Hox gene clusters are one of the most ancient and highly conserved multigene loci in the animal kingdom (1, 2). Tandem duplication and unequal crossing over in an ancestral organism created a cluster of Hox genes, which then underwent further duplication and divergence from the common ancestral cluster (3). Functional studies in a wide range of invertebrate and vertebrate species have underscored the conserved roles of the Hox family of transcription factors as central players in the regulation of axial patterning during elaboration of the basic body plan in the evolution of animals (2, 4–7). The series of genome duplications associated with vertebrate evolution have generated multiple Hox complexes and sets of paralogous genes within a species. This creates a situation whereby a subset of Hox genes or clusters may fulfill ancestral functions in axial patterning, whereas the others are available to evolve new roles or activities and become coupled to regulation of different developmental processes (1, 3, 8).

Many fundamental properties of the organization, regulation, and function of Hox gene clusters, such as colinearity, posterior prevalence, response to major signaling pathways, auto-, para-, and cross-regulation, and long-range or global regulation, appear to be common features of the regulatory landscape of Hox clusters among widely diverse species (2, 9–11). Studies on the expression, regulation, and evolutionary origins of Hox gene clusters have primarily focused on the protein-coding regions (12). However, recent advances in genomics have unearthed a treasure trove of transcribed sense and antisense noncoding sequences embedded within the vertebrate Hox clusters and their flanking regions (for example, Hotair, Hottip, Hobbit, Halr1, Hotdag, mir10, and mir196), providing insight into a new layer of regulatory inputs for temporally and spatially restricted patterns of Hox expression. Ironically, Lewis, in his original analysis, postulated that many cis-regulatory regions in the Drosophila bithorax complex were regulatory RNAs (7). This has proven to be correct, but for a long time, this concept received little attention from the community (13–15). This raises the intriguing question of whether noncoding transcripts are also common features of the ancestral Hox clusters or newly evolved properties of complex vertebrate genomes. Here, we will focus on the current state of knowledge of noncoding transcripts associated with mammalian Hox clusters, their regulation, and putative functions as a basis for thinking about their implications in development, disease, and evolution.

Extensive transcription of long intergenic noncoding transcripts (lincRNAs) is a characteristic of many multigene loci, including globin, immunoglobulin, and Hox gene clusters (16–19). These intergenic transcripts are implicated in activation and repression through opening large chromatin domains, maintenance of active chromatin state, or RNA interference–mediated silencing processes, as shown for the globin gene cluster (16, 18, 20). These lincRNAs can affect gene regulation through both cis and trans mechanisms on Hox and non-Hox genes (21–25). Mammalian and invertebrate Hox clusters show extensive transcriptional activity from both strands of coding and noncoding regions during development (19, 26–31). The functional significance of such noncoding transcription is beginning to emerge as evidence from several groups suggests the important roles of noncoding transcription in the regulation of Hox clusters (19, 26, 32–34) (Fig. 1). Intergenic transcripts are often associated with active Hox genes. Analysis of human Hox clusters identified 15 antisense transcribed regions that represent 38% of spliced transcripts from these clusters (38.46% for HOXA, 33.11% for HOXB, 13.16% for HOXC, and 34.84% for HOXD) (31). Figure 2 illustrates extensive syntenic or positional conservation of many noncoding transcripts, including Mirs between human and mouse Hox clusters. This suggests that there may be common functional roles for these transcripts.

There appear to be more noncoding transcripts both within and flanking the HoxA cluster relative to other clusters (Fig. 2). Positioned 50 kb 3′ of the HoxA cluster, in the intergenic region between Halr1 and Skap2, is a ~16-kb region (Heater) that gives rise to a large number of spliced and unspliced polyadenylated transcripts originating from both strands (Halr1 and Halr1os1) (30, 34, 35). These transcripts have multiple isoforms and epigenetic marks (H3K4Me3 and H3K27Me3), and occupancy of Pol II (RNA polymerase II) indicates that they arise from at least four different start sites. From a regulatory perspective, mouse halr1, halr1os1, and their isoforms are among the most rapidly induced transcripts upon retinoic acid (RA) treatment of embryonic stem (ES) cells and also respond to RA in developing embryos (30).

The Heater region may be important for potentiating the response of Hoxa1 to retinoids (Fig. 3A) because knockdown of three Halr1 isoforms leads to increased levels of Hoxa1 in uninduced ES cells (34). Halr1 interacts with PURB (purine-rich element binding protein B), a single-stranded DNA/RNA binding protein involved in transcriptional regulation (36), and knockdown of PURB leads to increased expression of Hoxa1. There appears to be a strict threshold on the number of Halr1 molecules per cell (<10 transcripts per cell) (34), and RA treatment of ES cells alters this relationship by increasing both the levels and composition of Halr1 and Halr1os1 transcripts through addition of new isoforms (30). These changes to the transcriptional repertoire have the potential to decouple interactions between Halr1 and PURB, alluring regulatory input on Hoxa1 (Fig. 3A). Retinoids appear to directly induce the Heater lincRNAs because two flanking regions (H-ARI and H-AR2)

1Stowers Institute for Medical Research, Kansas City, MO 64110, USA. *Department of Anatomy and Cell Biology, Kansas University Medical Center, Kansas City, KS 66160, USA.

*Corresponding author. E-mail: rek@stowers.org
contain multiple retinoic acid response elements (RAREs) that display dynamic occupancy of retinoic acid receptors [RARs and RXRs (retinoid X receptors)] (30). This suggests a model for how a key signaling pathway in development (retinoids) may regulate noncoding transcripts that, in turn, affect the expression of the adjacent Hoxa1 gene.

Within the human HOXA cluster, at the 3′ end between HOXA1 and HOXA2, is HOTAIR1, a lincRNA from the noncoding strand initially identified in association with myelopoiesis in humans (28). In mice, along with Hotair1, a new isoform and a novel transcript, Hotaira1, has been mapped to this region, and they display dynamic expression patterns during development (30). Intriguingly, these lincRNAs are also rapidly induced by RA in human myeloid lineages and during mouse ES cell differentiation and embryonic development (28, 30). This reveals conservation in both their syntenic position and regulatory response to retinoids. HOTAIR1 feedbacks into RA induced changes in gene expression because its knockdown in NB4 cells results in alterations related to RA-induced growth arrest at G1 and granulocytic maturation (37). This further highlights regulatory interactions between retinoids and multiple Hox lincRNAs.

Functionally, HOTAIR1 modulates gene expression in both cis and trans (Fig. 1). Reducing the levels of HOTAIR1 results in a loss of gene expression of 3′ HOXA cluster genes (cis) and alterations in β2-integrin signaling through CD11b and CD18 and in integrin switch mechanism involving CD11c and CD49d (trans) (28, 37). This implicates HOTAIR1 in cell cycle regulation through moderation of G1/S transition.

Further 5′ in the HoxA cluster, HoxA-AS2 is a lincRNA with several isoforms expressed from the noncoding strand between Hoxa3 and Hoxa4 (38). HoxA-AS2 is induced by RA, IFN-γ (interferon-γ), and TNF-α (tumor necrosis factor-α) and is functionally linked with the repression of apoptosis through modulation of the caspase 8 and 9 pathways (38). Overlapping with the promoter region of Hoxa11 is a conserved antisense RNA, HOX11AS (human) and Hoxa11os (mouse), which shows mutually exclusive expression throughout development with Hoxa11 (39, 40). This is illustrated by the expression of human HOX11AS during the menstrual cycle, which peaks at midproliferative stage in an inverse relationship to Hoxa11 expression (41). Mechanistically, ectopic expression of Hoxa11os does not down-regulate endogenous
**Distribution of HOX genes**

- **HOXA**:
  - Human: Hoxa11, Hoxb11, Hoxa12, Hoxa9, Hoxa13
  - Mouse: Hoxa11, Hoxb11, Hoxa12, Hoxa9, Hoxa13

- **HOXB**:
  - Human: Hoxb5, Hoxb4, Hoxb3, Hoxb2, Hoxb1
  - Mouse: Hoxb5, Hoxb4, Hoxb3, Hoxb2, Hoxb1

- **HOXC**:
  - Human: Hoxc9, Hoxc8, Hoxc7, Hoxc6, Hoxc5
  - Mouse: Hoxc9, Hoxc8, Hoxc7, Hoxc6, Hoxc5

- **HOXD**:
  - Human: Hoxd3, Hoxd10, Hoxd9, Hoxd8, Hoxd7
  - Mouse: Hoxd3, Hoxd10, Hoxd9, Hoxd8, Hoxd7

**Hoxa11 expression in murine uterus**, which appears to rule out potential degradation of Hoxa11 by sense-antisense pairing and raises the possibility of modulation via promoter interference (41). In addition to the examples above, there are an extensive series of lincRNAs embedded in and spread throughout the HoxA cluster in human and mouse that display varying degrees of syntenic conservation, but their functional significance is yet to be explored (Fig. 2).

The 5’ end of the HoxA cluster is also marked by two lincRNAs, HIT18844 and HOTTIP. HIT18844 contains a highly conserved 265-bp block in vertebrates that maps 1.8 kb upstream of Hoxa13 gene (29). HOTTIP is expressed from the noncoding strand 330 bp upstream of HOX A13, and the HOTTIP region displays both H3K4me3 and H3K27me3 epigenetic marks (bivalent) that change upon activation (22). Depletion of Hottip leads to shortening and bending of distal bony elements in the limb similar to the loss of Hoxa11 expression and regulation of noncoding transcripts originating from the four Hox clusters. Each Hox cluster is scaled on the basis of human coordinates with the gene name listed below. Relative positions of mouse (green) and human (red) non-coding transcripts are shown on the basis of human coding genes as landmark. Arrows indicate the direction of transcription.

- **Hoxa11** expression in murine uterus, which appears to rule out potential degradation of Hoxa11 by sense-antisense pairing and raises the possibility of modulation via promoter interference (41).
- In addition to the examples above, there are an extensive series of lincRNAs embedded in and spread throughout the HoxA cluster in human and mouse that display varying degrees of syntenic conservation, but their functional significance is yet to be explored (Fig. 2).
- The 5’ end of the HoxA cluster is also marked by two lincRNAs, HIT18844 and HOTTIP. HIT18844 contains a highly conserved 265-bp block in vertebrates that maps 1.8 kb upstream of Hoxa13 gene (29). HOTTIP is expressed from the noncoding strand 330 bp upstream of HOX A13, and the HOTTIP region displays both H3K4me3 and H3K27me3 epigenetic marks (bivalent) that change upon activation (22). Depletion of Hottip leads to shortening and bending of distal bony elements in the limb similar to the loss of Hoxa11 expression and regulation of noncoding transcripts originating from the four Hox clusters. Each Hox cluster is scaled on the basis of human coordinates with the gene name listed below. Relative positions of mouse (green) and human (red) non-coding transcripts are shown on the basis of human coding genes as landmark. Arrows indicate the direction of transcription.

**Hottip** has other functional roles outside of input into Hox regulation. In combination with microRNA mir-101, it regulates cartilage development through modulation of integrin-a1 by means of DNMT-3B-mediated epigenetic regulation (42).

**Many of these features on expression and regulation of noncoding RNAs are also observed to a lesser degree in other clusters, HoxD-As1 (Haglr), a noncoding RNA from the intergenic region of Hoxd1 and Hoxd3, is implicated in the regulation of RA-induced differentiation and activation by the PI3K (phosphatidylinositol 3-kinase)/AKT pathway. This RNA is implicated in metastasis through regulation of genes associated with angiogenesis and inflammation, on the basis of knockout analyses in SH-SY5Y cells (43).** An interesting feature of a lincRNA (Hobbit1) from the HoxB cluster transcribed from the sense strand between Hoxb4 and Hoxb5 is that it shares cis-regulatory elements with the adjacent coding genes (Figs. 2 and 3B). The unspliced and polyadenylated Hobbit1 transcript is expressed in developing embryos and rapidly induced during RA-mediated differentiation of murine ES cells (30). Consistent
with the kinetics of induction, there is a rapid gain of the H3K4Me3 mark, associated with gene activation. In mouse embryos, Hobbit1 expression is dependent on an RARE that plays a role in the regulation of multiple Hox genes (Fig. 3B) (30, 44). This opens the possibility that many noncoding RNAs embedded in the Hox clusters may share common regulatory components with the protein-coding genes.

LincRNAs may also be associated with long-range regulation and sharing through physical interactions. For example, a pair of noncoding transcripts, Hotdog (HoxD telomeric desert lincRNAs) and Tog (twin Hotdog), arise from the gene dessert downstream of the HoxD cluster. Hotdog and tog are transcribed from the noncoding strand and display restricted expression in developing cecum (45). Furthermore, their transcription start sites show enrichment of H3K4me3 and Pol II and display strong physical interactions with active HoxD genes (Hoxd4 or Hoxd11) in the cecum. Hotdog and Tog expression levels are completely abolished by deletion of the region from Hoxd9 to Hoxd11. Disruption of contact between Hotdog/Tog and HoxD genes by chromosomal inversion leads to complete loss of HoxD expression. This suggests a model of long-range enhancer sharing between lincRNAs (Hotdog and Tog) and HoxD genes (Fig. 3C) (45).

One of the most studied lincRNAs is HOTAIR (HOX transcript antisense intergenic RNA), which is a spliced, antisense, and polyadenylated transcript generated from the intergenic region between HOXC11 and HOXC12 (19). HOTAIR serves as a scaffold for interaction with PRC2 (Polycomb repressive complex 2) and LSD1 (lysine-specific demethylase 1A) complex. The interaction between HOTAIR and PRC2 modulates enzymatic activity, which is mediated by interplay between EZH2 (enhancer of zeste homolog 2), EED (embryonic ectoderm
development), and JARID2 (jumonji, AT-rich interactive domain 2). Thus, HOTAIR works in trans to play roles in development and disease by localizing the PRC2 complex on its genome-wide targets, which include posterior HoxD genes and WIF-1 (Wnt inhibitory factor-1) (46–49). Quantitative proteomic analysis following knockdown of HOTAIR in HeLa cells reveals differential expression of a large number of proteins (~170) involved in diverse cellular processes, including the dynamics of the cytoskeleton and mitochondrial structure and function (50). Overexpression of HOTAIR is a hallmark of many human cancers and is linked to aspects of carcinogenesis, including metastasis, epithelial-to-mesenchymal transition, invasion, aggression, and apoptosis (46). In mouse development, Hotair is expressed in limb buds and the posterior trunk, in an area corresponding to the future lumbosacral vertebra (47, 48). Deletion of Hotair in mice leads to vertebral transformations and abnormal development of metacarpal and carpels, including deletion and/or fusion of digit elements. These phenotypes are attributed to anterior expansion of Hoxd10 and Hoxd11 in the trunk and ectopic expression of the imprinted gene Dlk1. In addition, there is a large-scale derepression of target genes, further confirming its role as part of a repressor complex (48).

It is challenging to establish orthologous relationships between vertebrate lincRNAs because, in light of their noncoding nature, they display varying degrees of sequence conservation. HOTAIR1M, HoxA11as, and HOTAIR display some level of sequence similarity in mammals (51). For example, exon 1 of HOTAIR1M is conserved and displays similar expression profiles across all mammals. HoxA11as is highly conserved in eutherian mammals but shows less conservation in marsupials, suggesting that it arose after the eutherian-marsupial divide. In the case of HOTAIR, there are larger transcripts and different exon/intron organizations in human compared to other species. For example, the two exons of mouse Hotair match with exon 4 and exon 6 of the human transcript (47, 51). Exon 4 is highly conserved in all mammals, whereas exon 6 is conserved in eutherian mammals, but in marsupials there is a reduced degree of conservation.

In addition to the lincRNAs, the human and mouse Hox clusters are embedded with multiple microRNAs, including Mir196a/b, Mir10a/b, Mir615, and MIR3185. With the exception of MIR3185, these microRNAs are conserved between mouse and human and seen in syntenic regions (Fig. 2). Mir10 family members are positioned between the group 4 and 5 paralogous genes in the HoxB and HoxD clusters (Fig. 2). Their expression correlates with the adjacent Hox genes, including direction of transcription and response to RA and ethanol (52–56). As illustrated with Mir10a (Fig. 3B), this may reflect the role of shared regulatory elements in potentiating expression of nearby Hox genes, lincRNAs (Hobbit1), and Mirs. The Mir10 family of microRNAs regulate Hox genes in both cis and trans (Fig. 1) (56–59), and major signaling pathways (Wnt, Fgfs, and Notch) are key non-Hox targets of the Mir10 family (60). Because these signaling pathways are also targets of Hox genes, Mir10 RNAs have feedforward regulatory inputs into Hox gene regulatory networks.

Three Mir196 family members, Mir196a, Mir196b, and Mir196a-2, are present in the mouse and human HoxA, HoxB, and HoxC clusters (Fig. 2). These Mir196 paralogs also regulate the expression of Hox genes in cis and trans and play important roles in the patterning of mid-thoracic skeletal element through modulating Hox genes and Wnt signaling (61). Mir196 paralogs directly regulate the expression of Rarfβ, which, in turn, affects axial patterning (62). Transcriptome analyses suggest that MIRs in Hox clusters are functionally involved in the progression, metastasis, and prognosis of various diseases, including cancers. Mir10b, Mir196a, MIR196b, and MIR615 are up-regulated in Huntington’s disease (63), whereas MIR196a/b is up-regulated and MIR10 is down-regulated in head and neck cancer (64). Current evidence implies that integration of MIRs in Hox clusters provides another layer of regulation for Hox genes and their targets, in fine-tuning developmental processes under Hox control.

From an evolutionary perspective, microRNAs provide insight into exploration conservation of noncoding transcripts in Hox clusters. The Mir10 family is located near Hox paralogous group 4 in vertebrates and adjacent to Antennapedia in Drosophila (65, 66). Mir10s appear to be the most ancient microRNAs because they are present in the common ancestor of eumetazoa (67, 68). Mir10 in Nematostella vectensis indicates origins predating the cnidarian-bilaterian split (69–71). Mir10s are not only seen in conserved syntenic regions but also display sequence conservation among bilaterians. The loss of Mir10 family is linked to disintegration of anterior Hox genes as in the case of nematodes and tunicates (72). The presence of Mir196 paralogs in jawless lamprey but not in nonvertebrate chordates suggests their origin at the base of vertebrate evolution before initial cluster duplication.

With respect to evolutionary conservation of lincRNAs, in a manner analogous to the vertebrate Hox clusters, large numbers of intergenic noncoding transcripts (iabs) arise from Drosophila HOM-C and exhibit spatial colinearity in expression and function (7, 13–15). Their domains of expression are normally delimited by insulator elements, and if transcription proceeds through these elements, there is a loss of insulator function and associated segmental transformations (73–76). Sustained expression of these iabs also alters Polycomb-mediated repression and serves to maintain active chromatin states in the bithorax complex. Hence, they are linked with modulation of segment-specific expression of genes in the bithorax complex (74). In red flour beetle, Tribolium castaneum, the Hox cluster has three noncoding transcripts, one located in the first intron of Utx/Tc-Ubx and two positioned between ptr/Tc-Antp and Utx/Tc-Ubx (77). In hemichordates, in addition to the syntenically conserved position of Mir10, their single Hox cluster has two sense and antisense noncoding transcripts (78). Because some of the lincRNAs and Mirs in vertebrates show syntonic conservation in insects and hemichordates, noncoding transcripts within Hox clusters may be an ancestral feature with functional relevance. It will be interesting to see how intermingled the functional and regulatory relationships between Hox coding genes and lincRNAs are as we learn more about their roles in evolution, development, and disease. They may well be sides of the same coin in Hox-associated gene regulatory networks.

REFERENCES AND NOTES


HOXs and lincRNAs: Two sides of the same coin
Bony De Kumar and Robb Krumlauf

Sci Adv 2 (1), e1501402.
DOI: 10.1126/sciadv.1501402

ARTICLE TOOLS
http://advances.sciencemag.org/content/2/1/e1501402

REFERENCES
This article cites 77 articles, 21 of which you can access for free
http://advances.sciencemag.org/content/2/1/e1501402#BIBL

PERMISSIONS
http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service