PNAS RESEARCH ARTICLE | DEVELOPMENTAL BIOLOGY OPEN ACCESS

CTCF-dependent insulation of *Hoxb13* and the heterochronic control of tail length

Lucille Lopez-Delisle[a,1], Jozsef Zakany[b,1], Célia Bochaton[a,1], Pierre Osteil[a,2], Alexandre Mayran[a], Fabrice Darbellay[a,3], Bénédicte Mascrez[b], Hocine Rekaik[a,c,4], and Denis Duboule[a,b,c,4]

Affiliations are included on p. 10.

Contributed by Denis Duboule; received July 24, 2024; accepted October 12, 2024; reviewed by Terence D. Capellini and René Rezsohazy

Mammalian tail length is controlled by several genetic determinants, among which are *Hox13* genes, whose function is to terminate the body axis. Accordingly, the precise timing in the transcriptional activation of these genes may impact upon body length. Unlike other *Hox* clusters, *HoxB* lacks posterior genes between *Hoxb9* and *Hoxb13*, two genes separated by a ca. 70 kb large DNA segment containing a high number of CTCF sites, potentially isolating *Hoxb13* from the rest of the cluster and thereby delaying its negative impact on trunk extension. We deleted the spacer DNA to induce a potential heterochronic gain of function of *Hoxb13* at physiological concentration and observed a shortening of the tail as well as other abnormal phenotypes. These defects were all rescued by inactivating *Hoxb13* in-cis with the deletion. A comparable gain of function was observed in mutant Embryonic Stem (ES) cells grown as pseudoembryos in vitro, which allowed us to examine in detail the importance of both the number and the orientation of CTCF sites in the insulating activity of the DNA spacer. A short cassette containing all the CTCF sites was sufficient to insulate *Hoxb13* from the rest of *HoxB*, and additional modifications of this CTCF cassette showed that two CTCF sites in convergent orientations were already capable of importantly delaying *Hoxb13* activation in these conditions. We discuss the relative importance of genomic distance versus number and orientation of CTCF sites in preventing *Hoxb13* to be activated too early during trunk extension and hence to modulate tail length.

axial elongation | CTCF | *Hox* timer | temporal colinearity | regulatory heterochrony

Significance

Extension of the embryonic body axis is in part terminated by the activation of the most posterior group 13 *Hox* genes. Therefore, the timing of activation of these terminal genes may dictate the eventual length of the body. We addressed this hypothesis by engineering a *HoxB* gene cluster where *Hoxb13* is activated prematurely and show that it produces series of morphological anomalies, as well as a shorter tail. By using pseudoembryos, we also reveal that the time of activation of *Hoxb13* is delayed by the presence of several CTCF sites, which together prevent *Hoxb13* to be transcribed too early, and hence, these CTCF sites allow for a complete axial morphogenesis to occur before extension is terminated.

The body axis of most mammalian species including humans (1) terminates with a tail of a defined and characteristic length. Since the isolation of the T/Brachyury gene (2), several genetic determinants of tail variation have been described, and progress has been made in exploring the mechanisms of length variability during development (3), in wild mouse populations (4), or during human evolution (5). In this context, several studies reported a role for the most posterior *Hox13* genes in setting up tail length (4, 6), and a genetic modification in mice (7) produced tail overgrowth reminiscent of the targeted inactivation of *Hoxb13* (8, 9). Conversely, forced expression of either *Hoxb13* or of other group 13 genes induced variable, often dramatic vertebral column truncations (10). As a consequence, it was proposed that *Hox13* genes were collectively responsible for terminating the extension of the main body axis during embryogenesis (10–12), thereby setting the length of the tail through a dominant negative effect of their protein products (13). Recent results obtained using unbiased approaches in natural populations of either deer mice (4) or Chinese long-tailed sheep breeds (14, 15) have further identified *Hoxd13* and *Hoxb13* as candidate genes to regulate tail length.

Since HOX13 proteins participate to the termination of the major body axis, it is important that these genes are kept silent until the time and the place when they need to be implemented such as to avoid premature termination. This is partly controlled by a tight mechanism that involves the separation of group 13 genes from the rest of the *Hox* clusters due to the presence of a chromatin border between two topologically associating domains (TADs) (16, 17), which insulate *Hox13* genes from regulatory influences emanating from the neighboring TAD (18, 19). Such TAD borders are often organized by the presence of several CTCF binding sites (CBSs) with opposite orientations leading to separate tropisms for chromatin loop extrusion (see ref. 20). This somewhat generic organization is particularly visible and conserved within the *HoxA*, *HoxC,* and *HoxD* gene clusters (19, 21), which likely reflects an ancestral chromatin architecture present before the occurrence of the full genomic duplications that generated these multiple clusters.

The *HoxB* cluster, on the other hand, shows a somewhat distinct topology since it lacks the *Hoxb10*, *Hoxb11,* and *Hoxb12* genes, i.e., the region that contains several such CTCF

Author contributions: L.L.-D., J.Z., C.B., H.R., and D.D. designed research; L.L.-D., J.Z., C.B., P.O., A.M., F.D., and B.M. performed research; L.L.-D., J.Z., C.B., P.O., A.M., F.D., B.M., H.R., and D.D. analyzed data; and L.L.-D. and D.D. wrote the paper.

Reviewers: T.D.C., Harvard University; and R.R., Université Catholique de Louvain.

The authors declare no competing interest.

[1]L.L.-D., J.Z., and C.B. contributed equally to this work.

[2]Present address: Faculté de Médecine, CNRS, INSERM, Genetics, Reproduction and Development Institute, Université Clermont Auvergne, Clermont-Ferrand F-63000, France.

[3]Present address: Department of Medical Genetics, Faculty of Medicine, University of Geneva, Geneva 4 1211, Switzerland.

[4]To whom correspondence may be addressed. Email: Hocine.Rekaik@college-de-france.fr or Denis.duboule@epfl.ch.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2414865121/-/DCSupplemental.

Published November 5, 2024.

sites in the other three clusters. However, *Hoxb13* is not found at the usually close vicinity of its nearest neighbor *Hoxb9*, but instead it lies ca. 70 kb far from it (8), which explains why it was initially overlooked when cloning the *HoxB* clusters in humans and mice (22, 23). Furthermore, several CBSs are found regularly spaced within this DNA segment (Fig. 1*A*, CBS5 to CBS10), as if *Hox* genes had been deleted after genome duplications while leaving in place their associated CBSs (18, 19). This suggests that the length of the spacer DNA segment and/or its content in CBSs, participate in the insulation of *Hoxb13* and thus delays its timing

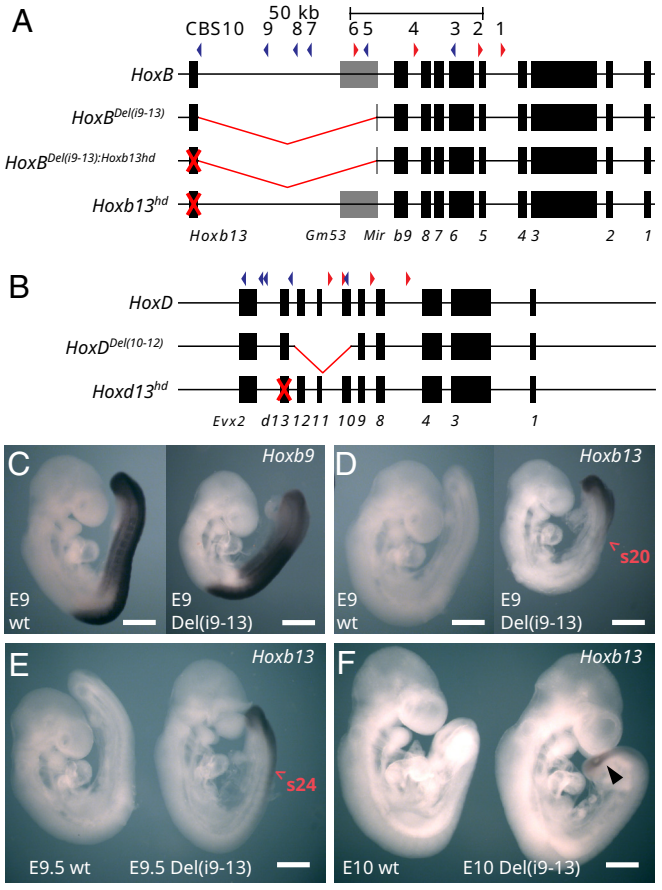

Fig. 1. Deletion of the *Hoxb9* to *Hoxb13* spacer DNA in the *HoxB* cluster. (*A*) Schematic representation of the *HoxB* alleles used in this work. On top, the CBSs are numbered 1 to 10 from *Hoxb1* to *Hoxb13,* and the colored arrowheads below indicate orientations. Below, the wt *HoxB* locus is shown. An approximately 70 kb large region separates *Hoxb9* from *Hoxb13*, without any protein-coding genes, whereas a LncRNA (*Gm53*) and the microRNA *Mir196a-1* ("Mir") are found within the 19 kb immediately flanking *Hoxb9*. The size of this spacer DNA is well conserved in all investigated species (*SI Appendix*, Fig. S1). As in all amniotes, *Hoxb10*, *11,* and *12* are absent. The extent of the induced *HoxB(i9-13)* deletion is shown below, as well as the loss of function $Hoxb13^{hd}$ allele generated either *in-cis* with the *HoxB(i9-13)* deletion, or *in-trans* (red crosses). (*B*) Schematic representation of the *HoxD* alleles used in this work (same scale as *HoxB*). On top, the wt *HoxD* locus is shown. The extent of the *Del(10-12)* deletion bringing *Hoxd13* in close proximity to *Hoxd9* (25) is shown below, as well as the loss of function $Hoxd13^{hd}$ allele (red cross) (26). (*C*) WISH using a *Hoxb9* probe at E9. *Hoxb9* expression pattern was indistinguishable between wt (*Left*) and homozygous $HoxB^{Del(i9-13)}$ mutant specimens (*Right*). (*D*) *Hoxb13* transcript accumulation at E9 in a homozygous $HoxB^{Del(i9-13)}$ mutant specimen (*Right*). At this stage, *Hoxb13* signal was undetectable in wt littermates (*Left*). The red arrowhead highlights the position of somite 20. (*E*) *Hoxb13* transcript accumulation in a E9.5 homozygous $HoxB^{Del(i9-13)}$ mutant specimen (*Right*), compared with a control littermate (*Left*) still negative for *Hoxb13* mRNAs. The red arrowhead highlights the position of somite 24. (*F*) *Hoxb13* expression in a homozygous $HoxB^{Del(i9-13)}$ mutant animals at E10 (*Right*) compared to a control littermate (*Left*). The expression of *Hoxb13* is highlighted by the black arrowhead. All specimens were treated in the same experiment in parallel, using the same stocks of reagents, incubations, and washes. The arrowhead shows expression in the mutant specimen only. (Scale bar: 500 μm.)

of activation. This hypothesis is supported by chromosome conformation capture experiments showing that this DNA spacer acts as a chromatin boundary between the *Hoxb1* to *Hoxb9* region, on the one hand, and *Hoxb13,* on the other hand (24), as expected from the orientations of five out of six CTCF sites present in the DNA spacer (19).

To verify the importance of this region in delaying *Hoxb13* expression in the elongating trunk, we shortened it from ca. 70 kb down to 6.6 kb such as to bring *Hoxb13* near *Hoxb9* by removing at the same time all CBSs present in this "spacer" DNA. We show that this recondensed *HoxB* cluster leads to mice with short tails assorted with additional thoracic and lumbar vertebra losses. A secondary targeted inactivation of *Hoxb13* in-*cis* rescues all vertebral column defects demonstrating that the gained HOXB13 protein is solely responsible for the abnormal phenotypes, while transcriptome studies suggest that tail shortening is due to the precocious activation of a *Hoxb13* functional program. To discriminate between the importance of the length of the DNA spacer versus the presence of multiple CBSs in the insulation of *Hoxb13*, we reproduced the same deletion in ES cell–derived pseudoembryos. Since premature *Hoxb13* activation was observed, a synthetic cassette containing all or some CTCF sites was recombined between *Hoxb9* and *Hoxb13* to see its impact upon keeping *Hoxb13* silent.

Results

We induced a 67.5 kb large deletion of the mouse *HoxB* intergenic spacer region by using CRISPR/Cas9 [the $HoxB^{Del(i9-13)}$ allele, Fig. 1*A*]. At the breakpoint near the *HoxB* cluster, the deletion included the *GM53* LncRNA of unknown function, yet the *Mir196a-1* was left in place due to its potential importance in regulating some *Hox* RNAs stability ((27) and ref. therein). A survey of 157 informative genome sequences revealed that this short distance between *Hoxb13* and *Hoxb9* (6.6 kb) is three times smaller than that observed in some rare fish genomes, while the shortest distances found in amniotes is six times this large in aves and at least eight times as large in placentalia (*SI Appendix*, Fig. S1*A*). Also, while the length of this spacer is globally maintained throughout mammals, the DNA sequence is poorly conserved (*SI Appendix*, Fig. S1*B*), even when mouse inbred strains are compared (*SI Appendix*, Fig. S1*C*), suggesting that a minimal DNA length between *Hoxb9* and *Hoxb13* had been selected. We also produced two control alleles where we inactivated *Hoxb13* function either in-*cis* with the spacer deletion [$HoxB^{Del(i9-13):Hoxb13hd}$], or in-*trans* ($Hoxb13^{hd}$, Fig. 1*A*). The related $HoxD^{Del(10-12)}$ and $Hoxd13^{hd}$ alleles (25, 26) were also investigated in this context to test for a potential synergy in phenotypes (Fig. 1*B*).

Heterochronic Shift of *Hoxb13* Expression in the Deletion Mutant. Whole mount in situ hybridization (WISH) on $HoxB^{Del(i9-13)}$ mutant embryos showed strong *Hoxb9* expression at E9 (Theiler stage 14) identical to the wild type (wt) expression (Fig. 1*C*), whereas *Hoxb13* transcripts were not yet detected (Fig. 1 *D*, *Left*). In contrast, expression of *Hoxb13* was detected as early as E9 in $HoxB^{Del(i9-13)}$ mutant embryos (Fig. 1 *D*, *Right*). The signal was localized to the posterior trunk, with an anterior limit approximately at somite 20 (Fig. 1*D*, red arrowhead), which contributes to the formation of prevertebra 16, i.e., thoracic vertebra 9 (T9). At E9.5, the anterior limit of *Hoxb13* signal had shifted to ca. somite number 24 (Fig. 1 *E*, *Right*). In formed epithelial somites, signal was weak, if any, compared to the posterior pole of the embryo including the presomitic mesoderm (PSM). In late E10 embryos, a weak ectopic *Hoxb13* signal was

still visible in the tail around the posterior neuropore (Fig. 1 *F*, *Right*, arrowhead). All these *Hoxb13* signals were clearly premature and ectopic. Indeed, the wt *Hoxb13* signal was first detected at somite level 45 at E11, Theiler stage 18, which corresponds to prevertebra 41 i.e., the eleventh caudal vertebra (9). These results were confirmed by RNA-sequencing analysis (see below).

Vertebral Column Malformations in Adult *HoxB*$^{Del(i9.13)}$ Mutant Mice. Gross observations of *HoxB*$^{Del(i9-13)}$ mutant specimens [or Del(i9-13)] revealed a tail shorter than in wt littermates, with measurements of the distance between the anus and the tail tip in F2 adults quantifying the extent of these distinct tail truncations. These observations were controlled by measures taken after skeletal preparations and µCT scans of the various alleles (Fig. 2*A*). To verify that the ectopic gain of *Hoxb13* expression was indeed responsible for tail shortening, we analyzed the two separate alleles [*HoxB*$^{Del(i9-13):Hoxb13hd1}$] and [*HoxB*$^{Del(i9-13):Hoxb13hd2}$] where a secondary mutation inactivated HOXB13 in *cis* with the *HoxB*$^{Del(i9-13)}$ deletion (Figs. 1*A* and 2*A*). Siblings from both breeding stocks were analyzed for their tail length past 8 wk of age and a clear rank of mean tail sizes was observed within the allelic series (Fig. 2*B*). As expected, there was no difference in tail length between the *HoxB*$^{Del(i9-13):Hoxb13hd}$ and the *Hoxb13*^{hd} alleles (Fig. 2*B* and *SI Appendix*, Fig. S2). This was not unexpected since the two *Hoxb13*^{hd} loss of function alleles had deletions in their DNA binding motives (*SI Appendix*, Table S1), which likely made both proteins transcriptionally inactive, in a way similar to the initial allele where a lacZ construct was introduced at about the same position (9). Consequently, all mice carrying a homozygous loss of function of *Hoxb13* had elongated tails, regardless of their complete genotype and in agreement with the initial report (9). As HOXB13 inactivation in-*cis* rescued the tail shortening effect, we concluded that reduced tail length directly resulted from the *Hoxb13* gain of function.

We used skeletal preparations of adult specimens from the same breeding stock to evaluate to what extent such variations in tail length were due to changes in the number of vertebral types. Control specimens displayed between 28 and 30 complete caudal vertebrae (Fig. 3*A*). In Del(i9-13) homozygous animals, this number was between 24 and 27, while heterozygous displayed between 24 and 29 vertebrae (Fig. 3*A*). Skeletal alterations were also scored outside the tail, for all Del(i9-13) homozygous mice and most heterozygous displayed a number of ribs bearing thoracic vertebrae reduced by one (Fig. 3 *A* and *B*), and a reduction in the number of lumbar vertebrae to L4 was sporadically observed (Fig. 3 *A*, *Right*). Therefore, the most affected individuals had a C7, T12, L4, S4, C25 vertebral formula, instead of the control C7, T13, L5, S4, C29 formula, prevalent in this background (*SI Appendix*, Fig. S3). The defects in homozygous animals were more prevalent than in heterozygous and with higher expressivity, consistent with the gene dosage effect already observed in the tail length phenotype.

This dosage effect was further assessed by adding to Del(i9-13) animals a *Hoxd13* gain of function produced by the deletion of *Hoxd10* through *Hoxd12*, i.e., a *HoxD* allele identical to the Del(i9-13) allele on the *HoxB* cluster, bringing *Hoxd13* next to *Hoxd9*. In this case, a transient gain of *Hoxd13* expression was observed (25). *HoxB*$^{Del(i9-13)}$:*HoxD*$^{Del(10-12)}$ compound mutants showed a further decrease in the number of caudal vertebra when compared to *HoxB*$^{Del(i9-13)}$ (*SI Appendix*, Fig. S2 *B* and *C*). While this decrease was minor, it remained statistically significant. We conclude that group 13 genes might share a function in limiting posterior growth zone elongation during generation of tail somites. However, the contribution of *Hoxb13* in this task was clearly more prominent than that of *Hoxd13*.

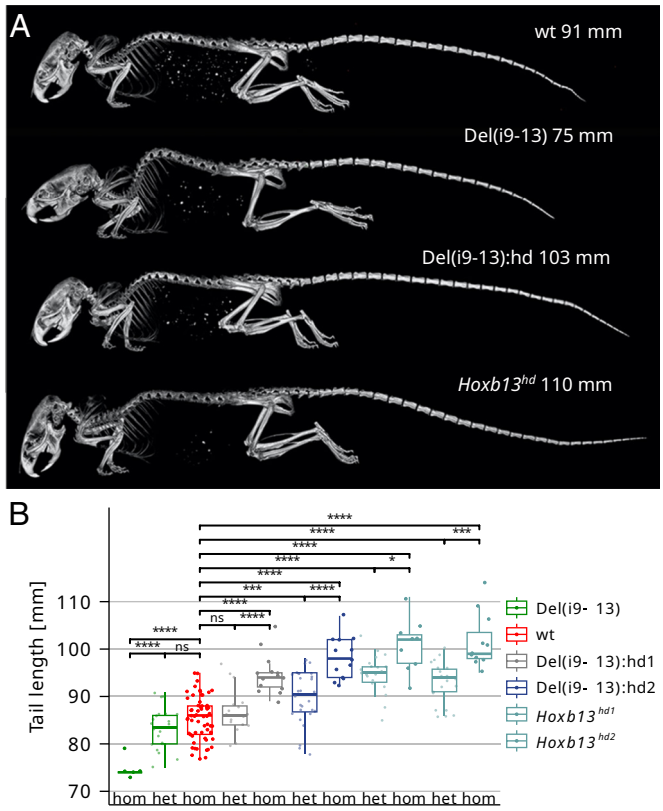

Fig. 2. Effects of *Hoxb13* gain and loss of function upon tail length. (*A*) Lateral views of µCT 3D reconstructions of skeletons of representative phenotypes of adult animals carrying the various *HoxB* alleles shown in Fig. 1*A*. On top is a control animal with a tail length of 91 mm. Below is a *HoxB*$^{Del(i9-13)}$ homozygous specimen with a tail reduced to 75 mm (minus 18%). The inactivation of *Hoxb13* *in-cis* with this deletion [below, *HoxB*$^{Del(i9-13):Hoxb13hd}$] displayed a longer tail (103 mm, plus 13%), equivalent in length to those of mice lacking *Hoxb13* function (*Bottom*). (*B*) Quantifications of tail lengths. While both homozygote *HoxB*$^{Del(i9-13)}$ (strong green) and *HoxB*$^{Del(i9-13):Hoxb13hd}$ [Del(i9-13):hd1, strong gray] and Del(i9-13):hd2 (strong blue) specimen displayed statistically significant deviations from controls (wt, red), their heterozygous versions (weak green and weak gray and blue, respectively) were much closer to wt counterparts. hd1 and hd2 correspond to two independent lines. All wt littermates of all mutant lines were pooled together. Significance is assessed by the two-sided Welsh's *t* test (ns: $P > 0.05$, *$P <= 0.05$, **$P <= 0.01$, ***$P <= 0.001$, and ****$P <= 0.0001$). Number of animals for each line (heterozygous and homozygous): Del(i9-13): 22 and 5, wt: 60, Del(i9-13):hd1: 17 and 14, Del(i9-13):hd2: 28 and 13, *Hoxb13*^{hd1}: 20 and 9, *Hoxb13*^{hd2}: 22 and 11. The lower and upper hinges correspond to the first and third quartiles. The upper whisker extends from the hinge to the largest value no further than 1.5 * IQR from the hinge.

Transcriptome Analyses. We looked at the impact of *Hoxb13* gain of function on the embryonic transcriptomes at E9.5 and E10.5. RNA was extracted from "posterior embryos," i.e., dissected from below the forelimb buds, with three samples of each genotype. *Hoxb13* transcript levels were dramatically increased up to 17 FPKM in E9 Del(i9-13) samples (adjusted *P*-value of 1e-66), as well as to ca. 4 FPKM in both E9.5 and E10.5 *HoxB*$^{Del(i9-13)}$ and *HoxB*$^{Del(i9-13):Hoxb13hd}$ (adjusted *P*-value of 0.02 and 0.004 respectively; *SI Appendix*, Fig. S4*A*). Principal component analysis confirmed that most of the variance was likely due to the developmental stage (*SI Appendix*, Fig. S4*B*). The distribution of samples along the second principal component is explained by genotypes. The relative position of the *HoxB*$^{Del(i9-13)}$ transcriptome was concordant between both stages and the *HoxB*$^{Del(i9-13):Hoxb13hd}$ samples were expectedly more similar to controls. Expression levels of all *Hox* genes confirmed that the assigned developmental stages were indeed correct and that no major change was scored beside *Hoxb13* RNAs (*SI Appendix*, Fig. S4*C*).

We next looked for RNAs which were modulated between wt and *HoxB*$^{Del(i9\text{-}13)}$ embryos at both stages and inversely regulated in the *HoxB*$^{Del(i9\text{-}13):Hoxb13hd}$ allele. We identified only two genes significantly up-regulated with *Hoxb13* gain of function (*Chl1*, *Baiap2l1*, *SI Appendix*, Fig. S5*A*), whereas a single gene was down-regulated (*Rbpj*, *SI Appendix*, Fig. S5*B*). The decrease in *Rbpj* mRNAs is noteworthy, for *Rbpj* loss of function mutants show a developmental arrest during early somitogenesis (28).

ES Cell–Derived Gastruloids as Proxies to Study *Hoxb13* Insulation. To study the importance of the DNA spacer length versus the presence and number of CBSs, in the necessary insulation of *Hoxb13,* we turned to gastruloids, i.e., mES cell–derived pseudoembryos (29, 30), which are excellent proxies to study the extending posterior part of mammalian embryos (31, 32). Gastruloids can be produced in large amount and are thus amenable to high-throughput analyses (19). We reproduced in ES cells the same Del(i9-13) deficiency and produced heterozygous gastruloids. We examined the expression of both *Hoxb9* and *Hoxb13* by WISH in control and mutant gastruloids and found no salient difference in the expression of *Hoxb9* at 120 h after aggregation (AA) (Fig. 4*A*), a stage most closely related to a E8.5 posterior mouse embryo (31). In both cases, the entire extending (posterior) part was positive with a clearly delimited spatial boundary at a more 'anterior' level. At this stage, as well as at 144 h AA, control gastruloids did not show any trace of *Hoxb13* mRNAs, yet a strong signal was detected in Del(i9-13) mutant specimen, with a full penetrance (Fig. 4 *A* and *B*). The position of the expression boundary was slightly more posterior than that of *Hoxd9* (Fig. 4*C*) in agreement with the colinear distribution of *Hox* transcripts previously observed in gastruloids (19, 31).

Same Cellular Populations in Both Control and *Hoxb13* Gastruloids. To look at which cells precisely expressed the gained *Hoxb13* mRNAs, as well as to evaluate any potential effects of this gain of function either upon gene expression or on the distribution of cell types in these mutant gastruloids, we carried out single-cell RNA analysis using 144 h gastruloids, a stage roughly corresponding to E9 embryos (31), where *Hoxb13* expression was well established in mutant specimens. The analysis of this single-cell RNA-sequencing (scRNA-seq) dataset revealed that the distribution of cellular clusters remained virtually unchanged between control and mutant gastruloids (*SI Appendix*, Fig. S6*A*). It also showed that gained *Hoxb13* mRNAs were mostly found within neuro-mesodermal progenitors (NMP) cells, as well as within the "Neural Tube 1" cluster, likely composed of early differentiating neuronal cells, adjacent to the NMP cluster (*SI Appendix*, Fig. S6*B*). Clustering based on all *Hox* gene expression throughout cell fates revealed that *Hoxb13*, which normally clusters with its paralogous *Hoxa13* gene, now clusters with group 9-10 *Hox* genes in the mutant gastruloids, i.e., *Hoxb13* now is expressed in cells with a general context related to the latter genes (*SI Appendix*, Fig. S6*C*, arrowheads).

Using baredSC (33), the exact coexpression of genes was assessed, and while no positive correlation was found with the expression of either *Hoxa9*, *Hoxc9,* and *Hoxd9*, a strong positive correlation was observed with the expression of *Hoxb9*, i.e., the new immediate neighbor of *Hoxb13* after deletion of the spacer DNA. Indeed, 54 percent (±14%) of cells positive for *Hoxb9* were also positive for the ectopic *Hoxb13* mRNAs (*SI Appendix*, Fig. S6 *D* and *E*), illustrating that the relocated *Hoxb13* gene was transcribed at the time and in the cells where an elusive *Hoxb10* gene would have likely been transcribed, if it was still present in the amniote *HoxB* cluster, as is the case for some anamniotes species like zebrafishes. This illustrates once more the decisive role played by the relative position of *Hox* genes in their respective cluster, rather than by their promoters, in the precision of their transcriptional regulation (references in ref. 34).

Altogether, the gain of *Hoxb13* expression did not drastically modify the transcriptional landscape of gastruloids at 144 h, when the analysis was carried out using control and mutant specimens

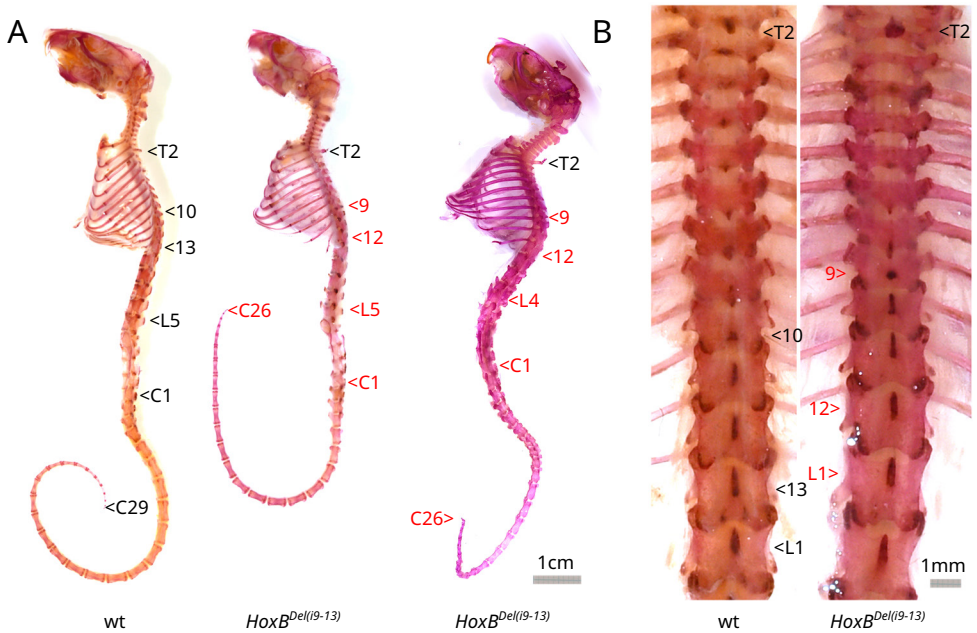

Fig. 3. *Hoxb13* gain of function-dependent vertebral anomalies. (*A*) Alizarin-stained skeletal preparations showing the reproducible vertebral formulae in both wt (*Left*) and *HoxB*$^{Del(i9\text{-}13)}$ mutants (*Middle* and *Right*). The wt skeletal pattern was composed of 7 cervical (C7), 13 thoracic (T13), 5 -rarely 6- lumbar (L5), 4 sacral (S4) and 29 -rarely 30- caudal (C29) vertebrae. The patterns most frequently observed in *HoxB*$^{Del(i9\text{-}13)}$ homozygous mutants are shown in the *Middle* and *Right*. (*B*) Magnification of the thoracic regions of a wt (*Left*) and *HoxB*$^{Del(i9\text{-}13)}$ mutant (*Right*). In several heterozygous and all homozygous mutants, T9 carried a neural spine typical of wt T10 (the so-called anticlinal vertebra). In such specimens, only 12 thoracic rib bearing vertebrae were scored and hence the T9-T10 exchange may be considered as a loss of normal T9 followed by a serial transformation. In the majority of hets and all homs, the caudal series was composed of less than 29 complete vertebrae. In the depicted mutant specimens, the number was 26, instead of 29. The corresponding quantifications are in *SI Appendix*, Fig. S2*B*.

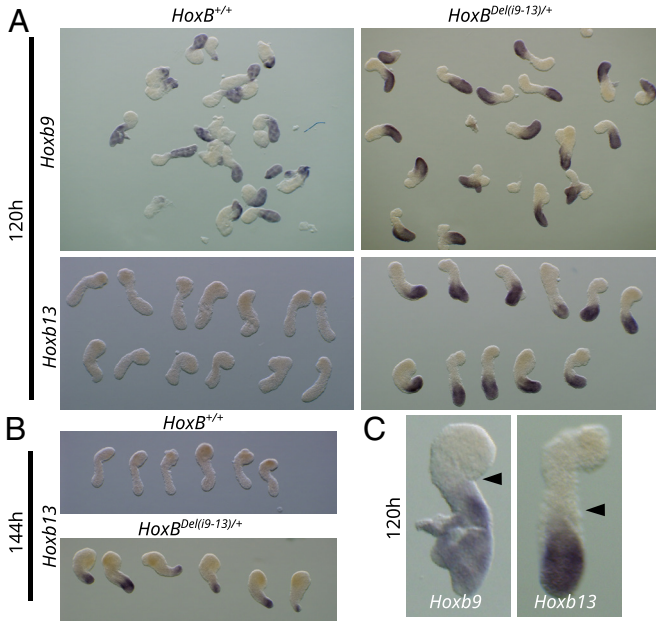

Fig. 4. *Hoxb9* and *Hoxb13* expression in control and mutant gastruloids. (*A*) Whole-mount in situ hybridization for *Hoxb9 Top* and *Hoxb13 Bottom* on either wt (*Left*) or $HoxB^{Del(i9\text{-}13)/+}$ (*Right*) 120 h gastruloids showing the clear and penetrant gain of *Hoxb13* in the mutant condition. (*B*) WISH of *Hoxb13* on wt (*Top*) and $HoxB^{Del(i9\text{-}13)/+}$ (*Bottom*) 144 h gastruloids. (*C*) Comparison of the pattern of expression of *Hoxb9* in wt and *Hoxb13* in $HoxB^{Del(i9\text{-}13)/+}$ gastruloids at 120 h, showing colinearity in the anterior–posterior (AP) expression border (arrowheads).

from the same batch. This observation agrees with our comparable analysis in mouse embryos (see above) and further confirmed that the effect of this gain of function is mostly quantitative, i.e., involving differences in transcription timing, rather than purely qualitative such as starting a distinct transcription program, in agreement with the phenotypes observed in mice.

Insulation of *Hoxb13* from the *HoxB* Cluster. Next, we carried out a series of analyses in 132 h gastruloids to better document the role of this spacer DNA in the isolation of *Hoxb13* from earlier and more "anterior" regulations. At this stage, the H3K27 acetylation (H3K27ac) not only covered the entire *Hoxb1* to *Hoxb9* region but also included the nearby located *Mir196a-1* locus and the *GM53* LncRNA, i.e., a region including both CBS5 and CBS6 (Fig. 5*A*), indicating that this immediate neighborhood was also actively transcribed. The ChIP profile of NIPBL, a factor that helps loading of the cohesin complex, revealed an enrichment throughout this H3K27ac positive region, which matched with the distribution of RAD21, a subunit of the cohesin complex that was enriched all over this region, with a strong accumulation at CBS7 and CBS8 (Fig. 5*A*). From this, we conclude that as in the case of the *HoxD* cluster (19), cohesin deposition and loop extrusion occurred in an asymmetric manner. Cohesin was already detected beyond the *Hoxb9* position and hence the main blockage to looping to *Hoxb13* was likely achieved by either CBS7 or CBS8, or both, despite their opposite orientations.

To verify this directionality in regulation, we analyzed a H3K27ac HiChIP dataset, which revealed that this entire transcribed region strongly interacted with sequences located mostly in 3′, i.e., within the same TAD, confirming that the early activation of *Hox* genes involves regulatory inputs coming from this TAD (19, 35), whereas the "posterior" TAD, including *Hoxb13*, was tightly insulated from these regulations with no signal involving *Hoxb13* itself (Fig. 5*B*).

To further assess the mechanism of *Hoxb13* insulation, in particular to discriminate between the effect of the linear genomic distance versus the presence and orientation of the six CBS localized between *Hoxb9* and *Hoxb13* (Figs. 4*A* and 6), we produced an ES cell clone deficient for one copy of the entire *HoxB* locus, i.e., carrying a targeted deletion including from *Hoxb1* to *Hoxb13* ($HoxB^{Del}$; Fig. 6*A*). A deletion of the DNA spacer was then induced on the other chromosome, thus leading to the $HoxB^{Del/Del(i9\text{-}13)}$ configuration (Fig. 6). Similar to the $HoxB^{Del(i9\text{-}13)/+}$ heterozygous gastruloids, their $HoxB^{Del/Del(i9\text{-}13)}$ mutant counterparts expressed *Hoxb13* prematurely and at an ectopic anterior position in 120 h gastruloids (Fig. 6), whereas *Hoxb9* expression remained unchanged (*SI Appendix*, Fig. S7). Noteworthy, $HoxB^{Del/Del(i9\text{-}13)}$ mutant gastruloids displayed some variability (up to three folds) in the strength of *Hoxb13* activation between different replicates (*SI Appendix*, Fig. S7).

We used the $HoxB^{Del/Del(i9\text{-}13)}$ ES cells as hosts to reintroduce the complete series of deleted CBSs, yet without the large DNA spacing in between. We used a synthetic DNA cassette containing the six CBSs initially identified in the spacer DNA (CBS 5 to CBS 10, Figs. 1 and 6*A*), but separated from one another by ca. 500 bp. The overall size of this cassette was thus of ca. 3 kb. Homology arms were introduced and the cassette was recombined between *Hoxb9* and *Hoxb13*, on top of the $HoxB^{Del(i9\text{-}13)}$ chromosome to produce the $HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10)}$ allele. The cassette contained all six CBSs in their native orientations and hence the number, orientation, and relative sequence of CBSs were as in the wt chromosome, except that the genomic distance was reduced to a size generally found between any neighboring *Hox* genes (Fig. 6 *A*, *Bottom*) and that *Mir-196a-1* was now 5′ of the cassette.

ChIPmentation (ChIPM-seq) was carried out using 96 h mutant gastruloids to check for the presence of both bound CTCF and the cohesin complex through its RAD21 subunit. Sequence reads were aligned onto an in silico reconstructed Del(i9-13):Ins(CBS5-10) mutant genome. The CTCF profile observed was comparable to the control situation in terms of relative intensities, with CBS5, 7, and 8 being prominent whereas CBS6 was barely scored (Fig. 6*B*, compare first and third profiles). Likewise, RAD21 mostly accumulated at CBS7 and CBS8, as in the control situation (Fig. 6*B*; second and fourth profiles), even though the global accumulation was weaker, in part due to the hemizygosity of the locus. The recombination of this CBS cassette between *Hoxb9* and *Hoxb13* in ES cells had no detectable effect upon the mRNA levels of either *Hoxb1* or *Hoxb9* in gastruloids derived thereof (Fig. 6 *C*, *Top*). In contrast, gastruloids generated from $HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10)}$ ES cells had no detectable expression of *Hoxb13* by WISH, and mRNA levels were poorly significant, either in 120 h gastruloids, or later at 144 h (Fig. 6 *C*, *Bottom* and *SI Appendix*, Fig. S7).

To further verify that the insertion of the cassette in this $HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10)}$ allele had not structurally affected the *Hoxb13* transcription unit in any way, the mutant locus was entirely DNA-sequenced by nCATS (36) and no obvious off-target modifications were scored (*SI Appendix*, Fig. S8 *A* and *B*). Furthermore, we deleted the cassette after its insertion [the $HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10:Del(CBS5\text{-}10))}$] and RNA-seq experiments comparing the $HoxB^{Del(i9\text{-}13)}$ mutant gastruloids at 120 h with the $HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10)}$ and the $HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10:Del(CBS5\text{-}10))}$ versions (with and without the CTCF cassette inserted) were carried out. Deletion of the cassette led to an upregulation of *Hoxb13* transcription close -if not equal (65%)- to that observed before its recombination (*SI Appendix*, Fig. S8*C*). From this, we conclude that this synthetic cassette was capable to achieve most of *Hoxb13* insulation in gastruloids, despite the short distance

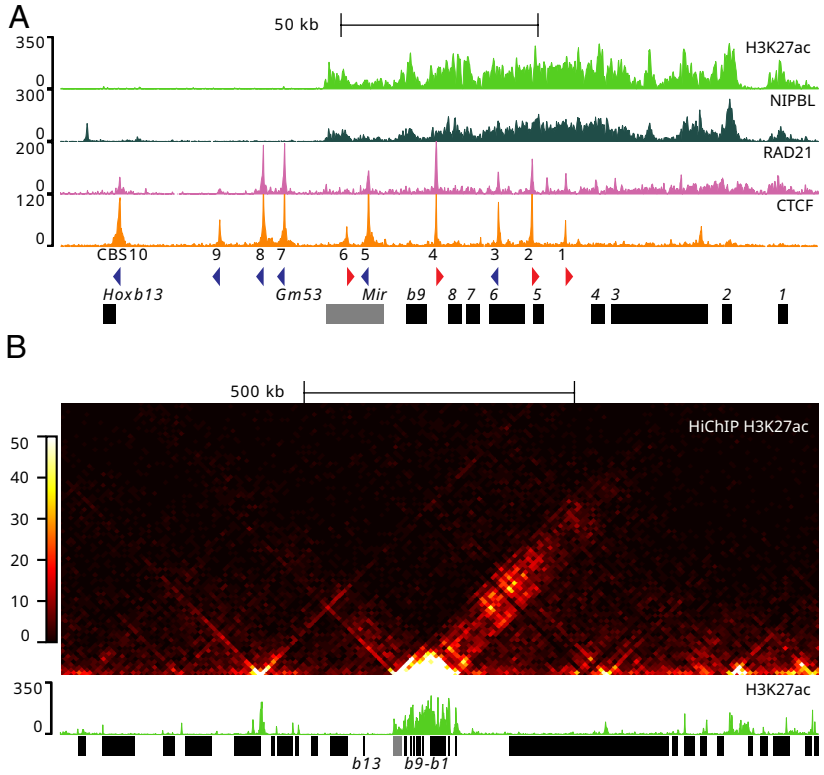

Fig. 5. Chromatin structure of the *HoxB* locus in late gastruloids. (*A*) Profiles of either H3K27ac (light green), NIPBL (dark green), and RAD21 (magenta) in 132 h gastruloids, or CTCF at 168 h over the *HoxB* locus (data extracted from ref. 19, GSE205779). CBSs are numbered 1 to 10 from *Hoxb1* to *Hoxb13* and the colored arrowheads below indicate orientations. *Hoxb* genes are shown as black rectangles below. The *HoxB* cluster is active until the *Mir196a-1* gene, while *Hoxb13* is tightly isolated. (*B*) HiChIP of H3K27ac over a Mb of DNA containing the *HoxB* cluster (indicated below), with the matching profile of H3K27ac ChIP-seq of 120 h gastruloids below (green) (data extracted from ref. 19, GSM6226299 and GSM6226246). Black rectangles below represent protein-coding genes. *Gm53* and *Mir196a-1* are in gray. The vast majority of contact between "active" regions, i.e., involving H3K27ac, are toward the TAD located 3′ of the cluster, whereas *Hoxb13* is not at all involved.

between *Hoxb9* and *Hoxb13* in the $HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10)}$ mutant allele.

CTCF and *Hoxb13* Insulation. While the CTCF ChIPM-seq profiles of both the control and the $HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10)}$ mutant allele indicated that all CBS can be bound (Fig. 6*B*), it did not reveal whether all (or several) sites must be bound simultaneously to achieve full insulation or, alternatively, if CTCF binding to only few sites is sufficient. We thus deleted either three or four consecutive CBSs from the cassette (Fig. 7) and look at the effect upon *Hoxb13* transcription both by WISH and RNA-seq. ES cells carrying these deletions were verified for the presence of bound CTCF on the remaining sites either by ChIPmentation or by CUT&RUN approaches (Fig. 7*A*). While the deletion of CBS8 to 10 [Fig. 7*A*; Del(CBS8-10)] did not elicit any detectable upregulation of *Hoxb13* (Fig. 7*B*), a slight though significant increase was scored when the four CBS7 to 10 were deleted [Fig. 7 *A* and *B*; Del(CBS7-10)]. The latter increase was nevertheless detectable by WISH as a weak but clear signal in most posterior gastruloid cells (Fig. 7*C*, arrowhead). However, as was the case for the $HoxB^{Del/Del(i9\text{-}13)}$ gastruloids, the response to this partial CBS deletion was somewhat variable among replicates and data in Fig. 7*C* show gastruloids with the highest amount of *Hoxb13* mRNAs. These results indicated that while only 2 CBSs with opposite orientations could not achieve full isolation, they were still capable to importantly delay ectopic *Hoxb13* activation.

Finally, we looked at the insulating capacity of the native CBSs reintroduced into the $HoxB^{Del/Del(i9\text{-}13)}$ chromosome though with inverted orientations. The inversion of the full reinserted cassette [$HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10):Inv(CBS5\text{-}10)}$] had an insulating capacity close to -if not identical with- the insertion of the cassette carrying the CBSs in their native orientation, suggesting that the orientation of the CTCF sites was not the main determinant of the insulation, at least in this particular topology (*SI Appendix*, Fig. S9). We also produced a $HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10):Del(CBS7\text{-}10)}$ allele where the remaining CBS5 and CBS6 had an inverted configuration. While gastruloids produced from this [$HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10):Inv(CBS5\text{-}10):Del(CBS7\text{-}10)}$] configuration showed a gain of *Hoxb13* expression (*SI Appendix*, Fig. S9 *A* and *B*), the leakage in insulation was in the same range than that observed before the inversion of these two CBSs.

Discussion

Past experimental approaches have suggested that *Hox* genes belonging to the paralogy group 13 control the end of axial extension during vertebrate development. Accordingly, their functional modulation could be determinant in species-specific body lengths (9–11). In addition, adaptive variations in tail lengths under natural conditions were associated with few QTLs in deer mice, among which, one affected *Hoxd13*. In this case, a lower amount of *Hoxd13* was associated with a larger PSM, from where somites derive, due to increased number of NMP cells (4). Furthermore, *Hoxb13* was recently identified as the main candidate gene for tail length in Chinese long-tailed sheep breeds (37). Here, we demonstrate that a physiological gain of function of *Hoxb13* leads to both a drastic variation in tail length and other vertebrate transformations at more rostral body levels. Such phenotypes associated with the ectopic expression of *Hox13* genes could explain the evolution of a TAD boundary always located between the latter genes and the rest of the gene clusters thereby insulating this terminal

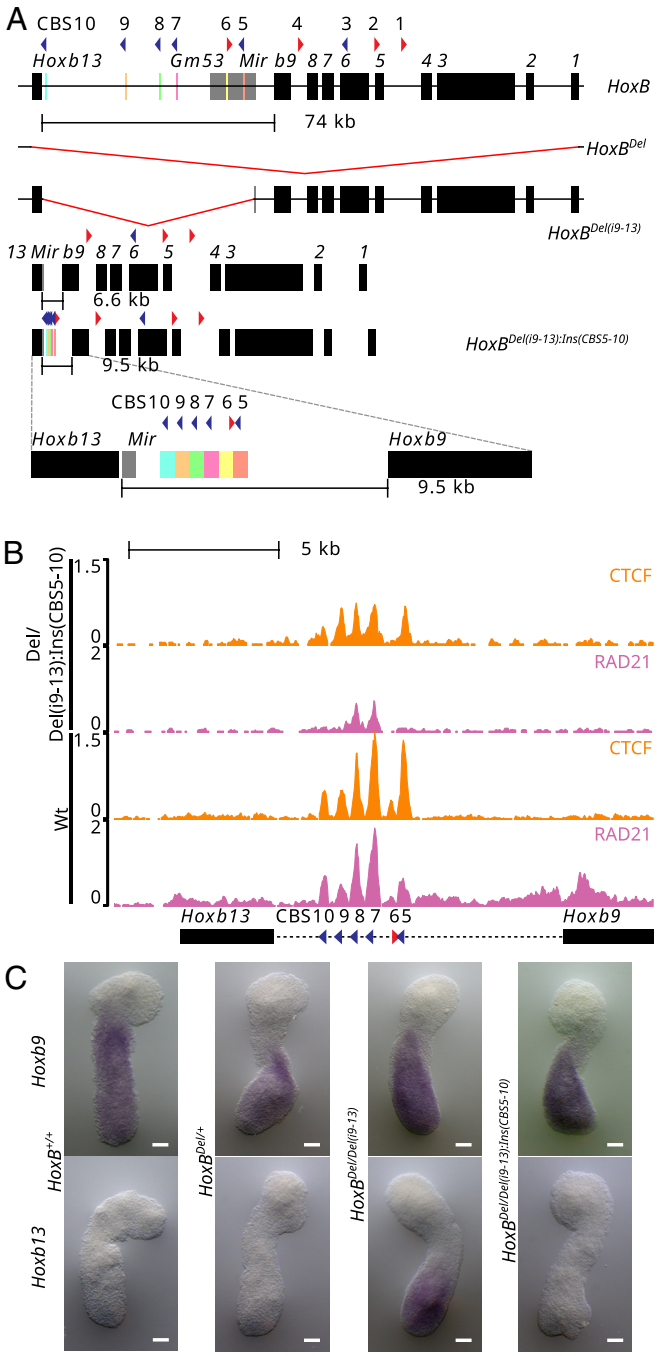

Fig. 6. Insertion of a CTCF cassette into the deleted spacer DNA. (*A*) Schematic of the various *HoxB* alleles with on top the wt *HoxB* locus along with the 74 kb large DNA spacer indicated as a black line. Below is shown the large deletion of the entire *HoxB* locus used to balance the other alleles. The third track shows the deletion of the (9-13) intergenic region, leading to the "consolidated" *HoxB* cluster shown in the fourth track. In this latter configuration, the distance between *Hoxb9* and *Hoxb13* is now 6.6 kb without any CBSs. The fifth track shows the allele after insertion of a cassette containing the 6 native CBS separated from one another by ca. 0.5 kb. The distance between *Hoxb9* and *Hoxb13* is now of 9.5 kb. This new intergenic region is magnified below, with each colored block containing a CBS corresponding to those of the wt allele, in the right order (see first track). (*B*) ChIP-seq profiles of CTCF (orange) and RAD21 (magenta) using the $HoxB^{Del/Del(i9-13):Ins(CBS5-10)}$ allele (the insertion of the cassette) on top and wt on bottom on 96 h gastruloids aligned on the *in silico* reconstructed Del(i9-13):Ins(CBS5-10) mutant genome. (*C*) Whole mount in situ hybridization with *Hoxb9* (*Top*) and *Hoxb13* (*Bottom*) probes on 120 h gastruloids with the following genotypes from *Left* to *Right*: $HoxB^{+/+}$, $HoxB^{Del/+}$, $HoxB^{Del/Del(i9-13)}$, and $HoxB^{Del/Del(i9-13):Ins(CBS5-10)}$. The insertion of the CTCF cassette fully suppresses the *Hoxb13* gain of function, even with a much shorter distance between *Hoxb9* and *Hoxb13*. For quantifications, see *SI Appendix*, Fig. S7. (Scale bar: 100 μm.)

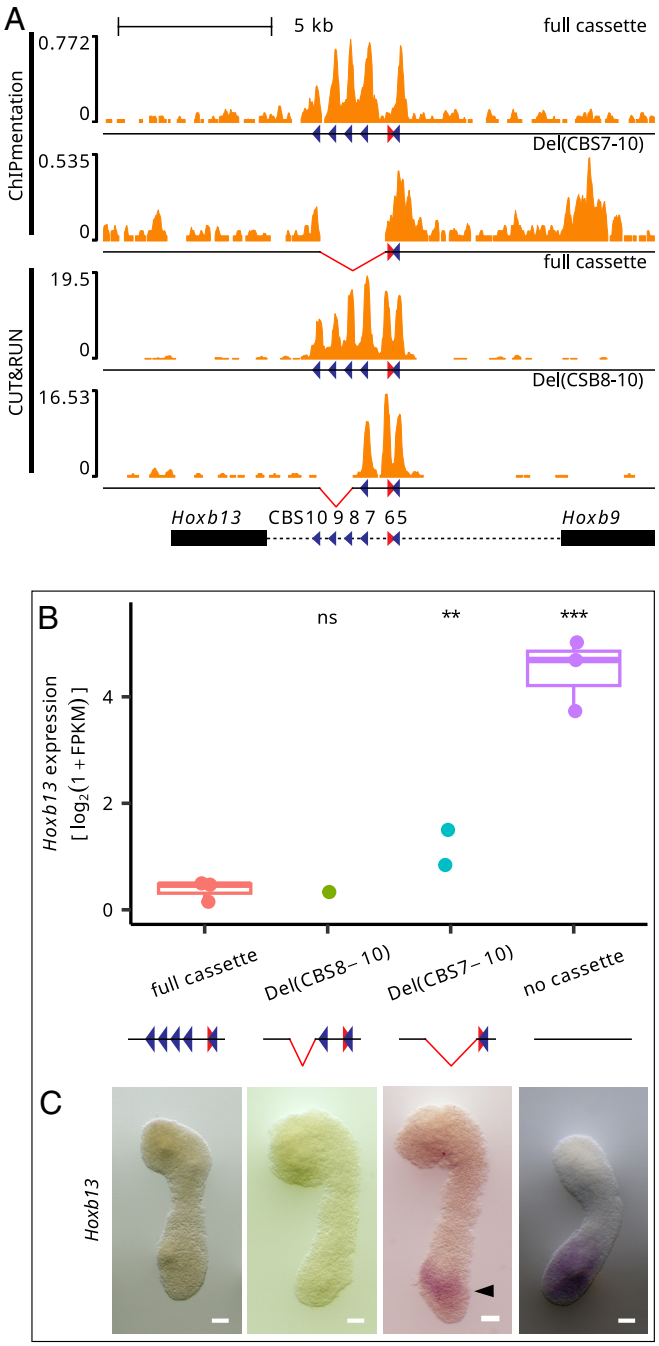

Fig. 7. Deletion of CBSs from the inserted CTCF cassette. (*A*) ChIPmentation on 96 h gastruloids (*Top* two profiles) and CUT&RUN on ES cells (two profiles on *Bottom*) of CTCF from the $HoxB^{Del/Del(i9-13):Ins(CBS5-10):Del(CBS7-10)}$ [*Top*, Del(CBS7-10)] and $HoxB^{Del/Del(i9-13):Ins(CBS5-10):Del(CBS8-10)}$ [*Bottom*, Del(CBS8-10)]. (*B*) Quantification of *Hoxb13* mRNAs in the two deletions shown in (*A*). While in Del(CBS8-10) *Hoxb13* is still tightly isolated, some transcripts start to appear in the Del(CBS7-10). The significance was assessed by the Wald test on counts with DESeq2 in comparison to the full cassette condition. The *P*-values were not corrected for multiple tests: ns: $P > 0.05$, *$P <= 0.05$, **$P <= 0.01$, and ***$P <= 0.001$. (*C*) WISH of *Hoxb13* transcripts in the two deletions shown under *A*. A weak signal appears at the right position in the Del(CBS7-10) gastruloids (black arrowhead), not detected in the Del(CBS8-10) counterparts. (Scale bar: 100 μm.)

function and delaying its implementation until the appropriate time has come (21). Despite its particular structure lacking most other posterior *Hox* genes, a strong and constitutive TAD boundary was maintained in *HoxB* (24) coinciding with both a large DNA spacer and the preservation of a number of CTCF sites, as in the three other *Hox* gene clusters (19).

A Physiological Gain of Function Causing a Heterochronic Phenotype. Removal of the DNA region between *Hoxb9* and *Hoxb13* induced premature anterior expression of *Hoxb13*, leading to vertebral column shortening, including vertebra loss from the lower thoracic, lumbar, and caudal regions. These defects were all rescued when the ectopic gained HOXB13 protein was inactivated in cis, demonstrating causality. These phenotypes were much weaker than those produced by using "transgenic" gain of function where neither the amount of ectopic protein, nor its precise timing of delivery can be precisely controlled (10, 38). In contrast, the extent of the morphological anomalies reported here is both dose dependent and in the range of what was previously reported using *on-site* genomic modifications for *Hox13* genes (e.g., refs. 9, and 39). Therefore, the dramatic developmental patterning defects observed upon ectopic expression of *Hox13* genes (6, 7, 38), while illustrating a genuine effect should be considered as an extreme outcome of overloading the system with HOX13 proteins, rather than as a physiological response to a slight variation in *Hox13* transcripts as it may naturally occur.

The deletion of the DNA spacer repositioned *Hoxb13* near *Hoxb9*, at the relative position where *Hoxb10*, absent from the amniote *HoxB* cluster, remains still in teleosts. As a consequence (40, 41), *Hoxb13* transcription was activated prematurely and its gain of expression was scored as early as E9 in the posterior growth zone. By late E10, signal intensity dropped significantly, resulting in a short window of *Hoxb13* exposure during somitogenesis. This was nevertheless enough to impact the morphology of the vertebral column at a level far more rostral than expected for *Hoxb13*, up to the ninth thoracic vertebra, reflecting an ectopic transitory expression in progenitors of somite 20 and more posterior. This thoracic 9th rostral limit of the $HoxB^{Del(i9\text{-}13)}$ phenotypes is of interest since in the absence of the *Hoxb1-Hoxb9* part of the cluster [$HoxB^{Del(1\text{-}9)}$], which left the *HoxB(i9-13)* spacer region untouched and thus likely preserved the insulation of *Hoxb13* from the influence of anterior enhancers, malformations were limited to thoracic vertebra 8 or more anterior (42). We conclude that the presence of the spacer region protects the lower thoracic, lumbar, and sacral midsegment of the main body axis from growth arrest by preventing premature *Hoxb13* expression in the posterior growth zone.

The phenotypes observed in the lumbar and the tail regions are notable, for L4 or L3 types of lumbar vertebral formulae had previously been reported only in very few stocks of laboratory mice, either in animals carrying the *Dominant hemimelia* (*Dh*) mutation (43), or in mice with premature expression of posterior *Hox* transgenes (44). Regarding tail length, the gain of function, rescue, and loss of function effects indicate that *Hoxb13* alone could account for at least one-third of length variation of the adult tail in laboratory mice. Premature presence of HOXB13 made tails shorter and the absence of the *Hoxb13* homeobox made tails longer, both in a gene dose-dependent manner. Combined gain of function mutations between *Hoxb13* and *Hoxd13* showed only a subtle increase in tail shortening despite recent results in deer mice where *Hoxd13* seems to be prominent (4).

Tail bud NMP cells display progenitor activity partly relying on *Hox13*, *Gdf11,* and *lin28* genes (6) and forced expression of *Lin28b* increased tail vertebra count by as much as five (7), a hypermorphosis that corresponds to the effect of *Hoxb13* inactivation. Conversely, both the hereby described *Hoxb13* gain of function and the *Lin28a* and *Lin28b* double knockouts reduced vertebra count by approximately four, suggesting a window of ca. 10 vertebrae where this system can operate. We did not detect any significant change in *Lin28a* or *Lin28b* transcript levels at E10.5 in our *Hoxb13* loss- or gain-of-function, which is consistent with *Hoxb13* being downstream a *Lin28/let-7* driven genetic regulatory circuit.

Transcriptome Analyses and Insulation by CTCF Sites. The gain of *Hoxb13* function had a moderate effect on global gene transcription, with only a handful of genes up- or downregulated. This was not unexpected when dealing with a heterochronic shift in an otherwise normal process, i.e., the termination of tail extension, a fortiori under physiological conditions. However, the downregulation of *Rbpj* is of interest, for transcripts accumulation in the tail of this effector of the *Notch* pathway is reminiscent of *Hoxb13* (*SI Appendix*, Fig. S3 *A–C*). Homozygous *Rbpj* loss of function mutants have severe arrest of early somitogenesis (28), which makes it possible that *Rbpj* mediates the effect of HOXB13 during vertebral column development, in both normal and ectopic contexts. Such potential interactions between *Hox* genes and the *Notch* pathway had been previously proposed (45).

Using gastruloids, we show that cells expressing ectopic *Hoxb13* mRNAs are a subset of those expressing *Hoxb*9 within the NMP population, further supporting the claim that *Hox* genes are expressed in time and space along the AP axis mostly in response to their relative position within the clusters rather than based on specific regulatory sequences (46). We also show that a small CTCF cassette reestablished most of the *Hoxb13* insulation, at least until the end of our gastruloid cultures (168 h) and hence that these sites are necessary and even sufficient in these conditions for tight insulation to occur. When the cassette was reduced to three or two CTCF sites, a leakage occurred suggesting that the reiteration of CBS in this spacer region was important. Recent work has shown that during gastruloid development, loop extrusion initially occurs asymmetrically upstream the *Hox* clusters (on the side of *Hox1*), as suggested by an enriched deposition of both RAD21 and NIPBL (19). In this view, these CBSs may represent as many potential blocks to loop extrusion, inducing transcription delays in an additive manner before reaching *Hoxb13*.

Two transcription units were nevertheless found directly 5′ of *Hoxb9* (*GM53* and *Mir196a-1*), which were also enriched in both cohesin and NIPBL thus adding 19 kb of chromatin potentially active in loop extrusion upstream *Hoxb9* and including both CBS5 and CBS6. Therefore, loop extrusion initiated from this part of the "spacer" DNA would be blocked or delayed mostly by CBS7 and CBS8. The orientation of these CBSs did not seem to be critical in their insulation capacity, either when the full cassette was inverted, or when only two occupied sites were left in either orientation, in agreement with previous work (20) and ref. therein.

If the insulation of *Hoxb13* mostly depends upon the presence of CBSs, then one may wonder why this large DNA spacer still exists in all vertebrates *HoxB* clusters examined thus far. The probability of reducing the length by spontaneous internal deletions not involving any CTCF site may be low, despite an enrichment in repeats (SINE, LINE, LTR) when compared to the rest of the cluster, which as all *Hox* clusters are generally depleted of such sequences (47). Another possibility is that other DNA sequences present in this spacer may play a role during development or adulthood related or unrelated to the "classical" function of *Hox* genes, which evolutionarily constrained this DNA interval. The presence of some DNA sequences conserved within all mammals may support this hypothesis, even though two of them precisely correspond to CBS7 and CBS8.

Materials and Methods

Animals. All experiments involving mice were approved and performed in compliance with the Swiss Law on Animal Protection (LPA) under license numbers GE 81/14 and VD2306.2 (to D.D.). All animals were kept in a continuous back cross with C57BL6 × CBA F1 hybrids. Mice were housed at the University of Geneva Sciences III animal colony with light cycles between 07:00 and 19:00 in the summer and 06:00 and 18:00 in winter, with ambient temperatures maintained between 22 and 23 °C and 45 and 55% humidity, the air was renewed 17 times per hour.

Generation of Alleles. We generated the *HoxB*$^{Del(i9\text{-}13)}$ allele by pronuclear injection of a mixture of two plasmids each containing the coding region for the Cas9 enzyme and a selected template for single guide RNA (*SI Appendix*, Table S2). The deficiency included 67,403 bp (chr11:96,197,567–96,264,470 mm10) in the *HoxB* cluster. Zygotes fertilized by Del(i9-13) homozygous mutant males and wt females were injected with a mix of two plasmids expressing specific guide RNAs corresponding to two nearby sequences flanking two-thirds of the homeodomain of *Hoxb13* (chr11:96,196,015 and chr11:96,196,172) (*SI Appendix*, Table S2). Animals were genotyped by PCR (*SI Appendix*, Table S3). In this way, we produced four different alleles where the *Hoxb13hd* mutant was isolated either *in cis* with *HoxB*$^{Del(i9\text{-}13)}$ [the *HoxB*$^{Del(i9\text{-}13):Hoxb13hd1}$ and *HoxB*$^{Del(i9\text{-}13):Hoxb13hd2}$ alleles], or in trans, i.e., on the wt chromosome (the *Hoxb13*^{hd1}and *Hoxb13*^{hd2} alleles). The sequences of both the various breakpoints and the mutated *Hoxb13* homeodomains are in *SI Appendix*, Tables S1 and S4, respectively. All four founders were crossed with wt breeders in three consecutive generations to segregate away potential CRISPRcas9 off-target events.

Other Mutant Mice. The mutant strains *HoxD*$^{Del(10\text{-}12)}$ and *Hoxd13*^{hd} were previously published (25, 26).

WISH. Whole mount in situ of mouse embryos was performed as described in ref. 48, using probes in *SI Appendix*, Table S5. Image acquisition was as in ref. 45.

Skeletal Preparations. Skeletal preparations were carried according to conventional methods (49). After Alizarin redS staining, the specimens were dehydrated into concentrated glycerol and photographed with a Nikon D810 camera, equipped with AF-S Micro NIKKOR 60 mm f/2.8G ED objective and Nikon SB-910 SPEEDLIGHT.

Micron-Scale Computed Tomography (μCT). Measurements of the vertebrae were carried out in the OsiriX MD v.10.0.1 software, in the 2D orthogonal MPR views of 3D reconstructions using WL/WW CT-Bone mode, displayed at 0.1 mm thick slab setting. Each measurement was taken when the object was approximately zoomed to five centimeters on the screen in one of the three windows, and the measure was manually defined using the length tool over the bone section. In all three windows, the longest head to tail length of all vertebrae was scrolled to, recorded, and the longest of the three values was used. Jpg images and movies were exported at preset modes.

Tail Length Measurements. For the tail length growth trajectories (*SI Appendix*, Fig. S2*A*), data were pooled from five litters of two separate stocks born from heterozygous parents. Eleven and nine homozygous individuals carrying either the *Hoxb13*^{hd1}or the *Hoxb13*^{hd2} alleles, respectively, and a total of 15 wt littermates were documented two to three times weekly, by taking direct measures of the length between the anus and the tip of the tail on a millimeter scale. The same measurement method was used as in Fig. 2*B*, but in Fig. 2*B*, each dot represents a different animal while in *SI Appendix*, Fig. S2*B* each dot is a measurement.

Production of Gastruloids. Gastruloids were produced exactly as described in ref. 19. Mouse embryonic stem (mES) cells were routinely cultured in gelatinized tissue-culture dishes with 2i (50). LIF DMEM medium composed of DMEM + GlutaMAX supplemented with 10% ES certified FBS, nonessential amino acids, sodium pyruvate, beta-mercaptoethanol, penicillin/streptomycin, 100 ng/mL of mouse LIF, 3 μM of GSK-3 inhibitor (CHIR99021) and 1 μM of MEK1/2 inhibitor (PD0325901). Cells were passaged every 3 d and maintained in a humidified incubator (5% CO2, 37 °C). The differentiation protocol for gastruloids was previously described (31). ES cells were collected after Accutase treatment, washed, and resuspended in prewarmed N2B27 medium (50% DMEM/F12 and 50% Neurobasal supplemented with 0.5× N2 and 0.5× B27 and with nonessential amino acids, sodium pyruvate, beta-mercaptoethanol, and penicillin/streptomycin). Then, 300 cells were seeded in 40 μL of N2B27 medium in each well of a low-attachment, rounded-bottom 96-well plate. 48 h after aggregation, 150 μL of N2B27 medium supplemented with 3 μM of GSK-3 inhibitor was added to each well. Then, 150 μL of medium was then replaced every 24 h.

Generation of Mutant ES Cells. As described in ref. 19, wt mES cells (EmbryoMax 129/SVEV) were used to generate different mutant lines following the CRISPR/Cas9 genome editing protocol described in ref. 51. sgRNA targeting guides (*SI Appendix*, Table S2) were cloned into a Cas9-T2A-Puromycin expressing plasmid containing the U6-gRNA scaffold (gift of A. Németh; Addgene plasmid, 101039). ES cells were transfected with 8 μg of plasmid using the Promega FuGENE 6 transfection kit and dissociated 48 h later for puromycin selection (1.5 μg mL^{-1}). Clone picking was conducted 5 to 6 d later, and positive ES cell clones were assessed by PCR screen using the MyTaq PCR mix kit (Meridian Bioscience) and specific primers surrounding the targeted region (*SI Appendix*, Table S3). Mutations were verified by Sanger sequencing (*SI Appendix*, Table S4). For the *HoxB*$^{Del(i9\text{-}13):Ins(CBS5\text{-}10)}$ the inserted CTCF cassette was synthesized (GeneArt, ThermoFisher). The sequence of the cassette was defined by selecting 500 bp around each of the six distinct CBS located in between *Hoxb9* and *Hoxb13*. In order to increase CRISPR guide efficiency, one base pair was added 10 bp after and before CBS9. The cassette was recombined between *Hoxb9* and *Mir196a-1* due to the lack of appropriate DNA sequences between *Mir196a-1* and *Hoxb13*.

WISH of Gastruloids. As described in ref. 19, gastruloids were collected at the indicated stage and processed following a previously reported WISH procedure (31). They were fixed overnight in 4% PFA at 4 °C and stored in methanol at −20 °C until ready for processing. Each sample was rehydrated and prepared with Proteinase K (EuroBio) at 2.5 μg/mL for 2 min. They were then incubated in a blocking solution at 68 °C for 4 h before incubation overnight with specific digoxigenin-labeled probes (*SI Appendix*, Table S5) at a final concentration of 100 to 200 ng/mL. The next day, samples were washed and incubated with an anti-DIG antibody coupled to alkaline phosphatase (Roche, 1:3,000). Staining was performed with BM-Purple (Roche). WISH of gastruloids was performed as in ref. 19.

Generation of In Silico Mutant Files. A fasta file containing the sequence of chromosome 11 including HoxBDel(i9-13):Ins(CBS5-10) and one with the sequence of chromosome 11 including *HoxB*^{Del} (renamed chr11_delB) were generated using the R package seqinr (52) with the sequence of chromosome 11 from mm10 from UCSC as a template. Both mutant chromosome 11 sequences available at https://doi.org/10.5281/zenodo.12723266 were concatenated with the sequences of all other autosomes, chr X, chr Y, and mitochondrial DNA from mm10 (UCSC).

CTCF and RAD21 ChIPmentation. ChIPmentation of CTCF and RAD21 was performed as described in ref. 19. Fastqs of the RAD21 ChIPmentation of 120 h wt gastruloids from ref. 19 were retrieved from SRA (SRR19601466) and processed with datasets generated in this study. For data analysis, adapter and bad quality bases were removed from fastq files using cutadapt version 4.1 (53) (-a CTGTCTCTTATACACATCTCCGAGCCCACGAGAC -A CTGTCTCTTATACACATCTGACGCTGCCGACGA -q 30 -m 15). Filtered reads were aligned on the in silico mutant genome using bowtie2 version 2.4.5 with default parameters (53). No filtering for mutlimap reads was performed, for sequences that are duplicated between both alleles of chromosome 11. Coverage was computed with macs2 version 2.2.7.1 (54) (--call-summits --format BAMPE -B) removing PCR duplicates and then normalized to the million fragments after filtering used by macs2.

CUT&RUN. Each ESC per genotype was washed into fresh PBS before dissociation. The dissociation was performed in 0.5 mL StemPro Accutase (Gibco A1110501), 5 min at 37 °C. Cells were resuspended into 2 mL of ESC media and span at 300 g for 2 min. Then, 0.5 × 10^6 cells were processed according to the CUT&RUN protocol (55). First, cells were resuspended into 1 mL of Wash Buffer [WB: 20 mM HEPES-KOH pH7.5, 150 mM NaCl, and 0.5 mM Spermidin (Sigma S2626)] and bound to Concanavalin A–coated beads (BioMagPlus 86057) into Digitonin Wash Buffer (DWB) using a final concentration of 0.02% digitonin (Apollo APOBID3301)

into WB. Second, cells were incubated with 0.5 μg/100 μL of anti-CTCF antibody (Active Motif 61311) in DWB at 4 °C for 2 h. The pA-MNase was produced by the Protein Production and Structure Core Facility at Ecole Polytechnique Fédérale de Lausanne (EPFL) and added at 0.5 μL/100 μL in Digitonin Wash Buffer for 1 h at 4 °C. Third, cells were digested in Low Calcium Buffer (100 mM CaCl2 into DWB). The reaction was stopped using cold 2× STOP Buffer [340 mM NaCl, 4 mM EGTA, 20 mM EDTA, 0.02% Digitonin, 50 μg RNase A (10 mg/mL), and 50 μG Glycogen (5 mg/mL)] for 5 min. Finally, targeted chromatin was released for 30 min at 37 °C, then precipitated into 70% EtOH and stored at −20 °C until sequencing libraries generation. Sequencing libraries were prepared with KAPA HyperPrep reagents (07962347001) with 2.5 μL of adapters at 0.3 μM and ligated for 1 h at 20 °C. Then, DNA was cleaned and size selected using 1:1 ratio of DNA:Ampure SPRI beads (A63881) followed by an additional 1:1 wash and size selection with HXB. The DNA was amplified for 16 cycles. Postamplified DNA was cleaned and size selected using 1:1 ratio followed by an additional 1:1 wash and size selection with HXB. HXB is equal parts 40% PEG8000 (Fisher FIBBP233) and 5 M NaCl. Libraries were sequenced on NextSeq 500 or NovaSeq 6000. For data analysis, adapter and bad quality bases were removed from fastq files using cutadapt version 4.1 (53) (-a GATCGGAAGAGCACACGTCTGAACTCCAGTCAC -A GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-q 30 -m 15). Filtered reads were aligned on the in silico mutant genome using bowtie2 version 2.4.5 (56) with adapted parameters (--very-sensitive --no-unal --no-mixed --no-discordant --dovetail -X 1000). No filtering for mutlimap reads was performed as sequences are duplicated between both alleles of chromosome 11. PCR duplicates were removed with Picard MarkDuplicates version 2.27.4 (57). Filtered BAM file was converted to BED with BEDTools version 2.30.0 (58). Coverage was computed with macs2 version 2.2.7.1 (54) (--nomodel --keep-dup all --shift -100 --extsize 200 --call-summits -B) and then normalized to the million reads in peaks.

HiChIP. The dataset was produced and extracted from ref. 19, GSM6226299.

RNA-seq. Murine E9.5 and E10.5 mutant or wt littermate embryos were dissected to remove the head. RNA was extracted using the RNeasy plus micro kit (Qiagen) following the manufacturer's recommendation. The E9.5 sample libraries are poly(A)-enriched, and E10.5 are ribo-depleted. Both were prepared with stranded RNA TruSeq kits (Illumina) and sequenced 100 bp single-read on a HiSeq 4000. 120 h and 144 h AA gastruloids were collected and RNA was extracted as in ref. 31. Libraries were prepared with ISML kit from Illumina, except for two replicates of $HoxB^{Del/Del(i9\text{-}13):Ins(CBS5\text{-}10):Inv(CBS5\text{-}10)}$ clone 1 at 120 h, which were prepared with the TruSeq Stranded mRNA kit from Illumina. Libraries were sequenced on a NextSeq 500 or NovaSeq 6000 or HiSeq 4000, paired-end 2 times 75 bp. Fastqs of the two replicates of 120 h and 144 h wt gastruloids from ref. 19 were retrieved from SRA (SRR19600485 and SRR19600486, SRR19600479 and SRR19600480) and processed with datasets generated in this study. For data analysis, adapter and bad quality bases were removed from fastq files using cutadapt version 1.18 (53) (-q 30 -m 15 and sequences for -a and -A were adapted to the library). Filtered reads were aligned on mm10 using STAR version 2.7.10a (59) with the ENCODE parameters and a custom gtf (https://doi.org/10.5281/zenodo.7510406). FPKM were computed with cufflinks version 2.2.1 (60, 61) using --max-bundle-length 10000000 --multi-read-correct --library-type "fr-firststrand" -b mm10.fa --no-effective-length-correction -M mm10_chrM.gtf. Counts from STAR and FPKM values were subsetted for protein-coding genes and genes on chrM were excluded. For the embryo datasets, genes on chromosomes X and Y were also excluded. Differential expression analysis was performed with DESeq2 version 1.38.0 (62) on R version 4.2.2. PCA was computed using log2(1 + FPKM) using the 500 most variant genes.

scRNA-seq. scRNA-seq of gastruloids at 144 h was performed as in ref. 63, with 10× Chemistry v3 as well as the analysis up to matrix generation. Then, analysis was performed using Seurat v4.3 (64) with R version 4.3.0. We first filtered out barcodes with less than 200 identified gene and genes identified in less than three cells. Low quality cells and potential doublets were removed by computing the mean UMI content and the percentage of mitochondrial genes and filtering out barcodes with less than 0.4 times the mean UMI or more than 2.5 times the mean UMI. Only barcodes between 0.05 and 8 percent of mitochondrial UMI were kept. Matrices were normalized, and the cell cycle score (using the 2019 updated gene list from Seurat) from these filtered libraries was computed. Then we merged the different samples using the merge command by Seurat. The combined Seurat object was normalized, 2,000 variable features were identified, and the data were scaled and regressed by cell cycle score and percentage of mitochondrial reads. Principal components were then computed using variable genes falling within the 5th and 80th percentile of expression to limit batch effect as performed in ref. 65. UMAP and k-nearest neighbors were computed using 30 principal components. Clustering was performed with a resolution 0.4 and cluster annotation was performed manually. The average expression of a gene in a cluster was computed as the sum of the raw counts from all cells divided by the sum of the total counts of all cells. This value was log transformed like with a pseudocount of 1 and a scale of 10^4. The inferred expression distribution of *Hoxb13* in mutant NMP and neural tube cells was obtained with baredSC version 1.1.2 (33) (--xmax 3 --minNeff 200 --minScale 0.1 for 1 to 4 Gaussians combined), for the joint inferred expression distribution of *Hoxb13* with *Hoxa9*, *Hoxb9*, *Hoxc9,* or *Hoxd* genes, similar parameters were used (--xmax 3 --ymax 3 --minNeff 200 --minScalex 0.1 --minScaley 0.1 for 1 to 4 Gaussians combined). Convergence of each MCMC was manually inspected using corner plots and ACF. Plots were realized with ggpubr (66).

minION Sequencing. nCATS was performed as in ref. 67 with a pool of 4 guides and a pool of 8 guides (*SI Appendix*, Table S2). Base calling was computed with Guppy version 5.0.16+b9fcd7b from Oxford Nanopore (--flowcell FLO-MIN106 --kit SQK-LSK109 --fast5_out). Mapping was performed on mm10 or on the in silico mutant genome using minimap2 version 2.28 (68) (-ax splice). Nonprimary alignments were removed with samtools version 1.16.1 (69) and coverage was generated by BEDTools version 2.30.0 (58).

Gene Distance Analysis. Gene distances were extracted from Ensembl Release 111. For each species of the database, genes with external gene names matching *Hoxb1*, *Hoxb9*, *Hoxb13,* or derivatives were extracted, as well as all genes registered as homologous to the three mouse genes. Only couple of genes which were on the same chromosome were considered and the distance corresponds to the intergenic distance.

Sequence Conservation. Sequence conservation between the mouse and other vertebrate *HoxB* loci was visualized using the Vertebrate Multiz Alignment & Conservation maf file provided by UCSC for mm10. The Repeat Mask was also provided by UCSC. The sequence conservation between *mus musculus* and other strains was visualized using the 21-way Enredo-Pecan-Ortheus multiple alignments provided by Ensembl. This sequence conservation was computed on mm39. To evaluate distance conservation between the various mouse strains, gene annotations were downloaded from Ensembl version 102 (mm10) and centered on *Hoxb9*.

Data Analysis. All NGS analyses were computed using the facilities of the Scientific IT and Application Support Center of EPFL. The genomic tracks were displayed using pyGenomeTracks version 3.9 (70, 71). The quantifications were plotted with R version 4.4.0 (https://www.R-project.org/) and ggpubr version 0.6.0 (66).

Data, Materials, and Software Availability. All scripts are available at https://github.com/lldelisle/scriptsForLopezDelisleZakanyBochatonEtAl2024 (72). All raw and processed datasets are available in the Gene Expression Omnibus repository under accession number GSE272483 (73).

ACKNOWLEDGMENTS. We thank all the colleagues from the Duboule laboratories for discussions, Dr. Christopher C. Bolt for his help, and Anne-Catherine Cossy for technical assistance. This work was supported in part using the resources and services for sequencing of the Gene Expression Research Core Facility at the School of Life Sciences of EPFL and the Geneva iGE3 Genomics Platform (University of Geneva). In addition, all computational work was performed using the facilities of the Scientific IT and Application Support Center of EPFL. This work was supported by funds from the EPFL (Lausanne), the European Research Council grant RegulHox (588029), and the Swiss National Science Foundation (CRSII5_189956 and 310030B_138662).

Author affiliations: [a]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland; [b]Department of Genetics and Evolution, University of Geneva, Geneva 1211, Switzerland; and [c]Center for Interdisciplinary Research in Biology, Collège de France, CNRS UMR 7241, INSERM U1050, Université Paris Sciences et Lettres, Paris 75231, France

1. H. Fol, Sur la queue de l'embryon humain. *Comptes Rendus Acad. Sci.*, 1469–1472 (1885).
2. A. Kispert, B. G. Herrmann, The Brachyury gene encodes a novel DNA binding protein. *EMBO J.* **12**, 3211–3220 (1993).
3. M. Mallo, The vertebrate tail: A gene playground for evolution. *Cell. Mol. Life Sci.* **77**, 1021–1030 (2020).
4. E. P. Kingsley *et al.*, Adaptive tail-length evolution in deer mice is associated with differential Hoxd13 expression in early development. *Nat. Ecol. Evol.* **8**, 791–805 (2024).
5. B. Xia *et al.*, On the genetic basis of tail-loss evolution in humans and apes. *Nature* **626**, 1042–1048 (2024).
6. R. Aires *et al.*, Tail bud progenitor activity relies on a network comprising Gdf11, Lin28, and Hox13 genes. *Dev. Cell* **48**, 383–395.e8 (2019).
7. D. A. Robinton *et al.*, The Lin28/let-7 pathway regulates the mammalian caudal body axis elongation program. *Dev. Cell* **48**, 396–405.e3 (2019).
8. L. Zeltser, C. Desplan, N. Heintz, Hoxb-13: A new Hox gene in a distant region of the HOXB cluster maintains colinearity. *Development* **122**, 2475–2484 (1996).
9. K. D. Economides, L. Zeltser, M. R. Capecchi, Hoxb13 mutations cause overgrowth of caudal spinal cord and tail vertebrae. *Dev. Biol.* **256**, 317–330 (2003).
10. T. Young *et al.*, Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev. Cell* **17**, 516–526 (2009).
11. N. Di-Poi *et al.*, Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature* **464**, 99–103 (2010).
12. C. Gomez *et al.*, Control of segment number in vertebrate embryos. *Nature* **454**, 335–339 (2008).
13. D. Duboule, G. Morata, Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.* **10**, 358–364 (1994).
14. T. Li *et al.*, Whole-genome scanning for selection signatures reveals candidate genes associated with growth and tail length in sheep. *Animals* **14**, 687 (2024).
15. D. K. Lagler *et al.*, Fine-mapping and identification of candidate causal genes for tail length in the Merinolandschaf breed. *Commun. Biol.* **5**, 918 (2022).
16. E. P. Nora *et al.*, Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
17. J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
18. A. R. Amândio *et al.*, Sequential in cis mutagenesis in vivo reveals various functions for CTCF sites at the mouse HoxD cluster. *Genes Dev.* **35**, 1490–1509 (2021), 10.1101/gad.348934.121.
19. H. Rekaik *et al.*, Sequential and directional insulation by conserved CTCF sites underlies the Hox timer in stembryos. *Nat. Genet.* **55**, 1164–1175 (2023).
20. C. Anania *et al.*, In vivo dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation. *Nat. Genet.* **54**, 1026–1036 (2022).
21. H. Rekaik, D. Duboule, A CTCF-dependent mechanism underlies the Hox timer: Relation to a segmented body plan. *Curr. Opin. Genet. Dev.* **85**, 102160 (2024).
22. E. Boncinelli *et al.*, Organization of human homeobox genes. *Hum. Reprod.* **3**, 880–886 (1988).
23. A. Graham, N. Papalopulu, R. Krumlauf, The murine and Drosophila homeobox gene complexes have common features of organization and expression. *Cell* **57**, 367–378 (1989).
24. B. Bonev *et al.*, Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017).
25. B. Tarchini, D. Duboule, Control of Hoxd genes' collinearity during early limb development. *Dev. Cell* **10**, 93–103 (2006).
26. F. Darbellay *et al.*, The constrained architecture of mammalian Hox gene clusters. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 13424–13433 (2019).
27. S. F. L. Wong *et al.*, Independent regulation of vertebral number and vertebral identity by microRNA-196 paralogs. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4884–E4893 (2015).
28. I. D. B. Barrantes *et al.*, Interaction between Notch signalling and Lunatic fringe during somite boundary formation in the mouse. *Curr. Biol.* **9**, 470–480 (1999).
29. D. A. Turner *et al.*, Anteroposterior polarity and elongation in the absence of extraembryonic tissues and spatially localised signalling in Gastruloids, mammalian embryonic organoids. *Development* **144**, 3894–3906 (2017).
30. S. C. van den Brink *et al.*, Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* **141**, 4231–4242 (2014).
31. L. Beccari *et al.*, Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* **562**, 272–276 (2018).
32. J. V. Veenvliet *et al.*, Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science* **370**, eaba4937 (2020).
33. L. Lopez-Delisle, J.-B. Delisle, baredSC: Bayesian approach to retrieve expression distribution of single-cell data. *BMC Bioinformatics* **23**, 36 (2022).
34. D. Duboule, The (unusual) heuristic value of Hox gene clusters; a matter of time? *Dev. Biol.* **484**, 75–87 (2022).
35. R. Neijts *et al.*, Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos. *Genes Dev.* **30**, 1937–1942 (2016).
36. T. Gilpatrick *et al.*, Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* **38**, 433–438 (2020).
37. P. Kalds, S. Huang, Y. Chen, X. Wang, Ovine HOXB13: Expanding the gene repertoire of sheep tail patterning and implications in genetic improvement. *Commun. Biol.* **5**, 1196 (2022).
38. N. Denans, T. Iimura, O. Pourquie, Hox genes control vertebrate body elongation by collinear Wnt repression. *Elife* **4**, e04379 (2015).
39. P. Tschopp, B. Tarchini, F. Spitz, J. Zakany, D. Duboule, Uncoupling time and space in the collinear regulation of Hox genes. *PLoS Genet.* **5**, e1000398 (2009).
40. D. Duboule, Temporal colinearity and the phylotypic progression: A basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development* **1994**, 135–142 (1994).
41. A. Durston, S. Wacker, N. Bardine, H. Jansen, Time space translation: A hox mechanism for vertebrate a-p patterning. *Curr. Genomics* **13**, 300–307 (2012).
42. O. Medina-Martinez, A. Bradley, R. Ramirez-Solis, A large targeted deletion of Hoxb1-Hoxb9 produces a series of single-segment anterior homeotic transformations. *Dev. Biol.* **222**, 71–83 (2000).
43. J. Suto, T. Wakayama, K. Imamura, S. Goto, K. Fukuta, Skeletal malformations caused by the Dh(Dominant hemimelia) gene in mice. *Exp. Anim.* **45**, 95–98 (1996).
44. J. Zakany, M. Gerard, B. Favier, S. S. Potter, D. Duboule, Functional equivalence and rescue among group 11 Hox gene products in vertebral patterning. *Dev. Biol.* **176**, 325–328 (1996).
45. J. Zakany, M. Kmita, P. Alarcon, J. L. de la Pompa, D. Duboule, Localized and transient transcription of Hox genes suggests a link between patterning and the segmentation clock. *Cell* **106**, 207–217 (2001).
46. F. van der Hoeven, J. Zákány, D. Duboule, Gene transpositions in the HoxD complex reveal a hierarchy of regulatory controls. *Cell* **85**, 1025–1035 (1996).
47. N. Di-Poï, J. Montoya-Burgos, D. Duboule, Atypical relaxation of structural constraints in Hox gene clusters of the green anole lizard. *Genome Res.* **19**, 602–610 (2009), 10.1101/gr.087932.108.
48. J. M. Woltering *et al.*, Axial patterning in snakes and caecilians: Evidence for an alternative interpretation of the Hox code. *Dev. Biol.* **332**, 82–89 (2009).
49. R. Behringer, *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, ed. 4, 2014).
50. J. Silva *et al.*, Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6**, e253 (2008).
51. G. Andrey, M. Spielmann, CRISPR/Cas9 genome editing in embryonic stem cells. *Methods Mol. Biol.* **1468**, 221–234 (2017).
52. M. Gouy, F. Milleret, C. Mugnier, M. Jacobzone, C. Gautier, ACNUC: A nucleic acid sequence data base and analysis system. *Nucl. Acids Res.* **12**, 121–127 (1984).
53. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* **17** (2011).
54. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
55. P. J. Skene, J. G. Henikoff, S. Henikoff, Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).
56. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
57. Broad Institute. Picard Tools, Broad Institute, GitHub, version 2.27.4. http://broadinstitute.github.io/picard/. Accessed 25 October 2024.
58. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
59. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
60. A. Roberts, C. Trapnell, J. Donaghey, J. L. Rinn, L. Pachter, Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* **12**, R22 (2011).
61. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
62. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
63. A. Mayran *et al.*, Cadherins modulate the self-organizing potential of gastruloids. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.11.22.568291 (Accessed 27 June 2024).
64. Y. Hao *et al.*, Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2023), 10.1038/s41587-023-01767-y.
65. C. Aztekin *et al.*, Identification of a regeneration-organizing cell in the *Xenopus* tail. *Science* **364**, 653–658 (2019).
66. A. Kassambara, ggpubr: 'ggplot2' Based Publication Ready Plots. R package, version 0.6.0. https://rpkgs.datanovia.com/ggpubr/. Accessed 25 October 2024.
67. C. C. Bolt *et al.*, Context-dependent enhancer function revealed by targeted inter-TAD relocation. *Nat. Commun.* **13**, 3488 (2022).
68. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
69. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
70. F. Ramirez *et al.*, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
71. L. Lopez-Delisle *et al.*, pyGenomeTracks: Reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).
72. L. Delisle. lldelisle/scriptsForLopezDelisleZakanyBochatonEtAl2024: Published version (Version v202409). Zenodo. https://doi.org/10.5281/zenodo.13990512. Deposited 24 October 2024.
73. C. Bochaton *et al.*, Data from CTCF-dependent insulation of Hoxb13 and the heterochronic control of tail length. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE272483. Deposited 18 July 2024.