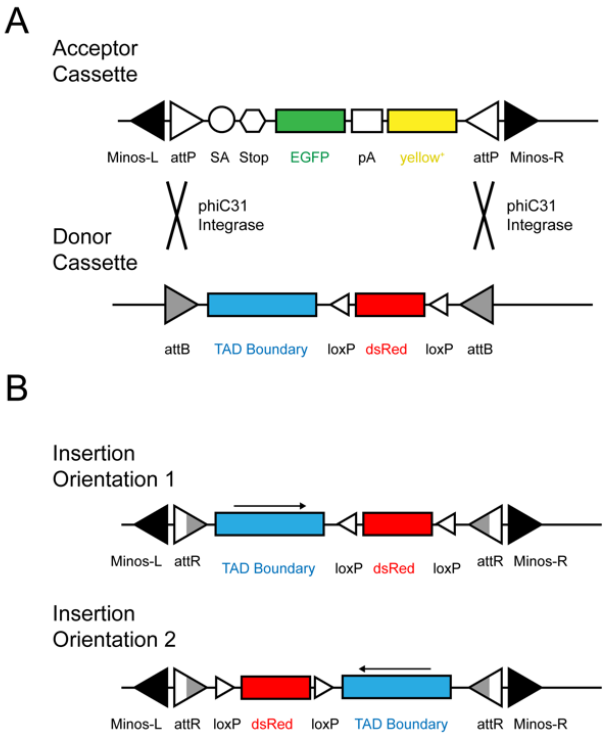

Figure 38: **Boundary insertion via direct injection of donor plasmid into Drosophila line carrying acceptor site.**
(A) Scheme representing RMCE using a donor plasmid carrying a DsRed marker.
(B) In this strategy, insertion of the donor cassette can occur in either orientation, randomly.
SA = Splicing acceptor site, STOP = stop codon, pA = polyadenylation site

6.1.3 Outline of the crossing-based strategy

Similarly to the injection-based strategy, we generated plasmids containing a different donor DNA cassette. This consisted in an endogenous TAD boundary (~2-3kb, see below) and a *mini-white* marker, both flanked by attB sites in an inverted orientation (see Methods) (Fig. 39A). Importantly, the acceptor cassette is flanked by FRT sites, and thus excisable by the Flp recombinase. The plasmid also contained sequences required for P-element transformation. Thus, upon injection the plasmid integrates randomly in the genome via P-element

transformation (Fig. 39A). This transgenic fly is then crossed to flies carrying the MiMIC acceptor site in the location of interest, a heat-shock inducible *Flp recombinase*, and the *phiC31* integrase. In the progeny, the donor cassette is excised from the genome upon heat-shock (due to expression of the Flp recombinase) (Fig. 39A). The donor cassette can then be exchanged with the acceptor cassette in a reaction catalyzed by the phiC31 integrase (Fig. 39B). Similarly to the injection-based strategy, the insertion can occur in either orientation randomly (Fig. 39C).

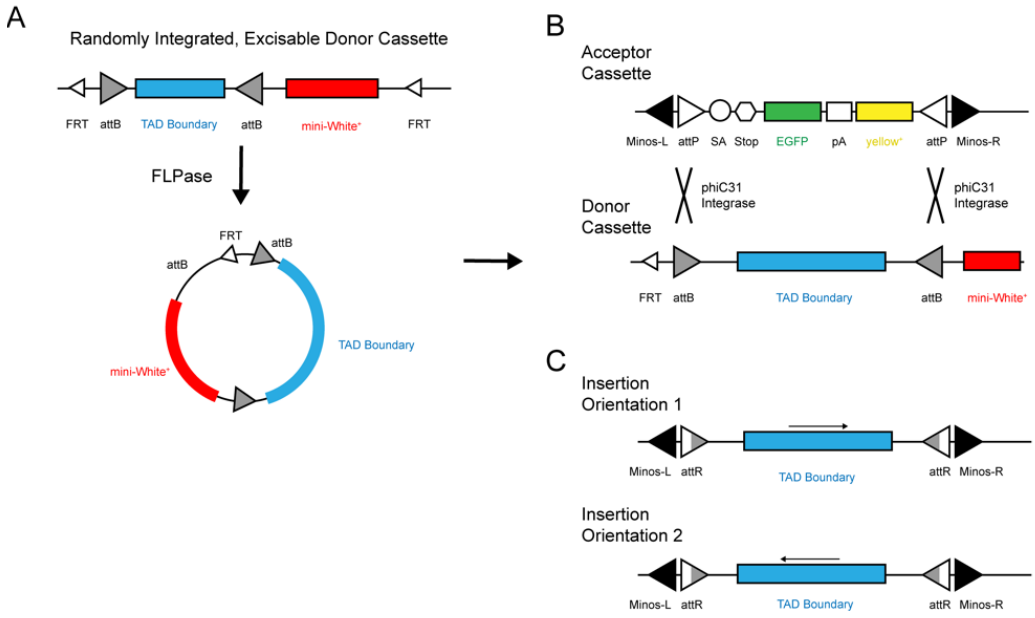

Figure 39: **Boundary insertion via a crossing-based strategy.**
(A) First, flies carrying an integrated version of the donor cassette are generated via P-element insertion. After crosses to combine this construct and the acceptor cassette in the same fly, the excision of the donor cassette can be stimulated via heat-shock induced Flp expression.
(B) Scheme representing RMCE using the excised donor plasmid.
(C) In this strategy, insertion of the donor cassette can occur in either orientation, randomly. Flies carrying the insertion can be identified by loss of the yellow marker (from the acceptor cassette).
SA = Splicing acceptor site, STOP = stop codon, pA = polyadenylation site

6.2 Selecting TAD boundaries and landing sites

To assess how the genomic context affects TAD boundary function, we searched for TAD boundaries and landing sites with heterogeneous properties. Specifically, we selected TAD boundaries based on their overlap with multiple different combinations of insulator

proteins and/or promoters, and acceptor sites within TADs displaying distinct properties related to size, gene content and insulator protein binding.

The selections were guided by chromatin topology data (Hi-C) and insulator protein binding data (ChIP-seq). We used high-resolution Hi-C data from 6-8h AEL embryos previously generated in the lab (unpublished). In order to map insulator protein binding specifically at the 6-8h AEL stage, we generated ChIP-seq data for the insulator proteins BEAF-32, CTCF, CP190 and Su(Hw) using staged WT embryos. We decided to focus on the 6-8h AEL stage because chromatin topology is more consolidated than in early embryos. For example, our DNA FISH experiments showed that the proximity between loop anchors is increased in 6-8h AEL embryos in comparison to early stages (NC14) (Fig. 17,18). Other studies also support the notion of chromatin topology becoming more robust after early stages (Hug et al., 2017; Ogiyama et al., 2018).

6.2.1 Computational classification of TAD boundaries into multiple categories

To select heterogeneous TAD boundaries to be inserted, we first computationally obtained the position of TAD boundaries based on their insulation score (see 3.9.2.4). In parallel, we called peaks from the 6-8h AEL insulator ChIP-seq data using MACS (see 3.7.4.1). Then, we classified boundaries into categories, based on whether they overlap with combinations of the following ChIP-seq peaks: (1) CP190, (2) BEAF-32, (3) CTCF or (4) Su(Hw). Since the majority of *Drosophila* TAD boundaries overlap with the insulator protein CP190 (88% or 1121/1270 boundaries in 6-8h AEL Hi-C), we decided to narrow our selection to TAD boundaries that overlap with this factor. Those categories were then sub-divided based on their overlap with a promoter (which could be a single or divergent promoter). This resulted in 25 categories, each with a different number of boundaries throughout the genome (Table 5).

Table 5: Division of TAD boundaries into 25 categories.

# of Boundaries	CP190	CTCF	BEAF-32	Su(Hw)	Promoter
18	■				
94	■				Single
133	■				Divergent
30	■	■			
87	■	■			Single
80	■	■			Divergent
6	■		■		

71					Single
177					Divergent
9					
19					Single
21					Divergent
12					
98					Single
180					Divergent
26					
23					Single
14					Divergent
0					
2					Single
2					Divergent
3					
11					Single
5					Divergent
149		?	?	?	?

6.2.2 Selection of TAD boundaries

Having defined the 25 boundary categories, I selected examples from the 9 best categories (defined below) to be cloned. I visually screened the boundaries from each category using a Hi-C genome browser (Juicebox) (Durand et al., 2016) loaded with the Hi-C and ChIP-seq data to select the candidates. This visual screening was important since boundary calling is sensitive to Hi-C resolution and the algorithm being used, which can sometimes lead to boundaries being called in unstructured regions of the genome. To overcome that, we manually selected boundary regions. Of note, for some categories, there was no convincing example of a TAD boundary, leading us to excluded those from the analyses. Using this approach, we selected 16 boundaries representing 9 different categories (Table 6).

Table 6: Coordinates of TAD boundaries selected for cloning.

chr	start	end	Boundary ID	CP190	CTCF	BEAF-32	Su(Hw)	Promoter
chr3L	10897481	10900569	14.2					Single
chr2R	21479673	21482576	3.1					Divergent
chrX	8572475	8575127	9.1					Divergent
chr3R	4820372	4822479	5.3					Single
chr3L	6993614	6996923	16.2					
chrX	17283782	17286719	11.2					Single
chr2L	808691	811086	11.3					Single

chr3R	17810546	17812852	12.1					Divergent
chr3L	17556384	17558888	5.2					Single
chr2R	14301700	14304020	14.1					Single
chr3R	20007647	20009928	6.1					Divergent
chr3R	24944115	24946221	6.2					Divergent
chr2R	7915420	7918221	15.1					Divergent
chr3L	12001928	12004244	15.2					Divergent
chr3R	5732341	5734625	11.1					Single
chr3L	5869139	5872197	17.1					Single

I then designed PCR primers to amplify DNA sequences spanning from 2 to 3 kb around the boundary. When present, insulator protein binding site(s) and promoter(s) were included in the amplicon. Two examples of selected TAD boundaries that overlap with distinct features are shown in (Fig. 40) and (Fig. 41). The Boundary 15.1 is occupied by BEAF-32, CTCF and CP190, overlaps with divergent promoters, and separates a gene-poor from a gene-rich domain (Fig. 40A). The cloned region is a 2.8 Kb region shown in (Fig. 40B). The Boundary 17.1 is occupied by CTCF, Su(Hw) and CP190 and overlaps with a single promoter (Fig. 41A). This boundary also separates a gene-poor domain from a gene-rich one. The cloned region (3.0 Kb) is shown in (Fig. 41B). The list of all cloned TAD boundaries, with their genomic coordinates (dm6) and associated features, is presented in Table 6.

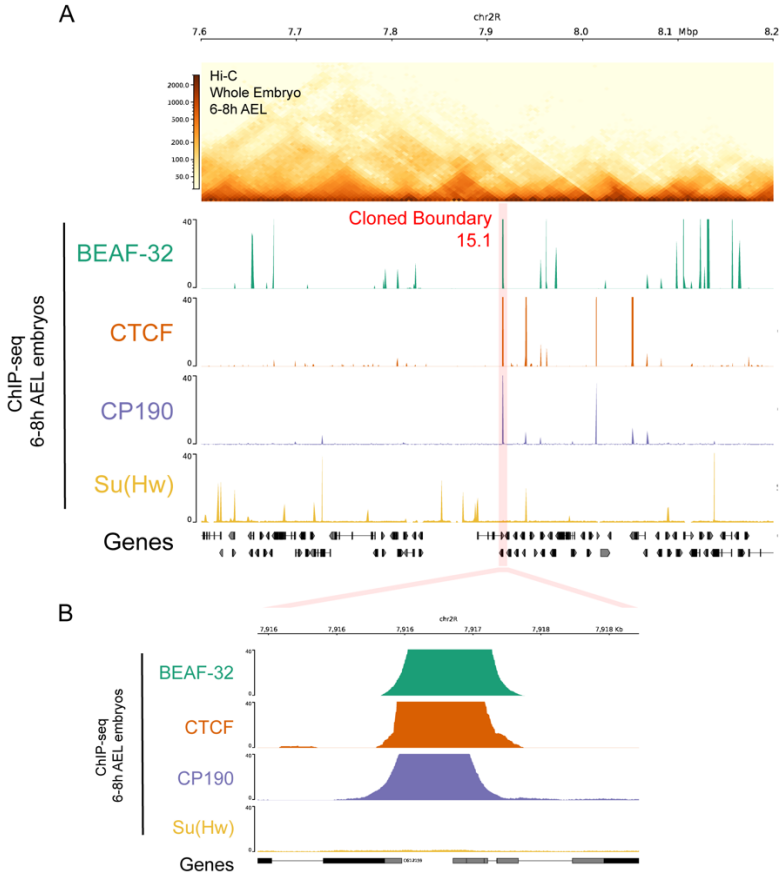

Figure 40: **An example of a cloned TAD boundary: 15.1.**
(A) Hi-C from 6-8h AEL embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins at the same stage. The cloned TAD boundary is highlighted in red.
(B) Snapshot of ChIP-seq tracks for insulator proteins from 6-8h AEL embryos, spanning the region cloned at the boundary 15.1

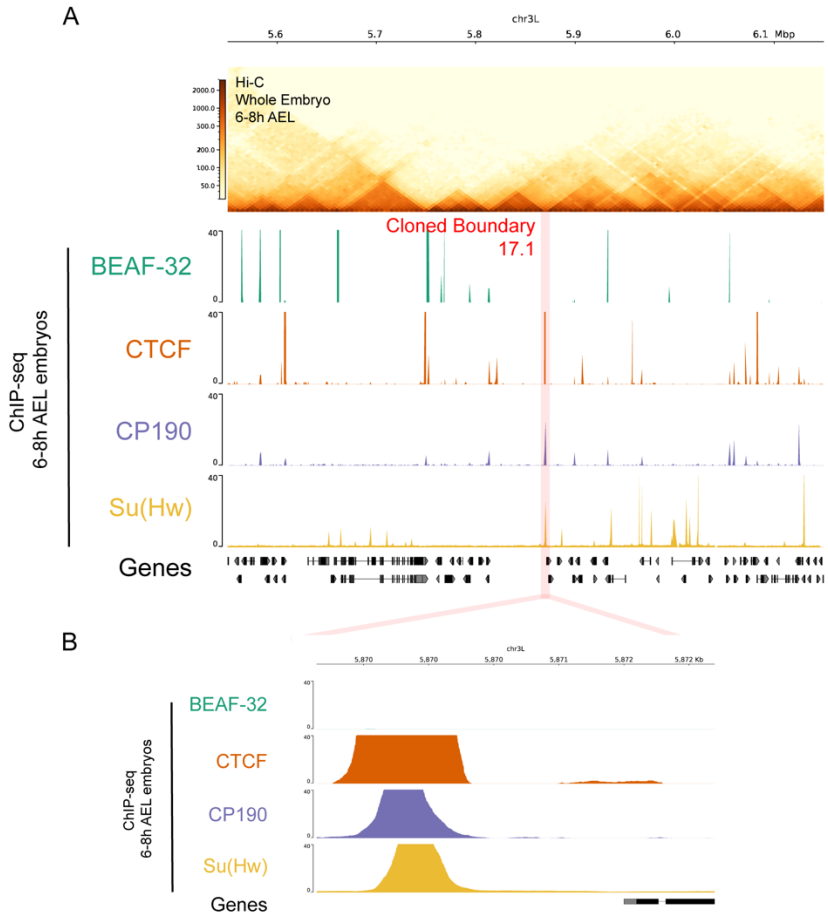

Figure 41: **An example of a cloned TAD boundary: 17.1.**
(A) Hi-C from 6-8h AEL embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins at the same stage. The cloned TAD boundary is highlighted in red.
(B) Snapshot of ChIP-seq tracks for insulator proteins from 6-8h AEL embryos, spanning the region cloned at the boundary 17.1.

6.2.3 Selection of landing sites

I used four main criteria for landing site selection:

(1) Size of the TAD. *Drosophila* TADs can vary in size, and are generally smaller than typical vertebrate TADs. Since we planned to analyze dozens of insertions, we needed to be able to resolve potential effects on chromatin topology with affordable readouts. Larger TADs would allow the use of lower resolution, cheaper readouts with both imaging and genomics experiments.

(2) Presence of developmentally-regulated genes. We expected that at least some TAD boundary insertions would perturb local chromatin topology, with potential effects on the transcription of local genes. We hypothesized that selecting loci containing dynamically-expressed genes (i.e. developmentally-regulated) controlled by one or multiple enhancers, would facilitate the detection of topological-induced transcriptional changes. Therefore, we selected loci containing genes expressed dynamically during embryonic development.

(3) Heterogeneous combination of insulator proteins at the TAD boundary. In order to assess the context-dependency of TAD boundary function, we searched for acceptor sites within TADs that have different combinations of insulator proteins binding at their boundaries. We hypothesized that some inserted boundaries would retain their activity in only certain contexts, which could be influenced by the presence of specific insulator proteins nearby.

(4) Availability of a fly line carrying a MiMIC acceptor site in that TAD, as we took advantage of commercially available fly lines carrying acceptor cassettes.

To search for the landing sites, we again visually inspected the *Drosophila* genome using the Juicebox Hi-C browser, looking for regions that fulfil the above criteria. We used the previously mentioned high-resolution Hi-C data from 6-8h AEL embryos generated in the lab, and our newly generated ChIP-seq data from the same stage.

The criteria mentioned above resulted in the selection of two different regions as landing sites (TAD A and B).

6.2.4 Properties of the selected landing sites

6.2.4.1 **TAD A**

The first locus, denominated TAD A, consists of a ~310 Kb TAD encompassing the genes *scylla* (*scyl*) and *charybde* (*chrb*) (Fig. 42A), and was also studied in Results I section. This TAD contains a three-way looping interaction: the left anchor is upstream of the gene *scyl*, the right anchor is at the gene *chrb*, while the middle anchor is in a non-coding region (Fig. 42A). Both left and right anchors are occupied by CTCF at 6-8h AEL, albeit the left anchor displaying much higher ChIP-seq enrichment (Fig. 42B). The MiMIC acceptor site is located in between the loop anchors (Fig. 42A). The looping interaction is embedded in a larger TAD structure, from which the left boundary is occupied by BEAF-32 and CTCF and overlaps

with divergent promoters of the genes *Ube3a* and *CG7600* (Fig. 42B). The right TAD boundary is occupied by CTCF, Su(Hw) and CP190, at the 3' UTR of the gene *Mob2* (Fig. 42B).

The genes *scyl* and *chrb* are paralogous, and both are broadly expressed in the developing embryo (Scuderi, Simin, Kazuko, Metherall, & Letsou, 2006). Knockout of either *scyl* or *chrb* does not affect embryo viability, while double knockout is embryonic lethal, due to defects in head formation (Scuderi et al., 2006). There are other 12 protein-coding genes at that locus, and multiple unannotated non-coding transcripts, all of which are either very lowly or not expressed during embryogenesis, and the function of these transcripts is unknown.

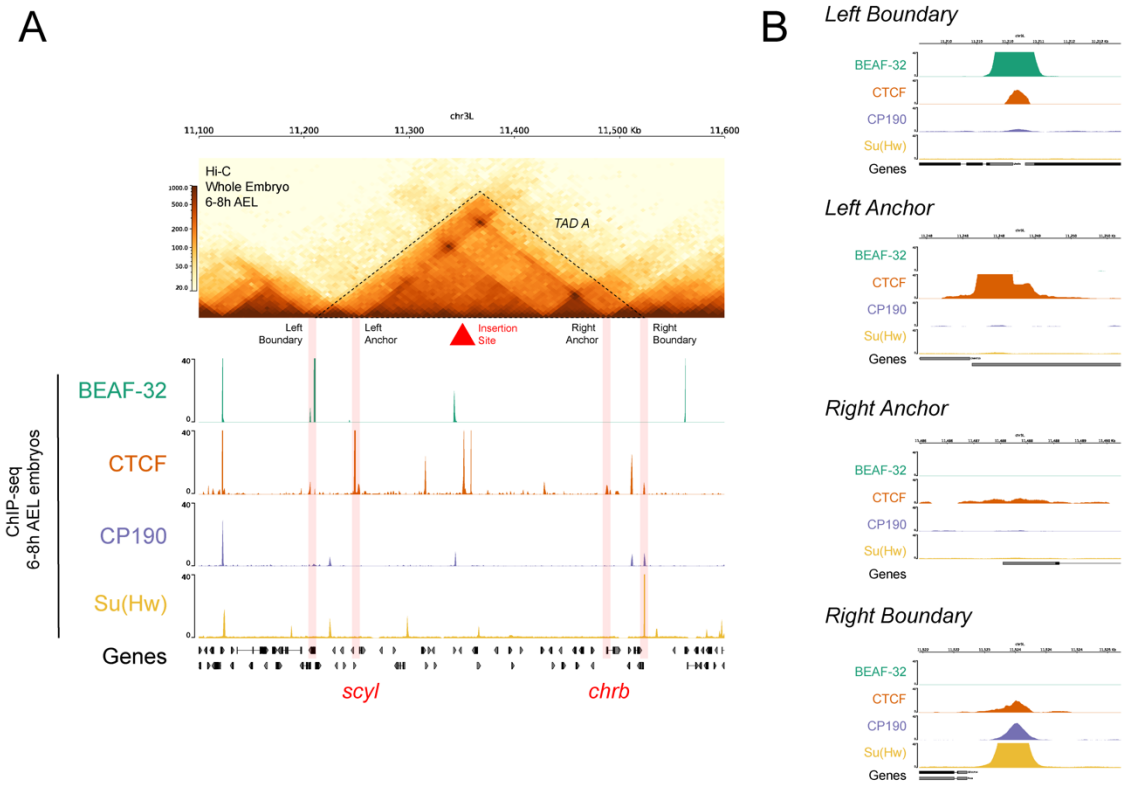

Figure 42: **The landing site at TAD A.**
(A) Hi-C from 6-8h AEL embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins at the same stage. The dashed triangle indicates TAD A. The acceptor site location is indicated by an red arrow. The TAD boundaries and loop anchors in TAD A are highlighted in red.
(B) Snapshot of ChIP-seq tracks for insulator proteins from 6-8h AEL embryos, spanning TAD A's boundaries and loop anchors.

6.2.4.2 **TAD B**

This locus contains the genes *escargot* (*esg*), *worniu* (*wor*) and *snail* (*sna*) within a ~230 Kb TAD (Fig. 43A). Contrary to TAD A, this TAD does not display high-frequency looping interactions detectable by Hi-C at the 6-8h AEL stage (Fig. 43A). The left TAD B boundary is occupied by CTCF, BEAF-32 and Su(Hw), and overlaps with the promoter of the gene *Cul3* (Fig. 43B). The right boundary is occupied by CTCF and CP190, and overlaps with the promoter of *lace* (Fig. 43B). The MiMIC landing site is located in the middle of the TAD, between the genes *esg* and *wor* (Fig. 43A).

The three genes (*esg, wor* and *sna*) are expressed in very specific patterns in the embryo, and are required for correct mesoderm- and neuro-development (Ashraf, Ganguly, Roote, & Ip, 2004; Hekmat-Scafe, Dang, & Tanouye, 2005; Simpson, 1983). There are 8 additional protein-coding genes at this TAD, which are not expressed during embryogenesis.

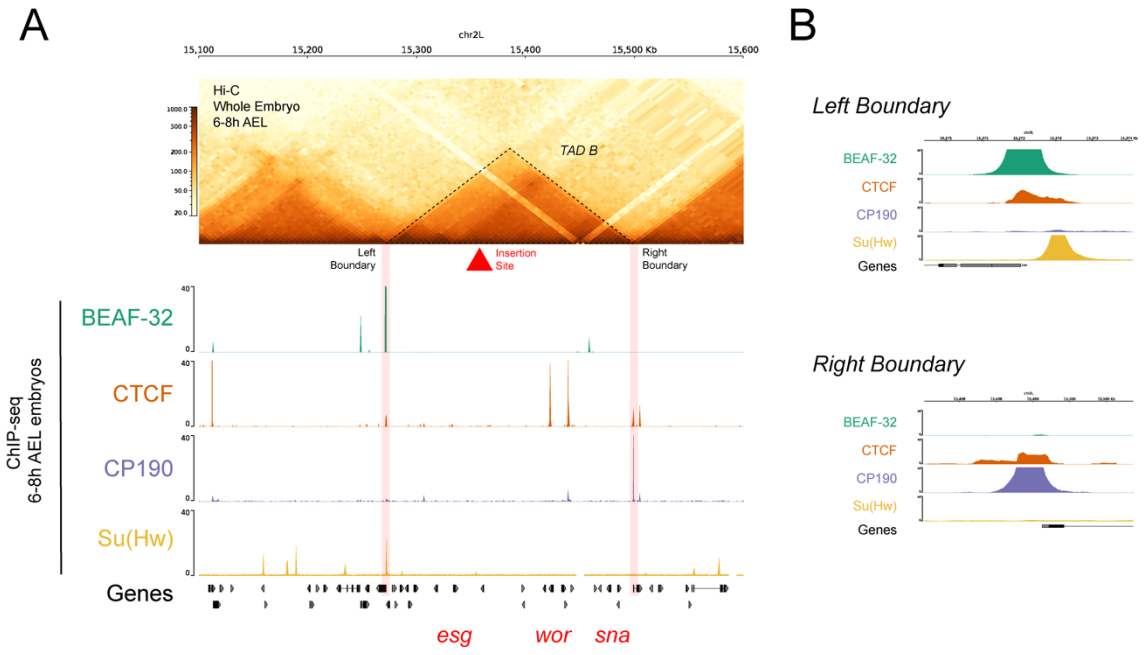

Figure 43: **The landing site at TAD B.**
(A) Hi-C from 6-8h AEL embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins at the same stage. The dashed triangle indicates TAD B. The acceptor site location is indicated by an red arrow. The TAD boundaries are highlighted in red.
(B) Snapshot of ChIP-seq tracks for insulator proteins from 6-8h AEL embryos, spanning TAD B's boundaries.

6.3 Efficiency and macro-phenotypes of boundary insertions

In order to obtain TAD boundary insertions, I tried both the crossing-based strategy and injection-based strategy, as described in sections 6.1.2 and 6.1.3.

I initially used the crossing-based strategy, described in the section 6.1.3. Flies carrying the donor cassettes integrated into random locations were generated by injection of the plasmids containing the donor cassette and P-element sequences. The crossing scheme of the crossing-based strategy is shown in (Fig. 11). Flies carrying the donor cassettes randomly integrated either in the 2nd or 3rd chromosomes were crossed to flies carrying the acceptor cassettes in the desired location (TAD A or TAD B), the heat-shock inducible *Flp* recombinase and the *phiC31* integrase. The progeny were then heat-shocked at the larval stage to stimulate the excision of the donor cassette. The loss of red eyes in adults indicated a high efficiency of donor cassette excision, due to loss of the *mini-white* marker associated with the donor cassette. Nearly all adult flies had white eyes, with only a few patches of red pigment. Unfortunately, the subsequent phiC31-mediated exchange between the acceptor and donor cassettes was very inefficient. We detected very few RCME events in a first trial experiment (see 3.1.4.1), and therefore decided to focus on the second strategy, based on the injection of plasmids containing the donor cassette and direct RCME.

The injections of the injection-based strategy were performed by Dr. Alessandra Reversi at the EMBL *Drosophila* Injection Facility. The crossing scheme of the injection-based strategy is shown in (Fig. 12). This strategy proved to be more efficient, as for most donor plasmids we obtained RCME events after one or two rounds of injection. However, since the orientation of insertion is random, in some cases we could not obtain the same insertion in both orientations so far. The genotyping strategy to confirm the location, identity and orientation of insertions is described in the sections 3.3.2 and 3.3.3.

6.3.1 Accomplished insertions

In total, we obtained 26 insertions into TAD A to date, out of the possible 32 (16 boundaries x 2 orientations). These comprise 15 different boundaries, from which 11 were achieved in two orientations, and 4 in a single orientation. We are currently working to obtain the other orientations of these 4 boundaries. All accomplished insertions are listed in Table 7.

In TAD B, we obtained 22 insertions so far. These correspond to 12 boundaries, from which 9 were on two orientations, and 4 in a single orientation.

The injections to obtain the missing boundaries and orientations in each TAD are being performed. All accomplished insertions are listed in Table 7.

Table 7: Accomplished insertions into TAD A and TAD B.

Boundary	Orientation	Obtained in TAD A?	Obtained in TAD B?
3.1	OR1	Yes	Yes
3.1	OR2		Yes
5.2	OR1	Yes	Yes
5.2	OR2		
5.3	OR1	Yes	
5.3	OR2	Yes	
6.1	OR1		Yes
6.1	OR2		Yes
6.2	OR1	Yes	
6.2	OR2	Yes	
9.1	OR1	Yes	Yes
9.1	OR2	Yes	Yes
11.1	OR1	Yes	Yes
11.1	OR2	Yes	Yes
11.2	OR1	Yes	Yes
11.2	OR2	Yes	Yes
11.3	OR1	Yes	Yes
11.3	OR2	Yes	Yes
12.1	OR1	Yes	Yes
12.1	OR2	Yes	Yes
14.1	OR1	Yes	Yes
14.1	OR2	Yes	Yes
14.2	OR1		
14.2	OR2	Yes	Yes
15.1	OR1	Yes	
15.1	OR2	Yes	Yes
15.2	OR1		
15.2	OR2	Yes	Yes
16.2	OR1	Yes	Yes
16.2	OR2	Yes	Yes
17.1	OR1	Yes	
17.1	OR2	Yes	

6.3.2 Macro-phenotypes of boundary insertions

Both TAD A and TAD B contain genes expressed during embryogenesis that are required for embryonic viability. We hypothesized that insertions affecting gene expression could compromise viability when in homozygosity. In order to characterize the phenotypic

consequences of boundary insertions, we tested if the *Drosophila* lines carrying TAD boundary insertions in TAD A or TAD B could be maintained in a homozygous state.

6.3.2.1 **TAD A – all insertions are homozygous viable**

All the 26 fly lines with boundary insertions into TAD A were fertile and could be maintained in a homozygous state (Table 8). Therefore, these boundary insertions do not recapitulate the lethality observed with double-knockout of *scyl* and *chrb*, but they may be consistent with the homozygous viability observed in single knockouts. Although the insertions are not embryonically lethal, it is possible that expression of either *scyl* or *chrb* is disrupted in specific tissues or stages. Moreover, not all genes in this locus are expressed during embryonic development, meaning we cannot assess the effect of their disruption.

6.3.2.2 **TAD B – half of the insertions are homozygous lethal**

From the 22 fly lines carrying boundary insertions, 11 were homozygous lethal (Table 8). Three boundary insertions caused lethality in an orientation-dependent manner (i.e. only lethal when in a specific orientation), while 3 other boundaries were lethal in an orientation-independent manner. In addition, 2 boundaries were lethal in one orientation, but the other orientation has not been generated yet. This suggests that the expression of one or more of the three essential genes in TAD B (*esg*, *wor* and *sna*) is disrupted in those 11 lines.

Table 8: Viability of TAD boundary insertions into TAD A and TAD B.

Boundary	Orientation	Homozygous Viable in TAD A?	Homozygous Viable in TAD B?
3.1	OR1	Viable	Lethal
3.1	OR2		Viable
5.2	OR1	Viable	Viable
5.2	OR2		
5.3	OR1	Viable	
5.3	OR2	Viable	
6.1	OR1		Lethal
6.1	OR2		Lethal
6.2	OR1	Viable	
6.2	OR2	Viable	
9.1	OR1	Viable	Viable
9.1	OR2	Viable	Viable
11.1	OR1	Viable	Lethal
11.1	OR2	Viable	Viable

11.2	OR1	Viable	Viable
11.2	OR2	Viable	Viable
11.3	OR1	Viable	Viable
11.3	OR2	Viable	Viable
12.1	OR1	Viable	Lethal
12.1	OR2	Viable	Lethal
14.1	OR1	Viable	Viable
14.1	OR2	Viable	Lethal
14.2	OR1		
14.2	OR2	Viable	Lethal
15.1	OR1	Viable	
15.1	OR2	Viable	Lethal
15.2	OR1		
15.2	OR2	Viable	Viable
16.2	OR1	Viable	Lethal
16.2	OR2	Viable	Lethal
17.1	OR1	Viable	
17.1	OR2	Viable	

6.4 Assessing topological consequences of boundary insertions

To investigate if inserting TAD boundaries leads to changes in local chromatin topology, we performed DNA FISH for TAD A or TAD B loci in embryos carrying TAD boundary insertions. In order to control for tissue and stage variation in chromatin topology, we acquired all our pictures in 6-8h AEL embryos. Moreover, we had selected the TAD boundaries for insertions and landing sites based on Hi-C and ChIP-seq data at that stage. In order to be consistent, we acquired images from the same region in all embryos: the outermost cells in the central dorsal region. We chose this region because of accessibility and visibility (cells at the embryo surface suffer less from refraction effects than in the interior of the embryo).

6.4.1 Boundary insertions in TAD A can disrupt chromatin topology in an orientation-dependent and -independent manner

We used DNA FISH probes targeting the anchors of the ~235kb loop between the *scyl* and *chrb* genes. This allowed us to directly measure the effect of the insertions in the proximity between the loop anchors (Fig. 44A).

So far, we measured by DNA FISH the effect of 24 boundary insertions (out of 26 obtained lines) in TAD A. In total, 10 insertions led to a detectable and significant increase in

the 3D distance between FISH probes (Fig. 44B). From all 10 insertions with a detectable effect, 9 displayed a substantial increase (median distance in WT - 248 nm; median distance in the 9 lines - 354nm), while 1 insertion had a milder, but still detectable effect (insertion 5.3 OR1 – median distance of 285nm) (Fig. 44B).

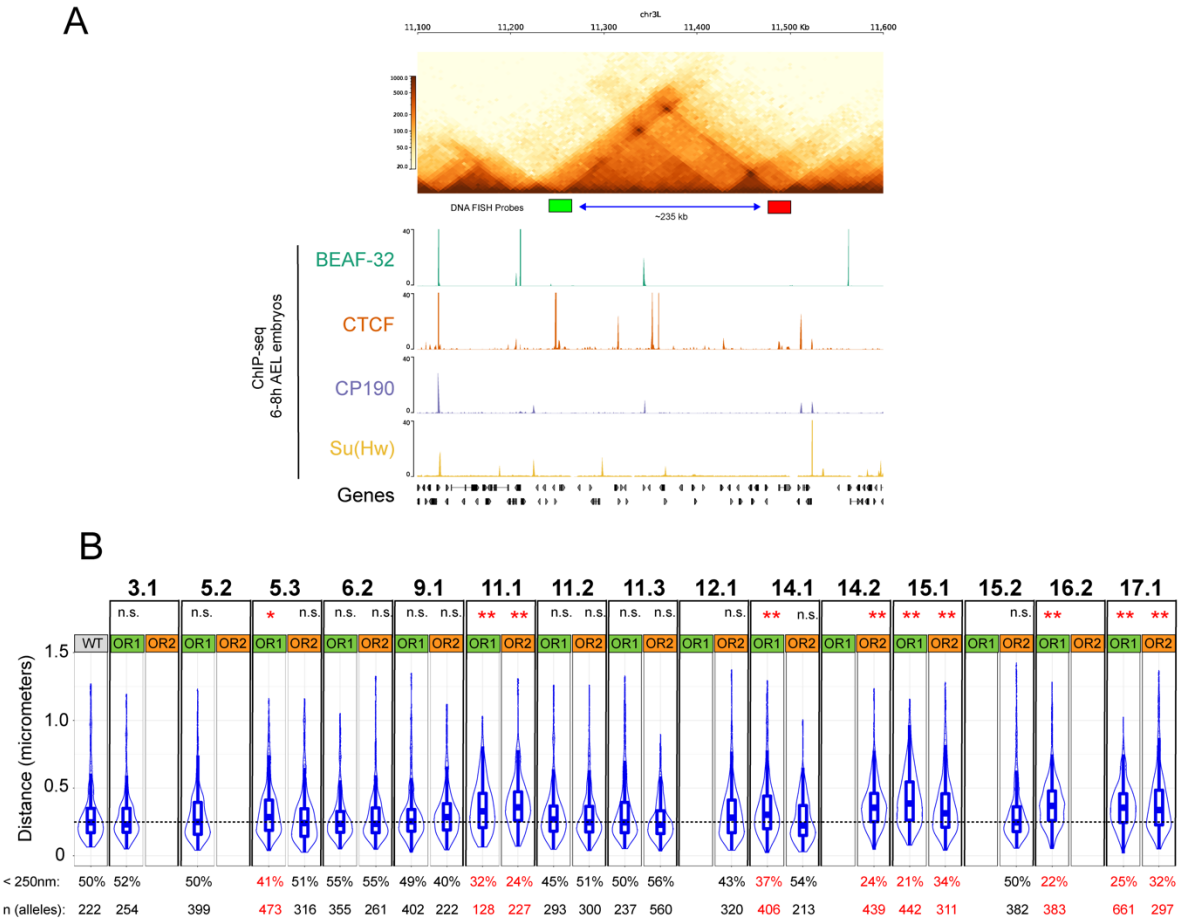

Figure 44: **DNA FISH to evaluate the effect of different TAD boundary insertions into the chromatin topology of TAD A.**
(A) Hi-C from 6-8h AEL WT embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins at the same stage. The position of the DNA FISH probes is shown below the Hi-C matrix.
(B) 3D distance distributions between DNA FISH probes shown in (A) across different lines with boundary insertions in two different orientations. Insertions for which we didn't obtain the other orientation are blank. Distributions significantly different from WT are highlighted in red.
n = number of alleles measured in at least 3 embryos. A Kolmogorov-Smirnov test was applied to evaluate statistical significance: n.s. = non-significant, *p < 0.01, **p < 0.001

The DNA FISH results in TAD A are summarized in (Fig. 45). The 10 disruptive insertions can be divided in 3 categories (Fig. 45A):

1) Two boundaries displayed an orientation-dependent disruption in loop-anchor proximity (i.e. one orientation had an effect while the other did not).

2) Three boundaries displayed an orientation-independent effect (i.e. both orientations disrupted loop-anchor proximity).

3) Two boundaries disrupted proximity in one orientation, but have not been tested in the other orientation yet.

Four boundaries did not disrupt proximity in either orientation (Fig. 45B). And finally, 4 boundaries did not disrupt proximity in one orientation, but have not been tested in the other orientation (Fig. 45B).

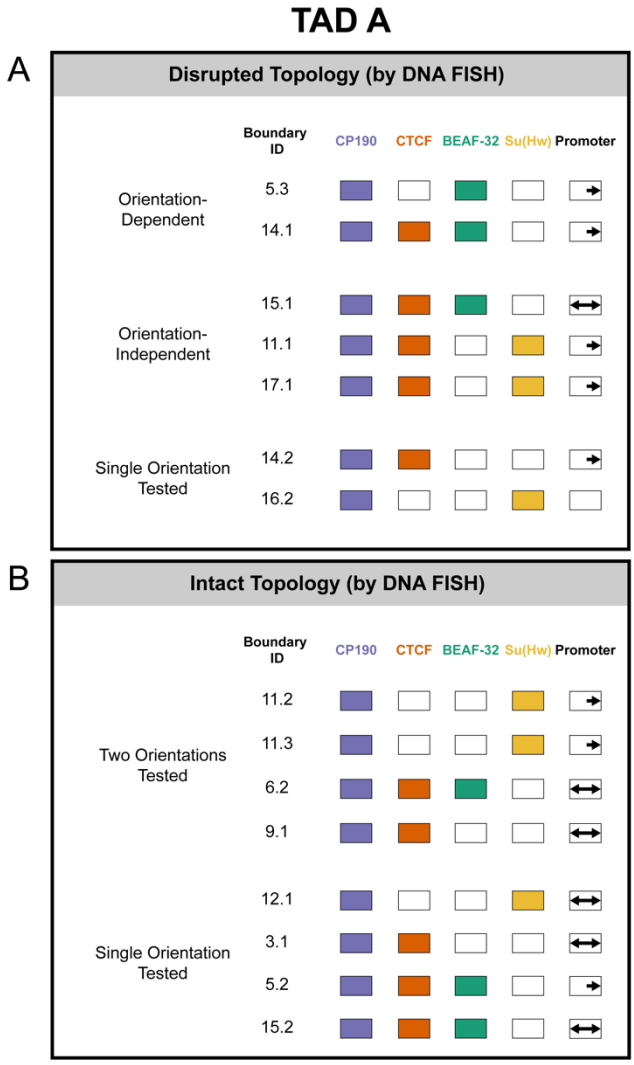

Figure 45: **Effects of TAD boundary insertions in the chromatin topology of TAD A.**

6.4.2 Multiple boundary insertions in TAD B disrupt chromatin topology

In TAD B we used DNA FISH probes targeting the TAD boundaries (Fig. 46A). Since the TAD B locus did not display focal high-frequency looping interactions as observed in TAD A in Hi-C maps from the 6-8h AEL stage (Fig. 42), we decided to use probes that cover a larger region of the locus (~80 kb each), and are centered at the boundaries (Fig. 46A). This allowed us to have a stronger fluorescence signal while still being able to evaluate 3D proximity between the boundary regions. Of note, for all embryos carrying insertions into TAD B and the WT control, we performed the DNA FISH in combination with RNA FISH to detected GFP expression (see 3.6.3). This was needed because some lines were homozygous lethal, and the allele with the insertion needed to be kept over a GFP-marked balancer chromosome. The combined DNA/RNA FISH experiment allowed us to distinguish which embryos were homozygous for the insertion (GFP-negative).

The 3D distances between DNA FISH probes in the TAD B locus were on average higher than in TAD A (WT median distances: 330 nm in TAD B versus 248 nm in TAD A). This occurred despite the TAD B probes targeting sites in a closer linear genomic distance (205kb versus 235kb between the center of the probes in TAD B and TAD A respectively). These larger 3D distances match differences in contact frequency from Hi-C experiments: while TAD A displays a high-frequency looping interaction, this is not observed between the boundaries of TAD B.

Next, we measured the effect of 15 boundary insertions on the proximity between the boundaries of TAD B. In total, 8 insertions out of 15 led to a detectable, significant increase in distance between DNA FISH probes (Fig. 46B). From the 8 disruptive insertions, 7 displayed a substantial increase in probe distances (median distances: 330 nm in WT versus 402 nm for the 7 insertions), while 1 insertion had a weaker effect (Boundary 11.3 OR2 – median distance of 369 nm) (Fig. 46B).

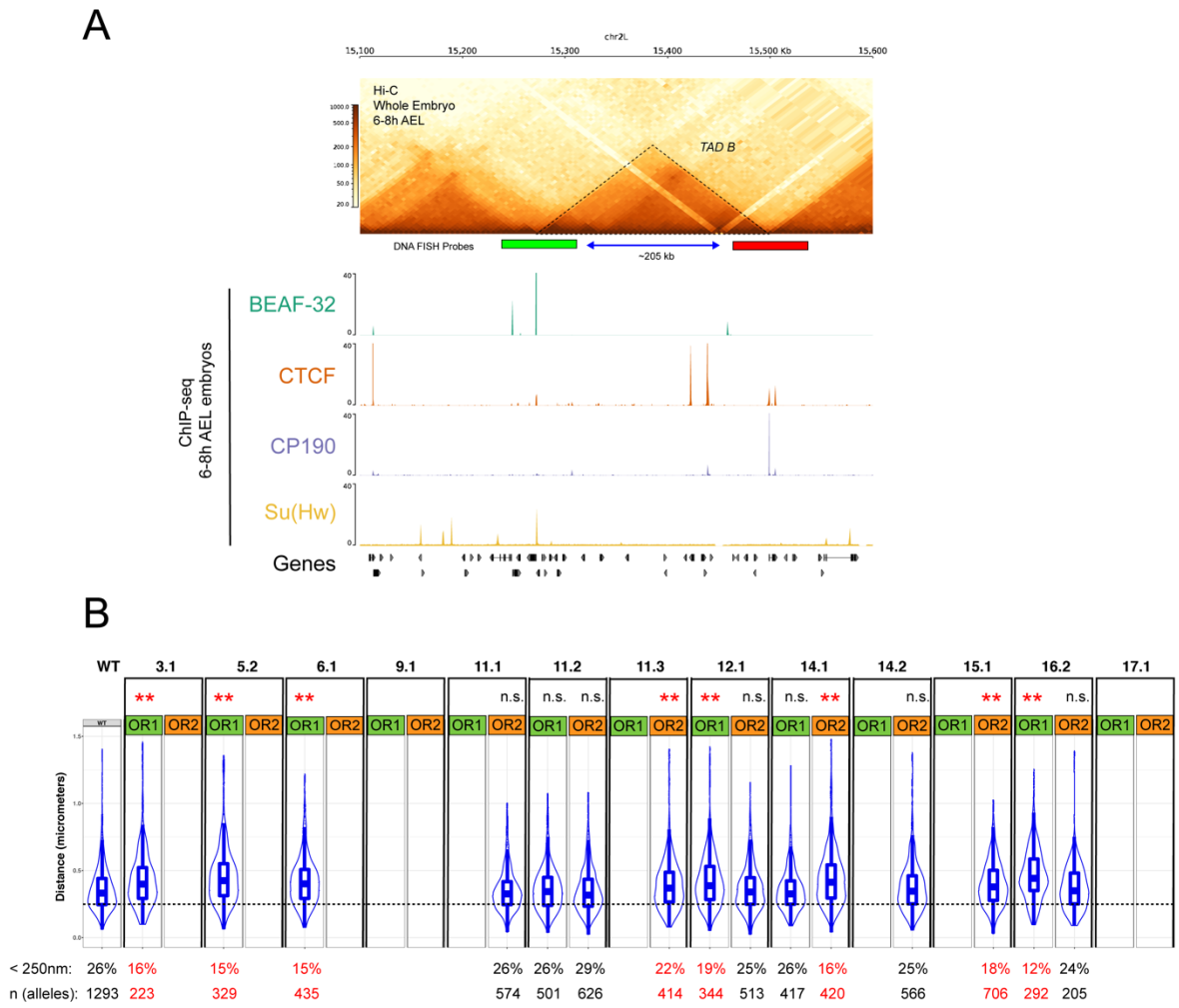

Figure 46: **DNA FISH to evaluate the effect of different TAD boundary insertions into the chromatin topology of TAD B.**

(A) Hi-C from 6-8h AEL WT embryos previously generated in the lab (unpublished), along ChIP-seq tracks of insulator proteins at the same stage. The position of the DNA FISH probes is shown below the Hi-C matrix.

(B) 3D distance distributions between DNA FISH probes shown in (A) across different lines with boundary insertions in two different orientations. Insertions for which we didn't obtain the other orientation are blank. Distributions significantly different from WT are highlighted in red.

n = number of alleles measured in at least 3 embryos. A Kolmogorov-Smirnov test was applied to evaluate statistical significance: n.s. = non-significant, *p < 0.01, **p < 0.001

The DNA FISH results in TAD B are summarized in (Fig. 47). The 8 disruptive insertions can be divided in 2 categories (Fig. 47A):

1) Four inserted boundaries led to an orientation-dependent disruption of TAD B boundary proximity (i.e. one orientation had an effect while the other did not)

2) Another 4 inserted boundaries led to disruption of TAD B boundary proximity when placed in one orientation (but have not been tested in the other orientation yet)

Three insertions did not disrupt topology in one orientation, but have not been tested in the other orientation) (Fig. 47 B).

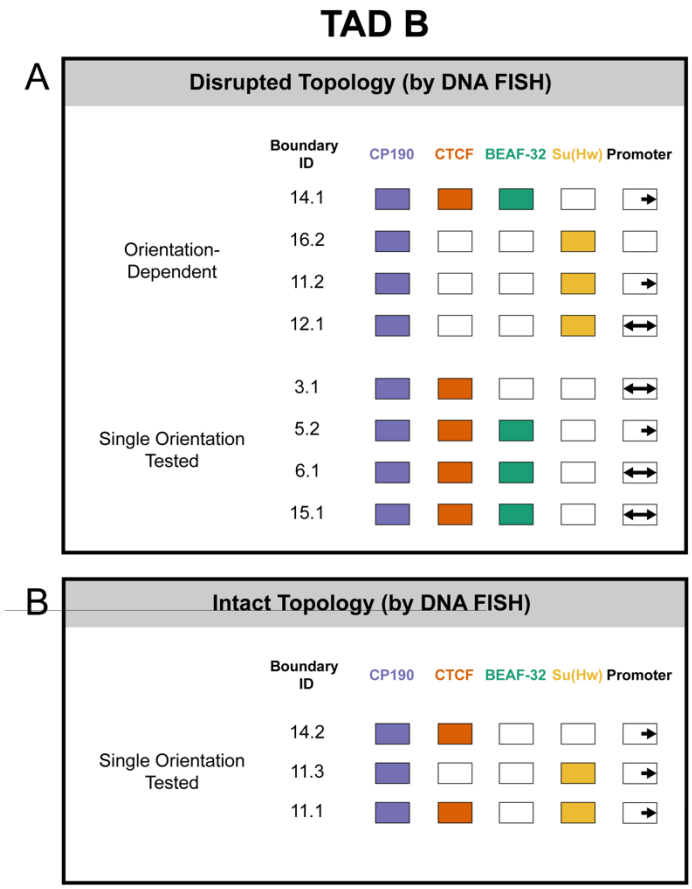

Figure 47: **Effects of TAD boundary insertions in the chromatin topology of TAD B.**

Our DNA FISH measurements indicate that boundaries can often perturb chromatin topology when placed in an ectopic genomic location. In TAD A and TAD B, approximately half of the insertions affected the proximity between DNA FISH probes targeting loop anchors (TAD A) or TAD boundaries (TAD B). Furthermore, topology disruption is orientation-dependent for some inserted boundaries. The orientation dependency of topological disruption varies between boundaries, suggesting a high-context specificity and potential different mechanisms by which boundaries can work.

7 Discussion and Perspectives II

7.1 What topology defects are caused by TAD boundary insertions?

What are the sequence dependencies of TAD boundaries and their local genomic context, and what combination of factors are required for their function, are two open questions. We addressed these by inserting multiple TAD boundaries into two 'acceptor' TADs which contain distinct features. By employing a DNA FISH strategy, we measured the effects of those multiple boundary insertions into "TAD A" and "TAD B". In both TADs, roughly half of the insertions caused an effect detectable by DNA FISH, suggesting that boundary function is highly context-specific. How can we interpret the DNA FISH measurements in terms of topological changes?

In TAD A, we used probes targeting the anchors of a high-frequency looping interaction (Fig. 44). This allowed us to measure the effect of the insertions directly on that looping interaction, which is located within a larger TAD. However, our measurements do not allow the distinction between disruption of the loop or the whole TAD structure (Fig. 36). The ectopic boundary may prevent loop formation while not splitting the TAD in two new domains, and vice-versa. Importantly, the 3D distances in the affected embryos are compatible with the scenario of a weakened loop, but maintenance of TAD structure. The 3D distance between FISH probes in the mutant embryos is still smaller than the distances in WT embryos at an earlier stage (NC14), when the loop is weak, but the TAD is present (median distances: 354nm in 6-8h AEL affected mutants versus 456nm in NC14 WT embryos).

In TAD B, we used large probes (~80 Kb), centered at the TAD boundary. This provided a strong DNA FISH signal that helped us to overcome difficulties of cloning smaller probes by PCR. Importantly, those larger probes also allowed us to evaluate TAD structure as they capture a large region instead of a single interaction between TAD boundaries. To confirm that this strategy is not affecting the sensitivity in detecting topological changes, we recently obtained small probes targeting the boundary and will perform DNA FISH measurements with those in a few lines.

To further evaluate the topological effects of boundary insertions, we are currently performing experiments that will allow us to probe interactions spanning the whole TAD A and TAD B loci. Those are detailed in section 7.4.

7.2 Lethality, gene expression defects and topological changes caused by TAD boundary insertions

Both loci where we inserted TAD boundaries contain genes required for embryonic viability. The insertions affected embryo viability in very different ways, which we discuss below.

In the TAD A, *scyl* and *chrb* are the only genes expressed during embryogenesis, and their double knockout impairs embryonic development, while their single knockouts are viable (Scuderi et al., 2006). While multiple insertions into TAD A altered the topology of that locus (Fig. 44,45), none of the insertions caused a lethal phenotype (Table 8). That does not rule out that the insertions caused gene expression changes, but if so, those are clearly not equivalent to a combined loss of both genes. Gene expression changes could still have occurred in a number of different scenarios:

1) Considering that single knockout of *scyl* or *chrb* does not lead to developmental phenotypes (Scuderi et al., 2006), the expression of either gene may have been completely lost;
2) Both genes could have been transcriptionally affected, but their levels remain enough to sustain their function;
3) Both genes might be transcriptionally affected in different stages or tissues in the embryo.

Both *scyl* and *chrb* have a very complex expression pattern in the developing embryo (Scuderi et al., 2006), which impeded DNA FISH measurements comparing cells that express versus those that do not express those genes.

In TAD B, *esg*, *wor* and *sna* are expressed during embryogenesis, and knockout of each gene impairs embryonic development (Ashraf et al., 2004; Buszczak et al., 2007; Simpson, 1983). Multiple insertions into TAD B were homozygously lethal. Importantly, not all lethal insertions caused a topological disruption detectable by DNA FISH. This apparent discrepancy can be explained by different possibilities:

1) We performed our DNA FISH measurements at a single stage and in a single population of cells in the embryo. The insertions for which we didn't observe an effect might affect topology in other tissues or stages in the developing fly;
2) Topological changes may occur only within TADs, not affecting TAD boundaries, what in principle would not be revealed by our DNA FISH strategy;
3) Insertions may not perturb the topology of the locus, and rather disturb transcription by other mechanisms (e.g. through recruitment of ectopic transcriptional regulators).

Although we have not performed gene expression analyses to date, we are currently working in collaboration with the lab of Dr. Alistair Boettiger to achieve that. Those analyses will help us to understand how chromatin topology disruptions affect local gene expression, and are detailed in section 4.

7.3 Deciphering the rules of boundary function in the Drosophila genome

Inserting TAD boundaries in two loci revealed that boundary function is highly context-specific. Boundary insertions can disrupt topology in an orientation-dependent or -independent manner, while some insertions do not cause an effect. Moreover, some insertions disrupt topology only in TAD A or TAD B, while others functioned in both.

What rules define if a boundary remains functional in an ectopic context? We compared boundaries that disrupted topology versus boundaries that did not in both TADs (Figs. 45,47), to start defining those rules.

7.3.1 TAD A

1) No single DNA-binding insulator proteins is present in all boundaries that disrupted topology. Therefore, neither BEAF-32, CTCF or Su(Hw) are essential to establish a functional boundary in that locus. Even though CTCF binds to both loop anchors on that locus, its binding between the anchors is not sufficient to disrupt loop anchor proximity;
2) Promoter activity is not necessary to disrupt topology, as one of the boundaries that functioned does not contain a promoter.

Interestingly, boundaries that did not disrupt topology were very similar to some boundaries that did. For example, insertion of boundary 6.2 (bound by CP190, CTCF, BEAF-32, and overlapping a divergent promoter) did not disrupt topology. Meanwhile, boundary 15.1 is bound by the same factors and overlaps a divergent promoter, but disrupted topology in an orientation-dependent manner. One explanation would be that sequence motif features (i.e. number, arrangement, strength, spacing, etc) may determine whether the boundaries with the same factors work or not. Alternatively, other factors (e.g. binding of additional insulator proteins, promoter strength, etc) may play a role.

7.3.2 TAD B

1) No single DNA-binding insulator protein is present in all boundaries that disrupted topology. Therefore, neither BEAF-32, CTCF or Su(Hw) are essential to establish a functional boundary in that locus;
2) Promoter activity is not necessary to form a boundary, as one of the boundaries that functioned does not contain a promoter.

Again, some of the boundaries that did not disrupt topology contained binding sites for the same factors of some of the boundaries that did. For example, insertion of the boundary 11.3 (containing CP190, Su(Hw) and a single promoter) did not disrupt topology. Meanwhile, boundary 15.1 contains binding sites for the same factors, and could function as a boundary in an orientation-dependent manner. This indicates that other factors are binding to these boundary elements, which we have yet to resolve.

7.3.3 TAD A and TAD B (Cross-TAD comparison)

Three boundaries were able to disrupt topology when inserted into either TAD A or TAD B (boundaries 14.1, 15.1 and 16.2). Interestingly, boundary 14.1 functioned in an orientation-dependent manner in either TAD. This boundary is bound by the factors CTCF, BEAF-32 and CP190, and overlaps with a single promoter. All three factors are present in either TAD A or TAD B, indicating a possible communication between the boundary with local insulators depending on their binding orientation.

The other two boundaries that worked in both TADs (15.1 and 16.2) haven't been tested in both orientations in the two TADs so far. Nevertheless, inserting the boundary 15.1 in TAD

A caused an orientation-independent disruption of topology, as opposed to the boundary 14.1 mentioned above. Since both boundaries are bound by the same factors, this may indicate a different mode of function.

We are in the process of obtaining all the desired insertions in the two orientations, to be able to better compare the effect of orientations in genome topology. Obtaining the missing orientations will allow us to achieve more meaningful comparisons across TADs. In parallel, we are currently investigating more features that could diversify the inserted boundaries, and may provide rules that predict their functioning in an ectopic context. This is described below.

7.4 Ongoing work to better understand mechanisms of boundary function

Even though DNA FISH enabled us to restrict our analyses to a homogeneous population of cells in a specific developmental stage, it only allowed the detection of the proximity between two genomic regions. For example, while some boundary insertions in TAD A led to detectable topological effects on a looping interaction, it is not clear whether the topology of the whole TAD is affected. Similarly, some boundary insertions led to alterations in the proximity between TAD B boundaries, but whether there are heterogeneous topological re-configurations is unclear. Therefore, we sought to have a more detailed view of how TAD A and TAD B were topologically affected by boundary insertions, e.g. as provided by Hi-C experiments. High-resolution Hi-C requires high-depth sequencing, and involves sequencing ligation events throughout the genome, while our boundary insertions target individual loci. Since we were interested in evaluating multiple insertions, this would dramatically increase sequencing costs, making such experiments unfeasible. Therefore, we decided to try a recently developed protocol, capture-Hi-C (Davies et al., 2016; Despang et al., 2019; Mifsud et al., 2015), which allows high-resolution assessment of contact frequencies in single loci, significantly reducing the sequencing costs while increasing resolution. This protocol involved performing a regular Hi-C experiment up to the generation of a Hi-C library, but with only ~3 PCR cycles, prior to sequencing library preparation. This “pre-Hi-C library” is then hybridized to a pool of oligonucleotide probes that span a whole TAD in addition to the two neighbouring TADs. This allows the capture of the genome region of interest in a Hi-C library, and at the same time allows the multiplexing of several samples, dramatically reducing the sequencing costs. We are currently performing the capture-Hi-C experiment on samples with boundary insertions in TAD A, and expect it to provide an assessment of topological changes orthogonal

to the DNA FISH experiments. Furthermore, this will allow us to distinguish between effects on TAD structure versus disruption of more focal looping interactions.

Changes in transcription and topology associated with the boundary insertions will be profiled as part of a recent collaboration with the laboratory of Dr. Alistair Boettiger, by using an imaging-based method (Mateo et al., 2019), which will allow spatial and quantitative changes in the mRNA distribution for the genes at the TAD A and TAD B loci.

Furthermore, we are starting a collaboration with the laboratory of Dr. Oliver Stegle to further explore the sequence features of TAD boundaries (both the inserted boundaries and the boundaries at the landing sites), such as the number of DNA sequence motifs for different insulator proteins, as well as the orientation of these motifs. With these analyses, we intend to reveal a combinatorial logic that would explain why some boundaries work in certain contexts but not in others.

7.5 Conclusions

Our results revealed a striking context-specificity of chromatin organization in the *Drosophila* genome. TAD boundaries can disrupt topology when transplanted into an ectopic location, but predicting the outcome of such insertions is not straightforward. For example, inserting boundaries bound by similar factors differentially impacts the topology/gene expression of a locus, indicating the presence of additional layers conferring specificity to boundary function. Moreover, TAD boundary insertions likely have a local effect on gene expression, as multiple insertions caused lethality when placed in a TAD containing essential developmental regulators. While finishing to obtain all desired insertions, we envisage exciting findings combining our current results with new readouts of chromatin topology and transcription.

8 Appendix

Supplementary Table 1: List of Drosophila lines used in this thesis.

Line Name	Full Genotype	Source	Origin ID	Description
WT	yw;;	BDSC	#6598	
WT	w^{1118};;	BDSC	#3605	
Mat-tub-gal4	w;; P{matα4-GAL-VP16}V37	BDSC	#7063	Carries one Gal4 transgene, allowing its expression during late female germline development
MTD-Gal4	P{w[+mC]=otu-GAL4::VP16.R}1, w[*]; P{w[+mC]=GAL4-nos.NGT}40; P{w[+mC]=GAL4::VP16-nos.UTR}CG6325[MVD1]	BDSC	#31777	Carries three Gal4 transgenes, allowing its expression throughout female germline development
CP190 shrRNA	y[1] sc[*] v[1] sev[21]; P{y[+t7.7] v[+t1.8]=TRiP.HMS00845}attP2	BDSC	#33903	Carries a shRNA against CP190
CP190 shrRNA	y[1] sc[*] v[1] sev[21]; P{y[+t7.7] v[+t1.8]=TRiP.HMS00895}attP2/TM3, Sb[1]	BDSC	#33944	Carries a shRNA against CP190
CP190 shrRNA	w;P{w+ VSH35077}	VDRC	#35077	Carries a shRNA against CP190
BEAF-32 shRNA	w;P{w+ VSH330274}/CyO	VDRC	#330274	Carries a shRNA against BEAF-32
BEAF-32 shRNA	y[1] sc[*] v[1] sev[21]; P{y[+t7.7] v[+t1.8]=TRiP.HM05202}attP2	BDSC	#29734	Carries a shRNA against BEAF-32
BEAF-32 shRNA	y[1] sc[*] v[1] sev[21]; P{y[+t7.7] v[+t1.8]=TRiP.GLV21006}attP2	BDSC	#35642	Carries a shRNA against BEAF-32
CTCF KO, UAS-Flp	y w; +; {UAS-FLP(w+)}VK33, CTCF[KO B](DsRed) / TM6B	Furlong lab	LT#88	To generate CTCF maternal/zygotic null embryos
CTCF KO, NGVP16	(y) w; {>CTCF(w+)}VK37 / (CyO); CTCF[KO A](DsRed), NGVP16(w+) / (TM6B)	Furlong lab	LT#89	To generate CTCF maternal/zygotic null embryos
CP190 FRT82B	FRT82B Cp190/TM3	Jordan Raff's lab		For CP190 germline clones
Zld FRT19A	zld^{294}FRT19A;;	Christine Rushlow's lab		For Zld germline clones
OvoD FRT82B	hs FLP; FRT82B ovoD1/TM3	Jordan Raff's lab		For CP190 germline clones
OvoD FRT19A	y w sn P{mini w+, ovoD1} FRT19A, hsFLP122;;	Christine Rushlow's lab		For Zld germline clones
Vasa-Cas9	w[1118];PBac{y[+mDint2]=vas-Cas9}	BDSC	#51324	To generate CRISPR deletions
Btsz_Promoter_Del	yw;Btsz-PromoterDel/ CyO GFP	This thesis		Carries a deletion at the btsz promoter
Btsz_Insulator_Del	yw;Btsz-InsulatorDel	This thesis		Carries a deletion at the btsz-insulator binding site
Btsz_Boundary_Del	yw;Btsz-BoundaryDel/ CyO GFP	This thesis		Carries a deletion at the btsz-boundary
MiMIC acceptor site (TAD A)	y[1] w[*]; Mi{y[+mDint2]=MIC}MI02296	BDSC	#33194	Carries an acceptor site at TAD A
MiMIC acceptor site (TAD B)	y[1] w[*]; Mi{y[+mDint2]=MIC}MI14731	BDSC	#60957	Carries an acceptor site at TAD B
Integrase B, Flp	P{ry[+t7.2]=hsFLP}12, y[1] w[*] M{vas-int.B}ZH-2A; Pri[1]/TM3, Sb[1]	BDSC	#33216	Carries integrase B and Flp transgene; those are combined with MiMIC acceptor sites to generate final acceptor lines
TAD A - 3.1 OR1	yw;;TAD A {Boundary 3.1 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 5.2 OR1	yw;;TAD A {Boundary 5.2 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 5.3 OR1	yw;;TAD A {Boundary 5.3 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 5.3 OR2	yw;;TAD A {Boundary 5.3 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 6.2 OR1	yw;;TAD A {Boundary 6.2 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 6.2 OR2	yw;;TAD A {Boundary 6.2 OR2-DsRed}	This thesis		Boundary insertion into TAD A

TAD A - 9.1 OR1	yw;;TAD A {Boundary 9.1 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 9.1 OR2	yw;;TAD A {Boundary 9.1 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 11.1 OR1	yw;;TAD A {Boundary 11.1 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 11.1 OR2	yw;;TAD A {Boundary 11.1 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 11.2 OR1	yw;;TAD A {Boundary 11.2 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 11.2 OR2	yw;;TAD A {Boundary 11.2 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 11.3 OR1	yw;;TAD A {Boundary 11.3 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 11.3 OR2	yw;;TAD A {Boundary 11.3 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 12.1 OR1	yw;;TAD A {Boundary 12.1 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 12.1 OR2	yw;;TAD A {Boundary 12.1 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 14.1 OR1	yw;;TAD A {Boundary 14.1 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 14.1 OR2	yw;;TAD A {Boundary 14.1 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 14.2 OR2	yw;;TAD A {Boundary 14.2 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 15.1 OR1	yw;;TAD A {Boundary 15.1 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 15.1 OR2	yw;;TAD A {Boundary 15.1 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 15.2 OR2	yw;;TAD A {Boundary 15.2 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 16.2 OR1	yw;;TAD A {Boundary 16.2 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 16.2 OR2	yw;;TAD A {Boundary 16.2 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 17.1 OR1	yw;;TAD A {Boundary 17.1 OR1-DsRed}	This thesis		Boundary insertion into TAD A
TAD A - 17.1 OR2	yw;;TAD A {Boundary 17.1 OR2-DsRed}	This thesis		Boundary insertion into TAD A
TAD B - 3.1 OR1	yw;TAD B {Boundary 3.1 OR1-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B
TAD B - 3.1 OR2	yw;TAD B {Boundary 3.1 OR2-DsRed}	This thesis		Boundary insertion into TAD B
TAD B - 5.2 OR1	yw;TAD B {Boundary 5.2 OR1-DsRed}	This thesis		Boundary insertion into TAD B
TAD B - 6.1 OR1	yw;TAD B {Boundary 6.1 OR1-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B
TAD B - 6.1 OR2	yw;TAD B {Boundary 6.1 OR2-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B
TAD B - 9.1 OR1	yw;TAD B {Boundary 9.1 OR1-DsRed}	This thesis		Boundary insertion into TAD B
TAD B - 9.1 OR2	yw;TAD B {Boundary 9.1 OR2-DsRed}	This thesis		Boundary insertion into TAD B
TAD B - 11.1 OR1	yw;TAD B {Boundary 11.1 OR1-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B
TAD B - 11.1 OR2	yw;TAD B {Boundary 11.1 OR2-DsRed}	This thesis		Boundary insertion into TAD B
TAD B - 11.2 OR1	yw;TAD B {Boundary 11.2 OR1-DsRed}	This thesis		Boundary insertion into TAD B
TAD B - 11.2 OR2	yw;TAD B {Boundary 11.2 OR2-DsRed}	This thesis		Boundary insertion into TAD B
TAD B - 11.3 OR1	yw;TAD B {Boundary 11.3 OR1-DsRed}	This thesis		Boundary insertion into TAD B
TAD B - 11.3 OR2	yw;TAD B {Boundary 11.3 OR2-DsRed}	This thesis		Boundary insertion into TAD B
TAD B - 12.1 OR1	yw;TAD B {Boundary 12.1 OR1-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B
TAD B - 12.1 OR2	yw;TAD B {Boundary 12.1 OR2-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B
TAD B - 14.1 OR1	yw;TAD B {Boundary 14.1 OR1-DsRed}	This thesis		Boundary insertion into TAD B

TAD B - 14.1 OR2	yw;TAD B {Boundary 14.1 OR2-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B
TAD B - 14.2 OR2	yw;TAD B {Boundary 14.2 OR2-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B
TAD B - 15.1 OR2	yw;TAD B {Boundary 15.1 OR2-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B
TAD B - 15.2 OR2	yw;TAD B {Boundary 15.2 OR2-DsRed}	This thesis		Boundary insertion into TAD B
TAD B - 16.2 OR1	yw;TAD B {Boundary 16.2 OR1-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B
TAD B - 16.2 OR2	yw;TAD B {Boundary 16.2 OR2-DsRed}/CyO, GFP	This thesis		Boundary insertion into TAD B

Supplementary Table 2: Primers used in this thesis

Primer_ID	Sequence	Purpose
scylchrb_DNAFISH_Probe1_Fw	GTCTACGCGATTCGGGACAT	DNA FISH probe cloning
scylchrb_DNAFISH_Probe1_Rv	CTTTGCGATACATGGCCAGC	DNA FISH probe cloning
scylchrb_DNAFISH_Probe2_Fw	GAGAGTTCACGGCCAAGGAA	DNA FISH probe cloning
scylchrb_DNAFISH_Probe2_Rv	CTTCCATCCGGGGGTTTTGA	DNA FISH probe cloning
scylchrb_DNAFISH_Probe3_Fw	CTCAGTGCCTAGTGCGGAAA	DNA FISH probe cloning
scylchrb_DNAFISH_Probe3_Rv	CCTTCGTGGCCAAAAGAACG	DNA FISH probe cloning
tshtio_DNAFISH_Probe1_Fw	GAATGTGACGAGCGATTGCC	DNA FISH probe cloning
tshtio_DNAFISH_Probe1_Rv	AGATTGCAAGGGGCGATTGA	DNA FISH probe cloning
tshtio_DNAFISH_Probe2_Fw	CCAGAGAGAACGCCGACTC	DNA FISH probe cloning
tshtio_DNAFISH_Probe2_Rv	GTCTCGTTTTCGGGTGTCGT	DNA FISH probe cloning
tshtio_DNAFISH_Probe3_Fw	GACCTTCTCGACAACAGGGG	DNA FISH probe cloning
tshtio_DNAFISH_Probe3_Rv	CTGTGGCCACGACCTTTCTT	DNA FISH probe cloning
LBR_DNAFISH_Probe1_Fw	CTTGGCATCGCTTTCAGACC	DNA FISH probe cloning
LBR_DNAFISH_Probe1_Rv	GCGCTTCATTGAACTCGTCC	DNA FISH probe cloning
LBR_DNAFISH_Probe2_Fw	TCATGATGCAGGCAGTCAGC	DNA FISH probe cloning
LBR_DNAFISH_Probe2_Rv	AGCGATCACCACTTACATGCC	DNA FISH probe cloning
LBR_DNAFISH_Probe3_Fw	ACACCCAGGAAACTGTTGTACTC	DNA FISH probe cloning
LBR_DNAFISH_Probe3_Rv	GAGTGGGCACTGGATCGTATC	DNA FISH probe cloning
Promoterdel-5-HA_Fw	ttgaactcgattgacTTTCATCAGTCGGGTCAGCTC	CRISPR cloning (Homology Arms)
Promoterdel-5-HA_Rv	gctatacgaagttatCTGGGTGGAGGAACTGAGTCA	CRISPR cloning (Homology Arms)
Promoterdel-3-HA_Fw	cagttggggcactacTTTCGGTGTTCAGCACAAACTC	CRISPR cloning (Homology Arms)
Promoterdel-3-HA_Rv	ggtgtcgcccttcgcTTCTCGCAGTGTCTCATACAGC	CRISPR cloning (Homology Arms)
Insulatordel-5-HA_Fw	ttgaactcgattgacTCTTAATCCGGGTCTTCATGCTC	CRISPR cloning (Homology Arms)
Insulatordel-5-HA_Rv	gctatacgaagttatCGAAGGATTCCAGAATTTGGTACC	CRISPR cloning (Homology Arms)
Insulatordel-3-HA_Fw	cagttggggcactacTACCCTCTATCCGCATCAGCA	CRISPR cloning (Homology Arms)
Insulatordel-3-HA_Rv	ggtgtcgcccttcgcGTACTCCTCTAATGCACAATCGC	CRISPR cloning (Homology Arms)
gRNA_oligo_promoterdel-5_S	CTTCGTGGAGGAACTGAGTCAGCG	gRNA oligos
gRNA_oligo_promoterdel-5_AS	AAACCGCTGACTCAGTTCCTCCAC	gRNA oligos
gRNA_oligo_promoterdel-3_S	CTTCGCTGGTCTGCGAACTGCTTT	gRNA oligos
gRNA_oligo_promoterdel-3_AS	AAACAAAGCAGTTCGCAGACCAGC	gRNA oligos

gRNA_oligo_insulatordel-5_S	CTTCGCGGTGGACTGTCGCCTCGA	gRNA oligos
gRNA_oligo_isulatordel-5_AS	AAACTCGAGGCGACAGTCCACCGC	gRNA oligos
gRNA_oligo_insulatordel-3_S	CTTCGATGCGGATAGAGGGTAGTT	gRNA oligos
gRNA_oligo_isulatordel-3_AS	AAACAACTACCCTCTATCCGCATC	gRNA oligos
Boundary_3.1_Fw	ccctcgaggaattccTCGAACCGAACTTGGAGTTCC	TAD Boundary cloning
Boundary_3.1_Rv	actctagaactagtcACTTGTGGATGTCGTCTTCGTG	TAD Boundary cloning
Boundary_5.2_Fw	ccctcgaggaattccCCAAGAAGTACGATGTCTGCG	TAD Boundary cloning
Boundary_5.2_Rv	actctagaactagtcTTTATGGTTAGAGGCAAGCGC	TAD Boundary cloning
Boundary_5.3_Fw	ccctcgaggaattccAAGAGGCACAGGTCCATAAGC	TAD Boundary cloning
Boundary_5.3_Rv	actctagaactagtcGCGGCTATTACTTCGCTGTC	TAD Boundary cloning
Boundary_6.1_Fw	ccctcgaggaattccAGTGACGCCCAAGCATAACTG	TAD Boundary cloning
Boundary_6.1_Rv	actctagaactagtcTGTTCATGACCTGCTGCTGAG	TAD Boundary cloning
Boundary_6.2_Fw	ccctcgaggaattccGCGGATTTGCTTGGGTTACAC	TAD Boundary cloning
Boundary_6.2_Rv	actctagaactagtcAGAAGGTACTTGACCGACTGG	TAD Boundary cloning
Boundary_9.1_Fw	ccctcgaggaattccGGTGACAATTGGAGCTACAGC	TAD Boundary cloning
Boundary_9.1_Rv	actctagaactagtcGATGATAGTGCCGTACACATCG	TAD Boundary cloning
Boundary_11.1_Fw	ccctcgaggaattccATCAGATACCCGTTACTCATCTGG	TAD Boundary cloning
Boundary_11.1_Rv	actctagaactagtcGGATCGGTCGCTTGAATCCTAC	TAD Boundary cloning
Boundary_11.2_Fw	ccctcgaggaattccTGAATGCTCGAATCAATTGCGG	TAD Boundary cloning
Boundary_11.2_Rv	actctagaactagtcTACGTGCCCAACATGGACAG	TAD Boundary cloning
Boundary_11.3_Fw	ccctcgaggaattccCTTTCAGTGCAGATGGCGTTG	TAD Boundary cloning
Boundary_11.3_Rv	actctagaactagtcGGGACAGACAGTTGAAGTTCCTG	TAD Boundary cloning
Boundary_12.1_Fw	ccctcgaggaattccTTGGAGCCGAAGTATCGCTC	TAD Boundary cloning
Boundary_12.1_Rv	actctagaactagtcTATGCCAAATGCTGCCATACG	TAD Boundary cloning
Boundary_14.1_Fw	ccctcgaggaattccTGGTCAATATGAGGCGTATGGAC	TAD Boundary cloning
Boundary_14.1_Rv	actctagaactagtcGAACCAATAGCAGAAGTGGCAC	TAD Boundary cloning
Boundary_14.2_Fw	ccctcgaggaattccCTAGACACTGGAAGAGAGCCAC	TAD Boundary cloning
Boundary_14.2_Rv	actctagaactagtcCTCTCTATCTGGTTGAGGTCATACG	TAD Boundary cloning
Boundary_15.1_Fw	ccctcgaggaattccTCACTATTCACGTCCTCGATGG	TAD Boundary cloning
Boundary_15.1_Rv	actctagaactagtcAGCTCGCTCAGGTATGACTTG	TAD Boundary cloning
Boundary_15.2_Fw	ccctcgaggaattccGCACCTATGATGTATCGCATGC	TAD Boundary cloning
Boundary_15.2_Rv	actctagaactagtcGGCACGAGATACAATGTGCAC	TAD Boundary cloning
Boundary_16.2_Fw	ccctcgaggaattccACCGATTGGATGGAGCAACTG	TAD Boundary cloning
Boundary_16.2_Rv	actctagaactagtcAAGTGCCTGAGCATGATCTCC	TAD Boundary cloning
Boundary_17.1_Fw	ccctcgaggaattccTCATCCGCAAGGATAACAGGC	TAD Boundary cloning
Boundary_17.1_Rv	actctagaactagtcACTGGTCGAAAGTTCATCTCCAC	TAD Boundary cloning
MiL.F	GCGTAAGCTACCTTAATCTCAAGAAGAG	TAD Boundary insertion genotyping
MiL.R	CGCGGCGTAATGTGATTTACTATCATAC	TAD Boundary insertion genotyping
pw35_seq10	GGAATTCCTCGAGGGATCCAAG	TAD Boundary insertion genotyping
TAD_A_Fw	CTTTCGTGCACGCCCAATTA	TAD Boundary insertion genotyping
TAD_A_Rv	TTTGCACGGAACAAACGACAA	TAD Boundary insertion genotyping

TAD_B_Fw	TCGACAGAGCTTGAGTTGGTTT	TAD Boundary insertion genotyping
TAD_B_Rv	GTCTGCCGCTATTTGTTGCATA	TAD Boundary insertion genotyping

9 References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., . . . Venter, J. C. (2000). The genome sequence of Drosophila melanogaster. *Science, 287*(5461), 2185-2195. doi:10.1126/science.287.5461.2185

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., . . . Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res, 44*(W1), W3-w10. doi:10.1093/nar/gkw343

Akdemir, K. C., Le, V. T., Chandran, S., Li, Y., Verhaak, R. G., Beroukhim, R., . . . Consortium, P. (2020). Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat Genet, 52*(3), 294-305. doi:10.1038/s41588-019-0564-y

Arzate-Mejia, R. G., Josue Cerecedo-Castillo, A., Guerrero, G., Furlan-Magaril, M., & Recillas-Targa, F. (2020). In situ dissection of domain boundaries affect genome topology and gene transcription in Drosophila. *Nat Commun, 11*(1), 894. doi:10.1038/s41467-020-14651-z

Ashraf, S. I., Ganguly, A., Roote, J., & Ip, Y. T. (2004). Worniu, a Snail family zinc-finger protein, is required for brain development in Drosophila. *Dev Dyn, 231*(2), 379-386. doi:10.1002/dvdy.20130

Bartkuhn, M., Straub, T., Herold, M., Herrmann, M., Rathke, C., Saumweber, H., . . . Renkawitz, R. (2009). Active promoters and insulators are marked by the centrosomal protein 190. *EMBO J, 28*(7), 877-888. doi:10.1038/emboj.2009.34

Beagan, J. A., & Phillips-Cremins, J. E. (2020). On the existence and functionality of topologically associating domains. *Nat Genet, 52*(1), 8-16. doi:10.1038/s41588-019-0561-1

Bell, A. C., West, A. G., & Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell, 98*(3), 387-396. doi:10.1016/s0092-8674(00)81967-4

Bintu, B., Mateo, L. J., Su, J. H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., . . . Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science, 362*(6413). doi:10.1126/science.aau1783

Blackwood, E. M., & Kadonaga, J. T. (1998). Going the distance: a current view of enhancer action. *Science, 281*(5373), 60-63. doi:10.1126/science.281.5373.60

Blythe, S. A., & Wieschaus, E. F. (2015). Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. *Cell, 160*(6), 1169-1181. doi:10.1016/j.cell.2015.01.050

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., . . . Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature, 529*(7586), 418-422. doi:10.1038/nature16496

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., . . . Cavalli, G. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell, 171*(3), 557-572.e524. doi:10.1016/j.cell.2017.09.043

Bonn, S., Zinzen, R. P., Perez-Gonzalez, A., Riddell, A., Gavin, A. C., & Furlong, E. E. (2012). Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. *Nat Protoc, 7*(5), 978-994. doi:10.1038/nprot.2012.049

Bushey, A. M., Dorman, E. R., & Corces, V. G. (2008). Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol Cell, 32*(1), 1-9. doi:10.1016/j.molcel.2008.08.017

Buszczak, M., Paterno, S., Lighthouse, D., Bachman, J., Planck, J., Owen, S., . . . Spradling, A. C. (2007). The carnegie protein trap library: a versatile tool for Drosophila developmental studies. *Genetics, 175*(3), 1505-1531. doi:10.1534/genetics.106.065961

Butcher, R. D., Chodagam, S., Basto, R., Wakefield, J. G., Henderson, D. S., Raff, J. W., & Whitfield, W. G. (2004). The Drosophila centrosome-associated protein CP190 is essential for viability but not for cell division. *J Cell Sci, 117*(Pt 7), 1191-1199. doi:10.1242/jcs.00979

Butler, J. E., & Kadonaga, J. T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev, 16*(20), 2583-2592. doi:10.1101/gad.1026202

Cai, H. N., & Levine, M. (1997). The gypsy insulator can function as a promoter-specific silencer in the Drosophila embryo. *EMBO J, 16*(7), 1732-1741. doi:10.1093/emboj/16.7.1732

Campos-Ortega, J. A., & Hartenstein, V. (2013). *The embryonic development of Drosophila melanogaster*: Springer Science & Business Media.

Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., & Zeitlinger, J. (2013). A global change in RNA polymerase II pausing during the Drosophila midblastula transition. *Elife, 2*, e00861. doi:10.7554/eLife.00861

Chen, Y., & Belmont, A. S. (2019). Genome organization around nuclear speckles. *Curr Opin Genet Dev, 55*, 91-99. doi:10.1016/j.gde.2019.06.008

Chodagam, S., Royou, A., Whitfield, W., Karess, R., & Raff, J. W. (2005). The centrosomal protein CP190 regulates myosin function during early Drosophila development. *Curr Biol, 15*(14), 1308-1313. doi:10.1016/j.cub.2005.06.024

Chou, T. B., & Perrimon, N. (1996). The autosomal FLP-DFS technique for generating germline mosaics in Drosophila melanogaster. *Genetics, 144*(4), 1673-1679.

Cremer, T., & Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet, 2*(4), 292-301. doi:10.1038/35066075

Cubenas-Potts, C., Rowley, M. J., Lyu, X., Li, G., Lei, E. P., & 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*(4), 1714-1730. doi:10.1093/nar/gkw1114

Davies, J. O., Telenius, J. M., McGowan, S. J., Roberts, N. A., Taylor, S., Higgs, D. R., & Hughes, J. R. (2016). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods, 13*(1), 74-80. doi:10.1038/nmeth.3664

De Renzis, S., Elemento, O., Tavazoie, S., & Wieschaus, E. F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS Biol, 5*(5), e117. doi:10.1371/journal.pbio.0050117

de Wit, E. (2020). TADs as the Caller Calls Them. *J Mol Biol, 432*(3), 638-642. doi:10.1016/j.jmb.2019.09.026

de Wit, E., Vos, E. S., Holwerda, S. J., Valdes-Quezada, C., Verstegen, M. J., Teunissen, H., . . . de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell, 60*(4), 676-684. doi:10.1016/j.molcel.2015.09.023

Dekker, J., & Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. *FEBS Lett, 589*(20 Pt A), 2877-2884. doi:10.1016/j.febslet.2015.08.044

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science, 295*(5558), 1306-1311. doi:10.1126/science.1067799

Denker, A., & de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev, 30*(12), 1357-1382. doi:10.1101/gad.281964.116

Despang, A., Schopflin, R., Franke, M., Ali, S., Jerkovic, I., Paliou, C., . . . Ibrahim, D. M. (2019). Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet, 51*(8), 1263-1271. doi:10.1038/s41588-019-0466-z

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., . . . Ren, B. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature, 518*(7539), 331-336. doi:10.1038/nature14222

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., . . . Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature, 485*(7398), 376-380. doi:10.1038/nature11082

Djekidel, M. N., Chen, Y., & Zhang, M. Q. (2018). FIND: difFerential chromatin INteractions Detection using a spatial Poisson process. *Genome Res, 28*(3), 412-422. doi:10.1101/gr.212241.116

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., . . . Xie, W. (2017). Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature, 547*(7662), 232-235. doi:10.1038/nature23263

Duncan, I. W. (2002). Transvection effects in Drosophila. *Annu Rev Genet, 36*, 521-556. doi:10.1146/annurev.genet.36.060402.100441

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst, 3*(1), 99-101. doi:10.1016/j.cels.2015.07.012

Eagen, K. P., Aiden, E. L., & Kornberg, R. D. (2017). Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc Natl Acad Sci U S A, 114*(33), 8764-8769. doi:10.1073/pnas.1701291114

Erdel, F., Rademacher, A., Vlijm, R., Tünnermann, J., Frank, L., Weinmann, R., . . . Rippe, K. (2020). Mouse Heterochromatin Adopts Digital Compaction States without Showing Hallmarks of HP1-Driven Liquid-Liquid Phase Separation. *Mol Cell, 78*(2), 236-249.e237. doi:10.1016/j.molcel.2020.02.005

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics, 32*(19), 3047-3048. doi:10.1093/bioinformatics/btw354

Field, A., & Adelman, K. (2020). Evaluating Enhancer Function and Transcription. *Annu Rev Biochem, 89*, 213-234. doi:10.1146/annurev-biochem-011420-095916

Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., . . . Lobanenkov, V. V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol, 16*(6), 2802-2813. doi:10.1128/mcb.16.6.2802

FOE, V. E. (1993). Mitosis and morphogenesis in the Drosophila embryo: point and counterpoint. *The development of Drosophila melanogaster*(1), 149-300.

Fraser, J., Williamson, I., Bickmore, W. A., & Dostie, J. (2015). An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiol Mol Biol Rev, 79*(3), 347-372. doi:10.1128/mmbr.00006-15

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep, 15*(9), 2038-2049. doi:10.1016/j.celrep.2016.04.085

Fujioka, M., Mistry, H., Schedl, P., & Jaynes, J. B. (2016). Determinants of Chromosome Architecture: Insulator Pairing in cis and in trans. *PLoS Genet, 12*(2), e1005889. doi:10.1371/journal.pgen.1005889

Furlong, E. E. M., & Levine, M. (2018). Developmental enhancers and chromosome topology. *Science, 361*(6409), 1341-1345. doi:10.1126/science.aau0320

Galan, S., Machnik, N., Kruse, K., Díaz, N., Marti-Renom, M. A., & Vaquerizas, J. M. (2020). CHESS enables quantitative comparison of chromatin contact data and automatic feature extraction. *Nat Genet, 52*(11), 1247-1255. doi:10.1038/s41588-020-00712-y

Gambetta, M. C., & Furlong, E. E. M. (2018). The Insulator Protein CTCF Is Required for Correct Hox Gene Expression, but Not for Embryonic Development in Drosophila. *Genetics, 210*(1), 129-136. doi:10.1534/genetics.118.301350

Gambetta, M. C., & Müller, J. (2014). O-GlcNAcylation prevents aggregation of the Polycomb group repressor polyhomeotic. *Dev Cell, 31*(5), 629-639. doi:10.1016/j.devcel.2014.10.020

Gaszner, M., & Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet, 7*(9), 703-713. doi:10.1038/nrg1925

Gerasimova, T. I., & Corces, V. G. (1996). Boundary and insulator elements in chromosomes. *Curr Opin Genet Dev, 6*(2), 185-192. doi:10.1016/s0959-437x(96)80049-9

Gerasimova, T. I., & Corces, V. G. (2001). Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annu Rev Genet, 35*, 193-208. doi:10.1146/annurev.genet.35.102401.090349

Geyer, P. K., & Corces, V. G. (1992). DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes Dev, 6*(10), 1865-1873. doi:10.1101/gad.6.10.1865

Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R. R., Korbel, J. O., & Furlong, E. E. M. (2019). Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet, 51*(8), 1272-1282. doi:10.1038/s41588-019-0462-3

Gratz, S. J., Rubinstein, C. D., Harrison, M. M., Wildonger, J., & O'Connor-Giles, K. M. (2015). CRISPR-Cas9 Genome Editing in Drosophila. *Curr Protoc Mol Biol, 111*, 31.32.31-31.32.20. doi:10.1002/0471142727.mb3102s111

Gyurkovics, H., Gausz, J., Kummer, J., & Karch, F. (1990). A new homeotic mutation in the Drosophila bithorax complex removes a boundary separating two domains of regulation. *EMBO J, 9*(8), 2579-2585.

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol, 19*(10), 621-637. doi:10.1038/s41580-018-0028-8

Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., & Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife, 6*. doi:10.7554/eLife.25776

Heitz, E. (1928). *Das heterochromatin der moose*: Bornträger.

Hekmat-Scafe, D. S., Dang, K. N., & Tanouye, M. A. (2005). Seizure suppression by gain-of-function escargot mutations. *Genetics, 169*(3), 1477-1493. doi:10.1534/genetics.104.036558

Hogga, I., Mihaly, J., Barges, S., & Karch, F. (2001). Replacement of Fab-7 by the gypsy or scs insulator disrupts long-distance regulatory interactions in the Abd-B gene of the bithorax complex. *Mol Cell, 8*(5), 1145-1151. doi:10.1016/s1097-2765(01)00377-x

Hsieh, T. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Rando, O. J., Tjian, R., & Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell, 78*(3), 539-553 e538. doi:10.1016/j.molcel.2020.03.002

Hug, C. B., Grimaldi, A. G., Kruse, K., & Vaquerizas, J. M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell, 169*(2), 216-228 e219. doi:10.1016/j.cell.2017.03.024

Hyle, J., Zhang, Y., Wright, S., Xu, B., Shao, Y., Easton, J., . . . Li, C. (2019). Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Res, 47*(13), 6699-6713. doi:10.1093/nar/gkz462

Iampietro, C., Cléard, F., Gyurkovics, H., Maeda, R. K., & Karch, F. (2008). Boundary swapping in the Drosophila Bithorax complex. *Development, 135*(24), 3983-3987. doi:10.1242/dev.025700

Jack, J., Dorsett, D., Delotto, Y., & Liu, S. (1991). Expression of the cut locus in the Drosophila wing margin is required for cell type specification and is regulated by a distant enhancer. *Development, 113*(3), 735-747.

Jackson, D. A., Hassan, A. B., Errington, R. J., & Cook, P. R. (1993). Visualization of focal sites of transcription within human nuclei. *EMBO J, 12*(3), 1059-1065.

Jiang, N., Emberly, E., Cuvier, O., & 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*(13), 3556-3568. doi:10.1128/mcb.01748-08

Kaaij, L. J. T., van der Weide, R. H., Ketting, R. F., & de Wit, E. (2018). Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell Rep, 24*(1), 1-10.e14. doi:10.1016/j.celrep.2018.06.003

Kellum, R., & Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. *Cell, 64*(5), 941-950. doi:10.1016/0092-8674(91)90318-s

Kellum, R., & Schedl, P. (1992). A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol, 12*(5), 2424-2431. doi:10.1128/mcb.12.5.2424

Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T. S., . . . Rando, O. J. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell, 78*(3), 554-565 e557. doi:10.1016/j.molcel.2020.03.003

Kyrchanova, O., & Georgiev, P. (2014). Chromatin insulators and long-distance interactions in Drosophila. *FEBS Lett, 588*(1), 8-14. doi:10.1016/j.febslet.2013.10.039

Kyrchanova, O., Mogila, V., Wolle, D., Deshpande, G., Parshikov, A., Cléard, F., . . . Georgiev, P. (2016). Functional Dissection of the Blocking and Bypass Activities of the Fab-8 Boundary in the Drosophila Bithorax Complex. *PLoS Genet, 12*(7), e1006188. doi:10.1371/journal.pgen.1006188

Kyrchanova, O., Sabirov, M., Mogila, V., Kurbidaeva, A., Postika, N., Maksimenko, O., . . . Georgiev, P. (2019). Complete reconstitution of bypass and blocking functions in a minimal artificial Fab-7 insulator from Drosophila bithorax complex. *Proc Natl Acad Sci U S A, 116*(27), 13462-13467. doi:10.1073/pnas.1907190116

Lajoie, B. R., Dekker, J., & Kaplan, N. (2015). The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods, 72*, 65-75. doi:10.1016/j.ymeth.2014.10.031

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., . . . Snyder, M. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res, 22*(9), 1813-1831. doi:10.1101/gr.136184.111

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods, 9*(4), 357-359. doi:10.1038/nmeth.1923

Lenhard, B., Sandelin, A., & Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet, 13*(4), 233-245. doi:10.1038/nrg3163

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., . . . de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet, 12*(14), 1725-1735. doi:10.1093/hmg/ddg180

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*, 323. doi:10.1186/1471-2105-12-323

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Liang, H. L., Nien, C. Y., Liu, H. Y., Metzstein, M. M., Kirov, N., & Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature, 456*(7220), 400-403. doi:10.1038/nature07388

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., . . . Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science, 326*(5950), 289-293. doi:10.1126/science.1181369

Lun, A. T., & Smyth, G. K. (2015). diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC Bioinformatics, 16*, 258. doi:10.1186/s12859-015-0683-0

Lupianez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., . . . Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell, 161*(5), 1012-1025. doi:10.1016/j.cell.2015.04.004

Ma, W., Noble, W. S., & Bailey, T. L. (2014). Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat Protoc, 9*(6), 1428-1450. doi:10.1038/nprot.2014.083

Mateo, L. J., Murphy, S. E., Hafner, A., Cinquini, I. S., Walker, C. A., & Boettiger, A. N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature, 568*(7750), 49-54. doi:10.1038/s41586-019-1035-4

Matthey-Doret, C., Baudry, L., Breuer, A., Montagne, R., Guiglielmoni, N., Scolari, V., . . . Cournac, A. (2020). Computer vision for pattern detection in chromosome contact maps. *Nat Commun, 11*(1), 5795. doi:10.1038/s41467-020-19562-7

McCord, R. P., Kaplan, N., & Giorgetti, L. (2020). Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Mol Cell, 77*(4), 688-708. doi:10.1016/j.molcel.2019.12.021

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., . . . Osborne, C. S. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet, 47*(6), 598-606. doi:10.1038/ng.3286

Muravyova, E., Golovnin, A., Gracheva, E., Parshikov, A., Belenkaya, T., Pirrotta, V., & Georgiev, P. (2001). Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science, 291*(5503), 495-498. doi:10.1126/science.291.5503.495

Nègre, N., Brown, C. D., Shah, P. K., Kheradpour, P., Morrison, C. A., Henikoff, J. G., . . . White, K. P. (2010). A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet, 6*(1), e1000814. doi:10.1371/journal.pgen.1000814

Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., . . . de Laat, W. (2011). Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat Cell Biol, 13*(8), 944-951. doi:10.1038/ncb2278

Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., . . . Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell, 169*(5), 930-944 e922. doi:10.1016/j.cell.2017.05.004

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., . . . Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature, 485*(7398), 381-385. doi:10.1038/nature11049

Northcott, P. A., Lee, C., Zichner, T., Stutz, A. M., Erkek, S., Kawauchi, D., . . . Pfister, S. M. (2014). Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature, 511*(7510), 428-434. doi:10.1038/nature13379

O'Farrell, P. H., Stumpff, J., & Su, T. T. (2004). Embryonic cleavage cycles: how is a mouse like a fly? *Curr Biol, 14*(1), R35-45. doi:10.1016/j.cub.2003.12.022

Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J. M., & Cavalli, G. (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Mol Cell, 71*(1), 73-88 e75. doi:10.1016/j.molcel.2018.05.032

Ozdemir, I., & Gambetta, M. C. (2019). The Role of Insulation in Patterning Gene Expression. *Genes (Basel), 10*(10). doi:10.3390/genes10100767

Paliou, C., Guckelberger, P., Schopflin, R., Heinrich, V., Esposito, A., Chiariello, A. M., . . . Andrey, G. (2019). Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proc Natl Acad Sci U S A, 116*(25), 12390-12399. doi:10.1073/pnas.1900672116

Peifer, M., & Bender, W. (1986). The anterobithorax and bithorax mutations of the bithorax complex. *EMBO J, 5*(9), 2293-2303.

Peters, J. M., Tedeschi, A., & Schmitz, J. (2008). The cohesin complex and its roles in chromosome biology. *Genes Dev, 22*(22), 3089-3114. doi:10.1101/gad.1724308

Phillips-Cremins, J. E., Sauria, M. E., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S., . . . Corces, V. G. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell, 153*(6), 1281-1295. doi:10.1016/j.cell.2013.04.053

Pombo, A., & Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol, 16*(4), 245-257. doi:10.1038/nrm3965

Postika, N., Metzler, M., Affolter, M., Muller, M., Schedl, P., Georgiev, P., & Kyrchanova, O. (2018). Boundaries mediate long-distance interactions between enhancers and promoters in the Drosophila Bithorax complex. *PLoS Genet, 14*(12), e1007702. doi:10.1371/journal.pgen.1007702

Ramani, V., Deng, X., Qiu, R., Gunderson, K. L., Steemers, F. J., Disteche, C. M., . . . Shendure, J. (2017). Massively multiplex single-cell Hi-C. *Nat Methods, 14*(3), 263-266. doi:10.1038/nmeth.4155

Ramirez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Gruning, B. A., Villaveces, J., . . . Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun, 9*(1), 189. doi:10.1038/s41467-017-02525-w

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res, 42*(Web Server issue), W187-191. doi:10.1093/nar/gku365

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., . . . Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell, 159*(7), 1665-1680. doi:10.1016/j.cell.2014.11.021

Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., . . . Aiden, E. L. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell, 171*(2), 305-320.e324. doi:10.1016/j.cell.2017.09.026

Rodríguez-Carballo, E., Lopez-Delisle, L., Zhan, Y., Fabre, P. J., Beccari, L., El-Idrissi, I., . . . Duboule, D. (2017). The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev, 31*(22), 2264-2281. doi:10.1101/gad.307769.117

Rowley, M. J., Lyu, X., Rana, V., Ando-Kuri, M., Karns, R., Bosco, G., & Corces, V. G. (2019). Condensin II Counteracts Cohesin and RNA Polymerase II in the Establishment of 3D Chromatin Organization. *Cell Rep, 26*(11), 2890-2903.e2893. doi:10.1016/j.celrep.2019.01.116

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., . . . Corces, V. G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell, 67*(5), 837-852 e837. doi:10.1016/j.molcel.2017.07.022

Roy, S., Gilbert, M. K., & Hart, C. M. (2007). Characterization of BEAF mutations isolated by homologous recombination in Drosophila. *Genetics, 176*(2), 801-813. doi:10.1534/genetics.106.068056

Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schäfer, C., . . . Reuter, G. (2007). Heterochromatin formation in Drosophila is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Mol Cell, 26*(1), 103-115. doi:10.1016/j.molcel.2007.02.025

Ruf, S., Symmons, O., Uslu, V. V., Dolle, D., Hot, C., Ettwiller, L., & Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nature Genetics, 43*(4), 379-386. doi:10.1038/ng.790

Sanborn, A. L., Rao, S. S., Huang, S. C., Durand, N. C., Huntley, M. H., Jewett, A. I., . . . Aiden, E. L. (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*(47), E6456-6465. doi:10.1073/pnas.1518552112

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., . . . Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods, 9*(7), 676-682. doi:10.1038/nmeth.2019

Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., . . . Ren, B. (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep, 17*(8), 2042-2059. doi:10.1016/j.celrep.2016.10.061

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet, 20*(8), 437-455. doi:10.1038/s41576-019-0128-0

Schulz, K. N., & Harrison, M. M. (2019). Mechanisms regulating zygotic genome activation. *Nat Rev Genet, 20*(4), 221-234. doi:10.1038/s41576-018-0087-x

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., . . . Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature, 551*(7678), 51-56. doi:10.1038/nature24281

Scuderi, A., Simin, K., Kazuko, S. G., Metherall, J. E., & Letsou, A. (2006). scylla and charybde, homologues of the human apoptotic gene RTP801, are required for head involution in Drosophila. *Dev Biol, 291*(1), 110-122. doi:10.1016/j.ydbio.2005.12.014

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., . . . Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell, 148*(3), 458-472. doi:10.1016/j.cell.2012.01.010

Simpson, P. (1983). Maternal-Zygotic Gene Interactions during Formation of the Dorsoventral Pattern in Drosophila Embryos. *Genetics, 105*(3), 615-632.

Spitz, F., & Furlong, E. E. (2012). Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet, 13*(9), 613-626. doi:10.1038/nrg3207

Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., . . . Graf, T. (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat Genet, 50*(2), 238-249. doi:10.1038/s41588-017-0030-7

Staller, M. V., Yan, D., Randklev, S., Bragdon, M. D., Wunderlich, Z. B., Tao, R., . . . Perrimon, N. (2013). Depleting gene activities in early Drosophila embryos with the "maternal-Gal4-shRNA" system. *Genetics, 193*(1), 51-61. doi:10.1534/genetics.112.144915

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., . . . Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res, 24*(3), 390-400. doi:10.1101/gr.163519.113

Tsai, A., Galupa, R., & Crocker, J. (2020). Robust and efficient gene regulation through localized nuclear microenvironments. *Development, 147*(19). doi:10.1242/dev.161430

Udvardy, A., Maine, E., & 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*(2), 341-358. doi:10.1016/0022-2836(85)90408-5

Ulianov, S. V., Khrameeva, E. E., Gavrilov, A. A., Flyamer, I. M., Kos, P., Mikhaleva, E. A., . . . Razin, S. V. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res, 26*(1), 70-84. doi:10.1101/gr.196006.115

Ursu, O., Boley, N., Taranova, M., Wang, Y. X. R., Yardimci, G. G., Stafford Noble, W., & Kundaje, A. (2018). GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs. *Bioinformatics, 34*(16), 2701-2707. doi:10.1093/bioinformatics/bty164

Vallot, A., & Tachibana, K. (2020). The emergence of genome architecture and zygotic genome activation. *Curr Opin Cell Biol, 64*, 50-57. doi:10.1016/j.ceb.2020.02.002

Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J. E., & Corces, V. G. (2012). Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res, 22*(11), 2176-2187. doi:10.1101/gr.136788.111

Vastenhouw, N. L., Cao, W. X., & Lipshitz, H. D. (2019). The maternal-to-zygotic transition revisited. *Development, 146*(11). doi:10.1242/dev.161471

Venken, K. J., Schulze, K. L., Haelterman, N. A., Pan, H., He, Y., Evans-Holm, M., . . . Bellen, H. J. (2011). MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nat Methods, 8*(9), 737-743. doi:10.1038/nmeth.1662

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., . . . Zhu, X. (2001). The sequence of the human genome. *Science, 291*(5507), 1304-1351. doi:10.1126/science.1058040

Wan, L. B., Pan, H., Hannenhalli, S., Cheng, Y., Ma, J., Fedoriw, A., . . . Bartolomei, M. S. (2008). Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development. *Development, 135*(16), 2729-2738. doi:10.1242/dev.024539

Wei, W., & Brennan, M. D. (2001). The gypsy insulator can act as a promoter-specific transcriptional stimulator. *Mol Cell Biol, 21*(22), 7714-7720. doi:10.1128/mcb.21.22.7714-7720.2001

Williamson, I., Kane, L., Devenney, P. S., Flyamer, I. M., Anderson, E., Kilanowski, F., . . . Lettice, L. A. (2019). Developmentally regulated Shh expression is robust to TAD perturbations. *Development, 146*(19). doi:10.1242/dev.179523

Yan, K. K., Yardimci, G. G., Yan, C., Noble, W. S., & Gerstein, M. (2017). HiC-spector: a matrix library for spectral and reproducibility analysis of Hi-C contact maps. *Bioinformatics, 33*(14), 2199-2201. doi:10.1093/bioinformatics/btx152

Yang, J., & Corces, V. G. (2012). Insulators, long-range interactions, and genome function. *Curr Opin Genet Dev, 22*(2), 86-92. doi:10.1016/j.gde.2011.12.007

Yang, T., Zhang, F., Yardımcı, G. G., Song, F., Hardison, R. C., Noble, W. S., . . . Li, Q. (2017). HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res, 27*(11), 1939-1949. doi:10.1101/gr.220640.117

Yardımcı, G. G., Ozadam, H., Sauria, M. E. G., Ursu, O., Yan, K. K., Yang, T., . . . Noble, W. S. (2019). Measuring the reproducibility and quality of Hi-C data. *Genome Biol, 20*(1), 57. doi:10.1186/s13059-019-1658-7

Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics, 31*(14), 2382-2383. doi:10.1093/bioinformatics/btv145

Yuen, K. C., Slaughter, B. D., & 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*(6), e1700191. doi:10.1126/sciadv.1700191

Zhang, Y., Li, T., Preissl, S., Amaral, M. L., Grinstein, J. D., Farah, E. N., . . . Ren, B. (2019). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet, 51*(9), 1380-1388. doi:10.1038/s41588-019-0479-7

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., . . . Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol, 9*(9), R137. doi:10.1186/gb-2008-9-9-r137

Zolotarev, N., Fedotova, A., Kyrchanova, O., Bonchuk, A., Penin, A. A., Lando, A. S., . . . Georgiev, P. (2016). Architectural proteins Pita, Zw5,and ZIPIC contain homodimerization domain and support specific long-range interactions in Drosophila. *Nucleic Acids Res, 44*(15), 7228-7241. doi:10.1093/nar/gkw371

Zolotarev, N., Maksimenko, O., Kyrchanova, O., Sokolinskaya, E., Osadchiy, I., Girardot, C., . . . Georgiev, P. (2017). Opbp is a new architectural/insulator protein required for ribosomal gene expression. *Nucleic Acids Res, 45*(21), 12285-12300. doi:10.1093/nar/gkx840

Zufferey, M., Tavernari, D., Oricchio, E., & Ciriello, G. (2018). Comparison of computational methods for the identification of topologically associating domains. *Genome Biol, 19*(1), 217. doi:10.1186/s13059-018-1596-9