The regulation of Hox gene expression during animal development

Moisés Mallo1,* and Claudio R. Alonso2,*

Summary
Hox genes encode a family of transcriptional regulators that elicit distinct developmental programmes along the head-to-tail axis of animals. The specific regional functions of individual Hox genes largely reflect their restricted expression patterns, the disruption of which can lead to developmental defects and disease. Here, we examine the spectrum of molecular mechanisms controlling Hox gene expression in model vertebrates and invertebrates and find that a diverse range of mechanisms, including nuclear dynamics, RNA processing, microRNA and translational regulation, all concur to control Hox gene outputs. We propose that this complex multi-tiered regulation might contribute to the robustness of Hox expression during development.

Key words: Hox genes, Gene regulation, Chromatin, RNA processing, MicroRNAs, Translation, Drosophila, Mouse

Introduction
Hox genes provide a paradigm for several areas in modern biology. First, from a developmental perspective, they constitute a genetic system involved in the allocation of segmental identity along animal body axes. As such, they offer an opportunity to investigate how transcription factors organise networks of subordinate genes to guide the behaviour of cell populations during morphogenesis. Second, given their remarkable evolutionary conservation across distant animal phyla, they represent an abstract system of cardinal information able to operate within a wide spectrum of invertebrates and vertebrates, bringing about the question of how the same set of molecular mechanisms involved in the generation of widely diverse developmental programmes. Third, their genomic organisation and molecular regulation is highly complex, opening up the possibility to look at the molecular mechanisms by which genomic information is extracted and subsequently converted into physiological and morphological processes as development progresses. This Review focuses on this latter gene regulatory dimension, aiming to provide an updated perspective on the variety of molecular mechanisms involved in Hox gene regulation. We consider that this is important given that models of how genes are molecularly regulated constrain our understanding of how they might exert their control over development in health and disease. For the developmental and evolutionary functions of Hox genes, the reader is referred to recent reviews (Duboule, 2007; Gehring et al., 2009; Lemons and McGinnis, 2006; Maeda and Karch, 2009; Mallo et al., 2010; Merabet et al., 2005; Pearson et al., 2005; Pick and Heffer, 2012).

This Review first provides some general features of the Hox genes and then uses this foundation to look at problems of chromosomal and chromatin regulation, transcriptional control, RNA processing and microRNA (miRNA) regulation before turning to the modulation of protein translation. The main discussion is set up using Drosophila and mouse as key experimental models, bringing in information from other systems where useful and possible. Owing to space limitations, we give priority to regulatory mechanisms other than classical transcriptional regulation, given that these have been well covered elsewhere (Alexander et al., 2009; Maeda and Karch, 2009; Tschopp and Duboule, 2011). This Review thus offers an account of the wide variety of molecular processes able to affect the regulation of Hox gene expression during animal development.

Key features of Hox genes
The Hox genes were discovered in Drosophila (Bridges and Morgan, 1923), where they exist in two separate gene clusters (Fig. 1): the Antennapedia and Bithorax complexes (ANT-C and BX-C, respectively) (Kauffman et al., 1980; Lewis, 1978; Lewis et al., 1980). Early genetic experiments in adult flies demonstrated that Hox genes are involved in the allocation of distinct morphological identities to each body segment: mutations affecting specific Hox genes typically lead to homeotic transformations, in which the morphology of a given segment is transformed into the likeness of another (Bateson, 1894). Further conceptual work on BX-C mutations by Ed Lewis showed that the gene order inferred from BX-C genetic maps was directly related to the anatomical areas influenced by the individual Hox genes: mutations in genes located at one end of the complex affected anterior larval structures, whereas lesions on genes at the other end of the cluster affected posterior larval patterning (Lewis, 1978). This pioneering work suggested that Hox genes provide a genetic coordinate system for the allocation of developmental identities in the fly. However, understanding the mechanisms that link Hox genes to their developmental roles was only possible when the relevant genes were cloned, and their expression domains in the fly embryo, as well as their regulation, clarified. Indeed, molecular cloning of the BX-C (Bender et al., 1983) followed by expression analysis showed that Hox genes are expressed in particular subdomains along the anteroposterior axis of the embryo (Akam, 1987; Harding et al., 1985). These observations also revealed that the order in which individual Hox genes are expressed along the head-to-tail axis of the embryo mirrors the physical order of the Hox genes within the Hox cluster (Akam, 1987; Harding et al., 1985), a characteristic generally known as spatial collinearity, which had a strong impact in further investigations on the molecular regulation of Hox genes.

By the end of the 1980s, Hox genes had also been identified in the mouse genome. This remarkable observation suggested that they were likely to be present in most other vertebrates. Molecular analyses revealed that the basic genomic organisation and
expression patterns of mouse and Drosophila Hox genes showed common features (Duboule and Dollé, 1989; Graham et al., 1989). In particular, it was shown that mouse Hox genes are organised in clusters and that their expression also follows the spatial collinearity principle. On the basis of these similarities, it was suggested that clustering is an intrinsic property of Hox genes that is indispensable for proper regulation of Hox gene expression. However, the cloning of Hox genes from other bilaterians, which revealed a variety of cluster structures as well as cases with a complete absence of clustering, raised some questions about the role of Hox gene clustering (Lemons and McGinnis, 2006; Duboule, 2007).

With the probable exception of lampreys (Smith et al., 2013), vertebrates possess at least four Hox clusters [teleost fishes have seven or eight clusters (Hurley et al., 2005)], possibly as a result of successive duplications of an ancestral cluster (Hurley et al., 2005) (Fig. 1). Although the configuration of vertebrate Hox clusters has been interpreted as a paradigmatic form of Hox gene organisation, it might instead represent a rather exceptional case of organisation and compaction (Duboule, 2007) that reflects intrinsic regulatory features of vertebrate Hox genes that are not necessarily present in other organisms (see below).

**Hox transcriptional regulation in space and time**
A rigorous analysis of gene interactions and Hox expression patterns during early Drosophila development (Akam, 1987) suggested that Hox expression domains are likely to be determined by at least three distinct regulatory inputs: transcriptional regulation...
from earlier segmentation genes (Irish et al., 1989); a cellular memory system based on the action of Polycomb (PcG)/trithorax (trxG) group proteins (Denell, 1978; Puro and Nygren, 1975; Wedeen et al., 1986); and cross-regulatory interactions among the Hox genes themselves (Morata and Kerridge, 1982; Struhl, 1982). Regarding the latter mechanism, it was observed that the more posterior Hox genes are able to repress the expression (and suppress the function) of more anterior genes, a process termed ‘posterior prevalence’ (Akam, 1987; Duboule and Morata, 1994; Hafen et al., 1984; Harding et al., 1985). Nonetheless, within single *Drosophila* segments the expression of each Hox gene is substantially modulated according to position and cellular type. Furthermore, in some developing structures, *Drosophila* Hox genes are activated in domains that are not discernibly related to the metameric specification of the animal (Akam, 1987; Akam and Martinez-Arias, 1985; White and Wilcox, 1985). These observations in the fly suggest that the control of Hox gene expression might have further dimensions in addition to the three main components mentioned above.

In vertebrates, Hox gene expression includes an additional regulatory step. During initial stages of development, Hox genes are kept globally silent and become progressively activated during development following a temporal sequence that correlates with the gene’s position within the cluster in a 3’ to 5’ direction (Kmita and Duboule, 2003). This property, commonly known as ‘temporal collinearity’, has been observed in both the vertebrate primary body axis and developing limb buds (Kmita and Duboule, 2003). After initial activation, domains of vertebrate Hox gene expression are subsequently refined to produce the characteristic spatial distribution observed in the different tissues of older embryos (Deschamps and Wijgerde, 1993; Alexander et al., 2009; Deschamps and van Nes, 2005). This latter and belated regulatory phase seems to rely on the activity of a second tier of control elements that produce specific Hox expression patterns, which are subsequently refined to produce the characteristic spatial distribution observed in the different tissues of older embryos. Hox gene transcriptional patterns, thus furthering our understanding of their mechanisms of activity (Akbari et al., 2006; Alexander et al., 2009). These studies revealed that transcriptional cis-regulation represents a major determinant of Hox gene spatial expression, that cis-regulatory elements are functionally autonomous, and that the activity of cis-activity regulatory elements is independent of them being inserted within a Hox cluster (Whiting et al., 1991; Charité et al., 1995; Hérault et al., 1998; Püschel et al., 1991). Therefore, despite the remarkable correlation between Hox expression domains along the anterior-posterior (AP) body axis and the position of the gene within the cluster (Harding et al., 1985; Akam, 1987; Graham et al., 1989; Duboule and Dollé, 1989) (Fig. 1), the spatial control of Hox gene expression seems to be mostly independent of genomic arrangement. This is consistent with the finding that in animals in which Hox genes are not clustered, Hox expression still maintains

![Fig. 2. Representation of different aspects of Hox gene regulation associated with chromatin/nuclear characteristics.](image) In the undifferentiated state (left), the Hox clusters are associated with nucleosomes that contain inactivating (H3K27m3) marks and remain in a compacted configuration within a silent nuclear domain (yellow). In this chromosomal configuration, the cis-regulatory elements involved in spatial control of Hox gene expression are kept inactive. Activation of the Hox clusters is progressive, following a 3’ to 5’ sequence (represented in the middle and right-hand diagrams). This activation is associated with a change in nucleosome modifications, which now include activating (H3K4m3) marks. During activation, the relevant areas of the cluster enter a nuclear territory of active transcription (orange). This relocation of Hox chromatin leads to exposure of cis-regulatory elements that can now be regulated by relevant transcription factors leading to Hox gene expression. Adapted from Noordermeer et al. (Noordermeer et al., 2011).
spatially collinear gene expression (considering the expression of paralogue groups) (Seo et al., 2004). Furthermore, the structure of the Hox cluster differs among species within the Drosophila genus; for example the cluster in D. buzzatii has two distinct splits compared with that of D. melanogaster, yet the changes do not seem to lead to any significant differences in Hox gene expression (Negre et al., 2005) arguing that the integrity of the gene complex is not an absolute requirement for the establishment of Hox expression patterns.

**The chromatin component of Hox gene regulation**

The functional autonomy of Hox cis-regulatory elements suggests that their activity with regards to broad transcriptional patterns of gene activation is largely independent of chromosome structural features that involve chromatin-based regulation. However, this notion does not hold true when looking at the molecular mechanisms underlying early phases of vertebrate Hox expression or those ensuring maintenance of appropriate Hox expression domains.

In vertebrates, the early global repression and subsequent collinear activation of Hox gene expression are closely associated with the physical characteristics of chromatin at the Hox clusters (Kmita and Duboule, 2003; Tschopp and Duboule, 2011). Indeed, a number of genomic manipulations in the mouse Hoxd locus indicate that temporal collinear gene activation is linked to its clustered structure (Tschopp and Duboule, 2011), which could explain the more compact structure of Hox loci in vertebrate genomes (Duboule, 2007).

Hox gene expression within the cluster correlates with the distribution of histone modifications associated with inactive (trimethylation at Lys27 of histone H3, H3K27m3) and active (trimethylation at Lys4 of histone H3, H3K4m3) chromatin (Schuettengruber et al., 2007) throughout the Hox cluster. In embryonic stem cells (ESCs), which are thought to represent the earliest stages of mammalian development and do not express Hox genes, Hox gene clusters are fully decorated by H3K27m3 and contain very low levels of H3K4m3, which is typically associated with transcriptional start sites in the configuration known as ‘bivalent’ chromatin (Bernstein et al., 2006; Soshnikova and Duboule, 2009; Eskeland et al., 2010). When the Hoxd cluster was analysed in the tail tip of mouse embryos before embryonic day (E)8.5 and E10.5 (representing different stages of temporal activation), Hox gene expression was closely followed by the disappearance of H3K27m3 and by a strong increase in H3K4m3 (Soshnikova and Duboule, 2009). A similar change in the H3K27m3 pattern was observed in Hoxb and Hoxd clusters of mouse ESCs undergoing retinoic acid (RA)-induced differentiation (Eskeland et al., 2010). This indicates that the temporal component of Hox gene activation is associated with the progressive change in chromatin marks that take the system from a silent to an active configuration (Fig. 2).

The correlation between H3K27m3 marks and inactive Hox clusters (Soshnikova and Duboule, 2009; Eskeland et al., 2010) suggests involvement of PcG in Hox gene silencing, and genetic analyses in mice seem to support this hypothesis. PcG activity involves sequential action of the polycomb repressive complexes (PRC) 2 and 1. PRC2 first interacts with target DNA to introduce H3K27m3 marks into the chromatin, which are then read by PRC1 to promote silencing of the associated chromatin, perhaps by the incorporation of additional modifications to histones, most notably a ubiquitin moiety at lysine 119 of histone H2A (H2AK119) (Schuettengruber et al., 2007). ESCs with an inactive PRC2 fail to introduce the H3K27m3 mark into Hox cluster chromatin (Eskeland et al., 2010; Shen et al., 2008). Similarly, active PRC1 is required for cluster silencing; ESCs lacking RING1B, a E3 ligase that introduces ubiquitin into H2AK119 (de Napoles et al., 2004), fail to repress Hox gene expression, which in those cells reaches activation levels close to those obtained upon differentiation of normal ESCs (Eskeland et al., 2010). Interestingly, the finding that Hox genes are expressed in Ring1B (Ring2 – Mouse Genome Informatics) mutant cells indicates that the presence of H3K27m3 alone in the Hox clusters is not sufficient to inactivate Hox gene expression because in these cells H3K27m3 distribution along the Hoxb or Hoxd clusters is largely not affected (Eskeland et al., 2010).

It has long been established that PcG is essential for maintaining appropriate Hox gene expression domains by a repressive mechanism (Lewis, 1978). It is now recognised that PcG activity is also closely linked to modulation of specific chromatin states, which are usually not restricted to specific genes, but rather affect large chromosomal domains (Schwartz et al., 2006; Schuettengruber et al., 2007). Repression of Drosophila Hox gene expression is also associated with such a mechanism involving modulation of chromatin conformations. Indeed, in the fly embryo, repressed Hox genes are contained within a single PcG structural domain regardless of whether they reside within BX-C or ANT-C, indicating that the two Hox complexes belong to the same functional PcG domain (Bantignies et al., 2011; Tolhuis et al., 2011), despite being separated by ~10 Mb.

**Adding active and inactive chromatin-associated marks to Hox chromatin**

The mechanisms responsible for targeting of PcG activity to specific loci seem to be different in Drosophila and vertebrates. In the fly, PRC2 complexes associate with cis-regulatory regions known as polycomb responsive elements (PREs) (Müller and Kassis, 2006; Schwartz et al., 2006). Recruitment of PcG activity to the Drosophila Hox genes has been extensively characterised (Akbari et al., 2006). Work in the late 1990s showed that coordinated PcG repression of BX-C and ANT-C genes is regulated by Fab-7, an element previously known to control PcG activity in BX-C (Mihaly et al., 1997). The finding that Fab-7 deletion mutants also affect PcG activity on ANT-C suggests the existence of long-range interactions between the two Drosophila Hox complexes (Bantignies et al., 2011) (see below).

In vertebrates, the mechanisms of PcG targeting are largely unknown so very little can be said about PcG targeting of vertebrate Hox genes (Delest et al., 2012). A few studies identified some regions within vertebrate Hox clusters that could be involved in ‘seeding’ PcG activity (Kim et al., 2006; Woo et al., 2010). Of these, the best characterised is a phylogenetically conserved region between HoxD11 and HoxD12 that seems to be involved in homing PRC2 to the HOXD cluster in human stem cells (Woo et al., 2010). This region contains several of the characteristic features of Drosophila PREs and is able to promote repression of reporter genes when stably integrated in the genome (Woo et al., 2010). However, its role in the control of Hox gene expression is still unclear because its removal from the mouse genome had no evident deleterious effects on development or on the expression of Hoxd genes in mouse embryos (Beckers and Duboule, 1998; Tschopp et al., 2009). Therefore, the mechanisms that direct PcG to the Hox clusters to (1) produce global silencing during early development and (2) secure proper spatial Hox gene expression at later developmental stages remain unclear.
The mechanisms controlling the sequential 3’ to 5’ loss of H3K27m3 from the Hox clusters are also far from being understood. One possibility is that this process results from regulated activity of specific demethylases. Two H3K27m3 demethylases, Kdm6b (also known as Jmjd3) and Kdm6a (also known as Utx) (Hong et al., 2007; Xiang et al., 2007), have been shown to interact with Hox genes and to modulate H3K27m3 levels at their promoters (Agger et al., 2007; Lan et al., 2007). Phenotypic analysis of genetically inactive and/or knock-down Kdm6 genes indicate that, of the two demethylases, Kdm6a emerges as a more promising Hox regulator given that experiments that remove or decrease its activity lead to Hox-like phenotypes (Lan et al., 2007; Lee et al., 2012; Satoh et al., 2010).

Hox cluster activation not only correlates with reduced H3K27m3 but also with a broad increase in H3K4m3, which seems to precede H3K27m3 disappearance in the mouse tail tip (Soshnikova and Duboule, 2009). Therefore, the regulated addition of chromatin activation marks could be a key driver of progressive Hox cluster activation. TrxG is a leading candidate to play a role in such process because its activity adds H3K4m3 marks to chromatin (Schuettengruber et al., 2007). So far the role of trxG in Hox cluster opening has not been directly evaluated and the available data from different trxG mutants do not allow a clear assessment of its role during early phases of vertebrate Hox expression (Ayton et al., 2001; Glaser et al., 2006; Stoller et al., 2010; Yu et al., 1995; Yu et al., 1998). Interestingly, Kdm6a has been shown to establish functional interactions with the trxG member Mll2 (Km2d – Mouse Genome Informatics), which encodes for a methyl transferase that produces H3K4m3 (Issaeva et al., 2007), suggesting that sequential activation of Hox clusters could result from a coordinated action of H3K4m3-promoting and H3K27m3-removing activities.

The mechanisms controlling the early phases of vertebrate Hox gene activation are less well known than those regulating late Hox gene expression in limbs and other domains (Alexander et al., 2009; Tschopp and Duboule, 2011). Coordinated activation of Hoxd genes in limb buds depends on specific global control regions (Tschopp and Duboule, 2011), although the mechanistic details of their activity are mostly unknown. Recent analyses detected physical interactions between these control regions and Hoxd genes (Montavon et al., 2011). However, the role of these interactions in Hox gene activation is not clear because they were also observed in tissues that do not express Hoxd genes. In the main body axis, Hox gene activation seems to depend on the coordinated activity of several signalling pathways, including those activated by retinoic acid, fibroblast growth factors (FGFs), Wnts and growth differentiation factor 11 (Gdf11) (Iimura and Pourquié, 2007; Alexander et al., 2009). In addition, a variety of genetic analyses suggest that Cdx transcription factors could be a core integrator of these signals in the growing posterior embryo (Young and Deschamps, 2009). However, although the impact of these factors on Hox gene expression is clear, their relationship to chromatin changes associated with vertebrate Hox gene activation remains to be clarified.

### 3D chromatin conformation and Hox gene expression

Recent data show that PcG (and possibly trxG) function is connected with specific 3D chromatin configurations within the nucleus. Analysis of the subnuclear localisation of PRC components revealed that PcG activity is confined to distinct nuclear regions termed polycomb (PC) bodies (Buchenau et al., 1998). In addition, extensive chromatin contacts have been detected among PcG-regulated areas, even if these are separated by considerable distances (Delest et al., 2012). *Drosophila* Hox genes provide a good example of this regulatory mechanism, showing that all inactive Hox genes nucleate at the same PC body and display extensive mutual physical interactions regardless of whether they are in the BX-C or the ANT-C complex (Bantignies et al., 2011; Tolhuis et al., 2011). These observations reinforce the notion of coordinated regulation of the *Drosophila* Hox genes by PcG, and show that this regulation involves arrangement of their associated chromatin within the same nuclear domain.

Specific 3D chromatin configurations are also likely to be involved in Hox gene regulation during early vertebrate embryo development. Analyses of the Hox clusters in undifferentiated human embryonal carcinoma and mouse ESCs using chromosome conformation capture approaches, revealed strong chromatin interactions spanning the whole cluster (Ferraiuolo et al., 2010; Noordermeer et al., 2011), indicating that inactive vertebrate Hox clusters might also be packed together within a specific nuclear compartment. A similar conclusion emerged from fluorescent in situ hybridisation (FISH) studies showing that 3’ and 5’ areas of the inactive Hoxb and Hoxd clusters localise together within their respective chromosomal territory (i.e. the area of the interphase nucleus occupied by each chromosome) (Chambeyron and Bickmore, 2004; Morey et al., 2007). Interestingly, these chromosome interactions are essentially restricted to the Hox cluster itself, with very little extension into adjacent chromatin (Noordermeer et al., 2011), suggesting the existence of a Hox cluster-specific marking mechanism that controls chromosomal configuration.

Studies using mutant ESC lines support the involvement of PcG in Hox cluster compaction. In particular, in ESCs deficient for either the PRC2 component Eed or the PRC1 component Ring1B, the Hoxb and Hoxd clusters adopted an open configuration (Eskeland et al., 2010). It was shown that undifferentiated ESCs also contain PC bodies that remodel with differentiation (Ren et al., 2008). Whether silent Hox clusters are also contained within those PC bodies is as yet unclear.

Activation of Hox gene expression both in differentiating cells and in the developing tail and limb buds is associated with major structural changes in the Hox cluster chromatin (Chambeyron and Bickmore, 2004; Chambeyron et al., 2005; Eskeland et al., 2010; Ferraiuolo et al., 2010; Montavon et al., 2011; Morey et al., 2007; Noordermeer et al., 2011). In particular, inactive and active areas of the Hox cluster segregate into separate domains. Inactive areas keep their mutual interactions but lose contact with active regions of the same cluster (Ferraiuolo et al., 2010; Noordermeer et al., 2011; Montavon et al., 2011). Interestingly, interactions were still observed throughout H3K4m3-rich active areas of the Hox clusters (Noordermeer et al., 2011; Montavon et al., 2011). A similar type of interaction between the H3K4m3-containing active areas of the HOXA cluster was also observed in primary human fibroblasts and in a macrophage cell line (Guenther et al., 2005; Wang et al., 2011), indicating that such interactions are a general characteristic of active Hox clusters. Therefore, active Hox chromatin might be precisely organised within a defined nuclear domain that is physically and functionally different to the nuclear domain containing the inactive Hox cluster.

This view is consistent with FISH analyses showing that gene activation within the *Hoxb* and *Hoxd* clusters is associated with the relocation of their active domain into a position within the nucleus that is physically separated from that occupied by the silent part of the cluster (Chambeyron and Bickmore, 2004; Chambeyron et al., 2009).
It was recently reported that chromatin in the interphase nucleus is organised into distinct globular compartments, in which active and inactive chromatin are segregated from each other (Lieberman-Aiden et al., 2009). The 3D structural characteristics of Hox clusters at different activation stages are consistent with a model in which active and inactive Hox chromatin resides separately within such globular compartments and the activation process requires a transit between them. Because the activation sequence of vertebrate Hox genes follows the position of the gene in the cluster, it can be hypothesised that the movement of Hox cluster DNA between compartments is produced by linear sliding of the chromatin fibre in a 3′-to-5′ direction. The characteristic knot-free organisation of fibres within the globular compartments (Lieberman-Aiden et al., 2009) would facilitate the linear unwinding and rewinding of chromatin required for this type of mechanism. Such a model also fits with the alterations observed in the temporal activation of mouse Hox genes when their distance to the 3′ end of the cluster was modified either by deletions or duplications in the cluster (Tschopp et al., 2009).

Although 3D organisation might follow similar general principles for all four mammalian Hox clusters, it is possible that there are also cluster- and tissue-specific features. For instance, global interactions within the inactive Hox cluster seem to differ slightly from those observed in the other clusters (Noordermeer et al., 2011). In addition, whereas active regions of the Hoxb and Hoxd clusters move to a position outside their chromosomal territory within the nuclei of ESCs and tail bud cells (Chambeyron and Bickmore, 2004; Chambeyron et al., 2005; Morey et al., 2007), in the limb bud cell nuclei active Hox genes separate from the inactive areas of the cluster but stay within their chromosomal territory (Morey et al., 2007).

In conclusion, several aspects of Hox gene expression are strongly connected to higher order structural features of chromatin, suggesting that a better understanding of the structural and functional characteristics of nuclear compartments and the dynamics of chromatin transit among them is likely to impact current models of Hox gene regulation.

Long non-coding RNAs in the regulation of Hox gene expression

Recent studies suggest that long non-coding RNAs (lncRNAs) could interact with transcription factors and chromatin modifiers to modulate gene function during development (Lee, 2012). In this context, one pervasive role of lncRNAs is to act as a device that targets PcG and trxG activity to specific functional targets.

The Drosophila Hox genes have been reported to be among the targets of lncRNA-mediated gene regulation. A well-studied case is that of the lncRNA bithoraxoid (bdx) in the control of Ubx transcription. bxd is expressed from the genomic region upstream of Ubx and is transcribed in the same direction as Ubx. The exact functional relationship between bxd and Ubx remains unclear, however, as two groups have reported conflicting reports on this issue (Petruk et al., 2006; Sanchez-Elsner et al., 2006). Nonetheless, in both cases bxd seems to recruit trxG proteins to the cis elements regulating expression of these RNAs. In one of the cases, it was proposed that bxd-mediated trxG recruitment promotes its own transcription, eventually resulting in the silencing of Ubx (Petruk et al., 2006). In the other study, it was argued that bxd brings trxG proteins to the Ubx promoter, thus promoting Ubx transcription (Sanchez-Elsner et al., 2006).

The lncRNA iab-8 has also been implicated in the control of Hox gene expression in Drosophila. Posterior expression of abd-A in the Drosophila embryonic central nervous system (CNS) is repressed by the lncRNA iab-8, which is transcribed from the genomic region between Abd-B and abd-A and seems to repress abd-A transcription in cis by a transcriptional interference mechanism (Gummalla et al., 2012). These observations led to the suggestion that iab-8 might play a role in the process of posterior prevalence (Gummalla et al., 2012).

lncRNAs have also been implicated in the control of vertebrate Hox gene expression. One of them, HOTTIP, is transcribed from a region 5′ of HOXA13, in the opposite direction to the HOXA genes (Wang et al., 2011). HOTTIP activity in human cells seems to be linked to interactions with the trxG protein WDR5, which recruits other trxG members to introduce the H3K4m3 mark into the 59 area of the HOXA cluster to keep it transcriptionally active (Wang et al., 2011). Experiments performed in chicken embryos also suggest that HOTTIP is required to induce and/or maintain expression of the 5′ Hoxa cluster genes in the distal part of developing limb buds (Wang et al., 2011). A regulatory role has also been described for another non-coding antisense transcript spanning the Hoxb5 and Hoxb6 loci (Hoxb5/b6as) (Dinger et al., 2008). Hoxb5/b6as has also been reported to interact with the trxG regulatory network in differentiating mouse embryoid bodies, most notably with Mll1 and with H3K4m3, either to activate or maintain the transcription of Hoxb5 and Hoxb6.

HOTAIR is another lncRNA that has been proposed to regulate vertebrate Hox genes. In human cells, HOTAIR is transcribed from the region between HOXCI11 and HOXCI2 (Rinn et al., 2007). Intriguingly, unlike other more characterised lncRNAs, HOTAIR does not seem to regulate the expression of adjacent transcripts, but it rather appears to block expression of genes within the posterior HOXD cluster. HOTAIR seems to work by coordinately targeting PRC2 and the histone demethylase LSD1 (KDM1A – Human Gene Nomenclature Database) (Shi et al., 2004) to specific areas of the HOXD complex (Rinn et al., 2007; Tsai et al., 2010). However, HOTAIR sequence conservation between human and mouse is surprisingly low, and its removal from the mouse genome has no detectable negative effects on embryonic development or on the distribution of the H3K27m3 mark throughout the Hoxd cluster (Schorderet and Duboule, 2011), suggesting that HOTAIR represents a novel mechanism that emerged in the human lineage and/or that its roles are compensated by other factors.

In summary, experiments so far indicate that, in Drosophila, lncRNAs do play a role in the regulation of Hox expression and developmental functions in embryos (Gummalla et al., 2012; Petruk et al., 2006; Sanchez-Elsner et al., 2006). In vertebrates, although growing evidence supports the notion that lncRNAs control Hox gene expression, their biological roles are still debated.

Hox gene regulation via RNA processing

Although transcriptional and chromatin-mediated regulation play a key role in gene expression, a range of co-transcriptional and post-transcriptional processes are also able to affect mRNA quality and/or quantity and thus represent an additional layer of gene control in eukaryotic organisms (Licatalosi and Darnell, 2010). For instance, differential RNA processing via alternative splicing, alternative polyadenylation, and/or alternative promoter usage can lead to the formation of mRNA transcripts that carry substantially different information and/or structures (Licatalosi and Darnell, 2010; Nilsen and Graveley, 2010; Proudfoot, 2011). In line with this general regulatory potential, RNA processing does play an important role in the regulation of Hox gene expression.
One of the best examples of the impact of alternative splicing on Hox function concerns the Drosophila Ubx gene (Lewis, 1978; Sánchez-Herrero et al., 1985). Drosophila Ubx produces six distinct mRNA isoforms (Kornfeld et al., 1989; O’Connor et al., 1988) via a recursive splicing mechanism (Hatton et al., 1998) that releases a family of transcripts bearing common 5’ and 3’ exons but carrying distinctive small exonic sequences (microexons m1 and m2) and/or a 9-aa extension to the 3’exon (the b element) (Kornfeld et al., 1989; O’Connor et al., 1988). Notably, the same spectrum of Ubx isoforms found in D. melanogaster is present in distantly related Drosophila species, implying that the distinctive pattern of Ubx splicing has been maintained for over 60 million years of independent evolution (Bomze and López, 1994). Furthermore, each alternatively spliced Ubx mRNA isoform displays a characteristic developmental expression pattern in D. melanogaster and such patterns are also evolutionarily conserved (Bomze and López, 1994). This strongly suggests functional implications. Yet, despite initial reports in cell culture, which highlighted specific functions for Ubx isoforms (Gavis and Hogness, 1991; Krasnow et al., 1989), early genetic experiments in flies suggested that all isoforms were largely equivalent in their developmental roles (Busturia et al., 1989). Nonetheless, further studies focused on specific developmental, cellular and molecular roles under Ubx control have consistently and conclusively demonstrated that alternatively spliced Ubx isoforms perform distinct functions during the development of the embryo and adult fly. For example, heat-shock induction of Ubx alternatively spliced isoforms leads to isoform-specific transformations in the peripheral nervous system (PNS). In particular, ectopic expression of the Ubx splicing isoform Ia but not Ubx-I1a transforms the PNS of thoracic parasegments into the PNS of an abdominal-like segment (Mann and Hogness, 1990). Further genetic work on the PNS showed that indeed Ubx-I1a cannot substitute functionally for other isoforms to promote normal development of the PNS (Subramaniam et al., 1994). Recent studies using the Drosophila UAS/Gal4 system further demonstrated the impact of Ubx alternative splicing on the activation of Ubx target genes during embryogenesis (Reed et al., 2010) and on the ability of Ubx to control the morphology and underlying gene networks of adult appendage development (de Navas et al., 2011).

Notably, the molecular characterisation of five of the remaining seven Drosophila Hox genes, proboscipedia (pb), Antennapedia (Antp), labial (lab), abdominal A ( Abd-A) and Abdominal-B ( Abd-B), revealed that they also undergo differential RNA processing either in the form of alternative splicing, alternative promoter usage and/or alternative polyadenylation (Celniker et al., 1989; Celniker et al., 1990; Kuziora and McGinnis, 1988; Laughon et al., 1986; Mlodzik et al., 1988; O’Connor et al., 1988; Sánchez-Herrero and Crosby, 1988; Schneuwly et al., 1986; Scott et al., 1983; Stroecher et al., 1986). Furthermore, inspection of recent RNA-sequencing data suggests that Sex combs reduced (Scr) might also undergo a form of RNA processing via alternative promoter usage (P. Patraquim and C.R.A., unpublished). Altogether, these observations indicate that RNA processing affects the expression of most, perhaps even all, Drosophila Hox genes.

By contrast, much less is known about the extent to which RNA processing affects the expression of mammalian Hox genes. A number of studies provide strong evidence that several rodent Hox genes undergo RNA processing, mostly via alternative splicing. For example, Hoxa1 was shown to produce two alternatively spliced transcripts that encode two different protein isoforms (LaRosa and Gudas, 1988). Similarly, several transcripts were described for Hoxa9, originating from both alternative splicing and differential promoter usage (Fujimoto et al., 1998; Rubin and Nguyen-Huu, 1990). The different mRNA species of Hoxa9 encode different versions of the Hoxa9 protein, two of them containing the same homeodomain but differing slightly upstream of this motif (Rubin and Nguyen-Huu, 1990), and another encoding a homeodomain-less protein (Fujimoto et al., 1998). Transcripts encoding different protein versions can be detected in embryonic tissue (LaRosa and Gudas, 1988; Fujimoto et al., 1998), suggesting that they could be functionally relevant, but a direct evaluation of this possibility remains to be performed. Interestingly, alternatively spliced mRNAs are also produced from the Hoxa10 locus, and encode a predicted protein that is essentially the homeodomain without any significant N-terminal attached to it (Benson et al., 1995). The role of such a molecule is still unclear, as functional analysis in transgenic mice indicate that it is not able to activate the typical Hoxa10 patterning programme in the axial skeleton or interfere with it (Guerreiro et al., 2012). Hoxb3 and Hoxa5 also produce different mRNA isoforms through a combination of alternative splicing and alternative promoter usage; interestingly, the resulting isoforms are expressed in different domains during embryonic development (Chan et al., 2010; Coulombe et al., 2010; Sham et al., 1992), once again suggesting that the resulting isoforms might carry out specific functions. Finally, a recent bioinformatic study (P. Patraquim and C.R.A., unpublished) focused on mouse Hox genes revealed that approximately a third of them produce a spectrum of mRNA isoforms generated by different types of RNA processing, many of which are evolutionarily conserved between mouse and human. These findings strongly suggest that RNA processing represents a widespread and somewhat overlooked form of gene regulation affecting mammalian Hox genes.

**Hox gene regulation via miRNAs**

Studies in mouse and Drosophila show that regulation via miRNAs represents another tier in the complex molecular regulatory circuit controlling Hox gene expression. miRNAs are short (22-mer) non-coding RNAs that repress gene expression by binding to complementary target sequences in 3’UTRs of mRNAs leading to mRNA degradation and/or translational repression (Bartel, 2009; Lee and Shin, 2012). Several miRNAs have been shown to regulate Hox gene expression. Notably, most examples available to date show that Hox-regulating miRNAs are encoded within the Hox clusters. This genomic arrangement might provide an effective mechanism to guide the production of miRNAs in the ‘same cells and at the same time’ as their Hox mRNA targets (Alonso, 2012). In vertebrates, these miRNAs include miR-196, miR-10 and miR-615 (Fig. 1).

The mouse has three miR-196 gene paralogues, miR-196b, miR-196a-1 and miR-196a-2, which are situated within Hox clusters a, b and c, respectively (Fig. 1). Intriguingly, the mouse miR-196 genes are all situated between the Hox9 and Hox10 genes, a configuration that is also observed for the miR-196 genes present in the frog (Xenopus laevis), zebrafish (Danio rerio) and human genomes. Assuming that sufficient evolutionary time has elapsed to probe different chromosomal orders, the absence of change in miRNA genomic location might reflect that the chromosomal position of this miRNA is essential for cell survival and/or normal development. In the case of miR-10, the mouse has two genes encoding this miRNA, miR-10a and miR-10b, which are located in Hox clusters a and d, respectively. Unlike miR-196, which is specific to chordates (Mansfield and McGinn, 2012), miR-10
represents a much more ancient miRNA system (see below). Remarkably, miR-10 genes are always found in between the same Hox genes [i.e. paralogue groups (PGs) 4 and 5]. Finally, miR-615 is confined to Eutherians (non-monotreme, non-marsupial mammals), in which it is produced from an intron in Hoxc5.

Regarding the regulatory roles of these miRNAs, miR-196 was the first miRNA to be experimentally shown to regulate the expression of a Hox gene (Hoxb8) (Yekta et al., 2004). Downregulation of miR-196 in the chick (McGlinn et al., 2009) and zebrafish (He et al., 2011) leads to Hox protein de-repression and the generation of homeotic transformations, highlighting the importance of this miRNA. Similarly, downregulation of miR-10 in zebrafish led to overexpression of hoxb1a and hoxb3a (Woltering and Durston, 2008), further confirming the notion that Hox complex miRNAs are important factors that modulate Hox protein expression during vertebrate development. Although miR-615 has been implicated in several pathologies affecting human liver function (El Tayebi et al., 2012), its roles in Hox gene expression have not yet been investigated.

In Drosophila, at least four miRNA genes are located within the Hox clusters: miR-iab-4, miR-iab-8, miR-10 and miR-993. Of these, miR-iab-4 and miR-iab-8 were shown to target several Hox genes including Ubx, Antp, abd-A and Abd-B (Bender, 2008; Ronshaugen et al., 2005; Stark et al., 2008; Thomsen et al., 2010; Tyler et al., 2008). Curiously, miR-iab-4 and miR-iab-8 are both encoded by the opposite DNA strands of the same locus, which is located between the BX-C genes abd-A and Abd-B (Bender, 2008; Gummalla et al., 2012; Stark et al., 2008; Thomsen et al., 2010; Tyler et al., 2008). Despite the fact that the Drosophila miR-iab-4/8 system and the mouse miR-196 are located in equivalent positions relative to orthologue Hox genes in each species, the murine and fly miRNAs show no sequence similarity with each other, arguing that these miRNAs have evolved independently in the insect and mammalian lineages (Lemons et al., 2012). This observation is remarkable in that it might imply the requirement of a Hox-controlling miRNA system to be present within these particular genomic locations for Hox expression to be adequately regulated.

Ectopic overexpression of miR-iab-4 in the Drosophila haltere disc leads to a mild, but clear, haltere-to-wing transformation closely reminiscent of a Ubx-like phenotype and it also produces a reduction in Ubx protein levels, indicating that this miRNA can indeed regulate Ubx gene expression during appendage development (Ronshaugen et al., 2005). However, genetic removal of the *miR-iab-4/8* locus does not lead to any evident homeotic transformation in the haltere or elsewhere in the developing fly (Bender, 2008; Thomsen et al., 2010). These experiments raise doubts about the biological significance of Hox regulation by miR-iab4/8. However, expression of Ubx protein within the embryonic CNS is markedly increased in *miR-iab-4/8* mutants suggesting that miR-iab4/8 might exert a biologically relevant regulatory function specifically within this tissue (Bender, 2008; Thomsen et al., 2010).

A possible explanation for the absence of biological interactions between miR-iab-4/8 and Ubx during early development emerged from the observation that several Drosophila Hox genes undergo developmentally regulated alternative polyadenylation so that longer 3′UTR isoforms (i.e. those containing more miRNA target sites) are only expressed within the CNS (Thomsen et al., 2010). This makes it likely that such longer forms are specifically subjected to miR-iab-4/8 repression, whereas shorter isoforms (those expressed at earlier developmental stages) might be immune to miR-iab-4/8 downregulation or segregated to distinct cellular expression domains (Thomsen et al., 2010) (Fig. 3). Notably, alternative polyadenylation of Hox transcripts is not an exclusive feature of *D. melanogaster*. A computational study of different *Drosophila* species (Patraquim et al., 2011) revealed that alternative Hox transcript polyadenylation patterns are maintained across *Drosophila* lineages that have evolved independently from each other for ~60 million years, implying that this is an ancestral feature of the *Drosophila* group with functional relevance in modern fruitflies.

The evolution of the *miR-10* gene is markedly different from that of the *miR-iab-4/8* locus. A very high level of sequence similarity between fly and mouse *miR-10* genes strongly suggests that this miRNA system was a feature of the ancient common ancestor between insects and mammals (Lemons and McGinnis, 2006). This substantial evolutionary conservation strongly suggests an important functional role for *miR-10* across the bilaterians. These considerations are, however, at odds with recent work in *Drosophila* embryos (Lemons et al., 2012), which showed that the two functional miRNAs derived from *miR-10* (i.e. *miR-10-5p* and *miR-10-3p*), predicted to target *Sex combs reduced* (*Scr*) and *Abdominal-b* (*Abd-B*) respectively, have no major effects on the expression of these Hox genes. Indeed, this study also failed to detect the interaction of miR-iab-4-5p with one of its Hox targets, *Antp*, at least during embryogenesis (Lemons et al., 2012). This suggests that whatever the biological roles of *miR-10* are during Hox expression, they might only be detectable in specific developmental contexts or come about as a result of combinatorial interactions with other RNA regulators or when coupled with selective RNA processing patterns of their targets.

Thus, it is still early days with regards to understanding the underlying mechanisms and biological roles of Hox regulation by miRNAs. Furthermore, although our discussion has been focused on Hox-embedded miRNAs, many other miRNA systems encoded outside the Hox clusters could potentially play key roles in the regulation of Hox gene expression.

**Translational control**

Examples of translational control of Hox protein production exist but are, to date, scarce. Yet this might not necessarily reflect the lack of relevance of this type of regulation but instead the absence of a systematic search for such translational regulatory effects on Hox gene expression.

The *Drosophila* genes *Antp* and *Ubx* for example, have been suggested to be under translational control during embryonic development (Oh et al., 1992; Ye et al., 1997) (Fig. 3). In particular, it has been shown that a subgroup of transcripts produced from the *Antp* and *Ubx* loci contain functional internal ribosomal entry sites (IRES) that allow their translation using a ‘cap’-independent mechanism (Oh et al., 1992; Ye et al., 1997) (Fig. 3). The IRES elements of both gene transcripts were reported to be functional in the *Drosophila* embryo and their activity to be developmentally regulated (Ye et al., 1997), providing patterns of activity that match the known expression domains of the endogenous Antp and Ubx proteins (Ye et al., 1997). This suggests that IRES-mediated translational regulation could indeed be involved in controlling the spatial and/or temporal production of the Antp and Ubx proteins during *Drosophila* development. Interestingly, IRES-mediated regulation of Antp must be coordinated with other regulatory processes at the transcriptional and splicing levels because production of the IRES-containing miRNA requires transcription from the alternative P2 promoter and a specific splicing scheme (Oh et al., 1992) (Fig. 3).
Translational regulation of Hox expression has also been documented in the mouse Hoxb4 a decade ago (Brend et al., 2003). Within the murine developing spinal cord, Hoxb4 transcripts are localised in broader domains than those of the corresponding protein. Interestingly, this regulation is tissue specific because in the paraxial mesoderm both Hoxb4 mRNA and protein were detected in similar areas. However, the mechanistic details and functional significance of this form of Hoxb4 regulation are still unknown.

More recently, cloning of the mutant allele of three mouse mutant strains, known as Tail short (Ts), Tail-short Shionogi (Tss) and Rabo torcido (Rbt), led to the identification of a new type of translational regulation that might impact the production of several Hox proteins (Kondrashov et al., 2011). The axial skeletons of these mutants presented a variety of homeotic transformations reminiscent of those associated with inactivating mutations of several Hox genes. All three strains are heterozygous for an inactivating mutation in the ribosomal protein Rpl38, an integral component of the 60S ribosomal subunit. The Rpl38 deficiency resulted in inefficient translation of a specific set of transcripts, without apparent negative effects on global translation. Transcripts for eight different Hox genes affected by the Rpl38 deficiency, supporting the hypothesis that the Rpl38 deficiency is a regulatory mechanism. If the latter case is true, it suggests that the molecular phenotype of Ts mice represents a hypomorphic phenotype with no relevance to physiological regulatory processes or if it has uncovered a new bona fide regulatory mechanism. The former case is true, it suggests that the molecular phenotype of Ts mice is vertebrate specific (or even mouse specific) or if they represent an evolutionarily conserved control mechanism.

Fig. 3. The effects of RNA processing on Hox gene expression. Variations in Hox RNA processing lead to the production of Hox mRNAs of different sizes that bear different sequences and secondary structure. The use of alternative promoters (left) leads to the production of mRNAs with distinct 5'UTRs, which in turn lead to differential translational efficiency. Co-transcriptional alternative splicing reactions (middle) affect the composition of different CDS, thus giving rise to different protein isoforms that can perform specific functions during development. The use of alternative polyadenylation signals (right) determines the release of mRNAs bearing different 3'UTR tails that contain distinct sets of miRNA target sequences, thus leading to differential regulation by miRNAs. CDS, coding DNA sequence; PAS, polyadenylation site.
Concluding remarks

Our discussion here reveals a remarkably wide spectrum of regulatory mechanisms involved in the control of Hox expression, many of which have been evolutionarily conserved between insects and mammals (Table 1). However, it must be noted that the presence of similar regulatory mechanisms in flies and mice does not necessarily imply a common origin and could instead be the result of convergent evolution. Although each Hox-regulating mechanism is likely to have its own intrinsic properties, it is important to note that many of the mechanisms presented seem to be linked to one another either by the regulatory input imparted by common factors or other forms of functional coupling.

Examples in *Drosophila* show, for instance, that the process of alternative promoter usage affecting the expression of mRNAs (Oh et al., 1992) dictate the type of 5'UTRs included in the mRNA, which, in turn, could affect protein translation. Definition of Hox mRNA 'ends' via the coupled reactions of cleavage, polyadenylation and transcriptional termination determines the nature of 3'UTR information on Hox transcripts, which will have direct impact on the set of regulatory interactions between Hox mRNAs and RNA regulators, such as miRNAs and RNA-binding proteins, involved in the control of mRNA degradation and translation efficiency (Thomsen et al., 2010; Alonso, 2012). Further examples of gene expression coupling include the effects of transcriptional elongation rates on the spectrum of alternatively spliced forms of *Ubx* (de la Mata et al., 2003) and the consequences of nonsense mutations on *Ubx* mRNA stability and alternative splicing (Alonso and Akam, 2003).

In vertebrates, Hox cluster miRNAs seem to follow a temporal expression pattern highly similar to that of their targets, leading to the suggestion that miRNAs could be an integral part of the mechanisms behind a long known, but still poorly understood, property of Hox genes: posterior prevalence (Yekta et al., 2008). miRNAs synthesised within the Hox clusters have also been suggested to play a role in posterior prevalence in *Drosophila* (Gummalla et al., 2012), thus indicating that similar mechanisms could be operating in vertebrates and flies. Interestingly, it has been shown that the role of miRNAs in posterior prevalence can be reinforced by a lncRNA transcribed from the same genomic area (Gummalla et al., 2012). Therefore, the coordinated evolution of the different regulatory processes might be a core feature underlying the robustness of the Hox expression programme.

In addition, we have discussed the impact of 3D chromatin conformation on several aspects of Hox gene regulation. Although it is still too early to draw a clear picture of this process, it is plausible that initial stages of vertebrate Hox gene activation involve a sequential relocation of Hox chromatin from transcriptionally silent nuclear domains into nuclear areas engaged in active transcription. Also, although it now seems clear that both Hox chromatin movements within the nucleus and Hox gene silencing/activation are connected to PcG and trxG activities, it will be important to understand how these processes are mechanistically linked together. In this context, it is conceivable that recently discovered regulatory players, including non-coding RNAs, such as lncRNAs or enhancer-associated transcripts, might influence chromatin structure and cluster-wide control mechanisms.

Yet these functional links across the many processes involved in gene expression are not a unique feature of the Hox genes. Indeed, more than a decade ago Tom Maniatis and Robin Reed suggested the existence of a pervasive level of functional coupling across all the molecular machines involved in gene expression control (Maniatis and Reed, 2002). However, most work addressing the mechanisms and biological implications of gene expression coupling has so far been conducted in mammalian cultured cells or yeast leaving the question of the extent to which these interconnections play a relevant role within the physiological context of development largely unexplored. In this context, Hox genes might provide an excellent system in which to investigate both the mechanisms and biological roles of the many interconnections across different gene regulatory levels that are traditionally studied in isolation.

A possible scenario with regards to the biological roles of a highly interconnected network of Hox-regulatory interactions is that this has emerged to increase robustness of the Hox expression programme, ensuring that correct spatial and temporal patterns of Hox expression are achieved despite intrinsic molecular variation and extrinsic
fluctuations in embryonic environment. More generally, the existence of this highly interconnected form of control might suggest that animal embryos must carefully control Hox gene expression in order to complete development successfully. This interlocked regulatory network arrangement might also provide an explanation for why core features of Hox expression, especially those related to expression domains, have been observed to be largely unchanged during the extended periods of bilaterian evolution.

Acknowledgements

C.R.A. wishes to thank Michael Akam, as well as past and present members of the C.R.A. lab for inspirational discussions about the workings of Hox genes.

Competing interests statement

The authors declare no competing financial interests.

References


