A COMPLEX REGULATORY LANDSCAPE INVOLVED IN THE DEVELOPMENT OF EXTERNAL GENITALS

Short title: *Hox* gene regulation during the development of genitals

Ana Rita Amândio[1], Lucille Lopez-Delisle[1], Christopher Chase Bolt[1], Bénédicte Mascrez[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, 30 quai Ernest Ansermet, 1211, Geneva, Switzerland. [3]Collège de France, Paris, France

Key words: TADs, *Hox* clusters, chromatin conformation, CTCF, developmental enhancers, loop extrusion model

*Corresponding author:
Denis Duboule
denis.duboule@epfl.ch
denis.duboule@unige.ch

ABSTRACT

In vertebrates, developmental genes are often controlled by large regulatory landscapes matching the dimensions of topologically associating domains (TADs). In various ontogenic contexts, the associated constitutive chromatin backbone is modified by fine-tuned specific variations in enhancer-enhancer and enhancer-promoter interaction profiles. In this work, we take one of the TADs flanking the *HoxD* gene cluster as a paradigm to address the question of how these complex regulatory architectures are formed and how they are de-constructed once their function has been achieved. We suggest that this TAD can be considered as a coherent functional unit in itself, with several regulatory sequences acting together to elicit a transcriptional response. With one notable exception, the deletion of each of these sequences in isolation did not produce any substantial modification in the global transcriptional outcome of the system, a result at odds with a conventional view of long-range enhancer function. Likewise, both the deletion and inversion of a supposedly critical CTCF site located in a region rich in such sequences did not affect transcription of the target gene. In the latter case, however, slight modifications were observed in interaction profiles *in vivo* in agreement with the loop extrusion model, despite no apparent functional consequences. We discuss these unexpected results by considering both conventional explanations and an alternative possibility whereby a rather unspecific accumulation of particular factors within the TAD backbone may have a global impact upon transcription.

INTRODUCTION

During mammalian development, the organization of body structures and their morphogenesis require the accurate transcriptional regulation of the *Hox* gene family of transcription factors. These proteins instruct progenitor cells, at different levels along the main anterior to posterior (AP), about their developmental fates. In addition to this ancient role in trunk patterning, subsets of the four *Hox* gene clusters were co-opted during evolution to promote the development of secondary body axes such as the limbs and the external genitalia (Dolle et al., 1991b). In the latter case, mice lacking both *Hoxa13* and *Hoxd13* function fail to develop external genitalia due to a complete agenesis of the genital tubercle (GT) (Kondo et al., 1997; Warot et al., 1997).

In the case of the *HoxD* cluster, the control of gene transcription in the emerging GT involves *cis*-regulatory sequences located in a 700kb regulatory landscape positioned 5' to the cluster, referred to as centromeric regulatory landscape (C-DOM). (Andrey et al., 2013; Montavon et al., 2011; Spitz et al., 2003). This landscape matches one of the two topologically-associating domains (TADs), which flank the gene cluster. The functional importance of the C-DOM was confirmed by *in vivo* chromosome engineering studies. For example, when this region was repositioned 3Mb away from *HoxD*, transcription of *Hoxd13* in the GT was almost entirely abolished (Tschopp and Duboule, 2011) and subsequent deletions spanning various parts of C-DOM supported this conclusion (Lonfat et al., 2014). Genetic and biochemical analyses have shown that this entire regulatory landscape is shared between GT and digits, and contains multiple enhancer sequences that are active in either both or in only one of these developing structures (Gonzalez et al., 2007; Lonfat et al., 2014; Montavon et al., 2011). Overall, it thus appears that within a large constitutive TAD structure, subtle yet specific modifications of chromatin architecture are formed either in GT or in digit cells (Lonfat and Duboule, 2015).

Unlike the opposite regulatory landscape (T-DOM), which includes a large variety of enhancers with distinct specificities regulating the 'anterior' part of the *HoxD* cluster, the C-DOM appears to be devoted to the control of the most posterior and distal terminal body structures by regulating mostly *Hoxd13* either in digit cells or in the GT. The tropism of C-DOM enhancers for *Hoxd13* results from the presence of a strong chromatin boundary between this target gene and the rest of the cluster (Rodriguez-Carballo et al., 2017). Over the past years, the importance of the C-DOM in controlling *Hoxd* genes expression has been clearly

97 demonstrated. However, both the dynamic behavior of such a regulatory landscape i.e. its
98 implementation and decommissioning, as well as the functional contribution of specific *cis*-
99 elements in these processes remained to be established in order to understand how an entire
100 TAD can be transcriptionally mobilized in different morphogenetic contexts to achieve similar
101 regulatory outcomes. A 'specific' view of the regulatory system would involve discriminative
102 factors, progressively building a tissue-specific chromatin context with a deterministic strategy.
103 Alternatively, a more generic process could be considered, where the accumulation of various
104 factors available in different tissues would elicit the same transcriptional response through
105 whichever chromatin configuration they would trigger.

106 In this work, we tackled these issues by studying both the *HoxD* locus chromatin
107 conformation dynamics during GT development, as well as the functional contribution of
108 specific *cis*-elements to *Hoxd* genes regulation. We observed that the gross chromatin
109 organization of C-DOM predates the appearance of the GT. As GT development progresses,
110 we observed a reduction in transcript levels correlating with a decrease in enhancer-promoter
111 chromatin loops within the C-DOM. This decrease occurred while maintaining a subset of
112 CTCF associated contacts, which are preserved independently from the transcriptional status
113 of the gene cluster. While both the deletion of the *Prox* enhancer and deletions of clusters of
114 enhancers severely affected *Hoxd* genes transcript levels, deletions of most other enhancers in
115 isolation had little (if any) effect on transcription in the GT. Moreover, the deletion of a
116 conserved CTCF site, the only one present in the central part of the regulatory landscape, did
117 not impact the transcriptional outcome, even though its inversion reallocated contacts in a
118 manner compatible with the loop extrusion model (Fudenberg et al., 2016; Rao et al., 2014;
119 Vian et al., 2018). These results point to a high resilience of the regulatory strategy at work in
120 this locus. They also suggest the existence in the same TAD of distinct mechanisms to control
121 target gene activation, either relying upon sequence specific enhancer-promoter interactions,
122 or involving less deterministic parameters and using the underlying chromatin structure.
123
124 **RESULTS**
125
126 ***Hox* genes and GT development**

127 To precisely assess *Hox* genes transcription during GT development, we initially
128 quantified their expression levels by using RNA-sequencing (RNA-seq). We analyzed datasets
129 from three different stages of GT embryonic development starting from embryonic day 12.5

(E12.5), E16.5 and E18.5. We observed that genes positioned in the 5' portion of both *HoxA* (*Hoxa7* to *Hoxa13*) and *HoxD* (*Hoxd8* to *Hoxd13*) clusters were expressed at all developmental stages (Figure 1A and Figure 1–figure supplement 1). Furthermore, with the exception of *Hoxc11* and *Hoxc10*, only basal levels of mRNAs were scored for the *HoxC* and *HoxB* clusters (Figure 1–figure supplement 1), consistent with previous observations (Hostikka and Capecchi, 1998; Montavon et al., 2008). Overall, we detected a general decrease in the amount of *Hox* mRNAs during GT development, in particular for *Hoxd12* and *Hoxd13* (Figure 1A).

To try and define the full dynamics of *Hoxd* transcript accumulation during GT development, we micro-dissected the cloaca region (CR) at E10.5, the major contributing embryonic tissue to the emergence of the GT (Georgas et al., 2015), as well as genital buds at E12.3, E13.5, E15.5, E16.5, E17.5 and E18.5. We performed RT-qPCR for *Hoxd13* and detected transcripts in the CR at E10.5 (Figure 1B), followed by a significant increase in transcript levels between the CR and the E12.5 GT ($p<0.0001$). The mRNA levels then remained constant between E12.5 and E13.5, whereas they were significantly reduced in E13.5 and E15.5 GTs ($p<0.0001$). After E15.5, the transcript levels continued to decrease yet to a lesser extent (between E15.5 and E18.5; $p= 0.0175$, Figure 1B), confirming the RNA-seq results (Figure 1A).

We next compared chromatin accessibility and selected histone modifications in three developmental stages to correlate with transcript levels. We used the CR at E10.5 (prior to GT formation, low *Hoxd13* expression), GT at E13.5 (early GT development, high *Hoxd13* expression) and GT at E17.5 (late GT development, low *Hoxd13* expression) and performed ATAC-seq and ChIP-seq for both H3K27ac and H3K27me3 chromatin marks. At E10.5, prior to GT formation, all *Hoxd* genes and *Evx2* were accessible as defined by ATAC-seq (Figure 1C). H3K27ac signals of moderate intensity were scored over the *Hoxd9* to *Evx2* interval as well as peaks on the promoters of *Hoxd1*, *Hoxd3, and Hoxd4* (Figure 1C), indicating a somewhat general activity of *Hoxd* genes in this region of the body axis. This was confirmed by a low coverage in H3K27me3 marks, which were detected mostly over the *Evx2* gene flanking the *Hox* cluster (Figure 1C, gray area).

At E13.5, in the growing genital bud, a different picture was observed with a whole inactivation of the cluster from *Hoxd1* to *Hoxd10-11*, as indicated by a robust coverage of this region by H3K27me3 marks and the disappearance of H3K27ac marks and ATAC-seq signals (Figure 1D). In contrast, ATAC-seq peaks remained in the *Hoxd11* to *Evx2* region, accompanied by a large increase in H3K27ac signals (Figure 1D) reflecting full transcription of the latter genes. At this stage, a clear separation of the cluster into two distinct epigenetic

domains was scored, reminiscent of the situation described in distal forelimb buds (Andrey et al., 2013). At E17.5, this clear dichotomy between epigenetic domains was still detected for H3K27ac signals, though at a lower magnitude, but started to vanish when H3K27me3 marks were considered, with their progressive spreading over the entire gene cluster. These data are in agreement with the analysis of mRNA levels as observed by both RNA-seq and RT-qPCR.

Implementation and decommissioning of a chromatin architecture

Hoxd genes are regulated in the developing GT by long-range acting sequences positioned within the flanking, centromeric-located TAD (C-DOM; Figure 2A). To assess the dynamics of the TAD structure during bud development, we used circularized chromosome conformation capture combined with high-throughput sequencing (4C-seq) to reveal the physical chromatin interactions established between *Hoxd13* and the C-DOM, at various developmental stages. *Hoxd13* was selected as a viewpoint since it is the highest expressed *Hoxd* gene in this tissue and because its disruption leads to alterations in external genitals (Dolle et al., 1993; Kondo et al., 1997; Warot et al., 1997). We micro-dissected CR at E10.5 and GTs at E12.5, E13.5, E15.5, and E17.5, and used forebrain at E12.5 as a control tissue lacking *Hoxd* mRNA. As a baseline to our temporal series, we used a mouse embryonic stem cells (mESC) dataset (Noordermeer et al., 2014) assuming that these cells somehow reflect the ground-state 3D architecture of the gene cluster.

In mESC, contacts between *Hoxd13* and the C-DOM were mainly scored in the island II and V regions. A large proportion of the interactions was scored in the cluster itself (Figure 2B, top, red lines) where they were likely driven by H3K27me3 marks (Vieux-Rochas et al., 2015). This 3D architecture was altogether quite comparable to that found in forebrain cells with discrete contacts established between *Hoxd13* and island II and V (Figure 2B, bottom, red lines). These two profiles likely reflected the 3D chromatin state of C-DOM in the complete absence of transcription. Upon transcriptional activation, however, frequencies of contacts with the C-DOM increased and interactions between *Hoxd13* and previously characterized enhancers (Prox, GT2) (Gonzalez et al., 2007; Lonfat et al., 2014) became visible (Figure 2B, second track, blue lines). Quantification of these interactions revealed a 22% increase in overall contacts over this regulatory region, when the CR at E10.5 was compared with ES cells (Figure 2B). This dataset showed a C-DOM specific chromatin architecture that is organized before the emergence of the genital bud.

In subsequent stages of GT development (E12.5 or E13.5), contacts between various enhancer regions and *Hoxd13* continued to increase to reach a maximum at E13.5 with an

198 additional 35% of overall interaction when compared to the CR sample (Figure 2B). As
199 development further progressed, contacts established between *Hoxd13* and C-DOM weakened.
200 From E13.5 to E17.5, there was a 28.1% decrease in interactions. At the latter stage the profile
201 observed was comparable to either forebrain cells or the mESC profiles, with a loss of contacts
202 with specific enhancers (Prox and GT2; Figure 2B). We quantified the percent of fragments
203 covering each regulatory island by using mESC as a reference (Figure 2–figure supplement
204 2A). The relative frequency of contacts with island II and island V remained fairly constant in
205 all samples analyzed. In contrast, the contacts between *Hoxd13* and either Prox or GT2
206 dramatically increased from the mESC to the E13.5 GT samples. The decrease in contacts
207 observed between E13.5 to E17.5 GTs correlated with a decrease in *Hoxd13* transcript levels.
208 Fetal forebrain cells, which do not express any *Hoxd* genes, showed the lowest values of
209 interactions between *Hoxd13* and either Prox or GT2 (Figure 2–figure supplement 2A).
210 To validate these results, we selected both the GT2 region, which displayed important
211 changes in interaction frequencies with *Hoxd13* during GT development, and the island V
212 region which showed more constitutive contacts, as viewpoints in 4C-seq experiments. We
213 used 4C-seq libraries for E12.5, E13.5, E15.5, and E17.5 GT cells and for E12.5 forebrain cells
214 as negative control. We confirmed that the interactions between GT2 and the *Hoxd13* region
215 substantially decreased from E13.5 to E17.5, whereas contact frequencies between island V
216 and *Hoxd13* was essentially stable, regardless of the stage and tissue analyzed (Figure 2–figure
217 supplement 2B). Therefore, as transcription decreased, some contacts established with C-DOM
218 were lost whereas others were maintained (island II and island V), indicating that at the time
219 transcription is switched off, C-DOM goes back to the pre-organized chromatin backbone that
220 characterizes tissues or cells that do not express any *Hox* genes. Of note, the constitutive contact
221 regions include binding sites occupied by CTCF (see below), a protein known to facilitate
222 enhancer-promoter contacts by DNA-looping (see (Ong and Corces, 2014).
223
224 **Dissecting the regulatory potential of the C-DOM TAD**
225 We next explored the functional dynamics of C-DOM during GT development. A
226 detailed analysis of our CR ATAC-seq and ChIP-seq datasets revealed several accessible
227 chromatin sites, some of which correspond to previously identified GT enhancers such as GT2
228 (Lonfat et al., 2014) (Figure 3A, black arrow). Noteworthy, the GT and limb enhancer sequence
229 Prox was not yet accessible at this stage (Figure 3A, red arrow). At E13.5, when C-DOM is
230 fully active, both chromatin accessibility peaks and H3K27ac marks were scored over
231 previously characterized enhancers within this region, including Prox and GT2 (Figure 3A).

As development progressed, in E17.5 GT, H3K27ac marks were lost in C-DOM (Figure 3A) correlating with the loss of both *Hoxd* transcripts and chromatin interactions (see above).

To evaluate the functional importance of sub-regions of C-DOM for the transcriptional control of *Hoxd* genes during GT development, we used a series of partial deletions, in particular the *Del(rel1-rel5), Del(rel5-SB)* and the *Del(SB-Atf2)* alleles (Figure 3A, bottom) as well as the *Del(IV-SB)* allele corresponding to a deletion between island IV and SB (Figure 3A, bottom). The latter allele, a 154 kb large deficiency, removed half of the regulatory region between the *rel5* and *SB* breakpoints and contained three GT regulatory regions, E1, IIIE and IVE (see below). We analyzed the effect of each of these four deletions on *Hoxd* genes transcription by RT-qPCR at E12.5.

In the *Del(rel1-rel5)* allele, one-third of C-DOM is removed, including two digit and/or GT enhancers (GCR and Prox) (Gonzalez et al., 2007; Spitz et al., 2003) (Figure 3A, bottom). In these mutant mice, a 47% reduction in *Hoxd13* mRNA levels was scored in the GT ($p=0.0005$), whereas, *Hoxd12*, *Hoxd11*, and *Hoxd10* were less affected (Figure 3B). The *Del(rel5-SB)* allele is a 300kb large deletion of C-DOM including the GT2, and island III, IV and V regulatory sequences. Mice carrying this deletion displayed a greater effect on the steady-state level *Hoxd13* mRNAs, which was reduced by 76% ($p<0.0001$). Again, *Hoxd12*, *Hoxd11* and *Hoxd10* were also affected, yet to a lower extent (Figure 3C). We next analyzed the *Del(IV-SB)* allele and noticed a 38% decrease in the amount of *Hoxd13* mRNAs ($p=0.0066$), yet no significant effect was detected for any other genes (Figure 3D). Finally, we looked at the *Del(SB-Atf2)* allele where the most centromeric part of the TAD had been deleted. In these mutant mice, we observed a slight but significant upregulation of *Hoxd13* mRNA levels ($p=0.003$) in the GT, whereas other genes were not affected (Figure 3E). Taken together, these results indicated that several non-overlapping regions located within C-DOM are required for the transcriptional activation of *Hoxd13* in the developing GT.

Deletion of the Prox enhancer sequence

Within the different DNA intervals delimited by our large deletions, we assessed the contribution of single regulatory elements to the control of *Hoxd13* transcription. We applied CRISPR/Cas9 genome editing to fertilized eggs and generated a series of alleles where these elements were either deleted or inverted. We initially focused on the region between the *rel1* and *rel5* breakpoints (Figure 3A, bottom). In this genomic interval the limb- and GT-specific Prox enhancer (Figure 4B) accounted for the majority of chromatin interactions with *Hoxd13*

and presented strong coverage by H3K27ac marks in the GT (Figure 3A). We generated the *Del(Prox)* allele, a micro-deletion of the Prox sequence (Figure 4A), and observed a 36% decrease in the expression of *Hoxd13* by qPCR in E12.5 GTs (p=0.006) (Figure 4C). This severe impact seemed to be exclusively quantitative, as the *Hoxd13* expression pattern detected by whole mount *in situ* hybridization (WISH) remained unchanged (Figure 4D). This result indicated that the Prox enhancer accounts for more than a third of the *Hoxd13* transcriptional efficiency and is thus a major contributor to this regulation in GT.

We then looked at whether this effect was 'enhancer-autonomous' or if it involved a significant reorganization of the entire C-DOM regulatory landscape by performing ATAC-seq and 4C-seq in both control and *Del(Prox)* mutant E13.5 GTs (Figure 4E-F). The ATAC-seq profiles revealed no obvious change in chromatin accessibility throughout the C-DOM after the deletion of Prox (Figure 4E). Likewise, when we examined the potential importance of Prox in building the C-DOM interaction landscape by 4C-seq using *Hoxd13* as a viewpoint, we only noticed minor alterations in the frequency of contacts between *Hoxd13* and discrete *cis*-regulatory elements (Figure 4F). We thus concluded that the Prox enhancer, while of critical importance for regulating *Hoxd13*, does not actively contribute to the general architectural organization of the locus.

Identification of GT-specific enhancers

In order to identify other elements acting in GT, we then focused on the genomic interval positioned between the *SB* and the *rel5* breakpoints (Figure 5A), since this region accounted for 76% of *Hoxd13* expression in the incipient bud (see Figure 3C). Based on ATAC-seq, H3K27ac ChIP-seq, 4C-seq datasets and on DNA sequence conservation, we selected five sub-regions of approximately 30kb in size and tested them for enhancer activity in transgenic assays (Figure 5B, C). Each region was cloned upstream of a *LacZ* reporter gene driven by a minimal *beta-globin* promoter and integrated at random positions in the mouse genome.

X-gal staining of E13.5 transgenic embryos revealed enhancer activity in the GT for the IIIE and IVE sequences (Figure 5C), in cellular territories included within the wider expression domain of *Hoxd13* in this tissue. These two sequences showed complementary specificities, with IIIE active in dorsal GT cells, whereas the IVE sequence strongly labelled the ventral half of the GT (Figure 5C). Embryos injected with the E1 sequence showed a weak only signal on the GT (Figure 5C) and no staining was scored either when using the VE, or the

E2 sequences (Figure 5C), despite their promising chromatin signatures. Of particular interest, the VE region includes a CTCF binding site. These elements are involved in facilitating enhancer-promoter contact by DNA-looping (e.g. (Long et al., 2016) and this particular CTCF binding sequence is the only strongly occupied site present in the ca 550kb-region between *Evx2* and island II.

Therefore, out of the five regions tested, only E1, IIIE, and IVE showed some activity in the developing GT. We also re-investigated the activity of the GT2 sequence in transgenic embryos and scored a strong staining throughout the bud (Figure 5C). These experiments highlighted the regulatory complexity of the C-DOM, with individual enhancer elements displaying distinct and complementary patterns of activity (e.g., IIIE and IVE), while others show largely overlapping domains of expression (e.g., GT2).

Serial deletions of single *cis*-regulatory elements

To further evaluate the regulatory potential of these DNA sequences, we generated deletion alleles for all suspected enhancers located between the *rel5* and *SB* breakpoints. When deleted, this region had the largest impact upon *Hoxd13* transcription (Figure 3C). Therefore, independent mouse strains were produced carrying either a *Del(GT2)*, *Del(IV)* or *Del(IIIE)* allele. In addition, to assess the importance of bound CTCF proteins within island V, we both deleted and inverted this region (*Del(V)* and *Inv(V),* respectively) (Figure 6A). As a read out, we quantified *Hoxd13* mRNA levels by RT-qPCR and examined the transcripts distribution by WISH. Unexpectedly, we did not detect any significant difference, either in transcript levels or in their spatial patterns, in any of the *Del(GT2)*, *Del(IV)*, *Del(IIIE)*, *Del(V)* and *Inv(V)* alleles (Figure 6B, C). Unlike the Prox sequence analyzed above, these results suggest that none of these sequences is in itself functionally important enough to elicit a visible transcriptional effect upon the main target gene, at least in the GT and at the developmental stage considered.

The lack of visible effect of the *Del(GT2)* allele was particularly surprising, for this sequence displayed a strong, highly specific and continuous staining in the GT in transgenic embryos and also because of the robust transcriptional down-regulation obtained when using a larger deletion including it. Consequently, we performed both 4C-seq and ATAC-seq in *Del(GT2)* homozygous GT at E13.5 to assess whether this deletion would at least impact the functional organization of the regulatory landscape (Figure 6–figure supplement 3). Except for the loss of a single accessibility peak located between GT2 and CsB in the *Del(GT2)* mutant allele (Figure 6–figure supplement 3A, black arrows), the distribution of accessible DNA sequences over C-DOM appeared to be independent from the GT2 element (Figure 6–figure

supplement 3). This absence of global impact of the GT2 deletion was confirmed when using a viewpoint on *Hoxd13* to evaluate by 4C-seq, potential reallocations of contacts in the mutant allele. There again, no salient change in the chromatin organization of C-DOM was observed (Figure 6–figure supplement 3B), further indicating that the deletion of GT2 in isolation had essentially no effect on the global architecture of the C-DOM landscape.

This absence of visible impact after deletion of a strong and specific enhancer can be due to a variety of reasons (see the discussion). Amongst them, the possibility that the functional contribution of GT2 is required at a particular stage of GT development, which was not considered in our analyses. To explore this possibility, we used RT qPCR to measure the *Hoxd13* mRNA level in the CR at E10.5, a developmental stage where this enhancer is already accessible, as seen in our ATAC-seq dataset (Figure 6–figure supplement 3C, black arrow), and capable of triggering *lacZ* transcription (Lonfat, 2013). At this early stage, we observed a slight (27%), but significant (p=0.0152) decrease in the expression of *Hoxd13* (Figure 6–figure supplement 3D), suggesting that GT2 alone may have a role in controlling *Hoxd13* expression prior to GT formation.

CTCF and C-DOM chromatin organization

Amongst the various sequences isolated in C-DOM, island V was shown to interact with *Hoxd13* in all tissues and developmental stages analyzed thus far. We used our *Del(V)* and *Inv(V)* alleles to evaluate the importance of this element in ensuring proper 3D-chromatin organization at the *HoxD* locus. We first defined the CTCF binding profile in wildtype and mutant E13.5 GTs, by using both ChIP-seq and Cut & Run (CnR). In the wildtype locus, our Chip-seq results showed several CTCF binding sites in the centromeric part of C-DOM, primarily between island II and *Atf2* and matching with other islands and 4C-seq peaks, close to the centromeric TAD boundary (Figure 7A). Of note, island V was the only region between *Evx2* and island II (approximately 550kb in linear distance) where a clear binding of this protein was detected (Figure 7A, arrow). A close examination of this element revealed a major CTCF binding site oriented towards the cluster and a weaker site observed nearby. In the *HoxD* cluster, the distribution of bound CTCF was as for limb buds cells (Lonfat, 2013), with a series of strong sites at its 5' extremity flanking *Hoxd13* and orientated towards C-DOM (Figure 7A).

We first verified the CTCF binding profiles in the two island V mutant alleles. As expected, the *Del(V)* allele showed a complete loss of CTCF associated with island V (Figure 7A). In contrast, when we analyzed CTCF occupancy in the *Inv(V)* allele in GT at E13.5, a

strong CTCF binding to the major peak was detected, indicating that the inversion of the site did not affect its binding capacity (Figure 7A). We next looked at the potential impact of either deleting or inverting this CTCF site on the remaining regulatory elements by performing ATAC-seq in wildtype, *Del(V),* and *Inv(V)* homozygous GT at E13.5. In mutant *Del(V)* GT cells, with the exception of the deleted region, we did not observe any change in the ATAC-seq profile (Figure 7–figure supplement 4A) when compared to control GT cells. Minor changes were not reproduced in replicates and were likely due to individual variation (Figure 7–figure supplement 4A). In the mutant *Inv(V)* GT cells, we observed the loss of one ATAC-seq peak located between the GT2 and CsB sequences (Figure 7–figure supplement 4A, black arrow), similar to what was scored in the *Del(GT2)* allele. Therefore, neither the deletion nor the inversion of this centrally-located CTCF site had any substantial effect on the accessibility of the remaining regulatory elements, corroborating the RT-qPCR results where expression of *Hoxd13* was unchanged in these two alleles (Figure 6).

The position and orientation of this CTCF site suggested that it may play a role in helping the central part of the C-DOM, rich in potential GT-specific elements, to reach *Hoxd13* through the formation of a large loop. We thus performed 4C-seq by using the *Del(V)* and *Inv(V)* mutant alleles on GT cells at E13.5 to investigate whether either the absence or the inversion of the CTCF site would affect the interaction landscape within C-DOM. When *Hoxd13* was taken as a viewpoint for the *Del(V)* allele, the global interaction profile between *Hoxd13* and C-DOM was virtually identical to control (Figure 7B). We confirmed this result by using a viewpoint positioned on island IV, at the vicinity of island V. Only a slight reduction in the frequency of contacts between island IV and *Hoxd13* was scored (Figure 7B, black arrow). Therefore, island V and its associated CTCF site have a marginal importance in maintaining the global chromatin structure of this regulatory landscape.

The majority of CTCF mediated chromatin loops are established between sites displaying opposite and convergent orientations (i.e. with CTCF motifs pointing toward each other) (de Wit et al., 2015; Rao et al., 2014; Vietri Rudan et al., 2015). Because our *Inv(V)* allele modified the orientation of this centrally-positioned CTCF site, we analyzed the impact of this inversion upon chromatin conformation. Qualitative analysis of the interaction profile generated using *Hoxd13* as a viewpoint revealed a slight disruption in the contacts between *Hoxd13* and island V (Figure 7C, red arrow). We validated this result by doing the reverse experiment and using a viewpoint on island V. In this set up, we observed a reduction in the overall frequency of interactions in the region between island V and *Hoxd13* thus confirming the previous result (Figure 7C and Figure 7–figure supplement 4B). Noteworthy, we observed

399 an increase in the interaction frequency in the region centromeric to island V up to island II
400 and island C, i.e. with the next CTCF sites displaying opposite and convergent orientations in
401 the mutant configuration (Figure 7C and Figure 7–figure supplement 4B). When we used island
402 IV as a viewpoint, we also observed a reduction in contacts with *Hoxd13* (Figure 7C, arrow).
403 Taken together, these results suggest that either the loss or the inversion of island V and its
404 associated CTCF site, had an effect on C-DOM chromatin structure. Nonetheless, this effect
405 did not greatly alter the regulatory landscape chromatin architecture, corroborating the lack of
406 impact on transcription.
407
408 **Group 13 HOX proteins access the TAD structure**
409 Our datasets on single enhancer deletions raise several potential hypotheses (see the
410 discussion). Amongst them the possibility that the transcriptional outcome of the C-DOM
411 regulation may rely upon an unspecific, global effect of accumulating various factors within
412 the landscape architecture, thus licensing the TAD for activation of the target genes. The same
413 C-DOM TAD was previously shown to regulate *Hoxd13* and neighboring *Hoxd* genes during
414 distal limb bud development, a structure that resembles in many respects the developing
415 genitals (Cobb and Duboule, 2005; Cohn, 2011; Infante et al., 2015; Tschopp et al., 2014). In
416 this case, the products of both *Hoxa13* and *Hoxd13* were shown to bind to most of those C-
417 DOM regulatory sequences specific for distal limb buds. From this observation, it was
418 concluded that HOX13 proteins themselves were instrumental in activating or re-enforcing
419 transcription of the *Hoxd13* gene in this developmental context, by accumulating at this
420 landscape and binding to many accessible sites due to their low binding specificity (Beccari et
421 al., 2016; Sheth et al., 2016).
422 In this context, we used an antibody against the HOXA13 product in a CnR approach,
423 with either CR cells at E10.5 or GT cells at E13.5, i.e. before GT formation and during its
424 emergence, respectively. Previous work has shown both redundancy of binding to limb
425 regulatory elements and similarity of DNA binding motifs between HOXA13 and HOXD13
426 (Sheth et al., 2016). As such, and because of the HOXA13 binding profile in our dataset, we
427 consider that this dataset reflects the binding of either HOXA13, HOXD13 or of both proteins
428 and is thus referred to as 'HOX13' (Figure 8). We detected enrichment of HOX13 binding
429 signals in both *Hoxd13* regulatory landscapes (Figure 8A; C-DOM and T-DOM) similar to
430 what was observed in distal forelimb at E12.5 (Beccari et al., 2016; Sheth et al., 2016). In CR
431 cells at E10.5, HOX13 binding was found in C-DOM at discrete positions corresponding to
432 previously described regulatory elements, in particular GT1, GT2, and Prox (Figure 8B). All

these binding sites and others, with the exception of the Prox enhancer, correlated with accessible chromatin sites as mapped by ATAC-seq (Figure 8B, arrow). In the case of Prox, HOX13 binding was scored before a clear ATAC-seq signal was detected, suggesting a potential role for HOX13 proteins in participating to making some of these sites accessible. The few strong ATAC-seq peaks, which were not matched by HOX13 binding corresponded to non-*Hox* gene promoters (Figure 8B, bottom line).

In E13.5 GT cells, as development progressed in parallel with C-DOM becoming fully active, an overall increase of HOX13 binding was scored over C-DOM (Figure 8B). While binding was strengthened at some sites bound at the earlier stage, other elements became both accessible and bound by HOX13 such as the islands II and III regions or a sequence located inside an intron of the *Lnpk* gene (Figure 8B). Overall, a good correlation was observed between the increase of *Hoxd13* transcript levels on the one hand, and both the activation of the C-DOM regulatory landscape and the binding of HOX13, on the other.

DISCUSSION

A preformed chromatin structure with multiple regulatory choices

In mammals, external genitals appear during fetal development as an overgrowth of a mesodermal territory surrounding the cloaca region (Georgas et al., 2015). In the absence of both *Hoxa13* and *Hoxd13* functions, this growth does not occur and the fetus displays a structure resembling that of a cloaca (Kondo et al., 1997; Warot et al., 1997), indicating that the proper transcriptional activation of these two genes in time and space is critical in this context. Studies of the *HoxD* cluster have provided some insights into this question (Lonfat et al., 2014) and suggested that the regulation of *Hoxd13* is primarily achieved by the C-DOM TAD, a large regulatory landscape flanking the gene cluster on its centromeric side, which also controls *Hoxd* gene activation in the developing digits. In the latter case, the chromatin interaction profile displayed some differences in transcriptionally active cells, even though the global TAD structure remained unchanged, suggesting that a C-DOM internal chromatin micro-organization had occurred due to the implementation of various digit-specific enhancers. Because of the close evolutionary neighborhood of digits and external genitals (Cobb and Duboule, 2005; Cohn, 2011; Tschopp et al., 2014), we examined this particular aspect of *Hoxd* gene regulation during the growth of the genital tubercle.

We looked at chromatin dynamics at the *Hoxd* locus and observed two types of chromatin interactions. On the one hand, we detected contacts associated with a pre-formed

structure, mainly linked to occupied CTCF sites. These contacts were observed independently of the transcriptional status of the cluster, as exemplified by island II and island V. On the other hand, we scored interactions present only when transcriptional activation had occurred such as the Prox and GT2 enhancer sequences. Our time-point series of interaction profiles revealed that the C-DOM TAD seems to be activated in a coordinated manner, with all specific contacts appearing mostly within the same developmental time window, suggesting that the TAD itself may be considered as a global regulatory unit (see below), rather than a field containing a range of disparate enhancers with specific features and acting at different times. Also, the chromatin architecture associated with this specific developmental context was already observed in the E10.5 CR, i.e. before the emergence of the GT. Therefore, this internal-TAD micro-organization predates the outgrowth of the GT structure, which suggests -but does not demonstrate- a causal relationship or at least a necessity for the TAD to be fully primed for the structure to develop.

Switching the TAD on and off to prevent regulatory leakages

Our time-series sampling gave us the unique opportunity to follow the C-DOM TAD dynamics in a developing system where most of the cells at E17.5 derive from a homogenous population of mesodermal cells in the nascent genital bud at E12.5, all expressing *Hoxd13*. The highest frequency of interactions with the C-DOM was scored in E12.5 and E13.5 GTs, which correlated with an increase in *Hoxd13* transcription, an enrichment of H3K27ac marks, and increase in binding of HOX13 proteins at discrete enhancer elements. After this time-point, a decrease in *Hoxd* transcript levels were scored in parallel with a reduction of all contacts associated with the active regulatory regions within the C-DOM. By E17.5 the C-DOM structure within the GT was reduced to a framework of constitutive interactions associated to CTCF binding sites, similar to the one observed in ES cells and fetal forebrain cells.

This global decrease, observed at a cell population level, can be explained either through a general decrease in transcription or through the selective transcriptional switch-off in some cell types along with their progressive differentiation thus leading to a dilution effect. Detailed ISH analyses (Dolle et al., 1991b; Warot et al., 1997) clearly favors the latter option, whereby some cell types differentiating from the early mesodermal GT precursors turn off C-DOM regulation, whereas others maintain this regulation. At E17.5 indeed, strong *Hoxd13* expression was scored in the anlagen of the *corpus cavernosum* while other cells of the tubercle became negative. Of note, this concentration of positive cells in the blastema and subsequent restriction to the periphery (Dolle et al., 1991b) resembles the situation for *Hoxd13* transcripts

during cartilage differentiation in developing digits. In support of this analogy, a penile bone (*baculum*) differentiates from this region in the mouse as in many other mammals.

The hereby described changes in the regulatory landscape architecture associated with transcriptional activity seem to be a pervasive feature during development (Andrey et al., 2017; Freire-Pritchett et al., 2017; Phillips-Cremins et al., 2013). We show that when these 'active' contacts disappear along with transcription being switched off, the TAD structure comes back to an inactive configuration. While such negative ground-state structures may simply reflect the absence of upstream factors and/or represent a scaffold to reinforce future enhancer-promoter contacts (Paliou et al., 2019), it may in our case be functionally required to prevent any transcriptional leakage of *Hoxd13*. HOX13 products are indeed potent dominant negative proteins (Darbellay et al., 2019; Villavicencio-Lorini et al., 2010) and their ectopic production in time and space must be prevented for proper development to be achieved (Young et al., 2009). A rapid return to an inactive chromatin conformation of C-DOM may help control this aspect, unlike other contexts where a particular chromatin topology is maintained for a long time (Fernandez-Albert et al., 2019).

Mechanism(s) of action of long-range enhancers

The complex pleiotropic expressions of vertebrate *Hox* genes, as well as of many other developmental genes, are usually controlled by multiple enhancers, either regulating subsets of the global pattern, or acting together in a partially redundant manner (Long et al., 2016; Montavon et al., 2011; Spitz and Furlong, 2012). We tested the potential function either of large DNA segments, or of shorter candidate regulatory regions within these segments and obtained different results depending on the position of the segment considered within C-DOM. When the *rel5* to *SB* DNA fragment was deleted, a substantial decrease in *Hoxd13* transcription was observed. However, the deletion of any single candidate sequence in isolation identified within this segment did not elicit any detectable decrease in transcription. This systematic analysis echoes previous studies where deleting a single and well-characterized enhancer did not have the expected effect upon its target gene (e.g. (Cretekos et al., 2008; Frankel et al., 2010; Osterwalder et al., 2018).

In contrast, the deletion of Prox resulted in a decrease in *Hoxd13* transcripts, which in itself could account for the decrease observed when the *rel1* to *rel5* DNA fragment was deleted. This occurred in the absence of any major reorganization either of the chromatin architecture, or of its accessibility to factors. Therefore, Prox seemed to act independently of the other elements in C-DOM, as initially expected for a 'classical' enhancer sequence. Concerning the

elements located within the *rel5* to *SB* central part of C-DOM whose deletions in isolation had no detectable effect, they could be functionally redundant with one another or, alternatively, compensatory mechanisms could be implemented for instance to re-direct the lost interactions towards another enhancer. Also, evolution might have selected regulatory processes to cope when facing particular conditions not necessarily tractable in laboratory conditions (Frankel et al., 2010; Hong et al., 2008). Our transgenic assays revealed that at least partial overlap in the functional domains was sometimes observed (GT2, GT1, Prox), whereas in other cases, transgenic sequences elicited complementary domain of expression (IIIE and IVE). Therefore, some functional overlap between enhancers may account for the absence of phenotype (Osterwalder et al., 2018). Finally, it is possible either that our experimental approach lacks the resolution required to discern mild alterations in gene expression, perhaps occurring in a subpopulation of cells, or that individual C-DOM enhancers elements may control gene expression at distinct developmental stages. In the latter scenario, we may have missed the enhancer function by focusing our analyses in only selected developmental time-point. Support for this alternative was provided by our results showing a decrease of *Hoxd13* mRNA in *del(GT2)* CR at E10.5.

Besides these potential explanations, the binding of HOX13 proteins to most -if not all- these C-DOM regulatory sequences raise yet another potential explanation related to recent work showing that phase-separation-induced condensates of RNA Pol II, transcription factors (TF) and the Mediator complex are present at particular enhancers leading to transcriptional activation (Boija et al., 2018; Hnisz et al., 2017; Sabari et al., 2018). In this view, condensate formation would be beneficial for transcriptional activation and could be promoted by the aggregation of protein containing intrinsically disordered regions (Kato et al., 2012). Both HOXD13 and HOXA13 contain long stretches of monotonic amino-acids (poly-Ala, Poly-Glu, Poly-Ser) (Akarsu, 1996; Mortlock and Innis, 1997; Muragaki et al., 1996), which could thus contribute to the building of this micro-environment by using the TAD as a scaffold. Naturally-occurring modifications in the lengths of these amino-acids repeats were shown to drastically affect the function of HOX13 proteins (Bruneau et al., 2001; Muragaki et al., 1996; Utsch et al., 2002). Yet their effects upon a potential regulatory structure has not yet been evaluated. Binding of HOX13 proteins over C-DOM involved most -yet not all- sequences determined accessible by ATAC-seq. In the case of the Prox sequence a robust association was detected by CnR before an ATAC-seq peak could be scored, in support of the idea that HOX13 protein may in some instances display a pioneer effect (Desanlis et al., 2019).

CTCF and the loop extrusion model *in embryo*

The *Rel5-SB* sub-region of C-DOM contains the largest series of defined GT regulatory sequences involved in *Hoxd13* regulation. Within this region lies island V, which contains the only occupied CTCF site in the central part of C-DOM. We thus assumed that this site would be instrumental to bring these enhancers towards the *HoxD* cluster through looping. Also, this element is one of the two constitutive contacts maintained in the absence of transcription (along with island II). After inversion of island V and the CTCF site contained within, the effects upon the global chromatin architecture were marginal. This result is in line with the lack of transcriptional decrease observed upon deleting this element. All other identified regulatory sequences located nearby were still able to contact *Hoxd13* with the same profile, suggesting that this CTCF site had no major role in securing interactions between these enhancers and *Hoxd13*, similar to what was suggested at another developmentally regulated locus (Williamson et al., 2019).

The inversion of island V and its CTCF site nevertheless resulted in a global decrease of interactions with *Hoxd13*, balanced by an increase in interactions with the centromeric region containing distal CTCF sits. After inversion, these CTCF sites were now facing the island V CTCF binding site and hence these partial redistributions of interactions are in agreement with the loop extrusion model (de Wit et al., 2015; Rao et al., 2014; Vietri Rudan et al., 2015). While the inversion of island V thus resulted in a slight reallocation of intra-TAD interactions, they were not sufficient to elicit changes in gene expression and had negligible impact on long-range regulation of *Hoxd* genes by C-DOM. Alternatively, we may be missing the time resolution to observe the impact of removing these sites.

REFERENCES

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**:W3–W10. doi:10.1093/nar/gkw343

Akarsu A. 1996. Genomic structure of HOXD13 gene: a nine polyalanine duplication causes synpolydactyly in two unrelated families. *Human Molecular Genetics* **5**:945–952. doi:10.1093/hmg/5.7.945

Amândio AR, Necsulea A, Joye E, Mascrez B, Duboule D. 2016. Hotair Is Dispensible for Mouse Development. *PLoS Genet* **12**:e1006232. doi:10.1371/journal.pgen.1006232

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

606 Andrey G, Schöpflin R, Jerković I, Heinrich V, Ibrahim DM, Paliou C, Hochradel M,
607 Timmermann B, Haas S, Vingron M, Mundlos S. 2017. Characterization of hundreds of
608 regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome*
609 *Res* **27**:223–233. doi:10.1101/gr.213066.116
610 Beccari L, Yakushiji-Kaminatsui N, Woltering JM, Necsulea A, Lonfat N, Rodriguez-
611 Carballo E, Mascrez B, Yamamoto S, Kuroiwa A, Duboule D. 2016. A role for HOX13
612 proteins in the regulatory switch between TADs at the HoxD locus. *Genes Dev* **30**:1172–86.
613 doi:10.1101/gad.281055.116
614 Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K,
615 Manteiga JC, Hannett NM, Abraham BJ, Afeyan LK, Guo YE, Rimel JK, Fant CB, Schuijers
616 J, Lee TI, Taatjes DJ, Young RA. 2018. Transcription Factors Activate Genes through the
617 Phase-Separation Capacity of Their Activation Domains. *Cell* **175**:1842-1855.e16.
618 doi:10.1016/j.cell.2018.10.042
619 Bruneau S, Johnson KR, Yamamoto M, Kuroiwa A, Duboule D. 2001. The mouse
620 Hoxd13(spdh) mutation, a polyalanine expansion similar to human type II synpolydactyly
621 (SPD), disrupts the function but not the expression of other Hoxd genes. *Dev Biol* **237**:345–
622 53. doi:10.1006/dbio.2001.0382
623 Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native
624 chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding
625 proteins and nucleosome position. *Nat Methods* **10**:1213–1218. doi:10.1038/nmeth.2688
626 Cobb J, Duboule D. 2005. Comparative analysis of genes downstream of the Hoxd cluster in
627 developing digits and external genitalia. *Development* **132**:3055–67. doi:10.1242/dev.01885
628 Cohn MJ. 2011. Development of the external genitalia: Conserved and divergent mechanisms
629 of appendage patterning. *Dev Dyn* **240**:1108–1115. doi:10.1002/dvdy.22631
630 Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini
631 LA, Zhang F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*
632 **339**:819–23. doi:10.1126/science.1231143
633 Cretekos CJ, Wang Y, Green ED, Martin JF, Rasweiler JJ th, Behringer RR. 2008.
634 Regulatory divergence modifies limb length between mammals. *Genes Dev* **22**:141–51.
635 doi:10.1101/gad.1620408
636 Darbellay F, Bochaton C, Lopez-Delisle L, Mascrez B, Tschopp P, Delpretti S, Zakany J,
637 Duboule D. 2019. The constrained architecture of mammalian Hox gene clusters. *Proc Natl*
638 *Acad Sci USA* **116**:13424–13433. doi:10.1073/pnas.1904602116
639 David FP, Delafontaine J, Carat S, Ross FJ, Lefebvre G, Jarosz Y, Sinclair L, Noordermeer
640 D, Rougemont J, Leleu M. 2014. HTSstation: a web application and open-access libraries for
641 high-throughput sequencing data analysis. *PloS one* **9**:e85879.
642 doi:10.1371/journal.pone.0085879
643 de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E,
644 Wijchers PJ, Krijger PH, de Laat W. 2015. CTCF Binding Polarity Determines Chromatin
645 Looping. *Molecular cell* **60**:676–84. doi:10.1016/j.molcel.2015.09.023
646 Desanlis I, Kherdjemil Y, Mayran A, Bouklouch Y, Gentile C, Sheth R, Zeller R, Drouin J,
647 Kmita M. 2019. HOX13-dependent chromatin accessibility modulates the target repertoires
648 of the HOX factors (preprint). Developmental Biology. doi:10.1101/789875
649 Dolle P, Dierich A, LeMeur M, Schimmang T, Schuhbaur B, Chambon P, Duboule D. 1993.
650 Disruption of the Hoxd-13 gene induces localized heterochrony leading to mice with neotenic
651 limbs. *Cell* **75**:431–41.
652 Dolle P, Izpisua-Belmonte JC, Boncinelli E, Duboule D. 1991a. The Hox-4.8 gene is
653 localized at the 5' extremity of the Hox-4 complex and is expressed in the most posterior
654 parts of the body during development. *Mech Dev* **36**:3–13.
655 Dolle P, Izpisua-Belmonte JC, Brown JM, Tickle C, Duboule D. 1991b. HOX-4 genes and

656 the morphogenesis of mammalian genitalia. *Genes Dev* **5**:1767–7.
657 Fernandez-Albert J, Lipinski M, Lopez-Cascales MT, Rowley MJ, Martin-Gonzalez AM, Del
658 Blanco B, Corces VG, Barco A. 2019. Immediate and deferred epigenomic signatures of in
659 vivo neuronal activation in mouse hippocampus. *Nat Neurosci* **22**:1718–1730.
660 doi:10.1038/s41593-019-0476-2
661 Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. 2010. Phenotypic robustness
662 conferred by apparently redundant transcriptional enhancers. *Nature* **466**:490–3.
663 doi:10.1038/nature09158
664 Freire-Pritchett P, Schoenfelder S, Várnai C, Wingett SW, Cairns J, Collier AJ, García-
665 Vílchez R, Furlan-Magaril M, Osborne CS, Fraser P, Rugg-Gunn PJ, Spivakov M. 2017.
666 Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic
667 stem cells. *eLife* **6**:e21926. doi:10.7554/eLife.21926
668 Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation
669 of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**:2038–49.
670 doi:10.1016/j.celrep.2016.04.085
671 Georgas KM, Armstrong J, Keast JR, Larkins CE, McHugh KM, Southard-Smith EM, Cohn
672 MJ, Batourina E, Dan H, Schneider K, Buehler DP, Wiese CB, Brennan J, Davies JA,
673 Harding SD, Baldock RA, Little MH, Vezina CM, Mendelsohn C. 2015. An illustrated
674 anatomical ontology of the developing mouse lower urogenital tract. *Development* **142**:1893–
675 1908. doi:10.1242/dev.117903
676 Gonzalez F, Duboule D, Spitz F. 2007. Transgenic analysis of Hoxd gene regulation during
677 digit development. *Dev Biol* **306**:847–59.
678 Hashimoto M, Takemoto T. 2015. Electroporation enables the efficient mRNA delivery into
679 the mouse zygotes and facilitates CRISPR/Cas9-based genome editing. *Scientific Reports* **5**.
680 doi:10.1038/srep11315
681 Herault Y, Hraba-Renevey S, van der Hoeven F, Duboule D. 1996. Function of the Evx-2
682 gene in the morphogenesis of vertebrate limbs. *EMBO J* **15**:6727–38.
683 Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. 2017. A Phase Separation
684 Model for Transcriptional Control. *Cell* **169**:13–23. doi:10.1016/j.cell.2017.02.007
685 Hong J-W, Hendrix DA, Levine MS. 2008. Shadow enhancers as a source of evolutionary
686 novelty. *Science* **321**:1314. doi:10.1126/science.1160631
687 Hostikka SL, Capecchi MR. 1998. The mouse Hoxc11 gene: genomic structure and
688 expression pattern. *Mechanisms of development* **70**:133–45.
689 Infante CR, Mihala AG, Park S, Wang JS, Johnson KK, Lauderdale JD, Menke DB. 2015.
690 Shared Enhancer Activity in the Limbs and Phallus and Functional Divergence of a Limb-
691 Genital cis-Regulatory Element in Snakes. *Developmental Cell* **35**:107–119.
692 doi:10.1016/j.devcel.2015.09.003
693 Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J,
694 Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R,
695 McKnight SL. 2012. Cell-free Formation of RNA Granules: Low Complexity Sequence
696 Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**:753–767.
697 doi:10.1016/j.cell.2012.04.017
698 Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate
699 alignment of transcriptomes in the presence of insertions, deletions and gene fusions.
700 *Genome biology* **14**:R36. doi:10.1186/gb-2013-14-4-r36
701 Kondo T, Zákány J, Innis JW, Duboule D. 1997. Of fingers, toes and penises. *Nature*
702 **390**:29–29. doi:10.1038/36234
703 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
704 **9**:357–9. doi:10.1038/nmeth.1923
705 Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA,

706 Copeland NG. 2001. A highly efficient Escherichia coli-based chromosome engineering
707 system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics*
708 **73**:56–65. doi:10.1006/geno.2000.6451
709 Lonfat N. 2013. An ancestral regulatory mechanism underlies Hoxd gene expression in both
710 developing genitals and digits. doi:10.5075/epfl-thesis-5997
711 Lonfat N, Duboule D. 2015. Structure, function and evolution of topologically associating
712 domains (TADs) at HOX loci. *FEBS letters* **589**:2869–76. doi:10.1016/j.febslet.2015.04.024
713 Lonfat N, Montavon T, Darbellay F, Gitto S, Duboule D. 2014. Convergent evolution of
714 complex regulatory landscapes and pleiotropy at Hox loci. *Science* **346**:1004–6.
715 doi:10.1126/science.1257493
716 Long HK, Prescott SL, Wysocka J. 2016. Ever-Changing Landscapes: Transcriptional
717 Enhancers in Development and Evolution. *Cell* **167**:1170–1187.
718 doi:10.1016/j.cell.2016.09.018
719 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
720 reads. *EMBnet.journal* **17**. doi:10.14806/ej.17.1.200
721 Mashiko D, Fujihara Y, Satouh Y, Miyata H, Isotani A, Ikawa M. 2013. Generation of
722 mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided
723 RNA. *Sci Rep* **3**:3355. doi:10.1038/srep03355
724 Meers MP, Tenenbaum D, Henikoff S. 2019. Peak calling by Sparse Enrichment Analysis for
725 CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**:42. doi:10.1186/s13072-019-
726 0287-4
727 Montavon T, Le Garrec J-F, Kerszberg M, Duboule D. 2008. Modeling Hox gene regulation
728 in digits: reverse collinearity and the molecular origin of thumbness. *Genes Dev* **22**:346–59.
729 doi:10.1101/gad.1631708
730 Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F,
731 Duboule D. 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell*
732 **147**:1132–45. doi:10.1016/j.cell.2011.10.023
733 Mortlock DP, Innis JW. 1997. Mutation of HOXA13 in hand-foot-genital syndrome. *Nat*
734 *Genet* **15**:179–80. doi:10.1038/ng0297-179
735 Muragaki Y, Mundlos S, Upton J, Olsen BR. 1996. Altered Growth and Branching Patterns
736 in Synpolydactyly Caused by Mutations in HOXD13. *Science* **272**:548–551.
737 doi:10.1126/science.272.5261.548
738 Noordermeer D, Leleu M, Schorderet P, Joye E, Chabaud F, Duboule D. 2014. Temporal
739 dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *eLife*
740 **3**:e02557. doi:10.7554/eLife.02557
741 Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D. 2011. The
742 dynamic architecture of Hox gene clusters. *Science* **334**:222–5. doi:10.1126/science.1207194
743 Ong CT, Corces VG. 2014. CTCF: an architectural protein bridging genome topology and
744 function. *Nat Rev Genet* **15**:234–46. doi:10.1038/nrg3663
745 Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA,
746 Zhu Y, Plajzer-Frick I, Pickle CS, Kato M, Garvin TH, Pham QT, Harrington AN, Akiyama
747 JA, Afzal V, Lopez-Rios J, Dickel DE, Visel A, Pennacchio LA. 2018. Enhancer redundancy
748 provides phenotypic robustness in mammalian development. *Nature* **554**:239–243.
749 doi:10.1038/nature25461
750 Paliou C, Guckelberger P, Schöpflin R, Heinrich V, Esposito A, Chiariello AM, Bianco S,
751 Annunziatella C, Helmuth J, Haas S, Jerković I, Brieske N, Wittler L, Timmermann B,
752 Nicodemi M, Vingron M, Mundlos S, Andrey G. 2019. Preformed chromatin topology assists
753 transcriptional robustness of *Shh* during limb development. *Proc Natl Acad Sci USA*
754 **116**:12390–12399. doi:10.1073/pnas.1900672116
755 Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT,

Hookway TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R, Dekker J, Taylor J, Corces VG. 2013. Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153**:1281–95.

Quinlan AR. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current protocols in bioinformatics / editoral board, Andreas D Baxevanis . [et al]* **47**:11 12 1-11 12 34. doi:10.1002/0471250953.bi1112s47

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**:1665–80. doi:10.1016/j.cell.2014.11.021

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

Rodriguez-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

Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, Li CH, Guo YE, Day DS, Schuijers J, Vasile E, Malik S, Hnisz D, Lee TI, Cisse II, Roeder RG, Sharp PA, Chakraborty AK, Young RA. 2018. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**. doi:10.1126/science.aar3958

Schmid M, Durussel T, Laemmli UK. 2004. ChIC and ChEC; genomic mapping of chromatin proteins. *Mol Cell* **16**:147–157. doi:10.1016/j.molcel.2004.09.007

Sheth R, Barozzi I, Langlais D, Osterwalder M, Nemec S, Carlson HL, Stadler HS, Visel A, Drouin J, Kmita M. 2016. Distal Limb Patterning Requires Modulation of cis-Regulatory Activities by HOX13. *Cell Rep* **17**:2913–2926. doi:10.1016/j.celrep.2016.11.039

Skene PJ, Henikoff JG, Henikoff S. 2018. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**:1006–1019. doi:10.1038/nprot.2018.015

Skene PJ, Henikoff S. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**. doi:10.7554/eLife.21856

Soshnikova N, Montavon T, Leleu M, Galjart N, Duboule D. 2010. Functional analysis of CTCF during mammalian limb development. *Dev Cell* **19**:819–30. doi:10.1016/j.devcel.2010.11.009

Spitz F, Furlong EE. 2012. Transcription factors: from enhancer binding to developmental control. *Nature reviews Genetics* **13**:613–26. doi:10.1038/nrg3207

Spitz F, Gonzalez F, Duboule D. 2003. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**:405–17.

Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL. 2015. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS ONE* **10**:e0124633. doi:10.1371/journal.pone.0124633

Tschopp P, Duboule D. 2011. A regulatory "landscape effect" over the HoxD cluster. *Developmental biology* **351**:288–96. doi:10.1016/j.ydbio.2010.12.034

Tschopp P, Sherratt E, Sanger TJ, Groner AC, Aspiras AC, Hu JK, Pourquie O, Gros J, Tabin CJ. 2014. A relative shift in cloacal location repositions external genitalia in amniote evolution. *Nature* **516**:391–4. doi:10.1038/nature13819

Utsch B, Becker K, Brock D, Lentze MJ, Bidlingmaier F, Ludwig M. 2002. A novel stable polyalanine [poly(A)] expansion in the HOXA13 gene associated with hand-foot-genital syndrome: proper function of poly(A)-harbouring transcription factors depends on a critical repeat length? *Hum Genet* **110**:488–494. doi:10.1007/s00439-002-0712-8

Vian L, Pękowska A, Rao SSP, Kieffer-Kwon K-R, Jung S, Baranello L, Huang S-C, El Khattabi L, Dose M, Pruett N, Sanborn AL, Canela A, Maman Y, Oksanen A, Resch W, Li X, Lee B, Kovalchuk AL, Tang Z, Nelson S, Di Pierro M, Cheng RR, Machol I, St Hilaire BG, Durand NC, Shamim MS, Stamenova EK, Onuchic JN, Ruan Y, Nussenzweig A, Levens D, Aiden EL, Casellas R. 2018. The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **175**:292–294. doi:10.1016/j.cell.2018.09.002

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 reports* **10**:1297–309. doi:10.1016/j.celrep.2015.02.004

Vieux-Rochas M, Fabre PJ, Leleu M, Duboule D, Noordermeer D. 2015. Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proceedings of the National Academy of Sciences of the United States of America* **112**:4672–7. doi:10.1073/pnas.1504783112

Villavicencio-Lorini P, Kuss P, Friedrich J, Haupt J, Farooq M, Turkmen S, Duboule D, Hecht J, Mundlos S. 2010. Homeobox genes d11-d13 and a13 control mouse autopod cortical bone and joint formation. *The Journal of clinical investigation* **120**:1994–2004. doi:10.1172/JCI41554

Warot X, Fromental-Ramain C, Fraulob V, Chambon P, Dolle P. 1997. Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts. *Development* **124**:4781–91.

Williamson I, Kane L, Devenney PS, Flyamer IM, Anderson E, Kilanowski F, Hill RE, Bickmore WA, Lettice LA. 2019. Developmentally regulated Shh expression is robust to TAD perturbations. *Development* **146**. doi:10.1242/dev.179523

Woltering JM, Noordermeer D, Leleu M, Duboule D. 2014. Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biol* **12**:e1001773. doi:10.1371/journal.pbio.1001773

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, Giron CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Keenan S, Lavidas I, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Nuhn M, Parker A, Patricio M, Pignatelli M, Rahtz M, Riat HS, Sheppard D, Taylor K, Thormann A, Vullo A, Wilder SP, Zadissa A, Birney E, Harrow J, Muffato M, Perry E, Ruffier M, Spudich G, Trevanion SJ, Cunningham F, Aken BL, Zerbino DR, Flicek P. 2016. Ensembl 2016. *Nucleic acids research* **44**:D710-6. doi:10.1093/nar/gkv1157

Young T, Rowland JE, van de Ven C, Bialecka M, Novoa A, Carapuco M, van Nes J, de Graaff W, Duluc I, Freund JN, Beck F, Mallo M, Deschamps J. 2009. Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Developmental cell* **17**:516–26. doi:10.1016/j.devcel.2009.08.010

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**:R137. doi:10.1186/gb-2008-9-9-r137

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

LEGENDS TO FIGURES

Figure 1: Transcription of *Hoxd* genes in developing GT. A) Quantification of *Hoxd* genes transcript levels by RNA-seq (FPKM values) in GT at E12.5 (Amândio et al., 2016), E16.5 and E18.5. **B)** RT-qPCR of *Hoxd13* mRNAs in different stages of GT development. The plotted values indicate the ratio of expression using the cloaca region (CR) as a reference (n≥3 biological replicates for each sample). A Welch's *t*-test was used to evaluate the putative significant changes in *Hoxd13* expression. Bars indicate mean with SD, ****p<0.0001, *p=0.0175. **C-E)** ATAC-seq (gray) and ChIP-seq profiles for H3K27ac (blue) and H3K27me3 (red) at the *HoxD* locus in E10.5 wildtype CR (**C**), E13.5 GT (**D**) and E17.5 GT (**E**). Coordinates (mm10): chr2:74637433-74775728. The gray box in track 3 indicates the enrichment of H3K27me3 at 5'-located *Hoxd* genes in the CR. The gray box in track 8 indicates the relative gain of H3K27me3 at 5'-located *Hoxd* genes in E17.5 GT when compared to the E13.5 GT sample.

Figure 2: Chromatin topology of C-DOM during GT development. A) Schematic representation of the two regulatory landscapes, with the centromeric (C-DOM) and telomeric (T-DOM) TADs flanking the *HoxD* cluster (black box), which acts as a boundary. Gray boxes represent non-*Hox* genes. The *cis*-regulatory elements involved in the control of *Hoxd* gene transcription in the GT are located in C-DOM (blue arrow). **B)** 4C-seq interactions profiles between the *Hoxd13* viewpoint (gray line) and both the *HoxD* cluster and the C-DOM. From top to bottom, 4C-seq profiles from mouse ES cells (mESC; track 1) (Noordermeer et al., 2014), E10.5 CR, E12.5 GT, E13.5 GT, E15.5 GT, E17.5 GT and fetal forebrain cells (track 7) are represented. Coordinates (mm10): chr2:73815520-74792376. A schematic representation of the *HoxD* cluster and the C-DOM is shown below, known enhancers are represented by black boxes. The percentages in gray represent the ratio, using mouse ES cells as a reference, of the sum of the fragments in the centromeric gene desert, divided by the sum of fragments that fall in a non-interacting region of the T-DOM (chr2:75166258-75571741). Blue lines highlight the changes in chromatin interactions between *Hoxd13* and Prox or GT2 in the different developmental stages and tissues analyzed. Red lines highlight that the contacts between *Hoxd13* and island II or island V remained fairly constant in all samples analyzed.

Figure 3: Various segments of C-DOM contribute to *Hoxd13* transcription in the GT. A) The gray tracks show ATAC-seq profiles of E10.5 CR (average of two biological replicates, track 1) and E13.5 GT (average of three biological replicates, track 2). The blue tracks are ChIP-seq for H3K27ac with E13.5 GT (track 3) and E17.5 GT (track 4). Coordinates (mm10): chr2: 73815520-74792376. A schematic representation of the *HoxD* cluster and the C-DOM is shown below, known enhancers are represented by black boxes. The red arrowheads represent the deletions breakpoints. The four large deletion alleles analyzed are depicted as gray dashed lines with scissors. **B-E)** RT-qPCR of *Hoxd* genes mRNAs for wildtype and homozygous mutant deletion alleles using E12.5 GT. The mutant allele is indicated on top of each plot. The values plotted indicate the ratio of mRNA levels using wildtype as a reference (black dots) (n=4 biologically independent wildtype or mutant GT). A Welch's *t*-test was used to evaluate the statistical significance of changes in gene expression. Bars indicate mean with SD, * $p \leq 0.02$; ** $p \leq 0.007$; *** $p \leq 0.0005$, **** $p \leq 0.0001$; ns= non-significant.

Figure 4: Deletion of the Prox enhancer. A) Schematic representation of the *HoxD* cluster and the C-DOM with the deletion of the Prox sequence leading to the *Del(Prox)* allele. **B)** X-gal staining showing the activity of the Prox enhancer. **C)** *Hoxd13* transcripts levels obtained by RT-qPCR using wildtype and homozygous mutant *Del(Prox)* GTs at E12.5. The values plotted indicate the ratio of expression using wildtype as a reference (black dots) (n=4 biologically independent WT or mutant GTs). A Welch's *t*-test was used to evaluate the statistical significance expression changes. Bars indicate mean with SD, ** p=0.006). **D)** WISH using the *Hoxd13* probe in both wildtype and mutant *Del(Prox)* E12.5 embryos. The *Hoxd13* expression pattern remained unchanged. **E)** ATAC-seq profiles covering C-DOM and *HoxD* in wildtype (top) and *Del(Prox)* mutant (bottom) E13.5 GTs. Coordinates (mm10): chr2:73815520-74792376. The wildtype profile is the average of three biological replicates whereas the *Del(Prox)* represents the average of two biological replicates. Peaks called using MACS2 are displayed under the corresponding tracks (vertical black lines) for each individual replicate. Black arrows highlight the deleted region. **F)** 4C-seq profiles (average of two biological replicates) of wildtype and mutant *Del(Prox)* E13.5 GTs. The *Hoxd13* viewpoint is shown as a gray line. The overlay of the two tracks wildtype (blue) and *Del(Prox)* (red) (bottom track) highlight the loss of the Prox enhancer in the *Del(Prox)* allele and the lack of major alterations in the frequency of contacts between *Hoxd13* and discrete *cis*-regulatory elements. Coordinates (mm10): chr2:73815520-74792376.

Figure 5: Activity of C-DOM regulatory elements *in vivo*. A) Schematic representation of C-DOM and the *HoxD* cluster. Previously characterized enhancers are shown as black boxes and red arrowheads point to the SB and rel5 breakpoints. **B)** ATAC-seq profile (top, average of three biological replicates) and H3K27ac ChIP-seq profile (bottom) of E13.5 GTs, focusing on the DNA interval between rel5 and SB (coordinates mm10: chr2:74084880-74432824). The vertical blue lines below the H3K27ac ChIP-seq profile represent the output of the MACS2 peak caller tool using the corresponding input as control. **C)** Enhancer transgene activity of all the individual regulatory sub-regions analyzed within the rel5 to SB interval. The gray dashed line boxes represent the tested sub-regions as well as the GT2 sequence. For each clone, a representative staining is shown at E13.5.

Figure 6: Serial deletions of single *cis*-regulatory elements. A) Schematic representation of the alleles generated by CRISPR-Cas9 editing *in vivo*. **B)** Relative expression of *Hoxd13* obtained by RT-qPCR of both wildtype control and the various mutant alleles using E12.5 GT cells. The values plotted indicate the ratio of expression using wildtype as a reference (black dots) for each gene (n≥3 biologically independent wildtype and mutant GTs). **C)** WISH using the *Hoxd13* probe and both wildtype and mutant E12.5 littermates. Both the mRNA levels and transcripts distribution remained globally unchanged.

Figure 7: Deletion and inversion of the island V CTCF site *in vivo*. A) CTCF ChIP-seq profiles of wildtype and *Del(V)* mutant E13.5 GTs. Cut & Run (CnR) of mutant *Inv(V)* E13.5 GT. The upper track shows the orientations of the CTCF motives (red and blue arrowheads). The black arrow indicates the major CTCF peak on island V. **B)** 4C-seq profiles (average of two biological replicates) of wildtype and mutant *Del(V)* E13.5 GTs. The positions of the *Hoxd13* (upper tracks) and island IV (lower track) viewpoints are shown with a gray line. The profiles are displayed as overlays of wildtype (blue) and *Del(V)* (red). The red arrow shows the deleted region and the black arrow points to the *Hoxd13* region. **C)** 4C-seq profiles (average of two biological replicates) of wildtype and mutant *Inv(V)* homozygous E13.5 GTs. Viewpoints are highlighted by a gray line. The profiles are shown as overlays of wildtype (blue) and *Inv(V)*(red). The red arrow shows the inverted region and the black arrows indicates the loss of contacts between island IV and the *Hoxd13* region in the *Inv(V)* sample. Percentages in blue (wildtype) and red (*InvV*) represent the proportion of the sums of interactions centromeric

or telomeric to island V. (coordinates (mm10) for the quantifications: centromeric: chr2:74015789-74276083; telomeric chr2:74332870-74671433). Coordinates (mm10): chr2:73815520-74792376

Figure 8: HOX13 protein binding in C-DOM. A) HOX13 CnR profiles using E10.5 CR cells (red) and GT cells at E13.5 (Blue). The blue box represents the *HoxD* cluster and gray boxes are non-*Hox* genes. The profile encompasses 4Mb and highlights the enrichment of HOX13 binding on both C-DOM and T-DOM *HoxD* regulatory landscapes. Coordinates (mm10): chr2: 72760109-76760109. **B)** ATAC-seq and HOX13 CnR profiles of E10.5 CR cells (red) and E13.5 GT cells (Blue). Close-up view of C-DOM and the *HoxD* cluster (coordinates in mm10: chr2:73815520-74792376). The arrows indicate that although the Prox enhancer is bound by HOX13 in the CR at E10.5, the chromatin is not yet accessible at this element.

MATERIAL AND METHODS

Mouse strains and genotyping

Genotyping of all alleles was done by PCR. Mouse tissue biopsies were lysed for 15' at 95ºC, 800rpm, in lysis buffer (50mM NaOH, 0.2mM EDTA). For all genotyping reactions PCR was performed with a standardized cycling protocol (1x(94°3'), 2x(94°1' .62°1',72°1'), 30x(94°30''.62°30'',72°30''), 1x(72°10')). The primers used to genotype the *Del(rel1-rel5), Del(rel5-SB)*, and *Del(SB-Atf2)* alleles can be found in (Montavon et al., 2011). Primers used to genotype the remaining alleles can be found in Table supplement 1.

CRISPR-Cas9

With the exception of the *Del(rel1-rel5), Del(rel5-SB),* and *Del(SB-Atf2)* alleles (Montavon et al., 2011), all mouse strains carrying deletions or inversions of the different regulatory regions were generated using CRISPR–Cas9 genome editing technology. Single guide RNAs (sgRNAs) were designed flanking the genomic regions of interest (5' and 3' to the regions of interest) using the crispr.mit.edu web tool (from the Zang laboratory) for the *Del(V), Inv(V),* and *Del(GT2)* alleles, or CCTop (Stemmer et al., 2015) for the *Del(III), Del(Pox),* and *Del(IV-SB)* alleles (Table supplement 2). All sgRNAs were cloned, as recommended in (Cong et al., 2013), into the BbsI site of the pX330:hSpCas9 (Addgene ID 42230) vector. The mouse strains *Del(V), Inv(V),* and *Del(GT2)* were produced by pronuclear

injection (Mashiko et al., 2013) of a mix of the two appropriate sgRNAs cloned into the pX330:hSpCas9 vector (sgRNA:pX330:hSpCas9) (25 ng/µl each). The mouse strains *Del(IV-SB), Del(IV), Del(III),* and *Del(Prox)* were produced by electroporation (Hashimoto and Takemoto, 2015) using a mix containing Cas9 mRNA (final concentration of 400ng/µl) and two sgRNAs (300ng/µl each) in Opti-MEM 1x injection buffer. PCR based genotyping was carried out with primers designed on both sides of sgRNAs targets, with an approximate distance of 150-300bp from the cutting site (Table supplement 1). Sanger sequence of positive PCR bands was used to identify and confirm the deletion or inversion breakpoints of the F0 funder animals (figure supplement 5).

Transgenic analysis

All mouse fosmid clones were obtained from BACPAC Resources Center (https://bacpacresources.org) (Table supplement 3). Their integrity was verified by Sanger sequence and restriction enzyme fingerprinting. The fosmids were introduced in EL250 cells (Lee et al., 2001) and targeted, by ET-recombineering, with a construct containing a *PI-SceI* restriction site, a *βglobin::LacZ* reporter gene with a FRT-flanked kanamycin selection marker, and flanked by 50 bp-long homology arms. The targeting constructs were produced by PCR amplification using the primers indicated in Table supplement 4 to introduce the homology arms. The WI1-D5 was shortened to remove the sequences that corresponded to island-IV. The targeted fosmids were selected at 30ºC on LB plates containing chloramphenicol and kanamycin. The integrity of each modified fosmid was verified by restriction enzyme fingerprinting, and the correct integration of the *βglobin::LacZ* reporter gene was confirmed by PCR and Sanger sequence. All fosmids were linearized with *PI-SceI* and micro-injected into mouse oocytes. Embryos were harvested at E13.5 and stained for β-galactosidase activity following standard procedures. A minimum of three transgenic animals with consistent staining were obtained per construct. The transgenic mouse embryos for either the Prox or GT2 were obtained as described in (Gonzalez et al., 2007; Lonfat et al., 2014). Embryos were stained using standard procedures. Whole embryos (E13.5) were fixed in 4% paraformaldehyde at 4°C for 35 min, stained in a solution containing 1 mg/ml X-gal at 37°C overnight, washed in PBS, imaged, and stored in 4% paraformaldehyde.

Whole-mount *in situ* hybridization

Whole-mount *in situ* hybridization (WISH) was performed according to (Woltering et al., 2014). Briefly, embryos were dissected in PBS and fixed overnight in 4% paraformaldehyde (PFA), washed in PBS, dehydrated, and stored in 100% methanol at -20°C. Rehydration was performed by a series of methanol/TBS-T washes, followed by a short digestion of Proteinase K, and re-fixation in 4% PFA. Pre-hybridization, hybridization, and post-hybridization steps were carried out at 67°C. For all genotypes, both mutant and control wildtype (E12.5) littermate embryos were processed in parallel to maintain identical conditions throughout the WISH procedure. DIG-labeled probes for *in situ* hybridizations were produced by *in vitro* transcription (Promega) and detection was carried out using an alkaline phosphatase conjugated anti-digoxigenin antibody (Roche). *Hoxd13* and *Evx2* WISH probes were previously described (Dolle et al., 1991a; Herault et al., 1996). For detection the chromogenic substrates NBT/BCIP or BM-purple were used.

RT-qPCR

Before processing, all tissues were stored at -80°C in RNA*later* stabilization reagent (Invitrogen). RNA was extracted from single micro-dissected GT (E12.5) or single cloaca region (CR) (E10.5), using Qiagen Tissue Lyser and RNeasy Plus kit (Qiagen), according to the manufacturer's instructions. RNA was reverse transcribed using Superscript III (Invitrogen) or Superscript IV (Invitrogen) and random hexamers. qPCR was performed on a CFX96 real-time system (BioRad) using GoTaq qPCR Master Mix (Promega). Primers were previously described in (Montavon et al., 2008). Three technical replicates were used per biological replicate. Relative gene expression levels were calculated by the $2^{-\Delta Ct}$ method using a reference gene. *Tubβ* was chosen as internal control and the mean of wildtype control samples was set as reference to calculate the ratio between the different samples. Graphical representation and statistical analysis were performed with GraphPad Prism 7.

4C-seq

Circular chromosome conformation capture (4C-seq) was performed as described in (Noordermeer et al., 2011). Briefly, tissues (20-40 GT or 40 CR) were isolated in PBS supplemented with 10% Fetal Calf Serum and dissociated to single cell by collagenase treatment. Samples were fixed in 2% formaldehyde, lysed, and stored at −80°C. Pools of between 20-40 GT or 40 CR were primarily digested with NlaIII (NEB, R0125L) followed by ligation under diluted conditions. After decrosslinking and DNA purification DpnII (NEB,

R0543M) was used for the second restriction. All ligation steps were performed using highly concentrated T4 DNA ligase (Promega, M1794). For each viewpoint approximately 1µg of DNA was amplified by using 12 individual PCR reactions. Libraries were constructed with inverse primers for different viewpoints (see Table supplement 5) containing Illumina Solexa adapter sequences and sequenced on an Illumina HiSeq 2500 sequencer, as single-end reads (read length 100 base pairs or 80 base pairs). In some samples 4-bp barcodes were added between the adapter and each specific viewpoint to allow sample multiplexing.

4C-seq reads were demultiplexed, mapped on GRCm38/mm10 mouse assembly, and analyzed using the 4C-seq pipeline of the Bioinformatics and Biostatistics Core Facility (BBCF) HTSstation (http://htsstation.epfl.ch) (David et al., 2014) or using a local version of it using the facilities of the Scientific IT and Application Support Center of EPFL. Profiles were normalized to a 5Mb region surrounding the *HoxD* cluster and smoothened using a window size of 11 fragments. C-DOM quantifications on Figure 2 were done by dividing the sum of the scores in the C-DOM (chr2:73921943–74648943) by the sum of the scores that fall in a non-interacting region of the T-DOM (chr2:75166258-75571741) (background local normalization). Signals falling either centromeric or telomeric to island V (in Figure 7 and Figure 7–figure supplement 4) were assessed by calculating the sum of the scores in the region of interest normalized by the sum of the scores in both regions (coordinates (mm10) for the quantifications: centromeric: chr2:74015789-74276083; telomeric chr2:74332870-74671433). Quantifications of the interactions established with the cis-regulatory elements in Figure 2–figure supplement 2 were calculated as a percentage of the sum of the scores of each element using the mESC sample as a reference.

ChIP-seq

Micro-dissected 35-40 GT or 70 CR were crosslinked in 1% formaldehyde/PBS for 20 min and stored at -80ºC until further processing. Chromatin was sheared using a water-bath sonicator (Covaris E220 evolution ultra-sonicator). Immunoprecipitation was done using the following antibodies, anti-CTCF (Active Motif, 61311), anti- H3K27ac (Abcam, ab4729), and H3K27me3 (Merck Millipore, 07-449). Libraries were prepared using the TruSeq protocol, and sequenced on the Illumina HiSeq system (100bp single-end reads) according to manufactures instructions.

ChIP-seq reads processing was done on the Duboule lab local Galaxy server (Afgan et al., 2016). Adapters and bad-quality bases were removed with Cutadapt version 1.16 (Martin,

2011) (options -m 15 -q 30 -a GATCGGAAGAGCACACGTCTGAACTCCAGTCAC). Reads were mapped to the mouse genome (mm10) using Bowtie2 (v2.3.4.1) (Langmead and Salzberg, 2012), with standard settings. The coverage was obtained as the output of MACS2 (v2.1.1.20160309) (Zhang et al., 2008). Peak calling in Figure 5 was done using MACS2 (v2.1.0.20160309) call peak (--gsize 1870000000) using the corresponding input data as control BAM (-c). CTCF motif orientation was assessed using the CTCFBSDB 2.0 database (Ziebarth et al., 2012), with EMBL_M1 identified motifs.

ATAC–seq

ATAC–seq was performed as described in (Buenrostro et al., 2013). Briefly, micro-dissected tissues (a poll of 2 GT or 2-3 CR) were isolated in PBS supplemented with 10% Fetal Calf Serum and dissociated to single cell by collagenase treatment. After isolation, 50,000 cells were lysed in 50 µl of lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% IGEPAL CA-630), nuclei were carefully resuspended in 50µl transposition reaction mix (25µl TD buffer, 2.5µl Tn5 transposase and 22.5µl nuclease-free water) and incubated at 37 °C for 30min. DNA was isolated with a MinElute DNA Purification Kit (Qiagen). Library amplification was performed by PCR (10 to 12 cycles) using NEBNext High-Fidelity 2x PCR Master Mix (NEB, M0541S). Library quality was checked on a fragment analyzer, and paired-end sequencing was performed on an Illumina NextSeq 500 instrument (read length 2 × 37 base pairs).

ATAC-seq reads processing was done on the Duboule lab local Galaxy server (Afgan et al., 2016). Reads were mapped to the mouse genome (mm10) using Bowtie2 (v2.3.4.1) (Langmead and Salzberg, 2012), (-I 0 -X 2000 --fr --dovetail --very-sensitive-local). Reads with mapping quality below 30, mapping to mitochondria, or not properly paired were removed from the analysis. PCR duplicates were filtered using Picard (v1.56.0). Peak calling was done using MACS2 (v2.1.0.20151222) call peak (--nomodel --shift -100 --extsize 200 --call-summits). The coverage was done using the center of the Tn5 insertion and extended on both sides by 20bp (script developed by L. Lopez-Delisle). When indicated, coverage profiles represent an average of the replicates, this was done by dividing each replicate by the number of million reads that fall within peaks in each sample (for normalization) and calculating the average coverage.

RNA-seq

Micro-dissected GT from different embryonic stages were individual 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. RNA quality was assessed using an Agilent 2100 Bioanalyser. Only samples with high RNA integrity number were used. Sequencing libraries were prepared according to TruSeq Stranded mRNA Illumina protocol, with polyA selection. RNA-seq libraries were sequenced on an Illumina HiSeq 2500 sequencer, as single-end reads (read length 100 base pairs).

Raw RNA-seq reads were aligned on the mouse mm10 genome assembly using TopHat 2.0.9 (Yates et al., 2016). Gene expression computations were performed using uniquely mapping reads extracted from TopHat alignments and genomic annotations from filtered gtf from Ensembl release 82 (Kim et al., 2013) as discribed in (Amândio et al., 2016). FPKM (fragments per kilo- base per million mapped fragments) expression levels for each gene were calculated using Cufflinks (Roberts et al., 2011).

Cut & Run

Cut & Run (Schmid et al., 2004; Skene and Henikoff, 2017) was performed as described in (Meers et al., 2019; Skene et al., 2018). Briefly, micro-dissected tissues (a set of 8 to 10 GT or CR) were isolated in PBS supplemented with 10% fetal calf serum and dissociated to single cell by collagenase treatment. After isolation, 500000 cells were washed and bond to concanavalin A-coated magnetic beads and permeabilized with wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, and Roche Complete protein inhibitor) containing 0.02% digitonin. Bonded cells were incubated with primary antibody (anti-HOXD13, AbCam ab19866; anti-CTCF, Active Motif, 6131) for 2h at room temperature. After washes the samples were incubated with Protein A-MNase (pA-MN) for 1 hour at 4ºC, then washed twice more with Wash Buffer. Samples were resuspended in low-salt rinse buffer (20 mM HEPES, pH7.5, 0.5 mM spermidine, and 0.125% Digitonin) and chilled to 0°C and the liquid was removed on a magnet stand. Ice-cold calcium incubation buffer (3.5 mM HEPES pH 7.5, 10 mM $CaCl_2$, 0.05% Digitonin) was added and samples were incubated on an ice-cold block for 30 min. STOP buffer (270 mM NaCl, 20 mM EDTA, 4 mM EGTA ,0.02% Digitonin, 50 µg glycogen, 50 µg RNase A) was added and samples were incubated at 37°C for 30 min, replaced on a magnet stand and the liquid was removed to a fresh tube. DNA was extracted by Phenol-Chloroform extraction and ethanol precipitation. Libraries were prepared as described in (Skene et al., 2018). Library quality was checked on a fragment analyzer, and paired-end

sequencing was performed on an Illumina NextSeq 500 instrument (read length 2 × 37 base pairs).

Cut & Run reads processing was done on the Duboule lab local Galaxy server (Afgan et al., 2016). Reads were mapped to the mouse genome (mm10) using Bowtie2 (v2.3.4.1) (Langmead and Salzberg, 2012), (-I 0 -X 1000 --fr --dovetail --very-sensitive). Reads with mapping quality below 30, mapping to mitochondria, or not properly paired were removed from the analysis. The output BAM file was converted to BED using bamtobed bedtools v2.18.2 (Quinlan, 2014). The coverage was obtained as the output of MACS2 (v2.1.1.20160309) (Zhang et al., 2008) (--format BED --keep-dup 1 --bdg --nomodel --extsize 200 --shift -100).

Ethics approval

All experiments were performed in agreement with the Swiss law on animal protection (LPA), under license No GE 81/14 (to DD).

Data availability

All raw and processed RNA-seq, 4C-seq, ChIP-seq, Cut & Run, and ATAC-seq datasets are available in the Gene Expression Omnibus (GEO) repository under accession number GSE138514.

Competing interests

The authors declare that they have no competing interests.

Funding

C.C.B is supported by the National Institute of Child Health & Human Development of the National Institutes of Health (NIH NICHD F32HD0935). This work was supported by funds from the École Polytechnique Fédérale (EPFL, Lausanne), the University of Geneva, the Swiss National Research Fund (No. 310030B_138662) and the European Research Council grants System*Hox* (No 232790) and Regul*Hox* (No 588029) (to D.D.). Funding bodies had no role in the design of the study and collection, analysis and interpretation of data and in writing the manuscript.

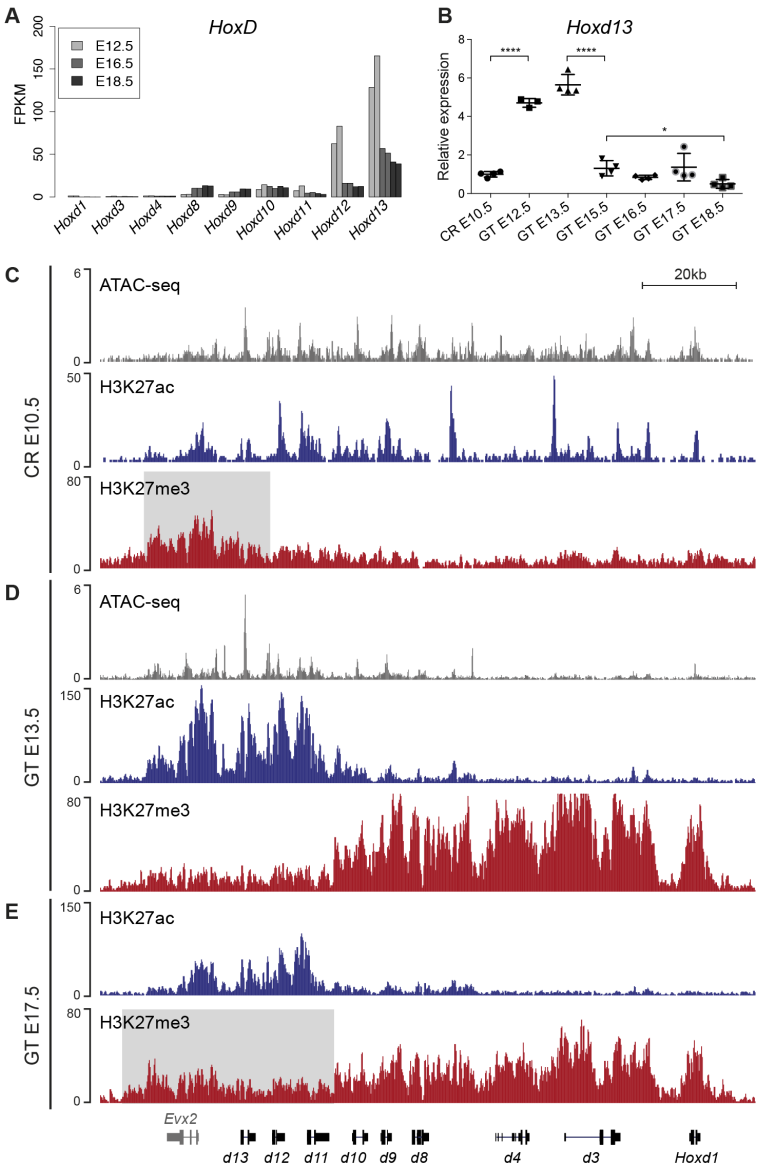
Author's contributions

Design of experiments, RA, CB, BM and DD; Bench work, RA, BM; Computing analysis, RA, LL-D; Analysis of results, RA, LL-D, DD; Manuscript writing, RA, CB, LL-D and DD; Funding acquisition, DD, CB. All authors read and approved the final manuscript

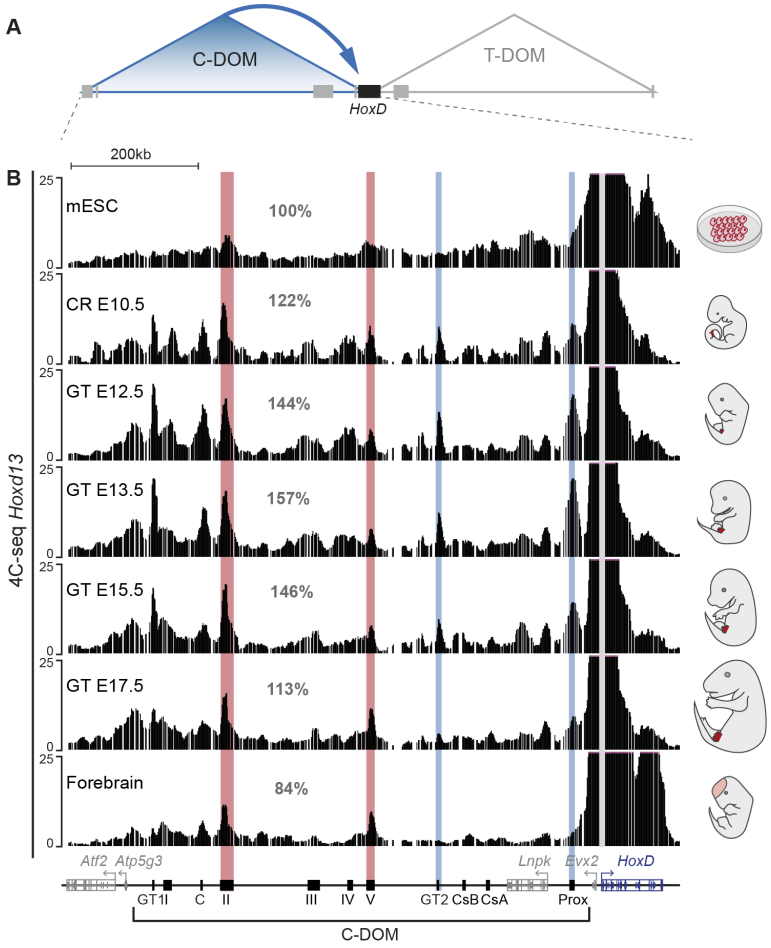
Acknowledgements

We thank all members of the Duboule laboratories for insightful comments and discussion. We are grateful to Sandra Gitto and Thi Hanh Nguyen Huynh for their help with mice breeding and genotyping, as well as Mylène Docquier, Brice Petit and Christelle Barraclough (University of Geneva) and Bastien Mangeat and the gene expression core facility (GECF, EPFL Lausanne) for DNA sequencing. We also thank Marion Leleu for her help with data analysis.

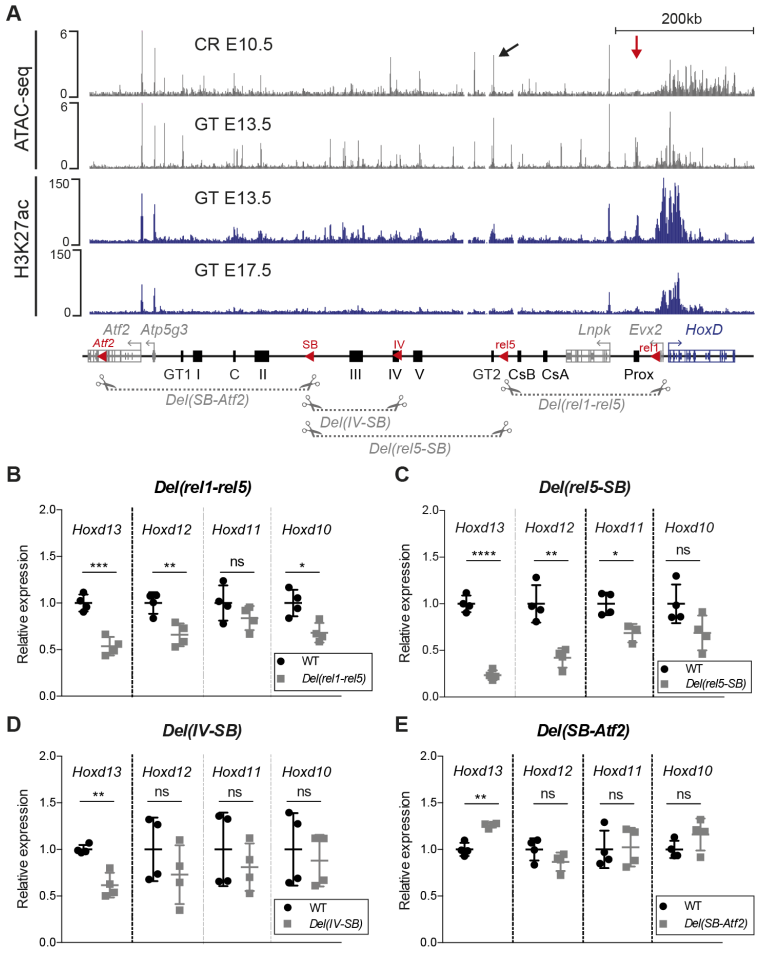
Amândio et al.
Figure 1



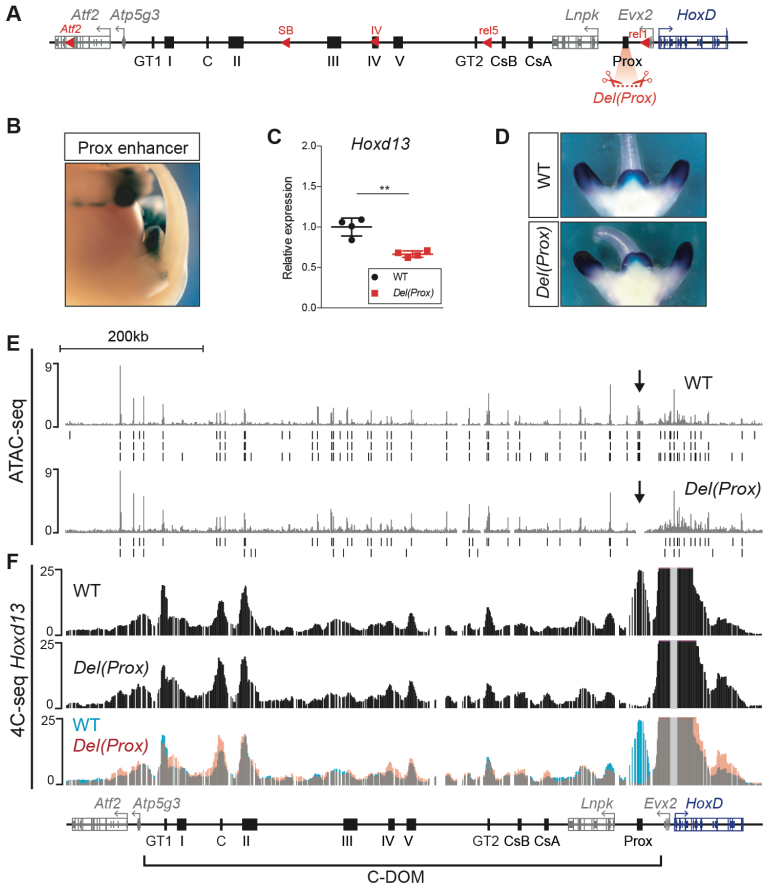
Amândio et al.
Figure 2



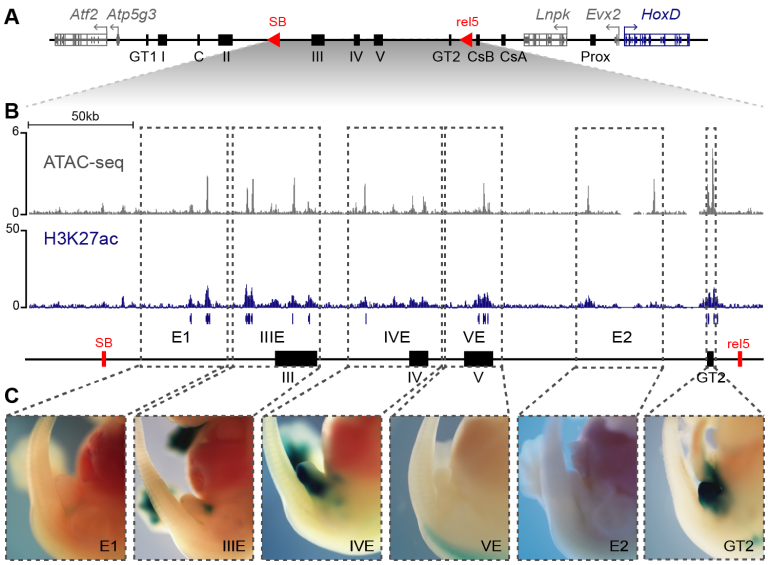
Amândio et al.
Figure 3



Amândio et al.
Figure 4

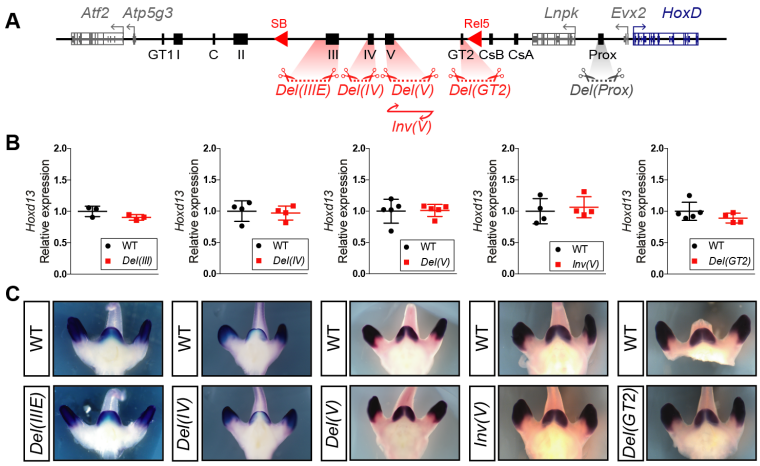


Amândio et al.
Figure 5

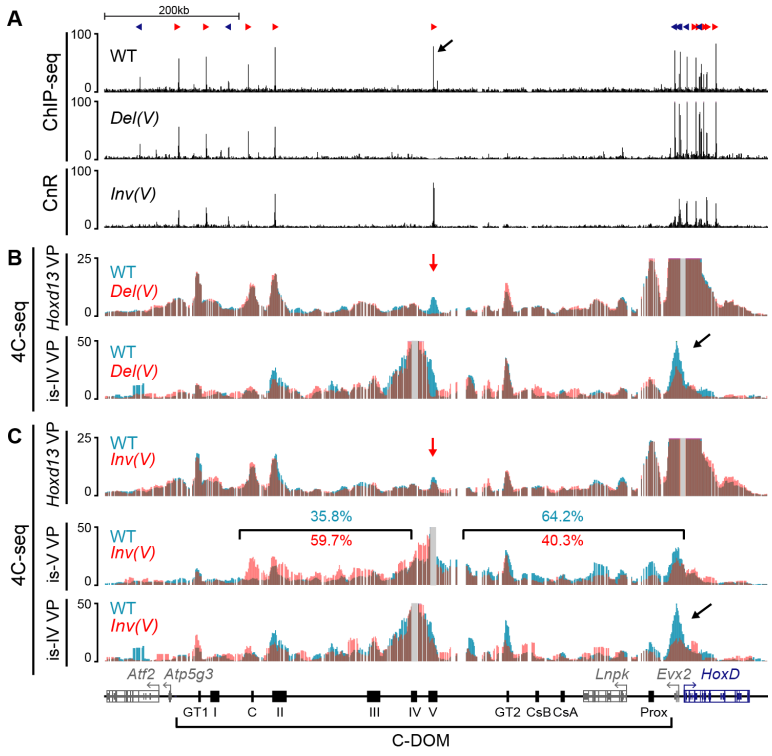


Amândio et al.

Figure 6



Amândio et al.
Figure 7



Amândio et al.
Figure 8