Cell, Vol. 113, 405–417, May 2, 2003, Copyright ©2003 by Cell Press

A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the *HoxD* Cluster

François Spitz, Federico Gonzalez, and Denis Duboule*
Department of Zoology and Animal Biology
and NCCR Frontiers in Genetics
University of Geneva
Sciences III
Quai Ernest Ansermet 30
1211 Geneva 4
Switzerland

Summary

During limb development, coordinated expression of several *Hoxd* genes is required in presumptive digits. We searched for the underlying control sequences upstream from the cluster and found *Lunapark* (*Lnp*), a gene which shares limb and CNS expression specificities with both *Hoxd* genes and *Evx2*, another gene located nearby. We used a targeted enhancer-trap approach to identify a DNA segment capable of directing reporter gene expression in both digits and CNS, following *Lnp*, *Evx2*, and *Hoxd*-specific patterns. This DNA region showed an unusual interspecies conservation, including with its pufferfish counterpart. It contains a cluster of global enhancers capable of controlling transcription of several genes unrelated in structure or function, thus defining large regulatory domains. These domains were interrupted in the *Ulnaless* mutation, a balanced inversion that modified the topography of the locus. We discuss the heuristic value of these results in term of locus specific versus gene-specific regulation.

Introduction

Mammalian *Hox* genes are essential during development (Krumlauf, 1994). Besides their role in organizing structures along the main body axis, *HoxA* and *HoxD* cluster genes are required for proper development of both limbs and external genitalia (Rijli and Chambon, 1997; Zákány and Duboule, 1999b). Genetic analyses have illustrated this crucial role, showing that in the absence of both *Hoxa11* and *Hoxd11* functions, forearms were virtually absent (Davis et al., 1995), whereas abrogation of both *Hoxa13* and *Hoxd13* induced a complete autopod (hands and feet) agenesis (Fromental-Ramain et al., 1996; Zákány et al., 1997). Five *Hox* genes are required for digit development, one member of the *HoxA* complex (*Hoxa13*) and four contiguous genes of the *HoxD* cluster, from *Hoxd10* to *Hoxd13*. In this latter case, all genes are coexpressed in presumptive digits, with virtually identical profiles, which led to the proposal that a single enhancer sequence would control the four transcription units, thus ensuring coordination in time and space (van der Hoeven et al., 1996). This idea was reinforced when both the number and order of *Hoxd* gene loci were modified and their response to the digit regulation analyzed (Zákány and Duboule, 1996, 1999a; Kondo and Duboule, 1999; Kmita et al., 2002a). These experiments tentatively localized a "digit enhancer" upstream from the cluster, a hypothesis subsequently confirmed by analyzing the expression of a human *HOXD* cluster in transgenic mice (Spitz et al., 2001).

A detailed characterization of this enhancer sequence would be of interest to understand both the ontogenesis of our limbs and their evolutionary history. *Hoxd* genes may be genetically downstream the *sonic hedgehog* (*shh*) signaling pathway (Laufer et al., 1994; Riddle et al., 1993), even though direct evidence is still lacking. Also, their expression patterns in limb buds suggest that the *Fgf* and *Wnt* pathways might be involved too (Capdevila and Izpisua Belmonte, 2001). To clarify these issues, it is important to identify the upstream regulators, hence to access the enhancer sequences. In an evolutionary context, mechanisms at work during digits ontogeny are likely the same than those involved in their emergence in ancestral tetrapods (Shubin et al., 1997). We previously showed that the "digit domain" of tetrapod *Hoxd* genes had no counterpart in developing zebrafish pectoral fins, unlike earlier and more proximal expression of the same genes (Sordino et al., 1995). This gave support to a neomorphic origin of digits, i.e., they appeared de novo as opposed to a transformation of preexisting structures. The agenesis of digits induced by combined *Hox* genes inactivations further supported this view, thus suggesting that posterior *Hoxd* genes were coopted along with the emergence of digits, as targets of a signaling pathway via the presence of a global enhancer sequence.

Recent experiments involving deletions and duplications of *Hoxd* genes have shown that such an enhancer sequence has little target specificity. Removal or addition of endogenous loci led to concurrent regulatory reallocations (Kmita et al., 2002a; Monge et al., 2003), whereas different promoters inserted within or upstream from the *HoxD* cluster readily became expressed in digits with the expected specificity (e.g., Hérault et al., 1999). Consequently, this hypothetical regulatory element seemed to display a large realm of action, leading to a delicate equilibrium between the various promoters located within this interval (Kmita et al., 2002a). However, direct evidence for both the existence and localization of this enhancer remained elusive.

In order to map this regulatory sequence, we embarked on an extensive study of the genomic region flanking the *HoxD* cluster in 5′ (centromeric). We constructed large mouse and human BAC contigs covering this region and could identify a previously unknown gene (*Lunapark*; *Lnp*) 90 kb upstream from the cluster. We show here that *Lnp* is also expressed in developing digits, with the same profile as *Hoxd* genes and *Evx2*, implying that the digit regulation not only concerns *Hoxd* genes and immediate surroundings, but also a phylogenetically unrelated gene lying at some distance.

Experiments using a BAC targeted enhancer-trap approach identified a 50 kb large DNA segment that dis-

*Correspondence: denis.duboule@zoo.unige.ch

played transcriptional enhancer activity corresponding to the expected profile in digits. This fragment showed an unusually high level of sequence identity with its human counterpart and a fair conservation with pufferfish. The functional importance of this region was further demonstrated by both transgenic analyses, which revealed a concentration of global enhancers, and the molecular resolution of the mouse *Ulnaless* (*Ul*) mutation (Hérault et al., 1997; Peichel et al., 1997). *Ul* is a large balanced inversion that disrupted the topographic relationship between this global control region and its target genes, thereby modifying the regulatory landscapes.

Results

Lunapark (*Lnp*)

In the course of our investigations of the genomic sequences determining *HoxD* gene expression, we used partial BAC DNA sequences to establish contigs covering the *HOXD* cluster. We searched for ESTs present in the human contig and overlapping ESTs identified a transcription unit. The 5′ end mapped to human BAC RPC11-514d19, which also contained posterior *HOXD11* genes, whereas the 3′ end was found on BAC RPCI-504o20, which extended further away from the *HOXD* cluster. We identified ESTs for the mouse counterpart, the 5′ end of which matched the sequence of BAC RPCI23-400h17, which also contained the *Hoxd* locus. Therefore, in both human and mouse genomes, the same gene was localized nearby the 5′ end of the *HoxD* cluster and transcribed from the opposite DNA strand (Figure 1A). In human, this gene (KIAA1715) was localized ~80 kb far from *HOXD*, whereas ~90 kb far upstream in the mouse (Figure 1A). No additional gene was found in this region.

The sequence of this transcript revealed 13 exons spanning about 100 kb of DNA (Figure 1A), with several alternative polyadenylation sites. It encodes a protein of unknown function (Figure 1B) and has orthologous counterparts in plants, fungi, and animals such as in *C. elegans*, *Drosophila*, and vertebrates. However, none of these products was reported so far to have a particular function. While interspecies alignments revealed domains of very high conservation amongst the members of this protein family (Figures 1C and 1D), none of them showed any clear-cut similarity with known domains, except for an atypical zinc finger and a peptide possibly related to a transmembrane domain. Because of the presence, in both vertebrates and arthropods, of the peptide LNPARK, we named this gene *Lunapark* (*Lnp*). We failed to find a second *Lnp*-related gene in either the human or the mouse genomes; hence, *Lnp* is a single copy gene in mammals.

Lnp Expression in Developing Digits and External Genitalia

In order to use information about *Lnp* expression to understand more about global regulation in this chromosomal region, we compared the developmental expression of *Lnp* with both *Evx2* and *Hoxd13* and found several similarities. Embryos stained with an *Lnp* RNA probe showed a weak ubiquitous signal. However, we detected a strong expression of *Lnp* in both limb and genital buds (Figure 1F), as is the case for *Evx2* (Figure 1G) and *Hoxd* genes, in particular *Hoxd13* (Figure 1H). In developing limb buds, *Lnp* transcription was first seen in 10.5 days old fetuses, in the posterior distal bud, to subsequently extend throughout the distal aspect, in presumptive digits. This expression was virtually identical to that of *Evx2* and posterior *Hoxd* genes (Figure 1J).

The expression of *Lnp* was further assessed by using an ES cell clone derived from a gene trap screen (Wiles et al., 2000). It contained a β-geo reporter gene, along with a splice acceptor site, into the first intron of *Lnp*. We injected these ES cells into blastocysts and chimeric embryos were stained for lacZ activity. We detected a robust signal in both limb and genital buds (Figure 1E). In addition, a clear signal was scored in the developing central nervous system (Figures 1E–1G; arrowhead). Therefore, *Lnp* is a gene expressed in most tissues at basal level, with reinforcement in distal limb buds, genital bud, and in parts of the CNS. Since the same structures express either *Evx2* or posterior *Hoxd* genes (or both), we concluded that these genes share enhancer sequences. *Lnp* was also strongly expressed in some domains, which were shared neither by *Evx2*, nor by *Hoxd* genes, such as the forebrain, the eyes, and the heart (Figures 1E–1F; red arrows).

To see whether this peculiar functional organization was specific to mammals, we cloned a partial chicken *Lnp* cDNA and found that it mapped very close to the *Hoxd* locus on a chick radiation hybrid map (M. Morisson and F.S., unpublished data). During chick limb development, *Lnp* was expressed with an early pattern at the posterior distal margin and subsequent extension over the distal limb bud (Figures 1K–1L). Therefore, in birds like in mammals, *Lnp* is closely linked to the *Hoxd* locus and coexpressed with the most posterior *Hoxd* genes during digits development. These results confirmed that in tetrapods, the digit enhancer would act over a rather large distance (at least from *Lnp* to *Hoxd10*, ~150 kb) and on a battery of genes not necessarily linked to each other: structurally, phylogenetically, or functionally.

Introduction of a human *HOXD* cluster in transgenic mice revealed that the digit enhancer was not located within the cluster itself (Spitz et al., 2001). We searched for this sequence by using ET recombination (Muyrers et al., 1999; Orford et al., 2000) to introduce lacZ reporter sequences into the human *HOXD11* gene, in a BAC containing the entire *LNP* to *HOXD* intergenic region (Figure 2). Three independent animals transgenic for this BAC were stained at day 12.5. While they all displayed strong lacZ signal in the posterior part of the trunk, as expected for *Hoxd11*, none of them expressed the transgene with the expected profile in digits, and only a posterior-located signal was scored in hindlimbs, as seen before with a short *Hoxd11lac* transgene (Gérard et al., 1993; Spitz et al., 2001). We analyzed five other independent transgenic animals at day 12.5 for *HOXD13* expression using a human-specific probe. While a human *HOXD13* signal was detected in the tail bud of transgenic animals, no signal was ever scored in limb buds of these specimens (except for a faint posterior signal in hindlimbs; Figure 2). We concluded that the digit enhancer was not present on this human BAC, suggesting a location centromeric to *Lnp*.

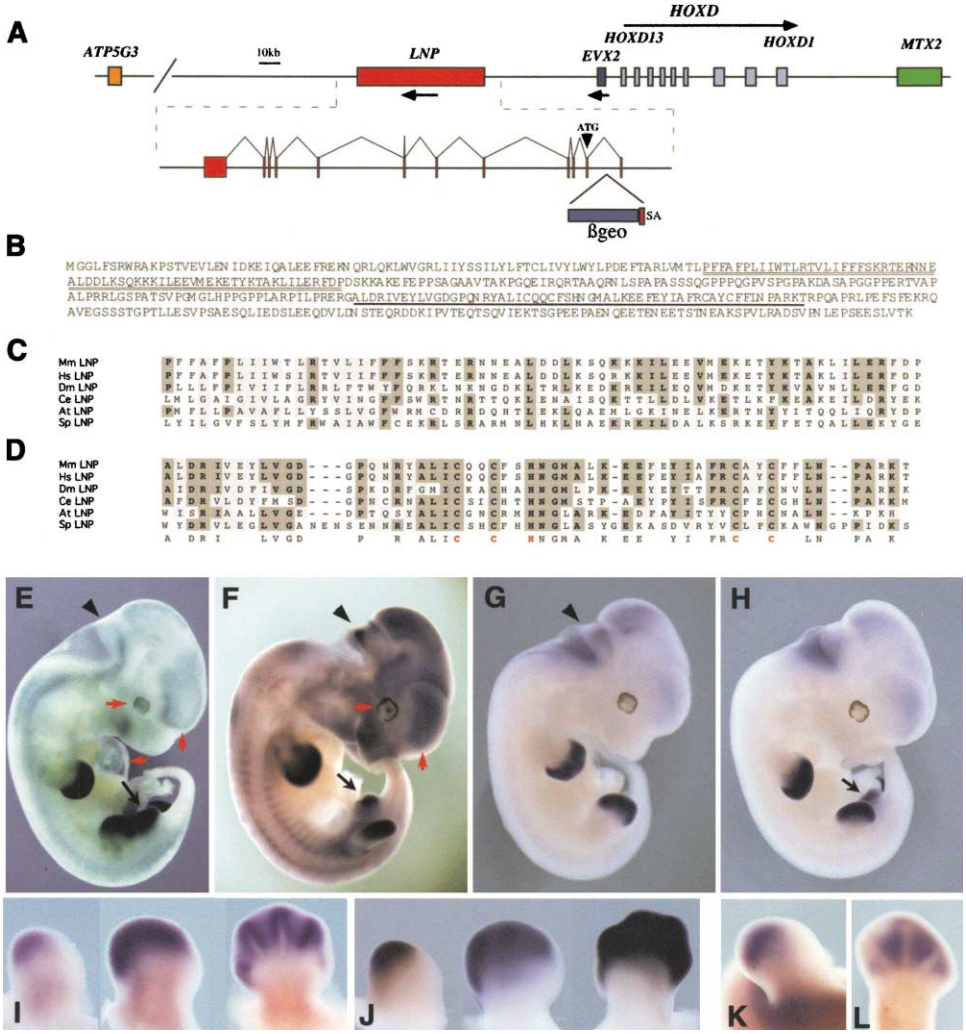

Figure 1. Cloning of *Lunapark*

(A) Organization of the *Hoxd* locus on mouse chromosome 2 (human 2q31), with the positions of *Lunapark* (*Lnp*), *Evx2*, *Mtx2*, and *Atp5g3*. In both species, *Lnp* is composed of 13 exons and is transcribed in the TEL to CEN orientation, as *Evx2* but unlike *Hoxd* genes. The insertion site of the β-geo transgene, along with a splice acceptor site (SA) in ES cells, is indicated between exon 1 and exon 2.

(B) DNA and amino acid sequences of the mouse *Lnp* coding region. Evolutionary conserved domains are underlined (double line for the domain shown in C; thick line for the domain shown in D).

(C and D) Partial comparison between the amino acid sequences of mouse, human, *Drosophila*, *C. elegans*, *Arabidopsis*, and *schizosaccharomymices Lnp*. A consensus sequence is indicated below and residues potentially involved in an atypical zinc finger are in red.

(F) Expression of *Lnp*, compared to that of *Evx2* (G) and *Hoxd13* (H), in day 11.5 embryos.

(E) β-gal staining of a chimeric embryo mostly composed of ES cells in which lacZ coding sequences are inserted within *Lnp*.

(I and J) Time course analysis of *Lnp* (I) and *Hoxd13* (J) expression in developing limb buds (left, day 10.5; middle, 11.5: right, 12.5). *Lnp* is coexpressed with *Evx2* and *Hoxd13* in distal limb buds and emerging external genitalia (black arrow). It is also expressed in the CNS (arrowhead), in a way highly similar to *Evx2*. Red arrows point to expression domains specific to *Lnp* (eyes, heart, and forebrain).

(K and L) Expression of chicken *Lnp* during wing bud development, showing the sharing of the digit enhancer between *Lnp* and *Hoxd* genes is conserved between mammals and birds.

Transposon-Based, Locus-Targeted Enhancer Trap

We designed an approach to allow for rapid labeling and transgenic screen of BACs, based on an enhancer-trap strategy targeted to a particular locus. As a reporter cassette, we used a lac*Z* gene driven by a minimal β-globin promoter (β-lac). This gene was inert, by itself, and required the presence of an active enhancer to become activated (Morgan et al., 1996). To insert this reporter cassette into BAC clones at random positions, we adapted the Tn7 transposition system (Biery et al., 2000). Transposition was induced in vitro, with a low target site preference, and a unique insertion event of the reporter cassette per BAC was routinely obtained.

We used a Tn7β-lac reporter transposon to tag our BACs through in vitro transposition. BACs carrying an insertion of the Tn7β-lac reporter were analyzed to determine both the number and position of the inserted transposon(s), as well as their integrity. More than 80 percent of the clones contained a single insertion and more than 95 percent had no visible genomic re-

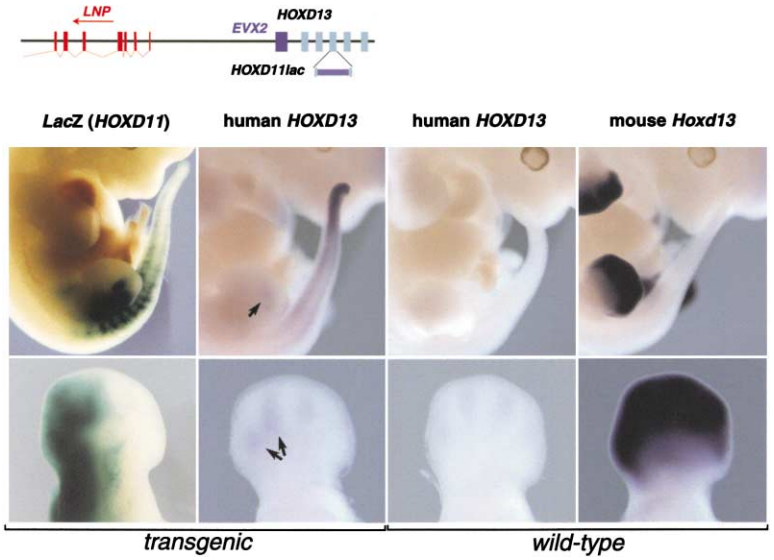

Figure 2. BAC Transgene Analysis

The *LNP-HOXD* intergenic region does not confer appropriate expression in distal limb buds. A human BAC (RPCI-11 514d19) containing lacZ reporter sequences inserted into *HOXD11* was injected into mouse-fertilized eggs. Transgenic animals were stained for β-gal activity and showed the expected *HOXD11* expression in the trunk (left). However, they failed to express this gene in the distal limb bud. Only a posteriorly restricted expression was detected in the hindlimb bud. Five other transgenic founders were analyzed with a human *HOXD13* specific probe and a faint posterior expression was detected in hindlimbs (middle images; arrows), in contrast to the strong distal expression detected by the mouse-specific probe (right).

arrangement. Insertion sites were randomly distributed along the BACs and, out of 20 clones analyzed in a pilot experiment, insertions were recovered nearly every 10 kb along a 175 kb large BAC. We used this system to randomly label BACs before introducing them into mice.

A "Digit Enhancer"

We assumed that the digit enhancer sequence, if present, would activate the reporter transgene due to its poor promoter specificity and its capacity to work at a distance. Because of the synteny with human chromosome 2, we used human BACs covering this interval, which allowed us to monitor the integrity of the inserted material. BACs covering 700 kb around the *HOXD* locus were tagged with the Tn7β-lac reporter and transposition sites determined. Selected clones were injected into fertilized oocytes and founders animals stained for lacZ expression.

As a pilot experiment, we injected a BAC with the β-lac reporter transgene inserted close to the *HOXD9* gene (Figure 3A; #437n19). This BAC extended 150 kb in 3′ (telomeric) of *HOXD1*, at the opposite end of the cluster. Two 12.5-day-old founders showed expression in the trunk, with an anterior limit close to that of *Hoxd9*. As expected, neither one was expressed in limbs (Figure 3A). This indicated that the β-lac reporter gene could be used to screen BACs centromeric to *LNP*.

Two tagged clones of BAC #504o20 with distinct insertions separated by 50 kb (Figure 3; #504o20, β-lac6 and β-lac20) were initially injected. Seven independent founders (four and three, respectively) were recovered and five analyzed at days 11.5 to 12.5. Two permanent lines were established for β-lac6. Transgenic mice carrying this BAC showed strong lacZ staining in both external genitalia and distal limb buds, regardless of the position of the transposon. Expression was also scored in the neural tube, forebrain and, in a majority of cases, in both heart (4 out of 7) and eyes (6 out of 7). Expression in branchial arches and proximal limbs was observed only once. The staining profiles of these transgenes were clearly reminiscent of either posterior *Hoxd* genes (limbs and genitalia), *Evx2* (limbs, genitalia, dorsal neural tube, and hindbrain), or *Lnp* (limbs, genitalia, neural tube, forebrain, heart, and eyes).

We used transgenic lines to analyze limb expression in details. The signal was first detected in limb buds at day 10.5 to 11, restricted to mesenchyme cells of the posterior margin (Figure 3B). It subsequently extended toward both the anterior part and its most distal aspect, along with outgrowth (day 11.5 to 12.5), to progressively cover the entire autopod, the expression being stronger in the mesenchyme immediately underlying both dorsal and ventral ectoderm. At later stages, expression was properly maintained in the autopod (Figure 3B). The initial posterior restriction followed by a progressive anterior and distal extension was exactly as expected for the limb distal domain (phase III in Nelson et al., 1996), characteristic of posterior *Hoxd* genes, *Evx2*, and *Lnp*, thus suggesting the presence on this BAC, of the corresponding enhancer. This pattern was nevertheless not strictly identical to that of *Hoxd13*, as this latter gene did not show reinforced expression in either dorsal or ventral mesenchyme.

Injection of the overlapping BAC #538a12 (Figure 3A) gave three founders with the same expression profile. A strong signal was observed around the proctodeum, as well as in a thin row of cells lying at the ventral aspect of limb bud, right at the junction with the trunk. It also labeled a column of cells below the otic vesicle. Therefore, while this BAC likely contained rather specific enhancers, none of them was related to the *Hoxd*, *Evx2*, or *Lnp* genes. Because no strong expression in distal limb bud was scored with either BAC #538a12 or #514d19, we concluded that the digit enhancer revealed by clone #504o20 was localized within the 100 kb fragment that did not overlap with the other two BACs.

This candidate region was further split into two main fragments with *Bsp*E1, one of which containing the reporter gene. We obtained three transgenic founders with lacZ patterns similar to those scored in transgenic animals containing the entire BAC (Figure 3; #504o20β-lacΔBspE1). In two such animals, integration of the proximal *Bsp*E1 fragment (the other part of the BAC) was not identified by PCR, whereas present in the third case.

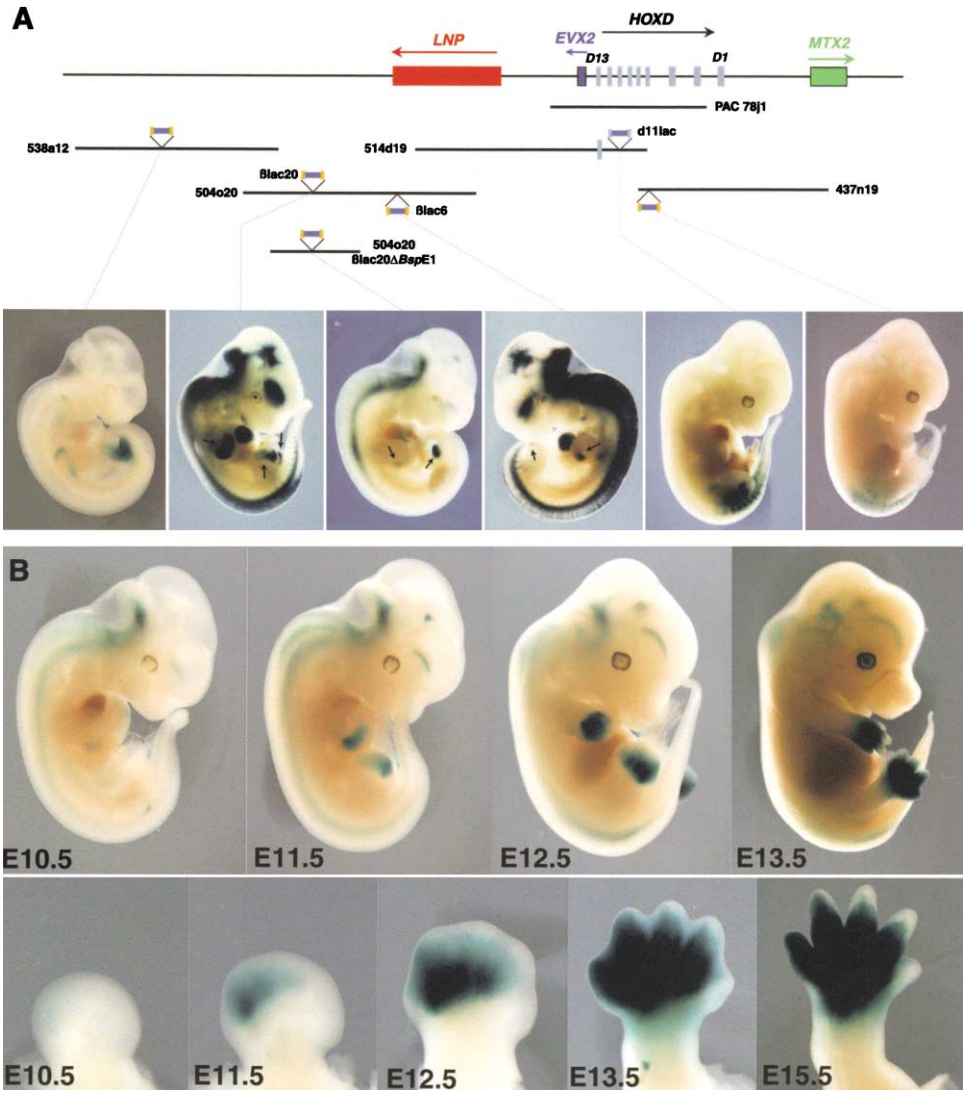

Figure 3. Targeted Enhancer-Trap Analysis of a 0.7 Mb Region around the *HOXD* Locus.

(A) BAC contig covering the human locus was established. PAC 78J1 was unable to recapitulate expression of either *HOXD11* or *HOXD13* in the autopod (Spitz et al., 2001). BAC #514d19 was labeled with lacZ into *HOXD11*. All other BACs were tagged with the Tn7β-lac transposon (flanked by yellow bars). Insertion sites of transposons are shown. Below are transgenic embryos representative of each labeled BAC. Only embryos with BACs #504o20β-lac6 and #504o20β-lac20 showed strong lacZ staining in both distal limbs and genital bud, similar to posterior *Hoxd* genes expression, indicating the presence of the corresponding enhancer sequence, centromeric to *LNP*. BAC fragment #504o20β-lac20Δ*Bsp*E1 also directed expression in these structure, albeit with a lower staining intensity likely due to the earlier developmental stage of the transgenic animals recovered.

(B) Activity of the digit enhancer during mouse development. Two permanent lines were established from independent #504o20β-lac6 transgenic animals. lacZ staining first appeared in 10.5 days old embryos, in ventral neural tube, hindbrain, and anterior dorsal neural tube. At this stage, expression in limb buds was restricted to the most posterior distal part. As the bud developed, expression was progressively reinforced to further extend to the entire autopod, as digit appeared, corresponding to the late phase of *Hoxd* gene activation (Dollé et al., 1989; Nelson et al., 1996), though with a slight delay.

This result indicated that the 62 kb large region containing the 3′ end of *LNP* was dispensable for both limb and genital expression profiles. Altogether, the enhancer region was assigned to a 54 kb large DNA fragment. Staining profiles obtained with BAC #504o20 were identical for either β-lac6 or β-lac20 insertions (Figure 3A).

Neural Enhancers

Lunapark is transcribed in a series of domains in the developing central nervous system, most clearly visible in chimeric animals derived from the *Lnp* gene-trap line (Figure 1A). Staining was scored in the ventral neural tube extending rostrally up to the upper part of the rhombencephalon, in the presumptive cerebellum, in dorsal neurons (Figures 1A, 1D, and 4A), as well as in columns of ventral interneurons (Figure 4E, arrow). Expression was also detected in some regions of both midbrain and forebrain. Some of these domains were strikingly similar to those described for *Evx2* expression (Dollé et al., 1994; Figures 4B and 4F), suggesting again that the associated enhancer sequences are shared between these genes. This was further supported by the

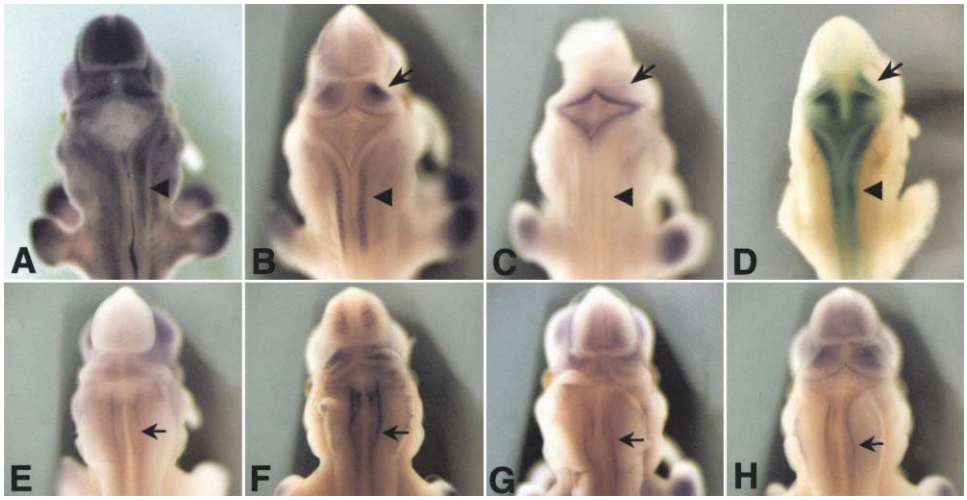

Figure 4. Expression of *Lnp* in CNS Depends in Part upon Enhancers Also Active on *Evx2*

Lnp (A and E) and *Evx2* (B and F) shared similar expression domains in the neural tube and hindbrain of E11.5 embryos. Embryos are shown either untouched (upper line), or with an open neural tube before staining to allow for better probe penetration (bottom line). Arrowheads mark cohorts of neurons from the dorsal neural tube, whereas arrows point to a thin row of cells (V0 interneurons) in the medial part of the neural tube. An enhancer sequence driving expression in dorsal neurons was localized upstream *Lnp*, in #BAC 504o20β-lac, as shown in a transgenic embryo (D). This enhancer was most likely required for *Evx2* expression, as these dorsal neurons are no longer stained for *Evx2* transcripts in *Ul*/Δ*Evx2-d11* embryos (C). In contrast, expression of *Evx2* in presumptive V0 inter-neurons was not abolished in *Ul*/Δ*Evx2-d11* embryos (G), when compared to either wild-type or +/Δ*Evx2-d11* embryos (F and H, respectively).

expression patterns of transgenes that had been previously relocated upstream *Evx2*, and also exhibited expression in some of these domains (Kondo and Duboule, 1999).

Therefore, as for digits and genitalia, neural enhancers with little promoter specificity appeared to lie in this region and regulate both *Lnp* and *Evx2*. Our targeted enhancer trap approach localized some of them, since embryos transgenic for both BAC #504o20 and its short (Δ*Bsp*E1) version showed reporter gene expression in *Evx2*/*Lnp* expression domains. These included dorsal neurons within the neural tube and the most anterior part of the hindbrain, as well as in midbrain and forebrain regions specific to *Lnp* (not shared by *Evx2*; Figures 1, 3, and 4D). In contrast, these regions were not stained in embryos carrying BAC #538a12. Therefore, we concluded that neural enhancers required to regulate both genes in this locus were located near the digit enhancer element. However, some aspects of *Evx2* (and *Lnp*) regulation were not recapitulated by the transgenes, such as in V0 interneurons and in the dorsal midbrain, indicating that the corresponding elements lie elsewhere.

Structure and Regulatory Potential of a Global Control Region (GCR)

We looked for human/mouse sequence conservation at this locus and alignment of the 700 kb region covered by our BAC contig showed strong blocks of conservation corresponding either to the EVX2-HOXD region, to the LNP exons, or to several noncoding sequences (Figure 5A). Interestingly, a cluster of highly conserved, nongenic sequences was found 3′ of LNP, in the presumptive enhancer region within BAC #504o20. Over 40 kilobases, sequence conservation was remarkably high, with two blocks at both extremities of the segment, 4 to 5 kb large each, displaying more than 80 percent identity, with large stretches above 95 percent (Figure 5B). In between, sequences were also well conserved, with gaps mostly due to repeated elements (Figure 5B; black bars).

We assessed the regulatory potential of this region by engineering a deletion of this 40 kb large fragment from BAC #504o20β-lac6, which was shown to contain the digit enhancer activity (Figures 5D and 5E). We obtained five independent transgenic mice and none of them showed any expression in limb buds (Figure 5E). Likewise, most of the expression domains previously scored with the native BAC in forebrain, midbrain, and dorsal neural tube were lost (Figures 5D and 5E). However, mice with the deleted BAC still showed expression in the heart (4 out of 5), the eyes (3 out of 5), and in two lateral spots within the genital bud (4 out of 5). Also, the most ventral aspect of the expression in developing spinal cord and rostral hindbrain was conserved (3 out of 5), indicating that CNS specific control sequences were multiple and not all located within the deleted segment. Because this region appeared to contain a cluster of enhancers capable of controlling several distinct genes over a large region, we referred to it as a global control region (GCR).

Evolutionary Conservation of the GCR

As this level of conservation is unusual for generic enhancer sequences, often composed of short motives displaying moderate conservation, we looked for a counterpart in the *HoxD* locus of teleost fishes, i.e., animals lacking both digits and external genitalia. The *Fugu rubripes* genome (Aparicio et al., 2002) contains a single *HoxD* cluster lacking *Hoxd13*, *Hoxd8*, and *Hoxd1*. However, upstream *Hoxd12*, *Evx2*, and *Lnp* orthologous

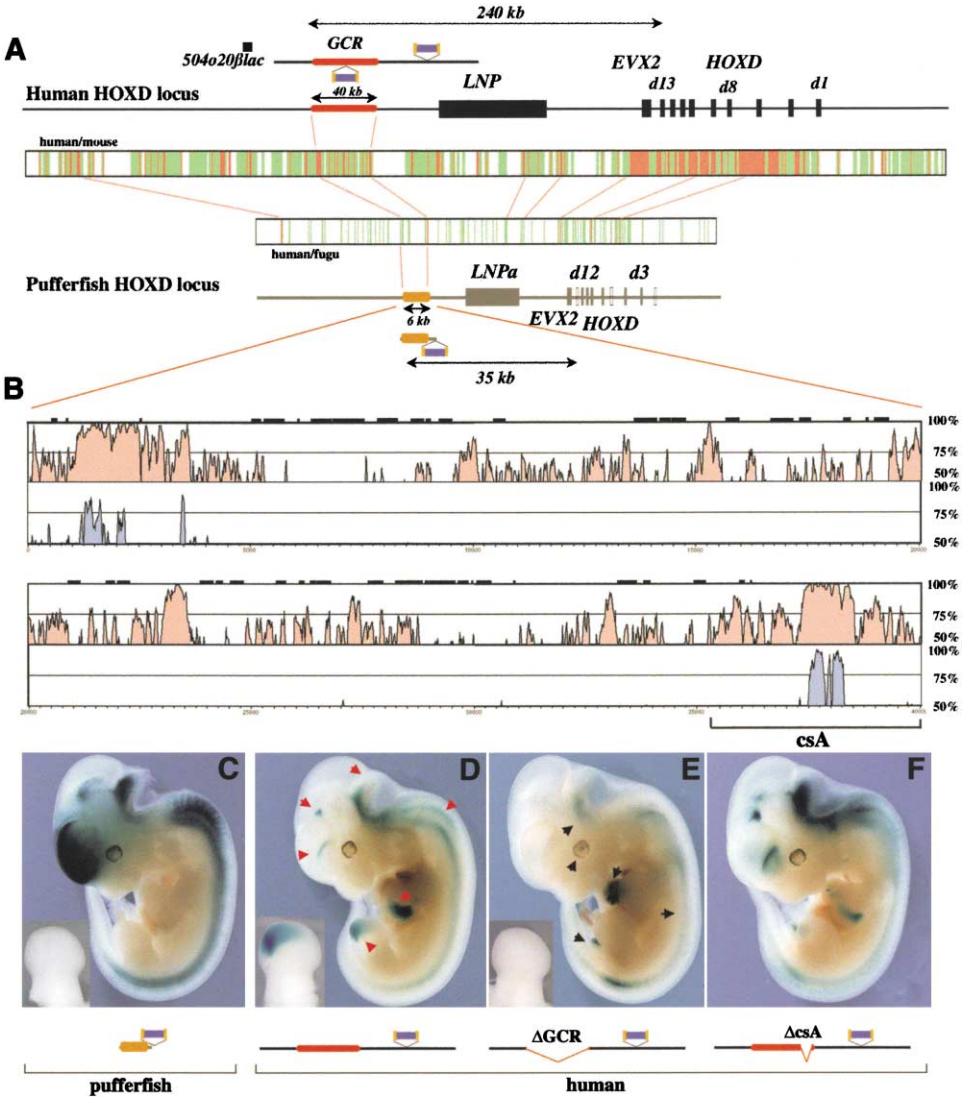

Figure 5. Cloning of the Global Control Region (GCR)

(A) Sequence comparison between the human and pufferfish (*Fugu* and *Tetraodon*) *Hoxd* loci, showed similar organization, despite the 5-fold compacted aspect of the pufferfish locus (not respected in the graphical representation for sake of clarity). Pufferfish genomes lacked both *Hoxd13*, *Hoxd8*, and *Hoxd1* counterparts. Between both loci, global human/mouse (top) and human/*Fugu* (bottom) sequence comparisons are shown, with red bars for a high similarity (at least 100 bp without gap and with at least 70% identity) and green bars for medium similarity. Thin red lines link mammalian and fish orthologous regions. Fragments used for human and T*etraodon* enhancer analysis in vivo are depicted on the map of the respective loci. The GCR is shown as a red (mammal) or orange (pufferfish) boxes. The different sizes and relative distances are indicated. The reporter gene (β-lac) is shown as a blue bar.

(B) Sequence comparison of the GCRs between human, mouse and *Fugu*. Alignments were performed with the mVISTA program (window size 50 bp, homology threshold 65 percent). Regions of homology between mammalian GCR sequences are in red, whereas in blue when indicating similarities between human and pufferfish. Black bars on the top point to the presence and extent of repeated elements. The region corresponding to the conserved sequence A (csA) is underlined.

(C–F) Mouse fetuses at day 12.5, transgenic for different human or pufferfish enhancer reporter constructs. The expression pattern of the GCR containing BAC #504o20β-lac6 BAC is shown in (D). Red arrows point to those expression domains, which were lost upon deletion of the GCR from the BAC (compare D with E; distal limb buds, dorsal neural tube, midbrain, forebrain, and anterior hindbrain). Black arrows in (E) point to expression domains scored with BAC #504o20β-lac, which did not require the GCR (ventral neural tube, heart, eyes and bilateral patches in the genital tubercle). Deletion of region A (F) did not detectably modify the staining profile.

(C) Transgenic embryo where the β-lac reporter gene was driven by a 7.8 kb fragment containing the *Tetraodon* GCR. Expression was similar to the GCR-specific expression pattern seen with the human BAC, except for the forebrain where the expression domain was enlarged, and in limb buds where staining was never observed.

genes were identified. Synteny with human chromosome 2 extended further until genes such as ATP5G3 and CHRNA1, on the centromeric side, or PDE11A on the telomeric side. We used sequences of the *Tetraodon nigroviridis* genome to build a contig covering the pufferfish *HoxD* locus and determined that it was organized as in mammals. In both pufferfish species, the size of the locus was expectedly smaller, with a compaction

factor of about 5- to 7-fold (Figure 5A). Sequence comparison unambiguously identified the GCR sequence at the expected position. Interestingly, however, while the pufferfish GCR contained sequences highly related to the two fully conserved segments used to define the ends of the GCR, the blocks of homology scattered in between were not found in pufferfish (Figure 5B). As a result, the fish GCR was 6 kb large, instead of 40, and mostly contained two blocks of high homology when compared with the mammalian counterparts.

We looked at the regulatory potential of an 8 kb large DNA fragment containing the *Tetraodon* GCR, linked to our β-lac reporter gene, in transgenic mice. Expression was clearly detected in the CNS, with a pattern reminiscent of that seen with the related human sequence, i.e., in rostral and caudal hindbrain, midbrain, dorsal neural tube, and ventral neurons (Figure 5C) and expression in forebrain was largely extended to involve the almost entire ventricle (Figures 5C and 5D). However, the pufferfish GCR was unable to drive expression in developing limb buds of these transgenic animals, suggesting that the digit enhancer activity associated with the tetrapod GCR was likely absent from the fish sequence.

Comparison between fish and mammalian GCRs revealed one block of particularly high sequence conservation (Figure 5B; csA). We assessed the potential of this DNA fragment by engineering a small deletion in BAC #504o20β-lac6 (Figures 5D and 5F), removing 5 kb of sequence at the telomeric end of the GCR. Three transgenic animals were obtained, in addition to one line, and the expression pattern of the reporter gene was not obviously different from that of the control BAC. All expression domains obtained with the 504o20β-lac6 transgenes were observed in the absence of csA (Figures 5D and 5F). We concluded that the CNS specificity associated with either the fish or mammalian GCR was likely linked to the other block of conservation, located at the centromeric end of the GCR.

The *Ulnaless* Mutation Is Allelic to *Lunapark*

The demonstration that the enhancers located in the GCR control this set of genes under physiological conditions in vivo was provided by the molecular resolution of the *Ulnaless* (*Ul*) mutation. *Ul* is as a semidominant, X-ray-induced mutation on mouse chromosome 2 near *HoxD* (Davisson and Cattanach, 1990; Peichel et al., 1996). Heterozygous mice show abnormal zeugopods, with an almost complete absence of ulna, a phenotype suggesting an allelism with the *HoxD*. While genomic analyses failed to reveal any rearrangement, *Hoxd* gene expression in mutant specimen uncovered a gain of expression of *Hoxd13* and *Hoxd12* in developing zeugopods, accompanied by their loss of expression in digits (Hérault et al., 1997; Peichel et al., 1997).

We reinvestigated the *HoxD* to *Lnp* DNA interval in *Ul* and control genomic DNA using molecular probes from the BAC contig. Probes covering the 5′ end of *Lnp* revealed several polymorphisms, suggesting DNA insertion within the transcription unit, between exon 3 and exon 4 (Figure 6A). We cloned and sequenced those *Ulnaless* fragments associated with these two exons and looked at their potential presence in the same chromosomal locus, using a YAC contig covering the 0.4 cM/1.7 Mb *Ulnaless* interval (Peichel et al., 1996). Both junction fragments localized at this locus, telomeric to *Hoxd* and within the *Ulnaless* interval (YAC 18F5; Figure 6B). While 7 nucleotides had been lost during rearrangement at the telomeric breakpoint, 8 bp were deleted at the centromeric breakpoint, within *Lnp* (Figure 6C). Therefore, *Ulnaless* is a balanced paracentric inversion of chromosome 2, with a centromeric breakpoint into *Lnp* and a telomeric breakpoint 770 kb away (Figure 6D). The inverted DNA included *Evx2*, the *HoxD* complex, *Mtx2*, as well as some pseudogenes.

Effect of the *Ulnaless* Inversion upon Gene Expression

Because the centromeric breakpoint of *Ulnaless* lies between the *Evx2*/*HoxD* cluster and the GCR, we could examine the in vivo relevance of this genomic topography. To overcome the difficulty to produce *Ul* homozygous specimen, we crossed *Ul* females with males carrying a deficiency including *Evx2* to *Hoxd11* (Δ*Evx2-d11*). We recovered animals carrying both the *Ulnaless* and the deletion, such that *Hoxd13* and *Evx2* could only be expressed from the *Ulnaless* chromosome. Both *Hoxd13* and *Evx2* expression was gained in the proximal zeugopod, whereas largely lost from both distal limb and genital buds (Figures 6E–6H; see Hérault et al., 1997; Peichel et al., 1997). Likewise, *Evx2* expression from the *Ulnaless* chromosome was lost in dorsal neurons and in the most anterior part of hindbrain, whereas still detected in midbrain and V0 interneurons progenitors (Figures 4C and 4G). These results revealed that *Hoxd* genes and *Evx2* would no longer respond to the GCR when repositioned more than 700 kb away, Therefore, the *Ulnaless* mutation provided a functional demonstration, in vivo, for the existence of a regulatory interaction between the GCR and the *Evx2*/*HoxD* cluster.

Discussion

The Digit Enhancer Is Part of a Global Control Region (GCR)

In this paper, we show that a DNA segment of approximately 40 kb contains a regulatory activity similar to that controlling both *Hoxd* and *Evx2* genes. *Lunapark*, a gene of unknown function present in this locus, is also under the control of the same sequences. Our molecular resolution of the *Ulnaless* mutation demonstrated that this enhancer was relevant for *Hoxd* gene function in vivo, as this large inversion separating the enhancer from the cluster, induced a downregulation of *Hoxd* genes expression in digits. Various evidences thus point to a genuine regulatory activity of this DNA fragment, under physiological conditions. Evolutionary conservation of this sequence was documented by a high level of conservation between human and mouse, with numerous interruptions of repeated sequences. However, two subregions displayed an almost uninterrupted 100 percent identity over several kilobases. Even though shorter sequences of similar conservation were previously reported, either within particular loci, or at a larger scale (Dermitzakis et al., 2002), the extent of sequence identity described here makes it unlikely that mere regulatory DNA/proteins interactions may impose such a selective pressure. Regarding the *HoxD* cluster, noncoding con-

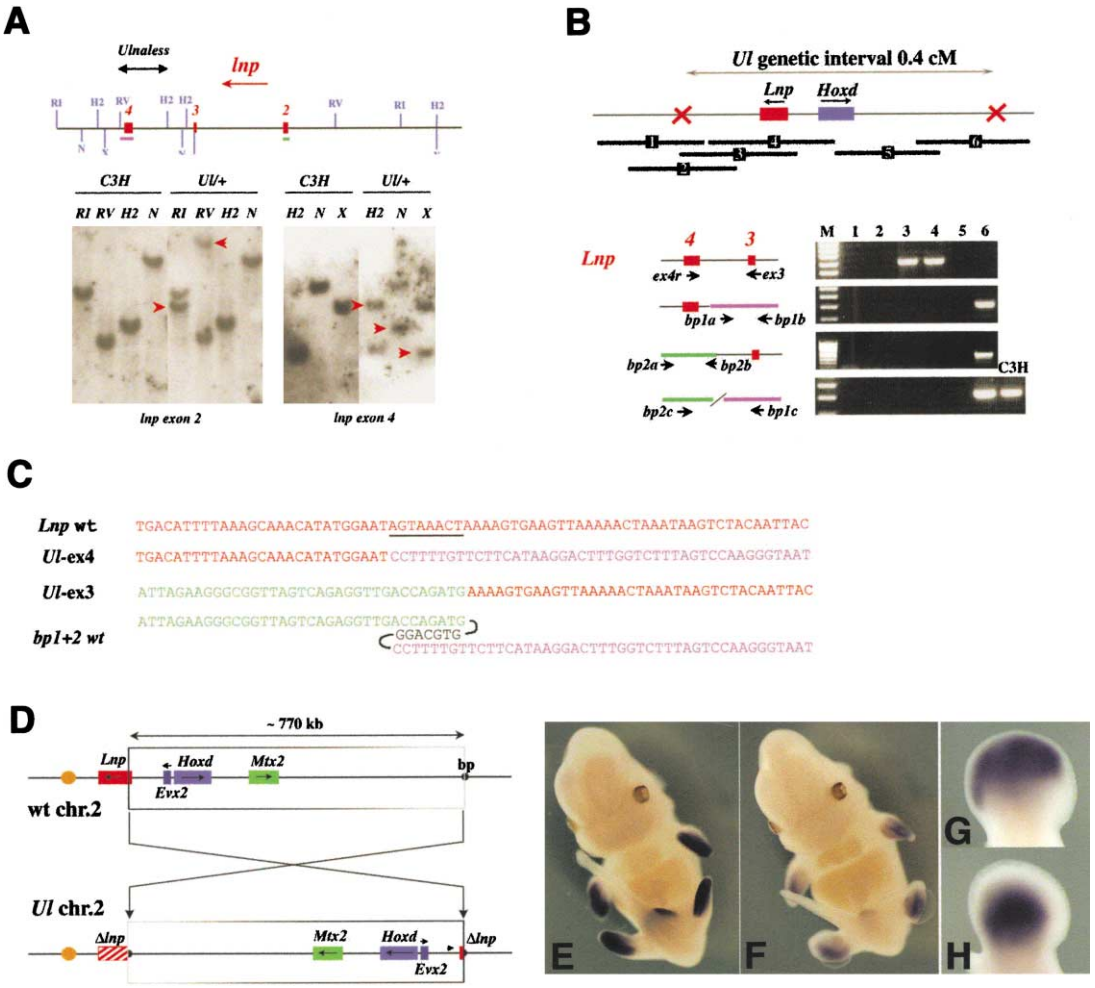

Figure 6. *Ulnaless* Is a Balanced Inversion of a 770 kb DNA Segment, which Disrupts *Lnp* and Increases the Distance between the *Hoxd* Cluster and the GCR

(A) Restriction map of the *Lnp* gene between exons 2 and 4. RI: EcoRI; RV: EcoRV; H2: Hinc2; N: Nco1; X: Xba1. *Lnp* exons are in red. The probes used for Southern blot are indicated in purple (exon 4) or green (exon 2). *Ul*/+ DNA showed polymorphisms (arrowheads), concomitantly with a decreased intensity of wild-type signals.

(B) Localization of the *Ul* breakpoints near *Hoxd*. Breakpoints were mapped by PCR on a YAC contig covering the *Hoxd*/*Ul* genetic interval (Peichel et al., 1997). *Lnp* primers ex4 and ex3 mapped in YAC 85E9 and 176C8 (upper image, lanes 3 and 4). The junction fragments associated with exon 4 (detected with primers bp1a/b) and exon 3 (primers bp2a/b) mapped both within YAC 18F5. When PCR was performed with a primer in both junction fragments (bp1c/bp2c), a specific band was obtained with the YAC 18F5 and C3H DNA samples (lower image), indicating that these two sequences were normally linked.

(C) Sequence of breakpoints. The wt *Lnp* sequence is in red, whereas sequences derived from the inversion are in purple and green. DNA sequence of the telomeric breakpoint showed that 7 bp (bp 1 + 2 wt, black) were deleted, whereas 8 bp were missing from the *Lnp* locus (*Lnp* wt, underlined).

(D) Schematic representation of a wt and *Ul* chromosome. The 770 kb segment, between 5′-*Lnp* and the telomeric breakpoint (bp) was inverted. As a consequence, *Lnp* was broken and posterior *Hoxd* genes moved away from sequences which are centromeric to the *Lnp* breakpoint, such that the GCR (yellow circle) was placed ~800 kb far from *Hoxd13*, instead of 200 kb. Gene expression in +/Δ*Evx2-d11* (E and G) and *Ul*/Δ*Evx2-d11* (F and H) embryos. Expression of *Hoxd13* (E and F) was downregulated in distal limb buds and genital bud of day 12 embryos, whereas gained in a proximal limb domain. The same regulatory alteration was observed for *Evx2* (G and H).

served sequences were previously identified (Beckers and Duboule, 1998; Hérault et al., 1998), but their genetic analysis failed to uncover essential function, except for one that displayed a boundary and/or enhancer positioning activity (Kmita et al., 2002a, 2002b).

The extent of sequence conservation was also remarkable between mammals and fishes, even though it involved only those two islands of maximal similarity in mammals, separated by only 6 kb in pufferfish. While the pufferfish GCR, when introduced into transgenic mice, was able to recapitulate most of the CNS expression specificities, expression in digits was never scored, at least at the developmental stages examined. This result indicated that CNS enhancers were conserved between these distant species and confirmed that they were located within the GCR. They also suggested that the tetrapod digit enhancer, likely absent from pufferfish, was perhaps not located within either one of

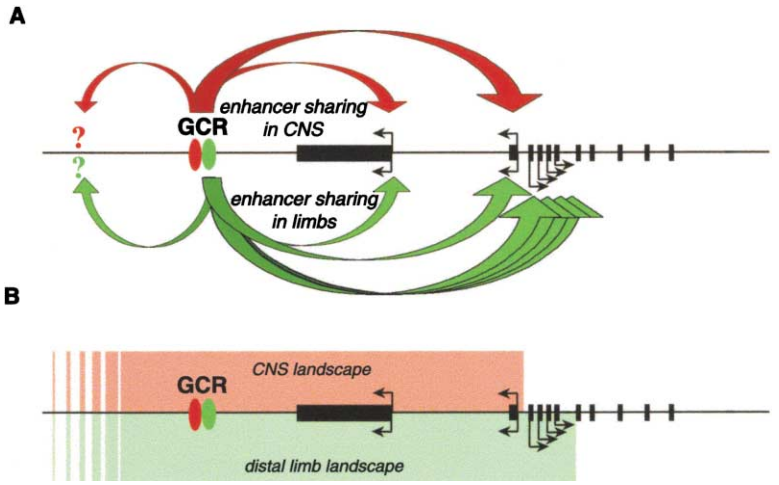

Figure 7. Regulatory Landscapes

A global control region (GCR), containing several enhancers, defines segments of the genome wherein genes are under the same regulatory influences. The extent of such landscapes may be comparable in different cell types, or different depending on both the peculiarities of given enhancers within the GCR and the existence of cell-type specific boundaries (CNS versus digits, in the case of the *Lnp-Evx2-Hoxd* locus). The underlying mechanisms may involve both global and local enhancer-promoter interactions (A), perhaps as a result of the formation of a functional or/and structural configuration facilitating long-range gene regulation from the GCR (B). In this view, the GCR would have concentrated, in the course of evolution, several important enhancers, due to an intrinsic property to work at a distance.

the two major regions of conservation in tetrapods, but instead, in between. This possibility was supported by deletion of region A from the GCR, which did not affect expression, neither in digits, nor in the CNS, suggesting that the CNS expression was likely driven by sequences located within region B.

Regulatory Landscapes

The localization of these enhancers raised several questions regarding both the organization and evolution of global regulatory controls. First, to which distance can this sequence work? Our results show that expression in digits can be obtained approximately 250 kb telomeric to the sequence. Assuming that this enhancer can also regulate genes that are centromeric, we evaluate the regulatory domain, i.e., the DNA interval within which the enhancer can influence transcription, to be about half a megabase large (Figure 7). The fact that posterior *Hoxd* genes no longer respond to this regulation in the *Ulnaless* chromosome indicates that a 0.7 Mb distance may not allow for proper regulation. Likewise, the next gene centromeric to *Lnp*, *ATP5G3*, is located more than 0.7 Mb away in human and was not expressed in developing limbs (data not shown). The size of this interval is nevertheless difficult to evaluate, as enhancer activity can be titrated out by genes located nearby (Kmita et al., 2002a; Monge et al., 2003), making it difficult to precisely assess this distance.

Regulation of target genes at such a distance is, by itself, not exceptional (see Kleinjan et al., 2001; Lettice et al., 2002; Pfeifer et al., 1999). In these cases, however, enhancer sharing was not reported, perhaps due to higher promoter specificity. So far, six genes were shown to respond to the digit enhancer; *Lnp*, *Evx2*, and *Hoxd13* to *Hoxd10*. This effect of a single regulatory element over many different genes spanning 250 kb is, to our knowledge, the first detailed example of such a large DNA region, where any gene present would likely adopt the global regulation at work, in addition to its own controls. As a consequence, localization of a gene within this domain (e.g., through chromosomal rearrangement) would provide opportunities for functional innovations to occur.

This situation is not analogous to well-known "enhancer sharing" mechanisms, which usually involves partners derived from tandem duplication of a common ancestor (e.g., *Hox*, Gould et al., 1997; *Myf5-MRF4*, Carvajal et al., 2001; *Dlx3-7*, Sumiyama et al., 2002; albumin-α fetoprotein, Jin et al., 1995), making it a legacy from an ancestral condition. The situation reported here involves genes that have initially nothing in common, except for their genomic location, and thus may be considered in the light of recent gene expression profiling studies, in human, *Drosophila*, and *C. elegans*, wherein a fair degree of genomic clustering was shown for coexpressed genes (Lercher et al., 2002; Spellman and Rubin, 2002; Roy et al., 2002). This raises the possibility that regulatory domains, or "landscapes" such as the one described here, are widespread in genomes.

Various mechanisms could account for this process, from an incidental activation of genes by regulatory elements required for a neighbor gene, to the spreading of an open chromatin configuration initiated from a single strongly transcribed gene (Oliver et al., 2002). In our case, the high degree of interspecies sequence conservation throughout the GCR, suggests that this DNA segment may have additional features, for instance in terms of general accessibility or nuclear topology. The apparent lack of effect, on reporter gene expression, of the deletion of region csA from the BAC may point to such a phenomenon. It also indicates that one should be cautious when systematically assigning particular transcriptional regulatory functions to evolutionary conserved sequences.

Neural enhancers driving expression of both *Lnp* and *Evx2* do not act upon *Hoxd* genes, illustrating that regulatory landscapes can be different in various cell types (Figure 7). Yet this did not reflect an intrinsic restriction in regulatory potential, as a short deletion of DNA between *Evx2* and *Hoxd* genes extended the "neural landscape" until the *HoxD* cluster (Kmita et al., 2002b). In this CNS landscape, the telomeric extremity is thus determined by a boundary element, rather than by distance or the presence of competing promoters. Therefore, cell-type specific factors may restrict the extent of a landscape by building up boundaries.

Collateral Effects

Which function was primarily used for the design and/or improvement of the digit enhancer? It is possible that

either *Evx2* or *Lnp* are the cause of this regulation. Yet we do not favor this hypothesis, as the function of *Evx2* in digits is not critical (Hérault et al., 1996). Regarding *Lnp*, even though a targeted inactivation is not available, the *Ul* mutation, which also inactivated *Lnp* function, suggested little function for *Lnp* in digits, if any. Indeed, *Ul/Ul* mice did not display a digit phenotype stronger than that obtained following loss-of-function of posterior *Hoxd* genes (Peichel et al., 1996). We conclude that *Lnp*, as *Evx2*, has no key function during digit development (unless they act downstream of *Hoxd* genes). Interestingly, however, large deletions in human, that include at least the *LNP* to *HOXD* interval showed digit defects stronger than after deletion of *HOXD* genes alone (Goodman et al., 2002), suggesting that an additional transcription unit located nearby (up to 5 Mb upstream from HOXD) may contribute to digit development.

If neither *Lnp*, nor *Evx2* have an obvious function during digit development, one may wonder why these genes remained under the influence of this enhancer during bird and mammalian evolution. Overlapping regulatory landscapes provides an answer to this question, as the constraint to keep these genes together may have been exerted on one particular landscape only. In this view, a coherent response to one type of regulation doesn't necessarily imply any functional coherence; *Lnp* and *Evx2* are closely linked, perhaps as they share CNS enhancers as a legacy of an ancestral GCR. They stayed nearby *Hoxd* genes because they were initially localized in between *Hoxd* and the GCR, thus explaining their expression in digits. Yet the fact that these former genes do not seem to have a "function", sensu stricto, in digits, does not exclude that they participated to their proper emergence, through their role in the overall distribution of enhancer activity, as the mere presence of promoters in this region likely impinged upon the contact between the enhancer and *Hoxd* promoters, through a titration mechanism (Kmita et al., 2002a; Monge et al., 2003). Separating the limb enhancer from *Lnp* and *Evx2* might have altered *Hoxd* genes activation and changed limb morphogenesis. These "collateral effects," whose importance should not be underestimated, suggest how, within a regulatory landscape, a group of neighbor genes may all coherently participate to the correct implementation of one functional task, even though several genes may not, by themselves, have a function in this task. It also indicates that genes showing identical expression domains should not necessarily be considered as parts of the same functional circuit, hence the notion of synexpression groups (Niehrs and Pollet, 1999) should be considered within the context of genomic topography.

Experimental Procedures

In Silico Construction of a BAC Contig and Cloning of *lunapark*

The BAC contig was constructed in silico with data from the human genome program available through the NCBI search interfaces in the GenBank HTGS and GSS databases (http://www.ncbi.nlm.nih.gov/) and BLAST algorithm (Altschul et al., 1990). Briefly, BAC ends sequence data (from GSS) were used to identify and anchor partially sequenced BACs (from HTGS), starting with BACs containing sequences from the human or and mouse *Hoxd* loci. These contigs were subsequently validated with the release of the human and mouse draft sequences. To identify genes around the *HOXD* locus, we searched for ESTs whose sequences matched with the genomic sequence from the BACs. Besides EST probably derived from transcriptionally active retrotransposons, we found several overlapping ESTs belonging to the UniGene clusters Hs.137200, Hs.69169, and Hs.118056 (now grouped into UniGene clusters Hs.209561, and Hs.355685). There were small (~50 bp large) overlaps between EST's from the different UniGene clusters, which matched the genomic sequence from BACs RPCI11-504o20 and 514d19. The mouse orthologous gene was identified by BLAST search (UniGene cluster Mm.37960) and ESTs from this cluster were fully sequenced. Related genes in other species were also identified by BLAST searching in sequence databases with the mouse LNP protein sequence.

BAC Modification and Transgenesis

Human BACs (Children's Hospital Oakland; BacPac Resources) were verified by PCR or Southern blot hybridization with locus-specific primers/probes and restriction enzyme fingerprinting. Modification of the BAC RPCI-11-514d19 was made as described previously (Spitz et al., 2001). A linearized mouse *Hoxd11*/lacZpAkanR cassette was electroporated into DH10B cells containing both the targeted BAC and the pGETrec plasmid after L-arabinose induction of the recombinases. The homology between the human and the mouse first exon of *Hoxd11* allow for recombination of the lacZkanR fragment in the human BAC *HOXD11* gene. Targeted BACs were recovered by selection on LB plates containing chloramphenicol and 20 μg/mL kanamycin. The correct integration and integrity of the BAC were verified.

The Tn7β-lac transposon was constructed by insertion of the −80 bp β-globin promoter/lacZSV40pA reporter gene (BGZ40 insert) from Tnpβ-globin-lacZ (Morgan et al., 1996; gift from R. Krumlauf) into the *Pme*1-*Not*1 sites of pGPS1 (New England Biolabs). This plasmid contained a R6K origin of replication and was maintained and propagated in π+ cells. To induce transposition, we mixed 40 ng of Tn7β-lac, 200 ng of the recipient BAC, 2μl of 10 × GPS buffer, and 1 μl of TnsABC* (GPS Kit, NEB) and H20 to a final volume of 18 μl. The reaction was incubated at 37°C for 10 min and after the additions of 2 μl of Start Solution for 1 hr at 37°C. The reaction was stopped by heat inactivation at 75°C for 10 min. One μl of the dialyzed reaction was electroporated in DH10B cells. Targeted BACs were selected on LB plates containing with chloramphenicol and kanamycin, at 37°C. The targeting efficiency was usually between 0.8 to 6%. Chl/Kan-resistant colonies were amplified in selective medium and the BAC purified from mini preps using alkaline lysis. Modified BACs were analyzed by Southern blot with lacZ and/or kanR probes to determine the number of Tn7β-lac integrations in the BAC, and by restriction enzyme fingerprinting to check for the integrity of the genomic insert. Only intact BACs with a single insertion were subsequently used. Integration sites were localized by mapping with restriction enzymes and, for some of them, by direct sequencing using appropriate primers.

PI-*Sce*1 linearized BACs were injected into mouse fertilized eggs (Spitz et al., 2001). F0 transgenic embryos were collected after 12 days (embryos were between E11.5 and E12.5). Transgenic embryos and adult mice were genotyped using membrane or tail DNA, by Southern blot with either a lacZ or a kanR probe, or by PCR using Tn7β-lac-specific primers (5′-GGTATGAGTCAGCAACACCTTCTTC-3′ and 5′-TGGGTAACGCCAGGGTTTTCCC-3′). The integrity of the BAC was checked using primers/probes corresponding BAC ends and some internal regions. In the case of BAC 504o20β-lac20Δ*Bsp*E1, we used internal markers to identify transgenic embryos which had integrated only the 60 kb *Bsp*E1 fragment containing the transposon, but lacking the more proximal fragment containing the 3′ end of *LNP*.

To engineer various deletions in BAC 504o20β-lac6, we electroporated a zeocin resistance cassette (Invitrogen) flanked by 50 bp of DNA fragments corresponding to both ends of the regions to be deleted, into EL350 competent cells containing the 504o20β-lac6 BAC, transiently induced to express the *gam-exo*-β-based recombination system (Lee et al., 2001). Briefly, for the deletion of the GCR (ΔGCR), PCR was used to flank the zeocin resistance gene with the sequences from the 504o20 BAC 5′-ATATTTGCAAGGTGAAGGGT GATACAATTATTCAGCTAGAGCATTCATTTCAG-3′ and 5′-GTGGGT GTGTATGCGCGTGTGTAGGGGGAAAGAATCTGTTTCTTGAAGGA TTG-3′, respectively. For the ΔcsA construct, the same zeocin resistance gene was flanked by sequences 5′-CCAGGTACACACATA TAAGAACATTACTTACCAAGCCTCAACTCTTATAAGGT and ATAT TTGCAAGGTGAAGGGTGATACAATTATTCAGCTAGAGCATTCAT

TTCAG-3′, respectively. BACs carrying the deletions were recovered by plating cells on zeocin (25μg/ml) containing LB agar plates at 30°C. Colonies were amplified and analyzed both by PCR and restriction fingerprinting to check for both the targeted deletion and the integrity of the BAC. BACs showing the expected deletions were prepared and injected into fertilized mouse eggs as previously described.

ES Cells, Chimeras, lacZ Staining, and WISH

The ES cell clone A20075 was obtained from the German Genetrap Consortium (GGTC) (Wiles et al., 2000). The 5′ trapped sequence corresponded to *Lnp* exon 1. After amplification of cells, the insertion site in the *Lnp* locus was controlled by Southern blot and detected in the first intron (data not shown). ES cells were injected into blastocysts and chimeric embryos were collected at 11.5 or 12.5 days. lacZ staining and whole-mount in situ hybridizations were performed. The human *HOXD13* specific probe, as well as both murine *Hoxd13* and *Evx2* probes, was as in Kmita et al. (2000) and Spitz et al. (2001). Mouse *Lnp* anti-sense transcripts were produced from EST vy09b06 (IMAGE:1294931), cloned into pSKII-Bluescript. The chicken *Lnp* and the mouse *Atp5g3* probes were made from EST clones DKHZp426K0110Q2 and IMAGp998N096796Q2, respectively, obtained from RZPD.

Lnp^{Ulnaless}

Ulnaless mice were purchased from the Jackson Laboratory and the strategy used to clone the breakpoints is available as Supplemental Data at http://www.cell.com/cgi/content/full/113/3/405/DC1.

The Pufferfish *Hoxd* Cluster and Enhancer Analysis

The sequence of the *Fugu rubripes Hoxd* locus was obtained from the *Fugu* genome project consortium (Aparicio et al., 2002). The region of the *Fugu* genome corresponding to our mammalian BAC contig was covered by scaffolds 281 and 449. This sequence data was used to construct a BAC contig of the corresponding *Tetraodon nigroviridis* sequence. These two pufferfishes showed a high degree of sequence identity; hence, we could screen in silico *Tetraodon* BACs by using *Tetraodon* BAC end sequences (http://www.genoscope.cns.fr/externe/tetraodon/Ressource.html) for comparison with the *Fugu Hoxd* locus. In this way, we identified BAC 49d22 from the *Tetraodon* Genoscope library A, which contained the entire *Lnp* to *HoxD* interval. Partial *Tetraodon* sequence was assembled from trace sequences generated by the *Tetraodon* sequencing consortium (Genoscope and Whitehead Institute) and used to generate primers to validate this BAC. A 7.8 kb *Sac*I fragment containing the sequence orthologous to the GCR was subcloned in pBluescript SKII. The β-lac reporter gene was inserted into the resulting plasmid, which was used to produce transgenic mice. Sequence alignments between mouse and human loci were made with the PipMaker and mVISTA programs (Schwartz et al., 2000; Mayor et al., 2000) through their respective websites (http://bio.cse.psu.edu/cgi-bin/pipmaker?advanced; http://www-gsd.lbl.gov/vista/).

Acknowledgments

We would like to thank M. Morisson (INRA Toulouse) for communicating result prior to publication; M.-T. Martinez-Pastor and R. Krumlauf for constructs; T. Floss and W. Würst (GGTC) for an ES cell clone; N. Copeland (NCI-Frederick) for the EL350 *E. coli* strain; H.R. Crollius (Genoscope) for *Tetraodon* BACs; T. Kondo for mice; and members of the Duboule laboratory for sharing reagents and for discussions. F.S. was supported by EMBO and HFSPO postdoctoral fellowships. This work was supported by funds from the Canton de Genève, the Swiss National Research Fund, the NCCR Frontiers in Genetics, the Claraz, Latsis, Cloetta, and Louis-Jeantet foundations.

Received: October 31, 2002
Revised: March 19, 2003
Accepted: April 8, 2003
Published: May 1, 2003

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Chistoffels, A., Rash, S., Hoon, S., Smit, A., et al. (2002). Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science *297*, 1301–1310.

Beckers, J., and Duboule, D. (1998). Genetic analysis of a conserved sequence in the *HoxD* complex: regulatory redundancy or limitations of the transgenic approach? Dev. Dyn. *213*, 1–11.

Biery, M.C., Stewart, F.J., Stellwagen, A.E., Raleigh, E.A., and Craig, N.L. (2000). A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. Nucleic Acids Res. *28*, 1067–1077.

Capdevila, J., and Izpisua Belmonte, J.C. (2001). Patterning mechanisms controlling vertebrate limb development. Annu. Rev. Cell Dev. Biol. *17*, 87–132.

Carvajal, J.J., Cox, D., Summerbell, D., and Rigby, P.W. (2001). A BAC transgenic analysis of the Mrf4/Myf5 locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development. Development *128*, 1857–1868.

Davis, A.P., Witte, D.P., Hsieh-Li, H.M., Potter, S.S., and Capecchi, M.R. (1995). Absence of radius and ulna in mice lacking hoxa-11 and hoxd-11. Nature *375*, 791–795.

Davisson, M.T., and Cattanach, B.M. (1990). The mouse mutation ulnaless on chromosome *2*. J. Hered. *81*, 151–153.

Dermitzakis, E.T., Reymond, A., Lyle, R., Scamuffa, N., Ucla, G., Deutsch, S., Stevenson, B., Flegel, V., Bucher, P., Jongeneel, C.V., and Antonarakis, S.E. (2002). Numerous potentially functional but non-genic sequences on human chromosome 21. Nature *420*, 578–582.

Dollé, P., Fraulob, V., and Duboule, D. (1994). Developmental expression of the mouse Evx-2 gene: relationship with the evolution of the HOM/Hox complex. Development (*suppl*), 143–153.

Dollé, P., Izpisua-Belmonte, J.C., Falkenstein, H., Renucci, A., and Duboule, D. (1989). Coordinate expression of the murine Hox-5 complex homoeobox-containing genes during limb pattern formation. Nature *342*, 767–772.

Fromental-Ramain, C., Warot, X., Messadecq, N., LeMeur, M., Dollé, P., and Chambon, P. (1996). Hoxa-13 and Hoxd-13 play a crucial role in the patterning of the limb autopod. Development *122*, 2997–3011.

Gérard, M., Duboule, D., and Zákány, J. (1993). Structure and activity of regulatory elements involved in the activation of the Hoxd-11 gene during late gastrulation. EMBO J. *12*, 3539–3550.

Goodman, F.R., Majewski, F., Collins, A.L., and Scambler, P.J. (2002). A 117-kb microdeletion removing HOXD9-HOXD13 and EVX2 causes synpolydactyly. Am. J. Hum. Genet. *70*, 547–555.

Gould, A., Morrison, A., Sproat, G., White, R.A., and Krumlauf, R. (1997). Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns. Genes Dev. *11*, 900–913.

Hérault, Y., Beckers, J., Gérard, M., and Duboule, D. (1999). Hox gene expression in limbs: colinearity by opposite regulatory controls. Dev. Biol. *208*, 157–165.

Hérault, Y., Beckers, J., Kondo, T., Fraudeau, N., and Duboule, D. (1998). Genetic analysis of a Hoxd-12 regulatory element reveals global versus local modes of controls in the HoxD complex. Development *125*, 669–677.

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.

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.

Jin, J.R., Wen, P., and Locker, J. (1995). Enhancer sharing in a

plasmid model containing the alpha-fetoprotein and albumin promoters. DNA Cell Biol. *14*, 267–272.

Kleinjan, D.A., Seawright, A., Schedl, A., Quinlan, R.A., Danes, S., and van Heyningen, V. (2001). Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. Hum. Mol. Genet. *10*, 2049–2059.

Kmita, M., Fraudeau, N., Hérault, Y., and Duboule, D. (2002a). Serial locus deletions and duplications in vivo suggest a mechanism for Hoxd genes colinearity in developing limbs. Nature *420*, 145–150.

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.

Kmita, M., Tarchini, B., Duboule, D., and Hérault, Y. (2002b). Evolutionary conserved sequences are required for the insulation of the vertebrate HoxD complex in neural cells. Development *129*, 5521–5528.

Kondo, T., and Duboule, D. (1999). Breaking colinearity in the mouse HoxD complex. Cell *97*, 407–417.

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

Laufer, E., Nelson, C.E., Johnson, R.L., Morgan, B.A., and Tabin, C. (1994). Sonic hedgehog and Fgf-4 act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. Cell *79*, 993–1003.

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.

Lercher, M.J., Urrutia, A.O., and Hurst, L.D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat. Genet. *31*, 180–183.

Lettice, L.A., Horikoshi, T., Heaney, S.J., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. USA *99*, 7548–7553.

Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., and Dubchak, I. (2000). VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics *16*, 1046–1047.

Monge, I., Kondo, T., and Duboule, D. (2003). An enhancer titration effect induces digit-specific regulatory alleles of the *Hoxd* cluster. Dev. Biol. *256*, 212–220.

Morgan, B.A., Conlon, F.L., Manzanares, M., Millar, J.B., Kanuga, N., Sharpe, J., Krumlauf, R., Smith, J.C., and Sedgwick, S.G. (1996). Transposon tools for recombinant DNA manipulation: characterization of transcriptional regulators from yeast, Xenopus, and mouse. Proc. Natl. Acad. Sci. USA *93*, 2801–2806.

Muyrers, J.P., Zhang, Y., Testa, G., and Stewart, A.F. (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res. *27*, 1555–1557.

Nelson, C.E., Morgan, B.A., Burke, A.C., Laufer, E., DiMambro, E., Murtaugh, L.C., Gonzales, E., Tessarollo, L., Parada, L.F., and Tabin, C. (1996). Analysis of Hox gene expression in the chick limb bud. Development *122*, 1449–1466.

Niehrs, C., and Pollet, N. (1999). Synexpression groups in eukaryotes. Nature *402*, 483–487.

Oliver, B., Parisi, M., and Clark, D. (2002). Gene expression neighborhoods. J Biol *1*, 4.

Orford, M., Nefedov, M., Vadolas, J., Zaibak, F., Williamson, R., and Ioannou, P.A. (2000). Engineering EGFP reporter constructs into a 200 kb human beta-globin BAC clone using GET recombination. Nucleic Acids Res. *28*, E84.

Peichel, C.L., Abbott, C.M., and Vogt, T.F. (1996). Genetic and physical mapping of the mouse Ulnaless locus. Genetics *144*, 1757–1767.

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.

Pfeifer, D., Kist, R., Dewar, K., Devon, K., Lander, E.S., Birren, B., Korniszewski, L., Back, E., and Scherer, G. (1999). Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. Am. J. Hum. Genet. *65*, 111–124.

Riddle, R.D., Johnson, R.L., Laufer, E., and Tabin, C. (1993). Sonic hedgehog mediates the polarizing activity of the ZPA. Cell *75*, 1401–1416.

Rijli, F.M., and Chambon, P. (1997). Genetic interactions of Hox genes in limb development: learning from compound mutants. Curr. Opin. Genet. Dev. *7*, 481–487.

Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. (2002). Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. Nature *418*, 975–979.

Shubin, N., Tabin, C., and Carroll, S. (1997). Fossils, genes and the evolution of animal limbs. Nature *388*, 639–648.

Sordino, P., van der Hoeven, F., and Duboule, D. (1995). Hox gene expression in teleost fins and the origin of vertebrate digits. Nature *375*, 678–681.

Spellman, P.T., and Rubin, G.M. (2002). Evidence for large domains of similarly expressed genes in the Drosophila genome. J. Biol *1*, 5.

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.

Schwartz, F., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., and Miller, W. (2000). PipMaker: a web server for aligning two genomic DNA sequences. Genome Res. *10*, 577–586.

Sumiyama, K., Irvine, S.Q., Stock, D.W., Weiss, K.M., Kawasaki, K., Shimizu, N., Shashikant, C.S., Miller, W., and Ruddle, F.H. (2002). Genomic structure and functional control of the Dlx3-7 bigene cluster. Proc. Natl. Acad. Sci. USA *99*, 780–785.

van der Hoeven, F., Zákány, J., and Duboule, D. (1996). Gene transpositions in the HoxD complex reveal a hierarchy of regulatory controls. Cell *85*, 1025–1035.

Wiles, M.V., Vauti, F., Otte, J., Fuchtbauer, E.M., Ruiz, P., Fuchtbauer, A., Arnold, H.H., Lehrach, H., Metz, T., von Melchner, H., and Wurst, W. (2000). Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. Nat. Genet. *24*, 13–14.

Zákány, J., and Duboule, D. (1996). Synpolydactyly in mice with a targeted deficiency in the HoxD complex. Nature *384*, 69–71.

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

Zákány, J., and Duboule, D. (1999b). Hox genes in digit development and evolution. Cell Tissue Res. *296*, 19–25.

Zákány, J., Fromental-Ramain, C., Warot, X., and Duboule, D. (1997). Regulation of number and size of digits by posterior Hox genes: a dose-dependent mechanism with potential evolutionary implications. Proc. Natl. Acad. Sci. USA *94*, 13695–13700.