Article

Co-option of an ancestral cloacal regulatory landscape during digit evolution

https://doi.org/10.1038/s41586-025-09548-0

Received: 13 March 2024

Accepted: 19 August 2025

Published online: 17 September 2025

Open access

Check for updates

Aurélie Hintermann[1,8,10], Christopher C. Bolt[2,9,10], M. Brent Hawkins[3,4,10], Guillaume Valentin[2], Lucille Lopez-Delisle[2], Madeline M. Ryan[3,4], Sandra Gitto[1], Paula Barrera Gómez[2], Bénédicte Mascrez[1], Thomas A. Mansour[5], Tetsuya Nakamura[5], Matthew P. Harris[3,4], Neil H. Shubin[6] & Denis Duboule[1,2,7]✉

The fin-to-limb transition in vertebrate evolution has been central to the study of how development underlies evolutionary change. In this context, the functional analysis of *Hox* gene regulation to infer evolutionary trajectories has been critical to explain the origin of new features. In tetrapods, the transcription of *Hoxd* genes in developing digits depends on a set of enhancers forming a large regulatory landscape[1,2]. The presence of a syntenic counterpart in zebrafish, which lacks digits, suggests deep homology[3] or shared developmental foundations underlying distal fin and limbs. However, how this regulatory program evolved has remained unresolved. We genetically evaluated the function of the zebrafish *Hoxd* regulatory landscapes by comparatively assessing the effects of their full deletions. We show that, unlike in mice, deletion of these regions in fish does not disrupt *hoxd* gene transcription during distal fin development. By contrast, we found that this deficiency leads to the loss of expression within the cloaca, a structure related by ancestry to the mammalian urogenital sinus, and that distal *hox13* genes are essential for correct cloacal formation. Because *Hoxd* gene regulation in the mouse urogenital sinus relies on enhancers located in this same chromatin domain that controls digit development, we propose that the current regulatory landscape active in distal limbs was co-opted as a whole in tetrapods from a pre-existing cloacal regulatory machinery.

The skeletal pattern of tetrapod limbs has been conserved since the Devonian Period, with a universal architecture of segments along the proximal-to-distal axis. The stylopod is a single bone (upper arm or thigh) attached at one end to the torso and at the other end to two zeugopod bones (forearm or leg), and the most distal are the mesopod (wrist and ankle) and autopod (hand and foot). The formation of this generic pattern began before the water-to-land transition, because sarcopterygian fishes display structures clearly related to proximal tetrapod limb structures[4]. However, when homologies are considered between fin structures and the most distal parts of tetrapod appendages (mesopod and autopod), the extent to which fishes possess homologous skeletal elements remains debatable. Although mesopodial elements and extensive distal segments are present in sarcopterygian fins, the presence of true digital homologues remains controversial[5,6].

Because the *HoxA* and *HoxD* gene clusters have been shown to be instrumental in making tetrapod limbs[7–10], their expression domains during fin development were used to infer the presence of an autopod-related structure in fishes. In particular, *Hoxa13* and *Hoxd13* have been studied because of their specific autopodial expression in tetrapod limbs[10] and because their combined inactivation in mice produces autopodial agenesis[7]. An analysis of *hox13* genes in the teleost fin suggested that a 'distal program' also exists in fishes, because their combined inactivation leads to the loss of distal structures, implying that a genetic regulatory network, or part thereof, would have preceded digit formation in tetrapods[11,12]. In such a scenario, the autopods of tetrapods are proposed to form from the postaxial vestiges of an ancestral sarcopterygian fin[13,14]. The partial retention of expression patterns that presage the emergence of digits in ray-finned and chondrichthyan fishes is nevertheless suggestive of a common regulatory program shared among vertebrates, the deployment of which, in different species, accompanied changes in form[13].

During tetrapod limb bud development, a series of enhancers within a large regulatory landscape positioned 3′ of the *HoxD* gene cluster (3DOM) control the transcription of *Hoxd* genes up to *Hoxd11* in a proximal expression domain. These expression domains encompass the tissue of the future stylopod (upper arm) and zeugopod (forearm) (Extended Data Fig. 1a (green and schemes on the left))[15]. Posterior–distal limb bud cells then switch off these enhancers and activate another large regulatory landscape (5DOM), located on the other side of (5′ to) the gene cluster. This region is enriched in conserved enhancer elements that have been found to control the formation of digits by

[1]Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland. [2]School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [3]Department of Genetics, Harvard Medical School, Boston, MA, USA. [4]Department of Orthopedic Research, Boston Children's Hospital, Boston, MA, USA. [5]Department of Genetics, Rutgers University, New Brunswick, NJ, USA. [6]Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL, USA. [7]Center for Interdisciplinary Research in Biology, Collège de France, CNRS, INSERM, Université PSL, Paris, France. [8]Present address: Stowers Institute for Medical Research, Kansas City, MO, USA. [9]Present address: Tessera Therapeutics, Somerville, MA, USA. [10]These authors contributed equally: Aurélie Hintermann, Christopher C. Bolt, M. Brent Hawkins. ✉e-mail: Denis.duboule@college-de-france.fr

activating *Hoxd13* and its closest neighbours (Extended Data Fig. 1a (blue)). Supporting this model, the deletion of 3DOM abrogated the expression of all *Hoxd* genes in the proximal limb domain[15], whereas the deletion of 5DOM removed all *Hoxd* messenger RNAs (mRNAs) from the forming autopod[2].

In teleosts, a further whole-genome duplication (TWGD) occurred, leading to the presence of two *hoxd* cluster orthologues. However, the *hoxdb* cluster was lost, resulting in a 1:1 orthology between the single remaining zebrafish *hoxda* cluster and the tetrapod *HoxD* cluster. The *hoxda* genes are also expressed during early fin bud development, with progressively nested expression domains comparable with the amniote situation[13,16]. At a later stage, transcription of both *hoxd9a* and *hoxd10a* persists in the 'preaxial' (anterior) part of the fin bud only (Extended Data Fig. 1b (magenta)), whereas *hoxd11a*, *hoxd12a* and *hoxd13a* transcripts are restricted to 'postaxial' (posterior) cells (Extended Data Fig. 1b (orange)), as is the case in the emerging fin bud[16]. For the latter genes, combined inactivation has revealed their function during distal fin skeletal development[11,17]. However, despite the identification of DNA elements orthologous to mouse distal enhancers[12], the functionality of the 3DOM and 5DOM regulatory landscapes in zebrafish has not yet been addressed. Owing to this gap in knowledge, the existence of comparable bimodal regulation of *Hoxd* gene expression has remained controversial.

To address these issues, we deleted both zebrafish *hoxda* regulatory landscapes to assess the extent to which appendage regulation by the orthologous 3DOM and 5DOM regions was conserved between zebrafish and mice. Embedding these results within a phylogenetic and genomic framework, we propose a new hypothesis of co-option for the origin of the distal regulation of appendages and genitalia in vertebrates.

Zebrafish *hoxda* locus

The zebrafish *hoxda* locus shares a high degree of synteny with the *HoxD* locus in mammals, reflecting broad conservation, given the key patterning role of this complex in the development of many axial structures. The gene cluster is flanked by two gene deserts: 3DOM (3′-located domain) and 5DOM (5′-located domain). As in mammals, the extent of both 3DOM and 5DOM corresponds to topologically associating domains (TADs), and 3DOM is split into two sub-TADs (Extended Data Fig. 2). This remarkable similarity in three-dimensional conformations, although with a 2.6-fold difference in size between the mouse versus zebrafish loci, is further supported by the conserved position and orientation of critical CTCF binding sites within the gene clusters and their enrichment at TAD and sub-TAD borders (Extended Data Fig. 2).

Interspecies genomic alignments revealed conserved sequences within 5DOM across vertebrates, whereas little conservation was scored in 3DOM (Extended Data Fig. 3a,b). Within the 5DOM comparison, we identified several previously annotated mouse enhancers in zebrafish[12,18] (see "Enhancers in the fish 5DOM"). Consistent with the apparent conservation of chromatin structure, we found the same global organization of both coding and non-coding elements as in the mouse landscape. When compared with the size of the *Hox* cluster, the relative sizes of both the 3′ and 5′ gene deserts were larger in mouse than in zebrafish, and the zebrafish 5DOM was found to be larger than 3DOM, in contrast to the mouse situation (Extended Data Fig. 3c). Because the overall genomic organization of both *HoxD* loci is well conserved between mammals and fishes, we concluded that these two flanking gene deserts and their TADs are ancestral features predating the divergence between ray-finned fishes and tetrapods, probably conserved because of important regulatory functions. Whether these domains have or retain *Hox* gene regulation, as initially defined at this locus in the mouse[1,19], remains unclear.

Zebrafish *hoxda* regulatory landscapes

To address the potential function(s) of zebrafish *hoxd* gene deserts, we explored the histone modification profiles for H3K27 acetylation and H3K27 trimethylation using the cleavage under targets and release using nuclease (CUT&RUN)[20] assay with the posterior trunk as a source of cells, that is, a domain where most *hox* genes are active. As a control sample, we used the corresponding dissected heads where *hoxda* genes were not expressed (Extended Data Fig. 4a,b). H3K27ac-positive marks were enriched over 3DOM, which was devoid of negative H3K27me3 marks (Extended Data Fig. 4c (green and magenta, respectively)), suggesting the strong involvement of 3DOM in transcriptional regulation, as seen within the *hoxda* cluster itself. By contrast, H3K27me3 marks were enriched over 5DOM. This analysis revealed that both zebrafish gene deserts might indeed serve as regulatory landscapes, with distributions of histone marks comparable with tetrapods in the same developmental context.

To assess the functional potential of both *hoxda* gene deserts, we generated zebrafish mutant lines carrying full deletions of either 5DOM ($hoxda^{del(5DOM)}$, referred to as *Del(5DOM)*) or 3DOM ($hoxda^{del(3DOM)}$ or *Del(3DOM)*), using CRISPR–Cas9 chromosome editing. We first examined the impact of these large deletions on *hoxd13a, hoxd10a* and *hoxd4a* expression using whole-mount in situ hybridization (WISH), spanning from 36 h post-fertilization (hpf), that is, from the onset of *hoxd13a* expression[16], to 72 hpf. In *Del(3DOM)* mutant embryos, the expression of both *hoxd4a* and *hoxd10a* completely disappeared from the pectoral fin buds (Fig. 1a (right and middle panels, arrowheads)). The same effect was observed at all stages analysed (Fig. 1a). These data are consistent with those of a similar analysis in mice, in which the limb proximal expression domain was no longer visible upon deletion of the 3DOM landscape[15]. This demonstrates that, similar to tetrapods, enhancers controlling the transcription of *hoxd3a* to *hoxd10a* during fin bud development are located in the adjacent 3′ landscape. Thus, 3DOM has an ancestral regulatory function in the development of proximal paired appendages. Expression of *hoxd13a* in postaxial cells, however, remained unchanged, with a global transcript distribution indistinguishable from that of wild-type fin buds (Fig. 1a (left panels, arrowheads)). These data indicate that the control of *hoxd13a* expression is distinct from that impacting *hoxd3a* to *hoxd10a*, as is also the case for tetrapods[2] (Extended Data Fig. 1a).

To determine whether *hoxd13a* transcription was controlled by enhancers present within 5DOM, we similarly analysed *Del(5DOM)* zebrafish embryos using WISH. Consistent with the regional control of *Hox* gene transcription, neither *hoxd4a* nor *hoxd10a* expression was affected in the mutant *Del(5DOM)* fin buds (Fig. 1b (arrowheads)). Up to 48 hpf, *hoxd13a* transcripts were virtually unaffected, with a pattern closely matching that of the control fin buds (Fig. 1b (arrowheads) and Supplementary Fig. 1). However, at later stages of development, the 5DOM deletion led to variably attenuated patterns of *hoxd13a* expression, whereas its overall mRNA distribution remained globally similar to that of the wild type. We confirmed this diminished expression in late-stage mutant fins using the hybridization chain reaction (HCR), but this trend was not significant (Extended Data Fig. 5). These results indicate that although some regulatory activity may exist within the 5DOM in fishes, the most significant regulatory inputs for *hoxd13a* are located outside the domain (Fig. 1c,d and Extended Data Fig. 5). This situation differs from that in mice, in which a similar deletion abrogates *hoxd13* expression entirely.

These two genomic regions also control expression in other axial systems in mice[21,22]. Thus, we extended our analysis to assess shared components of regulation between these regulatory landscapes. Mutant *Del(3DOM)* embryos did not reveal visible differences in expression in the trunk (*hoxd13a*, *hoxd10a* and *hoxd4a*), the pseudo-cloacal region (*hoxd13a*) or the branchial arches and rhombomeres (*hoxd4a*) (Extended Data Fig. 6a). *Del(5DOM)* embryos also showed comparable expression to wild-type controls, except for the complete disappearance of *hoxd13a* transcripts from the pseudo-cloacal region (Fig. 2). We noticed a temporary reduction in *hoxd13a* expression in the tailbud (Fig. 2a); yet, this deficit was no longer detectable at 36 hpf.

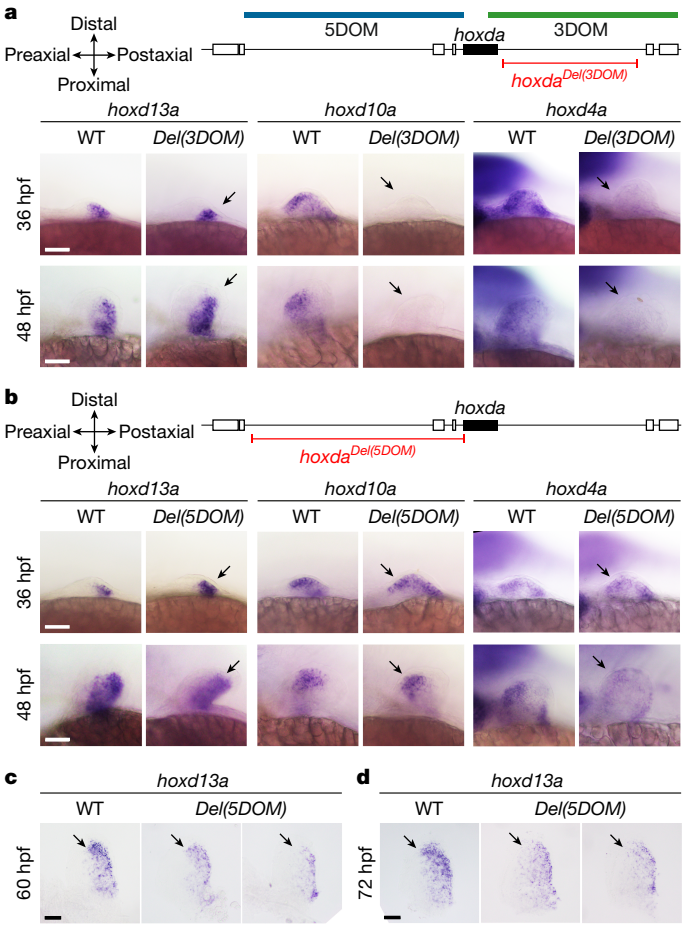


Fig. 1 | Regulation of *hoxda* genes in pectoral fins lacking the 3DOM and 5DOM regulatory landscapes. a–**d**, Expression of *hoxd13a*, *hoxd10a* and *hoxd4a* using WISH at 36 hpf, 48 hpf, 60 hpf and 72 hpf in zebrafish embryos with either the 3DOM (**a**) or 5DOM (**b**–**d**) regulatory landscape deleted. Wild-type and homozygous mutant embryos derived from the same cross are shown side by side. WISH was performed on at least five embryos for each combination of probes and genotypes, with four or more technical replicates, except for the *Hoxd10* probe, which had only one technical replicate. **a**, Expression of both *hoxd10a* and *hoxd4a* is completely lost in mutant fin buds lacking 3DOM (arrows), whereas expression of *hoxd13a* is identical to that of wild-type embryos (arrows). **b**, In fin buds lacking 5DOM, expression of *hoxd13a*, *hoxd10a* and *hoxd4a* is identical to matched wild-type embryos up to 48 hpf (arrows). **c**,**d**, However, at 60 hpf (**c**) and 72 hpf (**d**), a decrease in intensity was observed throughout, particularly marked in the distal aspect of the fin bud (arrows). The degree of this attenuation is variable among mutant fish (Supplementary Fig. 1). WT, wild type. Scale bars, 50 µm.

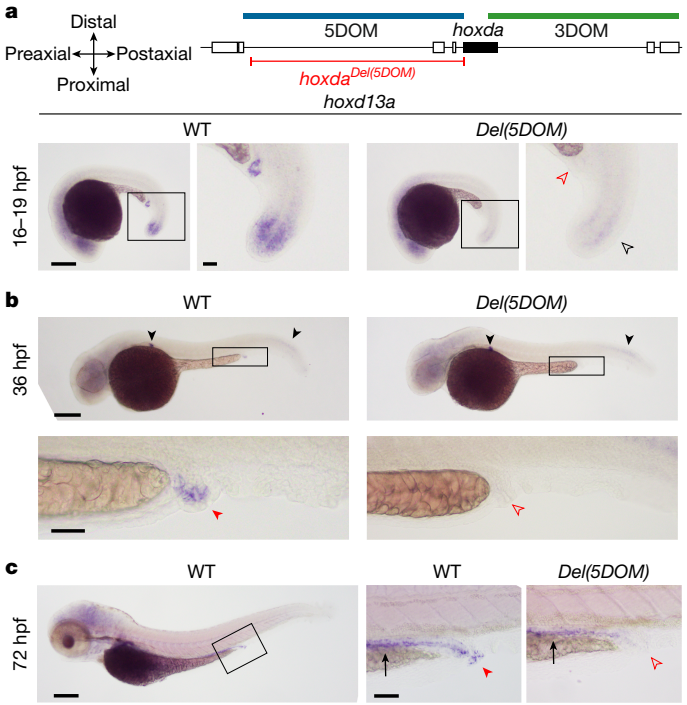


Fig. 2 | Effects of deleting 5DOM on *hoxd13a* regulation in the pseudo-cloacal region. a–**c**, Expression of *hoxd13a* is completely lost in the cloaca of 16-hpf (**a**), 36-hpf (**b**) and 72-hpf (**c**) embryos lacking 5DOM (open red arrowheads), whereas in embryos lacking 3DOM it is identical to controls (filled red arrowheads), indicating that 5DOM is required for *hoxd13a* activity in the pseudo-cloacal region. At 16 hpf, a temporary decrease in *hoxd13a* expression was observed in the tailbuds lacking 5DOM (open black arrowhead), but this effect was no longer observed at later stages (filled black arrowheads). WISH was performed on at least two embryos per condition. **b**, Enlargement of the cloacal region showing *hoxd13a* transcripts mostly lining the very end of the intestinal canal, converging towards the cloacal region. **c**, At 72 hpf, *hoxd13a* expression was detected in the posterior epithelial part of the gut in both control and mutant larvae (black arrow), indicating that expression in the cloacal region (red arrowheads) responds to a separate regulatory control. Scale bars, 200 µm (**a**–**c** (whole embryos)), 50 µm (**a**–**c** (enlarged views)).

These results revealed that in zebrafish, 5DOM-located enhancers regulate the *hoxd13a* genes in the cloacal area from its onset of expression until at least 72 hpf, whereas neither *hoxd10a* nor *hoxd4a* is expressed there (Extended Data Fig. 6). As previously reported, both *hoxd13a* and *hoxa13b*[23,24] transcripts were found in the region around the nascent pronephric ducts and the hindgut. These structures eventually converge towards a single pseudo-cloacal complex that exits the body with adjacent openings that never completely fuse. In 72-hpf larvae, *hoxd13a* mRNAs appeared in the posterior gut in both control and mutant samples. However, transcripts were still absent in the mutant cloacal region (Fig. 2c (black and red arrows, respectively)). This demonstrates that these two expression specificities are regulated separately.

The cloaca evolved at the base of the craniate lineage as a single orifice for the digestive, urinary and reproductive tracts, as found in birds and squamates. In mammals, a cloaca initially forms early in embryonic development, but as the embryo grows, it divides into different openings for the urogenital and digestive systems. To evaluate whether the observed 5DOM regulation of *hoxd13a* in the zebrafish pseudo-cloacal region is a derived or ancestral condition, we examined the developing mouse urogenital sinus (UGS), a structure derived from the mammalian embryonic cloacal area.

Hoxd gene regulation in the UGS

The UGS, positioned below the urinary bladder, is derived from a cloacal rudiment originating from the hindgut and ectodermal tissue[25,26]. During mid-gestation, as the nephric and Müllerian ducts grow towards the posterior end of the embryo, they meet and fuse with the invaginating cloaca. We performed WISH on dissected urogenital systems from control murine male and female embryos at embryonic day 18.5 (E18.5) (Fig. 3a,b). All genes tested, except *Hoxd13*, were detected in the anterior portions of the urogenital system, including the kidneys, uterus and deferens ducts[27,28] (Extended Data Fig. 7a,b). By contrast, *Hoxd13* expression was restricted to UGS in both male and female embryos, along with *Hoxd12*, *Hoxd11* and, to a weaker extent, *Hoxd10* (Fig. 3 and Extended Data Fig. 7). These are the same four genes that respond to both the digit and external genital long-range regulations exerted by 5DOM[29], thus suggesting a transcriptional control coming from this same 5′-located domain.

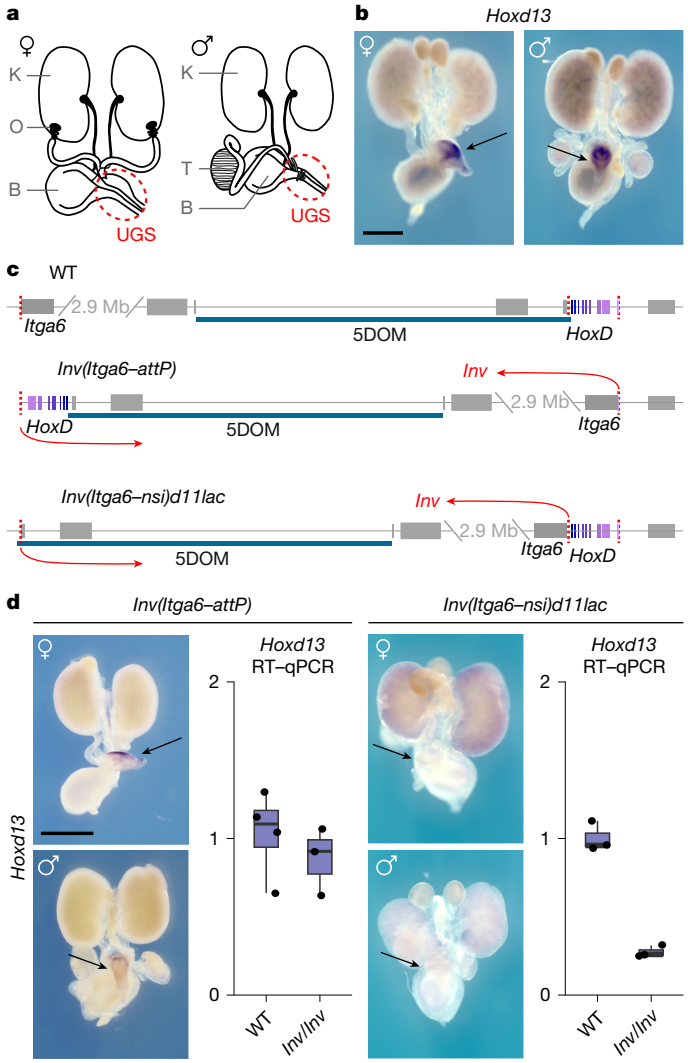


Fig. 3 | *Hoxd* gene expression in the mouse urogenital system. a, Schematic representations of male and female urogenital systems. UGS is indicated with a red circle. **b**, WISH of *Hoxd13* in representative female and male urogenital systems. *Hoxd13* is selectively expressed in UGS. **c**, Schematic representation of the two *HoxD* inversion alleles. The locations of the inversion breakpoints are depicted with red arrows. *Hox* genes shown in shades of purple. **d**, *Hoxd13* expression in urogenital systems of mice carrying the inversions (WISH, left panel; RT–qPCR, right panel). $HoxD^{Inv(Itga6-attP)}$: wild type ($n = 4$ samples) and inversion ($n = 3$ samples); $HoxD^{Inv(Itga6-nsi)d11lac}$: wild type ($n = 3$ samples) and inversion ($n = 3$ samples). Expression of *Hoxd13* in UGS was abolished when the target genes were disconnected from 5DOM. The boxes represent the interquartile range (IQR), with the lower and upper hinges denoting the first and third quartiles (25th and 75th percentiles). Whiskers extended from the hinges to the furthest data points within 1.5 times the IQR. The upper whisker reached the largest value within this range, whereas the lower whisker extended to the smallest value within 1.5 times the IQR from the hinge. Each experiment was repeated independently at least twice with similar results (**b**,**d**). B, bladder; K, kidney; O, ovary; T, testis. Scale bar, 1 mm (**b**,**d**).

We verified this using an engineered inversion that keeps the *HoxD* cluster linked with 5DOM but takes them far away from 3DOM (Fig. 3c; $HoxD^{Inv(Itga6-AttP)}$)[30]. In this allele, *Hoxd13* transcription in UGS was unaffected (Fig. 3d). We then tested a comparable inversion, yet with a breakpoint immediately 5′ to the *HoxD* cluster, thus disconnecting 5DOM from all *Hoxd* genes (Fig. 3c; $HoxD^{Inv(Itga6-nsi)d11lac)}$)[31]. This inversion led to a virtually complete loss of *Hoxd13* transcription (Fig. 3d), suggesting that most, if not all, UGS-specific enhancers were located within 5DOM. We confirmed this using a large BAC transgene containing only *HoxD* with no flanking sequences[30] (Extended Data Fig. 7c,d) introduced into mice lacking both copies of the *HoxD* locus[32] (Extended Data Fig. 7c). In this mutant line, *Hoxd13* transcription was not detected (Extended Data Fig. 7c (arrows)). Finally, we examined the β-galactosidase staining of a *LacZ* reporter integrated into the same BAC transgene. Whereas the reporter was strongly active in fetal kidneys as expected[22], UGS was not stained (Extended Data Fig. 7d). By contrast, a comparable *LacZ* enhancer–reporter transgene integrated into 5DOM within the inversion, separating 5DOM from the *HoxD* cluster (Extended Data Fig. 7d; $HoxD^{Inv(Itga6-nsi)d11lac)}$) robustly stained E18.5 UGS, again supporting the presence of UGS enhancers within 5DOM rather than within the *HoxD* cluster (Extended Data Fig. 7d).

We quantified the reduction in *Hoxd* gene expression in the $HoxD^{Inv(Itga6-nsi)d11lac}$ allele using RNA sequencing (RNA-seq) on E18.5 UGS of males and females. In both cases, *Hoxd13*, *Hoxd12* and *Hoxd10* transcription levels dropped abruptly when compared with the wild-type samples, whereas the transcription level of other *Hoxd* genes was not affected (Extended Data Fig. 8a). Altogether, these genetic configurations demonstrated that mammalian 5DOM contains UGS enhancers, similar to zebrafish 5DOM. It also showed that the *Hoxd* genes responsive to this regulation (*Hoxd13*–*Hoxd10*) are the same subgroup that responds to both digit and external genital enhancers.

Identification of mouse UGS enhancers

To identify UGS enhancers within the mouse 5DOM, we used three scanning deletion alleles covering 5DOM[2] (Fig. 4a (red)) and measured the change in expression by reverse transcription–quantitative polymerase chain reaction (RT–qPCR) (Extended Data Fig. 8b). In the $HoxD^{Del(Atf2-SB1)}$ allele, the most distal portion of 5DOM was removed, with no impact on *Hoxd* gene expression levels. However, when either the central ($HoxD^{Del(SB1-Rel5)}$) or the most proximal ($HoxD^{Del(Rel5-Rel1)}$) portions of 5DOM were removed, transcription of *Hoxd13*, *Hoxd12* and *Hoxd10* was significantly reduced, indicating that these two 5DOM intervals contain UGS enhancers (Extended Data Fig. 8b).

We then measured chromatin accessibility using the assay for transposase-accessible chromatin with sequencing (ATAC-seq) and profiled H3K27ac and H3K27me3 histone marks associated with either active or inactive chromatin, respectively, using chromatin immunoprecipitation followed by sequencing (ChIP–seq) on micro-dissected male UGSs (Fig. 4a). We identified a cluster of several conspicuous ATAC-seq and H3K27ac signals located approximately 200 kb upstream of *Hoxd13*, in a region encompassing the Rel5 breakpoint, that is, between the Del(SB1–Rel5) and the Del(Rel5–Rel1) deletions (Fig. 4a (dashed box)). Within this 67-kb large region, the ATAC and H3K27ac signals matched three elements previously characterized as enhancer sequences, yet with distinct tissue specificities. The GT2 and Island E sequences were identified as pan-dorsal and proximal-dorsal genital tubercle-specific enhancers, respectively[33,34], whereas the CsB element was reported as a neural and distal limb and fin enhancer element[1,12,35]. Within these two elements, the main ATAC peak was positioned in a small region that was relatively depleted for the H3K27ac mark (Extended Data Fig. 9), which is a hallmark of active enhancer elements[36]. We tested these three putative enhancers in an enhancer–reporter assay, and all three sequences were able to drive robust *lacZ* expression in UGS, closely matching the expression of posterior *Hoxd* genes in this area in both male and female specimens (Fig. 4b). This indicates that in mammals, 5DOM contains a set of multiple enhancer elements that control the transcriptional activation of *Hoxd* genes in UGS.

Enhancers in the fish 5DOM

Although the GT2 and CsB elements showed some islands of sequence conservation across bony fish *hoxda* loci, Island E was present only in mammals (Extended Data Fig. 9). Because the teleosts underwent

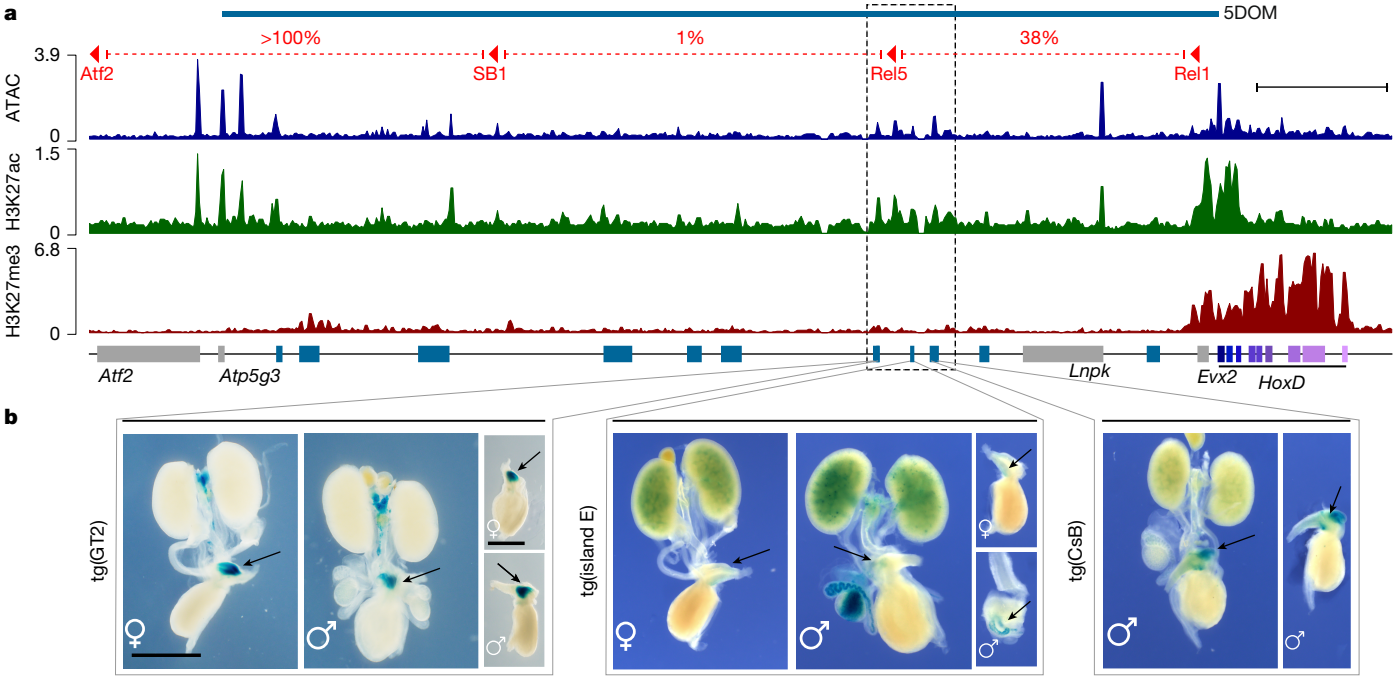


Fig. 4 | UGS enhancers located in 5DOM. a, Chromatin accessibility (ATAC-seq; blue track) and H3K27ac (green track) and H3K27me3 (red track) ChIP–seq profiles from micro-dissected male UGS at E18.5. The red lines on top delineate the three deletions within 5DOM with the percentage of *Hoxd13* expression left in UGS after each deletion (Extended Data Fig. 8). *Hoxd* genes are in purple. Blue rectangles indicate previously described 5DOM enhancers. The dashed box highlights an H3K27ac-positive cluster of ATAC-seq peaks lacking H3K27me3 and containing three enhancer sequences: GT2, Island E and CsB. ATAC-seq and H3K27ac, $n = 2$ experiments, one representative example is shown; H3K27me3, $n = 1$ experiment. **b**, Regulatory potential of the GT2, Island E and CsB elements when cloned into a *lacZ* reporter cassette. GT2 induces robust *lacZ* expression in UGS of both male and female embryos, whereas Island E shows weaker expression. The CsB transgene induces robust expression in males (no data available for females). Each experiment was repeated independently at least twice with similar results. Scale bars, 100 kb (**a**), 1 mm (**b**).

another whole-genome duplication (TWGD), such sequences may have been reshuffled to different genomic locations and hence might bias our interpretation. To assess the impact of TWGD on the structure of 5DOM, we aligned the genomes of 13 vertebrate species, including holosteans (gar and bowfin), cartilaginous fishes (skate and catshark) and the coelacanth, a lobe-finned fish that did not undergo further whole-genome duplication. First, we projected the alignment onto the zebrafish genome (Supplementary Fig. 2a). With one exception (Supplementary Fig. 2a (red arrow)), all sequences conserved between zebrafish and fugu, two distantly related teleosts, were conserved in the gar and bowfin genomes. This suggests that TWGD did not lead to widespread loss or rearrangement of conserved 5DOM elements despite the genetic redundancy introduced by duplication. This may be related to the loss of the *hoxdb* paralogue cluster, thus leading to the maintenance of all regulatory constraints that were initially applied to the *hoxda* locus before duplication.

In support of this conclusion, conservation blocks within zebrafish alignments largely covered the *hox* genes, which have many paralogues yet not the 5DOM regulatory landscape, demonstrating that the potential *hoxdb* counterpart of 5DOM was also lost after TWGD. Finally, although tetrapods and ray-finned fishes are more closely related to each other than to cartilaginous fishes, the 5DOM of ray-finned and cartilaginous fishes exhibited a higher degree of conservation, indicating sequence divergence in tetrapods following their split from the coelacanth lineage (Supplementary Fig. 2a (black arrows)).

To assess the impact of TWGD when a non-teleost species was used as a reference, we projected a 13-way alignment onto the mouse genome (Supplementary Fig. 2b). This revealed a strong positive correlation between conservation and phylogeny. Conservation blocks between the mouse and any species of fish (teleost or non-teleost) were largely the same, although with a few cases in which sequences were lost specifically in the zebrafish genome (Supplementary Fig. 2b (black arrows)). To confirm that the conservation blocks truly represent the corresponding loci across species, we repeated the alignments using the 5DOM region only in mouse, gar and zebrafish (Supplementary Fig. 2c). This confirmed that, at least at the *hoxda* locus, TWGD neither disrupted nor reshuffled any conserved sequence beyond what would be expected on the basis of phylogenic distance. We concluded that TWGD does not bias the analysis of 5DOM regulatory function.

Because of the low sequence conservation between the mammalian and fish 5DOM, we looked at potential regulatory sequences in the fish counterpart, either by producing ATAC-seq profiles (Supplementary Fig. 3 (three profiles on top)) or by mining single-cell ATAC-seq datasets[37,38] (Supplementary Fig. 3 (bottom four profiles)). The quality and reproducibility of these datasets can be verified by the signals recovered over the *hoxda* gene cluster itself. The range of accessibility peaks was scored in common with the micro-dissected head, cloaca and tailbud samples at 30 hpf, with little conservation with the main mouse sequences (such as for the limbs; bottom line). The extraction of cell clusters identified as tailbud mesoderm (24 hpf), predicted cloaca (14 hpf) and pectoral fins at both 48 hpf and 72 hpf from a previous study[38] (Methods) revealed common and specific peaks (Supplementary Fig. 3). Notably, although the predicted cloaca and pectoral fin cell types shared a few accessibility peaks, the profiles were clearly different, with many more peaks in the former than in the latter.

Enhancer deletion in zebrafish

Because of the transgenic activity of CsB during both fish and mouse appendage development[18,35], the ambiguity between the ATAC peak in pectoral fin cells and the possibility of a variable effect of the 5DOM deletion on *hoxd13a* expression during pectoral fin development, we sought to remove the CsB element from the zebrafish genome and observe the

effect on *hoxd13a* regulation. To have an efficient readout over developmental time and track potential expression changes with increased sensitivity, we generated an endogenous knock-in reporter line that expresses tdTomato from the *hoxd13a* locus (Extended Data Fig. 10a,b). We used CRISPR–Cas9 to delete CsB in *cis* to the *hoxd13a*$^{Tg(hsp70:tdTomato)}$ endogenous reporter and assessed its expression in the developing cloaca, fins and tail using confocal microscopy. Fish carrying the CsB deletion did not exhibit any decrease in fin signal compared with siblings with an intact CsB (Extended Data Fig. 10c), demonstrating that, unlike in the mouse, the zebrafish CsB sequence did not have a critical role in the regulation of *hoxd13a* in the developing appendage. This result agrees with previous findings that the pufferfish CsB sequence was unable to drive distal transgene expression in mouse[35], indicating that the teleost CsB element either does not contain distal fin regulatory functionality or that a potential functionality cannot be revealed by our deletion approach. Expression was also compared between the cloacal and tailbud regions of CsB-deleted and CsB-intact reporter animals, and no major differences in reporter signal were scored. If anything, the CsB-deleted animals exhibited slightly increased activity in the cloacal region at the 19-somite stage (Extended Data Fig. 10c–e), an observation reminiscent of the complex regulatory interactions occurring within this landscape, as reported in the mouse counterpart during the outgrowth of external genitals[21].

Ancestral cloacal regulatory landscape

The conservation for more than 350 million years of *Hoxd13* regulation in the cloacal region by 5DOM-located sequences suggests an ancestral role of *Hoxd13* in cloacal morphogenesis. This critical function was demonstrated in mice, in which the combined mutation of both *Hoxa13* and *Hoxd13* had a drastic effect on the development of the posterior part of the digestive and urogenital systems[28,39], causing an absence of any detectable UGS[28]. Previous studies have revealed the expression of most *hox13* genes in the intestine and cloacal regions of developing zebrafish (Supplementary Fig. 4). Additionally, *hoxa* genes are differentially regulated during patterning of the goby fish cloacal region[40]. Although these results indicate functional conservation[41], the requirement of *hox13* genes for cloaca development in fishes remains to be established. Therefore, we examined whether zebrafish *hox13* mutants exhibited cloacal phenotypes.

Wild-type zebrafish exhibited a pseudo-cloacal configuration, in which the hindgut and pronephric duct exited the trunk through separate but adjacent openings. The outlet of the hindgut was anterior to that of the pronephric duct, and a septum resided in between (Fig. 5a,e). Homozygous single mutants of *hoxa13a*, *hoxa13b* and *hoxd13a* were indistinguishable from the wild-type arrangement, as were animals triply heterozygous for these genes (Fig. 5b,f and Supplementary Fig. 5). However, the combined *hoxa13a;hoxa13b* double homozygous mutants exhibited connection between the hindgut and pronephric duct before exiting the body through a single opening (Fig. 5c,g and Supplementary Table 1). The loss of *Hoxa13* paralogues also affected the pronephric duct and hindgut length at the level of the median fin fold (Supplementary Fig. 5). A more severe phenotype was observed in *hoxa13a;hoxa13b;hoxd13a* triple mutants, in which the septum was dysmorphic and the hindgut and pronephric duct were fused, resulting in a large shared lumen and outlet (Fig. 5d,h and Supplementary Table 1). These results revealed a conserved requirement of *Hox13* function for the normal patterning of the termini of the digestive and urogenital systems across vertebrates.

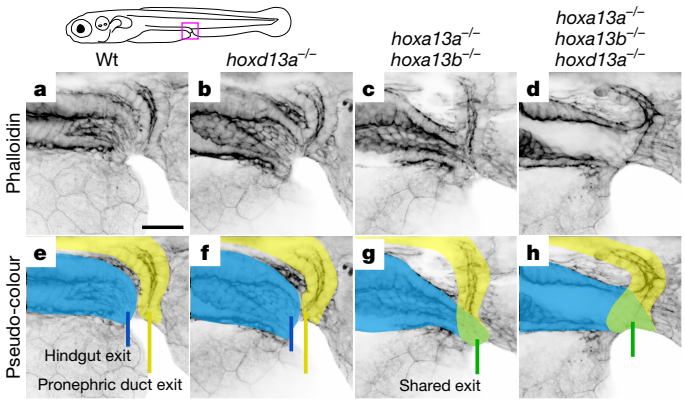


Fig. 5 | Loss of *hox13* paralogues in the zebrafish resulted in defects in the cloacal region. a–h, Confocal microscopy of phalloidin-labelled cloacal regions of wild-type and *hox13* mutant zebrafish at 6 days post-fertilization shown in a single channel (**a**–**d**) and with pseudo-colouring (**e**–**h**). Pseudo-colouring indicates hindgut (blue), pronephric duct (yellow) or fused ducts (green). The phenotypes reported here were consistent across individuals collected from four independent crosses of triply heterozygous parents. **a**,**e**, Wild-type fish have adjacent but distinct openings for the hindgut (blue line) and pronephric duct (yellow line) ($n = 10$), as do *hoxd13a* mutants ($n = 6$) (**b**,**f**). **c**,**g**, *hoxa13a;hoxa13b* double mutants exhibited fusion of the hindgut and pronephric duct and a single opening (green line) ($n = 4$). **d**,**h**, *hoxa13a;hoxa13b;hoxd13a* triple mutants showed connection of the hindgut and pronephric duct to form a large shared lumen (green) with a single opening (green line) ($n = 4$). The consistent phenotypes observed in the double and triple mutant classes were significant when each class was individually compared with wild type using Fisher's exact test (wild type, $n = 10$; mutant, $n = 4$; $P = 9.9 \times 10^{-4}$). Scale bar, 30 µm.

Discussion

Hox regulation and fin-to-limb transition

The expression and function of *Hoxd* and *Hoxa* genes have been central to hypotheses attempting to explain the evolutionary change from fins to limbs[5,11,13,14]. By comparing their complex transcription patterns across actinopterygian, chondrichthyan and sarcopterygian fishes, various efforts have been made to relate the two types of paired appendages. These analyses have led to the conclusion that despite being composed of different types of skeletons, the development of actinopterygian fin rays and digits has a common regulatory architecture[11,42]. Deletion of the two TADs flanking the zebrafish *hoxda* cluster has shown that the essential digit regulatory landscape characterized in tetrapods indeed has a structural counterpart in teleosts[43,44] (Extended Data Fig. 1). However, unlike in limbs, *hoxda* gene expression in distal fins is for the main part maintained in the absence of this regulatory landscape. Although a contribution from 5DOM to *hoxd13a* regulation probably exists[12], most of the regulatory control may reside within the gene cluster itself, probably in the vicinity of the *hoxd13a*, *hoxd12a* and *hoxd11a* genes, that is, the three genes sharing the same expression in postaxial cells[16].

This observation confirmed the results obtained when assaying fish 5DOM conserved sequences as transgenes, either in zebrafish or mice[12,18]. In addition, the lack of any strong effect of our deletion of the fish CsB, either in pectoral fins or in cloacal cells, supports the distinct functionalities of this enhancer sequence between mammals and fishes. CsB may thus play a role during zebrafish development that is only tangentially overlapping with that of its murine counterpart, in particular during differentiation of V0 interneurons, as driven by the gene *Evx2* (ref. 23) located nearby in 5DOM. This also explains why the zebrafish *lnpa* gene, which, like *Evx2*, is embedded into 5DOM, is not expressed in the emerging distal pectoral fin buds[16], whereas the mouse counterpart has a strong distal expression owing to enhancer hijacking[1].

The presence of a functionally truncated distal appendage 5DOM regulatory landscape in teleosts may potentially illustrate an intermediate step in the full co-option of this regulation, as achieved in tetrapods (see "Co-options of a whole regulatory landscape"). Alternatively, it may reflect a secondary loss of distal enhancers either in the actinopterygian lineage or in the aftermath of TWGD because such an event

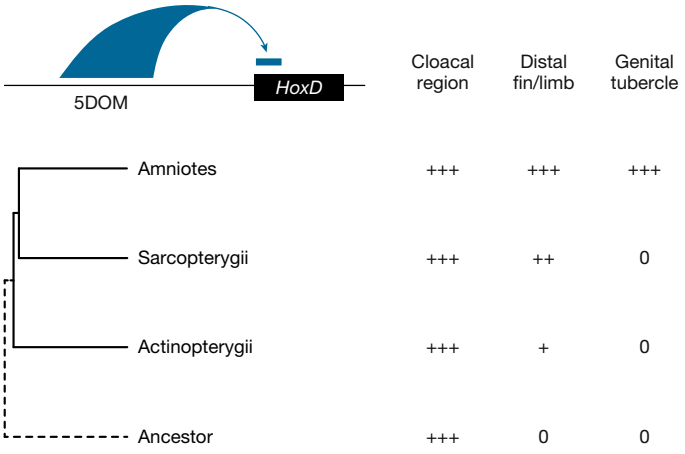


Fig. 6 | Evolutionary co-option of the *HoxD* 5DOM regulatory landscape. Schematic representation of posterior *Hoxd* gene regulation by the 5DOM regulatory landscape (top left) and (at least) three developmental contexts in which this landscape is functional (top right). On the left are shown the phylogenetic relationships between taxa where distal fins, distal limbs and external genitals emerged, whereas on the right, the corresponding 5DOM regulatory contributions to these structures are indicated; '0' denotes the absence of any given structure. In this hypothetical view, the 5DOM cloacal regulation is an ancestral feature. In actinopterygian fishes, 5DOM lightly contributes to *hoxda* gene regulation in the postaxial and distal territories of paired fin buds. The regulatory importance of 5DOM in distal fin territories has increased in sarcopterygian fishes. In amniotes, the 5DOM contribution expands to take over the entire regulation of posterior *Hoxd* genes in digits, as suggested by many enhancers with mixed specificities. Similarly, a distinct yet overlapping set of 5DOM-located enhancers entirely control *Hoxd* gene expression in the genital tubercle. It is difficult to infer the temporal sequence of the latter two co-options of this regulatory landscape. However, because genitalia are late amniote specializations, it is conceivable that elaboration of digital character arose initially, a sequence also supported by the appearance of digits in sarcopterygian fishes. A second co-option of this multifunctional regulatory landscape might have occurred along with the evolution of external genitals, facilitated both by the developmental proximity between external genitals and the embryonic cloacal region where posterior *Hox* genes are initially expressed[52], and by the tight developmental relationships between amniote limbs and genitals[27,52].

may lead to the reorganization of paralogous regulatory landscapes. However, our phylogenetic analysis of sequence conservation within 5DOM showed that the latter option is unlikely for this specific locus, which does not maintain any paralogous copy in zebrafish. In fact, the structural organization of 5DOM in zebrafish is very close to that of non-teleost fishes, such as the gar or bowfin, with very minor changes matching phylogenetic distances. Therefore, our conclusions are considered to apply to fishes in general rather than to teleosts only.

Ancestral cloacal regulation

Zebrafish *hoxa13a* and *hoxd13a*, as well as other *hox13* paralogues[37], are strongly expressed in and around the developing cloacal region[23,24]. This is an area where the extremities of both the gastro-intestinal tract and the reproductive and urinary systems come together, although their openings remain separated, unlike in some chondrichthyan fishes or other vertebrates, in which the tubes coalesce into a single opening (such as in sharks or birds). This pseudo-cloacal structure is disrupted in zebrafish carrying *hox13* mutant alleles, with an abnormal fusion between the intestinal and pronephric openings, thus giving rise to a single, abnormal, cloacal opening. Likewise, the developing murine UGS expresses *Hoxd13* (ref. 27), and double mouse *Hoxd13–Hoxa13* mutant animals have severely malformed posterior regions[28,39], with no distinguishable UGS[28], illustrating that the evolutionary conservation of this regulatory landscape is accompanied by shared functional effects.

We also documented that, as with zebrafish, the control of murine posterior *Hoxd* genes in the cloaca is achieved by enhancers located within the 5′ located regulatory landscape, that is, in the same genomic region that regulates expression in both digits and external genitals. In mice, several 5DOM enhancers are somewhat versatile, such as the GT2 sequence, which is both UGS and genitalia-specific[33], whereas CsB is UGS and digit-specific[1]. Other enhancer sequences, however, seem to have kept a unique specificity, such as 'Island 2', the strongest *Hox* digit enhancer identified thus far[45], which is located in a different area of 5DOM[2] and absent in zebrafish[12]. These observations illustrate the 'functional adaptation' of enhancers, which could be facilitated by spatial proximity within the same large chromatin domain, thus triggering the sharing of upstream factors[46]. Finally, all the regulatory specificities encoded in this 5DOM landscape control the same subset of posterior *Hoxd* genes in tetrapods (from *Hoxd13* to *Hoxd10*), suggesting that although groups of enhancers can be reused for new tissue types, there is a constraint on which genes they can target.

Co-options of a whole regulatory landscape

In vertebrates, *Hox13* genes are located within a TAD that is distinct from the one containing more anterior *Hox* genes and their regulations[15,47]. This condition prevents *Hox13* from being activated too early and hence too anteriorly in the body axis, which is detrimental to the embryo owing to the potent posteriorizing function of these proteins[48]. As a result, *Hoxd13* was probably the main target gene that triggered and stabilized the various evolutionary co-options of 5DOM regulations because of its location within the 5DOM TAD and through its function to organize posterior or distal body parts together with its *Hoxa13* paralogue[7,28,39]. Our results indicate that the initial functional specificity of this regulatory landscape was to organize a cloacal region, which is the posterior part of the intestinal and urogenital systems (Fig. 6). This conclusion is supported by the documented expression of *Hox13* paralogues in the cloacal regions of paddlefish[41,49], cat shark[49,50] and lampreys[51], suggesting that this pattern is characteristic of the common ancestor of craniates. Altogether, our results indicate that the regulatory landscape involved in the evolution of genitalia and limbs first arose to drive the formation of the cloaca and was co-opted, supporting the development and diversification of genital and digital morphologies.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09548-0.

1. Spitz, F., Gonzalez, F. & Duboule, D. A global control region defines a chromosomal regulatory landscape containing the *HoxD* cluster. *Cell* **113**, 405–417 (2003).
2. Montavon, T. et al. A regulatory archipelago controls *Hox* genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
3. Shubin, N., Tabin, C. & Carroll, S. Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009).
4. Thorogood, P. in *Developmental Patterning of the Vertebrate Limb* (eds Hinchliffe, J. R. et al.) 347–354 (Plenum, 1991).
5. Woltering, J. M. & Duboule, D. The origin of digits: expression patterns versus regulatory mechanisms. *Dev. Cell* **18**, 526–532 (2010).
6. Cloutier, R. et al. Elpistostege and the origin of the vertebrate hand. *Nature* **579**, 549–554 (2020).
7. Fromental-Ramain, C. et al. *Hoxa-13* and *Hoxd-13* play a crucial role in the patterning of the limb autopod. *Development* **122**, 2997–3011 (1996).
8. Dolle, P., Izpisua-Belmonte, J. C., Falkenstein, H., Renucci, A. & Duboule, D. Coordinate expression of the murine *Hox-5* complex homoeobox-containing genes during limb pattern formation. *Nature* **342**, 767–772 (1989).
9. Davis, M. C. The deep homology of the autopod: insights from Hox gene regulation. *Integr. Comp. Biol.* **53**, 224–232 (2013).
10. Zakany, J. & Duboule, D. The role of Hox genes during vertebrate limb development. *Curr. Opin. Genet. Dev.* **17**, 359–366 (2007).
11. Nakamura, T., Gehrke, A. R., Lemberg, J., Szymaszek, J. & Shubin, N. H. Digits and fin rays share common developmental histories. *Nature* **537**, 225–228 (2016).

12. Gehrke, A. R. et al. Deep conservation of wrist and digit enhancers in fish. *Proc. Natl Acad. Sci. USA* **112**, 803–808 (2015).
13. Sordino, P., van der Hoeven, F. & Duboule, D. *Hox* gene expression in teleost fins and the origin of vertebrate digits. *Nature* **375**, 678–681 (1995).
14. Woltering, J. M. et al. Sarcopterygian fin ontogeny elucidates the origin of hands with digits. *Sci. Adv.* **6**, eabc3510 (2020).
15. Andrey, G. et al. A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013).
16. Ahn, D. & Ho, R. K. Tri-phasic expression of posterior *Hox* genes during development of pectoral fins in zebrafish: implications for the evolution of vertebrate paired appendages. *Dev. Biol.* **322**, 220–233 (2008).
17. Hawkins, M. B., Henke, K. & Harris, M. P. Latent developmental potential to form limb-like skeletal structures in zebrafish. *Cell* **184**, 899–911 (2021).
18. Schneider, I. et al. Appendage expression driven by the Hoxd Global Control Region is an ancient gnathostome feature. *Proc. Natl Acad. Sci. USA* **108**, 12782–12786 (2011).
19. Bolt, C. C. & Duboule, D. The regulatory landscapes of developmental genes. *Development* **147**, dev171736 (2020).
20. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
21. Amândio, A. R., Lopez-Delisle, L., Bolt, C. C., Mascrez, B. & Duboule, D. A complex regulatory landscape involved in the development of mammalian external genitals. *eLife* **9**, e52962 (2020).
22. Di-Poi, N., Zakany, J. & Duboule, D. Distinct roles and regulations for *HoxD* genes in metanephric kidney development. *PLoS Genet.* **3**, e232 (2007).
23. Sordino, P., Duboule, D. & Kondo, T. Zebrafish *Hoxa* and *Evx-2* genes: cloning, developmental expression and implications for the functional evolution of posterior *Hox* genes. *Mech. Dev.* **59**, 165–175 (1996).
24. van der Hoeven, F., Sordino, P., Fraudeau, N., Izpisua-Belmonte, J. C. & Duboule, D. Teleost *HoxD* and *HoxA* genes: comparison with tetrapods and functional evolution of the *HOXD* complex. *Mech. Dev.* **54**, 9–21 (1996).
25. Georgas, K. M. et al. An illustrated anatomical ontology of the developing mouse lower urogenital tract. *Development* **142**, 1893–1908 (2015).
26. Liaw, A. et al. Development of the human bladder and ureterovesical junction. *Differentiation* **103**, 66–73 (2018).
27. Dolle, P., Izpisua-Belmonte, J. C., Brown, J. M., Tickle, C. & Duboule, D. HOX-4 genes and the morphogenesis of mammalian genitalia. *Genes Dev.* **5**, 1767–1767 (1991).
28. Warot, X., Fromental-Ramain, C., Fraulob, V., Chambon, P. & Dolle, P. 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**, 4781–4791 (1997).
29. Montavon, T., Le Garrec, J.-F., Kerszberg, M. & Duboule, D. Modeling *Hox* gene regulation in digits: reverse collinearity and the molecular origin of thumbness. *Genes Dev.* **22**, 346–359 (2008).
30. Schep, R. et al. Control of *Hoxd* gene transcription in the mammary bud by hijacking a preexisting regulatory landscape. *Proc. Natl Acad. Sci. USA* **113**, E7720–E7729 (2016).
31. Tschopp, P. & Duboule, D. A regulatory 'landscape effect' over the *HoxD* cluster. *Dev. Biol.* **351**, 288–296 (2011).
32. Spitz, F. et al. Large scale transgenic and cluster deletion analysis of the *HoxD* complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev.* **15**, 2209–2214 (2001).
33. Lonfat, N., Montavon, T., Darbellay, F., Gitto, S. & Duboule, D. Convergent evolution of complex regulatory landscapes and pleiotropy at *Hox* loci. *Science* **346**, 1004–1006 (2014).
34. Lonfat, N. *An Ancestral Regulatory Mechanism Underlies Hoxd Gene Expression in Both Developing Genitals and Digits*. PhD thesis, École Polytechnique Fédérale de Lausanne (2013).
35. Gonzalez, F., Duboule, D. & Spitz, F. Transgenic analysis of *Hoxd* gene regulation during digit development. *Dev. Biol.* **306**, 847–859 (2007).
36. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
37. Sur, A. et al. Single-cell analysis of shared signatures and transcriptional diversity during zebrafish development. *Dev. Cell* **58**, 3028–3047 (2023).
38. Sun, K. et al. Mapping the chromatin accessibility landscape of zebrafish embryogenesis at single-cell resolution by SPATAC-seq. *Nat. Cell Biol.* **26**, 1187–1199 (2024).
39. Kondo, T., Zákány, J., Innis, J. W. & Duboule, D. Of fingers, toes and penises. *Nature* **390**, 29–29 (1997).
40. Crow, K. D., Sadakian, A. & Kaslly, N. A. The role of the 5′ HoxA genes in the development of the hindgut, vent, and a novel sphincter in a derived teleost (bluebanded goby, *Lythrypnus dalli*). *J. Exp. Zool. B Mol. Dev. Evol.* **340**, 518–530 (2023).
41. Archambeault, S., Taylor, J. A. & Crow, K. D. HoxA and HoxD expression in a variety of vertebrate body plan features reveals an ancient origin for the distal Hox program. *EvoDevo* **5**, 44 (2014).
42. Tulenko, F. J. et al. *HoxD* expression in the fin-fold compartment of basal gnathostomes and implications for paired appendage evolution. *Sci. Rep.* **6**, 22720 (2016).
43. Franke, M. et al. CTCF knockout in zebrafish induces alterations in regulatory landscapes and developmental gene expression. *Nat. Commun.* **12**, 5415 (2021).
44. Woltering, J. M., Noordermeer, D., Leleu, M. & Duboule, D. Conservation and divergence of regulatory strategies at *Hox* loci and the origin of tetrapod digits. *PLoS Biol.* **12**, e1001773 (2014).
45. Bolt, C. C. et al. Context-dependent enhancer function revealed by targeted inter-TAD relocation. *Nat. Commun.* **13**, 3488 (2022).
46. Darbellay, F. & Duboule, D. Topological domains, metagenes, and the emergence of pleiotropic regulations at *Hox* loci. *Curr. Top. Dev. Biol.* **116**, 299–314 (2016).
47. Berlivet, S. et al. Clustering of tissue-specific sub-TADs accompanies the regulation of *HoxA* genes in developing limbs. *PLoS Genet.* **9**, e1004018 (2013).
48. Young, T. et al. *Cdx* and *Hox* genes differentially regulate posterior axial growth in mammalian embryos. *Dev. Cell* **17**, 516–526 (2009).
49. Tulenko, F. J. et al. Fin-fold development in paddlefish and catshark and implications for the evolution of the autopod. *Proc. R. Soc. B.* **284**, 20162780 (2017).
50. Freitas, R., Zhang, G. & Cohn, M. J. Biphasic *Hoxd* gene expression in shark paired fins reveals an ancient origin of the distal limb domain. *PLoS ONE* **2**, e754 (2007).
51. Kuraku, S. et al. Noncanonical role of Hox14 revealed by its expression patterns in lamprey and shark. *Proc. Natl Acad. Sci. USA* **105**, 6679–6683 (2008).
52. Tschopp, P. et al. A relative shift in cloacal location repositions external genitalia in amniote evolution. *Nature* **516**, 391–394 (2014).

Methods

Animal husbandry and ethics

All experiments using mice were approved and performed in compliance with the Swiss Law on Animal Protection (Loi fédérale sur la Protection des Animaux) under licence numbers GE45/20 and GE81/14. All animals were kept as a continuous backcross with C57BL6 × CBA F_1 hybrids. The 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. Temperatures were maintained between 22 °C and 23 °C, with humidity levels between 45% and 55%. The air was renewed 17 times per hour. Zebrafish (*Danio rerio*) were maintained according to standard conditions[53] under a 14 h/10 h on/off light cycle at 26 °C, with set points of 7.5 and 600 µS for pH and conductivity, respectively. All zebrafish husbandry procedures were approved and accredited either by the Federal Food Safety and Veterinary Office of the canton of Vaud, Switzerland (no. VD-H23), by the animal committees of Rutgers University under protocol no. 201702646 or under the guidance of the Institutional Animal Care and Use Committee (IACUC) of Boston Children's Hospital. AB, Tu and TL were used as wild-type strains and were obtained from the European Zebrafish Resource Center. The *hoxda*$^{Del(3DOM)}$ and *hoxda*$^{Del(5DOM)}$ mutants were generated for this study. Zebrafish embryos were derived from freely mating adults. Wild-type sibling *hoxda*$^{Del(3DOM)}$ and *hoxda*$^{Del(5DOM)}$ homozygous embryos were obtained by crossing the corresponding heterozygous mutant. Embryos were collected within 30 min after spawning and incubated at 28.5 °C in fish water, shifted to 20 °C after reaching 80% epiboly and grown at 28.5 °C to the proper developmental stage according to a previous study[54]. Pigmentation was prevented by treating the embryos with 0.002% *N*-phenylthiourea from 1 day post-fertilization (dpf) onwards. Sex was determined for animals used in the E18.5 UGS mouse experiments. Animals in other mouse experiments and in zebrafish experiments were not sexed. The sample size was not predetermined by a statistical test. Randomization and blinding were not conducted because the mutant and control animals were processed together in the same batch and grouped on the basis of their genotypes.

Generation of deletions in zebrafish

The *hoxda*$^{Del(3DOM)}$ and *hoxda*$^{Del(5DOM)}$ mutant alleles were generated using the CRISPR–Cas9 system described in a previous study[55]. The sequences of the CRISPR RNAs (crRNAs) used are listed in Supplementary Table 2. Loci were identified using the GRCz11 zebrafish genome assembly available on Ensembl. The corresponding genomic regions were amplified and sequenced from fin clips. Adults carrying verified target sequences were isolated and then selected for breeding to generate eggs for genome editing experiments. The guide RNA target sites were determined using the open-source software CHOPCHOP (http://chopchop.cbu.uib.no/index.php). Chemically synthesized Alt-R crRNAs and Alt-R trans-activating CRISPR RNAs (tracrRNAs) and the Alt-R Cas9 protein were obtained from Integrated DNA Technologies (IDT). To test the efficiency of these guide RNAs in generating the expected mutant alleles, we injected boluses ranging from 100 µm to 150 µm and containing 5 µM of the duplex crRNAs, tracrRNA and Cas9 ribonucleoprotein complex into the cytoplasm of one-cell-stage embryos. Injecting the ribonucleoprotein complex solution in a 100-µm bolus gave less than 5% mortality. With this condition, 30% of the embryos carried the 5DOM deletion and 15% carried the 3DOM deletion. For each condition, we extracted the genomic DNA of 20 individual larvae at 24 hpf for genotyping[56]. Identification of *hoxda*$^{Del(3DOM)}$ and *hoxda*$^{Del(5DOM)}$ mutants was performed using polymerase chain reaction (PCR). Amplification of *evx2* was used as a control to confirm the presence or absence of 5DOM. The PCR mix was prepared using Phusion High-Fidelity DNA Polymerase (New England Biolabs), and primer sequences are listed in Supplementary Table 2. In parallel, 120 larvae per allele were raised to adulthood. To identify founders, F_0 adults were outcrossed with wild type and 25 embryos were genotyped. Three and four independent founders were obtained for the *hoxda*$^{Del(5DOM)}$ allele and *hoxda*$^{Del(3DOM)}$, respectively. Two founders of each deletion were verified by Sanger sequencing (Supplementary Data 1) and used for further experiments.

Generation of knock-in reporter line

The endogenous *hoxd13a* reporter line (*hoxd13a*$^{Tg(hsp70:tdTomato)}$) was produced using a CRISPR–Cas9-mediated Gbait vector knock-in approach[57,58]. A guide targeting the coding region of exon 1 of *hoxd13a* (*hoxd13a*_KI_crRNA) was co-injected with a Gbait vector targeting guide (GFP_crRNA) and *Gbait:hsp70l:tdTomato* plasmid[17]. The injected embryos were screened for endogenous reporter RFP signal in expected *hoxd13a* expression domains, and positive individuals were raised to adulthood to outcross and recover F_1 germline founders. To verify vector insertion and orientation in founders, genomic primers (*hoxd13a*_KI_F and *hoxd13a*_KI_R) were each paired with primers internal to the insert (LacZ_F and hsp70_R) for PCR and Sanger sequencing. The vector was oriented in the reverse direction relative to the endogenous promoter in the *hoxd13a*$^{Tg(hsp70:tdTomato)}$ line, but reporter expression matched previously published in situ hybridization data and an *hoxd13a* knock-in line (*hoxd13a*^{egfp}) generated independently by another research group[59]. Genotyping primers are listed in Supplementary Table 2.

Removal of CsB in *cis* to *hoxd13a*$^{Tg(hsp70:tdTomato)}$

To delete the CsB sequence from the chromosome carrying the *hoxd13a*$^{Tg(hsp70:tdTomato)}$ endogenous reporter, each individual crRNA was duplexed with tracrRNA and injected at a final concentration of 6.25 µM with 1 µg Alt-R S.p. Cas9 Nuclease V3 (IDT). To estimate guide efficiency, DNA was extracted from four pools of three embryos each from 12 injected embryos and 12 control siblings and analysed using the T7 endonuclease 1 mismatch detection assay[60]. Embryos injected with efficient guides were raised to adulthood to outcross and identify founders. Guides flanking the CsB region (CsB_g1_crRNA and CsB_g2_crRNA) were injected into the *hoxd13a*$^{Tg(hsp70:tdTomato)}$ background. The injected embryos were sorted by RFP signal at 1 dpf, and 16 positive animals from each clutch were screened for CsB removal using PCR with deletion-spanning primers (CsB_g1_F and CsB_g2_R) that did not amplify the intact locus under short elongation conditions. Clutches exhibiting a high frequency of CsB removal were raised to adulthood, and individuals were outcrossed to T5D wild type to obtain embryos carrying *hoxd13a*$^{Tg(hsp70:tdTomato)}$-Del(CsB) chromosomes. To identify CsB deletions in *cis* to the reporter, outcrossed embryos were sorted for RFP and then genotyped for the CsB deletion. One F_0 injected parent (purple male 3) produced gametes with *hoxd13a*$^{Tg(hsp70:tdTomato)}$-Del(CsB) chromosomes at high frequency (approximately 25%), as well as gametes in which the CsB in *cis* to the reporter was left intact. Sanger sequencing of the deletion-spanning PCR product from 16 embryos revealed that each *hoxd13a*$^{Tg(hsp70:tdTomato)}$-Del(CsB) chromosome carried an identical deletion, suggesting clonality. Embryos resulting from outcrosses of this injected individual (purple male 3) were used in a subsequent expression analysis. The sequences of the crRNAs and genotyping primers used are listed in Supplementary Table 2. Sanger sequences of zebrafish founders are listed in Supplementary Data 1.

Quantification of *hoxd13a*$^{Tg(hsp70:tdTomato)}$ expression

Outcrossed progeny with RFP signal from the endogenous reporter were collected at the 19-somite stage and at 72 hpf. Embryos were fixed in 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS) (pH 7.4) for 2 h at room temperature with agitation in light-blocking containers, rinsed two times for 10 min each in PBS with 0.01% Tween 20 (PBST) and then incubated overnight at 4 °C in PBST with DAPI. The next day, the embryos were washed twice for 30 min each with PBST and processed for genotyping, as described above for the analysis

of cloacal morphology of *hox13* mutants, except that the head was removed for DNA extraction and the fins and trunk were retained for analysis. The embryos were imaged on a Zeiss LSM 800 confocal microscope to analyse *hoxd13a*$^{Tg(hsp70:tdTomato)}$ expression. The laser and filter settings were optimized individually for each stage and tissue type to be compared, and then these settings were kept constant across CsB-intact and CsB-deleted individuals. For the 19-somite stage, the cloaca and tailbud were imaged simultaneously as a single piece of trunk, but for the 72-hpf animals, the fins, cloaca and tail were dissected and imaged separately. Maximum projection images were produced from each scan and then exported as TIF files for analysis in ImageJ[61]. Each image was cropped to a specific region of interest (ROI) containing the specific expression domain, and ImageJ was used to measure the mean grey value for pixels in the region. The ROI size for each tissue was as follows: 75 µm × 75 µm for cloaca (19-somite and 72 hpf), 250 µm × 250 µm for 19-somite tailbud, 200 µm × 200 µm for 72-hpf tails and 100 µm × 250 µm for 72-hpf pectoral fins. For each tissue, the average mean grey value was calculated from CsB-intact individuals and used to normalize signal intensity values so that the average CsB-intact intensity for each tissue was equal to 1. The average relative intensities for CsB-intact and CsB-deleted tissues were compared using Welch's *t*-test in R (ref. 62).

Zebrafish *hox13* mutant lines

Frameshift loss-of-function alleles *hoxa13a*^{ch307}, *hoxa13b*^{ch308} and *hoxd13a*^{5bpins} were previously generated[11]. The zebrafish lines were propagated and maintained, as described in a previous study[63]. To generate compound *hox13* mutants, animals that were triple heterozygous for *hoxa13a*, *hoxa13b* and *hoxd13a* were intercrossed. The resulting larvae were fixed at 6 dpf in 4% PFA in PBS for 2 h at room temperature, with rocking agitation. After fixation, the larvae were rinsed twice for 5 min each in PBS with added 1% Triton X-100 (PBSX). To visualize the cloacal anatomy by labelling filamentous actin, the larvae were then incubated in PBSX with fluorophore-conjugated phalloidin (Sigma-Aldrich P1951; phalloidin-tetramethylrhodamine B isothiocyanate) added to a final concentration of 5 U ml^{-1} overnight at 4 °C, with rocking agitation. The larvae were then rinsed twice with PBSX for 1 h each.

For genotyping, the phalloidin-labelled larvae were cut in half, separating the head, yolk and pectoral fins from the cloaca and tail. The head half was used for genotyping, and the tail half was stored at 4 °C for later analysis. DNA was extracted from the head half by digesting tissue in proteinase K diluted to 1 mg ml^{-1} in 20 µl of 1× PCR buffer (10 mM Tris-HCl, 50 mM KCl and 1.5 mM $MgCl_2$) for 1 h at 55 °C, followed by heat inactivation at 80 °C for 20 min. The digested tissue was then subjected to brief vortexing, and then 1 µl was used directly as template for genotyping PCR, with primers listed in Supplementary Table 2. For thermocycling, after an initial step at 94 °C for 2 min, reactions were cycled 40 times (15 s at 94 °C, 15 s at 58 °C and 20 s at 72 °C) and finished with 5 min at 72 °C. The PCR products were then heteroduplexed on a thermocycler by heating to 95 °C for 10 min and then gradually cooled by 1 °C every 10 s until a final temperature of 4 °C was reached. Heteroduplexed PCR amplicons were then run on a high-percentage agarose gel to determine the genotype by product size.

To analyse cloacal morphology, fixed phalloidin-labelled tails were imaged using a Zeiss LSM 800 confocal microscope. After acquiring a full confocal stack through the cloacal region, a midline frame that demonstrated the hindgut and pronephric duct morphology was selected. In a separate set of quantifications, juveniles were photographed using a Leica M205 FCA stereotype microscope, PLANAPO 1.0× zoom lens and Leica MC170 HD camera. Using the pencil tool in Illustrator, we traced the internal lumen of the hindgut and pronephric duct complex from the level of the proximal end of the median fin fold to the terminal exit. A perpendicular line was then drawn to measure the width of the complex. The lengths of these lines were measured using Illustrator and were used for statistical analysis.

Mutant mouse stocks

The following mouse lines used in this study were previously reported: *Inv(Itga6-nsi)d11lac*[31], *Inv(Itga6–attP)* and *tgBAC(HoxD)*[30], *Del(HoxD)*[32] and *Del(Atf2–SB1)*, *Del(SB1–Rel5)* and *Del(Rel5–Rel1)*[2].

Whole-mount in situ hybridization

The zebrafish and mouse antisense probes used in this study are listed in Supplementary Data 2 and 3, respectively. For zebrafish, WISH was performed, as described[56], at 58 °C for all riboprobes (hybridization temperature and saline–sodium citrate washes). Whole-mount embryos were photographed using a compound microscope (SZX10; Olympus) equipped with a Nomarski optics and a digital camera (DP22; Olympus). Genotyping of individual embryos was performed after photographic documentation using the primers listed in Supplementary Table 2. Wild-type and mutant embryos originated from the same clutch of eggs produced by heterozygote crosses and underwent WISH in the same well. Details on the number of embryos per experiment and genotype are provided in Supplementary Table 3. Murine urogenital systems were isolated from E18.5 embryos and processed following a previously reported WISH procedure[64], with some specific adjustments. For proteinase K treatment, urogenital systems were incubated for 20 min in proteinase K diluted to 20 µg ml^{-1} in PBST. For the refixation step, a solution of 4% PFA containing 0.2% glutaraldehyde was used. The hybridization temperature was 69 °C, and the temperature of the post-hybridization washes was 65 °C. Staining was performed using BM-Purple (Roche; 11442074001) for approximately 4 h at room temperature.

Hybridization chain reaction

HCR in situ hybridization was performed, as previously described[65], with slight modifications. Embryos were fixed in 4% PFA in PBS at 4 °C overnight with rocking, washed three times for 5 min in PBS with 0.1% Tween (PBST) and then dehydrated in methanol washes (25%, 50% and 75% in PBST) for 3 min each, followed by five 10-min washes and one 50-min wash in 100% methanol. The embryos were stored at −20 °C in methanol for at least 48 h before starting the hybridization protocol. The embryos were rehydrated in methanol (75%, 50% and 25% methanol in PBST), washed twice with PBST and pre-hybridized in hybridization buffer (Molecular Instruments) at 37 °C for at least 1 h. The embryos were then incubated in 200 µl of a hybridization solution with *hoxd13a* probes (IDT oPools; Supplementary Table 4) at a concentration of approximately 65 nM each overnight at 37 °C. After 18–24 h in probe solution, the embryos were washed four times for 15 min each using a probe wash buffer (Molecular Instruments) at 37 °C. The embryos were then washed twice for 5 min each at room temperature with 5× SCCT on a rocker before incubation in amplification buffer (Molecular Instruments) for at least 1 h. The amplification solution with B2 546 amplifiers (Molecular Instruments) was prepared by heating 3 µl of hairpin 1 (3 µM) and 3 µl of hairpin 2 (3 µM) to 95 °C for 90 s, followed by snap-cooling. After 30 min, hairpins 1 and 2 were mixed and added to 200 µl of amplification buffer. The embryos were incubated in amplification solution overnight at room temperature on a rocker. After 18–24 h of incubation in amplification solution, the embryos were washed at least four times for 30 min each with 5× SCCT at room temperature on a rocker. The embryos were stored at 4 °C in 5× SCCT for 1 day until they were genotyped and mounted for confocal microscopy. Before genotyping, the embryos were washed in PBST with DAPI for 1 h. DNA was extracted from the dissected head of each embryo. The pectoral fins were then microdissected using tungsten needles and mounted in PBST for confocal imaging using an inverted Zeiss LSM 800. Wild-type fins were used to optimize the laser and filter settings, which were maintained across all samples during data collection. After image acquisition, post-processing was performed on the maximum-intensity projections of each sample to reduce non-specific

background signals. Specifically, the black value was changed from 0 to 50 uniformly for each image using the Zeiss Zen imaging software. These scans were then exported as TIF files for analysis in ImageJ[61]. The images were cropped to an ROI of 180 µm × 120 µm in size containing the *hoxd13a* fin expression domain. ImageJ was used to measure the mean grey value of ROI from each fin, and the average mean grey value was calculated from the wild-type fins. This average was used to normalize the signal intensity values such that wild-type fins had an average value of 1. The normalized relative intensities of wild-type and 5DOM deletion mutant fins were then compared using Welch's *t*-test in R[62].

Mouse genotyping

For extemporaneous genotyping, yolk sacs were collected and placed into 1.5-ml tubes containing rapid digestion buffer (10 mM EDTA (pH 8.0) and 0.1 mM NaOH) and then placed in a thermomixer at 95 °C for 10 min with shaking at 900 rpm. While the yolk sacs were incubating, the PCR master mix was prepared using Z-Taq (Takara; R006B) and primers (Supplementary Table 2) and aliquoted into PCR tubes. The tubes containing lysed yolk sacs were then placed on ice to cool briefly and quickly centrifuged at a high speed. The lysate (1 µl) was placed in the reaction tubes and cycled 32 times (2 s at 98 °C, 2 s at 55 °C and 15 s at 72 °C). The PCR reaction (20 µl) was loaded onto a 1.5% agarose gel, and electrophoresis was run at 120 V for 10 min. When samples could be kept for some time, a conventional genotyping protocol was applied using tail digestion buffer (10 mM Tris (pH 8.0), 25 mM EDTA (pH 8.0), 100 mM NaCl and 0.5% SDS) added to each yolk sac or tail clipping at 250 µl along with 4 µl of proteinase K at 20 mg ml^{-1} (Eurobio; GEXPRK01-15) and incubated overnight at 55 °C. The samples were incubated at 95 °C for 15 min to inactivate the proteinase K and stored at −20 °C until ready for genotyping. Genotyping primers (Supplementary Table 2) were combined with Taq polymerase (ProSpec; ENZ-308) in 25-µl reactions, cycled twice with Ta = 64 °C and then cycled 32 times with Ta = 62 °C.

Mouse RT–qPCR

UGSs were collected from E18.5 male embryos separately and placed in 1× diethyl pyrocarbonate–PBS on ice. A small portion of the remaining embryo was collected for genotyping. The UGSs were transferred into fresh 1× diethyl pyrocarbonate–PBS and then placed into RNAlater (Thermo Fisher Scientific; AM7020) for storage at −80 °C until processing. Batches of samples were processed in parallel to collect RNA using RNeasy extraction kits (QIAGEN; 74034). After isolating total RNA, first-strand complementary DNA (cDNA) was produced with SuperScript III VILO (Thermo Fisher Scientific; 11754-050) using approximately 500 ng of total RNA input. The cDNA was amplified with Promega GoTaq 2X SYBR Mix and quantified on a Bio-Rad CFX96 Real-Time System. Expression levels were determined by the difference between the cycle threshold (Ct) of the gene of interest (GOI) and the reference gene *Tbp*, calculated as dCt = Ct(GOI) − Ct(*Tbp*). They were normalized to 1 for each condition by subtracting each dCT from the mean dCT for each wild-type set. Finally, expression was evaluated by the power 2 minus this normalized dCT. Supplementary Table 2 contains the primer sequences used for quantification. RT–qPCR measurements were taken from distinct embryos. Box plots for expression changes and two-tailed unequal variance *t*-tests were produced in DataGraph 4.6.1. The boxes represent the IQR, with the lower and upper hinges denoting the first and third quartiles (25th and 75th percentiles). Whiskers extend from the hinges to the furthest data points within 1.5 times the IQR. The upper whisker reaches the largest value within this range, whereas the lower whisker extends to the smallest value within 1.5 times the IQR from the hinge.

Mouse RNA-seq

E18.5 male and female UGSs were collected by means of dissection separating the bladder from the UGS, including the proximal urethra in males and the vagina in females. Tissues were stored in RNAlater (Thermo Fisher Scientific; AM7020) and processed in parallel using RNeasy extraction kits (QIAGEN; 74034). RNA quality was assessed using an Agilent Bioanalyzer 2100 with RNA integrity number scores greater than 9.5. RNA sequencing libraries were prepared at the University of Geneva Genomics Platform using Illumina TruSeq Stranded Total RNA with Ribo-Zero Gold Ribo-deleted RNA kits to produce strand-specific 100-bp single-end reads on an Illumina HiSeq 2000. Raw RNA-seq reads were processed with Cutadapt v.4.1 (-a GATCGGAAGAGCACACGTCTGAACTCCAGTCAC -q 30 -m 15)[66] to remove TruSeq adapters and bad-quality bases. Filtered reads were mapped to the mouse genome mm39 using STAR v.2.7.10a[67] using ENCODE parameters with a custom gtf file[68] on the basis of Ensembl version 108. This custom GTF file was obtained by removing readthrough transcripts and all non-coding transcripts from a protein-coding gene. Fragments per kilobase of transcript per million mapped read values were evaluated using Cufflinks v.2.2.1 (refs. 69,70) with the options --max-bundle-length 10000000 --multiread-correct --library-type 'fr-firststrand' -b mm10.fa --no-effective-length-correction -M MTmouse.gtf -G. Box plots depicting expression levels in distinct embryos were generated using the same methodology as that used for RT–qPCR.

ATAC-seq

Mouse and fish tissues were isolated and placed into 1× PBS containing 10% fetal calf serum on ice. Collagenase (Sigma-Aldrich; C9697) was added to 50 µg ml^{-1} and incubated at 37 °C for 20 min with shaking at 900 rpm. Cells were washed three times in 1× PBS. The number of cells was counted, and viability was confirmed to be greater than 90%. An input of 50,000 cells was processed according to a previous description[36]. Sequencing was performed on École Polytechnique Fédérale de Lausanne (EPFL) Gene Expression Core Facility (GECF) using an Illumina NextSeq 500. We analysed in a manner similar to a previous study[71]. Raw ATAC-seq paired-end reads were processed with Cutadapt v.4.1 (-a CTGTCTCTTATACACATCTCCGAGCCCACGAGAC -A CTGTCTCTTATACACATCTGACGCTGCCGACGA -q 30 -m 15)[66] to remove Nextera adapters and bad-quality bases. Filtered reads were mapped on mm39 for mouse samples and danRer11 in which alternative contigs were removed for fish samples using Bowtie 2 v.2.4.5 (ref. 72) with the following parameters: --very-sensitive --no-unal --no-mixed --no-discordant --dovetail -X 1000. Only pairs mapping concordantly outside of mitochondria were kept (Samtools v.1.16.1) (ref. 73). The PCR duplicates were removed using Picard v.3.0.0 (http://broadinstitute.github.io/picard/index.html). The BAM files were converted to BED using bedtools v.2.30.0 (ref. 74). Peaks were called, and coverage was generated by MACS2 v.2.2.7.1 with --nomodel --keep-dup all --shift -100 --extsize 200 --call-summits -B. Coverages were normalized to million mapped reads.

ChIP–seq

Male UGSs were isolated and placed into 1× PBS containing 10% fetal calf serum on ice. ChIP–seq experiments were performed, as previously described[75]. Briefly, they were fixed for 10 min in 1% formaldehyde at room temperature, and the crosslinking reaction was quenched with glycine. Subsequently, nuclei were extracted, and chromatin was sheared using a water-bath sonicator (Covaris E220evolution ultrasonicator). Immunoprecipitation was performed using the following anti-H3K27ac (Abcam; ab4729) or anti-H3K27me3 (Merck Millipore; 07–449). Libraries were prepared using the TruSeq protocol and sequenced on an Illumina HiSeq 4000 (100-bp single-end reads) according to the manufacturer's instructions. CTCF was reanalysed using datasets from previous studies[43,71]. The accession numbers are listed in Supplementary Table 5. Raw ChIP–seq single-end or paired-end reads were processed using Cutadapt v.4.1 (-a GATCGGAAGAGCACACGTCTGAACTCCAGTCAC for single-end reads and -a CTGTCTCTTATACACATCTCCGAGCCCACGAGAC -A CTGTCTCTTATACACATCTGACGCTGCCGACGA -q 30 -m 15)[66] to remove TruSeq or Nextera adapters and bad-quality bases. Filtered reads were mapped on

mm39 for mouse samples and danRer11 in which alternative contigs were removed for reanalysis of fish samples using Bowtie 2 v.2.4.5 (ref. 72) with the default parameters. Only alignments with a mapping quality above 30 were kept (Samtools v.1.16.1)[73]. Peaks were called, and coverage was generated by MACS2 v.2.2.7.1 with --call-summits -B (and --nomodel --extsize 200 for single-end reads). Coverages were normalized to million mapped reads/pairs.

Mouse enhancer–reporter assay

Transgenic embryos were generated, as described[33]. Primers were designed to amplify genomic DNA from the region around the observed ATAC and H3K27Ac peaks (Supplementary Table 5). These primers included extra restriction sites for either *XhoI* or *SalI* at the 5′ ends. The PCR fragments were cleaned using a QIAGEN Gel Extraction Kit (28704). The PCR fragment and the pSKlacZ reporter construct (GenBank X52326.1)[75] were digested with *XhoI* or *SalI* and ligated together using the Promega 2X Rapid Ligation kit (C6711). Sanger sequencing confirmed that the correct sequences were inserted upstream of the promoter. Maxipreps of the plasmid were prepared and eluted in 1× IDTE (11-05-01-13). Pro-nuclear injections were performed, and embryos were collected at approximately E18.5 and stained for *lacZ*. UGSs were collected from E18.5 embryos in ice-cold 1× PBS in a 12-well plate. All steps were performed with gentle shaking on a rocker plate at room temperature. Tissues were fixed for 5 min at room temperature in freshly prepared 4% PFA. After fixing, the tissues were washed three times in 2 mM $MgCl_2$, 0.01% sodium deoxycholate, 0.02% Nonidet P-40 and 1× PBS for 20 min at room temperature. The wash solution was replaced with β-galactosidase staining solution (5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, 2 mM $MgCl_2$ hexahydrate, 0.01% sodium deoxycholate, 0.02% Nonidet P-40, 1 mg ml^{-1} of β-galactosidase and 1× PBS) for overnight incubation with the plate wrapped in aluminium foil to protect from light. The tissues were then washed three times in 1× PBS and fixed in 4% PFA for long-term storage. Images of embryos were collected using an Olympus DP74 camera mounted on an Olympus MVX10 microscope using Olympus cellSens Standard 2.1 software.

Mouse capture Hi-C sequencing

E18.5 male UGSs were collected, and collagenase-treated samples were crosslinked with 1% formaldehyde (Thermo Fisher Scientific; 28908) for 10 min at room temperature and stored at −80 °C until further processing, as previously described[76]. The SureSelectXT RNA probe design used for capturing DNA was performed using the SureDesign online tool by Agilent. Probes cover the region chr. 2: 72240000–76840000 (mm9) producing twice the coverage, with moderately stringent masking and balanced boosting. DNA fragments were sequenced on an Illumina HiSeq 4000 and processed with HiCUP v.0.9.2 on mm39 with --re1 ^GATC[77], Bowtie 2 v.2.4.5 (ref. 72) and Samtools v.1.16.1 (ref. 73). The output BAM was converted to a pre-juicer medium format with hic2juicer from HiCUP. The pairs with both mates on chr. 2: 72233000–76832000 were selected, sorted and loaded into a 10-kb bin matrix with cooler v.0.8.11 (ref. 78). The final matrix was balanced with the option --cis-only. TADs were computed using HiCExplorer hicFindTADs v.3.7.2 (refs. 79,80) with --correctForMultipleTesting fdr --minDepth 120000 --maxDepth 240000 --step 240000 --minBoundaryDistance 250000. Data were plotted on mm39 (chr. 2: 73600000–75550000).

Zebrafish Hi-C sequencing

The HiC profiles were derived from a reanalysis of data from previous studies[43,81]. The accession numbers are listed in Supplementary Table 5. Reads were mapped on danRer11 in which alternative contigs were removed, and no selection of reads were performed. Valid pairs were loaded into a 10-kb bins matrix. TAD calling parameters were adapted to the smaller size of the genome: --chromosomes "chr9" --correctForMultipleTesting fdr --minDepth 35000 --maxDepth 70000 --step 70000 --minBoundaryDistance 50000. Data were plotted on danRer11 (chr. 9: 1650000–2400000) and on an inverted *x* axis.

CUT&RUN

Zebrafish samples were processed using a final concentration of 0.02% digitonin (Apollo; APOBID3301). Approximately 0.5×10^6 cells were incubated with 0.1 µg (100 µl)$^{-1}$ of anti-H3K27ac antibody (Abcam; Ab4729) or 0.5 µg (100 µl)$^{-1}$ of anti-H3K27me3 (Merck Millipore; 07-449) in digitonin wash buffer at 4 °C. The protein A–micrococcal nuclease was kindly provided by the Henikoff Lab (batch 6) and added at 0.5 µl (100 µl)$^{-1}$ in digitonin wash buffer. Cells were digested in high-calcium buffer and released for 30 min at 37 °C. 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. The DNA was amplified for 14 cycles. Post-amplified DNA was cleaned and size selected using 1:1 ratio of DNA:AMPure SPRI beads (A63881) followed by an extra 1:1 wash and size selection with HXB. HXB is equal parts 40% polyethylene glycol 8,000 (Thermo Fisher Scientific; FIBBP233) and 5 M NaCl. Sequencing was performed at EPFL GECF on an Illumina HiSeq 4000. Raw CUT&RUN paired-end reads were processed with Cutadapt v.4.1 (-a GATCGGAAGAGCACACGTCTGAACTCCAGTCAC -A GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -q 30 -m 15) to remove TruSeq adapters and bad-quality bases[66]. Filtered reads were mapped on danRer11, in which alternative contigs were removed with Bowtie 2 v.2.4.5 (ref. 72) with the following parameters: --very-sensitive --no-unal --no-mixed --no-discordant --dovetail -X 1000. Only alignments with mapping quality above 30 were kept (Samtools v1.16.1) (ref. 73). PCR duplicates were removed by Picard v.3.0.0 (http://broadinstitute.github.io/picard/index.html). BAM files were converted to BED with bedtools v.2.30.0 (ref. 74). Peaks were called, and coverage was generated by MACS2 v.2.2.7.1 with --nomodel --keep-dup all --shift -100 --extsize 200 --call-summits -B. Coverages were normalized to million mapped reads.

Analyses of conserved sequences

Annotation of orthologous domains was performed using transcription start sites of orthologous genes, as reported in Supplementary Table 6. To identify conserved sequences between mouse and zebrafish, a pairwise alignment was done between the mouse genomic region chr. 2: 73600000–75550000 (mm39) and the zebrafish orthologous region chr. 9: 1650000–2400000 (danRer11) using discontinuous megablast. To reduce false positives, only reciprocal hits were considered. To display multispecies conservation levels, multiple alignment format files were generated between chr. 2 of the mouse genome (mm39) and contig chrUn_DS181389v1 of the platypus genome (ornAna2), chr. 7 of the chicken genome (galGal6), contig chrUn_GL343356 of the lizard genome (anoCar2), chr. 9 of the frog genome (xenTro10), contig JH127184 of the coelacanth genome (latCha1), chr. 9 of the zebrafish genome (danRer11), chr. 1 of the fugu genome (fr3) and the whole lamprey genome (petMar3). Details for the multiple alignment format generation are available on the GitHub repository (https://github.com/AurelieHintermann/HintermannBoltHawkinsEtAl2025; ref. 82). To facilitate visualization, a horizontal line was plotted for each species on each region.

Whole-genome alignments

Whole-genome alignments were performed using Progressive Cactus v.2.6.7 (ref. 83). The cactus command was used with default parameters to obtain the hierarchical alignment format. The hierarchical alignment was then projected on either zebrafish chr. 9 or mouse chr. 2 with cactus-hal2maf[84] using --chunkSize 500000 and --noAncestor. The genome assemblies are listed in Supplementary Table 6.

Single-cell assay for transposase-accessible chromatin sequencing

The single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) bigwig files were downloaded from a previous study[37]

(Gene Expression Omnibus (GEO) GSE243256) where annotations were available. To annotate the cells from the cloaca, the raw matrix of single-cell RNA sequencing (scRNA-seq) from a previous study[38] was downloaded from GEO (GSE223922) and stored in a Seurat object. Only 12,424 cells obtained at 14 hpf were kept. The data were normalized, and 3,000 variable features were extracted. Data were scaled. Uniform manifold approximation and projection (UMAP) and *t*-distributed stochastic neighbour embedding projections were calculated using the first 50 principal components. In parallel, scATAC-seq fragments of cells corresponding to 14 hpf were extracted from the general fragment file provided in a previous study[38] on GEO (GSE243256). A new ArchR gene annotation was generated using the Lawson gtf v.4.3.2 (ref. 85) to match the scRNA-seq data from a previous study[37], and the selected fragments were loaded into an ArchRProject with this genome. Iterative latent semantic indexing was computed with COR-Cut-off of 0.5. The clustering of scRNA-seq was then transferred to scATAC-seq using AddGeneIntegrationMatrix. The profile of the 38 cells whose transferred cluster corresponds to cloaca (endo.31) was generated with getGroupBW.

scRNA-seq

The matrix of the scRNA-seq atlas was downloaded from GEO (GSE223922; ref. 37) and the table with metadata. The matrix was loaded into a Seurat object using Seurat v.4.3.0 (ref. 86) in R v.4.3.0. Cells attributed to the 'tissue.name' 'endoderm' were selected. Normalization and principal component analysis were performed, as described in a previous study[37]. UMAP was performed on the top 70 principal component analyses and 50 nearest neighbours. UMAP coordinates and *hox13* normalized expression of endoderm cells were exported to a file and plotted using ggplot2 v.3.4.4.

Software

The phylogenic tree was generated with http://timetree.org using the following species: *Mus musculus*, *Protopterus*, *D. rerio*, *Carcharhinus leucas*, *Petromyzon marinus* and *Branchiostoma lanceolatum* and subsequently edited using SeaView 4.7. Genomic tracks from next-generation sequencing were plotted using pyGenomeTracks 3.8 using custom gene annotations available at https://doi.org/10.5281/ZENODO.7510796 (ref. 68; mm39) and https://doi.org/10.5281/zenodo.10283273 (ref. 87; danRer11). RT–qPCR, RNA-seq and domain size quantifications were plotted in R using the ggplot package.

Ethical statement

All experiments involving mice were performed in agreement with the Swiss Law on Animal Protection (Loi sur la Protection des Animaux) under licence no. GE 81/14. For zebrafish, work was carried out either under a general licence of EPFL granted by the Service de la Consommation et des Affaires Vétérinaires of the Canton of Vaud, Switzerland (no. VD-H23) or was either agreed upon by the animal committees of Rutgers University under protocol no. 201702646 or under guidance of the Institutional Animal Care and Use Committee of Boston Children's Hospital.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw and processed datasets are available in the GEO repository under accession no. GSE250267.

Code availability

All scripts necessary to reproduce figures from raw data are available at GitHub (https://github.com/AurelieHintermann/HintermannBoltHawkinsEtAl2025)[82].

53. Westerfield, M. *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio)* (Univ. of Oregon, 2000).
54. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995).
55. Hoshijima, K. et al. Highly efficient CRISPR-Cas9-based methods for generating deletion mutations and F0 embryos that lack gene function in zebrafish. *Dev. Cell* **51**, 645–657 (2019).
56. Narayanan, R. & Oates, A. C. Detection of mRNA by whole mount in situ hybridization and DNA extraction for genotyping of zebrafish embryos. *Bio. Protoc.* **9**, e3193 (2019).
57. Auer, T. O., Duroure, K., De Cian, A., Concordet, J.-P. & Del Bene, F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* **24**, 142–153 (2014).
58. Kimura, Y., Hisano, Y., Kawahara, A. & Higashijima, S. Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. *Sci. Rep.* **4**, 6545 (2014).
59. Tanaka, Y. et al. Anterior-posterior constraint on Hedgehog signaling by *hhip* in teleost fin elaboration. *Development* **151**, dev202526 (2024).
60. Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P. & Pruett-Miller, S. M. A survey of validation strategies for CRISPR-Cas9 editing. *Sci. Rep.* **8**, 888 (2018).
61. Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518–529 (2015).
62. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2018); https://www.r-project.org/.
63. Nüsslein-Volhard, C. & Dahm, R. R. *Zebrafish, A Practical Approach* (Oxford Univ. Press, 2002).
64. Woltering, J. M. et al. Axial patterning in snakes and caecilians: evidence for an alternative interpretation of the *Hox* code. *Dev. Biol.* **332**, 82–89 (2009).
65. Choi, H. M. T. et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).
66. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. Lopez-Delisle, L. Customized gtf file from Ensembl version 108 mm39. *Zenodo* https://doi.org/10.5281/ZENODO.7510796 (2023).
69. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
70. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. & Pachter, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* **12**, R22 (2011).
71. Hintermann, A. et al. Developmental and evolutionary comparative analysis of a regulatory landscape in mouse and chicken. *Development* **149**, dev200594 (2022).
72. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
73. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
74. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
75. Beccari, L. et al. A role for HOX13 proteins in the regulatory switch between TADs at the *HoxD* locus. *Genes Dev.* **30**, 1172–1186 (2016).
76. Yakushiji-Kaminatsui, N. et al. Similarities and differences in the regulation of *HoxD* genes during chick and mouse limb development. *PLoS Biol.* **16**, e3000004 (2018).
77. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* **4**, 1310 (2015).
78. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
79. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
80. Wolff, J. et al. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **48**, W177–W184 (2020).
81. Wike, C. L. et al. Chromatin architecture transitions from zebrafish sperm through early embryogenesis. *Genome Res.* **31**, 981–994 (2021).
82. Hintermann, A. & Delisle, L. AurelieHintermann/HintermannBoltHawkinsEtAl2025: publication final release. *Zenodo* https://doi.org/10.5281/ZENODO.16644564 (2025).
83. Armstrong, J. et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
84. Hickey, G., Paten, B., Earl, D., Zerbino, D. & Haussler, D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013).
85. Lawson, N. D. et al. An improved zebrafish transcriptome annotation for sensitive and comprehensive detection of cell type-specific genes. *eLife* **9**, e55792 (2020).
86. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
87. Hintermann, A. & Lopez-Delisle, L. Customized gtf file from Ensembl version 109 GRCz11 (danRer11). *Zenodo* https://doi.org/10.5281/zenodo.10283273 (2023).

Acknowledgements We thank colleagues of the Duboule laboratory for comments and discussions, the van der Goot laboratory for providing a fish cDNA library, M. Tanaka for the *hoxd10a* probe, J. Farrell for his advice and A. Oates for his support. Calculations were performed using the facilities of the Scientific IT & Application Support Center of EPFL. This study was supported in part using the resources and services of the GECF at the School of Life Sciences of EPFL, and mouse transgene injections were performed at the Transgenesis Core Facility of the University of Geneva, Medical School. We acknowledge the support of the Brinson Foundation (to N.H.S.). This study was supported by funds from EPFL, University of Geneva, Swiss National Research Fund (grant nos. 310030_196868 and CRSII5_189956), European Research Council grant Regul*Hox* (grant no. 588029) to D.D., NIH 1R01HD112906 to M.P.H. and NSF 2210072 to

T.N. The funders had no role in study design; data collection, analysis and interpretation; and writing of the paper.

Author contributions Conceptualization: A.H., C.C.B., D.D. Methodology: A.H., C.C.B., G.V., L.L.-D., B.M., M.B.H., M.M.R. Investigation: A.H., C.C.B., G.V., L.L.-D., S.G., P.B.G., B.M., T.N., T.A.M., M.B.H., M.M.R. Visualization: A.H., C.C.B., G.V., D.D., M.B.H., T.N., T.A.M. Funding acquisition: D.D., M.P.H., N.H.S. Project administration and supervision: A.H., C.C.B., D.D. Writing—original draft: A.H., C.C.B., D.D. Writing—review and editing: A.H., D.D., M.P.H., M.B.H., N.H.S.

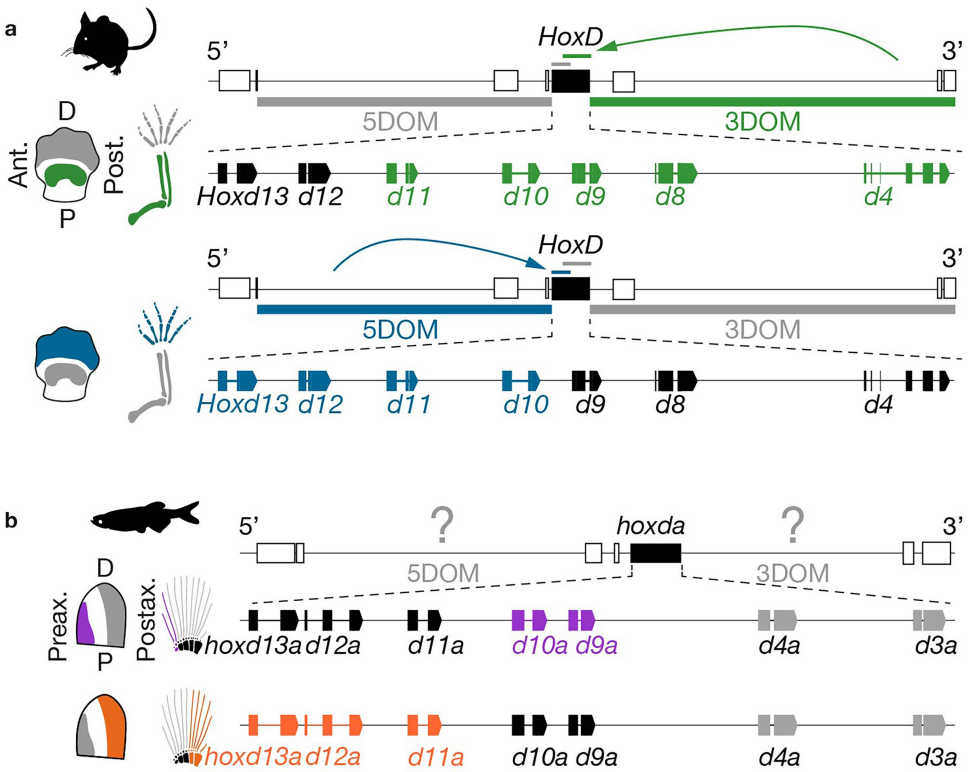
Competing interests The authors declare no competing interests.

Additional information

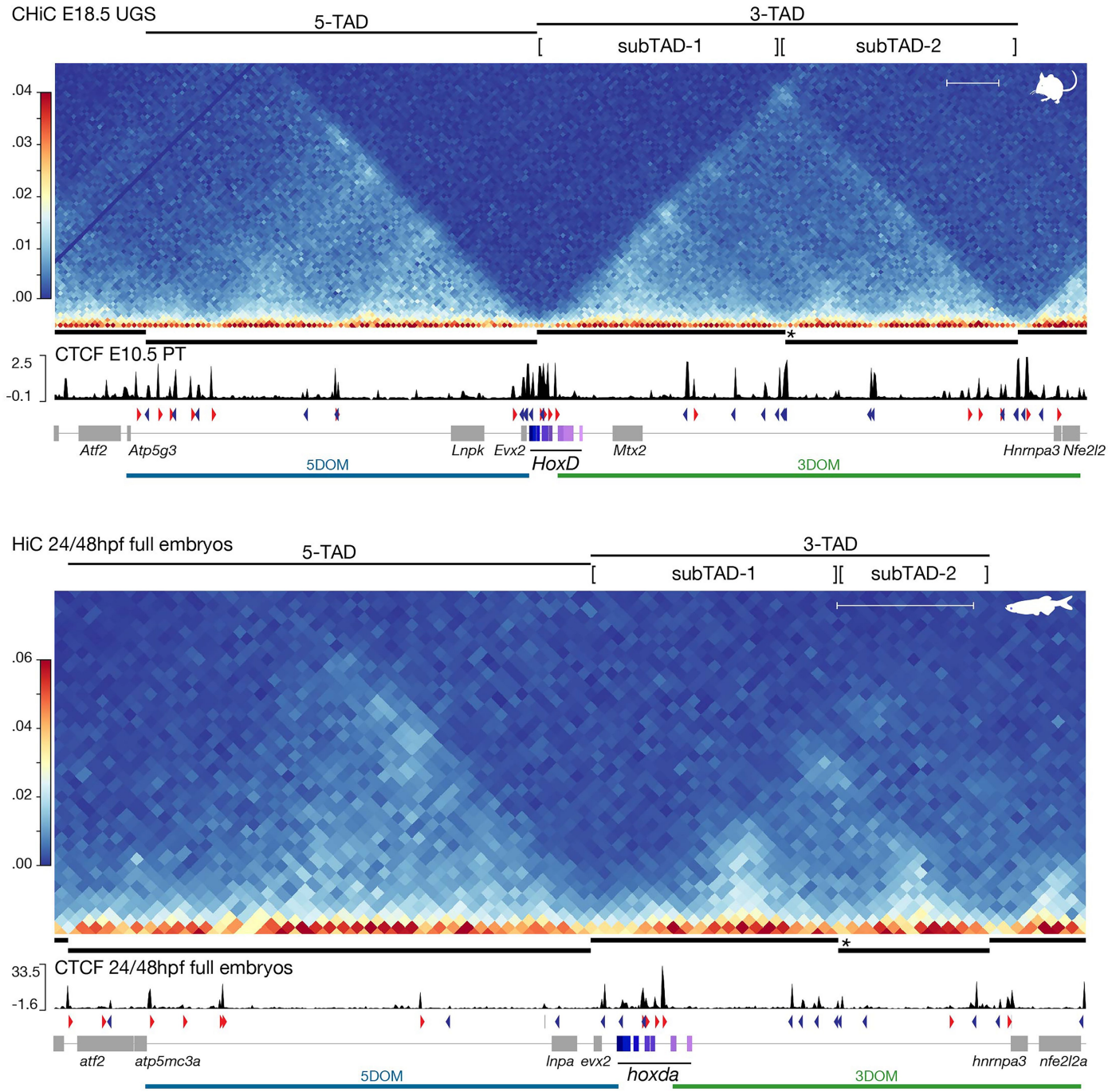
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09548-0.

Correspondence and requests for materials should be addressed to Denis Duboule.

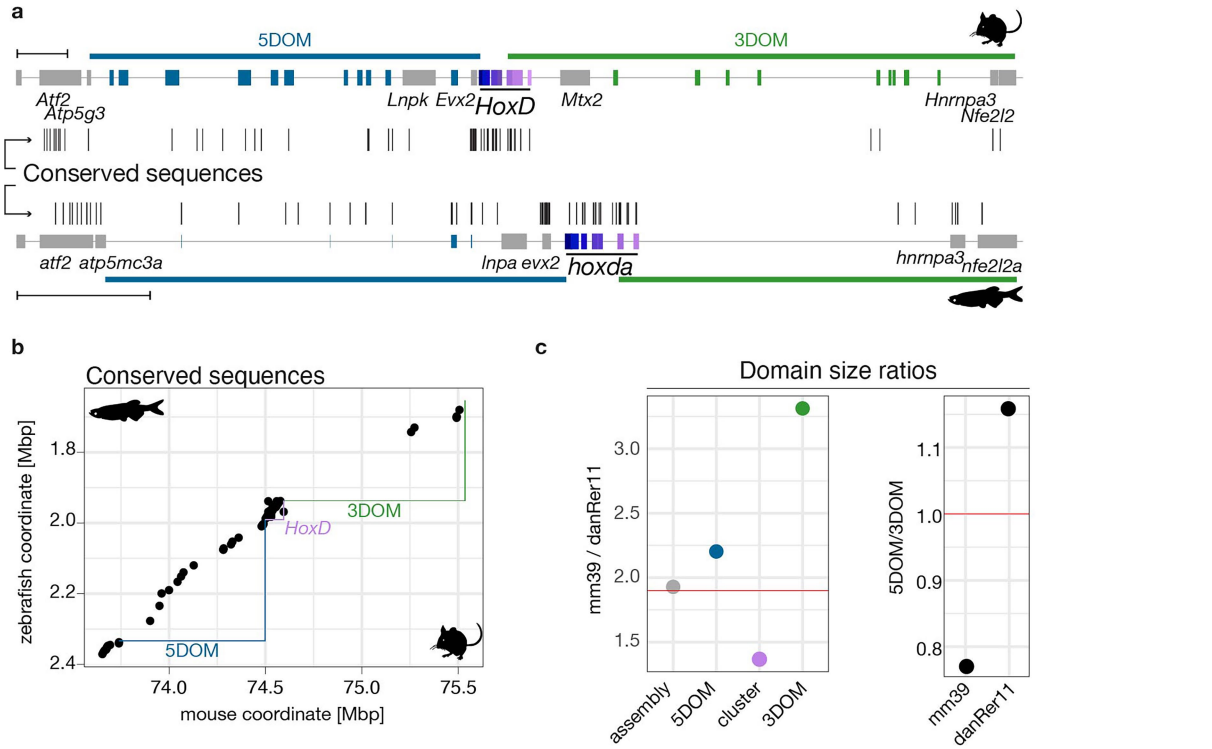
Peer review information *Nature* thanks Juan Tena and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



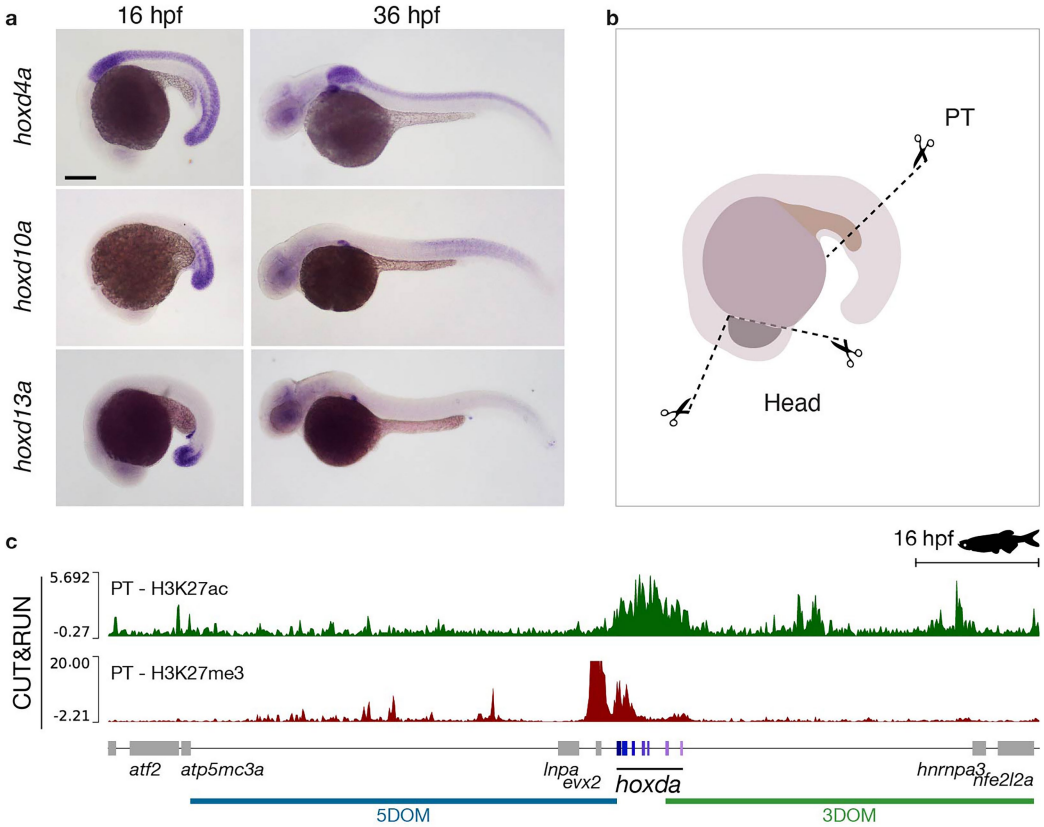
Extended Data Fig. 1 | Comparison of *HoxD* regulatory landscapes in mammals and fishes. a. *Hoxd* gene expression and regulation in mouse limb buds at E12.5. The *HoxD* cluster is flanked by two gene deserts, named according to their relative position (3′ or 5′) with respect to *Hoxd* gene orientation. The 3DOM regulatory landscape activates *Hoxd4* to *Hoxd11* in the proximal limb territory (green). The 5DOM activates *Hoxd10* to *Hoxd13* in the distal limb territory (blue). Schemes are based on ref. 15. **b**. Gene expression in fin buds at 40–60 hpf in the cognate zebrafish *hoxda* cluster. The fish cluster is also flanked by two gene deserts but their regulatory potentials are unknown (question marks). Fish *hoxd9a* to *hoxd11a* are expressed in the preaxial fin territory (purple) whereas *hoxd11a* to *hoxd13a* are expressed in a postaxial domain (orange). Schemes are inspired and WISH taken from[13,14,16].



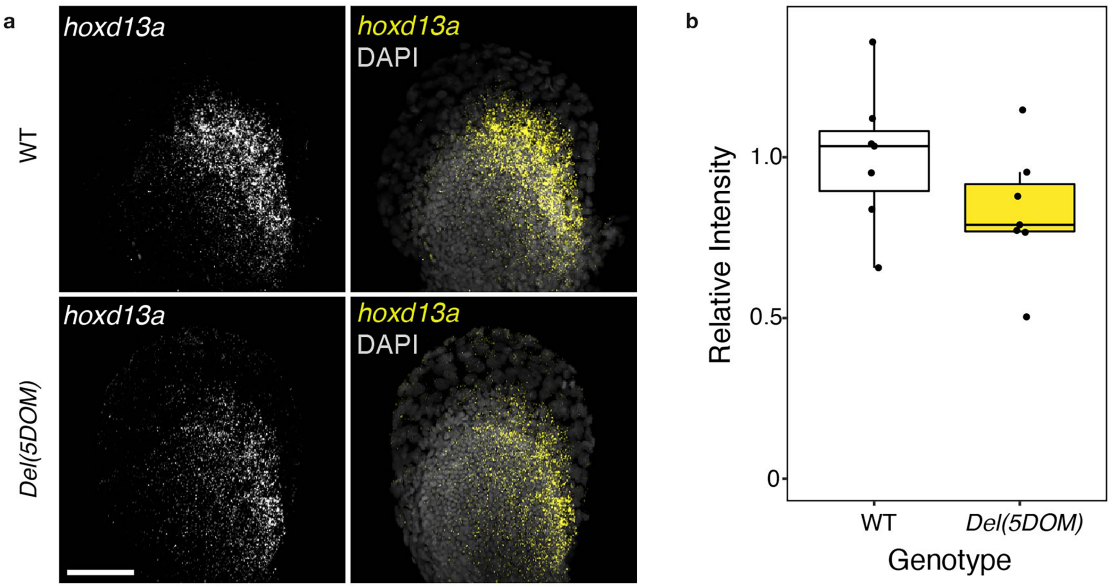
Extended Data Fig. 2 | 3D chromatin conformation at the mouse and fish *HoxD* loci. Contact frequency heatmaps at the mouse *HoxD* (E18.5 male UGS, one representative replicate out of two) and fish *hoxda* (24 hpf and 48 hpf total embryos[43,81]) loci (top and bottom, respectively). The similarities in the constitutive structural organization of the mouse and the fish loci are underlined either by the position and relative extents of TADs (thick black lines), the presence of a sub-TAD boundary within 3DOM (asterisk), as well as by the positions and orientation of CTCF binding sites (red and blue arrowheads). The fish CTCF profile is from 48hpf (GSM534493). *Hox* genes are in purple-scale rectangles and other genes are grey rectangles. Bin size is 10 kb. The scales on the *x* axes were adjusted to comparable sizes for ease of comparison, yet the fish locus is more compact. Scale bars in both cases: 100 kb.



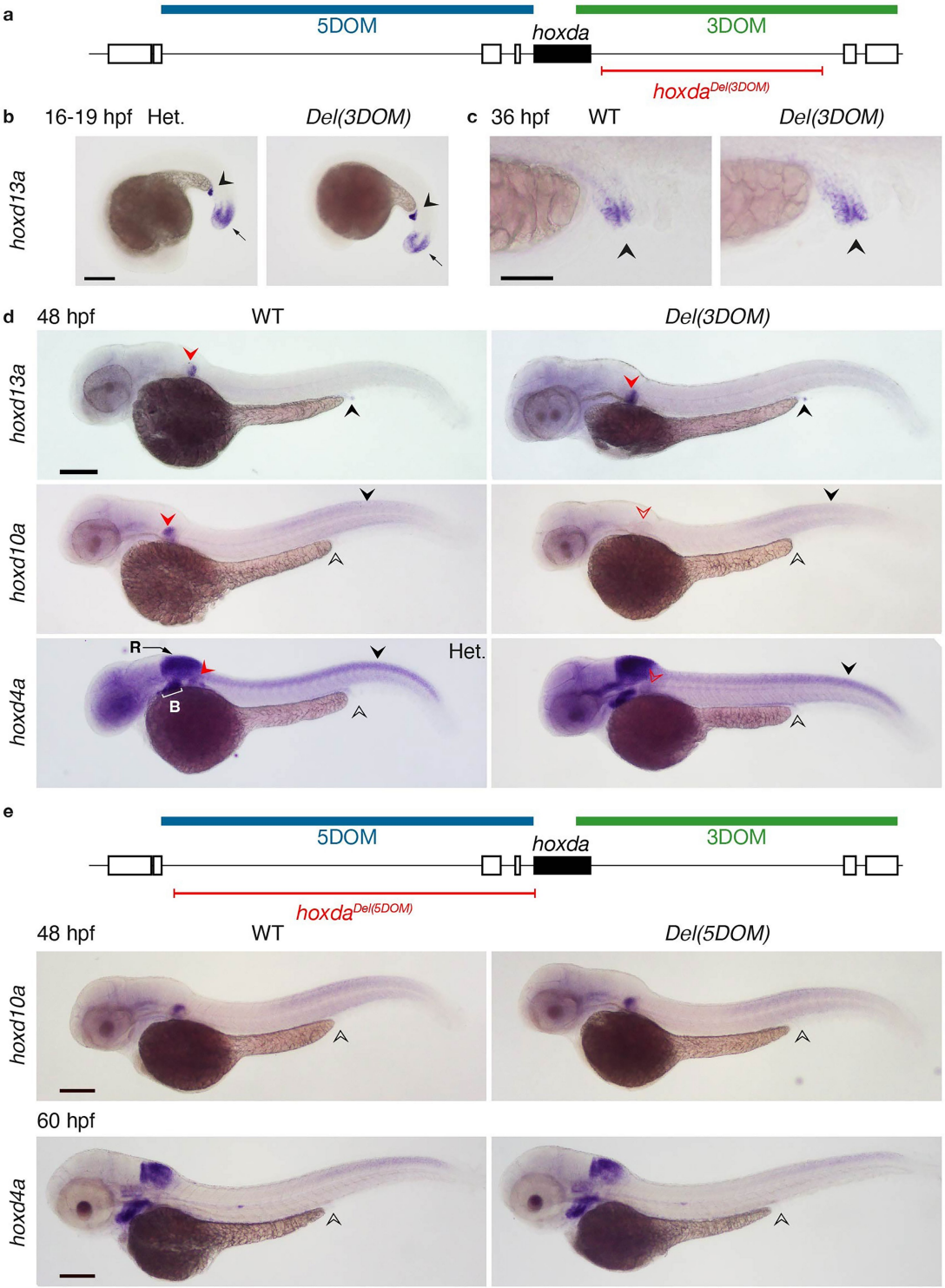
Extended Data Fig. 3 | The *HoxD* locus is part of a large syntenic interval. **a**. The mouse *HoxD* locus (mm39) is on top and the zebrafish *hoxda* locus (danRer11) is below. *Hox* genes are in purple-scale rectangles and annotated mouse enhancers are shown as either blue (5DOM) or green (3DOM) rectangles. Conserved sequences between the two gene deserts are shown as vertical black bars. Those conserved sequences overlapping with known murine enhancers were used to annotate the corresponding elements in zebrafish (blue rectangles). **b**. Synteny plot representing sequences conserved between the mouse and the zebrafish *HoxD* loci. On the *x* axis is the mouse locus (mm39, chr2:73600000-75670000) and on the *y* axis is the zebrafish locus (danRer11, chr9:1650000-2400000, inverted *y* axis). Despite a mouse locus that is in average 2.6 times larger than its zebrafish counterpart, the order of most conserved sequences is maintained, showing the absence of substantial genomic rearrangement at these gene deserts. **c**. Size comparisons between different regions of the zebrafish *hoxda* and the mouse *HoxD* loci. The left panel shows that the *Hox* clusters have maintained a similar size over time, while gene deserts have expanded in mouse and/or contracted in zebrafish. The right panel shows that the ratio between the sizes of 5DOM *versus* 3DOM is inverted in the two species.



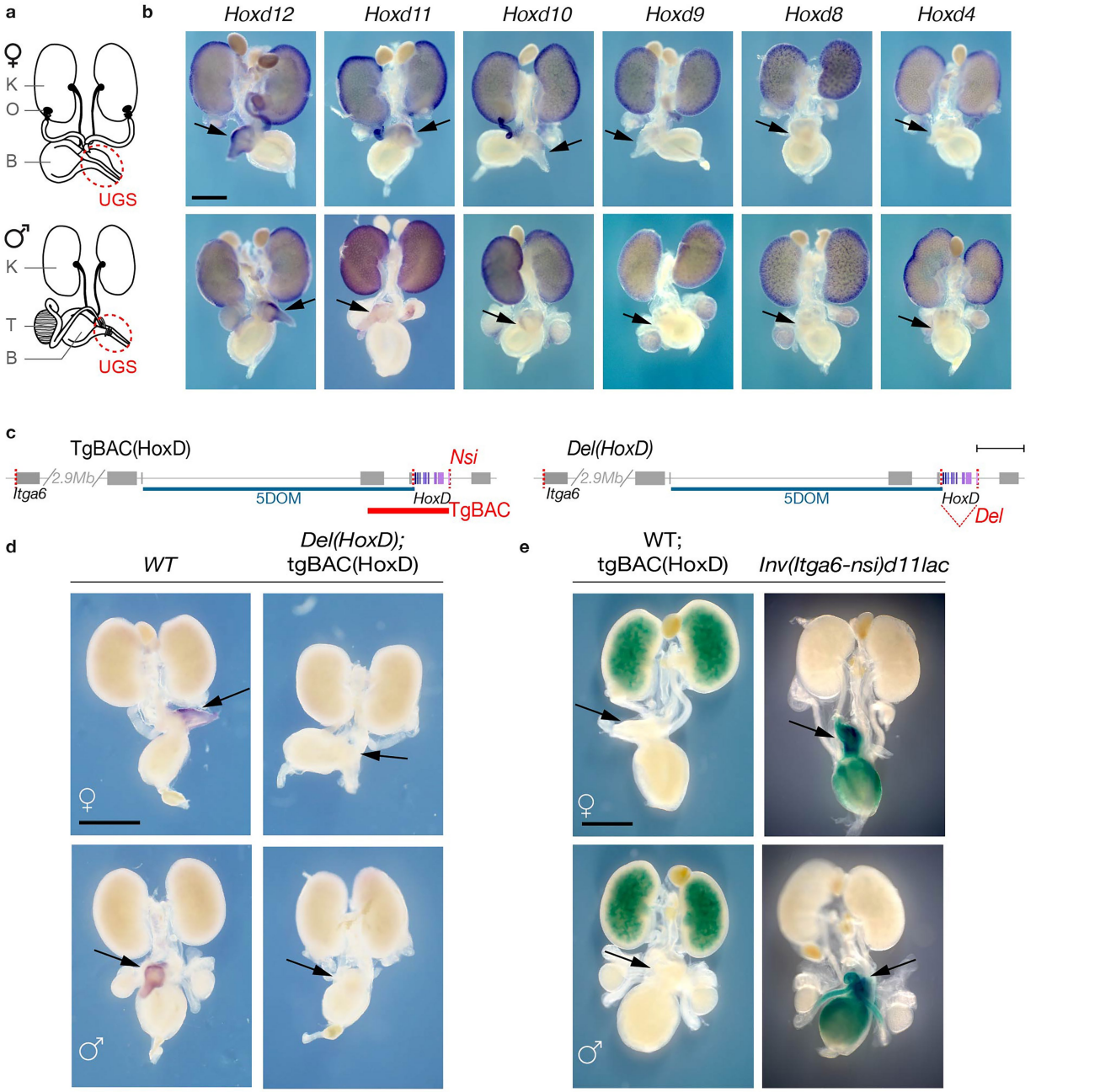
Extended Data Fig. 4 | Chromatin profiles in zebrafish embryos. **a**. Expression of *hoxd13a*, *hoxd10a* and *hoxd4a* in control zebrafish embryos by WISH. Stages are indicated on top of the panels. Scale bar: 200 µm. Each WISH experiment was repeated independently at least twice with similar results. **b**. Dissection plan used for panel (**c**). PT, posterior trunk. **c**. H3K27ac and H3K27me3 CUT&RUN profiles over the zebrafish *hoxda* locus in 16 hpf dissected posterior trunk cells (PT, blue, one representative replicate out of three). When compared to forebrain, the CUT&RUN profiles in posterior trunk cells show enrichment for H3K27ac (green coverage) on the central and anterior parts of the *hoxda* cluster, as well as on the 3DOM, while H3K27me3 (red coverage) is enriched on the posterior part and on *evx2*. H3K27ac: (n = 2 experiments), one representative example is shown; H3K27me3 (n = 1 experiment). Scale bar: 100 kb.



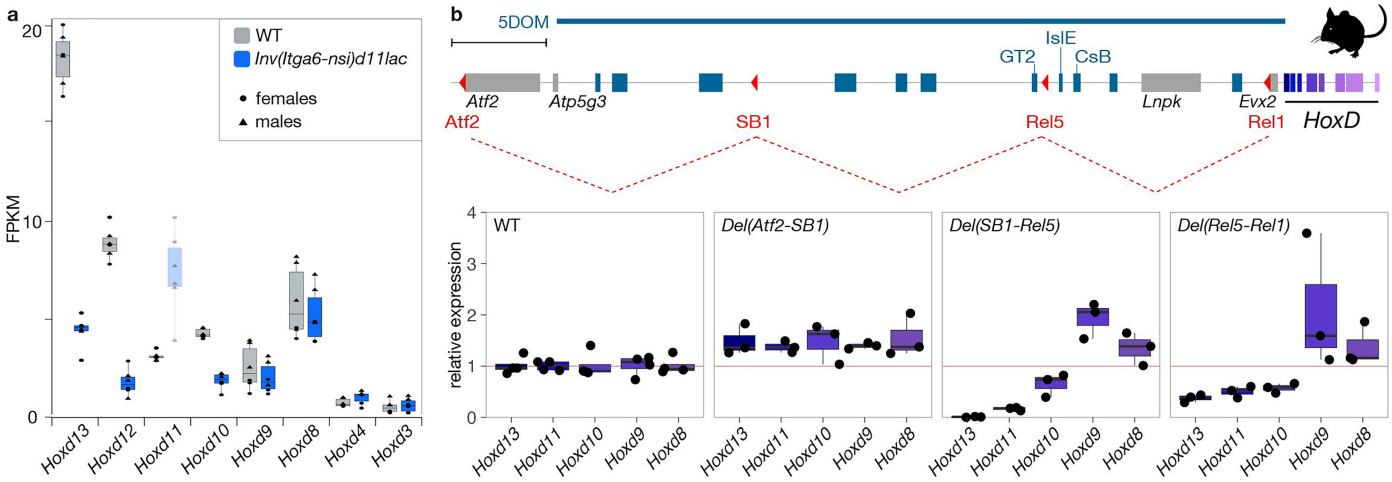
Extended Data Fig. 5 | Hybridization chain reaction (HCR) analysis of *hoxd13a* expression in the fins of control and 5DOM deletion animals. **a**. Maximum projection confocal images of 60 hpf pectoral fins from homozygous 5DOM deletion animals and wildtype siblings labeled for *hoxd13a* transcripts. Deletion fins express *hoxd13a* in the expected domain but at attenuated levels. **b**. Quantification of *hoxd13a* signal in wildtype and mutant fins (n = 7 each genotype). Relative intensity is normalized to the wildtype mean. While fin signal trended lower in mutant fins, this difference was not significant (Welch Two Sample t-test, $p = 0.1553$). The boxes represent the interquartile range (IQR), with the lower and upper hinges denoting the first and third quartiles (25^{th} and 75^{th} percentiles). Whiskers extend from the hinges to the furthest data points within 1.5 times the IQR. The upper whisker reaches the largest value within this range, while the lower whisker extends to the smallest value within 1.5 times the IQR from the hinge. Scale bar: 50 µm.



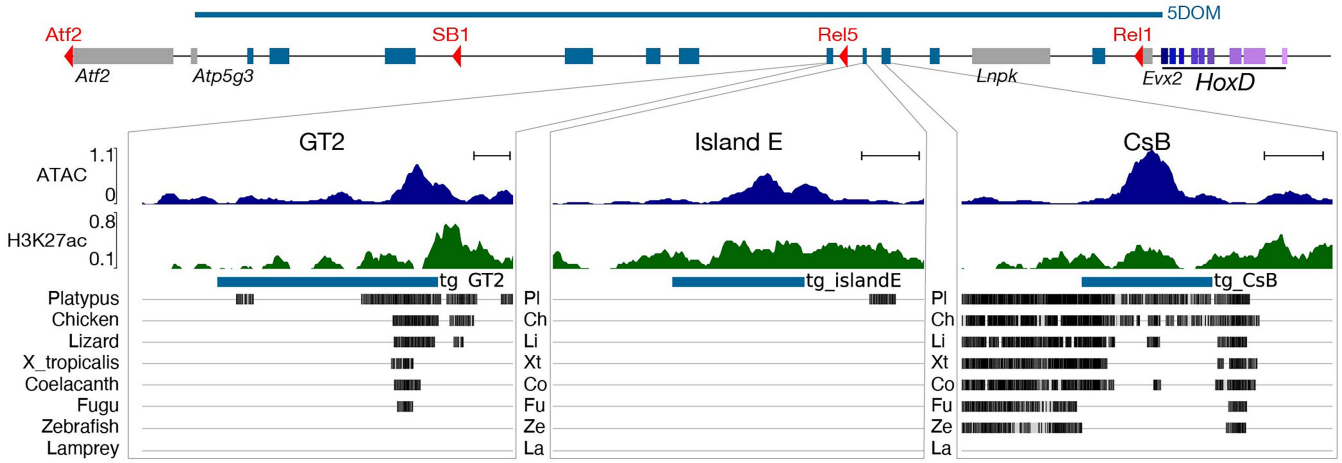
Extended Data Fig. 6 | WISH of *hoxd13a*, *hoxd10a* or *hoxd4a* in zebrafish embryos lacking either 3DOM (a-d), or 5DOM (e). a. Schematic of the 3DOM deletion. **b-d**. The genotypes (top) and genes analyzed (left) are shown as well as the stages (up left). Black arrowheads (empty for no expression and full for expression) indicate differential gene expression in the cloacal region, whereas red arrowheads (empty for no expression and full for expression) point to the pectoral fin buds. Control and homozygote mutant embryos are shown side by side for each condition, except for *hoxd4a* where a heterozygous (Het) mutant is shown. Wildtype and homozygous embryos originate from the same clutch of eggs and were processed together. Each experiment was done on two or more embryos except for *hoxd4a* in del(5DOM) at 48 hpf for which only one mutant embryo was obtained. **b, c**. In Del(3DOM) mutant embryos, *hoxd13a* transcripts in the cloaca are not affected. **d**. *Hoxd10a* and *hoxd4a* transcription is lost in fin buds. B, branchial arches; R, rhombomeres; Scale bars: 200 µm. **e**. In Del(5DOM) mutant embryos, *hoxd10a* (48 hpf) and *hoxd4a* (60 hpf) expressions are identical to that of wildtype embryos.



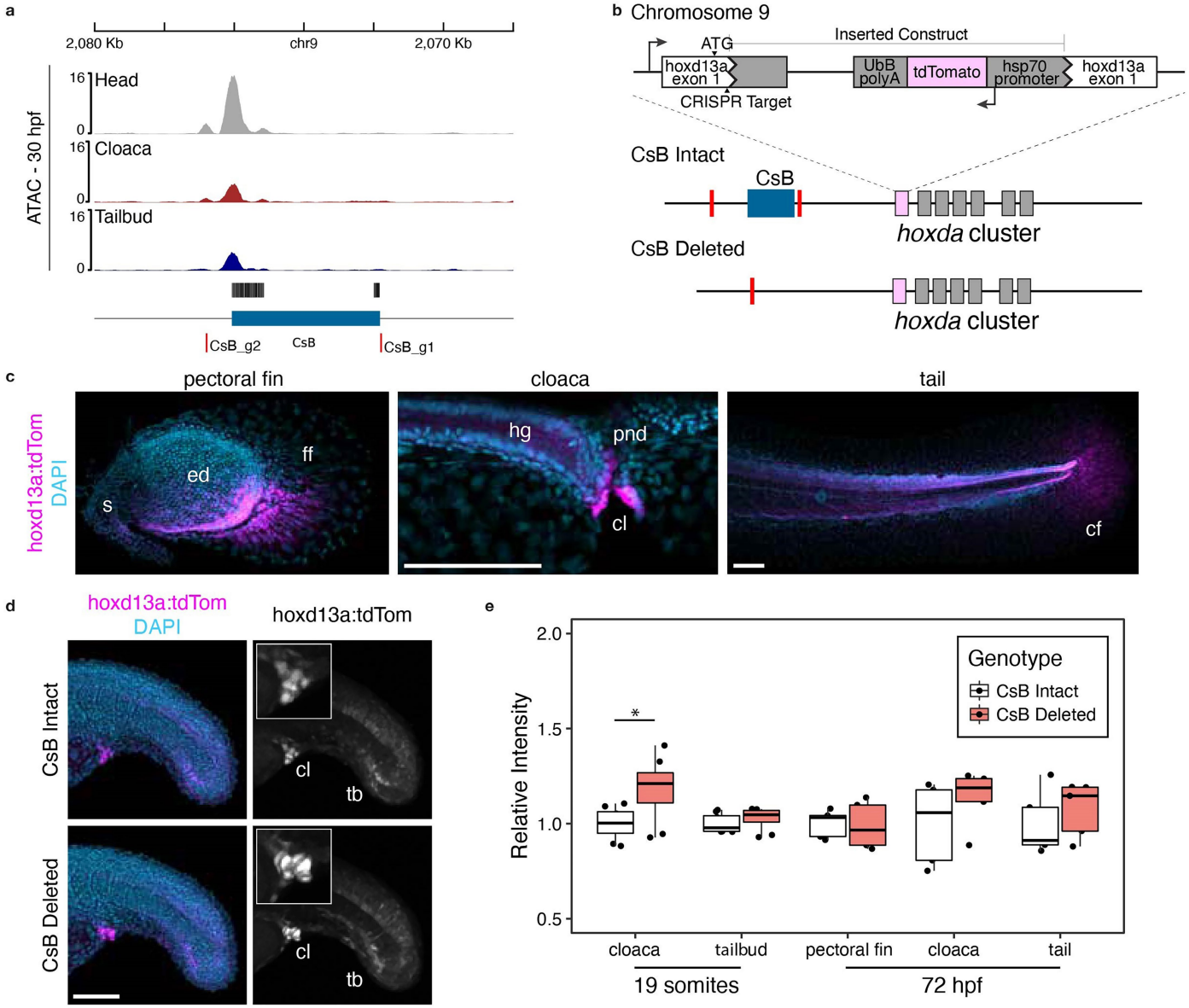
Extended Data Fig. 7 | *Hoxd* gene expression in the mouse urogenital system. **a**. Schematic representations of male and female urogenital systems. K: Kidney, B: Bladder, O: Ovary, T: Testis. The urogenital sinus (UGS) is indicated with a red circle. **b**. WISH of *Hoxd* genes in representative female and male urogenital systems. All *Hoxd* genes are expressed in anterior portions of the UGS including kidneys, the uterus, deferens ducts and the bladder, except *Hoxd13* transcripts, which are restricted to the UGS (see Fig. 3b). The UGS is indicated with a black arrow. **c**. Schematic representation of two *HoxD* genomic configurations, the first one is a random integration of a transgene carrying the same *HoxD* transgene plus some flanking sequences in 5' (thick red bar), whereas the second one is a deletion of the entire *HoxD* cluster. *Hox* genes as in shades of purple and the deletion breakpoints are shown as vertical dashed red lines. Scale bar: 100 kb. **d**. WISH of *Hoxd13* in UGS from a transgenic *HoxD* cluster (TgBAC), while lacking both endogenous copies of the *HoxD* cluster. Expression is not detected from the transgenic cluster. **e**. Likewise, beta-gal staining of UGS transgenic for the *HoxD* cluster containing a *LacZ* reporter displays no activity in the UGS. By contrast, *LacZ* staining of mutant *Inv(nsi-Itga6)d11lac* embryos, which also includes a *lacZ* reporter confirms that 5DOM is necessary and sufficient to drive expression in the UGS. Each WISH experiment and beta-gal staining was repeated independently at least twice with consistent results. Scale bar: 1 mm.



Extended Data Fig. 8 | Regulatory potential of sub-regions within 5DOM. **a**. RNA-seq FPKM values for various mouse *Hoxd* genes in E18.5 UGS obtained from either wildtype or *Inv(nsi-Itga6)d11lac* mutant embryos (schematic in Fig. 3c). Data are shown separately for females (n = 3, dots) and males (n = 3, triangles). Drastic decreases are observed for *Hoxd10*, *Hoxd12* and *Hoxd13* when 5DOM is disconnected from the *HoxD* cluster. *Hoxd11* could not be assessed due to the presence of a transgenic copy of this gene in the *LacZ* reporter cassette. **b**. On top is a scheme of the 5DOM regulatory landscape on mm39 with *Hox* genes in purple. Blue rectangles indicate previously described 5DOM enhancers. The red arrowheads delimit the serial deletion breakpoints. The three consecutive deletions are depicted by red dashed lines. Below are RT-qPCR quantifications of expression levels relative to wildtype (n = 4) in three mutant lines carrying serial deletions of 5DOM (n = 3). The horizontal red line represents the value of 1 for reference. Severe reductions are observed for both the *Del(Rel5-SB1)* and *Del(Rel1-Rel5)* conditions, unlike in the *Del(SB1-Atf2)* deletion. Scale bar: 100 kb. **a, b**. The boxes represent the interquartile range (IQR), with the lower and upper hinges denoting the first and third quartiles (25^{th} and 75^{th} percentiles). Whiskers extend from the hinges to the furthest data points within 1.5 times the IQR. The upper whisker reaches the largest value within this range, while the lower whisker extends to the smallest value within 1.5 times the IQR from the hinge.



Extended Data Fig. 9 | Sequence conservation of the GT2, islE and CsB UGS enhancers in vertebrates. All three sequences are comprised in the box highlighted in Fig. 4a. The ATAC-seq and H3K27ac ChIP-seq profiles are shown with, below, their sequence conservation from fishes and mammals. The thick blue lines below the H3k27ac profiles indicate the extent of the transgenes assayed in Fig. 4. Scale bar: 1 kb.



Extended Data Fig. 10 | Deletion of the conserved CsB sequence in *cis* with an *hoxd13a* reporter. a. Chromatin accessibility profiling of the CsB element in different tissues during zebrafish development. The extent of the sequence is shown below, with the red bars indicating CRISPR-Cas9 guide positions for the targeted deletion. **b**. Schematic of the *hoxd13a*$^{Tg(hsp70:tdTomato)}$ (*hoxd13a:tdTom*) endogenous reporter in CsB-intact and CsB-deleted chromosomes. **c**. Endogenous reporter activity in 72 hpf zebrafish recapitulates the *hoxd13a* expression pattern. Anterior is to the top and distal to the right in the fin panel, and anterior is to the left and dorsal to the top in the cloaca and tail panels. **d**. At the 19-somite stage CsB-deleted embryos show slightly increased reporter activity in the cloacal region relative to CsB-intact controls. **e**. Quantification of signal intensity in CsB-intact and CsB-deleted embryos at 19-somite (n = 8 each genotype) and 72 hpf (n = 5 each genotype) stages. Relative intensity is normalized to the wildtype mean. No significant differences were scored except in the cloaca at 19-somite stage ($p = 0.02$, two-sided Welch's t-test). The boxes represent the interquartile range (IQR), with the lower and upper hinges denoting the first and third quartiles (25th and 75th percentiles). Whiskers extend from the hinges to the furthest data points within 1.5 times the IQR. The upper whisker reaches the largest value within this range, while the lower whisker extends to the smallest value within 1.5 times the IQR from the hinge. Scale bar length is 100 µm in main panels and 25 µm in inset panels. cl, cloaca; cf, caudal fin; ed, endoskeletal disc; ff, fin fold; hg, hindgut; pnd, pronephric duct; s, shoulder; tb, tailbud.

nature portfolio

Corresponding author(s): Denis Duboule

Last updated by author(s): 4 August 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We confirm that all commercial and open source softwares used for image acquisition and CRISPR guide design are precisely described in the material and method section of the manuscript.
Data analysis	We confirm that all the scripts necessary to reproduce data analyses and figures are available at https://github.com/AurelieHintermann/HintermannBoltHawkinsEtAl2025 and the link is provided in the code availability section of the manuscript.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

We confirm that the GEO repository GSE250267 contains links to all raw and processed next-generation sequencing data of this study, including re-analysed and newly generated datasets, as well as replicates which are not shown in the figures. The link to the GEO repository is provided in the data availability section of the

manuscript. (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE250267).

Raw and quantified RT-qPCR and morphometric measurements are available in our github repository available at HintermannBoltHawkinsEtAl2025 and the link is provided in the code availability section of the manuscript.

There is no restriction on data availability and our study does not include any clinical dataset.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Our research does not include any human participant, data, or biological material.
Reporting on race, ethnicity, or other socially relevant groupings	Our research does not include any human participant, data, or biological material.
Population characteristics	Our research does not include any human participant, data, or biological material.
Recruitment	Our research does not include any human participant, data, or biological material.
Ethics oversight	Our research does not include any human participant, data, or biological material.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to predetermine sample size. Instead, sample sizes were determined to ensure they were sufficient for reliable analysis, while also considering factors such as animal usage and practical constraints related to experimental costs. The robustness of our data, achieved through the combination of various experimental approaches, has allowed us to establish that the described sample sizes are adequate for drawing meaningful conclusions.
Data exclusions	No data were excluded from the analysis.
Replication	Zebrafish: Whole-mount in situ hybridization (WISH) on zebrafish embryos were performed on five or more embryos per genotype and probe, combining developmental stages as no temporal differences were observed. HCR were performed and quantified for 7 individuals of each group. Phenotypes were recorded for each genotype in at least 4 individuals. ATAC-seq in posterior trunk was conducted in triplicates and compared to singleton ATAC-seq in three different Hox negative tissues (16 hpf head, 16 hpf rhombomeres and 30 hpf head), considered as three replicates of the negative condition. CUT&RUN experiments in samples collected on 16 hpf zebrafish samples were performed in duplicate for H3K27ac and singleton for H3K27me3. CUT&RUN results were consistent with ATAC-seq results and, when considered together, were sufficient to demonstrate regulatory activity. Mouse: All mouse urogetnial sinus (UGS) data were generated separately for males and females. WISH and β-gal staining in mouse UGS were done in two or more samples. ATAC-seq in mouse UGS were performed in two or more replicates in each sex. CHiC-seq and H3K27ac ChIP-seq were performed as singleton in each sex, H3K27me3 ChIP-seq as singleton in male only. All RT-qPCR and RNA-seq replicates (n>=3) were included in quantifications and no obvious outliers were found. Since mouse findings apply to both sexes, male and female samples can be considered replicates of each other. Since all samples consisted of a pool of three or more embryos, the resulting data represent averages across different biological samples. All attempts at replication were successful. Therefore, for the figures, one replicate was chosen as representative of the set. Raw and processed data of all sequencing replicates, including those not shown in the figures, are made available in the GEO repository (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE250267).
Randomization	No experiments required randomization as experimental groups are based on genotypes (zebrafish) and on genotypes and sex (mouse).
Blinding	Blinding is not relevant in this study because experimental groups are based on genotypes (zebrafish) and on genotypes and sex (mouse).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used: anti-H3K27ac (Abcam Ab4729), anti-H3K27me3 (Merck Millipore 07-449)

Validation: According to manufacturer, the anti-H3K27ac (Abcam Ab4729) antibody was verified by EpiCypher™ SNAP-ChIP™ Spike-in Controls to ensure target specificity and enrichment in ChIP. Species reactivity is shown in human and nematode.

According to manufacturer, the anti-H3K27me3 (Merck Millipore 07-449) is dot blot tested for trimethylated lysine 27 specificity and validated in western blot, immunocytochemistry and immunoprecipitation. Species reactivity is shown in human and mouse, and broad species cross-reactivity is expected.

In house validation was previously performed for both antibodies, with the quantification of ChIP-seq or Chipmentation signal intensity on loci known to be positively enriched (see Reporting Summary of 10.1038/s41588-023-01426-7).

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals: We confirm that all details about animal species, strain and handling are carefully listed both in the material and methods section of the manuscript and in the GEO repository (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE250267).

Wild animals: No wild animals were used in this study.

Reporting on sex: Sex has been considered in the mouse study design. Sex has been determined visually, based on the presence of ovaries or testis as depicted in Figure 3a. For the figures involving WISH and β-gal staining, one embryo of each sex was displayed. Mouse findings apply to both sexes, as no significant differences have been observed between female and male results. Therefore, only the male sequencing data is shown in the figures, but both male and female data are made available in the GEO repository (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE250267).

Sex has not been considered in the zebrafish study design because it is not possible to determine sex at embryonic stages.

Field-collected samples: No field-collected samples have been used in this research.

Ethics oversight: All experiments using mice were approved and performed in compliance with the Swiss Law on Animal Protection (LPA) under license numbers GE45/20 and GE81/14. All zebrafish husbandry procedures have been approved and accredited by the federal food safety and veterinary office of the canton of Vaud (VD-H23).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks: Our research does not involve the use of plants.

Novel plant genotypes: Our research does not involve the use of plants.

Authentication: Our research does not involve the use of plants.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

GEO repository GSE250267
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE250267

Files in database submission

Zebrafish CUT&RUN:
CnR_H3K27ac_danRer_16hpf_PT_rep1.bw
CnR_H3K27ac_danRer_16hpf_PT_rep1.narrowPeak.gz
CnR_H3K27ac_danRer_16hpf_PT_rep2.bw
CnR_H3K27ac_danRer_16hpf_PT_rep2.narrowPeak.gz
CnR_H3K27me3_danRer_16hpf_PT_rep1.bw
CnR_H3K27me3_danRer_16hpf_PT_rep1.narrowPeak.gz

Mouse ChIP-seq:
ChIP_H3K27ac_mm_E185_MUGS_rep1.bw
ChIP_H3K27ac_mm_E185_MUGS_rep1.narrowPeak.gz
ChIP_H3K27ac_mm_E185_FUGS_rep1.bw
ChIP_H3K27ac_mm_E185_FUGS_rep1.narrowPeak.gz
ChIP_H3K27me3_mm_E185_MUGS_rep1.bw
ChIP_H3K27me3_mm_E185_MUGS_rep1.narrowPeak.gz

Genome browser session
(e.g. UCSC)

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

CUT&RUN experiments in samples collected on 16 hpf zebrafish samples were conducted in duplicate for H3K27ac and singleton for H3K27me3. CUT&RUN results were consistent with ATAC-seq results and, when considered together, were sufficient to demonstrate regulatory activity.

Mouse H3K27ac ChIP-seq were performed as singleton in each sex. Since mouse findings apply to both sexes, male and female samples can be considered replicates of each other. Mouse H3K27me3 ChIP-seq was conducted as singleton in male embryos.

Sequencing depth

Zebrafish CUT&RUN:
CnR_H3K27ac_danRer_16hpf_PT_rep1: sequenced to 11.0821 million pairs of 76 and 76bp
CnR_H3K27ac_danRer_16hpf_PT_rep2: sequenced to 13.5354 million pairs of 76 and 76bp
CnR_H3K27me3_danRer_16hpf_PT_rep1: sequenced to 14.2077 million pairs of 76 and 76bp

Mouse ChIP-seq:
ChIP_H3K27ac_mm_E185_MUGS_rep1: sequenced to 46.1228 million reads of 105bp
ChIP_H3K27ac_mm_E185_FUGS_rep1: sequenced to 48.2158 million reads of 105bp
ChIP_H3K27me3_mm_E185_MUGS_rep1: sequenced to 35.287 million reads of 105bp

Antibodies: anti-H3K27ac (Abcam Ab4729), anti-H3K27me3 (Merck Millipore 07-449)

Peak calling parameters

We confirm that the command line program and parameters used for read mapping and peak calling, including the index files used and all aspects of data processing are available in our github repository.

Zebrafish CUT&RUN:

https://github.com/AurelieHintermann/HintermannBoltHawkinsEtAl2025/blob/main/CUTandRUN/CUTandRUN.sh

Mouse ChIP-seq:
https://github.com/AurelieHintermann/HintermannBoltHawkinsEtAl2025/blob/main/ChIP/01_ChIP_SR.sh

Data quality

While peaks were identified, we did not adjust the parameters for extended chromatin marks. This decision was made as we employed ChIP-seq in a descriptive context and chose not to display the peaks in the figures.

We report below the number of peaks above 5-fold enrichment (column 2) and the total number of peaks (column 3), for each dataset shown in the figures.
GSM7976437_CnR_H3K27me3_danRer_16hpf_PT_rep1 3549 17124
GSM7976438_CnR_H3K27ac_danRer_16hpf_PT_rep1 7126 45538
GSM7976439_CnR_H3K27ac_danRer_16hpf_PT_rep2 8571 60431
GSM7977989_ChIP_H3K27me3_mm_E185_MUGS_rep1 5982 40449
GSM7977991_ChIP_H3K27ac_mm_E185_FUGS_rep1 6789 43248
GSM7977990_ChIP_H3K27ac_mm_E185_MUGS_rep1 6105 32714

Software

All scripts necessary to reproduce figures from raw data are available at
https://github.com/AurelieHintermann/HintermannBoltHawkinsEtAl2025.