

UNIVERSITY OF GENEVA

MASTER'S THESIS IN BIOLOGY

***Study of the HoxD locus topological boundaries
inside and outside from their genomic context***

by

Andréa Willemin

July 26, 2020

DIRECTORS OF THE WORK

Professor Denis Duboule
Eddie Rodríguez-Carballo, DMD, PhD
Department of Genetics and Evolution
Faculty of Science (UNIGE)

JUROR

Professor David Shore
Department of Molecular Biology
Faculty of Science (UNIGE)

Table of Contents:

Résumé :	5
Summary:	7
Introduction:	9
1. The hierarchical organization of genomes at a glance	9
2. 3C-based techniques to study chromatin interactions	10
3. Topologically associating domains as units of genome organization	14
3.1 Properties and features associated with TAD boundaries	14
3.2 TADs and gene regulation by cis-acting elements	17
3.3 TADs in other contexts than cis regulation	19
4. Principles of chromatin folding	23
4.1 Of self-avoidance	23
4.2 When binders come into play	23
4.3 Putting the SBS model to the test	24
4.4 The loop extrusion model	25
4.5 Intra-TAD dynamics and relationships with compartments	27
4.6 Phase separation in genome's shape and gene expression control	30
5. Structure and function of the mammalian Hox gene clusters	33
5.1 Secondary axes development, long-range regulations and TAD organization at the HoxD locus	34
6. Scope of the thesis	38
Results:	39
1. Partial duplications of the HoxD cluster	39
1.1 Duplication of the region 8i-10	41
1.2 Duplication of the region 11-13	45
2. Ectopic boundary integrations	51
2.1 Preservation of contact directionality from an ectopic HoxD cluster	52
2.2 Relocating a boundary fragment	57
2.3 Region CS38-40 as an ectopic border	65
Discussion:	88
1. Modulating boundary potential at the endogenous HoxD cluster	88
2. Structure versus function at the HoxD locus	88
3. A possible contribution of promoter competition	90
4. Preservation of a HoxD-like contact directionality in a new context	92
5. Alterations of the local chromatin landscape at ectopic locations	93
6. Reconstitution of a functional sub-TAD boundary ectopically	94
7. Gene expression changes from the constructs and in the vicinity of ectopic boundary integrations	95
Conclusion:	98
Acknowledgements:	99
Material and Methods:	100
References:	113
Annexes:	136

Résumé :

Dans l'espace tridimensionnel du noyau cellulaire, le génome des mammifères est organisé en structures qui s'entrecroisent de manière hiérarchique à différentes échelles : des nucléosomes, au sein desquels la double hélice de l'ADN s'enroule autour des octamères d'histones, jusqu'aux territoires chromosomiques. Parmi cet éventail de structures de la chromatine se trouvent des unités d'organisation génomique d'une taille de l'ordre de la mégabase ou légèrement moins, appelés domaines d'association topologique (topologically associating domains) ou TADs. Les TADs se définissent comme domaines de la chromatine dans lesquels les interactions sont plus fréquentes qu'avec les régions situées en dehors de ces unités au sein d'une population de cellules. Il a été proposé que ces domaines endossent des fonctions importantes pour le génome, particulièrement dans le contexte de la régulation de l'expression des gènes en *cis*. En effet, il est communément admis que les éléments de régulation distants tels que les enhancers doivent être amenés en étroite proximité spatiale de leur promoteurs cibles de manière dynamique, suggérant de ce fait une fonction pour des conformations préexistantes de la chromatine telles que les TADs.

Au cours de ces dernières années, des efforts considérables ont été investis pour la caractérisation des TADs ainsi que dans la dissection des mécanismes par lesquels ces domaines se forment et sont maintenus. En particulier, des études basées sur l'utilisation combinée de techniques d'immunoprécipitation de la chromatine et de capture de conformation chromosomique ont identifié CTCF et le complexe cohésine comme étant des facteurs d'une importance majeure pour l'organisation du génome en TADs. Ces protéines architecturales sont absolument fondamentales pour l'extrusion de boucles (loop extrusion), qui représente le modèle prédominant pour expliquer la formation des boucles de la chromatine et des TADs. Dans cette optique, le complexe cohésine tire la fibre de chromatine de manière processive jusqu'à ce qu'il rencontre des sites CTCF en orientation convergente, formant ainsi une structure en boucle qui peut rester stable pendant un certain temps.

Dans le génome, les sites CTCF sont typiquement enrichis à la limite des TADs et, avec le complexe cohésine, il est à présent clair qu'ils participent à la formation des frontières entre les TADs. Néanmoins, des questions restent sans réponses dans ce contexte à l'heure actuelle. Typiquement, est-ce que la force d'une frontière pourrait être modulée en augmentant le nombre d'éléments de frontière qu'elle contient et est-ce qu'une frontière pourrait exercer sa fonction en dehors de son contexte d'origine ?

Dans cette étude, nous avons utilisé différentes régions du locus murin *HoxD* comme un paradigme de frontières topologiques avec pour objectif d'étendre la connaissance au sujet de la formation de ces frontières. Premièrement, nous avons tiré parti du complexe de gènes *HoxD*, qui se trouve lui-même à une frontière riche en CTCF et en gènes située entre deux TADs. Nous avons analysé des duplications internes du cluster augmentant le nombre d'éléments associés à la frontière *HoxD*, comprenant des sites de liaison à CTCF et des promoteurs de gènes. Avec ces duplications, nous avons observé une diminution des contacts établis entre les gènes proximaux à la frontière des TADs (qui sont capables d'interagir avec les deux TADs chez le *wild-type*) et le TAD situé au-delà du segment dupliqué, indiquant un renforcement de la frontière *in situ*.

Comme approche secondaire, nous avons analysé différentes intégrations ectopiques d'éléments de frontière provenant du locus *HoxD* dans le but de déterminer s'ils sont capables de fonctionner en dehors de leur contexte génomique originel. En bref, nous avons d'abord observé que les protéines architecturales sont recrutées au niveau de leurs sites de liaison sur les éléments de frontière déplacés. De plus, nous avons remarqué que des contacts se sont établis entre les éléments intégrés et les régions environnantes, typiquement restreints dans les limites des TADs dans lesquels les intégrations ont eu lieu. Ces contacts représentent la formation de nouvelles boucles dans la plupart des cas. En accord avec ces observations, nous avons noté une augmentation significative de l'insulation entre les régions à présent séparées par l'intégration comparé aux situations *wild-type* correspondantes. Collectivement, ces données suggèrent fortement que des éléments de frontière intégrés à des positions ectopiques sont capables de causer des altérations locales du paysage architectural dans leur nouveau contexte.

Summary:

Within the three-dimensional space of the nucleus, mammalian genomes are organized into structures that are intertwined in a hierarchical manner at different scales: from nucleosomes in which the DNA double-helix is wrapped around histone octamers to chromosome territories. Among this range of chromatin structures are sub-megabase or megabase-scale units of genomic organization, called topologically associating domains or TADs. TADs are defined as chromatin domains in which the interactions are more frequent than with regions located outside of such units within a cell population. These domains were proposed to hold important functions in the genome, particularly in the context of gene *cis* regulation. Indeed, it is generally understood that distant regulatory sequences such as enhancers need to come in close spatial proximity to their target promoters in a dynamic manner, hence suggesting a function for pre-existing chromatin conformations such as TADs.

Over the past years, tremendous efforts have been invested in the characterization of TADs as well as in the dissection of the mechanisms by which they are formed and maintained. In particular, studies using combinations of chromatin immunoprecipitation and chromosome conformation capture techniques found CTCF and the cohesin complex to be factors of major importance regarding the organization of the genome into TADs. These architectural proteins are fundamental for loop extrusion, which represents the prevailing model to explain the formation of chromatin loops and TADs. In this view, the cohesin complex pulls the chromatin fibre until it reaches convergent CTCF sites, thereby forming a loop that remains stable for a certain amount of time.

In the genome, CTCF sites are typically enriched at the limit between TADs and, together with the cohesin complex, they are known to participate in the formation of TAD boundaries. Nevertheless, it remained unclear whether the strength of a boundary could be modulated by increasing the number of the boundary elements it contains, and whether a boundary could function outside of its original context.

In the present study, we used different regions of the murine *HoxD* locus as a paradigm of topological borders with the aim of extending the knowledge about the formation of such boundaries. We first took advantage of the *HoxD* gene cluster, which is itself located at a CTCF and gene-rich boundary between two TADs. We analyzed intra-cluster duplications increasing the number of boundary-associated elements including CTCF-binding sites and gene promoters

within the *HoxD* TAD boundary. Upon duplication, we observed a decrease in the contacts established between genes that are proximal to the TAD border (which are able to interact with both TADs in the *wild-type* situation) and the TAD located beyond the duplicated segment, which is indicative of an increase in the strength of the boundary *in situ*.

As a second approach, we analyzed several ectopic integrations of boundary elements originating from the *HoxD* locus in order to determine if they are able to function outside of their original genomic context. In brief, we first observed that the architectural proteins were recruited at the level of their binding sites on the relocated boundary elements. In addition, we observed an establishment of contacts between the integrated elements and the surrounding regions, which were typically restricted to the welcoming TADs and most often represented the formation of new loops. Consistent with these observations, we noticed a significant increase in insulation between regions located across the integration sites compared to the corresponding *wild-type* situations. Taken together, these data strongly suggest that ectopically integrated boundary elements are able to cause local alterations of the architectural landscape in a novel context.

Introduction:

1. The hierarchical organization of genomes at a glance

Within the cell nucleus, mammalian genomes display a remarkably defined organization in interphase, with different levels of structures that are intertwined with respect to one another in a hierarchical manner (Fig. 1) (reviewed in Sivakumar, de las Heras, and Schirmer 2019). At the smallest scale, the wrapping of 146 base pairs (bp) of DNA around histone octamers produces nucleosomes. As the most fundamental units of the so-called chromatin fibre, nucleosomes are important for both the compaction and the regulation of the genome depending on various post-translational modifications of the histone tails or the incorporation of histone variants (reviewed in Bannister, Falcão, and Castelo-Branco 2017).

With the advent of microscopy and sequencing-based methods to study chromatin architecture (such as DNA-FISH and Hi-C, respectively; see further) (reviewed in Kempfer and Pombo 2020), it is now clear that the genome also displays a remarkable organization at larger scales, giving rise to structures that are collectively referred to as higher-order features of genome organization (Sivakumar, de las Heras, and Schirmer 2019). Typically, looping and condensation (e.g. phase separation; see further) events occur between regions of the chromatin fibre that can be quite distant on the linear genome (Finn and Misteli 2019). While some of these interactions are purely structural, like anchor points, others can be related to the gene-regulation process, as in the case of enhancer-promoter contacts (Bonev et al. 2017).

At the sub-megabase or megabase-scale, depending on the considered genomic region, the genome is partitioned in self-interacting segments, called topologically associating domains or TADs (Dixon et al. 2012; Nora et al. 2012; see further for details). At the next level, TADs sharing similar chromatin states (e.g. histone modifications status and chromatin opening, which relate to gene activity) tend to fold together within the nucleoplasm, forming the so-called compartments. In brief, two main compartments, called A and B, are formed by the association of active or inactive TADs, respectively (Lieberman-Aiden et al. 2009; Dixon et al. 2012; Eagen, Hartl, and Kornberg 2015).

Finally, at the largest possible scale, the chromosomes themselves are not randomly located within the nucleoplasm, but instead occupy their own preferred locations, referred to as chromosome territories (Bolzer et al. 2005; reviewed in Cremer and Cremer 2010).

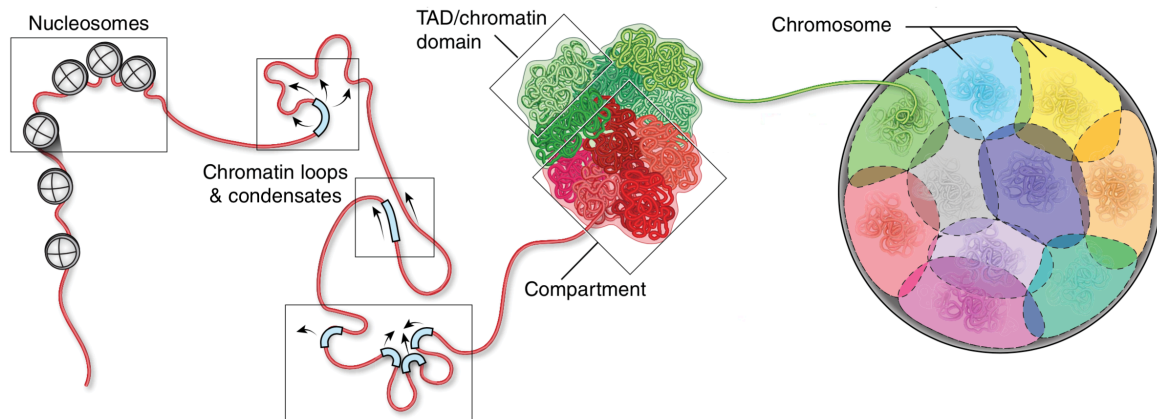


Figure 1: Modified from (Finn and Misteli 2019). Schematic representation of the various levels of genome organization. Note that the different features of genome architecture that are represented are not drawn to scale with respect to one another.

2. 3C-based techniques to study chromatin interactions

Given the central importance of understanding the structural framework underlying gene regulation, a large set of techniques are now available to study higher-order genome architecture (Kempfer and Pombo 2020). The majority of these techniques can be subdivided in two main categories, which are (1) imaging-based techniques, for instance DNA fluorescent *in situ* hybridization (DNA-FISH) (e.g. M. Cremer et al. 2008), and (2) chromosome conformation capture (3C)-based technologies (de Wit and de Laat 2012).

Imaging or microscopy-based techniques enable to characterize chromatin architecture at virtually any scale and in individual cells. Nevertheless, these techniques are generally limited to the study of a reduced number of chromatin regions at a time because of limitations in the different fluorescent probes that can be used simultaneously to visualize them (Mayer et al. 2005; M. Cremer et al. 2008; Ni et al. 2017). As a consequence, they are nowadays mainly applied to confirm the results obtained using other, higher throughput methods (Giorgetti and Heard 2016; e.g. Rodríguez-Carballo et al. 2017; Kempfer and Pombo 2020).

In contrast, recent developments of 3C such as Hi-C allow the detection of chromatin contacts genome-wide in a cell population, based on sequencing (Lieberman-Aiden et al. 2009; de Wit and de Laat 2012). All chromosome conformation capture techniques rely on a common procedure called proximity ligation. In brief, the chromatin is crosslinked by a formaldehyde treatment, which fixes protein-mediated interactions, digested using a restriction enzyme and ligated producing hybrid DNA molecules. The term “hybrid” means that these molecules are constituted by two parts, each of which originate from

regions that can be quite distant on the linear genome but were in close spatial proximity with one another (e.g. interacting chromatin segments). These DNA molecules are then reverse crosslinked before further processing and analysis by either qPCR or sequencing, depending on the technique (Dekker et al. 2002; Dostie et al. 2006; Zhao et al. 2006; Splinter et al. 2011; van de Werken et al. 2012; reviewed in Sati and Cavalli 2017).

In 3C-based techniques, the ligation events joining DNA fragments of different origins with respect to the linear genome are used to evaluate contact or interaction frequencies. The principle is the following: the more often two chromatin regions are actually interacting in the cell population under study, the more often their DNA will be crosslinked and proximity ligated, and the higher will be the observed interaction frequency. Hence, this signal is an average estimated probability of interactions between chromatin regions in the analyzed cell population (reviewed in de Wit and de Laat 2012; Giorgetti and Heard 2016).

The most commonly used 3C-based techniques are 4C-seq and Hi-C (Fig. 2) (Lieberman-Aiden et al. 2009; Splinter et al. 2011; van de Werken et al. 2012). 4C-seq stands for circular chromosome conformation capture (4C is also the acronym of 3C-on-chip; a version using microarrays) coupled with high-throughput sequencing (Zhao et al. 2006; Splinter et al. 2011; van de Werken et al. 2012). As its name suggests, 4C-seq is based on the production of circular hybrid DNA molecules thanks to an additional round of restriction-ligation, performed after reverse crosslinking. An inverse PCR (i.e. with primers in divergent orientation) is then carried out using primers recognizing a region of interest called the viewpoint. The PCR-amplified molecules, referred to as 4C-seq libraries, are finally analyzed by sequencing. Thereby, 4C-seq enables to score interactions taking place between the viewpoint and all other regions of the genome (“one *versus* all” strategy) (Splinter et al. 2011; van de Werken et al. 2012).

In contrast, Hi-C does not rely on the use of specific primers to score interactions between chromatin regions. In Hi-C, only one round of restriction-ligation is performed and includes the incorporation of biotinylated nucleotides at hybrid junctions. The hybrid molecules are then sheared in small fragments by sonication. The presence of biotin at the level of hybrid junctions enables to enrich the solution in hybrid DNA molecules so as to maximize their sequencing. In this way, Hi-C allows to detect the contacts occurring between any region of the genome and any other region (“all *versus* all” strategy) (Lieberman-Aiden et al. 2009).

Depending on what these techniques measure, the data can be visualized in different ways. Typically, because Hi-C scores contacts occurring between any pair of genomic regions, the data are generally represented in the form of a heatmap. In Hi-C heatmaps, the interactions are projected in diagonal and their relative frequencies are color-coded. Since 4C-seq measures interactions between the viewpoint and any other region of the genome, contact frequencies are represented in the form of a quantitative linear track (see examples on Fig. 2) (Ing-Simmons and Vaquerizas 2019).

Of particular interest, the recent identification of TADs and of compartments was made possible by the development of the Hi-C technique (Lieberman-Aiden et al. 2009; Dixon et al. 2012; Nora et al. 2012). As continuous domains of increased interaction frequencies, TADs appear as on-diagonal pyramids in Hi-C heatmaps. In contrast, compartments are discontinuous and therefore appear as an off-diagonal checkerboard pattern.

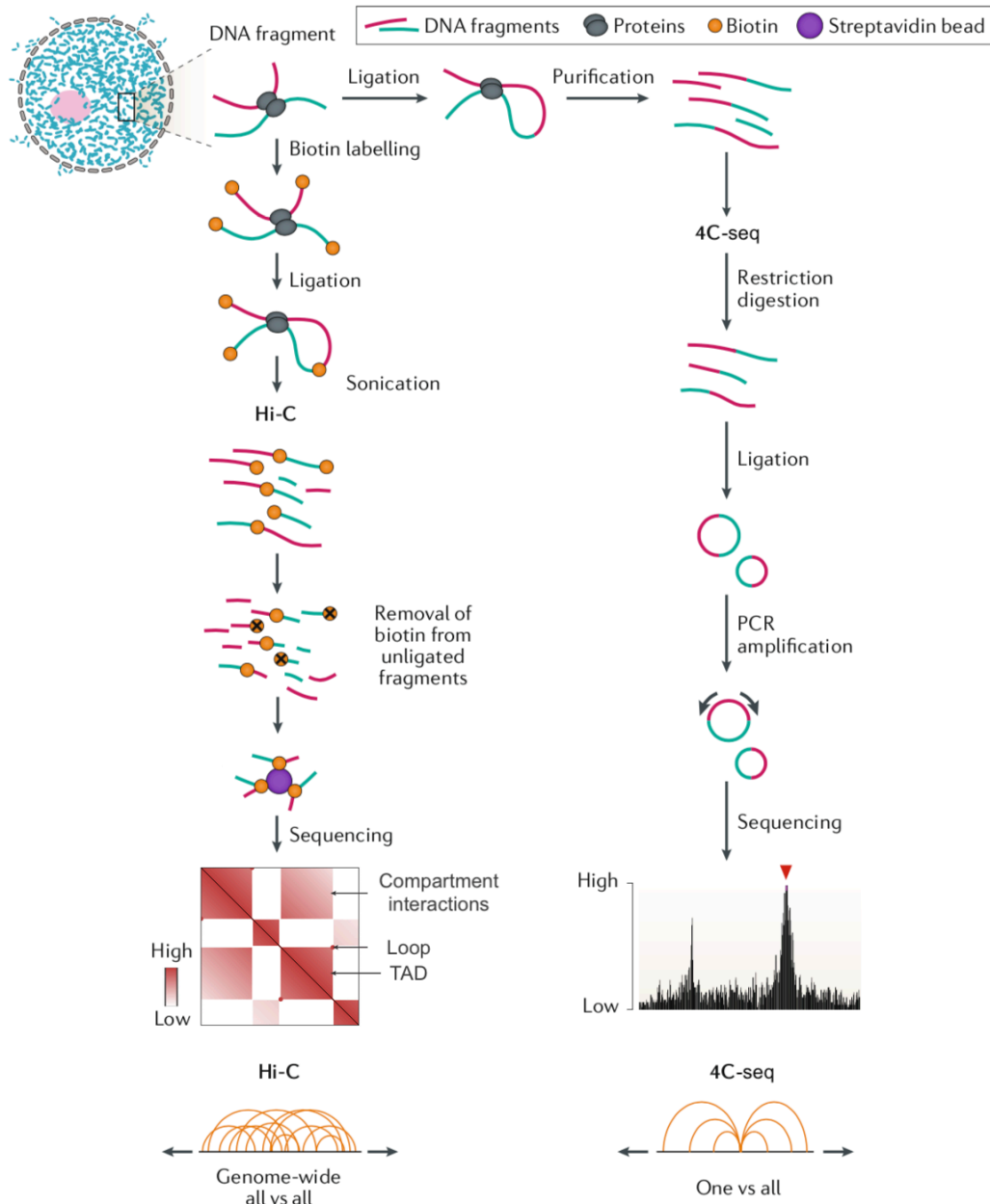


Figure 2: Modified from (Kempfer and Pombo 2020). Overview of Hi-C and 4C-seq. Crosslinked and restriction enzyme-fragmented chromatin serves as starting material for both techniques. In Hi-C, the ends of the DNA fragments are labelled with biotin. This is followed by ligation and sonication. The solution is then enriched in hybrid DNA molecules thanks to a streptavidin pull-down before sequencing. Below is shown an idealized heatmap representation of Hi-C data, where the color tone reflects the frequency of interaction; modified from (Eagen 2018). TADs appear as on-diagonal (continuous) pyramidal domains of increased interaction frequencies, constituted by many chromatin loops. Compartments form an off-diagonal (discontinuous) checkerboard pattern. In 4C-seq, an additional round of restriction-ligation is performed after reverse crosslinking, resulting in the production of circular hybrid DNA molecules. A viewpoint-directed inverse PCR is then carried out prior to sequencing. 4C-seq data are displayed in the form of a quantitative linear track, in which the height of the signal represents the interaction frequency. The position of the 4C-seq viewpoint is highlighted by a red arrowhead.

3. Topologically associating domains as units of genome organization

As previously mentioned, TADs are intermediate-scale units of genome organization, ranging from 40 kilobases to 3 megabases in size (Dixon et al. 2012; Nora et al. 2012; Rao et al. 2014). They are defined as domains of increased interaction frequencies among a cell population, in which each cell displays a particular set of loops that are highly dynamic over time and vary along differentiation pathways (Nagano et al. 2013; Dixon et al. 2015; Bonev et al. 2017; Finn et al. 2019). Beyond intra-domain variations (see intra-TAD dynamics further), the limits of TADs are mostly maintained among different cell types and are largely conserved across vertebrate species (Dixon et al. 2012; Vietri Rudan et al. 2015; Lazar et al. 2018).

3.1 Properties and features associated with TAD boundaries

The limits between TADs, which are most often referred to as TAD boundaries or borders, are defined as DNA regions that prevent most chromatin contacts from going across them (Dixon et al. 2012; Nora et al. 2012; Rao et al. 2014).

TADs and their boundaries are usually identified by either of two distinct computational approaches (Dixon et al. 2012; Crane et al. 2015; Kruse et al. 2016; Shin et al. 2016). Both kinds of TAD/boundary-calling algorithms start from genome-wide contact distributions, generally Hi-C matrices (e.g. Kruse et al. 2016). In the first variant, a drastic change in contact directionality (expressed by a “directionality index”) is used to identify the boundaries between adjacent TADs. The principle is the following: considering two TADs separated by a boundary region, a chromatin segment that is close on the left of the border will mostly interact with the left TAD, namely towards the left. The opposite contact directionality will be observed for a chromatin segment on the right-hand side of the boundary. The zone of contact directionality transition corresponds to the boundary (Dixon et al. 2012). The second variant is directly based on the definition of TADs as domains in which the interactions are more frequent than with regions located in a different TAD. In this approach, average contact frequencies are measured along a Hi-C matrix using sliding windows of given sizes, producing the so-called “insulation index”. Significant depletions of contacts, seen as a local minima of the insulation index, enable to identify the borders between adjacent domains (Crane et al. 2015; Shin et al. 2016).

Ever since the identification of TADs and of their boundaries, researchers have sought to determine what could be driving their formation. In this quest, it was shown that TAD boundaries are often associated with at least three different

features. These are (1) binding sites for the CTCF protein, (2) CpG islands and (3) housekeeping genes (Dixon et al. 2012; Nora et al. 2012; Sun et al. 2018).

The first two studies reporting the existence of topologically associating domains in the genome showed that most TAD boundaries are enriched for sites bound by a protein called CTCF (Dixon et al. 2012; Nora et al. 2012). As explained in detail further (see the loop extrusion model), CTCF and a ring-shaped complex called cohesin are known to participate in the formation of chromatin loops (Phillips-Cremins et al. 2013; Sanborn et al. 2015; Fudenberg et al. 2016). In the genome, chromatin loops can be located either within TADs, or span different domains. Nevertheless, consistent with the definition of TADs and the genome-wide distribution of CTCF sites, most chromatin loops are found within the limits of these domains, strongly suggesting their containment by the domains' boundaries (e.g. Dixon et al. 2012).

Intriguingly, only a small fraction of CTCF sites (about 15% in the pioneer study of Dixon and colleagues) is located at TAD boundaries, meaning that the majority are found within TADs. Hence, these results indicate that single or few CTCF sites are not sufficient to form a boundary in most cases, so that additional features need to be involved (Dixon et al. 2012; Nora et al. 2012). In this context, key studies reported that the density of CTCF sites and their orientation are of particular importance. Indeed, most boundary-located CTCF sites are densely arrayed and preferentially orientated towards the closest flanking domain (Rao et al. 2014; Zhan et al. 2017). In other words, considering a single TAD, boundary CTCFs are generally arranged in convergent (i.e. "facing") pairs with respect to the opposite limit of the domain (Rao et al. 2014). This arrangement was later proposed to anchor chromatin loops and favor boundary formation in what is referred to as the loop extrusion model (Sanborn et al. 2015; Fudenberg et al. 2016).

The existence of chromatin loops spanning adjacent TADs demonstrates that TAD boundaries do not act as completely impermeable barriers in most cases, but can instead show non-negligible leakage (Dixon et al. 2012; Chang, Ghosh, and Noordermeer 2020). In this context, one should bear in mind that TADs are cell population-averaged chromatin domains. Hence, if we consider two adjacent TADs, most of the cells will only display intra-TAD contacts. If the enrichment in intra-domain contacts is sufficient, two distinct TADs will be identified when the cell population is analyzed as a whole (like in bulk Hi-C experiments) (Lieberman-Aiden et al. 2009; Dixon et al. 2012; Nora et al. 2012; Rao et al. 2014). However, some cells will actually show interactions spanning

the two domains (inter-TAD contacts), implying that the latter are going through a region defined as a boundary (Nagano et al. 2013; Stevens et al. 2017; Finn et al. 2019).

Generally speaking, TAD boundaries display a variable strength, depending on the considered loci (Gong et al. 2018; Chang, Ghosh, and Noordermeer 2020). Interestingly, a recent study in which TAD boundaries were classified according to the insulation index found that the strength of a TAD boundary correlates with the level of CTCF binding at the boundary, as expressed by the height of the ChIP peaks (Gong et al. 2018). Since a higher ChIP signal is likely due to the binding of the protein in a larger number of cells rather than to a “stronger” binding at the single-cell level, these results also fall within the cell population-average perspective of TADs and of their boundaries. Indeed, they indicate a link between (1) the number of cells in which CTCF is occupying a given boundary and (2) the number of cells in which this boundary is actually preventing chromatin contacts from going through it (Gong et al. 2018). Apart from CTCF binding levels, it was shown that boundaries also correlate with other features, such as CpG islands and housekeeping genes (Dixon et al. 2012; Sun et al. 2018).

CpG islands (CGIs) are small DNA regions (one kilobase on average) that are particularly enriched for the 5'-Cytidine-phosphate-Guanosine-3', or CpG motif. In vertebrates' genomes, nearly half of all CpG islands are associated with sites of transcriptional initiation, which are hence referred to as CGI promoters, suggesting a role in the regulation of gene expression (reviewed in Deaton and Bird 2011). In particular, it is generally understood that CpG islands generate a chromatin state that is favorable for the initiation of transcription at CGI promoters, possibly by modulating nucleosome occupancy or via the recruitment of transcription factors. Conversely, when methylated at the level of their cytosines, transcriptional initiation would be prevented from occurring at CGI promoters (e.g. Deaton and Bird 2011; Jones 2012). As regards chromatin organization, it was shown that TAD boundaries match with regions of particularly high CpG islands' density compared to other, non-boundary regions (Sun et al. 2018). However, to date, the potential contribution of CpG islands to TAD boundary formation remains largely unexplored from a mechanistical point of view. One might even postulate that their enrichment at TAD boundaries could be linked to the regulation of housekeeping genes' expression (which most often carry CGI promoters and are enriched at TAD boundaries; see below) (Smale and Kadonaga 2003), rather than to boundary function.

As mentioned above, housekeeping genes are also enriched at TAD borders (Dixon et al. 2012). These genes hold fundamental functions in cells; they are for instance required for energy metabolism and proteins biosynthesis, and are therefore expressed in a stable and ubiquitous manner (Eisenberg and Levanon 2013). Hence, it has been proposed that a high and sustained transcriptional activity could participate in the establishment of TAD boundaries (Dixon et al. 2012). Nevertheless, a subsequent study found that forcing transcription to occur at given loci using a dead Cas9 fused to a strong activator domain failed to result in the formation of novel boundaries (Bonev et al. 2017). Therefore, the enrichment of housekeeping genes at the level of TAD boundaries is likely the consequence and not the cause of a boundary being present. Typically, since housekeeping genes most often rely on proximal regulatory elements (such as their proximal promoter), it is possible that they do not require scaffolding by TAD-like chromatin structures (see TADs and gene regulation by *cis*-acting elements below). This could possibly explain why housekeeping genes tend to occupy boundary regions in the genome.

3.2 TADs and gene regulation by *cis*-acting elements

Far more than a widespread format of DNA packaging within the nucleus, TADs are thought to be important functional units of the genome, especially for the regulation of gene expression. Indeed, because *cis* regulatory elements such as enhancers can be located as far as several hundreds of kilobases away from their target promoters (Gasperini et al. 2019; S. Kim and Shendure 2019), some mechanisms are required to bring them together in close spatial proximity in order to trigger gene expression (discussed in details in Furlong and Levine 2018).

In this view, it is generally understood that TADs (or at least the loops they contain) can act as structural scaffolds favoring interactions between distal regulatory elements and gene promoters, while ensuring the specificity of such interactions within specific “regulatory landscapes” or “insulated neighborhoods” (Fig. 3A) (Spitz, Gonzalez, and Duboule 2003; Montavon et al. 2011; Downen et al. 2014; Ji et al. 2016; also see van Arensbergen, van Steensel, and Bussemaker 2014). This is particularly important in situations in which transcriptional gene activation has to be tightly regulated by long-range *cis*-acting elements in both time and space, as in the case of developmental genes (Furlong and Levine 2018; Bolt and Duboule 2020).

Supporting this view, several studies reported that genetic alterations causing the disruption or reorganization of TADs, due to either deletions, inversions or duplications (referred to as structural variants) encompassing a

TAD boundary, can lead to the rewiring of enhancer-promoter interactions, which can in turn result in disease-causing gene misregulation (Lupiáñez et al. 2015; Franke et al. 2016; Hnisz et al. 2016; Rodríguez-Carballo et al. 2017; reviewed in Maeso, Acemel, and Gómez-Skarmeta 2017).

Importantly, the structural variants that were able to affect the overall TAD organization leading to altered gene expression were the ones including a TAD boundary in most of the cases (Lupiáñez et al. 2015; Franke et al. 2016; Northcott et al. 2017; Despang et al. 2019). For instance, deletions or inversions that were restricted within a TAD and not including one of its boundaries were unable to drive such changes, demonstrating how robust the structure and function of these domains can be as long as their boundaries are left intact (Lupiáñez et al. 2015; Despang et al. 2019).

Nevertheless, it should be noted that the propensity of TAD rearrangements to result in gene misregulation remains a matter of debate (Furlong and Levine 2018). Notably, key studies in which TADs were either disrupted or depleted genome-wide revealed only moderate transcriptional changes by RNA-sequencing, with the vast majority of genes displaying unaltered expression levels (Zuin et al. 2014; Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017; Ghavi-Helm et al. 2019). Moreover, in most cases, not every gene within a given domain displayed changes in expression upon TAD disruption or merging (Schwarzer et al. 2017; Rodríguez-Carballo et al. 2019). In this context, it is reasonable to think that only the expression of genes relying on regulation by distant enhancers (e.g. developmental genes) might be altered when interfering with TADs, whereas genes depending on very proximal regulatory elements (typically housekeeping genes) are likely to stay unaffected (Zuin et al. 2014; Rodríguez-Carballo et al. 2019).

It should be noted that gene expression analysis had only limited or sometimes even completely lacked spatiotemporally resolution in the abovementioned studies. Indeed, the experiments were either performed on cells in culture (Zuin et al. 2014; Nora et al. 2017; Rao et al. 2017), in a specific tissue (Schwarzer et al. 2017), or were conducted in a reduced number of developmental stages (Ghavi-Helm et al. 2019). Nevertheless, these results suggest that the rewiring of enhancer-promoter contacts upon alteration of TAD structure may reflect very particular situations rather than a general rule (Furlong and Levine 2018; Ghavi-Helm et al. 2019). Therefore, it raises the possibility for other features than TAD-related chromatin structure, for instance transcriptional condensates, to be able to drive appropriate enhancer-promoter

interactions and achieve gene activation (Hnisz et al. 2017; Boija et al. 2018; S. Kim and Shendure 2019; Shrinivas et al. 2019).

3.3 TADs in other contexts than *cis* regulation

In addition to their function in the regulation of gene expression by *cis* regulatory elements via “blocking activity” (Fig. 3B), at least in some specific cases as discussed above, TADs were also proposed to affect gene expression by different means as well as to function in other nuclear processes (reviewed in Dixon, Gorkin, and Ren 2016).

For instance, the genome may also benefit from TADs’ structure in the sense that their boundaries could prevent the excessive progression of active and repressive chromatin marks along the chromosomes (Fig. 3C) (Dixon et al. 2012; Nora et al. 2012; J. Wang et al. 2014; Narendra et al. 2015). The N-terminal tails of histone proteins that protrude out from the nucleosome core particle can be post-translationally modified by several enzymatic complexes, with important implications on chromatin compaction and gene activity (Bannister, Falcão, and Castelo-Branco 2017). Among the best described histone modifications are the trimethylation of histone H3 at the lysine residue K4 (H3K4me3), which is associated with active and poised genes, and trimethylation of the residue K27 of the same histone protein (H3K27me3), which marks facultative heterochromatin including poised and repressed genes (Bannister and Kouzarides 2005; Brookes et al. 2012; Li et al. 2018). Early evidences of topological boundaries being able to prevent the spread of histone marks in the genome come from genome-wide studies showing that domains marked by H3K27me3 abruptly stop at regions that were later shown to coincide with such boundaries (Barski et al. 2007; Cuddapah et al. 2009; Nora et al. 2012). Moreover, it was shown that the deletion of boundary-associated elements can result in the spread of either the active mark H3K4me3 (Narendra et al. 2015) or the repressive mark H3K27me3 (Luo et al. 2018) into the adjacent chromatin domain, correlating with increased or reduced expression of the nearby genes, respectively. Although these results suggest an additional way by which TAD borders could participate in the functional compartmentalization of the genome, the contribution of enhancer blocking activity to histone modifications landscapes and gene expression status was not addressed in the aforementioned studies. Hence, the relationship between TAD boundaries and histone modifications landscapes might be indirect. For instance, the spread of the histone mark H3K4me3 could be due to boundary-located genes becoming activated as a result of novel enhancer-promoter interactions (Narendra et al. 2015). Furthermore, as regards the H3K27me3 mark, it should be noted that

contradictory results were obtained after acute depletion of one of the major proteins that function at boundary elements, CTCF (Nora et al. 2017). Nevertheless, the absence of H3K27me3 spreading in these conditions could be due to an assessment within only a relatively short time period after the onset of depletion; a technical limitation associated with the fact that the CTCF protein is essential for cell survival (Sleutels et al. 2012; Nora et al. 2017). Indeed, such a timeframe could possibly not be sufficient for the spread to manifest, in contrast to situations in which boundary elements were stably deleted. Furthermore, this study was performed on embryonic stem cells and therefore could not capture the influence of CTCF in the dynamic changes in histone modifications landscapes that occur during differentiation, among which the expansion of H3K27me3-covered regions (Hawkins et al. 2010; Wang et al. 2014; Nora et al. 2017).

In addition, TADs would restrain the elongation of non-coding transcription along the chromatin, hence concomitantly contributing to the fine tuning of gene expression programs (Fig. 3D) (Austenaa et al. 2015). It is now well established that RNA polymerase II (RNAPII) can achieve bidirectional and upstream (i.e. 5') non-coding transcription at enhancers and gene promoters, respectively; a process that is thought to contribute to the activity of these regulatory regions (reviewed in Andersson and Sandelin 2020). Under physiological conditions, the non-coding transcripts that are thereby produced are usually short and are rapidly degraded (T.-K. Kim et al. 2010; Henriques et al. 2018). However, if transcription termination is defective, RNAPII has the potential to elongate transcription over long genomic distances, matching with an upregulation of the genes that are encountered. In this context, it was shown that such abnormal transcription elongation events are nevertheless constrained within the limits of TADs (Austenaa et al. 2015). Hence, these results suggest that TAD boundaries provide an additional layer of safety in preventing the spread of non-coding transcription and its potentially deleterious effects on fine-tuned gene expression programs along the genome.

Another important nuclear process that is thought to be influenced by TADs, or TAD-like structures, consists in the somatic rearrangements that generate receptors for antigen recognition during lymphocytic maturation in the adaptive immune system. In jawed vertebrates, T and B lymphocytes undergo irreversible somatic recombination of the loci encoding T cell receptors (TCR) and immunoglobulins (Ig), respectively (Schuurman and Quesniaux 1999). This represents the entry point in a series of events that ultimately gives rise to the impressive diversity of the antigen receptor repertoire, which allow the high

specificity and efficiency of adaptive immune responses against virtually any pathogen. Importantly, somatic recombination consists of the random retention and deletion of different possible DNA segments in the antigen receptors loci, known as V, D and J segments (reviewed in Nishana and Raghavan 2012). During this process, the DNA double-strand breaks are achieved by the Recombination Activating Genes protein complex RAG1/2. It was shown that RAG-mediated double-strand breaks for V(D)J rearrangements predominantly occur within TAD-like chromatin domains of about 2 Mb and mostly happen in the same orientation, suggesting a linear tracking mechanism by the RAG1/2 complex that would be restricted to the limits of the topological domain (Hu et al. 2015). Interestingly, previous studies also reported a correlation between such domains and the occurrence of Activation-induced cytidine deaminase (AID)-mediated double-strand breaks that lead to antibody class-switch recombination (CSR), which converts IgM antibodies to other types of soluble immunoglobulins such as IgG (Zarrin et al. 2007; Dong et al. 2015). Taken together, these results strongly suggest a role for TAD-related structural scaffolds in constraining the endonuclease activity of RAG and deaminase activity of AID to specific chromatin regions, thus preventing deleterious ectopic genomic rearrangements that could participate in lymphoma development (Dong et al. 2015; Hu et al. 2015).

It is noteworthy that, even though the aforementioned effects could appear to be unrelated at first sight, they are likely to arise from a single generic mode of action in which TAD boundaries would block the linear tracking of molecules beyond the limits of the domain. In this view, the processive travelling of either histone modifying enzymes, RNA polymerase II or recombination-inducing factors would be impaired at the level of TAD boundaries, possibly because they are densely occupied by architectural proteins (Dixon, Gorkin, and Ren 2016).

TADs were also proposed as important units for the control of DNA replication timing (Moindrot et al. 2012; Pope et al. 2014). The genome is segmented into regions displaying distinct replication timings, with some initiating replication earlier (early replication domains) than others (late replication domains) during the S phase of the cell cycle (Raghuraman et al. 2001; Yaffe et al. 2010; McGuffee, Smith, and Whitehouse 2013). It was shown that the size of these replication domains resembles that of TADs and that their limits closely correspond to the positions of TAD boundaries (Moindrot et al. 2012; Pope et al. 2014). Moreover, active TADs (as defined by transcription and histone modifications) often coincide with early-replicating domains, whereas inactive or repressed domains tend to initiate replication later (Pope et al. 2014). Finally, while TADs transiently disappear during mitosis, they are rapidly re-established

during early G1, which corresponds to the time when replication domains are established (Dileep et al. 2015; Nagano et al. 2017; H. Zhang et al. 2019). Taken together, these results suggest that the structure and internal characteristics of TADs might be important for controlling the dynamics of replication initiation and delimiting domains of distinct replication timings from one another (Moindrot et al. 2012; Pope et al. 2014). Nevertheless, this view is challenged by the fact that neither TAD boundary deletion nor the global depletion or disruption of TADs alter the timing of replication (Sima et al. 2019; Ma and Duan 2019; Oldach and Nieduszynski 2019; reviewed in Marchal, Sima, and Gilbert 2019). In particular, the fact that active TADs match with domains of early replication is likely to be the consequence of these domains containing elements known as ERCs (early replication control elements), which were shown to influence both replication timing and transcriptional status within a domain. At the same time, ERCs can participate in the structure of TADs and in their allocation to the active (A) compartment (Sima et al. 2019). Hence, these findings support a model in which ERCs simultaneously affect replication, transcription and structural features in the genome rather than a direct implication for TADs themselves in controlling replication dynamics.

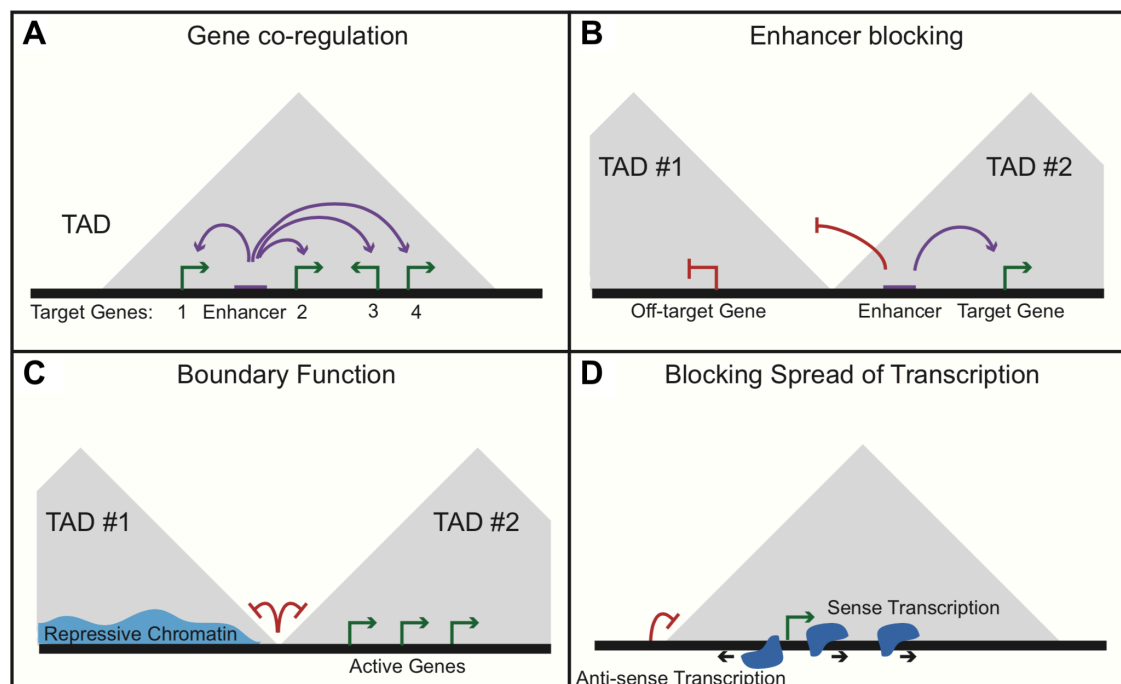


Figure 3: Modified from (Dixon, Gorkin, and Ren 2016). Proposed roles for TADs and their boundaries in the functional compartmentalization of the genome. (A) A TAD (gray pyramid) could favor regulatory interactions between enhancers and their target genes. In this example, several genes (green bended arrows) are under the control of the same enhancer (purple box). (B) Conversely, TADs can prevent abnormal communication between an enhancer and an off-target gene thanks to their localization in different TADs. (C and D) TADs could also limit the spread of chromatin marks (C) or non-coding transcription (D) along the genome.

4. Principles of chromatin folding

Over the past few decades, a number of polymer models were developed with the aim of capturing the behavior and organization of the genomic DNA. Below is presented a short historical overview of these models, ranging from the simplest to the most realistic, together with recent evidences that contribute to our understanding of the biophysics of chromatin organization and gene regulation.

4.1 Of self-avoidance

By essence, the DNA double-helix is a flexible polymer, meaning that its nucleotidic constituents are constrained in their movements due to their link with one another in the chain. One of the simplest ways to model the behavior of such a polymer is the so-called Brownian self-avoiding walk (SAW) (Doi and Edwards 1988; Binder and Heermann 2010). In this view, due to the molecular agitation (also known as Brownian motion) of its constituents and given the attachment of the latter along the chain, the DNA polymer is randomly moving and stretching within the 3D space while avoiding interactions with itself at a relatively small scale.

Nevertheless, this fundamentally random motion of the polymer, even though it is constrained by self-avoidance, is unable to reproduce on its own the features of genome organization observed in living cells, such as specific loops and invariant domains of increased interaction frequencies (Nicodemi and Prisco 2009).

4.2 When binders come into play

In order to get closer to reality, other players need to be taken into consideration. Indeed, the DNA is not alone in the nucleus, but is surrounded by proteins and RNA molecules that can affect its behavior. Among this range of particles, there are soluble DNA-binding factors that will be referred to as “binders” for sake of simplicity (Nicodemi and Prisco 2009; Barbieri et al. 2012).

The model described hereafter (Nicodemi and Prisco 2009), which was later termed the “strings and binders switch” (SBS) model (Barbieri et al. 2012), assumes a system constituted by a DNA polymer and binders (which are subjected to Brownian motion as well) that can dynamically bind specific sites present multiple times along the polymer. Importantly, the binders can attach to more than one site at a time, either directly given the presence of several DNA-binding domains on the binder itself, or indirectly thanks to adaptors that can bridge several binders, each of them attached to a single site on the polymer,

together in space (Nicodemi and Prisco 2009; Barbieri et al. 2012). Thereby, binders would be *a priori* able to bring several sites of the polymer in close spatial proximity.

Indeed, it was shown by *in silico* Monte Carlo simulations that in this system, above a threshold concentration of binders (C_{tr}) (Fig. 4) or affinity for their binding sites, the binders are able to organize the polymer into dynamic loops or stable densely folded domains (Nicodemi and Prisco 2009; Barbieri et al. 2012).

In particular, SBS-like models appear to be particularly well suited to explain the condensation of chromosomal compartments, which are supra-TAD chromatin structures (see further for details on compartments and their possible condensation by phase separation) (Erdel and Rippe 2018).

4.3 Putting the SBS model to the test

In 2012, the SBS model was directly tested against experimental data of chromatin folding (Barbieri et al. 2012). One crucial aspect of genome architecture is that different levels of compaction, from highly open euchromatin to very densely packed heterochromatin, can co-exist within the nucleus (Woodcock and Ghosh 2010). It was shown that the SBS model can indeed recapitulate a large proportion of the different possible folding states of the chromatin observed in cells by playing with the concentration of binders (C_m) or the number of their binding sites along the polymer (Fig. 4) (Barbieri et al. 2012).

However, the SBS model lacks some crucial aspects required to faithfully reflect chromatin folding at the TAD level. Indeed, it does not take into account that the binding sites of key factors involved in chromatin architecture (such as the CTCF protein) are not uniformly distributed along the DNA polymer (Dixon et al. 2012; Barbieri et al. 2012). Instead, as previously mentioned, these sites are actually enriched and cluster at the border between TADs, while also displaying particular orientations (Dixon et al. 2012; Zhan et al. 2017). Because these features are thought to be fundamental for the positioning of TAD boundaries and loop anchors, another model has been developed to take them into account (see the loop extrusion model below) (Sanborn et al. 2015; Fudenberg et al. 2016).

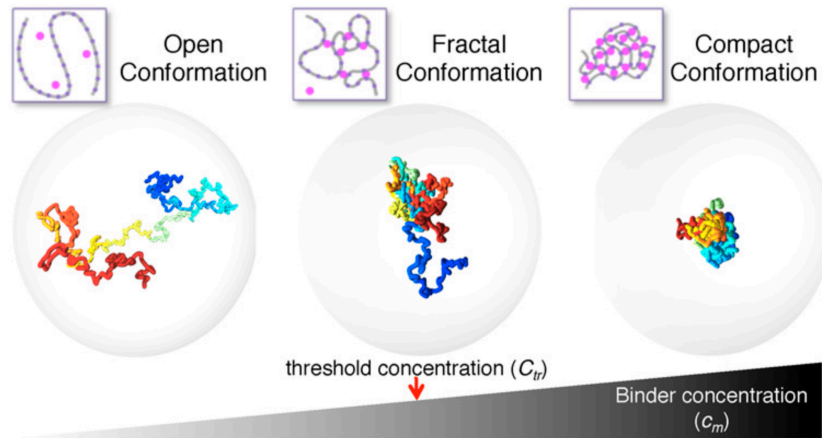


Figure 4: After (Barbieri et al. 2012). Simulations of the SBS model recapitulate a wide range of chromatin folding levels as a function of the concentration of binder molecules (C_m). The “switch” occurs at the threshold concentration (C_{tr}).

4.4 The loop extrusion model

To date, the prevailing model to explain the formation of chromatin loops and TADs is the so-called loop extrusion model (Fig. 5) (Sanborn et al. 2015; Fudenberg et al. 2016). Since the early characterization of TADs, the observed enrichment of most of their boundaries for the binding of a protein called CCCTC-binding factor (CTCF) (Dixon et al. 2012) raised the possibility for this factor to participate in the formation of the latter.

The CTCF protein was originally described as an evolutionary conserved transcriptional regulator, acting either as an activator or repressor of gene expression depending on the context, possibly based on varied DNA-binding configurations and interactions with cofactors (Klenova et al. 1993; Filippova et al. 1996). A study investigating enhancer blocking activity of insulator elements later showed that this protein is also the major insulator-associated factor in vertebrates (Bell, West, and Felsenfeld 1999). CTCF has non-palindromic binding sites onto the DNA, implying that they can display two different possible orientations. Importantly, chromatin loops and topologically associating domains are often delimited by pairs of convergently oriented CTCF sites; i.e. by sites that are facing each other onto the chromatin (Sanborn et al. 2015; Fudenberg et al. 2016).

The loop extrusion model was originally proposed by Kim Nasmyth in 2001 to explain the process of chromatin condensation resulting in the production of metaphase chromosomes (Nasmyth 2001). This model was later adapted to account for the formation of chromatin loops and TADs during interphase

(Sanborn et al. 2015; Fudenberg et al. 2016). The loop extrusion model assumes a system in which the so-called loop extruding factors pull a stretch of chromatin, making each extremities of an initial loop pass progressively through the extrusion complex, thereby making the loop grow in size (Nasmyth 2001; Sanborn et al. 2015; Fudenberg et al. 2016). Based on the aforementioned evidences about CTCF, the “interphase version” of the model considers CTCF sites as typical barriers to the loop extrusion process. In particular, the extrusion complex would stop when reaching CTCF sites, possibly due to protein-protein interactions between the extrusion machinery and a pair of DNA-bound CTCF proteins, preferentially when they are encountered in a convergent orientation (Fig. 5) (Sanborn et al. 2015; Fudenberg et al. 2016). In addition, a forming loop might be stalled when reaching the basis of another loop, resulting in the appearance of multiple neighboring, but non-overlapping looped structures (Fudenberg et al. 2016). CTCF-based loops would then disassemble after a certain time (estimated between 1 and 2 hours), as their constituents get released from the DNA (Hansen et al. 2017; Vian et al. 2018).

In this context, the most strongly supported loop extruder is the cohesin multiprotein complex (Fudenberg et al. 2017). Cohesin is a ring-shaped complex comprised of three major subunits: two Structural Maintenance of Chromosomes (SMC) proteins, SMC1 and SMC3, which form most of the ring, and a protein called RAD21 that closes the structure (reviewed in Gligoris and Löwe 2016; Yuen and Gerton 2018). The cohesin complex was first described for its key function in sister chromatid adhesion before anaphase during cell division (Michaelis, Ciosk, and Nasmyth 1997; Uhlmann, Lottspeich, and Nasmyth 1999; reviewed in Peters and Nishiyama 2012). It was latter shown that this complex is also able to extrude DNA loops in biochemical reconstruction experiments in the presence of ATP (Davidson et al. 2019), directly interacts with the CTCF protein *in vitro* (Y. Li et al. 2020) and physically binds to most CTCF sites on the chromatin in a CTCF-dependent manner *in vivo* (Parelho et al. 2008; Wendt et al. 2008; Busslinger et al. 2017). In addition, mutating the direct cohesin-binding interface on CTCF results in an almost complete loss of CTCF-based loops *in vivo*, strongly suggesting that its interactions with cohesin are required for either the establishment or the lifespan of these structures (Y. Li et al. 2020).

The cohesin-CTCF loop extrusion model is itself supported by several lines of evidence. First, *in silico* simulations of the model give rise to domains of increased interaction frequencies and faithfully set the position of their boundaries as compared to experimental data (Sanborn et al. 2015; Fudenberg et al. 2016). Second, targeted deletions of boundary-associated CTCF sites result

in a shift of the TAD border to the next CTCFs in convergent orientation relative to those of the opposite boundary (Sanborn et al. 2015; Fudenberg et al. 2016). Third, depletion of cohesin or CTCF have distinct effects on chromatin architecture. Whereas the loss of cohesin results in the weakening or loss of TADs genome-wide (characterized by decreased intra-domain interactions) (Zuin et al. 2014; Rao et al. 2017), similar to what is observed upon deletion of the cohesin loading factor NIPBL (Schwarzer et al. 2017), it does not result in increased inter-domain interactions. In contrast, the depletion of CTCF mostly causes an increase of inter-TAD contacts (Zuin et al. 2014; Nora et al. 2017). Finally, deletion of the cohesin release factor WAPL increases the size of CTCF-anchored loops and thereby promotes inter-TAD contacts, consistent with a dynamic cohesin-mediated extrusion mechanism, and considering the much faster binding/release kinetics of CTCF compared to those of cohesin (Haarhuis et al. 2017; Hansen et al. 2017).

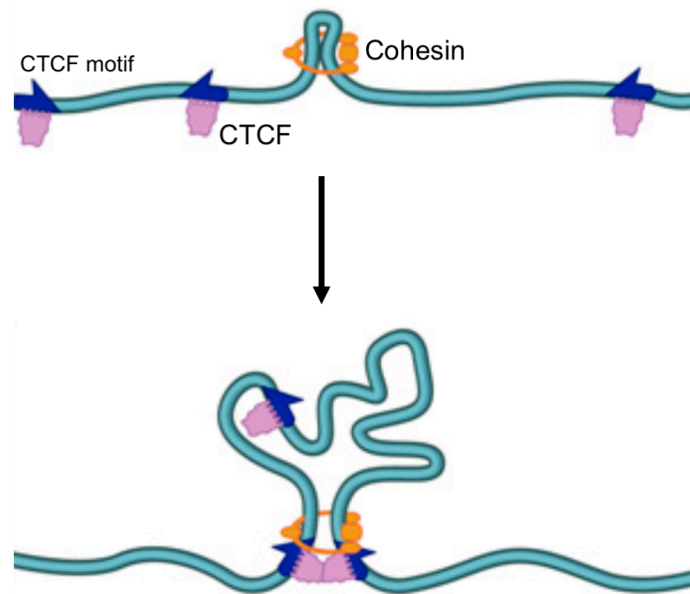


Figure 5: Modified from (Sanborn et al. 2015). Overview of the cohesin-CTCF loop extrusion process. After loading, the cohesin complex (yellow ring) extracts a stretch of chromatin (cyan string) and stops when reaching genomic sites bound by the CTCF protein (blue arrows and purple globules, respectively) in a convergent orientation, thereby forming a looped structure. If it is encountered, the basis of such a loop may also cause an adjacent forming loop to stall (also see Fudenberg et al. 2016).

4.5 Intra-TAD dynamics and relationships with compartments

As briefly mentioned above, TADs are generally maintained among cell types, meaning that the position of their boundaries is overall stable (Dixon et al. 2012; Nora et al. 2012; Dixon et al. 2015). Nevertheless, the internal organization

of these domains undergoes extensive changes during cell differentiation. Importantly, this gives rise to cell-type-specific intra-domain structures that were shown to correlate with the establishment of specific gene expression programs (Dixon et al. 2012; Nora et al. 2012; Dixon et al. 2015; Bonev et al. 2017).

The basis of cell-type or tissue-specific intra-TAD loops is highly enriched for CTCF sites showing convergent orientation (Phillips-Cremins et al. 2013; Beagan et al. 2017), raising the possibility of differential intra-TAD CTCF binding to be responsible for regulating the formation of these structures. In line with this hypothesis, important evolutionary divergence in the binding of CTCF at intra-TAD sites is associated with intra-domain variations between mammalian species, whereas CTCF binding is very well conserved at the level of TAD borders (Vietri Rudan et al. 2015; Lazar et al. 2018). Furthermore, cell-type-dependent dynamic changes in CTCF occupancy have been reported, with about 80% of sites showing variable CTCF binding among 40 different cell types in the largest study (H. Wang et al. 2012; Maurano et al. 2015; Beagan et al. 2017), strongly contrasting with previous reports of overall constitutive CTCF binding (T. H. Kim et al. 2007; Cuddapah et al. 2009; Y. Shen et al. 2012). Finally, it has been observed that some specific enhancer-promoter interactions can indeed be lost during the transition from a pluripotent to a more differentiated state concomitant with the loss of CTCF binding at the basis of the loops (Beagan et al. 2017; Bonev et al. 2017). Nevertheless, many cell-type or tissue-specific intra-TAD structures were shown not to be anchored by sites displaying cell-type-dependent CTCF binding, but by sites that are ubiquitously bound by CTCF (Phillips-Cremins et al. 2013; Beagan et al. 2017; Hanssen et al. 2017). Hence, differential CTCF binding cannot explain on its own the whole range of intra-domain architectural variation between cell types, and therefore other mechanisms need to participate in controlling the formation of these variable structures.

Interestingly, recent studies reported that a subset of contacts depends on the RNA-binding properties of the CTCF protein, suggesting another mean by which contacts could be modulated in different cells (Hansen et al. 2019; Saldaña-Meyer et al. 2019). In particular, about half of CTCF-mediated interactions in the genome were shown to depend on the interactions between CTCF and a variety of RNAs through regions of the protein that are distinct from the one responsible for its association with the DNA (Saldaña-Meyer et al. 2014; Hansen et al. 2019; Saldaña-Meyer et al. 2019). These studies also found that CTCF self-associates in an RNA-dependent manner, and that this behavior is

abolished upon deletion of the RNA-binding regions of the protein (Saldaña-Meyer et al. 2014; Hansen et al. 2019; Saldaña-Meyer et al. 2019). These results provide compelling evidences of a direct role for RNAs in favoring CTCF self-association, which then in turn affects CTCF function in some instances. Importantly, CTCF-RNA interactions were shown not to be required for the association of CTCF with the DNA in all cases and, upon deletion of the RNA-binding regions of CTCF, some chromatin loops can be lost even if the binding of CTCF on the DNA remains unchanged (Saldaña-Meyer et al. 2019). Hence, differential expression of RNAs controlling the behavior of CTCF could explain intra-TAD structural modulations between cell types without requiring changes in the binding of CTCF to the DNA.

The establishment, reinforcement or loss of interactions between gene promoters and *cis* regulatory elements can participate in the variation of intra-TAD structures between different cell types (Phillips-Cremins et al. 2013; Smith et al. 2016; Bonev et al. 2017), although the exact relationships between TADs, CTCF-mediated chromatin loops and enhancer-promoter interactions remain a matter of intense debate. Indeed, whereas some enhancer-promoter interactions are constitutively anchored by CTCF, as exemplified by the *Sonic hedgehog* (*Shh*) locus (Paliou et al. 2019; Williamson et al. 2019), others appear to be CTCF/cohesin-independent (Rao et al. 2017), strongly suggesting that they do not rely on structural scaffolding by TADs. Conversely, at least in some instances, topologically associating domains can form independently from enhancer-promoter interactions, as demonstrated by the preservation of a domain and of most of its internal organization upon removal of the major genes or enhancers it contains (Brown et al. 2018).

Beyond their internal dynamics, TADs also display cell-type-dependent allocation to chromosomal compartments (Lieberman-Aiden et al. 2009; Dixon et al. 2015). Compartments, which appear as off-diagonal checkerboard patterns in Hi-C heatmaps, arise from interactions occurring between multiple chromatin regions matching with TADs that share similar chromatin opening levels, histone modifications status and transcriptional activity (Lieberman-Aiden et al. 2009; Dixon et al. 2012; Rao et al. 2014; reviewed in Eagen 2018). The genome is subdivided in two major compartments, called A and B. Whereas the A compartment is formed by open and transcriptionally active chromatin, the B compartment tends to be more closed and transcriptionally silent (Lieberman-Aiden et al. 2009; Sivakumar, de las Heras, and Schirmer 2019). Importantly, a study comparing five different cell types including embryonic stem cells revealed that about 36% of all genomic regions switch from one compartment to the

other in at least one of the cell types under investigation. Moreover, most of the observed transitions concerned entire TADs and matched with transcriptional changes of the genes they contain; with either an upregulation of gene expression for B to A transitions, or a downregulation for A to B transitions (Dixon et al. 2015).

Although they are hierarchically interlinked, TADs and compartments are believed to be formed by distinct mechanisms; namely cohesin-driven loop extrusion (as explained above) and segregation of regions displaying different chromatin states from each other, respectively (Dixon et al. 2012; Fudenberg et al. 2017; Schwarzer et al. 2017). Indeed, while depleting cohesin from the chromatin results in the abolition of TADs, compartments do not disappear, demonstrating that they do not require TADs to form. On the contrary, compartments actually appear to be more pronounced under those conditions, indicating a competition in the formation of different layers of genome organization (Rao et al. 2017; Schwarzer et al. 2017; also see Haarhuis et al. 2017). In contrast, CTCF depletion does not result in compartmental reinforcement, suggesting that the act of cohesin-mediated extrusion itself competes with compartmental segregation, probably via the inclusion of segments displaying different chromatin states within the same loops (Nora et al. 2017; Mirny, Imakaev, and Abdennur 2019). Strongly supporting this notion, computer simulations of polymer models recapitulated interference between active loop extrusion and compartment-like segregation (Rao et al. 2017; Schwarzer et al. 2017; Nuebler et al. 2018). To date, surprisingly little is known about the identity of the molecular players underlying the formation of compartments, but there are mounting evidences that compartmental segregation is related to phase separation (see further) (Nora et al. 2017; Mirny, Imakaev, and Abdennur 2019).

4.6 Phase separation in genome's shape and gene expression control

Apart from chromatin looping, other molecular processes can lead to the establishment of higher-order features of genome organization and participate in the regulation of gene expression. In particular, there is a growing body of evidence that phase separation can drive transcriptional activation as well as heterochromatin formation (Hnisz et al. 2017; Erdel and Rippe 2018).

Phase separation is a phenomenon in which biomolecules undergo weak and dynamic interactions with one another; a process often referred to as molecular crowding (Machiyama et al. 2017). This results in the formation of supramolecular condensates that look and behave like liquid droplets. Such

condensates have to be distinguished from aggregates in the sense that they are reversible and constantly exchange material with the surrounding medium (Brangwynne et al. 2009; Banjade et al. 2015; reviewed in Bergeron-Sandoval, Safaee, and Michnick 2016). As they are able to concentrate sets of biomolecules, phase separated droplets can typically restrict particular biological processes within distinct locations in the cell without requiring demarcation by a membrane (reviewed in Banani et al. 2017).

In this context, it was shown that the recruitment of transcription factors (TFs) at either single or, more frequently, at clusters of enhancers (known as super-enhancers) (Whyte et al. 2013; Hnisz et al. 2013; reviewed in Peng and Zhang 2018) can induce the formation of phase separated globules or hubs, matching with the transcriptional activation of their target genes (Boija et al. 2018; Shrinivas et al. 2019). The activation domains of a vast majority of TFs are constituted by low complexity amino acid sequences, referred to as low complexity domains or LCDs, which are intrinsically disorganized in terms of secondary structure. This property enables them to engage in weak multivalent associations with one another and can drive the formation of reversible condensates (Chong et al. 2018; Boija et al. 2018). Remarkably, even though they can be formed by different sets of transcription factors depending on the cell type or locus, super-enhancers-based phase separated hubs recruit general transcriptional co-activators, such as the Mediator complex, supporting a model in which the establishment of active hubs can be achieved regardless of the identity of the TFs that are driving their formation (Boija et al. 2018; Shrinivas et al. 2019). In addition, a previous study highlighted direct interactions between the C-terminal domain (CTD) of RNA polymerase II (the enzyme responsible for mRNA production), which is itself disorganized, and the LCDs of transcription factors within droplets of the latter *in vitro* (Kwon et al. 2013). Together with the observations mentioned above, these results provide compelling evidence for hubs to be able to concentrate and stabilize the transcriptional machinery for productive gene expression *in vivo* (Kwon et al. 2013; Chong et al. 2018; Boija et al. 2018; Shrinivas et al. 2019).

Beyond transcriptional hubs, liquid-liquid phase separation has also been implicated in shaping the architecture and controlling the function of the genome at larger scales. In particular, heterochromatin, which is analogous to the B compartment, is thought to be established by phase separation processes involving the Heterochromatin protein 1 α (HP1 α) and the Chromobox protein 2 (CBX2) (Larson et al. 2017; Strom et al. 2017; Erdel and Rippe 2018; Tatavosian et al. 2019; Plys et al. 2019).

HP1 α is one of the major proteins associated with heterochromatin in mammalian cells (Hayakawa et al. 2003; Cheutin et al. 2003). It is homologous to *Drosophila*'s HP1a, which is required for heterochromatin formation and associated gene silencing (Eissenberg et al. 1992; Eissenberg and Elgin 2000). Interestingly, the proper accumulation of the HP1 α protein in heterochromatin depends upon the phosphorylation of its N-terminal tail, which is intrinsically disorganized (Hiragami-Hamada et al. 2011; Nishibuchi et al. 2014; Velez et al. 2015). HP1 α was later shown to oligomerize and form droplets in a manner that also requires phosphorylation of the N-terminal tail. Furthermore, HP1 α is able to compact double-stranded DNA into foci that closely resemble heterochromatin *in vitro* (Larson et al. 2017). A parallel study found that its *Drosophila* homolog HP1a establishes puncta that display all characteristics of liquid droplets *in vivo*, concomitant with heterochromatin formation (Strom et al. 2017). Altogether, these results support a model in which heterochromatin could form through a phase separation process involving the HP1 α protein. However, it should be noted that HP1 α is initially targeted to heterochromatin via direct interactions between its chromodomain (CD) and the histone modification H3K9me3 (Bannister et al. 2001; Lachner et al. 2001; Jacobs et al. 2001; Nielsen et al. 2002). Since H3K9me3 specifically marks constitutive heterochromatin (Nishibuchi and Déjardin 2017), the HP1 α -driven heterochromatin formation model cannot account for the formation of its facultative counterpart (see below).

As previously mentioned, facultative heterochromatin is enriched for the H3K27me3 mark. This histone modification is deposited by Polycomb repressive complex 2 (PRC2) (Chammas, Mocavini, and Di Croce 2020). It is now clear that PRC2 and the H3K27me3 mark form part of a positive feedback loop that also involves the other Polycomb repressive complex, PRC1. In brief, after H3K27me3 deposition by PRC2, PRC1 gets recruited to this mark. This is followed by monoubiquitination of histone H2A on its lysine residue 119 (H2AK119ub), which in turn recruits PRC2 (Aloia, Di Stefano, and Di Croce 2013; Blackledge, Rose, and Klose 2015). This mechanism results in the progressive formation of facultative heterochromatin domains in the genome. The subunits of the canonical PRC1 complexes that recognize the H3K27me3 mark belong to a chromodomain-containing protein family called CBX (Blackledge, Rose, and Klose 2015; Zhen et al. 2016). Of particular interest, one representative of this family, CBX2, was shown to drive PRC1 phase separation both *in vitro* and *in vivo* (Tatavosian et al. 2019; Plys et al. 2019). This potential depends on the intrinsically disorganized region of the CBX2 protein, as demonstrated by the absence of PRC1 droplets

upon mutation of conserved amino acids within this domain (Tatavosian et al. 2019; Plys et al. 2019). Importantly, previous studies found that mutating the same residues of CBX2 results in chromatin compaction defects and developmental alterations (Grau et al. 2011; Lau et al. 2017; Plys et al. 2019). Hence, these findings strongly suggest that similar mechanisms, namely the recruitment and phase separation of chromodomain-containing proteins, can drive the formation of both constitutive and facultative heterochromatin inside the nucleus.

5. Structure and function of the mammalian *Hox* gene clusters

Among the typical targets of Polycomb group proteins is the *Hox* gene family (e.g. Lewis 1978; Margueron and Reinberg 2011; Schorderet et al. 2013). *Hox* genes encode homeodomain-containing transcription factors that are important regulators of embryonic patterning in metazoans (Lewis 1978; Duboule and Dollé 1989; Wellik 2009). These genes were first described in the fruit fly *Drosophila melanogaster*, where their mutations resulted in changes in the identity of body segments; alterations that were termed homeotic transformations, hence the name of these genes (García-Bellido 1977; Lewis 1978; Gehring and Hiromi 1986). *Hox* genes were later shown to be overall conserved in their organization and function from flies to vertebrates (see below) (Duboule and Dollé 1989).

In vertebrates, *Hox* genes are highly organized within gene clusters, meaning that they are densely arrayed along the chromosomes and display the same orientation with respect to one another (Krumlauf 1994; Duboule 2007). As a result of the two rounds of whole-genome duplication that occurred during the evolution of vertebrates (Ohno 1970), there are four different *Hox* genes clusters in mammals, from *HoxA* to *HoxD*, each one located on a different chromosome (Krumlauf 1994).

During embryonic development, multiple different *Hox* gene products are sequentially expressed in partially overlapping domains along the anteroposterior or primary body axis (e.g. Gould et al. 1997; Deschamps and Duboule 2017). Such a segmented and progressive expression of *Hox* genes' combinations is fundamental for the patterning along the primary body axis, a process in which the different regions of the body acquire their respective positional identities (Krumlauf 1994; Kmita and Duboule 2003). HOX proteins perform this task by controlling the expression of multiple genes, which typically encode sets of transcription factors and signaling molecules (Akin and Nazarali 2005; Pearson, Lemons, and McGinnis 2005; Hueber and Lohmann 2008).

Importantly, the relative timing of expression of a given *Hox* gene during development corresponds to its relative physical position along the chromosome; a correspondence termed temporal collinearity (Pascal Dollé et al. 1989; Izpisua-Belmonte et al. 1991; Deschamps and van Nes 2005). Since development proceeds in an anterior-to-posterior fashion in vertebrates, the time of expression onset was proposed to participate in defining the anterior limit of expression of each gene, hence transforming a temporal cue into a spatial one (Duboule 1994; Deschamps and Duboule 2017). Indeed, the relative position of *Hox* genes' expression domains along the body axis also corresponds to their order within the gene clusters; a second correspondence referred to as spatial collinearity (Gaunt, Sharpe, and Duboule 1988). In brief, the situation can be summarized as follows: the farther on the 5' side of a *Hox* cluster a gene is located, the later it will start to be expressed during development and the more posterior will be its expression limit (Dollé et al. 1989; Duboule 1994; Deschamps and van Nes 2005).

The organization of *Hox* genes into clusters is thought to be itself of central importance for the collinear expression of these genes, at least in vertebrates (Gaunt, Sharpe, and Duboule 1988; Duboule 2007). This is for instance supported by the consequences of genetic alterations disrupting this organization regarding *Hox* gene expression (Zákány, Kmita, and Duboule 2004; Tschopp et al. 2009; Darbellay et al. 2019). In particular, it is likely that such a compacted configuration of the gene units has been maintained and consolidated by *cis* regulatory elements located inside or close to the gene clusters, which appear to be responsible for the collinear expression of *Hox* genes along the primary body axis (Spitz et al. 2001; Kmita and Duboule 2003; Deschamps and van Nes 2005; Tschopp et al. 2009). In turn, this robust and highly functional system may have enabled the emergence of long-range *cis* regulations (i.e. outside of the gene clusters) and the concomitant implementation of the *Hox* system to perform other developmental tasks (see long-range regulations at the *HoxD* locus and their implications in the development of secondary axial structures below) (reviewed in Darbellay and Duboule 2016).

5.1 Secondary axes development, long-range regulations and TAD organization at the *HoxD* locus

In addition to their role in implementing positional identities along the primary body axis, *Hox* genes have also been co-opted for the morphogenesis of secondary axial structures of the body, as exemplified by the external genitalia and the limbs (Lonfat et al. 2014; reviewed in Lonfat and Duboule 2015). The

HoxD paralog of the mammalian *Hox* gene clusters is among the systems that have been best studied in this respect (e.g. Dollé et al. 1989; Dollé et al. 1991; Kmita et al. 2002; Tarchini and Duboule 2006; Amândio et al. 2020). For more than 30 years now, the *HoxD* cluster has demonstrated its epistemic value concerning gene regulation in the context of development and, more recently, its intrinsic relationship with 3D chromatin structure (see below) (Lonfat and Duboule 2015; Darbellay and Duboule 2016).

The *HoxD* cluster consists of nine genes (*Hoxd1, 3, 4, 8, 9, 10, 11, 12* and *13*) (see Fig. 6) that are densely arrayed along a 100 kilobases interval on chromosome 2 (see Lonfat and Duboule 2015). Importantly, the regulation of *Hoxd* genes' expression during the development of secondary axes strongly contrasts with the anteroposterior regulation. Indeed, in the former case, it relies on multiple long-range *cis* regulatory elements located within large regulatory landscapes of approximately one megabase that are flanking the gene cluster on each side (Montavon et al. 2011; Andrey et al. 2013; reviewed in Montavon and Duboule 2013; Bolt and Duboule 2020). A similar situation can be observed at the *HoxA* locus (Lehoczky and Innis 2008; Dixon et al. 2012; Berlivet et al. 2013; Woltering et al. 2014; Lonfat et al. 2014), which encodes transcription factors that act in a partially redundant manner compared to those from *HoxD* (Fromental-Ramain et al. 1996; Kmita et al. 2005).

The action of these long-range regulatory sequences needs to be tightly coordinated for the proper spatiotemporal control of *Hoxd* genes' expression (Deschamps and Duboule 2017). In the context of vertebrate limb development, *Hoxd* genes are expressed following a bimodal strategy, first involving both the anterior (from *Hoxd1* to *4*) and the central (from *Hoxd8* to *11*) genes, and later on the central and the posterior (*Hoxd12* and *13*) genes (see below and Fig. 6) (reviewed in Lonfat and Duboule 2015).

In the first phase, enhancers located within the landscape standing on the telomeric side of the gene cluster (i.e. in 3') are activated and trigger the expression of *Hoxd1* to *Hoxd11* from the early limb bud onwards (Andrey et al. 2013; Yakushiji-Kaminatsui et al. 2018). This expression is maintained in the proximal part of the limbs, which will give rise to the arm and forearm (Tarchini and Duboule 2006). During the second or late phase, which is initiated in a very distal cell population of the growing limb bud, the telomeric landscape becomes repressed and gives way to its centromeric (i.e. 5') counterpart (Montavon et al. 2011; Andrey et al. 2013). Thus, the enhancers of the centromeric landscape, now active, drive the activation of *Hoxd9* to *Hoxd13* in the distal aspect of the

developing limb, which will generate the digits in the adult (Kmita et al. 2002; Spitz, Gonzalez, and Duboule 2003; Gonzalez, Duboule, and Spitz 2007; Montavon et al. 2011; Lonfat et al. 2014).

Of particular interest, it was shown that the regulatory landscapes hosting the *cis*-acting elements that control *Hoxd* gene expression during secondary axes development actually coincide with two TADs, called T-DOM (for telomeric domain) and C-DOM (for centromeric domain) (Andrey et al. 2013; Lonfat et al. 2014). Furthermore, the *HoxD* cluster was shown to constitute itself the boundary between these two TADs, since only a deletion of about 400 kilobases comprising the entire gene cluster resulted in the complete fusion of the domains (Rodríguez-Carballo et al. 2017).

Importantly, the *HoxD* cluster is subjected to a dynamic contact switch between the two TADs, meaning that the boundary limit can move along the cluster depending on the regulatory phase and corresponding tissue (Andrey et al. 2013; Rodríguez-Carballo et al. 2017). Hence, the interactions of the central *Hoxd* genes are either biased towards the T-DOM in the context of the proximal regulation or are reoriented towards the C-DOM for the distal regulation (see Fig. 6). In that sense, in addition to the differential activation of the regulatory landscapes (Rodríguez-Carballo et al. 2019), the *HoxD*-associated TADs would provide a proper structural framework favoring regulatory interactions between the *HoxD* cluster and the enhancers located within the flanking landscapes.

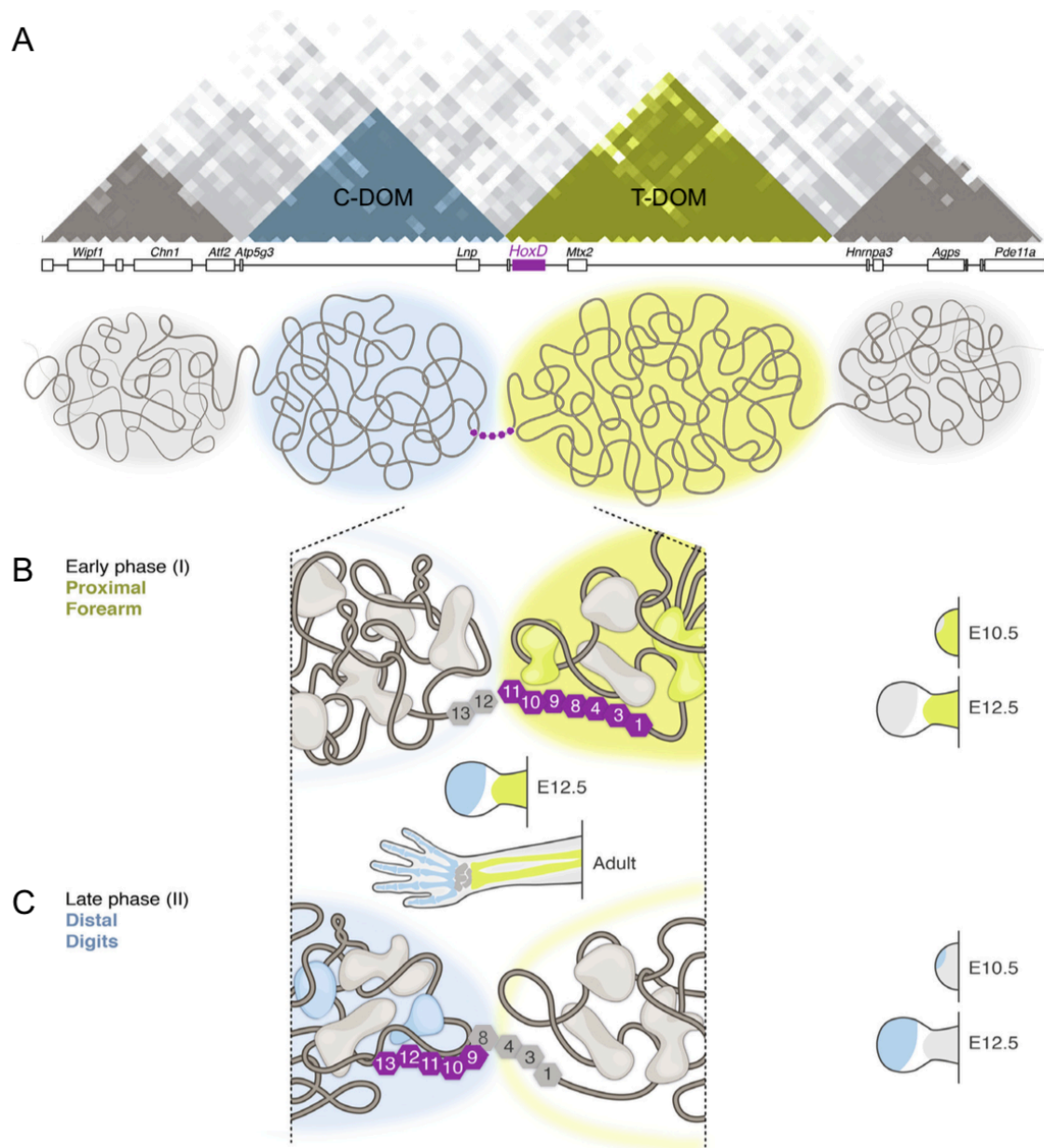


Figure 6: Modified from (Lonfat and Duboule 2015). TAD transition, *HoxD* cluster and limb development. (A) Hi-C heatmap of the *HoxD* locus on chr2. The two *HoxD*-associated TADs are highlighted in yellow (T-DOM) and blue (C-DOM). The position of the *HoxD* cluster at the border between these two TADs is indicated below. A schematic representation of the two TADs is shown at the bottom of panel (A). (B and C) Dynamic contact switch of the *HoxD* cluster from TAD to TAD. (B) During the early phase, which corresponds to the proximal regulation, the T-DOM is activated and triggers the expression of *Hoxd1* to *Hoxd11* in the early limb bud. Note that in this context, most *Hoxd* genes are interacting with the T-DOM. This expression is maintained in the proximal part of the limbs at embryonic day 12.5 (E12.5), which will give rise to the arm and the forearm. (C) During the late phase, the T-DOM is repressed and the central *Hoxd* genes are now interacting with the active C-DOM for the distal regulation. Consequently, genes from *Hoxd9* to *Hoxd13* are expressed in the distal aspect of the limb, which will form the digits in the adult. Active and inactive genes are depicted as purple and gray hexagons, respectively. The corresponding expression patterns in E10.5 limb buds and E12.5 limbs are also indicated.

6. Scope of the thesis

In this study, we focused on chromatin loops and topologically associating domains, as well as their boundaries, as features of interest with respect to genome organization. In particular, we used topological boundaries originating from the *HoxD* developmental locus of the mouse as a paradigm of architectural elements in order to extend the knowledge about their formation.

First, we made use of two intra-cluster duplications increasing the number of boundary elements, including CTCF-binding sites and gene promoters, within the *HoxD* gene complex in order to determine if this could reinforce the TAD boundary *in situ*.

As a second and alternative approach to study the formation of topological borders, we randomly integrated different DNA segments originating from the *HoxD* locus and containing several boundary elements. Our objective was to determine if they would be able to establish contacts within their new surroundings and alter the topological chromatin landscape around the integrations. Typically, we predicted that if these boundary elements are able to function outside of their genomic context, they should be able to disrupt a TAD ectopically.

Results:

1. Partial duplications of the HoxD cluster

As previously explained, the *HoxD* gene cluster is itself conforming a boundary between two topologically associating domains (TADs), called T-DOM and C-DOM (Figs. 6 and 7) (Andrey et al. 2013; Rodríguez-Carballo et al. 2017). Importantly, the *HoxD* cluster displays nine bound CTCF sites, most of which are concentrated at the posterior half or just outside of the gene cluster (from *Hoxd8* to *Evx2*) (Fig. 7). Interestingly, about half of these sites are found within the *Hoxd8-Hoxd11* interval and are mostly oriented towards the T-DOM, whereas the other half stand on the *Hoxd12-Evx2* interval and are all facing the C-DOM (Rodríguez-Carballo et al. 2017). In contrast, the CTCFs located within the *HoxD*-associated TADs are predominantly oriented towards the *HoxD* cluster (see Fig. 7). Hence, the above observations are highly consistent with the loop extrusion model (Sanborn et al. 2015; Fudenberg et al. 2016).

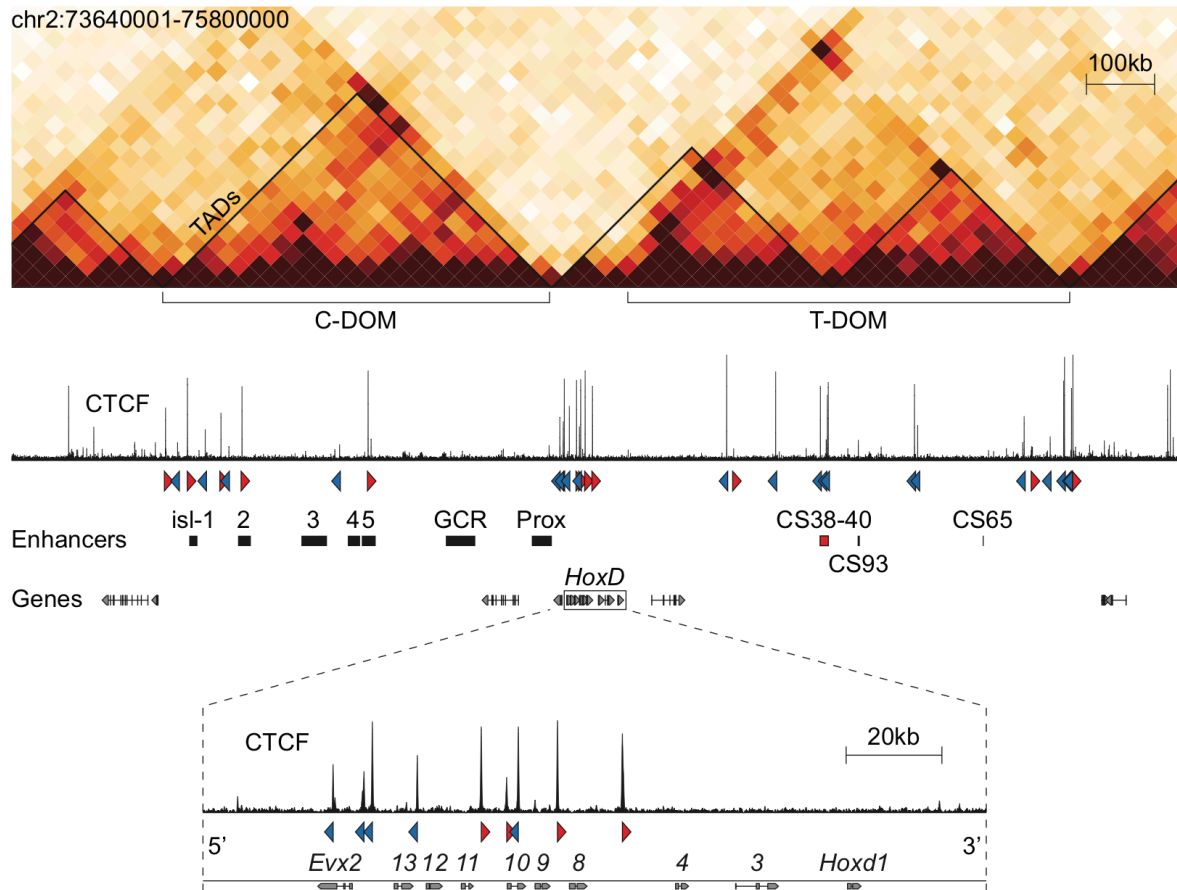


Figure 7: The *HoxD* cluster is a CTCF-rich topological boundary that segregates two TADs. On top is shown a Hi-C of the *HoxD* locus, with its two TADs (appearing as pyramids): T-DOM on the right and C-DOM on the left. Below is shown a CTCF ChIP-seq revealing a very high density of bound CTCF sites at the level of the *HoxD* cluster (orientation shown below by red or blue arrowheads, for forward or reverse sites, respectively). Note that many of the loops of the *HoxD*-associated TADs coincide with bound CTCF sites, particularly the ones showing convergent orientation. Magnification of the *HoxD* cluster revealing that most CTCF sites are concentrated towards the posterior (5') half of the gene cluster. Data are from (Rodríguez-Carballo et al. 2017) and correspond to the situation in the proximal part of *wild-type* forelimbs at embryonic day 12.5 (E12.5).

A previous publication from our group found that partial deletions of the gene cluster within the CTCF and gene-rich interval (i.e. from *Hoxd8* to *Hoxd13*) resulted in increased interactions between the remaining genes and the TAD located beyond the deleted segment, matching the altered expression of these genes expected from such interactions (Rodríguez-Carballo et al. 2017). As an example, the *Hoxd4* gene gained interactions with the C-DOM and became expressed in the distal part of the limbs. Hence, it appears that these alterations decreased the strength of the *HoxD* boundary, probably due to a reduction in the number of its potential boundary elements (namely CTCF-binding sites and active gene promoters).

Here, we first took the reverse approach, and asked whether the TAD boundary at the *HoxD* cluster could be reinforced by increasing the number of potential boundary elements it contains. In order to answer this question, we analyzed two mouse lines carrying partial duplications of the *HoxD* cluster located within the *Hoxd8-Hoxd13* interval that were previously established in our group, namely *Dup(8i-10)* (Figs. 8 and 9) (Tarchini and Duboule 2006) and *Dup(11-13)* (Figs. 10 and 11) (Kmita et al. 2002).

In particular, we predicted that if the strength of the boundary is increased by the duplications, the access of the central *Hoxd* genes to the TAD that is located beyond the duplicated segment should be impaired, whereas interactions established by genes or regulatory regions that are in principle mostly restricted within their respective domains (see *Evx2*, *Hoxd4*, island-2, island-4 and CS38 in Figs. 9 and 11) should remain unchanged or become even more secluded in the latter.

1.1 Duplication of the region 8i-10

In the *Dup(8i-10)* line, the duplication extends 36 kilobases (kb) from the *Hoxd4-Hoxd8* intergenic region to a breakpoint that is located upstream (i.e. in 5') of the *Hoxd10* gene (Tarchini and Duboule 2006). The duplicated segment comprises 5 CTCF-binding sites, among which 4 are oriented towards the T-DOM and one faces the C-DOM, along with 3 gene units (*Hoxd8*, *Hoxd9* and *Hoxd10*) (Fig. 8A).

It was previously shown that the expression of the *Hoxd11* gene is decreased in the proximal part of the limbs at E12 upon duplication of the region 8i-10 (Fig. 8B) (Tarchini and Duboule 2006), suggesting that the interactions established between this gene and the enhancers of the T-DOM (which are active in this tissue) (Tarchini and Duboule 2006; Andrey et al. 2013; Rodríguez-Carballo et al. 2019) might be impaired in this configuration. Importantly, the *Hoxd11* gene stands on the centromeric side of the duplicated segment and is not included in the latter (see Fig. 8A).

To evaluate changes in the interactions established by the *Hoxd11* gene in *Dup(8i-10)*, we performed 4C-seq experiments in E12.5 forelimb samples using the promoter of this gene as a viewpoint. By doing so, we observed a clear loss in the contacts between *Hoxd11* and regions of the T-DOM in the proximal part of E12.5 forelimbs upon duplication. This loss was particularly marked for regions CS38-40, CS93 and CS65 (Fig. 8C; see red arrows). In order to confirm these

results, we also performed a reciprocal assessment by taking CS38 as a 4C-seq viewpoint, which revealed a marked decrease in the contacts extending towards the posterior genes in *Dup(8i-10)* (Fig. 9B). We hypothesized that these structural changes might be due to the multiplication of telomeric-oriented CTCF sites between *Hoxd11* and the T-DOM. Moreover, as the CS39 (included in region CS38-40), CS93 and CS65 elements were previously shown to display strong limb enhancer activity (Andrey et al. 2013; Yakushiji-Kaminatsui et al. 2018), the limited access of *Hoxd11* to the latter could possibly explain the altered expression pattern observed for this gene (see Discussion).

Interestingly, a deeper analysis also revealed that the loss of contacts between *Hoxd11* and regions of the T-DOM correlated with a gain in the contacts established by this gene throughout the C-DOM (from a quantitative perspective, 11% of *Hoxd11* contacts switched from the T-DOM to the C-DOM in *Dup(8i-10)* compared to the *wild-type*; see Fig. 8C). These results suggest that the *Hoxd11* gene redirects part of its contacts towards the centromeric domain as its interactions are impaired on the telomeric side.

In contrast, in the distal part of the limbs at E12.5, the duplication had relatively weak effects on the contacts established between *Hoxd11* and the T-DOM, as expected (Fig. 9A). Indeed, in developing digits of *wild-type* embryos, central *Hoxd* genes such as *Hoxd11* display only few interactions with the T-DOM, as this domain is repressed in the aforementioned tissue (Andrey et al. 2013; Beccari et al. 2016; Rodríguez-Carballo et al. 2019). However, in *Dup(8i-10)* embryos, we could score a slight decrease in the interactions of *Hoxd11* towards region CS38-40 in this tissue, but this could be the result of including a fraction of proximal cells during the dissections.

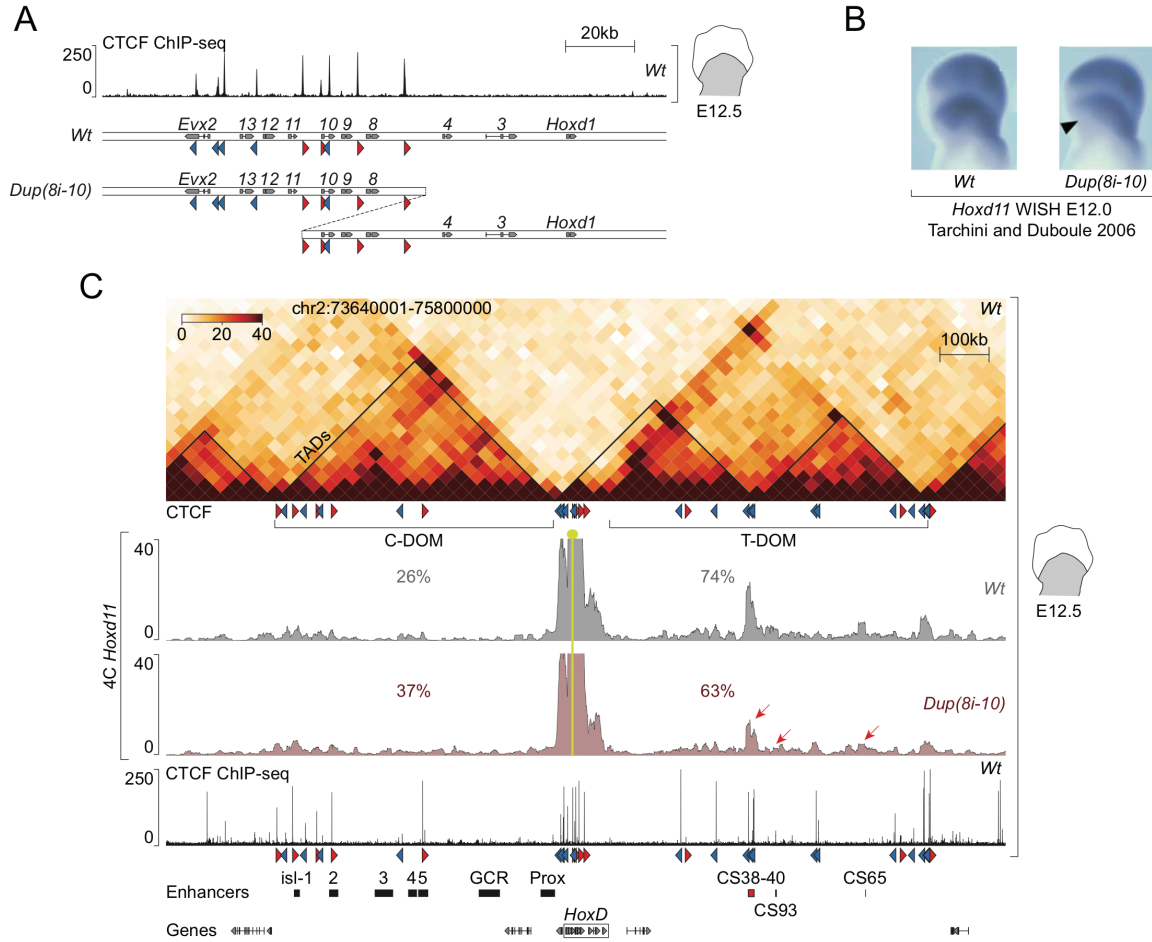


Figure 8: Altered gene expression correlates with structural changes upon duplication of the region *8i-10* (named *Dup(8i-10)*). (A) Configuration of *Dup(8i-10)* (bottom). The *wild-type* configuration and ChIP-seq of CTCF in *wild-type* proximal forelimbs (shaded area) at E12.5 are shown above for comparison. CTCF motifs are represented as red (forward) or blue (reverse) arrowheads. (B) Loss of proximal *Hoxd11* expression in *Dup(8i-10)* revealed by whole-mount *in situ* hybridizations (WISH) in forelimbs at E12.0. (C) Structural changes upon duplication of the region *8i-10*. The *wild-type* track is represented in gray and the *Dup(8i-10)* track in maroon. The region displayed is chr2:73640001-75800000 (mm10). All chromosome conformation capture experiments (Hi-C and 4C-seq) shown here were performed on proximal E12.5 forelimbs. The Hi-C heatmap shown on top is representative of the *wild-type* situation. The TADs (identified by the insulation-based TopDom algorithm with a window size of 240 kb; Shin et al. 2016) are depicted as black lines on the heatmap. Below are shown the results of 4C-seq experiments performed on *wild-type* (n=1) and *Dup(8i-10)* homozygous (n=2) samples. The yellow lollipop represents the position of the *Hoxd11* 4C-seq viewpoint. Red arrows highlight the loss of *Hoxd11* interactions towards regions CS38-40, CS93 and CS65. Quantifications of the contacts established by *Hoxd11* towards either of the two regulatory domains are indicated for both the *wild-type* and *Dup(8i-10)*. Hi-C and ChIP-seq data shown here are from (Rodríguez-Carballo et al. 2017); WISH pictures are modified from (Tarchini and Duboule 2006).

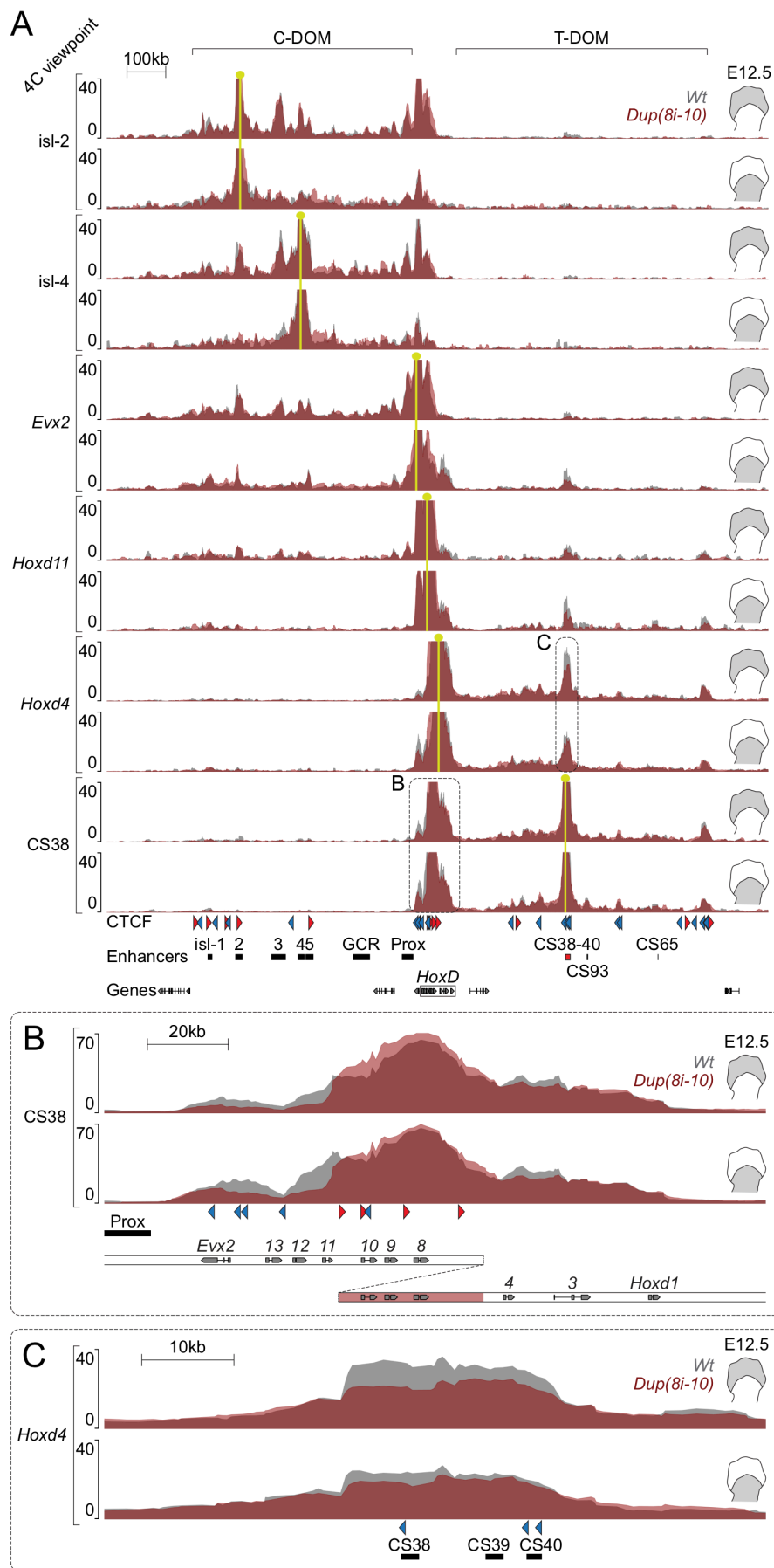


Figure 9: See caption on next page.

Figure 9: Global structural assessment in *Dup(8i-10)*. (A) Interactions established by viewpoints located within the C-DOM (*isl-2* and *4*), outside on the centromeric (*Evx2*) or on the telomeric (*Hoxd4*) side of the *HoxD* TAD boundary, as well as by a viewpoint in the T-DOM (CS38) in distal and proximal E12.5 forelimbs. Interactions were measured by 4C-seq. The *wild-type* tracks are represented in gray and the *Dup(8i-10)* tracks in maroon. *Hoxd11* contacts in both distal and proximal E12.5 forelimbs are also shown. The region displayed is the same as in Figure 8. The position of each 4C-seq viewpoint is represented by a yellow lollipop. Insets are magnified in (B) and (C) and correspond to regions chr2:74628369-74791953 and chr2:75105976-75176976 (mm10), respectively. (B) Contacts of CS38 towards the *HoxD* cluster. Note the decrease in contacts established with posterior *Hoxd* genes (here from *Hoxd11* to *Hoxd13*) in *Dup(8i-10)*. The increased signal over the *Hoxd8-Hoxd10* region is likely the consequence of the duplication of this segment in *Dup(8i-10)*, highlighted in maroon below. (C) Magnification of *Hoxd4* interactions with region CS38-40, which were decreased in *Dup(8i-10)*.

1.2 Duplication of the region 11-13

Based on our observations with *Dup(8i-10)*, we rationalized that if the strength of the boundary is increased upon duplication, we should observe a similar trend when taking the opposite view. This means that there should be a decrease in the contacts established towards the centromeric domain when using as a viewpoint a central *Hoxd* gene located on the telomeric side of an intra-cluster duplication.

To that end, we made use of another line called *Dup(11-13)* (Kmita et al. 2002), in which a segment of 22.5 kb including only one centromeric-oriented CTCF-binding site together with three gene units (*Hoxd11*, *Hoxd12* and *Hoxd13*) is duplicated (Fig. 10A).

A previous study from our group showed that the expression of the central gene *Hoxd10* is decreased in the distal part of E12.5 forelimbs in the duplication *11-13* (Fig. 10B) (Kmita et al. 2002). Knowing that the enhancers of the C-DOM are responsible for the expression of *Hoxd* genes in the distal portion of the limbs (Spitz, Gonzalez, and Duboule 2003; Gonzalez, Duboule, and Spitz 2007; Montavon et al. 2011; Lonfat et al. 2014), we hypothesized that a loss of contacts between *Hoxd10* and the enhancers of this domain could be responsible for the altered expression of the gene. Therefore, we decided to investigate in this direction.

To avoid possible effects on contact detection caused by the proximity between the 4C-seq viewpoint and the duplicated segment, we decided to use the promoter of a gene also located on the telomeric side of the duplication, but farther away from it than *Hoxd10*. We note that the gene to be chosen also had

to establish contacts towards both TADs in developing limbs and display a very similar regulation compared to that of *Hoxd10*. For this reason, we chose the promoter of *Hoxd9* rather than of *Hoxd10* as a viewpoint in *Dup(11-13)* (Kmita et al. 2002; Tarchini and Duboule 2006; Andrey et al. 2013). A similar strategy could not be carried out for *Dup(8i-10)*, due to the fact that the next gene farther on the centromeric side is *Hoxd12*. Importantly, based on its expression pattern, which is closer to that of *Hoxd13* than of *Hoxd11*, *Hoxd12* is thought to establish strong constitutive interactions with regions of the C-DOM that may prevent this gene from contacting the telomeric domain, similar to what is observed for the *Hoxd13* gene (Tschopp and Duboule 2011; Andrey et al. 2013; Rodríguez-Carballo et al. 2017).

By using *Hoxd9* as a 4C-seq viewpoint, we observed a marked decrease in the contacts established between this gene and the centromeric domain in the distal part of the limbs at E12.5 upon duplication. Remarkably, the loss of contacts concerned almost all known enhancers of the C-DOM; namely island-1 to 5 and Prox (Fig. 10C; see red arrows) (Gonzalez, Duboule, and Spitz 2007; Montavon et al. 2011; Lonfat et al. 2014); three of which display CTCF sites that are oriented towards the *HoxD* cluster (island-1, 2 and 5) (Rodríguez-Carballo et al. 2017). A similar effect is likely to be responsible for the previously observed loss of *Hoxd10* expression in the distal portion of the limbs (Kmita et al. 2002), as assessing the contacts of island-2 and 4 in presumptive digits revealed a decrease with a region including *Hoxd9* and *Hoxd10* (Fig. 11B).

Interestingly, the loss of contacts between *Hoxd9* and the C-DOM correlated with a gain of contacts between this gene and region CS38-40 on the telomeric side (Fig. 10C; see green arrow). Furthermore, quantifications of the contacts established with the whole regulatory domains revealed that about 16% of *Hoxd9* interactions were reoriented from the C-DOM to the T-DOM in *Dup(11-13)* compared to the *wild-type* (see Fig. 10C). Hence, these results indicate an increased insulation upon duplication resulting in contact redirection towards the TAD that is still accessible to the *Hoxd9* gene, in this case the T-DOM (see Discussion).

A complementary assessment of *Hoxd9* interactions in the proximal part of the limbs revealed almost no contacts with the C-DOM in both the *wild-type* and upon duplication of the region 11-13, consistent with this domain being inactive and separated from the central genes in this context (Andrey et al. 2013; Beccari et al. 2016; Rodríguez-Carballo et al. 2017; 2019). *Hoxd9* also gained interactions with region CS38-40 in the proximal limb portion upon duplication, although this

effect was less pronounced in this tissue compared to distal forelimbs (Fig. 11A). It is noteworthy that this is unlikely due to the formation of a novel CTCF-based loop between the duplicated segment and region CS38-40, as no telomeric-oriented CTCF is duplicated in *Dup(11-13)* (see Fig. 10A). Therefore, these results further support a model in which the increased insulation between *Hoxd9* and the C-DOM, caused by the duplication, would result in a contact switch by *Hoxd9* towards the T-DOM on the other side.

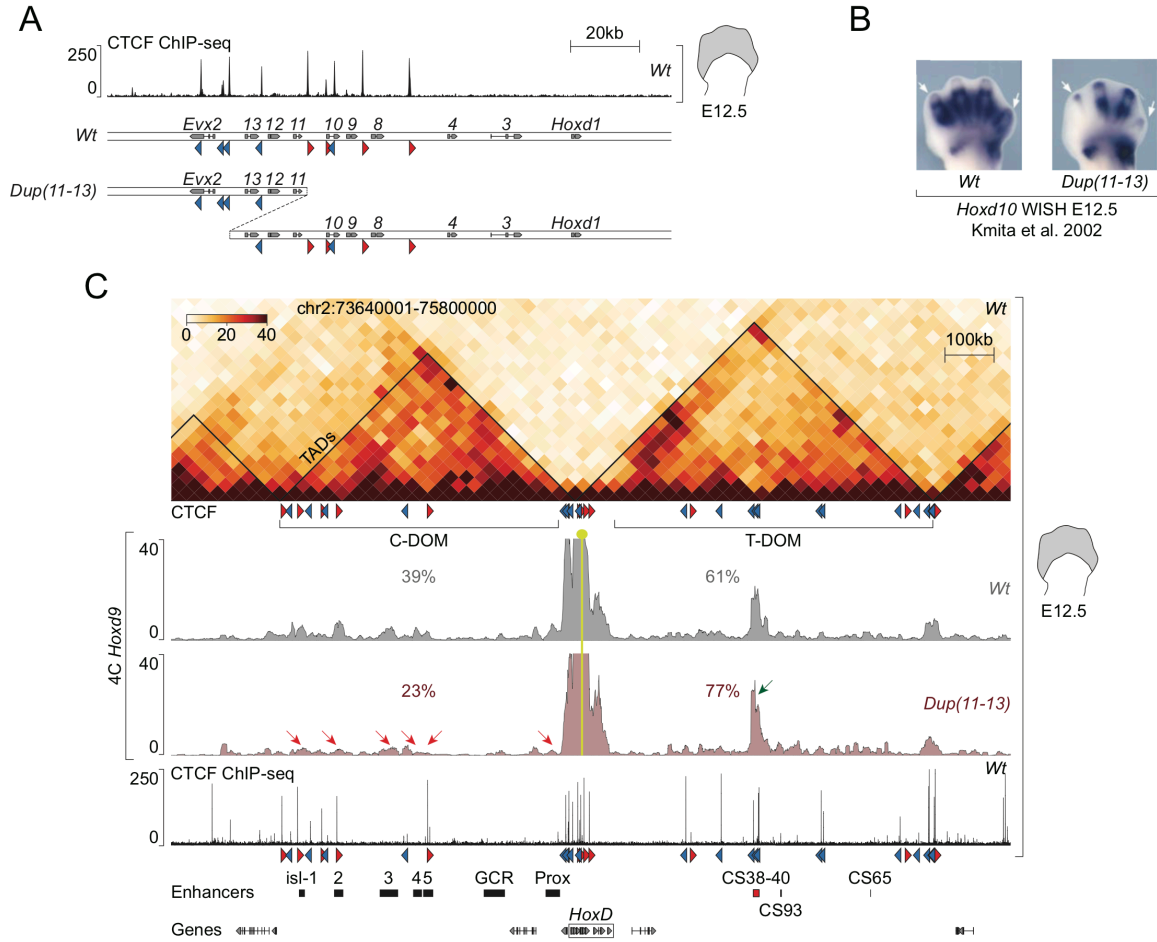


Figure 10: Structural and gene expression changes upon duplication of the region 11-13 (named *Dup(11-13)*). (A) Configuration of *Dup(11-13)* (bottom). The *wild-type* configuration and ChIP-seq of CTCF in *wild-type* distal forelimbs at E12.5 are shown above for comparison. (B) Loss of distal *Hoxd10* expression in *Dup(11-13)* revealed by whole-mount *in situ* hybridizations (WISH) in forelimbs at E12.5. (C) Structural changes upon duplication of the region 11-13. The *wild-type* track is represented in gray and the *Dup(11-13)* track in maroon. The region displayed is chr2:73640001-75800000 (mm10). All chromosome conformation capture experiments (Hi-C and 4C-seq) shown here were performed on distal E12.5 forelimbs. The Hi-C heatmap shown on top represents the *wild-type* situation. Below are shown the results of 4C-seq experiments on *wild-type* (n=2) and *Dup(11-13)* homozygous (n=2) samples. The yellow lollipop indicates the position of the *Hoxd9* 4C-seq viewpoint. Red arrows highlight the loss of *Hoxd9* interactions towards the enhancers of the C-DOM and the green arrow points to the gain of contacts with region CS38-40. Quantifications of the contacts established by *Hoxd9* with either of the two regulatory domains are indicated for both the *wild-type* and *Dup(11-13)*. Hi-C and ChIP-seq data shown here are from (Rodríguez-Carballo et al. 2017); WISH pictures are modified from (Kmita et al. 2002).

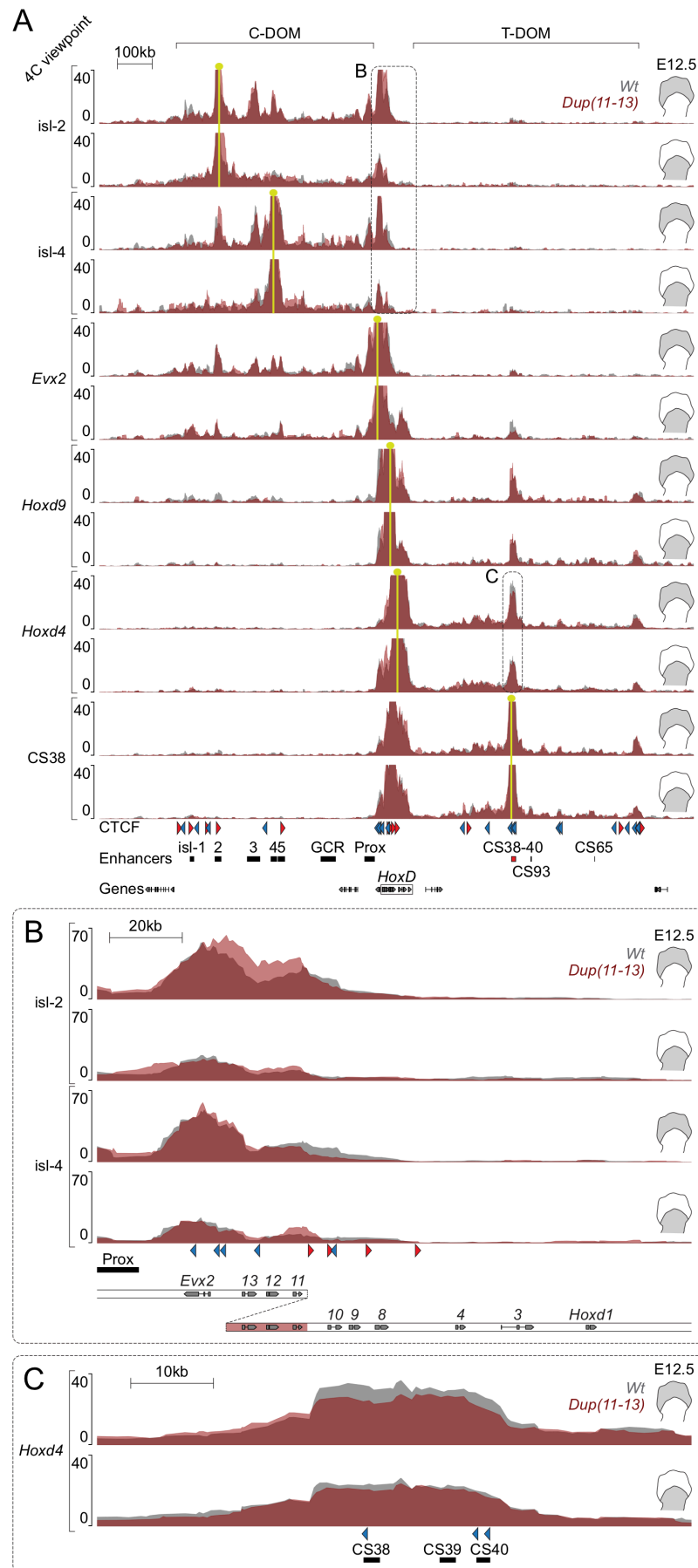


Figure 11: See caption on next page.

Figure 11: Global structural assessment in *Dup(11-13)*. (A) Interactions established by viewpoints located within the TADs, as in Figure 9. Interactions were measured by 4C-seq. The *wild-type* tracks are represented in gray and the *Dup(11-13)* tracks in maroon. *Hoxd9* contacts in both distal and proximal E12.5 forelimbs are also shown. The region displayed is the same as in Figure 10. The position of each 4C-seq viewpoint is represented by a yellow lollipop. Insets are magnified in (B) and (C) and correspond to regions chr2:74628369-74791953 and chr2:75105976-75176976 (mm10), respectively. (B) Contacts of island-2 and 4 towards the *HoxD* cluster. Note the decrease in contacts established with *Hoxd10* and *Hoxd9* upon duplication of the region 11-13 in distal forelimbs. The marked increase in signal over the *Hoxd11-Hoxd13* region is likely the consequence of the duplication of this region in *Dup(11-13)*, highlighted in maroon below. (C) Magnification of *Hoxd4* interactions with region CS38-40, which were decreased in *Dup(11-13)*.

As expected, interactions established by other genes (*Evx2* and *Hoxd4*) or regulatory regions (island-2, island-4 and CS38) that are in principle mainly restricted to their respective domains were largely unchanged in both duplications, confirming that the global structure of the two TADs was maintained (see Figs. 9A and 11A).

Intriguingly, when examining the contacts of *Hoxd4*, a gene that stands on the telomeric half of the cluster and whose expression is restricted to the proximal part of the limbs (Tarchini and Duboule 2006; Rodríguez-Carballo et al. 2017), we nonetheless noticed a decrease in its interactions with region CS38-40 in both duplications (Figs. 9C and 11C). This loss of contacts was more pronounced in the distal part of E12.5 limbs, where the interactions of *Hoxd4* towards region CS38-40 are more frequent than in the proximal portion (Rodríguez-Carballo et al. 2017). Hence, our results raise the possibility for these interactions to be more resilient to perturbations, although they are generally lower, when the T-DOM is active in proximal cells. Even though the loss of *Hoxd4* – CS38-40 contacts was observed in both duplications, it was stronger in *Dup(8i-10)*, suggesting a competition occurring between the promoter of *Hoxd4* and duplicated active genes' promoters for the access to region CS38-40 (see Discussion).

Taken together, our structural analyses of intra-cluster duplications demonstrate a limitation of the access of central *Hoxd* genes to the TAD that is located beyond the duplicated segment, strongly suggesting an increase in the strength of the *HoxD* TAD boundary *in situ*. Moreover, these findings provide possible explanations for the changes in gene expression that were observed in previous studies (Kmita et al. 2002; Tarchini and Duboule 2006).

2. Ectopic boundary integrations

An important question in the field of TAD boundary formation is whether a given boundary would be able to function outside of its original context, or, alternatively, would require specific features associated with this context such as its originally associated TADs. In particular, one would expect in the former case that the ectopic integration of a topological boundary in the genome would result in an establishment of contacts between the integrated elements and the surrounding chromatin regions (Fig. 12). Furthermore, if an ectopic boundary is meant to function in its new context, one could expect it to cause an increase in insulation, that is to prevent most contacts from going through it. One might even go a step further and hypothesize that this could cause a disruption of the TAD in which the boundary got inserted (welcoming TAD). Surprisingly, such evidences remain very limited (see Discussion for details) (Barutcu et al. 2018; Redolfi et al. 2019).

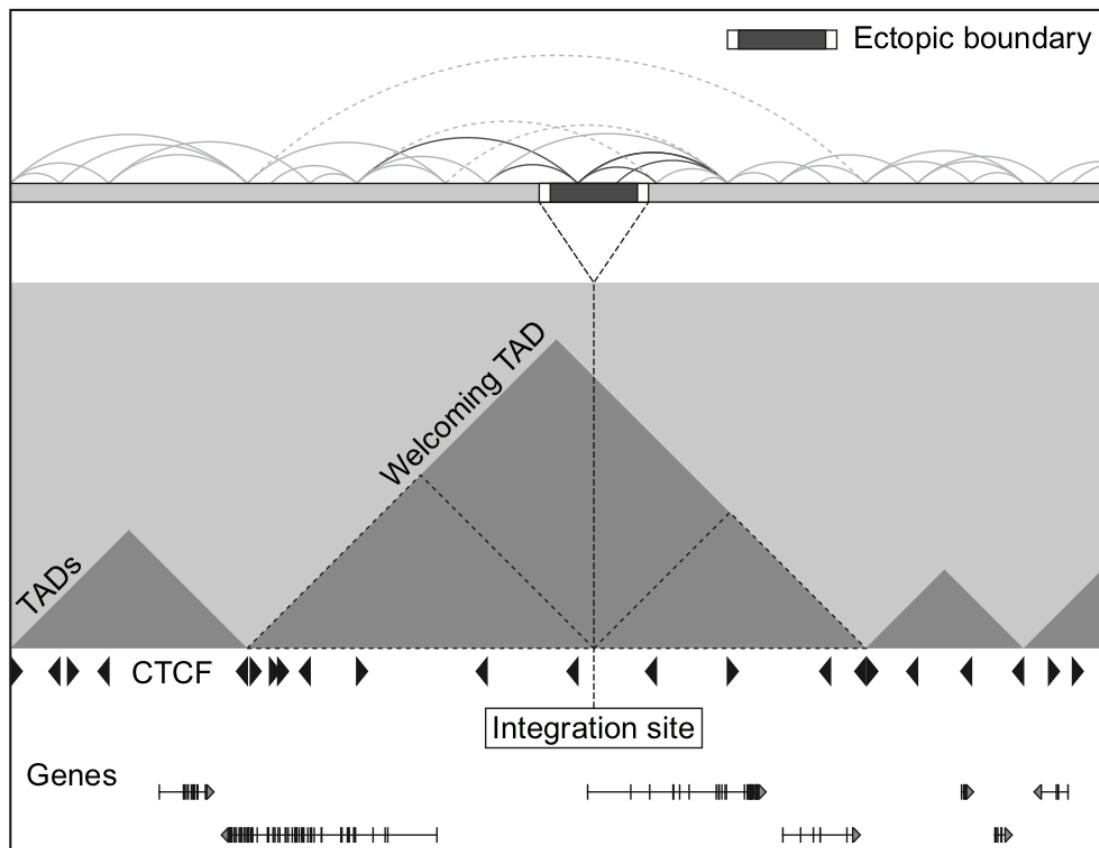


Figure 12: Ectopic boundary integrations. We sought to determine whether a topological boundary would be able to function outside of its original context. From a structural point of view, we expect in this case that (1) the elements of the ectopic boundary would be able to establish contacts with the surrounding regions (see the dark arches in the top panel) and (2) that the boundary would prevent contacts from going through it in a large fraction of cells (light dotted arches in the top panel). This could possibly cause a splitting of the TAD in which the integration occurred (represented as dotted triangles in the bottom panel).

2.1 Preservation of contact directionality from an ectopic *HoxD* cluster

Within the *HoxD* cluster, genes that are located at the extremities of the cluster display interactions that are biased towards the flanking TAD. For instance, *Hoxd13* and *Evx2*, which stand on the centromeric end of the cluster, establish strong constitutive contacts with the C-DOM, whereas *Hoxd1* and *Hoxd4*, which are located on the telomeric side, always interact with the T-DOM (Andrey et al. 2013; Rodríguez-Carballo et al. 2017). Based on this observation, we first asked whether genes that display such preferential contact directionality towards one domain in their original context would keep this behavior when disconnected from their originally associated TADs.

To that end, we took advantage of an insertion line previously established in our group, called *TgN(HoxD)*, in which the whole *HoxD* cluster was integrated in the form of a bacterial artificial chromosome (BAC) at an ectopic location in the genome (Guerreiro et al. 2016; Schep et al. 2016) (also see Material and Methods). The mouse BAC clone that was used spans 230 kb from a breakpoint located downstream (3') of the *Hoxd1* gene, includes the whole *HoxD* cluster together with the Prox element, and the first four exons of the *Lunapark* (*Ln Timer*) gene (Figs. 13A and S3A) (described in Spitz et al. 2001). The extent of this BAC approximately corresponds to the region that needs to be deleted for the merging of the two *HoxD*-associated TADs (region *attP-Rel5*) (Rodríguez-Carballo et al. 2017).

Prior to further analyses, we performed targeted locus amplification (TLA) (de Vree et al. 2014) to determine the genomic region where the *HoxD* BAC got integrated. Similar to 4C-seq, TLA is a proximity ligation and viewpoint-directed inverse PCR-based method. Both 4C-seq and TLA rely on the use of a restriction enzyme recognizing a DNA sequence of four base pairs (referred to as 4-cutter) as primary enzyme. Whereas 4C uses another 4-cutter as secondary enzyme, a 5-cutter recognizing a subset of the primary enzyme's restriction sites is used in TLA (see Material and Methods) (Splinter et al. 2011; van de Werken et al. 2012; de Vree et al. 2014). As a consequence, TLA produces larger hybrid DNA molecules, with a size of 2 kb on average, compared to 4C (~250 bp) (de Vree et al. 2014). Hence, TLA circles contain a higher diversity of hybrid junctions compared to 4C ligation products. Because TLA maximizes the chances to detect the breakpoints of genomic alterations from nearby viewpoints, it is suitable for characterizing integration sites of transgenes or chromosomal rearrangements (Figs. S1 to S5 and S8) (de Vree et al. 2014; Alimohamed et al. 2018; Laboulaye et al. 2018; L. O. Goodwin et al. 2019).

Using *Hoxd13*, *Hoxd12* and elements of the transgenic vector as TLA viewpoints, we could map the integration of the *HoxD* BAC to a region of chromosome 10 and confirmed the preservation of the construct's extent (Fig. S3B and C). In particular, the BAC inserted on the right half of a TAD of about 2 Mb in size (Fig. 13B). This TAD displayed a relatively low density of CTCF-binding sites and contained three protein-coding genes: *Lama2*, *Ptprk* and *Themis*.

A previous study from our group reported that the function of the *HoxD* TAD boundary is required for the preferential contact directionality of the genes located at the extremities of the *HoxD* cluster in its original context (Rodríguez-Carballo et al. 2017). Since the binding of architectural proteins gives the capacity to generate loops and potentially topological borders according to the loop extrusion model (Sanborn et al. 2015; Fudenberg et al. 2016), we asked whether the ectopic *HoxD* cluster would recapitulate normal CTCF and cohesin binding. To evaluate the recruitment of CTCF and of the cohesin subunit RAD21 on the relocated cluster, we carried out a low-input chromatin immunoprecipitation technique called ChIPmentation (see Material and Methods) (Schmidl et al. 2015). We performed this experiment on *TgN(HoxD)*-positive samples at around E10, which carried at the same time a deletion of the endogenous *HoxD* cluster on chromosome 2 (*del(attP-Rel5)d9lac*; see Fig. 13A) (Rodríguez-Carballo et al. 2017). This enabled us to exclude sequencing reads coming from the latter. By doing so, we observed a recruitment of CTCF and RAD21 to all sites that are bound on the endogenous (*wild-type*) cluster at the same stage, except for the one at the promoter of *Hoxd10* (Fig. 13A). Furthermore, TLA revealed the loss of this DNA segment from the construct and its replacement by an exogenous sequence (Table S5). Thus far, our results demonstrate the capacity of the *HoxD* cluster to recruit architectural proteins at known sites outside of its original context. We note that the ChIP signal at the CTCF sites upstream (reverse) and at the level of the *Hoxd9* promoter (cryptic forward) may arise at least in part from the *del(attP-Rel5)d9lac* genetic background. Indeed, in this allele, the entire endogenous cluster was replaced by a *Hoxd9/lacZ* reporter construct comprising these two sites (Fig. 13A; see gray shaded area) (Rodríguez-Carballo et al. 2017).

To determine whether *Hoxd* genes retain preferential contact directionality in their new genomic context, we performed 4C-seq experiments using the promoters of *Hoxd4* and *Hoxd13* as viewpoints in *TgN(HoxD)* (Fig. 13B) (Leonardo Beccari; unpublished results). For both viewpoints, the interactions were largely restricted to the TAD in which the transgene got integrated. Importantly, we observed that the *Hoxd13* gene engaged in stronger contacts

with the regions located on the left-hand side of the integration, with a maximum of contact frequency at the level of the forward CTCFs of the left TAD border. When examining the interactions of *Hoxd4*, we noticed an opposite trend: stronger contacts established with regions on the right-hand side of the integration, culminating at the reverse CTCFs of the right limit of the domain. Based on these observations, although this could not be confirmed by TLA (see Table S5), we deduced that the configuration of the *TgN(HoxD)* integration is likely to be the same as for the endogenous cluster, i.e. with the posterior part (*Hoxd13* side) of the cluster on the left and the anterior part (*Hoxd1* side) on the right. Starting from this assumption, our results strongly suggest that *Hoxd* genes retain preferential contact directionality in this new genomic context.

Although we did not evaluate changes in the local chromatin landscape using 4C-seq viewpoints surrounding the integration or by a Hi-C in the mutant, based on the evidences we have shown so far, we can speculate that the *HoxD* cluster retains boundary function outside of its original genomic context.

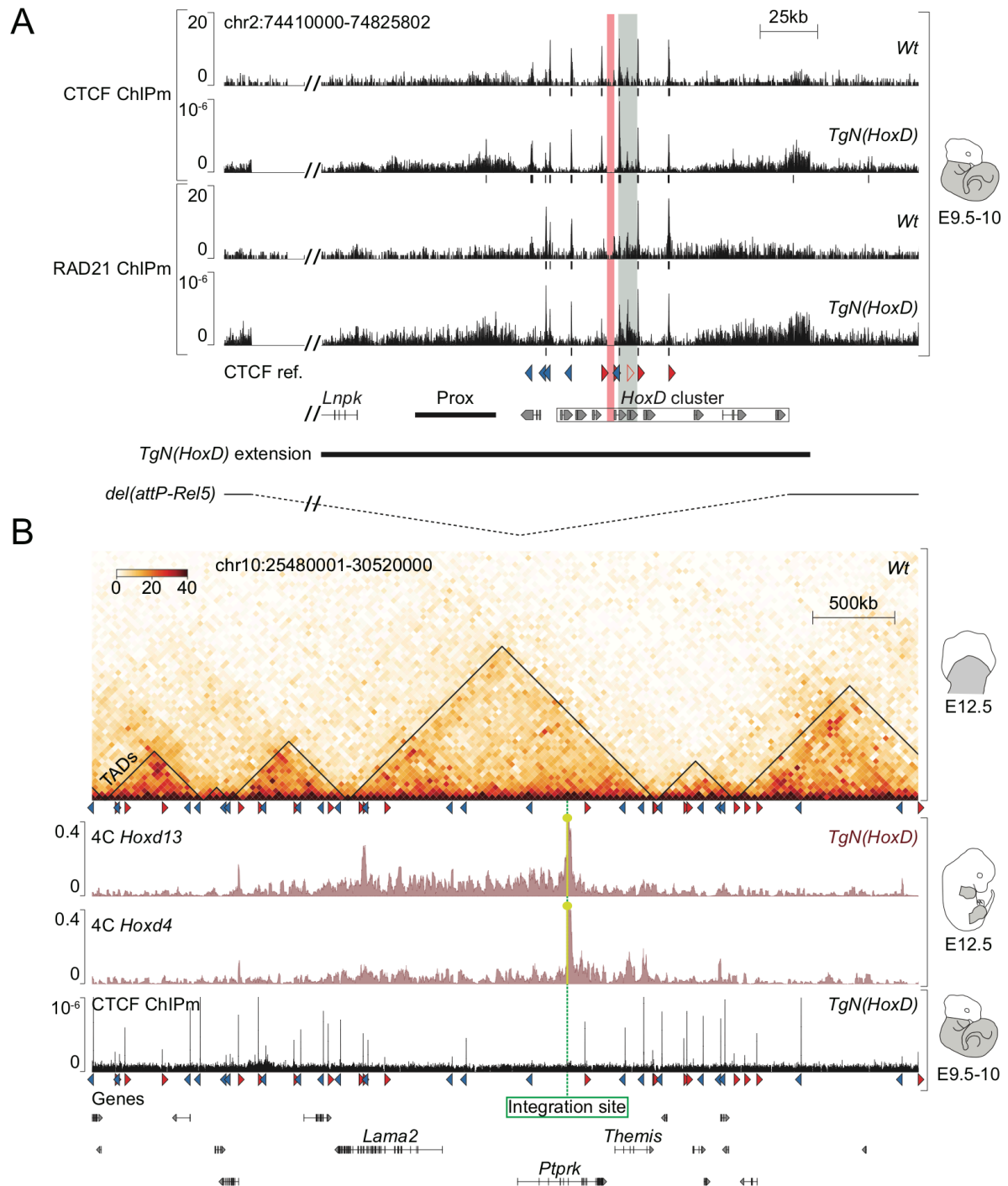


Figure 13: Preserved contact directionality and architectural proteins' recruitment by an ectopic *HoxD* cluster (named *TgN(HoxD)*). (A) Recruitment of CTCF and of the cohesin subunit RAD21 on *TgN(HoxD)*. The binding was assessed using ChIPmentation (ChIPm) on the endogenous *wild-type* cluster ($n=1$ for both CTCF and RAD21) or *TgN(HoxD)* ($n=3$ for CTCF; $n=2$ for RAD21) in E9.5-E10 headless embryos. The region displayed is chr2:74410000-74825802 (mm10), with an exclusion of chr2:74445394-74563131 (see //) to facilitate the visualization of the *HoxD* cluster. Under each track, the corresponding peak calling by the MACS2 algorithm (Y. Zhang et al. 2008; Feng et al. 2012) is represented as black boxes. A CTCF sites' reference (CTCF ref.) corresponding to *wild-type* E12.5 limbs is shown for comparison. The cryptic forward site at the promoter of *Hoxd9* is represented by an empty red arrowhead. The

extension of the *TgN(HoxD)* BAC (chr2:74563132-74777993) is represented below. For *TgN(HoxD)*, the ChIPm experiments were carried out on samples that were homozygous for the *del(attP-Rel5)d9lac* genetic background, which is shown at the bottom. The region that corresponds to the *Hoxd9/lacZ* construct of the *del(attP-Rel5)d9lac* allele is highlighted in gray, whereas the DNA sequence that was lost from the *TgN(HoxD)* construct is highlighted in red. (B) Directional contacts establishment from the ectopic *HoxD* cluster. 4C-seq experiments using the promoters of *Hoxd13* and *Hoxd4* as viewpoints in *TgN(HoxD)* (n=1) whole limbs at E12.5 (Leonardo Beccari; unpublished data). The region displayed is chr10:25480001-30520000 (mm10). The Hi-C heatmap shown on top is representative of the *wild-type* situation in proximal forelimbs at E12.5; data from (Rodríguez-Carballo et al. 2017). Of note, the central TAD of ~2 Mb was identified as three separated domains by TopDom with a window size of 240 kb. However, based on the observation that *Hoxd* interactions spread on the region that corresponds to the 2 Mb TAD, we considered it as a single domain and displayed here the results obtained with a window size of 400 kb. This did not cause any major difference in the identification of the neighboring TADs.

At last, we sought to determine whether *Hoxd* genes could respond to enhancers located in the vicinity of the integration. As expected from previous publications (Spitz et al. 2001; Tschopp et al. 2009; Guerreiro et al. 2016), the absence of its flanking regulatory domains almost completely restricted *HoxD* expression to the trunk and the neural tube along the anteroposterior axis, relying on enhancers located within the gene cluster (Fig. 14) (courtesy of Isabel Guerreiro). However, *Hoxd13* expression was observed in the most distal aspect of the limbs of *TgN(HoxD)* animals at E12.5. This is likely the consequence of the Prox element being included in the BAC, since the observed pattern closely resembled the region in which this element can drive the expression of a reporter gene at the same stage (Gonzalez, Duboule, and Spitz 2007). We also compared the patterns of *Hoxd* genes' expression to published data for the genes located within the *TgN(HoxD)*-containing TAD in chromosome 10, i.e. *Lama2*, *Ptprk* and *Themis*. Whereas *Lama2* is ubiquitously expressed (Schuler and Sorokin 1995), the expression of *Ptprk* is mostly restricted to the developing brain, liver and kidney (Jiang et al. 1993; P. Shen et al. 1999). For its part, *Themis* (also known as *Gasp*) displays a thymus-specific expression (Fu et al. 2009; Johnson et al. 2009; Lesourne et al. 2009; Patrick et al. 2009). Importantly, no ectopic expression of *Hoxd* genes in a pattern that would resemble those of the surrounding genes was detected, at least in tissues that are accessible by whole-mount *in situ* hybridization. These results support the notion that the capacity of gene promoters to respond to an enhancer is not solely due to the former being present in the same TAD than the latter.

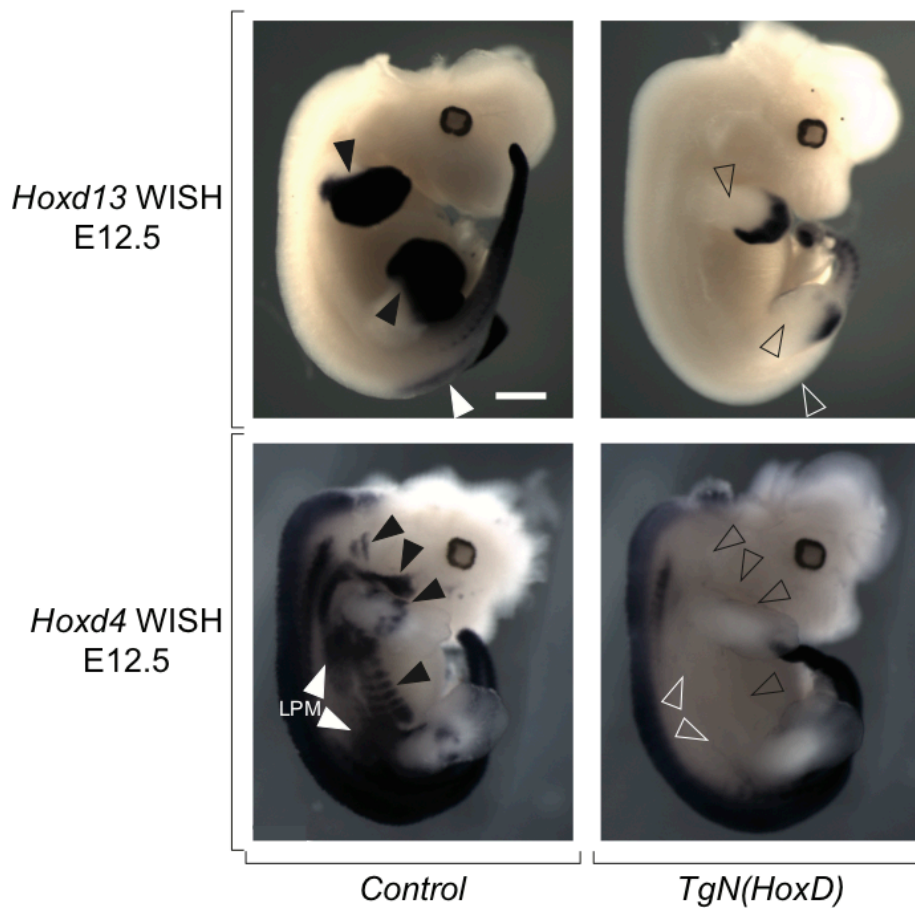


Figure 14: The expression of *Hoxd* genes is mostly restricted to the main (anteroposterior) body axis in *TgN(HoxD)*. Whole-mount *in situ* hybridization against *Hoxd13* and *Hoxd4* in *Control* (*HoxD^{del(1-13)}/+*) and *TgN(HoxD)* (*HoxD^{del(1-13)}/del(1-13)*; *chr10^{TgN(HoxD)}/+*) E12.5 embryos (courtesy of Isabel Guerreiro). In the *HoxD^{del(1-13)}* allele, all *Hoxd* genes (i.e. from *Hoxd1* to *Hoxd13*) are deleted and replaced by a *Hoxd11/lacZ* construct (Zákány et al. 2001). It was used as a genetic background enabling to balance *Hoxd* genes' dosage between conditions. Arrowheads and empty arrowheads indicate presence and loss of expression, respectively. LPM: lateral plate mesoderm. Scale bar: 1 mm.

2.2 Relocating a boundary fragment

As explained above, the *TgN(HoxD)* construct spans 230 kb and includes all genes of the *HoxD* cluster (Spitz et al. 2001; Guerreiro et al. 2016; Schep et al. 2016) (see Fig. 13A). Based on chromosome conformation and CTCF occupancy, previous publications from our group approximately positioned the *HoxD* TAD boundary within a dynamic interval at the posterior half of the gene cluster (Andrey et al. 2013; Rodríguez-Carballo et al. 2017). Hence, we wondered whether a smaller fragment, centered around the expected position of this border, would be able to exert boundary function outside of context.

To get insight into this question, we used a 37 kb fosmid extending from the 3' of *Hoxd11* to a breakpoint located within the third exon of *Evx2* to generate transgenic mice. The initial construct, which we termed *TgN(d11-Evx2)*, was designed so as to comprise four gene units, among which three were complete (*Hoxd11*, *Hoxd12* and *Hoxd13*) and one was partial (*Evx2*), together with three left-oriented CTCF sites (two sites between *Hoxd13* and *Evx2* and one between *Hoxd12* and *Hoxd13*) (Figs. 15A and S4A). The fosmid also contained a *loxP* site on one extremity (see Fig. S4A), allowing for removal of extra-copies in case of tandem integrations of the construct. This removal was achieved by crossing the initial *TgN(d11-Evx2)* transgenic male with a female mouse expressing the Cre recombinase in the germline (Tang et al. 2002) (see *Hprt^{cre}* in Material and Methods).

We first characterized the integration site and configuration of the *TgN(d11-Evx2)* fosmid by TLA using viewpoints distributed inside the construct. Both left and right integration breakpoints were located on a region of chromosome 7 (Figs. 15B, S4B and C). Based on the forward configuration of the construct sequences integrated at this location (see Fig. S4C), meaning that the positive strand of the transgene was connected to the positive strand of the chromosome, TLA suggested that the transgene was oriented with *Evx2* on the left and *Hoxd11* on the right. A 1 kb deletion of chromosomal sequence was also detected at the level of the integration (Fig. S4C). As regards the local chromatin landscape, the transgene got integrated on the left half of a TAD of about 2 Mb displaying a very low density of CTCF sites and containing a single protein-coding gene: *Hs3st4* (Fig. 15B). In particular, the integration occurred within the 400 kb-long intron of the *Hs3st4* gene.

Next, we assessed the recruitment of architectural proteins on the construct by ChIPmentation on E12.5 whole limbs. Strikingly, only the reverse CTCF site located between *Hoxd12* and *Hoxd13* was bound by CTCF and RAD21 at the level of the transgene (Fig. 15A). Similar to what was previously discussed for *TgN(HoxD)*, the ChIP signal at the CTCF sites upstream and at the level of the *Hoxd9* promoter originated from the *del(attP-Rel5)d9lac* genetic background (Rodríguez-Carballo et al. 2017), which was also used here. In addition, we noticed a depletion of ChIP (and total input) signal for the region located between the second exon of *Hoxd13* and the promoter of *Evx2* (Fig. 15A; see red shaded area), revealing an internal truncation of the construct. This was confirmed by the complete absence of signal over the integrated region when performing 4C-seq using the promoter of *Hoxd13* as a viewpoint after Cre recombination (data not shown). Although only a single CTCF-binding site from

the original construct was preserved (even though it might be present in two copies; see below), our results demonstrate its association with both CTCF and RAD21 at the new genomic location.

We also evaluated the number of copies of the transgene using the Control-FREE Copy number and allelic content caller (Control-FREEC) (Boeva et al. 2011; 2012). Control-FREEC is a quantification method that uses next-generation sequencing data to evaluate copy number profiles and allelic constitutions along genomic regions (Boeva et al. 2012). We applied Control-FREEC on the total input DNA data of our ChIPmentation experiments using the *TgN(d11-Evx2); del(attP-Rel5)/d9lac* sample as the “test” and a pool of samples that were *wild-type* for the endogenous *HoxD* cluster and negative for the *TgN(d11-Evx2)* integration as the “control”. In brief, for both the test and the control, the Control-FREEC signal was computed along a 7 Mb window of chromosome 2 including the *HoxD* locus. Then, we calculated the test/control signal ratio to evaluate in how many copies the *TgN(d11-Evx2)* fosmid was present (see Material and Methods). Control-FREEC estimated that the transgene was present in only one or two copies in *cis* at the level of this integration (Fig. S4D).

We asked whether the elements of the *TgN(d11-Evx2)* transgene engaged in interactions with regions surrounding the integration. To do so, we performed 4C-seq using the promoter of the *Hoxd12* gene, which is preserved in the construct, as a viewpoint. We observed that *Hoxd12* established relatively weak contacts with the surrounding regions in E10.5 headless embryos (Fig. 15B). Interestingly, most of the contacts were biased on the left-hand side of the integration and culminated at the promoter of the *Hs3st4* gene, about 20 kb aside from a forward CTCF (Fig. 15B). Hence, these results demonstrate an establishment of promoter-promoter contacts between *Hoxd12* and *Hs3st4*. Some small peaks of interactions also matched with the left and right limits of the welcoming TAD, as well as with a region located halfway between the *Hs3st4* promoter and the integration site (see further).

In order to investigate further, we sought to determine whether the integration of the *TgN(d11-Evx2)* construct caused changes in the local chromatin landscape surrounding the integration site. We note that, at least to our knowledge, studies trying to address this question remained largely lacking in the literature. In particular, before conducting the present study, only a single publication suggested that one CTCF-binding site and one gene promoter are not sufficient to alter the local chromatin landscape to the point of causing the

splitting of a TAD (Barutcu et al. 2018). However, these results are debatable from a methodological point of view (see Discussion).

To get insight into this question, we brought the *TgN(d11-Evx2)* integration into homozygosis and performed 4C-seq experiments using various viewpoints distributed on both sides of the integration site. We used the forward CTCFs at the left border of the welcoming TAD (CTCF-left 1) and upstream of the *Hs3st4* promoter (CTCF-left 2), as well as the promoter of the *Hs3st4* gene itself and two regions standing on the right of the integration site (right 1 and 2). In *wild-type* E12.5 whole limbs, we observed relatively weak contacts established by each viewpoint with the surrounding regions and most of these contacts were restricted within the limits of the domain (Fig. 15B; see gray tracks). In *TgN(d11-Evx2)*, we noticed a minor decrease in the contacts established by each viewpoint with regions located beyond the integration site, suggesting an increased insulation at the level of the integration (Fig. 15B; see maroon tracks). For instance, the interactions established by the forward CTCF site at the left border of the domain (CTCF-left 1 viewpoint), which stands on the left of the integration site, were generally decreased on the right of the integration. To summarize, our structural assessments by 4C-seq revealed that the local chromatin structure is modified, although not drastically, in *TgN(d11-Evx2)* transgenic animals. We note that this could be due to either a reduced number of ectopic boundary elements or to the intrinsic properties of the welcoming TAD, such as its strength (see Discussion). The above possibilities are also not mutually exclusive.

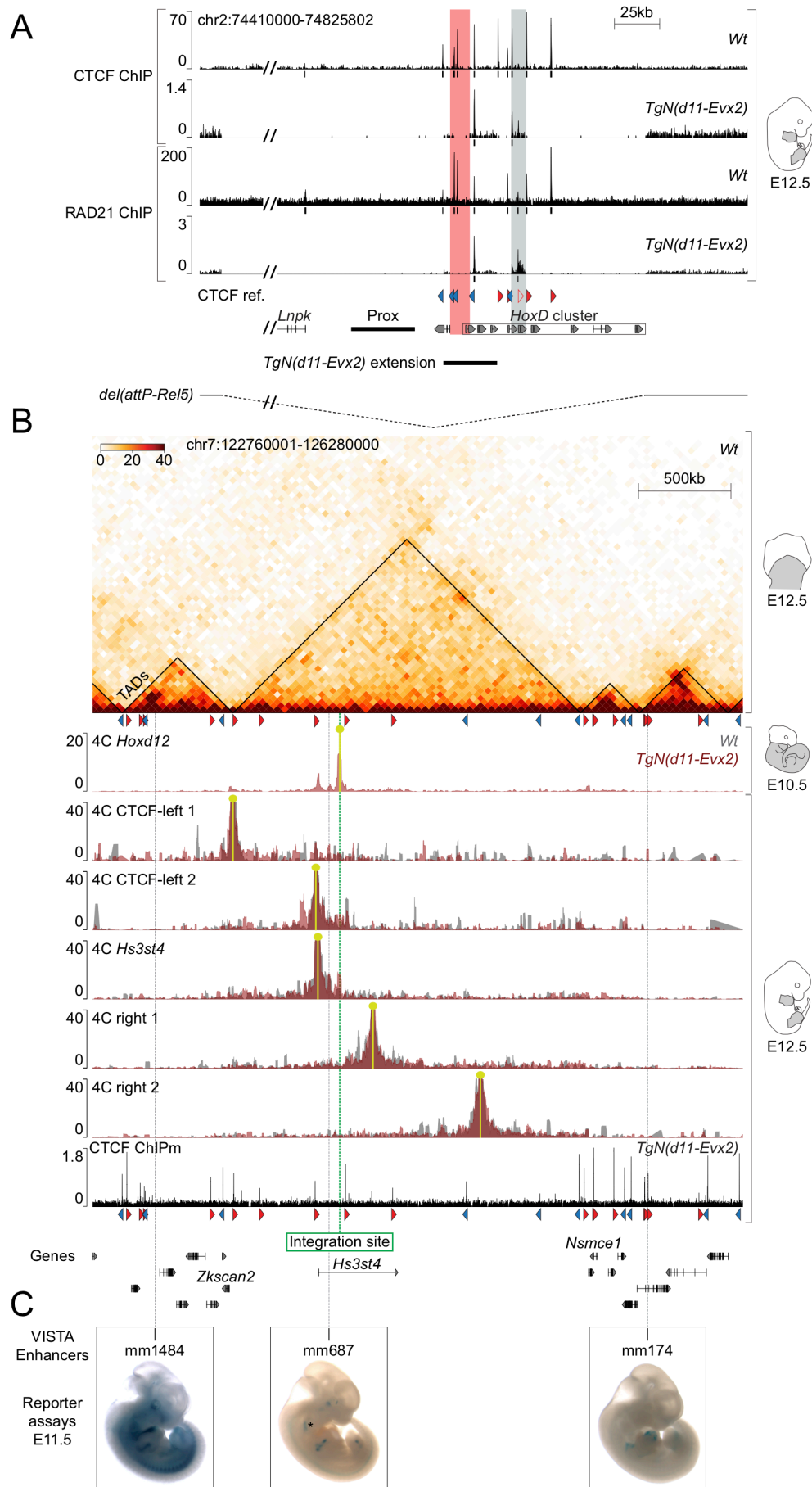


Figure 15: See caption on next page.

Figure 15: Minor structural changes upon integration of an internally truncated *TgN(d11-Evx2)* fosmid. (A) Recruitment of CTCF and of the cohesin subunit RAD21 on the intergenic *Hoxd12-Hoxd13* site of *TgN(d11-Evx2)*. The binding was assessed using ChIP-seq on the endogenous *wild-type* cluster (n=1 for both CTCF and RAD21; data from (Rodríguez-Carballo et al. 2017)), or ChIPmentation on *TgN(d11-Evx2)* (n=2 for both CTCF and RAD21), in whole limbs at E12.5. The region displayed is chr2:74410000-74825802 (mm10), excluding chr2:74445394-74563131 (see //) for an easier visualization of the *HoxD* cluster. Under each track, a MACS2 peak calling is represented as black boxes. The extension of the initial *TgN(d11-Evx2)* fosmid (chr2:74656027-74685767) is indicated below. For *TgN(d11-Evx2)*, the ChIPm experiments were carried out on samples that were homozygous for both the integration and the *del(attP-Rel5)/d9lac* genetic background. The region that corresponds to the *Hoxd9/lacZ* of *del(attP-Rel5)/d9lac* is highlighted in gray, whereas the DNA sequence that was lost from the construct is highlighted in red. (B) Establishment of intra-TAD contacts from the transgene and minor modifications of the local chromatin landscape. 4C-seq using the promoter of the *Hoxd12* gene as a viewpoint in *TgN(d11-Evx2)* E10.5 headless embryos (heterozygous samples). 4C-seq experiments using viewpoints surrounding the integration site in E12.5 whole limbs. The *wild-type* tracks are represented in gray and the *TgN(d11-Evx2)* tracks in maroon. The *TgN(d11-Evx2)* samples used here were homozygous for the integration. All 4C-seq experiments shown on this figure were performed in a single biological replicate. The region displayed is chr7:122760001-126280000 (mm10). The Hi-C heatmap shown on top represents the *wild-type* situation (proximal forelimbs E12.5; data from (Rodríguez-Carballo et al. 2017)). The TADs were identified by the TopDom algorithm with a window size of 240 kb. (C) Enhancers that were retrieved from the VISTA Enhancer Browser for the region displayed in (B). Representative pictures of *lacZ* reporter assays at E11.5 are shown below. These pictures were also taken from the VISTA Enhancer Browser. The asterisk on the mm687 reporter assay highlights enhancer activity in the semicircular canal of the ear.

At last, we asked whether the integration could impact on the expression of the genes surrounding the integration site, or, conversely, would trigger the ectopic activation of the genes included inside the construct. As mentioned above, the *TgN(d11-Evx2)* transgene got integrated in the intron of the *Hs3st4* gene (see Figs. 15B, S4B and C). Previous publications reported that *Hs3st4* expression is mostly brain-specific (Shworak et al. 1999; Lawrence et al. 2007). We performed WISH and found that *Hs3st4* was also expressed in the trunk (in what might be ganglia of the peripheral nervous system), in the developing semicircular canals of the ears and in some regions of the lateral plate mesoderm as well as of the limbs at E12.5 (Fig. 16A). As expected from the disruption of the gene by the integration, we observed that *Hs3st4* expression was completely lost in *TgN(d11-Evx2)* E12.5 embryos. Interestingly, we noticed that, in the *TgN(d11-Evx2)* specimen, the *Hoxd12* gene became expressed in two cervical stripes, at the level of the developing ears and in the anterior-proximal extremity of the forelimbs, closely resembling the *wild-type* *Hs3st4* expression pattern (Fig. 16B). A slight increase in the lateral plate mesoderm expression of *Hoxd12* was also detected.

On the basis of the above observations, we sought to determine whether a nearby enhancer could be responsible for the ectopic activation of *Hoxd12*. With that aim, we mined the publicly available data from the VISTA Enhancer Browser, which compiles experimentally validated enhancers as assessed by *lacZ* reporter assays *in embryo* (Visel et al. 2007). Three enhancers were described for the region displayed in Fig. 15B. One of them (mm687) was found inside the *TgN(d11-Evx2)*-containing TAD (Fig. 15C). This enhancer was located within the intron of the *Hs3st4* gene, on the left of the integration site. It displayed activity in the semicircular canals of the ears, in some regions of the limbs and in the trunk at E11.5, matching part of the *wild-type Hs3st4* expression domains (see Figs. 15C and 16A). Furthermore, the peak of *Hoxd12* interactions that was observed halfway between the promoter of *Hs3st4* and the integration site coincided with this enhancer (see Fig. 15B and C). Altogether, these results suggest that *Hoxd12* might have been activated by this enhancer in its new context. However, we cannot rule out a possible contribution of direct promoter-promoter communication between *Hoxd12* and *Hs3st4* to at least part of the ectopic expression pattern of the *Hoxd12* gene (see Discussion).

We also assessed the expression of genes located beyond the limits of the welcoming TAD in both *wild-type* and *TgN(d11-Evx2)* embryos. We focused our attention on two protein-coding genes, *Zkscan2* and *Nsmce1*, which are located just outside of the left and right limits of the welcoming TAD, respectively (see Fig. 15B). These genes were mostly used as controls, as they are located in a different TAD than the one in which the transgene integrated. In the *wild-type*, *Zkscan2* was mostly expressed in the distal part of the limbs and in the spinal cord at E12.5 (Fig. 16C). On its side, *Nsmce1* was expressed in the limbs, in a very similar pattern compared to that of *Zkscan2*, but also displayed a wide expression in the lateral plate mesoderm (Fig. 16D). In parallel, we looked for nearby *cis* regulatory elements possibly involved in controlling the expression of these genes. One enhancer (mm1484) was located on the left of *Zkscan2*, within the same TAD than the latter, and showed reproducible activity in the eyes, facial mesenchyme, heart and limbs at E11.5 (see Fig. 15C). Another enhancer (mm174) was found just outside on the right of the *Nsmce1* TAD and displayed heart-specific activity. Thus, the activity patterns of the mm1484 and mm174 enhancers did not show a high degree of correspondence with the expression domains of *Zkscan2* and *Nsmce1*, respectively. Hence, the expression of these genes might depend on *cis* regulatory elements that have not yet been identified or rely on more proximal regulations.

The expression of both genes was largely unchanged in *TgN(d11-Evx2)* transgenic animals. Only a subtle decrease of forelimb *Zkscan2* and mesodermal *Nsmce1* expression was observed (Fig. 16C and D; see arrowheads). We note that this could be due to slight differences in the developmental stage among embryos. Overall, these data suggest only minor or no effect on gene expression for genes located in a different TAD than the one in which the integration occurred. This is consistent with the notion that enhancers and the genes they regulate are generally present within the same TAD (e.g. Long, Prescott, and Wysocka 2016; Sikorska and Sexton 2020); a feature that could render their interactions resilient to perturbations occurring within other domains (also see Spielmann, Lupiáñez, and Mundlos 2018).

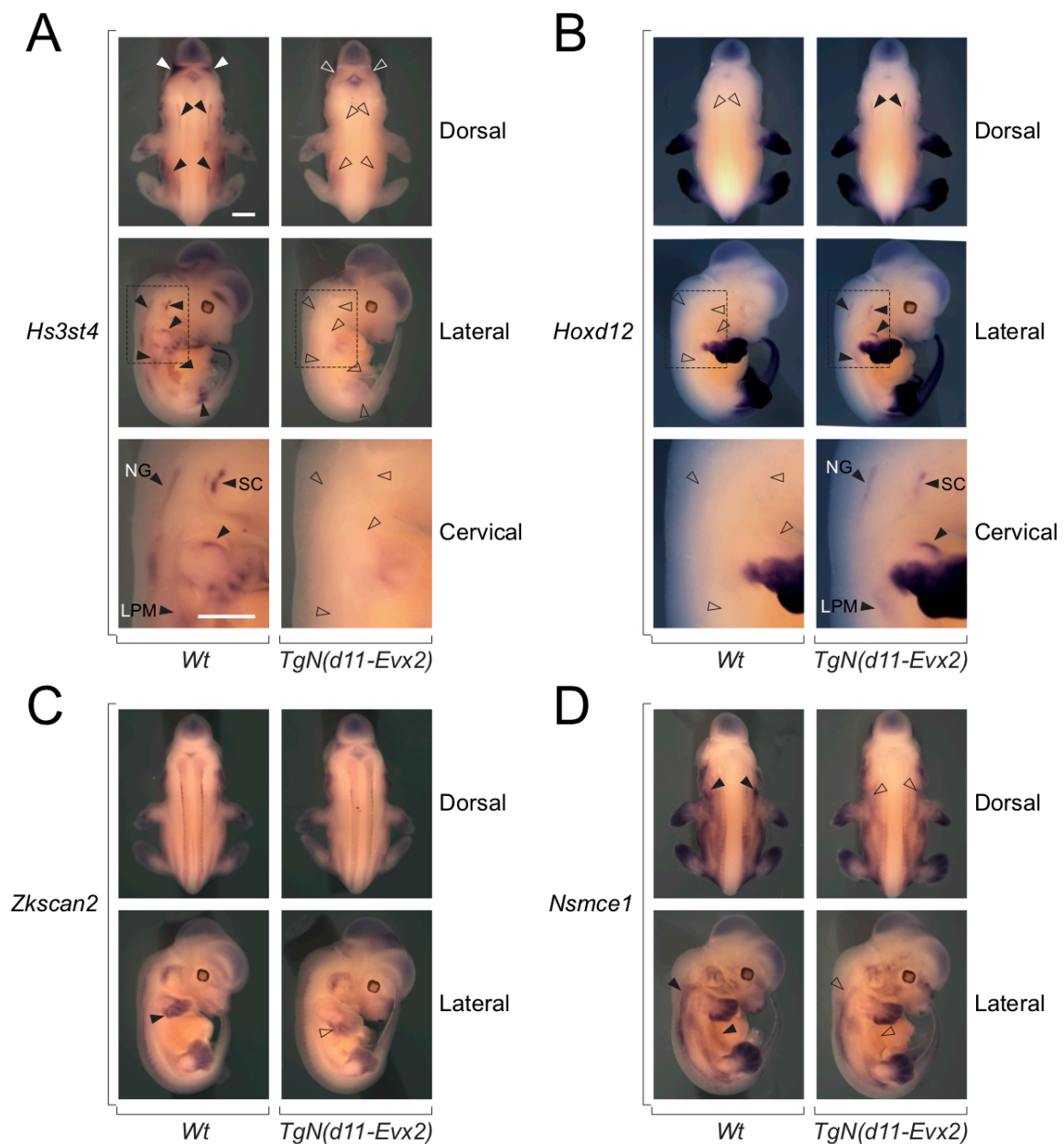


Figure 16: See caption on next page.

Figure 16: Gene expression analyses in *TgN(d11-Evx2)* by whole-mount *in situ* hybridization on E12.5 embryos. (A) *Hs3st4* expression. In the *wild-type*, *Hs3st4* was expressed in the brain, in two cervical stripes that might correspond to ganglia of the peripheral nervous system (NG, nervous ganglia), in the developing semicircular canals (SC) of the ears, as well as in some regions of the limbs and of the lateral plate mesoderm (LPM). In *TgN(d11-Evx2)*, *Hs3st4* expression was completely lost. (B) *Hoxd12* expression. The limb and trunk-specific expression of *Hoxd12* arose from the endogenous *wild-type* cluster. In *TgN(d11-Evx2)*, *Hoxd12* was ectopically expressed in the cervical region and to a lower extent in the lateral plate mesoderm, closely resembling the *wild-type* *Hs3st4* expression. Insets corresponding to the cervical region have been magnified for both *Hs3st4* and *Hoxd12* WISH. (C) *Zkscan2* expression. *Zkscan2* was mostly expressed in the limbs and in the spinal cord. A slight decrease of expression was observed in the forelimbs of the *TgN(d11-Evx2)* embryo. (D) *Nsmce1* expression. *Nsmce1* was highly expressed in the limbs and in the lateral plate mesoderm. Only a subtle decrease of lateral plate mesoderm expression could be observed in the *TgN(d11-Evx2)* embryo compared to the *wild-type*. Arrowheads indicate presence or gain of expression. Empty arrowheads indicate absence or loss of expression. All *TgN(d11-Evx2)* transgenic animals used here were homozygous for the integration and *wild-type* for the endogenous *HoxD* cluster. Scale bars: 1 mm.

2.3 Region CS38-40 as an ectopic border

We then asked whether another region of the *HoxD* locus displaying CTCF binding and holding insulation potential could function as a boundary at novel genomic locations.

With that aim in mind, we turned to a region called CS38-40 (Fig. 17). This region is located in the telomeric TAD of the *HoxD* locus (T-DOM) and displays three conserved non-coding sequences (CNS or CS): CS38, CS39 and CS40 (Andrey et al. 2013). Importantly, region CS38-40 is a relatively weak boundary that separates the T-DOM in two sub-domains, hence its status of sub-TAD border (Andrey et al. 2013; Rodríguez-Carballo et al. 2019; 2020). Indeed, the deletion of this region resulted in the merging of the T-DOM's sub-domains (Rodríguez-Carballo et al. 2020). CS38-40 is bound by CTCF on three sites that are oriented towards the *HoxD* cluster: one located at the level of CS38 and two at the level of CS40 (Rodríguez-Carballo et al. 2017). In particular, the function of the CS38-40 sub-TAD border relies on the binding of CTCF onto this region, as demonstrated by the fusion of the T-DOM's sub-TADs when only its associated CTCF sites were deleted (Rodríguez-Carballo et al. 2020).

In addition, region CS38-40 comprises a well-known limb enhancer, which corresponds to the CS39 element (Andrey et al. 2013; Beccari et al. 2016), as well as the transcription start sites of two divergently transcribed long non-coding RNAs (lncRNAs), called *Hotdog* (*Hog*) and *Twin of Hotdog* (*Tog*) (Delpretti et al. 2013). The latter transcripts were shown to be expressed at lower levels and in

less territories than *Hoxd* genes. Their expression is typically restricted to the developing cecum, to the proximal part of the developing limbs and to the mammary buds (Delpretti et al. 2013; Schep et al. 2016; Rodríguez-Carballo et al. 2017; 2019). Hence, it is noteworthy that *Hog* and *Tog*, with their shared transcription start site, may represent a less promiscuous transcriptional situation compared to that of the *HoxD* cluster. In the latter case, many genes are densely packed on the chromosome and are expressed in partially overlapping domains. This means that several promoters can be active simultaneously in the same cells. Since enhancer-promoter contacts could potentially affect the local chromatin landscape (possibly in a CTCF/cohesin-independent manner; see Rao et al. 2017), we asked whether structural changes would also be observed without such an array of gene promoters. For this reason, region CS38-40 seemed to be particularly well suited.

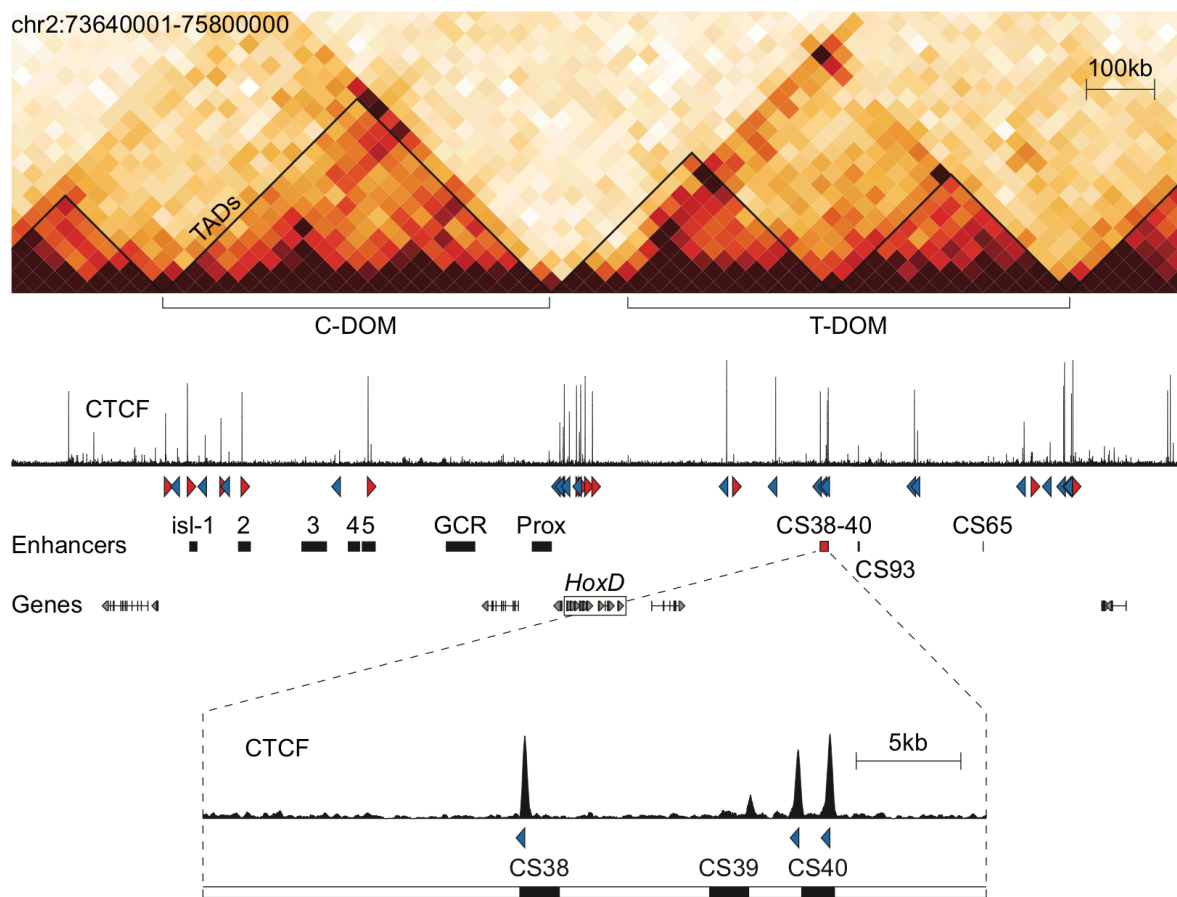


Figure 17: Region CS38-40 is a boundary that separates the T-DOM in two sub-domains. As for Figure 7, on top is shown a Hi-C of the *HoxD* locus with a CTCF ChIP-seq track below. Note the partitioning of the telomeric TAD of the *HoxD* locus (T-DOM) in two sub-TADs at the level of region CS38-40. Magnification of region CS38-40 reveals binding of CTCF at the level of three left-oriented sites: one at the level of the CS38 element and two at the level of CS40. Data are from (Rodríguez-Carballo et al. 2017) and correspond to the situation in the proximal part of *wild-type* forelimbs at E12.5.

We performed random transgenesis using a 45 kb fosmid comprising the entire region CS38-40, including its three left-oriented CTCF sites, which we termed *TgN(38-40)* (Figs. 18A, 21A, S5A and S8A). By doing so, we obtained two independent mouse lines carrying ectopic integrations of the construct on distinct chromosomes. Both initial transgenic males were crossed with Cre-expressing females for transgene extra-copy removal before performing the experiments.

For the first line, which we called *TgN(38-40)^{high}* (see below for details), TLA showed that the construct got integrated in a region of chromosome 3 (Figs. 18B, S5B and C). Identification of both integration breakpoints demonstrated that the transgene was present in a forward sequence configuration (i.e. with the CTCFs pointing towards the left). It also revealed a deletion of about 12 kb of chromosomal sequence at the level of the insertion site (see Fig. S5C).

Assessment of CTCF and RAD21 binding by ChIPmentation in *TgN(38-40)^{high}* E12.5 whole limbs confirmed the recruitment of both proteins on the relocated region CS38-40 (Fig. 18A). In particular, CTCF was present on all three binding sites at almost identical levels, whereas cohesin was highly enriched on the CS38 site, closely resembling the corresponding *wild-type* profiles. Similarly to what has been presented for previous lines, the samples that were used for ChIPmentation were homozygous for a specific deletion background (here a deletion of region CS38-40 on chromosome 2: *del(CS38-40)*; see Fig. 18A) (Schep et al. 2016), enabling us to distinguish between the endogenous and ectopic regions CS38-40.

Interestingly, we noticed a particularly high background level in the ChIP (and total input) sequencing signal within the region corresponding to the extent of the transgene, suggesting the presence of multiple tandem copies of the construct at the level of this integration (see Fig. 18A). In order to validate this finding, we applied the Control-FREEC quantification algorithm on the total input DNA data of the ChIPmentation experiment for this line. Control-FREEC estimated that the transgene was present in between 17 and 23 apparently entire copies in this line (Fig. S5D). In parallel and as an alternative approach, we performed qPCR on purified genomic DNA to evaluate the transgene copy number. Surprisingly, on average, qPCR supported the presence of only 4 copies of the construct, albeit with a substantial dispersion of the values among samples (Fig. S6). Although the results of Control-FREEC and qPCR-based quantifications showed discrepancies, they both support the presence of

multiple copies (4 or more) of the *TgN(38-40)* construct in this line, which is the reason why we termed it *TgN(38-40)^{high}*. We note that the retention of multiple in-cis copies of the transgene could be due to the loss of the *loxP* site from the construct, rendering it insensitive to the action of the Cre recombinase, or to an insufficient accumulation of Cre in the oocyte.

As regards the welcoming landscape, the precise TLA-identified *TgN(38-40)^{high}* integration site was found about 25 kb away from the right border of a 440 kb TAD (Fig. 18B). This TAD displayed only single CTCF sites at the level of its boundaries and established relatively frequent contacts with the next TAD on the right in the *wild-type*. We first assessed whether an element of the transgene engaged in interactions with the surrounding regions. To that end, we performed 4C-seq using a viewpoint close to the CTCF site of the CS38 element in E12.5 whole limbs (Fig. 18B; see 4C CS38). By doing so, we noticed that CS38 established contacts with regions of its new surrounding landscape. Interestingly, these contacts were mostly biased towards the left-hand side of the integration. Furthermore, the interactions of CS38 were largely restricted within the limits of the welcoming TAD and culminated at CTCF sites in convergent orientation relative to the ones of the transgene. These results demonstrate the ability of CS38 to engage in CTCF-delimited interactions in a novel genomic context. Surprisingly, we also noticed that some contacts extended up to the promoter of a gene called *Tbx15*, located at the left extremity of the left neighboring TAD (see below).

In order to evaluate changes in the local chromatin landscape upon integration, on the same basis as for *TgN(d11-Evx2)*, we designed 4C-seq viewpoints in the regions surrounding the integration site. In particular, we used the forward CTCF at the left edge of the welcoming TAD (CTCF-left), with which CS38 strongly interacted, the reverse CTCF at the right limit of the domain (CTCF-right), as well as the promoter of a gene located in the next TAD on the right, *Man1a2*. In *wild-type* limbs, each of these viewpoints established contacts with their surroundings, generally more frequently within the limits of their respective TAD (Fig. 18B; see gray tracks). Maximum contact frequencies were observed around TAD limits, as can be expected from the proximity of the endogenous boundaries. As regards *Man1a2*, the contacts established by its promoter were relatively low along the welcoming TAD, most probably owing to its location in the right neighboring TAD. Therefore, we considered this last viewpoint as an extra-TAD control for which the changes in contacts upon integration (if any) should be weaker than for the other viewpoints. In samples that were homozygous for the *TgN(38-40)^{high}* integration, we observed a clear

decrease in the contacts established by each viewpoint with regions located beyond the integration site, especially with the boundaries of the welcoming TAD (Fig. 18B; see maroon tracks). Interestingly, drops of contacts towards the domain's boundaries were of comparable amplitude regardless of the side considered relative to the integration. Indeed, when looking from the CTCF-left viewpoint, the decrease in the contacts established towards the right border of the welcoming TAD was very similar to that observed at the level of the left border when taking CTCF-right as a viewpoint (see Fig. 18B). Given the reverse orientation of the integrated CTCFs, the blockage of contacts spanning the integration in a right-to-left direction is likely the consequence of a loop-loop stalling effect (see Discussion). As expected, the drop in the contacts established by the promoter of *Man1a2* was less pronounced than for the CTCF-right viewpoint, located on the same side of the integration but belonging to the welcoming TAD. Although we did not perform a Hi-C of *TgN(38-40)^{high}*, the drastic structural changes that we observed in the local chromatin structure provide compelling evidence for a substantial increase in insulation caused by the ectopic insertion of the CS38-40 sub-TAD border in multiple copies.

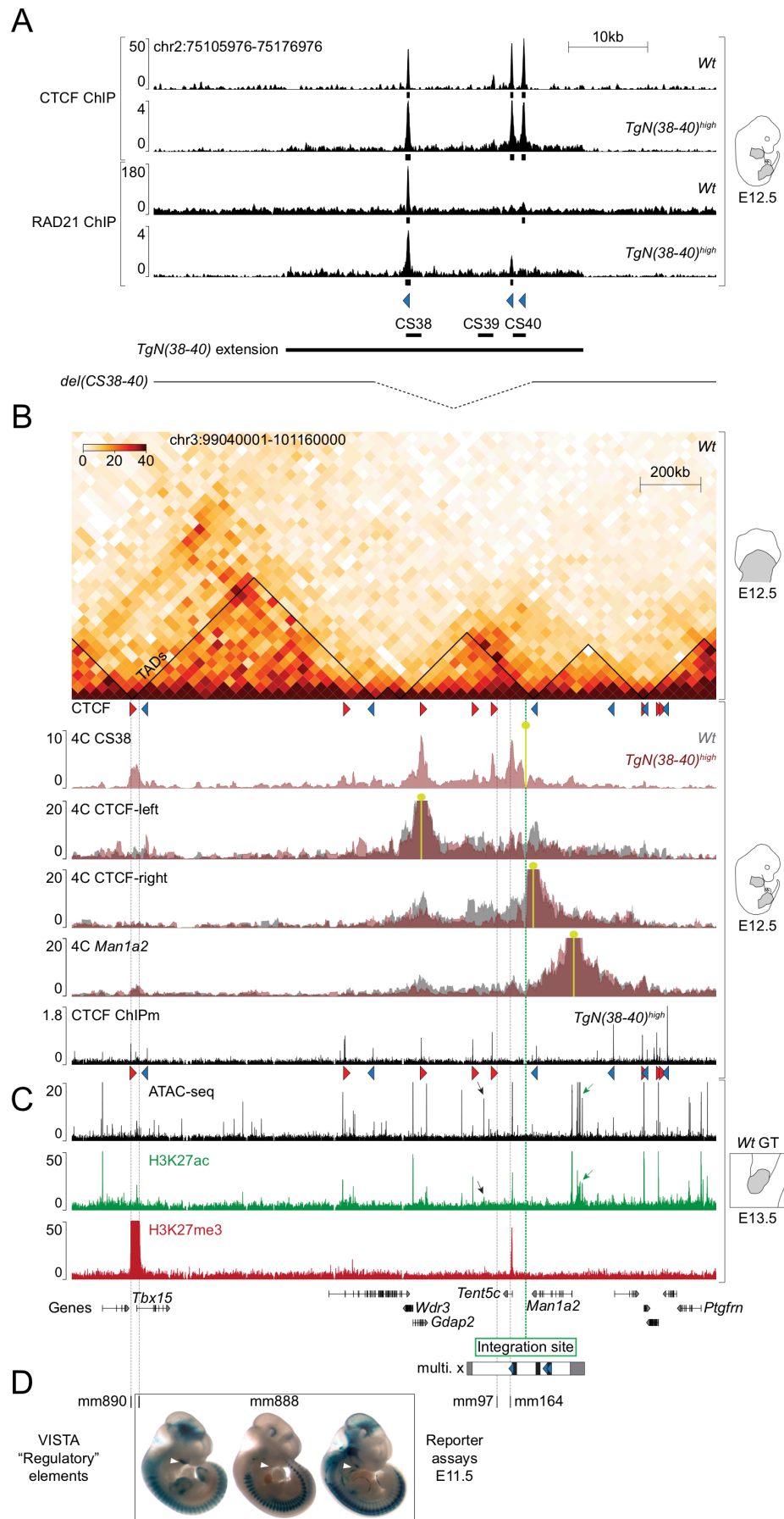


Figure 18: See caption on next page.

Figure 18: Drastic structural changes upon ectopic integration of region CS38-40 in multiple in-*cis* copies (*TgN(38-40)^{high}*). (A) Recruitment of CTCF and of the cohesin subunit RAD21 on the relocated region CS38-40. CTCF occupancy was similar for all three sites, whereas RAD21 was preferentially recruited at the level of CS38, closely resembling the *wild-type* profiles. Note the elevated background level in the signal within the region that corresponds to the transgene, indicating the presence of multiple copies of the latter. The binding was assessed using ChIP-seq on the endogenous (*wild-type*) region CS38-40 (n=1 for both CTCF and RAD21; data from (Rodríguez-Carballo et al. 2017)), or ChIPmentation on *TgN(38-40)^{high}* (n=2 for both CTCF and RAD21), in whole limbs at E12.5. The region displayed is chr2:75105976-75176976 (mm10). Under each track, a MACS2 peak calling is represented as black boxes. The extension of the initial *TgN(38-40)* fosmid (chr2:75122684-75160161) is indicated below. For *TgN(38-40)^{high}*, the ChIPm experiments were performed on samples that were homozygous for both the integration and the *del(CS38-40)* genetic background, which is shown at the bottom. (B) Establishment of welcoming TAD-restricted contacts from the transgene and major alterations of the local chromatin landscape. On top is shown a *wild-type* Hi-C of the welcoming region (proximal forelimbs E12.5; data from (Rodríguez-Carballo et al. 2017)). TAD identification was achieved using the TopDom algorithm with a window size of 240 kb. Below the Hi-C heatmap is shown a 4C-seq where region CS38 was used as a viewpoint in *TgN(38-40)^{high}* E12.5 whole limbs (heterozygous samples). Note that most contacts stopped at CTCF sites that are convergent relative to the ones of the transgene, whose orientation is shown at the bottom of panel (C). Some contacts also extended towards the promoter of the *Tbx15* gene on the far left. 4C-seq using viewpoints surrounding the integration site in E12.5 whole limbs. The *wild-type* tracks are represented in gray and the *TgN(38-40)^{high}* tracks in maroon. The *TgN(38-40)^{high}* samples used here were homozygous for the integration. All 4C-seq experiments were performed in a single biological replicate. The region displayed is chr3:99040001-101160000 (mm10). (C) The last three tracks represent chromatin opening (ATAC-seq in black; n=1) and coverage of the active H3K27ac (green) or repressive H3K27me3 (red) histone modifications (assessed by ChIP-seq; n=1) in *wild-type* genital tubercles (GT) at E13.5; data from (Amândio et al. 2020). Black and green arrows indicate a putative silencer (ATAC-seq+; H3K27ac-), or enhancers (ATAC-seq+; H3K27ac+), respectively. (D) Validated (mm888) or putative (mm890, 97 and 164) regulatory elements retrieved from the VISTA Enhancer Browser for the region shown in (B) and (C). Note that only the mm888 element displayed reproducible enhancer activity at E11.5 (branchial arches; see white arrowheads). Representative pictures of *lacZ* reporter assays were also retrieved from the VISTA Enhancer Browser.

We then asked whether the integration of region CS38-40 in multiple copies in this new landscape caused changes in the expression of genes located in the vicinity of the insertion site. We first probed the expression of *Wdr3*, a gene located at the left limit of the welcoming TAD (see Fig. 18B). In the *wild-type*, we found that *Wdr3* expression was mostly restricted to the limbs, the lateral plate mesoderm and the genital tubercle (GT) at E12.5 (Fig. 19A). Interestingly, both mesodermal and GT transcript levels appeared to be severely decreased in *TgN(38-40)^{high}*, suggesting an effect of the integration on the expression of this gene. Given the relatively frequent contacts occurring between the welcoming TAD and its right neighbor in the *wild-type* (see Fig. 18B), we postulated that, in

this particular situation, a change in the expression of a gene located in the latter domain would be possible. To evaluate this, we assessed the expression of the *Man1a2* gene, which is located beyond the right margin of the welcoming TAD. By doing so, we noticed that *Man1a2* displayed a very similar *wild-type* expression pattern compared to that of *Wdr3* (Fig. 19B). Interestingly, *Man1a2* expression seemed to be gained in the lateral plate mesoderm as well as in the genital tubercle of a *TgN(38-40)^{high}* embryo, on the contrary to what was observed for *Wdr3*. Given such an opposite effect for two different genes, the fact that the latter solely arose from stage differences among embryos is relatively unlikely. As regards another gene located within the welcoming TAD, *Gdap2*, its expression was almost completely restricted to the distal part of the limbs and the posterior half of the tail, and no particular changes could be observed between *wild-type* and *TgN(38-40)^{high}* individuals (Fig. 19C).

To our understanding, there are two main hypotheses which could explain the above observations concerning the *Wdr3* and *Man1a2* genes (Fig. 20). In the first case scenario, an enhancer triggering *Wdr3* expression in the *wild-type* mesoderm and GT would be located on the right-hand side of the integration site. Upon integration of CS38-40 in multiple copies, this enhancer would be disconnected from *Wdr3* and instead activate *Man1a2* in the mesoderm and the GT. Second, a silencer located on the left of the integration site would repress *Man1a2* in the *wild-type* mesoderm and GT. If the ectopic boundary is placed in between the gene and the silencer, the silencer would turn from *Man1a2* to *Wdr3* on the other side, resulting in the derepression of *Man1a2* and the *de novo* repression of *Wdr3*.

To try deciphering between these two possibilities, we first looked for potential *cis* regulatory elements within this region in the VISTA Enhancer Browser. Only a single experimentally validated enhancer was described for this region (mm888). This enhancer was located within the first intron of the *Tbx15* gene and showed reproducible activity in the branchial arches at E11.5 (Fig. 18D; see white arrowheads). Because of the large genomic distance, lack of chromatin interactions and activity pattern mismatches between the mm888 enhancer and *Wdr3/Man1a2*, it is unlikely that this regulatory region was in cause for any change in the expression of these genes (see Figs. 18B, 19A and B).

For that reason, we extended our analysis to putative regulatory elements (as defined by highly conserved non-coding sequences) from VISTA within the welcoming region (Mayor et al. 2000; Frazer et al. 2004; Visel, Bristow, and Pennacchio 2007). Two putative regulatory elements were found close to the

integration site (mm97 and mm164; see Fig. 18D), but they did not show reproducible enhancer activity at E11.5 (Visel et al. 2007). Nevertheless, it is possible that they act as enhancers at other developmental stages, or alternatively, that these sequences actually correspond to silencer elements. Therefore, in parallel, we mined chromatin accessibility and histone modifications data. Since the genital tubercle displayed changes in *Wdr3/Man1a2* transcripts levels, we took advantage of a recent publication from our group (Amândio et al. 2020) and focused our attention on this tissue. We found that none of these two elements showed increased chromatin accessibility (a prerequisite for the binding of most transcriptional regulators) (Tsompana and Buck 2014) as assessed by ATAC-seq in E13.5 genital tubercles (Fig. 18C). They were also devoid of H3K27ac (see Fig. 18C), a histone mark associated with active enhancers (Creyghton et al. 2010; Y. Shen et al. 2012; reviewed in Andersson and Sandelin 2020). Only considering chromatin opening and acetylation of H3K27, we noticed that some peaks were positive for both just upstream (i.e. in 5') of the *Man1a2* promoter (Fig. 18C; see green arrow). These sequences might correspond to enhancers, although this would require further assessment.

Since the presence of Polycomb repressive complexes (PRCs) can mediate transcriptional repression of their target genes (reviewed in Aloia, Di Stefano, and Di Croce 2013; Blackledge, Rose, and Klose 2015), we looked for the distribution of the repressive H3K27me3 mark, which is deposited by PRC2 (Chammas, Mocavini, and Di Croce 2020), in the *wild-type* region. We noticed that the region was globally devoid of H3K27me3, at the exception of the *Tbx15* gene and the promoter of a gene located on the left of the integration called *Tent5c* (Fig. 18C). Hence, these observations somehow disfavor alterations in Polycomb-mediated repression to be responsible for the observed changes in gene expression, at least in the genital tubercle.

To summarize, as regards the *Wdr3/Man1a2* pair, we favor the “Enhancer model” depicted in Figure 20. However, we note that other possibilities remain, such as local changes in repression by Polycomb/PRC2-independent mechanisms (Fig. 18C; e.g. sequence indicated by the black arrow).

At last, we assessed the expression of genes located at a larger distance from the *TgN(38-40)^{high}* integration site. In particular, we probed the expression of two genes, called *Tbx15* and *Ptgfrn*, which are located on the far left and right of the integration site, respectively (see Fig. 18B). As explained above, *Tbx15*, a gene standing at the left edge of a 800 kb TAD located on the left-hand side of

the *TgN(38-40)^{high}* integration, was contacted by the ectopic CS38 element. In line with previous publications (Agulnik, Papaioannou, and Silver 1998; Kuijper et al. 2005; Singh et al. 2005), we observed that *Tbx15* was expressed in the lateral plate mesoderm, in the limbs, in the proximal portion of the GT and in the mandibular zone at E12.5 (Fig. 19D). No obvious changes in expression could be observed in *TgN(38-40)^{high}* mutant animals. Furthermore, we noticed that this gene was highly covered by H3K27me3 in the GT. Together with the fact that *Tbx15* is located in a different TAD than our construct (see Fig. 18B and C), this raises the possibility that the transgene-*Tbx15* contacts might be mediated by Polycomb, at least in *Tbx15*-negative cells (see Fig. S7 and Discussion). On its side, the *Ptgfrn* gene was separated from the integration site by a complete TAD (see Fig. 18B). We noticed that *Ptgfrn* was overall ubiquitously expressed, without any apparent difference between genotypes (Fig. 19E). Similar to our previous conclusions from *TgN(d11-Evx2)*, these results indicate low or no effect on the expression of genes that are located at a great distance and are well-insulated from the transgene's welcoming site.

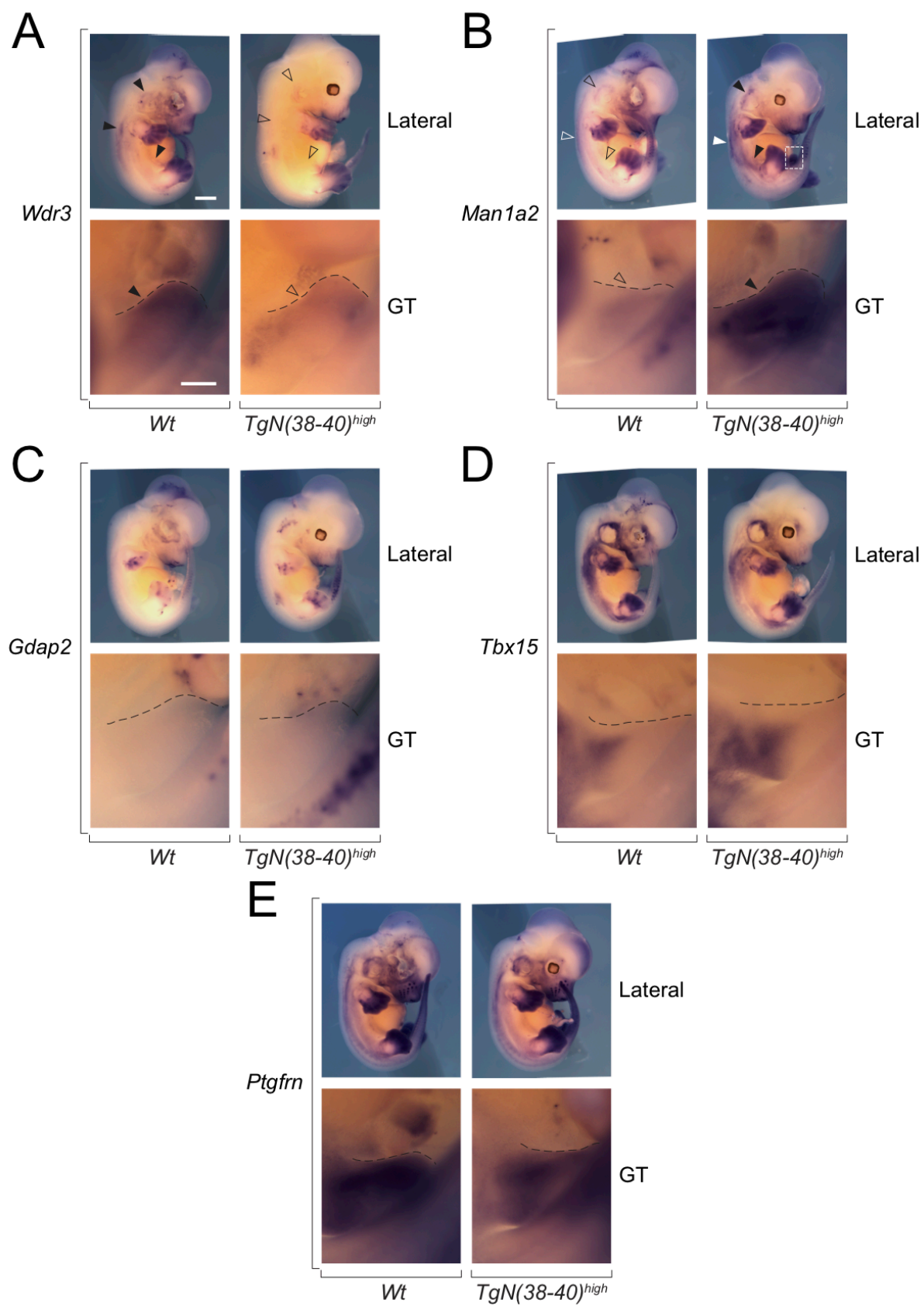


Figure 19: See caption on next page.

Figure 19: Gene expression analyses in *TgN(38-40)^{high}* by WISH on E12.5 embryos. (A) *Wdr3* expression. In the *wild-type*, the expression of *Wdr3* was mostly restricted to the limbs, the lateral plate mesoderm and the genital tubercle (GT; see magnifications below the lateral views). In *TgN(38-40)^{high}*, both mesodermal and GT *Wdr3* levels appeared greatly reduced. (B) *Man1a2* expression. The *wild-type* expression of *Man1a2* was very similar compared to that of *Wdr3*. On the contrary to *Wdr3*, *Man1a2* seemed to be gained in both the GT and the lateral plate mesoderm of a *TgN(38-40)^{high}* embryo. An inset indicates the region magnified for easier GT visualization. (C) *Gdap2* expression. *Gdap2* was almost exclusively expressed in the distal part of the limbs and posterior half of the tail. No particular changes in expression could be observed between *wild-type* and *TgN(38-40)^{high}* individuals. (D) *Tbx15* expression. In both *wild-type* and *TgN(38-40)^{high}* embryos, *Tbx15* was expressed in the lateral plate mesoderm, the limbs, the mandibular zone and in the proximal portion of the GT, with no obvious differences between genotypes. (E) *Ptgfrn* expression. Expression of *Ptgfrn* was overall ubiquitous and identical across genotypes. Arrowheads indicate presence or gain of expression, whereas empty arrowheads indicate lower or loss of expression. Dotted lines delimit the anterior end of the GTs in the magnifications. All *TgN(38-40)^{high}* embryos used here were homozygous for the integration. Scale bars: 1 mm (lateral views) and 250 μ m (GT magnifications).

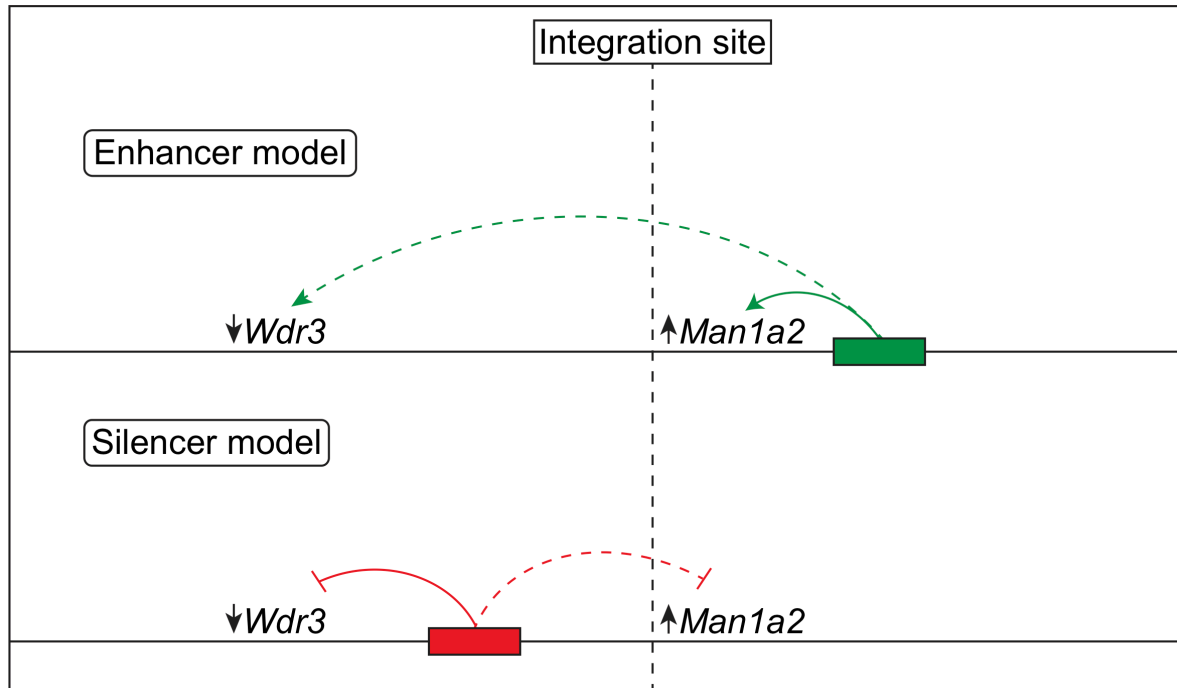


Figure 20: Possible models explaining the apparently opposite expression changes of *Wdr3* and *Man1a2* in the lateral plate mesoderm and the genital tubercle of *TgN(38-40)^{high}*. Top: In the “Enhancer model”, an enhancer (green box) located on the right-hand side of the integration triggers *Wdr3* expression in the *wild-type* (dotted green arrow). By placing a strong barrier to regulatory interactions between this element and *Wdr3*, this enhancer gets disconnected from *Wdr3* and instead activates *Man1a2* expression (solid green arrow). Bottom: In the “Silencer model”, a silencer (red box) located on the left of the integration site prevents *Man1a2* expression in the *wild-type* (dotted red inhibitory line). Upon integration of *TgN(38-40)^{high}* between the silencer and the gene, the silencer turns to the other side, causing *Man1a2* derepression and *de novo* repression of *Wdr3* (solid red inhibitory line). Black arrows pointing upwards or downwards indicate increased or decreased transcriptional activity of the genes in *TgN(38-40)^{high}*, respectively. The *TgN(38-40)^{high}* integration site is indicated by a vertical dotted line.

As previously explained, region CS38-40 was present in multiple copies at the level of the *TgN(38-40)^{high}* integration. This probably imposed very strong structural constraints onto the welcoming chromatin landscape, resulting in major depletions of contacts across the integration site (see Fig. 18B). Hence, the question of whether structural changes would also be observed with a greatly reduced number of copies of this region remained.

For this reason, we turned to another independent line transgenic for the same initial *TgN(38-40)* construct, but which carried the integration on a distinct chromosome. For this line, which we called *TgN(38-40)^{low}* (see below for details), TLA assigned the integration of the construct to a region of chromosome 10, with a forward sequence configuration (i.e. with the + strand of the transgene

connected to the + strand of the chromosome) (Fig. S8B and C). Importantly, analysis of the integration breakpoints also revealed that the right end of the transgene (i.e. the part displaying the transgene-to-chr10 topology) consisted of a partial fragment extending towards CS38 (Fig. S8C) (see below). In addition, it appeared that the integration of the construct generated a duplication of about 600 bp of chromosomal sequence, in the middle of which it got inserted (see Fig. S8C).

Similar to *TgN(38-40)^{high}*, ChIPmentation confirmed the binding of both CTCF and of the cohesin subunit RAD21 on *TgN(38-40)^{low}* in E12.5 limbs (Fig. 21A). In contrast to *TgN(38-40)^{high}*, when mapping the data on chromosome 2, we did not observe any major difference in the background level of the ChIP (and total input) sequencing signal between the extent of the transgene and the surrounding sequences, suggesting that the construct was present in a low copy number in this line. Furthermore, an approximately two times stronger signal for CTCF at the level of the CS38 site than for the other two (see Fig. 21A) led us to infer that the transgene was present in one and a half copies at the level of this integration. In line with this view, Control-FREEC supported the presence of two copies of the segment spanning from the beginning of the *TgN(38-40)* construct to the beginning of CS38 (Fig. S8D). In contrast, the algorithm estimated that the other parts of the construct, i.e. from the rest of CS38 to the end of the construct, were present in a single copy at the level of this integration. This finding was further validated by qPCR transgene quantification on genomic DNA (Fig. S9).

In order to fully confirm the topology and extent of the *TgN(38-40)^{low}* integration, we performed MinION long-read sequencing after release by CRISPR-Cas9 (nanopore Cas9-targeted sequencing or nCATS) (Gilpatrick et al. 2020). MinION is a next-generation sequencing method that uses biological nanopores to sequence DNA or RNA molecules (reviewed in Jain et al. 2016; Lu, Giordano, and Ning 2016). In brief, following adapter ligation, a DNA or RNA molecule is coupled to a motor protein that translocates it through a nanopore, which sits on an electrically resistant membrane. The sequence of the molecule is determined based on small variations of the current passing through the pore during translocation, which directly depends on the specific base occupying the pore at a given time (Lu, Giordano, and Ning 2016).

Of particular interest, MinION-based techniques are capable of sequencing reads of up to 45 kb with a relatively good accuracy (Lu, Giordano, and Ning 2016), although they generally show improved performances for reads not exceeding 25 kb (Goodwin et al. 2015; Laver et al. 2015). Therefore, such long-

read sequencing methodologies have been invaluable for the sequential determination of features along individual molecules, as exemplified by endogenous repetitive regions or tandem-integrated transgenes, which cannot be easily analyzed using conventional short-read sequencing approaches (Lu, Giordano, and Ning 2016; Giesselmann et al. 2019; Nicholls et al. 2019). In the nCATS derivative, genomic regions of interest are released by guide RNAs (gRNAs)-directed Cas9 cleavage prior to adapter ligation (see Material and Methods) (Gilpatrick et al. 2020). In this way, nCATS (or other similar methods) enables to enrich the solution in sequencing-competent DNA molecules corresponding to the region of interest for improved sequencing coverage without requiring PCR amplification (Gabrieli et al. 2018; Giesselmann et al. 2019; Gilpatrick et al. 2020).

Based on the aforementioned results and on a first attempt of *in silico* genome reconstruction, we designed single-guide RNAs (sgRNAs) in the chromosomal regions located both upstream and downstream of the integration site, as well as at regular (approximately 10 kb) intervals within the *TgN(38-40)* construct. Next, we released the sequences following a tiling approach (Fig. S10A), in which two distinct combinations of guides were used to cut overlapping fragments ranging from 9.5 to 23 kb in two independent reactions. Fragments originating from each of the two reactions were then pooled together and sequenced, allowing to properly characterize the mutant configuration from one end to the other. To facilitate analysis, the MinION data were displayed onto a mutant genomic configuration reconstructed from previous experiments (namely TLA, Control-FREEC and qPCR), consisting of one and a half times region CS38-40 (see Fig. S10A).

Analysis of the MinION sequencing coverage revealed a strong (about 30%) enrichment for the region of interest compared to the rest of the genome (Fig. S10B). Importantly, inspection of individual MinION reads confirmed the presence of a head-to-tail tandem configuration of the construct at the level of the *TgN(38-40)^{low}* integration (Fig. S10C). Moreover, the MinION results validated that the second partial copy of the transgene included the CTCF site of CS38. Hence, together with previous evidence, these results fully confirmed the presence of one and a half copies of the *TgN(38-40)* construct (i.e. 63.2 kb, including 4 left-oriented CTCF sites) in this particular integration.

From the structural point of view, the *TgN(38-40)^{low}* integration occurred within a 1.2 Mb TAD of chromosome 10 (Fig. 21B). This domain displayed a relatively low density of CTCF sites and contained only a single protein-coding

gene called *Btg1*. We first assessed whether the elements of the *TgN(38-40)* construct also engaged in interactions in this new landscape. The data were mapped onto the mutant genome reconstituted from previous assays for an easier interpretation of the results. Using both CS38 and CS40 as 4C-seq viewpoints in E12.5 limbs, we noticed an establishment of contacts within the welcoming region, which were more frequent within the limits of the *TgN(38-40)^{low}*-containing TAD and peaked at CTCF sites that predominantly displayed a convergent orientation relative to the ones of the transgene (Fig. 21B). The highest contact frequencies were observed at (or close to) the boundaries of the welcoming TAD. Since the CTCFs at the right border of the domain are in a reverse orientation, meaning in the same orientation than the ones of the transgene, the contacts observed between the transgene and the right TAD border probably resulted from the stalling of loops that formed on the right hand-side of the integration by other loops, anchored at the level of the transgene (see below and Discussion).

We then asked whether the integration of the *TgN(38-40)* construct in one and a half copies caused changes in local chromatin interactions within the welcoming region. To that end, we designed three different 4C-seq viewpoints: one at a forward CTCF of the left edge of the welcoming TAD (CTCF-left), another on the 3' UTR of the *Btg1* gene (3'*Btg1*; about 20 kb aside from a forward CTCF), and a last one at a reverse CTCF of the right border of the domain (CTCF-right). As expected, analysis of the interactions established by these viewpoints in *wild-type* limbs revealed that the contacts were almost exclusively restricted to the welcoming TAD (Fig. 21B). For both the CTCF-left and right viewpoints, the contacts spread throughout the domain and reached their maximum intensity at the CTCFs of the opposite limit of the TAD. On its side, the 3' UTR of the *Btg1* gene appeared to establish more discrete contacts, which barely extended towards the domain's boundaries. In samples that were homozygous for the *TgN(38-40)^{low}* integration, we noticed an establishment of contacts from both CTCF-left and right viewpoints towards the elements of *TgN(38-40)^{low}*. In contrast, only very few contacts towards *TgN(38-40)^{low}* were scored when using the 3'*Btg1* viewpoint, further supporting the notion that interactions involving both the transgene and the surrounding landscape depend on CTCF. Importantly, we noticed that the interactions established by each viewpoint were decreased beyond the integration site. For instance, when looking from the CTCF-left viewpoint, the contacts were impaired on the right hand-side of the integration; an effect that was particularly pronounced at the CTCFs of the right TAD limit. A similar trend, though inversed with respect to the affected side relative to the integration (as expected from the position of the viewpoint), was observed for

the CTCF-right viewpoint. Since this last viewpoint matched with a CTCF in reverse orientation, this observation further supports loop-loop stalling (see Discussion). Taken together, these results demonstrate that the novel contacts established between the transgene and its new landscape occurred at the expense of endogenous interactions, suggesting the reconstitution of a functional boundary (see below).

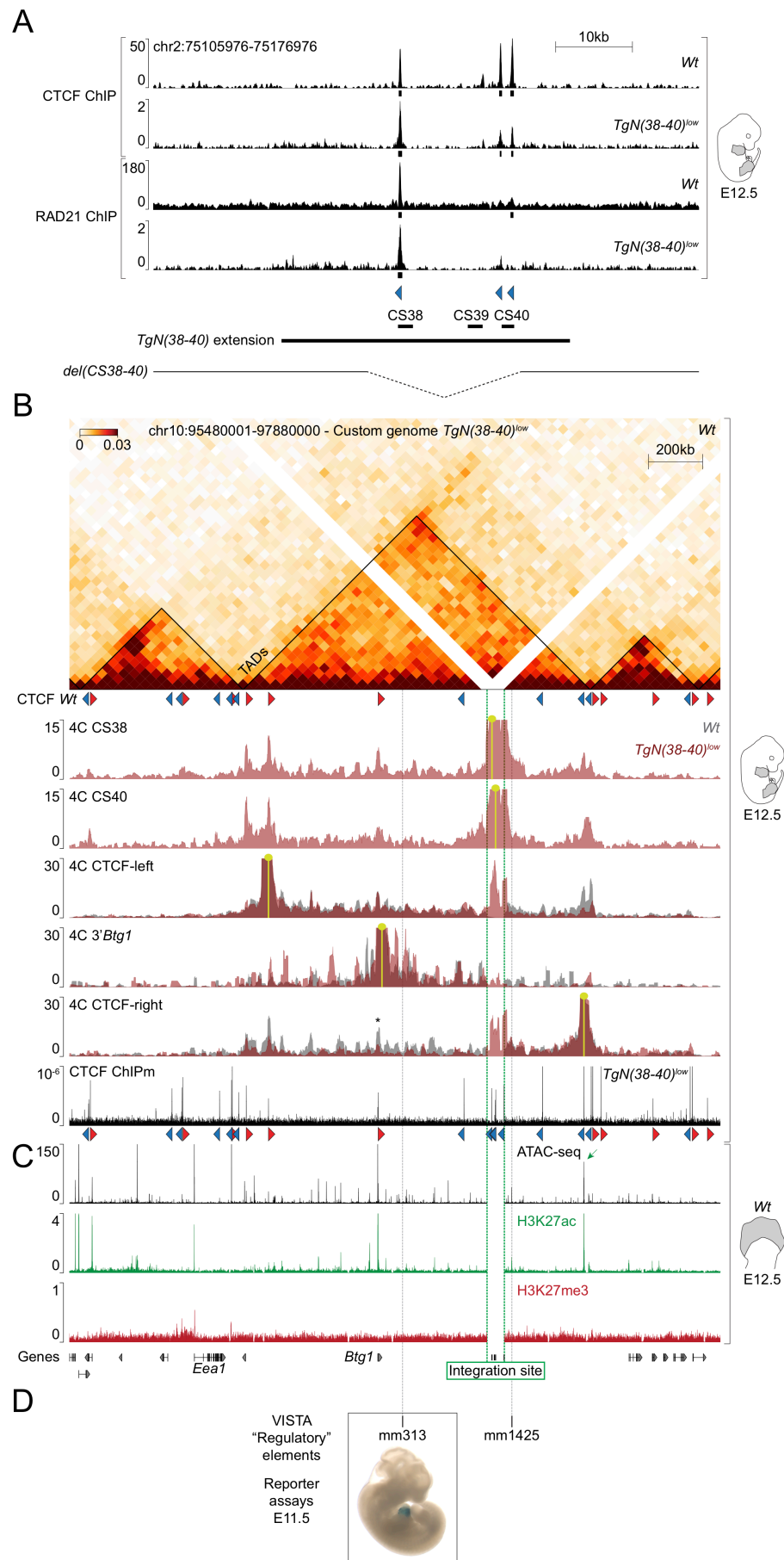


Figure 21: See caption on next page.

Figure 21: Structural changes upon integration of *TgN(38-40)^{low}* suggest retention of boundary function in a novel context. (A) Recruitment of CTCF and RAD21 on *TgN(38-40)^{low}*. Both CTCF and RAD21 signals were higher for the CTCF site of CS38. As regards CTCF, this indicated the presence of region CS38 in two copies at the level of this integration. The binding was assessed using ChIP-seq on the endogenous (*wild-type*) region CS38-40 (n=1 for both CTCF and RAD21; data from (Rodríguez-Carballo et al. 2017)), or ChIPmentation on *TgN(38-40)^{low}* (n=2 for both CTCF and RAD21), in whole limbs at E12.5. The region displayed is chr2:75105976-75176976 (mm10). Under each track, a MACS2 peak calling is represented as black boxes. The extension of the initial *TgN(38-40)* fosmid (chr2:75122684-75160161) is indicated below. For *TgN(38-40)^{low}*, the ChIPm experiments were performed on samples that were heterozygous for the integration and homozygous for the *del(CS38-40)* genetic background, which is shown at the bottom. (B) Welcoming TAD-restricted contacts from the transgene and alterations of the local chromatin landscape. All data of this panel were mapped onto a mutant genome carrying the insertion of one and a half copies of the *TgN(38-40)* construct at the indicated location. On top is shown a *wild-type* Hi-C of the welcoming region in whole limbs at E12.5. TAD identification was achieved using hicFindTADs with a window size of 240 kb. Of note, mapping of the *wild-type* Hi-C data on the mutant genome resulted in an artefactual splitting of the welcoming TAD at the level of the *Btg1* CTCF site due to the lack of mappable reads (see Fig. S11A, B and C). Therefore, we represented here the TAD calling obtained when mapping the same data onto the *wild-type* genome, adapted to the bin shift imposed by the integration. Below the Hi-C heatmap are shown the results of 4C-seq experiments using regions CS38 and CS40 as viewpoints in *TgN(38-40)^{low}* E12.5 whole limbs. Note that more frequent contacts occurred within the limits of the welcoming TAD. Highest contact frequencies were observed around the limits of the domain and at endogenous CTCFs that are convergent relative to the ones of the transgene, whose orientation is shown at the bottom of panel (B). 4C-seq using viewpoints within or at the limits of the welcoming TAD in E12.5 whole limbs. The *wild-type* tracks are in gray and the *TgN(38-40)^{low}* tracks in maroon. The asterisk highlights a decrease in the contacts established between the CTCF-right viewpoint and the forward CTCF of the *Btg1* gene promoter. All *TgN(38-40)^{low}* samples used for 4C-seq were homozygous for the integration and all 4C-seq experiments were performed in a single biological replicate. The region displayed is chr10:95480001-97880000 of the mm10+*TgN(38-40)^{low}* custom genome (extent corresponds to chr10:95480001-97816188 of *wild-type* mm10). (C) The last three tracks represent regions of open chromatin (ATAC-seq in black; n=1) and coverage of H3K27ac (green) or H3K27me3 (red) assessed by ChIP-seq (n=2 for H3K27ac; n=8 for H3K27me3) in *wild-type* distal forelimbs at E12.5. ATAC-seq data are unpublished and were shared by Christopher Chase Bolt. H3K27ac and H3K27me3 ChIP-seq data are from (Rodríguez-Carballo et al. 2019). The green arrow indicates a putative enhancer (ATAC-seq+; H3K27ac+), whose disconnection from *Btg1* might explain the changes in expression of this gene reported in Figure 23. (D) Validated (mm313) or putative (mm1425) regulatory elements annotated in the VISTA Enhancer Browser for the region shown in (B) and (C). Note that only the mm313 element displayed reproducible enhancer activity at E11.5 (heart; picture retrieved from the VISTA Enhancer Browser).

On the basis of the above observations, we sought to determine if the ectopic *TgN(38-40)^{low}* integration actually resulted in the splitting or disruption of the welcoming TAD. To that end, we performed Hi-C on E12.5 whole limb samples that were homozygous for the integration. By doing so, similar to the 4C-seq, we observed a clear depletion of contacts spanning the integration site in *TgN(38-40)^{low}* compared to the *wild-type* (Fig. 22). Nevertheless, although they were decreased, some interactions appeared to be maintained to a certain level even in this situation. Typically, a stripe of contacts established by the left boundary of the TAD towards the right limit could still be observed (Fig. 22B; see black arrowhead). To evaluate the significance of these results, we performed TAD identification based on insulation using the hicFindTADs tool of the HiCExplorer package (Ramírez et al. 2018). We first applied a window size of 240 kb, as previously used with TopDom (Shin et al. 2016; Ramírez et al. 2018). At these parameter values, a new boundary was identified precisely at the level of the integration site, resulting in the splitting of the welcoming TAD (see Figs. 22B and S11A). However, this behavior was not observed with larger window sizes (Fig. S11B, C and D). Hence, these results demonstrate the reconstitution of a relatively weak topological boundary at the level of the *TgN(38-40)^{low}* integration site (see Discussion), consistent with the definition of region CS38-40 as a sub-TAD border in its original context (Rodríguez-Carballo et al. 2019; 2020).

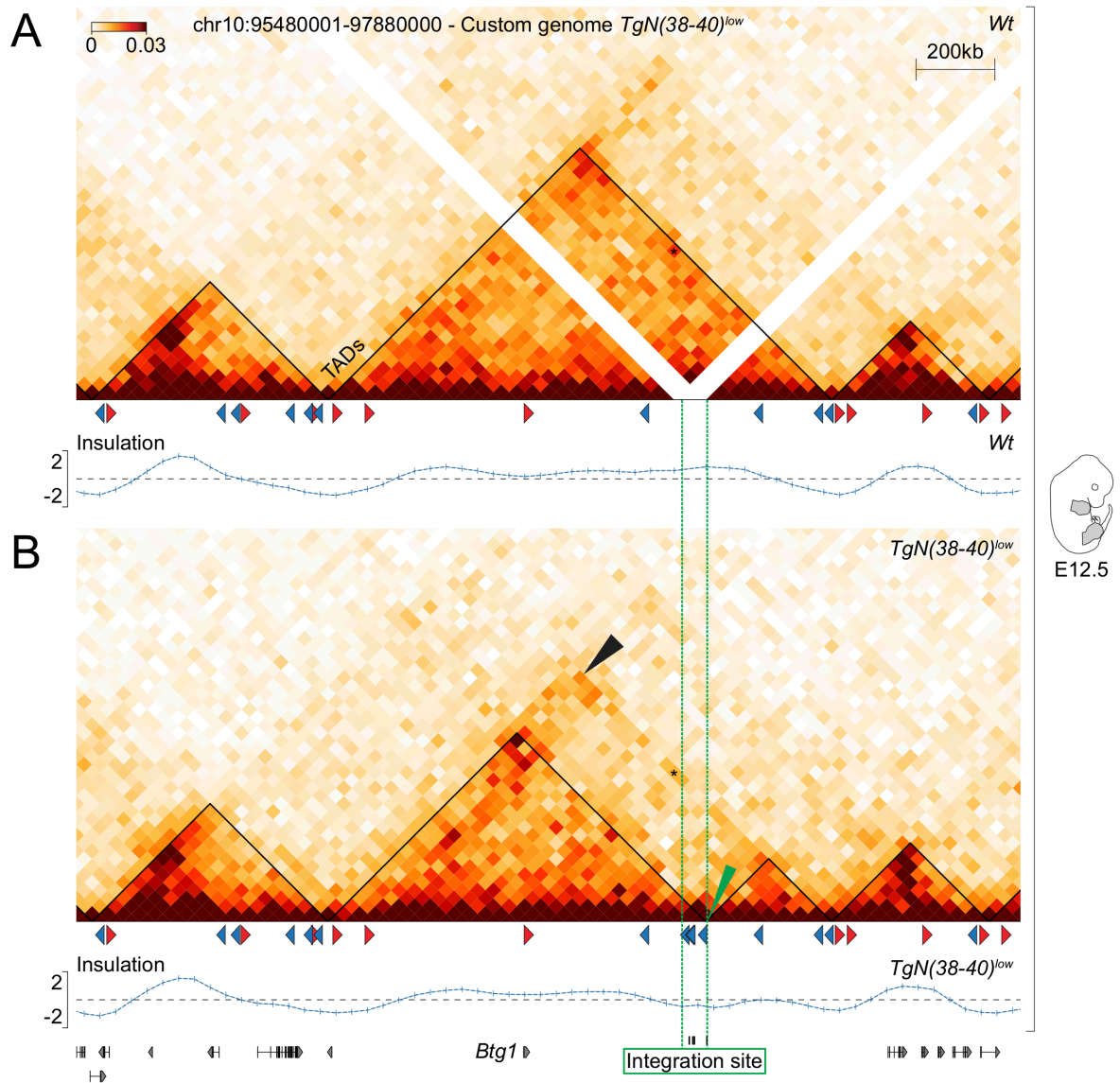


Figure 22: Ectopic sub-TAD boundary formation. (A) *Wild-type* Hi-C heatmap of whole limbs at E12.5. The data were displayed in the same manner as for Figure 21. The welcoming (*Btg1*-containing) TAD appears as a single domain. The position and orientation of the endogenous CTCFs is indicated below the *wild-type* Hi-C heatmap. Insulation from hicFindTADs is shown at the bottom of the pannel. (B) *TgN(38-40)^{low}* Hi-C heatmap of whole limbs at the same stage. Note that the welcoming TAD is splitted in two at the level of the integration site (green arrowhead), as a consequence of a local drop in the insulation score. The position and orientation of both endogenous and ectopic CTCFs is shown below the *TgN(38-40)^{low}* Hi-C heatmap. The black arrowhead indicates the stripe of contacts maintained in *TgN(38-40)^{low}*. The asterisk highlights contacts established between the reverse CTCFs at the right limit of the welcoming TAD and the forward CTCF of the *Btg1* gene promoter, which are decreased in *TgN(38-40)^{low}* (see text for details).

At last, we asked whether the integration of *TgN(38-40)^{low}* affected the expression of the only protein-coding gene located within the welcoming TAD, *Btg1* (see Figs. 21B and 22). The *Btg1* gene is implicated in the regulation of cell proliferation, probably by controlling the G1 to S phase transition of the cell cycle (Farioli-Vecchioli et al. 2012). By whole mount *in situ* hybridization, we observed that *Btg1* was overall ubiquitously expressed in E12.5 *wild-type* embryos (Fig. 23). The strongest transcript signals were observed in the facial region, in the distal part of the limbs as well as in the mammary buds. In an embryo that was homozygous for the *TgN(38-40)^{low}* integration, we noticed that the *Btg1* expression pattern was globally maintained (see Fig. 23). However, it appeared that the transcript levels were somehow decreased in the lateral plate mesoderm, in the mammary buds and to a lower extent in the forelimbs.

In order to get indication of whether a nearby *cis* regulatory elements could be disconnected from the *Btg1* gene upon integration of *TgN(38-40)^{low}*, possibly as a result of the TAD splitting, we went back to the VISTA Enhancer Browser. Two enhancer candidates were described within the welcoming region (mm313 and mm1425; see Fig. 21D), but only the mm313 element displayed reproducible enhancer activity in the VISTA reporter assays. Since the activity pattern of this enhancer was restricted to the heart at E11.5, not matching the apparent changes we observed, we evaluated candidate regulatory elements within this region based on chromatin accessibility and histone modifications. We took advantage of unpublished ATAC-seq (courtesy of Christopher Chase Bolt) and published H3K27ac and H3K27me3 ChIP-seq data (Rodríguez-Carballo et al. 2019) from our group in E12.5 distal forelimbs. Inspection of H3K27me3 in the *wild-type* revealed an almost complete absence of this chromatin mark from the welcoming region (Fig. 21C), hence strongly arguing against the implication of an endogenous Polycomb response element on the apparent changes in the expression of *Btg1*. Of particular interest, we noticed the presence of a highly accessible (ATAC-seq-positive) and strongly H3K27 acetylated peak at the right TAD border (Fig. 21C; see green arrow), which precisely matched with the CTCF-right viewpoint. We previously observed interactions between this viewpoint and the convergently oriented CTCF located at the promoter of the *Btg1* gene, and that these contacts were impaired upon integration (Figs. 21B and 22B; see asterisk). Therefore, it is tempting to speculate that the right limit of the TAD contains an enhancer that is disconnected from the *Btg1* gene upon integration and associated TAD splitting, explaining the apparent changes in the expression of this gene. Nevertheless, we note that this would require further investigation (see Discussion).

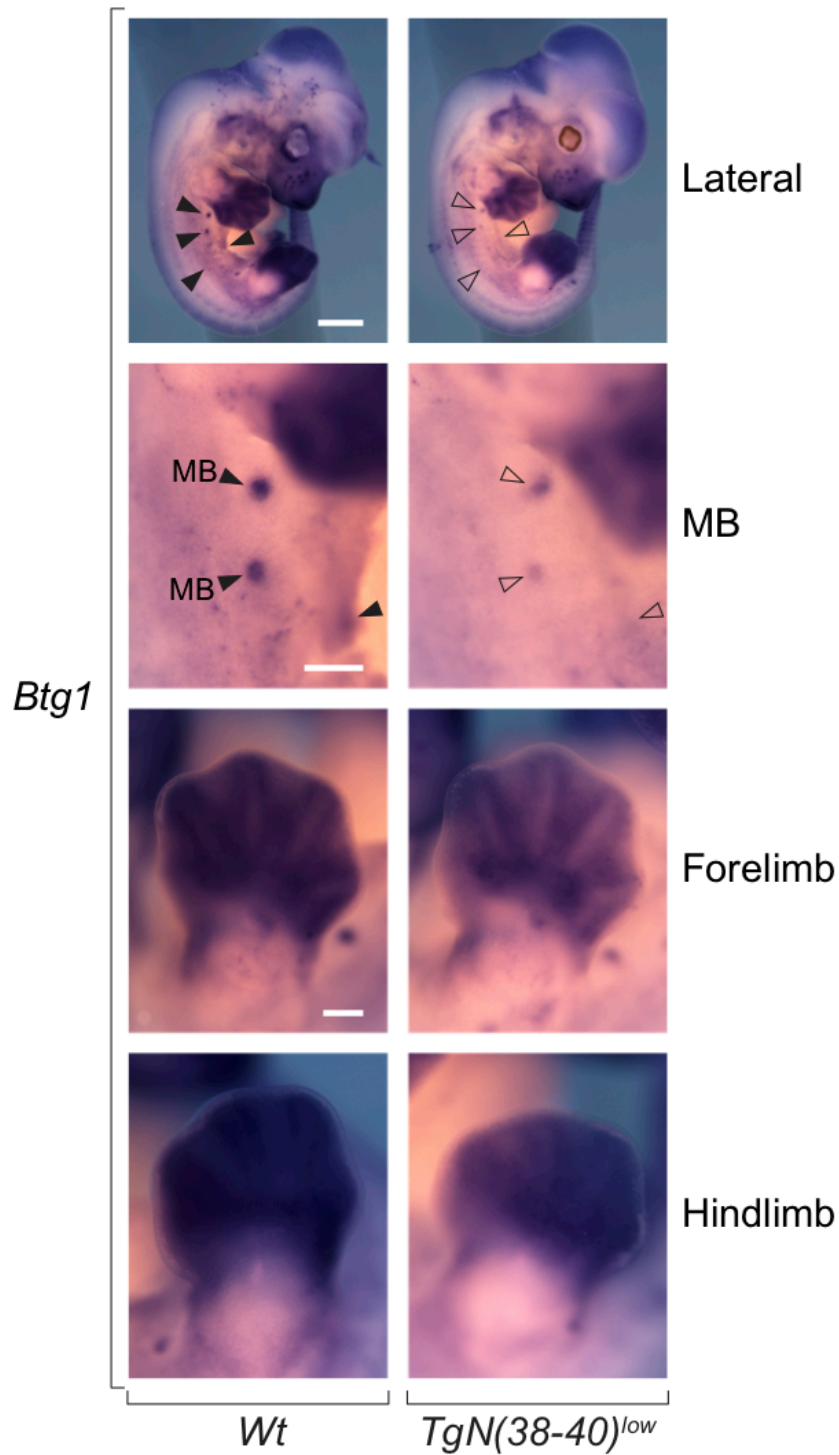


Figure 23: *Btg1* gene expression analysis in *TgN(38-40)^{low}* by WISH on E12.5 embryos. In the *wild-type*, *Btg1* was overall ubiquitously expressed, with stronger mRNA levels in the facial region, the distal part of the limbs and in the mammary buds (MB). This general expression pattern was maintained in a *TgN(38-40)^{low}* homozygous embryo, although with an apparent reduction of the transcript levels in the lateral plate mesoderm, mammary buds and to a lower extent in the forelimbs. Arrowheads indicate presence of expression whereas empty arrowheads indicate decreases in expression. Scale bars: 1 mm (lateral views) and 250 μ m (MB and limbs magnifications).

Discussion:

1. *Modulating boundary potential at the endogenous HoxD cluster*

Using intra-cluster duplications increasing the number of both CTCF-binding sites and gene promoters within *HoxD*, we first showed that these alterations resulted in a decrease of the contacts established by the central *Hoxd* genes and the TAD located beyond the duplicated segment. Together with previous evidence taking the reverse approach, in which partial deletions of the gene cluster resulted in an increase of interactions going through the boundary (Rodríguez-Carballo et al. 2017), the above results strongly suggest that the strength of the TAD boundary at the *HoxD* cluster can be modulated by playing with the number of boundary elements it contains.

Importantly, the structural modifications we observed could be the result of either an increase in the number of CTCF sites, of gene promoters, in the linear genomic distance, or a combination thereof. While it is now clear that CTCF participates in the formation of most TAD boundaries in the genome (see Dixon et al. 2012; Nora et al. 2012; Zuin et al. 2014; Sanborn et al. 2015; Fudenberg et al. 2016; Nora et al. 2017), the possible contribution of enhancer-promoter interactions to boundary function *per se* remains largely unexplored to date. Nevertheless, several studies reported that such regulatory interactions significantly contribute to the internal structure of TADs (Phillips-Cremins et al. 2013; Smith et al. 2016; Bonev et al. 2017; also see Pombo and Dillon 2015). Hence, this raises the possibility that regulatory interactions between gene promoters and *cis*-acting elements may hold topological and possibly boundary potential, at least to some extent.

In order to decipher between these different possible contributions, we note that another experimental system should be designed so as to increase either the number of CTCF-binding sites or of gene promoters at a given location, rather than both at a time.

2. *Structure versus function at the HoxD locus*

In the duplications analyzed in this work, the apparently decreased access of central *Hoxd* gene promoters to the enhancers located within the *HoxD*-associated TADs also correlated with expression changes at the level of the cluster reported in previous publications from our group (see below) (Kmita et al. 2002; Tarchini and Duboule 2006).

Importantly, the changes in expression precisely matched what was expected from such genetic alterations, based on knowledge about the

implications of the regulatory landscapes that are flanking the *HoxD* cluster, and in particular of the enhancers they contain, for the control of *Hoxd* genes' expression during limb development. Indeed, the expression of *Hoxd11* was reduced in the proximal part of the limbs upon duplication of the region *8i-10* (*Dup(8i-10)*) (Tarchini and Duboule 2006), located in-between this gene and the telomeric landscape, which is known to contain the enhancers responsible for the proximal regulation (Andrey et al. 2013; Yakushiji-Kaminatsui et al. 2018). In contrast, the expression of *Hoxd10* was only decreased in the distal part of the limbs in the duplication *11-13* (Kmita et al. 2002), designed on the centromeric side of the gene, probably by impairing the communication between the latter gene and the distal-specific enhancers of the centromeric landscape (Spitz, Gonzalez, and Duboule 2003; Gonzalez, Duboule, and Spitz 2007; Montavon et al. 2011; Lonfat et al. 2014).

In line with the above, the structural modifications we observed in both duplications were highly consistent with the previously reported changes in gene expression, suggesting the existence of a causal link. In particular, interactions between *Hoxd11* and the T-DOM's regions CS38-40 (which contains the CS39 enhancer), CS93 as well as CS65 were decreased in the proximal part of the limbs in *Dup(8i-10)*. Furthermore, as regards *Dup(11-13)*, the contacts between the central *Hoxd9* gene and almost all known enhancers of the C-DOM were reduced in the distal part of the limbs compared to the *wild-type*.

In addition, our observations are also in line with the results obtained in another study from our group, in which partial deletions of the *HoxD* cluster located within the *Hoxd8-Hoxd13* interval were used (Rodríguez-Carballo et al. 2017). Indeed, in this situation, the remaining genes gained interactions with the enhancers located beyond the deleted segment and became ectopically expressed in patterns that were expected from such communications.

From a quantitative perspective, the changes in contacts we observed between non-duplicated central *Hoxd* genes and the *HoxD*-associated TADs in both duplications were comprised in the range of 11 to 16%. In particular, we scored an 11% contact redirection from the T-DOM to the C-DOM when using *Hoxd11* as a viewpoint in the proximal limb portion of *Dup(8i-10)* individuals, against a 16% change, going from the C-DOM to the T-DOM, when *Hoxd9* was used in the presumptive digits of *Dup(11-13)* embryos. These results strongly indicate that the contacts were affected in only a relatively reduced proportion of limb cells.

Consistently, the observed changes in expression were also relatively mild, comprised in a similar range compared to the structural modifications. Hence, these results suggest that enhancer-promoter interactions are highly resilient, since they appear to largely fulfill their function (i.e. in most cells) even upon reinforcement of the TAD boundary at the *HoxD* cluster. In this particular context, comparable results were obtained when inverting the entire T-DOM, including its right border, while leaving the *HoxD* cluster in place (Rodríguez-Carballo et al. 2020). Indeed, despite the presence of a topological boundary between the gene cluster and the enhancers of the T-DOM in this configuration, enhancer-promoter communications appeared to be maintained to some extent. More precisely, while the onset of expression at the level of the *HoxD* cluster was severely impacted at around embryonic day 10, some enhancer-promoter contacts could still be scored. Furthermore, central *Hoxd* genes' transcripts were present in proximal limbs at E12.5 and only a minor truncation of their expression domains was observed (Rodríguez-Carballo et al. 2020).

Finally, it also appeared that the decrease in the contacts established with one TAD sometimes correlated with an increase in interactions towards the opposite domain, which could indicate a redirection of contacts resulting from the reinforcement of the boundary. Interestingly, this effect was particularly pronounced for *Hoxd9* – CS38-40 contacts in the distal part of the limbs upon duplication of the region 11-13, but was not evident in proximal cells. Knowing that the T-DOM is repressed in distal cells (Andrey et al. 2013), the gain of contacts between *Hoxd9* and region CS38-40 observed in this tissue might also reflect an increase in Polycomb-mediated long-range interactions, typically in *Hoxd9*-negative distal cells (Andrey et al. 2013; Denholtz et al. 2013; Schoenfelder et al. 2015; Rodríguez-Carballo et al. 2017).

3. A possible contribution of promoter competition

The evidence mentioned above indicates that the changes in gene expression might be due to an increased insulation, resulting from the duplications, between the relevant enhancer-promoter pairs. However, we cannot rule out the possible contribution of in-*cis* competition between gene promoters for enhancer access to the observed effects (Kmita et al. 2002).

Key studies on the globin gene clusters showed that gene promoters are able to compete with one another for the access to an enhancer located in *cis* and subsequent transcriptional activation (Jane et al. 1992; Yu et al. 2006; Rincón-Arano et al. 2009; Bartman et al. 2016). In brief, the more a promoter is competitive, probably depending on sequence features allowing the recruitment

of specific transcription factors (Jane et al. 1992), the more it can receive regulatory inputs from an enhancer at the expense of other genes within a given globin cluster. Of particular interest, a recent study reported that forcing interactions between the γ -globin gene and an enhancer shared with the β -globin gene significantly reduced the simultaneous expression of the latter from the same allele (Bartman et al. 2016). In our context, the decreased interactions between central *Hoxd* genes and the enhancers of the TAD located beyond the duplicated segment as well as the changes in expression might be, at least partially contributed by duplicated gene promoters (see below).

In both *Dup(8i-10)* and *Dup(11-13)*, three gene promoters are duplicated: from *Hoxd8* to *Hoxd10* for *Dup(8i-10)* and from *Hoxd11* to *Hoxd13* in *Dup(11-13)*. Considering *Dup(8i-10)*, as all genes from *Hoxd8* to *Hoxd10* are able to interact with the T-DOM enhancers during the proximal regulation (Andrey et al. 2013; Rodríguez-Carballo et al. 2017), a competition with the *Hoxd11* promoter could possibly explain the decrease of *Hoxd11* expression in the proximal part of the limbs. As regards *Dup(11-13)*, it is known that genes from *Hoxd11* to *Hoxd13* contact the C-DOM enhancers, either constitutively (*Hoxd13* and *Hoxd12*) or specifically for the distal regulation (*Hoxd11*) (Andrey et al. 2013; Rodríguez-Carballo et al. 2017; also see Lonfat and Duboule 2015), and hence could compete with *Hoxd9* and *10*, resulting in a decreased digital expression of the latter genes.

Interestingly, when examining the interactions established by the *Hoxd4* gene (which stands on the telomeric side of both duplicated alleles), although the contacts towards region CS38-40 were decreased in both duplications, this effect was more pronounced in *Dup(8i-10)* (see Figs. 9C and 11C). Since a larger number of gene promoters able to engage in strong interactions with region CS38-40 are duplicated in *Dup(8i-10)* (i.e. *Hoxd8*, *Hoxd9* and *Hoxd10*) compared to *Dup(11-13)* (only *Hoxd11*) (Andrey et al. 2013; Rodríguez-Carballo et al. 2017), these results suggest a competition occurring between the promoter of *Hoxd4* and duplicated gene promoters for the access to region CS38-40. However, the difference between the two duplications could also be related to the number or orientation of duplicated CTCF sites in each case. In addition, we note that the above explanations are not mutually exclusive. For instance, the duplication of telomeric-oriented CTCF sites in *Dup(8i-10)* might have an additive effect over promoter competition for the loss of *Hoxd4* – CS38-40 interactions. In this context, an assessment of *Hoxd4* expression would be needed to determine whether the functional output is actually affected in these conditions or not.

To fully elucidate the contribution of promoter competition to the observed structural and expression-related effects, a system enabling to specifically delete the duplicated gene promoters, for example based on small sequence differences between the duplicated versions of the genes, would be needed.

4. Preservation of a *HoxD*-like contact directionality in a new context

By analyzing a transgenic line in which the entire *HoxD* cluster was relocated to a region of chromosome 10 (*TgN(HoxD)*), we showed that genes located at the extremes of the cluster established directional interactions within the surrounding landscape, similar to their behavior at the endogenous *HoxD* locus (Andrey et al. 2013; Rodríguez-Carballo et al. 2017). Indeed, whereas the contacts of the *Hoxd13* gene promoter were stronger with regions located on the left-hand side of the integration, the ones of *Hoxd4* were higher towards the right-hand side. We note that this behavior likely relies on the relative position of the genes within the relocated cluster (see Darbellay et al. 2019), although this would require further validation, typically by engineering intra-cluster inversions in the new context.

Of particular interest, most contacts established by *Hoxd13* and *Hoxd4* were constrained within the limits of the TAD in which the *TgN(HoxD)* transgene integrated, consistent with the definition of TADs as domains of increased interaction frequencies (Dixon et al. 2012). Moreover, both the CTCF protein and the cohesin subunit RAD21 were recruited to their cognate binding sites on the *HoxD* cluster at the new location. In the light of the predictions from the loop extrusion model (Sanborn et al. 2015; Fudenberg et al. 2016), the above evidences indicate the reconstitution of a functional *HoxD* TAD boundary *ex situ*.

The fact that the *TgN(HoxD)* transgene is large (spanning 230 kb) and comprises all genes and almost all CTCFs (in both possible orientations) of the original *HoxD* cluster prompted us to follow a different approach in order to fully determine whether a boundary would be able to function ectopically. To that end, aiming at reducing possible confounding effects coming from a large insert size and numerous gene promoters, we turned to a fragment of the *HoxD* cluster, centered around the expected position of the dynamic *HoxD* boundary (i.e. the *Hoxd11-Evx2* segment) (Andrey et al. 2013; Rodríguez-Carballo et al. 2017), as well as to the CS38-40 sub-TAD border of the T-DOM (Andrey et al. 2013; Rodríguez-Carballo et al. 2019; 2020).

5. Alterations of the local chromatin landscape at ectopic locations

A previous attempt of local structural assessment upon ectopic TAD boundary integration was performed using the *Firre* locus, which encodes a long non-coding RNA (lncRNA) involved in the formation of chromosomal structures called super-loops that are specific to the inactive X chromosome (Xi) (Barutcu et al. 2018). The endogenous *Firre* locus displays a high density of CTCF-binding sites (i.e. 12 sites for the 72 kb of the gene) and is located at the border between two TADs. In this study, the authors generated a construct comprising the ca. 6 kb *Firre* cDNA, which contains only one CTCF-binding site, fused to a doxycycline-inducible promoter. This construct was used for random transgenesis in mice, after which they carried out Hi-C in order to evaluate structural changes. By doing so, no significant modifications of the structural landscape, such as TAD disruption, were scored among multiple integration events across the genome (Barutcu et al. 2018). In addition, inducing the expression of the ectopic *Firre* cDNAs using doxycycline did not alter TAD structure, in line with previous observations that forced gene expression is not sufficient to generate new boundaries (Bonev et al. 2017; Barutcu et al. 2018).

Nevertheless, it should be noted that the Hi-C experiments of (Barutcu et al. 2018) were performed on heterozygous samples, in which one homolog for each of the integrated chromosomes was devoid from integration and hence could not be distinguished from the one carrying the transgenic construct. Therefore, apart from the fact that only one CTCF-binding site and one promoter could possibly be insufficient to disrupt a TAD (see below), the superimposition of the Hi-C signal coming from the non-integrated homologs might also account for the preservation of TAD structure in their datasets.

Using samples that were homozygous for the *TgN(d11-Evx2)* integration, we observed only minor alterations of the surrounding chromatin landscape. Indeed, 4C-seq using viewpoints located within the welcoming TAD demonstrated that the amount of contacts spanning the integration site was very similar between the integrated and *wild-type* situations. However, an extensive analysis of this integration revealed that only the reverse *Hoxd12-Hoxd13* intergenic CTCF site was preserved from the original construct, together with the promoters of the *Hoxd12* and *Hoxd11* genes. Hence, such minor effects have at least two non-mutually exclusive explanations. Firstly, the intrinsic properties of the welcoming TAD could play a role in its response to the presence of ectopic boundary elements. In particular, the *TgN(d11-Evx2)* welcoming TAD in chromosome 7 was characterized by remarkably low contact frequencies, likely reflecting the presence of each of its loops in a small fraction of the cells

under study (see de Wit and de Laat 2012; Giorgetti and Heard 2016). Secondly, the greatly reduced number of ectopic boundary elements that were present in this situation might explain the mild structural changes, in line with the results of (Barutcu et al. 2018).

In contrast, however, when analyzing a line carrying an integration of region CS38-40 in multiple copies in *cis* ($TgN(38-40)^{high}$), representing an ectopic array of at least 12 CTCFs along a 180 kb genomic distance, we noticed drastic alterations of the local chromatin landscape. Indeed, most endogenous contacts going across the integration site were depleted in $TgN(38-40)^{high}$ homozygous samples, likely reflecting the blockage of chromatin loops by the integration in a vast majority of cells. This probably arose as a consequence novel chromatin loops established between the transgene on one side and discrete regions of the TAD on the other, particularly with CTCFs in convergent orientation relative to the ones of the transgene. It is noteworthy that the above can also account for the blockage of contacts coming from the right-hand side of the integration, since loops formed on this side are likely to be stalled by loops that are now anchored at the level of the transgene.

In order to get closer to the original situation of the three reverse CTCFs of region CS38-40 (Rodríguez-Carballo et al. 2017), we conducted structural assessments in the $TgN(38-40)^{low}$ line, in which four ectopic reverse CTCFs are present. This revealed similar structural modifications compared to $TgN(38-40)^{high}$, although with seemingly weaker depletions of contacts going through the integration site. It is noteworthy that the quantitative differences in the structural alterations observed between $TgN(38-40)^{high}$ and $TgN(38-40)^{low}$ are probably related to the different numbers of ectopic CTCF sites that are present in each case. Nevertheless, a complete validation of this hypothesis would require several mutant lines with different numbers of CTCF sites at the same ectopic location.

6. Reconstitution of a functional sub-TAD boundary ectopically

The architectural changes observed in $TgN(38-40)^{low}$ raised the question of whether the welcoming TAD was disrupted by the integration in this situation. Indeed, analysis of $TgN(38-40)^{low}$ homozygous embryos by Hi-C revealed the formation of a relatively weak topological boundary (i.e. only identified with the most stringent TAD calling parameters; see Fig. S11) at the level of the integration site, similar to that observed at the endogenous region CS38-40 (Andrey et al. 2013; Rodríguez-Carballo et al. 2019; 2020).

Of particular interest, in the course of the present study, another group found that the ectopic integration of an array of three CTCF binding sites (all in the same orientation) comprised within a 2.7 kb construct resulted in the formation of new chromatin loops with surrounding endogenous CTCFs (Redolfi et al. 2019). Furthermore, in some instances, the authors also noticed the formation of relatively weak boundaries at the level of the integrations. Nevertheless, their experiments were performed on cultured cells and lacked *in vivo* confirmation. Hence, to our knowledge, our results provide the first demonstration of the ectopic reconstitution of a functional sub-TAD boundary (i.e. retaining TAD splitting potential) in the genome of stable transgenic mice.

With our experimental setup, however, we could not dismiss the possible effects on the structural landscape of interactions arising from the CS39 enhancer of the *TgN(38-40)* construct. Nevertheless, the observation that the highest contact frequencies generally matched with endogenous CTCF sites, oriented convergently relative to the ones of the transgene in a majority of cases, strongly suggest a predominance of CTCF-dependent interactions to the observed structural changes. In addition, the rather similar results obtained by (Redolfi et al. 2019) when integrating an enhancer-free array of three CTCFs further support this hypothesis.

7. Gene expression changes from the constructs and in the vicinity of ectopic boundary integrations

In order to determine whether the genes of our transgenic constructs could respond to enhancers located within their new genomic landscapes, we performed whole-mount *in situ* hybridization experiments against *Hoxd* transcripts in both *TgN(HoxD)* and *TgN(d11-Evx2)* transgenic embryos.

As regards *TgN(HoxD)*, we found no ectopic expression of the *Hoxd13* and *Hoxd4* genes in a pattern that would resemble those of the genes surrounding the integration site, indicating that the presence of a gene in the same TAD than a given regulatory sequence is not sufficient for a productive communication between them. Nevertheless, the precise regulatory sequences of the surrounding genes remain to be identified in this case.

Strikingly, we noticed that the *Hoxd12* gene became expressed in a very similar cervical pattern compared to that of the gene that was disrupted by the integration of the *TgN(d11-Evx2)* construct, *Hs3st4*. Our structural investigations revealed that *Hoxd12* engaged in interactions with both an intragenic enhancer of *Hs3st4* and the promoter of the latter gene. Interestingly, this enhancer

showed reproducible activity in the same territories than the ones in which *Hoxd12* became ectopically expressed, indicating an enhancer-trap-like phenomenon (see Kvon 2015). However, deciphering between the contribution of promoter-promoter *Hs3st4-Hoxd12* and enhancer-promoter communication to the ectopic activation of *Hoxd12* would require specific genetic interventions, in which either the enhancer or the promoter of *Hs3st4* would be deleted.

Our experimental systems also allowed us to assess changes in the expression of genes located in the vicinity of the ectopic boundary integrations without confounding effects coming from a *wild-type* allele. By doing so, we generally observed little or no effect for genes that are located in different TADs than the ones in which the transgenes integrated, consistent with the notion that *cis* regulatory sequences and their target genes are generally present within the same TAD (e.g. Long, Prescott, and Wysocka 2016; Sikorska and Sexton 2020). However, it remains possible that gene expression was actually affected at other developmental stages.

For genes located within the welcoming TADs, or in a domain displaying relatively frequent interactions with a welcoming TAD, somehow different results were obtained. Surprisingly, we noticed apparently opposite changes in the expression of the *Wdr3* and *Man1a2* genes in *TgN(38-40)^{high}*. This could be explained by the partial reallocation of regulatory inputs from a putative *Wdr3* enhancer towards the *Man1a2* gene as a consequence of the integration. As regards the *TgN(38-40)^{low}* integration, only a mild decrease in the *Btg1* transcripts levels was scored, and might result from slightly modified enhancer-promoter communication as well.

For all cases above, we got indications about the potentially implicated *cis* regulatory elements by looking into publicly available *lacZ* reporter assays (Visel et al. 2007), or mining chromatin accessibility and histone modifications data (Rodríguez-Carballo et al. 2019; Amândio et al. 2020; Christopher Chase Bolt, unpublished data). However, definitive answers to whether these elements are implicated in the observed changes would come from their removal by targeted genetic interventions in each case.

At last, we did not observe any gain in the limb expression of genes surrounding the ectopic integrations of region CS38-40, despite the presence of the CS39 limb enhancer inside the construct (Andrey et al. 2013; Beccari et al. 2016). This was somehow surprising since clear interactions were scored between the *TgN(38-40)* elements and the surrounding gene promoters, which

were devoid from CTCF binding in some instances. These results suggest that other features than solely contacts participate in productive enhancer-promoter communication, such as biochemical compatibility (Long, Prescott, and Wysocka 2016). However, all the genes that we analyzed were already expressed in *wild-type* limbs at E12.5, making it difficult to unambiguously assess changes in expression by whole-mount *in situ* hybridization, so that a more quantitative assessment of transcripts levels using either RT-qPCR or RNA-seq would be needed for confirmation.

In addition, the possibility that the CS39 enhancer is inactive in the new contexts remains. Interestingly, we noticed that the *TgN(38-40)^{high}* construct engaged in long-range (ca. 1.3 Mb) interactions towards the promoter of the *Tbx15* gene, which is located in a different TAD than the one in which the integration occurred, without resulting in any detectable gain in the limb expression of *Tbx15*. Similar to what we observed in the genital tubercle, this gene also turned out to be covered by H3K27me3 marks in a large proportion of limb cells (observation on *wild-type*; Fig. S7). In particular, the pattern of H3K27me3 marks over the *Tbx15* gene promoter precisely matched the distant peak of interactions established by the transgene in *TgN(38-40)^{high}* embryos. This raises the possibility that the interactions scored between *Tbx15* and the transgene might be mediated by Polycomb-group proteins in *Tbx15*-negative cells (Denholtz et al. 2013; Schoenfelder et al. 2015), possibly involving an inactive CS39 enhancer (see Andrey et al. 2013; Rodríguez-Carballo et al. 2017), similar to what we previously speculated in the context of the genital tubercle.

Conclusion:

In the present study, taking the *HoxD* developmental locus of the mouse as a paradigm, we first showed that the strength of a TAD boundary can be modulated *in situ* by playing with the number of potential boundary elements it contains. In particular, intra-cluster duplications increasing the number of both CTCF-binding sites and gene promoters resulted in an increased isolation between the central *Hoxd* genes and the TAD located beyond the duplicated segment. Furthermore, the structural changes occurring between enhancer-promoter pairs in these configurations matched with modified expression patterns of *Hoxd* genes that were reported in previous publications from our group. In this context, however, a substantial proportion of regulatory interactions appeared to be maintained despite the presence of a duplicated segment in-between, as were the global expression patterns of the central *Hoxd* genes, indicating a high resilience of enhancer-promoter communication.

While it is now clear that CTCF and the cohesin complex hold key roles in shaping the genome into topologically associating domains, evidence for the capacity of a TAD boundary to function independently from its context of origin remained limited. Using various transgenic constructs containing *HoxD*-related boundary elements, we showed that the architectural proteins CTCF and RAD21 were recruited to their cognate binding sites even when the latter were relocated in the genome. Moreover, the elements of the transgenes established contacts with their new surroundings, which were more frequent within the limits of the welcoming TADs and peaked at CTCF sites in convergent orientation relative to the ones of the transgene in many instances. Finally, some of these integrations resulted in significant alterations of the local chromatin landscape surrounding the insertion sites, as demonstrated by the formation of a new boundary in at least one of the cases under study. Collectively, we believe that these results provide new insights into how topological boundaries function in the genome.

Acknowledgements:

I would like to thank Denis Duboule for hosting me as a Master student in his group and for his support. Working in his laboratories, which represent an amazing environment in which cutting-edge technologies are applied to extend our knowledge about gene regulation and development, was a great chance for me. I am grateful to David Shore for accepting to be the juror for the evaluation of the present work. I would like to thank Eddie Rodríguez-Carballo for the design and supervision of this project as well as for his day-by-day support in experimental procedures, but also in personal achievements. I am grateful to Lucille Lopez-Delisle for constant help with data analyses; Christopher Chase Bolt for help with both the Hi-C and MinION experiments; Imane El-Idrissi, Eddie Rodríguez-Carballo, Leonardo Beccari, Marie-Laure Gadolini, Isabel Guerreiro and Christopher Chase Bolt for sharing unpublished data; Aurélie Hintermann, Leonardo Beccari and Jozsef Zákány for their availability in answering my questions; as well as Sandra Gitto, Thi Hanh Nguyen Huynh, Bénédicte Mascrez and Marie-Laure Gadolini for mice care, breeding and genotyping. I would also like to thank all members of the Duboule laboratories for insightful comments and discussions about the project. I acknowledge financial support through the Excellence Master Fellowship of the Faculty of Science of the University of Geneva.

Material and Methods:

Genetically engineered mice strains

The *HoxD*^{Dup(8i-10)} (named *Dup(8i-10)*) and *HoxD*^{Dup(11-13)} (named *Dup(11-13)*) mouse lines have already been described in previous publications from our group (Tarchini and Duboule 2006; Kmita et al. 2002). They were obtained by targeted meiotic recombination (TAMERE) starting from parental lines carrying *loxP* sites at the desired duplication breakpoints (Hérault et al. 1998). The *TgN(Hoxd1lacZ/69)* (named *TgN(HoxD)*) line was obtained by Jozsef Zákány using the bacterial artificial chromosome (BAC) clone RP23-400H17 (accession number AC015584; described in Spitz et al. 2001) and has been previously reported in (Guerreiro et al. 2016; Schep et al. 2016). In brief, both a *lacZ* reporter gene and a ZeocinTM resistance cassette (derived from pEM7/Zeo, Thermo Fisher Scientific) were introduced in the BAC at the level of the HindIII site in the second exon of *Hoxd1* (Schep et al. 2016). Transgenesis was performed by pronuclear injection. The *TgN(d11-Evx2)* and *TgN(38-40)* constructs were designed by Eddie Rodríguez-Carballo and the corresponding random transgenic lines (i.e. *TgN(d11-Evx2)*, *TgN(38-40)*^{high} and *TgN(38-40)*^{low}) were produced by pronuclear injection at the transgenic core facility of the University of Geneva. Simplified schemes of all initial transgenic constructs that were used in the present study can be found in Figures S3A (*TgN(HoxD)*), S4A (*TgN(d11-Evx2)*), S5A and S8A (*TgN(38-40)*). 129S1/Sv-*Hprt*^{tm1(CAG-cre)Mnn/J} (short name *Hprt*^{cre}) mice were purchased from The Jackson Laboratory. These mice carried a Cre-expression cassette in the *Hprt* locus (active in the female germline) on chromosome X (Tang et al. 2002) and were used for transgene extra-copy removal thanks to a flanking *loxP* site on the *TgN(d11-Evx2)* (see Fig. S4A) or *TgN(38-40)* (see Figs. S5A and S8A) construct. The *HoxD*^{del(attP-Rel5)d9lac} (named *del(attP-Rel5)*) allele has already been described in (Rodríguez-Carballo et al. 2017). It was obtained by TAMERE starting from the *HoxD*^{attP} (Andrey et al. 2013) and *HoxD*^{Rel5-d9lac} (Montavon et al. 2011) alleles. The *HoxD*^{del(CS38-40)} (named *del(CS38-40)*) allele has been described in (Schep et al. 2016). It was produced by placing *loxP* sites on each sides of region CS38-40 through homologous recombination, followed by Cre-mediated excision of the *floxed* segment (see *Hprt*^{cre}). The *HoxD*^{del(1-13)d11lac} (named *HoxD*^{del(1-13)}) allele has been described in (Zákány et al. 2001). It was generated starting from an allele carrying a *Hoxd1/lacZ/loxP* construct at the endogenous position of *Hoxd1*, together with a *Hoxd11/lacZ/loxP* construct between *Hoxd13* and *Evx2*. The deletion of all sequences located between the two *loxP* sites was achieved by Cre-mediated recombination *in vitro* (Zákány and Duboule 1996; Zákány et al. 2001). All mouse lines used in this study were maintained in a heterozygous state on a

C57BL6xCBA background. Crosses between heterozygous individuals were performed to obtain homozygous embryos.

ChIPmentation

ChIPmentation was performed using the protocol of (Schmidl et al. 2015). Tissues were dissected in ice-cold cold 1X PBS and crosslinking was performed for 15 minutes at room temperature in a 1X PBS fixation solution containing 1% formaldehyde (Thermo Fisher Scientific, 28906). Crosslinking was stopped using Stop Solution (Active Motif, ChIP-IT HS Kit) and tissues were washed three times in ice-cold washing solution containing 1X PBS and Detergent (Active Motif, 1:20). Tissues were frozen in liquid nitrogen and stored at -80°C until further processing. For each genotype, either four E12.5 whole limbs or one E9.5-E10 headless embryo were used for the rest of the procedure. The tissues were resuspended in Chromatin Prep Buffer (Active Motif, ChIP-IT HS Kit) containing cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, 1 tablet in 10 mL), homogenized using a Polytron and nuclei were released using a Dounce homogenizer with type B (tight) pestle (Active Motif). Released nuclei were resuspended in sonication buffer (50mM Tris-HCl pH8, 10mM EDTA pH8, 0.25% SDS) containing protease inhibitors. Chromatin was fragmented using a Bioruptor® Pico sonication device (Diagenode), with the following settings. E12.5 limbs: 4-6 times (20 seconds ON, 30 seconds OFF). E10 headless embryos: 4 times (20 seconds ON, 30 seconds OFF). Sonication was controlled using agarose gel electrophoresis after reversing crosslinks and a fraction of the samples was preserved to evaluate the efficiency of the chromatin immunoprecipitation by qPCR prior to sequencing (see total input DNA control below). Dynabeads™ Protein A (Thermo Fisher Scientific, 10001D) were washed three times in RIPA with low salt concentration (LS, 140mM NaCl) containing BSA (1 mg/mL). Anti-CTCF (Active Motif, 61311) or Anti-RAD21 (Abcam, ab992) antibodies were then incubated with the beads for 3 hours on a rotating wheel at 4°C, followed by two washes in RIPA-LS containing BSA. Samples were incubated with the Antibody-coupled beads on a rotating wheel at 4°C overnight for chromatin immunoprecipitation. The following washes were performed in the presence of protease inhibitors: 2 washes in RIPA-LS, 2 washes in RIPA with high salt concentration (HS, 500 mM NaCl), 2 washes in RIPA containing LiCl, 1 wash in 10 mM Tris-HCl pH8. Tagmentation was performed using the Tagment Enzyme 1 (Tn5, Illumina) for 2 minutes at 37°C. Samples were kept on ice for 5 minutes in order to inactivate the Tn5 enzyme, followed by washes with RIPA-LS and TE (10 mM Tris-HCl pH8, 1 mM EDTA pH8); once each. ChIPed DNA was eluted using ChIP elution buffer (10 mM Tris-HCl pH8, 5 mM EDTA pH8, 300 mM NaCl, 0.4% SDS) containing proteinase K for 1 hour at 55°C, followed by 6 hours at 65°C. The

same previous step as well as all the following were also carried out on the total input DNA. DNA was purified using the MinElute PCR Purification Kit (Qiagen). ChIPmentation library amplification was performed using indexed Nextera custom primers (Buenrostro et al. 2013) and pre-heated KAPA HiFi HotStart ReadyMix (Roche) following the manufacturer's instruction (with an extension time of 30 seconds and annealing at 63°C) and applying a number of cycles determined empirically by qPCR on a fraction of the samples so as to minimize PCR duplicates. The resulting libraries were purified and size-selected using SPRI beads with magnetic separation. Enrichment for ChIPed DNA was assessed by qPCR using the total input DNA as a control. The ChIPmentation libraries were multiplexed and sequenced so as to obtain 50 bp single-end reads with an Illumina HiSeq. ChIPmentation data analysis was performed as described in (Rodríguez-Carballo et al. 2019). The data were demultiplexed, followed by adapter trimming and reads filtering using cutadapt (Martin 2011). Data mapping was achieved with Bowtie 2 (Langmead and Salzberg 2012) on either the mm10 reference genome, or on the mm10+*TgN(38-40)^{low}* custom genome. Only reads with a MAPping Quality (MAPQ) above 30 were kept when mapping on the mm10 genome, whereas all were kept for mm10+*TgN(38-40)^{low}* so that the reads mapping to the ~600 bp duplication observed in this line were not discarded. Prior to merging the replicates, the bedgraphs were normalized to the number of tags obtained after filtering. The merging of replicates was performed with the unionbedg tool of the BEDTools suite (Quinlan 2014). Peak calling of CTCF and RAD21 was achieved using the MACS2 algorithm (Y. Zhang et al. 2008; Feng et al. 2012) during data processing. CTCF site orientation was determined afterwards using CTCFBSDB 2.0 (<http://insulatorodb.uthsc.edu/>) (Ziebarth, Bhattacharya, and Cui 2013) with MIT_LM7 motif position weight matrix (PWM).

TLA

TLA was performed following the protocol published as a supplement to (de Vree et al. 2014). Tissues were dissected in ice-cold 1X PBS containing 10% fetal calf serum (FCS), followed by cell dissociation with collagenase type XI (Sigma-Aldrich, C7657) at 37°C while shaking for 30-45 minutes. This solution was strained using snap capped tubes with a nylon mesh (Falcon®, 352235). Chromatin was crosslinked for 10 minutes at room temperature in 1X PBS/10% FCS containing formaldehyde (Sigma-Aldrich, F8775) at a final concentration of 2%. Crosslinking was blocked by ice-cooling and ice-cold glycine addition. After centrifugation, the samples were lysed, centrifuged once more and frozen at -80°C until further processing. Transgene-positive samples were identified by PCR and either 2-4 E12.5 brains, 8 E12.5 whole limbs, or 1-2 E9.5-E10.5 whole embryos were used for the rest of the procedure. Nuclei pellets were

resuspended in 1X CutSmart® Buffer (New England Biolabs), followed by SDS addition to a final concentration of 0.28%. Samples were incubated at 37°C while shaking for 30 minutes, after which TritonX-100 was added to quench SDS. Crosslinked chromatin was primarily digested using NlaIII (4-cutter; New England Biolabs) overnight at 37°C while shaking. NlaIII digestion efficiency was assessed by agarose gel electrophoresis after reversing chromatin crosslinks using proteinase K. NlaIII was then inactivated by heating the samples at 65°C for 20 minutes. Ligation was performed in the presence of T4 DNA Ligase (Promega) for 2.5-3 hours at room temperature. Ligation efficiency was evaluated on agarose gel. Chromatin reverse crosslinking was performed by adding proteinase K to the samples, followed by incubation overnight at 65°C without agitation. Proteinase K removal was performed using phenol-chloroform DNA purification with ethanol precipitation. Secondary restriction was performed using NspI (5-cutter recognizing a subset of NlaIII sites; New England Biolabs) in 1X CutSmart® Buffer overnight at 37°C with agitation and was then assessed by agarose gel electrophoresis. NspI was heat-inactivated (65°C), followed by secondary ligation with T4 DNA Ligase at 16°C overnight in DNA-diluted conditions so as to favor intramolecular ligation events. After ethanol precipitation, DNA was purified using the QIAquick PCR Purification Kit (Qiagen), following the manufacturer's instructions. Enrichment for secondary ligation products was controlled on an agarose gel, followed by DNA quantification using a Qubit™ fluorometer (Thermo Fisher Scientific). TLA inverse PCR was performed using 800 ng template DNA (for 8 reactions in total), viewpoint-directed inverse PCR primers (a list of all TLA viewpoints used in this study is given in Table S4) and the Phire Hot Start II DNA Polymerase (Thermo Fisher Scientific) following the manufacturer's program with 2 minutes extension time and annealing at 55°C. TLA PCR products were purified using the QIAquick PCR Purification Kit, quantified using Qubit™ and quality checked by gel electrophoresis. TLA library preparation was achieved by tagmentation of sequencing adapters using the Nextera™ DNA Flex Library Prep (Illumina) protocol. Libraries were sequenced so as to obtain 100 bp single-end reads with an Illumina HiSeq. A detailed TLA data analysis workflow is given in Figure S2.

4C-seq

4C-seq was performed as in (Noordermeer et al. 2011). As for TLA, cells were dissociated with collagenase type XI, followed by chromatin crosslinking in a 2% formaldehyde solution and lysis. Samples were frozen at -80°C until further processing. For each genotype, samples corresponding to either 11 E12.5 PFL or DFL (*Dup(8i-10)* and *Dup(11-13)*), 12 E12.5 whole limbs (*TgN(HoxD)*; all *TgN(d11-Evx2)* viewpoints except *Hoxd12*; *TgN(38-40)^{high}*; *TgN(38-40)^{low}*) or 2 E10.5

headless embryos (*TgN(d11-Evx2)*, *Hoxd12* viewpoint) were used for the rest of the procedure. Nuclei pellets were resuspended in 1.2X CutSmart® Buffer (New England Biolabs), followed by SDS addition to a final concentration of 0.29%. The samples were incubated while shaking for 10 minutes at 55°C, followed by 50 minutes at 37°C. TritonX-100 was added in order to quench SDS. Crosslinked chromatin was primarily digested using *NlaIII* (4-cutter; New England Biolabs) for a minimum of 1 hour at 37°C while shaking before controlling the digestion. *NlaIII* digestion efficiency was assessed by agarose gel electrophoresis after reversing chromatin crosslinks using proteinase K. *NlaIII* digestion was continued overnight, followed by an additional digestion control. *NlaIII* was then inactivated by SDS addition and heating the samples at 65°C for 25 minutes. Ligation was performed in the presence of T4 DNA Ligase (Promega) for 4 hours at 16°C in diluted conditions so as to favor intra-complex ligation events. Chromatin reverse crosslinking was performed by adding proteinase K to the samples, followed by incubation overnight at 65°C without agitation. Proteinase K removal was achieved using phenol-chloroform DNA purification with ethanol precipitation. The concentration of the purified DNA was measured using Qubit™, ligation was assessed on agarose gel and the DNA was diluted to a concentration of 100 ng/μL in 1X NEBuffer™ DpnII. Secondary restriction was performed using *DpnII* (4-cutter; New England Biolabs) overnight at 37°C with agitation. *DpnII* was heat-inactivated (65°C), followed by phenol-chloroform DNA purification with ethanol precipitation. After this step, secondary restriction was assessed by agarose gel electrophoresis. Secondary ligation was performed with T4 DNA Ligase at 16°C for 4 hours in DNA-diluted conditions in order to favor intramolecular ligation events. Ligated DNA products were then extracted using phenol-chloroform with ethanol precipitation and purified using the QIAquick Nucleotide Removal Kit (Qiagen), following the manufacturer's instructions. 4C DNA was quantified using Qubit™ and enrichment for secondary ligation products was controlled on an agarose gel. An inverse PCR was performed using 100 ng template DNA (for 14 reactions in total), viewpoint-directed inverse PCR primers carrying Illumina Solexa sequencing adapters (a list of all 4C-seq viewpoints used in this study is given in Table S3) and the Expand™ Long Template PCR System polymerase (Roche) with the provided Buffer 1, following the manufacturer's protocol with 3 minutes extension time and annealing at 55°C. 4C PCR libraries were purified using the QIAquick PCR Purification Kit with a QIAvac vacuum manifold (Qiagen), quantified using Qubit™ and quality checked by gel electrophoresis. Libraries were multiplexed and sequenced so as to obtain 100 bp single-end reads with an Illumina HiSeq. 4C-seq data analysis was performed locally following the same procedure as previously available on the HTSstation web interface (David et al. 2014) (<http://htsstation.epfl.ch>). Raw sequencing reads were first demultiplexed

to assign each read to a viewpoint and experiment. For each read, the CATG sequence (NlaIII restriction site) and the following 26 base pairs (i.e. 30 bp in total) were mapped on either the mm10 reference genome or on the mm10+*TgN(38-40)^{low}* custom genome. Reads corresponding to self-ligation and non-digestion events as well as those that were not delimited by a NlaIII site on one end and a DpnII site on the other (e.g. with NlaIII sites on both extremities) were removed from the analysis. The signal was uniformly attributed to the 20 first hits of highest similarity and a coefficient allowing to adapt the score to repetitive regions was produced using RepeatMasker (Chen 2004). The region located within +/- (2.5-3.5 kb) around each viewpoint was also excluded from the analysis. The resulting scores were then normalized to the mean score of fragments mapping within +/- 5 Mb around the viewpoint. For each fragment, the signal was smoothened on 5 fragments preceding the fragment, this fragment and 5 fragments succeeding it (11 fragments smoothening).

Hi-C

Hi-C was performed following a modified version of the protocol from (Rao et al. 2014). As for TLA and 4C-seq, cells were dissociated with collagenase type XI, followed by chromatin crosslinking in a 2% formaldehyde solution, with the difference that they were frozen at -80°C before conducting the lysis. For each genotype (i.e. *wild-type* or homozygous *TgN(38-40)^{low}*), one sample corresponding to four E12.5 whole limbs was used for the rest of the procedure. Nuclei pellets were resuspended in lysis buffer and placed on a rotating wheel at 4°C for 30 minutes, centrifuged and resuspended in 1X NEBuffer™ 3.1 (New England Biolabs) containing SDS to a final concentration of 0.11%. The samples were incubated at 65° for 10 minutes, followed by ice-cooling and TritonX-100 addition for SDS quenching. Crosslinked chromatin was digested using DpnII (4-cutter; New England Biolabs) for 4 hours at 37°C while shaking before controlling the digestion by agarose gel electrophoresis after reversing chromatin crosslinks using proteinase K. DpnII digestion was continued overnight, followed by an additional digestion control. DpnII was then heat-inactivated (65°C) and samples were resuspended in 1X NEBuffer™ 3.1. Biotin fill-in was performed using biotin-14-dATP, non-biotinylated dCTP/dGTP/dTTP and DNA Polymerase I, Large (Klenow) Fragment with 3'->5' exonuclease activity (New England Biolabs, M0210). This reaction was incubated on a rotating wheel for 4 hours at room temperature. Blunt-end ligation was achieved with T4 DNA Ligase (Thermo Fisher Scientific, EL0011) on a rotating wheel for 4 hours at room temperature. NaCl and SDS were added to the samples to a final concentration of 260mM and 0.86%, respectively. Samples were incubated overnight at 68°C with agitation. Chromatin reverse crosslinking was performed by adding proteinase K to the

samples, followed by incubation at 56°C while shaking for 1 hour. Proteinase K removal was achieved through two rounds of phenol-chloroform DNA extraction with a final ethanol precipitation. The concentration of the purified DNA was measured using the Qubit™ dsDNA BR assay (Thermo Fisher Scientific) and ligation was assessed on agarose gel. Sonication was carried out using a Covaris® E220 sonication device with 175W peak incident power, 200 cycles per burst and for a treatment time of 90 seconds. A 1:8:1 size selection was performed using SPRI beads with magnetic separation. The concentration of the sonicated DNA was measured using Qubit™ dsDNA BR. The biotin pull-down was performed using Dynabeads™ MyOne™ Streptavidin C1 (Thermo Fisher Scientific) with magnetic separation, followed by quantification using the Qubit™ dsDNA HS assay (Thermo Fisher Scientific). Repair of DNA fragment ends was achieved using T4 Polynucleotide Kinase (New England Biolabs, M0201), T4 DNA Polymerase (New England Biolabs, M0203) and DNA Polymerase I, Large (Klenow) Fragment with 3'->5' exonuclease activity (New England Biolabs, M0210) at room temperature for 30 minutes with agitation during 10 seconds every 5 minutes. A-tailing was performed with a DNA Polymerase I, Large (Klenow) Fragment lacking both 5'->3' and 3'->5' exonuclease activities (New England Biolabs, M0212) in 1X NEBuffer™ 2 (Thermo Fisher Scientific) containing dATP for 30 minutes at 37°C. Ligation of TruSeq™ (Illumina) paired-end Y adapters including indexes for multiplexing was performed following the manufacturer's instructions using T4 DNA Ligase (Thermo Fisher Scientific, EL0011) at room temperature for 2 hours. Library amplification was performed by PCR using the Herculase II Fusion DNA Polymerase (Agilent, 600675) following the manufacturer's instructions (with an extension time of 15 seconds and annealing at 65°C) and applying a number of cycles determined empirically by conventional PCR on a fraction of the samples so as to minimize PCR bias and PCR duplicates. Libraries were purified using SPRI beads with magnetic separation, quantified with Qubit™ dsDNA HS and the sizes of the DNA fragments were measured on a Fragment Analyzer. Hi-C libraries were finally multiplexed and sequenced so as to obtain 75/75 bp paired-end reads with an Illumina NextSeq (first run, 80 million reads per sample) or HiSeq (second run, idem). Hi-C data analysis was achieved as in (Yakushiji-Kaminatsui et al. 2018), with modifications. After sequencing, the reads were processed with the HiCUP pipeline (version 0.7.3) (Wingett et al. 2015), working with Bowtie 2 (Langmead and Salzberg 2012) and SAMtools (H. Li et al. 2009). HiCUP was applied providing either the mm10 reference genome, or the mm10+*TgN(38-40)*^{low} custom genome. In both cases, GATC was defined as restriction site sequence (DpnII). For the mapping on the mm10+*TgN(38-40)*^{low} custom genome, the pipeline was adapted in order to prevent removal of reads mapping to the ~600 bp duplication observed in

TgN(38-40)^{low}. In contrast to (Yakushiji-Kaminatsui et al. 2018), all valid hybrid pairs were kept since no MAPping Quality (MAPQ) filter was applied. HiCUP assigned each read to a given genomic fragment, after which the coordinate of the middle of this fragment was retrieved. Analysis of the hybrid pairs resulted in a Hi-C matrix binned at 40 kb, which was further processed using cooler for balancing normalization. TAD identification in Figure 21, 22 and S11 was achieved using hicFindTADs from the HiCExplorer package (<https://github.com/deeptools/HiCExplorer>) (Ramírez et al. 2018). It was performed using the TopDom algorithm (Shin et al. 2016) in all other figures showing Hi-C heatmaps, whose data source is (Rodríguez-Carballo et al. 2017).

Nanopore Cas9-targeted sequencing (nCATS)

nCATS was performed as in (Gilpatrick et al. 2020), following the *Cas-mediated PCR-free enrichment* protocol (Oxford Nanopore Technologies Ltd) with a tiling approach, with the difference that single-guide RNAs (sgRNAs) were used for target enrichment. Multiple pairs of sgRNAs were designed on the target region (see Fig. S10A and Table S7) with Benchling (<https://www.benchling.com/>) and were converted into EnGen®-compatible DNA oligos using the NEBioCalculator (<http://nebiocalculator.neb.com/#!/sgrna>) (see Table S7). sgRNAs production was performed using the EnGen® sgRNA Synthesis Kit, *S. pyogenes* (New England Biolabs, E3322) following the manufacturer's instructions. The quality of the sgRNAs was evaluated by agarose gel electrophoresis and their concentrations were measured on a NanoDrop™ (Thermo Fisher Scientific). Two distinct pools of sgRNAs (see Fig. S10A) and Alt-R® *S. pyogenes* HiFi Cas9 nuclease V3 (IDT) were assembled into Cas9 ribonucleoprotein (RNP) complexes in 1X CutSmart® Buffer (New England Biolabs) at room temperature for 30 minutes. The efficiency of the sgRNA-directed Cas9 cleavage was evaluated by qPCR (with primer pairs spanning each cutting sites) after performing a "release test" on *TgN(38-40)^{low}* genomic DNA. High molecular weight genomic DNA (HMW gDNA) was prepared as follows, starting from a single E13.5 *TgN(38-40)^{low}* and *del(CS38-40)* double-homozygous headless embryo. The sample was proteinase K digested in digestion buffer (50 mM Tris-HCl pH8, 10 mM EDTA pH8, 200 mM NaCl, 0.5% SDS) at 55°C while shaking for 48 hours. No heat-inactivation of proteinase K was performed in order to preserve the integrity of large DNA molecules. HMW gDNA was purified with two successive rounds of phenol-chlorophorm extraction, followed by ethanol precipitation. The resulting gDNA was then quantified with a NanoDrop™ and quality checked on an agarose gel. Prior to following the standard protocol, a size selection was carried out so as to remove the smallest fragments from the solution. This was performed using 0.8 times the sample

volume of SPRI beads and magnetic separation. gDNA was quantified once more with a NanoDrop™ and enrichment for large molecules was assessed on a Fragment Analyzer. gDNA dephosphorylation was performed using Quick Calf Intestinal Phosphatase (CIP) (New England Biolabs, M0525) in 1X CutSmart® Buffer (New England Biolabs) at 37°C for 10 minutes, followed by inactivation at 80°C for 2 minutes. Cleavage and A-tailing of the dephosphorylated gDNA was performed using the Cas9-sgRNAs RNPs (for cleavage) together with Taq DNA Polymerase (for A-tailing) (New England Biolabs, M0273) in the presence of dATP within the same reaction mix. Two independent reactions corresponding to the two different pools of Cas9-sgRNAs RNPs were incubated for 1 hour at 37°C (cleavage), followed by 5 minutes at 72°C (Cas9 denaturation and A-tailing). A fraction of the samples was preserved at this step for an additional fragment size analysis. The products from both Cas9-mediated release reactions were pooled together and adapter ligation was performed using the AMX adapters from the Ligation Sequencing Kit (Oxford Nanopore Technologies Ltd, SQK-LSK109), following the manufacturer's instructions. The resulting library was then purified using SPRI beads and the Long Fragment Buffer (LFB) from the kit before loading on a MinION SpotON flow cell (Oxford Nanopore Technologies Ltd, FLO-MIN106D). The MinION output data in fast5 format were converted into the fastq format using the Guppy basecaller version 3.1.5 (Oxford Nanopore Technologies Ltd) and mapped on the mm10+*TgN(38-40)^{low}* custom genome using minimap2 version 2.15 (Oxford Nanopore Technologies Ltd). Only the best similarity hit was kept for each read. The format of the reads file was converted from bam to bed so as to obtain continuous reads despite the relatively high error rate of Nanopore sequencing. The MinION coverage shown on Figure S10B was extracted from the reads file in bed format. In parallel, reads mapping on both the integration site (chr10:97018026-97021028, mm10) and transgene (vector and chr2:75122684-75160161, mm10) sequence components of the *TgN(38-40)^{low}* mutant construction (see Fig. S10A) were converted from the fastq to the fasta format in order to produce the MinION reads dot plots displayed on Figure S10C. This was performed using a Perl script as in (Hughes et al. 2005; Nicholls et al. 2019).

In situ probe cloning and transcription

Total RNAs for cDNA production were extracted from various E13.5 embryonic tissues (including limbs, kidneys, forebrain and hindbrain) stored in RNAlater™ using the RNeasy® Micro Kit (Qiagen, 74004), following the manufacturer's instruction. Retrotranscription of RNAs into cDNAs was performed using random primers and the SuperScript® II Reverse Transcriptase (Promega, 18064014) in the presence of RNase inhibitors and MgCl₂. The

retrotranscription reaction was incubated for 5 minutes at 25°C, 1 hour at 42°C and 15 minutes at 70°C. Starting from the cDNAs, standard PCRs were performed using the TopTaq polymerase (Qiagen) and primers matching with either an exon or the 3' UTR of each mRNA of interest (see Table S8). The PCR products were migrated on an agarose gel and the band corresponding to the amplicon of interest was dissected under UV transillumination, followed by DNA purification using the QIAquick PCR Purification Kit (Qiagen). Each PCR products was independently cloned into the pGEM®-T Easy Vector (Promega, A1360), following the manufacturer's instructions, after which heat shock-driven transformation of chemically competent bacteria with the plasmid was performed. Bacteria were plated on LB-agar plates containing ampicillin. X-gal was added in order to distinguish bacteria carrying the empty vector (forming blue colonies) from the ones carrying an insert-containing plasmid (forming white colonies). Individual white colonies of bacteria were picked and used to inoculate a 2 mL LB + ampicillin liquid solution. Bacteria were grown overnight at 37°C while shaking, after which Miniprep plasmid DNA purification was performed. The purified plasmid DNA was analyzed using cleavage by restriction enzymes cutting at known sites inside the vector in order to determine the orientation of the insert, whereas another fraction of the DNA was Sanger sequenced. Bacterial cultures identified as carrying the correct insert within the pGEM®-T Easy Vector were used to inoculated 100 mL LB + ampicillin. After growing the bacteria overnight at 37°C, the bacteria were harvested to produce stocks in glycerol for long-term storage at -80°C on the one hand, and on the other hand to obtain highly concentrated purified plasmid DNA using the NucleoBond® Xtra Midi Kit (Macherey-Nagel). Plasmid DNA was linearized using the appropriate restriction enzyme and *in vitro* transcription was performed for 2 hours at 37°C using either the T7 or SP6 RNA Polymerase (Promega, P2075 and P1085), depending on the orientation of the insert in the vector, for antisense or sense probe production. Transcription was performed in the presence of DIG Labeling Mix (Roche, 11277073910) for digoxigenin (DIG) to be incorporated into the probes. The RNA probes were analyzed by agarose gel electrophoresis, purified using the RNeasy® Purification Kit (Promega) and quantified on a NanoDrop™.

Whole-mount in situ hybridization

Whole-mount *in situ* hybridization (WISH) was performed following the protocol published as a supplement to (Woltering et al. 2009). Embryos were isolated in ice-cold 1X PBS, collected in 1X PBS containing 4% paraformaldehyde (PFA) (Sigma-Aldrich, 16005) and incubated for 24 hours at 4°C for fixation. The embryos were then transferred to 100% methanol and stored at -20°C before

further processing. Sets of E12.5 embryos of comparable sizes and of all possible genotypes as regards a given transgene (i.e. *wild-type*, homozygous and heterozygous) were placed together in 7 mL plastic tubes. The samples were maintained in agitation in-between all following steps. Embryos were bleached in methanol containing 3% H₂O₂ for 20 minutes at room temperature. Rehydration was performed by transferring the embryos through successive solutions of decreasing methanol concentrations, ending with Tris-Buffered Saline (20 mM Tris Base, 137 mM NaCl, 2.7 mM KCl) containing 1:1000 Tween-20 (TBS-T). Permeabilization was achieved by treating the embryos with proteinase K at a final concentration of 10 ng/mL for 10 minutes at room temperature. After washing with TBS-T, proteinase K was inactivated by addition of acetic anhydride and 100 mM TEA pH7-8. The embryos were fixed one more time in 1X PBS containing 4% PFA for 20 minutes at room temperature, followed by TBS-T washes and incubation in a hybridization mix (containing RNA from Torula, formamide, SSC, Blocking Reagent, EDTA 0.5 M pH8, heparin and Tween-20) for 4 to 5 hours at 68°C. Hybridization was performed with DIG-labeled RNA probes (see *in situ* probe cloning and transcription above) at a final concentration of 100-200 ng/mL overnight at 68°C. Two washes were performed in hybridization mix without probe, followed by four washes in SSC containing Tween-20. Embryos were transferred in maleic acid buffer (150 mM maleic acid pH7.5, 100 mM NaCl) containing Tween-20 (MAB-T), and then to a Blocking buffer (MAB-T with 1% Blocking Reagent; Roche, 1096176) for 1.5 hours in order to equilibrate the samples. Incubation in Blocking buffer containing 1:3000 Anti-DIG antibodies coupled to alkaline phosphatase (Roche, 093274) was performed for 4 to 5 hours at room temperature. Five washes were done in MAB-T with 5 to 20 minutes waiting in-between. Another series of washes in MAB-T was carried out on the following day and was continued overnight. The embryos were equilibrated in alkaline phosphatase buffer (100 mM Tris-Base, 100 mM NaCl, 50 mM MgCl₂, 1:1000 Tween-20) and stained in BM-Purple (Roche, 11442074001) away from light for 8-14 hours (depending on the probe) at room temperature with moderate agitation. A last series of TBS-T washes were carried out before fixing the sample in 1X PBS containing 4% PFA. Pictures were taken with a Leica M205 FCA microscope equipped with a DFC 7000 T camera and were processed with Adobe Photoshop®.

qPCR transgene quantifications

Individual ear punches from adult mice were digested in proteinase K for 48 hours, followed by heat-inactivation of the enzyme at 96°C. Samples were cooled down, centrifuged and gDNA was purified using phenol-chloroform extraction and ethanol precipitation. qPCR was performed using PowerUp™

SYBR™ Green Master Mix (Thermo Fisher Scientific, A25742) on MicroAmp™ Optical 384-Well Reaction Plate (Thermo Fisher Scientific, 4309849) in a QuantStudio™ 5 Real-Time PCR device (Thermo Fisher Scientific). The primers used were either specific to the elements present in the *TgN(38-40)* construct (namely the CTCF site of CS38, CS39 and the first CTCF of CS40) (see Figs. S6 and S9), to control regions of the *HoxD* cluster (i.e. the intergenic *Hoxd4-Hoxd8* and *Hoxd8-Hoxd9* CTCF sites), or to a normalizing gene (*Aldehyde dehydrogenase 1 family, member A2* or *Aldh1a2*) (see Table S9). For each sample, the results were normalized to the value of *Aldh1a2* using the ΔC_t method. The qPCR quantifications plots shown on Figures S6 and S9 were produced using GraphPad Prism 8 and represent the values relative to the *wild-type* ($\Delta\Delta C_t$), for each qPCR target.

Sequencing data analysis and display

All chromosome conformation capture (4C-seq and Hi-C) data were analyzed through the Scientific IT and Application Support Center of the École polytechnique fédérale de Lausanne (EPFL). All chromatin immunoprecipitation (ChIP-seq and ChIPm) data were analyzed on our Galaxy platform (<https://galaxyduboule.epfl.ch/>). For all main figures in Results, data were plotted using the pyGenomeTracks software (<https://github.com/deeptools/pyGenomeTracks>) (Ramírez et al. 2018). For all supplemental figures, the genome browser tracks data were displayed using the UCSC Browser. Gene annotations were retrieved and adapted from UCSC (GRCm38.92-mm10). All figures were processed with Adobe Illustrator®.

Quantifications of 4C-seq contacts over the HoxD TADs

The quantifications shown on Figures 8 and 10 were performed by summing 4C-seq scores from bedgraphs over either the T-DOM (chr2:74781516-75600000) or C-DOM (chr2:73920001-74636454), while excluding the *HoxD* cluster (chr2:74636455-74781515). The resulting values were divided by the one obtained for both domains, and then multiplied by 100 to obtain percentage values. This analysis was performed with R (<https://www.r-project.org/>) using own scripts.

Control-FREEC

Control-FREE Copy number and allelic content calling (Control-FREEC) (Boeva et al. 2011; 2012) was performed using the software available at (<http://boevalab.com/FREEC/index.html>). For each transgenic line of interest, a “test” sample was generated by using the corresponding total input DNA data from the ChIPmentation experiment (see ChIPmentation above). When several

total input DNA datasets were available for the same line, they were merged prior to launch the Control-FREEC analysis using the unionbedg tool of the BEDTools suite (Quinlan 2014). Next, a “control” sample was created by pooling all available total input DNA data of samples that were negative for this transgene and *wild-type* for the corresponding deletion background. For example, for the Control-FREEC analysis of *TgN(d11-Evx2); del(attP-Rel5)*, the control sample included data from both *TgN(38-40)^{high}* and *TgN(38-40)^{low}*, since both the latter were negative for the *TgN(d11-Evx2)* and *wild-type* for the *del(attP-Rel5)* genetic background. For both the test and the control, the Control-FREEC signal, expressed as the number of reads scored for non-overlapping genomic windows of given sizes (Boeva et al. 2011), was computed along a 7 Mb region of chromosome 2 including the *HoxD* locus (chr2:71000000-78000000). The above analysis was carried out using three different window sizes: 1 kb, 2 kb and 5 kb (see Figs. S4D, S5D and S8D). Then, the software calculated the test/control signal ratio, i.e. the number of reads for the test divided by the number of reads for the control, for each window, multiplied by 2 in order to obtain absolute copy numbers corresponding to a diploid genome. At last, the Control-FREEC software evaluated copy numbers along the 7 Mb chr2 region starting from the test/control signal ratio by a maximum (log-)likelihood estimation (MLE) (Boeva et al. 2012).

TgN(38-40)^{low} mutant genome reconstruction

The *TgN(38-40)^{low}* mutant genome was built by inserting a 63812 bp sequence containing (1) the entire *TgN(38-40)* fosmid (chr2:75122684-75160161), (2) an additional fragment of *TgN(38-40)* extending towards the CTCF of CS38 (chr2:75122684-75147258) and (3) the ca. 600 bp duplicated region (chr10:97019222-97019824) inside chromosome 10 after the position chr10:97019824 using the SeqinR package (Charif and Lobry 2007) in R. The genome was completed by adding *wild-type* chromosomes of mm10 (retrieved from the UCSC Genome Browser), as well as the *del(CS38-40)* mutant chromosome 2 (Rodríguez-Carballo et al. 2020) available at (<https://zenodo.org/record/3826913#XxH985MzZTY>).

Ethical declaration on animal experimentation

All experiments were conducted in agreement with the Swiss Animal Welfare Act (LPA), under the animal experimentation license No GE 81/14 (to Professor Denis Duboule).

References:

- Agulnik, Sergei I., Virginia E. Papaioannou, and Lee M. Silver. 1998. "Cloning, Mapping, and Expression Analysis Of TBX15, a New Member of the T-Box Gene Family." *Genomics* 51 (1): 68–75. <https://doi.org/10.1006/geno.1998.5278>.
- Akin, Z. N., and A. J. Nazarali. 2005. "Hox Genes and Their Candidate Downstream Targets in the Developing Central Nervous System." *Cellular and Molecular Neurobiology* 25 (3): 697–741. <https://doi.org/10.1007/s10571-005-3971-9>.
- Alimohamed, Mohamed Z., Lennart F. Johansson, Eddy N. de Boer, Erik Splinter, Petra Klous, Mehmet Yilmaz, Anneke Bosga, et al. 2018. "Genetic Screening Test to Detect Translocations in Acute Leukemias by Use of Targeted Locus Amplification." *Clinical Chemistry* 64 (7): 1096–1103. <https://doi.org/10.1373/clinchem.2017.286047>.
- Aloia, Luigi, Bruno Di Stefano, and Luciano Di Croce. 2013. "Polycomb Complexes in Stem Cells and Embryonic Development." *Development* 140 (12): 2525–34. <https://doi.org/10.1242/dev.091553>.
- Amândio, Ana Rita, Lucille Lopez-Delisle, Christopher Chase Bolt, Bénédicte Mascrez, and Denis Duboule. 2020. "A Complex Regulatory Landscape Involved in the Development of Mammalian External Genitals." Edited by Job Dekker, Kevin Struhl, Job Dekker, and José Luis Gómez-Skármeta. *ELife* 9 (April): e52962. <https://doi.org/10.7554/eLife.52962>.
- Andersson, Robin, and Albin Sandelin. 2020. "Determinants of Enhancer and Promoter Activities of Regulatory Elements." *Nature Reviews Genetics* 21 (2): 71–87. <https://doi.org/10.1038/s41576-019-0173-8>.
- Andrey, Guillaume, Thomas Montavon, Bénédicte Mascrez, Federico Gonzalez, Daan Noordermeer, Marion Leleu, Didier Trono, François Spitz, and Denis Duboule. 2013. "A Switch between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs." *Science (New York, N.Y.)* 340 (6137): 1234167. <https://doi.org/10.1126/science.1234167>.
- van Arensbergen, Joris, Bas van Steensel, and Harmen J. Bussemaker. 2014. "In Search of the Determinants of Enhancer–Promoter Interaction Specificity." *Trends in Cell Biology* 24 (11): 695–702. <https://doi.org/10.1016/j.tcb.2014.07.004>.
- Austena, Liv M. I., Iros Barozzi, Marta Simonatto, Silvia Masella, Giulia Della Chiara, Serena Ghisletti, Alessia Curina, et al. 2015. "Transcription of Mammalian Cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination." *Molecular Cell* 60 (3): 460–74. <https://doi.org/10.1016/j.molcel.2015.09.018>.
- Banani, Salman F., Hyun O. Lee, Anthony A. Hyman, and Michael K. Rosen. 2017. "Biomolecular Condensates: Organizers of Cellular Biochemistry." *Nature Reviews Molecular Cell Biology* 18 (5): 285–98. <https://doi.org/10.1038/nrm.2017.7>.
- Banjade, Sudeep, Qiong Wu, Anuradha Mittal, William B. Peeples, Rohit V. Pappu, and Michael K. Rosen. 2015. "Conserved Interdomain Linker Promotes Phase Separation of the Multivalent Adaptor Protein Nck." *Proceedings of the National Academy of Sciences of the United States of America* 112 (47): E6426–6435. <https://doi.org/10.1073/pnas.1508778112>.
- Bannister, A. J., A. M. Falcão, and G. Castelo-Branco. 2017. "Chapter 2 - Histone Modifications and Histone Variants in Pluripotency and Differentiation." In *Chromatin Regulation and Dynamics*, edited by Anita Göndör, 35–64. Boston: Academic Press. <https://doi.org/10.1016/B978-0-12-803395-1.00002-2>.
- Bannister, Andrew J., and Tony Kouzarides. 2005. "Reversing Histone Methylation." *Nature* 436 (7054): 1103–6. <https://doi.org/10.1038/nature04048>.

- Bannister, Andrew J., Philip Zegerman, Janet F. Partridge, Eric A. Miska, Jean O. Thomas, Robin C. Allshire, and Tony Kouzarides. 2001. "Selective Recognition of Methylated Lysine 9 on Histone H3 by the HP1 Chromo Domain." *Nature* 410 (6824): 120–24. <https://doi.org/10.1038/35065138>.
- Barbieri, Mariano, Mita Chotalia, James Fraser, Liron-Mark Lavitas, Josée Dostie, Ana Pombo, and Mario Nicodemi. 2012. "Complexity of Chromatin Folding Is Captured by the Strings and Binders Switch Model." *Proceedings of the National Academy of Sciences* 109 (40): 16173–78. <https://doi.org/10.1073/pnas.1204799109>.
- Barski, Artem, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones, Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. 2007. "High-Resolution Profiling of Histone Methylations in the Human Genome." *Cell* 129 (4): 823–37. <https://doi.org/10.1016/j.cell.2007.05.009>.
- Bartman, Caroline R., Sarah C. Hsu, Chris C.-S. Hsiung, Arjun Raj, and Gerd A. Blobel. 2016. "Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping." *Molecular Cell* 62 (2): 237–47. <https://doi.org/10.1016/j.molcel.2016.03.007>.
- Barutcu, A. Rasim, Philipp G. Maass, Jordan P. Lewandowski, Catherine L. Weiner, and John L. Rinn. 2018. "A TAD Boundary Is Preserved upon Deletion of the CTCF-Rich Firre Locus." *Nature Communications* 9 (1): 1444. <https://doi.org/10.1038/s41467-018-03614-0>.
- Beagan, Jonathan A., Michael T. Duong, Katelyn R. Titus, Linda Zhou, Zhendong Cao, Jingjing Ma, Caroline V. Lachanski, Daniel R. Gillis, and Jennifer E. Phillips-Cremens. 2017. "YY1 and CTCF Orchestrate a 3D Chromatin Looping Switch during Early Neural Lineage Commitment." *Genome Research* 27 (7): 1139–52. <https://doi.org/10.1101/gr.215160.116>.
- Beccari, Leonardo, Nayuta Yakushiji-Kaminatsui, Joost M. Woltering, Anamaria Necsulea, Nicolas Lonfat, Eddie Rodríguez-Carballo, Benedicte Mascres, Shiori Yamamoto, Atsushi Kuroiwa, and Denis Duboule. 2016. "A Role for HOX13 Proteins in the Regulatory Switch between TADs at the HoxD Locus." *Genes & Development* 30 (10): 1172–86. <https://doi.org/10.1101/gad.281055.116>.
- Bell, A. C., A. G. West, and G. Felsenfeld. 1999. "The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators." *Cell* 98 (3): 387–96.
- Bergeron-Sandoval, Louis-Philippe, Nozhat Safaee, and Stephen W. Michnick. 2016. "Mechanisms and Consequences of Macromolecular Phase Separation." *Cell* 165 (5): 1067–79. <https://doi.org/10.1016/j.cell.2016.05.026>.
- Berlivet, Soizik, Denis Paquette, Annie Dumouchel, David Langlais, Josée Dostie, and Marie Kmita. 2013. "Clustering of Tissue-Specific Sub-TADs Accompanies the Regulation of HoxA Genes in Developing Limbs." *PLOS Genetics* 9 (12): e1004018. <https://doi.org/10.1371/journal.pgen.1004018>.
- Binder, Kurt, and Dieter Heermann. 2010. *Monte Carlo Simulation in Statistical Physics: An Introduction*. 5th ed. Graduate Texts in Physics. Berlin Heidelberg: Springer-Verlag. <https://doi.org/10.1007/978-3-642-03163-2>.
- Blackledge, Neil P., Nathan R. Rose, and Robert J. Klose. 2015. "Targeting Polycomb Systems to Regulate Gene Expression: Modifications to a Complex Story." *Nature Reviews Molecular Cell Biology* 16 (11): 643–49. <https://doi.org/10.1038/nrm4067>.

- Boeva, Valentina, Tatiana Popova, Kevin Bleakley, Pierre Chiche, Julie Cappel, Gudrun Schleiermacher, Isabelle Janoueix-Lerosey, Olivier Delattre, and Emmanuel Barillot. 2012. "Control-FREEC: A Tool for Assessing Copy Number and Allelic Content Using next-Generation Sequencing Data." *Bioinformatics* 28 (3): 423–25. <https://doi.org/10.1093/bioinformatics/btr670>.
- Boeva, Valentina, Andrei Zinovyev, Kevin Bleakley, Jean-Philippe Vert, Isabelle Janoueix-Lerosey, Olivier Delattre, and Emmanuel Barillot. 2011. "Control-Free Calling of Copy Number Alterations in Deep-Sequencing Data Using GC-Content Normalization." *Bioinformatics* 27 (2): 268–69. <https://doi.org/10.1093/bioinformatics/btq635>.
- Boija, Ann, Isaac A. Klein, Benjamin R. Sabari, Alessandra Dall'Agnese, Eliot L. Coffey, Alicia V. Zamudio, Charles H. Li, et al. 2018. "Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains." *Cell* 175 (7): 1842–1855.e16. <https://doi.org/10.1016/j.cell.2018.10.042>.
- Bolt, Christopher Chase, and Denis Duboule. 2020. "The Regulatory Landscapes of Developmental Genes." *Development* 147 (3). <https://doi.org/10.1242/dev.171736>.
- Bolzer, Andreas, Gregor Kreth, Irina Solovei, Daniela Koehler, Kaan Saracoglu, Christine Fauth, Stefan Müller, et al. 2005. "Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes." *PLOS Biology* 3 (5): e157. <https://doi.org/10.1371/journal.pbio.0030157>.
- Bonev, Boyan, Netta Mendelson Cohen, Quentin Szabo, Lauriane Fritsch, Giorgio L. Papadopoulos, Yaniv Lubling, Xiaole Xu, et al. 2017. "Multiscale 3D Genome Rewiring during Mouse Neural Development." *Cell* 171 (3): 557–572.e24. <https://doi.org/10.1016/j.cell.2017.09.043>.
- Brangwynne, Clifford P., Christian R. Eckmann, David S. Courson, Agata Rybarska, Carsten Hoege, Jöbin Gharakhani, Frank Jülicher, and Anthony A. Hyman. 2009. "Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation." *Science* 324 (5935): 1729–32. <https://doi.org/10.1126/science.1172046>.
- Brookes, Emily, Inês de Santiago, Daniel Hebenstreit, Kelly J. Morris, Tom Carroll, Sheila Q. Xie, Julie K. Stock, et al. 2012. "Polycomb Associates Genome-Wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs." *Cell Stem Cell* 10 (2): 157–70. <https://doi.org/10.1016/j.stem.2011.12.017>.
- Brown, Jill M., Nigel A. Roberts, Bryony Graham, Dominic Waithe, Christoffer Lagerholm, Jelena M. Telenius, Sara De Ornellas, et al. 2018. "A Tissue-Specific Self-Interacting Chromatin Domain Forms Independently of Enhancer-Promoter Interactions." *Nature Communications* 9 (1): 1–15. <https://doi.org/10.1038/s41467-018-06248-4>.
- Buenrostro, Jason D., Paul G. Giresi, Lisa C. Zaba, Howard Y. Chang, and William J. Greenleaf. 2013. "Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position." *Nature Methods* 10 (12): 1213–18. <https://doi.org/10.1038/nmeth.2688>.
- Busslinger, Georg A., Roman R. Stocsits, Petra van der Lelij, Elin Axelsson, Antonio Tedeschi, Niels Galjart, and Jan-Michael Peters. 2017. "Cohesin Is Positioned in Mammalian Genomes by Transcription, CTCF and Wapl." *Nature* 544 (7651): 503–7. <https://doi.org/10.1038/nature22063>.
- Chammas, Paul, Ivano Mocavini, and Luciano Di Croce. 2020. "Engaging Chromatin: PRC2 Structure Meets Function." *British Journal of Cancer* 122 (3): 315–28. <https://doi.org/10.1038/s41416-019-0615-2>.

- Chang, Li-Hsin, Sourav Ghosh, and Daan Noordermeer. 2020. "TADs and Their Borders: Free Movement or Building a Wall?" *Journal of Molecular Biology*, Perspectives on Chromosome Folding, 432 (3): 643–52. <https://doi.org/10.1016/j.jmb.2019.11.025>.
- Charif, Delphine, and Jean R. Lobry. 2007. "SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis." *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations, Biological and Medical Physics, Biomedical Engineering*, 207. https://doi.org/10.1007/978-3-540-35306-5_10.
- Chen, Nansheng. 2004. "Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences." *Current Protocols in Bioinformatics* 5 (1): 4.10.1-4.10.14. <https://doi.org/10.1002/0471250953.bi0410s05>.
- Cheutin, Thierry, Adrian J. McNairn, Thomas Jenuwein, David M. Gilbert, Prim B. Singh, and Tom Misteli. 2003. "Maintenance of Stable Heterochromatin Domains by Dynamic HP1 Binding." *Science* 299 (5607): 721–25. <https://doi.org/10.1126/science.1078572>.
- Chong, Shasha, Claire Dugast-Darzacq, Zhe Liu, Peng Dong, Gina M. Dailey, Claudia Cattoglio, Alec Heckert, et al. 2018. "Imaging Dynamic and Selective Low-Complexity Domain Interactions That Control Gene Transcription." *Science* 361 (6400). <https://doi.org/10.1126/science.aar2555>.
- Crane, Emily, Qian Bian, Rachel Patton McCord, Bryan R. Lajoie, Bayly S. Wheeler, Edward J. Ralston, Satoru Uzawa, Job Dekker, and Barbara J. Meyer. 2015. "Condensin-Driven Remodelling of X Chromosome Topology during Dosage Compensation." *Nature* 523 (7559): 240–44. <https://doi.org/10.1038/nature14450>.
- Cremer, Marion, Florian Grasser, Christian Lanctôt, Stefan Müller, Michaela Neusser, Roman Zinner, Irina Solovei, and Thomas Cremer. 2008. "Multicolor 3D Fluorescence in Situ Hybridization for Imaging Interphase Chromosomes." *Methods in Molecular Biology (Clifton, N.J.)* 463: 205–39. https://doi.org/10.1007/978-1-59745-406-3_15.
- Cremer, Thomas, and Marion Cremer. 2010. "Chromosome Territories." *Cold Spring Harbor Perspectives in Biology* 2 (3): a003889. <https://doi.org/10.1101/cshperspect.a003889>.
- Creyghton, Menno P., Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, et al. 2010. "Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State." *Proceedings of the National Academy of Sciences of the United States of America* 107 (50): 21931–36. <https://doi.org/10.1073/pnas.1016071107>.
- Cuddapah, Suresh, Raja Jothi, Dustin E. Schones, Tae-Young Roh, Kairong Cui, and Keji Zhao. 2009. "Global Analysis of the Insulator Binding Protein CTCF in Chromatin Barrier Regions Reveals Demarcation of Active and Repressive Domains." *Genome Research* 19 (1): 24–32. <https://doi.org/10.1101/gr.082800.108>.
- Darbellay, Fabrice, Célia Bochaton, Lucille Lopez-Delisle, Bénédicte Mascrez, Patrick Tschopp, Saskia Delpretti, Jozsef Zakany, and Denis Duboule. 2019. "The Constrained Architecture of Mammalian Hox Gene Clusters." *Proceedings of the National Academy of Sciences* 116 (27): 13424–33. <https://doi.org/10.1073/pnas.1904602116>.
- Darbellay, Fabrice, and Denis Duboule. 2016. "Chapter Sixteen - Topological Domains, Metagenes, and the Emergence of Pleiotropic Regulations at Hox Loci." In *Current Topics in Developmental Biology*, edited by Paul M. Wassarman, 116:299–314. Essays on Developmental Biology, Part A. Academic Press. <https://doi.org/10.1016/bs.ctdb.2015.11.022>.

- David, Fabrice P. A., Julien Delafontaine, Solenne Carat, Frederick J. Ross, Gregory Lefebvre, Yohan Jarosz, Lucas Sinclair, Daan Noordermeer, Jacques Rougemont, and Marion Leleu. 2014. "HTSstation: A Web Application and Open-Access Libraries for High-Throughput Sequencing Data Analysis." *PLOS ONE* 9 (1): e85879. <https://doi.org/10.1371/journal.pone.0085879>.
- Davidson, Iain F., Benedikt Bauer, Daniela Goetz, Wen Tang, Gordana Wutz, and Jan-Michael Peters. 2019. "DNA Loop Extrusion by Human Cohesin." *Science* 366 (6471): 1338–45. <https://doi.org/10.1126/science.aaz3418>.
- Deaton, Aimée M., and Adrian Bird. 2011. "CpG Islands and the Regulation of Transcription." *Genes & Development* 25 (10): 1010–22. <https://doi.org/10.1101/gad.2037511>.
- Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. 2002. "Capturing Chromosome Conformation." *Science* 295 (5558): 1306–11. <https://doi.org/10.1126/science.1067799>.
- Delpretti, Saskia, Thomas Montavon, Marion Leleu, Elisabeth Joye, Athanasia Tzika, Michel Milinkovitch, and Denis Duboule. 2013. "Multiple Enhancers Regulate Hoxd Genes and the Hotdog LncRNA during Cecum Budding." *Cell Reports* 5 (1): 137–50. <https://doi.org/10.1016/j.celrep.2013.09.002>.
- Denholtz, Matthew, Giancarlo Bonora, Constantinos Chronis, Erik Splinter, Wouter de Laat, Jason Ernst, Matteo Pellegrini, and Kathrin Plath. 2013. "Long-Range Chromatin Contacts in Embryonic Stem Cells Reveal a Role for Pluripotency Factors and Polycomb Proteins in Genome Organization." *Cell Stem Cell* 13 (5): 602–16. <https://doi.org/10.1016/j.stem.2013.08.013>.
- Deschamps, Jacqueline, and Denis Duboule. 2017. "Embryonic Timing, Axial Stem Cells, Chromatin Dynamics, and the Hox Clock." *Genes & Development* 31 (14): 1406–16. <https://doi.org/10.1101/gad.303123.117>.
- Deschamps, Jacqueline, and Johan van Nes. 2005. "Developmental Regulation of the Hox Genes during Axial Morphogenesis in the Mouse." *Development* 132 (13): 2931–42. <https://doi.org/10.1242/dev.01897>.
- Despang, Alexandra, Robert Schöpflin, Martin Franke, Salaheddine Ali, Ivana Jerković, Christina Paliou, Wing-Lee Chan, et al. 2019. "Functional Dissection of the Sox9 – Kcnj2 Locus Identifies Nonessential and Instructive Roles of TAD Architecture." *Nature Genetics* 51 (8): 1263–71. <https://doi.org/10.1038/s41588-019-0466-z>.
- Dileep, Vishnu, Ferhat Ay, Jiao Sima, Daniel L. Vera, William S. Noble, and David M. Gilbert. 2015. "Topologically Associating Domains and Their Long-Range Contacts Are Established during Early G1 Coincident with the Establishment of the Replication-Timing Program." *Genome Research* 25 (8): 1104–13. <https://doi.org/10.1101/gr.183699.114>.
- Dixon, Jesse R., David U. Gorkin, and Bing Ren. 2016. "Chromatin Domains: The Unit of Chromosome Organization." *Molecular Cell* 62 (5): 668–80. <https://doi.org/10.1016/j.molcel.2016.05.018>.
- Dixon, Jesse R., Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E. Antosiewicz-Bourget, Ah Young Lee, Zhen Ye, et al. 2015. "Chromatin Architecture Reorganization during Stem Cell Differentiation." *Nature* 518 (7539): 331–36. <https://doi.org/10.1038/nature14222>.
- Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. 2012. "Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions." *Nature* 485 (7398): 376–80. <https://doi.org/10.1038/nature11082>.

- Doi, M., and S. F. Edwards. 1988. *The Theory of Polymer Dynamics*. International Series of Monographs on Physics. Oxford, New York: Oxford University Press.
- Dollé, P., J. C. Izpisua-Belmonte, J. M. Brown, C. Tickle, and D. Duboule. 1991. "HOX-4 Genes and the Morphogenesis of Mammalian Genitalia." *Genes & Development* 5 (10): 1767–76. <https://doi.org/10.1101/gad.5.10.1767>.
- Dollé, Pascal, Juan-Carlos Izpisua-Belmonte, Hildegard Falkenstein, Armand Renucci, and Denis Duboule. 1989. "Coordinate Expression of the Murine Hox-5 Complex Homoeobox-Containing Genes during Limb Pattern Formation." *Nature* 342 (6251): 767–72. <https://doi.org/10.1038/342767a0>.
- Dong, Junchao, Rohit A. Panchakshari, Tingting Zhang, Yu Zhang, Jiazhi Hu, Sabrina A. Volpi, Robin M. Meyers, et al. 2015. "Orientation-Specific Joining of AID-Initiated DNA Breaks Promotes Antibody Class Switching." *Nature* 525 (7567): 134–39. <https://doi.org/10.1038/nature14970>.
- Dostie, Josée, Todd A. Richmond, Ramy A. Arnaout, Rebecca R. Selzer, William L. Lee, Tracey A. Honan, Eric D. Rubio, et al. 2006. "Chromosome Conformation Capture Carbon Copy (5C): A Massively Parallel Solution for Mapping Interactions between Genomic Elements." *Genome Research* 16 (10): 1299–1309. <https://doi.org/10.1101/gr.5571506>.
- Downen, Jill M., Zi Peng Fan, Denes Hnisz, Gang Ren, Brian J. Abraham, Lyndon N. Zhang, Abraham S. Weintraub, et al. 2014. "Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes." *Cell* 159 (2): 374–87. <https://doi.org/10.1016/j.cell.2014.09.030>.
- Duboule, D., and P. Dollé. 1989. "The Structural and Functional Organization of the Murine HOX Gene Family Resembles That of Drosophila Homeotic Genes." *The EMBO Journal* 8 (5): 1497–1505. <https://doi.org/10.1002/j.1460-2075.1989.tb03534.x>.
- Duboule, Denis. 1994. "Temporal Colinearity and the Phylotypic Progression: A Basis for the Stability of a Vertebrate Bauplan and the Evolution of Morphologies through Heterochrony." *Development* 1994 (Supplement): 135–42.
- . 2007. "The Rise and Fall of Hox Gene Clusters." *Development* 134 (14): 2549–60. <https://doi.org/10.1242/dev.001065>.
- Eagen, Kyle P. 2018. "Principles of Chromosome Architecture Revealed by Hi-C." *Trends in Biochemical Sciences* 43 (6): 469–78. <https://doi.org/10.1016/j.tibs.2018.03.006>.
- Eagen, Kyle P., Tom A. Hartl, and Roger D. Kornberg. 2015. "Stable Chromosome Condensation Revealed by Chromosome Conformation Capture." *Cell* 163 (4): 934–46. <https://doi.org/10.1016/j.cell.2015.10.026>.
- Eisenberg, Eli, and Erez Y. Levanon. 2013. "Human Housekeeping Genes, Revisited." *Trends in Genetics, Human Genetics*, 29 (10): 569–74. <https://doi.org/10.1016/j.tig.2013.05.010>.
- Eissenberg, J. C., G. D. Morris, G. Reuter, and T. Hartnett. 1992. "The Heterochromatin-Associated Protein Hp-1 Is an Essential Protein in Drosophila with Dosage-Dependent Effects on Position-Effect Variegation." *Genetics* 131 (2): 345–52.
- Eissenberg, Joel C, and Sarah CR Elgin. 2000. "The HP1 Protein Family: Getting a Grip on Chromatin." *Current Opinion in Genetics & Development* 10 (2): 204–10. [https://doi.org/10.1016/S0959-437X\(00\)00058-7](https://doi.org/10.1016/S0959-437X(00)00058-7).
- Erdel, Fabian, and Karsten Rippe. 2018. "Formation of Chromatin Subcompartments by Phase Separation." *Biophysical Journal* 114 (10): 2262–70. <https://doi.org/10.1016/j.bpj.2018.03.011>.

- Farioli-Vecchioli, Stefano, Laura Micheli, Daniele Saraulli, Manuela Ceccarelli, Sara Cannas, Raffaella Scardigli, Luca Leonardi, et al. 2012. "Btg1 Is Required to Maintain the Pool of Stem and Progenitor Cells of the Dentate Gyrus and Subventricular Zone." *Frontiers in Neuroscience* 6. <https://doi.org/10.3389/fnins.2012.00124>.
- Feng, Jianxing, Tao Liu, Bo Qin, Yong Zhang, and Xiaole Shirley Liu. 2012. "Identifying ChIP-Seq Enrichment Using MACS." *Nature Protocols* 7 (9). <https://doi.org/10.1038/nprot.2012.101>.
- Filippova, G. N., S. Fagerlie, E. M. Klenova, C. Myers, Y. Dehner, G. Goodwin, P. E. Neiman, S. J. Collins, and V. V. Lobanenko. 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." *Molecular and Cellular Biology* 16 (6): 2802–13.
- Finn, Elizabeth H., and Tom Misteli. 2019. "Molecular Basis and Biological Function of Variability in Spatial Genome Organization." *Science* 365 (6457). <https://doi.org/10.1126/science.aaw9498>.
- Finn, Elizabeth H., Gianluca Pegoraro, Hugo B. Brandão, Anne-Laure Valton, Marlies E. Oomen, Job Dekker, Leonid Mirny, and Tom Misteli. 2019. "Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization." *Cell* 176 (6): 1502-1515.e10. <https://doi.org/10.1016/j.cell.2019.01.020>.
- Franke, Martin, Daniel M. Ibrahim, Guillaume Andrey, Wibke Schwarzer, Verena Heinrich, Robert Schöpflin, Katerina Kraft, et al. 2016. "Formation of New Chromatin Domains Determines Pathogenicity of Genomic Duplications." *Nature* 538 (7624): 265–69. <https://doi.org/10.1038/nature19800>.
- Frazer, Kelly A., Lior Pachter, Alexander Poliakov, Edward M. Rubin, and Inna Dubchak. 2004. "VISTA: Computational Tools for Comparative Genomics." *Nucleic Acids Research* 32 (suppl_2): W273–79. <https://doi.org/10.1093/nar/gkh458>.
- Fromental-Ramain, C., X. Warot, S. Lakkaraju, B. Favier, H. Haack, C. Birling, A. Dierich, P. Doll E, and P. Chambon. 1996. "Specific and Redundant Functions of the Paralogous Hoxa-9 and Hoxd-9 Genes in Forelimb and Axial Skeleton Patterning." *Development* 122 (2): 461–72.
- Fu, Guo, Sébastien Vallée, Vasily Rybakin, Marielena V. McGuire, Jeanette Ampudia, Claudia Brockmeyer, Mogjiborahman Salek, et al. 2009. "Themis Controls Thymocyte Selection through Regulation of T Cell Antigen Receptor–Mediated Signaling." *Nature Immunology* 10 (8): 848–56. <https://doi.org/10.1038/ni.1766>.
- Fudenberg, Geoffrey, Nezar Abdennur, Maxim Imakaev, Anton Goloborodko, and Leonid A. Mirny. 2017. "Emerging Evidence of Chromosome Folding by Loop Extrusion." *Cold Spring Harbor Symposia on Quantitative Biology* 82 (January): 45–55. <https://doi.org/10.1101/sqb.2017.82.034710>.
- Fudenberg, Geoffrey, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny. 2016. "Formation of Chromosomal Domains by Loop Extrusion." *Cell Reports* 15 (9): 2038–49. <https://doi.org/10.1016/j.celrep.2016.04.085>.
- Furlong, Eileen E. M., and Michael Levine. 2018. "Developmental Enhancers and Chromosome Topology." *Science* 361 (6409): 1341–45. <https://doi.org/10.1126/science.aau0320>.
- Gabrieli, Tslil, Hila Sharim, Dena Fridman, Nissim Arbib, Yael Michaeli, and Yuval Ebenstein. 2018. "Selective Nanopore Sequencing of Human BRCA1 by Cas9-Assisted Targeting of Chromosome Segments (CATCH)." *Nucleic Acids Research* 46 (14): e87–e87. <https://doi.org/10.1093/nar/gky411>.

- García-Bellido, A. 1977. "Homoeotic and Atavic Mutations in Insects." *American Zoologist* 17 (3): 613–29.
- Gasperini, Molly, Andrew J. Hill, José L. McFaline-Figueroa, Beth Martin, Seungsoo Kim, Melissa D. Zhang, Dana Jackson, et al. 2019. "A Genome-Wide Framework for Mapping Gene Regulation via Cellular Genetic Screens." *Cell* 176 (1): 377–390.e19. <https://doi.org/10.1016/j.cell.2018.11.029>.
- Gaunt, Stephen J., Paul T. Sharpe, and Denis Duboule. 1988. "Spatially Restricted Domains of Homeo-Gene Transcripts in Mouse Embryos: Relation to a Segmented Body Plan." *Development* 104 (Supplement): 169–79.
- Gehring, W. J., and Y. Hiromi. 1986. "Homeotic Genes and the Homeobox." *Annual Review of Genetics* 20: 147–73. <https://doi.org/10.1146/annurev.ge.20.120186.001051>.
- Ghavi-Helm, Yad, Aleksander Jankowski, Sascha Meiers, Rebecca R. Viales, Jan O. Korb, and Eileen E. M. Furlong. 2019. "Highly Rearranged Chromosomes Reveal Uncoupling between Genome Topology and Gene Expression." *Nature Genetics* 51 (8): 1272–82. <https://doi.org/10.1038/s41588-019-0462-3>.
- Giesselmann, Pay, Björn Brändl, Etienne Raimondeau, Rebecca Bowen, Christian Rohrandt, Rashmi Tandon, Helene Kretzmer, et al. 2019. "Analysis of Short Tandem Repeat Expansions and Their Methylation State with Nanopore Sequencing." *Nature Biotechnology* 37 (12): 1478–81. <https://doi.org/10.1038/s41587-019-0293-x>.
- Gilpatrick, Timothy, Isac Lee, James E. Graham, Etienne Raimondeau, Rebecca Bowen, Andrew Heron, Bradley Downs, Saraswati Sukumar, Fritz J. Sedlazeck, and Winston Timp. 2020. "Targeted Nanopore Sequencing with Cas9-Guided Adapter Ligation." *Nature Biotechnology* 38 (4): 433–38. <https://doi.org/10.1038/s41587-020-0407-5>.
- Giorgetti, Luca, and Edith Heard. 2016. "Closing the Loop: 3C versus DNA FISH." *Genome Biology* 17 (1): 215. <https://doi.org/10.1186/s13059-016-1081-2>.
- Gligoris, Thomas, and Jan Löwe. 2016. "Structural Insights into Ring Formation of Cohesin and Related SMC Complexes." *Trends in Cell Biology* 26 (9): 680–93. <https://doi.org/10.1016/j.tcb.2016.04.002>.
- Gong, Yixiao, Charalampos Lazaris, Theodore Sakellaropoulos, Aurelie Lozano, Prabhanjan Kambadur, Panagiotis Ntziachristos, Iannis Aifantis, and Aristotelis Tsirigos. 2018. "Stratification of TAD Boundaries Reveals Preferential Insulation of Super-Enhancers by Strong Boundaries." *Nature Communications* 9 (1): 542. <https://doi.org/10.1038/s41467-018-03017-1>.
- Gonzalez, Federico, Denis Duboule, and François Spitz. 2007. "Transgenic Analysis of Hoxd Gene Regulation during Digit Development." *Developmental Biology* 306 (2): 847–59. <https://doi.org/10.1016/j.ydbio.2007.03.020>.
- Goodwin, Leslie O., Erik Splinter, Tiffany L. Davis, Rachel Urban, Hao He, Robert E. Braun, Elissa J. Chesler, et al. 2019. "Large-Scale Discovery of Mouse Transgenic Integration Sites Reveals Frequent Structural Variation and Insertional Mutagenesis." *Genome Research*, January, gr.233866.117. <https://doi.org/10.1101/gr.233866.117>.
- Goodwin, Sara, James Gurtowski, Scott Ethe-Sayers, Panchajanya Deshpande, Michael C. Schatz, and W. Richard McCombie. 2015. "Oxford Nanopore Sequencing, Hybrid Error Correction, and de Novo Assembly of a Eukaryotic Genome." *Genome Research*, October. <https://doi.org/10.1101/gr.191395.115>.

- Gould, A., A. Morrison, G. Sproat, R. A. White, and R. Krumlauf. 1997. "Positive Cross-Regulation and Enhancer Sharing: Two Mechanisms for Specifying Overlapping Hox Expression Patterns." *Genes & Development* 11 (7): 900–913. <https://doi.org/10.1101/gad.11.7.900>.
- Grau, Daniel J., Brad A. Chapman, Joe D. Garlick, Mark Borowsky, Nicole J. Francis, and Robert E. Kingston. 2011. "Compaction of Chromatin by Diverse Polycomb Group Proteins Requires Localized Regions of High Charge." *Genes & Development* 25 (20): 2210–21. <https://doi.org/10.1101/gad.17288211>.
- Guerreiro, Isabel, Sandra Gitto, Ana Novoa, Julien Codourey, Thi Hanh Nguyen Huynh, Federico Gonzalez, Michel C Milinkovitch, Moises Mallo, and Denis Duboule. 2016. "Reorganisation of Hoxd Regulatory Landscapes during the Evolution of a Snake-like Body Plan." Edited by Robb Krumlauf. *ELife* 5 (August): e16087. <https://doi.org/10.7554/eLife.16087>.
- Haarhuis, Judith H. I., Robin H. van der Weide, Vincent A. Blomen, J. Omar Yáñez-Cuna, Mario Amendola, Marjon S. van Ruiten, Peter H. L. Krijger, et al. 2017. "The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension." *Cell* 169 (4): 693-707.e14. <https://doi.org/10.1016/j.cell.2017.04.013>.
- Hansen, Anders S., Tsung-Han S. Hsieh, Claudia Cattoglio, Iryna Pustova, Ricardo Saldaña-Meyer, Danny Reinberg, Xavier Darzacq, and Robert Tjian. 2019. "Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF." *Molecular Cell* 76 (3): 395-411.e13. <https://doi.org/10.1016/j.molcel.2019.07.039>.
- Hansen, Anders S, Iryna Pustova, Claudia Cattoglio, Robert Tjian, and Xavier Darzacq. 2017. "CTCF and Cohesin Regulate Chromatin Loop Stability with Distinct Dynamics." Edited by David Sherratt. *ELife* 6 (May): e25776. <https://doi.org/10.7554/eLife.25776>.
- Hanssen, Lars L. P., Mira T. Kassouf, A. Marieke Oudelaar, Daniel Biggs, Chris Preece, Damien J. Downes, Matthew Gosden, et al. 2017. "Tissue-Specific CTCF–Cohesin-Mediated Chromatin Architecture Delimits Enhancer Interactions and Function in Vivo." *Nature Cell Biology* 19 (8): 952–61. <https://doi.org/10.1038/ncb3573>.
- Hawkins, R. David, Gary C. Hon, Leonard K. Lee, QueMinh Ngo, Ryan Lister, Mattia Pelizzola, Lee E. Edsall, et al. 2010. "Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells." *Cell Stem Cell* 6 (5): 479–91. <https://doi.org/10.1016/j.stem.2010.03.018>.
- Hayakawa, Tomohiro, Tokuko Haraguchi, Hiroshi Masumoto, and Yasushi Hiraoka. 2003. "Cell Cycle Behavior of Human HP1 Subtypes: Distinct Molecular Domains of HP1 Are Required for Their Centromeric Localization during Interphase and Metaphase." *Journal of Cell Science* 116 (16): 3327–38. <https://doi.org/10.1242/jcs.00635>.
- Henriques, Telmo, Benjamin S. Scruggs, Michiko O. Inouye, Ginger W. Muse, Lucy H. Williams, Adam B. Burkholder, Christopher A. Lavender, David C. Fargo, and Karen Adelman. 2018. "Widespread Transcriptional Pausing and Elongation Control at Enhancers." *Genes & Development* 32 (1): 26–41. <https://doi.org/10.1101/gad.309351.117>.
- Hérault, Yann, Minoo Rassoulzadegan, François Cuzin, and Denis Duboule. 1998. "Engineering Chromosomes in Mice through Targeted Meiotic Recombination (TAMERE)." *Nature Genetics* 20 (4): 381–84. <https://doi.org/10.1038/3861>.
- Hiragami-Hamada, Kyoko, Kaori Shinmyozu, Daizo Hamada, Yoshiro Tatsu, Koichi Uegaki, Shinsuke Fujiwara, and Jun-ichi Nakayama. 2011. "N-Terminal Phosphorylation of HP1α Promotes Its Chromatin Binding." *Molecular and Cellular Biology* 31 (6): 1186–1200. <https://doi.org/10.1128/MCB.01012-10>.

- Hnisz, Denes, Brian J. Abraham, Tong Ihn Lee, Ashley Lau, Violaine Saint-André, Alla A. Sigova, Heather A. Hoke, and Richard A. Young. 2013. "Super-Enhancers in the Control of Cell Identity and Disease." *Cell* 155 (4): 934–47. <https://doi.org/10.1016/j.cell.2013.09.053>.
- Hnisz, Denes, Krishna Shrinivas, Richard A. Young, Arup K. Chakraborty, and Phillip A. Sharp. 2017. "A Phase Separation Model for Transcriptional Control." *Cell* 169 (1): 13–23. <https://doi.org/10.1016/j.cell.2017.02.007>.
- Hnisz, Denes, Abraham S. Weintraub, Daniel S. Day, Anne-Laure Valton, Rasmus O. Bak, Charles H. Li, Johanna Goldmann, et al. 2016. "Activation of Proto-Oncogenes by Disruption of Chromosome Neighborhoods." *Science* 351 (6280): 1454–58. <https://doi.org/10.1126/science.aad9024>.
- Hu, Jiazhi, Yu Zhang, Lijuan Zhao, Richard L. Frock, Zhou Du, Robin M. Meyers, Fei-long Meng, David G. Schatz, and Frederick W. Alt. 2015. "Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes." *Cell* 163 (4): 947–59. <https://doi.org/10.1016/j.cell.2015.10.016>.
- Hueber, Stefanie D., and Ingrid Lohmann. 2008. "Shaping Segments: Hox Gene Function in the Genomic Age." *BioEssays* 30 (10): 965–79. <https://doi.org/10.1002/bies.20823>.
- Hughes, Jennifer F., Helen Skaletsky, Tatyana Pyntikova, Patrick J. Minx, Tina Graves, Steve Rozen, Richard K. Wilson, and David C. Page. 2005. "Conservation of Y-Linked Genes during Human Evolution Revealed by Comparative Sequencing in Chimpanzee." *Nature* 437 (7055): 100–103. <https://doi.org/10.1038/nature04101>.
- Ing-Simmons, Elizabeth, and Juan M. Vaquerizas. 2019. "Visualising Three-Dimensional Genome Organisation in Two Dimensions." *Development* 146 (19). <https://doi.org/10.1242/dev.177162>.
- Izpisúa-Belmonte, J.c., H. Falkenstein, P. Dollé, A. Renucci, and D. Duboule. 1991. "Murine Genes Related to the Drosophila AbdB Homeotic Genes Are Sequentially Expressed during Development of the Posterior Part of the Body." *The EMBO Journal* 10 (8): 2279–89. <https://doi.org/10.1002/j.1460-2075.1991.tb07764.x>.
- Jacobs, Steven A., Sean D. Taverna, Yinong Zhang, Scott D. Briggs, Jinmei Li, Joel C. Eissenberg, C. David Allis, and Sepideh Khorasanizadeh. 2001. "Specificity of the HP1 Chromo Domain for the Methylated N-Terminus of Histone H3." *The EMBO Journal* 20 (18): 5232–41. <https://doi.org/10.1093/emboj/20.18.5232>.
- Jain, Miten, Hugh E. Olsen, Benedict Paten, and Mark Akeson. 2016. "The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community." *Genome Biology* 17 (1): 239. <https://doi.org/10.1186/s13059-016-1103-0>.
- Jane, S.m., P.a. Ney, E.f. Vanin, D.l. Gumucio, and A.w. Nienhuis. 1992. "Identification of a Stage Selector Element in the Human Gamma-Globin Gene Promoter That Fosters Preferential Interaction with the 5' HS2 Enhancer When in Competition with the Beta-Promoter." *The EMBO Journal* 11 (8): 2961–69. <https://doi.org/10.1002/j.1460-2075.1992.tb05366.x>.
- Ji, Xiong, Daniel B. Dadon, Benjamin E. Powell, Zi Peng Fan, Diego Borges-Rivera, Sigal Shachar, Abraham S. Weintraub, et al. 2016. "3D Chromosome Regulatory Landscape of Human Pluripotent Cells." *Cell Stem Cell* 18 (2): 262–75. <https://doi.org/10.1016/j.stem.2015.11.007>.

- Jiang, Y. P., H. Wang, P. D'Eustachio, J. M. Musacchio, J. Schlessinger, and J. Sap. 1993. "Cloning and Characterization of R-PTP-Kappa, a New Member of the Receptor Protein Tyrosine Phosphatase Family with a Proteolytically Cleaved Cellular Adhesion Molecule-like Extracellular Region." *Molecular and Cellular Biology* 13 (5): 2942–51. <https://doi.org/10.1128/MCB.13.5.2942>.
- Johnson, Andy L., L. Aravind, Natalia Shulzhenko, Andre Morgun, See-Young Choi, Tanya L. Crockford, Teresa Lambe, et al. 2009. "Themis Is a Member of a New Metazoan Gene Family and Is Required for the Completion of Thymocyte Positive Selection." *Nature Immunology* 10 (8): 831–39. <https://doi.org/10.1038/ni.1769>.
- Jones, Peter A. 2012. "Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and Beyond." *Nature Reviews Genetics* 13 (7): 484–92. <https://doi.org/10.1038/nrg3230>.
- Kempfer, Rieke, and Ana Pombo. 2020. "Methods for Mapping 3D Chromosome Architecture." *Nature Reviews Genetics* 21 (4): 207–26. <https://doi.org/10.1038/s41576-019-0195-2>.
- Kim, Seungsoo, and Jay Shendure. 2019. "Mechanisms of Interplay between Transcription Factors and the 3D Genome." *Molecular Cell* 76 (2): 306–19. <https://doi.org/10.1016/j.molcel.2019.08.010>.
- Kim, Tae Hoon, Ziedulla K. Abdullaev, Andrew D. Smith, Keith A. Ching, Dmitri I. Loukinov, Roland D. Green, Michael Q. Zhang, Victor V. Lobanenko, and Bing Ren. 2007. "Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome." *Cell* 128 (6): 1231–45. <https://doi.org/10.1016/j.cell.2006.12.048>.
- Kim, Tae-Kyung, Martin Hemberg, Jesse M. Gray, Allen M. Costa, Daniel M. Bear, Jing Wu, David A. Harmin, et al. 2010. "Widespread Transcription at Neuronal Activity-Regulated Enhancers." *Nature* 465 (7295): 182–87. <https://doi.org/10.1038/nature09033>.
- Klenova, E. M., R. H. Nicolas, H. F. Paterson, A. F. Carne, C. M. Heath, G. H. Goodwin, P. E. Neiman, and V. V. Lobanenko. 1993. "CTCF, a Conserved Nuclear Factor Required for Optimal Transcriptional Activity of the Chicken c-Myc Gene, Is an 11-Zn-Finger Protein Differentially Expressed in Multiple Forms." *Molecular and Cellular Biology* 13 (12): 7612–24. <https://doi.org/10.1128/MCB.13.12.7612>.
- Kmita, Marie, and Denis Duboule. 2003. "Organizing Axes in Time and Space; 25 Years of Colinear Tinkering." *Science* 301 (5631): 331–33. <https://doi.org/10.1126/science.1085753>.
- Kmita, Marie, Nadine Fraudeau, Yann Hérault, and Denis Duboule. 2002. "Serial Deletions and Duplications Suggest a Mechanism for the Collinearity of Hoxd Genes in Limbs." *Nature* 420 (6912): 145–50. <https://doi.org/10.1038/nature01189>.
- Kmita, Marie, Basile Tarchini, Jozsef Zákány, Malcolm Logan, Clifford J. Tabin, and Denis Duboule. 2005. "Early Developmental Arrest of Mammalian Limbs Lacking HoxA / HoxD Gene Function." *Nature* 435 (7045): 1113–16. <https://doi.org/10.1038/nature03648>.
- Krumlauf, Robb. 1994. "Hox Genes in Vertebrate Development." *Cell* 78 (2): 191–201. [https://doi.org/10.1016/0092-8674\(94\)90290-9](https://doi.org/10.1016/0092-8674(94)90290-9).
- Kruse, Kai, Clemens B. Hug, Benjamín Hernández-Rodríguez, and Juan M. Vaquerizas. 2016. "TADtool: Visual Parameter Identification for TAD-Calling Algorithms." *Bioinformatics* 32 (20): 3190–92. <https://doi.org/10.1093/bioinformatics/btw368>.
- Kuijper, Sanne, Annemiek Beverdam, Carla Kroon, Antje Brouwer, Sophie Candille, Gregory Barsh, and Frits Meijlink. 2005. "Genetics of Shoulder Girdle Formation: Roles of Tbx15 and Aristaless-like Genes." *Development* 132 (7): 1601–10. <https://doi.org/10.1242/dev.01735>.

- Kvon, Evgeny Z. 2015. "Using Transgenic Reporter Assays to Functionally Characterize Enhancers in Animals." *Genomics*, Recent advances in functional assays of transcriptional enhancers, 106 (3): 185–92. <https://doi.org/10.1016/j.ygeno.2015.06.007>.
- Kwon, Iimin, Masato Kato, Siheng Xiang, Leeju Wu, Pano Theodoropoulos, Hamid Mirzaei, Tina Han, Shanhai Xie, Jeffery L. Corden, and Steven L. McKnight. 2013. "Phosphorylation-Regulated Binding of RNA Polymerase II to Fibrous Polymers of Low-Complexity Domains." *Cell* 155 (5): 1049–60. <https://doi.org/10.1016/j.cell.2013.10.033>.
- Laboulaye, Mallory A., Xin Duan, Mu Qiao, Irene E. Whitney, and Joshua R. Sanes. 2018. "Mapping Transgene Insertion Sites Reveals Complex Interactions Between Mouse Transgenes and Neighboring Endogenous Genes." *Frontiers in Molecular Neuroscience* 11. <https://doi.org/10.3389/fnmol.2018.00385>.
- Lachner, Monika, Dónal O'Carroll, Stephen Rea, Karl Mechtler, and Thomas Jenuwein. 2001. "Methylation of Histone H3 Lysine 9 Creates a Binding Site for HP1 Proteins." *Nature* 410 (6824): 116–20. <https://doi.org/10.1038/35065132>.
- Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. <https://doi.org/10.1038/nmeth.1923>.
- Larson, Adam G., Daniel Elnatan, Madeline M. Keenen, Michael J. Trnka, Jonathan B. Johnston, Alma L. Burlingame, David A. Agard, Sy Redding, and Geeta J. Narlikar. 2017. "Liquid Droplet Formation by HP1 α Suggests a Role for Phase Separation in Heterochromatin." *Nature* 547 (7662): 236–40. <https://doi.org/10.1038/nature22822>.
- Lau, Mei Sheng, Matthew G. Schwartz, Sharmistha Kundu, Andrej J. Savol, Peggy I. Wang, Sharon K. Marr, Daniel J. Grau, et al. 2017. "Mutation of a Nucleosome Compaction Region Disrupts Polycomb-Mediated Axial Patterning." *Science* 355 (6329): 1081–84. <https://doi.org/10.1126/science.aah5403>.
- Laver, T., J. Harrison, P. A. O'Neill, K. Moore, A. Farbos, K. Paszkiewicz, and D. J. Studholme. 2015. "Assessing the Performance of the Oxford Nanopore Technologies MinION." *Biomolecular Detection and Quantification* 3 (March): 1–8. <https://doi.org/10.1016/j.bdq.2015.02.001>.
- Lawrence, Roger, Tomio Yabe, Sassan Hajmohammadi, John Rhodes, Melissa McNeely, Jian Liu, Edward D. Lamperti, et al. 2007. "The Principal Neuronal GD-Type 3-O-Sulfotransferases and Their Products in Central and Peripheral Nervous System Tissues." *Matrix Biology* 26 (6): 442–55. <https://doi.org/10.1016/j.matbio.2007.03.002>.
- Lazar, Nathan H., Kimberly A. Nevenon, Brendan O'Connell, Christine McCann, Rachel J. O'Neill, Richard E. Green, Thomas J. Meyer, Mariam Okhovat, and Lucia Carbone. 2018. "Epigenetic Maintenance of Topological Domains in the Highly Rearranged Gibbon Genome." *Genome Research* 28 (7): 983–97. <https://doi.org/10.1101/gr.233874.117>.
- Lehoczky, Jessica A., and Jeffrey W. Innis. 2008. "BAC Transgenic Analysis Reveals Enhancers Sufficient for Hoxa13 and Neighborhood Gene Expression in Mouse Embryonic Distal Limbs and Genital Bud." *Evolution & Development* 10 (4): 421–32. <https://doi.org/10.1111/j.1525-142X.2008.00253.x>.
- Lesourne, Renaud, Shoji Uehara, Jan Lee, Ki-Duk Song, LiQi Li, Julia Pinkhasov, Yongqing Zhang, et al. 2009. "Themis, a T Cell-Specific Protein Important for Late Thymocyte Development." *Nature Immunology* 10 (8): 840–47. <https://doi.org/10.1038/ni.1768>.
- Lewis, E. B. 1978. "A Gene Complex Controlling Segmentation in *Drosophila*." *Nature* 276 (5688): 565–70. <https://doi.org/10.1038/276565a0>.

- Li, Feifei, Mian Wan, Binpeng Zhang, Yiran Peng, Yachuan Zhou, Caixia Pi, Xin Xu, Ling Ye, Xuedong Zhou, and Liwei Zheng. 2018. "Bivalent Histone Modifications and Development." *Current Stem Cell Research & Therapy* 13 (2): 83–90. <https://doi.org/10.2174/1574888X12666170123144743>.
- Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79. <https://doi.org/10.1093/bioinformatics/btp352>.
- Li, Yan, Judith H. I. Haarhuis, Ángela Sedeño Cacciatore, Roel Oldenkamp, Marjon S. van Ruiten, Laureen Willems, Hans Teunissen, et al. 2020. "The Structural Basis for Cohesin–CTCF-Anchored Loops." *Nature* 578 (7795): 472–76. <https://doi.org/10.1038/s41586-019-1910-z>.
- Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." *Science (New York, N.Y.)* 326 (5950): 289–93. <https://doi.org/10.1126/science.1181369>.
- Lonfat, Nicolas, and Denis Duboule. 2015. "Structure, Function and Evolution of Topologically Associating Domains (TADs) at HOX Loci." *FEBS Letters* 589 (20 Pt A): 2869–76. <https://doi.org/10.1016/j.febslet.2015.04.024>.
- Lonfat, Nicolas, Thomas Montavon, Fabrice Darbellay, Sandra Gitto, and Denis Duboule. 2014. "Convergent Evolution of Complex Regulatory Landscapes and Pleiotropy at Hox Loci." *Science* 346 (6212): 1004–6. <https://doi.org/10.1126/science.1257493>.
- Long, Hannah K., Sara L. Prescott, and Joanna Wysocka. 2016. "Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution." *Cell* 167 (5): 1170–87. <https://doi.org/10.1016/j.cell.2016.09.018>.
- Lu, Hengyun, Francesca Giordano, and Zemin Ning. 2016. "Oxford Nanopore MinION Sequencing and Genome Assembly." *Genomics, Proteomics & Bioinformatics*, SI: Big Data and Precision Medicine, 14 (5): 265–79. <https://doi.org/10.1016/j.gpb.2016.05.004>.
- Luo, Huacheng, Fei Wang, Jie Zha, Haoli Li, Bowen Yan, Qinghua Du, Fengchun Yang, et al. 2018. "CTCF Boundary Remodels Chromatin Domain and Drives Aberrant HOX Gene Transcription in Acute Myeloid Leukemia." *Blood* 132 (8): 837–48. <https://doi.org/10.1182/blood-2017-11-814319>.
- Lupíáñez, Darío G., Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, et al. 2015. "Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions." *Cell* 161 (5): 1012–25. <https://doi.org/10.1016/j.cell.2015.04.004>.
- Ma, Jian, and Zhijun Duan. 2019. "Replication Timing Becomes Intertwined with 3D Genome Organization." *Cell* 176 (4): 681–84. <https://doi.org/10.1016/j.cell.2019.01.027>.
- Machiyama, Hiroaki, Takamitsu J. Morikawa, Kazuko Okamoto, Tomonobu M. Watanabe, and Hideaki Fujita. 2017. "The Use of a Genetically Encoded Molecular Crowding Sensor in Various Biological Phenomena." *Biophysics and Physicobiology* 14 (August): 119–25. https://doi.org/10.2142/biophysico.14.0_119.
- Maeso, Ignacio, Rafael D. Acemel, and José Luis Gómez-Skarmeta. 2017. "Cis-Regulatory Landscapes in Development and Evolution." *Current Opinion in Genetics & Development* 43 (April): 17–22. <https://doi.org/10.1016/j.gde.2016.10.004>.

- Marchal, Claire, Jiao Sima, and David M. Gilbert. 2019. "Control of DNA Replication Timing in the 3D Genome." *Nature Reviews Molecular Cell Biology* 20 (12): 721–37. <https://doi.org/10.1038/s41580-019-0162-y>.
- Margueron, Raphaël, and Danny Reinberg. 2011. "The Polycomb Complex PRC2 and Its Mark in Life." *Nature* 469 (7330): 343–49. <https://doi.org/10.1038/nature09784>.
- Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.Journal* 17 (1): 10–12. <https://doi.org/10.14806/ej.17.1.200>.
- Maurano, Matthew T., Hao Wang, Sam John, Anthony Shafer, Theresa Canfield, Kristen Lee, and John A. Stamatoyannopoulos. 2015. "Role of DNA Methylation in Modulating Transcription Factor Occupancy." *Cell Reports* 12 (7): 1184–95. <https://doi.org/10.1016/j.celrep.2015.07.024>.
- Mayer, Robert, Alessandro Brero, Johann von Hase, Timm Schroeder, Thomas Cremer, and Steffen Dietzel. 2005. "Common Themes and Cell Type Specific Variations of Higher Order Chromatin Arrangements in the Mouse." *BMC Cell Biology* 6 (1): 44. <https://doi.org/10.1186/1471-2121-6-44>.
- Mayor, Chris, Michael Brudno, Jody R. Schwartz, Alexander Poliakov, Edward M. Rubin, Kelly A. Frazer, Lior S. Pachter, and Inna Dubchak. 2000. "VISTA: Visualizing Global DNA Sequence Alignments of Arbitrary Length." *Bioinformatics* 16 (11): 1046–47. <https://doi.org/10.1093/bioinformatics/16.11.1046>.
- McGuffee, Sean R., Duncan J. Smith, and Iestyn Whitehouse. 2013. "Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination." *Molecular Cell* 50 (1): 123–35. <https://doi.org/10.1016/j.molcel.2013.03.004>.
- Michaelis, Christine, Rafal Ciosk, and Kim Nasmyth. 1997. "Cohesins: Chromosomal Proteins That Prevent Premature Separation of Sister Chromatids." *Cell* 91 (1): 35–45. [https://doi.org/10.1016/S0092-8674\(01\)80007-6](https://doi.org/10.1016/S0092-8674(01)80007-6).
- Mirny, Leonid A, Maxim Imakaev, and Nezar Abdennur. 2019. "Two Major Mechanisms of Chromosome Organization." *Current Opinion in Cell Biology, Cell Nucleus*, 58 (June): 142–52. <https://doi.org/10.1016/j.ceb.2019.05.001>.
- Moindrot, Benoit, Benjamin Audit, Petra Klous, Antoine Baker, Claude Thermes, Wouter de Laat, Philippe Bouvet, Fabien Mongelard, and Alain Arneodo. 2012. "3D Chromatin Conformation Correlates with Replication Timing and Is Conserved in Resting Cells." *Nucleic Acids Research* 40 (19): 9470–81. <https://doi.org/10.1093/nar/gks736>.
- Montavon, Thomas, and Denis Duboule. 2013. "Chromatin Organization and Global Regulation of Hox Gene Clusters." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1620): 20120367. <https://doi.org/10.1098/rstb.2012.0367>.
- Montavon, Thomas, Natalia Soshnikova, Bénédicte Mascrez, Elisabeth Joye, Laurie Thevenet, Erik Splinter, Wouter de Laat, François Spitz, and Denis Duboule. 2011. "A Regulatory Archipelago Controls Hox Genes Transcription in Digits." *Cell* 147 (5): 1132–45. <https://doi.org/10.1016/j.cell.2011.10.023>.
- Nagano, Takashi, Yaniv Lubling, Tim J. Stevens, Stefan Schoenfelder, Eitan Yaffe, Wendy Dean, Ernest D. Laue, Amos Tanay, and Peter Fraser. 2013. "Single-Cell Hi-C Reveals Cell-to-Cell Variability in Chromosome Structure." *Nature* 502 (7469): 59–64. <https://doi.org/10.1038/nature12593>.
- Nagano, Takashi, Yaniv Lubling, Csilla Várnai, Carmel Dudley, Wing Leung, Yael Baran, Netta Mendelson Cohen, Steven Wingett, Peter Fraser, and Amos Tanay. 2017. "Cell-Cycle Dynamics of Chromosomal Organization at Single-Cell Resolution." *Nature* 547 (7661): 61–67. <https://doi.org/10.1038/nature23001>.

- Narendra, Varun, Pedro P. Rocha, Disi An, Ramya Raviram, Jane A. Skok, Esteban O. Mazzoni, and Danny Reinberg. 2015. "CTCF Establishes Discrete Functional Chromatin Domains at the Hox Clusters during Differentiation." *Science* 347 (6225): 1017–21. <https://doi.org/10.1126/science.1262088>.
- Nasmyth, K. 2001. "Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids during Mitosis and Meiosis." *Annual Review of Genetics* 35: 673–745. <https://doi.org/10.1146/annurev.genet.35.102401.091334>.
- Ni, Yanxiang, Bo Cao, Tszshan Ma, Gang Niu, Yingdong Huo, Jiandong Huang, Danni Chen, et al. 2017. "Super-Resolution Imaging of a 2.5 Kb Non-Repetitive DNA in Situ in the Nuclear Genome Using Molecular Beacon Probes." Edited by Lothar Schermelleh. *ELife* 6 (May): e21660. <https://doi.org/10.7554/eLife.21660>.
- Nicholls, Peter K., Daniel W. Bellott, Ting-Jan Cho, Tatyana Pyntikova, and David C. Page. 2019. "Locating and Characterizing a Transgene Integration Site by Nanopore Sequencing." *G3: Genes, Genomes, Genetics* 9 (5): 1481–86. <https://doi.org/10.1534/g3.119.300582>.
- Nicodemi, Mario, and Antonella Prisco. 2009. "Thermodynamic Pathways to Genome Spatial Organization in the Cell Nucleus." *Biophysical Journal* 96 (6): 2168–77. <https://doi.org/10.1016/j.bpj.2008.12.3919>.
- Nielsen, Peter R., Daniel Nietlispach, Helen R. Mott, Juliana Callaghan, Andrew Bannister, Tony Kouzarides, Alexey G. Murzin, Natalia V. Murzina, and Ernest D. Laue. 2002. "Structure of the HP1 Chromodomain Bound to Histone H3 Methylated at Lysine 9." *Nature* 416 (6876): 103–7. <https://doi.org/10.1038/nature722>.
- Nishana, Mayilaadumveetil, and Sathees C. Raghavan. 2012. "Role of Recombination Activating Genes in the Generation of Antigen Receptor Diversity and Beyond." *Immunology* 137 (4): 271–81. <https://doi.org/10.1111/imm.12009>.
- Nishibuchi, Gohei, and Jérôme Déjardin. 2017. "The Molecular Basis of the Organization of Repetitive DNA-Containing Constitutive Heterochromatin in Mammals." *Chromosome Research* 25 (1): 77–87. <https://doi.org/10.1007/s10577-016-9547-3>.
- Nishibuchi, Gohei, Shinichi Machida, Akihisa Osakabe, Hiromu Murakoshi, Kyoko Hiragami-Hamada, Reiko Nakagawa, Wolfgang Fischle, et al. 2014. "N-Terminal Phosphorylation of HP1 α Increases Its Nucleosome-Binding Specificity." *Nucleic Acids Research* 42 (20): 12498–511. <https://doi.org/10.1093/nar/gku995>.
- Noordermeer, Daan, Marion Leleu, Erik Splinter, Jacques Rougemont, Wouter de Laat, and Denis Duboule. 2011. "The Dynamic Architecture of Hox Gene Clusters." *Science* 334 (6053): 222–25. <https://doi.org/10.1126/science.1207194>.
- Nora, Elphège P., Anton Goloborodko, Anne-Laure Valton, Johan H. Gibcus, Alec Uebersohn, Nezar Abdennur, Job Dekker, Leonid A. Mirny, and Benoit G. Bruneau. 2017. "Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization." *Cell* 169 (5): 930–944.e22. <https://doi.org/10.1016/j.cell.2017.05.004>.
- Nora, Elphège P., Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, et al. 2012. "Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre." *Nature* 485 (7398): 381–85. <https://doi.org/10.1038/nature11049>.
- Northcott, Paul A., Ivo Buchhalter, A. Sorana Morrissy, Volker Hovestadt, Joachim Weischenfeldt, Tobias Ehrenberger, Susanne Gröbner, et al. 2017. "The Whole-Genome Landscape of Medulloblastoma Subtypes." *Nature* 547 (7663): 311–17. <https://doi.org/10.1038/nature22973>.

- Nuebler, Johannes, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid A. Mirny. 2018. "Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation." *Proceedings of the National Academy of Sciences* 115 (29): E6697–6706. <https://doi.org/10.1073/pnas.1717730115>.
- Ohno, Susumu. 1970. *Evolution by Gene Duplication*. Berlin Heidelberg: Springer-Verlag. <https://doi.org/10.1007/978-3-642-86659-3>.
- Oldach, Phoebe, and Conrad A. Nieduszynski. 2019. "Cohesin-Mediated Genome Architecture Does Not Define DNA Replication Timing Domains." *Genes* 10 (3): 196. <https://doi.org/10.3390/genes10030196>.
- Paliou, Christina, Philine Guckelberger, Robert Schöpflin, Verena Heinrich, Andrea Esposito, Andrea M. Chiariello, Simona Bianco, et al. 2019. "Preformed Chromatin Topology Assists Transcriptional Robustness of Shh during Limb Development." *Proceedings of the National Academy of Sciences* 116 (25): 12390–99. <https://doi.org/10.1073/pnas.1900672116>.
- Parelho, Vania, Suzana Hadjur, Mikhail Spivakov, Marion Leleu, Stephan Sauer, Heather C. Gregson, Adam Jarmuz, et al. 2008. "Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms." *Cell* 132 (3): 422–33. <https://doi.org/10.1016/j.cell.2008.01.011>.
- Patrick, Michael S., Hiroyo Oda, Kunihiro Hayakawa, Yoshinori Sato, Koji Eshima, Teruo Kirikae, Shun-ichiro Iemura, et al. 2009. "Gasp, a Grb2-Associating Protein, Is Critical for Positive Selection of Thymocytes." *Proceedings of the National Academy of Sciences* 106 (38): 16345–50. <https://doi.org/10.1073/pnas.0908593106>.
- Pearson, Joseph C., Derek Lemons, and William McGinnis. 2005. "Modulating Hox Gene Functions during Animal Body Patterning." *Nature Reviews Genetics* 6 (12): 893–904. <https://doi.org/10.1038/nrg1726>.
- Peng, Yanling, and Yubo Zhang. 2018. "Enhancer and Super-Enhancer: Positive Regulators in Gene Transcription." *Animal Models and Experimental Medicine* 1 (3): 169–79. <https://doi.org/10.1002/ame2.12032>.
- Peters, Jan-Michael, and Tomoko Nishiyama. 2012. "Sister Chromatid Cohesion." *Cold Spring Harbor Perspectives in Biology* 4 (11): a011130. <https://doi.org/10.1101/cshperspect.a011130>.
- Phillips-Cremins, Jennifer E., Michael E. G. Sauria, Amartya Sanyal, Tatiana I. Gerasimova, Bryan R. Lajoie, Joshua S. K. Bell, Chin-Tong Ong, et al. 2013. "Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment." *Cell* 153 (6): 1281–95. <https://doi.org/10.1016/j.cell.2013.04.053>.
- Plys, Aaron J., Christopher P. Davis, Jongmin Kim, Gizem Rizki, Madeline M. Keenen, Sharon K. Marr, and Robert E. Kingston. 2019. "Phase Separation of Polycomb-Repressive Complex 1 Is Governed by a Charged Disordered Region of CBX2." *Genes & Development* 33 (13–14): 799–813. <https://doi.org/10.1101/gad.326488.119>.
- Pombo, Ana, and Niall Dillon. 2015. "Three-Dimensional Genome Architecture: Players and Mechanisms." *Nature Reviews Molecular Cell Biology* 16 (4): 245–57. <https://doi.org/10.1038/nrm3965>.
- Pope, Benjamin D., Tyrone Ryba, Vishnu Dileep, Feng Yue, Weisheng Wu, Olgert Denas, Daniel L. Vera, et al. 2014. "Topologically Associating Domains Are Stable Units of Replication-Timing Regulation." *Nature* 515 (7527): 402–5. <https://doi.org/10.1038/nature13986>.

- Quinlan, Aaron R. 2014. "BEDTools: The Swiss-Army Tool for Genome Feature Analysis." *Current Protocols in Bioinformatics* 47 (1): 11.12.1-11.12.34. <https://doi.org/10.1002/0471250953.bi1112s47>.
- Raghuraman, M. K., Elizabeth A. Winzeler, David Collingwood, Sonia Hunt, Lisa Wodicka, Andrew Conway, David J. Lockhart, Ronald W. Davis, Bonita J. Brewer, and Walton L. Fangman. 2001. "Replication Dynamics of the Yeast Genome." *Science* 294 (5540): 115–21. <https://doi.org/10.1126/science.294.5540.115>.
- Ramírez, Fidel, Vivek Bhardwaj, Laura Arrigoni, Kin Chung Lam, Björn A. Grüning, José Villaveces, Bianca Habermann, Asifa Akhtar, and Thomas Manke. 2018. "High-Resolution TADs Reveal DNA Sequences Underlying Genome Organization in Flies." *Nature Communications* 9 (1): 1–15. <https://doi.org/10.1038/s41467-017-02525-w>.
- Rao, Suhas S. P., Su-Chen Huang, Brian Glenn St Hilaire, Jesse M. Engreitz, Elizabeth M. Perez, Kyong-Rim Kieffer-Kwon, Adrian L. Sanborn, et al. 2017. "Cohesin Loss Eliminates All Loop Domains." *Cell* 171 (2): 305-320.e24. <https://doi.org/10.1016/j.cell.2017.09.026>.
- Rao, Suhas S.P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian Sanborn, et al. 2014. "A Three-Dimensional Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping." *Cell* 159 (7): 1665–80. <https://doi.org/10.1016/j.cell.2014.11.021>.
- Redolfi, Josef, Yinxiu Zhan, Christian Valdes-Quezada, Mariya Kryzhanovska, Isabel Guerreiro, Vytautas Iesmantavicius, Tim Pollex, et al. 2019. "DamC Reveals Principles of Chromatin Folding in Vivo without Crosslinking and Ligation." *Nature Structural & Molecular Biology* 26 (6): 471–80. <https://doi.org/10.1038/s41594-019-0231-0>.
- Rincón-Arano, Héctor, Georgina Guerrero, Christian Valdes-Quezada, and Félix Recillas-Targa. 2009. "Chicken α -Globin Switching Depends on Autonomous Silencing of the Embryonic π Globin Gene by Epigenetics Mechanisms." *Journal of Cellular Biochemistry* 108 (3): 675–87. <https://doi.org/10.1002/jcb.22304>.
- Rodríguez-Carballo, Eddie, Lucille Lopez-Delisle, Nayuta Yakushiji-Kaminatsui, Asier Ullate-Agote, and Denis Duboule. 2019. "Impact of Genome Architecture on the Functional Activation and Repression of Hox Regulatory Landscapes." *BMC Biology* 17 (1): 55. <https://doi.org/10.1186/s12915-019-0677-x>.
- Rodríguez-Carballo, Eddie, Lucille Lopez-Delisle, Ye Zhan, Pierre J. Fabre, Leonardo Beccari, Imane El-Idrissi, Thi Hanh Nguyen Huynh, Hakan Ozadam, Job Dekker, and Denis Duboule. 2017. "The HoxD Cluster Is a Dynamic and Resilient TAD Boundary Controlling the Segregation of Antagonistic Regulatory Landscapes." *Genes & Development* 31 (22): 2264–81. <https://doi.org/10.1101/gad.307769.117>.
- Rodríguez-Carballo, Edgardo, Lucille Lopez-Delisle, Andréa Willemin, Leonardo Beccari, Sandra Gitto, Benedicte Mascres, and Denis Duboule. 2020. "Chromatin Topology And The Timing Of Enhancer Function At The HoxD Locus." *BioRxiv*, July, 2020.07.12.199109. <https://doi.org/10.1101/2020.07.12.199109>.
- Saldaña-Meyer, Ricardo, Edgar González-Buendía, Georgina Guerrero, Varun Narendra, Roberto Bonasio, Félix Recillas-Targa, and Danny Reinberg. 2014. "CTCF Regulates the Human P53 Gene through Direct Interaction with Its Natural Antisense Transcript, Wrap53." *Genes & Development* 28 (7): 723–34. <https://doi.org/10.1101/gad.236869.113>.
- Saldaña-Meyer, Ricardo, Javier Rodríguez-Hernaez, Thelma Escobar, Mayilaadumveetil Nishana, Karina Jácome-López, Elphege P. Nora, Benoit G. Bruneau, et al. 2019. "RNA Interactions Are Essential for CTCF-Mediated Genome Organization." *Molecular Cell* 76 (3): 412-422.e5. <https://doi.org/10.1016/j.molcel.2019.08.015>.

- Sanborn, Adrian L., Suhas S. P. Rao, Su-Chen Huang, Neva C. Durand, Miriam H. Huntley, Andrew I. Jewett, Ivan D. Bochkov, et al. 2015. "Chromatin Extrusion Explains Key Features of Loop and Domain Formation in Wild-Type and Engineered Genomes." *Proceedings of the National Academy of Sciences of the United States of America* 112 (47): E6456-6465. <https://doi.org/10.1073/pnas.1518552112>.
- Sati, Satish, and Giacomo Cavalli. 2017. "Chromosome Conformation Capture Technologies and Their Impact in Understanding Genome Function." *Chromosoma* 126 (1): 33-44. <https://doi.org/10.1007/s00412-016-0593-6>.
- Schep, Ruben, Anamaria Necsulea, Eddie Rodríguez-Carballo, Isabel Guerreiro, Guillaume Andrey, Thi Hanh Nguyen Huynh, Virginie Marcet, Jozsef Zákány, Denis Duboule, and Leonardo Beccari. 2016. "Control of Hoxd Gene Transcription in the Mammary Bud by Hijacking a Preexisting Regulatory Landscape." *Proceedings of the National Academy of Sciences* 113 (48): E7720-29. <https://doi.org/10.1073/pnas.1617141113>.
- Schmidl, Christian, André F. Rendeiro, Nathan C. Sheffield, and Christoph Bock. 2015. "ChIPmentation: Fast, Robust, Low-Input ChIP-Seq for Histones and Transcription Factors." *Nature Methods* 12 (10): 963-65. <https://doi.org/10.1038/nmeth.3542>.
- Schoenfelder, Stefan, Robert Sugar, Andrew Dimond, Biola-Maria Javierre, Harry Armstrong, Borbala Mifsud, Emilia Dimitrova, et al. 2015. "Polycomb Repressive Complex PRC1 Spatially Constrains the Mouse Embryonic Stem Cell Genome." *Nature Genetics* 47 (10): 1179-86. <https://doi.org/10.1038/ng.3393>.
- Schorderet, Patrick, Nicolas Lonfat, Fabrice Darbellay, Patrick Tschopp, Sandra Gitto, Natalia Soshnikova, and Denis Duboule. 2013. "A Genetic Approach to the Recruitment of PRC2 at the HoxD Locus." *PLOS Genetics* 9 (11): e1003951. <https://doi.org/10.1371/journal.pgen.1003951>.
- Schuler, F., and L. M. Sorokin. 1995. "Expression of Laminin Isoforms in Mouse Myogenic Cells in Vitro and in Vivo." *Journal of Cell Science* 108 (12): 3795-3805.
- Schuurman, Henk-Jan, and VFJ Quesniaux. 1999. "Development and Maturation of T and B Cells." In , 23-40.
- Schwarzer, Wibke, Nezar Abdennur, Anton Goloborodko, Aleksandra Pekowska, Geoffrey Fudenberg, Yann Loe-Mie, Nuno A. Fonseca, et al. 2017. "Two Independent Modes of Chromatin Organization Revealed by Cohesin Removal." *Nature* 551 (7678): 51-56. <https://doi.org/10.1038/nature24281>.
- Shen, P., P. D. Canoll, J. Sap, and J. M. Musacchio. 1999. "Expression of a Truncated Receptor Protein Tyrosine Phosphatase Kappa in the Brain of an Adult Transgenic Mouse." *Brain Research* 826 (2): 157-71. [https://doi.org/10.1016/S0006-8993\(99\)01179-8](https://doi.org/10.1016/S0006-8993(99)01179-8).
- Shen, Yin, Feng Yue, David F. McCleary, Zhen Ye, Lee Edsall, Samantha Kuan, Ulrich Wagner, et al. 2012. "A Map of the Cis-Regulatory Sequences in the Mouse Genome." *Nature* 488 (7409): 116-20. <https://doi.org/10.1038/nature11243>.
- Shin, Hanjun, Yi Shi, Chao Dai, Harianto Tjong, Ke Gong, Frank Alber, and Xianghong Jasmine Zhou. 2016. "TopDom: An Efficient and Deterministic Method for Identifying Topological Domains in Genomes." *Nucleic Acids Research* 44 (7): e70-e70. <https://doi.org/10.1093/nar/gkv1505>.
- Shrinivas, Krishna, Benjamin R. Sabari, Eliot L. Coffey, Isaac A. Klein, Ann Boija, Alicia V. Zamudio, Jurian Schuijers, et al. 2019. "Enhancer Features That Drive Formation of Transcriptional Condensates." *Molecular Cell* 75 (3): 549-561.e7. <https://doi.org/10.1016/j.molcel.2019.07.009>.

- Shworak, Nicholas W., Jian Liu, Lorin M. Petros, Lijuan Zhang, Masashi Kobayashi, Neal G. Copeland, Nancy A. Jenkins, and Robert D. Rosenberg. 1999. "Multiple Isoforms of Heparan Sulfate D-Glucosaminyl 3-O-Sulfotransferase ISOLATION, CHARACTERIZATION, AND EXPRESSION OF HUMAN CDNAs AND IDENTIFICATION OF DISTINCT GENOMIC LOCI." *Journal of Biological Chemistry* 274 (8): 5170–84. <https://doi.org/10.1074/jbc.274.8.5170>.
- Sikorska, Natalia, and Tom Sexton. 2020. "Defining Functionally Relevant Spatial Chromatin Domains: It Is a TAD Complicated." *Journal of Molecular Biology, Perspectives on Chromosome Folding*, 432 (3): 653–64. <https://doi.org/10.1016/j.jmb.2019.12.006>.
- Sima, Jiao, Abhijit Chakraborty, Vishnu Dileep, Marco Michalski, Kyle N. Klein, Nicolas P. Holcomb, Jesse L. Turner, et al. 2019. "Identifying Cis Elements for Spatiotemporal Control of Mammalian DNA Replication." *Cell* 176 (4): 816–830.e18. <https://doi.org/10.1016/j.cell.2018.11.036>.
- Singh, Manvendra K., Marianne Petry, Bénédicte Haenig, Birgit Lescher, Michael Leitges, and Andreas Kispert. 2005. "The T-Box Transcription Factor Tbx15 Is Required for Skeletal Development." *Mechanisms of Development* 122 (2): 131–44. <https://doi.org/10.1016/j.mod.2004.10.011>.
- Sivakumar, Aishwarya, Jose I. de las Heras, and Eric C. Schirmer. 2019. "Spatial Genome Organization: From Development to Disease." *Frontiers in Cell and Developmental Biology* 7. <https://doi.org/10.3389/fcell.2019.00018>.
- Sleutels, Frank, Widia Soochit, Marek Bartkuhn, Helen Heath, Sven Dienstbach, Philipp Bergmaier, Vedran Franke, et al. 2012. "The Male Germ Cell Gene Regulator CTCFL Is Functionally Different from CTCF and Binds CTCF-like Consensus Sites in a Nucleosome Composition-Dependent Manner." *Epigenetics & Chromatin* 5 (1): 8. <https://doi.org/10.1186/1756-8935-5-8>.
- Smale, Stephen T., and James T. Kadonaga. 2003. "The RNA Polymerase II Core Promoter." *Annual Review of Biochemistry* 72 (1): 449–79. <https://doi.org/10.1146/annurev.biochem.72.121801.161520>.
- Smith, Emily M., Bryan R. Lajoie, Gaurav Jain, and Job Dekker. 2016. "Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus." *The American Journal of Human Genetics* 98 (1): 185–201. <https://doi.org/10.1016/j.ajhg.2015.12.002>.
- Spielmann, Malte, Darío G. Lupiáñez, and Stefan Mundlos. 2018. "Structural Variation in the 3D Genome." *Nature Reviews Genetics* 19 (7): 453–67. <https://doi.org/10.1038/s41576-018-0007-0>.
- Spitz, F., F. Gonzalez, C. Peichel, T. F. Vogt, D. Duboule, and J. Zákány. 2001. "Large Scale Transgenic and Cluster Deletion Analysis of the HoxD Complex Separate an Ancestral Regulatory Module from Evolutionary Innovations." *Genes & Development* 15 (17): 2209–14. <https://doi.org/10.1101/gad.205701>.
- Spitz, François, Federico Gonzalez, and Denis Duboule. 2003. "A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the HoxD Cluster." *Cell* 113 (3): 405–17.
- Splinter, Erik, Elzo de Wit, Elphège P. Nora, Petra Klous, Harmen J. G. van de Werken, Yun Zhu, Lucas J. T. Kaaij, et al. 2011. "The Inactive X Chromosome Adopts a Unique Three-Dimensional Conformation That Is Dependent on Xist RNA." *Genes & Development* 25 (13): 1371–83. <https://doi.org/10.1101/gad.633311>.
- Stevens, Tim J., David Lando, Srinjan Basu, Liam P. Atkinson, Yang Cao, Steven F. Lee, Martin Leeb, et al. 2017. "3D Structures of Individual Mammalian Genomes Studied by Single-Cell Hi-C." *Nature* 544 (7648): 59–64. <https://doi.org/10.1038/nature21429>.

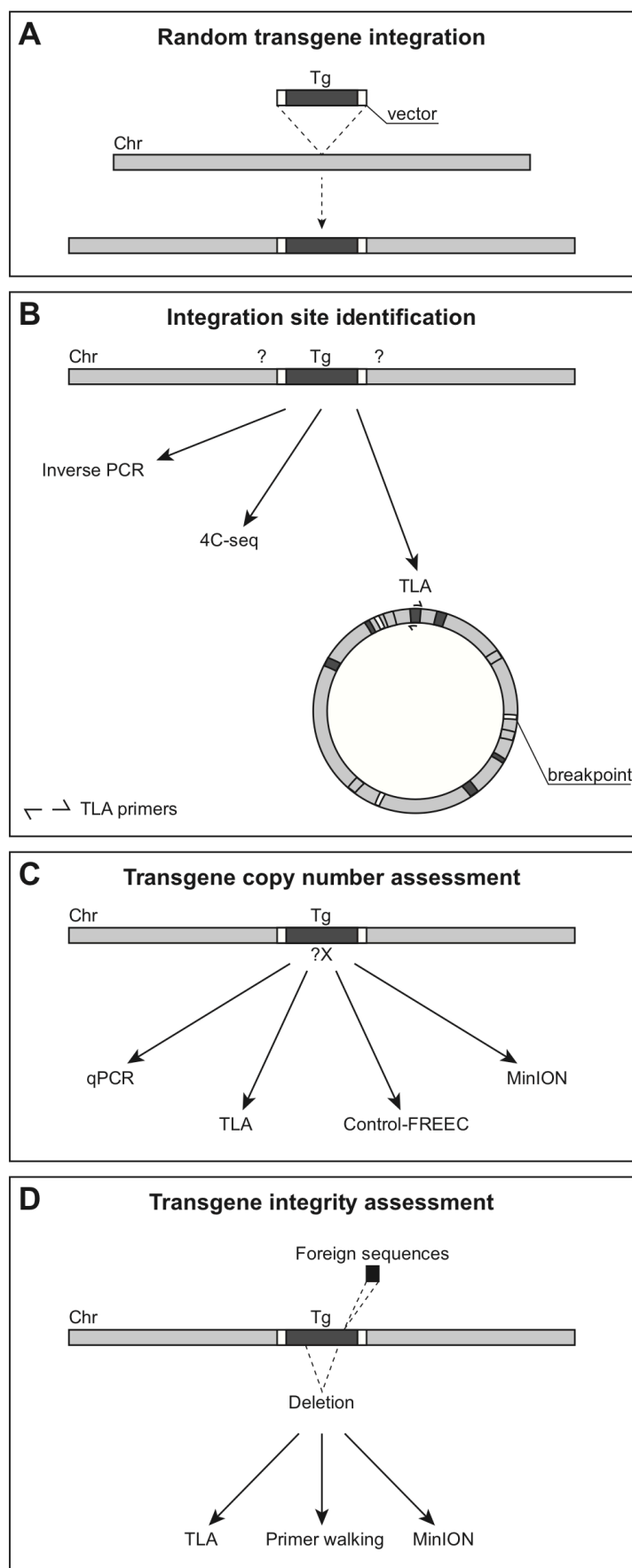
- Strom, Amy R., Alexander V. Emelyanov, Mustafa Mir, Dmitry V. Fyodorov, Xavier Darzacq, and Gary H. Karpen. 2017. "Phase Separation Drives Heterochromatin Domain Formation." *Nature* 547 (7662): 241–45. <https://doi.org/10.1038/nature22989>.
- Sun, James H., Linda Zhou, Daniel J. Emerson, Sai A. Phyo, Katelyn R. Titus, Wanfeng Gong, Thomas G. Gilgenast, et al. 2018. "Disease-Associated Short Tandem Repeats Co-Localize with Chromatin Domain Boundaries." *Cell* 175 (1): 224–238.e15. <https://doi.org/10.1016/j.cell.2018.08.005>.
- Tang, Shih-Huey E., Francisco J. Silva, Walter M. K. Tsark, and Jeffrey R. Mann. 2002. "A Cre/LoxP-Deleter Transgenic Line in Mouse Strain 129S1/SvImJ." *Genesis* 32 (3): 199–202. <https://doi.org/10.1002/gene.10030>.
- Tarchini, Basile, and Denis Duboule. 2006. "Control of Hoxd Genes' Collinearity during Early Limb Development." *Developmental Cell* 10 (1): 93–103. <https://doi.org/10.1016/j.devcel.2005.11.014>.
- Tatavosian, Roubina, Samantha Kent, Kyle Brown, Tingting Yao, Huy Nguyen Duc, Thao Ngoc Huynh, Chao Yu Zhen, Brian Ma, Haobin Wang, and Xiaojun Ren. 2019. "Nuclear Condensates of the Polycomb Protein Chromobox 2 (CBX2) Assemble through Phase Separation." *Journal of Biological Chemistry* 294 (5): 1451–63. <https://doi.org/10.1074/jbc.RA118.006620>.
- Tschopp, Patrick, and Denis Duboule. 2011. "A Regulatory 'Landscape Effect' over the HoxD Cluster." *Developmental Biology* 351 (2): 288–96. <https://doi.org/10.1016/j.ydbio.2010.12.034>.
- Tschopp, Patrick, Basile Tarchini, François Spitz, Jozsef Zakany, and Denis Duboule. 2009. "Uncoupling Time and Space in the Collinear Regulation of Hox Genes." *PLOS Genetics* 5 (3): e1000398. <https://doi.org/10.1371/journal.pgen.1000398>.
- Tsompana, Maria, and Michael J. Buck. 2014. "Chromatin Accessibility: A Window into the Genome." *Epigenetics & Chromatin* 7 (1): 33. <https://doi.org/10.1186/1756-8935-7-33>.
- Uhlmann, Frank, Friedrich Lottspeich, and Kim Nasmyth. 1999. "Sister-Chromatid Separation at Anaphase Onset Is Promoted by Cleavage of the Cohesin Subunit Scc1." *Nature* 400 (6739): 37–42. <https://doi.org/10.1038/21831>.
- Untergasser, Andreas, Harm Nijveen, Xiangyu Rao, Ton Bisseling, René Geurts, and Jack A. M. Leunissen. 2007. "Primer3Plus, an Enhanced Web Interface to Primer3." *Nucleic Acids Research* 35 (suppl_2): W71–74. <https://doi.org/10.1093/nar/gkm306>.
- Velez, Gabriel, Marisa Lin, Trace Christensen, William A. Faubion, Gwen Lomberg, and Raul Urrutia. 2015. "Evidence Supporting a Critical Contribution of Intrinsically Disordered Regions to the Biochemical Behavior of Full-Length Human HP1γ." *Journal of Molecular Modeling* 22 (1): 12. <https://doi.org/10.1007/s00894-015-2874-z>.
- Vian, Laura, Aleksandra Pękowska, Suhas S. P. Rao, Kyong-Rim Kieffer-Kwon, Seolkyoung Jung, Laura Baranello, Su-Chen Huang, et al. 2018. "The Energetics and Physiological Impact of Cohesin Extrusion." *Cell* 173 (5): 1165–1178.e20. <https://doi.org/10.1016/j.cell.2018.03.072>.
- Vietri Rudan, Matteo, Christopher Barrington, Stephen Henderson, Christina Ernst, Duncan T. Odom, Amos Tanay, and Suzana Hadjur. 2015. "Comparative Hi-C Reveals That CTCF Underlies Evolution of Chromosomal Domain Architecture." *Cell Reports* 10 (8): 1297–1309. <https://doi.org/10.1016/j.celrep.2015.02.004>.

- Visel, Axel, James Bristow, and Len A. Pennacchio. 2007. "Enhancer Identification through Comparative Genomics." *Seminars in Cell & Developmental Biology*, Model Systems for the Study of Cardiovascular Development and Disease, 18 (1): 140–52. <https://doi.org/10.1016/j.semcd.2006.12.014>.
- Visel, Axel, Simon Minovitsky, Inna Dubchak, and Len A. Pennacchio. 2007. "VISTA Enhancer Browser—a Database of Tissue-Specific Human Enhancers." *Nucleic Acids Research* 35 (suppl_1): D88–92. <https://doi.org/10.1093/nar/gkl822>.
- Vree, Paula J. P. de, Elzo de Wit, Mehmet Yilmaz, Monique van de Heijning, Petra Klous, Marjon J. A. M. Verstegen, Yi Wan, et al. 2014. "Targeted Sequencing by Proximity Ligation for Comprehensive Variant Detection and Local Haplotyping." *Nature Biotechnology* 32 (10): 1019–25. <https://doi.org/10.1038/nbt.2959>.
- Wang, Hao, Matthew T. Maurano, Hongzhu Qu, Katherine E. Varley, Jason Gertz, Florencia Pauli, Kristen Lee, et al. 2012. "Widespread Plasticity in CTCF Occupancy Linked to DNA Methylation." *Genome Research* 22 (9): 1680–88. <https://doi.org/10.1101/gr.136101.111>.
- Wang, Jiyong, Stephanie T. Lawry, Allison L. Cohen, and Songtao Jia. 2014. "Chromosome Boundary Elements and Regulation of Heterochromatin Spreading." *Cellular and Molecular Life Sciences* 71 (24): 4841–52. <https://doi.org/10.1007/s00018-014-1725-x>.
- Wellik, Deneen M. 2009. "Chapter 9 Hox Genes and Vertebrate Axial Pattern." In *Current Topics in Developmental Biology*, 88:257–78. Genes. Academic Press. [https://doi.org/10.1016/S0070-2153\(09\)88009-5](https://doi.org/10.1016/S0070-2153(09)88009-5).
- Wendt, Kerstin S., Keisuke Yoshida, Takehiko Itoh, Masashige Bando, Birgit Koch, Erika Schirghuber, Shuichi Tsutsumi, et al. 2008. "Cohesin Mediates Transcriptional Insulation by CCCTC-Binding Factor." *Nature* 451 (7180): 796–801. <https://doi.org/10.1038/nature06634>.
- Werken, Harmen J. G. van de, Gilad Landan, Sjoerd J. B. Holwerda, Michael Hoichman, Petra Klous, Ran Chachik, Erik Splinter, et al. 2012. "Robust 4C-Seq Data Analysis to Screen for Regulatory DNA Interactions." *Nature Methods* 9 (10): 969–72. <https://doi.org/10.1038/nmeth.2173>.
- Whyte, Warren A., David A. Orlando, Denes Hnisz, Brian J. Abraham, Charles Y. Lin, Michael H. Kagey, Peter B. Rahl, Tong Ihn Lee, and Richard A. Young. 2013. "Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes." *Cell* 153 (2): 307–19. <https://doi.org/10.1016/j.cell.2013.03.035>.
- Williamson, Iain, Lauren Kane, Paul S. Devenney, Ilya M. Flyamer, Eve Anderson, Fiona Kilanowski, Robert E. Hill, Wendy A. Bickmore, and Laura A. Lettice. 2019. "Developmentally Regulated Shh Expression Is Robust to TAD Perturbations." *Development* 146 (19). <https://doi.org/10.1242/dev.179523>.
- Wingett, Steven W., Philip Ewels, Mayra Furlan-Magaril, Takashi Nagano, Stefan Schoenfelder, Peter Fraser, and Simon Andrews. 2015. "HiCUP: Pipeline for Mapping and Processing Hi-C Data." *F1000Research* 4 (November): 1310. <https://doi.org/10.12688/f1000research.7334.1>.
- de Wit, Elzo, and Wouter de Laat. 2012. "A Decade of 3C Technologies: Insights into Nuclear Organization." *Genes & Development* 26 (1): 11–24. <https://doi.org/10.1101/gad.179804.111>.
- Woltering, Joost M., Daan Noordermeer, Marion Leleu, and Denis Duboule. 2014. "Conservation and Divergence of Regulatory Strategies at Hox Loci and the Origin of Tetrapod Digits." *PLOS Biology* 12 (1): e1001773. <https://doi.org/10.1371/journal.pbio.1001773>.

- Woltering, Joost M., Freek J. Vonk, Hendrik Müller, Nabila Bardine, Ioana L. Tuduce, Merijn A. G. de Bakker, Walter Knöchel, I. Ovidiu Sirbu, Antony J. Durston, and Michael K. Richardson. 2009. "Axial Patterning in Snakes and Caecilians: Evidence for an Alternative Interpretation of the Hox Code." *Developmental Biology*, Special Section: Evolution of Developmental Regulatory Systems, 332 (1): 82–89. <https://doi.org/10.1016/j.ydbio.2009.04.031>.
- Woodcock, Christopher L., and Rajarshi P. Ghosh. 2010. "Chromatin Higher-Order Structure and Dynamics." *Cold Spring Harbor Perspectives in Biology* 2 (5): a000596. <https://doi.org/10.1101/cshperspect.a000596>.
- Yaffe, Eitan, Shlomit Farkash-Amar, Andreas Polten, Zohar Yakhini, Amos Tanay, and Itamar Simon. 2010. "Comparative Analysis of DNA Replication Timing Reveals Conserved Large-Scale Chromosomal Architecture." *PLOS Genetics* 6 (7): e1001011. <https://doi.org/10.1371/journal.pgen.1001011>.
- Yakushiji-Kaminatsui, Nayuta, Lucille Lopez-Delisle, Christopher Chase Bolt, Guillaume Andrey, Leonardo Beccari, and Denis Duboule. 2018. "Similarities and Differences in the Regulation of HoxD Genes during Chick and Mouse Limb Development." *PLoS Biology* 16 (11). <https://doi.org/10.1371/journal.pbio.3000004>.
- Yu, Man, Hemei Han, Ping Xiang, Qiliang Li, and George Stamatoyannopoulos. 2006. "Autonomous Silencing as Well as Competition Controls γ -Globin Gene Expression during Development." *Molecular and Cellular Biology* 26 (13): 4775–81. <https://doi.org/10.1128/MCB.00406-06>.
- Yuen, Kobe C., and Jennifer L. Gerton. 2018. "Taking Cohesin and Condensin in Context." *PLOS Genetics* 14 (1): e1007118. <https://doi.org/10.1371/journal.pgen.1007118>.
- Zákány, József, and Denis Duboule. 1996. "Synpolydactyly in Mice with a Targeted Deficiency in the HoxD Complex." *Nature* 384 (6604): 69–71. <https://doi.org/10.1038/384069a0>.
- Zákány, József, Marie Kmita, Pilar Alarcon, José-Luis de la Pompa, and Denis Duboule. 2001. "Localized and Transient Transcription of Hox Genes Suggests a Link between Patterning and the Segmentation Clock." *Cell* 106 (2): 207–17. [https://doi.org/10.1016/S0092-8674\(01\)00436-6](https://doi.org/10.1016/S0092-8674(01)00436-6).
- Zákány, József, Marie Kmita, and Denis Duboule. 2004. "A Dual Role for Hox Genes in Limb Anterior-Posterior Asymmetry." *Science* 304 (5677): 1669–72. <https://doi.org/10.1126/science.1096049>.
- Zarrin, Ali A., Catherine Del Vecchio, Eva Tseng, Megan Gleason, Payam Zarin, Ming Tian, and Frederick W. Alt. 2007. "Antibody Class Switching Mediated by Yeast Endonuclease-Generated DNA Breaks." *Science* 315 (5810): 377–81. <https://doi.org/10.1126/science.1136386>.
- Zhan, Yinxu, Luca Mariani, Iros Barozzi, Edda G. Schulz, Nils Blüthgen, Michael Stadler, Guido Tiana, and Luca Giorgetti. 2017. "Reciprocal Insulation Analysis of Hi-C Data Shows That TADs Represent a Functionally but Not Structurally Privileged Scale in the Hierarchical Folding of Chromosomes." *Genome Research* 27 (3): 479–90. <https://doi.org/10.1101/gr.212803.116>.
- Zhang, Haoyue, Daniel J. Emerson, Thomas G. Gilgenast, Katelyn R. Titus, Yemin Lan, Peng Huang, Di Zhang, et al. 2019. "Chromatin Structure Dynamics during the Mitosis-to-G1 Phase Transition." *Nature* 576 (7785): 158–62. <https://doi.org/10.1038/s41586-019-1778-y>.

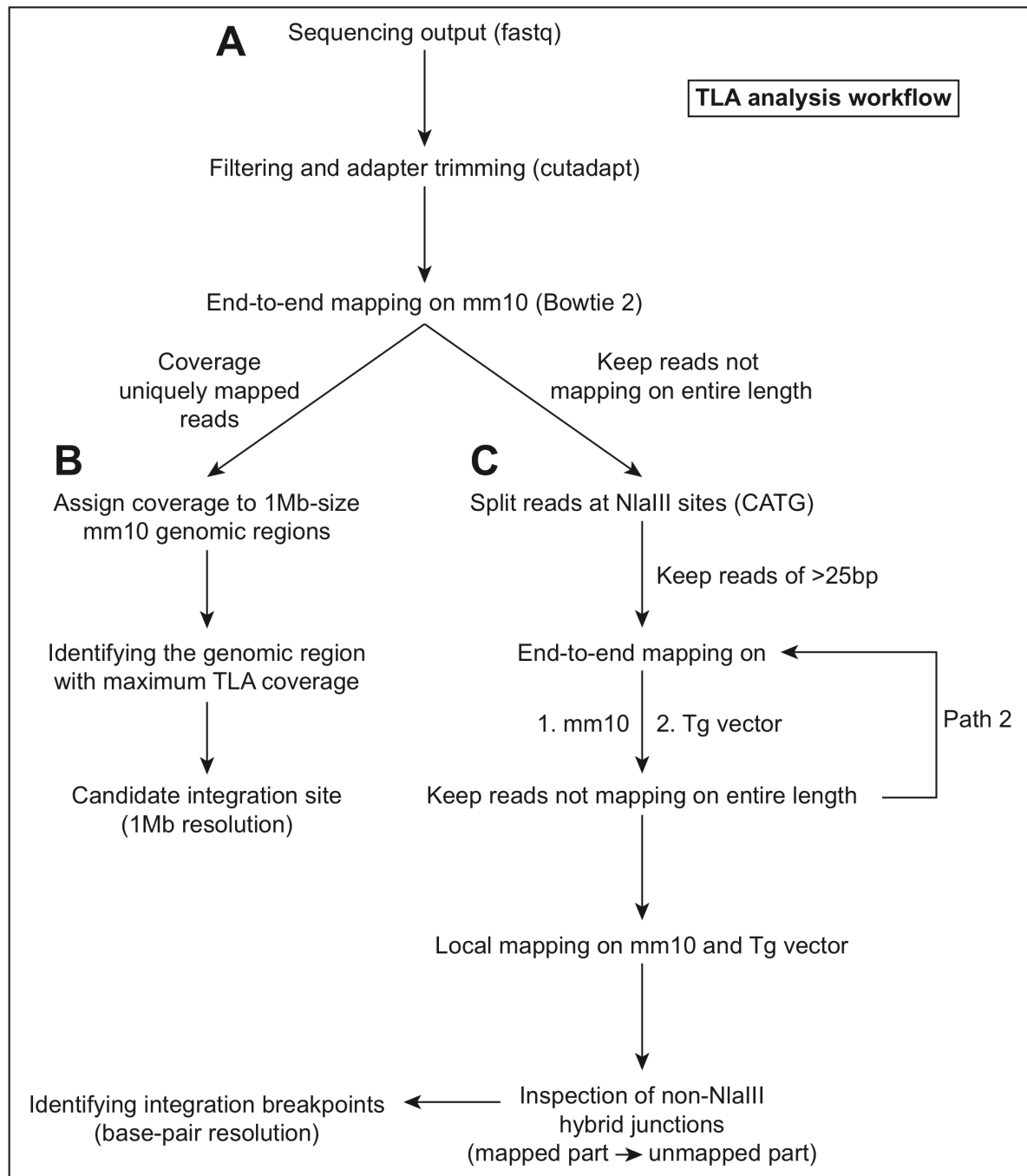
- Zhang, Yong, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoute, David S. Johnson, Bradley E. Bernstein, Chad Nusbaum, et al. 2008. "Model-Based Analysis of ChIP-Seq (MACS)." *Genome Biology* 9 (9): R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.
- Zhao, Zhihu, Gholamreza Tavoosidana, Mikael Sjölander, Anita Göndör, Piero Mariano, Sha Wang, Chandrasekhar Kanduri, et al. 2006. "Circular Chromosome Conformation Capture (4C) Uncovers Extensive Networks of Epigenetically Regulated Intra- and Interchromosomal Interactions." *Nature Genetics* 38 (11): 1341–47. <https://doi.org/10.1038/ng1891>.
- Zhen, Chao Yu, Roubina Tatavosian, Thao Ngoc Huynh, Huy Nguyen Duc, Raibatak Das, Marko Kokotovic, Jonathan B Grimm, et al. 2016. "Live-Cell Single-Molecule Tracking Reveals Co-Recognition of H3K27me3 and DNA Targets Polycomb Cbx7-PRC1 to Chromatin." Edited by Jerry L Workman. *ELife* 5 (October): e17667. <https://doi.org/10.7554/eLife.17667>.
- Ziebarth, Jesse D., Anindya Bhattacharya, and Yan Cui. 2013. "CTCFBSDB 2.0: A Database for CTCF-Binding Sites and Genome Organization." *Nucleic Acids Research* 41 (D1): D188–94. <https://doi.org/10.1093/nar/gks1165>.
- Zuin, Jessica, Jesse R. Dixon, Michael I. J. A. van der Reijden, Zhen Ye, Petros Kolovos, Rutger W. W. Brouwer, Mariëtte P. C. van de Corput, et al. 2014. "Cohesin and CTCF Differentially Affect Chromatin Architecture and Gene Expression in Human Cells." *Proceedings of the National Academy of Sciences of the United States of America* 111 (3): 996–1001. <https://doi.org/10.1073/pnas.1317788111>.

Annexes:



Supplemental Figure S1: See caption on next page.

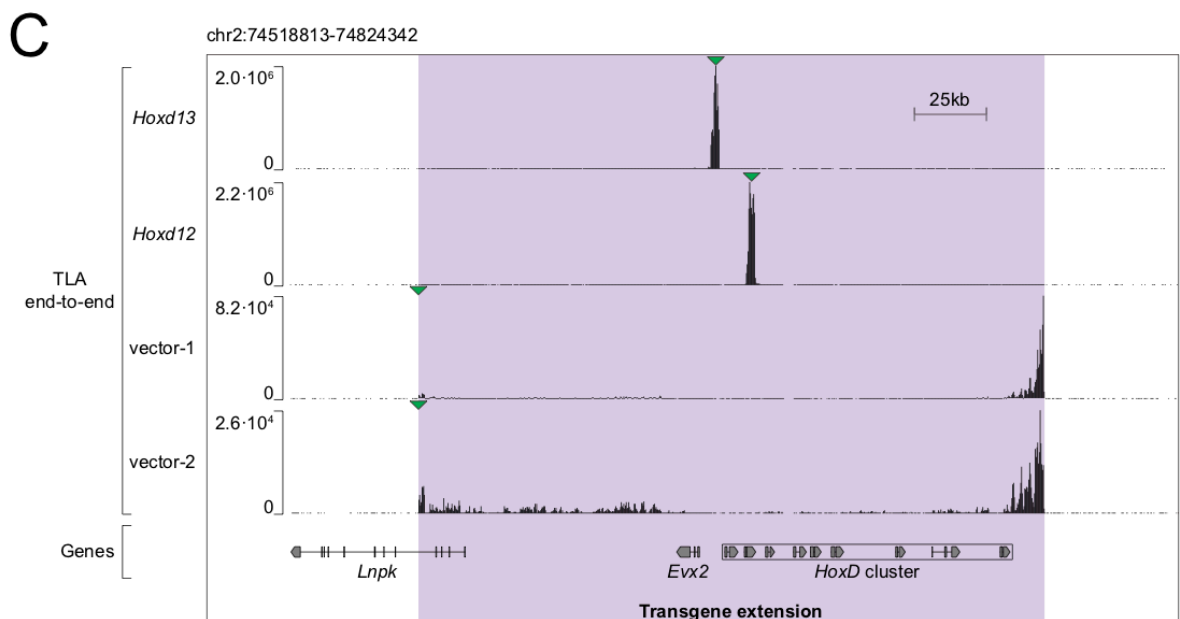
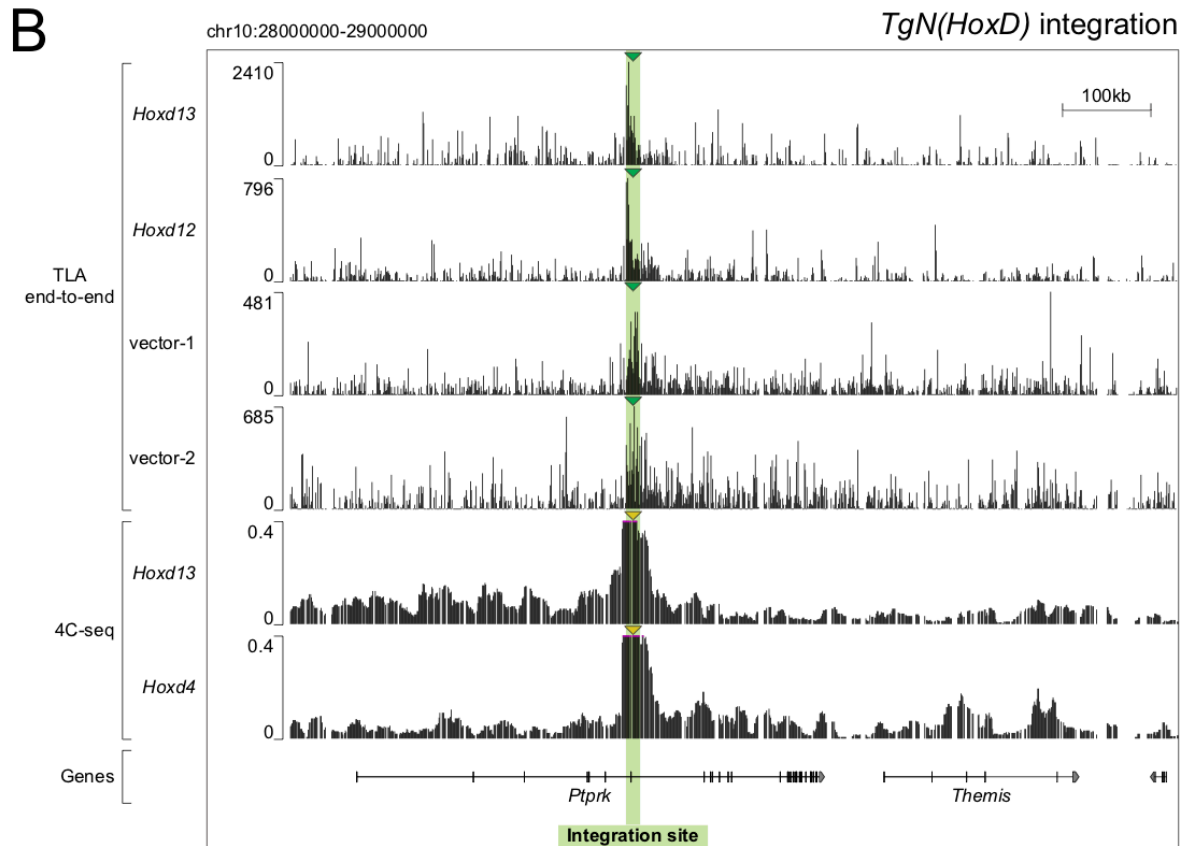
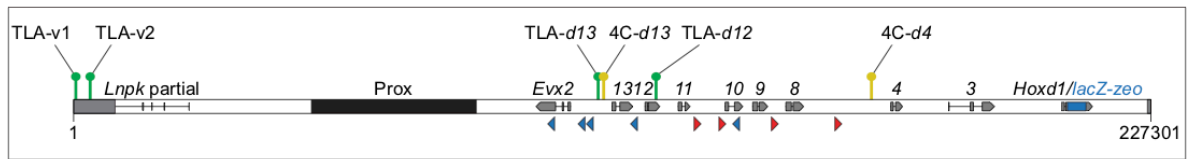
Supplemental Figure S1: Workflow for the extensive characterization of random transgene integrations. In the course of the present study, various aspects associated with the random integration of transgenes were analyzed by a combinatorial strategy involving multiple techniques. (A) Schematization of a random transgene integration event. The transgene (dark gray box), flanked by the 5' and 3' vector sequences (white boxes), integrated at a random genomic location (light gray box). (B) In order to identify a transgene's integration site, inverse PCR (data not shown), 4C-seq, and TLA were performed. While inverse PCR failed to identify the integration site, possibly due to the loss of a restriction site from the construct or because of only low-throughput results, 4C-seq enabled to locate it, but could not reach a base-pair resolution. In contrast to 4C-seq, TLA was suitable to determine the integration site (see breakpoint) with a base-pair resolution, most likely because a higher diversity of hybrid junctions was represented within the circularized TLA molecules (see TLA DNA circle below) compared to those of 4C-seq. (C) To determine in how many copies a given transgene was present, qPCR, TLA, Control-FREEC and MinION turned out to be the most informative techniques. While qPCR on genomic DNA enabled to evaluate copy number for specific DNA regions present within the transgene, TLA allowed us to detect head-to-tail tandem configurations, for example. Control-FREEC enabled to estimate copy number on the entire length of a transgene, based on total input DNA data from ChIPmentation experiments. On its side, MinION allowed us to distinguish between several copies of a transgene along individual molecules thanks to the sequencing of long reads. (D) Regarding the integrity of a transgene, TLA enabled to detect internal integrations of exogenous sequences as well as small deletions. Primer walking (data not shown) using overlapping combinations of primers along the transgene was suitable to detect deletions of large DNA segments (e.g. the deletion of the sequences comprised between the second exon of *Hoxd13* and the promoter of *Evx2* in *TgN(d11-Evx2)*; see Fig. 15A). MinION was also used to assess the preservation of a transgene's sequence, but the relatively high error rate inherent to this technique needs to be taken into consideration when using it for this purpose.



Supplemental Figure S2: TLA analysis workflow. (A) Raw TLA sequencing data in fastq format were filtered and adapter sequences were removed using cutadapt (Martin 2011). The TLA sequencing reads were mapped on their entire length (end-to-end) onto the mm10 reference genome using Bowtie 2 (Langmead and Salzberg 2012). Coverage from uniquely mapped reads and reads not mapping to mm10 on their entire length were used in the (B) and (C) branches of the pipeline, respectively. (B) The sequencing coverage from uniquely mapped reads was assigned to each non-overlapping 1 Mb-size regions of the mm10 genome. The candidate integration site was then identified as the 1 Mb genomic region displaying maximum TLA coverage (not considering the *HoxD* locus, from which all transgenes originated). (C) In parallel, starting from the end-to-end mapping on mm10, only the reads not mapping on their entire length onto the mm10 genome were retrieved. These reads were splitted at the level of NlaIII

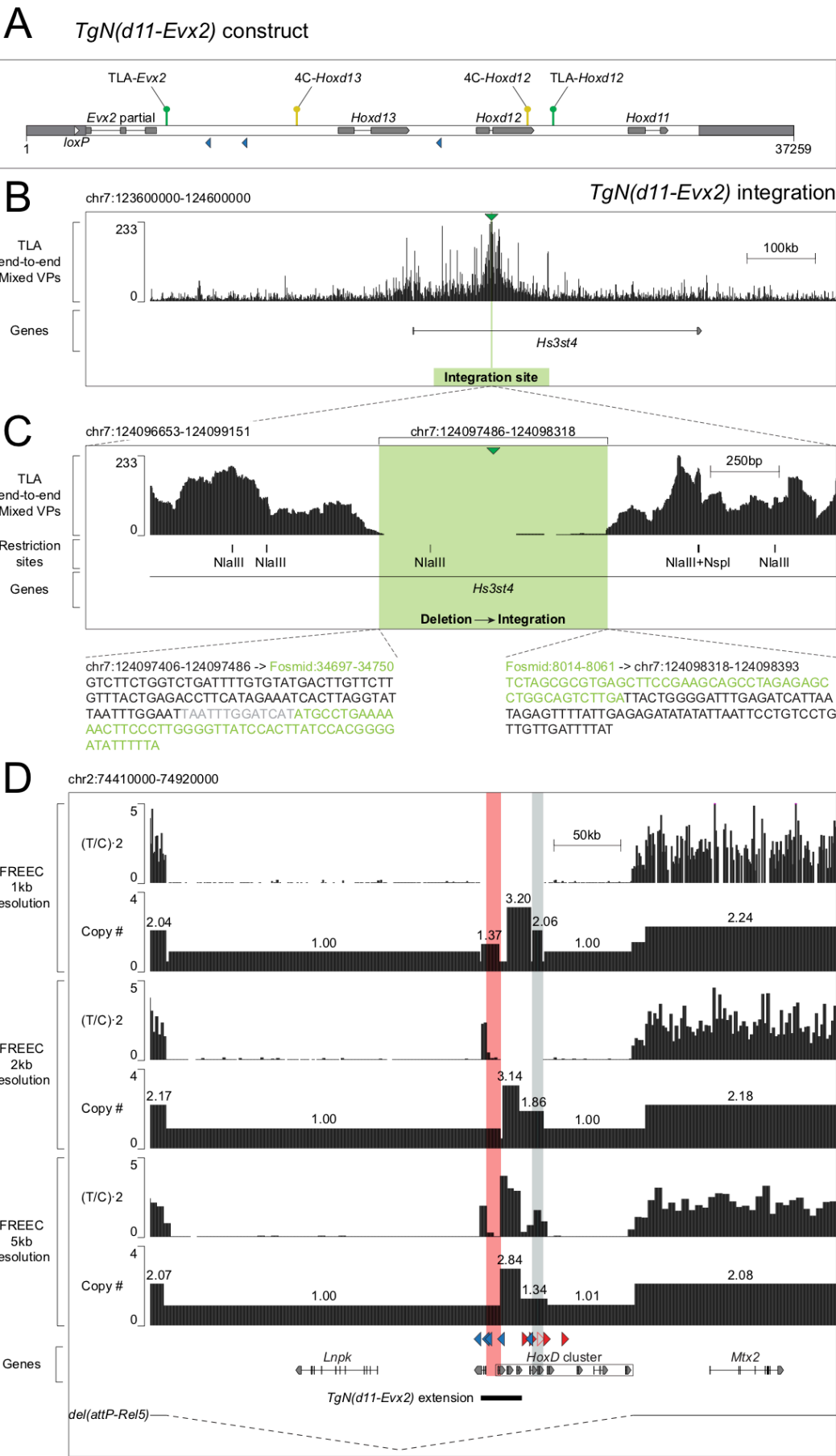
sites (CATG), for NlaIII-digestion + ligation events not to be considered as relevant hybrid junctions (NlaIII hybrid junctions). Only reads of more than 25 bp after splitting were considered informative and were kept for the next steps. Another round of end-to-end mapping on mm10 was performed (path 1). Only the reads not mapping end-to-end onto the mm10 genome were kept. These reads were mapped onto the transgene vector sequence (path 2) and only the reads not mapping end-to-end on the latter were retrieved. Hence, the resulting reads were those not corresponding to mm10, nor to the transgene vector on their entire length, and this could not be the result of NlaIII-digestion + ligation (non-NlaIII hybrid junctions). These reads were mapped locally, i.e. mapped while allowing a segment to not match, on both mm10 and the transgene vector. Next, non-NlaIII hybrid junctions were inspected, going from the mapped part (known) to the unmapped part (unknown), in order to characterize which sequences were brought together (e.g. chromosome on one side and transgene vector on the other). For more details, this analysis required the computation of an average hybrid sequence from all non-NlaIII hybrid reads displaying a particular connection. This enabled us to identify the transgene integration breakpoints with a base-pair resolution (see examples on Figs. S4C, S5C and S8C).

A *TgN(HoxD)* construct



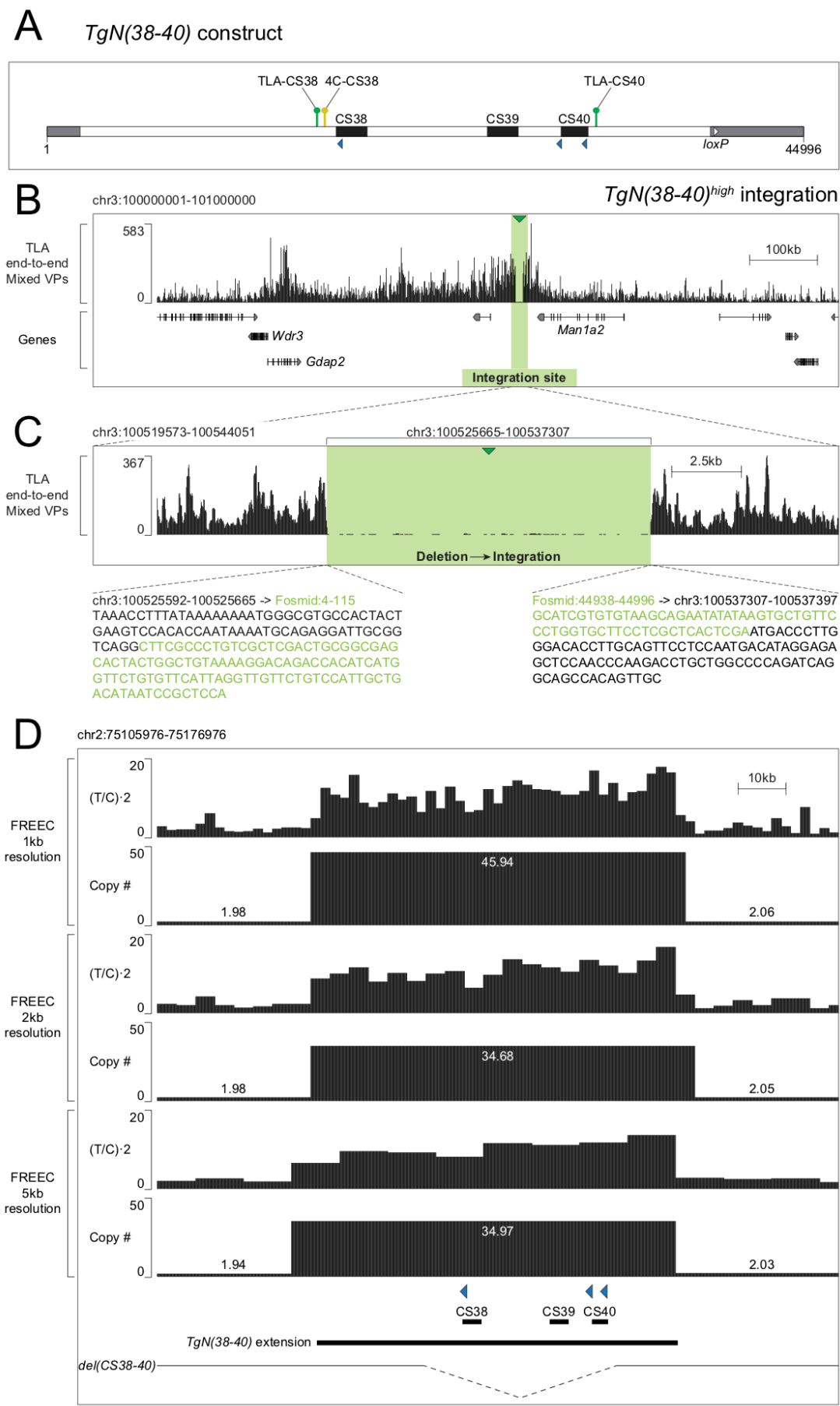
Supplemental Figure S3: See caption on next page.

Supplemental Figure S3: Characterization of the *TgN(HoxD)* integration by TLA and 4C-seq. (A) Scheme of the original *TgN(HoxD)* construct, spanning about 230 kb from part of *Lnpk* to the telomeric side of *Hoxd1/lacZ-zeo*. The transgene vector is indicated by dark gray boxes. The position of both TLA and 4C-seq viewpoints that were used for this line is also shown. (B) TLA and 4C-seq using viewpoints within the *TgN(HoxD)* construct. TLA was performed using the *Hoxd13*, *Hoxd12(-2)*, vector-1 and 2 viewpoints, whereas *Hoxd13* and *Hoxd4* were used as viewpoints for 4C-seq. The TLA tracks represent non-smoothed end-to-end coverages over the 1 Mb region of the mm10 genome that displayed the maximum signal intensity (not considering the *HoxD* cluster's region, from which the transgene originated). The 4C-seq tracks (Leonardo Beccari; unpublished results) represent the scores smoothened with a window of size corresponding to 11 fragments. Importantly, the highest TLA signals matched with the peaks of 4C-seq signals that could be expected in the direct vicinity of a viewpoint, strongly suggesting that the integration of the *TgN(HoxD)* construct occurred at the indicated level (green shaded area), within the *Ptprk* gene. The data were mapped onto the mm10 genome, so that the position of the TLA (green arrowheads) and 4C-seq (yellow arrowheads) viewpoints were artificially aligned to the candidate integration site. The region displayed is chr10:28000000-29000000 (mm10). (C) TLA signals over the *HoxD* cluster's region, from which the *TgN(HoxD)* BAC originated. The distribution of the TLA signals over the region confirmed the preservation of the BAC's extent (see coverages around the edges). The data were mapped onto the mm10 genome, but the signals originated from the ectopic integration since the samples used here were homozygous for a deletion of the endogenous *HoxD* cluster (*HoxD^{del(1-13)}*). The 4C-seq results are not displayed for this region because the samples that were used did not fulfill the abovementioned background requirements. The TLA viewpoints were positioned on this figure either according to their true locations on the region (*Hoxd13* and *Hoxd12*) or artificially at the level of the left end of the BAC (vector-1 and 2). The extension of the *TgN(HoxD)* BAC clone (chr2:74563132-74777993) is highlighted in purple. The region displayed is chr2:74518813-74824342 (mm10).



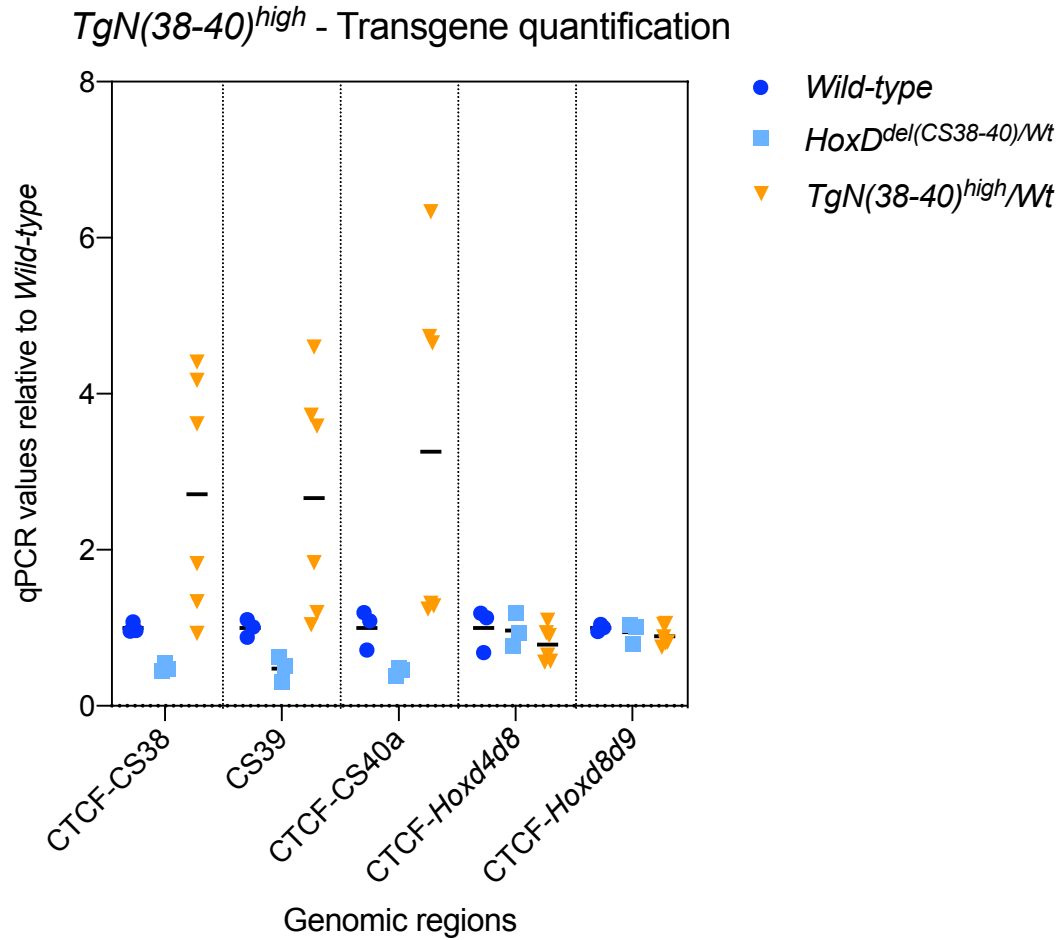
Supplemental Figure S4: See caption on next page.

Supplemental Figure S4: Characterization of the *TgN(d11-Evx2)* integration. (A) Scheme of the original *TgN(d11-Evx2)* construct, spanning about 37 kb from the third exon of *Evx2* to the telomeric side of *Hoxd11*. The transgene vector is indicated by dark gray boxes. The *loxP* site (light gray arrowhead) used for transgene extra-copy removal was located in the 5' part of the vector. The position of both on-construct TLA and 4C-seq viewpoints that were used for this line is also shown. (B and C) TLA using viewpoints within the *TgN(d11-Evx2)* construct. TLA was performed using the *Evx2* and *Hoxd12(-1)* viewpoints. Of note, TLA libraries were prepared starting from mixed samples, corresponding to different transgenic lines, and several viewpoints were used simultaneously. The signal was attributed to each line based on the knowledge of the approximative welcoming region (different for each line), which was identified during pilot 4C-seq experiments (data not shown). The data were mapped onto the mm10 genome, so that the position of the mixed TLA viewpoints (green arrowheads) was artificially aligned to the integration site (green shaded area). The region shown in (B) (chr7:123600000-124600000, mm10) displayed the maximum end-to-end TLA coverage that could be attributed to this line. The region magnified in (C) (chr7:124096653-124099151, mm10) comprises the integration site (chr7:124097486-124098318, mm10), located within the *Hs3st4* intron, where a ca. 1 kb deletion of chromosomal sequence was also detected. The average TLA hybrid junctions corresponding to both integration breakpoints are shown at the bottom of the panel. They joined chromosome 7 sequences (black) together with fosmid sequences (green), and vice versa. Both the topology and the coordinates support a forward sequence configuration of the construct at the level of this integration. The origin of the sequence in gray is unknown. (D) Control-FREEC analysis applied on total input DNA data from *TgN(d11-Evx2)* samples (homozygous for both the transgene and the *del(attP-Rel5)* background). The analysis was performed for non-overlapping windows of 3 different sizes (1, 2 and 5 kb). (T/C) stands for test (*TgN(d11-Evx2)*) over control (transgene-negative and *wild-type* for the endogenous *HoxD* cluster) signal ratio. It was multiplied by 2 in order to obtain values reflecting absolute copy numbers. Below are shown the corresponding copy number evaluations by the algorithm. Values around 2 were observed outside of the *del(attP-Rel5)* extension and at the level of the *Hoxd9/lacZ* of this allele (red shaded area) and correspond to the 2 copies of these DNA sequences at the endogenous *HoxD* locus, as expected. Importantly, a value around 3 for the *Hoxd12-Hoxd13* interval indicates the presence of this region in 1 or 2 *in-cis* copies at the level of the integration, since the *TgN(d11-Evx2)* samples that were used were homozygous for the transgene. Surprisingly, the algorithm returned a value of 1 instead of 0 for regions that were actually absent from the genome of the test samples, among which the sequence that was lost from the construct (red shaded area). The region displayed is chr2:74410000-74920000 (mm10).

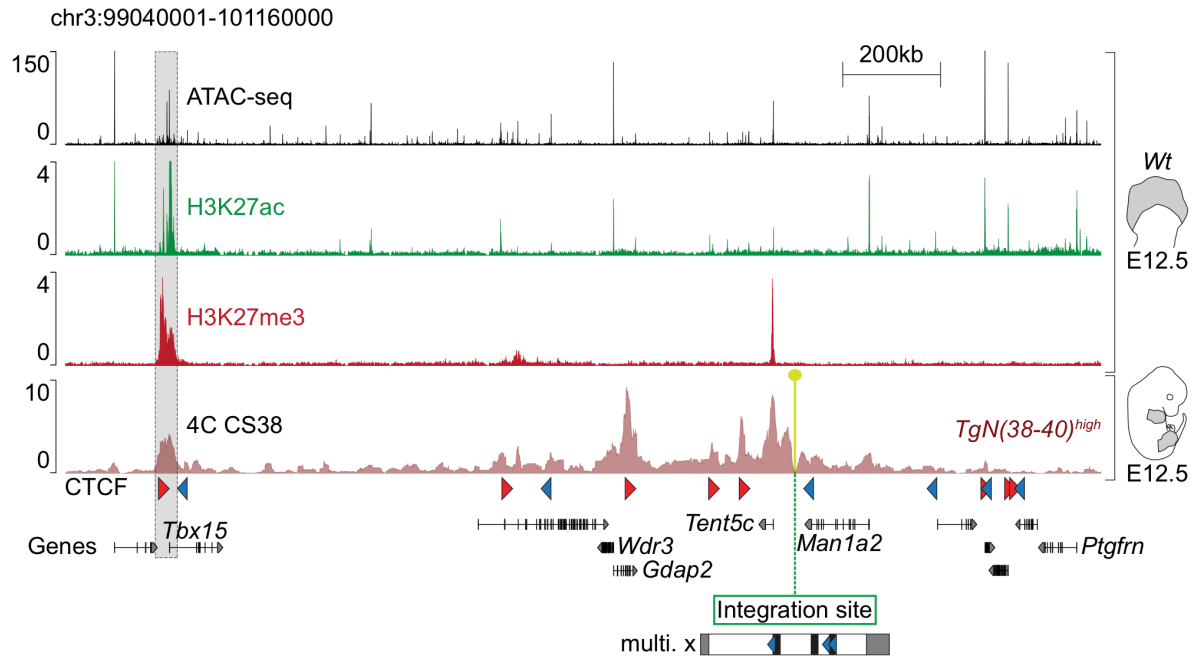


Supplemental Figure S5: See caption on next page.

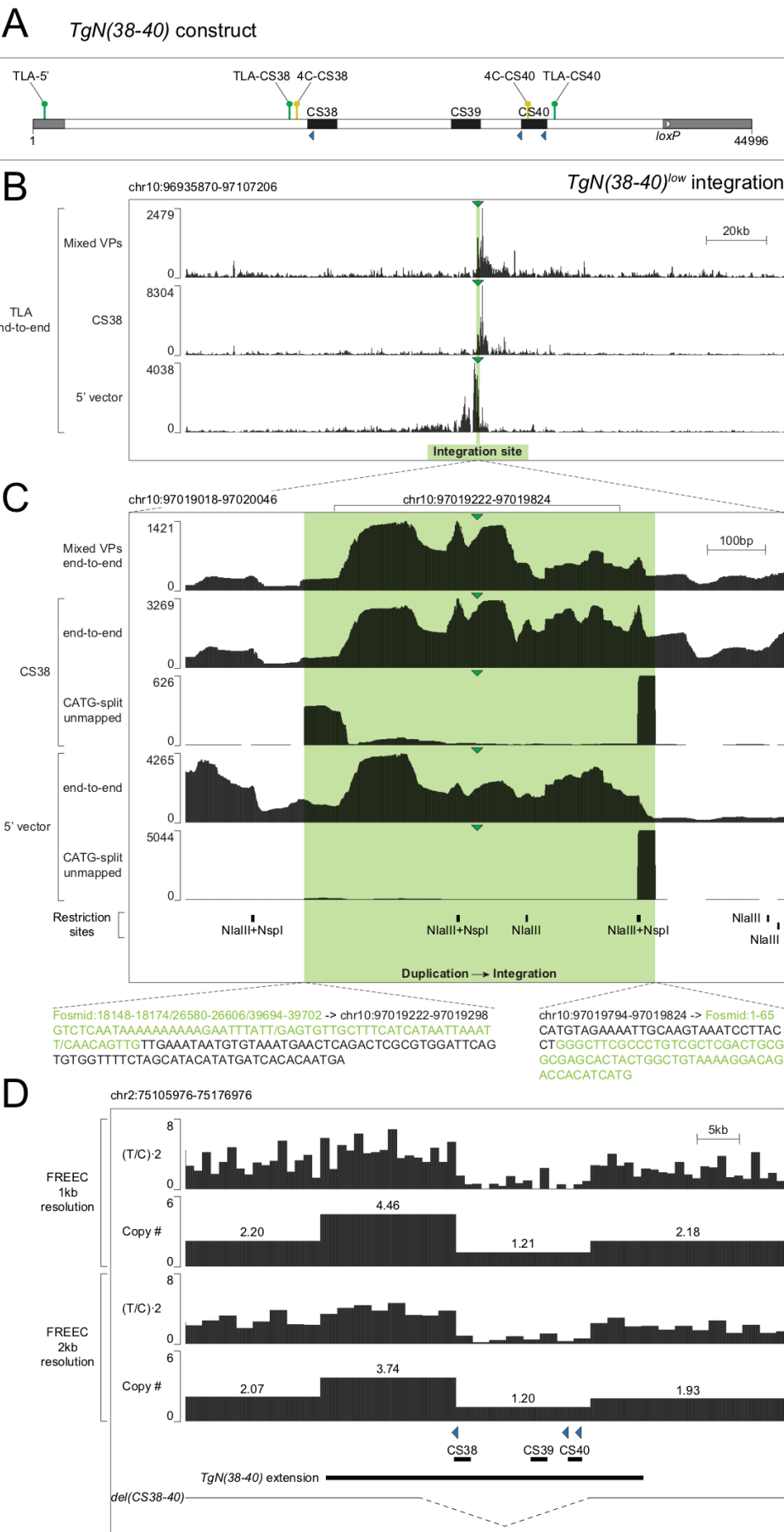
Supplemental Figure S5: Characterization of the *TgN(38-40)^{high}* integration. (A) Scheme of the original *TgN(38-40)* construct, spanning 45 kb and including regions CS38, CS39 and CS40. The transgene vector is indicated by dark gray boxes. The *loxP* site (light gray arrowhead) used for transgene extra-copy removal was located in the 3' (right) part of the vector. The position of both on-construct TLA and 4C-seq viewpoints that were used for this line is also shown. (B and C) TLA using viewpoints within the construct in *TgN(38-40)^{high}*. TLA was performed using the CS38 and CS40 viewpoints. TLA libraries were prepared with mixed samples, corresponding to different transgenic lines, and mixed viewpoints, following the same strategy as explained in the caption of Figure S4. The data were mapped onto the mm10 genome, so that the position of the mixed TLA viewpoints (green arrowheads) was artificially aligned to the integration site (green shaded area). The region shown in (B) (chr3:100000001-101000000, mm10) displayed the maximum end-to-end TLA coverage that could be attributed to this line. The region magnified in (C) (chr3:100519573-100544051, mm10) comprises the integration site (chr3:100525665-100537307, mm10), where a ca. 12 kb deletion of chromosomal sequence was also detected. The average TLA hybrid junctions corresponding to both integration breakpoints are shown at the bottom of the panel. They joined chromosome 3 sequences (black) together with fosmid sequences (green), and vice versa. Both the topology and the coordinates support a forward sequence configuration of the construct at the level of this integration. (D) Control-FREEC analysis applied on total input DNA data from *TgN(38-40)^{high}* samples (homozygous for both the transgene and the *del(CS38-40)* background). The analysis was performed for non-overlapping windows of 3 different sizes (1, 2 and 5 kb). (T/C) stands for test (*TgN(38-40)^{high}*) over control (transgene-negative and *wild-type* for the endogenous region CS38-40) signal ratio. It was multiplied by 2 in order to obtain values reflecting absolute copy numbers. Below are shown the corresponding copy number evaluations by the algorithm. Values around 2 were observed outside of the *del(CS38-40)* extension and correspond to the 2 copies of this sequence at the endogenous locus, as expected. Importantly, a value around 35-46 matching the extension of *TgN(38-40)* indicates the presence of this region in between 17 and 23 apparently entire *in-cis* copies at the level of the integration, since the *TgN(38-40)^{high}* samples that were used were homozygous for the transgene. The region displayed is chr2:75105976-75176976 (mm10).



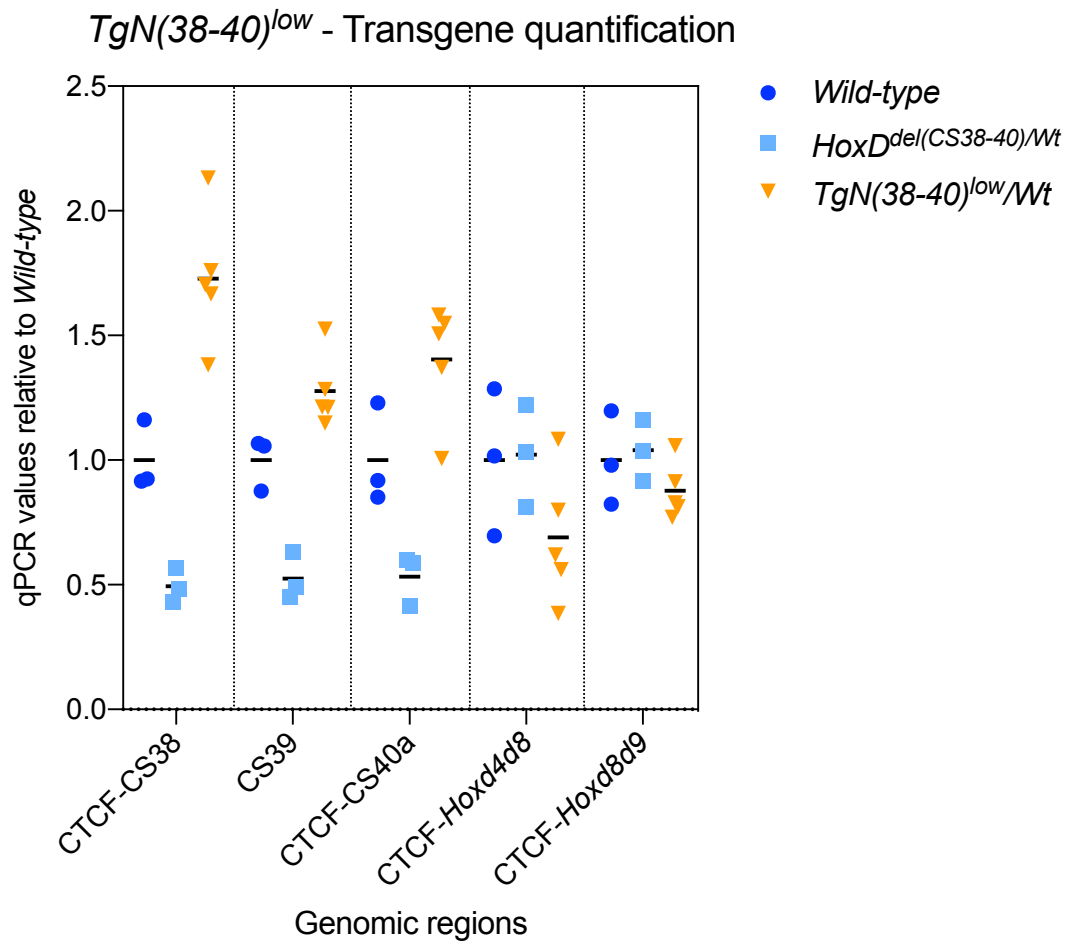
Supplemental Figure S6: qPCR transgene quantification of *TgN(38-40)^{high}*. This experiment was performed on *wild-type* (n=3, blue circles), *del(CS38-40)* heterozygous (n=3, light blue squares) and *TgN(38-40)^{high}* heterozygous (n=6, orange triangles) gDNA samples from adult mice (Eddie Rodríguez-Carballo and Marie-Laure Gadolini; unpublished results). For each sample, the results were normalized to the value of *Albh1a2*. In vertical is represented the qPCR value relative to the *wild-type*, for each qPCR target. CTCF-CS38, CS39 and CTCF-40a (first CTCF site of CS40) are regions present in the *TgN(38-40)* construct. CTCF-*Hoxd4d8* (intergenic CTCF site between *Hoxd4* and 8) and CTCF-*Hoxd8d9* (intergenic CTCF site between *Hoxd8* and 9) were used as control regions, since they are only present at the endogenous *HoxD* locus. In *TgN(38-40)^{high}*, a mean value around 3 (horizontal bar) for CTCF-CS38 represents an absolute copy number of 6. Since 2 of these copies correspond to the endogenous regions CS38-40, this represents 4 copies of CS38 at the level of the integration. Similar results were obtained for both CS39 and CTCF-CS40a, indicating the presence of multiple complete copies of CS38-40 in *TgN(38-40)^{high}*. As regards CTCF-*Hoxd4d8* and CTCF-*Hoxd8d9*, 2 absolute copies corresponding to the 2 *wild-type* alleles of the *HoxD* cluster were detected for all genotypes, as expected.



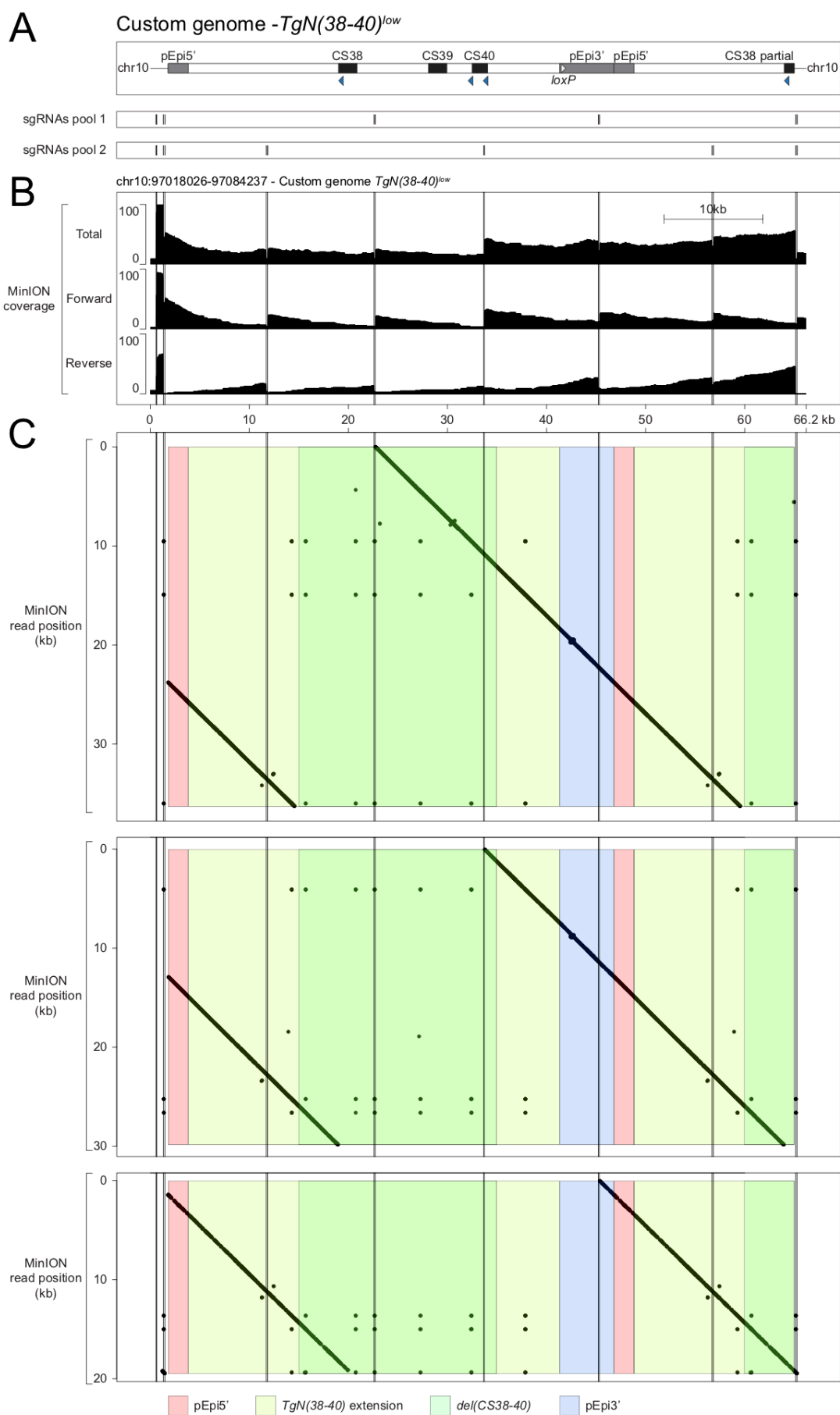
Supplemental Figure S7: Chromatin accessibility (ATAC-seq in black; $n=1$) and distribution of H3K27ac (green; $n=2$) and H3K27me3 (red; $n=8$) marks within the *TgN(38-40)^{high}* welcoming landscape in *wild-type* distal forelimbs at E12.5. Note that the *Tbx15* gene promoter is open and decorated by H3K27ac marks in a given fraction of cells, whereas it is covered by H3K27me3 marks in another fraction. Interactions established by the CS38 element in whole limbs of *TgN(38-40)^{high}* E12.5 embryos are shown below for comparison. Note that the contacts between CS38 and *Tbx15* precisely match the zone of H3K27me3 enrichment around the promoter of *Tbx15* (gray shaded area) (see Discussion). The region displayed is chr3:99040001-101160000 (mm10). ATAC-seq data are unpublished and were shared by Christopher Chase Bolt. H3K27ac and H3K27me3 ChIP-seq data are from (Rodríguez-Carballo et al. 2019).



Supplemental Figure S8: Characterization of the *TgN(38-40)^{low}* integration. (A) Scheme of the original *TgN(38-40)* construct, spanning 45 kb and including regions CS38, CS39 and CS40. The transgene vector is indicated by dark gray boxes. The *loxP* site (light gray arrowhead) used for transgene extra-copy removal was located in the 3' (right) part of the vector. The position of both on-construct TLA and 4C-seq viewpoints that were used for this line is also shown. (B and C) TLA using viewpoints within the construct in *TgN(38-40)^{low}*. TLA was performed using the 5' part of the vector, CS38 and CS40 as viewpoints. TLA libraries were either prepared with mixed samples and viewpoints (CS38 and CS40; see mixed VPs tracks), following the same strategy as explained in the caption of Figure S4, or individual libraries were produced to increase the resolution (see CS38 and 5' vector tracks). The data were mapped onto the mm10 genome, so that the position of the TLA viewpoints (green arrowheads) was artificially aligned to the integration site (green shaded area). The region shown in (B) (chr10:96935870-97107206, mm10) displayed the maximum end-to-end TLA coverages for this line. The region magnified in (C) (chr10:97019018-97020046, mm10) comprises the integration site (chr10:97019222-97019824, mm10), where a ca. 600 bp duplication of chromosomal sequence was also detected. For both the CS38 and 5' vector individual viewpoints, below the end-to-end coverage are shown the results of the local mapping of reads that were splitted at the level of NlaIII sites (CATG) but still did not map end-to-end on both the mm10 reference genome and the vector sequence (non-NlaIII hybrid reads). Drops of signal not coinciding with NlaIII sites are a typical signature of integration breakpoints. The determination of the part of these reads that did not map (i.e. beyond the drop) enabled to generate the average hybrid junctions corresponding to both integration breakpoints that are represented at the bottom of the pannel. They joined chromosome 10 sequences (black) together with fosmid sequences (green), and vice versa. Both the topology and the coordinates support that the transgene was present in a forward sequence configuration in-between the duplicated chromosomal sequences (chr10:97019824 – Transgene – duplication of chr10:97019222-97019824). Importantly, the part of the transgene displaying the transgene-to-chr10 topology consisted of a partial fragment extending towards CS38 (Fosmid:18148-18174), followed by short fused sequences of different origins with respect to the original construct (Fosmid:26580-26606, which originated from CS39 and Fosmid:39694-39702, which matched with the 3' part of the vector). (D) Control-FREEC analysis applied on total input DNA data from *TgN(38-40)^{low}* samples (heterozygous for the transgene and homozygous for the *del(CS38-40)* background). The analysis was performed for non-overlapping windows of 2 different sizes (1 and 2 kb). (T/C) stands for test (*TgN(38-40)^{low}*) over control (transgene-negative and *wild-type* for the endogenous region CS38-40) signal ratio. It was multiplied by 2 in order to obtain values reflecting absolute copy numbers. Below are shown the corresponding copy number evaluations by the algorithm. Values around 2 were observed outside of the *del(CS38-40)* extension, as expected (see caption of Fig. S5). Importantly, a value around 4 for the part of the *TgN(38-40)* sequence not matching with the deleted sequence of *del(38-40)* indicates the presence of this region in 2 copies at the level of the integration, since 2 copies can be attributed to the endogenous locus and taking into account that the *TgN(38-40)^{low}* samples that were used were heterozygous for the transgene. In contrast, a value around 1 was obtained for the rest of the transgene sequence, the majority of which was deleted at the endogenous locus, indicating the presence of only 1 copy of the corresponding regions at the integration. Hence, together with the TLA results, these data strongly suggest the presence of one entire copy of region CS38-40 followed by a partial copy extending towards region CS38 at the integration site. The region displayed is chr2:75105976-75176976 (mm10).

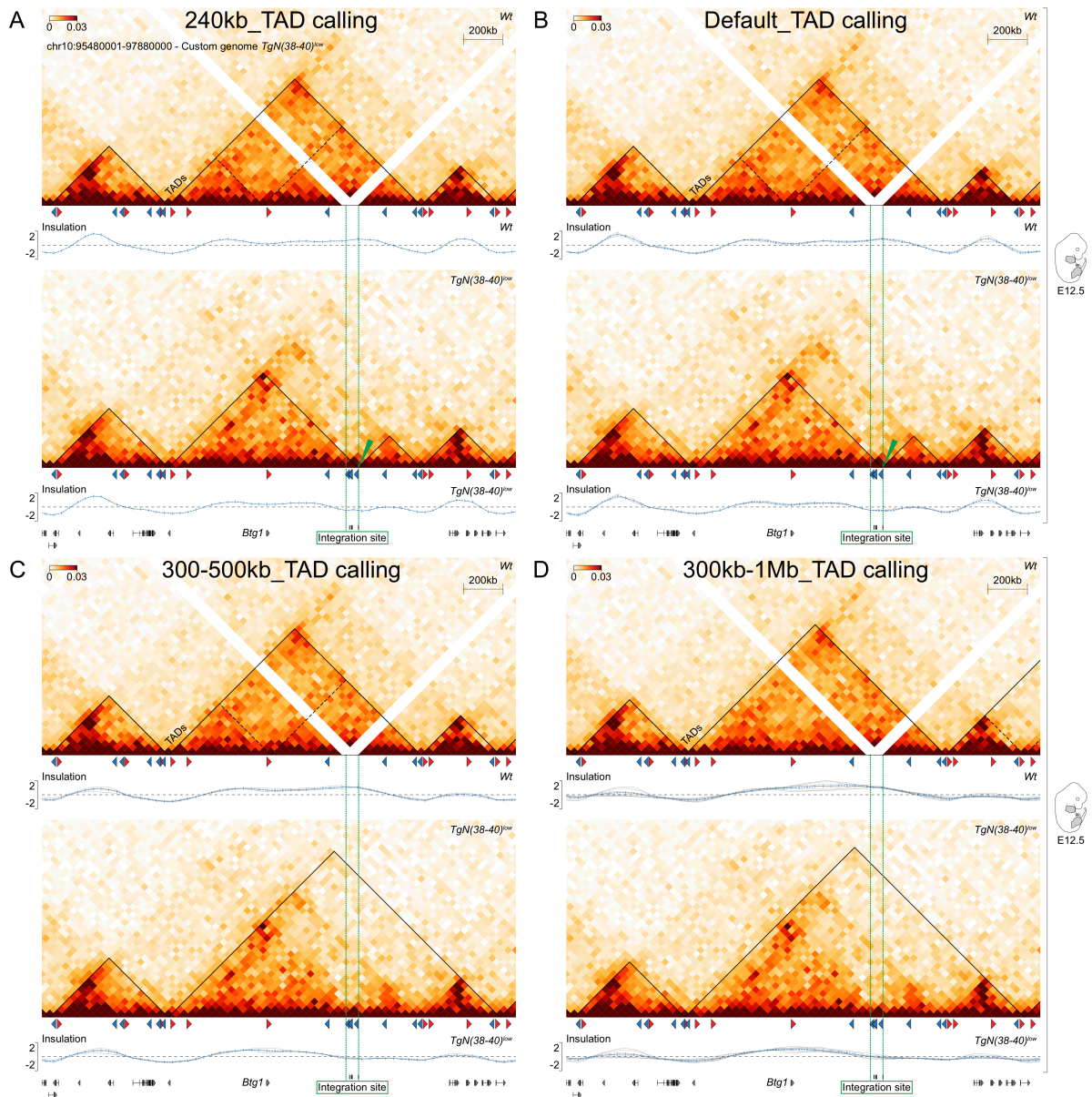


Supplemental Figure S9: qPCR transgene quantification of *TgN(38-40)^{low}*. This experiment was performed on *wild-type* (n=3, blue circles), *del(CS38-40)* heterozygous (n=3, light blue squares) and *TgN(38-40)^{low}* heterozygous (n=5, orange triangles) gDNA samples from adult mice (Eddie Rodríguez-Carballo and Marie-Laure Gadolini; unpublished results). For each sample, the results were normalized to the value of *Aldh1a2*. In vertical is represented the qPCR value relative to the *wild-type*, for each qPCR target. CTCF-CS38, CS39 and CTCF-40a (first CTCF site of CS40) are regions present in the *TgN(38-40)* construct. CTCF-*Hoxd4d8* (intergenic CTCF site between *Hoxd4* and 8) and CTCF-*Hoxd8d9* (intergenic CTCF site between *Hoxd8* and 9) were used as control regions, since they are only present at the endogenous *HoxD* locus. In *TgN(38-40)^{low}*, a mean value of 1.75 (horizontal bar) for CTCF-CS38 represents an absolute copy number of 3.5. Since 2 of these copies correspond to the endogenous regions CS38-40, this represents 1 or 2 copies of CS38 at the level of the integration. On the same basis, the results indicate that both CS39 and CTCF-CS40a were present in about 1 copy. The results obtained for both CTCF-*Hoxd4d8* and CTCF-*Hoxd8d9* are in line with the presence of 2 absolute copies corresponding to the 2 *wild-type* alleles of the *HoxD* cluster, although some discrepancies were observed for *TgN(38-40)^{low}*; namely a mean value of 0.75-0.8 compared to 1 for the other genotypes.



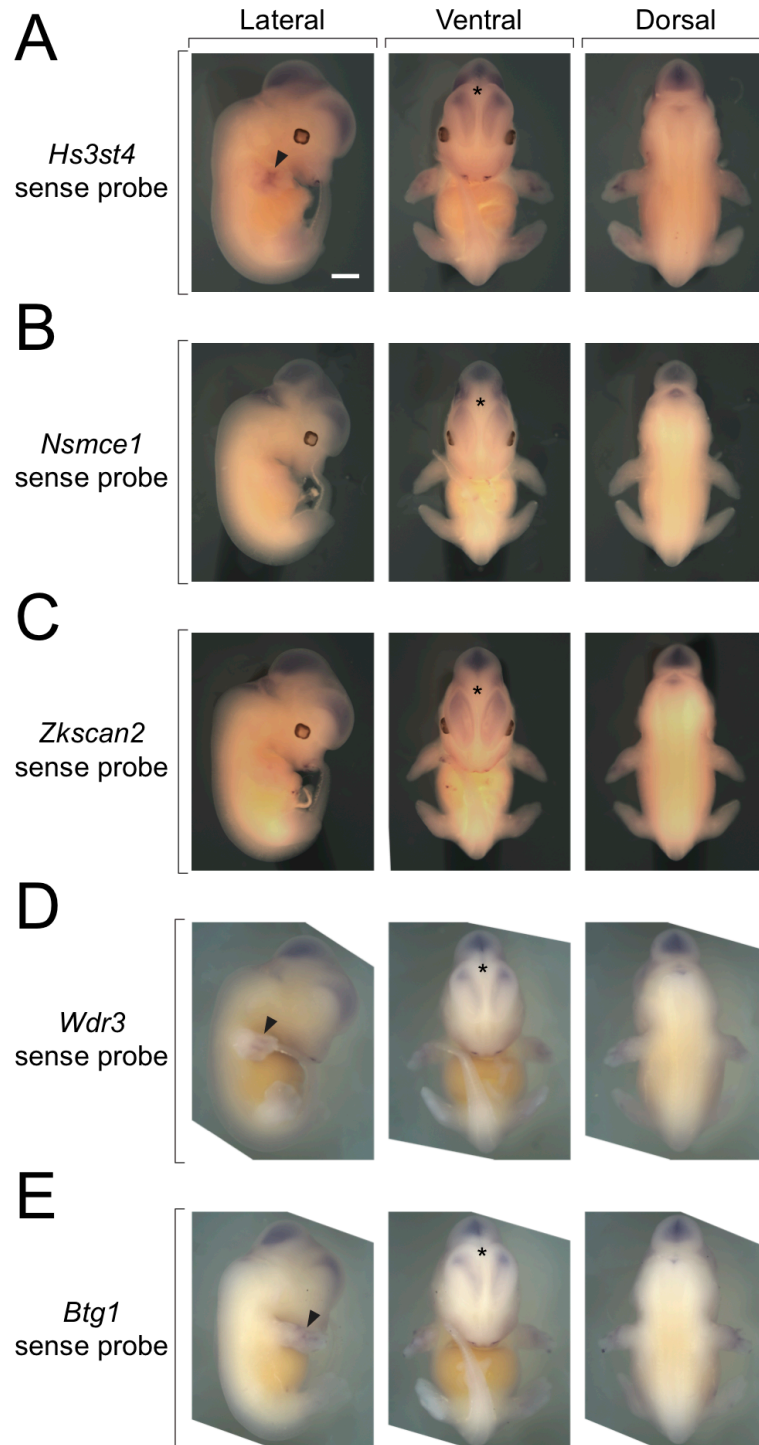
Supplemental Figure S10: See caption on next page.

Supplemental Figure S10: MinION long-read sequencing of *TgN(38-40)^{low}*. (A) Scheme of the *TgN(38-40)^{low}* mutant genomic configuration reconstructed from TLA, Control-FREEC and qPCR. It consists of one entire copy of region CS38-40 and of a partial fragment extending towards an additional region CS38 (63.2 kb in total), both inserted after position chr10:97019824 of the mm10 genome (see Material and Methods). For all panels, the region displayed is chr10:97018026-97084237 of the mm10+*TgN(38-40)^{low}* custom genome (chr10 extension corresponding to chr10:97018026-97020425 of *wild-type* mm10). Below is shown the position of the sgRNAs used for the Cas9-mediated release that enabled to enrich the solution in sequencing-competent DNA molecules originating from the region of interest. The release was performed in two independent reactions, using two distinct pools of sgRNAs so as to produce overlapping fragments. The resulting 9.5 to 23 kb fragments were pooled and sequenced together. The MinION data were displayed onto the *TgN(38-40)^{low}* configuration in panels (B) and (C). The sample used for this experiment was homozygous for both the transgene and the *del(CS38-40)* background (deletion of the endogenous region CS38-40 on chromosome 2). (B) MinION sequencing coverage on the *TgN(38-40)^{low}* mutant configuration. Total, forward and reverse strand sequencing coverages are shown. The MinION coverage revealed an about 30% enrichment for the region of interest compared to the rest of the genome and the segmented pattern demonstrated the efficiency of the release. Note that the quality of the sequencing was higher at the beginning of the reads, as revealed by opposite slopes between forward and reverse coverages. (C) Dot plot visualization of selected individual MinION sequencing reads. Each dot represents a complete identity along 20 bp between the read fasta file and the construct fasta file. Continuous oblique lines represent single MinION reads matching the expected genomic configuration. A difference between a read and the expected configuration would manifest by a line displaying either a discontinuity along the x-axis in case of insertion/deletion, or a bending causing the line to go backwards in case of duplication-inversion (e.g. head-to-head or tail-to-tail tandem). Importantly, no read fully spanned the mutant construction (i.e. displaying chr10 – Transgene – chr10), as expected from the total length and the release strategy. The longest read mapping on the construct spanned 37 kb and is shown on top. It confirmed the head-to-tail configuration of the integration and the preservation of all sequences along the indicated interval (from inter-CS38/CS40 of the first copy to a region located upstream of the second copy of CS38). The oblique line on the left corresponds to the same read, but automatically assigned to the first copy of CS38-40. Together, the second and third reads shown on this panel confirmed that the second (partial) copy of the construct was well preserved, that it included the CTCF site of CS38 and joined the partial CS38 with chromosome 10, as expected from the TLA results. Hence, the MinION results strongly support the mutant configuration shown in (A). The color-code for the region displayed is given at the bottom of the panel. pEpi5' and pEpi3' are the 5' and 3' parts of the vector, respectively.



Supplemental Figure S11: TAD callings in *TgN(38-40)^{low}* using various parameter values. TAD identification was achieved using hicFindTADs. (A) TAD calling with a window size of 240 kb, as for Figures 21 and 22. (B) TAD calling using default window sizes. Here, the algorithm computed insulation scores for windows of sizes 200 kb and 280 kb, calculated the mean insulation (see blue curve) and performed TAD identification starting from the mean. (C) TAD calling with arbitrary window sizes comprised between 300 kb and 500 kb. For more details, the algorithm computed insulation for windows of sizes 300, 340 and 413 kb. (D) TAD calling with arbitrary window sizes comprised between 300 kb and 1 Mb. More precisely, the algorithm computed insulation for windows of sizes 300, 340, 413, 508, 620, 747 and 888 kb. The integration-induced splitting of the welcoming TAD is indicated by a green arrowhead. Note that this was only observed for the more stringent 240 kb and Default TAD callings, indicating the formation of a relatively weak boundary (or sub-TAD boundary) at the level of the integration. In (A), (B) and (C), mapping of the *wild-type* Hi-C data on the mutant genome (dotted *wild-type* TADs) resulted in an artefactual splitting of the welcoming TAD at the level of the *Btg1* CTCF site due to the lack of mappable reads. Therefore, we also represented the

same data mapped onto the *wild-type* genome (solid *wild-type* TADs), adapted to the bin shift imposed by the integration. All TADs in the *TgN(38-40)^{low}* condition were identified starting from the mutant data mapped onto the mutant genome. The region displayed is chr10:95480001-97880000 of the mm10+*TgN(38-40)^{low}* custom genome (extent corresponds to chr10:95480001-97816188 of *wild-type* mm10).



Supplemental Figure S12: Selected sense probes controls of the whole-mount *in situ* hybridization (WISH) experiments. The embryos shown on this figure were processed following exactly the same protocol (see whole-mount *in situ* hybridization in Material and Methods) as for antisense whole-mount *in situ* hybridizations. (A) Sense probe *Hs3st4* control. BM-Purple staining was only observed in the limbs and the brain. (B and C) Sense probe controls for *Nsmce1* (B) and *Zkscan2* (C). Staining was only observed in the brain. (D and E) Sense probe controls for *Wdr3* (D) and *Btg1* (E). As in (A), signal was detected in both the brain and the limbs. Arrowheads and asterisks indicate signal in the limbs and the brain, respectively. Scale bar: 1 mm.

4C-seq viewpoint	Genotype	Tissue	Biological replicates
Island-2	<i>wild-type</i>	E12.5 PFL	1
		E12.5 DFL	1
	<i>Dup(8i-10)</i>	E12.5 PFL	1
		E12.5 DFL	1
	<i>Dup(11-13)</i>	E12.5 PFL	1
		E12.5 DFL	1
Island-4	<i>wild-type</i>	E12.5 PFL	5
		E12.5 DFL	5
	<i>Dup(8i-10)</i>	E12.5 PFL	2
		E12.5 DFL	2
	<i>Dup(11-13)</i>	E12.5 PFL	2
		E12.5 DFL	2
<i>Evx2</i>	<i>wild-type</i>	E12.5 PFL	3
		E12.5 DFL	3
	<i>Dup(8i-10)</i>	E12.5 PFL	2
		E12.5 DFL	2
	<i>Dup(11-13)</i>	E12.5 PFL	2
		E12.5 DFL	2
<i>Hoxd11</i>	<i>wild-type</i>	E12.5 PFL	1
		E12.5 DFL	1
	<i>Dup(8i-10)</i>	E12.5 PFL	2
		E12.5 DFL	2
<i>Hoxd9</i>	<i>wild-type</i>	E12.5 PFL	2
		E12.5 DFL	2
	<i>Dup(11-13)</i>	E12.5 PFL	2
		E12.5 DFL	2
<i>Hoxd4</i>	<i>wild-type</i>	E12.5 PFL	3
		E12.5 DFL	3
	<i>Dup(8i-10)</i>	E12.5 PFL	2
		E12.5 DFL	2
	<i>Dup(11-13)</i>	E12.5 PFL	2
		E12.5 DFL	2
CS38	<i>wild-type</i>	E12.5 PFL	5
		E12.5 DFL	5
	<i>Dup(8i-10)</i>	E12.5 PFL	2
		E12.5 DFL	2
	<i>Dup(11-13)</i>	E12.5 PFL	2
		E12.5 DFL	2

Supplemental Table S1: Biological replicates of the 4C-seq experiments performed in the duplications *Dup(8i-10)* and *Dup(11-13)* as well as the corresponding *wild-type* controls. DFL and PFL stand for proximal and distal forelimbs, respectively. Genotypes are expressed as alleles for sake of simplicity, but all experiments were performed on samples that were homozygous for the corresponding alleles. Genotypes are colored in the same way than the corresponding tracks of the main figures. The *wild-type* 4C-seq data for island-2, island-4, *Evx2*, *Hoxd4* and CS38 have been published in (Rodríguez-Carballo et al. 2017). One replicate from all 4C-seq experiments in *Dup(8i-10)* and *Dup(11-13)* performed in duplicates was contributed by Imane El-Idrissi.

Transgenic line	4C-seq viewpoint/Hi-C	Tissue	Genotype
<i>TgN(HoxD)</i>	<i>Hoxd13</i>	E12.5 WL	<i>TgN(HoxD)/–</i>
	<i>Hoxd4</i>	E12.5 WL	<i>TgN(HoxD)/–</i>
<i>TgN(d11-Evx2)</i>	<i>Hoxd12</i>	E10.5 HLE	<i>TgN(d11-Evx2)/Wt</i>
	CTCF-left 1	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(d11-Evx2)/TgN(d11-Evx2)</i>
	CTCF-left 2	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(d11-Evx2)/TgN(d11-Evx2)</i>
	<i>Hs3st4</i>	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(d11-Evx2)/TgN(d11-Evx2)</i>
	Right 1	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(d11-Evx2)/TgN(d11-Evx2)</i>
	Right 2	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(d11-Evx2)/TgN(d11-Evx2)</i>
<i>TgN(38-40)^{high}</i>	CS38	E12.5 WL	<i>TgN(38-40)^{high}/Wt</i>
	CTCF-left(h)	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(38-40)^{high}/TgN(38-40)^{high}</i>
	CTCF-right(h)	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(38-40)^{high}/TgN(38-40)^{high}</i>
	<i>Man1a2</i>	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(38-40)^{high}/TgN(38-40)^{high}</i>
<i>TgN(38-40)^{low}</i>	CS38	E12.5 WL	<i>TgN(38-40)^{low}/TgN(38-40)^{low}</i>
	CS40	E12.5 WL	<i>TgN(38-40)^{low}/TgN(38-40)^{low}</i>
	CTCF-left(l)	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(38-40)^{low}/TgN(38-40)^{low}</i>
	3' <i>Btg1</i>	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(38-40)^{low}/TgN(38-40)^{low}</i>
	CTCF-right(l)	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(38-40)^{low}/TgN(38-40)^{low}</i>
	Hi-C	E12.5 WL	<i>Wt/Wt</i>
		E12.5 WL	<i>TgN(38-40)^{low}/TgN(38-40)^{low}; HoxD^{del(CS38-40)/del(CS38-40)}</i>

Supplemental Table S2: Genotypes the 4C-seq and Hi-C samples used for all random transgenic lines. All 4C-seq and Hi-C experiments were performed in a single biological replicate. – indicates either *wild-type* or *TgN(HoxD)* allele (undetermined). (h) or (l) was added at the end of the CTCF-right and CTCF-left viewpoints names to distinguish the viewpoint surrounding the *TgN(38-40)^{high}* integration (h) from the one surrounding the *TgN(38-40)^{low}* integration (l). The *TgN(38-40)^{low}* Hi-C sample was homozygous for a deletion of the endogenous region CS38-40 (*HoxD^{del(CS38-40)}*) to ensure that signals mapping to CS38-40 only arose from the ectopic version of this region. Genotype are colored in the same way than the corresponding tracks of the main figures, if applicable. WL: whole limbs (both forelimbs and hindlimbs). HLE: headless embryos.

<i>HoxD</i>	Island-2	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGCATT CATCAAGCTGTGATTAGCA
		iR	CAAGCAGAAGACGGCATACGAAATCCATAATATGTAGACTGTAGTGTTC
	Island-4	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTTACAG CCTAGTCTTTTCTCATCACAT
		iR	CAAGCAGAAGACGGCATACGATGTAATTATTTTCAGGGTTGGAGTAGAATCA
	<i>Evx2</i>	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGGA AACCTGTGAGCCTAC
		iR	CAAGCAGAAGACGGCATACGAGGGAAGAAACCTACCACGACAC
	<i>Hoxd13</i>	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTAAAT CCTAGACCTGGTCATG
		iR	CAAGCAGAAGACGGCATACGAGGCGATGGTGTGTATAGG
	<i>Hoxd12</i>	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTAACAT CCCAAAGGCTCATG
		iR	CAAGCAGAAGACGGCATACGAGTGGTTACAGAGTCAGATC
	<i>Hoxd11</i>	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGCA TACTTCCTCAGAAGAGGCA
		iR	CAAGCAGAAGACGGCATACGACTAGGAAAATTCCTAATTTTCAGG
	<i>Hoxd9</i>	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTTCACA CTCCCGGGGCGAGCCA
		iR	CAAGCAGAAGACGGCATACGAGTCGGAGAGCCAGAGTCC
	<i>Hoxd4</i>	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGGA CAATAAAGCATCCATAGGCG
		iR	CAAGCAGAAGACGGCATACGATCCAGTGAATTGGGTGGGAT
	CS38	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTTTCCA AGGAGAAAGGTGTTGGTC
		iR	CAAGCAGAAGACGGCATACGACAGGGCGTTGGGTCACTCT
	CS40	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTAACAT TTTCCTGCTTCTTAGTC
		iR	CAAGCAGAAGACGGCATACGAAGCAAGACACAGAGAGATG
<i>TgN(d11-Evx2)-specific</i>	CTCF-left 1	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGGG TTAAGATATCTCCTGA
		iR	CAAGCAGAAGACGGCATACGATAAACAGAGCATTTACGTGATC
	CTCF-left 2	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTTT GGTTAGCATTAATTCCA
		iR	CAAGCAGAAGACGGCATACGAGCTGTGAGAATCAGGTATGA
	<i>Hs3st4</i>	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGTG TGGAAGACTCATG
		iR	CAAGCAGAAGACGGCATACGAGGTTGAAACTGCAGACTTT
	Right 1	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGGG TGTCTTGTCTCAT
		iR	CAAGCAGAAGACGGCATACGATTAGAGCTTGCCGTGTTATT
<i>TgN(38-40)^{high}-specific</i>	CTCF-left(h)	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGTA TAATTGGAGACGCAT
		iR	CAAGCAGAAGACGGCATACGAACTGTATCAACCTCTAGTGTG
	CTCF-right(h)	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGCT GCTCTAAGTCAACA
		iR	CAAGCAGAAGACGGCATACGAAACAACAACTGAGGAAATT
<i>TgN(38-40)^{low}-specific</i>	<i>Man1a2</i>	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTCA GCCTTAGAAATCAAGG
		iR	CAAGCAGAAGACGGCATACGAGCTGGAATGTGGGGTTCCT
	CTCF-left(l)	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTCC AGAATGTCCAGG
		iR	CAAGCAGAAGACGGCATACGATATCAGCCCCAGAGTAGATC
	3'Btg1	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTAATGG TATGACTACTTTAGCCA
		iR	CAAGCAGAAGACGGCATACGAAGAAAAGATAGCTCTCGGC
<i>TgN(38-40)^{low}-specific</i>	CTCF-right(l)	iF	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTACGTT TATAGTCTAATTGGGCA
		iR	CAAGCAGAAGACGGCATACGAATTGTGTTTTTCGGTTGCTTT

Supplemental Table S3: See caption on next page.

Supplemental Table S3: List of all 4C-seq primers used in this study. Illumina Solexa sequencing adapters are indicated in red (long adapter) for the inverse forward (iR) primers, or in blue (short adapter) for the inverse reverse (iR) primers. For some viewpoints, a 4 bp barcode (not represented) was inserted between the adapter and the rest of the primer in order to enable multiplexing of several 4C-seq libraries corresponding to the same viewpoint but belonging to different experiments. (h) or (l) was added at the end of the CTCF-right and CTCF-left viewpoints names to distinguish the viewpoint surrounding the *TgN(38-40)^{high}* integration (h) from the one surrounding the *TgN(38-40)^{low}* integration (l).

Transgenic line	TLA viewpoint		Sequence
<i>TgN(HoxD)</i>	TLA-vector-1	iF	CCTTGCGTATAATATTTGCCC
		iR	GGCGATTCAGGTCATCAT
	TLA-vector-2	iF	AGGATG TTCAGAATGAAACTCA
		iR	GAACTAACCCCCGATATCAG
	TLA- <i>Hoxd13</i>	iF	ATTTTCAAAGGGGTGGGAT
		iR	CCTCCGGGTTTGAATTTGA
	TLA- <i>Hoxd12-2</i>	iF	AAAGGAAAGCCATTGGACTT
		iR	GACTCAGACCCTAACACAAA
<i>TgN(d11-Evx2)</i>	TLA- <i>Evx2</i>	iF	TTCGTGGAGAAATGAGGAAG
		iR	TGACAGCTAACTTTAAACACAA
	TLA- <i>Hoxd12-1</i>	iF	ATCTATCTAGGAAAATTGGCCT
		iR	GAGGAGAGATCCAGAATCCT
<i>TgN(38-40)^{high}</i> and <i>TgN(38-40)^{low}</i>	TLA-CS38	iF	GGGCAGAACTCACTGATTAA
		iR	ATCATTTGGGGATTGCAGAA
	TLA-CS40	iF	GGGACAATAAAGTGGACCTT
		iR	GGGACTCCATCTAGACTTGT
<i>TgN(38-40)^{low}</i>	TLA-5' vector	iF	TCACAGGTATTTATTCGCGA
		iR	TTGTTACACCGTTTTCCATG

Supplemental Table S4: List of all TLA primers used in this study. The transgenic lines that were characterized by using the indicated viewpoints is given on the left. TLA experiments were performed in a single biological replicate. The TLA libraries were prepared with either mixed or individual samples and viewpoints (see Figs. S3, S4, S5 and S8 for details).

n�	Configuration	Supporting uniquely mapped TLA reads				
		Vector-1	Vector-2	Hoxd13	Hoxd12-2	Total
1	chr2:74777999 (telomeric to <i>Hoxd1</i> , end of the BAC clone) -> CCCATTGAATTCCGCGGATCCTCT (vector in 3', +strand)	4095	1185	54	50	5384
2	CTGTCTGTTCTGGAATTCGCGGA (<i>Lnpk</i> , beginning of the BAC clone, -strand) -> pbace3-6:2799 (vector in 5')	312	373	72	0	757
3	ATACAAACAGTGCACCGACATA (Prox, -strand) -> pbace3-6:11391	35	1	0	8	44
4	chr2:74645632 (between <i>Evx2</i> and Prox) -> CTTGCTGGACCCGACATAGATAAT (between <i>Evx2</i> and Prox, +strand, 12 bp duplication)	14	66	0	4	84
5	CTGAATATAGCTGCACCGACATA (telomeric to <i>Hoxd1</i> , 11 kb before the end of the BAC clone, -strand) -> pbace3-6:11391	249	318	36	47	650
6	GAATTCGATATCAAGCTTCCCCGG (Zeocin resistance, +strand) -> chr2:74764308 (second exon of <i>Hoxd1</i>)	95	176	819	239	1329
7	GGTGGATGGGCTTTGCTCCTACTA (similarity to <i>E. coli</i>) -> chr10:28377980 (deduced <i>TgN(HoxD)</i> integration site)	18	50	224	167	459
8	ACCGACGTTGACCGACCGCGTAAC (similarity to <i>E. coli</i>) -> pbace3-6:11303	43	78	331	137	589
9	pbace3-6:11370 -> GTAAGCGGGCATTCTTCTCTG (similarity to <i>E. coli</i>)	49	91	272	190	602
10	TATCAACGCTTATCCCAACAGCT (similarity to <i>E. coli</i>) -> chr2:74691986 (<i>Hoxd10</i> , third codon)	0	20	361	85	466

Supplemental Table S5: Selected examples of TLA-detected hybrid junctions for *TgN(HoxD)*. These junctions were identified using the TLA analysis pipeline presented in Figure S2, with both the mm10 genome and the sequence of the transgene vector (called pbace3-6) as references. They represent situations in which discontinuities within either mm10 or the transgene vector were detected. -> indicates a fusion between sequences of different origins. The origin of the sequences is given between parentheses. The number of uniquely mapped reads (resulting from local mapping, see Fig. S2B) from each TLA viewpoint supporting these configurations is given on the right. The hybrid configuration n 1 reveals a fusion between a sequence belonging to chromosome 2 of mm10 (in this case a sequence of the *HoxD* cluster located on the telomeric side of *Hoxd1*, which precisely matched with end of the BAC clone used to generate the *TgN(HoxD)* construct) and the transgene vector in 3', confirming the preservation of the BAC's extent on the telomeric side. Conversely, configuration n 2 confirmed the preservation of the BAC's extent on the centromeric side. The configurations n 3 to 5 revealed unexpected reorganization events between the BAC clone and the transgene vector (e.g. a junction between Prox and the vector; see configuration n 3), as well as within the BAC clone (e.g. a 12 bp duplication between *Evx2* and Prox; see configuration n 4). Configuration n 6 confirmed the presence of the Zeocin resistance cassette withing the *Hoxd1* gene (see genetically engineered mice strains in Material and Methods and Figure S3A for details). Of particular interest, configuration n 7 matched with the site of chromosome 10 where maximum TLA and 4C-seq coverages were observed (see Fig. S3B), likely reflecting the integration site, although only a sequence with similarity to *E. coli* could be detected at this level. Interspersed bacteria-related sequences were also detected within the transgene vector (see configurations n 8 and 9). At last, the configuration n 10 demonstrated the replacement of the 5' of *Hoxd10* by another exogenous sequence, matching with our observations by ChIPmentation (see Fig. 13A).

Experiment	Genotype	Tissue	Biological replicates
CTCF ChIP-seq	<i>Wild-type</i>	E12.5 PFL	1
		E12.5 DFL	1
		E12.5 WL	1
CTCF ChIPm	<i>Wild-type</i>	E9.5-10 HLE	1
	<i>TgN(HoxD)/-;</i> <i>HoxD^{del(attP-Rel5)/del(attP-Rel5)}</i>	E9.5-10 HLE	3
	<i>TgN(d11-Evx2)/TgN(d11-Evx2);</i> <i>HoxD^{del(attP-Rel5)/del(attP-Rel5)}</i>	E12.5 WL	2
	<i>TgN(38-40)^{high}/TgN(38-40)^{high};</i> <i>HoxD^{del(CS38-40)/del(CS38-40)}</i>	E12.5 WL	2
	<i>TgN(38-40)^{low}/Wt;</i> <i>HoxD^{del(CS38-40)/del(CS38-40)}</i>	E12.5 WL	2
RAD21 ChIP-seq	<i>Wild-type</i>	E12.5 WL	1
RAD21 ChIPm	<i>Wild-type</i>	E9.5-10 HLE	1
	<i>TgN(HoxD)/-;</i> <i>HoxD^{del(attP-Rel5)/del(attP-Rel5)}</i>	E9.5-10 HLE	2
	<i>TgN(d11-Evx2)/TgN(d11-Evx2);</i> <i>HoxD^{del(attP-Rel5)/del(attP-Rel5)}</i>	E12.5 WL	2
	<i>TgN(38-40)^{high}/TgN(38-40)^{high};</i> <i>HoxD^{del(CS38-40)/del(CS38-40)}</i>	E12.5 WL	2
	<i>TgN(38-40)^{low}/Wt;</i> <i>HoxD^{del(CS38-40)/del(CS38-40)}</i>	E12.5 WL	2
H3K27ac ChIP-seq	<i>Wild-type</i>	E13.5 GT	1
		E12.5 DFL	2
H3K27me3 ChIP-seq	<i>Wild-type</i>	E13.5 GT	1
		E12.5 DFL	8
ATAC-seq	<i>Wild-type</i>	E13.5 GT	1
		E12.5 DFL	1

Supplemental Table S6: Biological replicates of the ChIP-seq, ChIPmentation and ATAC-seq experiments. The *wild-type* CTCF and RAD21 E12.5 ChIP-seq data have been published in (Rodríguez-Carballo et al. 2017). H3K27ac, H3K27me3 and ATAC-seq data for E13.5 GT are from (Amândio et al. 2020). H3K27ac and H3K27me3 data for E12.5 DFL were retrieved from (Rodríguez-Carballo et al. 2019). ATAC-seq data for E12.5 DFL were shared by Christopher Chase Bolt. – indicates either *wild-type* or *TgN(HoxD)* allele (undetermined). H3K27ac and H3K27me3 ChIP-seq experiments are colored in the same way than the corresponding tracks of the figures. PFL: proximal forelimbs. DFL: distal forelimbs. WL: whole limbs. HLE: headless embryos. GT: genital tubercles.

EnGen®-compatible oligo	Sequence
EnGen_up_1	<u>TTCTAATACGACTCACTATA</u> G <u>GAGATGGTGAAAATGACCTGG</u> <u>GTTTTAGAGCTAGA</u>
EnGen_up_2	<u>TTCTAATACGACTCACTATA</u> G <u>GAGATTTGTTTGGGAGTGCGG</u> <u>GTTTTAGAGCTAGA</u>
EnGen_1/4_1	<u>TTCTAATACGACTCACTATA</u> G <u>TGGTGCAGGAGATACACAGA</u> <u>GTTTTAGAGCTAGA</u>
EnGen_1/4_2	<u>TTCTAATACGACTCACTATA</u> <u>GACTAGTAAAACAACTTGATG</u> <u>GTTTTAGAGCTAGA</u>
EnGen_1/2_1	<u>TTCTAATACGACTCACTATA</u> G <u>TGGTAATCTAGCAAGCTCTG</u> <u>GTTTTAGAGCTAGA</u>
EnGen_1/2_2	<u>TTCTAATACGACTCACTATA</u> G <u>AGTGCATAGAATCCCCAAAG</u> <u>GTTTTAGAGCTAGA</u>
EnGen_3/4_1	<u>TTCTAATACGACTCACTATA</u> <u>GTGATGGCTAGGTCACAGGAG</u> <u>TTTTAGAGCTAGA</u>
EnGen_3/4_2	<u>TTCTAATACGACTCACTATA</u> G <u>CTAGAAAGAGAAACCACAGA</u> <u>GTTTTAGAGCTAGA</u>
EnGen_pEpi3_1	<u>TTCTAATACGACTCACTATA</u> <u>GCTAGTGATAATAAGTGACTG</u> <u>GTTTTAGAGCTAGA</u>
EnGen_pEpi3_2	<u>TTCTAATACGACTCACTATA</u> <u>GCAAGAAGAAATATCCACCG</u> <u>TTTTAGAGCTAGA</u>
EnGen_down_1	<u>TTCTAATACGACTCACTATA</u> G <u>ATGTGTATTGCAGAAGTCAG</u> <u>GTTTTAGAGCTAGA</u>
EnGen_down_2	<u>TTCTAATACGACTCACTATA</u> G <u>TAAATAAAATTTATAATCTG</u> <u>GTTTTAGAGCTAGA</u>

Supplemental Table S7: List of all EnGen®-compatible DNA oligos used as templates for sgRNA production. The resulting sgRNAs were used to direct Cas9 cleavage in the nCATS experiment of *TgN(38-40)^{low}*. The name of the oligos indicate their approximate position on the *TgN(38-40)* transgene or surrounding regions (up=chromosome upstream of the transgene, 1/4=one quarter into the transgene, 1/2=halfway into the transgene, 3/4=three quarters into the transgene, pEpi3=transgene vector in 3', down=chromosome downstream of the transgene). _1 or _2 indexes are used to distinguish members of sgRNA pairs targeted to approximately the same region. The underlined sequence matches with the target DNA (followed by the PAM sequence, i.e. NGG, in the corresponding genomic region). The sequence highlighted in red is the T7 promoter for sgRNA production, whereas the one highlighted in cyan served as RNA scaffold for the Cas9 enzyme. A "G" has been darkened when not present in the original target sequence. It was added to ensure efficient sgRNA transcription.

Line	Primer	Sequence
<i>TgN(d11-Evx2)</i>	Hs3st4_ISH_Fw	TCAAGAAAGGAGGGACACGC
	Hs3st4_ISH_Rv	CACCCTCTTCCTGTTCCCAC
	Zkscan2_ISH_Fw	AGCAACAGTGGGGTTTGGAA
	Zkscan2_ISH_Rv	ATGGACGGCAGTTAACCAGG
	Nsmce1_ISH_Fw*	GGAGAGCAGGAGCCATGA
	Nsmce1_ISH_Rv*	GGTCAGCCCAGAACGATG
<i>TgN(38-40)^{high}</i>	Wdr3_ISH_Fw	CTTGCTTGCCACGGAAC
	Wdr3_ISH_Rv*	TGTGGCAGTCCCCAAAGT
	Man1a2_ISH_Fw*	GCAGCATCCTGGCTGAGT
	Man1a2_ISH_Rv*	CGCCAGAGAACAGCAGGT
	Gdap2_ISH_Fw*	TCCCAGTGCTTCTCGGTC
	Gdap2_ISH_Rv*	TGGAGTCCTGGGAAATGG
	Tbx15_ISH_Fw	CCCTTCAACTAATAATCAGC
	Tbx15_ISH_Rv	GAAGCCAAGTCCAGGTGTAGC
	Ptgfrn_ISH_Rw*	ATTCTCTCACCTCAGCATAGCC
	Ptgfrn_ISH_Rv*	AGACGAAAGGGAAAGGTAGGTC
<i>TgN(38-40)^{low}</i>	Btg1_ISH_Fw*	CTTTGGGTGGGCTCCTCT
	Btg1_ISH_Rv*	TGGTGTTTGTGGGAAAGA

Supplemental Table S8: List of all PCR primers used for *in situ* probe cloning. Primers indicated with an asterisk were retrieved from the Allen Brain Atlas (<https://portal.brain-map.org/>). All other primers were designed using Primer3Plus (Untergasser et al. 2007) (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>).

Line	Primer	Sequence
<i>TgN(38-40)^{high}</i> and <i>TgN(38-40)^{low}</i>	CTCF-CS38_qPCR_Fw	GCAAAGCTCGACCTACACTTC
	CTCF-CS38_qPCR_Rv	ATTTACAGTGCGAGGGTCTCC
	CS39_qPCR_Fw	TCACGCTGACACTGTGTTTG
	CS39_qPCR_Rv	GCAGGTTACGTAATGTGTGCTG
	CTCF-CS40a_qPCR_Fw	TGGTGGTGAAGAAGCGGTAAG
	CTCF-CS40a_qPCR_Rv	CGACAGCCTTTCCATCTGTTTG
	CTCF-Hoxd4d8_qPCR_Fw	TTCGGTTGTCTGGAGCTTTG
	CTCF-Hoxd4d8_qPCR_Rv	AGAAGAACGGGTGGCTCTTGTG
	CTCF-Hoxd8d9_qPCR_Fw	TCAGCGGTCCAAACCCAAGTCA
	CTCF-Hoxd8d9_qPCR_Rv	CTGCACCACAACGCTAGCTTTAC
	Aldh1a2_qPCR_Fw	CTCTTTGGCCATAAACGTTCCC
	Aldh1a2_qPCR_Rv	TAAAGTGACCGAGCAAGCAC

Supplemental Table S9: List of all qPCR primers used for the transgene quantifications of *TgN(38-40)^{high}* and *TgN(38-40)^{low}* shown in Figures S6 and S9, respectively.