Sequential in *cis* mutagenesis in vivo reveals various functions for CTCF sites at the mouse *HoxD* cluster

Ana Rita Amândio,[1,2,5] Leonardo Beccari,[2,4,5] Lucille Lopez-Delisle,[1] Bénédicte Mascrez,[2] Jozsef Zakany,[2] Sandra Gitto,[2] and Denis Duboule[1,2,3]

[1]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland; [2]Department of Genetics and Evolution, University of Geneva, 1211 Geneva, Switzerland; [3]Collège de France, 75231 Paris, France

Mammalian *Hox* gene clusters contain a range of CTCF binding sites. In addition to their importance in organizing a TAD border, which isolates the most posterior genes from the rest of the cluster, the positions and orientations of these sites suggest that CTCF may be instrumental in the selection of various subsets of contiguous genes, which are targets of distinct remote enhancers located in the flanking regulatory landscapes. We examined this possibility by producing an allelic series of cumulative in *cis* mutations in these sites, up to the abrogation of CTCF binding in the five sites located on one side of the TAD border. In the most impactful alleles, the global chromatin architecture of the locus was modified, yet not drastically, illustrating that CTCF sites located on one side of a strong TAD border are sufficient to organize at least part of this insulation. Spatial colinearity in the expression of these genes along the major body axis was nevertheless maintained, despite abnormal expression boundaries. In contrast, strong effects were scored in the selection of target genes responding to particular enhancers, leading to the misregulation of *Hoxd* genes in specific structures. Altogether, while most enhancer–promoter interactions can occur in the absence of this series of CTCF sites, the binding of CTCF in the *Hox* cluster is required to properly transform a rather unprecise process into a highly discriminative mechanism of interactions, which is translated into various patterns of transcription accompanied by the distinctive chromatin topology found at this locus. Our allelic series also allowed us to reveal the distinct functional contributions for CTCF sites within this *Hox* cluster, some acting as insulator elements, others being necessary to anchor or stabilize enhancer–promoter interactions, and some doing both, whereas they all together contribute to the formation of a TAD border. This variety of tasks may explain the amazing evolutionary conservation in the distribution of these sites among paralogous *Hox* clusters or between various vertebrates.

Embryonic development relies on complex and precise dynamics of gene activation and repression, driven in large part by the combined activity of multiple *cis*-regulatory elements (CREs) (Spitz and Furlong 2012; Long et al. 2016). In vertebrates, CREs can be located at long distances from their target genes and interact with them through the establishment of particular chromatin structures such as loops. Often, the same genomic region harbors multiple regulatory elements and several transcription units, raising the question of how specific enhancer–promoter interactions can be established without affecting neighboring genes.

The advent of chromosome conformation capture (3C) technologies (Dekker 2006) confirmed that the eukaryote genome is organized into several levels of folding with, at the megabase level, chromatin domains referred to as topologically associating domains (TADs) (Dixon et al. 2012; Nora et al. 2012; Sexton et al. 2012). TADs are domains where DNA sequences such as promoters and their enhancers interact more frequently than with regions located outside, independently from the linear distance (Dixon et al. 2012, 2016), and may thus constitute structural units in the organization of genomes associated

with particular functional tasks. Indeed, complex regulatory landscapes spread over large distances often match TADs (e.g., see Andrey et al. 2013). Even though a causal relationship remains to be fully clarified, TADs are thus thought to delimit functionally autonomous regions, somewhat channeling the activity of distal CREs (Sikorska and Sexton 2020) by reducing the search space between the enhancers and their promoters (Symmons et al. 2016). Accordingly, these domains tend to be evolutionarily conserved within large syntenic regions (Dixon et al. 2012; Krefting et al. 2018). In support of this view, the disruption of TAD borders was shown to lead to their loss of insulation and concurrent misregulation of genes due ectopic gene–enhancer interactions (Gómez-Marín et al. 2015; Lupiáñez et al. 2015; Ibn-Salem et al. 2017; Rodríguez-Carballo et al. 2017). In agreement, genomic rearrangements whereby TAD boundaries are placed between regulatory elements and their targets genes result in the down-regulation of the target genes along with TAD reorganization (Lupiáñez et al. 2015; Kraft et al. 2019; Willemin et al. 2021). However, how tissue- or gene-specific contacts can be established within one TAD is still elusive, in particular how distinct sets of enhancer sequences can interact with various subsets of transcription units and not others, in different cell types, while all located in the same chromatin domain.

Among those proteins that contribute to the establishment of the nuclear 3D chromatin organization, the CTCF zinc finger transcription factor plays an important role. It was initially described as a negative regulator of gene expression due to its capacity to repress transcription by blocking enhancer–promoter interactions, thus defining a category of CREs referred to as insulators (Lobanenkov et al. 1990; Chung et al. 1993; Bell et al. 1999; see Herold et al. 2012). CTCF recognizes a conserved GC-rich 20-nt-long consensus sequence (Renda et al. 2007; Nakahashi et al. 2013; Yin et al. 2017) and mediates loop formation in conjunction with cohesin, a protein complex with chromatin extruding activity (Merkenschlager and Odom 2013; Hansen et al. 2017; Sedeño Cacciatore and Rowland 2019). In the "loop extrusion" model, cohesin is loaded into the chromatin, where it forms a ring-shaped structure that moves along and progressively extrudes the DNA fiber until reaching CTCF-occupied sites where the CTCF N-terminal portion faces cohesin progression (Sanborn et al. 2015; Fudenberg et al. 2016; Stigler et al. 2016; Hansen et al. 2017; Davidson et al. 2019; Kim et al. 2019; Pugacheva et al. 2020; Xi and Beer 2021).

Accordingly, the orientation and location of the CTCF binding sites (CBSs) play a critical role in DNA loop formation and high-order chromatin organization. In fact, TAD and sub-TAD boundaries are enriched in CBSs, with usually several CTCF motifs displaying the same orientation, facing those sites located at the other extremity of the TAD (Kentepozidou et al. 2020; Huang et al. 2021). Despite their prominent role in the establishment of TAD boundaries, most CBSs are found outside these regions and are associated with a wide range of functions, including enhancer–promoter interaction, imprinting, and re-

combination (Guo et al. 2011; Phillips-Cremins et al. 2013; Franco et al. 2014; Gosalia et al. 2014). In agreement with this multifaceted role, CTCF depletion in the embryo resulted in the concomitant loss of TAD insulation and weakening of gene–enhancer interactions, sometimes with clearly documented effects upon gene transcription (e.g., see Paliou et al. 2019), whereas in other instances a more moderate and somewhat unpredictable impact was observed (Soshnikova et al. 2010; Nora et al. 2017; Luan et al. 2021). However, the potential importance of CTCF in helping tissue-specific enhancers to select the right promoter(s) and thus activate a subset of genes located within the same TAD remains to be assessed with precision. In this context, *Hox* gene clusters provide an excellent experimental paradigm. Indeed, besides their critical function in the organization of the major body axis (Wellik 2009), these genes are highly pleiotropic, as they are involved in the making of a range of organs and structures at various developmental times (e.g., see Krumlauf 1994). This is exemplified by the limbs, the external genitals, the urogenital and gastrointestinal tractus (Favier and Dollé 1997; Zakany and Duboule 2007; Mallo et al. 2010; Deschamps and Duboule 2017), and other endodermal organs, or teguments such as hair (Godwin and Capecchi 1998) and nails (Fernandez-Guerrero et al. 2020). The various enhancers necessary to achieve these widely diverse regulations are positioned on either side of the clusters and have been characterized in some details, in particular at the mouse *HoxD* locus.

This locus, which includes nine genes within an ~100-kb DNA segment, is positioned in between two large regulatory domains matching TADs (C-DOM and T-DOM) and contains in itself a strong chromatin boundary, a structure that leads to a differential tropism in enhancer–promoter interactions (Fig. 1A; Noordermeer et al. 2011; Andrey et al. 2013; Darbellay and Duboule 2016; Rodríguez-Carballo et al. 2017). While the "posterior" genes *Hoxd13* and *Hoxd12* mainly contact the C-DOM, the "anterior" part of the cluster (from *Hoxd1* to *Hoxd9*) preferentially interacts with the T-DOM, with *Hoxd10* and *Hoxd11* being more versatile in their interaction potential. This internal chromatin boundary is induced by the presence of a collection of nine CTCF sites, with an inversion of polarity in the middle, which positions the terminal genes *Hoxd12* and *Hoxd13* in a TAD (C-DOM) that is distinct from that containing the rest of the cluster (T-DOM) (Fig. 1C). While C-DOM contains several enhancers necessary to produce "terminal" structures, the hands and feet, and the genitals (Montavon et al. 2011; Amândio et al. 2020), T-DOM includes a range of enhancer sequences specific for various structures such as the proximal limbs, part of the intestines, mammary glands, or various head structures.

For each of these regulations, a distinct although overlapping set of contiguous *Hoxd* genes is selected as targets. In every case, the subgroup of genes responding to a given regulation is delimited by pairs of CTCF sites showing the same orientation, toward the TAD where the related enhancers are localized (Fig. 1). For example, digit enhancers located within C-DOM control *Hoxd13*
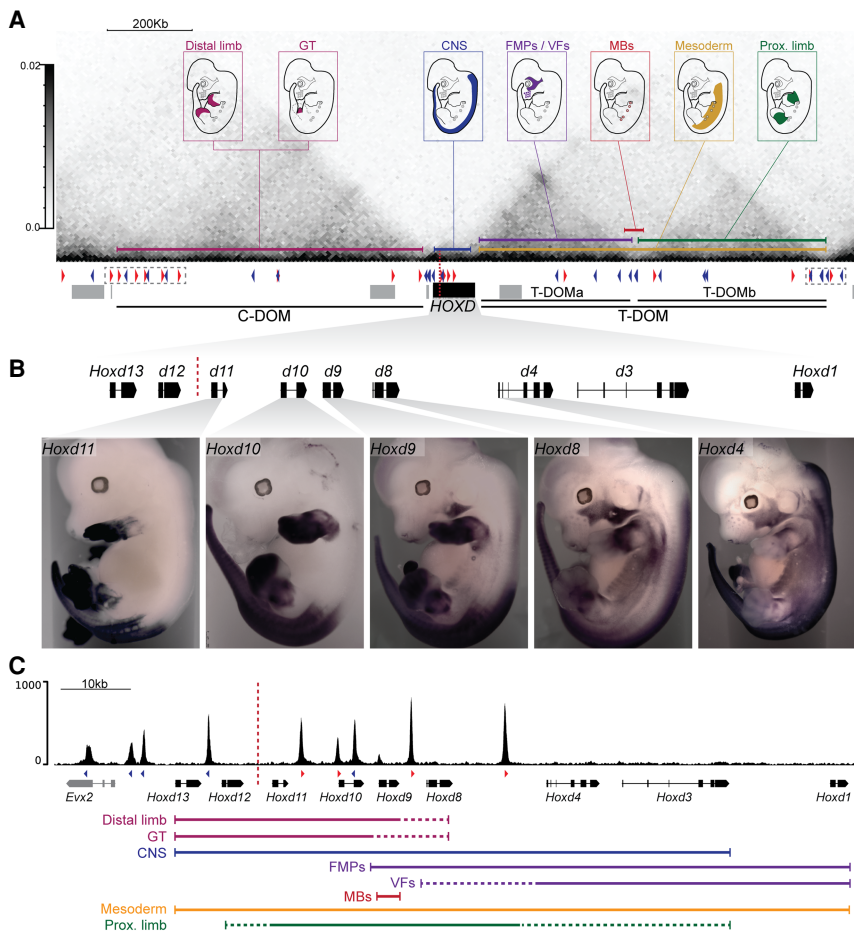
Amândio et al.

Figure 1. Distribution of CTCF binding sites at the *HoxD* locus and pleiotropic regulation. (*A*) Capture Hi-C profile of E9.5 trunks covering the *HoxD* cluster and flanking TADs (T-DOM and C-DOM; mm10: chr2: 73779626–75669724). The embryos on *top* illustrate various sites of *Hoxd* gene expression and are linked to colored horizontal bars indicating the DNA segments where the related enhancer sequences are positioned, either within the cluster itself (black box) or in the flanking TADs. Dashed gray boxes highlight dense arrays of CBSs at TAD borders. (GT) Genital tubercle, (CNS) central nervous system, (FMPs) facial muscle progenitors, (VFs) vibrissae follicles (whisker pads), (MB) mammary buds. The positions and orientations (red or blue) of CTCF binding sites (triangles) are indicated *below* the capture Hi-C heat map. Surrounding genes are shown as filled gray boxes. (*B*) WISH analysis showing the expression of *Hoxd11, Hoxd10, Hoxd9, Hoxd8,* and *Hoxd4* in E12.5 wild-type embryos to highlight the various overlapping expression patterns for subsets of these genes across distinct embryonic structures. (*C*) CTCF ChIP-M profile at *HoxD* using E10.5 trunks (mm10: chr2: 74650810–74767377). CBSs are shown as red and blue arrowheads as in *A*, with red for a telomeric orientation (pointing toward T-DOM) and blue for a centromeric orientation (pointing toward C-DOM). The dashed red line separates the two arrays of CBSs showing divergent orientations. The colored lines at the *bottom* indicate the subsets of *Hoxd* genes expressed in any given embryonic structures. The colors correspond to those delineating the related enhancer domains in *A*.

to *Hoxd10,* whereas forearm enhancers present in T-DOM mostly contact the *Hoxd11* to *Hoxd9* DNA interval. In both cases, the corresponding H3K27ac profiles over these different sets of target genes are delimited by different pairs of occupied CTCF sites (Rodríguez-Carballo et al. 2017).

This dense series of CTCF sites is spread ∼50–60 kb, with sites mostly found in between transcription units, a distribution that appeared conserved among the four *Hox* mammalian clusters as well as between tetrapod species (Yakushiji-Kaminatsui et al. 2018). This is suggestive of a strong selective pressure to maintain such an organization and thus of potentially important functions for these sites. Indeed, previous studies have revealed the role of individual CBSs in determining microboundaries, as exemplified with the *HoxA* and *HoxC* clusters (Luo et al. 2006; Narendra et al. 2015, 2016; Ghasemi et al. 2021; Su et al. 2021). In particular, Narendra et al. (2015, 2016) showed that the deletion of such sites would modify the extent of *Hox* genes expressed during the formation of the major body axis (in particular in motoneuron cultures) and hence that these sites may behave as insulator elements between neighboring genes. However, besides the

formation of the main body axis, the potential function of CTCF sites in the complex interactions between the large flanking regulatory landscapes and the various subsets of target genes remains to be determined.

In this study, we used a cumulative in *cis* CRISPR/Cas9 genome-editing strategy to disrupt the five CBSs located on one side of the TAD boundary; i.e., within the anterior and central part of the *HoxD* gene cluster. We report the analysis of mouse lines either carrying single mutated sites or the full series of combined mutations. This progressive allelic series starts with the most "anteriorly" located (closer to T-DOM) CBS and follows with the first two, three, four, and five CBSs in *cis*, the latter combination removing all those CTCF sites located on the telomeric side of the TAD border. We analyzed the impact of these various mutations both on *Hoxd* gene expression and chromatin architecture across different tissues, and described associated patterning defects along the major body axis. We conclude that CTCF sites within *Hox* clusters are important for the capacity of remote enhancers to select subgroups of target genes. However, not all CTCF sites share the same functional task and, while some sites appear to behave as insulators, preventing a gene to

respond to a particular enhancer, others seem to be necessary for the opposite function by allowing a particular gene to respond to a given regulation. Notably, some CBSs can display both activities in different tissues. Also, while the selective removal of all CTCF and RAD21 binding on one side of the TAD border certainly resulted in increased inter-TAD interactions, a TAD boundary was still clearly present, indicating that the series of remaining CTCF sites with the opposite orientation was sufficient to maintain the opposite tropism in enhancer–promoter interactions, even though it was weaker and less precise.

Results

Evolutionarily conserved CTCF sites in Hox clusters

During gastrulation, *Hox* genes are transcribed along the neural tube and paraxial and lateral mesoderm with expression boundaries that reflect their respective positions within the cluster (Fig. 1; Gaunt et al. 1988). Subsequently, various subsets of these genes are transcribed across a number of embryonic structures, as exemplified by the developing limbs where digit and forearm enhancers, located in opposite TADs, regulate partially overlapping subgroups of *Hoxd* genes (Fig. 1A; Andrey et al. 2013). In another context, *Hoxd1* to *Hoxd4* and *Hoxd1* to *Hoxd9* are expressed in the emerging vibrissae follicles (VFs) and in facial muscle progenitors (FMPs), respectively, driven by enhancers located within the T-DOM (Fig. 1A–C; Hintermann et al. 2021). Alternatively, a single gene can display one particular functionality such as *Hoxd9*, which is the only gene expressed in the mesenchymal condensates of the future embryonic mammary glands (Chen and Capecchi 1999), a specificity controlled by an enhancer also located within T-DOM, at the boundary between two sub-TADs: T-DOMa and T-DOMb (Fig. 1A–C; Schep et al. 2016).

To assess whether bound CTCF could be instrumental in the selection of distinct target promoters by such remote enhancers, we looked at CTCF occupancy in 10.5 embryonic trunks over the 2-Mb *HoxD* landscape, together with global chromatin interaction profiles derived from a capture Hi-C approach of E9.5 embryonic trunks (Fig. 1A,C). We scored nine occupied CTCF sites in the *HoxD* cluster and over the immediately adjacent *Evx2* gene (Fig. 1C), and two dense arrays of nine and seven CTCF peaks at the opposite borders of the C-DOM and T-DOM, respectively (Fig. 1A, dashed boxes). In contrast, CTCF peaks were less densely distributed within the latter two regulatory landscapes, although with a higher number in T-DOM than in C-DOM. Also, a comparison of CTCF occupancy across several embryonic structures revealed that CTCF binding over the entire *HoxD* genomic landscape was comparable in all tissues analyzed (Supplemental Fig. S1).

Motif analysis revealed that the orientations and positions of the various CBSs tightly correlated with the observed topology of interactions across these domains (Fig. 1A). For instance, within the sub-TAD T-DOMa,
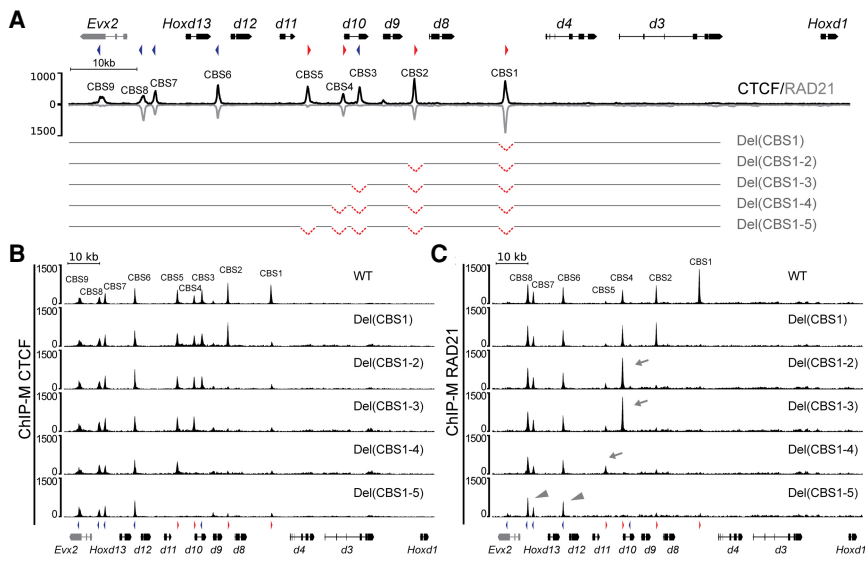
two divergent CBSs delimit a domain preferentially interacting with the *Hoxd1* gene, while the sub-TAD boundary region is enriched in CBSs with a negative orientation, defining the extent of the *Hoxd3–Hoxd8* preferential interactions (Fig. 1A). Furthermore, the CBSs located either in the center of C-DOM or within T-DOMa and T-DOMb correlate well with the extent of local interaction domains observed within their respective higher-order structures (Fig. 1A).

Within the *HoxD* cluster itself, CBSs are arranged in two arrays of motifs with divergent orientations. All but one CBS within the anterior and central portion of the cluster (from *Hoxd1* to *Hoxd11*) display a forward orientation and face convergent CBSs located in T-DOM, which are for the most part present in the reverse orientation. Instead, CBSs mapping between *Hoxd12* and *Evx2* have a reverse orientation coinciding with their interaction with CBSs located within or at the centromeric border of C-DOM (Fig. 1C). This inversion of CBS orientations within the *HoxD* cluster expectedly matches the TAD boundary in many cell types, even though this boundary was shown to slightly shift over the *Hoxd11* to *Hoxd10* genes in particular circumstances (Fig. 1; Andrey et al. 2013). Of note, this organization of the mouse *HoxD* cluster is largely maintained among the four *Hox* mammalian clusters (Supplemental Fig. S2), as well as in both the human and chicken orthologous *Hoxd* gene clusters (Supplemental Fig. S3), and is even similar to what was reported in squamates (Guerreiro et al. 2016). Therefore, the presence, distribution, and orientation of CBSs within *Hox* clusters are globally conserved across tetrapods, supporting the idea that they importantly contribute to some aspects of *Hox* gene regulation during development.

In addition, the analysis of RAD21 occupancy, a structural component of the cohesin complex, (Cheng et al. 2020 and references therein) revealed a substantial enrichment in several of these CBSs, either within the *HoxD* cluster or in the flanking TADs. Among the former, RAD21 accumulation was maximal at CBS1, CBS2, and CBS4, as well as in CBS6 to CBS8. Instead, weak or no RAD21 enrichment was observed in the centrally located CBS3 and CBS5 or at CBS9 (Fig. 2A). While this pattern of accumulation of RAD21 is compatible with the corresponding CTCF sites being involved in long-range interactions through loop formation, the heterogeneity in the RAD21 profile, at least in this tissue, suggests that not all occupied CTCF sites may share the exact same function.

An in cis allelic series for mice mutant for CTCF binding

To investigate the function of these CBSs either in the organization of the TAD boundary or during the implementation of remote enhancers, we generated an allelic series of mutant mice carrying homozygous microdeletions of all CBSs located on the T-DOM side of the TAD border (Fig. 2A, CBS1 to CBS5; Supplemental Fig. S4). We designed sgRNAs targeting the various CBSs identified within the ChIPmentation (ChIP-M) CTCF peaks (see the Materials and Methods; Supplemental Fig. S4) and

Amândio et al.

Figure 2. The CTCF in *cis* mutant allelic series. (*A*) CTCF (black) and RAD21 (gray) ChIP-M profiles at control and mutant *HoxD* loci, using dissected E10.5 trunks (mm10: chr2: 74650810–74767377). The CTCF binding site (CBS) number is indicated *above* the peaks with its orientation on *top*. A schematic representation of the CBS deletion alleles is shown *below*, only indicating the combined mutants in *cis*. Single CBS mutations are not shown except for Del(CBS1). (*B,C*) CTCF (*B*) and RAD21 (*C*) ChIP-M enrichments at the *HoxD* locus in E10.5 trunks of either control (WT) or Del (CBS1), Del(CBS1-2), Del(CBS1-3), Del (CBS1-4), or Del(CBS1-5) homozygous embryos. No significant difference in enrichment over the nonmutated CBSs was scored and no cryptic binding site was revealed. For RAD21, gray arrows indicate distinctive changes in RAD21 accumulation in some mutant alleles, whereas arrowheads highlight the stable enrichment of RAD21 over C-DOM-oriented CBSs (blue triangles), independently of any mutation (mm10: chr2: 74650810–74767377).

coelectroporated them with the Cas9 mRNA in fertilized mouse oocytes. After control by DNA sequencing, F0 animals were crossed to produce mutant lines (Fig. 2A; Supplemental Fig. S4). In this way, we generated from 6- to 78-bp-long microdeletions impacting the predicted CBSs (Supplemental Fig. S4B) and affecting those nucleotides required for CTCF binding (Lobanenkov et al. 1990; Hashimoto et al. 2017).

To produce the series of microdeletions in *cis*, we first obtained separately animals homozygous for the individual deletion of the CBS1, referred to as $HoxD^{Del(CBS1)-/-}$ or Del(CBS1), and the CBS2 [$HoxD^{Del(CBS2)-/-}$ or Del (CBS2)]. We then electroporated the sgRNA targeting CBS1 into zygotes heterozygous for Del(CBS2) and recovered the double mutant in *cis* $HoxD^{CBS(1-2)-/-}$ or Del (CBS1-2). This operation was reiterated three times by using the various sgRNAs on the newly produced strains carrying mutations in *cis* to eventually obtain the $HoxD^{CBS(1-5)-/-}$ mice or Del(CBS1-5), where the five contiguous CTCF sites located on the T-DOM side of the TAD boundary were mutated on the same chromosome. For each mutant and before processing to the next mutation in *cis*, we assessed CTCF binding by ChIP-M in the postoccipital region of E10.5 wild-type and homozygous mutant embryos. As expected, these mutations mostly abolished CTCF binding to the target sites (Fig. 2B). Of note, and in contrast to what was reported for CBS mutagenesis in other genomic contexts (Narendra et al. 2016; Paliou et al. 2019), we did not observe any cryptic CTCF binding site, which would have been revealed after the mutation of neighboring sites. Also, the binding enrichments observed at the remaining CTCF sites were not modified (Fig. 2B).

The effect of these microdeletions on CTCF binding was controlled by evaluating the accumulation of RAD21 using ChIP-M (Fig. 2C; Supplemental Fig. S5).

As expected, the disruption of the CTCF motifs resulted in the loss of RAD21 at the targeted CBSs. Furthermore, in the mutant conditions, RAD21 was redistributed and increased accumulations were observed at CBSs located next to the deleted site(s). For example, RAD21 accumulation was increased at CBS4 in the Del(CBS1-2) mutant, an increase reinforced in the Del(CBS1-3) allele (Fig. 2C, arrows; Supplemental Fig. S6). Changes in RAD21 enrichment were also observed at CBS5, a CTCF site that displayed virtually no accumulation of RAD21 in the wild-type condition (Fig. 2C, arrow; Supplemental Fig. S6). However, no difference in RAD21 enrichment was observed either at CBS6 to CBS9 or at CBS3; all of these CTCF sites displaying a reverse motif orientation (Fig. 2C, arrowheads; Supplemental Fig. S6) further supports an involvement of CBS1, CBS2, CBS4, and CBS5 in forming chromatin structures with CTCF sites located further telomeric, within T-DOM.

Impact of CTCF binding sites deletions in *cis* upon chromatin topology

We assessed the impact of these progressive deletions of CBSs upon the global TAD architecture by performing capture Hi-C over the region covering the *HoxD* cluster and flanking TADs. We used trunks of E9.5 control embryos where *Hox* genes are transcribed and compared the interaction profiles with those obtained from Del (CBS1), Del(CBS1-3), and Del(CBS1-5) homozygous fetuses. The *HoxD* cluster is positioned between the C-DOM and T-DOM TADs, which are insulated from one another by the CTCF-dependent boundary present within the gene cluster (Fig. 3A; Rodríguez-Carballo et al. 2017). In this embryonic material, the TAD separation score using the hicFindTADs algorithm identified the position of this
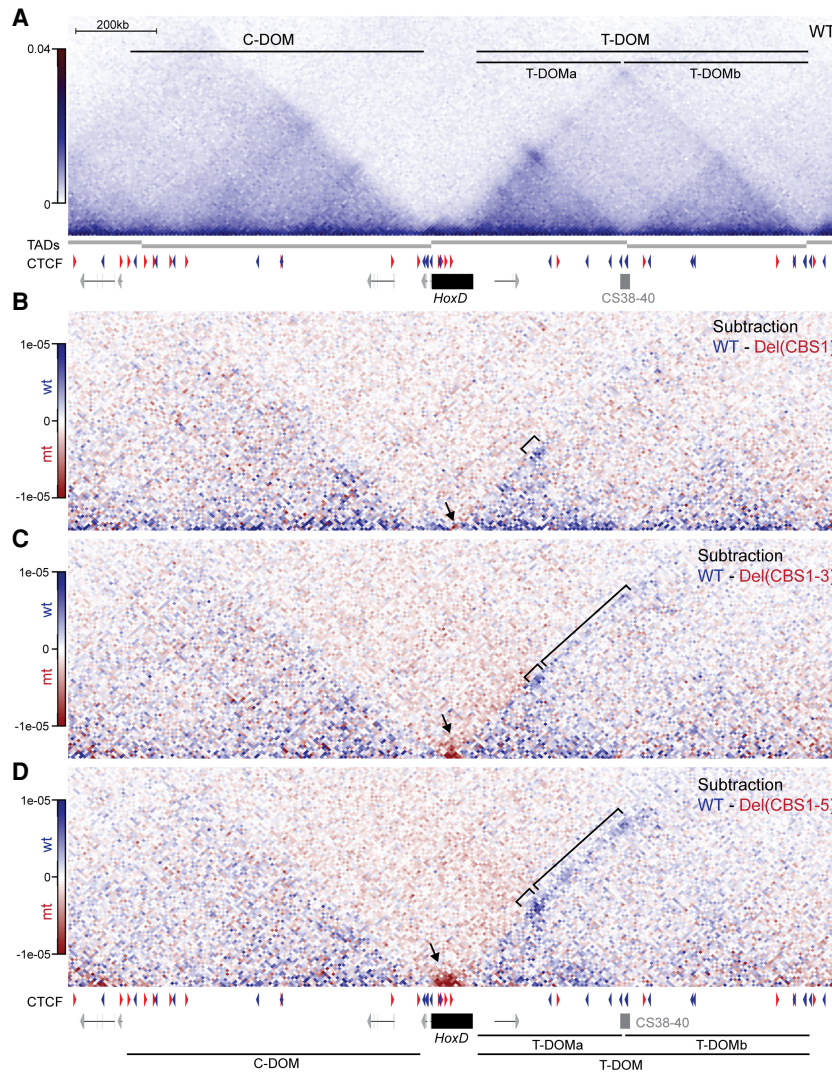
Figure 3. Changes in chromatin topology in CBS mutants in vivo. (*A*) Capture Hi-C map of control (WT) E9.5 trunks with the *HoxD* cluster (black rectangle) and flanking TADs (C-DOM and T-DOM; mm10: chr2: 73779626–75669724). TADs or sub-TADs (T-DOMa and T-DOMb) were identified using the hicFindTADs algorithm and are represented by gray bars *below* the heat map. Red and blue arrowheads *below* indicate the orientation of the CBSs. The sub-TAD boundary region CS38-40 is shown as a gray box. (*B–D*) The subtraction maps between the cHi-C profiles from control versus Del(CBS1), Del(CBS1-3), and Del (CBS1-5) homozygous embryos are displayed, with blue bins pointing to chromatin interactions that are more prevalent in control cells, while red bins represent interactions enriched in mutant alleles. The black arrows indicate the increase of intracluster interactions in mutant alleles. The small brackets point to a progressive decrease in interactions between the gene cluster and the CS38-40 region in the mutant alleles, whereas the large brackets indicate the loss of contact frequency between the *HoxD* cluster and T-DOMb in both the Del (CBS1-3) and Del(CBS1-5) alleles. C-DOM, T-DOM, T-DOMa, and T-DOMb are represented *below* as in *A* (mm10: chr2: 73779626–75669724).

boundary between *Hoxd13* and *Hoxd12* (Fig. 3A; Supplemental Fig. S7C).

For the three mutant conditions analyzed, subtractions were carried out from the control interaction profile and are displayed with gained interactions in red and lost interactions in blue (Fig. 3B–D). The abrogation of CBS1 resulted in two slight yet significant changes: First, an increase of self-interactions was observed between the *Hoxd1* and *Hoxd8* genes (Fig. 3B, arrow; Supplemental Fig. S7A,B) without altering the position of the TAD border (Supplemental Fig. S7C). Second, a decrease in interactions was scored between the gene cluster and the CS38-40 region (Fig. 3B, bracket), a region containing three CTCF sites oriented toward the *HoxD* cluster, which acts as a sub-TAD boundary within the T-DOM (Fig. 3A). An average signal quantification and virtual capture-C profiles using *Hoxd4* as a viewpoint confirmed this reduction (Supplemental Fig. S8A–C).

These two differences were strongly reinforced when the capture Hi-C profile of the Del(CBS1-3) was subtracted from the control counterpart. Indeed, a marked increase in

intracluster interactions was scored, which extended up to *Hoxd10* (Fig. 3C, arrow; Supplemental Fig. S7A,B). Furthermore, the TAD separation score analysis revealed a change in TAD border position that was now called at a more centromeric position, after the *Evx2* gene (Fig. 3C; Supplemental Fig. S7C). In addition, the frequencies of interactions between the gene cluster and T-DOM not only showed a loss in contacts with the CS38-40 region (Fig. 3C, small bracket), but also throughout the most telomeric sub-TAD T-DOMb (Fig. 3C, large bracket; Supplemental Fig. S8A,B). To facilitate data visualization, we generated virtual capture-C profiles using different viewpoints. This analysis showed a loss of interactions between the *Hoxd4* promoter and the CS38-40 region, as well as a reduction in contacts between the cluster and the T-DOMb, when using either *Hoxd9* or the T-DOMb 3′ border as viewpoints (Supplemental Fig. S8C).

This effect was again enhanced in the Del(CBS1-5); i.e., after the deletion of all CTCF sites oriented toward the T-DOM (Fig. 3D). The intracluster interactions were strengthened and extended up to *Hoxd13* (Fig. 3D, arrow;

Supplemental Fig. S7A,B), with a TAD boundary being called after the *Evx2* gene (Fig. 3D; Supplemental Fig. S7C). In addition, the interactions with the CS38-40 region were further lost when compared with the Del(CBS1-3) condition. Similarly, the decrease in global contact frequency between the *HoxD* cluster and the sub-TAD T-DOMb was further enhanced (Fig. 3D, large bracket; Supplemental Fig. S8A,B), with a loss of interactions even observed with the T-DOM telomeric TAD border. These results were confirmed by virtual capture-C using various viewpoints (Supplemental Fig. S8C). Therefore, the progressive deletions of intracluster CTCF sites clearly released long-range interactions between the gene cluster itself and various parts of T-DOM, where the cluster is normally anchored. Concomitant to this lack of long-range contacts, the cluster became more compact and thus increased its local interactions. Finally, none of the mutant interaction profiles showed any major differences in the contacts between the gene cluster and the C-DOM TAD.

Impact of alterations in chromatin structure upon Hoxd *genes expression*

T-DOM contains several enhancers necessary for the expression of *Hoxd* genes. Therefore, we assessed whether these alterations in contact distribution observed in the various CBS deletion alleles were paralleled by modifications in gene expression. We performed whole-mount RNA in situ hybridization (WISH) across our mutant lines and examined gene expression along the main body axis. While expression in the spinal cord appears to be regulated by intracluster control sequences (e.g., see Tschopp et al. 2012), transcription in mesoderm derivatives is at least partly regulated by elements located within T-DOM (Fig. 4A, brown). In Del(CBS1) mutant fetuses, we observed an anteriorization of the expression domain of *Hoxd8,* the gene positioned just 5′ of the deleted CTCF site (Fig. 4B, black arrowhead), whereas the expression domains of both *Hoxd4* and *Hoxd9* remained unaltered (Supplemental Fig. S9A,B). In the Del(CBS1-2) allele, we observed an anteriorization of the domain of expression of *Hoxd9* (Fig. 4B, black arrowhead) without affecting the expression of *Hoxd10* (Supplemental Fig. S9C). In turn, the domain of expression of the latter gene was anteriorized in Del(CBS1-3) mutant fetuses (Fig. 4B, black arrowhead), while *Hoxd11* transcripts remained indistinguishable from control littermates (Supplemental Fig. S9D). In both Del(CBS1-4) and Del(CBS1-5) homozygous embryos, both *Hoxd10* and *Hoxd11* domains were anteriorized as well (Fig. 4B, black arrowhead; Supplemental Fig. S9F). Although to a lesser extent, the expression territories of more anterior genes were still affected in these alleles, as exemplified by *Hoxd9* staining in the Del(CBS1-4) (Supplemental Fig. S9E).

Changes in axial patterning due to alterations in the expression of *Hox* genes have been widely described (see Kessel and Gruss 1991; Krumlauf 1994; Carapuco et al. 2005). We performed micro-CT scans of control and homozygous mutant adult mouse skeletons and expectedly observed alterations in the spine of mutant animals, asso-
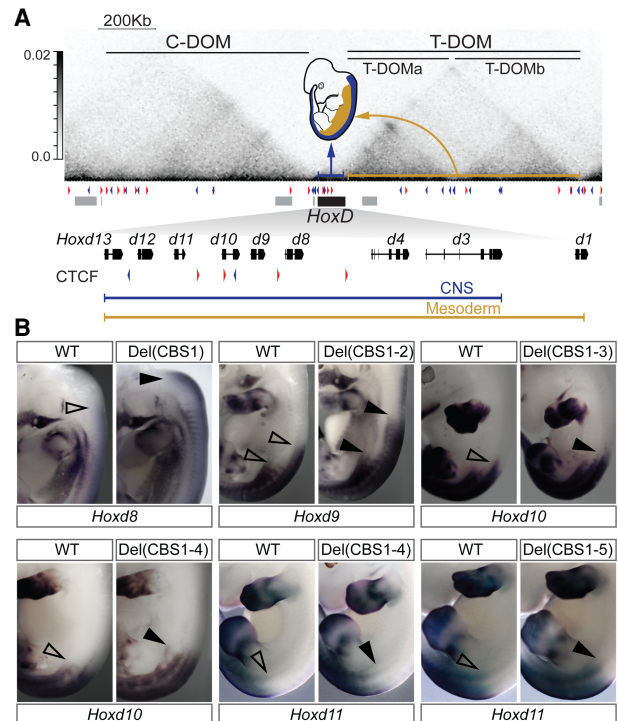

Figure 4. Alterations in *Hoxd* gene expression along the anterior-to-posterior axis. (*A*) cHi-C map from control E9.5 dissected trunk region showing the *HoxD* cluster and neighboring regulatory domains T-DOM and C-DOM (mm10: chr2: 73779626–75669724). The schematized embryo *above* the heat map represents the regulation of *Hoxd* gene expression in the spinal cord (CNS; blue) by enhancer elements located within the gene cluster (blue arrow), whereas gene transcription in axial mesoderm (yellow) is in large part controlled by regulatory elements located in T-DOM (yellow arrow). The positions of the *HoxD* cluster (black) and the surrounding genes (gray) are shown *below*. At the *bottom*, a zoom-in of the *Hoxd* genes and CBS is shown. The blue and yellow lines underline which *Hoxd* genes are expressed either in the CNS or in the mesoderm, respectively. (*B*) WISH analysis showing the expression of different *Hoxd* genes in the Del(CBS1), Del(CBS1-2), Del(CBS1-3), Del(CBS1-4), and Del(CBS1-5) E12.5 embryos with corresponding control littermates. An anteriorization of expression in the trunk was generally observed for the gene positioned immediately 5′ to the deleted CBS. Black arrowheads indicate an anterior shift of the expression domain of the displayed gene in mutant alleles. Hollow arrowheads indicate the rostral limits of these expression domains in control embryos.

ciated with changes in expression pattern (Supplemental Table S1). For example, 36% of Del(CBS1-4) mutant mice have four lumbar vertebrae (L4), instead of the L5 condition observed in this genetic background (Supplemental Table S1). Such lumbosacral defects were previously associated with alterations in *Hoxd11* expression (Gerard et al. 1996; Zákány et al. 1996; Wellik and Capecchi 2003). It is noteworthy that this morphological alteration was both less severe and less prevalent in Del(CBS1-5) animals (Supplemental Table S1).

More than half of the mutants carrying any of the three genotypes analyzed displayed defects at the atlanto–

occipital junction, involving the basioccipital bone, the atlas, and the axis (Supplemental Table S1). A minor proportion of both Del(CBS1-3) and Del(CBS1-4) mutant animals also showed the asymmetrical presence of cervical ribs, a malformation at the cervicothoracic transition (Supplemental Table S1). These results confirmed the physiological relevance of CTCF-mediated regulation of *Hoxd* gene expression along the A-P body axis and are in agreement with previous work involving the *HoxA* and *HoxC* clusters, where the local removal of CTCF led to such transformations (Narendra et al. 2016).

Differential impacts of CBS mutations on Hoxd gene regulation in mesoderm derivatives

The expression of *Hoxd1* to *Hoxd4* in the dermal papilla of vibrissae follicles (VFs) depends on CREs located within a region of T-DOMa that preferentially interacts with *Hoxd1* and the most "anterior" part of the *HoxD* cluster (Fig. 5A). Likewise, a set of enhancers located across the same T-DOMa drives the expression of *Hoxd1* to *Hoxd9* into FMPs (Fig. 5A; Hintermann et al. 2021). We asked whether the disruption of intracluster CBSs would change the allocations of target genes for these two regulatory specificities, which concern either genes located within the part of the gene cluster that is devoid of CTCF sites or genes extending slightly behind CBS1 and CBS2. In Del(CBS1) mutant embryos, *Hoxd8* was up-regulated in both VFs and FMPs, when compared with control littermates (Fig. 5B, arrowhead and asterisk), in contrast to *Hoxd9*, which was not affected (Supplemental Fig. S10A). Instead, *Hoxd9* was ectopically activated in these structures in Del(CBS1-2) mutant fetuses (Fig. 5B, arrowhead and asterisk), while *Hoxd10* remained inactive in these mutants (Supplemental Fig. S10A). It is noteworthy that a tissue-specific effect of the CBS deletions on *Hoxd* gene expression was observed. Indeed, while the deletion of CBS1 and CBS2 allowed for the extension of the gene subset expressed in VFs up to *Hoxd9*, the additional mutation of CBS3 triggered the transcription of *Hoxd10* in the FMPs, yet not in the VFs (Fig. 5B, asterisk). No changes in the transcription domains of *Hoxd* genes located 3′ to the mutated CBSs was observed, as exemplified by the analysis of *Hoxd1* and *Hoxd4* transcript distribution in the Del(CBS1) and Del(CBS1-2) mutants, respectively (Supplemental Fig. S10B). Also, we were unable to observe any change in the expression of *Hoxd9* to *Hoxd11* in Del(CBS1-5) mutant embryos (Supplemental Fig. S10C).

In summary, the disruption of intracluster CBSs generally led to the ectopic transcriptional activation of the gene located immediately centromeric to the deleted CBSs. This suggests that, in these contexts, CTCF sites operate as insulator elements constraining gene–enhancer interactions such as to delimit specific subsets of contiguous genes competent to respond to particular remote enhancers. However, the up-regulation of *Hoxd9* and *Hoxd10* in the FMPs was considerably weaker, if not absent, from Del(CBS1-4) and Del(CBS1-5) mutant embryos, in contrast to the Del(CBS1-2) and Del(CBS1-3) conditions
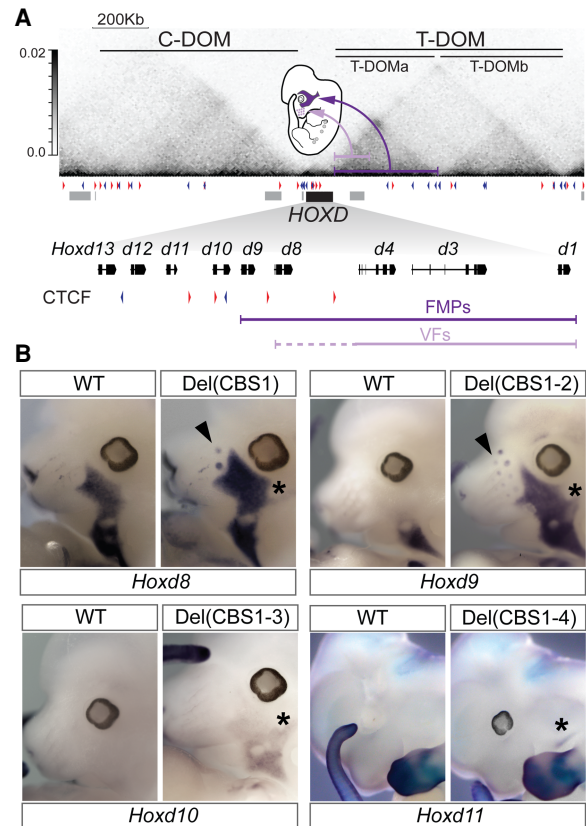

Figure 5. Alterations in *Hoxd* gene expression in facial structures. (*A*) Schematic representation of *Hoxd* gene regulation in both the vibrissae follicles (VFs; light purple) and some facial muscle precursors (FMPs; dark purple), as well as the DNA segments within T-DOM where the corresponding enhancers are located (Hintermann et al. 2021). A magnification of the gene cluster is shown *below* with the subsets of *Hoxd* genes expressed in each structure. CBSs and their orientations are indicated as red or blue triangles (*B*) WISH showing *Hoxd* gene expression in the VFs and FMPs of control (WT) or various CBS deletion alleles, as indicated in each panel. Black arrowheads and asterisks indicate either an up-regulation or an ectopic expression in the VFs and FMPs of each mutant allele, respectively.

(Fig. 5B; Supplemental Fig. S10C). Together with the loss of *Hoxd11* ectopic activation in Del(CBS1-5) embryos, this suggests that CBS4 and CBS5 may have an anchoring function for the FMPs and VF enhancers such that their mutations abrogate the gains of expression observed upon deletions of CBS1 to CBS3, the latter sites being used as insulators in these contexts.

Finally, mesoderm cells of the embryonic mammary bud contain high levels of *Hoxd9* mRNAs, in contrast to all other *Hoxd* genes, which are down-regulated in this structure around E12–E13 (Schep et al. 2016). In these cells, *Hoxd9* is controlled by a eutherian-conserved enhancer located within the CS38-40 sub-TAD boundary within T-DOM (Supplemental Fig. S10D). We looked at the effect of disrupting CBS2 in isolation (i.e., the CTCF site located between *Hoxd8* and *Hoxd9*) (Supplemental Fig. S10E), and the analysis of Del(CBS2) mutant embryos

revealed an up-regulation of *Hoxd9* in paraxial and lateral mesoderm, yet considerably weaker than that observed in Del(CBS1-2) mutant mice (Fig. 4B; Supplemental Fig. S10F). However, *Hoxd9* transcription in the mammary bud mesenchyme was no longer detected (Supplemental Fig. S10F), and *Hoxd10* was not ectopically activated in the mammary bud mesenchyme or in the main body axis of Del(CBS2) embryos (Supplemental Fig. S10G). These results indicate that CBS2 is required for the proper anchoring of the mammary bud enhancer with its target gene, while it operates as an insulator element in the main body axis. Together with the alterations observed in the VFs and FMPs, these data indicate that the disruption of CBS differentially impacts *Hoxd* gene transcription across tissues, with effects that cannot be solely attributed to their role as insulator elements.

Effects of CBS disruptions on Hoxd gene regulation in developing proximal limb buds

During limb buds development, *Hoxd* genes are regulated in a bimodal manner by the two TADs, with *Hoxd9* to *Hoxd11* initially under the control of forearm T-DOM enhancers, whereas *Hoxd10* to *Hoxd13* become subsequently controlled by C-DOM enhancers in digit cells (Figs 1A, 6A, 7A; Andrey et al. 2013; Beccari et al. 2016; Rodríguez-Carballo et al. 2017). To evaluate the function of CBSs both in the definition of these subsets of target genes and in the 3D chromatin conformation in limb bud cells, we performed RNA sequencing (RNA-seq) and capture Hi-C on microdissected proximal and distal E12.5 forelimb cells derived from both CBS mutant and control embryos.

Using the RNA-seq data sets, we assessed the effects of various CBS deletions first by performing a principal component analysis (PCA) considering the expression levels of the 500 most variant autosomal protein-coding genes. As expected, principal component 1 (PC1), which explained 83% of the total gene expression variance, separated the distal forelimb from the proximal forelimb (Supplemental Fig. S11A). Along PC2, which accounted for only 5% of the total variance, we observed that samples from the same litter tend to cluster together, illustrating a "litter effect" on the set of samples (Supplemental Fig. S11A). These observations were corroborated by the expression clustering based on pairwise Euclidean distances between samples (Supplemental Fig. S11B). These data indicated that the various CBS deletions have a negligible impact on the overall transcription profiles of the developing limbs, in agreement with the lack of major alterations in limb morphology in mutant animals throughout the allelic series.

We then conducted pairwise differential gene expression analyses (absolute $\log_2$ fold change >0.58, adjusted *P*-value < 0.05) of the proximal forelimb samples of control and various CBS mutant alleles. Using these parameters, we observed between six and 195 protein-coding genes differentially expressed in PFL (Supplemental Fig. S12A; Supplemental File 1), thus confirming the weak differences across samples observed in the PCA. However, the expression levels of anterior *Hoxd* genes (*Hoxd3*, *Hoxd4*, and *Hoxd8*) in proximal forelimb (PFL) were consistently de-

creased in all mutant alleles, with the exception of *Hoxd8* in Del(CBS1) (Fig. 6B). These differences could not be attributed to the litter effect, and the intersection of differentially expressed genes identified *Hoxd3* and *Hoxd4* as being the only misregulated genes in all alleles. Similarly, *Hoxd8* was the only differentially expressed gene in the Del(CBS1-2), Del(CBS1-3), Del(CBS1-4), and Del(CBS1-5) alleles (Supplemental Fig. S12A; Supplemental File 1). Finally, WISH analysis of *Hoxd8* transcripts in the Del(CBS1-2), Del(CBS1-3), and Del(CBS1-5) alleles confirmed this decrease in mRNAs levels (Supplemental Fig. S13A).

We then asked whether these changes in expression were due to reallocations in interactions between the genes and their regulatory landscapes. We performed capture Hi-C in microdissected proximal forelimb cells of the Del(CBS1), Del(CBS1-3), and Del(CBS1-5) alleles and compared them with control samples (Bolt et al. 2021). In the Del(CBS1) allele, we observed a minor loss of interactions with the CS38-40 T-DOM sub-TAD boundary and a gain of ectopic interactions with T-DOMa, as seen on the subtraction maps (Fig. 6C). When using *Hoxd4* as a viewpoint in virtual capture-C profiles, we observed a loss of interactions with the CS38-40 region in Del(CBS1) PFLs, which was not observed when using *Hoxd8* (Supplemental Fig. S14), in agreement with the changes in transcripts levels observed for these two genes. Instead, *Hoxd8* increased its interactions with T-DOMa in the Del(CBS1) allele (Fig. 6C, bracket x; Supplemental Fig. S14A,B), a gain that was reinforced in the Del(CBS1-3) allele (Fig. 6D, bracket x; Supplemental Fig. S14A,B). In the latter allele, a very significant decrease in interactions with the CS38-40 and the T-DOMb regions was scored (Fig. 6D; Supplemental Fig. S14A,B), correlating with the diminution of *Hoxd8* mRNA levels in the PFL of these mutants.

While virtual capture-C profiles confirmed the alterations in the interaction between *Hoxd8* and T-DOM (Fig. 6D; Supplemental Fig. S14C), the Del(CBS1-5) allele behaved somewhat unexpectedly, since the increase in anterior *Hoxd* interactions with T-DOMa was less pronounced, whereas the loss of interactions with CS38-40, T-DOMb, and the TAD boundary region was further accentuated when compared with the Del(CBS1) and Del(CBS1-3) alleles (Fig. 6E; Supplemental Fig. S14A–C). Furthermore, in the Del(CBS1-5) allele, we observed an increase in interactions between the anterior genes (e.g. *Hoxd4* and *Hoxd8*) and posterior *Hoxd* genes (*Hoxd13* to *Hoxd11*) (Fig. 6E, arrowhead; Supplemental Fig. S14C).

In parallel with their loss of contacts with T-DOMb, anterior genes gained some interactions with the centromeric end of C-DOM (Fig. 6E, bracket y), as if this part of the cluster was now detached from T-DOM and free to follow those contacts normally established between the 5′ C-DOM border and the CTCF sites with opposite orientations remaining in the "posterior" part of the *HoxD* cluster. In addition, T-DOMa itself seemed to follow this tendency and be dragged along by these remaining CTCF sites, as suggested by a general increase in interactions between T-DOMa and a region rich in CTCF sites in the 5′ part of C-DOM (Fig. 6E, bracket z).
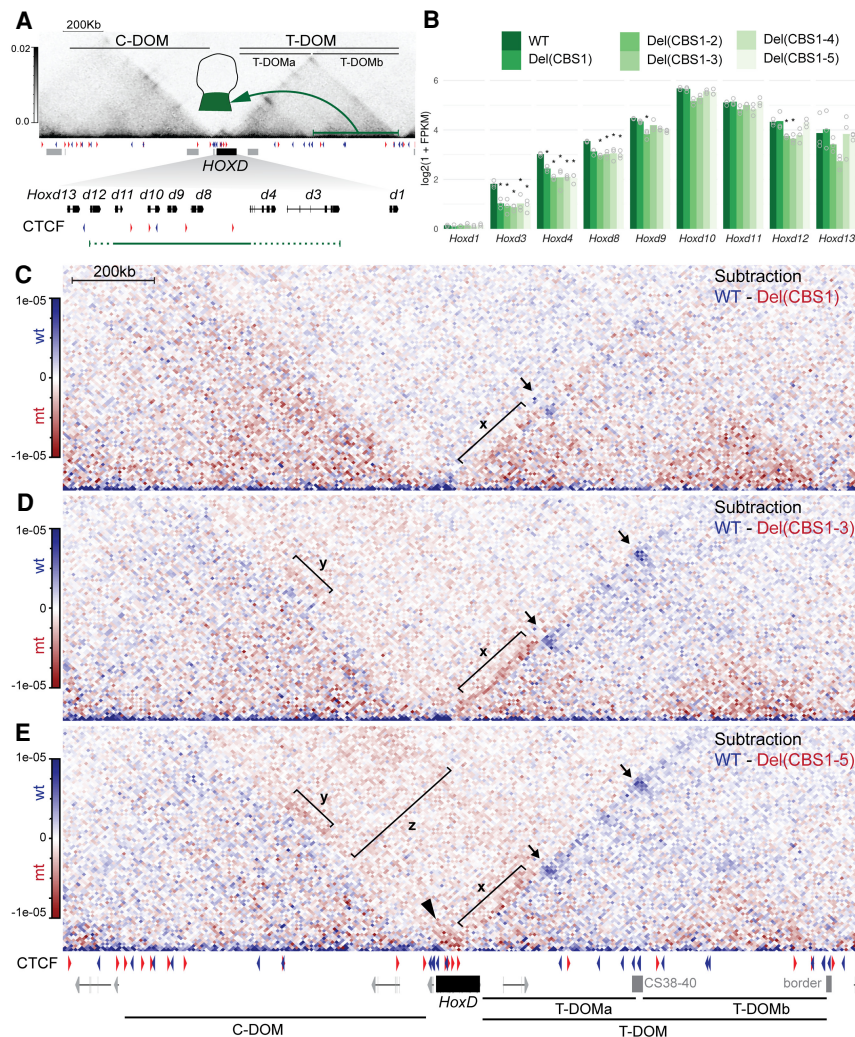
Figure 6. Impact of the CBS alleles on *Hoxd* gene regulation in proximal limb buds. (*A*) Capture Hi-C map of control E12.5 dissected proximal forelimb (PFL) cells with the *HoxD* cluster (black rectangle) and the neighboring T-DOM and C-DOM (mm10: chr2: 73779626–75669724). The schematic limb shows the T-DOM regulation in proximal cells (green) as a result of multiple enhancers located within T-DOMb (green line and arrow). *Below* is a magnification showing the CBSs and their orientations (blue and red arrowheads), as well as those *Hoxd* genes responding to this regulation (green line). (*B*) Bar plot representing *Hoxd* gene expression ($\log_2[1+FPKM]$) in PFL cells. Hollow circles show the replicates for each sample, and the asterisks point to those samples where significant differences in transcript levels were scored between control and mutant alleles (absolute $\log_2$ fold change >0.58, adjusted *P*-value < 0.05) (*C–E*) Capture Hi-C subtraction maps of control (blue) and mutant (red) E12.5 PFL cells. The subtracted alleles are indicated in the *top right* corners. The arrows represent the incremental loss of interactions between both the sub-TAD border region CS38-40 and the telomeric T-DOMb border region, and the part of the *HoxD* cluster where various CBSs were deleted in the mutant alleles. Bracket x indicates an increase in interactions between *HoxD* and T-DOMa, stronger in the Del (CBS1-3) than in the Del(CBS1-5). In contrast, intracluster interactions are stronger in the Del(CBS1-5) than in Del(CBS1-3) (arrowhead in *E*). Bracket y highlights a gain in interactions between the "anterior" genes and the centromeric end of C-DOM in Del (CBS1-3) and Del(CBS1-5). Bracket z points to a general increase in interactions between T-DOMa and a region rich in CTCF sites at the C-DOM centromeric border in Del (CBS1-5).

Effects of CBS disruptions on Hoxd gene regulation in developing distal limb buds

The general changes observed in gene expression and chromatin organization in distal forelimb (DFL) samples of our various mutant alleles were somewhat comparable with, yet slightly different from, those observed in PFL. Differential gene expression analyses identified between six and 282 protein-coding genes differentially expressed in distal forelimb cells (Supplemental Fig. S12B; Supplemental File 1). In the Del(CBS1-2), Del(CBS1-3), and Del (CBS1-4) alleles, we scored a mild yet significant decrease in mRNA levels of *Hoxd9*, *Hoxd10*, and *Hoxd11* (Fig. 7B). In agreement with RNA-seq data sets, WISH analysis showed decreased signal intensity of *Hoxd9* and *Hoxd11* in the DFLs of these alleles (Supplemental Fig. S13B). We also observed a recovery in transcript levels in the Del(CBS1-5) allele, both by RNA-seq and WISH (Fig. 7B; Supplemental Fig. S13C). Intersections between the Del (CBS1-2), Del(CBS1-3), and Del(CBS1-4) samples identi-

fied *Hoxd9*, *Hoxd10*, and *Hoxd11* as the only commonly differentially expressed genes in these samples (Supplemental Fig. S12B), highlighting again the local effect of these deletions on gene expression.

The subtractions of capture Hi-C data sets between DFL and control samples using the Del(CBS1-2), Del(CBS1-3), and Del(CBS1-5) mutant alleles revealed a progressive loss of interactions between the cluster and the CS38-40 region, T-DOMb, and the TAD border regions, similar to what was observed with PFL cells (Fig. 7C–E). Furthermore, we scored a gain of interactions with T-DOMa that was more pronounced in the Del(CBS1-3) allele, which seemed to slightly recover in the Del(CBS1-5) allele (Fig. 7D,E, bracket x). This loss of ectopic interactions with T-DOMa in the Del(CBS1-5) allele may influence the recovery of *Hoxd* gene expression in the Del(CBS1-5) allele.

While the interactions between the C-DOM and the 5′ part of the *HoxD* cluster did not seem to be much affected in the various mutant alleles, contacts were increased between the *Hoxd8* to *Hoxd4* region and C-DOM in the Del
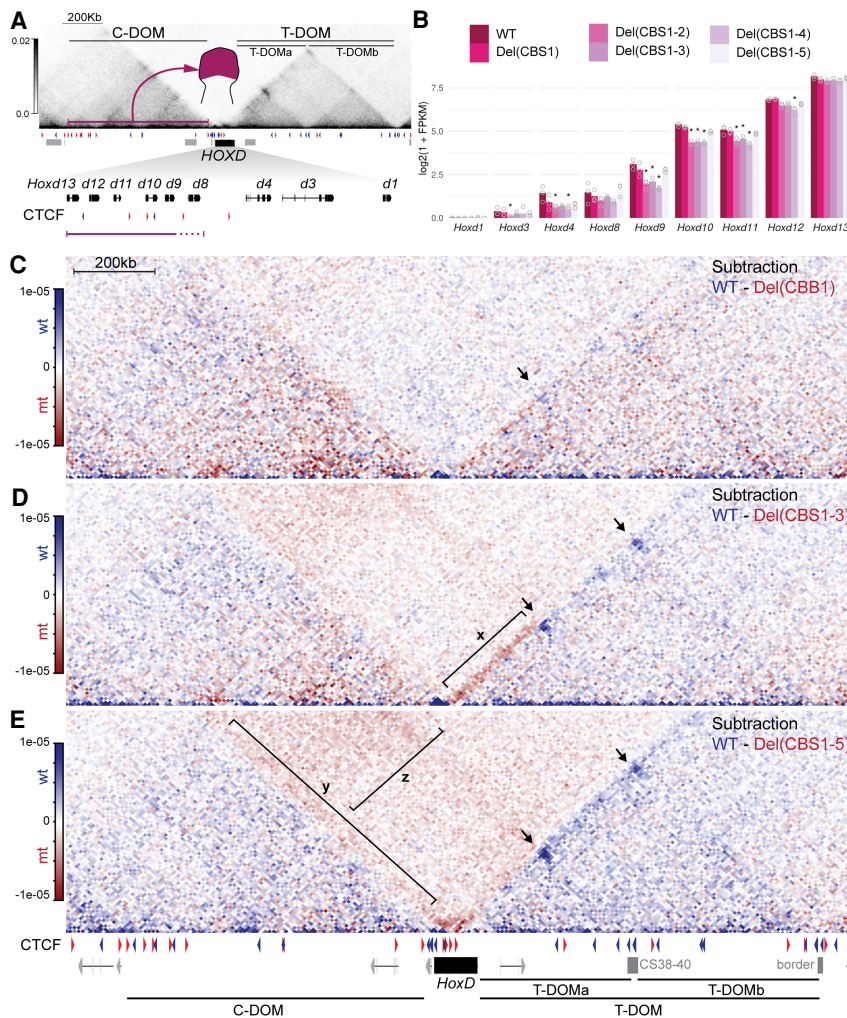
Amândio et al.

Figure 7. Impact of the CBS alleles on *Hoxd* gene regulation in distal limb buds. (*A*) Capture Hi-C map of control E12.5 dissected distal forelimb (DFL) cells with the *HoxD* cluster (black rectangle) and the neighboring T-DOM and C-DOM (mm10: chr2: 73779626–75669724). The schematic limb shows the C-DOM regulation in distal cells (pink) as a result of multiple enhancers spread over C-DOM (pink line and arrow). *Below* is a magnification showing the CBSs and their orientations (blue and red arrowheads), as well as those *Hoxd* genes responding to this regulation (pink line). (*B*) *Hoxd* genes expression ($\log_2[1+\text{FPKM}]$) in DFL. Hollow circles show the replicates for each sample, and the asterisks point to those samples where significant differences in transcript levels were scored between control and mutant alleles (absolute $\log_2$ fold change >0.58; adjusted *P*-value < 0.05). (*C–E*) Subtractions of cHi-C maps between control E12.5 DFL (blue) and various homozygous mutants (red). The subtracted alleles are indicated in the *top right* corners. Arrows indicate the loss of interactions in the mutant alleles between the "anterior" part of the *HoxD* cluster and both the sub-TAD boundary region CS38-40 and the telomeric T-DOMb border region. Bracket x points to increased interactions between the *HoxD* cluster and T-DOMa in the Del(CBS1-3) mutant (*D*), which is less pronounced in the Del(CBS1-5) allele (*E*). Bracket y indicates increased interactions observed in the Del(CBS1-5) mutant between anterior *Hoxd* genes and C-DOM, whereas bracket z highlights the loss of insulation between C-DOM and T-DOMa, which is most pronounced in this allele [compare with the Del(CBS1) in *C*].

(CBS1-5) (Fig. 7E, bracket y), as well as for the entire T-DOMa sub-TAD, as was observed in proximal cells (Fig. 7E, bracket z). However, this gain of contacts was not paralleled by a detectable gain of expression of anterior *Hoxd* genes in digits as revealed by RNA-seq and WISH analysis (Fig. 7B; Supplemental Fig. S13A). Overall, these results show that CBS deletions at the *Hoxd* gene locus result in changes in local chromatin architecture that lead to the down-regulation of *Hoxd* gene transcription in proximal and distal forelimbs. However, the observed changes in gene expression indicate a requirement of the intracluster CBSs to reach a proper level of transcription, distinct form the insulator or anchoring effects observed during the development of the main embryonic body axis.

Discussion

The mutation of CTCF binding sites to study the role of this protein in the organization of chromatin in 3D and in transcriptional insulation has been reported by using different genetic loci and experimental models. For example, the deletion of some CTCF sites at the mouse *Shh* locus weakened the formation of large chromatin loops between

the *Shh* gene and its ZRS enhancer (Paliou et al. 2019). Also, the importance of multiple CTCF sites for proper insulation at TAD borders through the formation of local chromatin domains has been assessed through mutagenesis of the *Sox2* locus in ES cells (Huang et al. 2021), as well as their cooperative and redundant functions when disposed as arrays at the *Pax3* locus (Anania et al. 2021). In this study, we genetically dissected a series of five CBSs, all located on one side of the strong TAD boundary positioned within the *HoxD* gene cluster. In addition to being involved in the making of a tight chromatin border (Rodríguez-Carballo et al. 2017), these sites delimit various subsets of contiguous *Hoxd* genes, which respond to distinct tissue-specific remote enhancers, thus raising the possibility that CBS-dependent microchromatin domains could be defined for each (series of) enhancer(s).

CTCF as moderator of spatial colinearity

We used our allelic series of mutations in *cis* to try to distinguish between a microinsulating effect (between genes), a macroinsulating effect (between TADs), and an anchoring function for these multiple CBSs. The

efficiencies of the mutations were checked using both CTCF and RAD21 in ChIPmentation experiments, which revealed that CTCF binding had been abrogated. While some very weak CTCF signals were still observed under certain experimental conditions, as a result of our strategy to induce a minimal disruption, RAD21 was no longer enriched at any of the sites. No cryptic CBS was revealed even in the absence of the five native CBSs, and the enrichment of RAD21 was redistributed toward the remaining CBSs, with a clear preference for the CBSs located close to the deleted series [e.g. CBS4 in the Del(CBS1-3) allele]. This indicated that cohesin-driven chromatin architectures were locally modified in these mutants, at least by using a mix of trunk cells as starting material. However, CBS5 displayed only a moderate increase in RAD21 signal upon deletion of its four 3′ neighboring CTCFs, suggesting that sequence- or context-dependent factors, other than motif orientation, can influence the capacity of CBSs to retain cohesin (Phillips-Cremins et al. 2013).

This observation was paralleled by a more condensed aspect of the cluster itself, as determined by chromosome capture, with intracluster interactions increasing along with the number of CTCF sites mutated and progressively extending toward the centromeric end of the gene cluster up to the TAD boundary in the full CTCF mutant. Concomitantly, long-range contacts between the cluster and T-DOM were lost, an effect particularly visible at positions corresponding to the presence of convergent CTCF sites such as the CS38-40 region, as well as other CBSs located further telomeric. We interpret these two phenomena as a single effect of removing all sites orientated toward T-DOM within the *Hox* cluster; the lack of long-range interactions with T-DOM somehow relaxed the architecture, thus allowing *Hoxd* genes to establish more local interactions, which are normally reduced whenever the cluster is "under tension" through its contacts with T-DOM.

How such an increase in local interactions may affect *Hox* gene transcription is more difficult to evaluate in our experimental paradigm. *Hox* genes are transcribed in successively more "posterior" combinations along with embryonic caudal trunk extension, with a correspondence between the gene's order in the clusters and the anterior to posterior (A-P) level where they become activated (Gaunt et al. 1988). Previous deletions on CTCF sites showed that this particular A-P level could be modified in some instances, leading to a local misexpression of neighboring genes (Narendra et al. 2016). In this view, particular CBSs could be considered as having an insulator function. Here, by using this complete allelic series, we show that this effect is systematically observed for genes located close to the deleted CBSs. However, while the positioning of the expression boundaries was abnormal in the multiple mutated Del(CBS1-4) and Del(CBS1-5) specimens, the colinear distribution of these boundaries was conserved, suggesting that while CTCF is required to fine-tune the A-P levels where particular genes will be switched on, it is not necessary for the implementation of spatial colinearity in itself. The proper adjustment of these expression domains is required for harmonious de-velopment, as shown by the phenotypic alterations accompanying these allelic series, which were all in agreement with previous results and general principles of homeosis (Kessel and Gruss 1991; Gerard et al. 1996; Tarchini et al. 2005; Narendra et al. 2015). Therefore, in this context, CTCF can be seen as a moderator of spatial colinearity rather than its organizer.

Importance of CTCF for selective regulations by remote enhancers

In the case of remote enhancer–promoter interactions, bound CTCFs were shown to act both as insulators and as anchoring elements. For example, a tandem insertion of CTCF sites could insulate the *Pcdh* promoter from an enhancer (Jia et al. 2020), and tissue-specific insulation of a CTCF-bound region was also reported to control the selective expression of human growth hormone hGH gene clusters (in the placenta) (Tsai et al. 2016). In contrast, other CTCF-bound regions can act as facilitators/tethering elements to bring enhancers located at long distances close to their promoters, as exemplified by the transcription of the *Shh* gene during limb development, which requires the proximity of the ZRS sequence, a task achieved either partly (Paliou et al. 2019) or fully (Ushiki et al. 2021) by the presence of CTCF sites.

Our mutant alleles revealed that, while CBSs at *HoxD* were generally used as insulators, the insulating function was in some cases tissue-specific, because the observed extension of the expression domain was not the same in all tissues analyzed. Some CBSs were also necessary to properly anchor enhancer–promoter interactions, as reported in other loci, such as the *Thy1* locus (Ren et al. 2017). For example, the clear gain of expression of *Hoxd9* in the FMPs observed in both Del(CBS1-2) and Del(CBS1-3) was reduced or mostly abrogated in the Del(CBS1-4) and Del(CBS1-5) alleles, respectively, suggesting that both CBS4 and CBS5 triggered the appropriate contact rather than insulating yet another transcription unit from remote enhancers. This hypothesis is further supported by the observation that although *Hoxd11* is only weakly up-regulated in the FMPs of Del(CBS1-4) embryos, correlating with the weak RAD21 enrichment of CBS5 in these mutants, its gain was completely abolished in Del(CBS1-5) homozygous mutant specimens. In fact, the anchoring and insulating functions could be exerted by the same site in different contexts, as shown with CBS2, which clearly acted as an anchor in mammary bud cells where the expression of *Hoxd9* was lost when CBS2 was mutated, whereas the same gene was up-regulated in both paraxial and lateral mesoderm of CBS2 mutant embryos. Such a dual role has been reported for other CTCF binding sites such as those flanking the HS5 enhancer of *Protocadherin a* cluster (Guo et al. 2012, 2015; Jia et al. 2020), and the insulator activity of CTCF-bound elements has been related to their CpG methylation and epigenetic modifier state in different contexts (Bell and Felsenfeld 2000; Guo et al. 2012).

The functional ambivalence of these CBSs may depend on several parameters such as the presence of other

tissue-specific factors, the local extrusion conditions, or the strength of interactions, to mention a few possibilities. In any case, the dual and context-dependent activity of CBSs that we report at this locus indicate that one should be careful when assessing the effects of mutating one or a few CTCF sites within a larger array of such sequences, because the overall effect of such arrays may not be extrapolated from the addition of single mutation events. The functional compensation that we observed when mutating the full series of CBSs [Del(CBS1-5)] illustrates this point.

CBSs in the making of the TAD boundary at the HoxD cluster

All four mammalian *Hox* clusters are partitioned in between two TADs, with a boundary systematically isolating the most posterior group 13 *Hox* gene from the others. Group 13 *Hox* genes are coding for proteins that participate in the morphogenesis of terminal structures, and hence the embryo must be protected from the premature action of such proteins, which can abrogate axial extension (Young et al. 2009) or affect proper morphogenesis (Darbellay et al. 2019; Bolt et al. 2021) through their dominant-negative function. This isolation is achieved through the positioning of *Hox13* genes into a different TAD, thus making them unresponsive to the numerous enhancer sequences acting on other *Hox* genes during development. As illustrated with *HoxD*, TAD boundaries at *Hox* clusters are remarkably tight (Rodríguez-Carballo et al. 2017). While, as for many other such chromatin borders, *Hox* loci display arrays of CBSs (Gómez-Marín et al. 2015; Vietri Rudan et al. 2015; Kentepozidou et al. 2020), the orientations of these sites are generally not as intermingled as is often observed (e.g., see Nanni et al. 2020; Anania et al. 2021). Instead, the CBSs are organized as two arrays of opposite orientations, thus defining a clear point of transition, tightly isolating *Hox13* from the gene cluster.

While several studies have revealed that the number of CBSs was important for insulation at TAD boundaries, likely through a synergistic effect (Huang et al. 2021), particular CBSs may be more important than others in achieving this task (Anania et al. 2021), and it appears that tandem distributions of CBSs may provide robustness to the TAD border. In our allelic series, the global insulation capacity of the TAD border was increasingly affected along with more CTCF sites being mutated, as best seen in the case of the developing limbs. However, this effect was weak and the TADs remained almost as in control cells, even when all CBSs with the telomeric orientation had been mutated. Strong variations were observed in the interactions between the cluster and T-DOM, yet the TAD border itself was only weakly affected, suggesting that a single array of CTCF sites with the same (centromeric) orientation is sufficient to provide isolation between the TADs. This was confirmed by the lack of observable gain of expression of *Hoxd13* whenever it was tested, in particular in proximal limb cells, and this even though the position of the TAD border was shifted toward this gene. This illustrates that the strong T-DOM enhancers, actively working on *Hoxd10* and *Hoxd11*, were still unable to reach *Hoxd13*, even in mutant Del(CBS1-5) limbs, and hence it suggests that a single array of CTCF sites with the same orientation is sufficient to reach a fair level of insulation, at least in this context. However, the shift of the TAD border toward *Hoxd13* may clearly weaken the isolation of this gene, and it is not impossible that its ectopic transcription in response to some T-DOM enhancers in a particular cellular context, during development or adulthood, has escaped our attention thus far.

These multiple functions for such CBS arrays at *Hox* loci may explain the remarkable evolutionary conservation in the global organization of such binding sites. Indeed, despite several hundred million years of separate evolution for these paralogous gene clusters (Holland et al. 1994) accompanied by drastic functional divergences (Soshnikova et al. 2013) and sometimes the loss of several genes (Hoegg and Meyer 2005), the general distribution of these two arrays of CTCF sites, as well as their opposed orientations, was conserved. Likewise, human and mouse *Hox* clusters are identical in this respect, unlike other situations where the function of CTCF sites was conserved, but not their exact locations (Ushiki et al. 2021). Furthermore, the fact that in all four clusters the first CTCF site is located right upstream of the group 4 *Hox* gene promoter demonstrates a function for these CBSs that is more precise and complex than making a mere TAD border.

The impact of CBSs for Hoxd gene regulation in limb buds

During limb development, different subsets of *Hoxd* genes respond first to T-DOM-located proximal enhancers, then to C-DOM-located digit enhancers (Montavon et al. 2011; Andrey et al. 2013). Because the two subgroups of target genes are delimited by CTCF sites, it was suggested that such CBSs may specify, through a 3D structure, particular enhancer–promoter interactions leading to these bimodal transcription patterns (Rodríguez-Carballo et al. 2020). However, the most visible effect observed across the different mutant lines generated in this study was more quantitative than qualitative, with no major alterations in limb morphology in mutant animals and no obvious misexpression of target gene subgroups.

Furthermore, all variations observed in the amounts of mRNAs in all mutant configurations tested were systematically losses of transcripts, when compared with control mRNA levels. Such losses could be explained when looking at variations in chromatin topology. For example, mutations in all five CTCF sites considerably weakened contacts between the cluster and T-DOMb, where most PFL enhancers are located (Andrey et al. 2013; Rodríguez-Carballo et al. 2020). These contacts were redirected toward T-DOMa, which is poor in PFL enhancers but rich in other regulations, leading to the ectopic expressions described above, concomitant to a loss of expression in proximal limbs.

While the overall weakening of contacts in T-DOMb was even more evident in Del(CBS1-5) mutants, this abrogation of all five CTCF sites resulted in a re-equilibration of contacts with T-DOMa. In addition, the increase in

local intra-TAD contacts observed in the C-DOM of Del (CBS1-3) mutants was also reverted in the distal forelimb of Del(CBS1-5) mutant embryos, likely accounting for the rescue in *Hoxd* gene expression levels observed in these mutants. These results indicate that CBS1 to CBS5 are required for the correct anchoring of *Hoxd* genes with their long-range enhancers situated in the flanking TADs, and the mutagenesis of some of these sites thus impacts on their interaction pattern, as well as on the internal TAD architecture. However, the removal of all sites resulted in a more homogeneous loss of contacts between *Hoxd* genes and their regulatory landscapes. Therefore, in this context at least, rather than being used to fine-select the appropriate sets of target *Hoxd* genes, the CTCF sites are required to secure the highest possible expression levels by triggering more robust interactions with T-DOMb-located long-range enhancers. In this view, CBSs may achieve a function in potentiating regulations rather than in their specific establishment.

A gene cluster under tension?

How such a potentiation effect could be achieved is elusive, and the abrogation of every single CTCF binding site taking part in these regulations might be of help in this context. It is likely that this effect results from a global reorganization of the chromatin landscape rather than punctual modifications. For example, in the trunk of a Del(CBS1-5) mutant specimen, anterior *Hoxd* genes (i.e., *Hoxd3* and *Hoxd4*) established ectopic interactions with more "posterior" genes (*Hoxd9* and *Hoxd10*). Concomitantly, the "anterior part" of the gene cluster lost many interactions with T-DOM, suggesting that the two patterns of interactions are exclusive and illustrate two global conformations of the 3D chromatin at the locus. As a consequence, it is possible that, in the normal situation, the contacts established with T-DOM are necessary to bring the gene cluster under some tension to open it such that it may respond to enhancers in an optimal manner. Physical forces have been proposed to play a role in the colinear regulation of *Hox* genes (Papageorgiou 2006). In this particular case, however, the forces would be generated by the local chromatin interactions themselves, rather than through an asymmetrically localized point of attachment to the nuclear environment.

In any case, our results indicate that CTCF may be considered as a factor used to maximize transcription through the reorganization of chromatin, rather than a key determinant for enhancer–promoter interaction necessary for transcription to occur. Our observations are in agreement with the somehow reduced effect of removing CTCF genome-wide on global transcription patterns, at least during development (Wan et al. 2008; Soshnikova et al. 2010; Nora et al. 2017; Kubo et al. 2021).

Materials and methods

Cloning of sgRNAs and generation of the CBS mutant stocks

The sgRNA targeting guides were generated by annealing complementary pairs of oligonucleotides (Supplemental Table S2)

and cloning into the pX330 vector as described in Darbellay et al. (2019). For sgRNA transcription, we PCR-amplified the sgRNA sequence cloned into the px330 plasmid using a T7 promoter containing primer and a universal reverse oligonucleotide (TAATACGACTCACTATAG). PCR products were gel-purified and transcribed in vitro using the HiScribe T7 high-yield RNA synthesis kit (NEB). Cas9 mRNA was synthesized from the pX330 plasmid using the mMessage mMachine T7 Ultra (Thermo Fisher) according to the manufacturer's instructions. The transcribed sgRNAs were purified using the RNeasy mini kit (Qiagen). The purified Cas9 mRNA and sgRNAs were coelectroporated in mouse fertilized oocytes. The sgRNAs used are listed in Supplemental Table S2.

With the exception of the Del(CBS1-4) line, F0 animals obtained upon Cas9/sgRNA electroporation were screened at weaning age using the Surveyor mutation detection kit and specific primers amplifying a 500- to 1000-bp fragment around the targeted CBS (Supplemental Table S3). The Del(CBS1-4) animals were screened by PCR (Supplemental Table S3). The first mutant strain obtained was the Del(CBS2). The Del(CBS1-2) strain was obtained on top of Del(CBS2) zygotes by introducing the CBS1′ mutation depicted in Supplemental Figure S4. The distinctive CBS1 mutation (Supplemental Fig. S4) was present only in the Del(CBS1) strain, produced independently. The Del(CBS1-3) strain was produced on top of Del(CBS1-2) zygotes, and the Del(CBS1-4) strain was produced on top of Del(CBS1-3) zygotes, with the mutation referred to as CBS4 in Supplemental Figure S4. The Del(CBS1-5) strain was produced on top of another Del(CBS1-4) allele containing the CBS4′ mutation referred to in Supplemental Figure S4.

For genotyping, specific primers were used to amplify a region of ~160–423 bp surrounding the newly mutated CBSs for each combined mutant strain generated, allowing us to discriminate the control from deleted alleles in 5% agarose gel electrophoresis (Supplemental Table S3). All mutant alleles were verified by Sanger sequencing (Supplemental Fig. S4). To evaluate whether any new mutated CBSs were positioned in *cis* or in *trans* with previous CBS mutation(s), several F0 specimens for the new allele were crossed over control animals and F1 progenies were genotyped to look at the segregation of mutations. In this way, mutant alleles were recovered with the newly mutated CBSs positioned either in *cis* or in *trans* with previous CBS mutations. All mice used either for zygote electroporation or for further breeding were from a mix Bl6XCBA background, as all *HoxD* alleles produced in the laboratory.

Whole-mount in situ hybridization (WISH)

WISH experiments were performed as described in Woltering et al. (2009). The *Hoxd11*, *Hoxd10*, *Hoxd9*, *Hoxd8*, and *Hoxd4* gene probes were described in Dolle et al. (1991) and Gerard et al. (1996).

Micro-CT scan analyses

Mouse adult skeletons were revealed after micro-CT scanning using the Quantum GX2 micro-CT imaging system (PerkinElmer) at the small animal preclinical imaging platform of the Medical Faculty of the University of Geneva. Images were analyzed using OsiriX MD v.10.0.1 software.

ChIPmentation (ChIP-M)

ChIP-M experiments for CTCF and RAD21 were performed according to Darbellay et al. (2019). Briefly, trunk regions of E10.5 embryos (from below the postoccipital region) derived from

trans-heterozygous crosses were individually dissociated into single cells with collagenase and fixed in 1% formaldehyde solution for 10 min at room temperature. The cross-linking reaction was stopped by adding glycine to a final concentration of 0.125 M, and the cell pellet was washed three times with cold PBS with protease inhibitors (Complete mini EDTA-free proteinase inhibitor cocktail; Roche). Fixed samples were stored at −80°C. Yolk sacs were used for genotyping. Samples from homozygous mutant embryos or control littermates were resuspended in sonication buffer (50 mM Tris HCl at pH 8.0, 10 mM EDTA, 0.25% SDS, protease inhibitors), and the chromatin was sheared in a Covaris S200 sonicator to an average fragment size of 250–300 bp. The sonicated chromatin was diluted 2.5 times with dilution buffer (20 mM HEPES at pH 7.3, 1 mM EDTA, 1% NP40, 150 mM NaCL, protease inhibitors). Antibodies against CTCF (4 g; Active Motif 61311) or RAD21 (5 µg; Abcam ab992) were conjugated with Dynabeads Protein G (Thermo Fisher) magnetic beads. The bead–antibody complexes were added to the diluted chromatin and incubated overnight at 4°C in a rotating platform. The day after, chromatin–antibody complexes were washed using a magnetic stand with RIPA buffer (10 mM Tris HCl at pH 8.0, 1 mM EDTA, 0.1% sodium deoxycholate, 1% TritonX-100, 140 mM NaCl; two washes), RIPA high-salt buffer (10 mM Tris HCl at pH 8.0, 1 mM EDTA, 0.1% sodium deoxycholate, 1% Triton X-100, 500 mM NaCl; two washes), LiCl buffer; (10 mM Tris HCl at pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% sodium deoxycholate, 0.5% NP40; two washes), and 10 mM Tris HCl (pH 8.0; two washes). The chromatin was then tagmented for 2 min with the Tn5 transposase (Illumina). Tagmented chromatin was eluted, reverse-cross-linked, and purified using Qiagen Minielute columns. Libraries were quantified using the Kapa library quantification kit and PCR-amplified using barcoded primers. ChIP-M libraries were pair-end-sequenced in a NextSeq 500 sequencer (PE 2x 37-43bp). Adapters and bad quality bases were removed from the Fastqs with CutAdapt version 1.16 (Martin 2011) (-a CTGTCTCTTATACACATCTCCGAGCCCACGAGACACTGT CTCTTATACACATCTGACGCTGCCGACGA -q 30 -m 15). Filtered reads were mapped to the mouse genome mm10 with Bowtie2 version 2.3.5 (Langmead and Salzberg 2012) with default options. Only alignments with a mapping quality >30 were kept (Danecek et al. 2021). PCR duplicates were removed with Picard version 2.19.0 (http://broadinstitute.github.io/picard). To decrease the importance of fragment length variation between libraries, only the first reads in pairs were kept, and BAM was converted to BED. Peak calling was run with a fixed fragment size of 200 bp with Macs2 Callpeak version 2.1.1.20160309 (–call-summits -f BED –nomodel –extsize 200 -B –keep-dup all). In order to normalize all ChIP-M despite different signal to noise ratios, we ran MAnorm version 1.1.4 (Shao et al. 2012) with -w 100 between each sample and the second replicate of the wild type. The M-A model coefficient was extracted and used to normalize each Bedgraph from Macs2 (see https://github.com/lldelisle/scriptsForAmandioEtAl2021 for all details). Replicates were then averaged at each base using Bedtools version 2.27.1 (Quinlan 2014). In Supplemental Figure S9, the coverages from Macs2 were normalized to the million tags. Then, wild-type replicates were averaged. CTCF site orientation was determined using CTCFBSDB 2.0 (http://insulatordb.uthsc.edu; Ziebarth et al. 2012) using the CTCFBS Prediction tool with sequences of 500 bp centered on each summit of each of the three wild-type replicates. For each sequence, the motif with the highest score among REN_20, LM2, LM7, and LM23 motifs was kept. Only motifs that were common between the three replicates are displayed. RAD21 MAnorm-normalized coverage was quantified on 500-bp regions centered on the CTCF motifs with multiBigwigSmmary from deepTools version 3.5 (Ramírez et al. 2016). The PWM motif lo-gos of Supplemental Figure S4 were generated using the STAMP (Mahony and Benos 2007; Mahony et al. 2007).

ChIP-seq

CTCF ChIP-seq experiments in Supplemental Figure S1 were performed as in Rodríguez-Carballo et al. (2017). For the posterior trunk samples, the portion corresponding to the tailbud, presomitic mesoderm, and two to three somite pairs, excluding as much as possible the region of the incipient hindlimb buds, was microdissected from ∼30 E9.75–E10 embryos. For the brain samples, a pool of six microdissected forebrains from E12.5 embryos was used. The ChIP-seq Fastqs of E12.5 distal and proximal forelimb samples were downloaded from GEO (GSM2713707 and GSM2713708, respectively). TrueSeq adapters were removed from single-read Fastqs with CutAdapt version 1.16 (-a CTGTCTCTTATACACATCTCCGAGCCCACGAGAC -q 30 -m 15) (Martin 2011). Filtered reads were mapped to the mouse genome mm10 with Bowtie2 version 2.3.5 (Langmead and Salzberg 2012) with default options. Only alignments with a mapping quality >30 were kept (Samtools version 1.9) (Danecek et al. 2021). Peak calling was run with a fixed fragment size of 200 bp with Macs2 Callpeak version 2.1.1.20160309 (–call-summits –nomodel –extsize 200 -B). Finally, the coverages were normalized to the million unique tags. For the published ChIP-seq used in Supplemental Figure S3, coverage files were downloaded from GEO as well as the corresponding peaks files. To identify CTCF orientation, the procedure was very similar to the analysis for ChIP-M except that for the humans there was no summit information so the peaks were scaled to 500 bp.

Capture Hi-C

Microdissected E12.5 proximal and distal forelimb pairs and individual E9.5 trunks were isolated in PBS supplemented with 10% fetal calf serum, dissociated into single cells by collagenase treatment, fixed in 1% formaldehyde, and stored at −80°C until further processing. Samples were genotyped by PCR as described above to select for homozygous mutant or control tissues. The SureSelectXT RNA probe design and capture Hi-C experiments were performed as described in Bolt et al. (2021). The first part of the data analysis was performed on our local galaxy server (Afgan et al. 2016). Raw reads were preprocessed with CutAdapt version 1.16 (-a AGATCGGAAGAGCACACGTCTGAACTC-CAGTCAC -A AGATCGGAAGAGCGTCGTGTAGGGAAA-GAGTGTAGATCTCGGTGGTCGCCGTATCATT –minimum-length = 15 –pair-filter = any –quality-cutoff = 30) (Martin 2011). Then, Hicup version 0.6.1 (Langmead and Salzberg 2012; Dryden et al. 2014) and Samtools 1.2 (Danecek et al. 2021) were used with default parameters. The BAM file was converted to a tabular file with a Python script (see https://github.com/lldelisle/scriptsForAmandioEtAl2021 for more details). The pairs were then loaded to 10-kb resolution matrices with Cooler version 0.7.4 (Abdennur and Mirny 2020). The heat maps in Figures 1, 3, 4, and 5 and Supplemental Figures S6, S7, and S9 are representations of the two replicates using HiCExplorer hicSumMatrices tool version 3.6 (Ramírez et al. 2018; Wolff et al. 2018, 2020) and Cooler balance. Subtraction maps were generated from balanced matrices using the HiCExplorer hicCompareMatrices tool version 3.6. For Supplemental Figure S6, a correction was applied to account for differences between matrices in their close distance-dependent signal using the HiCExplorer hicTransform tool version 3.6 (Ramírez et al. 2018; Wolff et al. 2020) with the obs_exp_non_zero transformation method and the perChromosome option. The TAD

separation scores were computed with HiCExplorer hicFind-TADs version 3.6. (Ramírez et al. 2018; Wolff et al. 2020) with a fixed window size of 240 kb. Heat maps were plotted with pyGenomeTracks 3.5 (Ramírez et al. 2018; Lopez-Delisle et al. 2021). Virtual C profiles were generated similarly to Despang et al. (2019), with a custom Python script available at https://github.com/lldelisle/scriptsForAmandioEtAl2021. The viewpoint coordinates (mm10) used were *Hoxd4*, chr2: 74721977; *Hoxd8*, chr2: 74704614; *Hoxd9*, chr2: 74697726; and TAD border: chr2: 75588758. The quantifications in Supplemental Figures S6, S7, and S13 were also carried out with a custom Python script.

RNA-seq

Microdissected individual pairs of either control or mutant E12.5 proximal and distal forelimbs were stored at –80°C in RNAlater stabilization reagent (Ambion) before further sample processing. Total RNA was extracted from tissues using Qiagen RNeasy Plus micro kit (Qiagen) after disruption and homogenization according to the manufacturer's instructions. RNA quality was assessed using an Agilent 2100 Bioanalyzer. The sequencing libraries were prepared according to TruSeq stranded mRNA Illumina protocol, with polyA selection. RNA-seq libraries were sequenced on an Illumina HiSeq 4000 sequencer, as single reads (read length 50 bp). Raw RNA-seq reads were processed with CutAdapt version 1.16 (-a CTGTCTCTTATACACATCTCC GAGCCCACGAGAC -q 30 -m 15) (Martin 2011) to remove TruSeq adapters and bad quality bases. Filtered reads were mapped on the mouse genome mm10 with STAR version 2.7.0.e (Dobin et al. 2013) with ENCODE parameters with a custom gtf file (https://doi.org/10.5281/zenodo.4596489) based on Ensembl version 102. This custom gtf file was obtained by removing read-through transcripts and all noncoding transcripts from a protein-coding gene. In addition, all genes with the same gene name that overlaps were merged under the same gene ID to avoid ambiguous reads. FPKM values were evaluated by Cufflinks version 2.2.1 (Trapnell et al. 2010; Roberts et al. 2011) with options –max-bundle-length 10000000 –multiread-correct –library-type "fr-firststrand" -b mm10.fa –no-effective-length-correction -M MTmouse.gtf -G. Counts from autosomal chromosomes were used for differential expression analysis with DESeq2 version 1.24.0 (Love et al. 2014) with R version 3.6.0 (http://www.r-project.org) with default parameters except for θ, which was fixed to 0.15 and 0.99. Only genes with absolute $\log_2$ fold change >0.58 and adjusted P-value > 0.05 were considered as significant. All the significant results are summarized in Supplemental File S1. PCA and clustering were performed on $\log_2(1 + FPKM)$ values of the 500 most variant genes from autosomal chromosomes. Gene annotations on the figures were plotted using a custom gtf file (https://doi.org/10.5281/zenodo.4596489) based on Ensembl version 102, and both *Hoxa3*-201 and *Hoxd3*-203 transcripts were removed to facilitate visualization.

Ethics approval

All experiments involving animals were performed in agreement with the Swiss Law on Animal Protection (LPA) under license number GE 81/14 (to D.D.).

Data availability

All raw and processed data sets are available in the Gene Expression Omnibus (GEO) repository under accession number GSE181387. All scripts necessary to reproduce figures from raw data are available at https://github.com/lldelisle/scriptsForAmandioEtAl2021.

References

Abdennur N, Mirny LA. 2020. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36:** 311–316. doi:10.1093/bioinformatics/btz540

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44:** W3–W10. doi:10.1093/nar/gkw343

Amândio AR, Lopez-Delisle L, Bolt CC, Mascrez B, Duboule D. 2020. A complex regulatory landscape involved in the development of external genitals. *Elife* **9:** e52962. doi:10.7554/eLife.52962

Anania C, Acemel RD, Jedamzick J, Bolondi A, Cova G, Brieske N, Kühn R, Wittler L, Real FM, Lupiáñez DG. 2021. In vivo dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation. bioRxiv doi:10.1101/2021.04.14.439779

Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. 2013. A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science* **340:** 1234167. doi:10.1126/science.1234167

Beccari L, Yakushiji-Kaminatsui N, Woltering JM, Necsulea A, Lonfat N, Rodriguez-Carballo E, Mascrez B, Yamamoto S, Kuroiwa A, Duboule D. 2016. A role for HOX13 proteins in the regulatory switch between TADs at the HoxD locus. *Genes Dev* **30:** 1172–1186.

Bell AC, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405:** 482–485. doi:10.1038/35013100

Bell AC, West AG, Felsenfeld G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98:** 387–396. doi:10.1016/S0092-8674(00)81967-4

Bolt CC, Lopez-Delisle L, Mascrez B, Duboule D. 2021. Mesomelic dysplasias associated with the hoxd locus are caused by regulatory reallocations. *Nat Commun* **12:** 5013. doi:10.1038/s41467-021-25330-y

Carapuco M, Novoa A, Bobola N, Mallo M. 2005. Hox genes specify vertebral types in the presomitic mesoderm. *Genes Dev* **19:** 2116–2121. doi:10.1101/gad.338705

Chen F, Capecchi MR. 1999. Paralogous mouse Hox genes, Hoxa9, Hoxb9, and Hoxd9, function together to control development of the mammary gland in response to pregnancy. *Proc Natl Acad Sci* **96:** 541–546. doi:10.1073/pnas.96.2.541

Cheng H, Zhang N, Pati D. 2020. Cohesin subunit RAD21: from biology to disease. *Gene* **758:** 144966. doi:10.1016/j.gene.2020.144966

Chung JH, Whiteley M, Felsenfeld G. 1993. A 5′ element of the chicken β-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* **74:** 505–514. doi:10.1016/0092-8674(93)80052-G

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10:** giab008. doi:10.1093/gigascience/giab008

Darbellay F, Duboule D. 2016. Topological domains, metagenes, and the emergence of pleiotropic regulations at Hox loci. *Curr Top Dev Biol* **116:** 299–314. doi:10.1016/bs.ctdb.2015.11.022

Darbellay F, Bochaton C, Lopez-Delisle L, Mascrez B, Tschopp P, Delpretti S, Zakany J, Duboule D. 2019. The constrained architecture of mammalian *Hox* gene clusters. *Proc Natl Acad Sci* **116:** 13424–13433. doi:10.1073/pnas.1904602116

Davidson IF, Bauer B, Goetz D, Tang W, Wutz G, Peters J-M. 2019. DNA loop extrusion by human cohesin. *Science* **366:** 1338–1345. doi:10.1126/science.aaz3418

Dekker J. 2006. The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods* **3:** 17–21. doi:10.1038/nmeth823

Deschamps J, Duboule D. 2017. Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes Dev* **31:** 1406–1416. doi:10.1101/gad.303123.117

Despang A, Schöpflin R, Franke M, Ali S, Jerković I, Paliou C, Chan W-L, Timmermann B, Wittler L, Vingron M, et al. 2019. Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet* **51:** 1263–1271. doi:10.1038/s41588-019-0466-z

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485:** 376–380. doi:10.1038/nature11082

Dixon JR, Gorkin DU, Ren B. 2016. Chromatin domains: the unit of chromosome organization. *Mol Cell* **62:** 668–680. doi:10.1016/j.molcel.2016.05.018

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21. doi:10.1093/bioinformatics/bts635

Dolle P, Izpisua-Belmonte JC, Brown JM, Tickle C, Duboule D. 1991. HOX-4 genes and the morphogenesis of mammalian genitalia. *Genes Dev* **5:** 1767–1767. doi:10.1101/gad.5.10.1767

Dryden NH, Broome LR, Dudbridge F, Johnson N, Orr N, Schoenfelder S, Nagano T, Andrews S, Wingett S, Kozarewa I, et al. 2014. Unbiased analysis of potential targets of breast cancer susceptibility loci by capture Hi-C. *Genome Res* **24:** 1854–1868. doi:10.1101/gr.175034.114

Favier B, Dollé P. 1997. Developmental functions of mammalian Hox genes. *Mol Hum Reprod* **3:** 115–131. doi:10.1093/molehr/3.2.115

Fernandez-Guerrero M, Yakushiji-Kaminatsui N, Lopez-Delisle L, Zdral S, Darbellay F, Perez-Gomez R, Bolt CC, Sanchez-Martin MA, Duboule D, Ros MA. 2020. Mammalian-specific ectodermal enhancers control the expression of *Hoxc* genes in developing nails and hair follicles. *Proc Natl Acad Sci* **117:** 30509–30519. doi:10.1073/pnas.2011078117

Franco MM, Prickett AR, Oakey RJ. 2014. The role of CCCTC-binding factor (CTCF) in genomic imprinting, development, and reproduction1. *Biol Reprod* **91:** 125. doi:10.1095/biolreprod.114.122945

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation of chromosomal domains by loop extrusion. *Cell Rep* **15:** 2038–2049. doi:10.1016/j.celrep.2016.04.085

Gaunt S, Sharpe PT, Duboule D. 1988. Spatially restricted domains of homeo-gene transcripts in mouse embryos: relation to a segmented body plan. *Development* **104:** 169–179. doi:10.1242/dev.104.Supplement.169

Gerard M, Chen JY, Gronemeyer H, Chambon P, Duboule D, Zakany J. 1996. In vivo targeted mutagenesis of a regulatory element required for positioning the Hoxd-11 and Hoxd-10 expression boundaries. *Genes Dev* **10:** 2326–2334. doi:10.1101/gad.10.18.2326

Ghasemi R, Struthers H, Wilson ER, Spencer DH. 2021. Contribution of CTCF binding to transcriptional activity at the HOXA locus in NPM1-mutant AML cells. *Leukemia* **35:** 404–416. doi:10.1038/s41375-020-0856-3

Godwin AR, Capecchi MR. 1998. Hoxc13 mutant mice lack external hair. *Genes Dev* **12:** 11–20. doi:10.1101/gad.12.1.11

Gómez-Marín C, Tena JJ, Acemel RD, López-Mayorga M, Naranjo S, de la Calle-Mustienes E, Maeso I, Beccari L, Aneas I, Vielmas E, et al. 2015. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc Natl Acad Sci* **112:** 7542–7547. doi:10.1073/pnas.1505463112

Gosalia N, Neems D, Kerschner JL, Kosak ST, Harris A. 2014. Architectural proteins CTCF and cohesin have distinct roles in modulating the higher order structure and expression of the CFTR locus. *Nucleic Acids Res* **42:** 9612–9622. doi:10.1093/nar/gku648

Guerreiro I, Gitto S, Novoa A, Codourey J, Nguyen Huynh TH, Gonzalez F, Milinkovitch MC, Mallo M, Duboule D. 2016. Reorganisation of Hoxd regulatory landscapes during the evolution of a snake-like body plan. *eLife* **5:** e16087. doi:10.7554/eLife.16087

Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng H-L, Hansen E, Despo O, Bossen C, Vettermann C, et al. 2011. CTCF-binding elements mediate control of V(D)J recombination. *Nature* **477:** 424–430. doi:10.1038/nature10495

Guo Y, Monahan K, Wu H, Gertz J, Varley KE, Li W, Myers RM, Maniatis T, Wu Q. 2012. CTCF/cohesin-mediated DNA looping is required for protocadherin promoter choice. *Proc Natl Acad Sci* **109:** 21081–21086. doi:10.1073/pnas.1219280110

Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, et al. 2015. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162:** 900–910. doi:10.1016/j.cell.2015.07.038

Hansen AS, Pustova I, Cattoglio C, Tjian R, Darzacq X. 2017. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6:** e25776. doi:10.7554/eLife.25776

Hashimoto H, Wang D, Horton JR, Zhang X, Corces VG, Cheng X. 2017. Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol Cell* **66:** 711–720.e3. doi:10.1016/j.molcel.2017.05.004

Herold M, Bartkuhn M, Renkawitz R. 2012. CTCF: insights into insulator function during development. *Development* **139:** 1045–1057. doi:10.1242/dev.065268

Hintermann A, Guerreiro I, Bolt CC, Gitto S, Duboule D, Beccari L. 2021. Developmental and evolutionary comparative analysis of a hoxd regulatory landscape in mammals and birds. *bioRxiv*doi:10.1101/2021.08.12.456039

Hoegg S, Meyer A. 2005. Hox clusters as models for vertebrate genome evolution. *Trends Genet* **21:** 421–424. doi:10.1016/j.tig.2005.06.004

Holland PW, Garcia-Fernandez J, Williams NA, Sidow A. 1994. Gene duplications and the origins of vertebrate development. *Dev Suppl* **1994:** 125–133.

Huang H, Zhu Q, Jussila A, Han Y, Bintu B, Kern C, Conte M, Zhang Y, Bianco S, Chiariello AM, et al. 2021. CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat Genet* **53:** 1064–1074. doi:10.1038/s41588-021-00863-6

Ibn-Salem J, Muro EM, Andrade-Navarro MA. 2017. Co-regulation of paralog genes in the three-dimensional chromatin architecture. *Nucleic Acids Res* **45:** 81–91. doi:10.1093/nar/gkw813

Jia Z, Li J, Ge X, Wu Y, Guo Y, Wu Q. 2020. Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer–promoter selection. *Genome Biol* **21:** 75. doi:10.1186/s13059-020-01984-7

Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, Roller M, Flicek P. 2020. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol* **21:** 5. doi:10.1186/s13059-019-1894-x

Kessel M, Gruss P. 1991. Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. *Cell* **67:** 89–104. doi:10.1016/0092-8674(91)90574-I

Kim Y, Shi Z, Zhang H, Finkelstein IJ, Yu H. 2019. Human cohesin compacts DNA by loop extrusion. *Science* **366:** 1345–1349. doi:10.1126/science.aaz4475

Kraft K, Magg A, Heinrich V, Riemenschneider C, Schöpflin R, Markowski J, Ibrahim DM, Acuna-Hidalgo R, Despang A, Andrey G, et al. 2019. Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat Cell Biol* **21:** 305–310. doi:10.1038/s41556-019-0273-x

Krefting J, Andrade-Navarro MA, Ibn-Salem J. 2018. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biol* **16:** 87. doi:10.1186/s12915-018-0556-x

Krumlauf R. 1994. Hox genes in vertebrate development. *Cell* **78:** 191–201. doi:10.1016/0092-8674(94)90290-9

Kubo N, Ishii H, Xiong X, Bianco S, Meitinger F, Hu R, Hocker JD, Conte M, Gorkin D, Yu M, et al. 2021. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene acti-

vation. *Nat Struct Mol Biol* **28:** 152–161. doi:10.1038/s41594-020-00539-5

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. doi:10.1038/nmeth.1923

Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, Goodwin GH. 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′-flanking sequence of the chicken c-myc gene. *Oncogene* **5:** 1743–1753.

Long HK, Prescott SL, Wysocka J. 2016. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167:** 1170–1187. doi:10.1016/j.cell.2016.09.018

Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, Ramírez F, Manke T. 2021. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37:** 422–423. doi:10.1093/bioinformatics/btaa692

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8

Luan J, Xiang G, Gómez-García PA, Tome JM, Zhang Z, Vermunt MW, Zhang H, Huang A, Keller CA, Giardine BM, et al. 2021. Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. *Cell Rep* **34:** 108783. doi:10.1016/j.celrep.2021.108783

Luo M, Yu Y, Kim H, Kudrna D, Itoh Y, Agate RJ, Melamed E, Goicoechea JL, Talag J, Mueller C, et al. 2006. Utilization of a zebra finch BAC library to determine the structure of an avian androgen receptor genomic region. *Genomics* **87:** 181–190. doi:10.1016/j.ygeno.2005.09.005

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161:** 1012–1025. doi:10.1016/j.cell.2015.04.004

Mahony S, Benos PV. 2007. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res* **35:** W253–W258. doi:10.1093/nar/gkm272

Mahony S, Auron PE, Benos PV. 2007. DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies. *PLoS Comput Biol* **3:** e61. doi:10.1371/journal.pcbi.0030061

Mallo M, Wellik DM, Deschamps J. 2010. Hox genes and regional patterning of the vertebrate body plan. *Dev Biol* **344:** 7–15. doi:10.1016/j.ydbio.2010.04.024

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17:** 10. doi:10.14806/ej.17.1.200

Merkenschlager M, Odom DT. 2013. CTCF and cohesin: linking gene regulatory elements with their targets. *Cell* **152:** 1285–1297. doi:10.1016/j.cell.2013.02.029

Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D. 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147:** 1132–1145. doi:10.1016/j.cell.2011.10.023

Nakahashi H, Kwon K-RK, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson S, Yamane A, et al. 2013. A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep* **3:** 1678–1689. doi:10.1016/j.celrep.2013.04.024

Nanni L, Ceri S, Logie C. 2020. Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. *Genome Biol* **21:** 197. doi:10.1186/s13059-020-02108-x

Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. 2015. CTCF establishes discrete functional chromatin domains at the *Hox* clusters during differentiation. *Science* **347:** 1017–1021. doi:10.1126/science.1262088

Narendra V, Bulajić M, Dekker J, Mazzoni EO, Reinberg D. 2016. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev* **30:** 2657–2662. doi:10.1101/gad.288324.116

Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D. 2011. The dynamic architecture of Hox gene clusters. *Science* **334:** 222–225. doi:10.1126/science.1207194

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485:** 381–385. doi:10.1038/nature11049

Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. 2017. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169:** 930–944.e22. doi:10.1016/j.cell.2017.05.004

Paliou C, Guckelberger P, Schöpflin R, Heinrich V, Esposito A, Chiariello AM, Bianco S, Annunziatella C, Helmuth J, Haas S, et al. 2019. Preformed chromatin topology assists transcriptional robustness of *Shh* during limb development. *Proc Natl Acad Sci* **116:** 12390–12399. doi:10.1073/pnas.1900672116

Papageorgiou S. 2006. Pulling forces acting on Hox gene clusters cause expression collinearity. *Int J Dev Biol* **50:** 301–308. doi:10.1387/ijdb.052034sp

Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong C-T, Hookway TA, Guo C, Sun Y, et al. 2013. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153:** 1281–1295. doi:10.1016/j.cell.2013.04.053

Pugacheva EM, Kubo N, Loukinov D, Tajmul M, Kang S, Kovalchuk AL, Strunnikov AV, Zentner GE, Ren B, Lobanenkov VV. 2020. CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc Natl Acad Sci* **117:** 2020–2031. doi:10.1073/pnas.1911708117

Quinlan AR. 2014. BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinforma* **47:** 11.12.1–11.12.34. doi:10.1002/0471250953.bi1112s47

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44:** W160–W165. doi:10.1093/nar/gkw257

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9:** 189. doi:10.1038/s41467-017-02525-w

Ren G, Jin W, Cui K, Rodriez J, Hu G, Zhang Z, Larson DR, Zhao K. 2017. CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. *Mol Cell* **67:** 1049–1058.e6. doi:10.1016/j.molcel.2017.08.026

Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, Felsenfeld G, Pedone PV. 2007. Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger–DNA interaction controls binding at imprinted loci. *J Biol Chem* **282:** 33336–33345. doi:10.1074/jbc.M706213200

Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. 2011. Improving RNA-seq expression estimates by correcting for fragment bias. *Genome Biol* **12:** R22. doi:10.1186/gb-2011-12-3-r22

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

Rodríguez-Carballo E, Lopez-Delisle L, Willemin A, Beccari L, Gitto S, Mascrez B, Duboule D. 2020. Chromatin topology and the timing of enhancer function at the *HoxD* locus. *Proc Natl Acad Sci* **117:** 31231–31241. doi:10.1073/pnas.2015083117

Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, et al. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci* **112:** E6456–E6465. doi:10.1073/pnas.1518552112

Schep R, Necsulea A, Rodríguez-Carballo E, Guerreiro I, Andrey G, Nguyen Huynh TH, Marcet V, Zákány J, Duboule D, Beccari L. 2016. Control of *Hoxd* gene transcription in the mammary bud by hijacking a preexisting regulatory landscape. *Proc Natl Acad Sci* **113:** E7720–E7729. doi:10.1073/pnas.1617141113

Sedeño Cacciatore Á, Rowland BD. 2019. Loop formation by SMC complexes: turning heads, bending elbows, and fixed anchors. *Curr Opin Genet Dev* **55:** 11–18. doi:10.1016/j.gde.2019.04.010

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

Shao Z, Zhang Y, Yuan G-C, Orkin SH, Waxman DJ. 2012. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol* **13:** R16. doi:10.1186/gb-2012-13-3-r16

Sikorska N, Sexton T. 2020. Defining functionally relevant spatial chromatin domains: it is a TAD complicated. *J Mol Biol* **432:** 653–664. doi:10.1016/j.jmb.2019.12.006

Soshnikova N, Montavon T, Leleu M, Galjart N, Duboule D. 2010. Functional analysis of CTCF during mammalian limb development. *Dev Cell* **19:** 819–830. doi:10.1016/j.devcel.2010.11.009

Soshnikova N, Dewaele R, Janvier P, Krumlauf R, Duboule D. 2013. Duplications of hox gene clusters and the emergence of vertebrates. *Dev Biol* **378:** 194–199 doi:10.1016/j.ydbio.2013.03.004

Spitz F, Furlong EE. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13:** 613–626. doi:10.1038/nrg3207

Stigler J, Çamdere GÖ, Koshland DE, Greene EC. 2016. Single-molecule imaging reveals a collapsed conformational state for DNA-bound cohesin. *Cell Rep* **15:** 988–998. doi:10.1016/j.celrep.2016.04.003

Su G, Wang W, Chen J, Liu M, Zheng J, Guo D, Bi J, Zhao Z, Shi J, Zhang L, et al. 2021. CTCF-binding element regulates ESC differentiation via orchestrating long-range chromatin interaction between enhancers and HoxA. *J Biol Chem* **296:** 100413. doi:10.1016/j.jbc.2021.100413

Symmons O, Pan L, Remeseiro S, Aktas T, Klein F, Huber W, Spitz F. 2016. The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev Cell* **39:** 529–543. doi:10.1016/j.devcel.2016.10.015

Tarchini B, Huynh TH, Cox GA, Duboule D. 2005. HoxD cluster scanning deletions identify multiple defects leading to paralysis in the mouse mutant Ironside. *Genes Dev* **19:** 2862–2876. doi:10.1101/gad.351105

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript

assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:** 511–515. doi:10.1038/nbt.1621

Tsai Y-C, Cooke NE, Liebhaber SA. 2016. Long-range looping of a locus control region drives tissue-specific chromatin packing within a multigene cluster. *Nucleic Acids Res* **44:** 4651–4664. doi:10.1093/nar/gkw090

Tschopp P, Christen AJ, Duboule D. 2012. Bimodal control of Hoxd gene transcription in the spinal cord defines two regulatory subclusters. *Development* **139:** 929–939. doi:10.1242/dev.076794

Ushiki A, Zhang Y, Xiong C, Zhao J, Georgakopoulos-Soares I, Kane L, Jamieson K, Bamshad MJ, Nickerson DA, University of Washington Center for Mendelian Genomics, et al. 2021. Deletion of CTCF sites in the SHH locus alters enhancer–promoter interactions and leads to acheiropodia. *Nat Commun* **12:** 2282. doi:10.1038/s41467-021-22470-z

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. 2015. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* **10:** 1297–1309. doi:10.1016/j.celrep.2015.02.004

Wan L-B, Pan H, Hannenhalli S, Cheng Y, Ma J, Fedoriw A, Lobanenkov V, Latham KE, Schultz RM, Bartolomei MS. 2008. Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development. *Development* **135:** 2729–2738. doi:10.1242/dev.024539

Wellik DM. 2009. Hox genes and vertebrate axial pattern. *Curr Top Dev Biol* **88:** 257–278. doi:10.1016/S0070-2153(09)88009-5

Wellik DM, Capecchi MR. 2003. Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. *Science* **301:** 363–367. doi:10.1126/science.1085672

Willemin A, Lopez-Delisle L, Bolt CC, Gadolini M-L, Duboule D, Rodríguez-Carballo E. 2021. Context-independent function of a chromatin boundary in vivo. *PLoS Genet* **17:** e1009691. doi:10.1371/journal.pgen.1009691

Wolff J, Bhardwaj V, Nothjunge S, Richard G, Renschler G, Gilsbach R, Manke T, Backofen R, Ramírez F, Grüning BA. 2018. Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res* **46:** W11–W16. doi:10.1093/nar/gky504

Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, Grüning BA. 2020. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res* **48:** W177–W184. doi:10.1093/nar/gkaa220

Woltering JM, Vonk FJ, Müller H, Bardine N, Tuduce IL, de Bakker MA, Knöchel W, Sirbu IO, Durston AJ, Richardson MK. 2009. Axial patterning in snakes and caecilians: evidence for an alternative interpretation of the Hox code. *Dev Biol* **332:** 82–89. doi:10.1016/j.ydbio.2009.04.031

Xi W, Beer MA. 2021. Loop competition and extrusion model predicts CTCF interaction specificity. *Nat Commun* **12:** 1046. doi:10.1038/s41467-021-21368-0

Yakushiji-Kaminatsui N, Lopez-Delisle L, Bolt CC, Andrey G, Beccari L, Duboule D. 2018. Similarities and differences in the regulation of HoxD genes during chick and mouse limb development. *PLoS Biol* **16:** e3000004. doi:10.1371/journal.pbio.3000004

Yin M, Wang J, Wang M, Li X, Zhang M, Wu Q, Wang Y. 2017. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res* **27:** 1365–1377. doi:10.1038/cr.2017.131

Young T, Rowland JE, van de Ven C, Bialecka M, Novoa A, Carapuco M, van Nes J, de Graaff W, Duluc I, Freund JN, et al. 2009. Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev Cell* **17:** 516–526. doi:10.1016/j.devcel.2009.08.010

Zakany J, Duboule D. 2007. The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* **17:** 359–366. doi:10.1016/j.gde.2007.05.011

Zákány J, Gérard M, Favier B, Potter SS, Duboule D. 1996. Functional equivalence and rescue among group 11 Hox gene products in vertebral patterning. *Dev Biol* **176:** 325–328. doi:10.1006/dbio.1996.0137

Ziebarth JD, Bhattacharya A, Cui Y. 2012. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res* **41:** D188–D194. doi:10.1093/nar/gks1165

Sequential in *cis* mutagenesis in vivo reveals various functions for CTCF sites at the mouse *HoxD* cluster

Ana Rita Amândio, Leonardo Beccari, Lucille Lopez-Delisle, et al.

Supplemental Material	http://genesdev.cshlp.org/content/suppl/2021/10/26/gad.348934.121.DC1
Related Content	**Dissecting CTCF site function in a tense HoxD locus** Benoit G. Bruneau Genes Dev. November , 2021 35: 1401-1402
References	This article cites 119 articles, 33 of which can be accessed free at: **http://genesdev.cshlp.org/content/35/21-22/1490.full.html#ref-list-1** Articles cited in: **http://genesdev.cshlp.org/content/35/21-22/1490.full.html#related-urls**
Creative Commons License	This article, published in *Genes & Development*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**