Trends in Developmental Biology
Vol. 8, 2014

Review

Hox gene regulation in vertebrates

João Leite-Castro[1], Pedro Nuno Rodrigues[1,2] and Renata Freitas[1,*]

[1]IBMC - Instituto de Biologia Celular e Molecular, 4150-180 Porto, Portugal;
[2]ICBAS - Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Portugal.

ABSTRACT

The formation of the body plan of bilaterians involves the determination of cell identity fields that ultimately result in the formation of specific anatomical structures. *Hox* genes have long been suggested to play a critical role in this process, encoding transcription factors strongly active during development that condition cellular differentiation. Their function seems to be tightly associated with their unique mode of transcription, where the order of a given *Hox* gene, within a cluster of related genes, defines the time and place of expression during development, a phenomenon known as colinearity. There are punctual exceptions to this transcriptional behaviour, adding complexity to the understanding of *Hox* gene's regulation and function. Nevertheless, the precise expression domains that these genes assume suggest that refined regulatory mechanisms, which change over time and place, take action to instigate a particular expression profile. The discovery of the principles that govern *Hox* regulation has been a challenge, which recently has gained new impetus with the increasing knowledge of the genome, its products, and its epigenome. A fundamental part of this research has been the identification of the "players" acting upon *Hox* regulation. This effort led to the discovery of genomic regions that act as promoters, enhancers or insulators of *Hox* gene's transcription. Additionally, micro- and long-non-coding RNAs, which contribute to the transcriptional or post-transcriptional control in specific cases, have been uncovered and the state-of-the-art is now heading towards understanding the impact of chromatin configuration and other epigenetic phenomena associated with the regulation of *Hox* gene expression and function. Here we review the current information available on the elements and mechanisms found so far to act on *Hox* gene regulation and discuss their hypothetical role in the evolution of the *Hox*-mediated developmental processes.

KEYWORDS: *Hox* genes, transcriptional regulation, development, evolution

ABBREVIATIONS

AP	-	anteroposterior
CpG	-	cytosine-phosphate-guanine dinucleotide
CRE	-	*cis*-regulatory elements
eRNAs	-	enhancer RNAs
lncRNAs	-	long non-coding RNAs
miRNAs	-	microRNAs
Mya	-	million years ago
ncRNAs	-	non-coding RNAs
PD	-	proximodistal
PG	-	paralogous group
piRNA	-	piwi-interacting RNA
PAR	-	promoter-associated RNAs
R	-	round
r	-	rhombomere
RA	-	retinoic acid
s	-	somitomere
siRNAs	-	small interfering RNAs
TF	-	transcription factor
WGD	-	whole genome duplication

*Corresponding author: renata.freitas@ibmc.up.pt

INTRODUCTION

Understanding the evolutionary process that led to the current multitude of animal body morphologies is, without a doubt, one of the most fascinating scientific endeavours. Regarding the molecular basis of body pattern establishment and evolution, few genes assume a role as central as the Hox family of transcriptional factors (Figure 1). The discovery of these genes allowed for the description of a general principle for the specification of the body plan, along the anteroposterior axis of animals [1]. In animal models as different as the fruit fly and the mouse, experiments causing deletion or ectopic expression of a given *Hox* gene are able to alter the anteroposterior identity of specific embryonic structures, causing a modification in axial morphology [1, 2]. Furthermore, these genes seem to have acquired novel developmental roles in the course of vertebrate´s evolution, becoming, for example, essential for the outgrowth and patterning of appendicular structures, both along the anteroposterior (AP) and the proximodistal (PD) axes [3-5].

Hox genes are present in all bilaterian genomes mapped up to date [6] and are characterized by a 180 bp long conserved motif, which is transduced into a highly conserved 60-aminoacid sequence, the "homeodomain" [7]. This domain of Hox proteins is responsible for their binding capacity to DNA and RNA, but it also seems to be important in the determination of protein-protein interactions [8, 9]. The exact way by which Hox homeodomains find their targets within a living cell, has been enigmatic. Nevertheless, the apparent stochastic

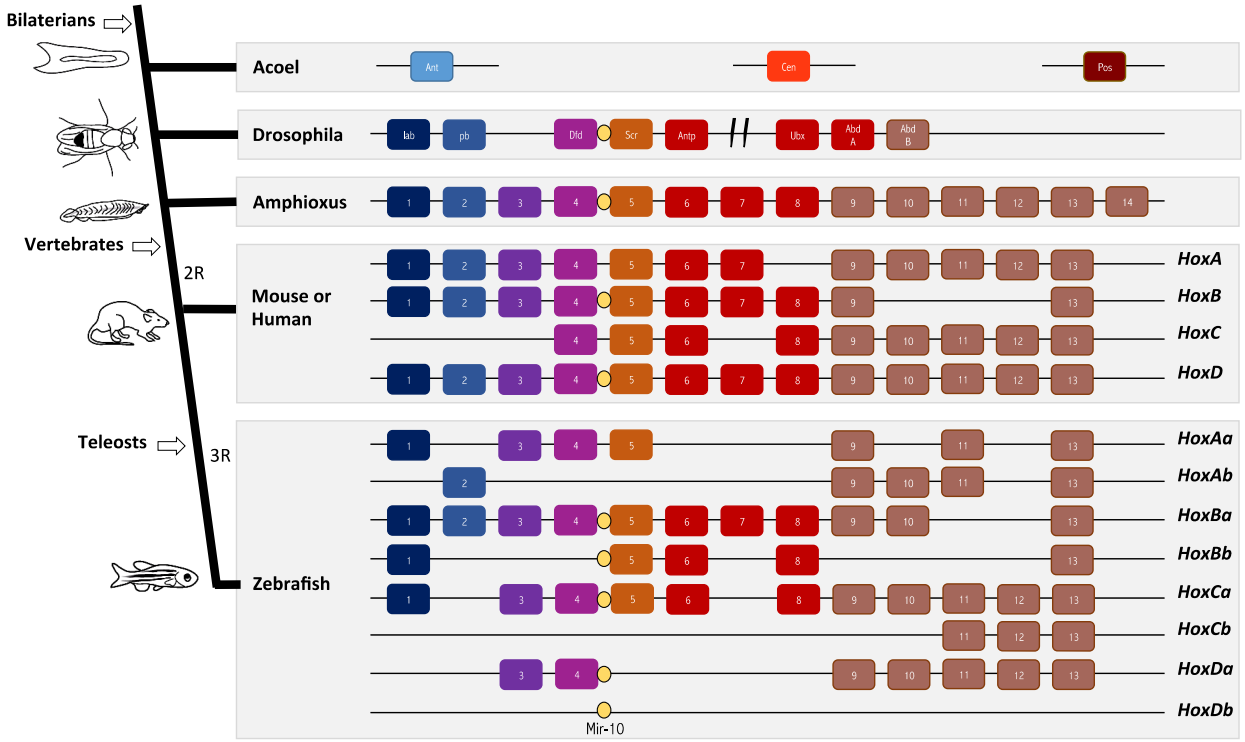

Figure 1. Dendrogram showing the repertoire of *Hox* clusters in representatives of bilaterians and illustrating their evolution (Based on references [10, 119-122]). *Hox* gene clusters are thought to have developed by a process of duplication and divergence from primordial homeobox genes with anterior (Ant), central (Cen) and posterior (Pos) expression domains along the body axis. Vertebrates, such as mouse and human, present 4 clusters that may have derived from a single basal deuterostome cluster that suffered two rounds of whole-genome duplication (2R). In certain lineages of vertebrates, a third round of whole genome duplication took place (3R), as in the teleost lineage that has a total of 8 clusters. Note the extensive gene loss that occurred in zebrafish and the simultaneous conservation of regions with regulatory function, such as Mir10.

binding to the chromatin seems to be maintained for longer time periods with particular target sequences to which Hox proteins, together with co-activators or co-repressors of gene expression, bind to exert their biological function [9].

Throughout the evolutionary history of deuterostomes, *Hox* genes suffered considerable modifications both in number and genomic organization, as revealed by analysis performed in invertebrates and basal vertebrate lineages [10]. Curiously, an evolutionary tendency to generate clustered *Hox* genes, with the same transcriptional orientation, seems to accompany the increase in body plan complexity (Figure 1). *Hox* gene evolution, in most vertebrate lineages, seems to have culminated in four clusters per genome, named as *HoxA*, *B*, *C* and *D* (Figure 1). It has long been hypothesized that these clusters appeared due to two rounds (1R/2R) of whole genome duplication (WGD) that occurred during early vertebrate evolution, over 525 Mya [11]. More recently, approximately 350 Mya, an extra third round of whole genome duplication (3R) seems to have occurred in the common ancestor of more than 26,000 species of teleost [12] (Figure 1). The same phenomenon seems to have happened independently in basal vertebrate lineages, such as the one comprising the lampreys [13].

Body regionalization is a fundamental biological function of *Hox* genes and they do so in a fashion that recapitulates their genomic organization, meaning that the most anterior genes in a cluster exert a major role in the patterning of the anterior portion of the body while the posterior genes are responsible for guiding the regionalization of the most distal parts [14]. For example, members of the *HoxPG1* (paralogous group) and *PG2* in mammals are expressed in the developing brain [15-17]; *HoxPG3, PG4* and *PG5* are involved in the patterning of the cervical skeleton [18, 19]; *HoxPG6* is responsible for controlling rib cage formation [20]; *HoxPG9* seem to be essential for the development of floating ribs [19]; *HoxPG10* inhibit rib formation; *HoxPG11* controls the development of the sacrum; and *HoxPG13* appear to be involved in the control of body length [17]. The *HoxA* and *HoxD 9-13* paralogues are also key players in the anteroposterior and proximodistal fin/limb patterning [21-23]. This characteristic expression pattern is denominated spatial colinearity and is a well-conserved characteristic of chordates [24-26]. However, some exceptions were detected in amphioxus where *Hox6* was proven to be expressed in a more rostral region than *Hox4* [27]. In addition to spatial colinearity, *Hox* genes also present temporal colinearity, meaning that the most 3' *HoxPG* are expressed, not only more anteriorly in the body, but also earlier than the 5' *HoxPG*. However, temporal colinearity is not as well conserved as spatial colinearity, as it seems that the highly organized vertebrate *Hox* cluster architecture is a requirement to maintain the sequential timing of gene expression [24, 26, 28]. At a cellular level, body patterning seems to result from the regulation of several cellular functions, directly by *Hox* genes or through their downstream targets. Studies performed in fruit flies and nematodes revealed that *Hox* genes are involved in the control of cell death, migration, affinity, size and proliferation. Through these processes they establish tissue specific characteristics [25].

Studying the influence that *Hox* genes have in body patterning during development brought to light the major importance of global mechanisms of transcriptional regulation. Indeed, the essential role of *Hox* genes in body patterning, with their characteristic temporal and spatial colinearity, the variable expression domains during different developmental stages and their function in adult tissues, indicate that a complex regulatory network, operating at a transcriptional level, must be in place to ensure proper *Hox* gene expression at each developmental stage [17, 25, 29, 30]. It is noteworthy that, although most studies focus on the transcriptional level of *Hox* gene regulation, there are also evidences of a translational control [31].

Thus, changes in these regulatory mechanisms might have led to profound variations in animal morphology during evolution [5]. An interesting example illustrating this idea was recently found in snakes. Contrasting with other tetrapods, these animals have no limbs and possess quite a uniform vertebral skeleton, which seems to have resulted, to some extent, from alterations that occurred throughout evolution in the *Hox cis*-regulatory regions [32, 33]. Indeed, a vast amount of *cis*- and *trans*-acting elements, together with an added level of epigenetic control, interact to

ensure the correct time and space boundaries of *Hox* gene transcription [34, 35]. Our understanding of *Hox* gene regulation is currently expanding, taking advantage from the emerging knowledge of the whole genome in several organisms, and the development of novel technologies that allow identification of DNA and protein interactions [36]. Driven by these questions, we propose to synthesize the latest research regarding the genetic and epigenetic control of *Hox* gene expression and explore how this impacts embryonic development and the genesis of morphological innovations throughout evolution.

1. *Cis*-regulation of the *Hox* clusters

In global terms, transcriptional regulation is mediated by *cis*-regulatory elements (CREs) that are bound by transcriptional factors able to interact with RNA polymerases and potentiate their interaction with gene promoter regions. Interestingly, *Hox* clusters have been one of the favoured systems to study these interactions and to explore the function of different CREs on gene regulation, a challenging task involving multiple experimental and computational tools [37]. Indeed, the global *cis*-transcriptional control of *Hox* clusters might explain their tendency to conserve their genomic arrangement and expression during evolution [38]. A considerable progress has been achieved concerning the identification and characterization of *Hox* cluster associated CREs that includes promoters, enhancers and insulators.

a) Promoters

Regarding the described *Hox* gene promoters (Table 1), cases were found where a single promoter is sufficient to activate transcription throughout development, such as in mouse *Hoxb1* [39], *Hoxc4* [40], *Hoxb6* [41], *Hoxa7* [42] and *Hoxd9* [43]. For this last gene, the promoter has a modular nature, being composed of at least two separated elements [43]. However, more than one promoter may regulate a particular gene, acting concomitantly or differentially in a specific tissue or developmental stage [44, 45]. As such, two promoters were found to act on mouse *Hoxb3* [46, 47], *Hoxa4* [48], *Hoxb4* [49], *Hoxd4* [50] and *Hoxc6* [50]. The same scenario was found for human *HOXD4* [45, 51, 52] and *HOXC5* [53], and also for zebrafish *hoxb4a* [54]. Further complexity was found for mouse *Hoxa5* that seems to be activated by three distinct promoters [44], a feature shared with the zebrafish *hoxb3a* [54]. Moreover, situations where distinct *Hox* genes share a promoter also occur, in the zebrafish *hoxb3a* and *hoxb4a*, which supposedly ensure the overlapping and tissue-specific expression of both genes in the posterior hindbrain and spinal cord [54].

The presence of multiple promoters (Table 1) hints that each may specify different expression domains by displaying different sensitivities to retinoic acid (RA), or binding to different transcriptional factors. This was indeed observed for mouse *Hoxd4* promoters, with Promoter 1 (P1) being the sole responsible for transcription in the most anterior domains of *Hoxd4* expression in the hindbrain (r6/7 boundary) while P2 mediated expression in the hindbrain is more caudal and diffuse [45]. These two promoters also display different sensitivity to RA [45]. The human *HOXD4* gene also exhibits two promoters, a distal one, located 1142 bp upstream of the transcription starting site, and a proximal one, that is distanced by 20 bp from the ATG site [45, 52]. Curiously, the human distal promoter shows conservation with the mouse P1 promoter, but no similarity was found between the human proximal promoter and the mouse regulatory sequences [45]. Also, sequence comparison revealed that the P1 promoter described in mice seems to be conserved in zebrafish [51]. Moreover, an auto-regulatory element (ARE) and two RA response elements (RARE) involved in *HOXD4* regulation are conserved, as indicated by sequence comparison, among human, mouse and zebrafish [45, 51]. The human promoters appear to be regulated in a tissue- and stage-specific manner, and display different sensitivity to RA induction [52, 55].

Other promoters were also shown to provide tissue/stage specific expression information, as is the case of *Hoxb6* promoter, which has the capacity of directing the expression in the anterior and posterior regions of hindlimbs, the posterior region of forelimbs, and the ventrolateral regions of the limb buds, during development [41]. The *Hox* gene expression specificity is also aided by promoter negative regulators, putative inhibitors,

Table 1. Information on the *Cis*-regulatory elements (CREs) found in zebrafish, mouse and human. Asterisk (*) indicates situations in which enhancer elements were suggested to also contribute to the expression profiles. References (Ref.) indicated in the last column.

Organism	Gene	*CREs*	Genomic location	Observations	Ref.
Mouse	*Hoxa4*	Proximal promoter	360 bp upstream the gene's starting site	Inducer of *Hoxa4* expression	[48]
		Distal promoter	2000 bp range upstream or in the 725 bp region downstream of the gene's starting site	Mediator of *Hoxa4* expression in response to RA	
Mouse	*Hoxa5*	Proximal promoter (P1)	Near the gene's starting site	Promoter of the most abundant transcripts	[44]
		Distal promoter (D1)	*Hoxa6*'s putative promoter	Inducer of bicistronic RNAs containing *Hoxa5* and *Hoxa6* information	
Mouse	*Hoxb3*	Proximal promoter (P1)	*Hoxb3* second intron	Specifies expression in neural tube	[46]
		Distal promoter (D1)	17 kb upstream of the proximal promoter	Specifies expression in neural tube, mesoderm and forelimb*	
Zebrafish	*hoxb3a*	P1	Upstream of *hoxb3a* exon 4	Mediator of expression within the central nervous system	[54]
		P2	Upstream of *hoxb4a* exon 1	Produce bicistronic RNA (*hoxb3a* and *hoxb4a*) and mediates expression within the spinal cord and in mesodermal cells	
		P3	Partially overlaps with *hoxb3a* exon 2	Mediates expression at r4.	
Zebrafish	*hoxb4a*	P1	Upstream of *hoxb4a* exon 2	Mediator of expression in the central nervous system and mesodermal cells	[54]
		P2	Upstream of *hoxb4a* exon 1	Produce bicistronic RNA (*hoxb3a* and *hoxb4a*) and mediates expression within the spinal cord and in mesodermal cells	
Mouse	*Hoxb6*	Promoter	3.6 kb range upstream of *Hoxb6* starting site	Specifies expression in hindlimbs and forelimbs	[41]
Mouse	*Hoxd4*	Promoter 1 (P1)	1.1 kb upstream of the coding region	Promoter of the most anterior domains of *Hoxd4* expression in the hindbrain	[45]
		Promoter 2 (P2)	5.2 kb upstream of the coding region	Mediation of expression is in more caudal and diffuse domains in comparison with P1.	
Human	*HOXD4*	Proximal	20 bp away from the starting site	Promoters are regulated in a tissue/stage specific manner; they have different sensitivity to RA induction	[52]
		Distal	1142 bp upstream of the gene's starting site		

as reported for *Hoxb4*. Besides the two promoters (P1 and P2), other *cis*-regulatory sequences were also identified near the promoters of *Hoxb4* gene and were denominated *a*, *b*, *c* and *d*. The *c* element was suggested to be a putative enhancer requested to activate the *Hoxb4* promoter, while *a*, *b* and *d* seem to act as negative regulators revealing cell specific activity [49]. The only other case known of a possible inhibitor comes from the *Hoxc8* gene, where a negatively acting element was shown to be able to suppress reporter gene expression in the developing spinal ganglia of transgenic mice [56].

Hox gene's promoter-mediated transcription has an additional degree of complexity, as above mentioned for the zebrafish, where some genes form transcriptional units by co-option of the same promoter. For instance, it was shown that in the case of *Hoxa5*, a proximal promoter (P1) drives expression of the most abundant transcript, and two distal promoters D1 (the *Hoxa6*'s putative promoter) and D2 (located downstream of the 3' end of *Hoxa7*) originate bicistronic RNAs, containing both *Hoxa5* and *Hoxa6* transcriptional information [44]. The human *HOXC4*, *-C5* and *-C6* have also been shown to form a transcriptional unit [57]. These three genes are simultaneously transcribed from the same major upstream promoter, into a polycistronic pre-mRNA, that is then alternatively spliced to produce the different mature mRNAs. However, this finding does not exclude the possibility that each gene possesses an individual promoter that mediates a transcriptional response in a more tissue- and/or stage-specific manner [57], as is the case of the identified *HOXC5* proximal promoter [53].

In summary, a striking observation comes from the above-mentioned studies: *Hox* gene promoters can account for the temporal and spatial specificity of *Hox* gene expression in vertebrates as divergent as mammals and teleosts. In fact, specific promoters, or a particular combination of them, were found to regulate *Hox* transcription in particular groups of cells, during precise developmental times. The challenge now is to discover how this promoter-based regulation evolved and how this relates to the morphological innovations of vertebrates. It would also be relevant to gain insight on how the recruitment of promoters is related to evolutionary changes in the *Hox* clusters themselves, for example by evaluating how they vary with modifications in the intergenic regions length that occurred in particular lineages of vertebrates.

b) Enhancers

Enhancers are DNA sequences that contain short motifs acting as recognition sites for transcriptional factors, which upon binding recruit other proteins that will modulate the transcriptional activity of the enhancer's target gene or genes [58]. These *cis*-acting elements act independently of distance and orientation with respect to the targets, and assume a modular nature, where each element contributes additively to the complete gene expression pattern. Furthermore enhancers are associated with nucleosome devoid chromatin and their flanking chromatin regions are usually enriched with the H3K4me1 (Histone 3 lysine 4 monomethylation) and H3K27ac (Histone 3 lysine 27 acetylation) histone marks [58].

The dynamic expression patterns that *Hox* genes present during embryonic development has long suggested sophisticated mechanisms of transcriptional regulation, able to modulate their levels over time with tissue-selectivity [2]. Transgenic assays, particularly in mice, revealed that indeed a quite complex enhancer network is in place during *Hox* gene transcription to direct their expression in a tissue and stage specific manner. For instance, three distinct enhancers were implied in the regulation of *Hoxa1* and *Hoxa2* during early murine craniofacial development [59]. Enhancer-I, located within *Hoxa2* region, seems to act upon *Hoxa1* transcription in its most anterior expression domain, the dorsal region of r4. This enhancer may also be implied in *Hoxa2* expression itself, directing its transcription in that rhombomere. However, it remains unknown if this *cis*-regulatory element acts on each gene independently or rather simultaneously [59]. Enhancer-II directs majorly *Hoxa2* expression in r2. However, it also seems to be responsible for both *Hoxa2* and *Hoxa1* transcriptional activation in the rostral somites. Curiously there is some evidence that enhancer-II has been conserved since the divergence of chordates, as this element, and possibly some of its upstream regulators, can direct *Hoxa2/probocipedia* expression in a restricted rostral region, in mouse and *Drosophila*. Enhancer-III is active in the notochord, floor plate and gut epithelium, and its characteristic RARE

element is essential to drive *Hoxa1* expression in r5, and possibly r6. Regarding the r2 boundary of *Hoxa2* expression, a very curious regulatory module was discovered. It contains five elements and two Sox protein binding regions. What is interesting in these CREs is the fact that they are located within the coding region of the *Hoxa2* second exon, which places the second exon under an evolutionary constrain for both a coding and regulatory function, possibly contributing to the high cross-species conservation detected for this particular enhancer [60].

Beside enhancers-I and -II, *Hoxa2* is regulated by other *cis*-acting elements, such as a compound enhancer, formed by four distinct interacting elements, that is responsive to the AP-2 family of transcriptional factors and directs *Hoxa2* expression in cranial neural crest cells [61]. Furthermore, Nonchev *et al.* described the presence of a Krox20-responsive enhancer located in the intergenic region between *Hoxa2* and *Hoxa3* that has the capacity of establishing the r3 an r5 expression domains of *Hoxa2* [62]. Curiously, a Krox20-responsive enhancer, that has the capacity of directing expression to the r3 and r5 domains, was also found in *Hoxb2* [63], denoting some regulatory similarity between paralogous *Hox* genes.

For *Hoxb1*, one of the *Hoxa1* paralogues, the enhancer-driven expression is set by a different complement of elements. In fact, the transcription of this gene is controlled by an enhancer located 5' to the transcription initiation site and a RARE element located 3' of the gene. The 5' enhancer was shown to have an *auto*- and *para*-regulatory binding site, where Hoxa1 and Hoxb1 binding help to delimit the r4 expression boundary while Hoxb1 alone is required to ensure high *Hoxb1* expression levels [64]. The RARE element seems to be essential to trigger a proper early *Hoxb1* transcription. Not surprisingly, a RARE is also present in the 3' region after the *Hoxa1* gene and mediates the early expression of this gene, which will then exert its *para*-regulatory function on *Hoxb1* [64]. Moreover, Hoxa1 and Hoxb1 also exert a *para*-regulatory role upon *Hoxb2*, through interaction with a *Hoxb2* enhancer located in the 5' region of its transcription starting site that possess a Pbx/Hox binding sequence. Notably, the *Hoxb2* response mediated by RA is achieved indirectly, as it is a consequence of *para*-regulation by the RA responsive *Hoxa1* and *Hoxb1* genes [65].

Hoxb3 has a quite complex *cis*-regulatory network that seems to comprise an additional promoter and four regulatory elements, including some previously implied in *Hoxb4* regulation, located between the promoters P1 and P2 [46]. These regulatory elements (I, III, IV, and V) are responsible for setting the different boundaries of *Hoxb3* expression. For instance, element Va is a major mesoderm-specific element, which has the capacity of directing *Hoxb3* expression in all somites, with the exception of the first pair, and establishes the vertebrae C1 expression domain at E12.5 in mice. Besides Va, elements IIIa and IVb also have activity in the paraxial mesoderm expression of *Hoxb3*, with IIIa controlling expression from the newly condensed somitomere up to the last 6 somites and IVb regulating expression up to somite 6 [46]. Elements Ib, IVb, and Va are also implied in limb development where each element establishes a specific expression pattern in a stage dependent manner. The regulatory elements Ib, IIIa, IVb, and Va are also important in establishing *Hoxb3* transcription in a wide range of mesodermal derived tissues that display *Hoxb3* endogenous expression. Moreover, the initial wave and anterior domains of *Hoxb3* transcription are set by elements IIIa, IVa and the element-I´s RA responsive early neural enhancer, which acts on P2 to establish the r6/7 anterior limit. The rhombomeric expression of *Hoxb4* is induced by the same RA responsive elements [66, 67]. The late stage and posterior expression is only determined by an autoregulatory element-I component, the late neural enhancer, also known as the CR3 element, which controls *Hoxb4* [46, 66]. Additionally, in mouse and zebrafish the expression of *Hoxb3/Hoxb3a* in the r5 boundary is controlled by an enhancer that contains binding sites for Krox-20 (KroxB), Kreisler and Pbx/Hox transcription factors [54].

Besides the promoters, ARE, and RARE that regulate *Hoxd4* transcription, several enhancers have been reported to be essential in the establishment of *Hoxd4* expression at the neuronal and mesodermal boundaries [68]. For instance, the human *HOXD4* neuronal expression limits are set by two enhancers, one located in the 5' direction of the ATG starting site that mediates neural-specific expression and a composed 3' located

enhancer that is essential to establish the r6/7 expression limits. This 3' enhancer is in turn formed by two elements, a 3' component that mediates neural specificity, and a 5' element that is responsible for setting the anterior *HOXD4* expression boundary at r6/7 [68]. The enhancers of the 3' region are conserved in mouse *Hoxd4* [68] and in zebrafish *Hoxd4a* [51] genes. Human, mouse and zebrafish 3' enhancers and human and mouse 5' enhancer possess an RA response element that is essential for initiation and maintenance of *Hoxd4* expression in the hindbrain [51, 68, 69]. Moreover, Morrison *et al.* reported the existence of other two enhancers in human and mice, flanking each side of the 3' neuronal enhancer, which are responsible for establishing the mesodermal expression of *HOXD4*. Both mesodermal enhancers possess the capacity to promote expression up to the proper s5/6 junction [68, 70].

In the mouse *Hoxb4* paralogue, a 3' region with a RARE and 5' region with neuronal enhancers were also identified, where they determine the r6/7 boundary and neuronal-specific *Hoxb4* expression, respectively [66, 68]. The 3' neural enhancer contains two conserved regions (CR2 and CR3). CR3 is the most downstream of the two elements and is *Hox*-responsive, responding to *auto*- and *para*-regulatory feedback from a range of HOX proteins. This enhancer is sufficient to direct the late phase of neural expression of *Hoxb4* and precisely defines the r6/7 expression position [66] and is also functionally conserved in zebrafish *Hoxb4a* [54]. An enhancer implied in the early expression of *Hoxb4* was also identified further downstream of CR3 and found to restrict the expression in the neural tube, extending up to the future r6/7 boundary and mirroring the early phase of *HOXB4* expression [66]. Both early and late phase enhancers are responsive to RA, but they show different sensitivity, with the early enhancer being rapidly induced in a transient manner and the late enhancer being activated later, but producing a sustained response [66]. No relevant sequence similarities were detected between *Hoxd4* and *Hoxb4* 3' enhancers, implying that, if the similar function of this 3' neuronal enhancer is derived from a common ancestral regulatory sequence, its current function is mediated by small dispersed motifs [68]. A single mesodermal enhancer was detected near the 3' neural enhancer, but other mesodermal enhancers were also identified elsewhere. One of them is located near the 5' neuronal enhancer and a third one is present in *Hoxb4* intronic region, which seems to regulate expression at both neural and somitic mesodermal limits (s6/7) [68].

Hoxa4 also possesses a 3' region with a RARE and 5' region neuronal enhancers. Although the 3' regulatory element does not have the capacity to specify the r6/7 limit of *Hoxa4* expression, being responsible for neuronal-specific *Hoxa4* expression, while the 5' enhancer provides the r6/7 expression limits information [68]. Two enhancers determine the *Hoxa4* mesodermal expression pattern: one located near the 3' neuronal enhancer and the other near the 5' neuronal enhancer. Both mesodermal enhancers have the capacity to specify the s7/8 limits of *Hoxa4* expression [68].

Hoxa7 expression boundaries are also determined by several enhancers, which act upon the promoter of this gene to restrict its expression to the appropriate tissues in a stage specific manner. Three enhancers were identified for this gene, referred to as A, B and C. Element A has a dual function as it seems that this enhancer is important to maintain high expression levels of *Hoxa7* and also to determine the anterior *Hoxa7* expression border. Element B is essential in the establishment of the posterior expression boundary in the mesoderm and in the ectoderm. This element also provides lineage restriction information, as it is responsible for limiting *Hoxa7* expression to sclerotomal cells and restricts *Hoxa7* expression to the ventral half of the neural tube. Finally, element C is responsible for confining *Hoxa7* expression in pre-vertebrae [42].

In a study done by Anand *et al.*, an interesting point was highlighted regarding *Hoxc8* early enhancers. This team showed that a *Hoxc8* enhancer controls the initiation and establishment of this gene's expression in the posterior regions of the neural tube and mesoderm [71]. In mouse, this enhancer is composed of five elements (A to E) that possess potential binding sites for caudal, Hox, fork-head and the high mobility group/box of transcriptional factors. The elements A to D seem to be well conserved in zebrafish and pufferfish, in contrast with the region that encompasses element E [71]. Reporter assays done in mouse showed some variations in the expression domain dictated by the enhancers of each species; although the mouse, zebrafish and

pufferfish enhancers drive *Hoxc8* expression in the posterior neural tube and mesoderm, they do so at different boundaries, adding evidence to the hypothesis that elaboration/remodelling of the CREs that control *Hox* gene expression was important to generate morphological diversity across species [71].

In the context of limb development, the *HoxD* cluster has been providing valuable information on how *cis*-regulatory elements located outside of the *Hox* cluster interfere with the expression of its constitutive genes. Indeed, during the development of these appendages, two phases of *HoxD* gene expression take place, which relate with the formation of distinct anatomical structures. The initial phase leads to the patterning of arm and forearm and is regulated by a 3' (telomeric) regulatory landscape (Figure 2). The late phase, crucial for autopod development, is characterized by the expression of *Hoxd10* to *Hoxd13* and is under control of 5' (centromeric) *cis*-regulatory elements [14, 72] (Figure 2). The most anterior and the most posterior *HoxD* genes only interact with their respective 3' and 5' regulatory landscapes while central genes, such as *Hoxd9*, *Hoxd10* and *Hoxd11*, can interact with both regulatory landscapes, depending on which is active at the time [14]. A similar bimodal regulatory system seems to be required to regulate the *HoxA* cluster during limb development [14]. Further studies in mouse revealed that the late phase of 5' *HoxD* gene expression is indeed mediated by five distinct regulatory islands, which contact with each other, and with the 5'*HoxD* genes, in particular time points [72]. The most downstream island (CsC), located between *Lnp* and *Evx2*, is a tetrapod specific enhancer. Regarding the remaining islands (I to V), it was found, so far, that islands II, III and V are widespread in vertebrates while islands I and IV are tetrapod specific [73]. This discovery is truly remarkable as it suggests that autopod evolution might have arisen as a result of the elaboration of *HoxD* and *HoxA* regulatory landscape [5, 14, 73] (Figure 2).

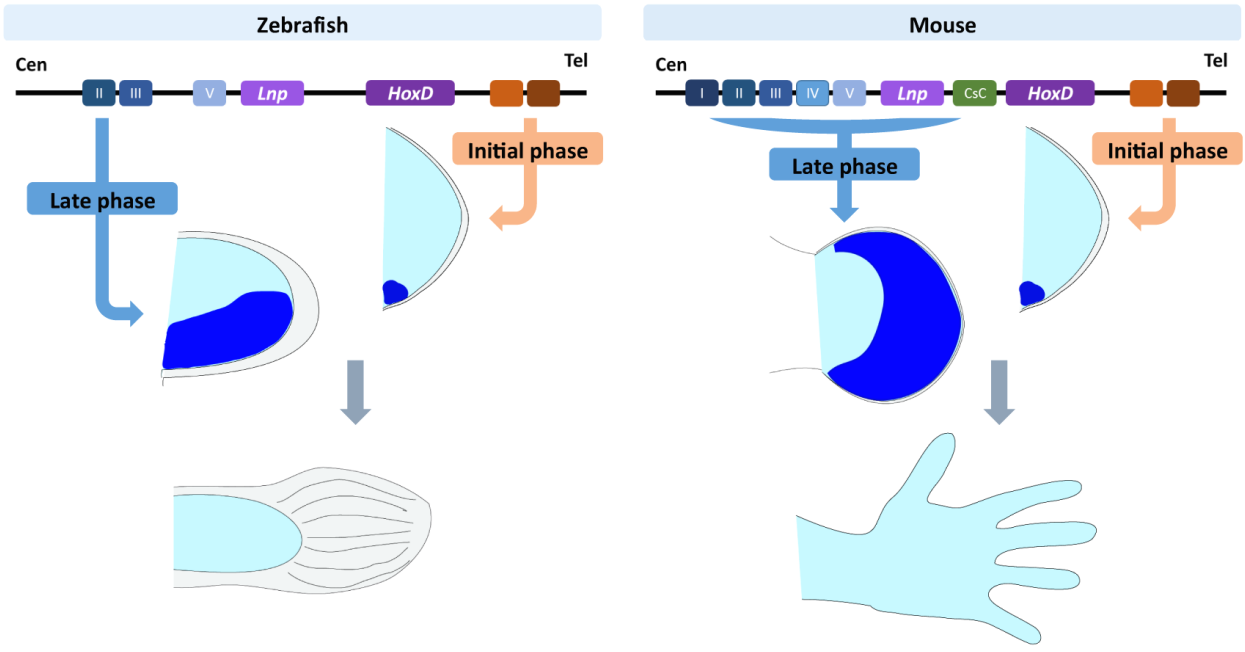

Figure 2. *Hoxd* regulatory landscapes involved in appendage development in zebrafish and mouse. Based on references [2, 73]. Expression of *Hoxd13* paralogues is represented in dark blue in the developing fins and limbs outlined. An initial phase of *HoxD* gene transcription, regulated by CREs in the telomeric (Tel) side of the cluster, occurs during the formation of fins and limbs. Then, CREs situated centromeric (Cen) to the cluster activate a late phase of expression, which has been associated with autopod development in mouse. Note the missing of *cis*-regulatory regions I, IV and CsC in zebrafish, which may account for the expression pattern variations detected between these two species.

Taking into account what is presented above, a new paradigm arises in *Hox*-related studies, which emphasizes the importance of their *cis*-regulation for the dynamic transcriptional modulation of these genes throughout development. This research will most likely clarify the impact that addition/ restructuration of *Hox* CREs had in the determination of the *Hox* expression patterns and on how variations in these regulatory processes contributed to the extant cross-species morphologic variability.

c) Insulators

Insulators are another example of *cis*-acting elements, however unlike enhancers, these elements are involved in the compartmentalization of chromosome regions into independent transcriptional units. This is the reason why insulators are also known as boundary elements [74]. The classic view on their mechanisms of action states that these elements block enhancers' activity, impeding their access to gene promoters [74].

Within the *Hox* genes, insulators have been described with more detail for the *HoxD* cluster, where they are implied in restricting the transcription to specific sets of *HoxD* genes within a tissue. An interesting example emerged with a study of *HoxD* genes in the context of limb development. Prior to the development of the autopod in these appendages, *Hoxd10-13* are expressed in the distal portion of the developing limb buds, however *Hoxd10* and *Hoxd11* are also required for proper development of the ileo-coecal sphincter [75]. These distinct expression domains of *Hoxd10* and *Hoxd11* are set by a polar insulator, discovered by inversion of the regions between *Hoxd13* and *Hoxd10* genes, which include *Hoxd11* and *Hoxd12* transcriptional units. This *cis*-acting element seems to be involved in the determination of the number of genes that respond to a given regulatory landscape [75].

Another insulator was suggested to be determinant in the time-space separation of *Hoxd13* and *Evx2* expression domains during the development of limbs, external genitalia and brain. This insulator, located within the *Evx2-Hoxd13* intergenic region is composed of two units: one that possesses the blocker activity and a second one that drives tissue specificity [74]. Those two units are conserved between mouse and chicken and at least one of them is also present in the zebrafish genome [76].

A third insulator unit, also located in the *Evx2-Hoxd13* intergenic region, was similarly implied in the distinction of *Evx2* and *Hod13* expression domains in the early central nervous system. This element shows an outstanding conservation in bilaterians, being detected in animals as divergent as humans and fruit flies. Its mechanism of action has been revealed in *Drosophila,* where it exerts its function through binding of the GAGA insulating factor to GA repeats [77]. The vertebrates' GAGA factor homolog also seems to be important in the establishment of the proper *Hox* gene expression domains, as recently demonstrated by high-resolution chromatin immuno-precipitation Chip-on-Chip tiling arrays. This technique revealed the presence of histone-3 free regions in almost all intergenic regions of the *Hox* clusters, which are associated with the vertebrate GAGA factor. These regions display significant enhancer-blocking activity in human cells, suggesting a scenario where *Hox* gene expression is regulated by the distinct chromatin domains of each gene [78].

In summary, several insulators have been shown so far to have a determinant impact on *Hox* gene transcription by interfering with the range of action of their enhancers. It is interesting to notice their involvement in the transcriptional regulation of 5' *HoxD* genes during tetrapod limb development, which leads to question what might have been the phylogenetic origin of these elements and if that relates to the emergence of limbs in the tetrapod lineage.

2. *Trans*-regulators of *Hox* gene transcription

The transcription of a gene is influenced not only by CREs, but also by the so-called *trans*-regulation, which is defined as the action that nuclear-based molecules exert upon the *cis* elements. Over the last few years, several *Hox* gene *trans*-regulators have started to be characterized [25]. Indeed, *Hox* genes can be regulated by variable transcriptional factors (Tfs), including KROX20, KRML1, AP-2 proteins, Cdx, Hox proteins themselves, and retinoic acid [25, 69, 79]. These molecules bind to specific promoter/enhancer sequences to direct *Hox* gene expression in the appropriate time/space domains during development. *Krox20* encodes a zinc-finger Tf that, through regulation of its downstream targets, plays a particularly significant role in the

establishment of odd-numbered rhombomeres, with a crucial function in the formation, and following delimitation, of r3 and r5 territories, as shown in chicken embryos [80]. It is in these rhombomere territories that Krox20 exerts its regulatory function upon *Hox* genes. This Tf acts as a positive regulator of *Hoxa2* [81] and *Hoxb2* [63] in mouse r3 and r5, and, by interaction with PIASxβ, regulates negatively *Hoxb1* as shown in chicken embryos [80]. Krml1, in turn, is a large Maf bZIP (basic leucine zipper) Tf equally involved in hindbrain segmentation that binds to enhancers that regulate *Hoxa3* expression in r5 and r6, and *Hoxb3* expression in r5 [82]. The AP-2 family of Tfs is yet another example of a *trans*-acting regulator of *Hox* genes. This family of proteins is required for proper craniofacial patterning [79] and three of their members (AP-2α, AP-2β and AP-2.2) are able to bind a *Hoxa2* enhancer and promote its expression in the cranial neural-crest cells, with a fundamental impact for the development of the jaws [61, 70].

As mentioned above, *Hox* genes are also regulated by homeobox-containing Tfs, such as Cdx and Hox proteins themselves [69, 83]. *Cdx* genes are important modulators of the AP identity along the body, with a crucial role in promoting posterior tissue expansion, as revealed by studies performed in mouse and zebrafish [83]. Part of the AP patterning mediated by *Cdx* genes is achieved through regulation of downstream central *Hox* genes. Indeed, Cdx regulates *Hoxb8* expression through interaction with its enhancer region [29] and *Hoxa7* expression through direct binding to its promoter [84]. However, the set of *Hox* genes regulated by *Cdx* may be larger and encompass most *Hox* genes involved in the AP patterning [84, 85]. The *Hox* regulation by Hox encoded proteins occurs in an *auto*- and/or *para*-regulatory fashion [69], as illustrated by the case of *Hoxb1* [64]. One of the enhancers of this *Hox* gene possess an auto- and para-regulatory binding site, where binding of Hoxa1 and Hoxb1 help to delimit the expression in the r4 boundary, while Hoxb1 alone is required to ensure high *Hoxb1* expression levels [64]. Examples of other genes regulated by Hox proteins are *Hoxb2*, *Hoxa4*, *Hoxb4* and *Hoxd4* [69]. It is noteworthy that the efficiency of most Hox proteins to bind DNA is enhanced by hetero-dimerization with members of the PBX/EXD family and several studies have revealed that *in vivo* functional Hox binding sites are co-localized with PBX-binding consensus sequences [29, 65].

As a fundamental morphogen during embryonic development, retinoic acid assumes a critical role in the regulation of *Hox* genes that ultimately specify the AP identity of the main body axis and the PD identity of secondary appendicular structures, such as limbs and genitalia [86-89]. This small lipophilic molecule regulates gene expression by binding to retinoic acid nuclear receptors (RAR), which in turn are bound to retinoic acid response elements (RARE). As previously mentioned, these RARES have been found to regulate several *Hox* genes such as *Hoxa1*, *Hoxb1*, *Hoxa4*, *Hoxb4*, and *Hoxd4* [69].

From an evolutionary perspective, it would be interesting in the future to gain further insight on the Hox *trans*-regulation in other vertebrate groups. For example, the recruitment of AP-2 family members as *trans*-regulators of *Hoxa2* expression might have been relevant for the origin of the jaws, a characteristic feature of the gnathostome lineage. Similarly, the involvement of Cdx proteins in the definition of the AP expression of some *Hox* genes may help to explain the diversification of the axial structures in vertebrates. Therefore, it might be quite informative to explore the phylogenetic origin of these *trans*-regulatory mechanisms.

3. Epigenetics and *Hox* gene regulation

Epigenetics is defined as a group of heritable changes that control gene expression by mechanisms that exclude the underlying DNA sequence. Most of these processes control the accessibility of the DNA to other proteins through implementation of a condensed or relaxed chromatic state that will repress or facilitate gene expression [90]. To date, several epigenetic processes have been identified, those being histone modification, chromatin remodelling, DNA methylation and gene nuclear positioning. Histone modification contributes to gene expression regulation through post-translational modifications that include methylation, acetylation, phosphorylation, ubiquitylation, and sumolyation, which cause different epigenetic responses [91]. Another fundamental epigenetic phenomenon during

development is chromatin remodelling, which is caused by histone modifications, several enzymes, microRNAs and small interfering RNAs [90, 91].

Given the complexity of the *Hox* expression patterns during development, with the required and characteristic temporal and spatial colinearity, it is not surprising the requirement of a proper epigenetic regulation to ensure appropriate levels of transcription. An interesting epigenetic event occurs during early development assuring *Hox* gene silencing that then maintains the cells in an undifferentiated state. This process has been associated with a "bivalent" chromatin configuration, which is characterized by the simultaneous presence of repressive (H3K27me) and permissive (H3K4me) histone marks [90]. When *Hox* expression is required, the bivalent chromatic signature is resolved. At that point, it was shown that permissive chromatic states are implemented by TrxG proteins, acting in parallel with UTX and JMJD3 to demethylate the repressor H3K27me3 while the repressive chromatic state is set by PRC2 mediated recruitment of RBP2 (JARID1), which will demethylate the permissive H3K4me3 histone modification [30, 90, 92]. *Trithorax* (TrxG) and *Polycomb* (PcG) groups of proteins are essential to establish the permissive and repressive chromatin states in the *Hox* clusters, a topic extensively reviewed lately [72, 90, 93].

As development progresses, other epigenetic events take place to assure the characteristic temporal and spatial colinearity of *Hox* gene expression and its function in cell differentiation. It consists of the sequential and unidirectional chromatin opening resulting in the expression of each *Hox* gene according to its physical location within the cluster [72, 90]. Such was observed, in mouse *in vivo*, for the *HoxD* cluster, where a loss of H3K27me3 and a simultaneous gain of H3K4me3 were shown to accompany the sequential *HoxD* gene activation during extension of the main body axis [72].

An additional level of epigenetic regulation was also described for *Hox* genes, related to chromatin organization and its impact in nuclear positioning of each gene [90]. For instance, DNA looping was shown to correlate with the temporal expression of the *HoxB* cluster. Reports have shown that *Hoxb1* loops out in cells of the r4 hindbrain segment, where its transcriptional activity is high [90]. However, such looping does not occur in the r5 domain where *Hoxb1* is not expressed. Looping out was also observed for genes in the 3' and 5' end of the *HoxB* cluster in the tail bud, where they are transcriptionally active. After increased exposure caused by looping, 3' *HoxB* genes were shown to assume their non-looped out state while 5' located genes remain with higher exposure [90]. A very interesting study revealed, for the *HoxA* cluster, that higher-order chromatin organization is regulated by the insulator protein CCCTC-binding factor (CTCF) and cohesin [94]. Kim *et al.* that CTCF is required for restriction of *HoxA* heterochromatic domains, and that loading of cohesin into CTCF-binding sites is necessary for higher-order loops formation and proper *HoxA* gene expression [94]. Histone H1 was also proven to be an important player in the control of *Hox* gene expression, through regulation of chromatin higher-order, as shown by *Hox* decreased expression when histone H1 is depleted from the mouse genome [95].

Gene regulation mediated by DNA methylation is achieved by cytosine's methylation in promoters. This process represses gene expression by inhibiting transcription factor binding or by promoting the recruitment of repressor proteins that contain methyl-binding domains [90]. The inverse, promoter demethylation and consequent facilitation of gene transcriptional activation, also occurs. Curiously, CpG methylation might also aid the transcriptional activity of a given gene, if such methylation occurs within the gene body, as a positive correlation between those events was found [96].

The role of DNA methylation in the control of *Hox* gene expression during development is still unclear. Although, in a paper published by Terranova *et al.* [97], where the developmental role of *Mll* was explored, this team reported the occurrence of several skeletal defects in ΔSET mutant mice. These defects are caused by altered *Hox* transcriptional levels, which in turn are associated with decreased H3K4me1 levels and altered DNA methylation patterns at the same *loci* [97]. Furthermore, data coming from oncobiology studies strongly suggest that this level of epigenetic control is indeed important in *Hox* gene regulation. For instance, in cell lines derived from mixed-lineage leukemia (MLL), the hypomethylation of

HoxA promoters has been associated with an induction of gene expression in that cluster [98]. In another study, promoter hypomethylation, open chromatin, and transcriptionally permissive histone modification were also proven to increase *HOXD9* expression in Melanoma Brain Metastasis (MBM) [99]. Hypermethylation is also involved in *Hox* gene expression control, as a study performed in oral squamous cell carcinoma (OSCC) reported the identification of hypermethylated *Hox* genes from all 4 human clusters. This hypermethylated state, for *HOXB* genes and *HOXB4* in particular, was associated with repression of expression [100]. Interestingly, Branciamore and colleagues explored the evolutionary pressure acting upon CpG islands located in the gene body, detecting very little CpG depletion for *Hox* genes and other transcriptional factors, in a marked contrast with what happened in most coding regions [96]. Therefore, they suggested that this pro-epigenetic selection of intragenic CpGs in *Hox* genes could be involved in important regulatory circuits during development [96].

Given the exposed above, it is not difficult to imagine a scenario where a collinear demethylation of *Hox* genes would accompany their expression during patterning of the AP body axis, acting in parallel with chromatin re-modulation, to ensure proper *Hox* gene expression, which gives us a hypothesis that remains to be tested in the future. Comparative studies regarding this issue would also be useful to explain the variable time of *Hox* expression found during the development of different vertebrate representatives [101].

4. Non-coding RNAs interfering with *Hox* function

Non-coding RNAs (ncRNAs), transcribed in the nucleus and processed in the cytoplasm, exert gene regulatory functions at many levels, such as by regulating chromatin structure [102] and through RNA interfering mechanisms [90]. To date, six different ncRNAs classes have been described, those being the long non-coding RNAs (lncRNAs), MicroRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), Small interfering RNAs (siRNAs), Enhancer RNAS (eRNAs) and Promoter-associated RNAs (PARs) [103]. A general feature of these ncRNAs is their low transcriptional level when compared to coding RNAs, which points to a regulatory role within the cells. This has received support from data implicating them in development and stress response regulation [103]. A curious observation is that, unlike coding RNAs that have remained relatively static among the animals studied so far, ncRNAs are highly divergent and their number per genome increases along with its evolutionary history [103]. Regarding the biologic function: piRNAs are known for their role in the suppression of transposon activity during germ line development; siRNAs are involved in post-transcriptional gene silencing; eRNAs are potential gene transcriptional activators; and it has been suggested that PARs can act as activators and repressors of gene expression [103]. Despite their biological importance, these four classes of ncRNAs have not yet been implied in the transcriptional or post-transcriptional regulation of *Hox* genes: as such, we will hereafter focus on lncRNAs and miRNAS.

a) Long non-coding RNAs (lncRNA)

Two regulatory strategies have been described for lncRNAs-mediated control of gene expression; those being, by promotion of chromatin conformation changes and by controlling the assembly or activity of transcription factor complexes [104]. *HOTAIR* (*Hox* Antisense Intergenic RNA) is a classic example of an lncRNA that is involved in the transcriptional control of human *HOXD* genes, by induction of a closed/condensed chromatin state. This lncRNA is transcribed in an anti-sense fashion from the *HOXC locus*, on human chromosome 12, and acts in *trans* to repress the *HOXD locus*, located at the human chromosome 2 [105]. A HOTAIR 5' domain was proven to bind to SUZ12, a subunit of PRC2 (Polycomb Repressive Complex 2), while a 3' domain has the ability to bind to LSD1, a member of the CoREST/REST complex. This led to the proposal that *HOXD* repression by *HOTAIR* in human cells results from the recruitment of PRC2 and CoREST/REST to its target location, promoting the tri-methylation of histone H3 at lysine 27 (H3K27me3) by PCR2 and demethylation of H3K4me2 by CoREST/REST. This indicate that *HOTAIR* has the ability to act as a scaffold by providing histone modification enzymes with a binding site, thus facilitating *HOXD* transcriptional silencing through the establishment of specific histone modifications [104, 106-108].

Other lncRNA that regulate human *HOX* gene expression is *HOTTIP*, a *HOXA* distal transcript antisense RNA [109]. This transcript is implied in the transcriptional activation of some of the *HOXA* genes in a way that seems to be dependent on gene position relatively to the *HOTTIP* location. This lncRNA facilitates chromatin de-condensation by interaction with WDR5-MILL complexes, which promote local histone H3 lysine 4 tri-methylation (H3K4me3) [104, 107, 108]. *HOTTIP* transcription has been reported in human foreskin and foot fibroblasts [109], in mouse limbs at E13.5, and in chick limb buds, exposing a preferential expression of *HOTTIP* at distal anatomical sites where *HoxA* genes are expressed [105].

Mistral is yet another lncRNA that is transcribed from the mouse *Hoxa locus* and *in vitro* studies have proven that this lncRNA has the ability to associate to MILL1 SET by a 3' located hairpin loop [105]. *Mistral* is transcribed from the spacer DNA region separating *Hoxa6* and *Hoxa7* in mouse embryonic stem cells. This ncRNA is implied in the transcriptional activation of those two genes [110] by a mechanism that seems to involve the recruitment of WDR5-MILL complexes and consequent histone H3 lysine 4 tri-methylation (H3K4me3) [107, 108]. A third lncRNA was identified in the human *HOXA* locus, named *HOTATRM1*, which acts in the differentiation process of NB4 cells into granulocytes and is also highly expressed in leukocytes during human haematopoiesis [105].

b) MicroRNAs (miRNAs)

Like siRNAs, miRNAs are also post-transcriptional regulators of gene expression. Moreover, this type of ncRNA has also been implied in the regulation of gene expression by promoter targeting and by induction of translational activation [103]. Two examples of miRNAs that act as regulators of *Hox* gene expression are mir-10 and mir-196, located within the *Hox* clusters. These miRNAs are highly conserved among distinct vertebrate genomes, and their evolutionary history resembles that of the *Hox* clusters, suggesting a functional significance for mir-10 and mir-196 [111]. Mir-10 and mir-196 are transcribed in the same orientation as *Hox* genes and are expressed in patterns that mimic the characteristic expression of those genes, meaning that mir-10 and mir-196 expression correlates with their position within the cluster [112].

In Zebrafish the mir-10 paralogues, near the 5' genomic region of *hox4* genes, have been implied in the repression of *hoxb1a* and *hoxb3a* in their posterior expression domains in the spinal cord [111, 113]. An interesting observation that compiles evidence for the importance of mir-10 is the discovery that, in spite the loss of the *hoxdb* cluster in zebrafish, mir-10 was retained (Figure 1; [114]).

Mir-196, in chicken embryos, acts as an inhibitor of *Hoxb8* in hindlimbs, preventing its induction by ectopic RA [112]. McGlinn *et al.*, also proved that downregulation of mir-196 in chicken embryos cause an anterior expansion of *Hoxb8* expression and a concomitant homeotic transformation of the last cervical vertebra into a thoracic identity [115]. This data indicates that miRNAs are an important element in the gene regulatory machinery. However, miRNAs seem to be implied in the fine-tuning of gene expression during development rather than being a determinant factor [115].

Mir-196 is also involved in the regulation of limb development in zebrafish. He *et al.* [114] showed that overexpression of mir-196 in zebrafish embryos led to the loss of the endochondral disc and scapulacoracoid at the adult fin stage, through inhibition of the RA signalling [114]. Besides abnormal fin bud development, mir-196 overexpression also caused loss of the 6th pharyngeal arch, loss of rostral vertebrae, homeotic aberration and reduced number of ribs and somites [114]. This miRNA is also implied in fine-tuning the anterior expression border of *hoxb5a*, *hoxb5b*, *hoxb6b* and *hoxc6a*, directly or by regulation of genes that are by its turn regulators of the *Hox* genes previously referred [114].

Additionally, other miRNAs have been implied in the regulation of *Hox* genes, such as mir-99 that regulates *Hoxa1* transcription [116] and mir-126 that targets the homeobox domain of *Hoxa9*, as proven in immortalized bone marrow cells [117]. The mir-99 family is one of the most ancient microRNAs families found and its origin seems to predate the bilaterian origin [116]. Chen *et al.* [116], reported that mir-99 target two binding sequences located at human *HOXA1* mRNA, leading to down-regulation of this gene [116].

The examples presented above show that *Hox* gene regulation is controlled, by ncRNAs, at various levels. Recently, in a report by Chan *et al.* [118], it is hinted that the transcriptional regulation exerted by ncRNAs might be more complex [118]. In the paper, this team explored the expression pattern of some of the seven mouse ncRNAs, derived only from the *Hoxb3 locus*, and found that they might have multiple functions, such as regulation of protein translation, control of gene transcription by binding to enhancer sequences, and post-transcriptional suppression of *Hoxb3* sense transcripts [118]. This report suggests that *Hox* gene regulation driven by ncRNAs is probably more complex than what was previously thought and sets the focus on the importance of their characterization in different animals and/or developmental processes.

5. Concluding remarks

The data collected from multiple studies on the role of *Hox* genes during development demonstrate their unequivocal involvement in the anteroposterior determination of bilaterian structures [25]. These transcription factors act directly or indirectly upon other genes/signalling pathways modulating crucial cellular events such as cell death, affinity and proliferation, which in turn lead to the determination of body structure. Given the major importance of *Hox* genes for embryonic development, the existence of a precise transcriptional and post-transcriptional regulation, acting in a time and space specific manner is not surprising.

The first efforts to understand these mechanisms started in the late 1980s and lead to the discovery of various regulatory mechanisms, as shown in this review, that are comprised of: *cis*-regulatory elements (that act either as promoters, enhancers or insulators); *trans*-acting factors (comprising several transcription factors, *Hox* genes included, and RA); epigenetic mechanisms (such as histone modifications, chromatin remodelling, DNA methylation and nuclear positioning); and ncRNAs (such as lncRNAs and miRNAs). However, our understanding of the above-mentioned mechanisms is still incomplete, lacking comprehensive information on their mode of activation, coordination and function. Moreover, the genomic localization of the players involved in these regulatory processes remains available only for a select number of model organisms.

Concerning the *cis*-regulatory elements, the most striking observation coming out of the functional studies presented here is the complexity of these regulatory networks, which need to act in a strict coordinated way to regulate the expression of *Hox* genes throughout development. Indeed, sets of promoters and enhancers are selectively recruited in each tissue, probably involving the action of insulators, to drive expression of a particular *Hox* gene or group of genes. Exploring the *cis*-regulatory architecture of *Hox* genes across vertebrates will certainly yield invaluable insight on how *Hox*-associated CREs contributed to the current morphological characteristics of each lineage, and consequent habitat adaptation, as hinted by the works of Anand *et al.* and Di-Poi *et al.* [33, 71].

Additionally several *trans*-regulators, epigenetic mechanisms and ncRNAs have been described as crucial orchestrators of the transcriptional or post-transcriptional regulation of *Hox* genes within each cell. The information presented in this review, with a focus on vertebrates, show how these *trans*-regulatory mechanisms contribute to the expression pattern of *Hox* genes along animal body and appendages. However, it remains unexplored at what phylogenetic timepoints each of these *trans*-regulatory elements was recruited for the regulation of *Hox* genes. The characterization of putative transcriptome changes caused by different *trans*-regulatory mechanisms, may contribute to our understanding of the evolution of vertebrate features, such as the axial skeleton, jaws or limbs. It is relevant to point out the particular importance that ncRNAs may have assumed during the evolution of vertebrates, not only because of their multifaceted regulatory function, but also because of their high divergence even among related species, providing a mechanistic way to explain small body variations among closely related taxa. Moreover, epigenetic events such as DNA methylation, demethylation and remodelling may provide explanations for the collinear behaviour of the *Hox* clusters found in bilaterians.

Taken together, the information presented in this review, reveals that research on the regulation of

Hox genes has largely contributed to our overall understanding of the gene regulatory networks. This review also highlights the importance of gene regulatory mechanisms in the evolution of the developmental processes. In the future, greater understanding of their intricate regulation will certainly yield new insights on the evolution of the morphological complexity of vertebrates.

ACKNOWLEDGMENTS

This work was funded by FEDER through the Operational Competitiveness Programme – COMPETE, and by National Funds through FCT – Fundação para a Ciência e a Tecnologia, under the project FCOMP-01-0124-FEDER-041095 (EXPL/BEX-BID/0801/2013).

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to disclose.

REFERENCES

1. Lewis, E. B. 1978, Nature, 276, 565-570.
2. Montavon, T. and Soshnikova, N. 2014, Semin. Cell. Dev. Biol., 34, 76-84, doi:10.1016/j.semcdb.2014.06.005.
3. Freitas, R., Zhang, G. J. and Cohn, M. J. 2006, Nature, 442, 1033-1037.
4. Zakany, J. and Duboule, D. 2007, Curr. Opin. Genet. Dev., 17, 359-366.
5. Freitas, R., Gomez-Marin, C., Wilson, J. M., Casares, F. and Gomez-Skarmeta, J. L. 2012, Dev. Cell, 23, 1219-1229.
6. Ponting, C. P. 2008, Nat. Rev. Genet., 9, 689-698.
7. Gehring, W. J. 1993, Gene, 135, 215-221.
8. Lynch, V. J., Roth, J. J. and Wagner, G. P. 2006, BMC Evol. Biol., 6, 86.
9. Gehring, W. J. 2011, Biol. Aujourdhui, 205, 75-85.
10. Pascual-Anaya, J., D'Aniello, S., Kuratani, S. and Garcia-Fernandez, J. 2013, BMC Dev. Biol., 13, 26.
11. Dehal, P. and Boore, J. L. 2005, PLoS Biology, 3, e314.
12. Martin, K. J. and Holland, P. W. 2014, Mol. Biol. Evol., 10, 2592-2611, doi:10.1093/molbev/msu202.
13. Mehta, T. K., Ravi, V., Yamasaki, S., Lee, A. P., Lian, M. M., Tay, B. H., Tohari, S., Yanai, S., Tay, A., Brenner, S. and Venkatesh, B. 2013, P. Proc. Natl. Acad. Sci. USA, 110, 16044-16049.
14. Woltering, J. M., Noordermeer, D., Leleu, M. and Duboule, D. 2014, PLoS Biology, 12, e1001773.
15. Gouti, M., Briscoe, J. and Gavalas, A. 2011, Stem Cells, 29, 858-870.
16. Di Bonito, M., Narita, Y., Avallone, B., Sequino, L., Mancuso, M., Andolfi, G., Franze, A. M., Puelles, L., Rijli, F. M. and Studer, M. 2013, PLoS Genetics, 9, e1003249.
17. Mallo, M., Wellik, D. M. and Deschamps, J. 2010, Dev. Bio., 344, 7-15.
18. Horan, G. S., Ramirez-Solis, R., Featherstone, M. S., Wolgemuth, D. J., Bradley, A. and Behringer, R. R. 1995, Genes & Dev., 9, 1667-1677.
19. McIntyre, D. C., Rakshit, S., Yallowitz, A. R., Loken, L., Jeannotte, L., Capecchi, M. R. and Wellik, D. M. 2007, Development, 134, 2981-2989.
20. Vinagre, T., Moncaut, N., Carapuco, M., Novoa, A., Bom, J. and Mallo, M. 2010, Dev. Cell., 18, 655-661.
21. Yu, H., Lindsay, J., Feng, Z. P., Frankenberg, S., Hu, Y., Carone, D., Shaw, G., Pask, A. J., O'Neill, R., Papenfuss, A. T. and Renfree, M. B. 2012, BMC Genomics, 13, 251.
22. Villavicencio-Lorini, P., Kuss, P., Friedrich, J., Haupt, J., Farooq, M., Turkmen, S., Duboule, D., Hecht, J. and Mundlos, S. 2010, J. Clin. Invest., 120, 1994-2004.
23. Freitas, R., Gomez-Skarmeta, J. L. and Rodrigues, P. N. 2014, J. Exp Zool. Part B, 322(7), 540-552, doi:10.1002/jez.b.22563.
24. Monteiro, A. S. and Ferrier, D. E. 2006, Int. J. Biol. Sci., 2, 95-103.
25. Sanchez-Herrero, E. 2013, Scientifica, 2013, 738257.
26. Tschopp, P., Tarchini, B., Spitz, F., Zakany, J. and Duboule, D. 2009, PLoS Genetics, 5, e1000398.
27. Pascual-Anaya, J., Adachi, N., Alvarez, S., Kuratani, S., D'Aniello, S. and Garcia-Fernandez, J. 2012, EvoDevo, 3, 28.
28. Monteiro, A. S., Freitas, R. and Palmeirim, I. 2014, J. Appl. Ichthyol., 30, 689-695.

29. Charite, J., de Graaff, W., Consten, D., Reijnen, M. J., Korving, J. and Deschamps, J. 1998, Development, 125, 4349-4358.
30. Soshnikova, N. 2014, Dev. Dyn., 243, 49-58.
31. Kondrashov, N., Pusic, A., Stumpf, C. R., Shimizu, K., Hsieh, A. C., Xue, S., Ishijima, J., Shiroishi, T. and Barna, M. 2011, Cell, 145, 383-397.
32. Woltering, J. M. 2012, Curr. Genomics, 13, 289-299.
33. Di-Poi, N., Montoya-Burgos, J. I., Miller, H., Pourquie, O., Milinkovitch, M. C. and Duboule, D. 2010, Nature, 464, 99-103.
34. Chambeyron, S. and Bickmore, W. A. 2004, Genes & Dev., 18, 1119-1130.
35. Morey, C., Da Silva, N. R., Perry, P. and Bickmore, W. A. 2007, Development, 134, 909-919.
36. van Dijk, E. L., Auger, H., Jaszczyszyn, Y. and Thermes, C. 2014, Trends Genet., 30(9), 418-426, doi:10.1016/j.tig.2014.07.001.
37. Andrey, G. and Duboule, D. 2014, Cell, 156, 856-856 e851.
38. Duboule, D. 2007, Development, 134, 2549-2560.
39. Popperl, H., Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., Mann, R. S. and Krumlauf, R. 1995, Cell, 81, 1031-1042.
40. Geada, A. M., Gaunt, S. J., Azzawi, M., Shimeld, S. M., Pearce, J. and Sharpe, P. T. 1992, Development, 116, 497-506.
41. Schughart, K., Bieberich, C. J., Eid, R. and Ruddle, F. H. 1991, Development, 112, 807-811.
42. Puschel, A. W., Balling, R. and Gruss, P. 1991, Development, 112, 279-287.
43. Renucci, A., Zappavigna, V., Zakany, J., Izpisua-Belmonte, J. C., Burki, K. and Duboule, D. 1992, The EMBO Journal, 11, 1459-1468.
44. Coulombe, Y., Lemieux, M., Moreau, J., Aubin, J., Joksimovic, M., Berube-Simard, F. A., Tabaries, S., Boucherat, O., Guillou, F., Larochelle, C., Tuggle, C. K. and Jeannotte, L. 2010, PLoS One, 5, e10600.
45. Folberg, A., Kovacs, E. N. and Featherstone, M. S. 1997, J. Biol. Chem., 272, 29151-29157.
46. Kwan, C. T., Tsang, S. L., Krumlauf, R. and Sham, M. H. 2001, Dev. Biol., 232, 176-190.
47. Sham, M. H., Hunt, P., Nonchev, S., Papalopulu, N., Graham, A., Boncinelli, E. and Krumlauf, R. 1992, The EMBO Journal, 11, 1825-1836.
48. Galliot, B., Dolle, P., Vigneron, M., Featherstone, M. S., Baron, A. and Duboule, D. 1989, Development, 107, 343-359.
49. Gutman, A., Gilthorpe, J. and Rigby, P. W. 1994, Mol. Cell Biol., 14, 8143-8154.
50. Coletta, P. L., Shimeld, S. M., Chaudhuri, C., Muller, U., Clarke, J. P. and Sharpe, P. T. 1991, Mech. Dev., 35, 129-142.
51. Nolte, C., Amores, A., Nagy Kovacs, E., Postlethwait, J. and Featherstone, M. 2003, Mech. Dev., 120, 325-335.
52. Cianetti, L., Di Cristofaro, A., Zappavigna, V., Bottero, L., Boccoli, G., Testa, U., Russo, G., Boncinelli, E. and Peschle, C. 1990, Nucleic. Acids Res., 18, 4361-4368.
53. Arcioni, L., Simeone, A., Guazzi, S., Zappavigna, V., Boncinelli, E. and Mavilio, F. 1992, The EMBO Journal, 11, 265-277.
54. Hadrys, T., Punnamoottil, B., Pieper, M., Kikuta, H., Pezeron, G., Becker, T. S., Prince, V., Baker, R. and Rinkwitz, S. 2006, Dev. Biol., 297, 26-43.
55. Mavilio, F., Simeone, A., Giampaolo, A., Faiella, A., Zappavigna, V., Acampora, D., Poiana, G., Russo, G., Peschle, C. and Boncinelli, E. 1986, Nature, 324, 664-668.
56. Bieberich, C. J., Utset, M. F., Awgulewitsch, A. and Ruddle, F. H. 1990, P. Proc. Natl. Acad. Sci. USA, 87, 8462-8466.
57. Simeone, A., Pannese, M., Acampora, D., D'Esposito, M. and Boncinelli, E. 1988, Nucleic Acids Res., 16, 5379-5390.
58. Shlyueva, D., Stampfel, G. and Stark, A. 2014, Nat. Rev. Genet., 15, 272-286.
59. Frasch, M., Chen, X. and Lufkin, T. 1995, Development, 121, 957-974.
60. Tumpel, S., Cambronero, F., Sims, C., Krumlauf, R. and Wiedemann, L. M. 2008, P. Proc. Natl. Acad. Sci. USA, 105, 20077-20082.
61. Maconochie, M., Krishnamurthy, R., Nonchev, S., Meier, P., Manzanares, M., Mitchell, P. J. and Krumlauf, R. 1999, Development, 126, 1483-1494.

62. Nonchev, S., Maconochie, M., Vesque, C., Aparicio, S., Ariza-McNaughton, L., Manzanares, M., Maruthainar, K., Kuroiwa, A., Brenner, S., Charnay, P. and Krumlauf, R. 1996, P. Proc. Natl. Acad. Sci. USA, 93, 9339-9345.
63. Sham, M. H., Vesque, C., Nonchev, S., Marshall, H., Frain, M., Gupta, R. D., Whiting, J., Wilkinson, D., Charnay, P. and Krumlauf, R. 1993, Cell, 72, 183-196.
64. Studer, M., Gavalas, A., Marshall, H., Ariza-McNaughton, L., Rijli, F. M., Chambon, P. and Krumlauf, R. 1998, Development, 125, 1025-1036.
65. Maconochie, M. K., Nonchev, S., Studer, M., Chan, S. K., Popperl, H., Sham, M. H., Mann, R. S. and Krumlauf, R. 1997, Genes & Dev., 11, 1885-1895.
66. Gould, A., Itasaki, N. and Krumlauf, R. 1998, Neuron, 21, 39-51.
67. Gould, A., Morrison, A., Sproat, G., White, R. A. and Krumlauf, R. 1997, Genes & Dev., 11, 900-913.
68. Morrison, A., Ariza-McNaughton, L., Gould, A., Featherstone, M. and Krumlauf, R. 1997, Development, 124, 3135-3146.
69. Zhang, F., Nagy Kovacs, E. and Featherstone, M. S. 2000, Mech. Dev., 96, 79-89.
70. Pasqualetti, M., Ori, M., Nardi, I. and Rijli, F. M. 2000, Development, 127, 5367-5378.
71. Anand, S., Wang, W. C., Powell, D. R., Bolanowski, S. A., Zhang, J., Ledje, C., Pawashe, A. B., Amemiya, C. T. and Shashikant, C. S. 2003, P. Proc. Natl. Acad. Sci. USA, 100, 15666-15669.
72. Montavon, T. and Duboule, D. 2013, Philos. Trans. R. Soc. Lond. B Biol. Sci., 368, 20120367.
73. Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F. and Duboule, D. 2011, Cell, 147, 1132-1145.
74. Yamagishi, T., Ozawa, M., Ohtsuka, C., Ohyama-Goto, R. and Kondo, T. 2007, PLoS One, 2, e175.
75. Kmita, M., Kondo, T. and Duboule, D. 2000, Nature Genetics, 26, 451-454.
76. Kmita, M., Tarchini, B., Duboule, D. and Herault, Y. 2002, Development, 129, 5521-5528.
77. Vasanthi, D., Anant, M., Srivastava, S. and Mishra, R. K. 2010, Development, 137, 4239-4247.
78. Srivastava, S., Puri, D., Garapati, H. S., Dhawan, J. and Mishra, R. K. 2013, Epigenetics Chromatin, 6, 8.
79. Li, H., Sheridan, R. and Williams, T. 2013, Human Molecular Genetics, 22, 3195-3206.
80. Desmazieres, A., Charnay, P. and Gilardi-Hebenstreit, P. 2009, J. Biol. Chem., 284, 10831-10840.
81. Nonchev, S., Vesque, C., Maconochie, M., Seitanidou, T., Ariza-McNaughton, L., Frain, M., Marshall, H., Sham, M. H., Krumlauf, R. and Charnay, P. 1996, Development, 122, 543-554.
82. Manzanares, M., Cordes, S., Ariza-McNaughton, L., Sadl, V., Maruthainar, K., Barsh, G. and Krumlauf, R. 1999, Development, 126, 759-769.
83. Young, T., Rowland, J. E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J. N., Beck, F., Mallo, M. and Deschamps, J. 2009, Dev. Cell, 17, 516-526.
84. Subramanian, V., Meyer, B. I. and Gruss, P. 1995, Cell, 83, 641-653.
85. van den Akker, E., Forlani, S., Chawengsaksophak, K., de Graaff, W., Beck, F., Meyer, B. I. and Deschamps, J. 2002, Development, 129, 2181-2193.
86. Duester, G. 2008, Cell, 134, 921-931.
87. Shimozono, S., Iimura, T., Kitaguchi, T., Higashijima, S. and Miyawaki, A. 2013, Nature, 496, 363-366.
88. Zhao, X., Sirbu, I. O., Mic, F. A., Molotkova, N., Molotkov, A., Kumar, S. and Duester, G. 2009, Curr. Biol., 19, 1050-1057.
89. Lara-Ramirez, R., Zieger, E. and Schubert, M. 2013, Int. J. Bioch. & Cell Biol., 45, 1302-1313.
90. Barber, B. A. and Rastegar, M. 2010, Ann. Anat., 192, 261-274.
91. Weinhold, B. 2006, Environ. Health Perspect., 114, A160-167.
92. Li, X., Isono, K., Yamada, D., Endo, T. A., Endoh, M., Shinga, J., Mizutani-Koseki, Y., Otte, A. P., Casanova, M., Kitamura, H., Kamijo, T., Sharif, J., Ohara, O., Toyada, T.,

Bernstein, B. E., Brockdorff, N. and Koseki, H. 2011, Mol. Cell Biol., 31, 351-364.
93. Casaca, A., Santos, A. C. and Mallo, M. 2014, Dev. Dyn., 243, 24-36.
94. Kim, Y. J., Cecchini, K. R. and Kim, T. H. 2011, P. Proc. Natl. Acad. Sci. USA, 108, 7391-7396.
95. Zhang, Y., Liu, Z., Medrzycki, M., Cao, K. and Fan, Y. 2012, PLoS One, 7, e38829.
96. Branciamore, S., Chen, Z. X., Riggs, A. D. and Rodin, S. N. 2010, P. Proc. Natl. Acad. Sci. USA, 107, 15485-15490.
97. Terranova, R., Agherbi, H., Boned, A., Meresse, S. and Djabali, M. 2006, P. Proc. Natl. Acad. Sci. USA, 103, 6629-6634.
98. Ng, R. K., Kong, C. T., So, C. C., Lui, W. C., Chan, Y. F., Leung, K. C., So, K. C., Tsang, H. M., Chan, L. C. and Sham, M. H. 2014, J. Pathol., 232, 65-74.
99. Marzese, D. M., Scolyer, R. A., Huynh, J. L., Huang, S. K., Hirose, H., Chong, K. K., Kiyohara, E., Wang, J., Kawas, N. P., Donovan, N. C., Hata, K., Wilmott, J. S., Murali, R., Buckland, M. E., Shivalingam, B., Thompson, J. F., Morton, D. L., Kelly, D. F. and Hoon, D. S. 2014, Hum. Mol. Genet., 23, 226-238.
100. Xavier, F. C., Destro, M. F., Duarte, C. M. and Nunes, F. D. 2014, Arch. Oral Biol., 59, 783-789.
101. Sakamoto, K., Onimaru, K., Munakata, K., Suda, N., Tamura, M., Ochi, H. and Tanaka, M. 2009, PLoS One, 4, e5121.
102. Saxena, A. and Carninci, P. 2011, BioEssays, 33, 830-839.
103. Kaikkonen, M. U., Lam, M. T. and Glass, C. K. 2011, Cardiovasc. Res., 90, 430-440.
104. Kugel, J. F. and Goodrich, J. A. 2012, Trends Biochem. Sci., 37, 144-151.
105. Wang, X. Q., Crutchley, J. L. and Dostie, J. 2011, Curr. Genomics, 12, 307-321.
106. Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E. and Chang, H. Y. 2010, Science, 329, 689-693.
107. Faust, T., Frankel, A. and D'Orso, I. 2012, Transcription, 3, 78-86.
108. Hu, W., Alvarez-Dominguez, J. R. and Lodish, H. F. 2012, EMBO Rep., 13, 971-983.
109. Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B. R., Protacio, A., Flynn, R. A., Gupta, R. A., Wysocka, J., Lei, M., Dekker, J., Helms, J. A. and Chang, H. Y. 2011, Nature, 472, 120-124.
110. Bertani, S., Sauer, S., Bolotin, E. and Sauer, F. 2011, Mol. Cell, 43, 1040-1046.
111. Olena, A. F. and Patton, J. G. 2010, J. Cell. Phys., 222, 540-545.
112. Yekta, S., Tabin, C. J. and Bartel, D. P. 2008, Nat. Rev. Genet., 9, 789-796.
113. Woltering, J. M. and Durston, A. J. 2008, PLoS One, 3, E1396.
114. He, X., Yan, Y. L., Eberhart, J. K., Herpin, A., Wagner, T. U., Schartl, M. and Postlethwait, J. H. 2011, Dev. Biol., 357, 463-477.
115. McGlinn, E., Yekta, S., Mansfield, J. H., Soutschek, J., Bartel, D. P. and Tabin, C. J. 2009, P. Proc. Natl. Acad. Sci. USA, 106, 18610-18615.
116. Chen, D., Chen, Z., Jin, Y., Dragas, D., Zhang, L., Adjei, B. S., Wang, A., Dai, Y. and Zhou, X. 2013, PLos One, 8, E80625.
117. Shen, W. F., Hu, Y. L., Uttarwar, L., Passegue, E. and Largman, C. 2008, Mol. Cell. Biol., 28, 4609-4619.
118. Chan, K. T., Qi, J. and Sham, M. H. 2010, Bioch. Bioph. Res. Comm., 398, 153-159.
119. Chopra, V. S. 2011, Cell. Mol. Life Sciences, 68, 977-990.
120. Prince, V. E. and Pickett, F. B. 2002, Nat. Rev. Genetics, 3, 827-837.
121. Amemiya, C. T., Prohaska, S. J., Hill-Force, A., Cook, A., Wasserscheid, J., Ferrier, D. E., Pascual-Anaya, J., Garcia-Fernandez, J., Dewar, K. and Stadler, P. F. 2008, J. Exp. Zool. Part B, 310, 465-477.
122. Tehler, D., Hoyland-Kroghsbo, N. M. and Lund, A. H. 2011, RNA Biology, 8, 728-734.