6 Global Control Regions and Regulatory Landscapes in Vertebrate Development and Evolution

Francois Spitz* **and Denis Duboule**[†,‡]

*Developmental Biology Unit, EMBL, 69117 Heidelberg, Germany
[†]NCCR "Frontiers in Genetics" and Department of Zoology and Animal Biology, University of Geneva, 1211 Geneva, Switzerland
[‡]School of Life Sciences, Federal Polytechnic School, CH-1015 Lausanne, Switzerland

Advances in Genetics, Vol. 61
0065-2660/08 $35.00
DOI: 10.1016/S0065-2660(07)00006-5

ABSTRACT

During the course of evolution, many genes that control the development of metazoan body plans were co-opted to exert novel functions, along with the emergence or modification of structures. Gene amplification and/or changes in the *cis*-regulatory modules responsible for the transcriptional activity of these genes have certainly contributed in a major way to evolution of gene functions. In some cases, these processes led to the formation of groups of adjacent genes that appear to be controlled *by both* global and shared mechanisms.

I. INTRODUCTION

The proper implementation of the genetic program controlling cell differentiation and, ultimately, metazoan development requires the highly coordinated actions of multiple genes. Consequently, these genes need to be tightly regulated in both time and space, and with respect to their quantitative outputs. Small changes in a single gene expression pattern can lead to severe morphological alterations, as exemplified by haplo-insufficiencies in many human syndromes (e.g., Guris *et al.*, 2006; Slager *et al.*, 2003), ectopic gene expression in cancers (e.g., Hayday *et al.*, 1984), or even slight heterochronic variations (e.g., Juan and Ruddle, 2003; Zakany *et al.*, 1997). Yet gene products rarely act alone, but usually interact with partners either to form large multiproteins complexes (e.g., Polycomb-group complexes, muscle contractile apparatus) or to be part of a sequential metabolic pathway (e.g., retinoic acid). Therefore, it is important that genes whose products are part of the same functional pathway are expressed at the same time and in the same cells.

In 1960, François Jacob and Jacques Monod showed that *LacZ*, *LacY*, and *LacA*, the genes required to degrade lactose in *Escherichia coli*, are aligned sequentially along the bacterial chromosome. Their expression is controlled by a repressor molecule produced by the *lacI* gene which blocks their transcriptional activity by binding to an operator element localized upstream of this small gene cluster (Jacob *et al.*, 1960). Furthermore, these genes are transcribed as a single RNA molecule, reenforcing their tight coordination. Such an organization of genes within "operons" provided a paradigm for the integration of gene expression, function, and structural organization, and indeed, many prokaryotic genes are organized similarly.

Operons are also present in eukaryotes, even though they usually differ slightly from the canonical prokaryotic version. For instance, up to 15% of *Caenorhabditis elegans* genes are organized as operons (Blumenthal *et al.*, 2002). They are transcribed as polycistronic pre-mRNAs that are subsequently processed by polyadenylation and *trans*-splicing to form mature, single gene coding

mRNAs. As for prokaryotes, some of these "operons" may contain genes whose products are involved in the same biological process (Blumenthal *et al.*, 2002). However, it is unclear how ancestral these operons are, since they do not seem to be conserved in another nematode (Lee and Sommer, 2003). In mammals, only a few bicistronic systems have been described and it is as yet unclear whether the functions of the genes concerned are related in any way (Gray *et al.*, 1999; Lee, 1991). Therefore, our current view of eukaryotic gene regulation tends to consider genes as individual units, generally controlled by private sets of regulators. However, if we consider the original concept of the operon, initially regarded as "a grouping of adjacent structural genes controlled by a common operator" (Francois Jacob, in his Nobel Lecture), several recently described eukaryotic genetic loci may fulfill such a definition. Indeed, recent approaches involving either specific loci or genome-wide gene analysis tools tell us that a significant proportion of adjacent eukaryotic genes display related expression patterns, suggesting shared regulatory mechanisms.

In eukaryotes, many functionally related genes are found in clusters. In most cases, these situations derive from tandem duplications of an ancestral gene, leading to series of contiguous genes structurally related to each other, rather than distinct genes displaying related functions. In many instances, genes belonging to such clusters are expressed similarly, regardless of their general functional classification as "structural" or "regulatory" genes. For example, clustering may concern not only genes such as keratins, olfactory receptors, or protocadherins (Glusman *et al.*, 2001; Hesse *et al.*, 2004; Wu and Maniatis, 1999), but also transcription factors such as homeobox genes (Brooke *et al.*, 1998; Duboule *et al.*, 1986; Jagla *et al.*, 2001) or signaling molecules such as Wnts and Fgfs (Katoh, 2002; Nusse, 2001). While some of these clusters emerged recently during evolution, others can be traced down to the common metazoan ancestor and have since been maintained. This conservation of clustering illustrates the existence of constraints, one of them likely being the shared regulation of several gene members in a given cluster, suggesting that potent regulatory mechanisms act globally to impose a common regulation at the level of the cluster, rather than acting upon individual genes. Recent work on different model systems has identified such global mechanisms of gene expression, and these are discussed below.

Using genome-wide approaches, several studies have reported that adjacent, yet otherwise unrelated, genes share common expression specificities, leading to large co-expression territories (Boutanaev *et al.*, 2002; Spellman and Rubin, 2002). In vertebrates, while global surveys have mainly revealed clusters of housekeeping genes (Lercher *et al.*, 2002), it is quite clear that similar situations occur around developmental genes, as illustrated by a growing list of examples (Crackower *et al.*, 1996; Holmes *et al.*, 2003; Maas and Fallon, 2004; Spitz *et al.*, 2003; Zuniga *et al.*, 2004). While it is as yet unclear whether "regulatory landscapes" fulfill particular biological function and/or correspond

to a functional partition of the genome, these findings have important practical and conceptual implications for our understanding of the relationships between gene regulation and genome structural organization and evolution.

Comparisons between different metazoan genomes have revealed that most of the genes involved in generating animal body forms were already present in their common ancestors (e.g., Carroll, 2006), indicating that evolution proceeded by redeploying a rather limited repertoire of genes. The co-option of genes [or of whole regulatory circuitries (Davidson, 2006)] in parallel with the evolution of novel tasks, and the resulting pleiotropy, was naturally accompanied by what can be seen as side effects, such as compensatory mechanisms and redundancy (Duboule and Wilkins, 1998; Kirschner and Gerhart, 1998). While most gene families were arguably present in basal animals, the number of available genes subsequently varied. In vertebrates, for instance, this repertoire expanded following both whole genome and local duplications. These two successive rounds of genome duplications account for the observation that many genes are found in three to four copies in mammals, whereas they are unique in lophotrochozoans and ecdysozoans (Levine and Tjian, 2003; McLysaght *et al.*, 2002). Alternatively, or in parallel, gene families have also been expanded through tandem duplications, leading to the formation of multigenic clusters. Some of these clusters are rather young and have evolved rapidly, such as pheromone receptors (Rodriguez, 2005), whereas others are found conserved in insects and mammals, suggesting that these peculiar organizations are functionally constrained.

II. GLOBAL CONTROLS

A. Clusters of co-expressed developmental genes

Developmental genes are often grouped into multigenic complexes containing two or more closely related genes. Interestingly, this seems to mostly involve transcription factors [e.g., the *Fox* (Wotton and Shimeld, 2006), *Hox* (Krumlauf, 1994), *Irx* (Gomez-Skarmeta and Modolell, 2002), *Six* (Gallardo *et al.*, 1999), *NK-Lbx* (Luke *et al.*, 2003) gene families], though some signaling molecules, such as *Fgf3/4/19* and *Wnt1/6/10* (Katoh, 2002; Nusse, 2001), are also organized in clusters. Quite often, genes within a cluster tend to share some expression features during embryogenesis or in adult tissues.

The *Irx/Iroquois* gene clusters are rather typical in this respect. *Irx/Iroquois* genes encode homeobox-containing transcription factors and are organized in clusters, containing three genes separated by large intergenic regions (from 300 kb to more than 1 Mb in human). In mammals, two *Irx* clusters

have been reported (*Irx1/2/4* and *Irx3/5/6)* (Peters *et al.*, 2000), which may have evolved independently from the single one described in *Drosophila* (*Iro-C*) (Gomez-Skarmeta and Modolell, 2002). The expression of vertebrate *Irx* genes has been studied in several vertebrate species (Christoffels *et al.*, 2000; Houweling *et al.*, 2001; Lecaudey *et al.*, 2005; Mummenhoff *et al.*, 2001; Zulch *et al.*, 2001) and revealed that, within a cluster, adjacent genes are expressed in similar structures during embryogenesis. For example, both *Irx1* and *Irx2* show strong and comparable expression domains during mouse brain development (ventral mesencephalon and telencephalon, otic vesicle), in the neural tube and in the condensing cartilage of developing limb buds, as well as in other sites (Houweling *et al.*, 2001). Likewise, *Irx3* and *−5* are co-expressed in domains that may overlap with those of *Irx1/2*, yet with important differences in the neural tube, in motor neurons as well as in developing limbs. In both clusters, the gene located at the 3′ extremity (*Irx4* or *Irx6)* displays usually a more divergent pattern, which does not correlate with a greater intergenic spacing, for the distance between *Irx1* and *Irx2* is larger than that between *Irx2* and *Irx4*. Some expression domains are nonetheless shared by all three members of each cluster, such as, for example, in kidneys or in mammary glands (Houweling *et al.*, 2001). Both the structural organization and some of these expression features are conserved between mammalian, amphibian, and teleost *Irx* genes.

Co-expression of tandemly duplicated genes can derive from different processes. For example, an ancestral gene and its associated regulatory elements (promoter, enhancers, and silencers) are duplicated all together, leading to a situation where both copies have their own regulatory elements. As these elements have a common origin, the duplicated genes may control expression in a similar way, before each copy evolves separate regulatory capacities. Such a situation does not require genes to maintain a topographic link. Alternatively, coding regions may be duplicated without some of their associated regulatory regions, or one set of the regulatory regions duplicated together with the coding regions might be subsequently lost, before evolving unique properties that would favor its preservation (Force *et al.*, 1999). In such cases where co-expression of the duplicated genes in a subset of domains is of selective advantage, regulations must be shared, and therefore the tandem organization of genes and *cis*-regulatory elements will be maintained over long evolutionary periods. In this context, several blocs of highly conserved DNA sequences have been found within the large intergenic regions of the *Irx* gene clusters (de la Calle-Mustienes *et al.*, 2005; McEwen *et al.*, 2006). Some of these potential *cis*-regulatory elements have been shown to drive reporter gene expression in domains that are shared between adjacent genes (de la Calle-Mustienes *et al.*, 2005; Pennacchio *et al.*, 2006). Interestingly, sequence comparisons have revealed that among these enhancers, some of them are related to each other, indicating that they were probably part of the duplicated material (Fig. 6.1).

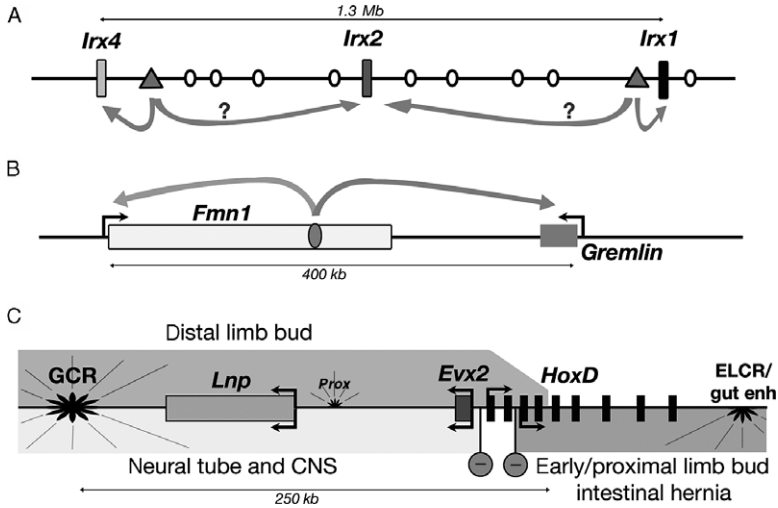

Figure 6.1. Global and shared regulations. (A) The *Irx124* cluster. Two homologous evolutionarily conserved sequences (triangles) were described within the complex and shown to activate gene expression in domains related to *Irx1*, *−2*, *and −4* expression patterns (de la Calle-Mustienes *et al.*, 2005). These elements probably derived from an ancestral enhancer that was duplicated along with the rearrangement(s) that led to the formation of the cluster as related elements are also found at orthologous positions in the sister *Irx356* cluster. However, because of the large number of other associated enhancers present within the cluster and the possibility of redundancy, it is unclear whether these duplicated elements are shared between the different *Irx* genes and account for their highly similar expression profiles. (B) The *Formin1–Gremlin* locus. The limb bud expression of *Gremlin* is controlled by an enhancer localized within the adjacent *Formin1* gene (Zuniga *et al.*, 2004). This enhancer also activates *Fmn1* in the limb bud, even though no function has been associated with this evolutionarily conserved expression. Both *Gremlin* and *Formin1* are also expressed autonomously in distinct domains. (C) The *Lnp–Evx2–Hoxd* regulatory landscapes. Several overlapping regulatory landscapes (shaded regions) coexist around the *Lunapark* (*Lnp*), *Evx2*, and *Hoxd* gene cluster. *Lnp*, *Evx2*, and the posterior *Hoxd* genes are included within a large region where all genes are expressed in the distal part of the limb and genital buds. This landscape is defined by a remote Global Control Region (GCR) which reinforces the intrinsically more restricted action of another enhancer element (*Prox*) (Gonzalez *et al.*, 2007; Spitz *et al.*, 2003). The GCR also contributes to the neural tube and CNS expression domains shared by *Lnp* and *Evx2*. At the other end of the *Hoxd* complex, a gut enhancer and an Early Limb Control Region (ELCR) localized 3′ of the cluster activate the *Hoxd1-Hoxd11* genes in the intestinal hernia, and in early stages and proximal parts of the limb, respectively (Spitz *et al.*, 2005; Zakany *et al.*, 2004). The activities of the GCR in the neural/CNS and of the gut enhancer are somehow restricted by boundary elements, whereas in the distal limb bud, it progressively declines with the distance to the 5′ end of the *Hoxd* cluster (Kmita *et al.*, 2000b, 2002a,b).

Therefore, the *Irx* clusters may illustrate both processes by which co-expression of tandemly duplicated genes can be maintained, with the sharing of regulatory elements, and duplicate ones, all of them contributing to co-expression of the *Irx* genes.

While several gene clusters show co-expression among neighboring genes, it is noteworthy that others do not, as exemplified by the Krüppel associated box (KRAB) gene family. These genes encode zinc-finger transcription factors and are organized in clusters (Huntley *et al.* 2006). They form a rather young and dynamic gene family, since mouse and human clusters arose probably from independent duplication events in rodents and primates (Huntley *et al.*, 2006; Shannon *et al.*, 2003). Despite this recent evolution, mammalian KRAB genes within a given cluster do not clearly share expression domains, indicating that tandem duplication does not necessarily imply the subsequent use of the same regulations. These young KRAB clusters might still be evolving, with their fate (either breaking down to individual genes or developing shared regulation) yet to be established.

B. Shared regulatory elements between neighboring genes

The observation that a given enhancer may be shared by neighboring genes calls for more than the mere contemplation of expression domains, requiring that experimental investigations be carried out *in vivo*. Using homologous recombination in ES cells and production of genetically modified mice, the endogenous locus can be deleted, so as to observe the effects upon both, or all potential target genes. Alternatively, the locus of interest can be transferred onto a large "transgene" such as Bacterial Artificial Chromosomes (BACs). The increasing availability of BAC resources (a list can be accessed at http://www.genome.gov/10001852), their accessibility to manipulation, and the development of rapid protocols to modify them at will have greatly facilitated the studies of large-scale gene regulation over the past years. BACs can contain large genomic fragments (up to 200 kb) and are thus able to incorporate entire loci or multigenic clusters. In addition, they are less prone to uncontrolled recombination than YACs and easier to expand and purify. The introduction of "recombineering strategies" in *E. coli* (Lee *et al.*, 2001; Muyrers *et al.*, 1999; Warming *et al.*, 2005) has opened the door to all kinds of modifications, such as the insertion of reporter genes, the deletion of *cis*-regulatory elements, or the introduction of point mutations. In this way, shared regulatory elements were successfully identified, for example, in the β-globin (see chapter 4), the *Il4–13–5* (Loots *et al.*, 2000), the *Dlx* (Sumiyama *et al.*, 2002), the *Myf5/Mrf4* (Carvajal *et al.*, 2001), or the *Hox* gene clusters (Chiu *et al.*, 2000; Spitz *et al.*, 2001, 2003). In this context, *Hox* clusters are of special interest as they illustrate all possible mechanisms that may be involved in the global regulation of gene clusters during development.

Mammals have 39 *Hox* genes, present in four different complementation groups, each containing from 9 to 11 genes. These four *Hox* clusters (*HoxA* to *HoxD*) are derived from an elusive cluster present in an early chordate, as may possibly be illustrated by the single *Hox* cluster described in the cephalochordate *Amphioxus* [see, e.g., Garcia-Fernandez (2005)]. A cluster of *Hox* genes must have existed in ancestral bilaterian animals, at the base of the radiation between vertebrates, lophotrochozoans, and ecdysozoans, and remnants of such a cluster can be found in various species within these large groups (Lemons and McGinnis, 2006). Generally, the expression domains of *Hox* genes along the anterior to posterior (AP) axis provide positional cues to the developing embryo which contribute to defining the fate of axial structures (Krumlauf, 1994). Interestingly, the anterior most limits of these expression domains in somitic mesoderm and in the neural tube correlate with the relative position of the genes within their clusters, such that genes localized at the 3′ end of the cluster are activated first and in anterior structures, whereas more 5′ located genes are expressed later and in more posterior domains. Different models have been proposed to account for this phenomenon of "colinearity," that is, to explain how the topography of these genes may relate to their expressions in time and space (Kmita and Duboule, 2003).

During mammalian development, *Hox* gene expression is controlled by multiple mechanisms. For example, Gould, Sharpe, Krumlauf, and colleagues reported the presence of enhancers within the *HoxB* cluster and shared between two adjacent genes (Gould *et al.*, 1997). They showed that the adjacent *Hoxb3* and *Hoxb4* genes are co-expressed in the hindbrain, with a sharp boundary between rhombomeres 6 and 7, and identified a conserved DNA sequence between these two genes that conferred hindbrain expression up to the r6 to r7 level (Gould *et al.*, 1997). When this enhancer was deleted from the transgene, expression no longer reached this rostral level, suggesting that both *Hoxb3* and *Hoxb4* require this regulation to extend to their proper expression boundary. A similar analysis on the adjacent *Hoxb4*, *Hoxb5*, *and Hoxb6* region used a two-reporters transgenic strategy, allowing the transcriptional activities of neighboring genes to be monitored in parallel (Sharpe *et al.*, 1998). In this way, several tissue-specific enhancer elements were localized within this short genomic interval, including a shared enhancer that activates simultaneously *Hoxb4* and *Hoxb5* in somites 7 to 8. In contrast, other enhancers appeared to act on a single gene only, either due to promoter selectivity or due to promoter competition. Enhancer sharing among adjacent mammalian *Hox* genes was further demonstrated by targeted alterations of the *HoxD* cluster. Deletion of several evolutionarily conserved intergenic elements led to selective alterations in the expression of several genes, either in time or in space (Gerard *et al.*, 1996; Zakany *et al.*, 1997). This was particularly well illustrated by the deletion of the RXI element, leading to a concomitant posteriorization of both *Hoxd10* and *Hoxd11* expressions in the trunk of embryos (Gerard *et al.*, 1996).

C. Global regulation in *Hox* cluster

Extensive sharing of local *cis*-regulatory elements between adjacent groups of *Hox* genes provides an attractive model to account for both the related expression patterns and the maintenance of clustered organization (Duboule, 1998). However, studies on the *HoxD* cluster showed that this model is incomplete, with other elements and mechanisms being involved. In particular, the early phase of *Hox* genes expression could not be properly reproduced with transgenes, which often showed delayed activation. Accordingly, a large transgene containing most of the human *HOXD* cluster was shown to contain those sequences required for collinear expression in late phases, but could not clearly reproduce early expression patterns (Spitz *et al.*, 2001). Since several local *Hoxb* enhancers are controlled by HOX products themselves (Gould *et al.*, 1997; Maconochie *et al.*, 1997; Popperl and Featherstone, 1992; Tumpel *et al.*, 2007), it is possible that shared local enhancers are used to maintain, reenforce, or secure collinear gene expression, through auto- and cross-regulatory loops. The lack of maintenance of appropriate expression patterns associated with a large deletion within the *HoxD* cluster (Spitz *et al.*, 2001) indeed suggests that sequences within the cluster are important to properly maintain a collinear pattern possibly setup by controls localized outside of the cluster itself.

1. Shared remote enhancers acting on contiguous set of *Hox* genes

In addition to their expression in somitic mesoderm and in the neural tube, *Hox* genes have acquired a variety of functional tasks along with the emergence of vertebrates and the concomitant cluster amplification. As a consequence, while all four clusters are functionally important in the ancestral expression domains, more recent functionalities are associated with either one or a few specific clusters. Good examples of this are given by the *HoxD* cluster, where the seven most anterior genes (from *Hoxd1* to *Hoxd11)* are co-expressed in the intestinal hernia, a region of the gut that corresponds to the future transition between the ileocaecal region and the colon and which develops outside the fetal abdomen (Roberts *et al.*, 1995; Zakany and Duboule, 1999). Likewise, the four contiguous genes localized 5′-most in the cluster (from *Hoxd13* to *Hoxd10)* are co-expressed in future digits of the developing limb buds (Dolle *et al.*, 1989). Two genes, *Hoxd11* and *Hoxd10*, thus show expression in both the developing limbs and the intestinal hernia. However, neither transgenes containing a single *Hoxd* gene and its associated promoter region nor an entire cluster were able to recapitulate expression in any of these domains (Gerard *et al.*, 1993; Herault *et al.*, 1998; Renucci *et al.*, 1992; Spitz *et al.*, 2001; van der Hoeven *et al.*, 1996). Conversely, when the full *HoxD* cluster was deleted but replaced by a single reporter gene, the

latter displayed expression in these structures, indicating that elements localized inside the cluster are dispensable for these particular expression domains (Spitz *et al.*, 2001).

These results suggest that *Hox* genes respond to regulatory cues that are in part imposed by elements localized outside the clusters (remote enhancers), and furthermore, such regulations appear to bear little promoter specificity, since expression of a given gene largely depends upon its relative position within a cluster (Herault *et al.*, 1999; Tarchini and Duboule, 2006; van der Hoeven *et al.*, 1996; Zakany *et al.*, 2004). Accordingly, whenever foreign transcription units were inserted within the *HoxD* cluster, they adopted an expression pattern corresponding to the site of insertion (Herault *et al.*, 1999; Kmita *et al.*, 2000a; van der Hoeven *et al.*, 1996).

2. Controlling enhancer activity: relay, silencer, boundary, and tethering elements

Which mechanisms can control or distribute such global activities? Experimental approaches *in vivo* have shown that the mere relative position of a gene within the cluster, that is, its relative distances from either the 3′ or the 5′ ends are important cues (Tarchini and Duboule, 2006), possibly reflecting a progressive reduction of enhancer efficiency as the distance (number of interspersed promoters) increases. Yet, regardless of how such basic mechanism is implemented, it is helped or refined by elements of well-defined functions in fine-tuning the responses of particular (groups of) genes to this global effect. For example, any *Hoxd* gene localized naturally or artificially at the 5′ extremity of the cluster is always expressed with maximal efficiency when compared with its 3′ located neighbors in developing digits. This regulatory preference for a gene localized at the end of the cluster is a consequence of both the proximity and the presence of a helper element located nearby (Kmita *et al.*, 2002a).

Likewise, the intestinal hernia enhancer does not regulate the two most-posterior *Hoxd* genes because of the presence of an element downstream of *Hoxd12* that isolates *Hoxd12* and *Hoxd13* from this global regulation. This block has a polarity since, upon inversion of this element *in vivo*, *Hoxd12* and *Hoxd13* are activated in the hernia, whereas the isolation is observed on regulations coming from the other side of the cluster (Kmita *et al.*, 2000b). Therefore, the relative extent of global enhancer activities clearly depends on the intricate interplay between the position of the gene within the cluster (distance from 5′ and 3′ ends and the number of promoters in between), and a combination of local tethering or boundary elements that will either favor or restrict the action of these enhancers to a specific set of target genes.

III. CO-EXPRESSION CHROMOSOMAL TERRITORIES, REGULATORY LANDSCAPES, AND GLOBAL CONTROL REGIONS

A. Extending away from the *Hox* cluster: *Evx2* and *Lnp*

One way to precisely assign global enhancer sequences to one or other side of a gene cluster relies upon splitting it into two subclusters. This was done via an engineered chromosomal inversion that separated the *HoxD* cluster into two independent subclusters (Spitz *et al.*, 2005). Expression analyses of embryos carrying such inversions demonstrated that the digit and the genital bud enhancers lie centromeric to (5′ from-) the *HoxD* cluster, whereas the intestinal hernia enhancer lies telomeric to *HoxD* (Spitz *et al.*, 2005). Interestingly, *Evx2* and *Lunapark* (*Lnp*), two genes located centromeric to *HoxD*, are also expressed in developing digits and genital bud with the same specificity as the neighboring posterior *Hoxd* genes (Bastian *et al.*, 1992; Dolle *et al.*, 1994; Spitz *et al.*, 2003), suggesting an extensive sharing of regulatory elements among them.

Evx2 codes for a homeobox-containing transcription factor somewhat related to HOX proteins (Gauchat *et al.*, 2000), and its physical association with *Hox* genes must be an ancestral feature since it is observed in both *HoxA* and *HoxD* vertebrate clusters (Dolle *et al.*, 1994), in the lancelet (Minguillon and Garcia-Fernandez, 2003), and in coral (Miller and Miles, 1993). However, the murine *Evx* genes do not seem to endorse any *Hox*-related function, but rather seem to be required for the specification of specific pools of interneurons within the developing neural tube (Herault *et al.*, 1996; Moran-Rivard *et al.*, 2001).

The case of the *Lnp* gene is even more striking as co-expression with *Hox* genes is observed despite a larger genomic distance. *Lnp* is indeed further away from the *HoxD* cluster than *Evx2* (90 kb against 8 kb), and encodes a protein of unknown function that is conserved in plants, yeasts, and all animals. It is structurally unrelated to homeobox-containing transcription factors, hence *Lnp* bears no relationship with *Hox* genes whatsoever, other than being adjacent to the *HoxD* cluster. Yet *Lnp* is expressed in developing limbs of mice and chicken in a distal posterior domain virtually identical to those of either *Hoxd13* or *Evx2* (Spitz *et al.*, 2003). Besides limbs, *Hoxd*, *Evx2*, and *Lnp* are also co-expressed in the developing external genital organs, and *Evx2* and *Lnp* have overlapping expressions in several domains of the developing central nervous system (CNS). The expression patterns of these genes are nevertheless not fully equivalent in all tissues and, unlike *Hox* genes, neither *Evx2* nor *Lnp* are transcribed in the trunk. Likewise, *Hoxd* genes are not expressed in those neural derivatives where both *Evx2* and *Lnp* are transcribed. Furthermore, *Lnp* is expressed both in the heart and in the eyes, where neither *Evx2* nor any *Hoxd* gene was shown expressed (Spitz *et al.*, 2003).

The existence of such global expression patterns on the top of more "gene-specific" features suggests that the *cis*-acting elements underlying the former may confer the associated regulation in a complete gene-independent fashion, that is, to any transcription unit lying within the realm of action of a hitherto qualified "global enhancer." The extent of the DNA interval within which various promoters will respond to a given global regulation was defined as a "regulatory landscape" (Spitz *et al.*, 2003). In the example described above, two overlapping regulatory landscapes are considered: a "limb landscape," which encompasses *Lnp*, *Evx2* and the posterior part of the *HoxD* cluster, and a "CNS" landscape restricted to *Lnp* and *Evx2* (Fig. 6.1C).

The search for, and identification of, *cis*-acting sequences involved in such global regulations is complicated by their intrinsic property to work at a distance. One method for identifying these elements was designed so as to isolate the enhancer responsible for the above-mentioned limb regulatory landscape. A contig of BAC clones covering the *HoxD* locus and flanking sequences was used as a starting point for the random transposition of a Tn7-based transposon containing a minimal promoter-reporter system (Spitz *et al.*, 2003). In this way, a region was identified that conferred expression to the reporter gene in both distal limb mesenchyme, the genital bud, and the neural tubes, that is, domains where either *Lnp*, *Evx2*, or *Hoxd* genes (or a combination thereof) are normally expressed. This region lies at ca. 200 kb from the *HoxD* cluster, beyond *Lnp*, at the border of a gene desert.

The molecular understanding of the mouse *Ulnaless* (*Ul*) mutation (Davisson and Cattanach, 1990) provided further evidence for the role of this element in controlling gene expression within these regulatory landscapes. *Ulnaless* is an X-ray induced inversion of a ca. 770-kb fragment that includes both *Evx2* and the *HoxD* cluster, with a breakpoint within *Lnp* (Spitz *et al.*, 2003). As a consequence of this inversion, the enhancer element is separated from both *Evx2* and *Hoxd* genes. As expected, these genes show a loss of expression in distal limb mesenchyme as well as in some dorsal neurons, all structures where the enhancer was shown active in a transgenic context (Herault *et al.*, 1997; Peichel *et al.*, 1997; Spitz *et al.*, 2003). Altogether, these various forms of evidence showed that this region contained various regulatory elements, defining at least two overlapping (limb and neural) regulatory landscapes. The term "Global Control Region (GCR)" was created to describe such DNA regions containing several long-range regulations acting on multiple genes (Spitz *et al.*, 2003).

B. *Gremlin/Formin*

Subsequently, regulatory elements of similar nature have been reported, as exemplified by studies of the *limb deformity* (*ld*) locus (Zuniga *et al.*, 2004). The *ld* mutation (Woychik *et al.*, 1985), as well as several spontaneous or

engineered mutant stocks allelic to *ld*, shows a reduction in the number and size of digits, a concurrent loss of digit identities, and a fusion between the ulna and the radius. Positional cloning narrowed down the *ld* gene to a small region on mouse chromosome 2, and in several *ld* alleles, alterations were identified at the 3′ end of the *formin* gene (the gene formerly identified to cause the *ld* phenotype), leading to presumptive proteins with a truncated C-terminus (Mass *et al.*, 1990). *Formin* is indeed expressed in distal limb mesenchyme and its expression is diupted in several *ld* alleles, supporting a role in limb morphogenesis (Zeller *et al.*, 1989). It was subsequently established that the BMP-antagonist *Gremlin*, a critical factor for limb development involved in both *Shh* and *Fgf* signaling feedbacks, was absent in *ld* mutant limbs. Analysis of *Gremlin* expression in control and *ld* animals, as well as genetic analysis, supported a model whereby *Formin/ld* controls the expression of *Gremlin* (Zuniga *et al.*, 1999).

However, additional engineered alleles of the *formin* gene with deletion of the 5′ region failed to reproduce the *ld* phenotypes (Michos *et al.*, 2004) and, conversely, some *ld* alleles were shown to be associated with mutations in the *Gremlin* gene, rather than in *formin*, the two genes lying 80 kb from one another (Zuniga *et al.*, 2004). This contrasting picture was solved by using BAC transgenes, showing that the *ld* alleles affecting the 3′ part of the *formin* gene were not only deleting the protein encoded by this gene but also removing a critical *cis*-acting element required for the expression of *Gremlin* in distal limb mesenchyme (Zuniga *et al.*, 2004). Therefore, much like in the case of the *HoxD* limb regulatory landscape, a *cis*-acting element can work concomitantly on different promoters over a distance of ca. 400 kb, associated either with a gene involved in actin skeleton remodeling (*formin*) or with a BMP antagonist (*gremlin*) (Fig. 6.1B).

C. Co-expression territories

In addition to the *Lnp/Evx2/Hoxd* and *Fmn1/Gremlin* regulatory landscapes, cases have been described of co-expression of adjacent genes that are otherwise not related in any way. For example, *Lmx1b* and its immediate neighbor ALC (Holmes *et al.*, 2003), or the clustered *Dlx5-Dlx6* and adjacent *Dss1* (Crackower *et al.*, 1996) genes, are co-expressed. In the case of the *hGH* and *CD79b* genes, this co-expression was also shown to be imposed by a shared enhancer sequence (Cajiao *et al.*, 2004). However, the question as to whether regulatory landscapes and co-expression territories are widespread, or are restricted to a few anecdotal situations, remains to be firmly established. The development of various technologies to study genome-wide gene expression patterns, such as microarrays, SAGE, or MPSS libraries, has fostered various studies looking for correlation between gene expression profiling on the one hand and chromosomal position on the other. In *Drosophila* and *C. elegans*, several studies (Boutanaev *et al.*, 2002;

Roy *et al.*, 2002; Spellman and Rubin, 2002) have shown that adjacent genes often share expression specificities, forming what Spellman and Rubin described as "co-expression territories."

It is as yet unclear how such co-expression territories relate to regulatory landscapes and how many of them are defined by sequences analogous to GCRs. A key difference between the two notions is that co-expression territories are defined by similarities in expression across multiple tissues or experimental conditions, using semiquantitative approaches, whereas regulatory landscapes are defined by co-expression, as assayed by qualitative *in situ* hybridization, within one tissue or embryonic structure. Therefore, these terms may correspond to rather different situations, underlined by distinct molecular mechanisms.

In vertebrates, only a few co-expression territories have been described (Lercher *et al.*, 2002; Su *et al.*, 2004), and most of them correspond either to clusters of housekeeping genes or to tandem arrays of duplicated genes expressed in a very few tissues, such as clusters of olfactory receptors. This failure to detect tissue-specific co-expression territories and/or regulatory landscapes by genome-wide approaches, for example, in human or mice, may not reflect their absence but, instead, the inadequacy of the currently used "global profiling" methodology. As noted above, control of gene expression in mammals is mostly the result of modular, *cis*-acting elements. Consequently, GCRs are expected to work on top of other gene-specific expression specificities, and the use of "global co-expression" as a criterium would certainly disqualify the *HoxD/Lnp* regulatory landscape because differences between expression patterns would overcome the expression similarities in the limb. We conclude that global profiling may be quite effective in detecting co-expression of genes displaying either a broad (ubiquitous) or restricted (e.g., a single tissue) transcriptional activity. In contrast, it might fall short in cases involving genes sharing a global enhancer sequence for only one particular aspect of otherwise quite divergent expression patterns.

IV. MECHANISMS OF UNDERLYING GLOBAL REGULATION

Gene expression is generally seen as a rather tightly controlled process, yet the observation that adjacent genes often share expression specificities raises interesting questions regarding the mechanism(s) that could potentially be involved (Fig. 6.2).

The description of large co-expression territories in flies, *C. elegans*, and human has led to some mechanistic proposals: for example, in nematodes, most clusters are associated with multicistronic transcripts (operons *sensu stricto*) and are thus defined by the cotranscription, on a single precursor RNA molecule, of several genes (Lercher *et al.*, 2003). In *Drosophila*, co-expression of adjacent genes often involves one strongly expressed gene, whereas others are expressed at much

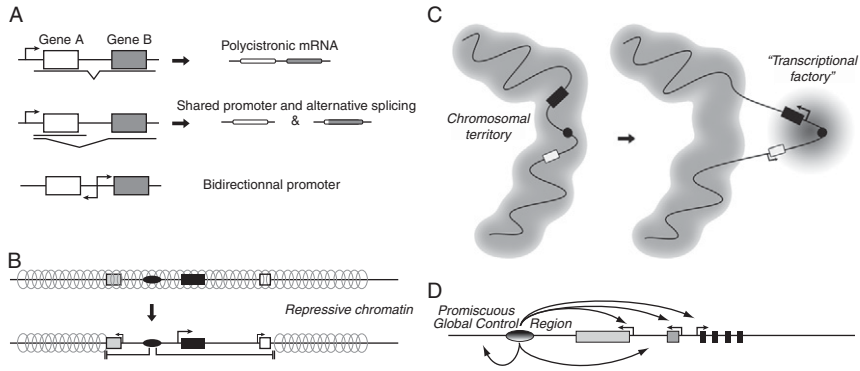

Figure 6.2. Mechanisms associated with co-expression of adjacent genes. (A) Local action. Through polycistronic transcription, alternative splicing, or even bidirectional promoters, two adjacent genes can share regulatory control by the same transcription factors. (B) Large-scale remodeling of chromatin structure. Enhancers can bind transcription factors which then recruit chromatin remodeling complexes that could act over relatively large distances. Genes in their vicinity might become more accessible to the basal transcription machinery, leading to their loose expression. (C) Global relocation of locus into the nuclear space. Enhancers, together with their target gene, might move away from compacted and repressive chromosomal territories to reach "transcription factories," and thus drag along the adjacent genes. (D) Promiscuous global—and usually long-range—enhancers contact all available promoters within their range of action. These elements might act in combination with other enhancers to drive expression in a variety of tissues. Their action depends on the architecture of the locus (distances, number of genes), and might be limited by boundary elements. These different schematized models are not mutually exclusive, and probably correspond to different facets of the mechanisms involved in global and long-range regulation of gene expression.

lower levels (Spellman and Rubin, 2002). Accordingly, it was proposed that these latter genes are transcribed simply because of their close proximity to a strongly transcribed unit, thus appearing as a bystander effect. Such bystander regulation may either involve a shared enhancer, like the GCR, which may contact promoters surrounding the strongly transcribed unit eagerly, with no restricted specificity, or from the spreading of transcriptionally permissive chromatin structures around active genes.

Genomic DNA is usually packaged around histones as well as in larger chromatin structures. It is generally believed that this compaction of DNA into dense chromatin structures is associated with repression of gene expression by prohibiting transcription factors and RNA polymerases to access promoter and enhancer sequences. A release in chromatin structure is necessary for gene activation, and consequently, chromatin found around active genes may be somewhat relaxed, thus increasing the chance for basal transcription factors to access promoter regions, leading to basal transcription of those genes located in

the proximity of strongly expressed genes (for a more complete discussion about this topic, see Sproul *et al.*, 2005). Differential posttranslational modifications of histones, in particular through acetylation or methylation of different residues, have been associated with either an increase or a decrease of transcriptional activity (for a recent review, see Li *et al.*, 2007). Propagation of repressive chromatin from a nucleation center was shown to spread from specific DNA sequences [e.g., Polycomb Response Elements (Papp and Muller, 2006; Schwartz and Pirrotta, 2007), see also Rest/NRSF elements (Lunyak *et al.*, 2002)], and the counteraction of these mechanisms, for example via trithorax group proteins, might appear to be sufficient to keep the locus in a poised/permissive configuration that could allow basal transcription to proceed, not only for the "main" gene but also for the adjacent transcription units. However, genome-wide studies have revealed that these histone marks associated with gene activity are usually confined to discrete *cis*-acting regions (promoter, enhancer), rather than being spread over broad domains, with the striking exception of *Hox* clusters (Bernstein *et al.*, 2005). There is, in general, therefore no straightforward link between large-scale co-expression loci and locus-wide distribution of a specific chromatin/histone pattern, and the possible underlying mechanisms/factors required remain, in most cases, elusive (Sproul *et al.*, 2005).

Several reports have also underlined that the position of a given locus within the nucleus might impact on its expression. As initially shown in the yeast and, subsequently, in vertebrates (reviewed in Kosak and Groudine, 2004), chromosomal regions containing transcriptionally active genes tend to loop out of the condensed chromosomal territory, coming close to the periphery of the nucleus in so-called "transcription factories," that is, regions of the nucleus enriched in components of the transcription machinery. Genes that are co-expressed, but separated by several megabases, have indeed been shown to colocalize in these discrete nuclear compartments (Osborne *et al.*, 2004). It is conceivable that genes localized around actively transcribed genes would be carried along and "passively" benefit from the changes induced, leading to their spurious transcriptional activation. Several loci have evolved boundary elements that limit such bystander effects. For example, the formation of the Active Chromatin Hub at the β-globin locus serves a dual function. It favors an efficient interaction between remote enhancer elements and the associated globin genes (see chapter 4), but it also excludes the neighboring olfactory receptor genes from the influence of this interactive hub (Bulger *et al.*, 2000, 2003). While boundary elements, or other regions with chromatin organizing activity, provide some molecular mechanisms to compartment the genome in distinct functional domains, the extensive embedding of enhancers and genes over large regions might somehow increase the complexity in setting up such ordered organization, and would therefore allow leaky expression of more or less large sets of adjacent genes.

Bidirectional promoters may also lead to the co-expression of neighboring genes, and several eukaryotic promoters have this property (Engstrom *et al.*, 2006). In the majority of cases, bona fide mRNAs are produced together with noncoding, likely nonfunctional, transcripts (Engstrom *et al.*, 2006). However, such bidirectional activity sometimes correlates with genuine co-expression of functional transcripts (Bellizzi *et al.*, 2007; Chen *et al.*, 2007). As recent surveys of gene transcription start sites suggest the existence of a much more complex and extended organization than anticipated, bidirectional promoter sharing might have been overlooked and may be one cause of co-expression (Carninci *et al.*, 2006; Engstrom *et al.*, 2006).

However, promoter sharing can hardly account for those coregulated genes that are organized in a tail-to-tail manner, for example, the *Formin–Gremlin* pair, or various pairs of *Dlx* genes. Likewise, it cannot explain the coregulation of the *Lnp–Evx2–Hoxd* genes in developing limbs. While in the latter example bidirectional promoters likely do exist in the DNA interval between the genes (Sessa *et al.*, 2007), experimental data argue against such a simple mechanism playing a really prominent role. First, the deletion of the region where such a promoter should lie (between *Evx2* and *Hoxd13)* did not impair the activity of the remaining genes (Kmita *et al.*, 2002a). *Hox* gene co-expression has also been proposed to arise through multicistronic or alternatively spliced transcripts. Indeed, some evidence for such phenomena have been described in crustaceans (Shiga *et al.*, 2006), and in mammals, several hybrid transcripts can be found for all four vertebrate *Hox* clusters in ESTs databases. Nevertheless, transgene insertions targeted into this DNA region are still expressed in digits regardless of their orientation (Herault *et al.*, 1999; van der Hoeven *et al.*, 1996), and an engineered inversion of the *Hoxd12* to *Hoxd11* segment did not impact greatly upon the expression of these two genes in developing limbs and genitals (Kmita *et al.*, 2000b). Therefore, within the *Lnp–Hoxd* regulatory landscapes, shared expression patterns are observed irrespective of transcriptional orientation.

In this case, the GCR and associated elements behave rather like a promiscuous enhancer, which is able to contact various promoters within the same cell at the same time. It is as yet unclear whether all the genes in this landscape are transcribed simultaneously or sequentially, as shown for the β-globin genes (Dillon *et al.*, 1997; Wijgerde *et al.*, 1995). Also, the associated chromatin structure has not been studied in detail, even though it appears that chromatin around the *Lnp/Hoxd* interval is decondensed in E9.5 limb buds (Morey *et al.*, 2007). While the GCR is not yet fully active in limbs at this stage (*Lnp*, *Hoxd13*, or GCR-linked reporter genes start to be expressed at E10.5), these results suggest that some global chromatin modifications may be necessary beforehand. Whether such modifications are determined by the GCR is unknown; however, it is clear that the GCR does not only work via a global

modification of the surrounding chromatin, and that its activity is required for proper gene transcription, as observed from the effect of either the deletion or the addition of genes in the landscape (Kmita *et al.*, 2002a; Monge *et al.*, 2003). In such cases, the resulting up- and downregulation of resident genes of the landscape indicate transcriptional reallocations that are at odds with a mere modification of the chromatin structure. Instead, this competition in transcription argues for an active role of the GCR in contacting the different promoters. The identification of the transcription factors bound to the GCR and of the protein complexes involved in the long-range action of this element would help to decipher the precise molecular mechanisms at play in these complex gene regulations.

V. CO-EXPRESSION CHROMOSOMAL TERRITORIES, REGULATORY LANDSCAPES: BYSTANDER EFFECTS OR FUNCTIONAL OPERONS?

Regardless of the nature of the underlying mechanisms, the existence of co-expressed gene clusters raises the issue of both their biological significance and their functional relevance. In this context, clusters containing similar genes, that is, derived from ancestral gene duplication event(s), and clusters containing genes unrelated in their phylogenies may have to be considered separately.

In the former case, for example, the *Hox* or *Dlx* gene clusters, genes may often encode proteins containing similar functional domains, which are thus more susceptible to endorse related—if not similar—functions, such as activating or repressing the same set of target genes. For example, the *Dlx5* and *Dlx6* genes are functionally redundant and mice with a targeted inactivation of either the one or the other have normal limbs, whereas the combined deletion led to a severe ectrodactyly (Robledo *et al.*, 2002). Likewise, the inactivation of *Hoxd13*, *Hoxd12*, or *Hoxd11* in mice led to malformed hands and feet, implying that all these genes are somewhat involved in the patterning of appendages (Davis and Capecchi, 1994; Dolle *et al.*, 1993; Favier *et al.*, 1995; Kondo *et al.*, 1996). In this case however, the defects were much more pronounced whenever *Hoxd13* was mutated, indicating that within the cluster, the functionalities of each gene are not equal, either for extrinsic reasons, such as their expression levels, or for intrinsic reasons, such as their various capacities to compensate for one another's function (Kmita *et al.*, 2002a; Tarchini *et al.*, 2006). This example shows that genes within a co-expression cluster can be involved in similar pathways, yet with very different functional impacts.

In the case of clustering of unrelated genes, there is little evidence that co-expression is associated with related functions and, beyond some anecdotal situations [e.g., Ellis-van Creveld syndrome (Ruiz-Perez *et al.*, 2003)], the function associated with the global expression is generally achieved by only one gene,

the others being dispensable. Neither the *Lnp*, *Evx2* nor *Formin* genes seem to have important function in limb morphogenesis despite their strong expression in the developing limb bud, which results from their position within either the *Hoxd* or the *Gremlin* regulatory landscapes, respectively (Herault *et al.*, 1996; Spitz *et al.*, 2003; Trumpp *et al.*, 1992; Zuniga *et al.*, 2004). In the *hGF/CD79b* locus, both genes are transcriptionally active in the pituitary, yet only hGF was detected at the protein level, while *CD79b* is translated in B cells, where it is important for signal transduction from the B-cell receptor, but not in the pituitary (Cajiao *et al.*, 2004). In this case, expression was proposed to be a nonfunctional consequence of the genes' localization within the highly acetylated chromatin domain defined by the enhancer of the growth-hormone genes. These examples suggest that in most reported cases, co-expression may be a collateral effect of the mechanism implemented to ensure proper transcription in a given place and time of the one gene functionally important within the locus. The observation that the immediate neighboring genes of *Lmx1b* in both mice and chick are expressed in dorsal limb mesenchyme much like *Lmx1b* despite the fact that the genes are not orthologous between these two species (Holmes *et al.*, 2003) also argues that bystander gene activation might occur in most co-expression regions.

If co-expressed clustered genes do not share any functional outcome, which constraint kept them together throughout evolution? If co-expression is judged based on either EST or SAGE or microarrays data, and considering the possible bias of the data set, as mentioned previously, it seems that these gene clusters were maintained more than it should be expected either by chance or by the inertia of the system (Semon and Duret, 2006; Singer *et al.*, 2005).

As discussed previously, clustering may reflect the highly intertwined organization of genes and their associated regulatory elements, which make rearrangements of the cluster difficult. In the cases of the *Formin/Gremlin* and *Lnp/HoxD* gene clusters, breaking the genomic linkage between these pairs of genes would also lead to the separation of a remote enhancer and the gene it must control. Such cases where synteny is conserved because of the complex architecture of the loci are frequent and do not necessarily associate with co-expression, as exemplified with the *Pax6*, *Myf5*, or *Shh* loci (Hadchouel *et al.*, 2000; Kleinjan *et al.*, 2002; Lettice *et al.*, 2002). In these cases, co-expression may occur, in particular if it is a parcimonious way of building the architecture and if expression of the neighboring gene is not deleterious to the embryo. In addition, this would save the organism the difficult task of evolving complex arrays of silencers and boundaries elements to restrict the action of global regulatory elements.

On the other hand, repressive or tethering elements may be used to fine-tune gene expression in complex loci, although these delicate mechanisms might not be as frequent as anticipated, since the potential benefit of their

implementation has to be balanced by their own side effects. The expression of the *Shh* gene in the posterior limb bud, which defines the anterior to posterior asymmetry of the limbs, is controlled by an enhancer localized ca. 800 kb away, lying within the intron of the *Lmbr1* gene (Lettice *et al.*, 2003; Sagai *et al.*, 2005). In early stages of limb development, neither *Lmbr1* nor *Rnf32*, a gene located in between *Shh* and *Lmbr1*, is expressed in the Zone of Polarizing Activity (ZPA), implying that the enhancer is specifically acting on *Shh*. However, at later stages, *Lmbr1* expression is detected, at least in chicken, in the posterior limb together with *Shh* (Maas and Fallon, 2004), suggesting a relaxation of the tethering mechanisms that ensure a preferential interaction between this enhancer and the *Shh* gene during early limb development.

It is also possible that the maintenance of co-expression be functionally constrained, despite the absence of a functional protein produced from the adjacent genes. This is conceivable in the case where the function of a remote enhancer might be facilitated whenever immediate neighboring genes are transcriptionally active. This may help to provide a permissive chromatin configuration around the enhancer and make this region more accessible to transcription factors. In this view, promoter regions of genes localized in between a remote enhancer and a functional target might act as relays to help in setting up the proper interaction.

Besides such purely mechanistic help provided by co-expressed genes, the presence of neighboring promoters may be important to fine-tune the expression level of the functionally active target gene, not by acting in *trans* but rather by interactions in *cis*, for instance through the titration of the regulatory input. As mentioned earlier, the deletion or addition of genes within the *Hoxd* cluster lead to reallocation of the GCR activity and preventing the action of the GCR on *Lnp* and *Evx2* might also redistribute the regulation on the remaining *Hoxd* genes in a way deleterious for limb development. A similar argument might account for the expression of *Lmbr1* under the control of the *Shh* enhancer in chicken limbs, despite any detectable function. In such a view, the apparent absence of *Lmbr1* expression in the mouse ZPA raises the possibility that this differential behavior contributes to the different morphologies of rodent and bird autopods.

Various examples of loss or gain of expression have been reported following transgene insertion at the proximity of regulatory elements (Olson *et al.*, 1996; Qin *et al.*, 2004; Sharpe *et al.*, 1999), probably caused by complex titration-like mechanisms of either positive or negative regulations. In addition, few examples exist of gain of expression caused either by a deletion near a gene promoter (Kmita *et al.*, 2002a) or by a microdeletion removing several hundred kilobases (Kokubu *et al.*, 2003). Therefore, the importance of these titration mechanisms in the regulation of our transcriptome is still elusive and additional examples are required to draw up general rules on this issue.

VI. EVOLUTIONARY IMPLICATIONS OF GLOBAL GENE CONTROL

Such mechanisms may also have contributed to the evolution of gene function. The number and content of gene families are not grossly different among various metazoans, which indicates that novel function were generally not associated with the evolution of novel genes, but rather emerged by the functional redeployment of preexisting transcription units. Novel functionalities may have resulted from changes in the coding sequences of a gene by altering the properties of the protein or by modifying those regulatory modules that control the expression of a given gene. The latter possibility, suggested 30 years ago by King and Wilson (1975), has now received ample experimental validation in different systems (reviewed by Carroll, 2005).

These changes generally occurred either by slight alterations of existing regulations or by the loss of ancestral *cis*-acting elements, and the dissection of the regulatory mechanisms that control highly multifunctional genes has revealed that their expression usually results from the addition of small independent modules. These modules individually define specific aspects of the overall expression pattern and can be localized quite far away from the promoter they will interact with. Altogether, this suggests that newly evolved enhancer sequences generally correspond to novel functions; however, the mechanisms underlying the emergence of these sequences and their recruitment by target promoters remain unknown.

In this respect, regulatory landscapes (as defined by the presence of global enhancers) might provide a context wherein regulatory innovations may occur at an interesting frequency. First, GCR-like sequences (DNA segments containing several enhancer sequences) already have the capacity to interact with many genes, over large distances. Should a GCR evolve a modified activity, for example, through random drift of its DNA sequence, this novel activity could be tested on a variety of potential target genes. Also, some GCRs seem to synergize with other enhancers' sequences, which themselves may have more limited ranges of action either in terms of distance, number of target genes, or in their sensitivity to boundary elements (Gonzalez *et al.*, 2007), to further extend the GCR's realm of action. It is therefore conceivable that regulatory landscapes, while not favoring the emergence of enhancers, can facilitate their further association with an appropriate target. Some steps in the evolution of *Hoxd* gene regulation in tetrapods may have followed such a scenario, taking advantage of the intrinsic long-range property of a remote enhancer, originally acting in the CNS to control *Lnp* and *Evx2* expression, to gradually evolve limb enhancer activity, leading to the progressive establishment of a limb-specific regulatory landscape (Gonzalez *et al.*, 2007; Spitz *et al.*, 2003).

The promiscuous action of GCRs, or of other long-range acting enhancers, can thus contribute significantly to the evolution of gene function by increasing the opportunities to select for novel expression specificities.

Consequently, co-expression territories or regulatory landscapes may represent rather unstable situations (in an evolutionary time frame), prone to evolve into more specific gene-enhancer interactions, as suggested by both the persistence of the phenomenon and the relative lack of maintenance of these territories in different species. In this respect, the highly intermingled distribution of genes and enhancers along chromosomes perhaps testifies to the successive ancestral regulatory landscapes and promiscuous interactions that subsequently evolved into more exclusive interactions.

The situation may have been quite different when clusters of evolutionary related genes are considered. Indeed, in this case the enhancer likely existed prior to the duplication that led to the formation of the cluster and horizontal gene duplication could generate slightly different proteins that may be controlled in a similar way. Tandem duplications, by releasing some constraints on the duplicated genes, offer the possibility to have either one gene or both, evolving a new function, whereas the ancestral function is either retained by one of the duplicates or distributed between all copies. In such a situation, keeping duplicated genes under the same control mechanism provides the possibility to use their differentially evolved properties in a coherent manner.

The control of expression from a single shared regulatory mechanism, by distributing its activity either homogeneously or sequentially in time and space, can both provide this coherence and prevent potential dosage effects associated with overexpression, as might occur in the case where all gene duplicates were to have their own enhancer. An example is given by the *Dlx5–Dlx6* gene-pair: while the genes appear functionally redundant, the proteins evolved such that the domains required for chondrogenesis are significantly different. They might assemble on different target genes or, alternatively, form different complexes on the same targets (Hsu *et al.*, 2006), thus leading to near equivalent outputs. Such a situation provides robustness to the skeletal differentiation program by transforming an ancestral circuit into two parallel ones that use different plugs to connect on the same targets.

At the β-globin locus, the locus control region (LCR) is used to switch from an embryonic to adult isoform of a protein. Here the same mechanism is used both to ensure a suitable level of expression for the most appropriate isoform at a given time and to turn down the other genes. Similar temporal switches, implemented by a global mechanism, have been reported to involve a shared mechanism, even though the temporal overlap between the different genes of the cluster may be reduced to the minimal [e.g., α-fetoprotein/albumin, *Rhox* gene cluster (Maclean *et al.*, 2005)]. The *Hox* clusters are another example of this regulatory logic, whereby the distribution of the genes along the chromosomes is translated into a coherent spatiotemporal pattern of activity following the action of global mechanisms. Similarly, though at the single cell level, the exclusive expression of a single olfactory receptor gene in a given neuron from the

olfactory bulb may also depend upon the use of a shared enhancer (Lomvardas *et al.*, 2006) that can act not only on the cluster of receptors localized nearby on the same chromosome but also on those localized on different chromosomes.

VII. GLOBAL REGULATION, CHROMOSOMAL ARCHITECTURE, AND GENETIC DISORDERS

Because GCRs can act over long distances and on several genes, they likely contributed to the preservation of syntenies along the chromosomes of various different animals. Likewise, chromosomal rearrangements occurring within a regulatory landscape might affect genes at a distance from the breakpoint, potentially leading to both loss of function, whenever genes are moved away from the GCR, and gain of function, whenever genes are brought within the zone of influence of a particular GCR. In humans, several genetic disorders are associated with recurrent rearrangements, for example, the deletion between low copy repeats (Ji *et al.*, 2000). In such cases, attention is generally focused on the genes directly affected by the rearrangements, which may lead to an incorrect understanding of the underlying mechanistic cause, as illustrated by the case of the *ld* mutation. The *formin* gene was considered as causative of the *ld* phenotype because chromosomal rearrangement disrupted this gene in multiple *ld* alleles, a hypothesis strengthened by the observed expression of *formin* in the embryonic structures affected by the mutation. Subsequently, it was shown that the gene responsible for the *ld* phenotype (*Gremlin*) is located 80 kb away from the various mutagenic events (Zuniga *et al.*, 2004). Similarly, the dominant *Ul* mutation is primarily a disruption of the *Lnp* gene through the presence of an inversion breakpoint. Because this inversion disrupts the GCR–*Lnp*–*Evx2*–*HoxD* regulatory landscape, it also involves tissue-specific loss of function of *Hoxd* and *Evx2*, as well as gain of function of *Hoxd* genes in other domains of the developing limbs (Spitz *et al.*, 2003). This latter case is a paradigm of the multiple, complex defects that can be obtained whenever regulatory landscapes are affected by large chromosomal rearrangements. It also nicely exemplifies the great difficulty in understanding the phenotype–genotype relationships of such complex mutations outside the context of global regulations.

VIII. CONCLUDING REMARKS

Over the past 30 years, our thinking of regulation of gene expression has been very much influenced by the pioneering work done on viral and bacterial systems, as well as in cultured eukaryotic cells. The general understanding was that genes and their regulations could be considered as separate units, and that

the interactions between these units were mostly occurring in *trans*. In this ultimate genetic reductionism view, every gene is associated with (or responsible for) one particular function. This superficial view of gene function and regulation has persisted, due to both its mechanistic simplicity and paradigmatic value to account for orthodox (gradualist) evolutionary modifications; any character may indeed be "improved" in isolation since the underlying genetic circuitry can be taken in isolation.

However, recent progress in mouse molecular genetics and human genetics, following the sequencing of several genomes, is now revealing a slightly different picture, where various components are more interdependent on each other than previously anticipated, and where genes and their regulations can no longer be considered in isolation, out of their genomic context. While a large proportion of gene regulatory events are likely to occur in *trans*, several phenomena testify to the existence of global *cis*-mechanisms, which we will need to integrate in our future thinking of biological systems. Beside the mere mechanistic interest, both human genetics and evolutionary biology may rapidly benefit from these developments. While the concept of regulatory landscapes may help understand genetic syndromes for which no simple molecular etiology is available, it may also give us new avenues to think about the evolution of regulations, and hence the emergence of evolutionary novelties.

References

Bastian, H., Gruss, P., Duboule, D., and Izpisua-Belmonte, J. C. (1992). The murine even-skipped-like gene Evx-2 is closely linked to the Hox-4 complex, but is transcribed in the opposite direction. *Mamm. Genome* **3,** 241–243.

Bellizzi, D., Dato, S., Cavalcante, P., Covello, G., Di Cianni, F., Passarino, G., Rose, G., and De Benedictis, G. (2007). Characterization of a bidirectional promoter shared between two human genes related to aging: SIRT3 and PSMD13. *Genomics* **89,** 143–150.

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., III, Gingeras, T. R., Schreiber, S. L., and Lander, E. S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120,** 169–181.

Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W. L., Duke, K., Kiraly, M., and Kim, S. K. (2002). A global analysis of Caenorhabditis elegans operons. *Nature* **417,** 851–854.

Boutanaev, A. M., Kalmykova, A. I., Shevelyov, Y. Y., and Nurminsky, D. I. (2002). Large clusters of co-expressed genes in the *Drosophila* genome. *Nature* **420,** 666–669.

Brooke, N. M., Garcia-Fernandez, J., and Holland, P. W. (1998). The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* **392,** 920–922.

Bulger, M., Bender, M. A., van Doorninck, J. H., Wertman, B., Farrell, C. M., Felsenfeld, G., Groudine, M., and Hardison, R. (2000). Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse beta-globin gene clusters. *Proc. Natl. Acad. Sci. USA* **97,** 14560–14565.

Bulger, M., Schubeler, D., Bender, M. A., Hamilton, J., Farrell, C. M., Hardison, R. C., and Groudine, M. (2003). A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. *Mol. Cell. Biol.* **23,** 5234–5244.

Cajiao, I., Zhang, A., Yoo, E. J., Cooke, N. E., and Liebhaber, S. A. (2004). Bystander gene activation by a locus control region. *EMBO J.* **23,** 3854–3863.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A., Taylor, M. S., Engström, M. S., Frith, M. C., Forrest, A. R., Alkema, W. B., *et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38,** 626–635.

Carroll, S. B. (2005). Evolution at two levels: On genes and form. *PLoS Biol.* **3,** e245.

Carroll, S. B. (2006). "Endless Forms Most Beautiful: The New Science of Evo Devo." W. W. Norton, New York.

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.

Chen, P. Y., Chang, W. S., Chou, R. H., Lai, Y. K., Lin, S. C., Chi, C. Y., and Wu, C. W. (2007). Two non-homologous brain diseases-related genes, SERPINI1 and PDCD10, are tightly linked by an asymmetric bidirectional promoter in an evolutionarily conserved manner. *BMC Mol. Biol.* **8,** 2.

Chiu, C. H., Amemiya, C. T., Carr, J. L., Bhargava, J., Hwang, J. K., Shashikant, C. S., Ruddle, F. H., and Wagner, G. P. (2000). A recombinogenic targeting method to modify large-inserts for cis-regulatory analysis in transgenic mice: Construction and expression of a 100-kb, zebrafish Hoxa-11b-lacZ reporter gene. *Dev. Genes Evol.* **210,** 105–109.

Christoffels, V. M., Keijser, A. G., Houweling, A. C., Clout, D. E., and Moorman, A. F. (2000). Patterning the embryonic heart: Identification of five mouse Iroquois homeobox genes in the developing heart. *Dev. Biol.* **224,** 263–274.

Crackower, M. A., Scherer, S. W., Rommens, J. M., Hui, C. C., Poorkaj, P., Soder, S., Cobben, J. M., Hudgins, L., Evans, L., and Tsui, L. C. (1996). Characterization of the split hand/split foot malformation locus SHFM1 at 7q21.3-q22.1 and analysis of a candidate gene for its expression during limb development. *Hum. Mol. Genet.* **5,** 571–579.

Davidson, E. H. (2006). "The Regulatory Genome: Gene Regulatory Networks in Development and Evolution". Academic Press/Elsevier, San Diego.

Davis, A. P., and Capecchi, M. R. (1994). Axial homeosis and appendicular skeleton defects in mice with a targeted disruption of hoxd-11. *Development* **120,** 2187–2198.

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

de la Calle-Mustienes, E., Feijoo, C. G., Manzanares, M., Tena, J. J., Rodriguez-Seguel, E., Letizia, A., Allende, M. L., and Gomez-Skarmeta, J. L. (2005). A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. *Genome Res.* **15,** 1061–1072.

Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P., and Grosveld, F. (1997). The effect of distance on long-range chromatin interactions. *Mol. Cell* **1,** 131–139.

Dolle, 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.

Dolle, P., Dierich, A., LeMeur, M., Schimmang, T., Schuhbaur, B., Chambon, P., and Duboule, D. (1993). Disruption of the Hoxd-13 gene induces localized heterochrony leading to mice with neotenic limbs. *Cell* **75,** 431–441.

Dolle, P., Fraulob, V., and Duboule, D. (1994). Developmental expression of the mouse Evx-2 gene: Relationship with the evolution of the HOM/Hox complex. *Dev* (Suppl.), 143–153.

Duboule, D. (1998). Vertebrate hox gene regulation: Clustering and/or colinearity? *Curr. Opin. Genet. Dev.* **8,** 514–518.

Duboule, D., and Wilkins, A. S. (1998). The evolution of 'bricolage'. *Trends. Genet.* **14,** 54–59.

Duboule, D., Baron, A., Mahl, P., and Galliot, B. (1986). A new homeo-box is present in overlapping cosmid clones which define the mouse Hox-1 locus. *EMBO J.* **5,** 1973–1980.

Engstrom, P. G., Suzuki, H., Ninomiya, N., Akalin, A., Sessa, L., Lavorgna, G., Brozzi, A., Luzi, L., Lam Tan, S., Yang, L., Kunarso, G., Ng, E. L.-C., *et al.* (2006). Complex loci in human and mouse genomes. *PLoS Genet.* **2,** e47.

Favier, B., Le Meur, M., Chambon, P., and Dolle, P. (1995). Axial skeleton homeosis and forelimb malformations in Hoxd-11 mutant mice. *Proc. Natl. Acad. Sci. USA* **92,** 310–314.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151,** 1531–1545.

Gallardo, M. E., Lopez-Rios, J., Fernaud-Espinosa, I., Granadino, B., Sanz, R., Ramos, C., Ayuso, C., Seller, M. J., Brunner, H. G., Bovolenta, P., and Rodríguez de Córdoba, S. (1999). Genomic cloning and characterization of the human homeobox gene SIX6 reveals a cluster of SIX genes in chromosome 14 and associates SIX6 hemizygosity with bilateral anophthalmia and pituitary anomalies. *Genomics* **61,** 82–91.

Garcia-Fernandez, J. (2005). The genesis and evolution of homeobox gene clusters. *Nat. Rev. Genet.* **6,** 881–892.

Gauchat, D., Mazet, F., Berney, C., Schummer, M., Kreger, S., Pawlowski, J., and Galliot, B. (2000). Evolution of Antp-class genes and differential expression of Hydra Hox/paraHox genes in anterior patterning. *Proc. Natl. Acad. Sci. USA* **97,** 4493–4498.

Gerard, M., Duboule, D., and Zakany, 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.

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

Glusman, G., Yanai, I., Rubin, I., and Lancet, D. (2001). The complete human olfactory subgenome. *Genome Res.* **11,** 685–702.

Gomez-Skarmeta, J. L., and Modolell, J. (2002). Iroquois genes: Genomic organization and function in vertebrate neural development. *Curr. Opin. Genet. Dev.* **12,** 403–408.

Gonzalez, F., Duboule, D., and Spitz, F. (2007). Transgenic analysis of Hoxd gene regulation during digit development. *Dev. Biol.* **306**(2), 847–859.

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.

Gray, T. A., Saitoh, S., and Nicholls, R. D. (1999). An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proc. Natl. Acad. Sci. USA* **96,** 5616–5621.

Guris, D. L., Duester, G., Papaioannou, V. E., and Imamoto, A. (2006). Dose-dependent interaction of Tbx1 and Crkl and locally aberrant RA signaling in a model of del22q11 syndrome. *Dev. Cell* **10,** 81–92.

Hadchouel, J., Tajbakhsh, S., Primig, M., Chang, T. H., Daubas, P., Rocancourt, D., and Buckingham, M. (2000). Modular long-range regulation of Myf5 reveals unexpected heterogeneity between skeletal muscles in the mouse embryo. *Development* **127,** 4455–4467.

Hayday, A. C., Gillies, S. D., Saito, H., Wood, C., Wiman, K., Hayward, W. S., and Tonegawa, S. (1984). Activation of a translocated human c-myc gene by an enhancer in the immunoglobulin heavy-chain locus. *Nature* **307,** 334–340.

Herault, 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.

Herault, Y., Fraudeau, N., Zakany, 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.

Herault, 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,** 1669–1677.

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

Hesse, M., Zimek, A., Weber, K., and Magin, T. M. (2004). Comprehensive analysis of keratin gene clusters in humans and rodents. *Eur. J. Cell Biol.* **83,** 19–26.

Holmes, G., Crooijmans, R., Groenen, M., and Niswander, L. (2003). ALC (adjacent to LMX1 in chick) is a novel dorsal limb mesenchyme marker. *Gene Expr. Patterns* **3,** 735–741.

Houweling, A. C., Dildrop, R., Peters, T., Mummenhoff, J., Moorman, A. F., Ruther, U., and Christoffels, V. M. (2001). Gene and cluster-specific expression of the Iroquois family members during mouse development. *Mech. Dev.* **107,** 169–174.

Hsu, S. H., Noamani, B., Abernethy, D. E., Zhu, H., Levi, G., and Bendall, A. J. (2006). Dlx5- and Dlx6-mediated chondrogenesis: Differential domain requirements for a conserved function. *Mech. Dev.* **123,** 819–830.

Huntley, S., Baggott, D. M., Hamilton, A. T., Tran-Gyamfi, M., Yang, S., Kim, J., Gordon, L., Branscomb, E., and Stubbs, L. (2006). A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res.* **16,** 669–677.

Jacob, F., Perrin, D., Sanchez, C., and Monod, J. (1960). Operon: A group of genes with the expression coordinated by an operator. *C. R. Hebd. Seances Acad. Sci.* **250,** 1727–1729.

Jagla, K., Bellard, M., and Frasch, M. (2001). A cluster of *Drosophila* homeobox genes involved in mesoderm differentiation programs. *Bioessays* **23,** 125–133.

Ji, Y., Eichler, E. E., Schwartz, S., and Nicholls, R. D. (2000). Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res.* **10,** 597–610.

Juan, A. H., and Ruddle, F. H. (2003). Enhancer timing of Hox gene expression: Deletion of the endogenous Hoxc8 early enhancer. *Development* **130,** 4823–4834.

Katoh, M. (2002). WNT and FGF gene clusters (review). *Int. J. Oncol.* **21,** 1269–1273.

King, M. C., and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science* **188,** 107–116.

Kirschner, M., and Gerhart, J. (1998). Evolvability. *Proc. Natl. Acad. Sci. USA* **95,** 8420–8427.

Kleinjan, D. A., Seawright, A., Elgar, G., and van Heyningen, V. (2002). Characterization of a novel gene adjacent to PAX6, revealing synteny conservation with functional significance. *Mamm. Genome* **13,** 102–107.

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

Kmita, M., van Der Hoeven, F., Zakany, J., Krumlauf, R., and Duboule, D. (2000a). Mechanisms of Hox gene colinearity: Transposition of the anterior Hoxb1 gene into the posterior HoxD complex. *Genes Dev.* **14,** 198–211.

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

Kmita, M., Fraudeau, N., Herault, Y., and Duboule, D. (2002a). Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs. *Nature* **420,** 145–150.

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

Kokubu, C., Wilm, B., Kokubu, T., Wahl, M., Rodrigo, I., Sakai, N., Santagati, F., Hayashizaki, Y., Suzuki, M., Yamamura, M., Abe, K., and Imai, K. (2003). Undulated short-tail deletion mutation in the mouse ablates Pax1 and leads to ectopic activation of neighboring Nkx2–2 in domains that normally express Pax1. *Genetics* **165,** 299–307.

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

Kosak, S. T., and Groudine, M. (2004). Gene order and dynamic domains. *Science* **306,** 644–647.

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

Lecaudey, V., Anselme, I., Dildrop, R., Ruther, U., and Schneider-Maunoury, S. (2005). Expression of the zebrafish Iroquois genes during early nervous system formation and patterning. *J. Comp. Neurol.* **492,** 289–302.

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

Lee, S. J. (1991). Expression of growth/differentiation factor 1 in the nervous system: Conservation of a bicistronic structure. *Proc. Natl. Acad. Sci. USA* **88,** 4250–4254.

Lee, K. Z., and Sommer, R. J. (2003). Operon structure and trans-splicing in the nematode Pristionchus pacificus. *Mol. Biol. Evol.* **20,** 2097–2103.

Lemons, D., and McGinnis, W. (2006). Genomic evolution of Hox gene clusters. *Science* **313,** 1918–1922.

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.

Lercher, M. J., Blumenthal, T., and Hurst, L. D. (2003). Coexpression of neighboring genes in Caenorhabditis elegans is mostly due to operons and duplicate genes. *Genome Res.* **13,** 238–243.

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

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12,** 1725–1735.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. *Nature* **424,** 147–151.

Li, B., Carey, M., and Workman, J. L. (2007). The role of chromatin during transcription. *Cell* **128,** 707–719.

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell* **126,** 403–413.

Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M., and Frazer, K. A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288,** 136–140.

Luke, G. N., Castro, L. F., McLay, K., Bird, C., Coulson, A., and Holland, P. W. (2003). Dispersal of NK homeobox gene clusters in amphioxus and humans. *Proc. Natl. Acad. Sci. USA* **100,** 5292–5295.

Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S.-H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., Mandel, G., and Rosenfeld, M. G. (2002). Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* **298,** 1747–1752.

Maas, S. A., and Fallon, J. F. (2004). Isolation of the chicken Lmbr1 coding sequence and characterization of its role during chick limb development. *Dev. Dyn.* **229,** 520–528.

Maclean, J. A., Chen, M. A., II, Wayne, C. M., Bruce, S. R., Rao, M., Meistrich, M. L., Macleod, C., and Wilkinson, M. F. (2005). Rhox: A new homeobox gene cluster. *Cell* **120,** 369–382.

Maconochie, M. K., Nonchev, S., Studer, M., Chan, S. K., Popperl, H., Sham, M. H., Mann, R. S., and Krumlauf, R. (1997). Cross-regulation in the mouse HoxB complex: The expression of Hoxb2 in rhombomere 4 is regulated by Hoxb1. *Genes Dev.* **11,** 1885–1895.

Mass, R. L., Zeller, R., Woychik, R. P., Vogt, T. F., and Leder, P. (1990). Disruption of formin-encoding transcripts in two mutant limb deformity alleles. *Nature* **346,** 853–855.

McEwen, G. K., Woolfe, A., Goode, D., Vavouri, T., Callaway, H., and Elgar, G. (2006). Ancient duplicated conserved noncoding elements in vertebrates: A genomic and functional analysis. *Genome Res.* **16,** 451–465.

McLysaght, A., Hokamp, K., and Wolfe, K. H. (2002). Extensive genomic duplication during early chordate evolution. *Nat. Genet.* **31,** 200–204.

Michos, O., Panman, L., Vintersten, K., Beier, K., Zeller, R., and Zuniga, A. (2004). Gremlin-mediated BMP antagonism induces the epithelial-mesenchymal feedback signaling controlling metanephric kidney and limb organogenesis. *Development* **131,** 3401–3410.

Miller, D. J., and Miles, A. (1993). Homeobox genes and the zootype. *Nature* **365,** 215–216.

Minguillon, C., and Garcia-Fernandez, J. (2003). Genesis and evolution of the Evx and Mox genes and the extended Hox and ParaHox gene clusters. *Genome Biol.* **4,** R12.

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.

Moran-Rivard, L., Kagawa, T., Saueressig, H., Gross, M. K., Burrill, J., and Goulding, M. (2001). Evx1 is a postmitotic determinant of v0 interneuron identity in the spinal cord. *Neuron* **29,** 385–399.

Morey, C., Da Silva, N. R., Perry, P., and Bickmore, W. A. (2007). Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development* **134,** 909–919.

Mummenhoff, J., Houweling, A. C., Peters, T., Christoffels, V. M., and Ruther, U. (2001). Expression of Irx6 during mouse morphogenesis. *Mech. Dev.* **103,** 193–195.

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.

Nusse, R. (2001). An ancient cluster of Wnt paralogues. *Trends Genet.* **17,** 443.

Olson, E. N., Arnold, H. H., Rigby, P. W., and Wold, B. J. (1996). Know your neighbors: Three phenotypes in null mutants of the myogenic bHLH gene MRF4. *Cell* **85,** 1–4.

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W., and Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* **36,** 1065–1071.

Papp, B., and Muller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* **20,** 2041–2054.

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.

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., *et al.* (2006). *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444,** 499–502.

Peters, T., Dildrop, R., Ausmeier, K., and Ruther, U. (2000). Organization of mouse Iroquois homeobox genes in two clusters suggests a conserved regulation and function in vertebrate development. *Genome Res.* **10,** 1453–1462.

Popperl, H., and Featherstone, M. S. (1992). An autoregulatory element of the murine Hox-4.2 gene. *EMBO J.* **11,** 3673–3680.

Qin, Y., Kong, L. K., Poirier, C., Truong, C., Overbeek, P. A., and Bishop, C. E. (2004). Long-range activation of Sox9 in Odd Sex (Ods) mice. *Hum. Mol. Genet.* **13,** 1213–1218.

Renucci, A., Zappavigna, V., Zakany, J., Izpisua-Belmonte, J. C., Burki, K., and Duboule, D. (1992). Comparison of mouse and human HOX-4 complexes defines conserved sequences involved in the regulation of Hox-4.4. *EMBO J.* **11,** 1459–1468.

Roberts, D. J., Johnson, R. L., Burke, A. C., Nelson, C. E., Morgan, B. A., and Tabin, C. (1995). Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. *Development* **121,** 3163–3174.

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

Rodriguez, I. (2005). Remarkable diversity of mammalian pheromone receptor repertoires. *Proc. Natl. Acad. Sci. USA* **102,** 6639–6640.

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.

Ruiz-Perez, V. L., Tompson, S. W. J., Blair, H. J., Espinoza-Valdez, C., Lapunzina, P., Silva, E. O., Hamel, B., Gibbs, J. L., Young, I. D., Wright, M. J., and Goodship, J. A. (2003). Mutations in two nonhomologous genes in a head-to-head configuration cause Ellis-van Creveld syndrome. *Am. J. Hum. Genet.* **72,** 728–732.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132,** 797–803.

Schwartz, Y. B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nat. Rev. Genet.* **8,** 9–22.

Semon, M., and Duret, L. (2006). Evolutionary origin and maintenance of coexpressed gene clusters in mammals. *Mol. Biol. Evol.* **23,** 1715–1723.

Sessa, L., Breiling, A., Lavorgna, G., Silvestri, L., Casari, G., and Orlando, V. (2007). Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. *RNA* **13,** 223–239.

Shannon, M., Hamilton, A. T., Gordon, L., Branscomb, E., and Stubbs, L. (2003). Differential expansion of zinc-finger transcription factor loci in homologous human and mouse gene clusters. *Genome Res.* **13,** 1097–1110.

Sharpe, J., Nonchev, S., Gould, A., Whiting, J., and Krumlauf, R. (1998). Selectivity, sharing and competitive interactions in the regulation of Hoxb genes. *EMBO J.* **17,** 1788–1798.

Sharpe, J., Lettice, L., Hecksher-Sorensen, J., Fox, M., Hill, R., and Krumlauf, R. (1999). Identification of sonic hedgehog as a candidate gene responsible for the polydactylous mouse mutant Sasquatch. *Curr. Biol.* **9,** 97–100.

Shiga, Y., Sagawa, K., Takai, R., Sakaguchi, H., Yamagata, H., and Hayashi, S. (2006). Transcriptional read-through of Hox genes Ubx and Antp and their divergent post-transcriptional control during crustacean evolution. *Evol. Dev.* **8,** 407–414.

Singer, G. A., Lloyd, A. T., Huminiecki, L. B., and Wolfe, K. H. (2005). Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. *Mol. Biol. Evol.* **22,** 767–775.

Slager, R. E., Newton, T. L., Vlangos, C. N., Finucane, B., and Elsea, S. H. (2003). Mutations in RAI1 associated with Smith-Magenis syndrome. *Nat. Genet.* **33,** 466–468.

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 Zakany, J. (2001). Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev.* **15,** 2209–2214.

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

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

Sproul, D., Gilbert, N., and Bickmore, W. A. (2005). The role of chromatin structure in regulating the expression of clustered genes. *Nat. Rev. Genet.* **6,** 775–781.

Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M. P., Walker, J. R., *et al.* (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **101,** 6062–6067.

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.

Tarchini, B., and Duboule, D. (2006). Control of Hoxd genes' collinearity during early limb development. *Dev. Cell* **10,** 93–103.

Tarchini, B., Duboule, D., and Kmita, M. (2006). Regulatory constraints in the evolution of the tetrapod limb anterior-posterior polarity. *Nature* **443,** 985–988.

Trumpp, A., Blundell, P. A., de la Pompa, J. L., and Zeller, R. (1992). The chicken limb deformity gene encodes nuclear proteins expressed in specific cell types during morphogenesis. *Genes Dev.* **6,** 14–28.

Tumpel, S., Cambronero, F., Ferretti, E., Blasi, F., Wiedemann, L. M., and Krumlauf, R. (2007). Expression of Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1. *Dev. Biol.* **302,** 646–660.

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

Warming, S., Costantino, N., Court, D. L., Jenkins, N. A., and Copeland, N. G. (2005). Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33,** e36.

Wijgerde, M., Grosveld, F., and Fraser, P. (1995). Transcription complex stability and chromatin dynamics *in vivo*. *Nature* **377,** 209–213.

Wotton, K. R., and Shimeld, S. M. (2006). Comparative genomics of vertebrate Fox cluster loci. *BMC Genomics* **7,** 271.

Woychik, R. P., Stewart, T. A., Davis, L. G., D'Eustachio, P., and Leder, P. (1985). An inherited limb deformity created by insertional mutagenesis in a transgenic mouse. *Nature* **318,** 36–40.

Wu, Q., and Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* **97,** 779–790.

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

Zakany, J., Gerard, M., Favier, B., and Duboule, D. (1997). Deletion of a HoxD enhancer induces transcriptional heterochrony leading to transposition of the sacrum. *EMBO J.* **16,** 4393–4402.

Zakany, J., Kmita, M., and Duboule, D. (2004). A dual role for Hox genes in limb anterior-posterior asymmetry. *Science* **304,** 1669–1672.

Zeller, R., Jackson-Grusby, L., and Leder, P. (1989). The limb deformity gene is required for apical ectodermal ridge differentiation and anteroposterior limb pattern formation. *Genes Dev.* **3,** 1481–1492.

Zulch, A., Becker, M. B., and Gruss, P. (2001). Expression pattern of Irx1 and Irx2 during mouse digit development. *Mech. Dev.* **106,** 159–162.

Zuniga, A., Haramis, A. P., McMahon, A. P., and Zeller, R. (1999). Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. *Nature* **401,** 598–602.

Zuniga, A., Michos, O., Spitz, F., Haramis, A.-P. G., Panman, L., Galli, A., Vintersten, K., Klasen, C., Mansfield, W., Kuc, S., Duboule, D., Dono, R., *et al.* (2004). Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev.* **18,** 1553–1564.