Cell Reports
Article

Cell PRESS
Open ACCESS

Multiple Enhancers Regulate *Hoxd* Genes and the *Hotdog* LncRNA during Cecum Budding

Saskia Delpretti,[1] Thomas Montavon,[1,3] Marion Leleu,[1] Elisabeth Joye,[1] Athanasia Tzika,[2] Michel Milinkovitch,[2] and Denis Duboule[1,2,*]
[1]School of Life Sciences, Ecole Polytechnique Fédérale, Lausanne CH-1015, Switzerland
[2]Department of Genetics and Evolution, University of Geneva, 1211-Geneva 4, Switzerland
[3]Present address: Department of Epigenetics, Max Planck Institute of Immunobiology and Epigenetics, Stübeweg 51, 79108 Freiburg, Germany
*Correspondence: denis.duboule@epfl.ch
http://dx.doi.org/10.1016/j.celrep.2013.09.002

SUMMARY

Hox **genes are required for the development of the intestinal cecum, a major organ of plant-eating species. We have analyzed the transcriptional regulation of *Hoxd* genes in cecal buds and show that they are controlled by a series of enhancers located in a gene desert flanking the *HoxD* cluster. The start site of two opposite long noncoding RNAs (lncRNAs), *Hotdog* and *Twin of Hotdog*, selectively contacts the expressed *Hoxd* genes in the framework of a topological domain, coinciding with robust transcription of these genes during cecum budding. Both lncRNAs are specifically transcribed in the cecum, albeit bearing no detectable function in *trans*. Hedgehogs have kept this regulatory potential despite the absence of the cecum, suggesting that these mechanisms are used in other developmental situations. In this context, we discuss the implementation of a common "budding toolkit" between the cecum and the limbs.**

INTRODUCTION

The gastrointestinal tract is composed of a series of morphological subdivisions specialized in various functions associated with food intake. In mammals, the diet considerably differs among species, depending on particular environmental adaptations. Although most sources of nutrients can be hydrolyzed by enzymes produced by the digestive system itself, cellulose carbohydrates, which can represent as much as 40% of a vegetarian diet, require a specialized symbiotic bacteria that secretes cellulase. This microbial digestion is slow and necessitates prolonged contacts between microorganisms and their substrate. Many mammalian species achieve this through their cecum, a specialized intestinal organ located at the transition between the small and the large intestines (Figure 1A). This blind-end diverticulum, absent from carnivorous species, retains fluid and microorganisms and is thus a critical organ in herbivorous and omnivorous species like rodents.

The mechanisms underlying cecum positioning, budding, and extension are poorly understood. *Fgf9* is expressed in the gut epithelium and in the mesoderm of the second intestinal loop at embryonic day 10.5 (E10.5) prior to cecum budding. Its removal from the mesoderm results in a hypoplastic cecum with a narrow epithelial invagination into the mesoderm (Al Alam et al., 2012), whereas its deletion leads to cecum agenesis (Zhang et al., 2006). As *Fgf10* expression is absent from these mutant ceca, it was proposed that epithelial FGF9 signaling to the mesoderm triggers the initial budding and activates *Fgf10* in concert with *Fgfr2b* (Burns et al., 2004; Fairbanks et al., 2004), initiating the outgrowth of the organ. *Bmp4* and *Pitx2*, both expressed in the mesodermal bud, lie at intermediate positions within this signaling cascade (Al Alam et al., 2012; Burns et al., 2004; Fairbanks et al., 2004; Nichol and Saijoh, 2011; Zhang et al., 2006).

As the intestinal tract has an anterior-to-posterior polarity, the role of *Hox* genes in its compartmentalization has been proposed early on. Mammals have 39 *Hox* genes encoding transcription factors and clustered at four genomic loci (*HoxA* to *HoxD*), and these genes are transcribed sequentially according to their respective position within the clusters (Kmita and Duboule, 2003; Krumlauf, 1994). In the developing gut, systematic expression studies (Kawazoe et al., 2002; Pitera et al., 1999; Sekimoto et al., 1998), as well as loss- and gain-of-function approaches (e.g., Aubin et al., 2002; Boulet and Capecchi, 1996), have revealed their importance in gut regionalization.

Several consecutive *Hoxd* genes are expressed in the developing cecum, where they play a particular role during budding (Figure 1B) (Zacchetti et al., 2007). For instance, mouse fetuses carrying a deletion from *Hoxd4* to *Hoxd13* display an ill-formed ileocecal transition and absence of the associated sphincter (Zákány and Duboule, 1999). Also, mice homozygous for a deletion of either the *Hoxd1* to *Hoxd10* or the *Hoxd4* to *Hoxd11* DNA segments show a partial or complete cecal agenesis (Zacchetti et al., 2007), due to the induced gain of function of *Hoxd12* and its concurrent negative effect over other HOX proteins via posterior prevalence (Duboule, 1994; Duboule and Morata, 1994). Therefore, although several *Hoxd* genes play a positive role during cecal development, the transcription of both *Hoxd12* and *Hoxd13* must be excluded from this bud.

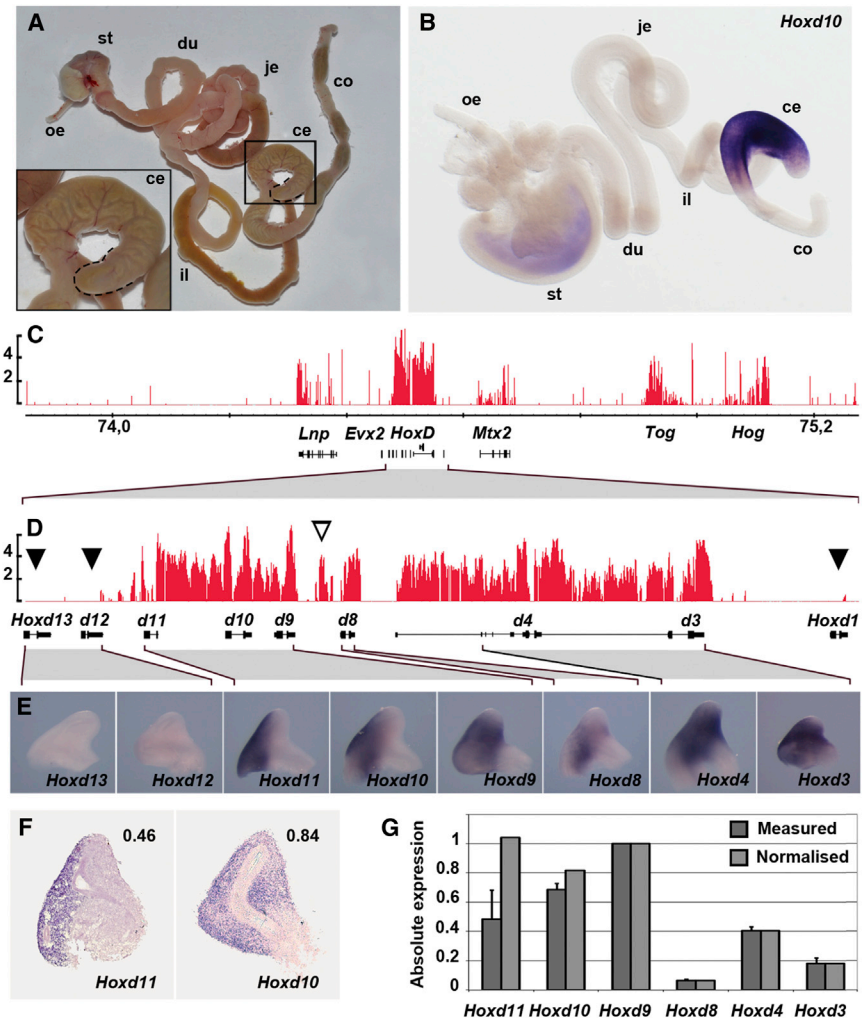

Figure 1. *Hoxd* Transcripts in the Cecum

(A) An adult mouse gastrointestinal tract; esophagus (oe), stomach (st), duodenum (du), jejunum (je), ileum (il), cecum (ce), and colon (co). The tip of the cecum is underlined with a dashed line.

(B) *Hoxd10* expression at E13.

(C and D) Cecum transcription profile over the *HoxD* locus and flanking gene deserts. The y axis is the $\log_2$ ratio of ds cDNA/genomic DNA, and the x axis indicates chromosomal coordinates in Mb (UCSC 2008 assembly, mm9). (C) Transcriptional activity is observed around the position of *Lnp*, the *HoxD* complex, and *Mtx2*, as well as over large regions (*Hog* and *Tog*), in the gene desert telomeric to *Mtx2*. (D) Transcription is seen as two large blocks separated by *Hoxd8*, which is flanked by an antisense transcript (open arrowhead). Silent genes are indicated with black arrowheads.

(E) *Hoxd* gene expression territories in E12.5 ceca. Craniocaudal axis from right to left.

(F) *Hoxd10* and *Hoxd11* gene expression in histological sections of E12.5 ceca. Top right: the relative expression territory for both genes was computed from serial histological sections.

(G) Absolute quantification (dark gray) of *Hoxd* mRNAs in E12.5 ceca was performed by parallel real-time PCR amplification of ceca together with known amounts of the various mRNAs (see Extended Experimental Procedures). *Hoxd10* and *Hoxd11* levels were normalized for the relative size of their expression territories (light gray). *Hoxd9* expression level was set to one. Error bars indicate SD (n = 3).

See also Figure S1.

Here, we studied how this fine-tuned transcriptional regulation is organized during cecum development. We identified an unusual number of enhancer sequences, with comparable specificities, located within a gene desert flanking the *HoxD* cluster. Within this large DNA interval, we found two long noncoding RNAs (lncRNAs), *Hotdog* and *Twin of Hotdog*, which are specifically expressed in the growing cecum at midgestation. These lncRNAs bear no detectable function in *trans* and might be involved in *cis* in the transcriptional regulation of centrally located *Hoxd* genes via physical contacts with their shared transcription start site. We propose a model for the regulation of these genes in the cecum and for the successive acquisition of multiple enhancer sequences with comparable specificities, triggered by the existence of a topological domain where these regulatory interactions occur. Finally, we show that in animals lacking a cecum such as hedgehogs, the orthologous DNA region can still trigger this regulation in transgenic mice, suggesting that these enhancers are used for distinct purposes in various developmental contexts. We propose that cecum and limb budding largely rely on the use of the same genetic toolkit.

RESULTS

Hox Transcription in the Cecum

We produced transcription profiles for all *Hox* loci by using E13.5 embryonic ceca total RNAs (Figure S1). In all clusters, we observed robust signals covering *Hox* genes essentially from paralogy groups 2 to 11. In contrast, transcripts were never scored over groups 12 and 13 (Figure S1). The steady-state level of RNAs in the cecum was significantly higher for *Hoxd* genes than for other *Hox* clusters. This and their expression patterns in time and space (Figure 1) suggested that *Hoxd* genes are important players in making the cecum, along with their *Hoxa* paralogs.

Transcription over the *HoxD* locus and its flanking gene deserts was further assessed with high-resolution tiling microarrays. E13.5 ceca single-stranded or double-stranded complementary DNAs (cDNAs) revealed specific transcriptional activity at and around the *HoxD* complex (Figures 1C and 1D). Although the centromeric gene desert was largely devoid of transcripts, we observed significant signals over two large regions located hundreds of kilobases telomeric from *Hoxd1*. Comparisons between single-stranded and double-stranded cDNAs revealed the presence of two lncRNAs, encoded by opposite DNA strands and sharing the same (or closely located) transcription start site(s). The largest lncRNA was referred to as "*HoxD* telomeric desert lncRNA operating in the gut" or

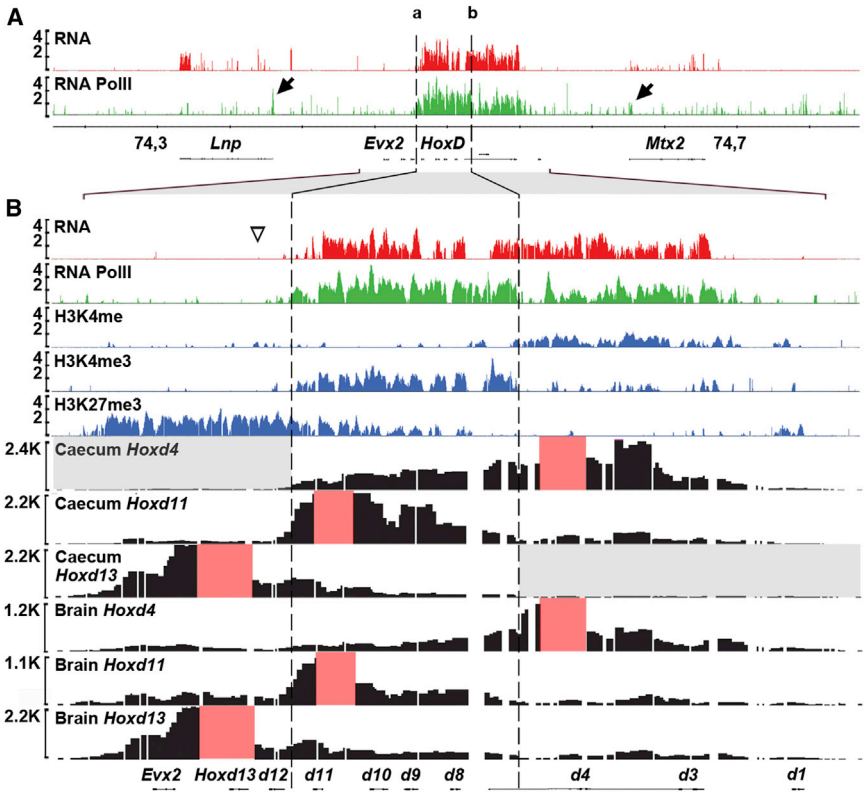

Figure 2. Active and Inactive Chromatin Domains

(A and B) Chromatin marks and 3D conformation of the *HoxD* gene cluster. The y axis indicates the $\log_2$ ratio of ds cDNA/genomic DNA, the $\log_2$ ratio of ChIP-enriched DNA/input, and the ratio of 4C-amplified DNA/input (4C-seq). Transcript profiles (red) are aligned with Pol II distribution (green), the H3K4me1, H3K4me3, and H3K27me3 profiles (blue), as well as the 4C profiles (black). The brain was used as a control for 4C. Signal intensities were adapted to allow for comparisons between various profiles. The arrows in (A) point to the promoters of *Lnp* and *Mtx2*. The baits used for 4C analyses are indicated with red boxes. The vertical dashed lines delineate the distinct chromatin domains within the *HoxD* cluster, as identified by different marks on H3 tail, with the inactive compartment on the left of dashed line (open arrowhead). Regions with little or no interactions are gray-shaded.
See also Figure S2.

Hotdog (*Hog*), and the opposite RNA as *Twin of Hotdog* (*Tog*) (see below).

Little (if any) signals were scored over *Hoxd12* and *Hoxd13* (Figure 1D, black arrowheads). Instead, transcription of *HoxD* in the cecum was scored from exon 1 of *Hoxd11* continuously up to *Hoxd9*, until a 2 kb large antisense transcript present within the *Hoxd9* to *Hoxd8* intergenic region (Figure 1D, white arrowhead). Transcripts were also detected over *Hoxd8*, followed by a block covering the *Hoxd4* to *Hoxd3* region. *Hoxd3* transcripts extended from the end of the ileum to the beginning of the colon, with rather loose expression boundaries (not shown), consistent with previous observations (Zacchetti et al., 2007). *Hoxd4*, *Hoxd8*, and *Hoxd9* all showed a sharp anterior expression limit at the ileocecal transition up to the start of the colon. *Hoxd10* expression boundary was shifted caudally, whereas *Hoxd11* transcripts were detected in the posterior half of the cecum only (Figures 1E and 1F). Quantitative PCR (qPCR) indicated that *Hoxd9*, *Hoxd10*, and *Hoxd11* messenger RNAs (mRNAs) were in higher amounts than *Hoxd3* and *Hoxd4* (Figure 1G) and these figures were normalized by the percent of cells expressing *Hoxd10* and *Hoxd11* (Figure 1F). Consequently, *Hoxd9*, *Hoxd10*, and *Hoxd11* appeared as the strongest-expressed *Hoxd* genes in the growing cecum (Figure 1G).

Keeping *Hoxd12* and *Hoxd13* Silent

To understand how high transcription of these three genes can occur while their immediate *Hoxd12* and *Hoxd13* neighbors are excluded, we analyzed the effects of targeted deletions within the *HoxD* cluster (Extended Results; Figure S2). Since several such deletions induced the ectopic expression of *Hoxd12* (Figure S2), we concluded that this latter gene has the capacity to respond to the cecum regulation(s), yet its relative position within the gene cluster normally prevents its activation. In contrast, *Hoxd13* remained silent in all these genetic conditions (Figure S2).

However, *Hoxd13* was expressed in the cecum after the targeted inversion of *Hoxd11* and *Hoxd12* (Kmita et al., 2000). There, its promoter remained in place and was activated together with the inverted *Hoxd11* and *Hoxd12* transcription units (Extended Results; Figure S2). In this inversion, *Hoxd11* was transcribed from the opposite DNA strand and extended its transcripts over the *Hoxd13* promoter, which might have allowed for transcripts to initiate from this promoter (Figure S2). From these results, we concluded that both the *Hoxd13* and *Hoxd12* promoters have the intrinsic capacity to respond to the cecum regulation(s), despite a strong silencing under normal conditions, which prevent their deleterious transcription in this developing organ.

Chromatin Domains and Interaction Profiles

To assess whether this repression was associated with a particular chromatin state, we microdissected E13.5 ceca and looked at both RNA Pol II occupancy and various chromatin marks associated either with regulatory sequences (H3K4me1; Heintzman et al., 2007), with actively transcribing promoters (H3K4me3), or with inactive genes such as H3K27me3. As expected, RNA Pol II and H3K4me3 were detected as narrow peaks over the promoters of both *Lnp* and *Mtx2* (Figure 2A, arrows). In addition, signals were broadly distributed over the *HoxD* locus, except for the *Hoxd12* to *Hoxd13* region (Figure 2B, white arrowhead). Two large domains of RNA Pol II were scored,

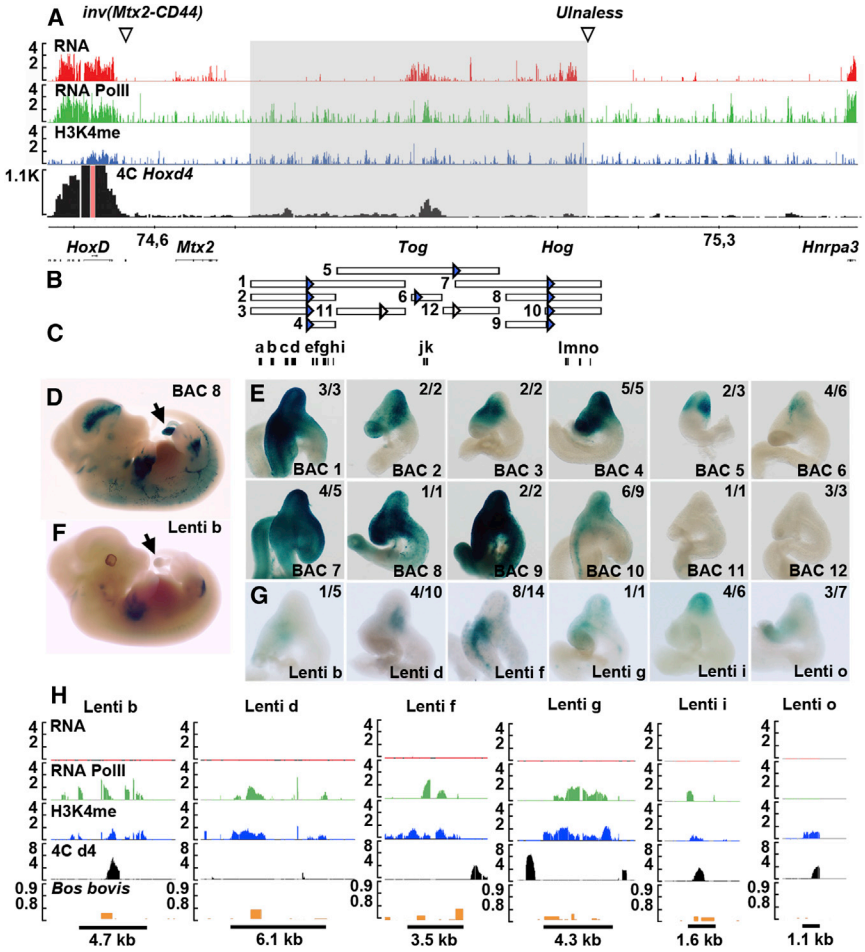

Figure 3. Multiple Enhancers Regulate *Hoxd* Genes in the Cecum

(A) A 420 kb large regulatory interval (gray box), spanning from *Mtx2* up to the *Ulnaless* breakpoint (open arrowhead). The y axis is as for Figure 2, with transcripts in red, aligned with Pol II signals (green), H3K4me1 (blue), and 4C interactions (black). The *Hoxd4* 4C bait is indicated with a light-red box. The left open arrowhead points to the *inv(Mtx2-CD44)* centromeric breakpoint (see also Figures 5A and 5B). *Hoxd4* shows preferential interaction with the telomeric gene desert.

(B) Positions of the native BACs and their deleted versions used for transgenesis. The *LacZ* reporter is indicated with triangles, which is blue should the BAC be active in the cecum.

(C) Candidate enhancer sequences used in transgenesis.

(D and E) *LacZ* staining at E12.5 after BAC transgenesis. The arrow in (D) points to the cecum.

(E) Dissected ceca stained after BAC transgene integration. The BAC number is as under (B) and the number of embryos with stained cecum over total is indicated on the upper right corners (see Table S1).

(F and G) *LacZ* staining in E12.5 ceca after lentiviral transgene integrations of enhancer sequences. The arrow in (F) points to the cecum.

(G) Dissected ceca transgenic for enhancer sequences b, d, f, g, i and o (as in C).

(H) Alignment of various enhancer sequences (black bars below) with Pol II, H3K4me1 modifications, and 4C interaction peaks (4C on array). Sequence conservation with the cow genome (orange boxes) is indicated as a relative value.

See also Table S1 and Figure S3.

from *Hoxd11* till an upstream promoter of *Hoxd3* (Figure 2B, between dashed lines a and b) and from the promoter of *Hoxd4* to the end of *Hoxd3* (Figure 2B, right of dashed line b).

These distinct domains showed different levels of H3K4 methylation, with H3K4me1 covering the latter, whereas H3K4me3 was scored on the former (Figure 2B). In contrast, H3K27me3 chromatin marks were spread over the silenced part of the cluster (Figure 2B), peaking over *Evx2*, *Hoxd13* and *Hoxd12* with weaker signals over *Hoxd11* and *Hoxd10* likely due to the presence of negative cells in dissected ceca (see Figure 1F). Therefore, sustained transcription occurred in the budding cecum from the *Hoxd11* gene to a distal *Hoxd3* promoter (Figure 2B, between both dashed lines), whereas the *Hoxd4* and proximal *Hoxd3* promoters were not equally active, indicating that two subgroups of genes responded differently to the cecum regulation.

To see whether these distinct chromatin domains would reflect 3D regulatory structures, we established the interaction profiles of both a silenced (*Hoxd13*) and an active (*Hoxd4*) gene, using chromosome conformation capture sequencing (4C-seq). When using *Hoxd4* as bait, signals extended from *Hoxd3* to *Hoxd11*, showing a preferential interaction with other transcribed *Hoxd* genes, rather than with silenced loci (Figure 2B). This *Hoxd4* interaction domain precisely matched the addition of both H3K4me1 and H3K4me3 domains. The same result was obtained with *Hoxd9* and *Hoxd11* (Figure 2B), supporting the presence of this ‘‘positive’’ interaction domain.

In contrast, when *Hoxd13* was used as bait, signals were distributed mostly around the *Evx2* to *Hoxd13* region, overlapping with the H3K27me3 domain. *Hoxd13* thus preferentially interacted with silenced DNA in the cecum (Figure 2B). When brain tissue was used, where all *Hoxd* genes are silent, the interaction profiles covered the entire *HoxD* cluster regardless of the bait, as previously reported in other instances (Montavon et al., 2011; Noordermeer et al., 2011), indicating a globular-like negative 3D structure (Figure 2B).

A Range of Enhancer Sequences

In addition to these interactions, active *Hoxd* genes established significant contacts with the telomeric gene desert. This 1 Mb large region downstream *Mtx2* (Figure 3) thus appeared as a potential reservoir for cecum regulatory sequences, a hypothesis supported by two targeted rearrangements. First, when the *HoxD* cluster is deleted and replaced by a *Hoxd11/LacZ* transgene, *LacZ* expression is scored in the cecum indicating that enhancers are located outside the cluster itself (Spitz

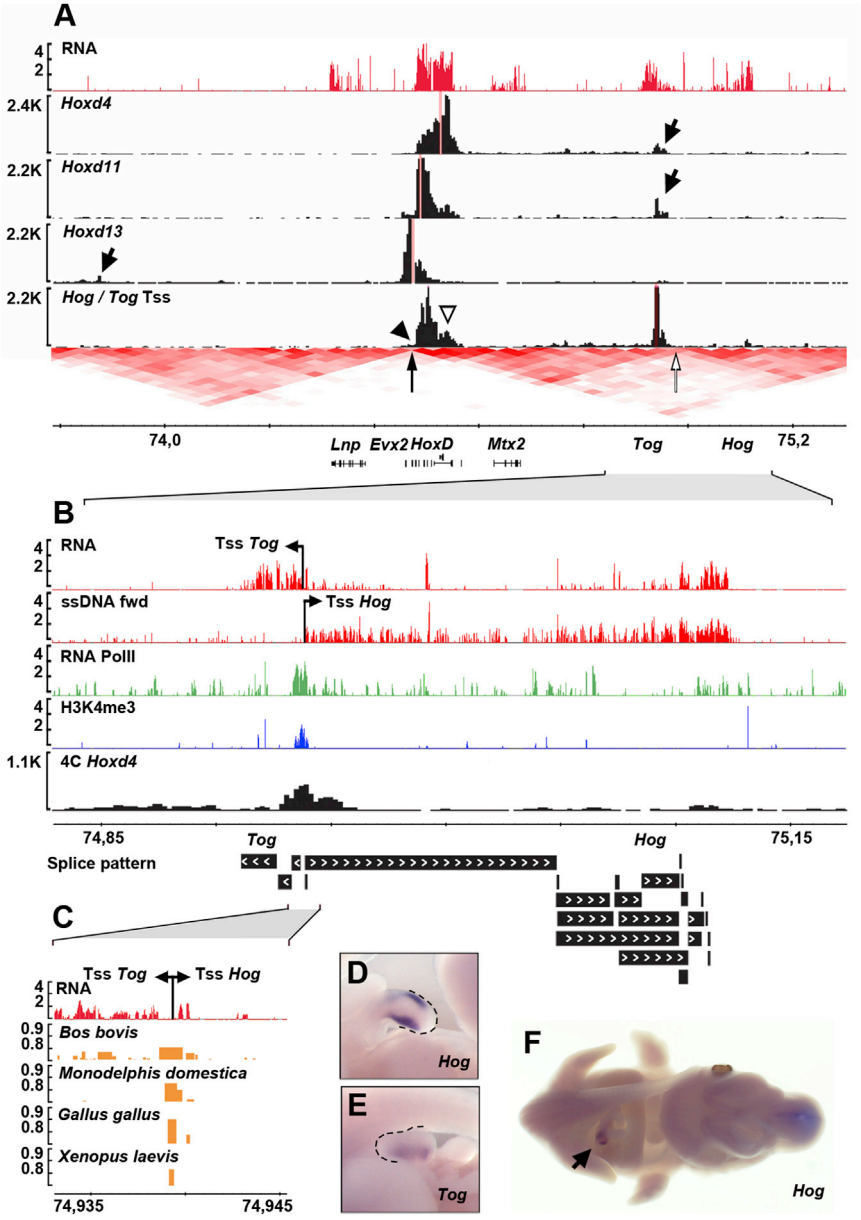

Figure 4. Contacts between *Hoxd* Genes and the *Hog* and *Tog* Transcription Start Site

(A) Cecum RNAs profile (red) covering the *HoxD* locus and flanking gene deserts. Below are 4C interaction profiles using either the active *Hoxd4* and *Hoxd11* loci as baits or the inactive *Hoxd13* locus. The arrows indicate strong long-range contacts, either in the telomeric gene desert (*Hoxd4*, *Hoxd11*) or with the centromeric side (*Hoxd13*). Below is the 4C profile when using the telomeric interacting sequence itself as bait (indicated as Tss in B). In this latter case, only the active domain of the *HoxD* cluster is contacted, with weaker contacts over anterior genes (open arrowhead), whereas *Hoxd12* and *Hoxd13* are not contacted (black arrowhead). The extent of the two reported topological domains (Dixon et al., 2012) is shown in red below, with a vertical black arrow to indicate the centromeric border of the telomeric domain. The vertical open arrow shows the border between two subdomains, with the Tss located right centromeric to this border.

(B) Enlargement of (A) with the RNA profile covering the telomeric desert. The large transcribed area is aligned with strand specific cDNAs (below, red), thus identifying two distinct lncRNAs, *Hotdog* (*Hog*) and *Twin of Hotdog* (*Tog*), with opposite transcriptional directions, sharing the same start site (Tss; arrows). This DNA region is enriched in both Pol II occupancy (green) and H3K4me3 (blue). This Tss maps with the highest peak of interaction as detected by 4C (black). *Hotdog* shows a complex splicing pattern, as revealed by using paired-end sequencing.

(C) The Tss of both *Hog* and *Tog* locates in a region highly conserved throughout vertebrates. DNA sequence conservation is shown in orange as a relative value.

(D–F) In E12.5 embryos, the expression of both *Hog* (D and F) and *Tog* (E) is restricted to the developing cecum.

See also Table S2 and Figure S4.

et al., 2001). Second, when the cluster is disrupted into two independent pieces via an inversion breakpoint, expression in the cecum is maintained for those genes associated with the telomeric neighborhood (Spitz et al., 2005).

To further reduce this potential regulatory interval, we used *Ulnaless* (*Ul*) mutant mice, a 770 kb large inversion including the *HoxD* locus and adjacent telomeric DNA (Figure S3) (Hérault et al., 1997; Peichel et al., 1997; Spitz et al., 2003). Because 570 kb of telomeric DNA are inverted along with the gene cluster, enhancers present within this interval maintain their neighborhood with *Hoxd* genes (Figure 3A, white arrowhead). Although *Hoxd13* was not transcribed in *Ul* mutant cecum, *Hoxd11* was transcribed as in wild-type controls (Figure S3), indicating that cecum enhancers had been inverted along with their targets and are thus present in the 570 kb flanking the gene cluster, in the region containing the *Hog* and *Tog* lncRNAs (Figure 3A).

H3K4me1 was found throughout the gene desert along with range of highly conserved DNA sequences (Lee et al., 2006a), but an alignment with the interaction profiles revealed enriched contacts up until the position of the *Tog* transcript, ca. 400 kb far from *Hoxd1* (Figures 3, 4, and S4). Consequently, we selected the 420 kb large region from *Mtx2* to the *Ulnaless* breakpoint for subsequent transgenic analyses (Figure 3A, gray shaded area). We used three overlapping bacterial artificial chromosomes (BACs) covering this interval (Figure 3B, BACs 1, 5, and 7) and recombined a *LacZ* reporter gene (Figure 3B, arrowheads). In transgenic fetuses, both BACs 1 and 7 elicited a strong *LacZ* expression in the cecum (Figure 3E; Table S1). Whereas the anterior limit of expression for BAC 1 was similar to *Hoxd9,* expression appeared broader in the midgut for BAC 7, somewhat comparable to *Hoxd3*. On the other hand, BAC 5 induced

significant *LacZ* staining in the tip of the cecum only (Figure 3E). We concluded that at least two distinct regulatory sequences were present, in BACs 1 and 7.

We shortened both BACs and used a fosmid clone such that five contiguous DNA segments were assayed (Figure 3B, BACs 2, 11, 12, 8, and clone 6). BACs 2, 8, and fosmid 6 gave staining in the cecum (Figures 3D and 3E). As expected, the former showed a *Hoxd9*-like expression, whereas BAC 8 mimicked *Hoxd3*. Fosmid 6 stained the tip of the cecum, resembling BAC 5. We further shortened BACs 2 and 8 (BACs 3, 4, 9, and 10; Figure 3B) and all these shorter fragments unexpectedly gave *LacZ* expression in the cecum (Figure 3E), again within territories related to those of either *Hoxd9* or *Hoxd3*. Therefore, four distinct BACs and one fosmid revealed the presence of at least five enhancer sequences.

Within these large DNA fragments, we identified evolutionary conserved sequences and aligned them with RNA Pol II occupancy, H3K4me1 marks, and the 4C interaction profiles. Fifteen candidate enhancers were assayed in a lentiviral-based transgenic system (Figure 3C) (Friedli et al., 2010). Although signals were fainter than with BACs (Figures 3F and 3G), six sequences gave *beta-gal* activity in the cecum (Figure 3G; Table S1). These sequences were devoid of any detectable transcripts, yet they were covered with RNA Pol II (except transgene "o") and H3K4me1 marks (Figure 3H). Peaks of interaction with *Hoxd4* were also present near five out of the six DNA fragments used as transgenes. Finally, all of them but sequence o showed conservation with the cow genome. Expression patterns of the transgenes varied qualitatively and quantitatively, yet each one reflected part of the wild-type *Hoxd* expression territory.

In *cis* Interaction with the *Hog* and *Tog* Transcription Start Site

The strongest interaction peak between active genes (*Hoxd4* or *Hoxd11*) and the regulatory landscape was observed with the lncRNAs (Figures 4A and S4, arrows), whereas *Hoxd13* did not significantly contact this region (Figure 4A). Using single-stranded cDNA, we identified a shared start site (Tss) for the *Hotdog* and *Twin of Hotdog* RNAs. The ca. 200 kb large *Hog* transcript extended telomeric, whereas *Tog* was encoded by the opposite strand. Paired-end sequencing indicated that *Hog* had a complex splicing pattern with multiple alternative exons (Figure 4B).

The interaction peak over the *Hog* and *Tog* transcription start site overlapped with both H3K4me3 marks and a clear RNA Pol II enrichment (Figure 4B). Of note, this region is highly conserved in vertebrates, including amphibians (Figure 4C), and contains range of conserved binding sites for transcription factor (Table S2). In turn, when the *Hog* and *Tog* Tss was used as bait in 4C-seq, the strongest contacts were observed with the *Hoxd9* to *Hoxd11* region, whereas weaker contacts were scored over the *Hoxd1* to *Hoxd8* interval (Figure 4A, white arrowhead), thus matching the transcription profile of *Hoxd* genes in the cecum. The *Evx2-Hoxd13* region was as expected not contacted at all (Figure 4A, black arrowhead). Also, few contacts were observed telomeric to the Tss, i.e., toward the region where the *Hog* transcript elongates. When the Tss was used as bait, only 25% of all contacts were scored telomeric, whereas 75% concerned the DNA interval from *Hog* to *Hoxd11*. This biased distribution exactly matched the presence of a topological domain (Dixon et al., 2012), with *Hoxd11* and the *Hog* Tss labeling its borders (Figure 4, bottom; black and white arrows).

Both lncRNAs were transcribed in the developing cecum, whereas no other expression site could be identified at least at this stage and using current procedures (Figures 4D–4F). This exquisite tissue specificity suggested that *Hog*, *Tog*, and *Hoxd* genes might all be controlled by the same enhancer sequences. To assess whether the contacts between the *Hog* and *Tog* Tss and the *HoxD* cluster were necessary to trigger lncRNA transcription, we engineered a 28 Mb large inversion between *Mtx2* and *CD44*. In this *inv(Mtx2-CD44)* allele, *Hog* and *Tog* were displaced 28 Mb further away (Figure 5A). Mice carrying this inversion no longer displayed any *Hoxd* expression in the cecum, demonstrating that the enhancers were located telomeric to *HoxD* (Figures 5A and 5B). In contrast, both *Hog* and *Tog* were still expressed due to their unchanged proximity to the various enhancers (Figures 5C and 5D), illustrating their transcriptional independency from *Hoxd* genes.

We checked whether cross-regulatory effects such as enhancer sharing with *Hog* and *Tog* may influence the transcription of *Hoxd* genes in the cecum. We used embryos carrying deletions at the *HoxD* locus to evaluate a potential impact upon the transcription of the lncRNAs via enhancer-promoter reallocation. When animals lacking from *Hoxd1* to near *Hoxd8* [*del(1–4i)*] were used, the overall transcription profile of both lncRNAs remained unchanged (Figure 5E). However, when the three most transcribed *Hoxd* genes in the cecum were removed [*del(9–11)*], *Hog* transcription was virtually abolished despite its unchanged proximity to the requested enhancers (Figures 5E–5G).

A Function for *Hog* and *Tog* in *trans*?

This unexpected result was controlled by using other deletions removing various numbers of posterior *Hoxd* genes and we consistently observed a large decrease in steady state levels of *Hog* RNAs in all deletions containing gene members of the preferential targets (from *Hoxd9* to *Hoxd11*; Figure 5G). In contrast, the deletion of both *Hoxd12* and *Hoxd13* (i.e., genes that are inactive in the cecum and not contacted by the *Hog* and *Tog* Tss) induced a slight loss of *Hog* RNAs only (Figure 5G). This suggested that in the absence of the major target genes, a reorganization had occurred in the interaction between *HoxD* and the lncRNAs, which lead to a downregulation of *Hog* (Figure 5H).

Besides this *cis* effect, we investigated a possible function for these noncoding RNAs in *trans* by comparing ceca transcriptomes obtained for two mutant conditions affecting similarly the transcription of *Hoxd* genes, but either with or without *Hog* and *Tog* transcripts (Figures 6A–6C). We used the 525 kb large [*del(65-TpSB2)*] deletion removing most of the telomeric gene desert (Andrey et al., 2013), including all identified cecum enhancers and both *Hog* and *Tog* (Figure 6B). *Hoxd* transcripts level dropped down to ca. 30% of wild-type levels (Figure 6D), indicating that additional enhancers locate further telomeric to the deletion breakpoint, in addition to those reported in this work. As a reciprocal mutant condition, we used the 28 Mb large

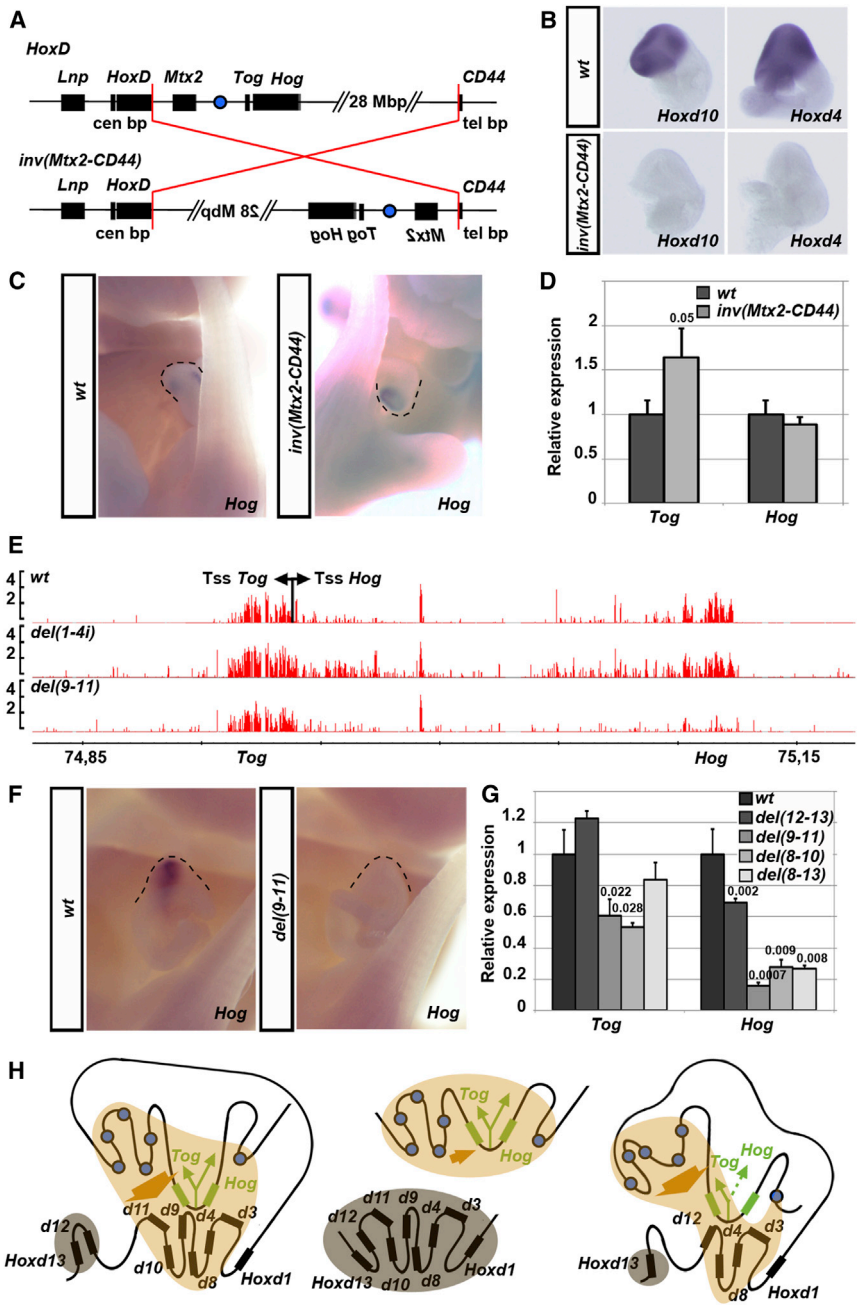

Figure 5. Genetic Interactions in *cis* between *Hoxd* Genes and *Hog* and *Tog*

(A) A 28 Mb large inversion (in red) separates the *HoxD* cluster from its telomeric neighborhood. Enhancer sequences are represented with a single blue circle to indicate their relative position after inversion.

(B and C) In this allele, *Hoxd* gene expression is abrogated from the cecum, whereas expression of *Hog* and *Tog* is maintained.

(D) RT-PCR comparisons of *Hog* and *Tog* RNA steady-state levels in E13.5 ceca in wild-type and inverted mutant animals (n = 3). Significant p value is given on top of error bar.

(E) RNA profiles of *Hog* and *Tog* in either *del(1–4i),* or *del(9–11)* (see schemes in Figure S2) mutant ceca. The start sites (Tss) are indicated with arrows. *Hog* is severely downregulated in *del(9–11)* ceca, yet it remains transcribed in *del(1–4i)*.

(F) *Hog* RNAs are virtually abolished in *del(9–11)* mutant ceca. The tip of the cecum is underlined with a dashed line.

(G) Steady-state levels of *Hog* and *Tog* RNAs in E13.5 ceca (n = 3) of animals carrying various deletions within the *HoxD* cluster. Significant p values are given on top.

(H) Model of *Hoxd* gene regulation in the developing cecum. Wild-type on the left, with *Hoxd* genes (black rectangles) from the central part of the cluster (*Hoxd9* to *Hoxd11*) strongly interacting (large orange arrow) with both the *Hog* and *Tog* Tss (green) and the various enhancers (blue). A large domain of interactions is formed (orange cloud), which corresponds to both H3K4me1 and H3K4me3 chromatin domains and involves the telomeric gene desert. *Hoxd13* and *Hoxd12* are excluded from this domain (gray cloud). Middle: when the cluster is separated from the gene desert, all *Hoxd* genes are in a negative domain, whereas both *Hog* and *Tog* are actively transcribed. Right: a deletion of the major *Hoxd* target genes reallocates the enhancers contacts toward *Hoxd12* and thus changes the overall spatial configuration, impinging over the transcription of *Hog*.

inversion between *Mtx2* and *CD44* (Figure 6C). In this *inv(Mtx2-CD44)* allele, *Hoxd* transcription was virtually abolished, whereas *Hog* and *Tog* remain transcribed (Figure 6E). Cross-comparisons among these various transcriptomes thus allowed to discriminate between genes misregulated due to the knock down of *Hoxd* genes (indirect targets) from genes directly affected by the absence of *Hog* and *Tog*. Very few genes were found up- or downregulated in this latter category (Figure 6F; Table S3), suggesting that *Hog* and *Tog* have little (if any) function in controlling transcription in *trans*, at least in such in vivo conditions.

Cecum-Specific Transcripts

We also compared transcripts sequenced from microdissected embryonic ileum, cecum, and colon to extract cecum-specific RNAs as potential targets of *Hog* and *Tog* (Figure 7A). We found that more than 1,800 genes were specifically upregulated at least by a factor of two in the budding cecum, when compared either to the ileum or to the colon. A total of 429 genes were similarly upregulated in the cecum in both comparisons (Figures 7B and 7C) with *Hoxd9*, *Hoxd10*, and *Hoxd11* on top of the list, as expected. We used gene ontologies with a minimal 5-fold enrichment and six out of the first seven terms were related to limb or appendage morphogenesis (Table S4; Figure 7D). We validated some of these genes by expression studies and found that *Gdf5*, a gene known for its major role in joint formation (Storm et al., 1994), was transcribed in cells forming a ring at the basis of

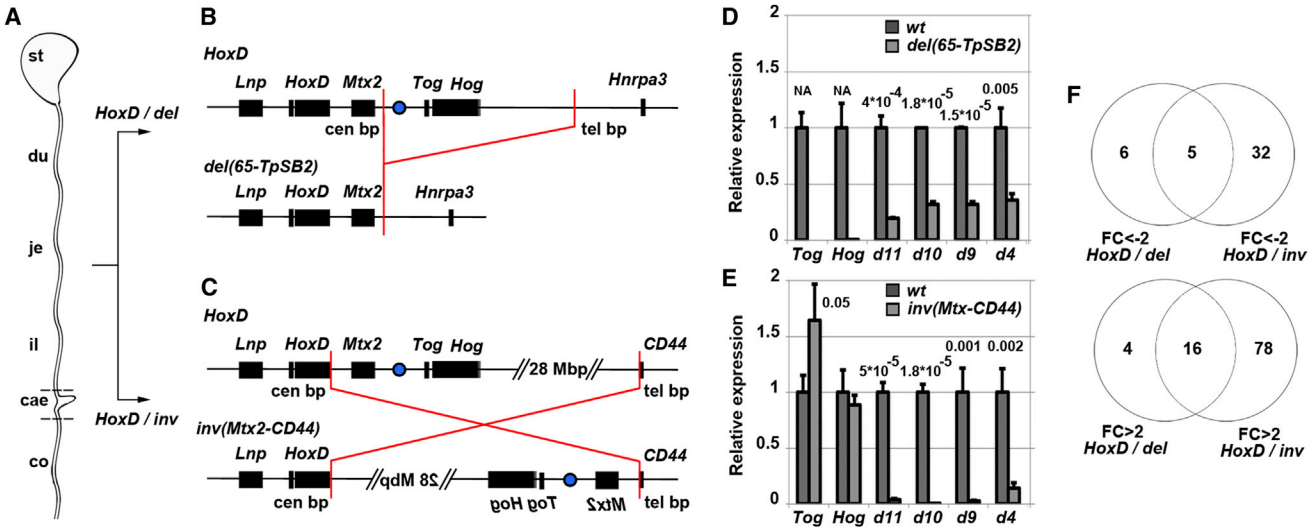

Figure 6. Potential Function of *Hog* and *Tog* in *trans*

(A–C) Dissection used for RNA-seq and comparison between the wild-type, *del(65-TpSB2)*, and *inv(attP-CD44)* alleles. In *del(65-TpSB2)*, the lncRNAs *Hog* and *Tog*, as well as all identified cecum enhancers, are removed in a large 528 kb deletion.

(D and E) Quantitative PCR analyzes of selected *Hoxd* genes, as well as the lncRNAs *Hog* and *Tog*. In *del(65-TpSB2)*, *Hoxd* transcripts are reduced to one-third of their wild-type levels, whereas the *Hog* and *Tog* locus is deleted (D). In *inv(attP-CD44)*, *Hoxd* transcript levels are very low, whereas lncRNAs *Hog* and *Tog* are still expressed under the control of cecum enhancers (E).

(F) Genes downregulated (upper panel) or upregulated (lower panel) in either the deletion or the inversion compared to wild-type (respectively left and right Venn diagram). Analysis of differential gene expression (fold changes greater than two and significant p value) indicates very few genes affected in the deletion, yet not the inversion, a condition reflecting a potential function for *Hog* and *Tog* in *trans*.

See also Table S3.

the budding cecum, labeling the future joint between the ileum and the colon (i.e., right at the presumptive place of the ileocecal valve) (Figures 7E and 7F). Of note, this gene was also found downregulated in the *del(65-TpSB2)* and *inv(Mtx2-CD44)* alleles.

Hoxd Regulation in a Gut without Cecum

We assessed whether this complex regulation was conserved in the four-toed hedgehog (*Atelerix albiventris*), an omnivorous species that lacks a cecum. The hedgehog gastrointestinal tract undergoes a smooth and progressive morphological transition between the ileum and colon, with no clear boundary between the small and large intestines (Figure S5). We collected hedgehog embryos at day 22 to 23 postcoitum and looked at *Hoxd10* expression. Signals were found in the limbs, the genital bud, and the trunk, as for the mouse, with in addition a lateral expression domain covering the flanks of the hedgehog embryo (Figures 8A and 8B). However, neither *Hoxd10* nor *Hoxd11* appeared to be highly transcribed in any localized part of the developing gut, as in mice (Figures 8C–8F). Other *Hoxd* genes, as well as several *Hoxa* genes, were comparatively analyzed with qPCR after dissecting developing guts into several segments, approximately matching the intestinal segments found in mice (Figure S5). Murine *Hoxd3* to *Hoxd4* were detected along the whole intestine, whereas *Hoxd9, Hoxd10 and Hoxd11* were expressed robustly in the cecum and, to a lower extent in the colon (Figure S5). In this most posterior segment, *Hoxd12* and *Hoxd13* were also expressed, in agreement with previous results (Kondo et al., 1996).

In the hedgehog gut, both the distribution and the amount of *Hox* transcripts were comparable, except for the lack of the cecum-associated expression peak of group 9 to 11 genes (Figure S5). We identified and sequenced twelve partially overlapping hedgehog BACs spanning the *HoxD* cluster and telomeric neighborhood and confirmed *Hoxd* gene transcription by sequencing RNA extracted from both the ileum and colon of *A. albiventris* (Figures 8G–8J). A specific transcriptional activity was mapped around the transcription start site of *Tog* and *Hog*, yet these transcripts did not elongate, as observed in the mouse ileum and colon (Figures 8K and 8L). We thus concluded that the absence of cecum in hedgehogs was due to neither the lack of *Hox* gene products during the early phase of gut regionalization nor a genomic rearrangement in the telomeric neighborhood of the hedgehog *HoxD* cluster.

We selected one of the hedgehog BAC clones covering a region orthologous to murine BAC 2, which gave high *beta-gal* activity in the cecum (see Figures 3 and S5). We recombined a *LacZ* reporter cassette at a site comparable to the mouse cognate, and introduced it into transgenic mice. The expression pattern was similar to that produced by mouse BAC2, with some signal in the proximal limbs, in migrating crest cells and along various structures in the trunk (Figures 8M and 8N). In addition, the cecum was clearly stained (Figures 8O and 8P). Using our *A. albiventris* contig assembly, we identified sequences related to the mouse enhancer sequences b, d, f, and g (Figure 3) and tested them in our lentiviral transgenic system. Three out of the four sequences were capable of driving *LacZ* expression in the cecum. Staining was also scored in the proximal limb, along the trunk, and in crest cells (Figure S5).

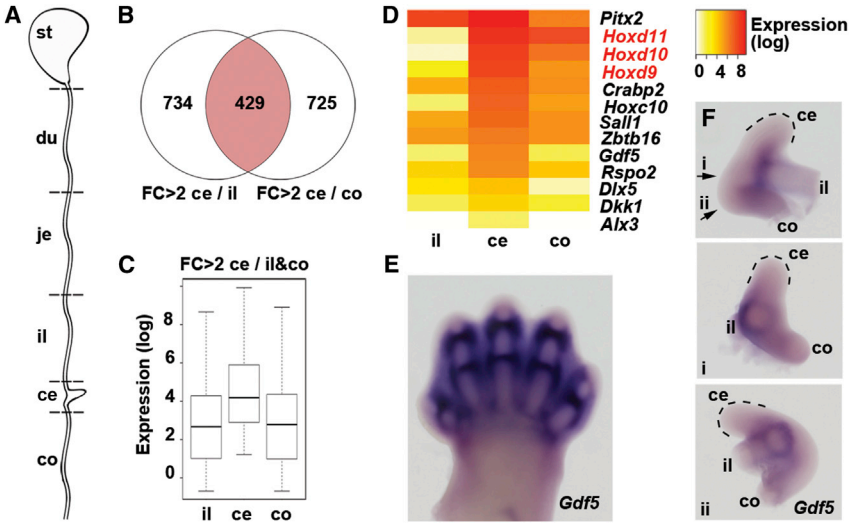

Figure 7. Differential Gene Expression along the Gut

(A) Gut dissections used for RNA sequencing.

(B and C) Venn diagram and boxplots of genes upregulated in the cecum (ce), when compared to either the ileum (il) or the colon (co), with a fold change greater than two and a significant p value. A total of 429 genes have significant increased expression in the cecum.

(D) Gene ontology terms associated with those genes upregulated in the cecum (enrichment over five, significant p value; see also Table S4) reveal several groups related with limb development and morphology, as exemplified with "limb development" (D).

(E and F) *Gdf5* gene expression in E13.5 old ceca and distal limb. *Gdf5* expression is observed in presumptive cells forming the ileocecal sphincter, the junction between the ileum and the colon, as well as in future digital joints.

See also Table S4.

DISCUSSION

Origin of the Cecum: a Budding Toolkit?

The emergence of the cecum and its capacity to digest cellulose was certainly an important step in the adaptation of animals to a vegetarian diet and hence in the colonization of novel biotopes. The presence of a cecum is closely associated with a regionalization of the intestines as it buds out exactly at the transition between the small and the large intestines. *Hox* genes are likely involved in this transition, as suggested by their strong expression right at this ileocolonic boundary. A particular combination of *Hox* proteins there could directly or indirectly trigger the activation of a signaling system, leading to the outgrowth of a bud. Since *Fgf10* and *Fgfr2b* are no longer expressed in *Pitx2* null mice, which are lacking a cecum (Nichol and Saijoh, 2011), it is possible that a local combination of HOX proteins activates *Pitx2*, which may subsequently control the production of *Fgf10*. In support of this, ectopic *Hoxd12* expression, which is known to act as a dominant negative over more anterior HOX proteins, leads to cecum agenesis along with the disappearance of both *Pitx2* and *Fgf10* transcripts (Zacchetti et al., 2007).

In our transcriptome analyses, many of the genes upregulated in the cecum were previously known for their important functions during limb bud development. Besides *Hoxd9*, *Hoxd10*, and *Hoxd11*, these included *Pitx2* (Marcil et al., 2003), *Crabp2* (Lampron et al., 1995), *Dlx5* (Robledo et al., 2002), *Zbtb16* (*Plzf*) (Barna et al., 2000), *Sall1* (Kawakami et al., 2009), or the mouse *R-spondin2* gene (*Rspo2*), known to be required for the maintenance of the limb apical ectodermal ridge (AER) during budding (Nam et al., 2007). Genes involved in retinoic acid signaling were also scored (Table S4). Of note, *Gdf5*, known for its specific function in appendicular joint formation (Storm et al., 1994) is expressed in cells labeling the future junction between the ileum and the colon, the presumptive ileo-cecal valve.

These unexpected similarities may reflect the sharing of basic processes. Limb buds are indeed initially positioned along the body axis by precise combinations of HOX proteins. Also, although limb budding is induced at least in part by triggering the *Fgf* signaling pathway, their growth and patterning largely depend upon distinct HOX proteins. Early limbs require the activity of *Hoxd9*, *Hoxd10*, and *Hoxd11*, whereas both *Hoxd12* and *Hoxd13* must be kept silenced (Andrey et al., 2013). This mechanistic parallelism between two fundamentally different structures may illustrate the existence of a budding toolkit for the developing embryo implemented during the emergence of various "appendages."

We also show that differential *Hox* gene expression exists along the gut of hedgehogs, which are omnivorous animal and which do not display a clear morphological regionalization of their intestinal tract. Although we cannot exclude that the exact HOX combination necessary to elicit the signaling response leading to the budding of the cecum is not found in hedgehogs, the difference observed in the segmental organization of its gut may instead rely on different interpretation of this patterning signal rather than on the signal itself. For instance, the absence of cecum in hedgehogs may be caused by a variation in any of the intermediate factors that would interpret the initial HOX combination to give the signal for budding, such as members of the *Fgf* or *Bmp* families, or like *Pitx2*.

A Domain-Specific Regulation

Enhancers controlling *Hoxd* genes in the cecum are located telomeric to the gene cluster, again similar to our budding limbs (Andrey et al., 2013; Spitz et al., 2005), even though regulatory modalities are clearly distinct. 4C experiments defined strong contact points within the gene desert, but only for genes transcribed in the cecum. In contrast, the silenced *Hoxd13* contacted the other side of the locus. Therefore, active and inactive genes occupy different spatial domains, as in the trunk (Noordermeer et al., 2011) or in digits (Montavon et al., 2011). When the strongest positive contact point was used as bait, it faithfully respected the boundary between active and inactive genes (i.e., between *Hoxd12* and *Hoxd11*), illustrating the robustness of the mechanism that prevents both *Hoxd12* and *Hoxd13* to be expressed in the cecum.

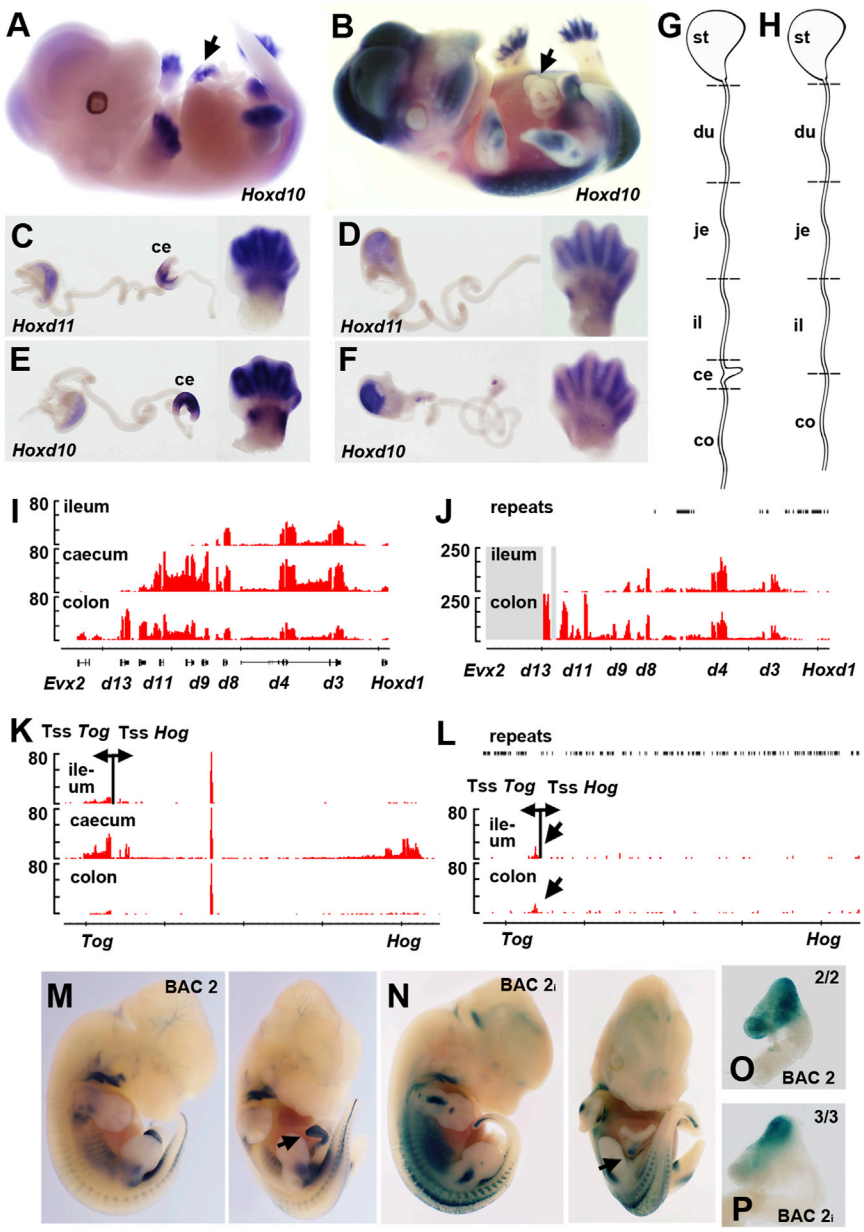

Figure 8. *Hoxd* Genes in the Hedgehog Gut and Transgenic Hedgehog Sequences

(A–F) *Hoxd10* and *Hoxd11* expression in mouse (A, C, and E) and hedgehog (B, D, and F) embryos at comparable developmental stages. *Hoxd10* is detected in the mouse limbs, the genital bud, the trunk, and the cecum (arrow in A). Hedgehog embryos are also positive for these expression sites, except for the gut (arrow in B), where neither *Hoxd10* nor *Hoxd11* is expressed at detectable levels. Limb buds are used as positive controls.

(G–L) Schemes of gut dissections and RNA profiles over the *HoxD* cluster (I and J) and around the start sites of *Hog* and *Tog* (arrows) (K and L) for the developing mouse (I and K) and hedgehog (J and L) gut samples; stomach (st), duodenum (du), jejunum (je), ileum (il), cecum (ce), and colon (co). Similar *HoxD* profiles are scored for the ileum and for the colon samples in both species, suggesting that *Hoxd* gene colinearity is implemented in the developing hedgehog gut. Although the strong signals found over *Hog* and *Tog* lncRNAs in the mouse cecum are not detected in any of the dissected piece of the hedgehog gut, significant signals in the hedgehog "ileum" and "colon" (L), arrow) indicate that at least the *Tog* lncRNA or Tss must be present in this species. Hedgehog-specific repeats are indicated with black boxes, genomic sequences unavailable for RNA detection in gray. See also Extended Experimental Procedures.

(M–P) Comparison in *LacZ* staining between either the mouse (M) or the hedgehog (N) BAC clones when introduced into transgenic mice. The position of the mouse BAC 2 clone (M) and comparable position of the orthologous hedgehog BAC 2i clone (N) are shown in Figure S5. Expression patterns are comparable, including in proximal limbs and in the cecum (arrow in M and N), as exemplified with the enlargements of ceca staining for the mouse (O) and hedgehog (P) BACs.

See also Figure S5 and Table S1.

Both the extent of the telomeric interaction domain and its boundaries around *Hoxd11* and the *Hog* and *Tog* Tss correspond to a mapped topological domain (Andrey et al., 2013; Dixon et al., 2012), as defined by large regions where enhancers-promoters interactions are privileged (Nora et al., 2013; Nora et al., 2012; Shen et al., 2012). This was confirmed by the interaction profiles, which revealed biased interactions within this domain. Although the *Hog* Tss interacted mostly with its centromeric neighborhood, its transcript elongated toward the telomeric side, into a distinct topological domain. This suggests that these domains restrict 3D interactions but do not delineate the resulting transcriptional activity.

The silencing of posterior genes coincides with the presence of H3K27me3 marks covering this part of the cluster, whereas active genes were decorated with H3K4me3. Whether H3K27me3 marks play an active role in preventing gene activation or instead are a mere readout of the repression is unclear. However, the fact that the *Hoxd13* promoter responded to these enhancers after the inversion of both *Hoxd11* and *Hoxd12* (Kmita et al., 2000) indicated that the repressive mechanism is rather flexible and can be alleviated for example by having transcripts reading through this promoter on the opposite strand and clearing out repressor molecules (Bertani et al., 2011; Petruk et al., 2006; Wang et al., 2011). Cecum enhancers may thus preferentially target their activities on the *Hoxd9* to *Hoxd11* interval due to a local chromatin microarchitecture, which may act as a "pocket" rather than solely to sequence-specific contacts with promoter regions.

Hog, *Tog*, and the Cecum Regulatory Landscape

The combination of ChIP and 4C data, phylogenetic footprints and transgenic approaches allowed us to isolate at least eight

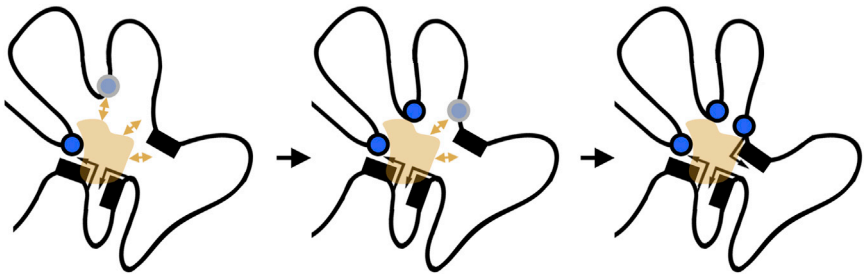

Figure 9. A Model for the Evolution of Multiple Cooperative Enhancers

The occurrence of a stable or dynamic preformed chromatin microarchitecture, including part of the *HoxD* cluster and the neighboring telomeric gene desert (topological domain in Dixon et al., 2012), provides opportunities for a transcription factor, or complex thereof (orange), recruited to an original enhancer sequence (blue), to establish a proximity with a low-affinity sequence (left; shaded blue). Additional contacts may result with surrounding DNA sequences (orange arrows). If a productive interaction is established, for example by further enhancing the transcriptional outcome, the maximization of this interaction may be selected, leading to the progressive individualization of yet additional genuine enhancer sequences (middle and right). This may lead to the stabilization and re-enforcement of the domain and may help recruit more transcription units by inducing slight changes in the overall architecture (right). In this view, the interaction domain is a playground for the emergence of novel enhancer sequences.

candidate cecum enhancer sequences. The signal was stronger with a BAC than with an isolated sequence and the expression intensities seemed to vary along various parts of the cecum, suggesting that a combination of enhancers is required to express *Hoxd* genes both at the right places and in the appropriate amounts. Comparable results were obtained when the cognate hedgehog sequences were used. The multiplicity of structures requiring *Hoxd* function and controlled from this neighborhood of the gene cluster (Spitz et al., 2005) could explain the functional conservation of these enhancers and thus illustrate the parsimony of regulatory circuitries. It is thus plausible that hedgehogs have the necessary regulatory circuitry to control *Hoxd* genes in a developing cecum, should the appropriate upstream signals be released.

The strongest point of contact with active genes was the shared transcription start site of the *Hog* and *Tog* lncRNAs. Yet neither *Hog* nor *Tog* required the *cis* proximity of the *HoxD* cluster to respond to the appropriate cecum enhancers. However, when the main *Hoxd9* to *Hoxd11* targets were deleted, the transcription of *Hog* was almost entirely abolished despite the presence of enhancers, demonstrating a strong interaction in *cis* between these two genetic loci. We propose that active *Hoxd* genes interact with several enhancers as well as with the *Hog* and *Tog* Tss (Figure 5H, left), which may be used as a pioneer contact with *HoxD*. This structure exists also in nonexpressing cells, though not with the exact same microarchitecture, such that upstream factors would consolidate or slightly modify a preexisting spatial conformation. In this view, enhancers would mostly interact with *Hoxd9* to *Hoxd11* and keep weaker contacts with those *Hoxd* genes labeled by H3K4me1, and the posterior *Hoxd12* and *Hoxd13* are not engaged into these interactions (Figure 5H, left).

Upon deletion of the *Hoxd* targets, the interactions with the lncRNAs Tss and the enhancers are perturbed, leading to a spatial reorganization with a strong contact now established with *Hoxd12*, a reorganization leading to the downregulation of *Hog* (Figure 5H, right). In this view, the deleted *HoxD* cluster would act as a "regulatory dominant-negative" configuration impairing the transcription of *Hotdog*. When the regulatory landscape is separated from the *HoxD* cluster, *Hoxd* genes become silenced whereas the Tss can still interact with the enhancer sequences (Figure 5H, middle). In this model, the interaction domain (Figure 5H, orange) can either represent direct physical contacts between the various enhancers and their target promoters or a more diffuse platform recruiting the necessary factors and thus increasing their concentration around target promoters.

Alternatively, *Hog* and *Tog* may be required in *trans* to enhance *Hoxd* gene transcription, even though this property may necessitate genomic proximity in *cis* (Ørom et al., 2010; Ørom and Shiekhattar, 2011), or genome-wide at multiple loci. Our comparative transcriptome results do not favor the latter possibility, since the absence of both *Hog* and *Tog* did not drastically change the transcription landscape in the budding cecum. Although a targeted and localized effect can never be ruled out, the function of *Hog* and *Tog* as global regulators of gene expression during cecal development is thus unlikely.

The Evolution of Enhancer Cooperativity

We show that multiple enhancers distributed over hundreds of kilobases are all capable, to some extent, of inducing expression in the cecum. These and recent results (Marinić et al., 2013; Montavon et al., 2011) raise the question of how such a cooperative system could evolve. The fact that topological domains are observed even in the absence of active transcription (Andrey et al., 2013; Dixon et al., 2012; Montavon et al., 2011; Nora et al., 2012) may increase the probability for a (group of) transcription factor(s) to come to the vicinity of both a target promoter and a particular sequence within the interaction domain. Once a productive interaction is established, every variation within this environment, which would increase the general transcriptional readout of this interaction domain, may be selected, and hence topological domains could be a playground for the emergence of novel regulatory sequences by facilitating the selection of novel protein-DNA contacts (Figure 9). Such a process may drive the evolution of new enhancer sequences displaying moderate activities when isolated from their contexts, but with a cooperative potential when active together. This is what our different transgenic constructs suggest, with BAC clones generating a more robust and complete expression pattern than the isolated individual enhancer sequences.

EXPERIMENTAL PROCEDURES

Mouse Strains

The origin and references to all mutant alleles are given in the Extended Experimental Procedures. The *Ulnaless* strain was purchased from the Jackson Laboratory (http://jaxmice.jax.org/strain/000557). Mice were handled following

the guidelines of the Swiss law on animal protection, with the requested authorization (to D.D.).

In Situ Hybridization

Whole-mount and on-section in situ hybridizations were performed according to standard protocols. *Hoxd* and *Evx2* probes were described before (*Hoxd3*, Condie and Capecchi, 1993; *Hoxd4*, Featherstone et al., 1988; *Hoxd8*, Izpisùa-Belmonte et al., 1990; *Hoxd9*, Zappavigna et al., 1991; *Hoxd10* and *Hoxd11*, Gérard et al., 1996; *Hoxd12*, Izpisúa-Belmonte et al., 1991; *Hoxd13*, Dollé et al., 1991; *Evx2*, Hérault et al., 1996). *Gdf5* was a gift from A. Kan. For other probes, see the Extended Experimental Procedures.

Reverse Transcription and qPCR

Embryonic tissues were dissected and stored in RNAlater reagent (QIAGEN) until genotyped. After disruption with PT1200E Polytron (Kinematica), RNA was isolated using the RNeasy micro- and minikit (QIAGEN). All tissues were reverse transcribed using random primers (QIAGEN) and SuperScript III RT (Invitrogen). Real-time PCR primers were designed using Primer Express 2.0 software (Applied Biosystems) and cDNA was PCR amplified using EXPRESS SYBR GreenER (Invitrogen) with a CFX96 Real-Time System (Bio-Rad). Expression changes were normalized to *Rps9* and analyzed using a Student's t test. For *Evx2*, *Hoxd9* to *Hoxd13*, and the corresponding standard, primers were described before (Montavon et al., 2008). See also the Extended Experimental Procedures.

RNA Sequencing

For cecum transcriptome profiling, total RNA was depleted of rRNA using Ribo-Zero rRNA Removal Kit (Epicenter). For comparisons of transcriptomes along the ileum, cecum, and colon of mice and hedgehogs, polyadenylated RNA was prepared using TruSeq RNA Sample Preparation (Illumina). Libraries and clusters were prepared and sequenced with Illumina Genome Analyzer IIx and paired-end sequencing of 80 for cecum transcriptome profiling to 100 bp for comparisons of transcriptomes along the gut of mice and hedgehogs. Mouse reads were mapped to the mouse genome (mm9) using HTSstation (http://htsstation.vital-it.ch/), with the maximum number of multiple hits limited to five. Hedgehog reads were mapped onto a chimeric genome consisting of *E. europaeus* Scaffolds and GeneScaffolds (from Ensembl), together with the genomic sequence of the *HoxD*, *Hotdog*, and *Twin of Hotdog* loci (de novo sequencing of *A. albiventris* BACs). Reads mapping within hedgehog-specific repeats (RepeatMaker) were removed along with 50 bp of adjacent DNA to remove unspecific transcription signals. Cecum transcriptome data were analyzed with SOAPsplice (http://soap.genomics.org.cn/ soapsplice.html) to uncover splicing junctions. The number of threads was set to 10, the insert length of paired-end reads to 190 and the minimum distance between paired-end reads to 80. Splicing junctions were verified using RLM-RACE. See also the Extended Experimental Procedures.

ChIP on chip

Cecum and brain were dissected from E13.5 embryos, fixed in 1%formaldehyde for 15 min at room temperature, and washed three times with cold phosphate buffer solution (PBS). Pools of fifty ceca or of two brains were used for each experiment. Chromatin immunoprecipitation (ChIP) was performed according to Lee et al. (2006b) with 2–6 μg of RNAPII (8WG16, Covance), H3K4me1 (ab8895, Abcam), H3K4me3 (07-473, Millipore), or H3K27me3 (17-622, Millipore) antibodies and EZview Red protein G/A Affinity Gel (Sigma) /Dynabeads M280 Sheep anti-mouse immunoglobulin G (Invitrogen). Immunoprecipitated chromatin and whole-cell extract DNA (input) was amplified using ligation-mediated PCR (Lee et al., 2006b). PCR fragments were fragmented, labeled, and hybridized to custom tiling arrays (Soshnikova and Duboule, 2009). For each sample and antibody, ChIP-chip was repeated once. Array data were normalized with input replicates and scaled to feature intensity of 100 using TAS software (Affymetrix). PM-MM pairs mapping within a sliding window of 250 bp was used.

Chromosome Conformation Capture

Chromosome conformation capture was performed according to Hagège et al. (2007) and Simonis et al. (2006). Nuclei from pools of 100 ceca or two brains were digested with NlaIII (New England Biolabs) and ligated with T4 DNA ligase HC (Promega) in diluted conditions to promote intramolecular ligation. Sequences ligated to fragments of interest were digested again with DpnII (New England Biolabs), ligated with T4 DNA ligase HC (Promega) in diluted conditions and amplified using AmpliTaq DNA polymerase (Applied Biosystems) and inversed PCR primers flanked with adaptors allowing multiplexing. *Hoxd4*, *Hoxd9*, *Hoxd11*, and *Hoxd13* PCR primers were described before (Noordermeer et al., 2011). 3C PCR fragments were sequenced with Illumina. See also the Extended Experimental Procedures.

BAC Transgenesis

Mouse BAC RP23-211O14, RP23-417J8, RP24-302M7, fosmid WI1-606G14, and *A. albiventris* BACs LB4-247G8 were purchased form BACPAC Resources Center (http://bacpac.chori.org/). We used EL250 bacteria (Lee et al., 2001) and recombined a *LacZ* reporter placed under the control of the minimal promoter of the *betaglobin* gene, followed by an *Ampicillin* resistance cassette and its adjacent FRT sequences, into the BACs using 1-kb-long homology arms. The FRT Ampicillin cassette was removed by inducing *Flp* expression with 0.1% L-arabinose for 1 hr. Mutant BACs were produced using an *Ampicillin* cassette flanked with mutated FRT sequences. Homologous recombination was verified by PCR analysis and BACs were analyzed with restriction enzyme fingerprinting. BACs were isolated with QIAGEN Large-Construct Kit and purified with Microcon YM-30 (Amicon) and Spin-X columns (Costar). BACs were injected into fertilized oocytes. Embryos were collected at E12.5 and stained for β-galactosidase activity.

Lentivirus-Mediated Transgenesis

An *attR1-ccdB-Chlo-attR2* cassette was PCR amplified from Gateway pAD-Dest vector (Invitrogen) with the following primers (underscored) flanked with an XhoI restriction site (bold): GW-pAD-Dest forward: 5′-AAA**CTCGAGAAA**ACAAGTTTGTACAAAAAAGCTG-3′ reverse: 5′-AAA**CTCGAGAAA**ACCACTTTGTACAAGAAAGCTG-3′. The XhoI-digested cassette was inserted into the unique XhoI site in pRRLbLac vector (Friedli et al., 2010) to create a Gateway-adapted pRRLbLac. DNA segments were PCR amplified from BACs using primers flanked with *attB1/2* sequences and the Expand Long Template PCR system (Roche). *att*B-flanked PCR products were initially transferred to Gateway pDONR221 vector (Invitrogen) with BP Clonase Enzyme mix (Invitrogen) to create a Gateway entry clone, then to the Gateway-adapted pRRLbLac with LR Clonase Enzyme mix (Invitrogen). Virus production and injection into fertilized mouse oocytes was performed as in (Barde et al., 2010; Friedli et al., 2010). Embryos were collected at E12.5 and stained for β-galactosidase activity.

ACCESSION NUMBERS

Sequencing and microarray data have been deposited to GenBank under accession number GSE45242. The *Hog* and *Tog* lncRNA sequences have been deposited to GenBank under accession numbers BK008787 and BK008788, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Results, Extended Experimental Procedures, five figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.09.002.

AUTHOR CONTRIBUTIONS

S.D., T.M., and D.D. designed the experiments, which were carried out by S.D., with the help of E.J. M.L. helped with bioinformatics analyses. A.T. and M.M. produced and contributed hedgehog material and embryos. S.D. and D.D. wrote the paper.

ACKNOWLEDGMENTS

We thank U. Gunthert for the *CD44* mutant allele, Laurie Thevenet for the *inv(Mtx2-CD44)* line, and A. Kan for the *Gdf5* probe. We thank B. Mascrez,

T. Nguyen Huynh, and S. Gitto for their help with mice as well as D. Noordermeer for the design of 4C primers and other members of the Duboule laboratories for discussions and reagents. We also thank I. Barde for her help with transgenics and J. Deschamps, A. Necsuela, and J. Zakany for their comments on the manuscript. This work was supported by funds from the University of Geneva, the Ecole Polytechnique Fédérale, Lausanne, the Swiss National Research Fund, the National Research Center (NCCR) "*Frontiers in Genetics*," the European Research Council (ERC), and the FP7 program IDEAL.

Received: February 6, 2013
Revised: August 7, 2013
Accepted: September 5, 2013
Published: September 26, 2013

REFERENCES

Al Alam, D., Sala, F.G., Baptista, S., Galzote, R., Danopoulos, S., Tiozzo, C., Gage, P., Grikscheit, T., Warburton, D., Frey, M.R., and Bellusci, S. (2012). FGF9-Pitx2-FGF10 signaling controls cecal formation in mice. Dev. Biol. *369*, 340–348.

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

Aubin, J., Déry, U., Lemieux, M., Chailler, P., and Jeannotte, L. (2002). Stomach regional specification requires Hoxa5-driven mesenchymal-epithelial signaling. Development *129*, 4075–4087.

Barde, I., Salmon, P., and Trono, D. (2010). Production and titration of lentiviral vectors. Curr. Protoc. Neurosci. *Chapter 4*, Unit 4.21.

Barna, M., Hawe, N., Niswander, L., and Pandolfi, P.P. (2000). Plzf regulates limb and axial skeletal patterning. Nat. Genet. *25*, 166–172.

Bertani, S., Sauer, S., Bolotin, E., and Sauer, F. (2011). The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. Mol. Cell *43*, 1040–1046.

Boulet, A.M., and Capecchi, M.R. (1996). Targeted disruption of hoxc-4 causes esophageal defects and vertebral transformations. Dev. Biol. *177*, 232–249.

Burns, R.C., Fairbanks, T.J., Sala, F., De Langhe, S., Mailleux, A., Thiery, J.P., Dickson, C., Itoh, N., Warburton, D., Anderson, K.D., and Bellusci, S. (2004). Requirement for fibroblast growth factor 10 or fibroblast growth factor receptor 2-IIIb signaling for cecal development in mouse. Dev. Biol. *265*, 61–74.

Condie, B.G., and Capecchi, M.R. (1993). Mice homozygous for a targeted disruption of Hoxd-3 (Hox-4.1) exhibit anterior transformations of the first and second cervical vertebrae, the atlas and the axis. Development *119*, 579–595.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dollé, P., Izpisúa-Belmonte, J.C., Boncinelli, E., and Duboule, D. (1991). The Hox-4.8 gene is localized at the 5′ extremity of the Hox-4 complex and is expressed in the most posterior parts of the body during development. Mech. Dev. *36*, 3–13.

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. Dev. Suppl., 135–142.

Duboule, D., and Morata, G. (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. Trends Genet. *10*, 358–364.

Fairbanks, T.J., Kanard, R.C., De Langhe, S.P., Sala, F.G., Del Moral, P.M., Warburton, D., Anderson, K.D., Bellusci, S., and Burns, R.C. (2004). A genetic mechanism for cecal atresia: the role of the Fgf10 signaling pathway. J. Surg. Res. *120*, 201–209.

Featherstone, M.S., Baron, A., Gaunt, S.J., Mattei, M.G., and Duboule, D. (1988). Hox-5.1 defines a homeobox-containing gene locus on mouse chromosome 2. Proc. Natl. Acad. Sci. USA *85*, 4760–4764.

Friedli, M., Barde, I., Arcangeli, M., Verp, S., Quazzola, A., Zakany, J., Lin-Marq, N., Robyr, D., Attanasio, C., Spitz, F., et al. (2010). A systematic enhancer screen using lentivector transgenesis identifies conserved and non-conserved functional elements at the Olig1 and Olig2 locus. PLoS ONE *5*, e15741.

Gérard, M., Chen, J.Y., Gronemeyer, H., Chambon, P., Duboule, D., and Zákány, J. (1996). In vivo targeted mutagenesis of a regulatory element required for positioning the Hoxd-11 and Hoxd-10 expression boundaries. Genes Dev. *10*, 2326–2334.

Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat. Protoc. *2*, 1722–1733.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Hérault, Y., Hraba-Renevey, S., van der Hoeven, F., and Duboule, D. (1996). Function of the Evx-2 gene in the morphogenesis of vertebrate limbs. EMBO J. *15*, 6727–6738.

Hérault, Y., Fraudeau, N., Zákány, J., and Duboule, D. (1997). Ulnaless (Ul), a regulatory mutation inducing both loss-of-function and gain-of-function of posterior Hoxd genes. Development *124*, 3493–3500.

Izpisùa-Belmonte, J.C., Dollé, P., Renucci, A., Zappavigna, V., Falkenstein, H., and Duboule, D. (1990). Primary structure and embryonic expression pattern of the mouse Hox-4.3 homeobox gene. Development *110*, 733–745.

Izpisúa-Belmonte, J.C., Falkenstein, H., Dollé, P., Renucci, A., and Duboule, D. (1991). Murine genes related to the Drosophila AbdB homeotic genes are sequentially expressed during development of the posterior part of the body. EMBO J. *10*, 2279–2289.

Kawakami, Y., Uchiyama, Y., Rodriguez Esteban, C., Inenaga, T., Koyano-Nakagawa, N., Kawakami, H., Marti, M., Kmita, M., Monaghan-Nichols, P., Nishinakamura, R., and Izpisua Belmonte, J.C. (2009). Sall genes regulate region-specific morphogenesis in the mouse limb by modulating Hox activities. Development *136*, 585–594.

Kawazoe, Y., Sekimoto, T., Araki, M., Takagi, K., Araki, K., and Yamamura, K.-i. (2002). Region-specific gastrointestinal Hox code during murine embryonal gut development. Dev. Growth Differ. *44*, 77–84.

Kmita, M., and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. Science *301*, 331–333.

Kmita, M., Kondo, T., and Duboule, D. (2000). Targeted inversion of a polar silencer within the HoxD complex re-allocates domains of enhancer sharing. Nat. Genet. *26*, 451–454.

Kondo, T., Dollé, P., Zákány, J., and Duboule, D. (1996). Function of posterior HoxD genes in the morphogenesis of the anal sphincter. Development *122*, 2651–2659.

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell *78*, 191–201.

Lampron, C., Rochette-Egly, C., Gorry, P., Dollé, P., Mark, M., Lufkin, T., LeMeur, M., and Chambon, P. (1995). Mice deficient in cellular retinoic acid binding protein II (CRABPII) or in both CRABPI and CRABPII are essentially normal. Development *121*, 539–548.

Lee, E.C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D.A., Court, D.L., Jenkins, N.A., and Copeland, N.G. (2001). A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics *73*, 56–65.

Lee, A.P., Koh, E.G., Tay, A., Brenner, S., and Venkatesh, B. (2006a). Highly conserved syntenic blocks at the vertebrate Hox loci and conserved regulatory elements within and outside Hox gene clusters. Proc. Natl. Acad. Sci. USA *103*, 6994–6999.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006b). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. *1*, 729–748.

Marcil, A., Dumontier, E., Chamberland, M., Camper, S.A., and Drouin, J. (2003). Pitx1 and Pitx2 are required for development of hindlimb buds. Development *130*, 45–55.

Marinić, M., Aktas, T., Ruf, S., and Spitz, F. (2013). An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. Dev. Cell *24*, 530–542.

Montavon, T., Le Garrec, J.-F., Kerszberg, M., and Duboule, D. (2008). Modeling Hox gene regulation in digits: reverse collinearity and the molecular origin of thumbness. Genes Dev. *22*, 346–359.

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., and Duboule, D. (2011). A regulatory archipelago controls Hox genes transcription in digits. Cell *147*, 1132–1145.

Nam, J.S., Park, E., Turcotte, T.J., Palencia, S., Zhan, X., Lee, J., Yun, K., Funk, W.D., and Yoon, J.K. (2007). Mouse R-spondin2 is required for apical ectodermal ridge maintenance in the hindlimb. Dev. Biol. *311*, 124–135.

Nichol, P.F., and Saijoh, Y. (2011). Pitx2 is a critical early regulatory gene in normal cecal development. J. Surg. Res. *170*, 107–111.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011). The dynamic architecture of Hox gene clusters. Science *334*, 222–225.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nora, E.P., Dekker, J., and Heard, E. (2013). Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods? Bioessays *35*, 818–828.

Ørom, U.A., and Shiekhattar, R. (2011). Noncoding RNAs and enhancers: complications of a long-distance relationship. Trends Genet. *27*, 433–439.

Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46–58.

Peichel, C.L., Prabhakaran, B., and Vogt, T.F. (1997). The mouse Ulnaless mutation deregulates posterior HoxD gene expression and alters appendicular patterning. Development *124*, 3481–3492.

Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W., and Mazo, A. (2006). Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. Cell *127*, 1209–1221.

Pitera, J.E., Smith, V.V., Thorogood, P., and Milla, P.J. (1999). Coordinated expression of 3′ hox genes during murine embryonal gut development: an enteric Hox code. Gastroenterology *117*, 1339–1351.

Robledo, R.F., Rajan, L., Li, X., and Lufkin, T. (2002). The Dlx5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. Genes Dev. *16*, 1089–1101.

Sekimoto, T., Yoshinobu, K., Yoshida, M., Kuratani, S., Fujimoto, S., Araki, M., Tajima, N., Araki, K., and Yamamura, K. (1998). Region-specific expression of murine Hox genes implies the Hox code-mediated patterning of the digestive tract. Genes Cells *3*, 51–64.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. *38*, 1348–1354.

Soshnikova, N., and Duboule, D. (2009). Epigenetic temporal control of mouse Hox genes in vivo. Science *324*, 1320–1323.

Spitz, F., Gonzalez, F., Peichel, C., Vogt, T.F., Duboule, D., and Zákány, J. (2001). Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. Genes Dev. *15*, 2209–2214.

Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell *113*, 405–417.

Spitz, F., Herkenne, C., Morris, M.A., and Duboule, D. (2005). Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. Nat. Genet. *37*, 889–893.

Storm, E.E., Huynh, T.V., Copeland, N.G., Jenkins, N.A., Kingsley, D.M., and Lee, S.J. (1994). Limb alterations in brachypodism mice due to mutations in a new member of the TGF beta-superfamily. Nature *368*, 639–643.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Zacchetti, G., Duboule, D., and Zakany, J. (2007). Hox gene function in vertebrate gut morphogenesis: the case of the caecum. Development *134*, 3967–3973.

Zákány, J., and Duboule, D. (1999). Hox genes and the making of sphincters. Nature *401*, 761–762.

Zappavigna, V., Renucci, A., Izpisúa-Belmonte, J.C., Urier, G., Peschle, C., and Duboule, D. (1991). HOX4 genes encode transcription factors with potential auto- and cross-regulatory capacities. EMBO J. *10*, 4177–4187.

Zhang, X., Stappenbeck, T.S., White, A.C., Lavine, K.J., Gordon, J.I., and Ornitz, D.M. (2006). Reciprocal epithelial-mesenchymal FGF signaling is required for cecal development. Development *133*, 173–180.