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

Saskia Delpretti,1 Thomas Montavon,1,3 Marion Leleu,1 Elisabeth Joyce,1 Athanasia Tzika,2 Michel Milinkovitch,2 and Denis Duboule1,2,*

1School of Life Sciences, Ecole Polytechnique Fédérale, Lausanne CH-1015, Switzerland
2Department of Genetics and Evolution, University of Geneva, 1211-Geneva 4, Switzerland
3Present 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
This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

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., 2004), as well as loss- and gain-of-function approaches (e.g., Aubin et al., 2002; Boulet and Capecci, 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 (Zákany 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.
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 non-coding 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 IncRNAs, encoded by opposite DNA strands and sharing the same (or closely located) transcription start site(s). The largest IncRNA was referred to as “HoxD telomeric desert IncRNA operating in the gut” or
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 activated. 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,
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...
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) (Héralt 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.

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 IncRNAs, Hotdog (Hog) and Twin of Hotdog (Tog), with opposite transcriptional directions, harboring 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.
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 IncRNAs (Figures 4A and 4S, 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 IncRNAs 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 IncRNA 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 IncRNAs via enhancer-promoter reallocation. When animals lacking from Hoxd1 to near Hoxd8 [del(1–4i)] were used, the overall transcription profile of both IncRNAs 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 IncRNAs, 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 transcripts 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–ToSB2)] deletion removing most of the telomeric gene desert (Andrey et al., 2013), including all identified cecum enhancers and both Hog and Tog (Figure 6B). Hog transcripts level dropped down to ca. 30% of wild-type levels (Figure 6D), indicating that additional enhancers localize 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
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
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 specimens between the ileum and colon, with no clear boundary between the small and large intestines (Figure S5). We identified and sequenced twelve partially overlapping 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 8G–8I). A specific transcriptional activity was mapped around the transcription start site of *Hog* and *Tog*, 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).
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 Sawai, 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 (Pitz) (Barna et al., 2000), Safl1 (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.
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
hancers, demonstrating a strong interaction in cis of Hog Hoxd two genetic loci. We propose that active the main Hoxd9 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 (Montavon et al., 2008). See also the Extended Experimental Procedures.

Reverse Transcription and qPCR

Embryonic tissues were dissected and stored in RNAlater reagent (QIAGEN) until genotyped. After disruption with PT1200 Polytron (Kinematica), RNA was isolated using the RNaseasy 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 (Epicerent). 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 (Illumina). RNA was prepared using TruSeq RNA Sample Preparation (Illumina). Libraries and clusters were prepared and sequenced with Illumina Genome Analyzer IIx (Illumina).

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 (8W16, 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 data were 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 inverted 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 RP23-211014, RP23-417J8, RP24-302M7, fosmid W1-606G14, and A. albiventris BACs LB4-247G8 were purchased form BACPAC Resources Center (http://bapac.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-Chio-attR2 cassette was PCR amplified from Gateway pAD-Dest vector (Invitrogen) with the following primers (underscored) flanked with an Xhol restriction site (bold): GW-pAD-Dest forward: 5’-AAAATCGAGAAAAAC AAGTTGTACAAAGGGTG-3’ reverse: 5’-AAGCTGAAAAACGACCT TTGTCGAGAACGCTG-3’. The XhoI-digested cassette was inserted into the unique Xhol site in pRRBlac vector (Friedli et al., 2010) to create a Gateway-adapted pRRBlac. DNA segments were PCR amplified from BACs using primers flanked with attR1/2 sequences and the Expand Long Template PCR system (Roche). attB-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 pRRBlac 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 IncRNA 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 bioinformatic 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. Masciez,
REFERENCES


