ELSEVIER

Seminars in

CELL & DEVELOPMENTAL BIOLOGY

HOX GENES: THE ORIGINAL BODY BUILDERS

Guest Editor: RICHARD S MANN

Volume 152-153
January/February 2024
ISSN 1084-9521

seminars in CELL & DEVELOPMENTAL BIOLOGY

Editorial correspondence should be addressed to: *Seminars in Cell & Developmental Biology*, Elsevier, Stover Court, Bampfylde Street, Exeter, EX1 2AH, UK, e-mail: semcell@elsevier.com.

Please note that this journal is a commissioned review journal and does not accept unsolicited articles. If you are interested in acting as a guest editor, please contact the Editorial Office: semcell@elsevier.com

For a full and complete Guide for Authors, please go to: http://www.elsevier.com/locate/semcdb

Author inquiries: You can track your submitted article at http://www.elsevier.com/track-submission. You can track your accepted article at http://www.elsevier.com/trackarticle. You are also welcome to contact Customer Support via http://service.elsevier.com

Abstracted/indexed in *Index Medicus,* MEDLINE on the MEDLARS system. Also covered in the abstract and citation database Scopus®. Full text available on ScienceDirect®

The home page for *Seminars in Cell & Developmental Biology* can be found at http://www.elsevier.com/locate/semcdb

Processed at Thomson Digital, Gangtok (India)

Seminars in Cell and Developmental Biology has no page charges.

Seminars in Cell and Developmental Biology 152-153 (2024) 1–3

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Hox genes: The original body builders

As biologists, we are constantly fascinated by the diversity and complexity of living organisms. To understand the origins of diversity in the animal kingdom we must understand animal development, and perhaps the group of genes that has most consistently captured the attention of developmental biologists since they were discovered are the Hox genes. These genes, which encode highly conserved homeobox-containing transcription factors, are present in a wide range of organisms, from fruit flies to humans. First discovered in the fruit fly *Drosophila melanogaster*, they were originally appreciated as playing a crucial role in determining the body plan of an organism [1]. And yet, research on Hox genes has gone far beyond animal development by informing at least two additional areas of biology. For one, they are critical drivers of animal evolution: changes to how and when they are deployed, and in the networks of downstream genes they regulate, have facilitated changes to animal body plans [2,3]. Hox gene research has also shed light on how families of related transcription factors with very similar DNA binding specificities can carry out distinct functions in vivo [4]. Arguably, no other set of genes have had such an important impact on such disparate and important areas of biology.

In this Special Issue, we present a wide range of articles that reflect all three of the fields on which Hox research has had a profound impact: animal development, animal evolution, and transcription factor mechanisms. For evolutionary insights, we have three fascinating articles. The first, by Mulhair and Holland [5, this issue], builds on the intriguing observation that most Hox genes are clustered in animal genomes and that their expression along the main body axis correlates with their position within these clusters. Mulhair and Holland's contribution is a tour-de-force effort that uses publicly available genome sequences for no less than 243 insects, representing 13 orders, to analyze trends in the cluster-level organization of these genes. Large order-specific differences in Hox cluster size, organization, and the duplication, loss, and emergence of new homeobox genes (e.g. the explosion of *zen* orthologs in *Lepidoptera*) suggests that Hox genes have many species-specific functions and modes of regulation that are yet to be discovered. The article by Wanninger [6, this issue] addresses the evolutionary origins of Hox genes and the relationship of Hox gene number to animal complexity. Wanninger first provides an analysis of when Hox genes emerged and lost during evolution by depicting several different scenarios that can account for the currently available sequence data. One conclusion is that rather than relying only on gene expression to determine the evolution of morphological characters, it is better to include datasets of comparative morphology and gene-gene interactions. Third, Turetzek et al. [7, this issue] take a deep dive into the organization and expression of Hox genes in spiders. Spiders, with their distinct body plans relative to better studied arthropods such as fruit flies, provide the opportunity to ask if body plan modifications do indeed correlate with changes in Hox expression and gene number. The answer, based largely on the observation that spiders have two Hox clusters and multiple divergent Hox expression patterns, is almost certainly yes. Despite these differences, some Hox-dependent functions, for example, suppression of legs in the abdomen, are likely to be conserved between spiders and flies.

Hox genes were originally called homeotic genes when first described in *Drosophila* due to their dramatic ability to transform an entire appendage or segment from one identity to another [8]. These transformations of body parts were reminiscent of natural variations in animal body morphologies first described by William Bateson, who initially coined the term 'homeosis' [9]. For example, the *Drosophila* Hox gene *Ultrabithorax* (*Ubx*) normally dictates the fate of the third thoracic segment and, when its function is removed genetically, that segment is transformed to a second copy of the second thoracic segment, resulting in flies with two pairs of wings instead of one (as shown in the cover image of this issue) [10]. Remarkably, analogous transformations of segment identity have been observed in many other animals where loss-of-function genetics is feasible [11–13], suggesting that Hox segment identity functions are ancient and highly conserved. However, what was not initially obvious from these dramatic transformations is that Hox genes also play a pivotal role in specifying the identities of non-ectodermal tissues, including the nervous system and mesoderm. Three articles in this issue explore these Hox-dependent functions. In the article by Pinto et al. [14, this issue], an approach based largely on genome-wide studies is used to compare the targets, cofactors, and specificity mechanisms (see more below) used by Ubx in the *Drosophila* mesoderm and ectoderm. Two articles examine the role of Hox genes in the nervous system of two very different model systems: vertebrates and the worm, *C. elegans*. In the article by Smith and Kratsios [15, this issue], multiple Hox-dependent neuronal fate examples are described in the worm that lead to novel insights. For example, in addition to their role in neuron specification and development, Hox function is also required in terminally differentiated neurons to maintain their fates. At the opposite end of the complexity spectrum, Miller and Dasen [16, this issue] summarize the current state of our understanding for how Hox genes are themselves regulated in vertebrates, initially by broad gradients of retinoic acid and fibroblast growth factors, and later maintained by the Polycomb Group (PcG) of regulators. Notably, the PcG-mediated maintenance of Hox expression patterns in vertebrates fits well with the terminal selector functions highlighted by Smith and Kratsios in worms.

Hox proteins have also been illustrative in explaining how individual members of transcription factor families can execute highly distinct

https://doi.org/10.1016/j.semcdb.2023.05.005

Available online 7 June 2023

functions in vivo while having very similar DNA binding specificities in vitro, a phenomenon that has been called the transcription factor specificity paradox. One solution to this paradox is that Hox proteins bind with cofactors that increase differences in DNA binding preferences [4]. The article by Pinto et al. [14, this issue], suggest that cofactor-based mechanisms to achieve specificity may differ between tissue types, such as mesoderm and ectoderm, and argue that there are likely many more Hox cofactors than we currently know about. The article by Merabet and Carnesecchi [17, this issue] highlight another, not mutually exclusive, mechanism that is also likely address the Hox specificity paradox, namely, that differences in Hox expression levels can impact their function. Although it has been long appreciated that Hox function is impacted by gene dose (a proxy for expression level), this article makes the point that this is a much more widespread phenomenon than was generally appreciated and could contribute to the evolution of many Hox-regulated morphologies. In addition to providing an updated review of the cofactor model for Hox specificity, the article by Bobola and Sagerström [18, this issue] underscores the important point that the classic Hox cofactors – the TALE (three amino acid loop extension) homeodomain proteins such as Meis and Pbx – also cooperate with many non-Hox transcription factors. Although it was long known that TALE proteins have non-Hox functions [19–21], Bobola and Sagerström suggest that it may be appropriate to flip the traditional view: it may be more accurate to say that Hox proteins are cofactors for the TALE transcription factors. One of the arguments in favor of this view is that the TALE factors are bound to chromatin prior to Hox protein expression, suggesting that the TALE factors may be acting as pioneer transcription factors, which are able to bind to sites initially made inaccessible by nucleosomes. Interestingly, and along the same lines, the article by Paul et al. [22, this issue] points out that some Hox proteins can themselves act as pioneer factors, and that differences in the pioneering ability between Hox proteins may also impact binding specificity in vivo. Notably, and consistent with the view put forth by Bobola and Sagerström, the pioneering activity of some Hox proteins is dependent on TALE transcription factors. Last but not least, the article by Salomone et al. [23, this issue] also summarizes the cofactor models for Hox specificity, but emphasizes that there is often a trade-off between specificity and affinity: the most specific Hox-TALE binding sites tend to be low affinity. This raises the question of how low affinity binding sites can be sufficiently bound in nuclei, which typically do not have high concentrations of transcription factors. One likely solution is that the distribution of Hox proteins within nuclei – as probably the case for most transcription factors – is non-uniform and concentrated in local hubs, providing high local concentrations. For many non-Hox transcription factors, these hubs have been shown to form via interactions between intrinsically disordered regions (IDRs) that can form liquid-liquid phase separated condensates [24]. Interestingly, as pointed out by Salomone et al., intriguing evidence that this mechanism is relevant to Hox proteins comes from the characterization of Hox mutations that lead to both alterations in their IDRs and to human diseases, such as cancer and synpolydactyly.

As a group, these ten articles summarize multiple important and novel insights stemming from research into Hox genes that help inform a wide variety of questions currently debated by biologists. Although many questions remain and will require future research to answer, it is particularly noteworthy and satisfying that whatever insights have been obtained thus far – ranging from deep evolutionary questions to the mechanisms underlying human diseases – ultimately stem from the instinctive curiosity of *Drosophila* geneticists who first discovered these fascinating genes and who never could have imagined where their initial discoveries would lead.

Cover image caption:

As first discovered by E. B. Lewis, loss of function mutations in the Hox gene *Ubx* results in flies with a nearly complete duplication of the second thoracic segment, resulting in two pairs of wings (right), in contrast to wild type flies (left) with only one pair of wings. Photograph credit to Nicholas Gompel (http://gompel.org/).

References

[1] E.B. Lewis, A gene complex controlling segmentation in Drosophila, Nature 276 (1978) 565–570, https://doi.org/10.1128/AAC.03728-14.

[2] M. Mallo, Shaping Hox gene activity to generate morphological diversity across vertebrate phylogeny, Essays Biochem 66 (2022) 717–726, https://doi.org/10.1042/EBC20220050.

[3] E. Jarvis, H.S. Bruce, N.H. Patel, Evolving specialization of the arthropod nervous system, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 10634–10639, https://doi.org/10.1073/pnas.1201876109.

[4] S. Merabet, R.S. Mann, To be specific or not: the critical relationship between hox and TALE proteins, Trends Genet. 32 (2016) 334–347, https://doi.org/10.1016/j.tig.2016.03.004.

[5] P.O. Mulhair, P.W.H. Holland, Evolution of the insect Hox gene cluster: comparative analysis across 243 species, Semin. Cell Dev. Biol. (2022), https://doi.org/10.1016/j.semcdb.2022.11.010 (This issue).

[6] A. Wanninger, Hox, homology, and parsimony: an organismal perspective, Semin. Cell Dev. Biol. (2023), https://doi.org/10.1016/j.semcdb.2023.01.007 (This issue).

[7] N. Turetzek, M. Pechmann, R. Janssen, N.M. Prpic, Hox genes in spiders: their significance for development and evolution, Semin. Cell Dev. Biol. (2023), https://doi.org/10.1016/j.semcdb.2022.11.014 (This issue).

[8] W.J. Gehring, Y. Hiromi, Homeotic genes and the homeobox, Ann. Rev. Genet. 20 (1986) 147–173. www.annualreviews.org (accessed May 8, 2023).

[9] W. Bateson, C.J. Clay, F.H. Pierpont, Materials for the Study of Variation Treated with Especial Regard to Discontinuity in the Origin of Species, Macmillan, London, 1894. ⟨https://www.biodiversitylibrary.org/item/16204⟩.

[10] G. Halder, P. Callaerts, W.J. Gehring, Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila, Science 267 (1995) 1788–1792, https://doi.org/10.1126/science.7892602.

[11] D.M. Liubicich, J.M. Serano, A. Pavlopoulos, Z. Kontarakis, M.E. Protas, E. Kwan, S. Chatterjee, K.D. Tran, M. Averof, N.H. Patel, Knockdown of Parhyale Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 13892–13896, https://doi.org/10.1073/pnas.0903105106.

[12] Tomoyasu Yoshinori, Wheeler Scott, Denell Robin, Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum, Nature 433 (2005) 638–643, https://doi.org/10.1038/nature03127.1.

[13] D.M. Wellik, Hox patterning of the vertebrate axial skeleton, Dev. Dyn. 236 (2007) 2454–2463, https://doi.org/10.1002/dvdy.21286.

[14] P.B. Pinto, K. Domsch, I. Lohmann, Hox function and specificity – a tissue centric view, Semin. Cell Dev. Biol. (2022), https://doi.org/10.1016/j.semcdb.2022.11.011 (This issue).

[15] J.J. Smith, P. Kratsios, Hox gene functions in the C. elegans nervous system: from early patterning to maintenance of neuronal identity, Semin. Cell Dev. Biol. (2022), https://doi.org/10.1016/j.semcdb.2022.11.012 (This issue).

[16] A. Miller, J.S. Dasen, Establishing and maintaining Hox profiles during spinal cord development, Semin. Cell Dev. Biol. (2023), https://doi.org/10.1016/j.semcdb.2023.03.014 (This issue).

[17] S. Merabet, J. Carnesecchi, Hox dosage and morphological diversification during development and evolution, Semin. Cell Dev. Biol. (2022), https://doi.org/10.1016/j.semcdb.2022.11.009.

[18] N. Bobola, C.G. Sagerström, TALE transcription factors: cofactors no more, Semin. Cell Dev. Biol. (2022), https://doi.org/10.1016/j.semcdb.2022.11.015.

[19] S. Aldaz, G. Morata, N. Azpiazu, Patterning function of homothorax/extradenticle in the thorax of Drosophila, Development 132 (2005) 439–446, https://doi.org/10.1242/dev.01585.

[20] F. Casares, R.S. Mann, Control of antennal versus leg development in Drosophila, Nature 392 (1998) 723–726, https://doi.org/10.1038/33706.

[21] C.B. Moens, L. Selleri, Hox cofactors in vertebrate development, Dev. Biol. 291 (2006) 193–206, https://doi.org/10.1016/j.ydbio.2005.10.032.

[22] R. Paul, R. Peraldi, M. Kmita, The pioneering function of the hox transcription factors, Semin. Cell Dev. Biol. (2023), https://doi.org/10.1016/j.semcdb.2022.11.013 (This issue).

[23] J. Salomone, E. Farrow, B. Gebelein, Homeodomain complex formation and biomolecular condensates in Hox gene regulation, Semin. Cell Dev. Biol. (2022), https://doi.org/10.1016/j.semcdb.2022.11.016 (This issue).

[24] S. Brodsky, T. Jana, N. Barkai, Order through disorder: the role of intrinsically disordered regions in transcription factor binding specificity, Curr. Opin. Struct. Biol. 71 (2021) 110–115, https://doi.org/10.1016/j.sbi.2021.06.011.

Richard S. Mann[a,b,c,*], William J. Glassford[a,b,c]

[a] *Department of Biochemistry and Molecular Biophysics, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, United States*

[b] *Department of Neuroscience, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, United States*

[c] *Department of Systems Biology, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, United States*

[*] Correspondig author at: Department of Biochemistry and Molecular Biophysics, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, United States.
E-mail address: rsm10@columbia.ecu (R.S. Mann).

Seminars in Cell and Developmental Biology 152-153 (2024) 85–92

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

The pioneering function of the hox transcription factors

Rachel Paul [a,b,1], Rodrigue Peraldi [a,c,1], Marie Kmita [a,b,c,*]

[a] Genetics and Development Research Unit, Institut de Recherches Cliniques de Montréal, Montréal, Québec, Canada
[b] Department of Experimental Medicine, McGill University, Montréal, Québec, Canada
[c] Department of Medicine, Université de Montréal, Montreal, Quebec, Canada

ARTICLE INFO

Keywords:
Hox genes
Pioneer factor
Patterning
Chromatin accessibility
Embryonic development

ABSTRACT

Ever since the discovery that the Hox family of transcription factors establish morphological diversity in the developing embryo, major efforts have been directed towards understanding Hox-dependent patterning. This has led to important discoveries, notably on the mechanisms underlying the collinear expression of Hox genes and Hox binding specificity. More recently, several studies have provided evidence that Hox factors have the capacity to bind their targets in an inaccessible chromatin context and trigger the switch to an accessible, transcriptional permissive, chromatin state. In this review, we provide an overview of the evidences supporting that Hox factors behave as pioneer factors and discuss the potential mechanisms implicated in Hox pioneer activity as well as the significance of this functional property in Hox-dependent patterning.

1. Introduction

The Hox family of transcription factors have a pivotal role in establishing morphological diversity along the main axis of developing embryos in all animal species with bilateral symmetry [1–3]. Hox genes were first identified in Drosophila, with the study of mutations where body structures were developing at the wrong position. In 1978, Ed Lewis discovered that Hox genes are clustered and that their order on the chromosome is collinear with their functional domain along the anterior-posterior axis [4]. Hox collinearity was subsequently uncovered in other animals, including vertebrates [5–7]. In addition to the spatial Hox collinearity, in animals developing with an anterior to posterior temporal progression, such as vertebrates, Hox genes are sequentially activated in time, from one end of the cluster (3′ end) to the other end (5′end). This phenomenon has been referred to as Hox temporal collinearity [8,9]. Such temporal sequence of Hox gene activation allows a coordination between the progressive formation of axial tissues and Hox expression [10–12]. Reminiscent of the situation in the main body axis, the collinear expression of Hox genes also contribute to the proper patterning of appendages, as exemplified with the critical role of the HoxA and HoxD gene clusters in the morphogenesis of the tetrapod limb [reviewed in e.g. 13].

While the differential expression of *Hox* genes represents the initial step in the Hox patterning process, an equally important aspect is the binding specificity of the highly conserved HOX transcription factors (TFs). Strikingly, series of in vitro experiments revealed that HOX TFs recognize similar AT-rich motifs [14–17]. Such discrepancy between the similar DNA-binding motif uncovered in vitro and the specific functions observed in vivo, referred to as the HOX paradox, has been reconciled with the evidence that cofactors contribute to Hox binding specificity [16,18]. The best characterized HOX cofactors are the TALE (three amino acid loop extension) homeodomain proteins [19–27]. In Drosophila, TALE cofactors are Extradenticle (Exd) and Homothorax (Hth) and in the mouse, the Exd-related proteins are Pbx1–4 and the Hth-related proteins are Meis1–3 and Prep1–2 [16]. These proteins have the ability to bind with Hox proteins in a highly cooperative manner. The identification and characterization of HOX-response elements associated with auto-, para-, and cross-regulatory interactions between HOX proteins have revealed that bipartite HOX-PBX sites are commonly used for HOX binding and functional activities [16,17,28–31]. Slattery and colleagues [17], using SELEX-seq experiment (Systematic Evolution of Ligands by Exponential Enrichment with massively parallel sequencing) to analyze the binding preferences of the eight Drosophila Hox proteins revealed that when monomers are compared two by two, they bind to the same sequences with relatively similar affinity while heterodimerization with Exd results in distinct binding specificities. Such cofactor dependent Hox binding specificity was referred to as latent specificity [17].

* Corresponding author at: Genetics and Development Research Unit, Institut de Recherches Cliniques de Montréal, Montréal, Québec, Canada.
E-mail address: Marie.Kmita@ircm.qc.ca (M. Kmita).
[1] Equal contribution

https://doi.org/10.1016/j.semcdb.2022.11.013
Received 31 August 2022; Received in revised form 13 November 2022; Accepted 30 November 2022
Available online 12 December 2022

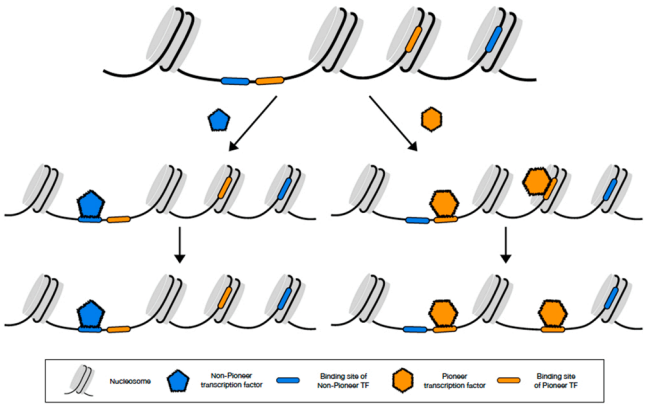


Fig. 1. Difference between pioneer and non-pioneer transcription factors in transcriptional regulation. (Top) Regulatory region containing binding sites for pioneer (orange) and non-pioneer transcription (blue) factors, located in accessible chromatin region and in nucleosome inclusive region. (Middle) While non-pioneer transcription factors can only access their binding sites in accessible regions (nucleosome depleted), pioneer factors can bind their targets both in nucleosome-depleted and nucleosome-inclusive region. (Bottom) After binding, only pioneer factors can locally open previously inaccessible chromatin, leading to new nucleosome-depleted regions. Chromatin opening is achieved either by nucleosome eviction or nucleosome sliding.

Binding affinity also contribute to Hox binding specificity. Transcription factors display a wide range of affinities to their cognate DNA binding motifs, usually including both the high affinity sites and a wide spectrum of lower affinity sites. It was shown that low-affinity binding sites for the Hox transcription factor Ultrabithorax (Ubx) in the Drosophila *shavenbaby* enhancers confer specificity to Ubx binding in vivo [32]. Indeed, substituting these low affinity sites for high-affinity sites led to ectopic expression in embryonic tissues that do not express Ubx, most likely through the binding of other homeodomain-containing transcription factors. High-affinity Hox binding sites can be bound by multiple Hox TFs and potentially by many homeodomain-containing genes [14,15]. In contrast, low-affinity binding sites provide specificity for individual HOX factors [32].

Another facet to consider in understanding how each Hox TF controls a specific set of targets, is the chromatin accessibility landscape. Indeed, transcription factors usually bind their target in an accessible chromatin state, with the exception of pioneer factors that can bind DNA wrapped around nucleosomes, subsequently leading to chromatin remodeling (eviction or displacement of nucleosomes), which in turn renders DNA accessible to the binding of other TFs and the transcriptional machinery [33,34] (Fig. 1). Over the past few years, increasing evidence suggest that HOX TFs act as pioneer factors, which provide an additional mechanism through which the HOX factors implement morphological diversity within the developing embryo. In this review, we present an overview of the studies that provided evidence for HOX pioneer activity and discuss the potential mechanisms underlying this novel functional property as well as its implications in our comprehension of HOX-dependent patterning.

2. Discovery of Hox pioneer activity

The first indication that HOX transcription factors may have pioneer activity came from a study analyzing the genome-wide binding of three Drosophila Hox factors Ubx, AbdA and AbdB [35]. Upon transient transfection of these factors in Drosophila Kc167 cells, the authors uncovered that Ubx and AbdA binding strongly correlated with pre-existing accessible loci. In marked contrast, a significant proportion of genomic loci specifically bound by AbdB were DNAse1 inaccessible. This result led the authors to propose that 1) the differential capacity to bind genomic loci with low chromatin accessibility represents an additional mechanism contributing to variation in the target repertoire of the different Drosophila HOX TFs and 2) the capacity of AbdB to bind its targets in a 'inaccessible' chromatin state could reflect the ability of AbdB to function as a pioneer factor. In a subsequent study, Porcelli et al. [36] provided evidence that the exogenous expression of AbdB in Kc167 cells resulted in a switch from inaccessible to accessible chromatin at AbdB-specific targets, thereby substantiating Beh et al. initial hypothesis. Porcelli et al. also extended the analysis to the eight Drosophila HOX TFs and showed that HOX-bound regions with high motif enrichment, observed for Lab, Pb, Dfd, and AbdB, largely correspond to genomic loci in inaccessible chromatin state. This result indicated that sites that are highly discriminating (i.e. sites bound by only one Hox member) are less accessible compared to sites bound by several different Hox proteins, which led to the notion that increased selectivity is associated with lower chromatin accessibility [36]. Further analysis of Dfd revealed that its binding at low accessible chromatin ultimately resulted in a switch of chromatin accessibility state, comparable to the effect observed for AbdB [36].

HOX binding at sites with low chromatin accessibility was also uncovered for mammalian HOX factors. In a study aimed at better understanding how HoxC factors establish distinct spinal cord fates in the developing embryo [37–40], Bulajic et al., analyzed the genome-wide binding of HoxC6, C8, C9, C10 and C13. Using embryonic stem cell (ES) lines expressing individual HoxC gene upon doxycycline treatment and in vitro motor neuron (MN) differentiation, the authors identified both common and specific targets. Interestingly, HoxC9 and HoxC13 specific targets were enriched at chromatin inaccessible regions [41]. The same trend was observed in undifferentiating cells, prompting the authors to suggest that the ability to bind inaccessible chromatin is likely an intrinsic property rather than the consequence of MN progenitor-specific features. Importantly, comparing the chromatin accessibility in the course of MN differentiation revealed a switch from inaccessible to accessible chromatin primarily at HoxC9- and HoxC13-bound loci, suggesting that HoxC9 and HoxC13 have the ability to promote chromatin accessibility [41]. Another study, aimed at understanding the specific function of HoxA13 and HoxD13 (Hox13 hereafter) in triggering digit development, identified a significant number of Hox13 targets requiring Hox13 function for their chromatin accessibility in the developing limb [42]. ATAC-seq at single cell resolution further showed that the chromatin accessibility specific to the distal limb (i.e. the presumptive digit domain) coincides with the Hox13 dependent chromatin accessibility [42]. Thus, while several 5′HoxA and HoxD genes are expressed in the developing distal limb [reviewed in e.g. 13], only the Hox13 factors have the capacity to establish the distal-limb (digit) specific chromatin accessibility landscape. Based on the complete digit agenesis resulting from Hox13 inactivation [43], the Hox13 pioneering activity has likely a pivotal role in the initiation of the digit developmental program. Consistent with the pioneering activity of the Hox13 TFs uncovered in the developing limb, Amandio et al. showed that in the genital bud, Hox13 binds the Prox regulatory element at a developmental stage that precedes the chromatin opening at this enhancer [44]. Together these studies provided in vivo evidence that Hox13 TFs are able to bind inaccessible chromatin and are required for the switch from inaccessible to accessible chromatin state at their targets.

Genome-wide binding comparison between HoxA11 (expressed in the presumptive zeugopod) and Hox13 (expressed in the presumptive autopod) showed that the Hox13 pioneer sites are exclusively bound by Hox13 [42]. Interestingly, ectopic expression of HoxA11 in the distal limb bud resulted in its binding at Hox13-specific sites. However, upon inactivation of Hox13, these sites are in an inaccessible state and no binding of HoxA11 can be detected [42] suggesting that target specificity for HoxA11 and Hox13 in the developing distal limb is primarily associated with their differential ability to bind loci with low chromatin accessibility. This observation, together with the finding that Hox selectivity in Drosophila Kc167 cells is associated with lower chromatin

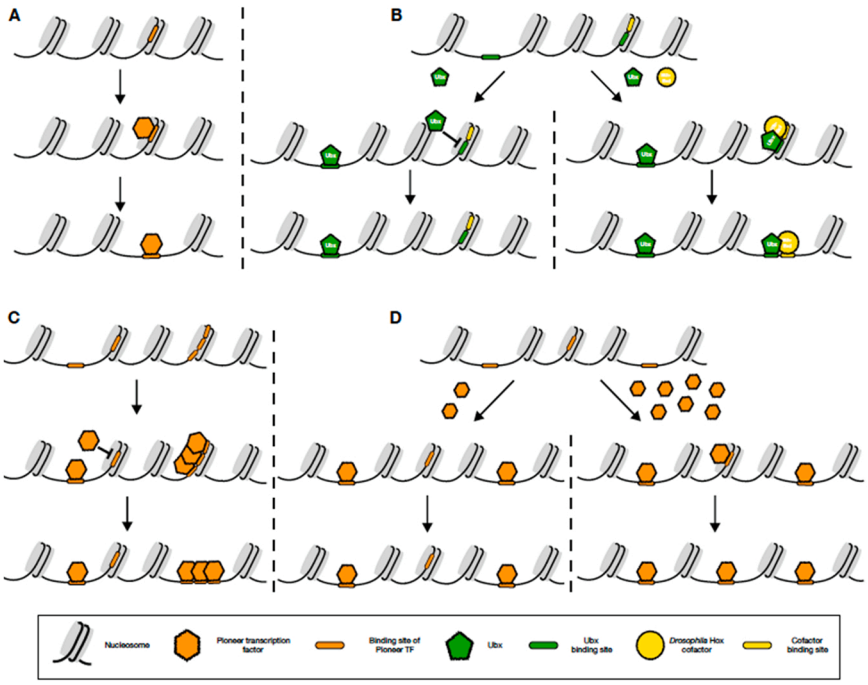


Fig. 2. Mechanisms of chromatin opening by HOX proteins. (a) HOX proteins can bind nucleosomal DNA regions and trigger chromatin opening, as exemplified by Abd-B. (b) Left: In the cell types investigated so far, a subset of HOX factors, such as Ubx, bind DNA accessible regions but not inaccessible regions. Right: In the presence of specific cofactors, such as Hth or Exd in *Drosophila*, these HOX factors become able to bind their targets in an inaccessible chromatin environment, which eventually leads to chromatin opening. (c) The number of binding motifs favors pioneer factor binding at inaccessible chromatin. Sites in nucleosome-free region can be bound even if only 1 binding motif is present. Higher number of binding motifs could favor binding at nucleosomal regions by increasing binding affinity and/or stabilizing the binding of the HOX factors. (d) Pioneer factor dosage has an effect on the capacity to bind inaccessible DNA. Left: Pioneer factors expressed at low level preferentially bind nucleosome-free regions. Right: Higher levels of expression are needed for pioneer factors to engage with inaccessible chromatin, resulting in chromatin opening.

accessibility, raise the question of how Hox binding specificity at inaccessible chromatin is achieved and why some Hox factors appear more efficient than others at binding inaccessible chromatin?

The aforementioned studies were performed in particular cell types: Drosophila Kc167 cells, motor neurons and limb cells, respectively. This limits the possibility to decipher whether the HOX proteins that were considered as poorly capable of binding inaccessible chromatin, could efficiently do so in other tissue contexts / cell types. For instance, in motor neuron progenitors, HOXC6, HOXC8 and HOXC10 failed to engage inaccessible chromatin [41]. However, this could be because the majority of their target sites are already accessible in these cells, and genome-wide studies could have hindered the detection of rare binding events at inaccessible chromatin sites. Yet, as far as HOXC6 and HOXC9 are concerned, their differential ability to bind inaccessible chromatin was observed both in motor neuron progenitors and undifferentiated cells, indicating that the difference between HOXC6 and HOXC9 is unrelated to the specific context of motor neuron progenitors [41]. This result supports the existence of intrinsic differences between HOX factors in their capacity to bind nucleosomal DNA but does not exclude that this capacity varies depending on the tissue/cell type context.

3. Possible mechanisms

3.1. Binding inaccessible chromatin

The DNA in eukaryotic cells is wrapped approximately twice around an octamer composed of the four core histones [45]. In the resulting nucleosome arrays, the portion of DNA facing the globular domains of the histones is sterically hidden, limiting the access of transcription factors to DNA. However, transcription factors referred to as pioneer factors have the capacity to target DNA wrapped around nucleosomes, subsequently triggering the displacement or eviction of nucleosomes, which in turns renders DNA accessible to other transcription factors and the transcriptional machinery (Fig. 1). In this respect, pioneer factors are essential factors for modifying genetic networks during cell fate transition [e.g. 34].

The first step of the pioneering activity is the binding to target DNA wrapped around nucleosomes. Numerous studies aimed at understanding how pioneer factors bind nucleosomal DNA have uncovered different classes of pioneer factors based on their binding characteristics [34]. For instance, FoxA, a pioneer factor essential for liver development [46–48], recognizes its full DNA binding motif and bind to DNA periodically around the nucleosome [34]. Interestingly, the 'winged helix' DNA binding domain of FoxA has a structure reminiscent of the linker histone H1 [49,50] and series of in vitro and in vivo studies indicate that FoxA binding displaces linker histones [51,52]. In addition, the FoxA C-terminal domain contains a core histone binding motif interacting with H3 and H4, and to a lesser extent to H2B [53]. This C-terminal domain together with FoxA high affinity DNA binding domain mediate chromatin opening by FoxA [53].

In contrast to FoxA, other pioneer factors bind smaller/degenerate motifs. When DNA is wrapped around nucleosomes only half of the double-helix is accessible, as the other half is facing the core histones, and binding to smaller or degenerate motifs allows pioneer factors to bind their target sequence in a context compatible with nucleosomal DNA. This binding strategy was uncovered, for instance, for the reprogramming factors Oct4, Sox2/3 and Klf4. At its nucleosome enriched binding sites, Sox2 recognizes a motif that is one nucleotide smaller than the canonical motif allowing it to bind less distorted DNA, which better fills the curvature of DNA wrapped around the histone octamere [54]. Oct4 contains a bipartite POU domain and when both domains are bound simultaneously to the full motif less than a quarter of the DNA circumference remains accessible which is thus incompatible with nucleosome binding due to steric hindrance [54]. In contrast, at its nucleosome enriched binding sites Oct4 uses only one of its two POU domains to bind DNA which accommodates less than half of the DNA surface of the double helix [54]. Klf4 uses only two of its three zinc-finger domains to recognize a hexameric motif occupying one side of the DNA double helix [54]. This contrasts with the binding strategy of another pioneer factor Pax7 [55,56], for which its two DNA binding domains (a paired domain and a homeodomain) are required for Pax7 pioneering activity [57].

How HOX TFs bind nucleosomal DNA remains unclear but a number of features associated with HOX binding at inaccessible chromatin have been reported. The analysis of the Drosophila HOX factors revealed that binding affinity at inaccessible chromatin regions is higher than at accessible loci [36]. As high binding affinity provides a means for transcription factors to compete with nucleosomes, it is likely to be an

important determinant underlying HOX ability to bind their target in an inaccessible chromatin environment (Fig. 2a). Comparison of AbdB pioneer and non-pioneer sites revealed distinct predicted DNA shape, with pioneer sites being associated with narrower minor grove [35]. Interestingly, narrow minor grooves, which allow for more robust electrostatic interactions, has been associated with high HOX binding affinity [58]. Statistical machine learning approaches also uncovered that narrow minor grooves improve HOX binding specificity [59], which likely contribute to the higher selectivity observed at HOX pioneer sites [36]. Binding affinity can also be enhanced by the presence of multiple binding motifs at a given locus. Interestingly, AbdB binding at pre-existing accessible sites compared to binding at inaccessible sites revealed a higher number of binding motifs present at inaccessible sites [35]. Thus, at least for AbdB, the higher number of binding motifs may contribute to its capacity to bind targets with low chromatin accessibility (Fig. 2c). The correlation between the number of binding motifs and transcription factors binding inaccessible chromatin was also observed for non-HOX transcription factors. For instance, reprogramming of embryonic fibroblasts towards endoderm progenitors was initially proposed to rely on the sequential activity of the pioneer factor FoxA1 followed by the non-pioneer HNF4A [60]. However, expression of these factors in K562 cells, either alone or in combination revealed that FoxA1 and HNF4A can bind inaccessible chromatin independently of each other [61]. The only noticeable difference between these two factors in engaging inaccessible chromatin is that HNF4A required approximately two times more motif count, which led the authors to propose that the ability of transcription factors to target DNA sites in inaccessible chromatin depend on the affinity of interaction with the DNA sequence [61].

The expression level of HOX factors may also impact their ability to engage nucleosomal DNA (Fig. 2d). Indeed, targets in an accessible chromatin state by being more open than targets in inaccessible chromatin, could taper-off the available Hox factors thereby impeding binding to nucleosomal DNA. In support of this concept, it was shown that by reducing the level of FOXA1 and HNF4A proteins, both factors were bound predominantly at accessible loci while at higher concentration their binding was increased at inaccessible sites [61]. Interestingly, measurements of transcript copy number of the 5′HoxD genes in the distal limb bud revealed that HoxD13 expression is more than two times higher than the other 5′HoxD genes [62] and Hox13 are the only Hox factors that have the capacity to trigger the distal limb-specific chromatin accessibility [42]. This potential dosage effect raises also the possibility that a given Hox factor, by having distinct expression level in different domains of the developing embryo, could exhibit differential ability to engage inaccessible chromatin.

While, in an in vivo context, expression level could contribute to differences between HOX factors in their ability to bind targets with low chromatin accessibility, it cannot account for the differences observed using cell culture assays, where experiments were performed with transfected cells bearing the same amount of HOX factors [35,36,41]. These latter experiments revealed distinct intrinsic abilities of the Hox factors in binding inaccessible chromatin. Interestingly, while Ubx transfection alone resulted in its binding exclusively at accessible sites, co-transfection with its cofactor Hth resulted in Ubx binding at inaccessible loci. Importantly, this was not due to Hth pre-opening these loci, thereby suggesting that Ubx binding inaccessible chromatin resulted from a Ubx-Hth cooperative effect [Fig. 2b, 35]. Based on these data, HOX may primarily differ in their intrinsic ability to bind loci with low chromatin accessibility rather than their capacity to promote the transition from inaccessible to accessible chromatin conformation per se. It was proposed that, reminiscent of the higher number of AbdB binding motif at pioneer sites, at Ubx-Hth composite sites, the presence of both Ubx and Hth allows for a more efficient competition with nucleosomes [35]. Consistent with this model, SOX2-OCT4 heterodimers have a higher nucleosome-binding ability compared to the individual ability of SOX2 and OCT4 [63]. Nonetheless, as far as the most posterior HOX TFs are concerned (i.e. the Drosophila AbdB and the mouse HoxC9, HoxC13, HoxA13 and HoxD13), motif analysis at pioneer sites did not provide evidence for potential cofactors, suggesting that there might be intrinsic differences among HOX factors in their ability to bind inaccessible chromatin [36,41,42]. In agreement with this model, homeodomain swapping between the non-pioneer HoxC10 and the pioneer HoxC13, provided evidence that the homeodomain identity impacts on the ability of Hox factors to bind inaccessible chromatin [41].

3.2. *Opening inaccessible chromatin*

Once bound to their targets in inaccessible chromatin, pioneer factors trigger the switch from inaccessible to accessible state, ultimately allowing these targets to be transcriptionally active. Studies of the mechanism underlying this step have brought forward distinct mechanisms depending on the pioneer factor considered. For instance, the DNA binding domain of FoxA, structurally similar to the linker histone H1, contributes to displace linker histones upon FoxA binding, thereby rendering the locus accessible [52]. Accordingly, FoxA can open chromatin in absence of ATP-dependent enzymes [53]. In contrast, several pioneer factors trigger the opening of inaccessible chromatin via the recruitment of chromatin remodelers. Chromatin remodelers are ATP-dependent multi-subunit complexes that can be divided in four main families: SWI/SNF, ISWI, INO80 and NURD [reviewed in 64]. These complexes remodel the chromatin through different ways such as nucleosome spacing, repositioning, ejection, eviction, and nucleosome editing through histone exchange [reviewed in 65]. Such remodeling creates nucleosome depleted regions, i.e. accessible regions. In ES cells the reprogramming factor OCT4 establishes chromatin accessibility by recruiting BRG1, one of the two catalytical subunits of the SWI/SNF complex [66]. Similarly, several other pioneer factors, such as Ils1 and GATA 3, have been shown to require BRG1 for their pioneer activity [67,68]. However, chromatin remodelers are not necessarily recruited by the pioneer factors themselves. For instance, Pax7, a well-known pioneer factor involved in the specification of the melanotrope lineage [55], binds on its own to its targets located in inaccessible chromatin [56] but the opening of these regions requires the recruitment of the non-pioneer factor Tpit [69]. It was proposed that it is actually Tpit that recruits chromatin remodelers [69].

While we start to understand how HOX factors bind to inaccessible chromatin, the mechanism by which this binding results in the switch to an accessible chromatin state remains unknown. Unlike FoxA, there is no obvious domain within the HOX proteins to suggest an intrinsic ability of Hox factors to remodel chromatin. In contrast, several screens have identified chromatin remodeling factors as potential HOX interacting partners [reviewed in 70]. It is thus tempting to speculate that HOX pioneer activity is mediated by the recruitment of chromatin remodelers. HOX interaction with their cofactors should also be taken into account as one of the mechanisms leading to chromatin opening. Indeed, in *Drosophila* Kc167 cells, the presence of the cofactors Exd and Hth resulted in increased chromatin opening compared to either HOX or cofactor alone, as measured by median ATAC-seq score [36]. HOX-cofactor complexes bound at inaccessible chromatin could create a more robust allosteric effect contributing to destabilizing DNA-nucleosome interaction, thereby contributing to DNA accessibility. Alternatively, HOX-cofactor complexes may be more efficient in recruiting chromatin remodelers that HOX factors alone.

4. Anti-pioneer activity

A well-known example of homeotic transformation is the four-winged fly, in which Ubx loss of function results in the transformation of the third thoracic segment, characterized by the presence of halteres, into a second thoracic segment bearing wings [4,71,72]. In a study aimed at understanding how Ubx triggers haltere development, Loker and colleagues identified specific changes in chromatin accessibility

mediated by Ubx [73]. ATAC-seq comparison between the wing and haltere imaginal disc revealed that most of the differences in chromatin accessibility between these two tissues correspond to Ubx-dependent changes in chromatin accessibility. Interestingly, the authors found that depending on the domain along the proximal-distal axis of the haltere imaginal disc, Ubx binding either promoted or reduced chromatin accessibility at its targets [73]. This result led the authors to propose that Ubx acts both as pioneer and anti-pioneer factor to set up the haltere-specific chromatin accessibility landscape. Similarly, while many HOX13-bound loci lose accessibility in Hox13-/- limb buds, a small subset of the HOX13 direct targets were associated with a gain of accessibility [42; Fig.S5b]. This latter finding suggests that, similar to Ubx, HOX13 may act both as pioneer and anti-pioneer factors but additional studies are needed to test this hypothesis.

How these opposite effects of Hox factors on chromatin accessibility is implemented remains to be investigated. The simplest explanation would rely on HOX interaction with distinct partners. The Ubx-Hth-Exd complex, allowing Ubx to bind inaccessible chromatin (and subsequent chromatin opening), has also been identified at anti-pioneer sites [73], which excludes Hth-Exd function as being a landmark of Ubx pioneer activity. Nonetheless, some data from non-HOX pioneer factors are consistent with the implication of distinct sets of interacting partners in the switch towards an anti-pioneer activity. For instance, FoxA1, which was demonstrated to act as a pioneer factor [53], is also able to recruit Grg3, a member of the Groucho related co-repressors, and this recruitment results in a significantly more DNAse1 resistant structure [74]. The authors proposed that this chromatin condensation might be due to individual Grg3 subunit interacting at the same time with the Histone H3 tails of adjacent nucleosomes or with a structure at the interface of neighboring nucleosomes [74]. Similarly, a mutation that impairs FoxA2 ability to bind nucleosomes and open chromatin in vitro (ΔHx mutant, 10-amino acid deletion that spans the α-helical domain), results in about 30% of sites with decrease accessibility, a majority of sites (~50%) with no difference compared to the WT and ~17% of sites showing an increased chromatin accessibility [75; in Fig.5]. Interestingly, these latter sites are enriched for FOX motif as well as motifs for SOX and MSX which are known to interact with repressive complexes [76–78]. These examples where pioneer factor interaction with some cofactors results in increased chromatin inaccessibility suggests that the function of pioneer factors may vary depending on the target site (i.e. which cofactor binding site is present) and also on the cell environment (i.e. which cofactor is expressed). However, the process can be more complex as exemplified with FOXD3 function in embryonic stem cells. Indeed, in these cells, FoxD3 binding at enhancers first recruits the BRG1-containing SWI/SNF complex, which leads to nucleosome removal, but subsequently recruits histone deacetylases to inhibit maximal activation of these enhancers [79].

The anti-pioneer activity might be as important as the pioneer activity itself in controlling cell fate. Indeed, while the classical view of pioneer factor function is to open new regulatory elements thereby leading to the transition towards a distinct genetic program, closing other set of regulatory elements is likely to be equally important in this process, or at least allow for a more drastic change in the genetic program. While the large majority of studies have so far focused on the mechanisms leading to the switch from inaccessible to accessible chromatin, it will be important to extend the analyses to mechanisms underlying the anti-pioneering activity.

5. Implications of Hox pioneer activity in embryonic patterning

The chromatin accessibility landscape defines the set of regulatory elements permissive to transcriptional activity. It is thus an important parameter in the process leading to the implementation of a cell type/tissue-specific genetic program. The capacity of the Hox factors to bind their targets in an inaccessible chromatin state and trigger the switch to an accessible state represents an additional mechanism by which Hox TFs control cell fate. During embryonic development, the sequential activation of the Hox genes along the A-P axis of the trunk and A-P as well as P-D axes in appendages, implies a sequential change in the chromatin accessibility landscape, which either contributes to or be decisive in the implementation of distinct developmental programs within the developing embryo.

The study of all Drosophila *Hox* genes provided evidence that selectivity, i.e. sites that are bound by a single HOX factor, are less accessible than sites bound by several HOX members [36]. This result raises the possibility that the mechanism whereby a given HOX establishes a specific body structure, is primarily associated with binding specificity at targets that were in an inaccessible state prior to the expression of this Hox gene. Consistent with this view, comparison between HoxA11 and Hox13 targets in the developing limb revealed that the Hox13-specific targets are the ones whose chromatin accessibility is dependent on Hox13 function, which correspond to the distal limb-specific chromatin accessibility landscape [42]. This however does not preclude that HOX binding specificity at accessible chromatin regions contributes to the specific developmental programs triggered by individual *Hox* genes. Assessing the precise contribution of HOX pioneer activity in establishing morphological diversity during embryonic development is challenging. Indeed, it requires abrogating the pioneer activity without interfering with the other properties of the HOX TFs. The studies performed so far have uncovered some features associated with HOX binding their targets in the nucleosomal DNA context. These include: 1) higher number of binding motifs, 2) high binding affinity, 3) requirement of cofactors, at least for some Hox factors (e.g. Ubx), 4) homeodomain identity and 5) Hox expression level (see paragraph 3.1). Based on the diversity of features associated with binding inaccessible chromatin, assessing the precise role of HOX pioneer activity in the establishment of morphological identity might be more feasible through the disruption of Hox-dependent chromatin remodelling.

An important question regarding HOX pioneer activity is how this functional property is reconcilable with the fact that a given *Hox* gene can contribute to the development of very different structures. For instance, HoxC9 induces a thoracic fate to the spinal cord neurons [39, 80] and is also required for fore-arm development [81], thus it seems unlikely that HoxC9 pioneer activity is identical in these two different developmental contexts. Interestingly, for the pioneer factor Pax7, whose function is critical in myogenic progenitors [reviewed in 82] as well as in pituitary development [55], it was found that a subset of muscle-specific Pax7 targets are not accessible in pituitary cells [69]. It was proposed that heterochromatin histone modifications could constitute a barrier to the recruitment of Pax7 and pioneer factors in general [83]. This model is in agreement with the discovery that H3K9me3-marked heterochromatin prevents the initial recruitment of pluripotency factors [84]. Whether heterochromatin histone modifications also act as a roadblock to HOX binding at nucleosomal DNA and contribute to tissue/cell type differences in HOX pioneering activity remains to be investigated. An alternative, yet non-exclusive possibility is that HOX cofactors refine the repertoire of HOX targets in the context of inaccessible chromatin, resulting in the opening of different regulatory elements depending on the cell type/tissue. In a recent study by Feng and colleagues, they characterized chromatin accessibility in two different Drosophila leg imaginal discs, one whose identity is controlled by Ubx (T3 leg) and the other by Scr (T1 leg) [85]. In contrast to Locker et al., this latter study revealed that differences in accessibility do not correlate with differential binding (~8% of all binding events) and the authors suggest that chromatin accessibility is neither altered by Hox expression nor can it account for paralog-specific Hox–DNA binding. It is conceivable that differences in leg identity are sufficiently minor to be achieved in absence of differential chromatin accessibility and relies on multiple, context-specific Hox cofactors eventually leading to distinct transcriptional outputs and leg identities.

Hox pioneer activity is likely to have important morphological consequences in situation where mutations lead to the expression of a

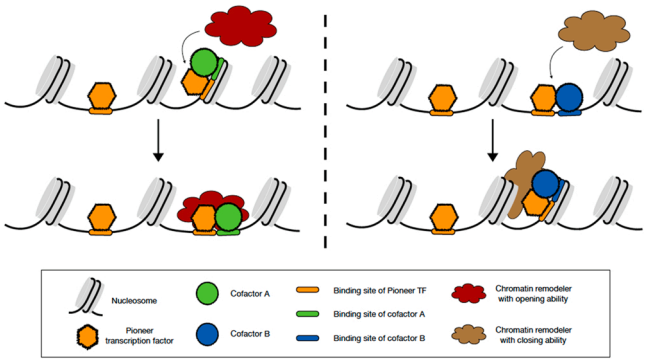


Fig. 3. Pioneer versus anti-pioneer activity of Hox factors. (Left) In the presence of binding motifs for a specific HOX cofactor (cofactor A, green), the HOX and cofactor proteins can bind a nucleosome-containing site and recruit specific chromatin remodeler with " opening " ability, resulting in the switch to accessible chromatin. This can be achieved by either nucleosome eviction or nucleosome sliding (here only nucleosome eviction is represented). (Right) In the presence of binding motifs for a different HOX cofactor (cofactor B, blue), both proteins bind to a nucleosome-free region and recruit a chromatin remodeler with " closing " ability.

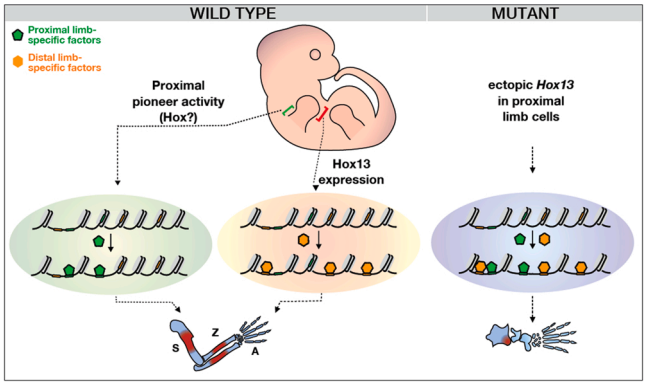


Fig. 4. Schematic representation of our working hypothesis regarding Hox pioneering activity in developing limbs. (Left) During wild type limb development, the pioneer activity of proximal-specific factors (green hexagones; Hox?) trigger chromatin accessibility at their targets in the proximal limb domain (green bracket). As a result, cells acquire a proximal fate, eventually leading to the formation of the proximal limb segments, i.e. stylopod (S) and zeugopod (Z). (Middle) In the distal limb bud (orange bracket), HOX13 factors are expressed (orange hexagones) and trigger chromatin opening at their previously inaccessible targets [42]. These cells acquire an autopod (A) fate, ultimately forming the digits. Upon ectopic expression of HOX13 in proximal limb bud cells (Right), the development of proximal limb segments is severely impaired [89,90]. We propose that such ectopic expression of HOX13 factors results in ectopic chromatin accessibility of distal-specific regulatory elements (orange rectangle), resulting in the ectopic expression of (at least some) distal (digit)-specific transcription factors. Such aberrant gene expression eventually impairs the implementation of proximal fate within the developing proximal limb bud, thereby disrupting the formation of proximal limb segments.

Hox gene in cells where this gene is normally not expressed as it implies that a series of regulatory elements become aberrantly permissive to transcriptional activity. There are numerous examples where ectopic expression of a *Hox* genes results in drastic morphological alterations. For instance, in Drosophila, *abd-A* specifies the second to fourth abdominal segments and an enhancer point mutation resulting in *abd-A* ectopic expression partially transforms the third thoracic segment into a second abdominal segment [86]. Similarly, in mice, while the paralogous *Hox10* genes are required to establish the lumbar identity [87], the ectopic expression of *Hoxa10* in progenitor cells normally forming thoracic vertebrae results in a thorax-to-lumbar transformation [88]. Future studies should be dedicated to assessing the role of Hox pioneer activity in Hox-associated congenital malformation/diseases.

A recent study provided evidence that Ubx-dependent formation of the Drosophila halteres is based on both pioneer and anti-pioneer activities of Ubx [73]. While the other published studies investigating the relationship between HOX factors and chromatin accessibility have focused on chromatin opening, it is likely that other HOX factors can also behave as anti-pioneer factors. Consistent with this hypothesis, a subset of Hox13 direct targets was found to gain accessibility upon Hox13 inactivation in the developing digits [42]. It is likely that the capacity of closing genomic regions, notably regulatory elements, is part of the mechanism underlying HOX-dependent patterning. Actually, HOX anti-pioneer activity might be as important as their pioneer activity in controlling cell fate, notably when drastically different fates are implemented. In turn, this raises the question of how HOX factors exert such opposite effects on chromatin accessibility. Intuitively, interaction with distinct protein partners may be at least one of the mechanisms accounting for this functional difference (Fig. 3).

6. Conclusion

The discovery that HOX factors behave as pioneer factors provides a new mechanism through which HOX TFs establish morphological diversity in the developing embryo and generate malformations/diseases when aberrantly expressed (Fig. 4). The property of HOX factors to modify the chromatin accessibility landscape has important implications as it contributes to defining the sets of regulatory elements transcriptionally permissive. In this view, HOX pioneer activity may represent a pivotal step in the implementation of distinct transcriptional programs within the developing embryo. The evidence that increased selectivity is associated with lower chromatin accessibility [36] supports this hypothesis. However, definitive assessment of the significance of Hox pioneer activity in Hox-dependent patterning will require experimental disruption of the pioneer activity of HOX factors without interfering with their non-pioneer activity, which may prove to be challenging.

Conflicts of Interests

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the Canadian Institutes of Health Research (grant number CHIR-174989 and CIHR-162143. R. Paul was supported by IRCM Jean-Coutu scholarship and R. Peraldi was supported by IRCM Jean-Coutu scholarship.

References

[1] J.C. Pearson, D. Lemons, W. McGinnis, Modulating Hox gene functions during animal body patterning, Nat. Rev. Genet 6 (12) (2005) 893–904.
[2] J. Castelli-Gair Hombría, C. Sánchez-Higueras, E. Sánchez-Herrero, Control of organogenesis by hox genes, in: J. Castelli-Gair Hombría, P. Bovolenta (Eds.), Organogenetic Gene Networks: Genetic Control of Organ Formation, Springer International Publishing, Cham, 2016, pp. 319–373.
[3] D. Duboule, Guidebook to the Homeobox Genes, Oxford University Press, Oxford, 1994.
[4] E.B. Lewis, Gene complex controlling segmentation in drosophila, Nature 276 (5688) (1978) 565–570.
[5] S.J. Gaunt, P.T. Sharpe, D. Duboule, Spatially restricted domains of homeo-gene transcripts in mouse embryos - relation to a segmented body plan, Development 104 (1988) 169–179.
[6] A. Graham, N. Papalopulu, R. Krumlauf, The murine and drosophila homeobox gene complexes have common features of organization and expression, Cell 57 (3) (1989) 367–378.
[7] D. Duboule, P. Dolle, The structural and functional-organization of the murine hox gene family resembles that of drosophila homeotic genes, Embo J. 8 (5) (1989) 1497–1505.

[8] P. Dolle, et al., Coordinate expression of the murine hox-5 complex homoeobox-containing genes during limb pattern-formation, Nature 342 (6251) (1989) 767–772.
[9] J.C. Izpisuabelmonte, et al., Murine genes related to the drosophila abdb homeotic gene are sequentially expressed during development of the posterior part of the body, EMBO J. 10 (8) (1991) 2279–2289.
[10] R. Neijts, et al., Cdx is crucial for the timing mechanism driving colinear Hox activation and defines a trunk segment in the Hox cluster topology, Dev. Biol. 422 (2) (2017) 146–154.
[11] R. Neijts, et al., Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos, Genes Dev. 30 (17) (2016) 1937–1942.
[12] J. Deschamps, D. Duboule, Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock, Genes Dev. 31 (14) (2017) 1406–1416.
[13] J. Zakany, D. Duboule, The role of Hox genes during vertebrate limb development, Curr. Opin. Genet Dev. 17 (4) (2007) 359–366.
[14] M.F. Berger, et al., Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences, Cell 133 (7) (2008) 1266–1276.
[15] M.B. Noyes, et al., Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites, Cell 133 (7) (2008) 1277–1289.
[16] R.S. Mann, K.M. Lelli, R. Joshi, Chapter 3 hox specificity: unique roles for cofactors and collaborators. Current Topics in Developmental Biology, Academic Press,, 2009, pp. 63–101.
[17] M. Slattery, et al., Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins, Cell 147 (6) (2011) 1270–1282.
[18] S. Merabet, R.S. Mann, To be specific or not: the critical relationship between hox and TALE proteins, Trends Genet 32 (6) (2016) 334–347.
[19] C.B. Moens, L. Selleri, Hox cofactors in vertebrate development, Dev. Biol. 291 (2) (2006) 193–206.
[20] M. Peifer, E. Wieschaus, Mutations in the drosophila gene extradenticle affect the way specific homeo domain proteins regulate segmental identity, Genes Dev. 4 (7) (1990) 1209–1223.
[21] E. Ferretti, et al., Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins, Development 127 (1) (2000) 155–166.
[22] C. Rauskolb, M. Peifer, E. Wieschaus, Extradenticle, A regulator of homeotic gene activity, is a homolog of the homeobox-containing human protooncogene PBX1, Cell 74 (6) (1993) 1101–1112.
[23] C. Rauskolb, et al., Extradenticle determines segmental identities throughout drosophila development, Development 121 (11) (1995) 3663–3673.
[24] L. Selleri, et al., The TALE homeodomain protein Pbx2 is not essential for development and long-term survival, Mol. Cell. Biol. 24 (12) (2004) 5324–5331.
[25] A. Vitobello, et al., Hox and Pbx factors control retinoic acid synthesis during hindbrain segmentation, Dev. Cell 20 (4) (2011) 469–482.
[26] A.J. Waskiewicz, et al., Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning, Development 128 (21) (2001) 4139–4151.
[27] A.J. Waskiewicz, H.A. Rikhof, C.B. Moens, Eliminating zebrafish Pbx proteins reveals a hindbrain ground state, Dev. Cell 3 (5) (2002) 723–733.
[28] T. Alexander, C. Nolte, R. Krumlauf, Hox genes and segmentation of the hindbrain and axial skeleton, Annu. Rev. Cell Dev. Biol. 25 (2009) 431–456.
[29] R.S. Mann, S.K. Chan, Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins, Trends Genet. 12 (7) (1996) 258–262.
[30] M. Manzanares, et al., Independent regulation of initiation and maintenance phases of Hoxa3 expression in the vertebrate hindbrain involve auto- and cross-regulatory mechanisms, Development 128 (18) (2001) 3595–3607.
[31] S. Tumpel, L.M. Wiedemann, R. Krumlauf, Hox genes and segmentation of the vertebrate hindbrain, in: O. Pourquie (Ed.), Hox Genes, Elsevier Academic Press Inc, San Diego, 2009, pp. 103–137.
[32] J. Crocker, et al., Low affinity binding site clusters confer hox specificity and regulatory robustness, Cell 160 (1–2) (2015) 191–203.
[33] M. Iwafuchi-Doi, K.S. Zaret, Pioneer transcription factors in cell reprogramming, Genes Dev. 28 (24) (2014) 2679–2692.
[34] K.S. Zaret, Pioneer transcription factors initiating gene network changes, Annu Rev. Genet 54 (2020) 367–385.
[35] C.Y. Beh, et al., Roles of cofactors and chromatin accessibility in Hox protein target specificity, Epigenet. Chromatin 9 (2016) 1.
[36] D. Porcelli, et al., Chromatin accessibility plays a key role in selective targeting of Hox proteins, Genome Biol. 20 (1) (2019) 115.
[37] Dasen, J.S. and T.M. Jessell, Hox networks and the origins of motor neuron diversity, in Hox Genes, O. Pourquie, Editor. 2009. p. 169–16.
[38] J.S. Dasen, J.P. Liu, T.M. Jessell, Motor neuron columnar fate imposed by sequential phases of Hox-c activity, Nature 425 (6961) (2003) 926–933.
[39] H. Jung, et al., Global control of motor neuron topography mediated by the repressive actions of a single hox gene, Neuron 67 (5) (2010) 781–796.
[40] X.D. Wu, et al., Cooperation between EZH2, NSPc1-mediated histone H2A ubiquitination and Dnmt1 in HOX gene silencing, Nucleic Acids Res. 36 (11) (2008) 3590–3599.
[41] M. Bulajic, et al., Differential abilities to engage inaccessible chromatin diversify vertebrate Hox binding patterns, Development 147 (22) (2020).
[42] I. Desanlis, et al., HOX13-dependent chromatin accessibility underlies the transition towards the digit development program, Nat. Commun. 11 (1) (2020) 2491.
[43] C. Fromental-Ramain, et al., Hoxa-13 and Hoxd-13 play a crucial role in the patterning of the limb autopod, Development 122 (10) (1996) 2997–3011.
[44] A.R. Amandio, et al., A complex regulatory landscape involved in the development of mammalian external genitals, Elife (2020) 9.
[45] R.D. Kornberg, Y.L. Lorch, Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome, Cell 98 (3) (1999) 285–294.
[46] R. Gualdi, et al., Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control, Genes Dev. 10 (13) (1996) 1670–1682.
[47] P. Bossard, K.S. Zaret, Repressive and restrictive mesodermal interactions with gut endoderm: possible relation to Meckel's Diverticulum, Development 127 (22) (2000) 4915–4923.
[48] C.S. Lee, et al., The initiation of liver development is dependent on Foxa transcription factors, Nature 435 (7044) (2005) 944–947.
[49] K.L. Clark, et al., CO-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone-H5, Nature 364 (6436) (1993) 412–420.
[50] V. Ramakrishnan, et al., Crystal-structure of globular domain of histone H5 and its implications for nucleosome binding, Nature 362 (6417) (1993) 219–223.
[51] J.H. Taube, et al., Foxa1 functions as a pioneer transcription factor at transposable elements to activate Afp during differentiation of embryonic stem cells, J. Biol. Chem. 285 (21) (2010) 16135–16144.
[52] M. Iwafuchi-Doi, et al., The pioneer transcription factor foxa maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation, Mol. Cell 62 (1) (2016) 79–91.
[53] L.A. Cirillo, et al., Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4, Mol. Cell 9 (2) (2002) 279–289.
[54] A. Soufi, et al., Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming, Cell 161 (3) (2015) 555–568.
[55] L. Budry, et al., The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling, Genes Dev. 26 (20) (2012) 2299–2310.
[56] A. Mayran, et al., Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate, Nat. Genet 50 (2) (2018) 259–269.
[57] A. Pelletier, et al., Pax7 pioneer factor action requires both paired and homeo DNA binding domains, Nucleic Acids Res 49 (13) (2021) 7424–7436.
[58] R. Rohs, et al., The role of DNA shape in protein-DNA recognition, Nature 461 (7268) (2009) 1248–1253.
[59] N. Abe, et al., Deconvolving the recognition of DNA shape from sequence, Cell 161 (2) (2015) 307–318.
[60] K. Horisawa, et al., The dynamics of transcriptional activation by hepatic reprogramming factors, Mol. Cell 79 (4) (2020), 660-66.
[61] J.L. Hansen, K.J. Loell, B.A. Cohen, A test of the pioneer factor hypothesis using ectopic liver gene activation, Elife (2022) 11.
[62] T. Montavon, et al., Modeling Hox gene regulation in digits: reverse collinearity and the molecular origin of thumbness, Genes Dev. 22 (3) (2008) 346–359.
[63] A.K. Michael, et al., Mechanisms of OCT4-SOX2 motif readout on nucleosomes, Science 368 (6498) (2020), 1460-146.
[64] L. Ho, G.R. Crabtree, Chromatin remodelling during development, Nature 463 (7280) (2010) 474–484.
[65] C.R. Clapier, et al., Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes, Nat. Rev. Mol. Cell Biol. 18 (7) (2017) 407–422.
[66] H.W. King, R.J. Klose, The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells, Elife (2017) 6.
[67] R. Gao, et al., Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate, Cell Res. 29 (6) (2019) 486–501.
[68] M. Takaku, et al., GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler, Genome Biol. 17 (2016) 16.
[69] A. Mayran, et al., Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening, Nat. Commun. 10 (1) (2019) 3807.
[70] R. Rezsohazy, et al., Cellular and molecular insights into Hox protein action, Development 142 (7) (2015) 1212–1227.
[71] W. Bender, et al., Molecular-genetics of the bithorax complex in drosophila-melanogaster, Science 221 (4605) (1983) 23–29.
[72] S.B. Carroll, S.D. Weatherbee, J.A. Langeland, Homeotic genes and the regulation and evolution of insect wing number, Nature 375 (6526) (1995) 58–61.
[73] R. Loker, J.E. Sanner, R.S. Mann, Cell-type-specific Hox regulatory strategies orchestrate tissue identity, Curr. Biol. 31 (19) (2021) 4246–4255, e4.
[74] T. Sekiya, K.S. Zaret, Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo, Mol. Cell 28 (2) (2007) 291–303.
[75] M. Iwafuchi, et al., Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones, Nat. Genet 52 (4) (2020) 418–427.
[76] R.B. Koumangoye, et al., SOX4 interacts with EZH2 and HDAC3 to suppress microRNA-31 in invasive esophageal cancer cells, Mol. Cancer 14 (1) (2015) 24.
[77] N. Rave-Harel, et al., The Groucho-related gene family regulates the gonadotropin-releasing hormone gene through interaction with the homeodomain proteins MSX1 and OCT1, J. Biol. Chem. 280 (35) (2005) 30975–30983.
[78] J. Wang, et al., The Msx1 homeoprotein recruits polycomb to the nuclear periphery during development, Dev. Cell 21 (3) (2011) 575–588.
[79] R. Krishnakumar, et al., FOXD3 regulates pluripotent stem cell potential by simultaneously initiating and repressing enhancer activity, Cell Stem Cell 18 (1) (2016) 104–117.
[80] H. Jung, et al., Evolving hox activity profiles govern diversity in locomotor systems, Dev. Cell 29 (2) (2014) 171–187.
[81] B. Xu, D.M. Wellik, Axial Hox9 activity establishes the posterior field in the developing forelimb, Proc. Natl. Acad. Sci. USA 108 (12) (2011) 4888–4891.

[82] M. Buckingham, F. Relaix, The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions, Annu. Rev. Cell Dev. Biol. 23 (2007) 645–673.
[83] A. Balsalobre, J. Drouin, Pioneer factors as master regulators of the epigenome and cell fate, Nat. Rev. Mol. Cell Biol. 23 (7) (2022) 449–464.
[84] A. Soufi, G. Donahue, K.S. Zaret, Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome, Cell 151 (5) (2012) 994–1004.
[85] S. Feng, et al., Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors, Nat. Commun. 13 (1) (2022) 3808.
[86] M.J. Shimell, et al., Enhancer point mutation results in a homeotic transformation in drosophila, Science 264 (5161) (1994) 968–971.
[87] D.M. Wellik, M.R. Capecchi, Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton, Science 301 (5631) (2003) 363–367.
[88] M. Carapuco, et al., Hox genes specify vertebral types in the presomitic mesoderm, Genes Dev. 19 (18) (2005) 2116–2121.
[89] M.E. Williams, J.A. Lehoczky, J.W. Innis, A group 13 homeodomain is neither necessary nor sufficient for posterior prevalence in the mouse limb, Dev. Biol. 297 (2) (2006) 493–507.
[90] C.C. Bolt, et al., Mesomelic dysplasias associated with the HOXD locus are caused by regulatory reallocations, Nat. Commun. 12 (1) (2021) 5013.

Seminars in Cell and Developmental Biology 152-153 (2024) 76–84

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

TALE transcription factors: Cofactors no more

Nicoletta Bobola [a,*], Charles G. Sagerström [b,*]

[a] School of Medical Sciences, University of Manchester, Manchester, UK
[b] Section of Developmental Biology, Department of Pediatrics, University of Colorado Medical School, Aurora, CO, USA

ARTICLE INFO

Keywords:
TALE
HOX
Transcription factors
Chromatin
Embryo

ABSTRACT

Exd/PBX, Hth/MEIS and PREP proteins belong to the TALE (three-amino-acid loop extension) superclass of transcription factors (TFs) with an atypical homedomain (HD). Originally discovered as "cofactors" to HOX proteins, revisiting their traditional role in light of genome-wide experiments reveals a strong and reproducible pattern of HOX and TALE co-occupancy across diverse embryonic tissues. While confirming that TALE increases HOX specificity and selectivity in vivo, this wider outlook also reveals novel aspects of HOX:TALE collaboration, namely that HOX TFs generally require pre-bound TALE factors to access their functional binding sites in vivo. In contrast to the restricted expression domains of HOX TFs, TALE factors are largely ubiquitous, and PBX and PREP are expressed at the earliest developmental stages. PBX and MEIS control development of many organs and tissues and their dysregulation is associated with congenital disease and cancer. Accordingly, many instances of TALE cooperation with non HOX TFs have been documented in various systems. The model that emerges from these studies is that TALE TFs create a permissive chromatin platform that is selected by tissue-restricted TFs for binding. In turn, HOX and other tissue-restricted TFs selectively convert a ubiquitous pool of low affinity TALE binding events into high confidence, tissue-restricted binding events associated with transcriptional activation. As a result, TALE:TF complexes are associated with active chromatin and domain/lineage-specific gene activity. TALE ubiquitous expression and broad genomic occupancy, as well as the increasing examples of TALE tissue-specific partners, reveal a universal and obligatory role for TALE in the control of tissue and lineage-specific transcriptional programs, beyond their initial discovery as HOX co-factors.

1. Introduction

Transcription factors usually belong to large protein families that share high sequence identity in their DNA binding domains. This means that members of the same TF family have similar DNA binding characteristics and that there must be mechanisms to expand TF binding preferences so that individual family members can control distinct gene expression programs. One striking example of this is the HOX TF family where there are at least 39 HOX genes in mouse and human, and as many as 48 in zebrafish, all of which contain related HDs that mediate DNA binding. In the case of HOX TFs, proteins of the TALE family have been implicated as "cofactors" that interact with HOX TFs and expand their binding repertoire. Initial examinations of TALE and HOX interactions relied on genetic, in vitro, and reporter-based assays and established a framework for TALE function. While highly informative, these studies left several questions unanswered. First, these experiments were by necessity restricted to a small number of examples, raising the question if they are generalizable. Second, it is unclear how well the in vitro findings reflect in vivo function. Third, it remains possible that cofactor functions extend beyond HOX proteins. In this review, we will focus on how recent applications of in vivo genome-wide approaches have begun providing answers to these questions. For more comprehensive reading on TALE TFs, we refer the reader to recent reviews [1, 2].

2. TALE proteins as HOX "cofactors"

HOX TFs are present as multiple family members in all bilaterian animals, where they establish the identities of different structures along the antero-posterior axis, as well as control aspects of limb and organ formation, during embryogenesis [3–5]. Beyond development, HOX TFs are also involved in tissue maintenance and homeostasis and are frequently misexpressed in cancers [6]. HOX genes are found in genomic clusters arranged in such a way that more anteriorly expressed genes are

Abbreviations: TF, Transcription factors; HD, Homeodomain.

* Corresponding authors.
E-mail addresses: Nicoletta.Bobola@manchester.ac.uk (N. Bobola), Charles.sagerstrom@cuanschutz.edu (C.G. Sagerström).

https://doi.org/10.1016/j.semcdb.2022.11.015
Received 19 July 2022; Received in revised form 27 October 2022; Accepted 30 November 2022
Available online 9 December 2022

located at the 3′ end of the cluster and posteriorly expressed ones at the 5′ end. Due to genome duplications in the vertebrate lineage, there are four copies of each cluster and HOX genes located at the analogous position in each cluster are referred to as paralogs. HOX TFs are defined by the HD – a highly conserved DNA binding moiety consisting of three alpha helices with helix three making key contacts in the major grove of DNA – that is shared by hundreds of TFs [7]. Individual HOX TFs are presumed to act by directing specific transcriptional programs in vivo [8–10], yet all HOX HDs recognise highly similar sequences, characterized by a four-base-pair TAAT sequence core [11,12]. This discrepancy is widely known as "the HOX specificity paradox" [13]. A partial explanation for this paradox was discovered in the form of HOX "cofactors". The first HOX cofactor identified was *Drosophila* Extradenticle (Exd), whose mutations phenocopy the types of homeotic transformations observed in HOX mutants [14–16]. *HOX* gene expression is unaffected in *exd* mutants, suggesting that Exd assists HOX proteins in directing the appropriate transcriptional program. A second *Drosophila* mutant, *homothorax* (*hth*) was subsequently shown to have similar characteristics to Exd and to also act as a HOX cofactor [17–19]. Both Exd and Hth have orthologs in other species and their functions appear to be broadly conserved. PBX proteins, the vertebrate orthologs of Exd, were originally discovered as translocation partners with E2a in pre-B-cell leukemias [20,21]. Hth is most closely related to vertebrate MEIS proteins, which were first identified as required for HOX-mediated transformations in a mouse model of leukemia [22,23]. PREP proteins (also known as PKNOX) were identified as a DNA binding activity at the uPA enhancer and subsequently shown to be related to MEIS/Hth [24]. TALE proteins are present in vertebrates and other insects, though PREP appears to have been lost in *Drosophila* [25], and they are broadly expressed across cell and tissue types (Fig. 1). The family of vertebrate TALE orthologs has since grown to contain four PBX, three MEIS and two PREP proteins in mouse and human. Collectively, PBX/Exd are known as PBC proteins and Hth/MEIS/PREP as Meinox proteins. They all contain a HD of a slightly variant type, with a three amino acid extension of the loop between helices one and two. This identifies them as members of the TALE (three amino acid loop extension) subfamily of HD proteins.

2.1. TALE proteins modulate the sequence selectivity of HOX TFs in vitro

The presence of a HD in TALE proteins indicates that they function as DNA-binding TFs. Extensive work has demonstrated that Exd/PBX binds DNA cooperatively with HOX TFs [26–34] and that formation of TALE:HOX complexes confers distinct sequence target specificities to HOX proteins [34–39]. Specifically, in vitro studies established that, while

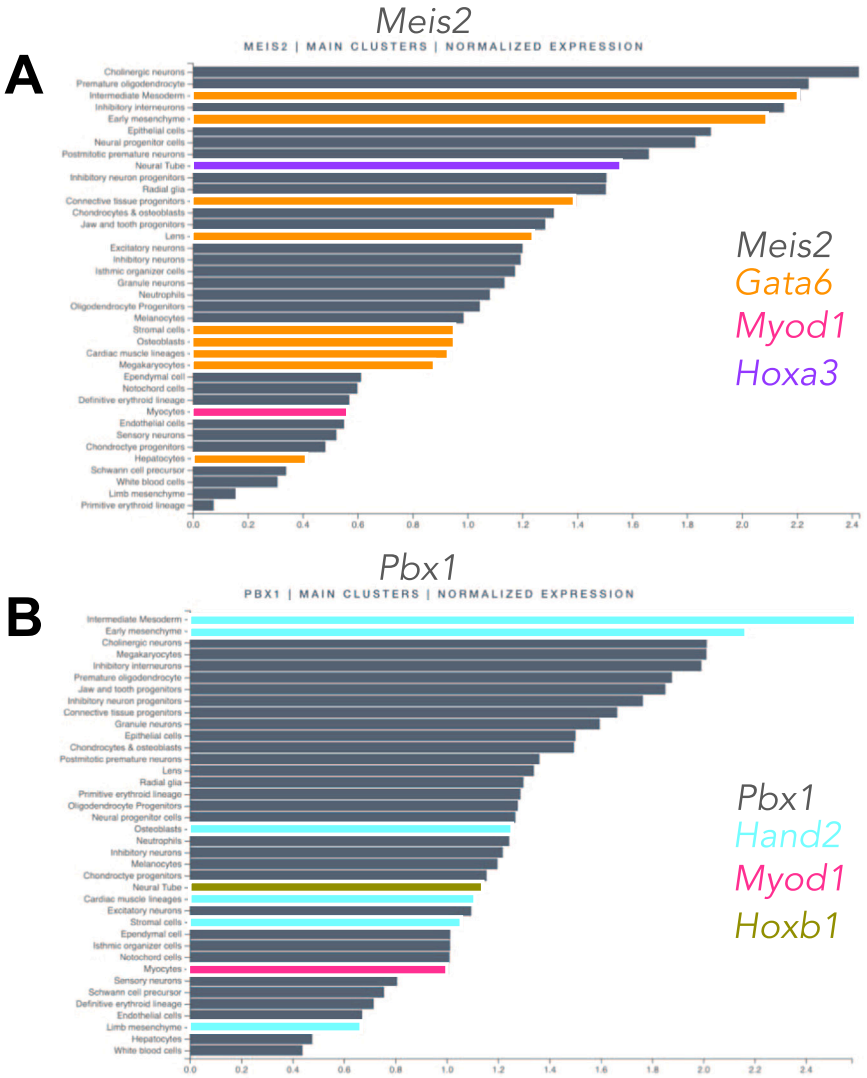


Fig. 1. Broadly ubiquitous expression of TALE relative to cooperating TFs. *Meis2* (A) and *Pbx1* (B) expression from the Mouse Organogenesis Cell Atlas (MOCA), a single cell analysis of mouse embryogenesis [120]. The main cell clusters are described on the left. The length of the bars corresponds to the normalised expression values for *Meis2* (A) and *Pbx1* (B). Cell clusters with normalised expression of *Hox* and other tissue-specific TFs > 0.01 are colour-coded for each TF.

monomeric HOX TFs prefer a TAAT recognition motif, heterodimers between Exd/PBX and HOX bind an extended TGATNNAT motif. SELEX-seq experiments have revealed that the sequence of the variable residues in this consensus motif is determined by the specific HOX protein present in a TALE:HOX complex, i.e. HOX proteins obtain novel recognition properties when they bind DNA as dimers with Exd [35]. Hth/MEIS/PREP proteins, which bind TGACAG motifs as monomers [24,40,41], also form heterodimers with Exd/PBX and this dimeric form prefers a TGATTGACAG recognition site [40,42]. Because Exd/PBX employs different protein domains to interact with Hth/MEIS/PREP versus HOX factors, this permits the formation of trimeric complexes [43–48]. Lastly, there is evidence that Hth/MEIS/PREP can interact with HOX proteins to form heterodimers independently of Exd/PBX [41]. Importantly, conclusions drawn from early in vitro binding studies are supported by functional analyses of HOX regulated enhancers in vivo. Both the mouse *Hoxb1* and *Hoxb2* loci contain PBX:HOX and MEIS/-PREP motifs bound by TALE-HOX heterotrimeric complexes that are required for enhancer activity in vivo [44,47,49]. Similarly, an autoregulatory element in the *Drosphila labial* enhancer requires binding of a Hth:Hox:Exd trimeric complex for in vivo function [45]. These studies demonstrate that TALE TFs form highly diverse heteromeric complexes with HOX proteins. Formation of such complexes enables HOX to recognise longer sequences (thereby solving the problem of selecting functional targets using a short TAAT target motif) and reveals a latent specificity of different HOX family members for different sequences (thereby solving the problem of conferring unique target specificities on individual HOX family members).

2.2. TALE occupancy overlaps extensively with HOX binding across the genome

Much of the initial characterization of TALE:HOX binding to DNA was carried out in vitro. This raises the question how TALE-HOX complexes function in vivo, where DNA recognition is challenged by large genomes, the compaction of DNA into chromatin and nuclei crowded with many other sequence-specific TFs and nuclear proteins. This issue has been addressed by ChIP-seq analyses defining HOX and TALE genomic occupancy in embryonic tissues and cultured cells. Consistent with the original discovery of MEIS proteins as involved in HOX-mediated leukemogenesis, there is extensive genomic co-occupancy between HOXB4 and MEIS at progressive stages of haematopoiesis [50] and HOXA9 binding overlaps with MEIS1 binding in transformed bone marrow cells [51]. In their physiological domains of expression in embryogenesis, HOXA2 and HOXA3 peaks almost completely coincide with regions occupied by MEIS and PBX in developing branchial arch 2 and the posterior branchial arches, respectively [52–54], and overlap between HOXC10 and MEIS peaks was observed in the developing mouse limb [55]. Whole genome binding studies in *Drosophila* revealed tissue-specific occupancy of Hth and the HOX protein Ubx in two tissues, haltere and T3 leg imaginal discs [56], and a close similarity between Hth and Ubx binding profiles was observed in the haltere disc [57]. Exd is also necessary for the HOX protein Sex combs reduced (Scr) to bind many of its paralog-specific targets in the first (T1) leg imaginal disc, whose final identity is specified by Scr [58]. All of these ChIP-seq analyses reveal an enrichment for the expected PBX:HOX and MEIS motifs at the ChIP peaks. Accordingly, PBX, MEIS and PREP ChIP-seq analyses of mouse embryo trunk [59], as well as PREP ChIP-seq in post-gastrula zebrafish embryos [60], identified both MEIS/PREP and PBX:HOX motifs associated with the ChIP peaks, further reinforcing an extensive overlap of TALE and HOX genomic occupancy in the developing embryo. Taken together, genome wide analyses in vivo largely confirm the in vitro studies by demonstrating a strong and reproducible pattern of HOX co-occupancy with TALE across the genome and by revealing binding at motifs containing the sequences defined in vitro.

2.3. TALE TFs modulate HOX binding site selectivity in vivo

The fact that HOX TFs bind broadly with TALE TFs genome-wide, raises the question if TALE TFs contribute to HOX binding specificity in a similar manner in vivo as was reported in vitro. The in vitro studies found that TALE:HOX complexes recognise longer, more specific sequences than HOX monomers (TGATNNAT versus TAAT) and this seems to be the case also in vivo because the extended PBX:HOX motif is the highest enriched motif observed in HOX ChIP-seq peaks. In vitro studies also revealed latent specificities of HOX TFs for slightly variant target sequences. In agreement, comparing HOXA2 and HOXA3 binding profiles in the mouse embryo highlights recognition of unique variants of the PBX:HOX motif by HOXA2 and HOXA3 [54]. Additional determinants of HOX paralog binding in vivo include differential affinity at shared PBX:HOX motifs and TALE:HOX cooperation with tissue-specific TFs [54]. This suggests that direct cooperativity of HOXA2 and HOXA3 with MEIS and PBX leads HOXA2 and HOXA3 to selectively target distinct subsets of binding sites, effectively directing HOXA2 and HOXA3 to different targets. In Drosophila leg imaginal discs, the Hox proteins Scr and Ubx share many sites across the genome. Similar to mouse HOX proteins, paralog (Scr)-binding specificity is mediated by Exd and by a non-TALE tissue-restricted TF, namely the homeodomain protein Distal-less (Dll), expressed in the distal leg domain [58]. Additional support for TALE increasing HOX binding selectivity comes from ChIP-seq data in mouse ESCs that were engineered for inducible HOXA1 expression [61]. Induction of HOXA1 revealed occupancy at thousands of regions, but HOXA1 binding levels were low and associated with open chromatin, likely indicating low affinity binding. When ESCs were induced to differentiate by retinoic acid (RA) treatment, a marked change in HOXA1 binding ensued such that most regions occupied in undifferentiated ESCs showed reduction or loss of binding, a subset retained occupancy, and a group of newly bound sites appeared in differentiated cells. Since RA treatment induces TALE expression, and the new group of HOXA1 sites was highly enriched in the TGATNNAT motif preferred by TALE:HOX complexes, it appears that TALE TFs can shift low affinity HOX binding to more spatially restricted and higher affinity binding in vivo.

The ability of TALE TFs to modulate HOX binding stems from HOX paralog proteins possessing distinct amino acid sequences immediately N-terminal to the HD. Upon heterodimerization with Exd/PBX, this region makes unique contacts in the minor grove of DNA, thereby providing a measure of selectivity [34,35]. PBX:HOX interactions involve a short HOX protein motif located upstream of the HD termed the hexapeptide [62]. While this motif is essential for HOX function in some in vivo [63–65] contexts, HOX proteins also use additional interaction motifs to interact with TALE, suggesting that the conformation of specific TALE:HOX complexes may be highly variable and could lead to additional diversification of DNA binding sites in a paralog as well as context-specific manner in vivo [66,67]. Further versatility derives from the ability of PBX and MEIS to participate in the formation of heterotrimeric complexes with HOX TFs as non-DNA-binding partners. In this case one TALE TF (as well as HOX) binds its recognition site on DNA, while the second TALE TF is recruited to the complex by protein-protein interactions [68]. This last observation could provide a partial explanation to the finding that the HD is dispensable for a substantial number of Hth functions during *Drosophila* development [69].

Functional association of HOX and TALE not only diversify TF:DNA interactions by providing each HOX with the ability to discriminate specific pool of targets. It can also underlie the evolution of new functions, as exemplified by the functional diversification of mouse HOX1 genes [70]. HOXA1 has retained essential ancestral functions of *Drosphila* Labial, while HOXB1 and HOXD1 have diverged. Genome wide analyses demonstrate similar DNA-binding patterns for HOXA1 and Labial in mouse cells, while HOXB1 binds to distinct targets. HOXA1 shows a greater degree of co-occupancy with PBX proteins on target sites and exists in the same complex with PBX on chromatin. It appears that

HOXB1 has evolved new functions by partly evading interaction with PBX. Interestingly, PBX1a, MEIS1a and HOXB1 form trimeric complexes on the *Hoxb2* and *Hoxb1* enhancers and, in the absence of MEIS, PBX1a and HOXB1 appear to form a stronger dimeric complex [44,47]. This may suggest that HOXB1 retains its ability to interact with PBX, but has perhaps evolved newer, higher affinity interactions with other factors (not available in the in vitro conditions tested). Overall, these analyses suggest that TALE directs HOX to its specific sites and contributes to HOX paralog-selective binding in vivo, largely in agreement with the models developed based on early in vitro work.

2.4. TALE TFs may facilitate HOX access to genomic binding sites

In addition to providing higher DNA binding selectivity to HOX proteins, TALE TFs may also act to facilitate HOX access to genomic binding sites in vivo. In support of this possibility, PREP and PBX TFs seem capable of accessing their binding sites in nucleosome-occupied DNA in zebrafish embryos [60] and breast cancer cells [71]. In addition, while the precise order of events responsible for loading TALE:HOX complexes onto chromatin has not been assessed comprehensively, it appears that TALE occupancy may precede HOX binding in vivo. Direct temporal analysis of TALE vs HOX occupancy at the *hoxb1a* enhancer during zebrafish embryogenesis demonstrated that PBX and PREP bind this element prior to HOX [72]. Once the TALE factors are bound, they may facilitate subsequent HOX binding, as indicated by the observation that MEIS occupies the same regions in HOX-negative (BA1) and HOX-positive (BA2) areas of the embryo [52]. Since PBX and PREP expression temporally precedes MEIS expression in different systems, PBX and PREP may be the first determinants for TALE binding in vivo. For instance, PBX and PREP are maternally expressed in zebrafish, while MEIS expression is observed after zygotic genome activation [60]. Similarly, PBX is expressed in mouse and human pluripotent embryonic stem cells, while MEIS expression appears later when cells commit to different fates [73,74]. PBX binding may help recruit MEIS, which is facilitated by 5-hydroxymethylcytosine (5hmC) modifications associated with chromatin opening and enhancer priming [75]. While protein-protein interactions likely underlie the recruitment of HOX to TALE-occupied sites, such interactions cannot be sufficient because HOX TFs appear to select only a subset of the sites that are occupied by TALE TFs and that contain a HOX recognition motif [52]. It is therefore likely that TALE TFs access their binding sites prior to the binding of HOX TFs – possibly in nucleosomal DNA — and provide additional functions required for HOX occupancy, such as local modifications to the chromatin state – as discussed further below.

While various genome wide analyses indicate that HOX can mainly (or only) bind in vivo targets together with TALE TFs, some HOX binding selectivity can be observed in the absence of TALE TFs and is strongly associated with chromatin accessibility [76]. In TALE-free nuclei (in the absence of Hth, which also prevents Exd nuclear entry), most HOX binding is detected on accessible chromatin, except for posterior HOX (Abd) that can occupy closed chromatin. Similarly, induction of HOXA1 by itself leads to low level binding at accessible chromatin regions in ESCs, whereas co-induction of TALE expression by treatment with RA leads to a gain in HOXA1 binding specificity characterized by a lower degree of occupancy (fewer regions bound) and with higher binding levels [61]. In the case of HOXA1, disrupting its in vivo interaction with PBX produces the same phenotype as deletion of the HOXA1 gene [65], suggesting that, at least in this case, TALE-independent binding is non-functional and may reflect non-specific interaction with accessible AT-rich DNA sites. Nevertheless, examples of functional HOX monomeric binding exist. Ubx binds as a monomer in a spatially restricted domain of the haltere disc, where it acts to decrease chromatin accessibility and repress transcription [77]. In line with these findings, Ubx directly represses the spalt (sal) gene in the absence of Exd and Hth and repression requires multiple monomer binding sites [78]. Similar to what was observed for HOXA1, Ubx monomer binding displays a weaker ChIP signal compared to Ubx binding in complex with Hth-Exd [77], suggesting that HOX monomeric binding is unstable unless reinforced by several, closely spaced motifs.

3. A broader role for TALE in cooperating with tissue-restricted TFs

The expression of TALE TFs not only precedes HOX TFs, but it is also broader – PBX and PREP family TFs are ubiquitously expressed and MEIS family TFs have been detected in most cells and tissues. These observations suggest that TALE TFs can also function independently of HOX. Indeed, genetic interactions in *Drosophila* have uncovered several instances of Exd/Hth HOX-independent function, including patterning of the notum [79] and induction of the antennae [80]. TALE TFs are also evolutionarily ancient [25] and may therefore have been available to function with other TFs during the evolutionary selection for tissue-restricted gene expression programs. This raises the possibility that, while TALE TFs were initially identified as HOX cofactors, they may actually act universally to facilitate the function of many (tissue restricted) TFs. Accordingly, TALE TFs are expressed in HOX-less tissues and many instances of TALE cooperation with non HOX TFs have been documented in various systems. PBX1 and MEIS2 bind with pancreas-specific PDX1 to a composite PBX:PDX1 site and increase PDX1-dependent transactivation upon transient transfection [81]. PBX1 also binds genomic targets with Estrogen Receptor (ER) in a breast cancer cell line [71]. In the muscle lineage, PBX is constitutively bound at the inactive *Myog* promoter and stabilizes binding of the muscle-specific MYOD TF via protein-protein interactions, leading to induction of *Myog* expression [82]. Indeed, the ability to bind DNA cooperatively with PBX:MEIS/PREP is a general property of all myogenic bHLH:E2a heterodimers [83]. In the mouse brain, MEIS2 interacts with non-HOX HD TFs. Specifically, in the developing midbrain, MEIS2, PAX3 and PAX7 are found in a complex whose functional importance is supported by genetic interaction [84]. In the subventricular zone, MEIS2 interacts with PAX6 and DLX2 [85]. Intriguingly, in the chick tectal anlage, MEIS2 binds to OTX2 in the absence of DNA and competes with the co-repressor TLE4, arguing for DNA-independent controlled assembly and disassembly of transcription regulator complexes by MEIS TFs [86]. No HOX proteins have been implicated in *Drosophila* eye development. Instead, Hth functions in a complex with Eyeless (ey) and Teashirt (Tsh) to promote cell proliferation in the eye discs and at the same time to prevent premature differentiation by repressing downstream TFs [87]. This function is fully integrated with the Hippo pathway, which controls organ proliferation by regulating nuclear availability of the transcriptional coactivator Yorkie (Yki) [88]. Yki is normally directed to DNA by its DNA binding partner Scalloped (Sd). However, Ykie uses Hth instead of Sd in the anterior eye disc and Hth, Tsh and Yki interact to directly regulate expression of the bantam microRNA, a Hippo pathway target.

While these individual examples of functional interactions between TALE and other TFs are informative, it remains unclear how broadly TALE proteins interact with non-HOX TFs. Recent genome-wide analyses have begun to reveal the full extent of such interactions. Mapping Yki, Sd, and Hth in the progenitors of the eye-antenna structures and the wing of the fly corroborates their genetic interactions and uncover integrated regulatory strategies for Sd and Hth, one shared between wing and eye and associated with Yki to control cell proliferation, the other TF- and tissue-specific to control developmental patterning [89]. Genetic evidence indicates an interaction between PBX1 and HAND2 in the developing mouse hindlimb [90]. Accordingly, ChIP-seq analysis revealed that HAND2 peaks largely overlap with PBX1 peaks, are enriched in HAND and PBX binding motifs, and are associated with accessible chromatin and H3K27ac enrichment [90]. Genomic regions co-occupied by HAND2 and PBX1 are also associated with genes regulating limb development and morphogenesis, indicating that HAND2:PBX1 cooperation selectively instructs a distinct developmental

program. Other genome wide analyses introduce further variations on the theme of TALE acting with various individual tissue specific TFs. The HOXA:MEIS1 transcriptional program plays a key role in MLL-mediated leukemogenicity, but, in this context, the hematopoietic regulator IKAROS displays functional cooperativity and extensive chromatin co-occupancy with MEIS1:HOXA9–10. Proximity-dependent biotin identification (BioID) identified a striking overlap of direct and proximal interacting proteins of IKAROS, MEIS1 and HOXA10, indicating that IKAROS is physically associated with TALE and HOX TFs in a larger complex [91]. In contrast with the examples above – which describe a positive cooperation of TALE with other TFs – the overlapping genomic occupancy of MEIS1 and NKX2–5 reflects mutually exclusive binding. This results in spatial and temporal synchronization of binding at a common pool of targets between NKX2–5 and MEIS1 [92]. In this scenario, early binding by MEIS1 would activate genes expressed in cardiac progenitors. Subsequently, as cardiac precursors integrate into the developing heart and NKX2–5 expression is activated, replacement of MEIS1 by NKX2–5 would repress these progenitor genes, allowing cardiomyocyte differentiation. Lastly, ChIP-seq in early zebrafish embryos revealed PBX and PREP co-occupancy at genomic sites prior to zygotic genome activation. At this stage, PBX:PREP bind a heterodimeric motif that is not associated with HOX binding sites. Instead, the TALE-occupied sites are associated with binding sites for NF-Y, a ubiquitously expressed TF [60,93]. TALE:NF-Y co-occupied genomic elements appear to function as enhancers since they are associated with the H3K27ac active chromatin mark and drive tissue-specific GFP expression in zebrafish embryos [93]. As a group, these enhancers are not enriched for tissue-specific TF binding sites, but since each individual enhancer displays a distinct expression pattern, it is plausible that TALE:NF-Y is associated with a distinct, spatially restricted TF at each enhancer tested.

TALE interactions with HOX TFs affect HOX binding specificity by modulating binding of the HOX N-terminal arm in the minor groove of DNA. There is limited evidence that interactions with TALE proteins enhance the binding specificity of non-HOX TFs, like they do for HOX TFs. This is based on structural analysis of MEIS1 and DLX3 bound to their identified recognition site. Interactions between MEIS1 and DLX3 are predominantly mediated by DNA, with recognition of a composite site that is different from the motifs preferred by either TF alone [94]. There is also emerging evidence that TALE TFs provide access for non-HOX TFs at genomic sites, similar to what was observed for HOX TFs. In a breast cancer model, genomic binding of PBX1 precedes Estrogen Receptor (ER) recruitment to DNA and PBX1 depletion blocks ER binding [71]. Similarly, PBX marks myogenic genes prior to MYOD binding [82], demonstrating that TALE binding precedes, and likely facilitates, binding also of non-HOX TFs. Accordingly, PBX:PREP complexes are observed genome wide in the zebrafish embryo at the time of zygotic gene activation, which is well before the expression of most tissue-specific TFs [60]. The model that emerges from these and other studies is that TALE occupancy precedes both HOX and non-HOX TF binding such that TALE TFs access their binding site in nucleosomal DNA and facilitate subsequent binding of other TFs. TALE:TF association is therefore likely to broadly inform DNA binding selectivity in vivo.

4. TALE factors may prime chromatin for activation of lineage-specific programs

The existing data support a model in which domain- and lineage-restricted TFs (HOX and non-HOX) require pre-bound TALE TFs to carry out their function. In this scenario, initial TALE binding may represent a key priming event that promotes a permissive chromatin state for binding of tissue-restricted TFs. The mechanisms whereby TALE factors achieve such a permissive chromatin state appear to be several-fold. First, TALE TFs can bind chromatin modifying enzymes such as HDACs, HATs and CBP, where the specific enzyme bound appears to depend on the composition of the TALE complex [72,95–97]. Accordingly, TALE:TF complex binding is associated with open chromatin regions that bear active histone marks in vivo. In transformed bone marrow cells, HOXA9 and MEIS1 co-occupied sites are enriched in active chromatin modifications and overlap binding of p300 and CBP [51], as well as C/EBPα [98]. Similarly, PBX1:ER complexes are bound at open chromatin regions in breast cancer cells [71] and TALE:HOXA1 complexes are associated with open chromatin bearing active histone marks in differentiated ES cells engineered to express HOXA1 [61]. MEIS1 binding is also associated with 5-hydroxymethylcytosine (5hmC) at distal enhancers, a DNA modification which dynamically correlates with cell-specific enhancer activity [99]. As expected, such TALE:TF complexes appear to be active since almost 50% of HOXA2:MEIS high confidence binding in the branchial arches is associated with genes downregulated in HOXA2 mutants [100]. Lastly, depletion of TALE TFs reduces active chromatin marks at enhancers during zebrafish development [60,72], further indicating that TALE factors recruit chromatin modifying enzymes in vivo. Second, TALE proteins may cooperate with other ubiquitous TFs to promote accessible chromatin. In the early zebrafish embryo, 25–30% of all TALE binding sites are co-occupied by NF-Y, a trimeric TF that possesses the ability to bind and open nucleosomal DNA [101], and depletion of NF-Y leads to reduced H3K27ac levels at TALE:NF-Y co-occupied regions [60].

The fact that TALE TFs are capable of recruiting chromatin modifying enzymes and co-regulators raises the question of what role is provided by the tissue-restricted TFs that are part of the TALE:TF complex. One possibility is that TALE occupancy is not sufficient for full activity, and that recruitment of tissue-restricted TFs may provide the missing component to initiate/sustain active transcription. Indeed, temporal analysis of TALE:HOX complex assembly on a zebrafish enhancer indicated that TALE factors recruit histone modifying enzymes and the RNAPII machinery in vivo [72]. This is likely mediated, at least in part, by MEIS interactions with PARP1, which can actively displace histone H1 and its associated nucleosome from promoters to facilitate loading of RNAPII [102]. Such TALE-mediated RNAPII recruitment may not be sufficient for optimal transcription until a HOX protein also binds, which triggers P-TEFb-mediated transitioning of RNAPII to the serine 2-phosphorylated form and efficient transcription [72]. It is unclear how the HOX protein mediates this change, but many HOX proteins possess transcription activation domains [103–105] that may act to recruit additional components. Complex formation may also permit HOX activation domains to cooperate with similar domains reported in TALE TFs. Accordingly, the MEIS1A C terminus harbours a transcriptional activation domain that is indispensable for its oncogenic properties and addition of the MEIS1 C-terminal domain transforms PREP1 into a HOXA9- collaborating oncoprotein [106]. A second, not mutually exclusive, possibility is that the tissue-restricted TFs help stabilize the TALE:TF complex. Indeed, in addition to TALE TFs promoting HOX binding, HOX TFs exert reciprocal effects on TALE TFs. Therefore, relative to TALE only, TALE:TF complexes display not only the addition of HOX (or non HOX) TFs, but also a reinforced, more stable binding of TALE. MEIS TFs bind broadly and to largely overlapping locations across different branchial arches [52,54], but enhanced TALE binding is observed close to HOXA2-occupied sites in the branchial arch 2 and these sites are highly enriched in MEIS and HOX recognition motifs. Further, the addition of HOXA2 to HOX-negative cells reinforces MEIS binding at selected sites [100]. Similar observations extend to non HOX TFs: while PBX bind to largely overlapping locations across different embryonic domains, enhanced PBX peaks in the limb overlaps HAND2-occupied sites [90]. In contrast to ubiquitous low-level TALE occupancy, high MEIS and PBX binding levels in a tissue are associated with distinct biological processes and are highly predictive of increased enhancer activity in the same tissue [54]. High TALE peaks are also enriched in tissue-specific motifs, raising the possibility that the function of HOX (and tissue-specific TFs) is to selectively convert the ubiquitous pool of low affinity TALE binding events into high confidence, tissue-specific binding events [54,107]. In this scenario, the increase in

TALE residence time on DNA, driven by cooperativity with tissue-specific TFs, may represent the switch for transcriptional activation.

The model that emerges from these and other studies is that TALE TFs create a permissive chromatin platform that is selected by tissue-restricted TFs for binding. In turn, HOX and tissue-restricted TFs selectively convert a ubiquitous pool of low affinity TALE binding events into high confidence, tissue-restricted binding events associated with transcriptional activation. As a result, TALE:TF complexes are associated with active chromatin and domain/lineage-specific gene activity.

5. A working model for TALE TF function in vivo

As summarized in this review, genome-wide in vivo analyses extend prior in vitro work by demonstrating a role for TALE TFs beyond their initial characterization as HOX cofactors. Not only do TALE TFs act together with many tissue-restricted TFs in addition to HOX TFs, but they also appear to fulfil important roles in establishing a permissive chromatin environment for the activity of such TFs. We propose a working model (Fig. 2) where TALE TFs access their binding sites in nucleosomal DNA (top panel). TALE can also recruit key transcriptional co-regulators (not shown) but initial binding is suboptimal and is not sufficient for transcription. This initial TALE binding becomes stabilized when a tissue-restricted TF binds an adjacent site on DNA. Because different cells express different lineage-restricted TFs, TALE:TF complexes become stabilized at different genomic locations in different cell types. A more stable TALE:TF binding equates to longer TF residence time on chromatin, thereby facilitating the recruitment of general co-activators and active transcription. The individual contributions of TALE and lineage-restricted TFs to transcriptional activation are not yet resolved. In this model, tissue-restricted TFs cannot efficiently access and bind their targets in the absence of the TALE TF, making initial TALE binding an essential priming event.

6. Controversial and unexplored aspects of TALE function

Our working model accounts for data available from in vitro and in vivo analyses and point to a general role for TALE TFs in facilitating chromatin access for other TFs and initiating lineage-specific transcriptional programs. However, several aspects of this model remain unresolved and controversial. Here we briefly summarise some areas that need to be explored further.

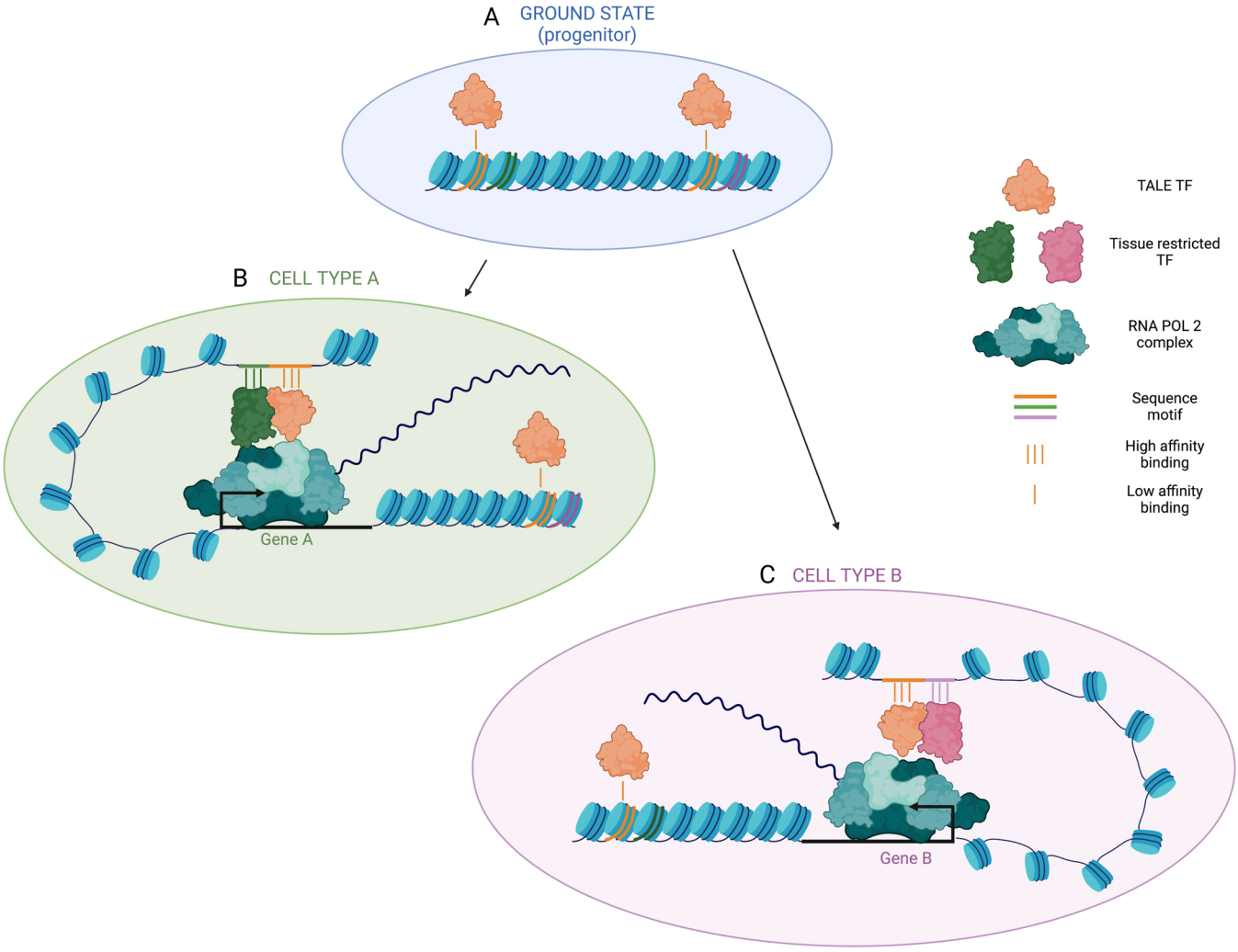


Fig. 2. Working model for TALE function. **A.** In the progenitor state, TALE TFs (orange) access their binding sites in nucleosomal DNA with low affinity (single vertical bar). **B, C.** TALE TFs facilitate recruitment of distinct tissue-restricted TFs in different cells (green in B, pink in C) to nearby binding sites. This stabilizes binding to DNA at co-occupied sites (three vertical bars), but TALE remains weakly bound at sites that lack a tissue-restricted TF. Stabilized TALE:TF complexes recruit the transcription machinery to associated promoters and initiate transcription, but weakly bound TALE TFs cannot support transcription. Thus, tissue-restricted TFs harness TALE TFs to drive tissue-specific gene expression programs.

6.1. Does TALE:TF complex composition affect binding preference and function?

Different TALE family members recruit different co-activators or co-repressors [72,95–97,108], suggesting that the function of TALE:TF complexes may vary depending on their composition. In mouse embryonic fibroblasts, *MEIS1* acts as an oncoprotein while *PREP1* exerts a tumor-suppressive function [109–111]. These opposite behaviours are likely to reflect differential abilities to activate transcription. Indeed, addition of the MEIS1 activation domain converts PREP1 into an oncogene [106]. Also supporting MEIS ability to activate transcription, MEIS expression appears as the embryonic genome starts to be transcribed [112–114]. In addition, as expression of TALE TFs is regulated in some systems, the composition of TALE:TF complexes may be under temporal control. It is therefore highly likely that variations in composition and function in TALE:TF complexes are important for activity in vivo. Accordingly, TALE TFs interact with numerous auxiliary factors (reviewed in [115,116]), but the functional consequences of these interactions have not been examined comprehensively.

Diverse composition of TALE complexes may also underlie the choice of binding site. TALE TFs bind DNA poorly as monomers in vitro, suggesting that this may not be the favoured arrangement in vivo. However, examination of PBX and MEIS ChIP-seq in diverse embryonic tissues identifies the monomeric TGACAG motif as the most enriched site. TGACAG tends to occur in clusters in MEIS and PBX peaks and could therefore support the formation of dimeric (MEIS-MEIS, PBX:PBX, MEIS:PBX) complexes. In contrast, ChIP-seq analyses of PREP and PBX in the early zebrafish embryo (prior to the onset of tissue-restricted gene expression) revealed occupancy primarily at heterodimeric TGATTGACAG sites, with limited contribution by monomeric sites [60,93]. This observation is suggestive of heterodimeric sites possibly playing a special role in the early binding of PBX:PREP during embryogenesis, but this needs to be examined in additional systems and at additional stages of embryogenesis. Coupled with the significant diversity within the TALE family, TALE factors can be present either as monomers or heterodimers when complexed with other TFs, indicating that many different types of TALE:TF complexes could potentially be present in vivo. Addressing each individual contribution of PBX, PREP and MEIS in the same context (organism, cell type) is a key requisite to understand TALE role in transcription networks.

6.2. Do TALE TFs modulate the chromatin state?

There has been recent debate whether an ability to access binding sites in nucleosomal DNA is a unique property of a specialized class of TFs (termed ‘pioneer factors’) with the ability to initiate opening of closed chromatin. This topic has been reviewed recently [117] and it has been suggested that TALE TFs belong to this class [118]. While the precise definition of pioneer factors is still under debate, there is clear evidence that TALE TFs occupy their binding sites prior to other TFs, and that they recruit chromatin modifying enzymes. This makes them strong candidates for a pioneering role, but the mechanistic details of this role remain to be uncovered. For instance, it is unclear if this property is a prerogative of PBX and PREP, which bind together at early embryonic stages, or if it is fully shared by MEIS.

Cellular chromatin is organised into compartments and domains, which facilitate (or prohibit) enhancer-enhancer and enhancer-promoter interactions, thereby providing an important regulatory control of lineage-specific transcription. Similar to TALE occupancy, chromatin domains are largely pre-established and independent of the specific cell or tissue. Given the apparent ability of TALE to provide broad chromatin scaffolds that are recognised by domain- or tissue-restricted TFs, future investigations should explore TALE binding in a 3D context (using 3 C-based methods) and the effects of TALE binding levels on enhancer-to-enhancer and enhancer-to-promoter interactions. Based on the higher occupancy of PBX and PREP at promoters, while MEIS occupies mainly distal regions an intriguing possibility is that MEIS/PBX dimerization (or a switch from PBX:PREP to PBX:MEIS) could link enhancers with their target promoters, a necessary step in transcriptional activation.

6.3. Do non-HOX TFs stabilize TALE DNA binding?

The functional repertoire of TALE TFs clearly extends beyond its originally proposed HOX “cofactor” role, but it remains unclear if TALE plays the same role when complexed with other TFs. Complex formation between TALE and HOX TFs leads to stabilization of DNA binding. This is apparent when comparing overall MEIS occupancy – which is shared across different branchial arches – to high MEIS binding levels, which are branchial arch-specific. While this stabilization coincides with the expression of different HOX TFs in different branchial arches, it also reflects the distribution of other TFs across these tissues [52,54,119]. These observations suggest that binding of tissue-specific TFs, in addition to HOX, can locally increase MEIS binding levels, although the ability of non HOX TFs to influence MEIS binding requires a more direct assessment. Specifically, it will be important to establish whether cooperative binding of TALE with non HOX-TFs requires direct protein-protein interaction or, alternatively, if it can be mediated by indirect cooperativity (when TFs bind to closely spaced sites, they derive a mutual advantage in displacing local nucleosomes, or changes in DNA conformation after binding of one TF may facilitate binding of additional TFs). A role for indirect cooperativity could be particularly important because bypassing the requirement for protein-protein interactions would significantly expand TALE cooperativity with any TF family.

7. Conclusions

The parameters of TALE:HOX complex formation and DNA binding specificity established in vitro are highly informative and have been largely confirmed on a genome-wide scale. Nevertheless, recent genome-wide analyses in vivo have also significantly expanded our understanding of the full repertoire of TALE activities. This includes cooperation with non-HOX tissue-restricted TFs and a broader role for TALE TFs in priming chromatin for the activation of tissue- and lineage-specific transcriptional programs. Describing TALE TFs as “HOX cofactors” is therefore too restrictive since it does not convey the fact that TALE TFs act broadly with non-HOX TFs and contribute functions that are equally important as the functions of other TFs in TALE:TF complexes.

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council grant BB/T007761/1 to NB and National Institutes of Health grants R01NS038183, R01NS124166 and R01GM142158 to CGS.

Declaration of Competing Interest

The authors have no competing interests to declare.

Data availability

No data was used for the research described in the article.

Acknowledgements

We are grateful to members of our laboratories for helpful discussions and apologize to colleagues whose work was not discussed due to space limitations.

Data access statement

No data were created or analysed in this study.

References

[1] D. Schulte, D. Geerts, MEIS transcription factors in development and disease, Development 146 (2019) 16.
[2] L. Selleri, V. Zappavigna, E. Ferretti, 'Building a perfect body': control of vertebrate organogenesis by PBX-dependent regulatory networks, Genes Dev. 33 (5–6) (2019) 258–275.
[3] C. McQueen, M. Towers, Establishing the pattern of the vertebrate limb, Development 147 (2020) 17.
[4] J.C. Hombria, B. Lovegrove, Beyond homeosis–HOX function in morphogenesis and organogenesis, Differentiation 71 (8) (2003) 461–476.
[5] Z. Afzal, R. Krumlauf, Transcriptional regulation and implications for controlling hox gene expression, J. Dev. Biol. 10 (2022) 1.
[6] R. Morgan, K. Hunter, H.S. Pandha, Downstream of the HOX genes: explaining conflicting tumour suppressor and oncogenic functions in cancer, Int J. Cancer 150 (12) (2022) 1919–1932.
[7] T.R. Burglin, M. Affolter, Homeodomain proteins: an update, Chromosoma 125 (3) (2016) 497–521.
[8] J.C. Pearson, D. Lemons, W. McGinnis, Modulating Hox gene functions during animal body patterning, Nat. Rev. Genet 6 (12) (2005) 893–904.
[9] H.J. Parker, R. Krumlauf, A Hox gene regulatory network for hindbrain segmentation, Curr. Top. Dev. Biol. 139 (2020) 169–203.
[10] H.J. Parker, I. Pushel, R. Krumlauf, Coupling the roles of Hox genes to regulatory networks patterning cranial neural crest, Dev. Biol. 444 (Suppl 1) (2018) S67–S78.
[11] M.B. Noyes, et al., Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites, Cell 133 (7) (2008) 1277–1289.
[12] M.F. Berger, et al., Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences, Cell 133 (7) (2008) 1266–1276.
[13] R.S. Mann, K.M. Lelli, R. Joshi, Hox specificity unique roles for cofactors and collaborators, Curr. Top. Dev. Biol. 88 (2009) 63–101.
[14] M. Peifer, E. Wieschaus, Mutations in the Drosophila gene extradenticle affect the way specific homeo domain proteins regulate segmental identity, Genes Dev. 4 (7) (1990) 1209–1223.
[15] C. Rauskolb, M. Peifer, E. Wieschaus, extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene pbx1, Cell 74 (6) (1993) 1101–1112.
[16] C. Rauskolb, E. Wieschaus, Coordinate regulation of downstream genes by extradenticle and the homeotic selector proteins, EMBO J. 13 (15) (1994) 3561–3569.
[17] G.E. Rieckhof, et al., Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein, Cell 91 (2) (1997) 171–183.
[18] E. Kurant, et al., Dorsotonals/homothorax, the Drosophila homologue of meis1, interacts with extradenticle in patterning of the embryonic PNS, Development 125 (6) (1998) 1037–1048.
[19] C.Y. Pai, et al., The Homothorax homeoprotein activates the nuclear localization of another homeoprotein, extradenticle, and suppresses eye development in Drosophila, Genes Dev. 12 (3) (1998) 435–446.
[20] M.P. Kamps, et al., A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL, Cell 60 (4) (1990) 547–555.
[21] J. Nourse, et al., Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor, Cell 60 (4) (1990) 535–545.
[22] J.J. Moskow, et al., Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice, Mol. Cell Biol. 15 (10) (1995) 5434–5443.
[23] T. Nakamura, N.A. Jenkins, N.G. Copeland, Identification of a new family of Pbx-related homeobox genes, Oncogene 13 (10) (1996) 2235–2242.
[24] J. Berthelsen, et al., Prep1, a novel functional partner of Pbx proteins, EMBO J. 17 (5) (1998) 1423–1433.
[25] K. Mukherjee, T.R. Burglin, Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution, J. Mol. Evol. 65 (2) (2007) 137–153.
[26] S.K. Chan, et al., The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein, Cell 78 (4) (1994) 603–615.
[27] M.A. van Dijk, C. Murre, extradenticle raises the DNA binding specificity of homeotic selector gene products, Cell 78 (4) (1994) 617–624.
[28] C.P. Chang, et al., Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins, Genes Dev. 9 (6) (1995) 663–674.
[29] P.S. Knoepfler, M.P. Kamps, The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1, Mol. Cell Biol. 15 (10) (1995) 5811–5819.
[30] Q. Lu, et al., Both Pbx1 and E2A-Pbx1 bind the DNA motif ATCAATCAA cooperatively with the products of multiple murine Hox genes, some of which are themselves oncogenes, Mol. Cell Biol. 15 (7) (1995) 3786–3795.
[31] D.E. Piper, et al., Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation, Cell 96 (4) (1999) 587–597.
[32] J.M. Passner, et al., Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex, Nature 397 (6721) (1999) 714–719.
[33] N.A. LaRonde-LeBlanc, C. Wolberger, Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior, Genes Dev. 17 (16) (2003) 2060–2072.
[34] R. Joshi, et al., Functional specificity of a Hox protein mediated by the recognition of minor groove structure, Cell 131 (3) (2007) 530–543.
[35] M. Slattery, et al., Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins, Cell 147 (6) (2011) 1270–1282.
[36] C.P. Chang, et al., Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus, Mol. Cell Biol. 16 (4) (1996) 1734–1745.
[37] W.F. Shen, et al., Hox homeodomain proteins exhibit selective complex stabilities with Pbx and DNA, Nucleic Acids Res 24 (5) (1996) 898–906.
[38] Q. Lu, M.P. Kamps, Heterodimerization of Hox proteins with Pbx1 and oncoprotein E2a-Pbx1 generates unique DNA-binding specifities at nucleotides predicted to contact the N-terminal arm of the Hox homeodomain–demonstration of Hox-dependent targeting of E2a-Pbx1 in vivo, Oncogene 14 (1) (1997) 75–83.
[39] S.T. Neuteboom, C. Murre, Pbx raises the DNA binding specificity but not the selectivity of antennapedia Hox proteins, Mol. Cell Biol. 17 (8) (1997) 4696–4706.
[40] C.P. Chang, et al., Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins, Mol. Cell Biol. 17 (10) (1997) 5679–5687.
[41] W.F. Shen, et al., AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins, Mol. Cell Biol. 17 (11) (1997) 6448–6458.
[42] P.S. Knoepfler, et al., Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1, Proc. Natl. Acad. Sci. USA 94 (26) (1997) 14553–14558.
[43] J. Berthelsen, et al., The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity, EMBO J. 17 (5) (1998) 1434–1445.
[44] Y. Jacobs, C.A. Schnabel, M.L. Cleary, Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity, Mol. Cell Biol. 19 (7) (1999) 5134–5142.
[45] H.D. Ryoo, R.S. Mann, The control of trunk Hox specificity and activity by Extradenticle, Genes Dev. 13 (13) (1999) 1704–1716.
[46] N. Vlachakis, D.R. Ellstrom, C.G. Sagerstrom, A novel pbx family member expressed during early zebrafish embryogenesis forms trimeric complexes with Meis3 and Hoxb1b, Dev. Dyn. 217 (1) (2000) 109–119.
[47] E. Ferretti, et al., Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins, Development 127 (1) (2000) 155–166.
[48] W.F. Shen, et al., HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells, Mol. Cell Biol. 19 (4) (1999) 3051–3061.
[49] H. Popperl, et al., Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx, Cell 81 (7) (1995) 1031–1042.
[50] R. Fan, et al., Dynamic HoxB4-regulatory network during embryonic stem cell differentiation to hematopoietic cells, Blood 119 (19) (2012) e139–e147.
[51] Y. Huang, et al., Identification and characterization of Hoxa9 binding sites in hematopoietic cells, Blood 119 (2) (2012) 388–398.
[52] S. Amin, et al., Hoxa2 selectively enhances Meis binding to change a branchial arch ground state, Dev. Cell 32 (3) (2015) 265–277.
[53] I.J. Donaldson, et al., Genome-wide occupancy links Hoxa2 to Wnt-beta-catenin signaling in mouse embryonic development, Nucleic Acids Res. 40 (9) (2012) 3990–4001.
[54] L. Bridoux, et al., HOX paralogs selectively convert binding of ubiquitous transcription factors into tissue-specific patterns of enhancer activation, PLoS Genet 16 (12) (2020), e1009162.
[55] I. Delgado, et al., Control of mouse limb initiation and antero-posterior patterning by Meis transcription factors, Nat. Commun. 12 (1) (2021) 3086.
[56] M. Slattery, et al., Genome-wide tissue-specific occupancy of the Hox protein Ultrabithorax and Hox cofactor Homothorax in Drosophila, PLoS One 6 (4) (2011), e14686.
[57] S.W. Choo, R. White, S. Russell, Genome-wide analysis of the binding of the Hox protein Ultrabithorax and the Hox cofactor Homothorax in Drosophila, PLoS One 6 (4) (2011), e14778.
[58] S. Feng, et al., Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors, Nat. Commun. 13 (1) (2022) 3808.
[59] D. Penkov, et al., Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins, Cell Rep. 3 (4) (2013) 1321–1333.
[60] F. Ladam, et al., TALE factors use two distinct functional modes to control an essential zebrafish gene expression program, Elife (2018) **7**.
[61] B. De Kumar, et al., HOXA1 and TALE proteins display cross-regulatory interactions and form a combinatorial binding code on HOXA1 targets, Genome Res 27 (9) (2017) 1501–1512.
[62] (In) P.M. der Rieden, et al., Homeodomain to hexapeptide or PBC-interaction-domain distance: size apparently matters, Trends Genet 20 (2) (2004) 76–79.
[63] O. Medina-Martinez, R. Ramirez-Solis, In vivo mutagenesis of the Hoxb8 hexapeptide domain leads to dominant homeotic transformations that mimic the loss-of-function mutations in genes of the Hoxb cluster, Dev. Biol. 264 (1) (2003) 77–90.
[64] S. Merabet, et al., The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions, Dev. Cell 4 (5) (2003) 761–768.

[65] S. Remacle, et al., Loss of function but no gain of function caused by amino acid substitutions in the hexapeptide of Hoxa1 in vivo, Mol. Cell Biol. 24 (19) (2004) 8567–8575.
[66] A. Dard, et al., Human HOX Proteins Use Diverse and Context-Dependent Motifs to Interact with TALE Class Cofactors, Cell Rep. 22 (11) (2018) 3058–3071.
[67] B. Hudry, et al., Hox proteins display a common and ancestral ability to diversify their interaction mode with the PBC class cofactors, PLoS Biol. 10 (6) (2012), e1001351.
[68] K. Shanmugam, et al., PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins, Mol. Cell Biol. 19 (11) (1999) 7577–7588.
[69] B. Noro, et al., Distinct functions of homeodomain-containing and homeodomain-less isoforms encoded by homothorax, Genes Dev. 20 (12) (2006) 1636–1650.
[70] N.P. Singh, R. Krumlauf, Diversification and Functional Evolution of HOX Proteins, Front Cell Dev. Biol. 10 (2022), 798812.
[71] L. Magnani, et al., PBX1 genomic pioneer function drives ERalpha signaling underlying progression in breast cancer, PLoS Genet 7 (11) (2011), e1002368.
[72] S.K. Choe, F. Ladam, C.G. Sagerstrom, TALE factors poise promoters for activation by Hox proteins, Dev. Cell 28 (2) (2014) 203–211.
[73] K.K. Chan, et al., KLF4 and PBX1 directly regulate NANOG expression in human embryonic stem cells, Stem Cells 27 (9) (2009) 2114–2125.
[74] S. Cinghu, et al., Intragenic Enhancers Attenuate Host Gene Expression, Mol. Cell 68 (1) (2017) 104–117, e6.
[75] E.A. Mahe, et al., Cytosine modifications modulate the chromatin architecture of transcriptional enhancers, Genome Res 27 (6) (2017) 947–958.
[76] D. Porcelli, et al., Chromatin accessibility plays a key role in selective targeting of Hox proteins, Genome Biol. 20 (1) (2019) 115.
[77] R. Loker, J.E. Sanner, R.S. Mann, Cell-type-specific Hox regulatory strategies orchestrate tissue identity, Curr. Biol. 31 (19) (2021) 4246–4255, e4.
[78] R. Galant, C.M. Walsh, S.B. Carroll, Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites, Development 129 (13) (2002) 3115–3126.
[79] S. Aldaz, G. Morata, N. Azpiazu, Patterning function of homothorax/extradenticle in the thorax of Drosophila, Development 132 (3) (2005) 439–446.
[80] F. Casares, R.S. Mann, Control of antennal versus leg development in Drosophila, Nature 392 (6677) (1998) 723–726.
[81] Y. Liu, R.J. MacDonald, G.H. Swift, DNA binding and transcriptional activation by a PDX1.PBX1b.MEIS2b trimer and cooperation with a pancreas-specific basic helix-loop-helix complex, J. Biol. Chem. 276 (21) (2001) 17985–17993.
[82] C.A. Berkes, et al., Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential, Mol. Cell 14 (4) (2004) 465–477.
[83] P.S. Knoepfler, et al., A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1, Nucleic Acids Res 27 (18) (1999) 3752–3761.
[84] Z. Agoston, et al., Genetic and physical interaction of Meis2, Pax3 and Pax7 during dorsal midbrain development, BMC Dev. Biol. 12 (2012) 10.
[85] Z. Agoston, et al., Meis2 is a Pax6 co-factor in neurogenesis and dopaminergic periglomerular fate specification in the adult olfactory bulb, Development 141 (1) (2014) 28–38.
[86] Z. Agoston, D. Schulte, Meis2 competes with the Groucho co-repressor Tle4 for binding to Otx2 and specifies tectal fate without induction of a secondary midbrain-hindbrain boundary organizer, Development 136 (19) (2009) 3311–3322.
[87] J. Bessa, et al., Combinatorial control of Drosophila eye development by eyeless, homothorax, and teashirt, Genes Dev. 16 (18) (2002) 2415–2427.
[88] H.W. Peng, M. Slattery, R.S. Mann, Transcription factor choice in the Hippo signaling pathway: homothorax and yorkie regulation of the microRNA bantam in the progenitor domain of the Drosophila eye imaginal disc, Genes Dev. 23 (19) (2009) 2307–2319.
[89] M. Slattery, et al., Divergent transcriptional regulatory logic at the intersection of tissue growth and developmental patterning, PLoS Genet 9 (9) (2013), e1003753.
[90] Losa, M., et al., *A spatio-temporally constrained gene regulatory network directed by PBX1/2 acquires limb patterning specificity* via *HAND2.* bioRxiv, 2022.
[91] B.J. Aubrey, et al., IKAROS and MENIN coordinate therapeutically actionable leukemogenic gene expression in MLL-r acute myeloid leukemia, Nat. Cancer 3 (5) (2022) 595–613.
[92] L. Dupays, et al., Sequential Binding of MEIS1 and NKX2-5 on the Popdc2 Gene: A Mechanism for Spatiotemporal Regulation of Enhancers during Cardiogenesis, Cell Rep. 13 (1) (2015) 183–195.
[93] W. Stanney 3rd, et al., Combinatorial action of NF-Y and TALE at embryonic enhancers defines distinct gene expression programs during zygotic genome activation in zebrafish, Dev. Biol. 459 (2) (2020) 161–180.
[94] A. Jolma, et al., DNA-dependent formation of transcription factor pairs alters their binding specificity, Nature 527 (7578) (2015) 384–388.
[95] H. Asahara, et al., Pbx-Hox heterodimers recruit coactivator-corepressor complexes in an isoform-specific manner, Mol. Cell Biol. 19 (12) (1999) 8219–8225.
[96] M. Saleh, et al., Cell signaling switches HOX-PBX complexes from repressors to activators of transcription mediated by histone deacetylases and histone acetyltransferases, Mol. Cell Biol. 20 (22) (2000) 8623–8633.
[97] S.K. Choe, et al., Meis cofactors control HDAC and CBP accessibility at Hox-regulated promoters during zebrafish embryogenesis, Dev. Cell 17 (4) (2009) 561–567.
[98] C. Collins, et al., C/EBPalpha is an essential collaborator in Hoxa9/Meis1-mediated leukemogenesis, Proc. Natl. Acad. Sci. USA 111 (27) (2014) 9899–9904.
[99] A.A. Serandour, et al., Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers, Nucleic Acids Res. 40 (17) (2012) 8255–8265.
[100] S. Amin, et al., Hoxa2 Selectively Enhances Meis Binding to Change a Branchial Arch Ground State, Dev. Cell 32 (3) (2015) 265–277.
[101] A.J. Oldfield, et al., Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors, Mol. Cell 55 (5) (2014) 708–722.
[102] A.C. Hau, et al., MEIS homeodomain proteins facilitate PARP1/ARTD1-mediated eviction of histone H1, J. Cell Biol. 216 (9) (2017) 2715–2729.
[103] G. Di Rocco, F. Mavilio, V. Zappavigna, Functional dissection of a transcriptionally active, target-specific Hox-Pbx complex, EMBO J. 16 (12) (1997) 3644–3654.
[104] I. Rambaldi, E.N. Kovacs, M.S. Featherstone, A proline-rich transcriptional activation domain in murine HOXD-4 (HOX-4.2), Nucleic Acids Res 22 (3) (1994) 376–382.
[105] M.A. Vigano, et al., Definition of the transcriptional activation domains of three human HOX proteins depends on the DNA-binding context, Mol. Cell Biol. 18 (11) (1998) 6201–6212.
[106] R. Bisaillon, et al., C-terminal domain of MEIS1 converts PKNOX1 (PREP1) into a HOXA9-collaborating oncoprotein, Blood 118 (17) (2011) 4682–4689.
[107] M. Phuycharoen, et al., Uncovering tissue-specific binding features from differential deep learning, Nucleic Acids Res 48 (5) (2020), e27.
[108] E. Longobardi, et al., Prep1 (pKnox1)-deficiency leads to spontaneous tumor development in mice and accelerates EmuMyc lymphomagenesis: a tumor suppressor role for Prep1, Mol. Oncol. 4 (2) (2010) 126–134.
[109] F. Blasi, et al., A tale of TALE, PREP1, PBX1, and MEIS1: Interconnections and competition in cancer, Bioessays 39 (2017) 5.
[110] L. Dardaei, E. Longobardi, F. Blasi, Prep1 and Meis1 competition for Pbx1 binding regulates protein stability and tumorigenesis, Proc. Natl. Acad. Sci. USA 111 (10) (2014) E896–E905.
[111] L. Dardaei, et al., The deficiency of tumor suppressor prep1 accelerates the onset of meis1- hoxa9 leukemogenesis, PLoS One 9 (5) (2014), e96711.
[112] C.G. Sagerstrom, et al., Isolation and characterization of posteriorly restricted genes in the zebrafish gastrula, Dev. Dyn. 220 (4) (2001) 402–408.
[113] T. Zerucha, V.E. Prince, Cloning and developmental expression of a zebrafish meis2 homeobox gene, Mech. Dev. 102 (1–2) (2001) 247–250.
[114] A. Salzberg, et al., A Meis family protein caudalizes neural cell fates in Xenopus, Mech. Dev. 80 (1) (1999) 3–13.
[115] R. Rezsohazy, et al., Cellular and molecular insights into Hox protein action, Development 142 (7) (2015) 1212–1227.
[116] F. Ladam, C.G. Sagerstrom, Hox regulation of transcription: more complex(es), Dev. Dyn. 243 (1) (2014) 4–15.
[117] A. Balsalobre, J. Drouin, Pioneer factors as master regulators of the epigenome and cell fate, Nat. Rev. Mol. Cell Biol. 23 (7) (2022) 449–464.
[118] M. Iwafuchi-Doi, K.S. Zaret, Cell fate control by pioneer transcription factors, Development 143 (11) (2016) 1833–1837.
[119] M. Losa, et al., A tissue-specific, Gata6-driven transcriptional program instructs remodeling of the mature arterial tree, Elife (2017) **6**.
[120] J. Cao, et al., The single-cell transcriptional landscape of mammalian organogenesis, Nature 566 (7745) (2019) 496–502.

Seminars in Cell and Developmental Biology 152-153 (2024) 24–34

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Hox genes in spiders: Their significance for development and evolution

Natascha Turetzek [a,1], Matthias Pechmann [b,2], Ralf Janssen [c,3], Nikola-Michael Prpic [d,*,4]

[a] Ludwig-Maximilians-University Munich, Biocenter, Faculty of Biology, Evolutionary Ecology, Großhaderner Straße 2, 82152 Planegg-Martinsried, Germany
[b] University of Cologne, Institute for Zoology, Biocenter, Zuelpicher Str. 47b, 50674 Cologne, Germany
[c] Uppsala University, Department of Earth Sciences, Palaeobiology, Villavägen 16, 75236 Uppsala, Sweden
[d] Justus-Liebig-Universität Gießen, Institut für Allgemeine und Spezielle Zoologie, AG Zoologie mit dem Schwerpunkt Molekulare Entwicklungsbiologie, Heinrich-Buff-Ring 38, 35392 Gießen, Germany

ARTICLE INFO

Keywords:
Araneae
Hox genes
Evolution of development
Arthopod bodyplan
Arthropoda

ABSTRACT

Hox genes are known for their role in the specification of typical body plan features in animals. Evolutionary changes in Hox gene function are believed to be involved in the emergence of the diverse body plans we observe in animals today. Spiders share many body plan features with other arthropods, but also have numerous unique traits of their own. Studies of spider Hox genes have already provided insights into evolutionarily conserved and derived features of the spider body plan and their genetic basis. However, many aspects of Hox gene biology have been insufficiently studied in spiders so far. In this review, we highlight previous comparative studies of Hox genes in spiders and their significance for our understanding of the evolution of the spider body plan. We also identify aspects of Hox gene biology that need to be studied in greater detail. Many spider Hox genes have not been investigated beyond their mRNA expression patterns, and the role of Hox genes with apparently plesiomorphic or dual functions, like *ftz* and *Hox3* is still unclear. Spiders have a duplicated Hox gene cluster, but possible sub- or neofunctionalisation of duplicates have not yet been studied systematically. Future research should therefore focus on these issues, in addition to the role of *Polycomb* and *trithorax*-mediated regulation, the identification of regulatory regions, cofactors or spider-specific target genes, and the significance of non-coding RNAs transcribed from within the Hox cluster and even from the antisense strand of particular Hox genes.

1. Introduction

Hox genes have initially been discovered for the ability of Hox gene mutations to dramatically change the insect bodyplan. Famous examples include flies with four wings instead of only two [1,2] or with legs instead of antennae on the head [3]. It soon became clear that Hox genes are not restricted to insects, and that these genes are important targets for evolutionary processes to produce novel animal bodyplans. No wonder that they became prime candidates for comparative studies in evolutionary developmental biology ("evo-devo"). One major question was: how do the Hox genes of the various animal groups differ in expression and function to produce all the different bodyplans that we can observe in nature?

Spiders (Araneae) are a species-rich arthropod group which derived their body plan from ancestral chelicerates living in the Cambrian seas approximately 500 million years ago [4]. The spider bodyplan deviates from the body plan of insects in many ways (Fig. 1). The body of both groups is subdivided into body segments, but the organization of these segments into larger functional units (tagmata) is quite different. Insects have a tripartite body, comprising head, thorax and abdomen, whereas spiders only comprise two tagmata: an anterior prosoma and a posterior opisthosoma. Also, the number and types of appendages are diversified: spiders do not have antennae or stout mouthparts like mandibles, but instead bear chelicerae ("fangs") and pedipalps, as well as four walking leg pairs instead of three leg pairs in insects. The hind body (abdomen and opisthosoma, respectively) is leg-less in both groups, but spiders

* Corresponding author.
E-mail addresses: zhang@bio.lmu.de (N. Turetzek), pechmanm@uni-koeln.de (M. Pechmann), ralf.janssen@geo.uu.se (R. Janssen), nprpi@zo.jlug.de (N.-M. Prpic).
[1] 0000-0002-5681-2177
[2] 0000-0002-0043-906X
[3] 0000-0002-4026-4129
[4] 0000-0002-0027-2175

https://doi.org/10.1016/j.semcdb.2022.11.014
Received 15 August 2022; Received in revised form 13 November 2022; Accepted 30 November 2022
Available online 13 December 2022

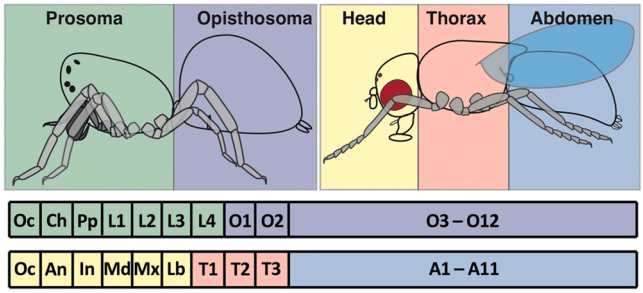


Fig. 1. Comparative overview of spider and insect body plans and segment composition. The spider body (top left) comprises only two parts (tagmata), the anterior prosoma (green) and the posterior prosoma (violet). The insect body (exemplified by a fly; top right) comprises three tagmata, head (yellow), thorax (red) and abdomen (blue). All tagmata in spiders and insects are themselves subdivided in body segments (lower part of figure). The spider prosoma comprises an ocular region (Oc) and the cheliceral (Ch), pedipalpal (Pp), and four consecutive leg-bearing segments (L1 - L4). The spider opisthosoma comprises 12 segments, however, these are fused and reduced in the adult. The insect head comprises an ocular region (Oc) and the antennal (An), intercalary (In), mandibular (Md), maxillary (Mx) and labial (Lb) segments, the insect thorax comprises three leg-bearing segments (T1 - T3), the insect abdomen comprises 11 segments. The anterior segments are generally regarded as orthologous (i.e. to trace from a common ancestor and align in a 1:1 fashion in both groups), but the orthology of the posterior segments is unclear (segment borders therefore omitted).

have several unique appendage types on the opisthosoma, for example respiratory appendages and spinnerets to produce spider silk. Are these differences between spider and insect bodyplans related to changes in their Hox genes?

Hox genes encode homeobox transcription factors, that contain a 60 amino acid DNA binding motif [5], the homeodomain. As indicated by the homeotic transformation mutant phenotypes and their expression along the anterior posterior body axis, Hox genes function as control genes for segment identity and thus morphological diversification of the animal body [6,7]. Most animals have several Hox genes that have been duplicated from one ancestral Hox gene by a "copy and paste" mechanisms, termed tandem duplication [8]. For example, the vinegar fly *Drosophila melanogaster* has eight canonical Hox genes which are crucial for providing positional identities to the distinct segments [reviewed in [9]]: *labial (lab)*, *proboscipedia (pb)*, *Deformed (Dfd)*, *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal A* (*abdA*) and *Abdominal B* (*AbdB*). In addition, *Drosophila melanogaster* has four additional genes derived from Hox genes, *fushi tarazu* (*ftz*), *zerknüllt* (*zen*), *zerknüllt 2* (*zen2*), and *bicoid* (*bcd*), that have lost their plesiomorphic Hox-like function and have been re-purposed for other functions during embryogenesis (overview in [9]). Spiders have orthologs of all of these Hox genes and we review here the current knowledge about their roles in the development of the spider embryo and our present understanding of their relevance for the evolution of the spider bodyplan and unique spider morphologies.

2. Tandem duplications and clustering

Clustering of tandem duplicates is a common property of Hox genes in many animal genomes. Although there are examples of dispersed Hox genes [e.g. 10–13], Hox gene clustering is believed to be the ancestral state in the metazoans (reviewed in [14,15]). Early studies of Hox genes in spiders have discovered Hox genes from all Hox classes [16,17], indicating that tandem duplicates are present in spider genomes as well, but it was long unclear if these duplicates also resided in a cluster of tightly linked genes. With the rising number of sequenced spider genomes, it is now clear that Hox gene clustering is also present in spiders [18–21].

It also became evident that spiders have not one Hox cluster, but two clusters (denoted as cluster A and cluster B) that likely originated in a whole genome duplication event (see chapter below) [18–21]. In the spider model species *Parasteatoda tepidariorum* both clusters are organised, i.e. all Hox genes are transcribed in the same direction, and both clusters contain additional interspersed genes. Altogether 98 predicted protein coding genes and 4 microRNA coding genes have been annotated interspersed in the two Hox clusters of *P. tepidariorum* [18]. The two clusters in this spider are complete, with the exception of the Hox gene *ftz*, which is present in one cluster only [18]. In another spider species, however, the wasp spider *Argiope bruennichi*, one cluster is complete and well organized, while the other cluster presents a reversal of *Ubx* and *abdA*, a potentially fused *ftz-Antp* sequence, and an overlapping *Dfd-Hox3-pb* sequence [20]. Similarly, in the related spider species *Trichonephila antipodiana*, one cluster is complete and well organized, while the other apparently lacks *Hox3*, *ftz*, *Ubx* and *abdA* [19]. Although genomic data from the classic spider model *Cupiennius salei* are not available, combined analyses show that at least one set of Hox genes appears to be complete, while the other may lack some genes [16,22–25].

Additional gene losses affecting at least one of the clusters have occurred in specific spider lineages. One copy of the *Hox3* gene is lost in all studied members of the so-called RTA clade ("higher spiders"), and most Araneae lack one copy of *ftz* (summarized in [21]). The genome of the social velvet spider *Stegodyphus mimosarum* lacks a second *Scr* gene, and the genome of the tarantula *Acanthoscurria geniculata* [26] apparently lacks a second copy of *pb*, *Scr*, *ftz*, and *AbdB*, but it is unclear whether this is an artefact from incomplete sequence assembly.

3. Spatial and temporal collinearity

The functional significance of Hox gene clustering might be linked with the phenomena of temporal and spatial collinearity. Collinearity means that certain properties of gene regulation in a gene cluster correlate with the order of genes in the cluster (reviewed in [27,28]). Temporal collinearity is the phenomenon that the timing of expression of a gene reflects its position in the cluster, i.e. anterior genes are activated earlier than posterior genes. Spatial collinearity is the phenomenon, that the location of gene expression along the anterior-posterior body axis mirrors its location in the cluster, i.e. anterior genes are expressed in the anterior portion of the body, whereas posterior genes are expressed towards the posterior end. For most spider species studied so far only a subset of the Hox genes of both clusters and/or only a limited set of stages has been studied, making the assessment of collinearity difficult (e.g. [16,17,22–24]). Only in *P. tepidariorum* the temporal expression profile has been established for all Hox genes in both clusters and across all embryonic stages [18]. Based on these gene expression data obtained by whole-mount in situ hybridization, there is only limited temporal collinearity in cluster A (Fig. 2C) and no unequivocal temporal collinearity in cluster B (Fig. 2D). Interestingly, however, if both clusters are considered together (i.e. only scoring the first expressed Hox paralog in any cluster), temporal collinearity is more evident and only *pb*, *Hox3* and *AbdB* do not obey temporal collinearity (Fig. 2B). Apparently, temporal collinearity in *P. tepidariorum* is not so much a phenomenon on the level of individual Hox gene clusters, but rather a phenomenon on the level of the Hox genes as a whole. Note, however, that the expression onset data obtained by transcriptome sequencing (RNAseq) somewhat deviates from the data obtained by whole-mount in situ hybridisation: the RNAseq expression data shown by Leite et al. [29] indicate large numbers of transcripts for *lab-B* and *Dfd-A* from stage 1 onwards, whereas whole-mount in situ hybridisation does not detect these transcripts in these early stages; the reason for this discrepancy is currently not clear.

Spatial collinearity, on the other hand, is very pronounced in *P. tepidariorum* and in all other spider species studied. This leads to segmentally staggered expression profiles of the Hox genes along the

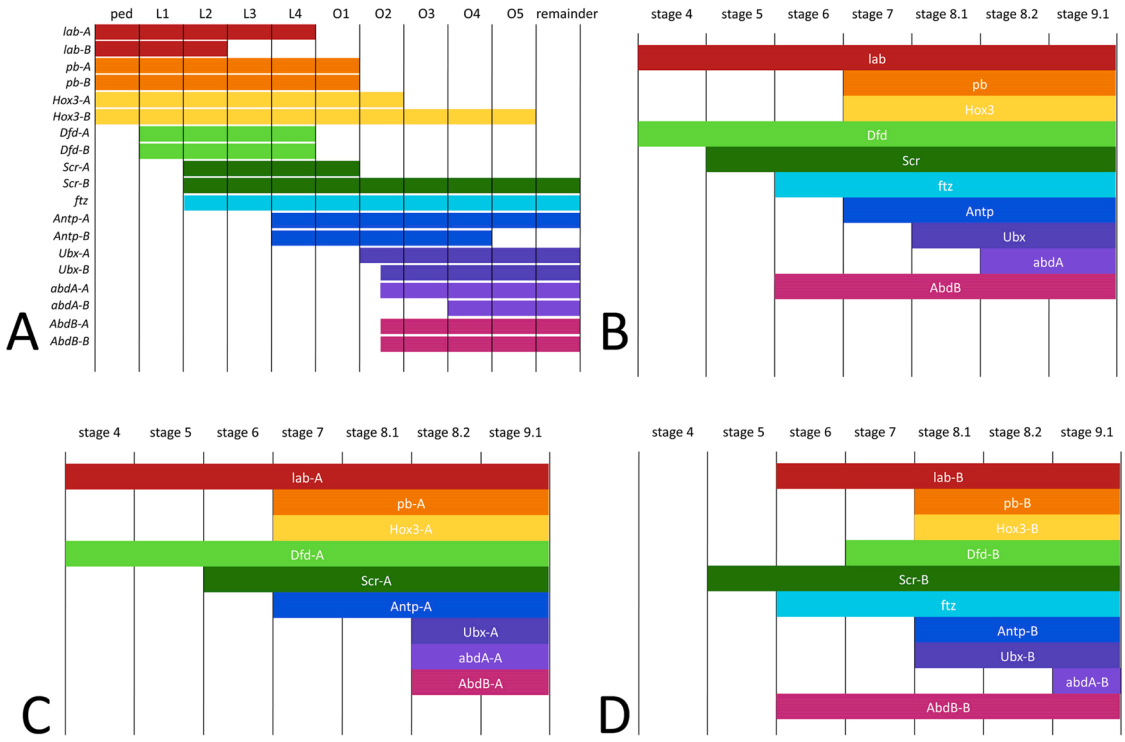


Fig. 2. Collinearity of Hox gene expression in spiders, based on whole-mount in situ hybridisation gene expression data in *P. tepidariorum*. (A) Spatial collinearity. Simplified scheme combining all expression loci, including low or late expression. (B-D) Temporal collinearity, stages are indicated at the top. Temporal collinearity is evident if each Hox gene is represented by its first-expressed ohnolog (B), but also if only Hox genes of cluster A are considered (C), but is less evident if only genes of cluster B are considered (D). Abbreviations: ped, pedipalpal segment; L1-L4, walking leg segments 1–4; O1-O5, opisthosomal segments 1–5.

spider anterior-posterior body axis [16,18,23] (Fig. 2A). Already the first studies of Hox gene expression in spiders and other chelicerate model systems, were able to identify strong spatial collinearity, although operating with an incomplete set of Hox genes and a few developmental stages only. These initial results have helped to answer the longstanding question of the homology of individual anterior body segments between spiders and other arthropods [16,30]. At the time, the prevailing hypothesis proposed that in spiders (and chelicerates in general) one head segment, the tritocerebral segment, was missing. Hox gene expression domains in chelicerates and insects only align, however, if one assumes that no segment is missing in chelicerates. Apparently, spatial collinearity and the exact position of the anterior borders of many Hox gene expression domains are strongly conserved in all arthropods (e.g. [31–36]; earlier work reviewed in [9]). This conclusion is today supported by expression data from all major arthropod groups and the idea of body segment homologisation using the staggered expression of the Hox genes has even been extended to other segmented animals like onychophorans (velvet worms) [37–39], tardigrades (water bears) [40] and annelids (segmented worms) (e.g. [41–44]).

These initial studies on Hox gene expression have also suggested that the domains in spiders obey the border between prosoma and opisthosoma, and a possible role of the Hox genes in establishing this fundamental border of the spider bodyplan has been assumed [16,17,23,45]. Later studies, however, that have incorporated all Hox genes and all developmental stages, have revealed that especially the posterior expression borders of most Hox genes (and also the anterior border of some Hox genes) are more dynamic than previously thought [23]. Apart from the posterior expression borders of *lab* and *Dfd*, no other Hox gene expression domain obeys the prosomal-opisthosomal boundary, if the entire temporal expression profile across all stages is taken into consideration [18]. Thus, a general role of the Hox genes in specifying the prosomal-opisthosomal border becomes less likely; however, some Hox genes like *lab* and *Dfd* may still play a role in this process, and other Hox genes like *pb*, *Hox3* and *Antp* initially obey the border and cross it only later in development, when the dynamic expression pattern expands [18] and their function might change.

4. Cluster duplication and sub-/neofunctionalisation

As already mentioned above, the Hox cluster of spiders has been duplicated. Therefore, spiders have two separate Hox clusters (denoted as cluster A and cluster B) that usually contain the full complement of 10 Hox genes, with a few exceptions already mentioned above: the loss of one copy of *ftz* appears to be a general feature of spiders, and the loss of one copy of *Hox3* might characterize the so-called RTA clade ("higher spiders") (summarized in [21]). However, these observations are still based on a few species only, and there is still insufficient data to draw any conclusions from these apparent gene losses, e.g. in terms of gene function or taxonomic importance. The significance of the gene losses is therefore currently unclear. What is very evident, however, is that in all cases the two genes of a pair of Hox duplicates (so-called ohnologs, coined after the geneticist and evolutionary biologist Susumu Ohno (1928–2000)) differ significantly in their regulation. The two ohnologs may differ in the timepoint of their activation, the general expanse of their expression domains along the body, their level of expression, their temporal dynamics after activation, and their tissue specificity [16–18, 23]. This strongly suggests that the two copies underwent evolutionary changes in their role during development, i.e. they underwent subfunctionalisation (the two copies divide the function of the original gene among themselves) or neofunctionalisation (one copy retains the original function, the other copy obtains a new function). Unfortunately, this aspect of Hox gene biology has not yet been studied in any detail. Only few functional studies have involved the two duplicates of a given spider Hox gene, but in these studies the function of at least one of the duplicates was inconclusive. For example, the study of *Antp* function using RNA interference (RNAi) produced visible phenotypes after *Antp-A*

RNAi, but not after *Antp-B* RNAi [46]. This, of course, raises the question of whether *Antp-B* has any function at all. It is conceivable that some of the regulatory differences observed between the Hox duplicates in spiders do not indicate subfunctionalisation or neofunctionalisation, but may be the product of a narrow specialisation, functional marginalisation or pseudogenisation of one gene copy.

5. Regulation of Hox gene transcription in spiders

An important aspect of Hox gene function is the phenomenon of posterior prevalence/phenotypic suppression, which is the hierarchical dominance of posterior Hox genes over the function (but not necessarily over the expression) of anterior Hox genes (e.g. [47–50]). In simple terms, the result of posterior prevalence is that in areas of co-expression of several Hox genes, the most posterior one of these will virtually "override" the function of most/all of the other Hox genes. One consequence of posterior prevalence are the staggered expression patterns of Hox genes along the anterior-posterior body axis, to create areas of the body where one or a few Hox genes can be expressed alone to "escape" from functional repression by the more posterior Hox genes. The achievement of posterior prevalence and the staggered expression patterns is complex and involves regulation of the Hox genes via diverse factors, including microRNAs from within the Hox cluster itself (e.g. [51–54]), and mutual transcriptional repression between the Hox genes (e.g. [55–58]). Although these processes have not been studied in spiders yet, there is convincing evidence for posterior prevalence in spiders as well. The staggered anterior expression borders in spiders, that already have been discussed above and that are very similar to the expression borders in other arthropod groups (e.g. [18,23]), strongly suggest that they are brought about by posterior prevalence in spiders, too. In addition, a few spider Hox genes have been functionally tested using RNAi and show a behaviour consistent with posterior prevalence. The loss of *Antp-A* expression in *P. tepidariorum* leads to the posterior expansion of *Dfd-A* and *Scr-B* expression into the first opisthosomal segment (O1), and *Dfd-A* expression even expands into the O2 segment if *Ubx-A* expression is impaired as well [46]. These results suggest that *Antp-A* normally represses *Dfd-A* and *Scr-B* transcription in the O1 segment, and *Antp-A* and *Ubx-A* together repress *Dfd-A* transcription in the O2 segment (Fig. 3). Thus, this is evidence for cross-regulatory interactions among these Hox genes in spiders. The loss of *Antp-A* expression also leads to the de-suppression of leg formation in this segment, but fails to de-suppress leg development in any of the other opisthosomal segments, unless *Ubx-A* expression is lost as well [46]. This suggests that *Antp-A* may have a general role of leg suppression in opisthosomal segments, but this role is "masked" by the co-expression of the more posterior Hox genes, that override this function by posterior prevalence and reveal the role only if they are impaired simultaneously with *Antp-A*. A functional test of this hypothesis, however, has been hampered so far by the fact that simultaneous RNAi with more than two Hox genes remains technically challenging.

In another RNAi experiment, the loss of *Dfd-A* expression in *P. tepidariorum* led to the upregulation of *lab-A* and *Hox3-B* transcription in the first walking leg segment (L1, where they are normally transcribed only weakly) [59] (Fig. 3). This suggests that *Dfd-A* normally represses the expression of *lab-A* and *Hox3-B*, but when it fails, then *lab-A* and *Hox3-B* can expand towards posterior until they are stopped again by posterior prevalence, presumably by *Scr* and/or *ftz* (although this has not been tested) (Fig. 3). Interestingly, the loss of *Dfd-A* only allows *lab-A* but not *lab-B* expression (Fig. 3 A and D) to expand into L1 [59]. This suggests the possibility that posterior prevalence is a cluster-specific phenomenon, i.e. posterior Hox genes can only dominate anterior Hox genes of the same cluster. As mentioned above, each cluster in *P. tepidariorum* contains a number of microRNAs, and microRNAs have been implicated in conveying (at least partially) posterior prevalence in other arthropods. Although speculative at the moment, this could implicate microRNAs and other non-coding RNAs residing in the spider Hox clusters in regulating posterior prevalence specifically within the cluster of their own origin.

Apart from this evidence for transcriptional regulation of spider Hox genes by other Hox genes, there is not much data about other factors regulating Hox gene expression in spiders. In the fly *D. melanogaster*, at least two further major inputs into the regulation of Hox transcription are present: first, transcriptional regulation by members of the segmentation gene cascade that act earlier in development (e.g. [60–69]), and second, an intricate epigenetic cell-memory system involving proteins of the Polycomb (PcG) and trithorax (trxG) groups and a number of non-coding RNAs (reviewed in [70]).

Although homologs of many members of the segmentation gene cascade have been studied in spiders (e.g. [25,45,71–75]) only for *hunchback* (*hb*) has the influence on the Hox genes been studied [76]. This work showed that the loss of *hb* expression in *P. tepidariorum* leads to a strong gap gene-like phenotype, but has no specific influence on the transcription of the Hox genes [76]. In vertebrates, the segmentation of the body does not involve a segmentation gene cascade as in *D. melanogaster*. Instead dynamic networks combining Wnt signalling, Notch signalling, and *caudal* (*cad/Cdx*) expression that operate in the posterior end of the embryo. The activation of posterior Hox genes in vertebrates is at least partially dependent on these dynamic networks (e.g. [77–81]). Body segmentation and posterior development in spiders is similar to the vertebrate situation in many aspects, and thus, it is conceivable that also the regulatory function of *cad/Cdx* and Wnt signalling on the posterior Hox genes may be conserved in spiders [73,82,83]. Indeed, the loss of posterior Wnt signalling leads to an upregulation of the expression of *Hox3-B*, and a significant downregulation of *Antp-A* and *Ubx-B* expression [83].

Although nuclear dynamics and chromatin remodelling via PcG and trxG members is known to have a major impact on the expression of Hox genes, this aspect of the biology of Hox genes has not yet been studied in spiders. However, we have detected two RNAs that derive from within the transcription unit of the *Ubx-A* and *Ubx-B* gene (Fig. 4), respectively, that are coming from the antisense strand of the respective transcription unit, and we consider them therefore as noncoding RNAs. Noncoding RNAs have been implicated in Polycomb-mediated epigenetic regulation of Hox genes [84] and in *D. melanogaster* there are two noncoding antisense RNAs transcribed from within the *Ubx* transcription unit [85,86], although their function is not clear yet. Also in myriapods, such antisense transcripts of *Ubx* have been detected [31,87]. The fact that apparently noncoding antisense transcripts of *Ubx* occur in distantly related arthropod groups, and these transcripts show very restricted and dynamically regulated expression during embryogenesis, suggests that these transcripts might be important for embryonic development and are therefore evolutionarily conserved.

6. Translation, co-factors and targets of Hox proteins

After transcription, Hox gene derived RNA may undergo further processing and translational control. Differences in 3 UTR modifications may lead to differential regulation of mRNAs via microRNAs, and differences in promoter usage may lead to differences in the efficiency of translation of mRNAs into Hox proteins (reviewed in [88]). However, in spiders no Hox promoters have yet been identified and microRNAs have been studied only in terms of the total repertoire and gene duplication [89], but not with respect to their role in regulating Hox gene function. The fact that of only 148 identified microRNAs in the entire genome of the spider *P. tepidariorum*, four are located within the two Hox gene clusters, might point to a special relationship between Hox gene regulation and the regulation of these microRNAs [18,89]. In addition, all four spider microRNAs in the clusters are located at the same positions as their *D. melanogaster* counterparts: *mir-993* is located between the *Hox3* homolog and *Dfd*, *mir-10* is located between *Dfd* and *Scr*, and the *mir-iab-4/8* homologs are located between *abdA* and *AbdB* [18,90]. This strong evolutionary conservation of the relative position of the Hox

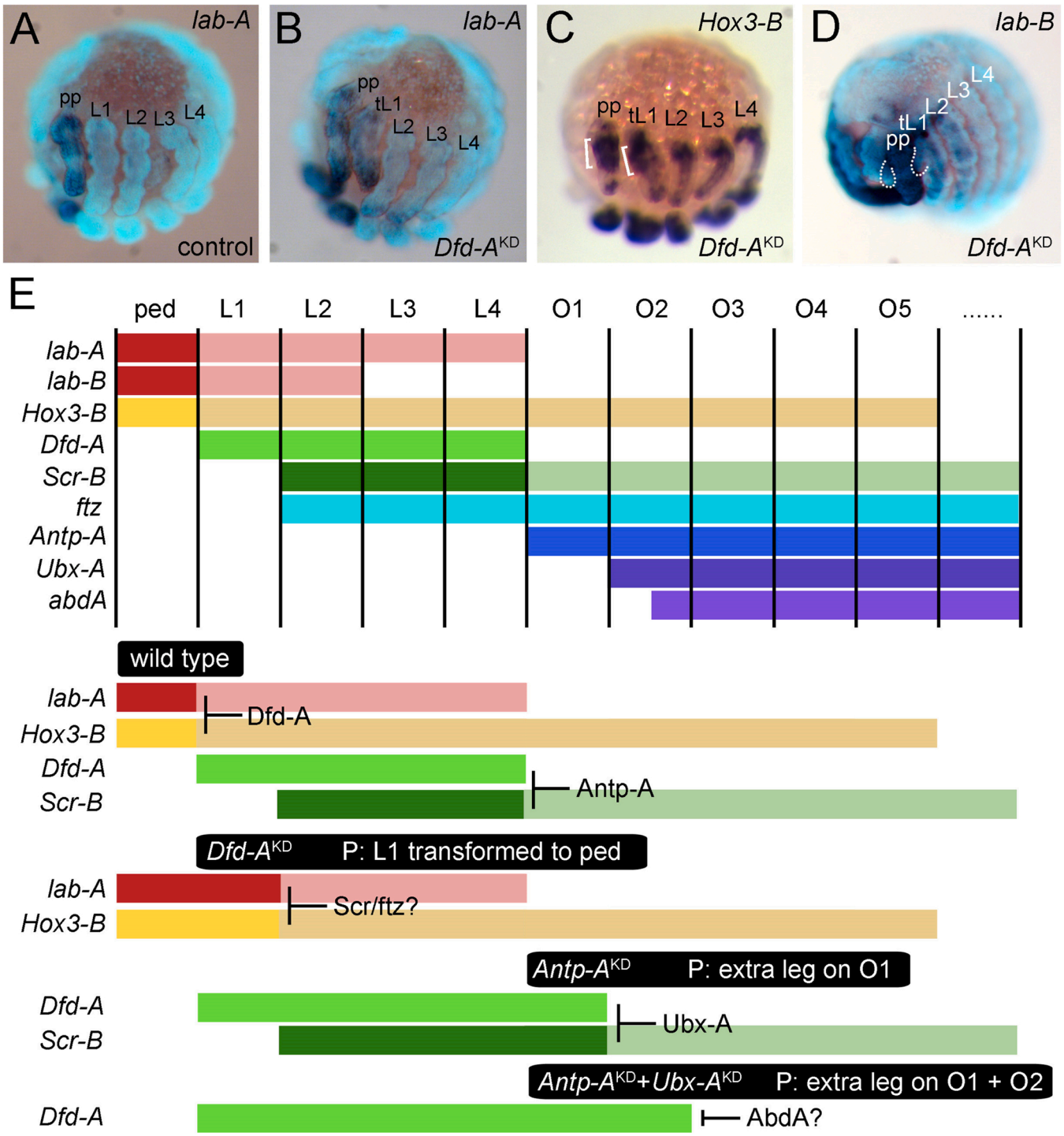


Fig. 3. Cross-regulatory interactions between Hox genes in spiders. (A-D) Expression of *lab-A* (A, B), *Hox3-B* (C) and *lab-B* (D) in control (A) and *Dfd-A* knockdown (B-C) embryos of *P. tepidariorum*. A-C stage 10, D stage 13 (staging after [120]). Embryos in A, B, D are counterstained with a fluorescent nuclear dye. In wild type embryos *lab-A* is strongly expressed in the pedipalps (A). *Dfd-A* knockdown leads to the transformation of the first walking leg towards pedipalpal identity (also see [59]) and RNAi embryos show strong *lab-A* (B) and *Hox3-B* (C, brackets), but not *lab-B* (D) expression in the transformed L1 segment. Gnathendites on the pedipalp and the transformed L1 appendage are marked by the dotted lines in (D). RNAi experiments and embryonic in situ hybridisation were performed as described in Pechmann et al. [59]. (E) Simplified schematic of Hox gene expression domains and interactions wild type, *Dfd-A*, *Antp-A* and *Antp-A/Ubx-A* knockdown embryos of *P. tepidariorum*. Please note: only relevant Hox genes are included in the schematics, and expression domain of *abdA* is the combined domain for both ohnologs. T-bars indicate potential repressive interactions between posteriorly and more anteriorly expressed Hox genes. Light colours indicate later/weaker expression in these segments. Abbreviations: pp/ped: pedipalp, L1-L4: walking legs 1–4, tL1: transformed L1 (towards pedipalp identity), O1-O5: opisthosomal segments 1–5, P: phenotype, KD: knockdown via parental RNA interference.

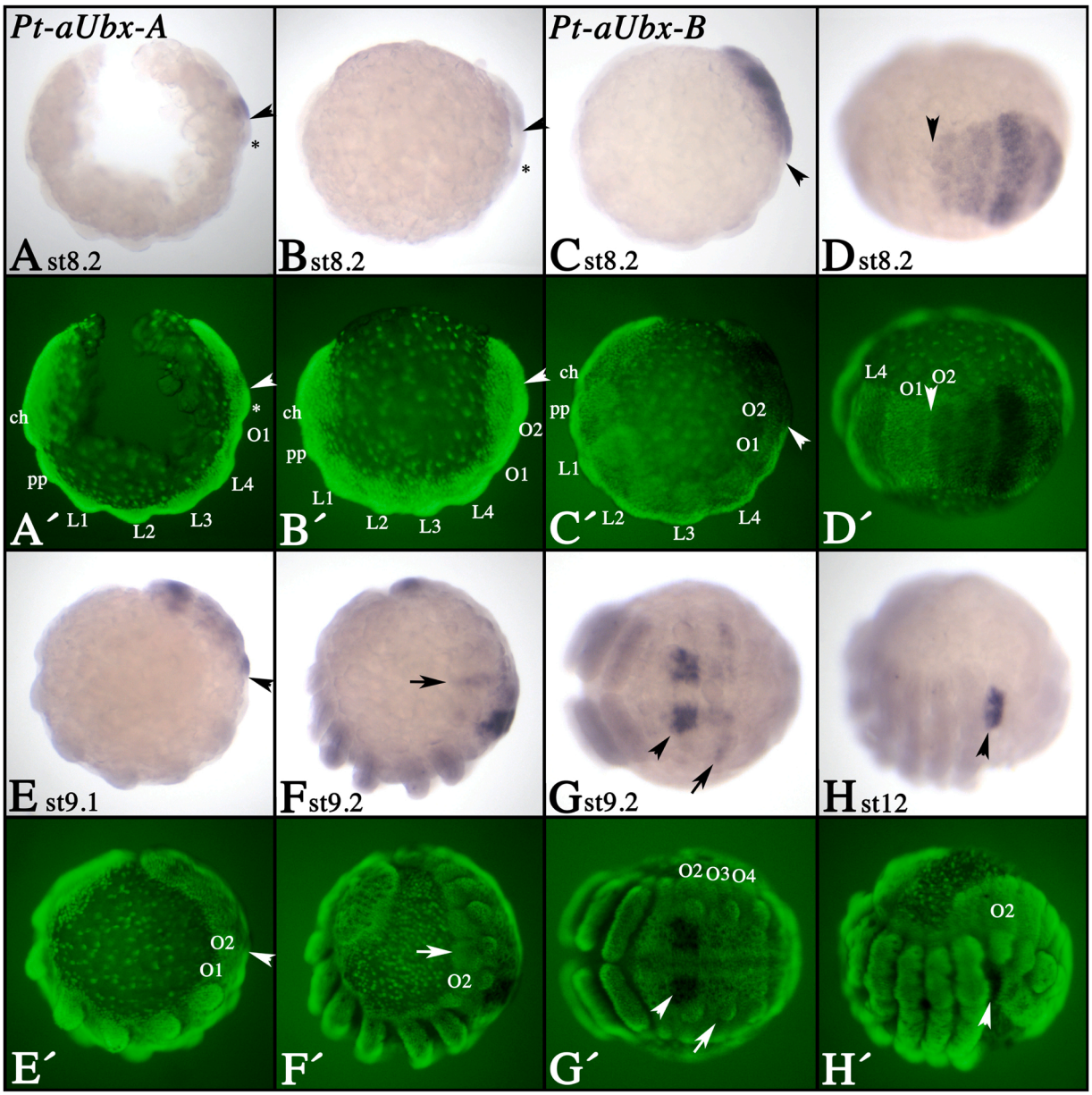


Fig. 4. Antisense RNA expression from the *Ubx* transcriptional units in *Parasteatoda tepidariorum*. Antisense RNA expression from the *Ubx-A* transcription unit (*Pt-aUbx-A*) is shown in panels A, B, E, F. Antisense RNA expression from *Ubx-B* transcription unit (*Pt-aUbx-B*) is shown in panels C, D, G, H. Embryonic in situ hybridisation were performed as described in Pechmann et al. [59]. The panels A´to H´show the same embryo as in the main panels A to H, but with a nuclear counterstain (SYBR-Green), to better reveal morphological details. Arrowheads point to anterior expression borders in all panels, except for panels A and B where a separate and weak anterior expression domain exists (indicated by asterisks (*)). Stages according to Mittmann and Wolff [120] are indicated bottom left. Arrows point to additional expression domains in the presumptive lateral and dorsal ectoderm. Abbreviations: ch, cheliceral segment; pp, pedipalpal segment; L1-L4, walking leg segments 1–4; O1-O4, opisthosomal segments 1–4.

genes and their interspersed microRNAs suggests an essential relevance of this arrangement that cannot easily be altered during evolution.

The Hox proteins are transcription factors, i.e. they bind to regulatory elements of their target genes. However, the binding specificity of the individual Hox proteins alone is very low (e.g. [91]), they require additional factors to increase the specificity of their action upon the regulatory sequences of their targets (e.g. [88,92–95]). In *D. melanogaster* and in other animal model species, a number of co-factors and targets for the different Hox proteins have been described (e.g. [52,96–99]). In spiders, however, information about possible co-factors and target genes is very limited. As already discussed above, some anterior spider Hox genes apparently are target genes of more posterior Hox genes, consistent with cross-regulatory interactions in posterior prevalence. Apart from these cross-regulatory interactions, only few studies have focused on possible target genes of Hox gene expression in spiders. The down-regulation of *Antp-A* via RNAi and subsequent posterior expansion of *Dfd-A* and *Scr-B* in *P. tepidariorum* leads to the formation of walking leg structures on the O1 segment, and this includes the ectopic activation of the entire gene network responsible for walking leg development, including *Distal-less*, *dachshund-1*, *homothorax-1*, *extradenticle-1* and *Sp6–9* [46,100], but it is unclear if these genes are direct regulatory targets of the ectopic expression of *Dfd-A* and *Scr-B*.

The Hox gene *lab-A* is strongly expressed in the pedipalpal segment of spiders and has been shown to be required for normal development of this segment [59]. A study in *P. tepidariorum* has therefore tested genes

known from other species to be either co-factors or targets of Hox genes orthologous to *labial*. Of 75 genes that have been studied, only three showed an expression pattern compatible with a possible connection to *lab-A* function [101]. This study concluded that the search for Hox gene co-factors and regulatory targets in spiders cannot be based solely on comparisons with previously known factors from other species, but also needs to apply *de novo* gene identification methods, e.g. comparative genome wide assays for transposase-accessible chromatin (ATAC-seq), that have successfully been used to identify candidate genes that exhibit locally-restricted expression in early embryonic stages of *P. tepidariorum* [102].

7. Considerations of the individual Hox genes in spiders

7.1. The anterior Hox genes (class-1 and class-2 Hox genes): labial and proboscipedia

All spider *labial* homologs studied so far in different species show the strongest expression in the pedipalpal segment [16,18,59], suggesting a significant role of *labial* in the specification of this segment. Indeed, the loss of *lab-A* in *P. tepidariorum* leads to a reduction of the pedipalpal segment and the loss of the pedipalp appendages [59]. Intriguingly, this reduced morphology of the pedipalpal segment is reminiscent of the reduced morphology of the homologous segment in insects, the intercalary segment. However, in insects the function of *labial* is required to produce this specifically reduced morphology (e.g. [103]), whereas in spiders *lab-A* is required to prevent a reduced morphology; apparently, the developmental significance of *labial* for segment morphology has been completely reversed between the evolutionary lineages of insects and spiders. The loss of *lab-A* in *P. tepidariorum* RNAi embryos does not lead to an anterior expansion of the expression of the more posterior Hox gene *Dfd-A*, nor does it affect the expression of *lab-B* in the (now reduced) pedipalpal segment. This indicates that *lab-A* is not necessary to repress the anterior expansion of *Dfd-A*, probably because *lab-B* is still expressed there and might be able to prevent *Dfd-A* expansion. If *Dfd-A* is impaired via RNAi, then *lab-A* but not *lab-B* expands into the first walking leg segment (Fig. 3), and at the same time the first walking leg pair is fully transformed into a pair of pedipalps [59]. This strongly suggests that the expression of *lab-A* is sufficient to confer pedipalp identity, without the input of *lab-B*.

RNAi experiments with *lab-B* in *P. tepidariorum* did not produce any observable phenotype in the embryos, thus the developmental role of this gene is still unclear [59]. The same is true for the role of the weak expression of both ohnologs, *lab-A* and *lab-B*, in the walking leg segments. The expression in these segments is much weaker and appears later than the expression in the pedipalpal segment, but no apparent phenotype was observed in the walking leg segments after RNAi knockdown of either ohnolog in *P. tepidariorum*.

The function of *pb* has not yet been studied via RNAi, thus any ideas about the role of the two ohnologs is currently based solely on gene expression patterns. In *P. tepidariorum*, *pb-A* is strongly expressed in the pedipalps and the walking legs, and weaker and partially delayed expression is seen in the neuroectoderm of the pedipalpal segment, all four leg-bearing segments and in the first opisthosomal segment [18]. However, earliest and strongest expression is in the pedipalpal segment, thus indicating a possible role of *pb-A* in the development of this segment, similar to (or in concert with) the *lab-A* gene. The *pb-B* gene expression is similar to *pb-A* expression, but appears much later und is very weak; any functional significance of this gene is therefore unclear. The expression of single *pb* genes has also been studied in the spider species *Cupiennius salei* and *Steatoda triangulosa* [17,23], where they are strongly expressed in the pedipalps and legs, suggesting that the studied *pb* gene is the ortholog of the *pb-A* gene in both cases.

7.2. The class-3 Hox genes

The class-3 Hox genes in insects are special, because they have no Hox-like function, i.e. they are not involved in the specification of positional information along the anterior-posterior body axis. Instead, they have been duplicated and the resulting genes (*zerknüllt* (*zen*), *zerknüllt-2* (*zen2*) and *bicoid* (*bcd*)) are important factors for the specification of the anterior germ band and extraembryonic tissues (e.g. [104,105]). It was, therefore, a significant finding when the expression pattern of spider homologs of *Hox3* were shown to have spatially restricted expression domains along the anterior-posterior body [17,22]. This indicated that the spider *Hox3* genes are not involved in anterior germ band development or extraembryonic tissue specification, and rather have retained their original Hox-like function along the anterior-posterior body axis. However, no functional studies have been performed to date to further corroborate a plesiomorphic Hox function for the *Hox3* genes in spiders, although an interesting experiment has already been proposed by Damen and Tautz [22] who write: "It would thus be interesting to test whether the spider *Hox3* gene can replace the function of a mouse *Hox3* gene and whether this capacity has been lost for the *zen* genes [of the insects]". The data shown in Fig. 3 suggest a Hox-like function of *Hox3-B*: it is upregulated in the proximal portion of the transformed L1 leg in *P. tepidariorum* after *Dfd-A* RNAi. This suggests that *Dfd-A* normally represses the expression level of *Hox3-B* in the L1 segment, and that strong *Hox3-B* expression in the proximal portion of the prosomal appendages might be involved in specifying pedipalpal identity over walking leg identity.

7.3. The central Hox genes (class-4 to class-6–9 Hox genes): Deformed, Sex combs reduced, fushi tarazu, Antennapedia, Ultrabithorax, and abdominal A

Both *Dfd* ohnologs are expressed solely in the walking leg-bearing segments L1 to L4 [16,18,23]. This already suggests a role of *Dfd* in specifying walking leg identity. Indeed, loss of *Dfd-A* in *P. tepidariorum* leads to de-repression of *Hox3-B* and *lab-A* in the L1 segment, and effectively transforms the L1 walking leg pair into a complete pair of pedipalps (Fig. 3) [59]. In addition, loss of *Antp-A*, and loss of *Antp-A* together with *Ubx-A*, leads to an expansion of *Dfd-A* expression into the O1 and O1 plus O2 segments, respectively, and initiates leg development in these normally leg-less body segments [46]. These data indicate that expression of *Dfd-A* is both necessary and sufficient to control walking leg formation in a body segment, and makes at least the *Dfd-A* ohnolog a good candidate for a walking-leg identity "master switch".

The expression of the ohnologs of *Scr* has been documented for *C. salei* [23] and *P. tepidariorum* [18]. Both copies are expressed in the walking legs L2, L3 and L4. Expression in L2 is weak, whereas expression in L3 and L4 is stronger and forms unique expression patterns of rings and a distal cap for each walking leg. A similar pattern was observed in *S. triangulosa* using an antibody against SCR protein [17], but in addition a very faint signal was detected in the L1 legs; it is unclear whether this difference is related to species-specific expression, or is caused by the differences in the detection method (mRNA vs. protein). In *P. tepidariorum* additional late expression aspects have been described [18]: in stages during germ band inversion, expression of *Scr-A* extends into the O1 segment and *Scr-B* is expressed ubiquitously but faintly in the entire opisthosoma. No comparable expression has been described for other spider species and the significance of the opisthosomal *Scr* expression in *P. tepidariorum* is not known. No functional data beyond gene expression is available for a *Scr* gene in any spider species.

The *ftz* gene is another Hox gene apart from *Hox3* that has no typical Hox-like role in insects, and instead it is involved in the segmentation gene cascade that subdivides the body into segments (reviewed in [106,107]). In *C. salei* the *ftz* ortholog is initially expressed in a Hox-like domain that starts at the border between the L1 and L2 segment and ends in the L4 segment [25]. Later on, it is expressed in the leg tips of L2,

L3 and L4 and also in a small number of neuroectodermal cells in all opisthosomal segments. This expression pattern suggested a Hox-like role, and no function in segmentation in spiders. However, a more detailed analysis of *ftz* expression across all developmental stages in *P. tepidariorum* showed, that *ftz* is expressed in a dynamic striped pattern at the posterior end of the germband, much like other typical segmentation genes, in addition to the Hox-like expression in the walking leg segments [18]. These results indicate that *ftz* in spiders actually combines both functions, and thus possibly provides an example for a Hox gene in the state of functional diversification, in that it still plays the plesiomorphic role in anterior-posterior patterning, but already has acquired a new role in body segmentation.

The remaining central Hox genes, *Antp*, *Ubx* and *abd-A* are known for their roles in the evolutionary diversification of tagmosis (i.e. the evolution of functional body subdivisions in arthropod bodyplans). Numerous studies have provided intriguing examples of evolutionary changes in the regulation of these three Hox genes and their interactions with cofactors or regulatory targets to result in body plan changes like e.g. the extent of the region bearing feeding appendages in crustaceans [108,109] or the distinction between fore- and hindwings in insects [6,110–112]. Indeed, also in spiders these three central Hox genes appear to be responsible for one main aspect of the spider bodyplan, namely the leg-less opisthosoma. The three Hox genes *Antp*, *Ubx*, and *abd-A* together cover the entire opisthosoma with their combined expression pattern (*Antp* later expands anterior into L4, but the significance of this expression is not known) [16,17]. Of these, *Antp* is the only Hox gene expressed in O1, it is joined by *Ubx* at the border between O1 and O2, and both are joined by *abd-A* in the posterior portion of O2. Work in *P. tepidariorum* has shown that *Antp-A* normally suppresses leg development in O1, and *Antp-A* together with *Ubx-A* suppress leg development in O2 [46]. The role of *abd-A* could not be tested, because triple RNAi proved technically challenging, but it is likely that all three central Hox genes work together to suppress leg development in the entire spider opisthosoma. It is important to note that, although the general role of these three central Hox genes in shaping important aspects of tagmosis appears to be conserved among arthropods, the details of gene regulation and cofactor/target gene usage are likely to be very different. For example, *Antp* in insects promotes leg development (hence the name: misexpression of *Antp* in the antennae transforms them into legs) and is expressed strongly in the leg-bearing tagma (the thorax) [113,114]. In spiders, *Antp* appears to have the opposite role: it suppresses leg development and is expressed strongly in the leg-less tagma of the body (the opisthosoma) [46]. However, the basis for this differential influence of *Antp* on leg formation does not reside in the ANTP proteins themselves, because Khadjeh et al. [46] have shown that spider *Antp* transforms *D. melanogaster* antennae towards legs very much like endogenous *D. melanogaster Antp*. Thus, the differences in the mode of operation between spider and insect *Antp* are to be found in the cellular environment of the ANTP protein, and this explains why it would be so important to identify spider specific cofactors or target genes of *Antp*, in order to understand and reconstruct the functional evolution of Hox genes in arthropods.

7.4. The posterior Hox genes (class-9–13(15) Hox genes): Abdominal B

In insects *Abd-B* is the most posteriorly expressed Hox gene and has a dual role: first it specifies the posterior end of the germ band, second it is critical for the specification and development of the primary genitalia (i.e. genital structures associated with the opening of the reproductive system) [115,116]. This is possible, because insect primary genitalia develop at the posterior end of the abdomen, and thus entirely in the expression domain of insect *Abd-B*. In spiders, however, primary genitalia and the opening of the reproductive system are located in O2 and thus in the anterior portion of the opisthosoma [117]. It has been argued previously [118], that *Abd-B* in spiders also has the dual role in posterior specification and genitalia development. This would explain the expression pattern of spider *Abd-B*, because in spiders this is not restricted to the very posterior end of the opisthosoma, but instead both ohnologs are expressed in the posterior portion of O2 and from there throughout the opisthosoma (including the posterior end). Indeed, one of the ohnologs even has a separate spot of expression in the O2 segment, and this separate expression locus has previously been linked to the developing genital opening [18,118]. However, this expression in the O2 segment is located in the posterior portion of the O2 limb buds, and this portion has been shown to form the primordium of the book lungs (see the striped *engrailed* expression pattern in [119]). Thus, a role of *Abd-B* in the specification of the posterior end of the germ band is likely, but the role of the *Abd-B* expression in O2 might be linked to the development of the book lungs, rather than genitalia formation. Clearly, more work on the function of *Abd-B* in spiders is necessary to establish its true function in the different opisthosomal segments and their organs.

8. Conclusions

Hox genes play an important role in the development of the body plan in animals, and evolutionary changes in their regulation and function contributed to the diversity in animal shape and form observed today. The study of spider Hox genes since the initial publication by Damen et al. [16] has contributed to our understanding how Hox genes are involved in producing the spider body plan: for example, it has been revealed that spiders retain an arthropod typical head segmentation including a deutocerebral segment, that spiders use the posterior Hox genes to repress leg formation on their abdomen (opisthosoma), and that at least the Hox gene *lab* plays the role of a "master switch" for the identity of the pedipalp appendage. However, many aspects of Hox gene biology have been insufficiently studied in spiders so far. Many spider Hox genes have still not been studied beyond their mRNA expression patterns, and especially the role of those Hox genes with apparently plesiomorphic or dual functions, like *ftz* and *Hox3* is still unclear. Spiders have a duplicated Hox gene cluster, but the roles and interrelationships among the duplicates (possible sub- or neofunctionalisations) have not yet been investigated systematically. Other aspects of Hox gene biology have been neglected entirely in spiders, for example the role of *Polycomb* and *trithorax* in spider Hox gene regulation, the identification of regulatory regions, cofactors or even spider-specific target genes, and the significance of non-coding RNAs transcribed from within the Hox cluster and even from the antisense strand of particular Hox genes. Future research should therefore focus on these insufficiently studied aspects of the role of spider Hox genes to provide more detailed insight into bodyplan evolution and plasticity in spiders and other arthropod groups as well.

Funding

This work has been partially funded by the Deutsche Forschungsgemeinschaft within the priority program SPP 2349 (project number 503325252).

Conflict of interests

The authors declare no conflict of interests.

Acknowledgements

We thank Daniel J. Leite (Oxford Brookes University) for help with assigning the Hox ohnologs of *Parasteatoda tepidariorum* to the correct clusters. The authors would also like to thank Richard S. Mann (Columbia University) for the opportunity to contribute to this special issue.

References

[1] E.B. Lewis, Genes and developmental pathways, Am. Zool. 3 (1963) 33–56.
[2] G. Morata, A. Garcia-Bellido, Developmental analysis of some mutants of the bithorax system of Drosophila, Wilhelm. Roux' Arch. 179 (1976) 125–143, https://doi.org/10.1007/BF00848298.
[3] S. Schneuwly, R. Klemenz, W.J. Gehring, Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia, Nature 325 (1987) 816–818, https://doi.org/10.1038/325816a0.
[4] C. Aria, J.-B. Caron, A middle Cambrian arthropod with chelicerae and proto-book gills, Nature 573 (2019) 586–589, https://doi.org/10.1038/s41586-019-1525-4.
[5] W.J. Gehring, Homeotic genes, the homeo box, and the genetic control of development, Cold Spring Harb. Symp. Quant. Biol. 50 (1985) 243–251, https://doi.org/10.1101/SQB.1985.050.01.031.
[6] E.B. Lewis, A gene complex controlling segmentation in Drosophila, Nature 276 (1978) 565–570, https://doi.org/10.1038/276565a0.
[7] R. Loker, J.E. Sanner, R.S. Mann, Cell-type-specific Hox regulatory strategies orchestrate tissue identity, e4, Curr. Biol. 31 (2021) 4246–4255, https://doi.org/10.1016/j.cub.2021.07.030.
[8] J. Zhang, M. Nei, Evolution of antennapedia-class homeobox genes, Genetics 142 (1996) 295–303, https://doi.org/10.1093/genetics/142.1.295.
[9] C.L. Hughes, T.C. Kaufman, Hox genes and the evolution of the arthropod body plan1, Evol. Dev. 4 (2002) 459–499.
[10] B. Negre, J.M. Ranz, F. Casals, M. Cáceres, A. Ruiz, A new split of the hox gene complex in drosophila: relocation and evolution of the gene labial, Mol. Biol. Evol. 20 (2003) 2042–2054, https://doi.org/10.1093/molbev/msg238.
[11] R.M. Pace, M. Grbić, L.M. Nagy, Composition and genomic organization of arthropod Hox clusters, EvoDevo 7 (2016) 11, https://doi.org/10.1186/s13227-016-0048-4.
[12] M.A. Hoy, R.M. Waterhouse, K. Wu, A.S. Estep, P. Ioannidis, W.J. Palmer, A. F. Pomerantz, F.A. Simão, J. Thomas, F.M. Jiggins, T.D. Murphy, E.J. Pritham, H. M. Robertson, E.M. Zdobnov, R.A. Gibbs, S. Richards, Genome sequencing of the phytoseiid predatory mite Metaseiulus occidentalis reveals completely atomised Hox genes and super-dynamic intron evolution, Genome Biol. Evol. (2016), https://doi.org/10.1093/gbe/evw048.
[13] R. Szabó, D.E.K. Ferrier, Two more Posterior Hox genes and Hox cluster dispersal in echinoderms, BMC Evolut. Biol. 18 (2018) 203, https://doi.org/10.1186/s12862-018-1307-x.
[14] N. Hajirnis, R.K. Mishra, Homeotic Genes: Clustering, Modularity, and Diversity, Frontiers in Cell and Developmental Biology. 9 (2021). https://www.frontiersin.org/articles/10.3389/fcell.2021.718308 (accessed August 13, 2022).
[15] D. Duboule, The (unusual) heuristic value of Hox gene clusters; a matter of time, Dev. Biol. 484 (2022) 75–87, https://doi.org/10.1016/j.ydbio.2022.02.007.
[16] W.G. Damen, M. Hausdorf, E.-A. Seyfarth, D. Tautz, A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider, Proc. Natl. Acad. Sci. 95 (1998) 10665–10670.
[17] A. Abzhanov, A. Popadic, T.C. Kaufman, Chelicerate Hox genes and the homology of arthropod segments, Evol. Dev. 1 (1999) 77–89, https://doi.org/10.1046/j.1525-142x.1999.99014.x.
[18] E.E. Schwager, P.P. Sharma, T. Clarke, D.J. Leite, T. Wierschin, M. Pechmann, Y. Akiyama-Oda, L. Esposito, J. Bechsgaard, T. Bilde, A.D. Buffry, H. Chao, H. Dinh, H. Doddapaneni, S. Dugan, C. Eibner, C.G. Extavour, P. Funch, J. Garb, L.B. Gonzalez, V.L. Gonzalez, S. Griffiths-Jones, Y. Han, C. Hayashi, M. Hilbrant, D.S.T. Hughes, R. Janssen, S.L. Lee, I. Maeso, S.C. Murali, D.M. Muzny, R. Nunes da Fonseca, C.L.B. Paese, J. Qu, M. Ronshaugen, C. Schomburg, A. Schönauer, A. Stollewerk, M. Torres-Oliva, N. Turetzek, B. Vanthournout, J.H. Werren, C. Wolff, K.C. Worley, G. Bucher, R.A. Gibbs, J. Coddington, H. Oda, M. Stanke, N.A. Ayoub, N.-M. Prpic, J.-F. Flot, N. Posnien, S. Richards, A.P. McGregor, The house spider genome reveals an ancient whole-genome duplication during arachnid evolution, BMC Biol. 15 (2017), https://doi.org/10.1186/s12915-017-0399-x.
[19] Z. Fan, T. Yuan, P. Liu, L.-Y. Wang, J.-F. Jin, F. Zhang, Z.-S. Zhang, A chromosome-level genome of the spider Trichonephila antipodiana reveals the genetic basis of its polyphagy and evidence of an ancient whole-genome duplication event, Gigascience 10 (2021), https://doi.org/10.1093/gigascience/giab016.
[20] M.M. Sheffer, A. Hoppe, H. Krehenwinkel, G. Uhl, A.W. Kuss, L. Jensen, C. Jensen, R.G. Gillespie, K.J. Hoff, S. Prost, Chromosome-level reference genome of the European wasp spider *Argiope bruennichi*: a resource for studies on range expansion and evolutionary adaptation, giaa148, GigaScience 10 (2021), https://doi.org/10.1093/gigascience/giaa148.
[21] A. Harper, L.B. Gonzalez, A. Schönauer, R. Janssen, M. Seiter, M. Holzem, S. Arif, A.P. McGregor, L. Sumner-Rooney, Widespread retention of ohnologs in key developmental gene families following whole genome duplication in arachnopulmonates, G3 Genes=Genomes|Genet. (2021), https://doi.org/10.1093/g3journal/jkab299.
[22] W.G.M. Damen, D. Tautz, A Hox class 3 orthologue from the spider Cupiennius salei is expressed in a Hox-gene-like fashion, Dev. Gene Evol. 208 (1998) 586–590, https://doi.org/10.1007/s004270050218.
[23] E.E. Schwager, M. Schoppmeier, M. Pechmann, W.G. Damen, Duplicated Hox genes in the spider Cupiennius salei, Front. Zool. 4 (2007) 10, https://doi.org/10.1186/1742-9994-4-10.
[24] W.G.M. DAMEN, D. TAUTZ, Comparative molecular embryology of arthropods: the expression of Hox genes in the spider Cupiennius salei, Invertebr. Reprod. Dev. 36 (1999) 203–209, https://doi.org/10.1080/07924259.1999.9652701.
[25] W.G. Damen, R. Janssen, N.-M. Prpic, Pair rule gene orthologs in spider segmentation, Evol. Dev. 7 (2005) 618–628.
[26] K.W. Sanggaard, J.S. Bechsgaard, X. Fang, J. Duan, T.F. Dyrlund, V. Gupta, X. Jiang, L. Cheng, D. Fan, Y. Feng, L. Han, Z. Huang, Z. Wu, L. Liao, V. Settepani, I.B. Thøgersen, B. Vanthournout, T. Wang, Y. Zhu, P. Funch, J.J. Enghild, L. Schauser, S.U. Andersen, P. Villesen, M.H. Schierup, T. Bilde, J. Wang, Spider genomes provide insight into composition and evolution of venom and silk, Nat. Commun. 5 (2014) 3765, https://doi.org/10.1038/ncomms4765.
[27] R. Krumlauf, Hox genes, clusters and collinearity, Int J. Dev. Biol. 62 (2018) 659–663, https://doi.org/10.1387/ijdb.180330rr.
[28] S.J. Gaunt, Hox cluster genes and collinearities throughout the tree of animal life, Int J. Dev. Biol. 62 (2018) 673–683, https://doi.org/10.1387/ijdb.180162sg.
[29] D.J. Leite, L. Baudouin-Gonzalez, S. Iwasaki-Yokozawa, J. Lozano-Fernandez, N. Turetzek, Y. Akiyama-Oda, N.-M. Prpic, D. Pisani, H. Oda, P.P. Sharma, Homeobox gene duplication and divergence in arachnids, Mol. Biol. Evol. (2018).
[30] M.J. Telford, R.H. Thomas, Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment, Proceedings of the National Academy of Sciences. 95 (1998) 10671–10675.
[31] C. Brena, A.D. Chipman, A. Minelli, M. Akam, Expression of trunk Hox genes in the centipede Strigamia maritima: sense and anti-sense transcripts, Evol. Dev. 8 (2006) 252–265, https://doi.org/10.1111/j.1525-142X.2006.00096.x.
[32] M. Manuel, M. Jager, J. Murienne, C. Clabaut, H.L. Guyader, Hox genes in sea spiders (Pycnogonida) and the homology of arthropod head segments, Dev. Genes Evol. 216 (2006) 481–491, https://doi.org/10.1007/s00427-006-0095-2.
[33] R. Janssen, W.G.M. Damen, The ten Hox genes of the millipede Glomeris marginata, Dev. Genes Evol. 216 (2006) 451–465, https://doi.org/10.1007/s00427-006-0092-5.
[34] P.P. Sharma, E.E. Schwager, C.G. Extavour, G. Giribet, Hox gene expression in the harvestman Phalangium opilio reveals divergent patterning of the chelicerate opisthosoma, Evol. Dev. 14 (2012) 450–463, https://doi.org/10.1111/j.1525-142X.2012.00565.x.
[35] P.P. Sharma, E.E. Schwager, C.G. Extavour, W.C. Wheeler, Hox gene duplications correlate with posterior heteronomy in scorpions, 20140661–20140661, Proc. R. Soc. B: Biol. Sci. 281 (2014), https://doi.org/10.1098/rspb.2014.0661.
[36] J.M. Serano, A. Martin, D.M. Liubicich, E. Jarvis, H.S. Bruce, K. La, W.E. Browne, J. Grimwood, N.H. Patel, Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiensis, Dev. Biol. 409 (2016) 297–309, https://doi.org/10.1016/j.ydbio.2015.10.029.
[37] B.J. Eriksson, N.N. Tait, G.E. Budd, R. Janssen, M. Akam, Head patterning and Hox gene expression in an onychophoran and its implications for the arthropod head problem, Dev. Genes Evol. 220 (2010) 117–122, https://doi.org/10.1007/s00427-010-0329-1.
[38] R. Janssen, B.J. Eriksson, N.N. Tait, G.E. Budd, Onychophoran Hox genes and the evolution of arthropod Hox gene expression, Front Zool. 11 (2014) 22.
[39] G. Mayer, F.A. Franke, S. Treffkorn, V. Gross, I. de Sena Oliveira, Onychophora, in: A. Wanninger (Ed.), Evolutionary Developmental Biology of Invertebrates 3: Ecdysozoa I: Non-Tetraconata, Springer, Vienna, 2015, pp. 53–98, https://doi.org/10.1007/978-3-7091-1865-8_4.
[40] F.W. Smith, T.C. Boothby, I. Giovannini, L. Rebecchi, E.L. Jockusch, B. Goldstein, The compact body plan of tardigrades evolved by the loss of a large body region, Curr. Biol. 26 (2016) 224–229, https://doi.org/10.1016/j.cub.2015.11.059.
[41] M.J. Kourakis, V.A. Master, D.K. Lokhorst, D. Nardelli-Haefliger, C.J. Wedeen, M. Q. Martindale, M. Shankland, Conserved anterior boundaries of Hox gene expression in the central nervous system of the leech Helobdella, Dev. Biol. 190 (1997) 284–300, https://doi.org/10.1006/dbio.1997.8689.
[42] S.Q. Irvine, M.Q. Martindale, Expression patterns of anterior Hox genes in the polychaete Chaetopterus: correlation with morphological boundaries, Dev. Biol. 217 (2000) 333–351, https://doi.org/10.1006/dbio.1999.9541.
[43] P.R.H. Steinmetz, R.P. Kostyuchenko, A. Fischer, D. Arendt, The segmental pattern of otx, gbx, and Hox genes in the annelid Platynereis dumerilii: the segmental pattern of otx, gbx, and Hox genes, Evol. Dev. 13 (2011) 72–79, https://doi.org/10.1111/j.1525-142X.2010.00457.x.
[44] M. Endo, C. Sakai, T. Shimizu, Embryonic expression patterns of Hox genes in the oligochaete annelid Tubifex tubifex, Gene Expr. Patterns 22 (2016) 1–14, https://doi.org/10.1016/j.gep.2016.09.002.
[45] W.G. Damen, Parasegmental organization of the spider embryo implies that the parasegment is an evolutionary conserved entity in arthropod embryogenesis, Development 129 (2002) 1239–1250.
[46] S. Khadjeh, N. Turetzek, M. Pechmann, E.E. Schwager, E.A. Wimmer, W. G. Damen, N.-M. Prpic, Divergent role of the Hox gene Antennapedia in spiders is responsible for the convergent evolution of abdominal limb repression, Proc. Natl. Acad. Sci. 109 (2012) 4921–4926.
[47] G. Struhl, Role of the esc+ gene product in ensuring the selective expression of segment-specific homeotic genes in Drosophila, J. Embryol. Exp. Morphol. 76 (1983) 297–331.
[48] A. González-Reyes, G. Morata, The developmental effect of overexpressing a Ubx product in Drosophila embryos is dependent on its interactions with other homeotic products, Cell 61 (1990) 515–522, https://doi.org/10.1016/0092-8674(90)90533-k.
[49] R.S. Mann, D.S. Hogness, Functional dissection of ultrabithorax proteins in D. melanogaster, Cell 60 (1990) 597–610, https://doi.org/10.1016/0092-8674(90)90663-Y.
[50] D. Duboule, G. Morata, Colinearity and functional hierarchy among genes of the homeotic complexes, Trends Genet 10 (1994) 358–364, https://doi.org/10.1016/0168-9525(94)90132-5.

[51] S. Yekta, C.J. Tabin, D.P. Bartel, MicroRNAs in the Hox network: an apparent link to posterior prevalence, Nat. Rev. Genet 9 (2008) 789–796, https://doi.org/10.1038/nrg2400.
[52] V.S. Chopra, R.K. Mishra, "Mir"acles in hox gene regulation, Bioessays 28 (2006) 445–448, https://doi.org/10.1002/bies.20401.
[53] D.M. Tyler, K. Okamura, W.-J. Chung, J.W. Hagen, E. Berezikov, G.J. Hannon, E. C. Lai, Functionally distinct regulatory RNAs generated by bidirectional transcription and processing of microRNA loci, Genes Dev. 22 (2008) 26–36, https://doi.org/10.1101/gad.1615208.
[54] A. Stark, N. Bushati, C.H. Jan, P. Kheradpour, E. Hodges, J. Brennecke, D. P. Bartel, S.M. Cohen, M. Kellis, A single Hox locus in Drosophila produces functional microRNAs from opposite DNA strands, Genes Dev. 22 (2008) 8–13, https://doi.org/10.1101/gad.1613108.
[55] E. Hafen, M. Levine, W.J. Gehring, Regulation of Antennapedia transcript distribution by the bithorax complex in Drosophila, Nature 307 (1984) 287–289, https://doi.org/10.1038/307287a0.
[56] G. Struhl, R.A. White, Regulation of the Ultrabithorax gene of Drosophila by other bithorax complex genes, Cell 43 (1985) 507–519, https://doi.org/10.1016/0092-8674(85)90180-1.
[57] S.B. Carroll, R.A. Laymon, M.A. McCutcheon, P.D. Riley, M.P. Scott, The localization and regulation of Antennapedia protein expression in Drosophila embryos, Cell 47 (1986) 113–122, https://doi.org/10.1016/0092-8674(86)90372-7.
[58] K. Zhu, H.P. Spaink, A.J. Durston, Collinear Hox-Hox interactions are involved in patterning the vertebrate anteroposterior (A-P) axis, PLoS One 12 (2017), e0175287, https://doi.org/10.1371/journal.pone.0175287.
[59] M. Pechmann, E.E. Schwager, N. Turetzek, N.-M. Prpic, Regressive evolution of the arthropod tritocerebral segment linked to functional divergence of the Hox gene labial, Proc. R. Soc. B. 282 (2015) 20151162. https://doi.org/10.1098/rspb.2015.1162.
[60] P.W. Ingham, A. Martinez-Arias, The correct activation of Antennapedia and bithorax complex genes requires the fushi tarazu gene, Nature 324 (1986) 592–597, https://doi.org/10.1038/324592a0.
[61] R.A. White, R. Lehmann, A gap gene, hunchback, regulates the spatial expression of Ultrabithorax, Cell 47 (1986) 311–321, https://doi.org/10.1016/0092-8674(86)90453-8.
[62] P.W. Ingham, D. Ish-Horowicz, K.R. Howard, Correlative changes in homoeotic and segmentation gene expression in Krüppel mutant embryos of Drosophila, EMBO J. 5 (1986) 1659–1665, https://doi.org/10.1002/j.1460-2075.1986.tb04409.x.
[63] P.D. Riley, S.B. Carroll, M.P. Scott, The expression and regulation of Sex combs reduced protein in Drosophila embryos, Genes Dev. 1 (1987) 716–730, https://doi.org/10.1101/gad.1.7.716.
[64] U. Nauber, M.J. Pankratz, A. Kienlin, E. Seifert, U. Klemm, H. Jäckle, Abdominal segmentation of the Drosophila embryo requires a hormone receptor-like protein encoded by the gap gene knirps, Nature 336 (1988) 489–492, https://doi.org/10.1038/336489a0.
[65] T. Jack, M. Regulski, W. Mcginnis, Pair-rule segmentation genes regulate the expresion of the homeotic selector gene, Deformed, Genes & Development, GENE Dev. 2 (1988) 635–651, https://doi.org/10.1101/gad.2.6.635.
[66] A. Macías, S. Pelaz, G. Morata, Genetic factors controlling the expression of the abdominal-A gene of Drosophila within its domain, Mech. Dev. 46 (1994) 15–25, https://doi.org/10.1016/0925-4773(94)90034-5.
[67] R.S. Mann, Engrailed-mediated repression of Ultrabithorax is necessary for the parasegment 6 identity in Drosophila, Development 120 (1994) 3205–3212, https://doi.org/10.1242/dev.120.11.3205.
[68] F. Casares, E. Sánchez-Herrero, Regulation of the infraabdominal regions of the bithorax complex of Drosophila by gap genes, Development 121 (1995) 1855–1866, https://doi.org/10.1242/dev.121.6.1855.
[69] J. Castelli-Gair, The lines gene of Drosophila is required for specific functions of the Abdominal-B HOX protein, Development 125 (1998) 1269–1274, https://doi.org/10.1242/dev.125.7.1269.
[70] J.A. Kassis, J.A. Kennison, J.W. Tamkun, Polycomb and Trithorax Group Genes in Drosophila, Genetics 206 (2017) 1699–1725, https://doi.org/10.1534/genetics.115.185116.
[71] W.G.M. Damen, M. Weller, D. Tautz, Expression patterns of hairy, even-skipped, and runt in the spider Cupiennius salei imply that these genes were segmentation genes in a basal arthropod, Proc. Natl. Acad. Sci. 97 (2000) 4515–4519, https://doi.org/10.1073/pnas.97.9.4515.
[72] M. Schoppmeier, W.G.M. Damen, Expression of Pax group III genes suggests a single-segmental periodicity for opisthosomal segment patterning in the spider Cupiennius salei, Evol. Dev. 7 (2005) 160–169, https://doi.org/10.1111/j.1525-142X.2005.05018.x.
[73] A.P. McGregor, M. Pechmann, E.E. Schwager, N.M. Feitosa, S. Kruck, M. Aranda, W.G.M. Damen, Wnt8 is required for growth-zone establishment and development of opisthosomal segments in a spider, Curr. Biol. 18 (2008) 1619–1623, https://doi.org/10.1016/j.cub.2008.08.045.
[74] M. Pechmann, A.P. McGregor, E.E. Schwager, N.M. Feitosa, W.G.M. Damen, Dynamic gene expression is required for anterior regionalization in a spider, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 1468–1472. https://doi.org/10.1073/pnas.0811150106.
[75] R. Janssen, N. Turetzek, M. Pechmann, Lack of evidence for conserved parasegmental grooves in arthropods, Dev. Genes Evol. (2022), https://doi.org/10.1007/s00427-022-00684-5.
[76] E.E. Schwager, M. Pechmann, N.M. Feitosa, A.P. McGregor, W.G.M. Damen, hunchback functions as a segmentation gene in the spider achaearanea tepidariorum, Curr. Biol. 19 (2009) 1333–1340, https://doi.org/10.1016/j.cub.2009.06.061.
[77] A. Hubaud, O. Pourquié, Signalling dynamics in vertebrate segmentation, Nat. Rev. Mol. Cell Biol. 15 (2014) 709–721, https://doi.org/10.1038/nrm3891.
[78] R. Neijts, S. Amin, C. van Rooijen, J. Deschamps, Cdx is crucial for the timing mechanism driving colinear Hox activation and defines a trunk segment in the Hox cluster topology, Dev. Biol. 422 (2017) 146–154, https://doi.org/10.1016/j.ydbio.2016.12.024.
[79] B.-K. Liao, A.C. Oates, Delta-Notch signalling in segmentation, Arthropod Struct. Dev. 46 (2017) 429–447, https://doi.org/10.1016/j.asd.2016.11.007.
[80] R. Kageyama, H. Shimojo, A. Isomura, Oscillatory control of notch signaling in development, Adv. Exp. Med Biol. 1066 (2018) 265–277, https://doi.org/10.1007/978-3-319-89512-3_13.
[81] K.F. Sonnen, V.M. Lauschke, J. Uraji, H.J. Falk, Y. Petersen, M.C. Funk, M. Beaupeux, P. François, C.A. Merten, A. Aulehla, Modulation of phase shift between wnt and notch signaling oscillations controls mesoderm segmentation, e12, Cell 172 (2018) 1079–1090, https://doi.org/10.1016/j.cell.2018.01.026.
[82] A. Schönauer, C.L.B. Paese, M. Hilbrant, D.J. Leite, E.E. Schwager, N.M. Feitosa, C. Eibner, W.G.M. Damen, A.P. McGregor, The Wnt and Delta-Notch signalling pathways interact to direct pair-rule gene expression via *caudal* during segment addition in the spider *Parasteatoda tepidariorum*, Development 143 (2016) 2455–2463, https://doi.org/10.1242/dev.131656.
[83] E.V.W. Setton, P.P. Sharma, A conserved role for arrow in posterior axis patterning across Arthropoda, Dev. Biol. 475 (2021) 91–105, https://doi.org/10.1016/j.ydbio.2021.02.006.
[84] J.L. Rinn, M. Kertesz, J.K. Wang, S.L. Squazzo, X. Xu, S.A. Brugmann, L. H. Goodnough, J.A. Helms, P.J. Farnham, E. Segal, H.Y. Chang, Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs, Cell 129 (2007) 1311–1323, https://doi.org/10.1016/j.cell.2007.05.022.
[85] B. Pease, A.C. Borges, W. Bender, Noncoding RNAs of the Ultrabithorax domain of the Drosophila bithorax complex, Genetics 195 (2013) 1253–1264, https://doi.org/10.1534/genetics.113.155036.
[86] A. Hermann, D. Kosman, W. McGinnis, E. Tour, The expression of Drosophila melanogaster Hox gene Ultrabithorax is not overtly regulated by the intronic long noncoding RNA lncRNA:PS4 in a wild-type genetic background, G3 (Bethesda). 12 (2021) jkab374. https://doi.org/10.1093/g3journal/jkab374.
[87] R. Janssen, G.E. Budd, Gene expression suggests conserved aspects of Hox gene regulation in arthropods and provides additional support for monophyletic Myriapoda, Evodevo 1 (2010) 4, https://doi.org/10.1186/2041-9139-1-4.
[88] M. Mallo, C.R. Alonso, The regulation of Hox gene expression during animal development, Development 140 (2013) 3951–3963, https://doi.org/10.1242/dev.068346.
[89] D.J. Leite, M. Ninova, M. Hilbrant, S. Arif, S. Griffiths-Jones, M. Ronshaugen, A. P. McGregor, Pervasive microRNA Duplication in Chelicerates: Insights from the Embryonic microRNA Repertoire of the Spider *Parasteatoda tepidariorum*, Genome Biol. Evol. 8 (2016) 2133–2144, https://doi.org/10.1093/gbe/evw143.
[90] D.L. Garaulet, E.C. Lai, Hox miRNA regulation within the Drosophila Bithorax Complex: patterning behavior, Mech. Dev. 138 (2015) 151–159, https://doi.org/10.1016/j.mod.2015.08.006.
[91] S. Feng, C. Rastogi, R. Loker, W.J. Glassford, H. Tomas Rube, H.J. Bussemaker, R. S. Mann, Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors, Nat. Commun. 13 (2022) 3808, https://doi.org/10.1038/s41467-022-31501-2.
[92] T. Montavon, N. Soshnikova, Hox gene regulation and timing in embryogenesis, Semin. Cell Dev. Biol. 34 (2014) 76–84, https://doi.org/10.1016/j.semcdb.2014.06.005.
[93] J. Crocker, N. Abe, L. Rinaldi, A.P. McGregor, N. Frankel, S. Wang, A. Alsawadi, P. Valenti, S. Plaza, F. Payre, R.S. Mann, D.L. Stern, Low affinity binding site clusters confer hox specificity and regulatory robustness, Cell 160 (2015) 191–203, https://doi.org/10.1016/j.cell.2014.11.041.
[94] S. Merabet, R.S. Mann, To be specific or not: the critical relationship between hox and TALE proteins, Trends Genet 32 (2016) 334–347, https://doi.org/10.1016/j.tig.2016.03.004.
[95] C. Sánchez-Higueras, C. Rastogi, R. Voutev, H.J. Bussemaker, R.S. Mann, J.C.-G. Hombría, In vivo Hox binding specificity revealed by systematic changes to a single cis regulatory module, Nat. Commun. 10 (2019) 3597, https://doi.org/10.1038/s41467-019-11416-1.
[96] U. Löhr, L. Pick, Cofactor-interaction motifs and the cooption of a homeotic hox protein into the segmentation pathway of drosophila melanogaster, Curr. Biol. 15 (2005) 643–649, https://doi.org/10.1016/j.cub.2005.02.048.
[97] S.W. Choo, R. White, S. Russell, Genome-wide analysis of the binding of the Hox protein Ultrabithorax and the Hox cofactor Homothorax in Drosophila, PLoS One 6 (2011), e14778, https://doi.org/10.1371/journal.pone.0014778.
[98] M. Baek, J. Enriquez, R.S. Mann, Dual role for Hox genes and Hox co-factors in conferring leg motoneuron survival and identity in Drosophila, Development 140 (2013) 2027–2038, https://doi.org/10.1242/dev.090902.
[99] A.D. Buffry, A.P. McGregor, Micromanagement of drosophila post-embryonic development by hox genes, J. Dev. Biol. 10 (2022) 13, https://doi.org/10.3390/jdb10010013.
[100] E.V.W. Setton, P.P. Sharma, Cooption of an appendage-patterning gene cassette in the head segmentation of arachnids, Proc. Natl. Acad. Sci. USA 115 (2018) E3491–E3500, https://doi.org/10.1073/pnas.1720193115.
[101] C. Schomburg, N. Turetzek, N.-M. Prpic, Candidate gene screen for potential interaction partners and regulatory targets of the Hox gene labial in the spider

Parasteatoda tepidariorum, Dev. Genes Evol. (2020), https://doi.org/10.1007/s00427-020-00656-7.
[102] S. Iwasaki-Yokozawa, R. Nanjo, Y. Akiyama-Oda, H. Oda, Lineage-specific, fast-evolving GATA-like gene regulates zygotic gene activation to promote endoderm specification and pattern formation in the Theridiidae spider, (2022) 2022.06.10.495620. https://doi.org/10.1101/2022.06.10.495620.
[103] R.J. Diederich, V.K. Merrill, M.A. Pultz, T.C. Kaufman, Isolation, structure, and expression of labial, a homeotic gene of the Antennapedia Complex involved in Drosophila head development, Genes Dev. 3 (1989) 399–414, https://doi.org/10.1101/gad.3.3.399.
[104] M. Stauber, H. Jäckle, U. Schmidt-Ott, The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene, Proc. Natl. Acad. Sci. USA 96 (1999) 3786–3789, https://doi.org/10.1073/pnas.96.7.3786.
[105] . Schmidt-Ott, A.M. Rafiqi, S. Lemke, Hox3/zen and the evolution of extraembryonic epithelia in insects., Advances in Experimental Medicine and Biology. (2010). https://doi.org/10.1007/978–1-4419–6673-5_10.
[106] W.G.M. Damen, fushi tarazu: a Hox gene changes its role, BioEssays 24 (2002) 992–995, https://doi.org/10.1002/bies.10180.
[107] L. Pick, Hox genes, evo-devo, and the case of the ftz gene, Chromosoma 125 (2016) 535–551, https://doi.org/10.1007/s00412-015-0553-6.
[108] T. Copf, N. Rabet, M. Averof, Knockdown of spalt function by RNAi causes de-repression of Hox genes and homeotic transformations in the crustacean Artemia franciscana, Dev. Biol. 298 (2006) 87–94, https://doi.org/10.1016/j.ydbio.2006.07.024.
[109] A. Pavlopoulos, Z. Kontarakis, D.M. Liubicich, J.M. Serano, M. Akam, N.H. Patel, M. Averof, Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean, Proc. Natl. Acad. Sci. 106 (2009) 13897–13902, https://doi.org/10.1073/pnas.0902804106.
[110] W. Bender, M. Akam, F. Karch, P.A. Beachy, M. Peifer, P. Spierer, E.B. Lewis, D. S. Hogness, Molecular genetics of the bithorax complex in drosophila melanogaster, Science 221 (1983) 23–29, https://doi.org/10.1126/science.221.4605.23.
[111] Y. Tomoyasu, S.R. Wheeler, R.E. Denell, Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum, Nature 433 (2005) 643–647, https://doi.org/10.1038/nature03272.
[112] S.-J. Fu, J.-L. Zhang, S.-J. Chen, H.-H. Chen, Y.-L. Liu, H.-J. Xu, Functional analysis of Ultrabithorax in the wing-dimorphic planthopper Nilaparvata lugens (Stål, 1854) (Hemiptera: Delphacidae, Gene 737 (2020), 144446, https://doi.org/10.1016/j.gene.2020.144446.
[113] S. Schneuwly, A. Kuroiwa, W.J. Gehring, Molecular analysis of the dominant homeotic Antennapedia phenotype, EMBO J. 6 (1987) 201–206.
[114] M. Levine, E. Hafen, R.L. Garber, W.J. Gehring, Spatial distribution of Antennapedia transcripts during Drosophila development, EMBO J. 2 (1983) 2037–2046, https://doi.org/10.1002/j.1460-2075.1983.tb01697.x.
[115] E. Sánchez-Herrero, I. Vernós, R. Marco, G. Morata, Genetic organization of Drosophila bithorax complex, Nature 313 (1985) 108–113, https://doi.org/10.1038/313108a0.
[116] S. Tiong, L.M. Bone, J.R. Whittle, Recessive lethal mutations within the bithorax-complex in Drosophila, Mol. Gen. Genet 200 (1985) 335–342, https://doi.org/10.1007/BF00425445.
[117] R.F. Foelix, Biology of Spiders, Oxford University Press, Oxford, 1996.
[118] W.G. Damen, D. Tautz, Abdominal-B expression in a spider suggests a general role for Abdominal-B in specifying the genital structure, J. Exp. Zool. 285 (1999) 85–91, https://doi.org/10.1002/(sici)1097-010x(19990415)285:1<85::aid-jez10>3.0.co;2-n.
[119] M. Pechmann, S. Khadjeh, F. Sprenger, N.-M. Prpic, Patterning mechanisms and morphological diversity of spider appendages and their importance for spider evolution, Arthropod Struct. Dev. 39 (2010) 453–467, https://doi.org/10.1016/j.asd.2010.07.007.
[120] B. Mittmann, C. Wolff, Embryonic development and staging of the cobweb spider Parasteatoda tepidariorum C. L. Koch, 1841 (syn.: Achaearanea tepidariorum; Araneomorphae; Theridiidae), Dev. Genes Evol. 222 (2012) 189–216, https://doi.org/10.1007/s00427-012-0401-0.

Seminars in Cell and Developmental Biology 152-153 (2024) 58–69

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Hox gene functions in the *C. elegans* nervous system: From early patterning to maintenance of neuronal identity

Jayson J. Smith [a,b,*], Paschalis Kratsios [a,b,*]

[a] Department of Neurobiology, University of Chicago, 947 East 58th Street, Chicago, IL 60637, USA
[b] University of Chicago Neuroscience Institute, 947 East 58th Street, Chicago, IL 60637, USA

ARTICLE INFO

Keywords:
Hox transcription factors
C. elegans
Neural progenitors
Cell survival
Cell migration
Neuronal terminal identity

ABSTRACT

The nervous system emerges from a series of genetic programs that generate a remarkable array of neuronal cell types. Each cell type must acquire a distinct anatomical position, morphology, and function, enabling the generation of specialized circuits that drive animal behavior. How are these diverse cell types and circuits patterned along the anterior-posterior (A-P) axis of the animal body? Hox genes encode transcription factors that regulate cell fate and patterning events along the A-P axis of the nervous system. While most of our understanding of Hox-mediated control of neuronal development stems from studies in segmented animals like flies, mice, and zebrafish, important new themes are emerging from work in a non-segmented animal: the nematode *Caenorhabditis elegans*. Studies in *C. elegans* support the idea that Hox genes are needed continuously and across different life stages in the nervous system; they are not only required in dividing progenitor cells, but also in post-mitotic neurons during development and adult life. In *C. elegans* embryos and young larvae, Hox genes control progenitor cell specification, cell survival, and neuronal migration, consistent with their neural patterning roles in other animals. In late larvae and adults, *C. elegans* Hox genes control neuron type-specific identity features critical for neuronal function, thereby extending the Hox functional repertoire beyond early patterning. Here, we provide a comprehensive review of Hox studies in the *C. elegans* nervous system. To relate to readers outside the *C. elegans* community, we highlight conserved roles of Hox genes in patterning the nervous system of invertebrate and vertebrate animals. We end by calling attention to new functions in adult post-mitotic neurons for these paradigmatic regulators of cell fate.

1. Introduction

Building a nervous system is a multi-step process. Following gastrulation, the ectoderm generates neural stem cells which give rise to specified neuronal and glial precursor cells [1]. These precursors then divide to produce post-mitotic neurons and glia. After exiting the cell cycle, newborn cells must differentiate and migrate to specific positions along the anterior-posterior (A-P) axis of the nervous system. Once settled into a defined location, neurons acquire region-specific identities and often undergo dramatic morphological changes to generate axons and dendrites. Finally, neurons establish specific connections (synapses) with other neurons to form neural circuits. The function of these circuits throughout life relies on the ability of post-mitotic neurons to acquire and maintain terminal identity features, such as expression of neurotransmitter (NT) biosynthesis proteins, NT receptors, neuropeptides, and ion channels [2,3]. In this review, we collectively refer to the series of events from gastrulation to the generation of axons and dendrites as *early* steps of nervous system development. We refer to the subsequent processes of synapse formation and the control of neuronal terminal identity features as *late* steps of neuronal development.

Both early and late steps of nervous system development must be precisely controlled to generate distinct cell types and circuits at specific positions along the A-P axis. Studies in all major model organisms have provided compelling evidence that the Hox gene family plays fundamental roles during the early steps of neuronal development (reviewed in [4–6]). However, it remains largely unknown whether Hox genes are essential for the execution of later steps of neuronal development. In this review, we specifically focus on the function of Hox genes in the nervous system of the nematode *Caenorhabditis elegans*. Like in all other model organisms, Hox genes in *C. elegans* play critical roles during early patterning events of the nervous system by controlling cell proliferation, survival, and migration. However, the ease of conducting temporally

* Corresponding authors at: Department of Neurobiology, University of Chicago, 947 East 58th Street, Chicago, IL 60637, USA.
E-mail addresses: jaysonsmith@uchicago.edu (J.J. Smith), pkratsios@uchicago.edu (P. Kratsios).

https://doi.org/10.1016/j.semcdb.2022.11.012
Received 13 September 2022; Received in revised form 14 November 2022; Accepted 30 November 2022
Available online 7 December 2022

controlled gene inactivation studies in *C. elegans* has enabled researchers to bypass early pleiotropies and discover new (non-canonical) roles for Hox genes during late developmental and adult stages. Here, we begin with a brief introduction to Hox genes. In subsequent sections, we highlight the early (Section 3) and late (Section 4) roles of Hox genes in the *C. elegans* nervous system.

2. Hox genes: spatial collinearity and homeotic transformations

Hox genes encode conserved transcription factors renowned for their roles in A-P body patterning and segmentation [7]. Hox mutations were first discovered in the fruit fly *Drosophila melanogaster* in the early 20th century by Calvin Bridges, a student of Thomas Hunt Morgan. Bridges began his foundational studies of Hox genes following reports by William Bateson in the late 19th century of "freak" animals, such as a moth with its legs transformed into wings. Bateson coined the term "homeosis" to describe this kind of transformation of one body part into the likeness of another, from the Greek *homoios*, meaning "similar" or "same". Bridges' discovery of fly mutants with striking homeotic transformations inspired later work on Hox genes [8]. Specifically, Bridges identified flies with duplicated thoracic segments, whose halteres (structures essential for flight) were transformed into a second pair of wings. The genes that control this homeotic transformation belong to the *bithorax* complex (BX-C) of genes. Genetic studies performed by Bridges and Morgan provided the foundation for Edward Lewis and Thomas Kaufman to later characterize systematically the functions of Hox genes of the BX-C complex. Lewis, along with Christiane Nusslein-Volhard and Eric Wieschaus, determined that Hox genes are required to pattern and segment the fly body, a discovery for which they shared the 1995 Nobel Prize in Physiology or Medicine.

Given this rich history, it is not surprising that *Drosophila* has been the premier model to investigate the role of Hox genes in body patterning and segmentation along the A-P axis [9]. Much of our understanding of Hox-mediated control of neuronal development has emerged from studies in the ventral nerve cord of *Drosophila*, including the discovery of body region-specific neuromuscular networks (Reviewed in [10,11]). In vertebrates, studies of the spinal cord, hindbrain, and limb systems have also provided key insights into the function of Hox genes in body patterning, segmentation, and nervous system development [5,6,12]. Three highly conserved themes have emerged from all these studies of bilaterian animals: (1) Hox genes are organized into chromosomal clusters, (2) the sequence of expression domains of each Hox gene along the A-P body axis matches the order of the genes in the Hox cluster (*spatial collinearity)*, and (3) mutations in Hox genes can have dramatic (e.g., homeotic transformations) as well as subtle effects in animal development.

This review specifically focuses on the function of Hox genes in the *C. elegans* nervous system for two reasons. First, we compare and contrast *C. elegans* studies with observations in *Drosophila* and vertebrates, aiming to highlight the remarkable conservation of Hox gene functions in early patterning of the nervous system. Second, we call attention to novel functions of Hox genes recently discovered in adult *C. elegans* neurons, i.e., a continuous requirement to maintain cellular identity. Such new functions may be conserved in other cell types and organisms.

2.1. Hox genes in C. elegans

Although Hox gene functions have been traditionally studied in animals with segmented structures within the nervous system (e.g., rhombomeres in the vertebrate hindbrain [13–15]), the non-segmented nervous system of the nematode *C. elegans* has also been an invaluable platform. The nematode nervous system is simple and well-characterized, offering single-cell resolution. The adult hermaphrodite and male contain 302 and 387 neurons, respectively. All neurons in both sexes have been described and named, and their complete lineages are known [16–21]. Furthermore, all neuronal connections (synapses) have been mapped for both nematode sexes [22] along with the transcriptional profiles of all neurons in the mature hermaphrodite [23]. Finally, the nematode has (a) a short lifespan of about three weeks, (b) is amenable to powerful genetic approaches, and (c) only six Hox genes (compared to eight in *Drosophila* and thirty-nine in mice) are embedded in its genome.

Hox genes encode homeodomain proteins – transcription factors defined by the presence of a 60 amino acid-long homeodomain essential for DNA contact [24]. In total, there are 102 homeodomain transcription factors in *C. elegans*. Only 6 of these 102 are Hox transcription factors, and these 6 Hox proteins are the focus of this review. We refer readers interested in the remaining 96 homeodomain transcription factors to another comprehensive review [25].

The six genes in the *C. elegans* Hox cluster span 5 Mb of chromosome III (Fig. 1 A). Unlike other bilaterians, *C. elegans* only contains 4 Hox ortholog groups: Hox1, Hox5, Hox6–8, and Hox9–13 [26]. There is a single anterior Hox gene *ceh-13* (Lab/Hox1), two midbody Hox genes *lin*-39 (Scr/Dfd/Hox3–5) and *mab-5* (Antp/Hox6–8), and three posterior Hox genes *egl*-5 (Abd-A/Abd-B/Hox9–13), *nob-1* (Abd-B/Hox9–13) and *php-3* (Abd-B/Hox9–13). Except for *ceh-13*, which is positioned between *lin-39* and *mab-5* due to a genomic inversion event, spatial collinearity is observed for all *C. elegans* Hox genes both in embryonic (Fig. 1 B) and postembryonic tissues (Fig. 1 C) [26–28]. Over the past three decades, the nematode *C. elegans* has been a prime model to systematically investigate the function of Hox genes in neurodevelopment.

3. Hox gene functions in early patterning of the *C. elegans* nervous system

Unlike other model organisms, single (null) mutant analysis for five of the six *C. elegans* Hox genes (*lin-39, mab-5, egl-5, nob-1, php-3*) indicates that they are not required for organismal survival [29–34]; only the anterior Hox gene *ceh-13* ("***C. elegans* h**omeobox"-13) is required for survival to adulthood [35]. Although double *nob-1; php-3* mutants are lethal [36], double and triple null mutants for midbody Hox genes *lin-39* and *mab-5* and the posterior Hox gene *egl-5* are viable. Hence, *C. elegans* provides the opportunity to investigate Hox gene function both in early and late stages of nervous system development. To date, much of what we know about Hox genes in the *C. elegans* nervous system is derived from single and double mutant analysis. In the next three Sections (3.1 - 3.3), we highlight key roles for Hox in early patterning of the *C. elegans* nervous system.

3.1. Control of neuroblast cell divisions

Hox genes influence the timing and number of cell divisions in *Drosophila* and vertebrate neural stem cells [37–40]. Because the timing and location of all cell divisions are well-defined in *C. elegans*, it has been possible to perform in-depth investigations into the functions of Hox genes in neuroblast divisions.

A null mutation in *ceh-13* or a mutation that eliminates expression of both *nob-1* and *php-3* cause severe defects in the organization of the nervous system, intestine, body wall muscle, and epidermis, leading to early lethality [35,36,41,42]. The nervous system is especially affected in these Hox mutants relative to other tissues. For example, the most severe organizational defects in *ceh-13* mutants occur in the DA and SAB motor neuron classes, which become anteriorly displaced by several cell diameters [42]. Morphological defects in both *ceh-13* and *nob-1;php-3* mutants were initially attributed to abnormal cell migration [35,36], but recent work has also revealed a role for these Hox genes in the control of embryonic cell divisions, especially in the nervous system [42]. For example, the precursor cells that normally differentiate to generate the sheath (glial) cells that wrap around the ADE sensory neuron fail to exit the cell cycle in *ceh-13* mutants [42]. These animals also have delayed cell divisions in the neuroblasts that generate the DA, DD, and SAB

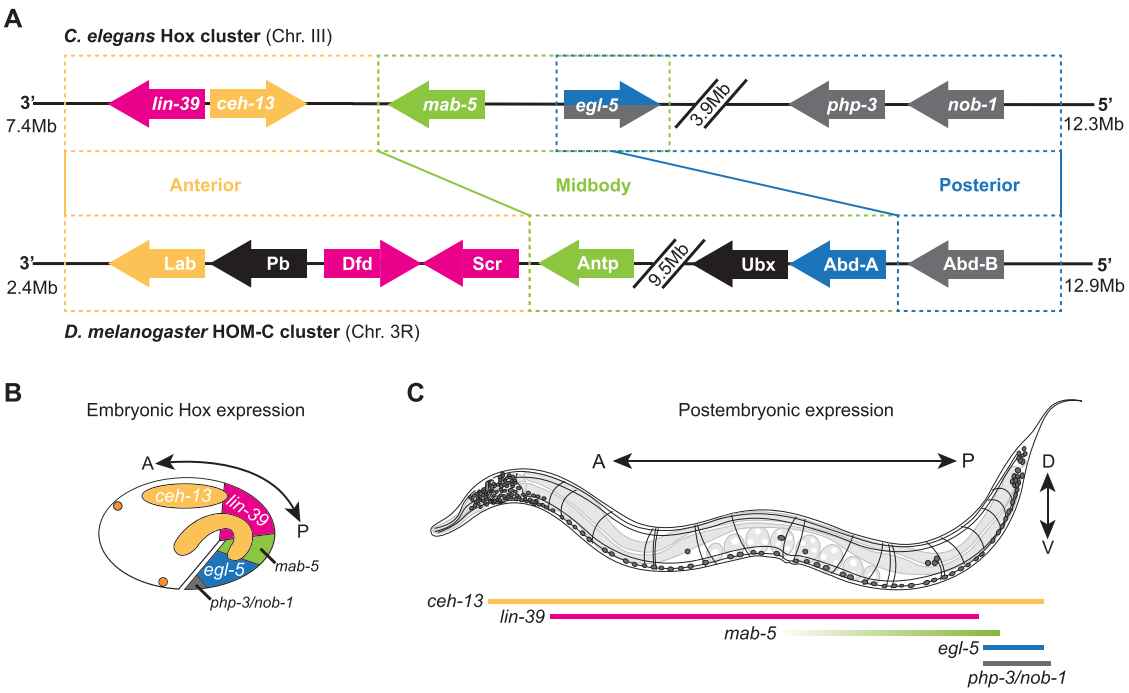


Fig. 1. Spatial collinearity is conserved in the *C. elegans* Hox cluster. (A) Depiction of Hox clusters in *C. elegans* (top) and *Drosophila* (bottom). Colors are used to depict the closely related orthologs. **(B)** Expression of Hox genes in *C. elegans* 1.5-fold embryo. **(C)** Expression of Hox genes in larval and adult stage *C. elegans*. A, Anterior; P, posterior; D, dorsal; V, ventral.

motor neurons. *nob-1; php-3* mutants also have delayed cell divisions; these occur in posterior neuroblasts that generate interneurons (PVQ class), sensory neurons (PHB class), and motor neurons (HSN class) [42]. Additionally, the neuroblast that normally generates the PLM and ALN sensory neurons in the posterior fails to divide entirely [42].

3.2. Control of neuronal lineage and cell survival

A distinguishing feature of *C. elegans* neurogenesis is that neurons emerge non-clonally from independent lineages or sub-lineages spanning the A-P axis. While some sub-lineages exclusively generate neurons, others also generate non-neuronal cells like muscle and hypodermis. Within a sub-lineage, neurons often do not share common features, i.e., they do not use the same neurotransmitter (NT) nor perform the same functions [43]. In *C. elegans*, studies on *ceh-13* and *mab-5* suggest that Hox gene expression is controlled, at least in part, by a lineage-based molecular mechanism; specific cell lineages and sub-lineages autonomously activate *ceh-13* and *mab-5* genes independently of cell position along the A-P axis [41,44]. In addition, Wnt signaling is necessary for Hox gene expression in *C. elegans* lineages [45–47]. Wnt signaling controls Hox expression in *Drosophila* and vertebrate embryos as well [48,49], constituting a conserved control mechanism of Hox gene expression. Further, neuronal lineages along the A-P axis of the nematode are controlled in part by Hox gene activity. This has been most evident from extensive studies on the patterning of neuroectodermal blast cell lineages in both *C. elegans* sexes (Section 3.2.1) and sensory neurons of the male tail (Section 3.2.2).

3.2.1. Lineage transformations in the neuroectoderm

The twelve ventral neuroectodermal blast cells called P1–12 that neatly align along the A-P axis of the *C. elegans* body constitute one of the most well-studied examples of neuronal lineages controlled by Hox gene activity (Fig. 2A). Early in L1, P cells divide to produce two cells: the anterior daughter cell (Pn.a) is a neuroblast, whereas the posterior daughter (Pn.p) is an epidermal cell. Each P cell will ultimately produce one epidermal and several distinct neuronal descendants (Fig. 2 B) [50]. Anterior descendants of P (Pn.a) become motor neurons (AS, VA, VB, VC, and VD classes) of the ventral nerve cord [50], an analogous structure to the vertebrate spinal cord. The motor neurons of the ventral nerve cord provide a straightforward context within which to interpret lineage transformations in Hox mutants for several reasons. First, the expression of each Hox gene is generally uniform in P descendants within each Hox expression domain [51,52]. Second, Hox expression domains establish variations in the patterning of fates along the ventral nerve cord motor neurons [51,53,54]. Third, ventral nerve cord motor neurons are linearly arrayed along the A-P axis and have region-specific characteristics and identities.

Here, we discuss key roles for midbody (*lin-39*, *mab-5*) and posterior (*egl-5)* Hox genes in the establishment of cell fate in P descendants. Mutations in *lin-39* were initially recovered in genetic screens for genes involved in cell survival [29,30]. In the *C. elegans* hermaphrodite, lack of *lin-39* gene activity leads to cell death of the sex-specific VC neurons that are used for egg-laying. Specifically, the midbody P descendants P3-P8. aap that normally generate VC neurons adopt instead the anterior or posterior Pn.aap fate of programmed cell death (Fig. 2C) [29,50,54–56]. Importantly, the same phenotype is observed in animals lacking *ceh-20* [57], a Hox cofactor of the extradenticle/Pbx class of homeodomain proteins [58]. Mechanistically, LIN-39 and CEH-20 form a complex that directly represses transcription of the proapoptotic gene *egl-1* (Bcl-2 family member), thereby ensuring survival of the VC neurons [59]. Cooperative DNA-binding of Hox proteins with their cofactors is a conserved strategy to increase Hox binding affinity and specificity [60].

Like in hermaphrodites, *lin-39* and *ceh-20* are necessary for survival of midbody P descendants in the *C. elegans* male [55,61]. That is, *lin-39* mutants exhibit fate transformations in the midbody P lineage that generates the male-specific CA and CP neurons: P3-P6.aap cells adopt the anterior P1.aap fate of programmed cell death, while P7-P8.aap survive but produce CAs and CPs that resemble the posterior P9-P11. aap fate, and thereby fail to produce serotonin – the normal NT of CA and CP neurons (Fig. 2 D) [54,55,61]. Thus, *lin-39* is required, in both *C. elegans* sexes, to distinguish motor neuron lineages in the midbody from lineages of motor neurons anterior or posterior to P3-P8 (Fig. 2A).

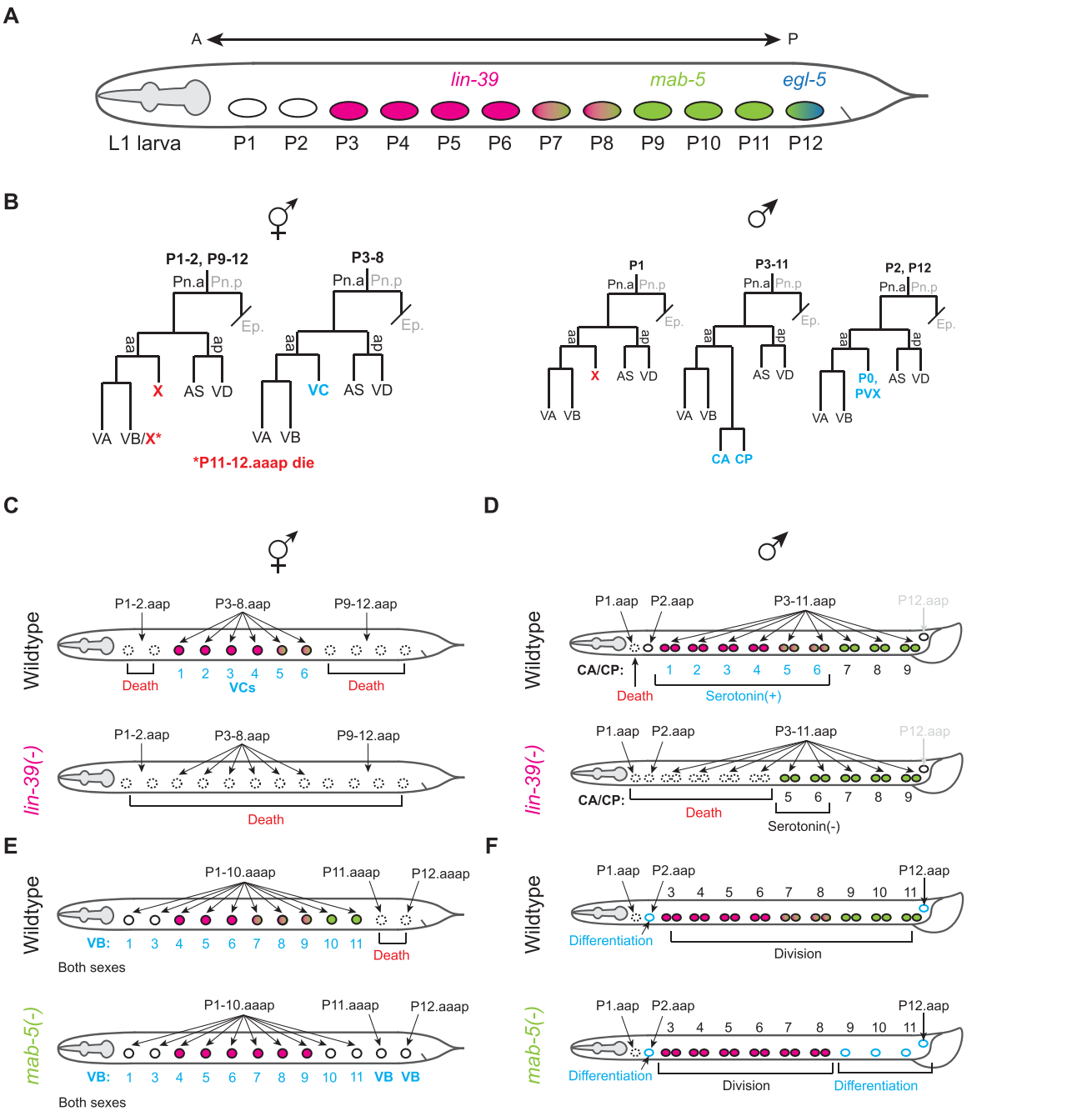


Fig. 2. Hox genes pattern the neuroectoderm in *C. elegans*. (A) Expression of Hox genes in the P cell lineage at the L1 stage. **(B)** Lineage diagrams indicating the descendants from each P-derived neuroblast along the A-P axis of hermaphrodites (left) and males (right). Posterior P lineage (Pn.p) generates cells of the epidermis and is excluded. Red X indicates programmed cell death. Blue text indicates neurons discussed in C-F. **(C)** LIN-39 patterns sex-specific VC neurons in the hermaphrodite midbody P lineage. **(D)** LIN-39 patterns sex-specific CA/CP neurons in the male anterior and midbody P lineage. **(E)** MAB-5 patterns the VB motor neurons in both sexes and the posterior P12.aap lineage in males **(F)**.

Lineage transformations have also been observed in animals lacking *mab*-5, another midbody Hox gene. Posterior to the *lin-39* expression domain (Fig. 2A), *mab*-5 has a role in determining the cell fate and survival of P descendants. MAB-5 protein is expressed in P7-P12 blast cells (Fig. 2A) [54]. In both hermaphrodite and male *C. elegans*, the anterior P1–10.aaap cells become the ventral B-type (VB) motor neurons, while the posterior P11–12.aaap cells undergo programmed cell death (Fig. 2 E). In *mab-5* mutants, all Pn.aaap cells survive and at least one of the P descendants that normally dies, P11.aaap, generates a putative VB neuron [62]. Hence, *mab-5* promotes cell death in posterior P11–12.aaap cells. Mechanistically, MAB-5 induces cell death by forming a complex with CEH-20 (extradenticle/Pbx) to activate transcription of the proapoptotic gene *egl-1* [57].

In addition to promoting cell death in P11.aaap cells of both *C. elegans* sexes, *mab-5* also controls patterning of neuronal cell lineages in Pn.aap cells in the *C. elegans* male. In wildtype males, the Pn.aap cells have definite fates based on position along the A-P axis: P1.aap undergoes cell death, P2.aap and P12.aap differentiate, and P(3−11).aap divide before differentiation. In *mab-5* mutants, the anterior cells P(1−8).aap are unaffected, while the posterior P(9−11).aap differentiate without division and morphologically resemble P2.aap cells (Fig. 2F) [62].

The posterior Hox gene *egl*-5 (Abd-A/Abd-B/Hox9–13) is expressed just posterior to the expression domain of *mab-5* (Fig. 2A). *egl-5*

regulates lineage patterning in the most posterior P cell P12; In *egl*-5 mutants of both sexes, P12 adopts a P11.p-like lineage [32]. Such lineage transformation is not observed in *ceh-20* mutants, suggesting CEH-20 (extradenticle/Pbx) may not act as cofactor for the posterior Hox protein EGL-5 [57]. Interestingly, work in *Drosophila* and vertebrates also suggested that Abd-B proteins, like EGL-5, can function independently of Hox cofactors [63,64].

The neuroectodermal blast cells P11 and P12 begin as developmentally equivalent left-right homologs at the L1 stage [32,65]. This symmetry is broken when each cell migrates to a specific location in the midline during the L1 stage, after which the terminal P cells produce distinct lineages [65]. In wildtype hermaphrodites, P11–12 divide to produce anterior neuroblasts P11–12.a, which undergo three more rounds of division to each produce three motor neurons of the ventral cord: P11.a produces VA11, AS11, VD12 and P12.a produces VA12, VD13, and PDB [16,66]. In *egl-5* mutants, the P12.a descendant VA12 neuron ectopically expresses molecular markers of anterior VA neurons, indicating transformation of a posterior to an anterior cell fate [67]. Such transformations are also widely observed in various cell types of flies and mice lacking Hox gene activity, and can be, at least partially, explained by posterior Hox genes repressing expression of anterior ones [5,68]. That is indeed the case for *egl-5*; it represses *lin-39* (Scr/Dfd/Hox3–5) in the P12.a descendant VA12 neuron [67].

Hox genes often function cell-autonomously to determine lineage patterns within their domain of expression [67,69–71]. However, the underlying mechanism is more complex when one considers interactions between multiple Hox genes. For example, the overlapping of *lin-39* and *mab-5* expression domains in the hermaphrodite posterior midbody gives rise to specific P cell lineages, distinct from those in more anterior or posterior domains where either *lin-39* or *mab-5* is exclusively expressed [54–56,72]. In wildtype males, LIN-39 specifies a serotonergic motor neuron fate in the descendants of P3-P8, while MAB-5 activity specifies a non-serotonergic interneuron fate in the descendants of P9-P11. In the absence of LIN-39, P7 and P8 generate non-serotonergic interneurons, reflecting the underlying activity of MAB-5 [54,55]. Thus, in male-specific neurons of the ventral nerve cord, LIN-39 limits the activity but *not* the expression of *mab-5* [53].

An additional facet to Hox function in lineage patterning is temporal segregation of Hox expression, i.e., some lineages require different Hox genes at different times. Both *mab-5* and *egl-5* expression are required in the P12 lineage (discussed above) but at different times [32,53,62]; *mab-5* expression is required in P12 and its early descendants, but its expression terminates at the onset of *egl-5* expression midway through P12 lineage progression [73]. This repression of *mab-5* expression is *egl-5*-dependent [54], suggesting that Hox expression can be dynamic within lineages, consistent with studies in *Drosophila*, mouse, and zebrafish.

3.2.2. Patterning the sensory neurons of the male tail

In the *C. elegans* male, the V and T cell lineages generate the lateral epidermis and contribute neurons and other cells to sensory rays, the tail structures essential for copulation (Fig. 3A). The anterior V cells (V1-V4) generate the seam cells that produce part of the cuticle, while the posterior V cells (V5-V6) generate the neuroblast lineages that produce cells for the sensory rays (Fig. 3 B-C). Each ray is a specialized sensillum that contains two sensory neurons, RnA and RnB, with distinct NT identity and connectivity (Fig. 3 D) [50,74–77]. Rays also differentially express *mab-5* and *egl-5* (Fig. 3 E). In *egl-5* mutants, posterior ray neurons (R3–5A, R3–5B) undergo a posterior-to-anterior transformation as they adopt the morphology and NT identity of anterior ray neurons (Fig. 3F) [77]. Consistent with neuronal transformations in flies and mice, it was proposed that this transformation occurs due to *egl-5* (posterior Hox) repressing *mab-5* (mid-body Hox). In *mab-5* mutants, ray neurons are absent altogether; V5-V6 descendants adopt the fate of their anterior homologs and produce the non-ray structures generated by V1-V4 cells (Fig. 3F) [32,62,72]. In addition, overexpression of either *mab-5* or *egl-5* also leads to anterior-to-posterior transformations in ray identities (Fig. 3F) [72,78–80].

These studies support a role for Hox genes as 'selectors' - genes required to pattern the specific identity of a tissue or organ, a concept introduced by Antonio Garcia-Bellido in 1975 based on work in *Drosophila* wing disc development [81]. In the *C. elegans* male tail, *egl*-5 and *mab-5* behave as selector genes of ray identity by controlling the region-specific features and the NT identity of sensory rays. The selector function of Hox in the nervous system appears highly conserved, as Hox1–4 also function as selectors of hindbrain patterning in vertebrates [82,83].

3.3. Control of cell migration during nervous system development

During development, neuroblasts and post-mitotic neurons often migrate to reach their final destination. In *C. elegans*, the Q neuroblasts and their descendants (sensory and interneurons) represent a prime model for the study of Hox genes in cell migratory behavior. In late embryos and young larvae (larval stage 1, L1), two Q neuroblasts occupy the right (QR neuroblast) and left (QL) sides of the *C. elegans* body at a similar position along the A-P axis (Fig. 4A). Descendants of each neuroblast migrate to a stereotypical position in the body, but in opposite directions. QL and its descendants migrate toward the posterior, whereas QR and its descendants migrate toward the anterior. The directionality of migration in Q neuroblast descendants depends on the activity of the midbody Hox genes *lin-39* and *mab-5*, as demonstrated by seminal studies from the labs of Cynthia Kenyon and Bob Horvitz (Fig. 4 **B**).

The descendants of QR (hereafter collectively referred to as QR.x) but not those of QL (QL.x) require *lin-39* to successfully complete their anterior migration [55,56]. Loss of *lin-39* has no apparent effect on the posterior migration of QL.x. Rather, in QL.x, *mab-5* is the key regulator and is both necessary and sufficient for posterior migration in these cells (Fig. 4 B) [84,85]. How do the QR and QL neuroblasts come to express different Hox genes despite occupying similar positions along the A-P axis? The canonical Wnt/Beta-catenin signaling pathway becomes activated early and specifically in QL, ultimately leading to the activation of *mab-5* [86–93]. QL requires an EGL-20/Wnt diffusion gradient along the A-P axis to activate *mab-5* [87,94]. This strategy appears specific to *mab-5*, as *lin-39* regulates QR.x migration through a Wnt-independent mechanism [95].

The anterior Hox *ceh-13* is also implicated in Q neuroblast migration. Unlike *lin-39* and *mab-5*, *ceh-13* mutants display defects in the migratory behaviors of both Q neuroblasts and their descendants [27]. In *ceh-13* mutants, the migration defects are subtle in QL.x, but QR.x terminate their anterior migration early (Fig. 4C) [27]. What explains the differential severity of these migratory defects in *ceh-13* mutants? Two Hox cofactors, *ceh-20* (extradenticle/Pbx) and *unc-62* (homothorax/Meis), are expressed only in QR.x, allowing QR.x descendants to respond to a migratory cue provided by the transmembrane protein MIG-13, an ortholog of Lrp12 in vertebrates [95,96]. CEH-13 activates *mig-13* in the anterior, while *mab-5* inhibits its expression in the posterior (Fig. 4C) [27,96]. Interestingly, the *mig-13* homolog in mice Lrp12 is expressed in migrating neurons during cortex development [97].

In addition to neuroblast migration, Hox genes also control the migration of post-mitotic neurons. The hermaphrodite-specific neurons (HSN class) are a bilaterally symmetric pair of motor neurons (HSNL/R) generated in the tail of the embryo [98]. Shortly after their birth, the HSNs begin anterior migration toward the hermaphrodite vulva, about halfway up the A-P axis. The posterior Hox protein EGL-5 is a key regulator of this process as well as many other aspects of HSN differentiation (Fig. 4 D) [98]. EGL-5 activates the expression of the zinc finger transcription factor *ham-2* (**H**SN **a**bnormal **m**igration-2) [99]. HSNs lacking *ham-2* terminate migration halfway to their final destination near the vulva [99]. Additionally, HSNs of *egl-5* mutants fail to down regulate the expression of the pro-migratory zinc finger

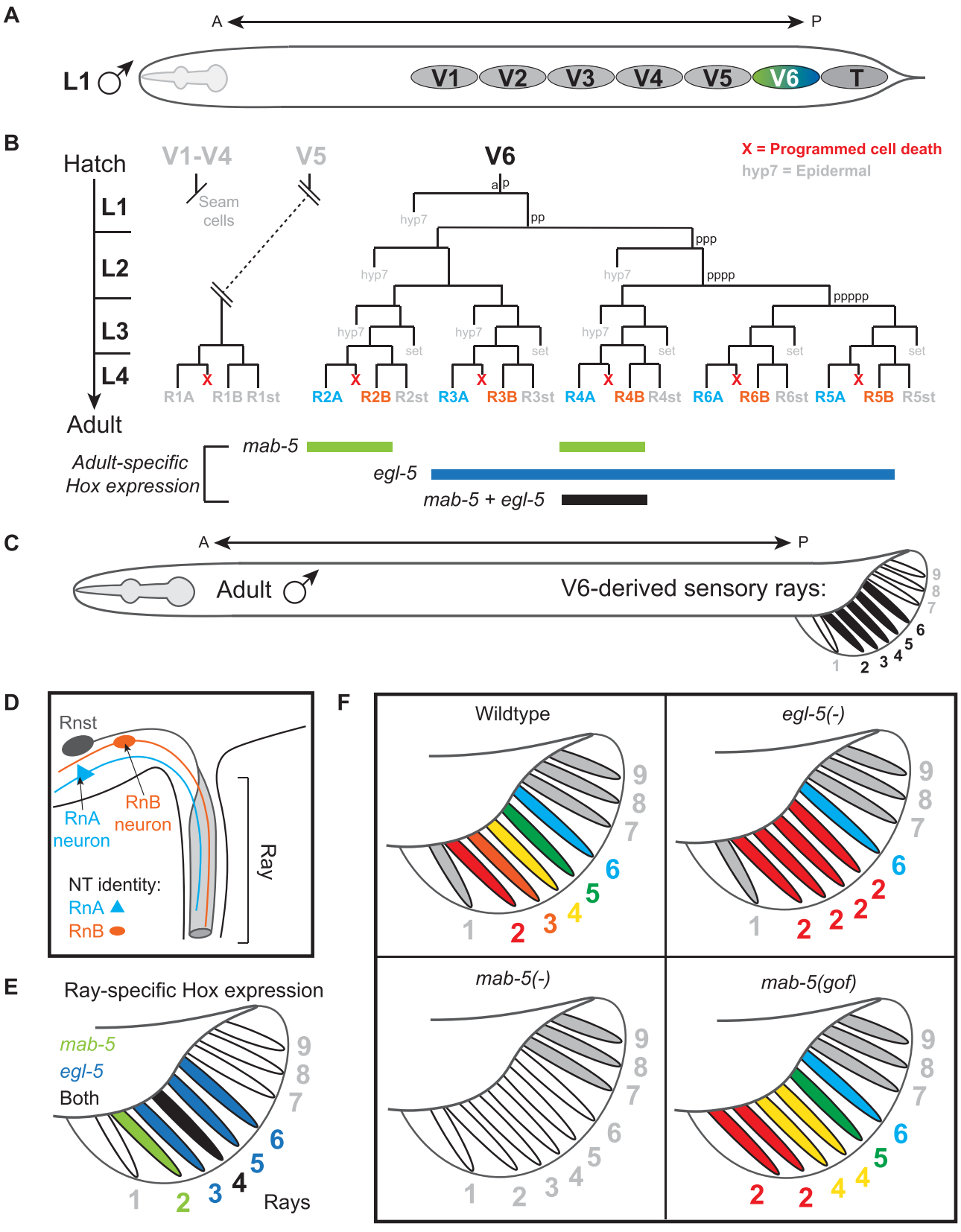


Fig. 3. Posterior Hox genes are selectors of sensory rays in the male tail. (A) Schematic of the V cells and T cell which generate the sensory rays in the copulatory male tail. **(B)** (top) Lineage diagrams of V5, which generates the R1A-R1B ray neurons, V6, which generates the R2A-R6A and R2B-R2A ray neurons; (bottom) Summary of hox expression across post-mitotic ray neurons in the adult. **(C)** Adult *C. elegans* male with V6-derived rays (ray 2–6) indicated in black. **(D)**The anatomy of a ray, including structure cell (Rnst) and both A- and B-type ray neurons color-coded blue and orange, respectively. **(E)** Expression of posterior Hox genes in V6-derived rays. **(F)** Summary of NT identity of rays in wildtype (top left), *egl-5* (top right) and *mab-5* (bottom left) null mutants, and in *mab-5* gain-of-function (*gof*) mutants (bottom right).

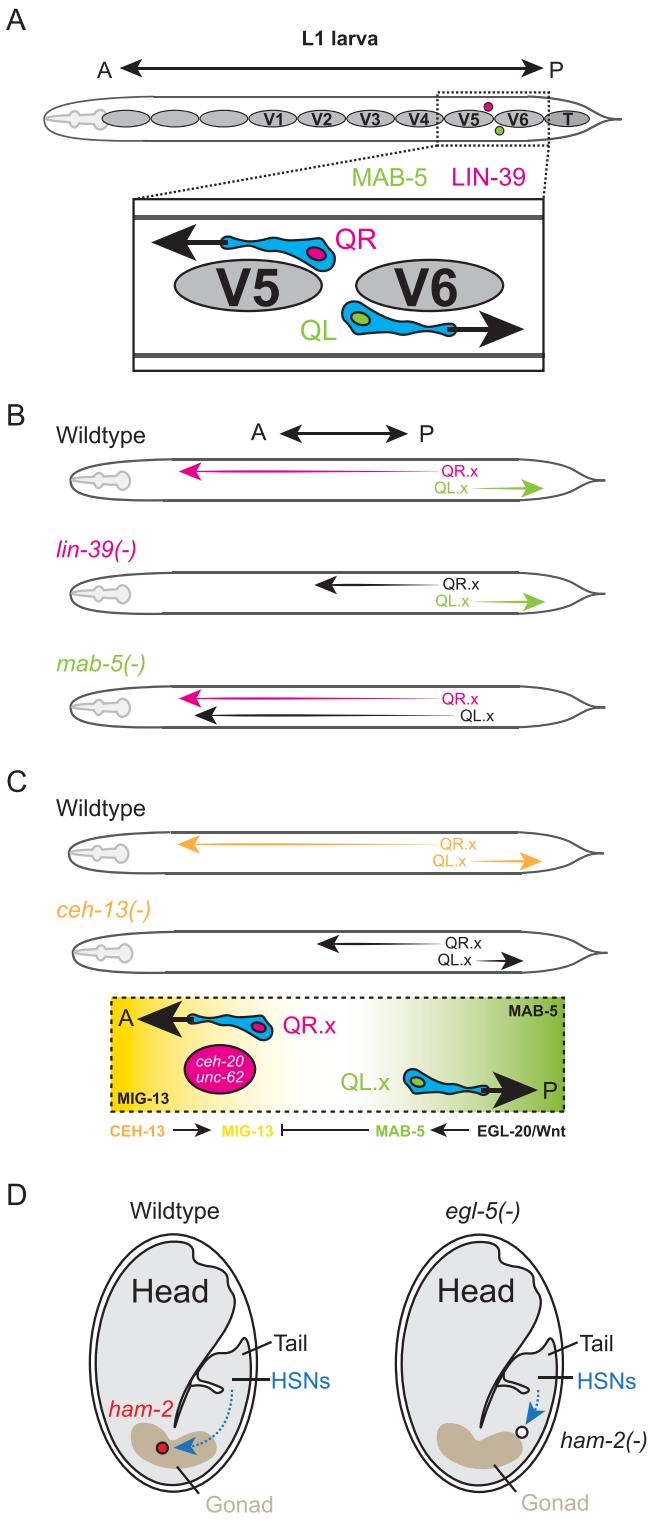


Fig. 4. Hox genes promote cell migration in the *C. elegans* nervous system. **(A)** Starting position of the QR (anteriorly polarized) and QL (posteriorly polarized) neuroblasts between V5 and V6 at L1 stage. **(B)** Summary of QR.x/QL.x neuroblast migration in wildtype (top) and in *lin-39* and *mab-5* null mutants (bottom). **(C)** Migration defects in *ceh-13* null mutants and depiction of opposing gradients of MIG-13 (anterior) and MAB-5 (posterior) that guide the migration of QR.X and QL.X. **(D)** Summary of HSN migration path in wildtype (left) and *egl-5* null mutants (right).

transcription factor, *egl-43*/EVI-1 [99]. Genes initially identified from Hox studies in *C. elegans* HSN migration (e.g., *egl-43/evi-1*, *epi-1*/laminin α5, *unc-71/adam-13*, and *ham-1/stox1*) have since been found to be necessary for migration of neural crest cells in vertebrates, suggesting the molecular mechanisms driving cell migration are deeply conserved [13,14,100].

4. Hox functions in late steps of *C. elegans* nervous system development

A striking one third of *C. elegans* neurons (93 of 302 neurons) maintain expression of at least one Hox gene in late developmental and adult stages [52]. By synthesizing information from multiple studies [23,52,67,101,102], we mapped the expression of each *C. elegans* Hox gene in the mature (L4) hermaphrodite nervous system with single-cell resolution (Fig. 5). All six *C. elegans* Hox genes are expressed in the nervous system. Their maintained expression suggests a continuous requirement for Hox in the nervous system at late developmental and adult stages of life. In fact, *C. elegans* has been instrumental in testing post-mitotic neuronal requirements for Hox function due to its short lifespan and powerful genetic tools, enabling Hox gene inactivation across different life stages. Such Hox requirements are less explored in flies and vertebrates. Below, we highlight recent studies on synapse maturation and neuronal terminal identity, which critically extend the Hox functional repertoire beyond early patterning.

4.1. Control of synapse formation/maturation in C. elegans

DA9, the ninth member of the DA class of cholinergic motor neurons, has been a powerful model to study synapse formation in *C. elegans*. The DA9 neuron is located close to the tail and its axon extends circumferentially to reach dorsal body wall muscles and form en passant neuromuscular synapses. In animals lacking activity of the posterior Hox gene *egl-5*, these DA9 synapses are generated onto more anteriorly located muscles when compared to wildtype animals, suggesting a synaptic specificity defect [67]. Split GFP reporter technology (GRASP) also revealed that the DA9 neurons of *egl-5* mutant animals fail to maintain synaptic inputs from the AVG interneurons. Importantly, the AVG inputs are properly established at early larval stages but fail to be maintained in adult *egl-5* mutants, indicating a Hox requirement in synapse maintenance. Together, these findings suggest that the posterior Hox gene *egl-5* controls both synaptic input and output of a posterior cholinergic motor neuron (DA9) in *C. elegans*. These observations are reminiscent of recent findings in *Drosophila* [103,104] and mice [105], suggesting a conserved role for Hox proteins in the formation and maintenance of neuronal synapses.

4.2. Control of neuronal terminal identity by C. elegans Hox genes

The function of every neuronal circuit critically relies on the ability of its constituent neurons to communicate with each other via neurotransmitters (NTs) and/or neuropeptides, as well as to display neuron type-specific morphological and electrophysiological signatures. These abilities are defined by the continuous expression of NT biosynthesis proteins, ion channels, neuropeptides, NT receptors, gap junction proteins, and cell adhesion molecules. Genes coding for such proteins have been termed "**terminal identity genes**" [106,107]. Because they are expressed continuously, from late developmental stages through adulthood, terminal identity genes determine the final (mature) identity and thus function of each neuron type. Emerging evidence suggests that all six *C. elegans* Hox genes are involved in the control of terminal identity of various neuron types [52,61,108,109]. Here, we specifically focus on touch receptor neurons (section 4.2.1) and nerve cord motor neurons (section 4.2.2) because mechanistic studies have been performed on these cells, strongly supporting the idea that Hox genes are continuously required to establish (during development) and maintain (in the adult)

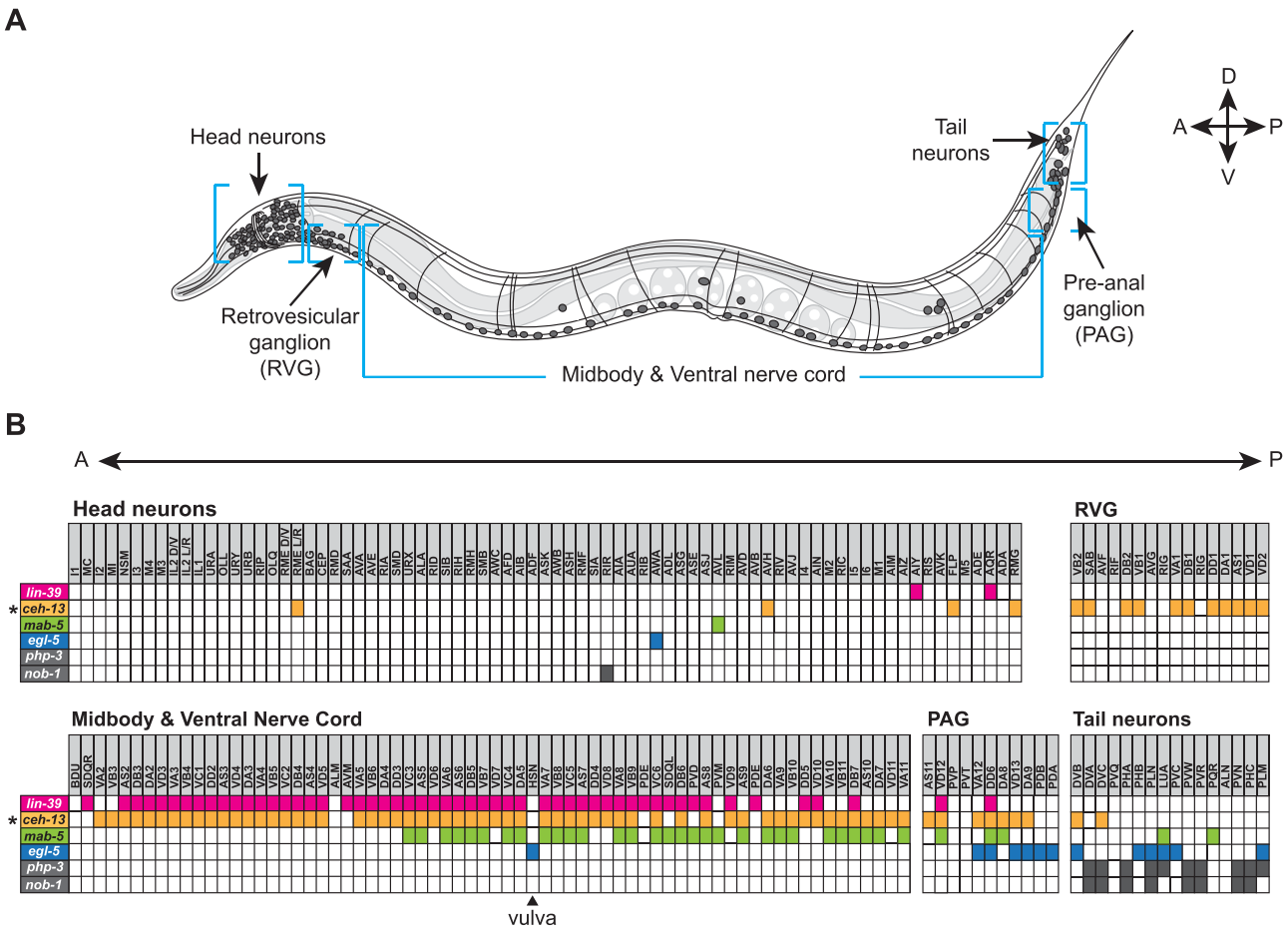


Fig. 5. Hox expression in the mature *C. elegans* nervous system. (A) Anatomy of the mature *C. elegans* nervous system. **(B)** Expression matrix of all 6 Hox genes (rows) in every neuron (columns) in the mature nervous system, ordered from anterior to posterior. Left-right and dorsal-ventral pairs of neurons are merged into one column when Hox gene expression is identical. Matrices are broken into the anatomical groups in A. **ceh-13* expression pattern is inferred from data in [101]; single-cell expression pattern is pending for this gene.

neuronal terminal identity features.

4.2.1. Establishment of touch receptor terminal identity

Six touch receptor neurons (TRNs) mediate sensory responses to light touch in *C. elegans*. TRNs are classified into four subtypes (classes). The bilaterally symmetrical pairs of ALM and PLM neurons are located in the midbody and tail region, respectively, whereas single AVM and PVM neurons are located in the midbody (Fig. 6A). The TRNs synapse onto and provide input to various command interneuron classes (PVC, AVB, AVD, AVA), which stimulate downstream motor neurons, thus generating touch reflex responses.

C. elegans animals lacking gene activity of the posterior Hox gene *egl-5* are touch-insensitive at the tail, suggesting *egl-5* controls the development of posteriorly located TRNs, the PLM neurons [110]. Indeed, *egl-5* controls PLM morphological characteristics, such as neurite length, by repressing anterior Hox genes (*lin-39, mab-5*) and the Hox cofactors *ceh-20* (extradenticle/Pbx) and *unc-62* (homothorax/Meis) [111]. Further, *egl-5* is necessary for the terminal identity and function of PLM neurons by activating the expression of various terminal identity genes, such as the gap junction-encoding gene *inx-13* [111,112] (Fig. 6 B). The case of *egl-5* highlights a recurring theme of Hox gene function across model systems: posterior Hox genes repress the expression of anterior Hox to generate a distinct (novel) cell fate. This simple mechanism is often the underlying cause of cell fate transformations in Hox mutant animals, providing a conceptual framework for the evolution of novel cell types in the nervous system [5,113]. Consistent with this idea, the PLM neurons in *egl-5* mutants do acquire molecular and morphological features of anteriorly located TRNs, called ALM neurons [71].

In more anteriorly located TRNs, the ALM neurons, the anterior Hox gene *ceh-13* regulates terminal identity; expression of a handful of ALM terminal identity genes (e.g., *mec-4/SCNN1 sodium channel, mec-17/ATAT1 tubulin acetyltransferase*) is reduced in *ceh-13* mutant animals (Fig. 6 B) [111,114]. Mechanistically, CEH-13 in ALM and EGL-5 in PLM act indirectly by controlling the levels of expression of *mec-3*, a LIM homeodomain transcription factor [111,114]. MEC-3 is a terminal selector for both ALM and PLM neurons (Fig. 6 B) [115,116]. Terminal selectors are transcription factors that determine the identity and function of specific neuron types by directly activating the expression of multiple terminal identity genes (e.g., NT biosynthesis proteins, ion channels, neuropeptides) [2]. The Hox proteins CEH-13 and EGL-5 increase the probability of transcriptional activation of the terminal selector gene *mec-3* in ALM and PLM neurons, respectively, ensuring robustness of TRN terminal differentiation. This mechanism is also relevant for the problem of neuronal subtype diversification, which is evident in every nervous system. A common TRN fate is controlled by the terminal selector MEC-3, but specific ALM and PLM terminal identities are established through the activities of anterior (CEH-13) and posterior (EGL-5) Hox proteins (Fig. 6 B) [115,116]. It is important to note that, in the context of TRNs, current evidence suggests that Hox proteins do not act as terminal selectors because it remains unclear whether they directly control terminal identity genes [111,114].

4.2.2. Establishment and maintenance of motor neuron terminal identity

Nine classes of motor neurons (MNs) are found in the *C. elegans* nerve

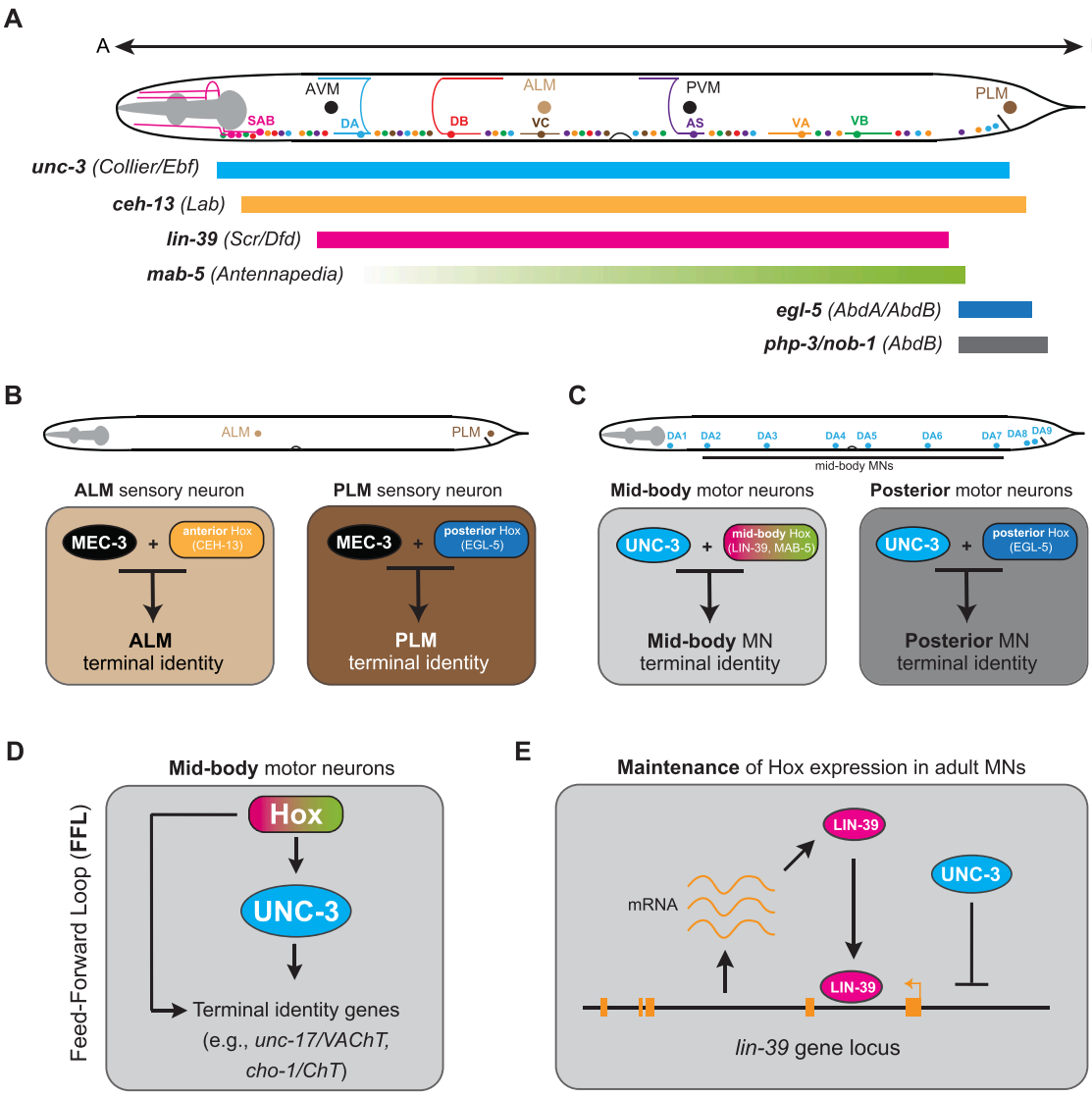


Fig. 6. Hox genes control neuronal terminal identity. (A) (Top) Schematic of mature *C. elegans* with color-coded motor neurons and TRNs. Text label colors of motor neuron classes correspond to circles/cell bodies on the ventral surface of the animal. (Bottom) Hox expression domains are indicated with color code that is consistent with previous figures. **(B)** Schematic depicting anterior (CEH-13) and posterior (EGL-5) Hox genes collaborating with A/PLM terminal selector MEC-3 in sensory neurons to determine neuronal terminal identity. **(C)** Midbody (LIN-39, MAB-5) and posterior (EGL-5) Hox genes collaborate with UNC-3 to co-activate terminal identity genes in ventral cord motor neurons. **(D)** Midbody Hox genes and UNC-3 operate in a positive feedforward loop (FFL) to ensure robust expression of terminal identity genes in midbody motor neurons. **(E)** Hox (LIN-39) expression is maintained in motor neurons throughout life via positive autoregulation, which is balanced by negative UNC-3 feedback.

cord of hermaphrodite animals. Based on neurotransmitter usage, they can be classified into cholinergic (SAB, DA, DB, VA, VB, AS, VC) and GABAergic (DD, VD) MNs (Fig. 6A). The SAB, DA, DB, and DD neurons are generated embryonically, whereas the VA, VB, VC, VD, and AS neurons are generated post-embryonically [117]. The terminal identity of most cholinergic MN classes in the nerve cord (SAB, DA, DB, VA, VB, AS) critically depends on the terminal selector UNC-3, member of the conserved family of Collier/Olf/Ebf (COE) family of TFs [118–120]. Mechanistically, UNC-3 binds directly to the *cis*-regulatory region of multiple terminal identity genes (e.g., acetylcholine [ACh] biosynthesis proteins, ion channels, neuropeptides) and activates their transcription. The homeodomain TF UNC-30 (PITX) acts in analogous manner in GABAergic (DD, VD) MNs [121,122].

Like anterior and posterior TRN subtypes, the study of nerve cord MNs offered critical insights into the role of Hox genes in neuronal subtype diversification. For example, the DA class consists of nine cholinergic MNs, which can be subdivided into three groups based on cell body position: (1) the anterior DA1 neuron is located anteriorly (at the retrovesicular ganglion), (2) the midbody DA2–7 neurons are located along the nerve cord, and (3) the posterior DA8–9 neurons are located at the posterior (preanal) ganglion (Fig. 6C). Neurons of the remaining cholinergic (DB, VA, VB, AS, VC) and GABAergic (DD, VD) classes are organized in a similar manner along the nerve cord (Fig. 6A). Moreover, anterior (e.g., DA1), midbody (e.g., DA2–7), and posterior (e.g., DA8–9) neurons do show distinct connectivity and expression profiles of terminal identity genes.

Hox genes control the terminal identity of mid-body and posterior cholinergic MNs via an intersectional strategy that involves the terminal selector UNC-3. For example, UNC-3 is expressed in all 9 DA neurons but collaborates with mid-body Hox genes *lin-39* and *mab-5* and the Hox cofactor *ceh-20* (extradenticle/Pbx) in mid-body DA2–7 neurons to control expression of multiple terminal identity genes specific to these neurons (Fig. 6C). Similarly, UNC-3 and the posterior Hox gene *egl-5* determine the terminal identity of posterior DA9 neurons by co-

activating a different set of terminal identity genes (Fig. 6 C). Biochemical evidence suggests that LIN-39 and MAB-5 – like UNC-3 – act directly by binding on the *cis*-regulatory region of terminal identity genes (*unc-129, del-1, acr-2, dbl-1, unc-77, slo-2)*. The direct mode of action combined with their continuous expression in MNs during developmental and adult stages support the idea that, in mid-body cholinergic MNs (DA, DB, VA, VB, AS classes), the Hox proteins LIN-39 and MAB-5 act as *bona fide* terminal selectors. The anterior Hox gene *ceh-13* is expressed in anterior MNs, but it remains unknown whether it controls of MN terminal identity.

A defining feature of terminal selectors is continuous requirement throughout life [106]. Are Hox genes required during adulthood to maintain the terminal identity and thereby the continuous functionality of nerve cord MNs? Protein depletion experiments using the auxin inducible degradation (AID) system demonstrated that the midbody Hox protein LIN-39 is indeed required in adult life to maintain expression of various terminal identity genes (e.g., ACh biosynthesis proteins, ion channels) of mid-body MNs [102,123,124]. Importantly, the terminal selectors LIN-39 and UNC-3 (Collier/Ebf) operate in a positive feedforward loop to ensure continuous and robust expression of terminal identity genes (Fig. 6 D). Of note, the continuous LIN-39 requirement in adult *C. elegans* MNs demonstrates a new function for Hox proteins, beyond their textbook roles in developmental patterning. Supporting this new role, a recent study in *Drosophila* demonstrated that the Hox gene *Ultrabithorax (Ubx)* is necessary for dopaminergic neuron identity and function in the adult [125].

How is Hox gene expression maintained in adult MNs? Feng et al. [124] identified a two-component mechanism for homeostatic control of *lin-39* expression in adult MNs: (1) *lin-39* positively and directly regulates its own expression, and (2) the transcriptional autoregulation of *lin-39* is counterbalanced by negative UNC-3 feedback (Fig. 6 E). The same mechanism applies to *mab-5* [124], the other mid-body Hox gene expressed in *C. elegans* nerve cord MNs. Future work in *C. elegans* and other model systems will determine whether transcriptional autoregulation and maintenance of neuronal terminal identity are conserved features of Hox gene function in the nervous system.

Recent work demonstrated that Hox genes can also act as terminal selectors in other MN classes that do not express *unc-3*. In GABAergic MNs, the midbody Hox genes *lin-39* and *mab-5* collaborate with the terminal selector *unc-30* to control terminal identity gene expression during development [102]. In cholinergic VC neurons, LIN-39 is required not only to establish during development but also maintain in the adult the expression of multiple terminal identity genes (e.g., ACh biosynthesis components, NT receptors), corroborating the emerging notion that Hox proteins can act as terminal selectors in specific neuron types.

5. The ancestral function of Hox genes is likely neuronal

In 1998, Jean Deutsch and Herve Le Guyader proposed that the primordial function of Hox genes is to design and pattern the nervous system [126]. This hypothesis is largely based on (a) comparisons of Hox gene expression across species and (b) a lack of correlation between the number of Hox genes and the increase in morphological diversity (e.g., segmental differentiation) during evolution. In support of the first point, all extant bilaterians studied to date express Hox genes in the central nervous system (CNS). Further, some distantly related organisms, like leeches and amphioxus, which are separated in evolution by more than 500 million years, express Hox genes *only* in the CNS [126–129]. Consequently, this hypothesis postulates that the well-known functions of Hox genes in patterning various tissues outside the CNS along the A-P axis of the bilaterian body plan are derived functions [126,130]. In their second point, Deutsch and Le Guyader acknowledge that the number and diversity of Hox genes can be similar between morphologically complex organisms (like vertebrates) and morphologically simpler species (like amphioxus). In this review, we explored the functions of Hox in nervous system development of an unsegmented (morphologically simple) animal, the nematode *C. elegans*. Like findings in complex model organisms, Hox genes are essential for neural patterning in *C. elegans*, lending support to the hypothesis that the ancestral function of Hox is neuronal. Importantly, Hox genes are involved at every level of nervous system development in *C. elegans*, from positioning neuroblasts in the embryo to maintaining neuronal terminal identity in the adult. Future studies are needed to determine whether the theme of a continuous requirement of Hox in the nervous system is widely applicable across bilaterians.

Declaration of Competing Interest

The authors declare no interests or relationships - financial, personal, or otherwise - that influence or could influence the work reported in this paper.

Acknowledgements

We would like to thank several members of the Kratsios lab (Manasa Prahlad, Honorine Destain, Weidong Feng, Filipe Marques) for helpful discussions and edits pertaining to this manuscript. This work was supported by two grants to P.K (R01 NS116365-01, R01 NS118078-01) from the National Institute of Neurological Disorders and Stroke (NINDS).

References

[1] M.W. Zappaterra, M.K. Lehtinen, The cerebrospinal fluid: regulator of neurogenesis, behavior, and beyond, Cell. Mol. Life Sci. 69 (17) (2012) 2863–2878.
[2] O. Hobert, P. Kratsios, Neuronal identity control by terminal selectors in worms, flies, and chordates, Curr. Opin. Neurobiol. 56 (2019) 97–105.
[3] O. Hobert, Chapter twenty-five - terminal selectors of neuronal identity, in: P. M. Wassarman (Ed.), Current Topics in Developmental Biology, Academic Press, 2016, pp. 455–475.
[4] M. Di Bonito, J.C. Glover, M. Studer, Hox genes and region-specific sensorimotor circuit formation in the hindbrain and spinal cord, Dev. Dyn. 242 (12) (2013) 1348–1368.
[5] P. Philippidou, J.S. Dasen, Hox genes: choreographers in neural development, architects of circuit organization, Neuron 80 (1) (2013) 12–34.
[6] H.J. Parker, R. Krumlauf, A. Hox, Gene regulatory network for hindbrain segmentation, Curr. Top. Dev. Biol. 139 (2020) 169–203.
[7] D. Lemons, W. McGinnis, GenomIc evolution of hox gene clusters, Science 313 (5795) (2006) 1918–1922.
[8] E.B. Lewis, A. Gene, Complex controlling segmentation in drosophila, in: H. D. Lipshitz (Ed.), Genes, Development and Cancer: The Life and Work of Edward B. Lewis, Springer US, Boston, MA, 2004, pp. 205–217.
[9] P.A. Lawrence, G. Morata, Homeobox genes: their function in drosophila segmentation and pattern formation, Cell 78 (2) (1994) 181–189.
[10] A. Estacio-Gómez, F.J. Díaz-Benjumea, Roles of Hox genes in the patterning of the central nervous system of Drosophila, Fly 8 (1) (2013) 26–32.
[11] M. Gummalla, S. Galetti, R.K. Maeda, F. Karch, Hox gene regulation in the central nervous system of Drosophila, Front. Cell. Neurosci. 8 (2014).
[12] C. Nolte, R. Krumlauf, Expression of Hox genes in the Nervous System of Vertebrates, Springer, New York, NY, 2007, pp. 14–41.
[13] V. Prince, A. Lumsden, Hoxa-2 expression in normal and transposed rhombomeres: independent regulation in the neural tube and neural crest, Development 120 (4) (1994) 911–923.
[14] J.R. Saldivar, C.E. Krull, R. Krumlauf, L. Ariza-McNaughton, M. Bronner-Fraser, Rhombomere of origin determines autonomous versus environmentally regulated expression of Hoxa-3 in the avian embryo, Development 122 (3) (1996) 895–904.
[15] P.A. Trainor, R. Krumlauf, Patterning the cranial neural crest: hinbrain segmentation and hox gene plasticity, Nat. Rev. Neurosci. 1 (2) (2000) 116–124.
[16] D.H. Hall, R.L. Russell, The posterior nervous system of the nematode Caenorhabditis elegans: serial reconstruction of identified neurons and complete pattern of synaptic interactions, J. Neurosci. 11 (1) (1991), 1 LP-22.
[17] S. Ward, N. Thomson, J.G. White, S. Brenner, Electron microscopical reconstruction of the anterior sensory anatomy of the nematode Caenorhabditis elegans, J. Comp. Neurol. 160 (3) (1975) 313–337.
[18] R.W. Ware, D. Clark, K. Crossland, R.L. Russell, The nerve ring of the nematode Caenorhabditis elegans: sensory input and motor output, J. Comp. Neurol. 162 (1) (1975) 71–110.
[19] J.G. White, E. Southgate, J.N. Thomson, S. Brenner, The structure of the nervous system of the nematode caenorhabditis elegans, Philos. Trans. R. Soc. Lond. Biol. Sci. 314 (1165) (1986) 1–340.

[20] R.M. Durbin, Studies in the Development and Organisation of the Nervous System of Caenorhabditis elegans, 1987, pp. 121–121.
[21] B.L. Chen, D.H. Hall, D.B. Chklovskii, Wiring optimization can relate neuronal structure and function, Proc. Natl. Acad. Sci. USA 103 (12) (2006) 4723–4728.
[22] S.J. Cook, T.A. Jarrell, C.A. Brittin, Y. Wang, A.E. Bloniarz, M.A. Yakovlev, K.C. Q. Nguyen, L.T.H. Tang, E.A. Bayer, J.S. Duerr, H.E. Bülow, O. Hobert, D.H. Hall, S.W. Emmons, Whole-animal connectomes of both Caenorhabditis elegans sexes, Nature 571 (7763) (2019) 63–71.
[23] S.R. Taylor, G. Santpere, A. Weinreb, A. Barrett, M.B. Reilly, C. Xu, E. Varol, P. Oikonomou, L. Glenwinkel, R. McWhirter, A. Poff, M. Basavaraju, I. Rafi, E. Yemini, S.J. Cook, A. Abrams, B. Vidal, C. Cros, S. Tavazoie, N. Sestan, M. Hammarlund, O. Hobert, D.M. Miller 3rd, Molecular topography of an entire nervous system, Cell 184 (16) (2021) 4329–4347, e23.
[24] T.R. Bürglin, M. Affolter, Homeodomain proteins: an update, Chromosoma 125 (3) (2016) 497–521.
[25] O. Hobert, Homeobox genes and the specification of neuronal identity, Nat. Rev. Neurosci. 22 (10) (2021) 627–636.
[26] A. Aboobaker, M. Blaxter, Hox gene evolution in nematodes: novelty conserved, Curr. Opin. Genet. Dev. 13 (6) (2003) 593–598.
[27] B. Tihanyi, T. Vellai, Á. Regős, E. Ari, F. Müller, K. Takács-Vellai, The C. elegans Hox gene ceh-13 regulates cell migration and fusion in a non-colinear way. Implications for the early evolution of Hoxclusters, BMC Dev. Biol. 10 (1) (2010), 78-78.
[28] J. Sulston, Z. Du, K. Thomas, R. Wilson, L. Hillier, R. Staden, N. Halloran, P. Green, J. Thierry-Mieg, L. Qiu, S. Dear, A. Coulson, M. Craxton, R. Durbin, M. Berks, M. Metzstein, T. Hawkins, R. Ainscough, R. Waterston, The C. elegans genome sequencing project: a beginning, Nature 356 (1992) 37–41.
[29] H.M. Ellis, H.R. Horvitz, Genetic control of programmed cell death in the nematode C. elegans, Cell 44 (6) (1986) 817–829.
[30] W. Fixsen, P. Sternberg, H. Ellis, R. Horvitz, Genes that affect cell fates during the development of Caenorhabditis elegans, Cold Spring Harb. Symp. Quant. Biol. 50 (1985) 99–104.
[31] J. Hodgkin, Male phenotypes and mating efficiency in caenorhabditis elegans, Genetics 103 (1) (1983) 43–64.
[32] A. Chisholm, Control of cell fate in the tail region of C. elegans by the gene egl-5, Development 111 (4) (1991) 921–932.
[33] M.D. Nelson, E. Zhou, K. Kiontke, H. Fradin, G. Maldonado, D. Martin, K. Shah, D. H.A. Fitch, A. Bow-Tie, Genetic architecture for morphogenesis suggested by a genome-wide RNAi screen in Caenorhabditis elegans, PLOS Genet. 7 (3) (2011), e1002010.
[34] S. Hori, S. Oda, Y. Suehiro, Y. Iino, S. Mitani, OFF-responses of interneurons optimize avoidance behaviors depending on stimulus strength via electrical synapses, PLOS Genet. 14 (6) (2018), e1007477.
[35] K. Brunschwig, C. Wittmann, R. Schnabel, T.R. Burglin, H. Tobler, F. Muller, Anterior organization of the Caenorhabditis elegans embryo by the labial-like Hox gene ceh-13, Development 126 (7) (1999) 1537–1546.
[36] K. Van Auken, D.C. Weaver, L.G. Edgar, W.B. Wood, Caenorhabditis elegans embryonic axial patterning requires two recently discovered posterior-group Hox genes, Proc. Natl. Acad. Sci. USA 97 (9) (2000) 4499–4503.
[37] C. Maurange, A.P. Gould, Brainy but not too brainy: starting and stopping neuroblast divisions in Drosophila, Trends Neurosci. 28 (1) (2005) 30–36.
[38] M. Gouti, A. Gavalas, Hoxb1 controls cell fate specification and proliferative capacity of neural stem and progenitor cells, Stem Cells 26 (8) (2008) 1985–1997.
[39] B.C. Bello, F. Hirth, A.P. Gould, A pulse of the drosophila hox protein abdominal-a schedules the end of neural proliferation via neuroblast apoptosis, Neuron 37 (2) (2003) 209–219.
[40] K.D. Economides, L. Zeltser, M.R. Capecchi, Hoxb13 mutations cause overgrowth of caudal spinal cordand tail vertebrae, Dev. Biol. 256 (2) (2003) 317–330.
[41] C. Wittmann, O. Bossinger, B. Goldstein, M. Fleischmann, R. Kohler, K. Brunschwig, H. Tobler, F. Muller, The expression of the C. elegans labial-like Hox gene ceh-13 during early embryogenesis relies on cell fate and on anteroposterior cell polarity, Development 124 (21) (1997) 4193–4200.
[42] J.I. Murray, E. Preston, J.P. Crawford, J.D. Rumley, P. Amom, B.D. Anderson, P. Sivaramakrishnan, S.D. Patel, B.A. Bennett, T.D. Lavon, E. Hsiao, F. Peng, A. L. Zacharias, The anterior Hox gene ceh-13 and elt-1/GATA activate the posterior Hox genes nob-1 and php-3 to specify posterior lineages in the C. elegans embryo, PLoS Genet. 18 (5) (2022) e1010187-e1010187.
[43] J.E. Sulston, E. Schierenberg, J.G. White, J.N. Thomson, The embryonic cell lineage of the nematode Caenorhabditis elegans, Dev. Biol. 100 (1) (1983) 64–119.
[44] D. Cowing, C. Kenyon, Correct Hox gene expression established independently of position in Caenorhabditis elegans, Nature 382 (6589) (1996) 353–356.
[45] Y. Arata, H. Kouike, Y. Zhang, M.A. Herman, H. Okano, H. Sawa, Wnt Signaling and a Hox Protein Cooperatively Regulate PSA-3/Meis to Determine Daughter Cell Fate after Asymmetric Cell Division in C. elegans, Developmental Cell 11(1) (2006) 105–115.
[46] J.A. Wagmaister, J.E. Gleason, D.M. Eisenmann, Transcriptional upregulation of the C. elegans Hox gene lin-39 during vulval cell fate specification, Mech. Dev. 123 (2) (2006) 135–150.
[47] N. Ji, Teije C. Middelkoop, Remco A. Mentink, Marco C. Betist, S. Tonegawa, D. Mooijman, Hendrik C. Korswagen, A. van Oudenaarden, Feedback control of gene expression variability in the Caenorhabditis elegans Wnt pathway, Cell 155 (4) (2013) 869–880.
[48] U. Nordström, E. Maier, T.M. Jessell, T. Edlund, An early role for wnt signaling in specifying neural patterns of Cdx and hox gene expression and motor neuron subtype identity, PLOS Biol. 4 (8) (2006), e252.
[49] A. Streit, R. Kohler, T. Marty, M. Belfiore, K. Takacs-Vellai, M.A. Vigano, R. Schnabel, M. Affolter, F. Müller, Conserved Regulation of the Caenorhabditis elegans labial/Hox1 Gene ceh-13, Dev. Biol. 242 (2) (2002) 96–108.
[50] J.E. Sulston, H.R. Horvitz, Post-embryonic cell lineages of the nematode, Caenorhabditis elegans, Dev. Biol. 56 (1) (1977) 110–156.
[51] C.J. Kenyon, J. Austin, M. Costa, D.W. Cowing, J.M. Harris, L. Honigberg, C. P. Hunter, J.N. Maloof, M.M. Muller-Immerglück, S.J. Salser, D.A. Waring, B. B. Wang, L.A. Wrischnik, The dance of the hox genes: patterning the anteroposterior body axis of Caenorhabditis elegans, Cold Spring Harb. Symp. Quant. Biol. 62 (1997) 293–305.
[52] C. Zheng, H.M.T. Lee, K. Pham, Nervous system-wide analysis of Hox regulation of terminal neuronal fate specification in Caenorhabditis elegans, PLOS Genet. 18 (2) (2022) e1010092-e1010092.
[53] S.J. Salser, C. Kenyon, Patterning C. elegans: homeotic cluster genes, cell fates and cell migrations, Trends Genet. 10 (5) (1994) 159–164.
[54] S.J. Salser, C.M. Loer, C. Kenyon, Multiple HOM-C gene interactions specify cell fates in the nematode central nervous system, Genes Dev. 7 (9) (1993) 1714–1724.
[55] S.G. Clark, A.D. Chisholm, H.R. Horvitz, Control of cell fates in the central body region of C. elegans by the homeobox gene lin-39, Cell 74 (1) (1993) 43–55.
[56] B.B. Wang, M.M. Müller-Immergluck, J. Austin, N.T. Robinson, A. Chisholm, C. Kenyon, A homeotic gene cluster patterns the anteroposterior body axis of C. elegans, Cell 74 (1) (1993) 29–42.
[57] H. Liu, T.J. Strauss, M.B. Potts, S. Cameron, Direct regulation of egl-1 and of programmed cell death by the Hox protein MAB-5 and by CEH-20, a C. elegans homolog of Pbx1, Development 133 (4) (2006) 641–650.
[58] T.R. Bürglin, Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals, Nucleic Acids Res. 25 (21) (1997) 4173–4180.
[59] M.B. Potts, D.P. Wang, S. Cameron, Trithorax, Hox, and TALE-class homeodomain proteins ensure cell survival through repression of the BH3-only gene egl-1, Dev. Biol. 329 (2) (2009) 374–385.
[60] R.S. Mann, K.M. Lelli, R. Joshi, Chapter 3 hox specificity: unique roles for cofactors and collaborators. Current Topics in Developmental Biology, Academic Press, 2009, pp. 63–101.
[61] A.K. Kalis, D.U. Kissiov, E.S. Kolenbrander, Z. Palchick, S. Raghavan, B. J. Tetreault, E. Williams, C.M. Loer, J.R. Wolff, Patterning of sexually dimorphic neurogenesis in the Caenorhabditis elegans ventral cord by Hox and TALE homeodomain transcription factors, Dev. Dyn. 243 (1) (2014) 159–171.
[62] C. Kenyon, A gene involved in the development of the posterior body region of C. elegans, Cell 46 (3) (1986) 477–487.
[63] M.L. Rivas, J.M. Espinosa-Vázquez, N. Sambrani, S. Greig, S. Merabet, Y. Graba, J. C.-G. Hombría, Antagonism versus cooperativity with TALE cofactors at the base of the functional diversification of Hox protein function, PLoS Genet. 9 (2) (2013), e1003252.
[64] W.-F. Shen, S. Rozenfeld, H.J. Lawrence, C. Largman, The Abd-B-like hox homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target *, J. Biol. Chem. 272 (13) (1997) 8198–8206.
[65] J.E. Sulston, J.G. White, Regulation and cell autonomy during postembryonic development of Caenorhabditis elegans, Dev. Biol. 78 (2) (1980) 577–597.
[66] L.I. Jiang, P.W. Sternberg, Interactions of EGF, Wnt and HOM-C genes specify the P12 neuroectoblast fate in C. elegans, Development 125 (12) (1998) 2337–2347.
[67] P. Kratsios, S.Y. Kerk, C. Catela, J. Liang, B. Vidal, E.A. Bayer, W. Feng, E.D. De La Cruz, L. Croci, G.G. Consalez, K. Mizumoto, O. Hobert, An intersectional gene regulatory strategy defines subclass diversity of C. elegans motor neurons, eLife 6 (2017) e25751-e25751.
[68] J.C.-G. Hombría, B. Lovegrove, Beyond homeosis—HOX function in morphogenesis and organogenesis, Differentiation 71 (8) (2003) 461–476.
[69] I. Miguel-Aliaga, S. Thor, Segment-specific prevention of pioneer neuron apoptosis by cell-autonomous, postmitotic Hox gene activity, Development 131 (24) (2004) 6093–6105.
[70] H. Becker, S. Renner, G.M. Technau, C. Berger, Cell-autonomous and non-cell-autonomous function of hox genes specify segmental neuroblast identity in the gnathal region of the embryonic CNS in drosophila, PLoS Genet. 12 (3) (2016), e1005961.
[71] C. Zheng, M. Diaz-Cuadros, M. Chalfie, Hox genes promote neuronal subtype diversification through posterior induction in Caenorhabditis elegans, Neuron 88 (3) (2015) 514–527.
[72] K.L. Chow, S.W. Emmons, HOM-C/Hox genes and four interacting loci determine the morphogenetic properties of single cells in the nematode male tail, Development 120 (9) (1994) 2579–2592.
[73] H.B. Ferreira, Y. Zhang, C. Zhao, S.W. Emmons, Patterning of Caenorhabditis elegans posterior structures by the abdominal-bhomolog,egl-5, Dev. Biol. 207 (1) (1999) 215–228.
[74] R. Lints, S.W. Emmons, Patterning of dopaminergic neurotransmitter identity among Caenorhabditis elegans ray sensory neurons by a TGFbeta family signaling pathway and a Hox gene, Development 126 (24) (1999) 5819–5831.
[75] K. Schinkmann, C. Li, Localization of FMRF amide-like peptides in Caenorhabditis elegans, J. Comp. Neurol. 316 (2) (1992) 251–260.
[76] J.E. Sulston, D.G. Albertson, J.N. Thomson, The Caenorhabditis elegans male: postembryonic development of nongonadal structures, Dev. Biol. 78 (2) (1980) 542–576.
[77] R. Lints, L. Jia, K. Kim, C. Li, S.W. Emmons, Axial patterning of C. elegans male sensilla identities by selector genes, Dev. Biol. 269 (1) (2004) 137–151.

[78] S.J. Salser, C. Kenyon, A C. elegans Hox gene switches on, off, on and off again to regulate proliferation, differentiation and morphogenesis, Development 122 (5) (1996) 1651–1661.
[79] H. Zhang, S.W. Emmons, The novel C. elegans gene sop-3 modulates Wnt signaling to regulate Hox gene expression, Development 128 (5) (2001) 767–777.
[80] H. Zhang, R.B.R. Azevedo, R. Lints, C. Doyle, Y. Teng, D. Haber, S.W. Emmons, Global regulation of Hox gene expression in C. elegans by a SAM domain protein, Dev. Cell 4 (6) (2003) 903–915.
[81] A. Garcia-Bellido, Genetic control of wing disc development in drosophila, Ciba Found. Symp. (1975) 161–182.
[82] Y. Narita, F.M. Rijli, Chapter 5 hox genes in neural patterning and circuit formation in the mouse hindbrain. Current Topics in Developmental Biology, Academic Press, 2009, pp. 139–167.
[83] M. Rossel, M.R. Capecchi, Mice mutant for both Hoxa1 and Hoxb1 show extensive remodeling of the hindbrain and defects in craniofacial development, Development 126 (22) (1999) 5027–5040.
[84] S.J. Salser, C. Kenyon, Activation of a C. elegans Antennapedia homologue in migrating cells controls their direction of migration, Nature 355 (6357) (1992) 255–258.
[85] J. Harris, L. Honigberg, N. Robinson, C. Kenyon, Neuronal cell migration in C. elegans: regulation of Hox gene expression and cell position, Development 122 (10) (1996) 3117–3131.
[86] J.N. Maloof, J. Whangbo, J.M. Harris, G.D. Jongeward, C. Kenyon, A. Wnt, signaling pathway controls hox gene expression and neuroblast migration in C. elegans, Development 126 (1) (1999) 37–49.
[87] J. Whangbo, C. Kenyon, A. Wnt, Signaling system that specifies two patterns of cell migration in C. elegans, Mol. Cell 4 (5) (1999) 851–858.
[88] H.C. Korswagen, M.A. Herman, H.C. Clevers, Distinct β-catenins mediate adhesion and signalling functions in C. elegans, Nature 406 (6795) (2000) 527–532.
[89] M. Herman, C. elegans POP-1/TCF functions in a canonical Wnt pathway that controls cell migration and in a noncanonical Wnt pathway that controls cell polarity, Development 128 (4) (2001) 581–590.
[90] H.C. Korswagen, D.Y.M. Coudreuse, M.C. Betist, S. van de Water, D. Zivkovic, H. C. Clevers, The Axin-like protein PRY-1 is a negative regulator of a canonical Wnt pathway in C. elegans, Genes Dev. 16 (10) (2002) 1291–1302.
[91] T. Walston, C. Guo, R. Proenca, M. Wu, M. Herman, J. Hardin, E. Hedgecock, mig-5/Dsh controls cell fate determination and cell migration in C. elegans, Dev. Biol. 298 (2) (2006) 485–497.
[92] T. Oosterveen, D.Y.M. Coudreuse, P.-T. Yang, E. Fraser, J. Bergsma, T.C. Dale, H. C. Korswagen, Two functionally distinct Axin-like proteins regulate canonical Wnt signaling in C. elegans, Dev. Biol. 308 (2) (2007) 438–448.
[93] M. Silhankova, H.C. Korswagen, Migration of neuronal cells along the anterior–posterior body axis of C. elegans: Wnts are in control, Curr. Opin. Genet. Dev. 17 (4) (2007) 320–325.
[94] D.Y.M. Coudreuse, G. Roël, M.C. Betist, O. Destrée, H.C. Korswagen, Wnt gradient formation requires retromer function in Wnt-producing cells, Science 312 (5775) (2006) 921–924.
[95] L. Yang, M. Sym, C. Kenyon, The roles of two C. elegans HOX co-factor orthologs in cell migration and vulva development, Dev. (Camb., Engl.) 132 (6) (2005) 1413–1428.
[96] M. Sym, N. Robinson, C. Kenyon, MIG-13 positions migrating cells along the anteroposterior body axis of C. elegans, Cell 98 (1) (1999) 25–36.
[97] S. Schneider, A. Gulacsi, M.E. Hatten, Lrp12/Mig13a reveals changing patterns of preplate neuronal polarity during corticogenesis that are absent in reeler mutant mice, Cereb. Cortex 21 (1) (2011) 134–144.
[98] C. Desai, G. Garriga, S.L. McLntire, H.R. Horvitz, A genetic pathway for the development of the Caenorhabditis elegans HSN motor neurons, Nature 336 (6200) (1988) 638–646.
[99] P.D. Baum, C. Guenther, C.A. Frank, B.V. Pham, G. Garriga, The Caenorhabditis elegans gene ham-2 links Hox patterning to migration of the HSN motor neuron, Genes Dev. 13 (4) (1999) 472–483.
[100] Y. Kee, B.J. Hwang, P.W. Sternberg, M. Bronner-Fraser, Evolutionary conservation of cell migration genes: from nematode neurons to vertebrate neural crest, Genes Dev. 21 (4) (2007) 391–396.
[101] M.B. Reilly, C. Cros, E. Varol, E. Yemini, O. Hobert, Unique homeobox codes delineate all the neuron classes of C. elegans, Nature 584 (7822) (2020) 595–601.
[102] W. Feng, Y. Li, P. Dao, J. Aburas, P. Islam, B. Elbaz, A. Kolarzyk, A.E. Brown, P. Kratsios, A terminal selector prevents a Hox transcriptional switch to safeguard motor neuron identity throughout life, Elife 9 (2020).
[103] J. Friedrich, S. Sorge, F. Bujupi, M.P. Eichenlaub, N.G. Schulz, J. Wittbrodt, I. Lohmann, Hox function is required for the development and maintenance of the drosophila feeding motor unit, Cell Rep. 14 (4) (2016) 850–860.
[104] R. Joshi, R. Sipani, A. Bakshi, Roles of drosophila hox genes in the assembly of neuromuscular networks and behavior, Front Cell Dev. Biol. 9 (2021), 786993.
[105] W. Feng, Y. Li, P. Kratsios, Emerging roles for hox proteins in the last steps of neuronal development in worms, flies, and mice, Front. Neurosci. 15 (801791) (2022).
[106] O. Hobert, Regulatory logic of neuronal diversity: terminal selector genes and selector motifs, Proc. Natl. Acad. Sci. USA 105 (51) (2008) 20067–20071.
[107] O. Hobert, P. Kratsios, Neuronal identity control by terminal selectors in worms, flies, and chordates, Curr. Opin. Neurobiol. 56 (2019) 97–105.
[108] A.K. Kalis, M.C. Sterrett, C. Armstrong, A. Ballmer, K. Burkstrand, E. Chilson, E. Emlen, E. Ferrer, S. Loeb, T. Olin, K. Tran, A. Wheeler, J. Ross Wolff, Hox proteins interact to pattern neuronal subtypes in Caenorhabditis elegans males, Genetics 220 (4) (2022).
[109] M.B. Reilly, T. Tekieli, C. Cros, G.R. Aguilar, J. Lao, I.A. Toker, B. Vidal, E. Leyva-Díaz, A. Bhattacharya, S.J. Cook, J.J. Smith, I. Kovacevic, B. Gulez, R. W. Fernandez, E.F. Bradford, Y.H. Ramadan, P. Kratsios, Z. Bao, O. Hobert, Widespread employment of conserved C. elegans homeobox genes in neuronal identity specification, PLOS Genet. 18 (9) (2022), e1010372.
[110] A. Chisholm, Control of cell fate in the tail region of C. elegans by the gene egl-5, Development 111 (4) (1991) 921–932.
[111] C. Zheng, M. Diaz-Cuadros, M. Chalfie, Hox genes promote neuronal subtype diversification through posterior induction in Caenorhabditis elegans, Neuron 88 (3) (2015) 514–527.
[112] A.S. Toker, Y. Teng, H.B. Ferreira, S.W. Emmons, M. Chalfie, The Caenorhabditis elegans spalt-like gene sem-4 restricts touch cell fate by repressing the selector Hox gene egl-5 and the effector gene mec-3, Development 130 (16) (2003) 3831–3840.
[113] P. Arlotta, O. Hobert, Homeotic transformations of neuronal cell identities, Trends Neurosci. 38 (12) (2015) 751–762.
[114] C. Zheng, F.Q. Jin, M. Chalfie, Hox proteins act as transcriptional guarantors to ensure terminal differentiation, Cell Rep. 13 (7) (2015) 1343–1352.
[115] J.C. Way, M. Chalfie, mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in C. elegans, Cell 54 (1) (1988) 5–16.
[116] A. Duggan, C. Ma, M. Chalfie, Regulation of touch receptor differentiation by the Caenorhabditis elegans mec-3 and unc-86 genes, Development 125 (20) (1998) 4107–4119.
[117] S.E. Von Stetina, M. Treinin, D.M. Miller 3rd, The motor circuit, Int Rev. Neurobiol. 69 (2006) 125–167.
[118] P. Kratsios, B. Pinan-Lucarre, S.Y. Kerk, A. Weinreb, J.L. Bessereau, O. Hobert, Transcriptional coordination of synaptogenesis and neurotransmitter signaling, Curr. Biol. 25 (10) (2015) 1282–1295.
[119] B. Prasad, O. Karakuzu, R.R. Reed, S. Cameron, unc-3-dependent repression of specific motor neuron fates in Caenorhabditis elegans, Dev. Biol. 323 (2) (2008) 207–215.
[120] B.C. Prasad, B. Ye, R. Zackhary, K. Schrader, G. Seydoux, R.R. Reed, unc-3, a gene required for axonal guidance in Caenorhabditis elegans, encodes a member of the O/E family of transcription factors, Development 125 (8) (1998) 1561–1568.
[121] C. Eastman, H.R. Horvitz, Y. Jin, Coordinated transcriptional regulation of the unc-25 glutamic acid decarboxylase and the unc-47 GABA vesicular transporter by the Caenorhabditis elegans UNC-30 homeodomain protein, J. Neurosci. 19 (15) (1999) 6225–6234.
[122] Y. Jin, R. Hoskins, H.R. Horvitz, Control of type-D GABAergic neuron differentiation by C. elegans UNC-30 homeodomain protein, Nature 372 (6508) (1994) 780–783.
[123] Y. Li, A. Osuma, E. Correa, M.A. Okebalama, P. Dao, O. Gaylord, J. Aburas, P. Islam, A.E. Brown, P. Kratsios, Establishment and maintenance of motor neuron identity via temporal modularity in terminal selector function, Elife 9 (2020).
[124] W. Feng, H. Destain, J.J. Smith, P. Kratsios, Maintenance of neurotransmitter identity by Hox proteins through a homeostatic mechanism, Nature, Communications 13 (1) (2022) 6097.
[125] C.R.A.A. Raouf Issa, A novel post-developmental role of the Hox genes underlies normal adult behaviour, bioRxiv (2021).
[126] J. Deutsch, H. Le Guyader, The neuronal zootype. An hypothesis, Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie 321(9) (1998) 713–719.
[127] J.W. Wysocka-Diller, G.O. Aisemberg, M. Baumgarten, M. Levine, E.R. Macagno, Characterization of a homologue of bithorax-complex genes in the leech Hirudo medicinalis, Nature 341 (6244) (1989) 760–763.
[128] V. Wong, G. Aisemberg, W. Gan, E. Macagno, The leech homeobox gene Lox4 may determine segmental differentiation of identified neurons, J. Neurosci. 15 (8) (1995) 5551–5559.
[129] P.W. Holland, L.Z. Holland, N.A. Williams, N.D. Holland, An amphioxus homeobox gene: sequence conservation, spatial expression during development and insights into vertebrate evolution, Development 116 (3) (1992) 653–661.
[130] J.S. Deutsch, Hox genes: studies from the 20th to the 21st century. Preface, Adv. Exp. Med Biol. 689 (2010) ix–xi.

Seminars in Cell and Developmental Biology 152-153 (2024) 70–75

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Hox dosage and morphological diversification during development and evolution

Samir Merabet *, Julie Carnesecchi

IGFL, UMR5242, CNRS, ENS, University Lyon I, 32/34 Av. Tony Garnier, 69007 Lyon, France

ARTICLE INFO

Keywords:
Hox
TALE
Dose
Transcription
Morphogenesis

ABSTRACT

Hox genes encode for evolutionary conserved transcription factors that have long fascinated biologists since the observation of the first homeotic transformations in flies. Hox genes are developmental architects that instruct the formation of various and precise morphologies along the body axes in cnidarian and bilaterian species. In contrast to these highly specific developmental functions, Hox genes encode for proteins that display poorly selective DNA-binding properties *in vitro*. This "Hox paradox" has been partially solved with the discovery of the TALE-class cofactors, which interact with all Hox members and form versatile Hox/TALE protein complexes on DNA. Here, we describe the role of the Hox dosage as an additional molecular strategy contributing to further resolve the Hox paradox. We present several cases where the Hox dosage is involved in the formation of different morphologies in invertebrates and vertebrates, with a particular emphasis on flight appendages in insects. We also discuss how the Hox dosage could be interpreted in different types of target enhancers within the nuclear environment *in vivo*. Altogether our survey underlines the Hox dosage as a key mechanism for shaping Hox molecular function during development and evolution.

1. Introduction

Hox genes are evolutionarily conserved regulators of morphological diversity in animals. They encode for homeodomain (HD)-containing transcription factors (TFs) which act by recognizing specific enhancers to regulate the expression of large numbers of downstream target genes [1,2]. Hox proteins have also recently been shown to control tissue-specific gene expression at the mRNA splicing regulatory level, enlarging their molecular repertoire for the fine tuning of gene expression *in vivo* [3,4].

A general paradox concerning the molecular mode of action of Hox proteins lies in the fact that their specific transcriptional programs *in vivo* contrast with their ability to recognize highly similar DNA-binding sites as monomers *in vitro*. The discovery of the generic TALE-class cofactors greatly contributed to better understand this *in vivo/in vitro* paradox. Two types of TALE cofactors interact with Hox proteins: the PBC and MEIS proteins. These proteins are highly conserved during evolution, with the presence of one (such as Extradenticle (Exd) or Homothorax (Hth) in *Drosophila*) or more (such as PBX1–4 or MEIS1–2 in human) representatives. PBC and MEIS cofactors interact on DNA with all Hox members and modulate both their DNA-binding properties and trans-regulatory activities [5,6]. In most cases, PBC/MEIS form trimeric complexes with the Hox protein, and the assembly of Hox/PBC/MEIS complexes has been described to rely on diverse Hox protein motifs in several instances [6]. This versatility is conserved from cnidarians to bilaterians and has been proposed to serve as a molecular scaffold for diversifying and specifying Hox patterning functions along longitudinal axes during animal evolution [7]. Beyond the TALE cofactors, many other types of TFs have been described to interact with Hox proteins [8]. Although it is expected that several of them could be Hox-specific, *in vivo* analyses showed that *Drosophila* TFs had a tendency to interact with two or more different Hox proteins [9]. The overall combination of interactions was however different for each *Drosophila* Hox protein, suggesting that Hox-specific transcriptional activity could result from the assembly of Hox-specific interactomes and not from individual Hox-specific cofactors [9].

Given their generic role as Hox cofactors, Hox-TALE-DNA interactions have been the subject of a number of studies. In particular, high throughput Selex-seq (Systematic Evolution of Ligands by Exponential Enrichment with massively parallel sequencing) and structural biochemistry showed that the TALE cofactors could help in revealing a "latent-specificity" in Hox DNA-binding properties [10]. This property

* Corresponding author.
E-mail address: samir.merabet@ens-lyon.fr (S. Merabet).

https://doi.org/10.1016/j.semcdb.2022.11.009
Received 14 July 2022; Received in revised form 15 October 2022; Accepted 30 November 2022
Available online 5 December 2022

was associated with the recognition of specific DNA shapes and binding to divergent nucleotide sequences that were not necessarily of high affinity [10,11]. These non-consensus or so-called low affinity DNA-binding sites were also found to be distinctly enriched in genome-wide binding profiles of Hox proteins from different tissues [10]. In addition, their presence in several target enhancers (see below) reinforces their key contribution for Hox functional specificity *in vivo*. Importantly, the existence of low-affinity DNA-binding sites raised the question of the impact of the Hox expression level, or Hox dosage, in their selective recognition genome wide.

Surprisingly, the role of the Hox dosage on Hox function remains a poorly investigated issue. Here, we present studies which exemplify the importance of the Hox dosage for controlling specific morphologies during development. In particular, we describe the role of the Hox dosage in the context of flight appendage formation during development and evolution in insects. We also discuss and speculate on the transcriptional readout of Hox dosage when considering the nuclear and chromatin environment *in vivo*.

1.1. Hox dosage in animal development

The first study describing the importance of the Hox dosage during development is the pioneer genetic analysis of Ed Lewis on *Bithorax-Complex* (*BX-C*) mutant phenotypes in the fruit fly *Drosophila melanogaster* [12]. This work showed that mutations affecting the expression level of the Hox gene *Ultrabithorax* (*Ubx)*, as well as other so-called "BX-C substances") were responsible of specific phenotypes in the *Drosophila* larva or adult [12].

Hox mutant phenotypes that were dose-dependent were later observed in vertebrates. One of these studies related to the analysis of various mutant combinations for the group 4 of HOX genes in mice [13]. More particularly, the authors compared the skeletal phenotype of single, *versus* double or triple mutant combinations for *Hoxa4*, *Hoxb4* and *Hoxd4*. They observed a clear dosage-dependent phenotype in the double and triple mutants, explaining the functional redundancy between the paralogs [13]. Previous analysis of other Hox mutant mice did not reveal this aspect of redundancy and dosage-dependent phenotype [14,15]. Given that *Hoxa4*, *b4* and *d4* encode for Hox proteins with highly similar HDs, it was suggested that this dosage dependency could reflect similar transcriptional activities on common downstream target genes [13].

Another study described dose-dependent effects of the Hox mode of action for controlling the number and size of digits. This effect implied the posterior genes *Hoxd11*, *Hoxd12*, *Hoxd13* and *Hoxa13* as major determinants of digit morphogenesis [16]. Importantly, progressive decrease of dosage with those Hox genes induced an increased severity in digit size and number defects in the mouse, highlighting a common Hox-dose dependent mechanism for controlling the size and number of digits. A dosage-dependent phenotype with the same four Hox genes was also described for external genitalia formation [17]. Based on developmental and phylogenetic arguments, the authors proposed an attractive model for the dose-dependent involvement of posterior Hox genes of the clusters A and D in diversifying more largely the length and number of digits during vertebrate evolution [16].

Interestingly, Hox dosage is known to be instrumental for different aspects of leg morphogenesis in insects. One case-model is the role of the Hox gene *Ubx* in repressing the formation of non-sensory microtrichiae, or trichomes, on the posterior femur of the second and third legs in different *Drosophila* species [18]. The pattern of trichomes distribution was not identical between *Drosophila* species that display different Ubx expression patterns. In addition, this pattern was shown to be dose-dependent and to require high levels of *Ubx* expression for efficient repression [18]. Interestingly, *Ubx* has also been described to modulate leg length depending on its expression level in the water strider *Limnoporus dissortis* [19]. This species is characterized by longer legs on the second thoracic segment (T2) than on the third thoracic segment (T3). This morphological difference was shown to depend on a low *Ubx* expression level in the T2 leg, which promotes growth, while a high *Ubx* expression level in the T3 leg has an opposite repressive role on growth. Decreasing *Ubx* expression levels shortened or lengthened the T2-leg or T3-leg, respectively. These observations underline that different tissues respond distinctly to Hox dose variations depending on their respective endogenous Hox expression level.

1.2. Hox dosage and flight appendage morphogenesis in insects during development and evolution

Flying insects cover thousands of different species which have been classified in several major orders [20]. The ancestral state in insects consisted of two pairs of similar wings on their second (T2 forewing, FW) and third (T3 hindwing, HW) thoracic segments [20,21]. Most existing lineages, such as the Odonata, Hymenoptera or Hemiptera, retain the ancestral state. Still, flight appendages diverged over the course of evolution, not only between species, but also between the T2 and T3 flight appendages of the same species. These modifications applied to the shape, size, and/or color pattern. For example, the FW and HW can be of different shapes and color patterns in butterflies (Lepidoptera order), while it is mostly the size that distinguishes the two pairs of wings in bees (Hymenoptera order). In more extreme cases, the wing evolved into a different flight organ, as exemplified with the FW transformed into a thick protective envelop called elytron in the Coleoptera, or the HW transformed into a small dumbbell-shaped organ called haltere in the Diptera (Fig. 1).

For a long time, the formation of FW in insects has been considered as a Hox-independent process, while HW formation and diversification was shown to result from the activity of *Ubx* in the third thoracic segment [21]. These distinct Hox contributions on the T2 or T3 segment were deduced from early work showing the absence of any obvious expression and function of the Hox gene *Antennapedia* (*Antp*) in the FW primordia of the beetle *Tribolium castaneum* [22] and the fruit fly *Drosophila melanogaster* [23]. Of note, the absence of any contribution of *Antp* for wing formation was not completely evident in *Drosophila*, since *Antp* mutant clones were described to generate subtle wing phenotypes in pioneer genetic analyses [24,25].

In addition, the HW was transformed into an elytron upon the loss of *Ubx* in the T3 segment in the beetle, and the same type of haltere-to-wing transformation was observed when affecting *Ubx* expression in the T3 segment in the fruit fly [22,23]. Ubx was also shown to be expressed in the wing primordia in two different ant species [26], the silkmoth *Bombyx mori* [27], the butterflies *Bicyclus anynana* [28,29] and *Junonia coenia* [28] and the honeybee *Apis mellifera* [27], raising the question of how a highly conserved Hox protein could have diversified its activity to trigger the formation of various flight appendages during insect evolution.

The advent of new genetic tools and more sensitive antibodies demonstrated that *Ubx* was in fact not the only Hox gene involved in flight appendage formation in insects. More particularly, *Antp* was shown to be required for proper elytron and HW formation in *Tribolium* [30], and to be expressed and required for the formation of the FW and HW in *Bombyx* [30]. In addition, analysis in *Drosophila* revealed the contribution of *Antp* for proper FW formation [31]. Interestingly, Antp is dynamically expressed at a low level in the region of the wing primordium that gives rise to the distal wing in the *Drosophila* adult [31]. Surprisingly, *Antp* can replace *Ubx* and rescue haltere formation in a *Ubx* mutant background when expressed at high *Ubx*-like doses [31]. Conversely, decreasing the Ubx dose in the haltere into an Antp-like dose led to haltere-to-wing transformation [31]. These results underlined that the dose, instead of the nature of the Hox protein, is decisive for making a wing or a haltere in *Drosophila*.

Observations in other insect species further illustrated a striking correlation between the expression level of both Antp and Ubx, and the morphological similarities or differences between the FW and HW in

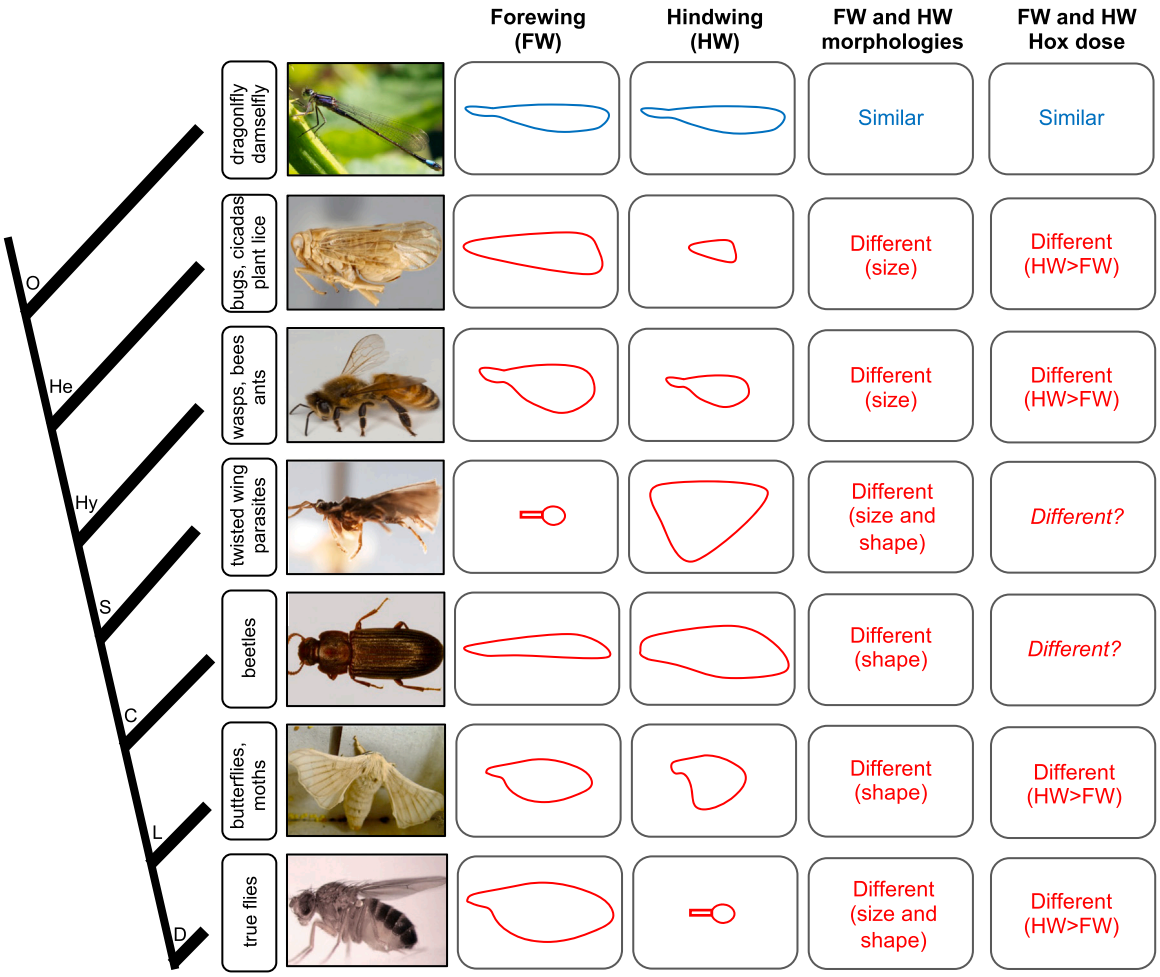


Fig. 1. Hox dosage and the diversification of flight appendages in insects: correlations and speculations. A simplified evolutionary tree of insects is shown with a representative species for each order. Pictures are not at scale and were obtained from WordPress (https://wordpress.org/). Cartoons schematize the forewing (FW) and hindwing (HW) morphology, and the Hox dose includes both Antp and Ubx expression levels in the FW and HW primordium of each corresponding species. This Hox level is not known in the male Strepsiptera (S) representative *Xenos vesparum* or in the Coleoptera (C) representative *Tribolium castenum*. Other branch orders and representative species are: Odonata (O; damselfly *Ischnura elegans*); Hemiptera (He; planthopper *Nilaparvata lugens*); Hymenoptera (Hy; honeybee *Apis melifera*); Lepidoptera (L; silk moth *Bombyx mori*); Diptera (D; fruit fly *Drosophila melanogaster*).

four wings insect species. For example, the damselfly *Ishnura elegans*, which has almost identical FW and HW, displays similar levels of Antp and Ubx in the FW and HW primordia [31]. In contrast, higher levels of Antp and Ubx were measured in the HW primordia of *Apis*, which gives rise to a smaller wing than the FW in the adult [31]. A differential Hox expression level was also observed between the FW and HW primordia in *Bombyx*, which has FW and HW of different shapes [31].

The role of the Hox dose has also been linked to the nutritional status in the rice planthopper insect *Nilaparvata lugens* (*Nl*), which expresses *Ubx* in both the FW and HW primordia [32,33]. Importantly, the level of *Ubx* expression is impacted by the quality of the diet: high quality induced its expression, which led to the short wing form, while low-quality ripe plants had the opposite effects with a reduction of its expression and the formation of long-wing form.

Altogether, these observations suggest that the Hox dosage (here considering Antp and Ubx) could be widely used for shaping flight appendages in insects. In particular, there is a recurrent observation that higher doses could trigger size reduction. Extreme cases are the flies, with the transformation of HW into small balancing organs. Whether the transformation of FW into elytra in the beetle could also result from a particular Hox expression level remains to be determined. Moreover, Hox dosage could explain the inverted pattern of the haltere and wing in the free-living male of the endoparasitic Strepsiptera insects when compared to the flies [34]. Instead of deploying ectopic expression of *Ubx*, as previously suggested [35], the modulation of the Hox dosage, with high Antp and low Ubx, could be sufficient to induce the formation of halteres or wings in the T2 and T3 segment, respectively (Fig. 1). This phenotype could be artificially reproduced in *Drosophila*, demonstrating that Hox dose variation in specific places of the body is compatible with the overall development of the animal.

Classically, the functional outcome of Hox gene activity was considered as resulting from specific spatial and temporal expression profiles, thus underestimating the key role of protein dosage. Now, the importance of protein dosage has been underestimated. Altogether, the presented studies showed that the Hox dosage is a molecular strategy promoting a diversity of morphological variations, from subtle wing size and/or shape modification to the formation of a completely new organ. In this context, we propose a speculative model based on Antp and Ubx expression levels to explain flight appendage diversification in insects (Fig. 1).

1.3. Molecular aspects of the Hox dosage at the transcriptional level

The *Drosophila* haltere disc can be divided into three main regions along the proximal-distal axis (Fig. 2). These regions give rise to different structures in the adult haltere: (i) the capitellum, which

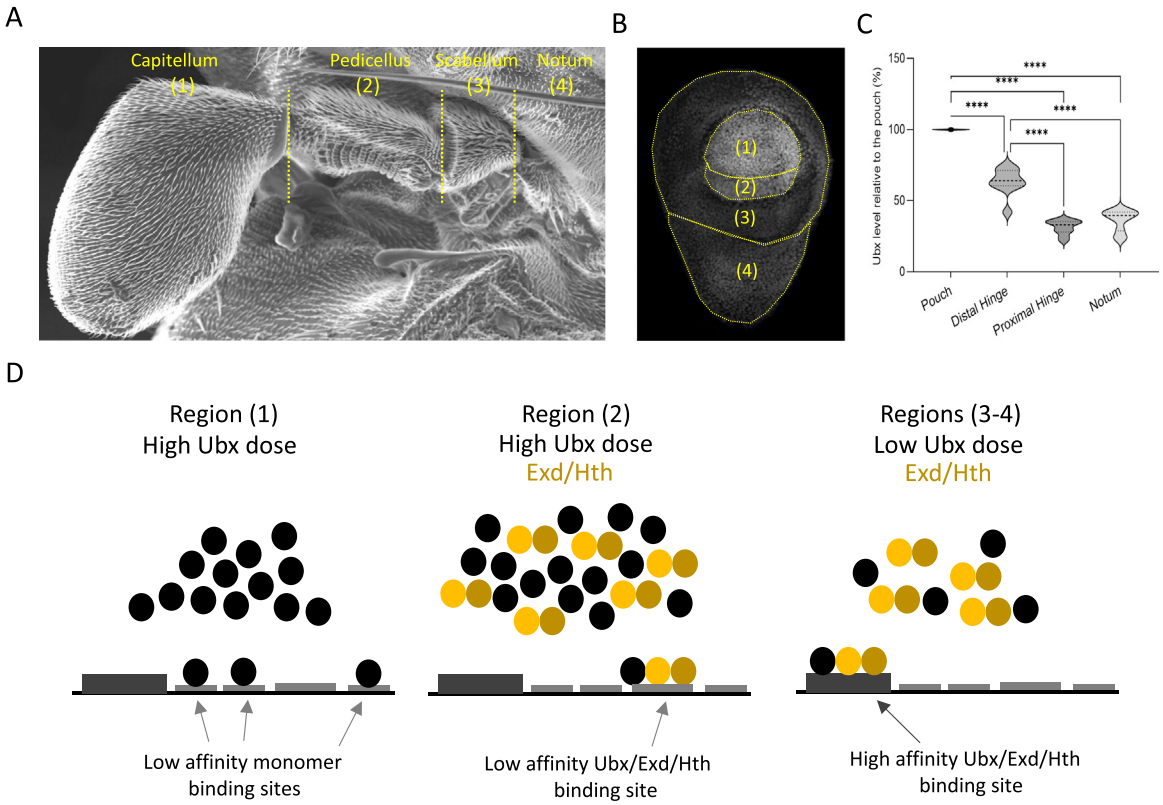


Fig. 2. Ubx dosage and the specification of different structures in the *Drosophila* haltere. A. SEM acquisition of adult wild type haltere with its four main regions (from distal to proximal): the capitellum (1), the pedicellus (2), the scabellum (3) and the notum (4). B. Haltere imaginal disc stained for Ubx (gray). The Ubx expression level if high in the pouch (1), which does not express the Exd and Hth cofactors, and which gives rise to the capitellum in the adult haltere. The Ubx expression level is also high in the distal hinge region (2), although to a less extent than in the pouch. This region expresses Exd and Hth and gives rise to the pedicellus in the adult haltere. The Ubx expression level is low in the proximal hinge (3) and notum (4), which give rise to the scabellum and notum in the adult haltere. These regions express Exd and Hth. C. Quantification of the Ubx expression level in the different regions of the haltere imaginal disc. The quantification is represented as a relative percentage of the level measured in the pouch in each individual disc (100%). One-way Anova test shows the statistical significance between the different regions (**** pvalue≤0,0001). D. Speculative model on the Ubx dosage and the recognition of different types of DNA-binding sites in the genome. High levels allow the recognition of low affinity binding sites, without (region 1) or with (region 2) the Exd and Hth cofactors. Low levels of Ubx with Exd and Hth restrict the recognition mode to highly-affinity/consensus DNA-binding sites (regions 3–4). Compared to high affinity binding sites, low affinity binding sites are less enriched in high throughput *in vitro* DNA-binding experiments and display divergent nucleotide sequences [10]. See also the main text.

originates from the most distal region called the pouch, (ii) the pedicillus and scabellum, which respectively originate from the distal and proximal hinge region, and (iii) the notum, which originates from the most proximal region of the haltere disc (Fig. 2). Interestingly, these different regions express different levels of Ubx, with a high-to-low gradient from the distal to the most proximal part of the disc (Fig. 2). Thus, there is a striking correlation between the fate of the different regions of the haltere disc and the dose of Ubx.

In addition to a specific pattern of Ubx doses, the haltere disc is characterized by a specific expression pattern for the Exd and Hth co-factors [36]. In particular, Exd and Hth cofactors are absent in the distal region (pouch) which expresses the highest Ubx level (see below). Notably, the activity of an autoregulatory module of *Ubx* was shown to be specifically controlled by low levels of Ubx with Exd and Hth [36], and this regulation occurred through low-affinity Ubx-Exd binding sites (see also below). These results demonstrated that the regulation of a specific target cis-regulatory module associates with a particular Ubx dosage.

Previous work identified the bulk transcriptome and genome-wide binding occupancy of Ubx in the whole haltere disc, however impairing the making of a direct correlation with specific Ubx doses genome-wide [37–39]. A recent study tackled this issue by considering more specifically regions expressing or not the Exd and Hth cofactors [40]. The analysis was also performed by doing a time-sensitive knockdown of Ubx in the distal domain and the results were systematically compared to the wing disc, which does not express Ubx. This work revealed that haltere-specific chromatin accessibility was different depending on the presence or not of the Exd/Hth cofactors: transcriptional repression and reduction of chromatin accessibility was associated with Hox monomer binding while the presence of Exd/Hth correlated with transcriptional activation and increase of chromatin accessibility [40]. These different effects were also observed in regions expressing different doses of Ubx (Fig. 2). In particular, the region expressing high level of Ubx with Exd/Hth was associated with atypical/low-affinity DNA-binding sites (as defined from SELEX-seq assays [10]), whereas the proximal regions expressing lower levels of Ubx were associated with consensus/high-affinity Hox/Exd DNA-binding sites [40]. Based on these observations, we propose a speculative model associated with the Ubx dose: high Ubx dose in the absence (in the pouch) or presence (in the distal hinge) of Exd/Hth allows for the recognition of low-affinity binding sites. In contrast, low Ubx dose with Exd/Hth (in the proximal hinge and notum) would more likely recognize high affinity DNA-binding sites (Fig. 2). Thus, the Ubx dose could be tightly linked to the DNA-binding site recognition mode and chromatin accessibility in the haltere disc.

Noteworthy, low-affinity binding sites have been described for Ubx in several enhancers of the target gene *shavenbaby* (*svb*) [11]. This gene is repressed by Ubx in the abdominal segments of the *Drosophila* embryo and this repression was found to rely on different redundant enhancers which contained several low-affinity DNA-binding sites for Ubx/TALE

complexes. Changing these DNA-binding sites into consensus/high-affinity DNA-binding sites led to ectopic activation of *svb* by different Hox proteins in the *Drosophila* embryo, showing that low-affinity DNA-binding sites are pivotal for conferring Hox specificity *in vivo* [11]. In addition, multiple low-affinity DNA-binding sites are important for stabilizing enhancer activation against Ubx dose variation or environmental stress, demonstrating a role in transcriptional robustness [11]. Interestingly, the different enhancers of *svb* are regrouped into specific microenvironments in the nucleus, allowing the formation of hubs that are enriched for Ubx and Hth [41]. These observations highlighted that Hox dosage could be locally controlled at specific loci in the nucleus, with the formation of local spots of concentration ensuring robust and active transcription with low-affinity enhancers [42]. This mechanism probably involves the local recruitment of other TFs and chromatin modifiers. Along this line, a similar mechanism has been described for the TF Bicoïd (Bcd), which is distributed in nuclear hubs in the posterior region of the embryo (characterized by a globally low concentration level; [43]). These local enrichments of Bcd molecules are dependent on the pioneer TF Zelda that is also distributed in nuclear hubs of high concentration (Zld, [43–45]). These nuclear hubs were proposed to reduce the time of Bcd DNA-binding site occupancy on target enhancers. More generally, the formation of micro-environments illustrates the importance of the nuclear architecture for controlling TFs and chromatin distribution. Along the same line, the nuclear architecture was recently described to increase in heterogeneity as embryonic development progresses, a phenomenon called “nuclear morphogenesis” which was interpreted as reflecting an increase in the complexity of gene regulation during development [46]. Considering the abovementioned dose-dependent role of Ubx on chromatin accessibility in the haltere disc, we suggest that Hox dosage could also impact on enhancer localization for proper regulation with other TFs in specific domains of the nucleus. This hypothesis could be tested by modifying artificially the level of Ubx in the proximal (with increasing doses) or distal (with decreasing doses) region of the haltere disc and assess whether these modifications could mimic the nuclear pattern of hubs normally observed in the distal or proximal region, respectively.

In addition, nuclear hubs could also serve as reservoirs of TFs. These reservoirs could be used to buffer strong Hox dose variations (up or down) under stress condition, allowing the maintenance of specific Hox transcriptional programs. Finally, nuclear hubs could also serve as a point source for the diffusion of TFs towards distant promoters. In this scenario, it will be interesting to analyze whether local nuclear hubs are enriched in enzymes like acetylases, which have been proposed to promote the diffusion of TFs from distant enhancers to the target promoter [47].

2. Conclusion

The Hox paradox has long been investigated by studying the DNA-binding and protein-protein interaction properties of Hox proteins with their generic TALE cofactors without considering the role of the dose [48]. In this review, we presented several lines of evidence which clearly demonstrate the importance of the Hox dosage, both at the macroscopic/phenotypic and molecular level. We proposed speculative modes of action linking the Hox dosage and the TALE partnership with the recognition of different types of DNA-binding sites and distinct impacts on chromatin accessibility. We also emphasized how the Hox dosage could be micro-shaped by the nuclear architecture to form local hubs with specific enhancers. Our model is based on the existence of low-affinity DNA-binding sites. It is important to stress out that other molecular mechanisms could exist depending on the Hox protein, tissue-type and/or the presence of non-TALE cofactors. For exemple, Deformed (Dfd) and Sex combs reduced (Scr) have recently been shown to recognize non-canonical Hox/Exd binding sites that were of high affinity [49,50]. These sites are still able to respond to different Hox levels [50]. It has been proposed that variations in DNA-binding affinity could be linked to cell-type complexity of the enhancer expression pattern [49].

In the future, the role of the Hox dose will certainly be investigated in a variety of model systems. For example, it will be interesting to know whether Hox dosage variation and local nuclear hubs could be a general principle associated with Hox function across the animal phyla (from Cnidaria to Bilateria). Along the same line, the Hox dosage could also potentially impact on their numerous molecular activities other than transcriptional regulation, such as genome repair [51,52], DNA replication [53,54], RNA splicing [4] or translation [55]. An important challenge will be to understand the impact of the Hox dose at both large-scale and single molecule resolution levels. The advent of MERFISH technology coupled to single cell RNA-seq [56] could provide an unprecedented resolution of the impact of the Hox dose on hundreds of target genes in the same cell. In addition, approaches like single particle tracking PhotoActivated Light Microscopy (sptPALM, [57]) could offer a precise measurement of both the enrichment and the dynamics of individual Hox or Hox-TALE interactions *in vivo*. Deciphering the Hox dosage at these resolution scales will strongly contribute to our understanding of the mechanistic of Hox function and certainly reveal novel and unexpected molecular facets of Hox proteins.

Conflict of interest

Authors declare no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgments

We thank Abderrahman Khila and René Rezsohazy for helpful comments on the manuscript. Work in the laboratory of S.M. is supported by the Center National de la recherche Scientifique (CNRS), Ecole Normale Supérieure (ENS) de Lyon, Région Auvergne-Rhône-Alpes, Institut National du Cancer (InCA) and European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement (RNA-NetHOX, 101024467) for JC.

References

[1] Lohmann, I.; Mcginnis, W. Hox genes: it ’ s all a matter of context. 2002, 12, 514–516.
[2] S.D. Hueber, I. Lohmann, Shaping segments: hox gene function in the genomic age, BioEssays 30 (2008) 965–979.
[3] J. Carnesecchi, G. Sigismondo, K. Domsch, C.E.P. Baader, M.R. Rafiee, J. Krijgsveld, I. Lohmann, Multi-level and lineage-specific interactomes of the Hox transcription factor Ubx contribute to its functional specificity, Nat. Commun. 11 (2020), https://doi.org/10.1038/s41467-020-15223-x.
[4] J. Carnesecchi, P. Boumpas, Y. van Nierop, P. Sanchez, K. Domsch, H.D. Pinto, L. I. Borges Pinto P, The hox transcription factor ultrabithorax binds RNA and regulates co-transcriptional splicing through an interplay with RNA polymerase II, Nucleic Acids Res. 50 (2022) 763–783, https://doi.org/10.1093/nar/gkab1250.
[5] R.S. Mann, K.M. Lelli, R. Joshi, Chapter 3 Hox specificity. unique roles for cofactors and collaborators, Curr. Top. Dev. Biol. 88 (2009) 63–101.
[6] S. Merabet, R.S. Mann, To be specific or not: the critical relationship between hox And TALE proteins, Trends Genet. 32 (2016) 334–347.
[7] B. Hudry, M. Thomas-Chollier, Y. Volovik, M. Duffraisse, A. Dard, D. Frank, U. Technau, S. Merabet, Molecular insights into the origin of the Hox-TALE patterning system, eLife 3 (2014), e01939, https://doi.org/10.7554/eLife.01939.
[8] R. Rezsohazy, A.J. Saurin, C. Maurel-Zaffran, Y. Graba, Cellular and molecular insights into Hox protein action, Development 142 (2015) 1212–1227, https://doi.org/10.1242/dev.109785.
[9] M. Baeëza, S. Viala, M. Heim, A. Dard, B. Hudry, M. Duffraisse, A. Rogulja-Ortmann, C. Brun, S. Merabet, Inhibitory activities of short linear motifs underlie hox interactome specificity in vivo, eLife (2015) 2015, https://doi.org/10.7554/eLife.06034.001.
[10] M. Slattery, T. Riley, P. Liu, N. Abe, P. Gomez-alcala, I. Dror, T. Zhou, R. Rohs, B. Honig, H.J. Bussemaker, R.S. Mann, Cofactor binding evokes latent differences in DNA binding specificity between hox proteins, Cell 147 (2011) 1270–1282, https://doi.org/10.1016/j.cell.2011.10.053.

[11] J. Crocker, N. Abe, L. Rinaldi, A.P. McGregor, N. Frankel, S. Wang, A. Alsawadi, P. Valenti, S. Plaza, F. Payre, R.S. Mann, D.L. Stern, Low affinity binding site clusters confer hox specificity and regulatory robustness, Cell (2015) 191–203, https://doi.org/10.1016/j.cell.2014.11.041.
[12] E.B. Lewis, A gene complex controlling segmentation in drosophila, Nature 276 (1978) 565–570, https://doi.org/10.1038/276565a0.
[13] Gerald S.B. Horan, Ramiro Ramirez-Solis, Mark S. Featherstone, Debra J. Wolgemuth, Allan Bradley, Richard R. Behringer, Compound mutants for the paralogous hoxa-4, hoxb-4, and hoxd-4 genes show more complete homeotic transformations and a dose-dependent increase in the number of vertebrae transformed, Genes Dev. 9 (1995) 1667–1677.
[14] B.G. Condie, M.R. Capecch, Mice homozygous for a targeted disruption of Hoxd-3 (Hox-4.1) exhibit anterior transformations of the first and second cervical vertebrae, the atlas and axis, Development 119 (1993) 579–595.
[15] D. Rancourt, T. Tsuzuki, M. Capecchi, Genetic interaction between hoxb-5 and hoxb-6 is revealed by nonallelic noncomplementation, Genes Dev. 9 (1995) 108–122.
[16] J. Zákány, C. Fromental-Ramain, X. Warot, D. Duboule, Regulation of number and size of digits by posterior Hox genes: a dose-dependent mechanism with potential evolutionary implications, Proc. Natl. Acad. Sci. USA 94 (1997) 13695–13700, https://doi.org/10.1073/pnas.94.25.13695.
[17] F. van der Hoeven, J. Zakany, D. Duboule, Gene transpositions in the HoxD complex reveal a hierarchy of regulatory controls, Cell 85 (1996) 1025–1035.
[18] D.L. Stern, A role of Ultrabithorax in morphological differences between Drosophila species, Nature 396 (1998) 463–466. *Dec3*.
[19] P.N. Refki, D. Armisén, A.J.J. Crumière, S. Viala, A. Khila, Emergence of tissue sensitivity to Hox protein levels underlies the evolution of an adaptive morphological trait, Dev. Biol. 392 (2014) 441–453, https://doi.org/10.1016/j.ydbio.2014.05.021.
[20] B. Misof, S. Liu, K. Meusemann, R.S. Peters, A. Donath, C. Mayer, P.B. Frandsen, J. Ware, T. Flouri, R.G. Beutel, et al., Phylogenomics resolves the timing and pattern of insect evolution, Science 346 (2014) 763–767, https://doi.org/10.1126/science.1257570.
[21] Y. Tomoyasu, Ultrabithorax and the evolution of insect forewing/hindwing differentiation, Curr. Opin. Insect Sci. 19 (2017) 8–15.
[22] Y. Tomoyasu, S.R. Wheeler, R.E. Denell, Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum, Nature 433 (2005) 643–647, https://doi.org/10.1038/nature03272.
[23] S.B. Carroll, S.D. Weatherbee, J.A. Langeland, Homeotic genes and the regulation and evolution of insect wing number, Nature 375 (1995) 58–61, https://doi.org/10.1038/375058a0.
[24] G. Struhl, A homoeotic mutation transforming leg to antenna in Drosophila, Nature 292 (1981) 635–638, https://doi.org/10.1038/292635a0.
[25] G. Struhl, Genes controlling segmental specification in the Drosophila thorax, Proc. Natl. Acad. Sci. USA 79 (1982) 7380–7384, https://doi.org/10.1073/pnas.79.23.7380.
[26] E. Abouheif, G.A. Wray, Evolution of the gene network underlying wing polyphenism in ants, Science 297 (2002) 249–252.
[27] N. Prasad, S. Tarikere, D. Khanale, S.L. Habib F, A comparative genomic analysis of targets of Hox protein Ultrabithorax amongst distant insect species, Sci. Rep. (2016) 27885.
[28] X. Tong, S. Hrycaj, O. Podlaha, A. Popadic, A. Monteiro, Over-expression of Ultrabithorax alters embryonic body plan and wing patterns in the butter fl y Bicyclus anynana, Dev. Biol. 394 (2014) 357–366, https://doi.org/10.1016/j.ydbio.2014.08.020.
[29] S.V. Saenko, M.S.P. Marialva, P. Beldade, Involvement of the conserved Hox gene Antennapedia in the development and evolution of a novel trait, Evodevo 2 (2011), https://doi.org/10.1186/2041-9139-2-9.
[30] C. Fang, Y. Xin, T. Sun, A. Monteiro, Z. Ye, F. Dai, C. Lu, X. Tong, The Hox gene Antennapedia is essential for wing development in insects, Development 149 (2022), https://doi.org/10.1242/dev.199841.
[31] R. Paul, G. Giraud, K. Domsch, M. Duffraisse, F. Marmigère, S. Khan, S. Vanderperre, I. Lohmann, R. Stoks, L.S. Shashidhara, S. Merabet, Hox dosage contributes to flight appendage morphology in Drosophila, Nat. Commun. 12 (2021) 2892.
[32] F. Liu, X. Li, M. Zhao, M. Guo, K. Han, X. Dong, J. Zhao, W. Cai, Q. Zhang, H. Hua, Ultrabithorax is a key regulator for the dimorphism of wings, a main cause for the outbreak of planthoppers in rice, Natl. Sci. Rev. (2020), https://doi.org/10.1093/nsr/nwaa061.
[33] S.J. Fu, J.L. Zhang, S.J. Chen, H.H. Chen, Y.L. Liu, H.J. Xu, Functional analysis of Ultrabithorax in the wing-dimorphic planthopper Nilaparvata lugens (Stål, 1854) (Hemiptera: Delphacidae), Gene 737 (2020), https://doi.org/10.1016/j.gene.2020.144446.
[34] Foottit, R.G.; H. Adler, P.; Kathirithamby, J. Biodiversity of Strepsiptera. In Insect Biodiversity; 2018; pp. 673–704.
[35] M.F. Whiting, W.C. Wheeler, Insect homeotic transformation, Nature 368 (1994) 696.
[36] R.K. Delker, V. Ranade, R. Loker, R. Voutev, R.S. Mann, Low affinity binding sites in an activating CRM mediate negative autoregulation of the Drosophila Hox gene Ultrabithorax, PLOS Genet. 15 (2019), https://doi.org/10.1371/journal.pgen.1008444.
[37] M. Slattery, L. Ma, N. Négre, K.P. White, R.S. Mann, Genome-wide tissue-specific occupancy of the hox protein ultrabithorax and hox cofactor homothorax in Drosophila, PLOS One 6 (2011), https://doi.org/10.1371/journal.pone.0014686.
[38] S.W. Choo, R. White, S. Russell, Genome-wide analysis of the binding of the Hox protein Ultrabithorax and the Hox cofactor Homothorax in Drosophila, PLOS One 6 (2011), e14778, https://doi.org/10.1371/journal.pone.0014778.
[39] P. Agrawal, F. Habib, R. Yelagandula, L.S. Shashidhara, Genome-level identification of targets of Hox protein Ultrabithorax in Drosophila: novel mechanisms for target selection, Sci. Rep. 1 (2011) 205, https://doi.org/10.1038/srep00205.
[40] Ryan Loker, Jordyn E. Sanner, Richard S. Mann, Cell-type-specific Hox regulatory strategies orchestrate tissue identity, Curr. Biol. 31 (2021) 4246–4255, https://doi.org/10.1016/j.cub.2021.07.030.
[41] A. Tsai, A.K. Muthusamy, M.R. Alves, L.D. Lavis, R.H. Singer, D.L. Stern, J. Crocker, Nuclear microenvironments modulate transcription from low-affinity enhancers, eLife 6 (2017), https://doi.org/10.7554/elife.28975.
[42] Judith F. Kribelbauer, Chaitanya Rastogi, Harmen J. Bussemaker, Richard S. Mann, Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes, Annu. Rev. Cell Dev. Biol. 35 (2019) 357–379.
[43] M. Mir, A. Reimer, J.E. Haines, X.Y. Li, M. Stadler, H. Garcia, M.B. Eisen, X. Darzacq, Dense Bicoid hubs accentuate binding along the morphogen gradient, Genes Dev. 31 (2017) 1784–1794, https://doi.org/10.1101/gad.305078.117.
[44] J. Dufourt, A. Trullo, J. Hunter, C. Fernandez, J. Lazaro, M. Dejean, L. Morales, S. Nait-Amer, K.N. Schulz, M.M. Harrison, C. Favard, O. Radulescu, M. Lagha, Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs, Nat. Commun. 9 (2018) 5194, https://doi.org/10.1038/s41467-018-07613-z.
[45] M. Mir, M.R. Stadler, S.A. Ortiz, C.E. Hannon, M.M. Harrison, X. Darzacq, M. Eisen, Cite share dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos, eLife 7 (2018), https://doi.org/10.7554/eLife.40497.
[46] A. Tsai, A.K. Muthusamy, M.R. Alves, L.D. Lavis, R.H. Singer, D.L. Stern, J. Crocker, Nuclear microenvironments modulate transcription from low-affinity enhancers, eLife 6 (2022), https://doi.org/10.7554/eLife.28975.
[47] J.P. Karr, J.J. Ferrie, R. Tjian, X. Darzacq, The transcription factor activity gradient (TAG) model: contemplating a contact-independent mechanism for enhancer-promoter communication, Genes Dev. 36 (2022) 7–16, https://doi.org/10.1101/gad.349160.121.
[48] S. Merabet, R.S. Mann, To be specific or not: the critical relationship between Hox And TALE proteins, Trends Genet. xx (2016) 1–14, https://doi.org/10.1016/j.tig.2016.03.004.
[49] S. Feng, C. Rastogi, R. Loker, W.J. Glassford, H. Tomas Rube, H.J. Bussemaker, R. S. Mann, Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors, Nat. Commun. 13 (2022) 3808, https://doi.org/10.1038/s41467-022-31501-2.
[50] P.B. Pinto, K. Domsch, X. Gao, M. Wölk, J. Carnesecchi, I. Lohmann, Specificity of the Hox member Deformed is determined by transcription factor levels and binding site affinities, Nat. Commun. 13 (2022) 5037, https://doi.org/10.1038/s41467-022-32408-8.
[51] E. Rubin, X. Wu, T. Zhu, J.C. Cheung, H. Chen, A. Lorincz, R.K. Pandita, G. G. Sharma, H.C. Ha, J. Gasson, L.A. Hanakahi, T.K. Pandita, S. Sukumar, E. Rubin, X. Wu, T. Zhu, J.C. Cheung, H. Chen, A. Lorincz, R.K. Pandita, G.G. Sharma, H. C. Ha, J. Gasson, L.A. Hanakahi, T.K. Pandita, S. Sukumar, A role for the HOXB7 homeodomain protein in DNA repair, Cancer Res. 67 (2007) 1527–1535.
[52] B.C. Feltes, Architects meets repairers: the interplay between homeobox genes and DNA repair, DNA Repair (2019) 34–48.
[53] L. Comelli, L. Marchetti, D. Arosio, S. Riva, G. Abdurashidova, F. Beltram, A. Falaschi, The homeotic protein HOXC13 is a member of human DNA replication complexes, Cell Cycle 8 (2009) 454–459.
[54] V. Salsi, S. Ferrari, R. Ferraresi, A. Cossarizza, A. Grande, V. Zappavigna, HOXD13 binds DNA replication origins to promote origin licensing and is inhibited by geminin, Mol. Cell. Biol. 29 (2009) 5775–5788.
[55] I. Topisirovic, A. Kentsis, J.M. Perez, M.L. Guzman, C.T. Jordan, Borden, L. B. Katherin, Eukaryotic translation initiation factor 4E activity is modulated by HOXA9 at multiple levels, Mol. Cell. Biol. 25 (2005) 1100–1112, https://doi.org/10.1128/MCB.25.3.1100-1112.2005.
[56] D. Gao, J. Ning, G. Liu, S. Sun, X. Dang, SpatialMap: spatial mapping of unmeasured gene expression profiles in spatial transcriptomic data using generalized linear spatial models, Front. Genet. 26 (2022), 893522.
[57] S. Manley, J.M. Gillette, J. Lippincott-Schwartz, Single-particle tracking photoactivated localization microscopy for mapping single-molecule dynamics, Methods Enzym. 475 (2010) 109–120.

Seminars in Cell and Developmental Biology 152-153 (2024) 16–23

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Hox, homology, and parsimony: An organismal perspective

Andreas Wanninger [1]

University of Vienna, Department of Evolutionary Biology, Unit for Integrative Zoology, Djerassiplatz 1, 1030 Vienna, Austria

ARTICLE INFO

Keywords:
Evodevo
Phylogeny
Complexity
Ground pattern
Pleiotropy
Hox
Homology
Morphology
Hox interaction
MorphoEvoDevo

ABSTRACT

Hox genes are important regulators in animal development. They often show a mosaic of conserved (e.g., longitudinal axis patterning) and lineage-specific novel functions (e.g., development of skeletal, sensory, or locomotory systems). Despite extensive research over the past decades, it remains controversial at which node in the animal tree of life the Hox cluster evolved. Its presence already in the last common metazoan ancestor has been proposed, although the genomes of both putative earliest extant metazoan offshoots, the ctenophores and the poriferans, are devoid of Hox sequences. The lack of Hox genes in the supposedly "simple"-built poriferans and their low number in cnidarians and the basally branching bilaterians, the xenacoelomorphs, seems to support the classical notion that the number of Hox genes is correlated with the degree of animal complexity. However, the 4-fold increase of the Hox cluster in xiphosurans, a basally branching chelicerate clade, as well as the situation in some teleost fishes that show a multitude of Hox genes compared to, e.g., human, demonstrates, that there is no per se direct correlation between organismal complexity and Hox number. Traditional approaches have tried to base homology on the morphological level on shared expression profiles of individual genes, but recent data have shown that, in particular with respect to Hox and other regulatory genes, complex gene-gene interactions rather than expression signatures of individual genes alone are responsible for shaping morphological traits during ontogeny. Accordingly, for sound homology assessments and reconstructions of character evolution on organ system level, additional independent datasets (e.g., morphological, developmental) need to be included in any such analyses. If supported by solid data, proposed structural homology should be regarded as valid and not be rejected solely on the grounds of non-parsimonious distribution of the character over a given phylogenetic topology.

1. Introduction

Hox genes are homeodomain transcription factors that are involved in the control of a number of key processes in metazoan development such as the specification of tissue (including neural) identity, determination of the anterior-posterior body axis, segmentation, as well as appendage and limb formation [24,43,45,59,62,63,64,68,76]. They are typically divided into four classes, the anterior-, group 3-, central-, and posterior-class Hox genes (see, e.g., [31,30] for review). The number of Hox genes varies among metazoans and ranges from zero in poriferans and ctenophores, a minimum of two in cnidarians (note that the number of cnidarian Hox genes is highly variable, but since two can be considered as orthologs of respective other metazoan Hox sequences, this number is often assumed as ancestral for cnidarians), 9–15 in many invertebrates, to several dozens in teleost fishes. Its supposed sister group, the ParaHox genes, ancestrally come as a cluster of three [30,31,44]. The Hox and ParaHox genes most likely evolved together from a so-called Proto-Hox cluster that gave rise to both gene families [10,31,30], but see [25] and below for an alternative view. Given the synchronicity of their evolution under this scenario, there once was an early animal that already had both, Hox and ParaHox genes in its genome.

Hox genes have originally been described in *Drosophila* as a cluster of genes that regulate anterior-posterior axis development [57]. Subsequent studies on vertebrates have found that the order of the respective Hox genes on the genome is rather conserved between these distant lineages and that, during ontogeny, anteriorly positioned Hox genes become active before the middle ones and these in turn are turned on earlier than the posterior Hox genes. Moreover, anterior Hox genes (i.e., Hox sequences closer to the 3′ region of the genome) appear to be primarily responsible for patterning of anterior body regions, while the middle and the posterior ones act in the differentiation of the subsequent areas along the longitudinal body axis. This correlation between physical location on the chromosome, temporal expression, and spatial distribution of their transcripts has been termed "temporal-spatial

E-mail address: andreas.wanninger@univie.ac.at.
[1] ORCID ID: 0000–0002-3266–5838

https://doi.org/10.1016/j.semcdb.2023.01.007

Available online 18 January 2023

collinearity“ (e.g., [12,33,75]).

Early studies on *Drosophila* have already shown that the Hox genes are divided into two speparate subclusters, the Antennapedia and the Bithorax complexes [15,32,50,57]. Subsequent comparative analyses revealed that a fully intact Hox cluster as well as collinear expression most likely presents the exception rather than the rule among bilaterians. As such, numerous bilaterian animals show a more or less fragmented Hox cluster [98,3,67,74,87,89,90,95]. This may account for non-collinear and non-staggered expression of Hox genes in the respective lineages, albeit cases are known where species with a fragmented Hox cluster do show spatial (but not temporal) collinearity [69]. Although functional studies are still largely lacking for the vast majority of the parahoxozoan lineages (Cnidaria + Placozoa + Bilateria; [86]), these altered expression patterns often go hand in hand with putative gain-of-functions in taxa with a broken Hox cluster. Accordingly, invertebrates that exhibit a lack of temporal-spatial collinearity often show clade-specific expression patterns confined to well-defined morphological structures such as skeletal, sensory, defensive, and locomotory systems [111,112,115,55,88,89,90]. This illustrates the high degree of functional plasticity of Hox family genes. The combination of their crucial ancestral function in axial patterning at the dawn of Bilateria, together with their lineage-specific functions, that most likely contributed to the emergence of morphological novelties in numerous clades, renders Hox genes one of the most fascinating and relevant groups of regulatory genes for studies into the evolutionary origins of tissues, organs, and animal bodyplans.

2. Ghosts of the past: Origin of Hox and ParaHox genes

Hox genes have unequivocally been identified in Bilateria (including its proposed earliest extant offshoot, Xenacoelomorpha) and Cnidaria, and thus were already present in the bilaterian-cnidarian last common ancestor (LCA) (Fig. 1). Some recent data indicate that Placozoa – flat animals with a dorsal and ventral epithelial layer that lack decisive morphological body axes and symmetry planes (but see [22] – may constitute the sister taxon to Cnidaria [54,53], thus including the placozoans in the group of animals that derive from a Hox/ParaHox-bearing ancestor. Irrespective of whether a Cnidaria + Placozoa clade indeed exists or whether the cnidarians alone form the sister taxon to the bilaterians, earlier genomic studies already suggested that – despite lacking distinct Hox genes – placozoans do have a putative ParaHox gene, the proposed *Gsx*-homolog *Trox2* [22,46,86,91,92]. In light of the ProtoHox hypothesis, this provides indirect evidence for the presence of Hox gene(s) in the placozoan lineage and thus in the last common ancestor of Parahoxozoa (Fig. 1).

But did Hox genes really originate at the dawn of Parahoxozoa? To this end, the situation in the two remaining non-bilaterian metazoan clades, Ctenophora and Porifera, appears crucial. Both phyla are the prime candidates for the earliest extant offshoot within the metazoan tree of life, and the Ctenophora-first versus Porifera-first hypotheses are still hotly debated (see, e.g., [70,7,82,113,114,38,26,54,49,84]). With regard to the Hox (and ParaHox) complement, there is little doubt that ctenophores lack Hox genes and that they thus derived from a Hox/ParaHox-less LCA. The situation appears somewhat more contested for the poriferans, for which some reports claimed the putative existence of Hox or ParaHox genes in extant species [20,28]. This was refuted by subsequent studies, and the current notion is that sponge genomes – while possessing non-Hox homeobox genes (e.g., [52,28]) - are indeed devoid of distinct Hox and ParaHox genes (e.g., [52,86,77]). However, comparative gene synteny analyses have revealed that both, poriferans and placozoans, have Hox/ParaHox-free sites on their genomes that are flanked by sequences that show similarities to those that house Hox and ParaHox genes in bilaterians, thus proposing homology of these “ghost loci“ to the respective Hox/ParaHox-bearing positions in bilaterians [25,83]. This has opened new avenues into speculations as to what degree the mere existence of homologous sites on the genome may argue in favor of an ancestral occupation of these sites by the respective Hox and ParaHox genes. Proponents of the “ghost locus hypothesis“ have strongly argued for an existence of Hox and ParaHox genes in the LCA of poriferans [25,83]. Assuming the Porifera-first hypothesis is correct, this would push back the origin of Hox and ParaHox genes right to the urmetazoan and well into the pre-Cambrian, probably to at least 600mya (Fig. 1A-C).

However, such an early appearance of Hox and ParaHox genes does not remain undisputed. Recent studies, and especially a re-analyses of the dataset on which the proposed existence of a poriferan *Cdx* ParaHox gene was proposed [28], have found no evidence for Hox or ParaHox homologs in Porifera [77]. Even if the flanking regions (i.e., the arrangement of genes directly adjacent to Hox/ParaHox loci in parahoxozoans; see [83]) of bilaterian Hox genes show high sequence similarity (or even homology) to sites on poriferan genomes, does this imply that the “locus“ between these regions was necessarily occupied by the respective homologous genes in an ancestor long gone? If so, this would imply that the Hox genes and their flanking regions either co-evolved with each other over hundreds of millions of years and that the flanking regions remained relatively conserved, even in lineages that have secondarily lost the respective Hox genes, or that the Hox genes originated on a genomic site different to that of the flanking regions and were secondarily transposed between these areas. An alternative scenario, however, is that the flanking regions evolved prior to the Hox genes in early, non-Hox-bearing metazoans, and that the Hox genes themselves originated at the base of the parahoxozoans between the respective flanking sites. Under this latter assumption, the shared existence of the flanking regions between Hox/ParaHox-bearing and Hox/ParaHox-less animals would not be indicative of secondary loss of Hox and ParaHox genes in pre-parahoxozoans. It should be noted that claiming the historical existence of an ancestral character that is absent in its extant descendants is highly problematic and requires well-founded indirect evidence. In light of the ghost locus scenario, the likelihood of Hox/-ParaHox gene loss inferred by sequence similarity of the flanking regions depends on two issues: (1) the correct homology assessment of the flanking regions across Hox/ParaHox-bearing and Hox/ParaHox-less metazoans and (2) the number of cases (lineages, individual Hox genes) for which homology of flanking sequences can be claimed with a high degree of certainty.

In order to make a strong case of homology of the respective flanking regions, reconstruction of the ancestral Hox/ParaHox flanking regions in the last common parahoxozoan ancestor would be an important step forward. Assuming the ProtoHox hypothesis is correct, there were one “primordial“ Hox and ParaHox gene cluster each (cf. [31,30]). Both these clusters had distict flanking sequences and probably each Hox/-ParaHox gene within its cluster was again flanked by (non-coding) sequences. Obviously, just like the Hox/ParaHox genes themselves, these flanking sequences evolved over time and thus underwent modifications in their nucleotide sequences along the various parahoxozoan lineages. However, although often a difficult undertaking, assessing character evolution requires the reconstruction of ground patterns of the respective traits. These can then be mapped on a phylogenetic tree and compared to, e.g., patterns observed in the outgroup of interest. Accordingly, for the clusters and the individual Hox/ParaHox genes themselves, the hypothetical ancestral flanking sequences should be reconstructed for the LCA of Parahoxozoa. To this end, the genes found in the Hox/ParaHox flanking regions should be identified across the parahoxozoans and rigorously scrutinized for orthology. This should result in the reconstruction of an ancestral parahoxozoan gene sequence for each Hox/ParaHox flanking region of interest. Using such proposed flanking region ground patterns, non-Hox/ParaHox-bearing metazoans (poriferans and ctenophores) can then be specifically screened for the presence of putative homologous loci for the respective Hox/ParaHox genes or (sub)clusters. This could potentially result in considerably more, hitherto unknown ghost loci, as if only the flanking regions of distantly related extant species are compared to each other (e.g.,

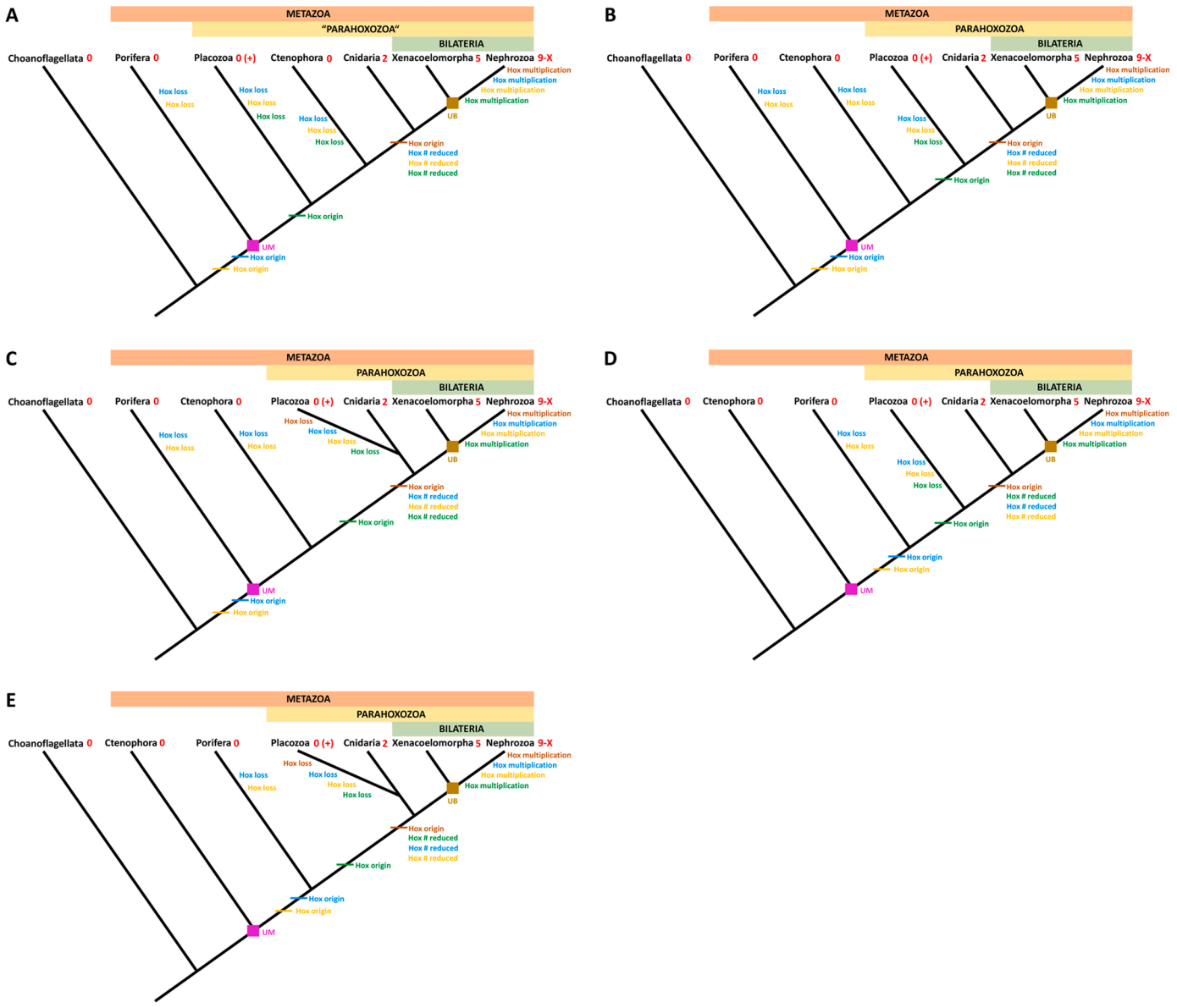


Fig. 1. Hypotheses on Hox gene evolution contingent on the underlying phylogenetic scenario and the presence of Hox genes in extant lineages. Scenarios considering the ghost locus theory (i.e., ancestral presence of Hox genes along the line towards Porifera) are also depicted. The common notion that the Hox and ParaHox clusters evolved simultaneously from a common "ProtoHox" gene or cluster, as well as independent evolution of Hox and ParaHox genes, are both taken into account. Orthology (single evolutionary origin, i.e., homology) of individual Hox genes between taxa is assumed in all cases. Red numbers denote number of Hox genes identified in given lineages (note that cnidarians have varying numbers of Hox genes, but one anterior and one posterior sequence are usually considered orthologs of respective bilaterian Hox genes), (+) indicates indirect evidence for Hox presence in the placozoan lineage due to the presence of a ParaHox gene, UB marks the last common ancestor of Bilateria ("urbilaterian"), and UM the last common metazoan ancestor ("urmetazoan"). Color code of Hox character changes reflects the following likely scenarios for Hox gene evolution: Blue: The ghost locus hypothesis suggests presence of Hox genes prior to the emergence of extant poriferans based on sequence homology of "flanking regions". Green: Traditional hypothesis taking into account the indirect evidence of an ancestral presence of Hox genes in the line towards Placozoa due to the existence of the ParaHox gene *Trox2* in *Trichoplax*. Yellow: Independent evolution of the Hox and ParaHox clusters in the light of the ghost locus hypothesis. Brown: Independent origin of Hox and ParaHox genes in traditional view (i.e., presence of Hox genes only considered where they occur in recent species). Note that the presence of Hox genes in the urmetazoan is only supported under the simultaneous acceptance of the Porifera-first scenario and the ghost locus hypothesis. In all other cases the Hox cluster originated after the emergence of the last common metazoan ancestor. A. Porifera-first hypothesis, traditional view, with Placozoa being the second metazoan offshoot, followed by Ctenophora. Note that "Parahoxozoa" is paraphyletic in this scenario (unless Ctenophora is included). Three independent events of losses of the entire Hox complement occurred under the ghost locus scenario (blue, yellow), two in the traditional view that considers ancestral Hox presence in Placozoa (green), and none if Hox and ParaHox genes evolved independently and the ghost locus hypothesis is not considered (brown). B. Porifera-first hypothesis with Ctenophora as second metazoan offshoot. Again, the ghost locus scenarios (blue, yellow) require three independent Hox gene loss events. C. Porifera-first with Cnidaria + Placozoa forming a monophyletic sister clade to Bilateria. Ghost locus scenarios (blue, yellow) again require three Hox losses, both other views (green, brown) only one (along the line towards Placozoa). D-E. The Ctenophora-first hypothesis requires only two Hox losses under the ghost locus scenario (blue, yellow) and one in the traditional view (green). D. No loss-of-Hox cluster event occurred in the Placozoa + (Cnidaria + Bilateria) clade in the traditional scenario, provided that Hox and ParaHox genes originated independently (brown). E. One loss-of-Hox cluster event in the topology suggesting ((Cnidaria + Placozoa) + Bilateria) in the traditional view under the assumption of independent evolution of Hox and ParaHox genes (brown).

poriferans or placozoans to human; cf. [83]). In case such homologous flanking regions/loci are identified, the probability that non-parahoxozoans indeed lost the genes of a proposed primordial (or fully differentiated) Hox/ParaHox cluster increases the more such homologous sites are found (unless the respective "flanking regions" were already present prior to the existence of the (Proto)Hox/ParaHox genes themselves; see above). Novel in silico approaches involving microsynteny analyses might provide valuable insights into such ancestral evolutionary events of small genomic regions in the future (e.g., [116]). It must be kept in mind, however, that, if a fully or partially intact Hox/ParaHox cluster was present in the metazoan LCA, such losses occurred not only in poriferans, but also in ctenophores and, at least partly, in cnidarians, placozoans, and possibly xenacoelomorphs (Fig. 1). This again implies that the respective ghost loci should also be present and detectable in these clades.

But why should we care about the exact emergence of Hox and ParaHox genes in ancestral animals? What are the implications whether or not the metazoan LCA already had (functional) Hox and ParaHox genes in its genome? The answer to that lies in the manifold crucial functions these genes have in animal ontogeny and most likely also in phenotypic diversification during evolution. In addition, Hox genes are often associated with "bodyplan complexity" that increased after Hox cluster evolution. Thereby, it is often assumed that the number of Hox genes present in the genome of a metazoan lineage positively correlates with cell type number and degree of tissue and organ system differentiation in the respective species ([99,108,5,56]; see below). Accordingly, reconstructing the full genome including the exact Hox/ParaHox complement of the LCA of all animals could allow for indirect indication concerning the degree of complexity of early-branching metazoans.

3. Hox genes and animal complexity

The drivers behind the evolution and ontogenetic mechanisms that govern animal complexity is one of the key topics in biology. Yet, a commonly accepted definition as to how complexity in biology is defined or even measured is still lacking (see, e.g., [37]). One problem of the term lies in its relative nature: What, in biology, is "complex" relative to something that, by comparison, is "simple"?

Complexity may be assessed on multiple levels, e.g., on the molecular (gene numbers, gene and protein interactions) and cellular (number of cell types), as well as the morphological (number of tissue types, organ systems) and behavioral or cognitive level (interactions with other animals and the environment, signal procession and learning) [1,27,37]. When comparing levels of morphological complexity among metazoan animals, the absolute number of cells and the number of distinct cell types are often considered as (relative) proxies for complexity [99,108,5,56]. However, a number of exceptions are known, particularly from microscopic invertebrates that exhibit rather complex bodyplans with an astonishing low number of cells (e.g., [5,71]).

A superficial view on Hox gene distribution among major bilaterian phyla suggests some correlation between the number of Hox genes and the morphological complexity using the above-mentioned parameters of its representatives (Fig. 1). Ancestrally, one anterior- and one posterior-class Hox gene was likely present in cnidarians, and traditional views claimed that an additional, central-class Hox gene belonged to the ground pattern of Xenacoelomorpha, the earliest bilaterian offshoot [17,42,69]. However, subsequent studies on acoels, nemertodermatids, and *Xenoturbella* identified 2–3 additional Hox homologs, suggesting that (at least) 5 Hox genes were present in the LCA of Xenacoelomorpha and hence Bilateria (Fig. 1; [48,29,104,9,93]). There has been some discussion as to whether a *Hox3* gene was present in the LCA of Xenacoelomorpha (and Bilateria), but to date no *Hox3* gene has been described from any xenacoelomorph species. This suggests that *Hox3* is indeed an autapomorphy of Nephrozoa (all bilaterians except Xenacoelomorpha; cf., e.g., [31,42]).

After the xenacoelomorph-nephrozoan split, the anterior, central, and posterior Hox genes increased in number on the line leading to the nephrozoans, most likely by tandem gene duplications [30]. This goes hand in hand with an observed increase in the number of cell and tissue types, elaboration of the nervous system, de novo evolution of excretory systems, as well as exo- and endoskeletal structures in various nephrozoan lineages. While most invertebrates only have one set of Hox genes (but see below for exceptions), vertebrates have multiplied this number. Accordingly, many vertebrates have four or more Hox clusters that resulted from several independent Hox cluster and/or whole genome duplication events [108,63]. It thus seems tempting to propose a correlation between the evolution of vertebrate complexity and their increase in Hox gene numbers. However, the situation found in teleost fishes with up to 14 copies of Hox genes, while mammals including humans only have four, contradicts this notion [108,4,65,66]. Accordingly, there is no absolute 1:1 correlation between the mere number of Hox genes and the observed relative morphological complexity of the respective animal taxa. This is congruent with the situation in, e.g., Chelicerata, where the early branching Xiphosura (horseshoe crabs) likely underwent two whole genome duplication (WGD) events, resulting in 2–4 copies of each Hox gene [51]. Independent of that, one WGD event occurred in the LCA of arachnids and scorpions [94,97]. While these two latter clades have evolved a number of evolutionary novelties that add to their overall bodyplan complexity and morphological diversity (e.g., pectines, book lungs, venom glands, specialized features of the opisthosoma), this does not apply to the xiphosurans with their rather uniform lifestyle and morphology involving only a low degree of appendage diversification and a comparatively simple nervous system. Nevertheless, the xiphosurans have preserved their extended Hox gene complement together with their conserved morphology over hundreds of millions of years. This raises fundamental questions about the functional roles and implications of this enlarged Hox gene set in these animals.

Interestingly, Hox/genome duplications have not been found in any of the genomes of cephalopod mollusks that rival the vertebrates with respect to neural complexity and sophisticated behavioral patterns. In Mollusca, the only WGD event unequivocally recorded to date stems from two species of giant *Achatina* land snails, making these the mollusks with the highest number of Hox genes known [58]. The potential role of this extra set of Hox genes in these snails, that by no means match the neuroanatomical or behavioral complexity of cephalopods, however, remains unresolved.

Altogether, there is increasing evidence that claiming a direct correlation between the number of Hox genes and the degree of complexity in bilaterians is an oversimplification of the actual factors that govern animal bodyplan evolution. While such a correlation may be true for deeper phylogenetic nodes (e.g., at the cnidarian-bilaterian and the xenacoelomorph-nephrozoan split or at the emergence of the vertebrates), a number of cases, particularly from protostomes, are known that do not match such a scenario. This suggests that (1) other and/or additional factors are important for the evolution of metazoan complexity and (2) Hox genes, together with other duplicated genes, were independently recruited into lineage-specific processes other than bodyplan complexity, such as, e.g., the evolution of terrestrialization or other ecological adaptations [58,94].

4. Hox gene expression signatures, structural homology, and the evolution of morphological diversity

As genes that encode for transcription factor proteins that are capable of regulating (i.e., activating or repressing) other developmental genes, members of the Hox family play a pivotal role in animal ontogeny. Their regulative capacities are likely also important drivers in the evolution of novel morphological traits [112,35,55,80,87,88,89,90]. Accordingly, the expression of Hox genes provides important insights into the molecular machinery that underlies organ system development and evolution, including key traits such as (parts of) the nervous system,

anterior-posterior axis determination, arthropod segment identity, or appendage and tissue patterning [2,59,60,79].

The advent of comparative gene expression analyses has marked a turning point for classical approaches to reconstruct organismal evolutionary history, because shared gene expression signatures have been used for proposing homology of given morphological structures (see [72,103] for discussion). Early approaches have tried to assess homology on the structural (tissue, organ, or organ system) level based on commonalities of regional expression of one or few genes alone (e.g., [40,39,34,13]). However, it has subsequently become clear that the mere shared expression pattern of individual genes is not conclusive to infer homology on tissue or organ system level (see [72] for discussion and examples). This is, because individual genes may be recruited into manifold developmental processes that govern structures that are clearly not homologous (pleiotropy) (see, e.g., [78] for extensive treatment of the subject). At the same time, apparent homologous structures may considerably differ in the underlying cellular, developmental, or molecular mechanisms (including the genes involved) that shape them during ontogeny, as, e.g., described in nematodes, where the formation of an obvious homologous character, the vulva, is underlain by different signaling pathways ([101,21,23]; see also [36]). This phenomenon, commonly known as developmental systems drift (DSD; [105,85,109, 102,36]), is the result of the fact that evolutionary change occurs on all levels of organismal organisation, including the genes that are expressed during ontogeny of the given phenotypic traits. As such, it is not only the morphology of a character that may undergo a significant makeover over millions of years, but also the genes and other developmental mechanisms that contribute to its very formation during ontogeny.

Is it then all in vain to use gene expression data for homology assessments? Not quite. In principle, the same rules for formulating homology hypotheses apply to molecular as they do to morphological data: The more shared subtraits concerning the character of interest are found between species, the higher the likelihood of shared ancestry and hence homology of the character itself. The underlying conjecture here is that increased similarity is likely due to increased complexity in the molecular and genetic pathways, up to shared gene regulatory networks (“character identity network“ sensu [106,107]) that underlie the ontogenetic formation of the trait in question. Characters that are underlain by more complex developmental and molecular patterns are likely more stable over time than traits governed by a simpler machinery. Thus, while not decisive by themselves, shared gene expression signatures during ontogeny may provide a first hint towards putative homology of the structures in question. In other words, an initial finding that homologous genes are similarly expressed in a given morphological trait of different species should call for further analyses that may result in a novel primary homology hypothesis about this character.

Although homology assessments based on gene expression profiles alone remain problematic, comparative analyses of Hox gene expression may still lead to crucial insights into the evolutionary transformation and diversification of bodyplan traits that emerged from a common ancestral scheme. For example, while the individual rays in the fins of teleost fishes cannot be directly homologized with skeletal elements of tetrapod limbs, both types of vertebrate appendages arose from an ancestral pectoral fin-like structure (e.g., [41]). While the distal regions of fins and tetrapod limbs are both dependent on *Hox13* expression, the zeugopod, that forms the “middle” or “intermediate” region of the tetrapod arm/leg (consisting of ulna and radius versus tibia and fibula, respectively), is under *Hox11* regulation, and loss of *Hoxa11* and *Hoxd11* expression results in significantly shorter zeugopods, as the respective ulna/radius and tibia/fibia fail to form long bones [8,19,41]. Interestingly, however, experimentally induced *Hox11* activity in the developing teleost fin results in de novo formation of additional “intermediate” skeletal elements [41]. This points towards an ancient, inherent developmental program with the capability to form tri-partite vertebrate limbs already in the last common ancestor of teleosts and tetrapods [41,6]. While this program only became live in the tetrapods, resulting in elongated zeugopods as distinct morphological novelties, its underlying Hox-dependent regulatory mechanism appears to have been inhereted from a distant vertebrate ancestor long before the emergence of the tetrapods.

A complex interplay of Hox gene expression, including mutual activation and repression, has also been shown to be a major driver during development of the highly diverse arthropod appendages, including legs, antennae, mouthparts, and others (see, e.g., [44]). Here, again, the emergence of morphological diversity and novelty from a common groundplan (the arthropod limb) is mediated by dynamic interactions of Hox genes along the anterior-posterior body axis. In *Drosophila*, *Antennapedia* (*Atp*) expression may result in the transformation of antennae into legs [44], while *Ultrabithorax* (*Ubx*) activity may change halteres into wings [61]. Hox gene interactions are possible because the expression of individual Hox genes is usually not confined to distinct segments along the anterior-posterior axis, but instead shows overlapping domains. The resulting sites of Hox gene co-expression are then destined to produce different derivatives of the arthropod limb [44]. In the amphipod crustacean *Parhyale hawaiensis*, multi-level interactions of Hox genes have been identified as main mediators in the formation of the highly diverse appendages of the crustacean bodyplan [62,96]. Through a series of experiments using CRISPR/Cas9 editing, a sophisticated system of “cross regulatory interactions” between *Ubx, abdA*, and *AbdB* was revealed, that is responsible for assigning distinct identities to the appendages on the respective segments along the longitudinal body axis during develoment of *Parhyale* [47]. It thus appears that, in order to enhance our understanding of the evolution of the grand diversity of arthropod (and possibly also other bilaterian) body plans, focus should be on revealing the intricate and distinct interactions between Hox genes rather then comparing the expression signatures of individual Hox genes alone.

The examples above show that, in order to fully appreciate the role of Hox genes during animal ontogeny and evolution, we need to shift our thinking and analytical approaches towards Hox gene interactions rather than relying on mere comparisons of expression patterns of individual genes. This should enable us to reconstruct the genetic basis of morphological transformation series of key metazoan traits from common, ancestral character states, and will help uncover the mechanisms that underlie animal phenotypic diversity.

5. Beyond (Hox) genes: Reconstructing evolution and avoiding the parsimony trap

When reconstructing scenarios of character evolution, three major hypothesis-based analyses are considered intimately linked to each other: homology, phylogeny, and parsimony (see [100] for extensive treatment of the subject). Characters or character states are mapped on a phylogenetic tree and their evolutionary history (emergence, modifications, losses) are assessed using parsimony as an underlying principle. Thereby, if mapped on a phylogeny, it is assumed that the least amount of changes in the state of a given character is the most likely one to explain its evolutionary history. On the other hand, phylogenetic trees are also used to formulate so-called secondary homology assumptions [119,11]. To this end, character (states) are considered homologous if their distribution on the phylogeny complies with the most parsimonious scenario. There are, however, several problems associated with such an unscrutinized application of the parsimony principle to evolutionary reconstructions.

Characters – on whatever level – are homologous (have shared evolutionary ancestry) or not, independent of the phylogenetic trees they are plotted on, and irrespective of whether such homology complies with parsimony or not. If the data that argue for homology are convincing (e.g., by combining morphological, ontogenetic, and (Hox) gene expression data), then the respective characters should be accepted to be homologous (until compelling counter evidence becomes available). By all what we know at least since Darwin’s Origin of Species [18]

and Owen's first application of the term to biology [73], homology – the common descent of characters from one ancestral character (state) - is a fact, a biological reality. Parsimony, on the contrary, is not. Character evolution may follow parsimony in some cases and may not in others, and thus it should not be considered the only alternative for reconstructing evolutionary transformations (e.g., [14,16], see also discussion in [100]). Certainly, parsimony should not be used as the deciding argument as to whether traits are homologous or not, because a parsimony analysis is not based on an intrinsic biological "logic". Phylogenetic trees are prone to change, and what appears parsimonious in one scenario may not do so in a different phylogenetic setting. But the – recognized or not – homology of the character in question is, per se, independent of the relatedness and phylogenetic distance of the organisms that exhibit this very trait. In other words: A character becomes not "less (likely) homologous" simply because a novel topology suggests a greater phylogenetic distance of two taxa exhibiting the trait. It is the inherent biological data related to the character (e.g., similarity in subordinate traits) that reflects its evolutionary heritage and thus is decisive for character homology across taxa. If reconstructing the evolution of a given character (trait) based on an underlying phylogeny requires, for example, non-parsimonious multiple losses or modifications of this character, then this should be accepted as the most likely scenario and not be dismissed (albeit maybe re-assessed) based on the non-parsimonious situation that results from the phylogenetic tree used. Quite to the contrary, such non-parsimonious character distributions may call into question the actual correctness of the phylogenetic scenario (the tree or cladogram) rather than homology of the characters for which strong evidence is available. We should trust thoroughly executed homology assessments rather than dismissing them simply because they do not comply with a parsimonious scenario of the phylogenetic framework used.

6. Outlook

Even after decades of research, Hox genes do not cease to fascinate evolutionary biologists. Their involvement in a multitude of ontogenetic processes that have been conserved over hundreds of millions of years suggests that they already occupied a central role in shaping the bodyplans of early animals. In addition, Hox genes have been coopted into numerous taxon-specific functions and therefore are likely important drivers of animal morphological diversity. Novel experimental approaches including RNAi, CRISPR/Cas9, or single cell RNA sequencing, as well as the ever-improved quality and number of metazoan genomes available, will continue to provide relevant insights into tracing evolutionary patters. If embedded in a holistic context that includes morphological and phylogenetic data, Hox gene distribution and temporal-spatial activity across metazoan lineages may ultimately aid in reconstructing molecular and morphological groundpatterns at crucial nodes in metazoan evolution ("MorphoEvoDevo", cf. [110]). Caution must be taken, however, to avoid common traps and flaws when formulating homology hypotheses on any level of biological organisation. Theoretical frameworks such as the parsimony principle must not outweigh the actual, biological data when assessing character evolution.

Acknowledgements

The author is grateful to Richard Mann (New York, USA) for the invitation to contribute to this topical edition. Simon Sprecher (Fribourg, Switzerland) and Pedro Martinez (Barcelona, Spain) are thanked for insightful discussions on the putative Hox gene complement in Xenacoelomorpha. Research in the Wanninger lab has been funded by the Austrian Science Fund (grant number P29455-B29).

References

[1] C. Adami, What is complexity, BioEssays 24 (2002) 1085–1094, https://doi.org/10.1002/bies.10192.
[2] M. Akam, Hox genes and the evolution of diverse body plans, Philos. Trans. R. Soc. Lond. B 349 (1995) 313–319, https://doi.org/10.1098/rstb.1995.0119.
[3] C.B. Albertin, O. Simakov, T. Mitros, Z.Y. Wang, J.R. Pungor, E. Edsinger-Gonzales, S. Brenner, C.W. Ragsdale, D.S. Rokhsar, The octopus genome and the evolution of cephalopod neural and morphological novelties, Nature 524 (2015) 220–224, https://doi.org/10.1038/nature14668.
[4] A. Amores, A. Force, Y.L. Yan, C. Amemiya, A. Fritz, R.K. Ho, J. Langeland, V. Prince, Y.L. Wang, M. Westerfield, M. Ekker, J.H. Postlethwait, Genome duplications in vertebrate evolution: evidence from zebrafish Hox clusters, Science 282 (1998) 1711–1714, https://doi.org/10.1126/science.282.5394.1711.
[5] G. Bell, A.O. Mooers, Size and complexity among multicellular organisms, Biol. J. Linn. Soc. 60 (1997) 345–363, https://doi.org/10.1111/j.1095-8312.1997.tb01500.x.
[6] X. Bi, K. Wang, L. Yang, H. Pan, H. Jiang, Q. Wei, M. Fang, H. Yu, C. Zhu, Y. Cai, Y. He, X. Gan, H. Zeng, D. Yu, Y. Zhu, H. Jiang, Q. Qiu, H. Yang, Y.E. Zhang, W. Wang, M. Zhu, S. He, G. Zhang, Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes, Cell 184 (2021) 1377–1391, https://doi.org/10.1016/j.cell.2021.01.046.
[7] M.L. Borowiec, E.K. Lee, J.C. Chiu, D.C. Plachetzki, Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa, BMC Genom. 16 (2015) 987, https://doi.org/10.1186/s12864-015-2146-4.
[8] A.M. Boulet, M.R. Capecchi, M.R. multiple roles of Hoxa11 and Hoxd11 in the formation of the mammalian forelimb zeugopod, Development 131 (2004) 299–309.
[9] M. Brauchle, A. Bilican, C. Eyer, X. Bailly, P. Martinez, P. Ladurner, R. Bruggmann, S.G. Sprecher, Xenacoelomorpha survey reveals that all 11 animal homeobox gene classes were present in the first bilaterians, Genome Biol. Evol. 10 (2018) 2205–2217, https://doi.org/10.1093/gbe/evy170.
[10] N.M. Brooke, J. Garcia-Fernandez, P.W. Holland, The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster, Nature 392 (1998) 920–922, https://doi.org/10.1038/31933.
[11] A.V.Z. Brower, V. Schawaroch, Three steps of homology assessment, Cladistics 12 (1996) 265–272, https://doi.org/10.1111/j.1096-0031.1996.tb00014.x.
[12] A.C. Burke, C.E. Nelson, B.A. Morgan, C. Tabin, Hox genes and the evolution of vertebrate axial morphology, Development 121 (1995) 333–346, https://doi.org/10.1242/dev.121.2.333.
[13] P. Callaerts, G. Halder, W.J. Gehring, PAX6 in development and evolution, Annu. Rev. Neurosci. 20 (1997) 483–532, https://doi.org/10.1146/annurev.neuro.20.1.483.
[14] J.H. Camin, R.R. Sokal, A method for deducing branching sequences in phylogeny, Evolution 19 (1965) 311–326, https://doi.org/10.2307/2406441.
[15] S.B. Carroll, Homeotic genes and the evolution of arthropods and chordates, Nature 376 (1995) 479–485, https://doi.org/10.1038/376479a0.
[16] L.L. Cavalli-Sforza, A.W.F. Edwards, Phylogenetic analysis. Models and estimation procedures, Am. J. Hum. Genet. 19 (1967) 233–257.
[17] C.E. Cook, E. Jimenez, M. Akam, E. Salo, The Hox gene complement of acoel flatworms, a basal bilaterian clade, Evol. Dev. 6 (2004) 154–163, https://doi.org/10.1111/j.1525-142X.2004.04020.x.
[18] C. Darwin, On the origin of species by means of natural selection, Murray, London, 1859, https://doi.org/10.4324/9780203509104.
[19] A.P. Davis, D.P. Witte, H.M. Hsieh-Li, S.S. Potter, M.R. Capecchi, Absence of radius and ulna in mice lacking hoxa-11 and hoxd-11, Nature 375 (1995) 791–795, https://doi.org/10.1111/j.1096-0031.1991.tb00045.x.
[20] B.M. Degnan, S.M. Degnan, A. Giusti, D.E. Morse, A hox/hom homeobox gene in sponges, Gene 155 (1995) 175–177, https://doi.org/10.1016/0378-1119(94)00908-B.
[21] M. Delattre, M.-A. Félix, Polymorphism and evolution of vulval precursor cell lineages within two nematode genera, Caenorhabditis and Oscheius, Curr. Biol. 11 (2001) 631–643, https://doi.org/10.1016/s0960-9822(01)00202-0.
[22] T.Q. DuBuc, J.F. Ryan, M.Q. Martindale, "Dorsal–ventral" genes are part of an ancient axial patterning system: evidence from Trichoplax adhaerens (Placozoa), Mol. Biol. Evol. 36 (2019) 966–973, https://doi.org/10.1093/molbev/msz025.
[23] M. Felix, Cryptic quantitative evolution of the vulva intercellular signaling network in Caenorhabditis, Curr. Biol. 17 (2007) 103–114, https://doi.org/10.1016/j.cub.2006.12.024.
[24] W. Feng, H. Destain, J.J. Smith, P. Kratsios, Maintenance of neurotransmitter identity by Hox proteins through a homeostatic mechanism, bioRxiv (2022), https://doi.org/10.1101/2022.05.05.490791.
[25] D.E.K. Ferrier, The origin of the Hox/ParaHox genes, the Ghost Locus hypothesis and the complexity of the first animal, Brief. Funct. Genom. 15 (2016) 333–341, https://doi.org/10.1093/bfgp/elv056.
[26] R. Feuda, M. Dohrmann, W. Pett, H. Philippe, O. Rota-Stabelli, N. Lartillot, G. Wörheide, D. Pisani, Improved modeling of compositional heterogeneity supports sponges as sister to all other animals, Curr. Biol. 27 (2017) 3864–3870, https://doi.org/10.1016/j.cub.2017.11.008.
[27] B.J. Finlay, G.F. Esteban, Can biological complexity be rationalized, Bioscience 59 (2009) 333–340, https://doi.org/10.1525/bio.2009.59.4.11.
[28] S.A.V. Fortunato, M. Adamski, O.M. Ramos, S. Leininger, J. Liu, D.E.K. Ferrier, M. Adamska, Calcisponges have a ParaHox gene and dynamic expression of

dispersed NK homeobox genes, Nature 514 (2014) 620–623, https://doi.org/10.1038/nature13881.
[29] G. Fritzsch, M.U. Böhme, M. Thorndyke, H. Nakano, O. Israelsson, T. Stach, M. Schlegel, T. Hankeln, P.F. Stadler, PCR survey of Xenoturbella bocki Hox genes, J. Exp. Zool. B Mol. Dev. Evol. 310 (2008) 278–284, https://doi.org/10.1002/jez.b.21208.
[30] J. Garcia-Fernàndez, Hox, ParaHox, ProtoHox: facts and guesses, Heredity (Edinb) 94 (2005) 145–152, https://doi.org/10.1038/sj.hdy.6800621.
[31] J. Garcia-Fernàndez, The genesis and evolution of homeobox gene clusters, Nat. Rev. Genet. 6 (2005) 881–892, https://doi.org/10.1038/nrg1723.
[32] J. Garcia-Fernàndez, P.W. Holland, Archetypal organization of the amphioxus Hox gene cluster, Nature 370 (1994) 563–566, https://doi.org/10.1038/370563a0.
[33] S.J. Gaunt, The significance of Hox gene collinearity, Int. J. Dev. Biol. 59 (2015) 159–170, https://doi.org/10.1387/ijdb.150223sg.
[34] W.J. Gehring, The master control gene for morphogenesis and evolution of the eye, Genes Cells 1 (1996) 11–15, https://doi.org/10.1046/j.1365-2443.1996.11011.x.
[35] W.J. Glassford, W.C. Johnson, N.R. Dall, S.J. Smith, Y. Liu, W. Boll, M. Noll, M. Rebeiz, Co-option of an ancestral Hox-regulated network underlies a recently evolved morphological novelty, Dev. Cell 34 (2015) 520–531, https://doi.org/10.1016/j.devcel.2015.08.005.
[36] E.S. Haag, J.R. True, Developmental system drift, in: L.N. de la Rosa, G.B. Müller (Eds.), Evolutionary Developmental Biology: A Reference Guide, Springer Nature, Switzerland, 2021, pp. 99–110, https://doi.org/10.1007/978-3-319-32979-6_83.
[37] P.F. Hagolani, R. Zimm, R. Vroomans, I. Salazar-Ciudad, On the evolution and development of morphological complexity: a view from gene regulatory networks, in: PLoS Comput. Biol, 17, 2021, e1008570, https://doi.org/10.1371/journal.
[38] K.M. Halanych, N.V. Whelan, K.M. Kocot, A.B. Kohn, L.L. Moroz, Miscues misplace sponges, Proc. Natl. Acad. Sci. USA 113 (2016) E946–E947, https://doi.org/10.1073/pnas.1525332113.
[39] G. Halder, P. Callaerts, W.J. Gehring, New perspectives on eye evolution, Curr. Opin. Genet. Dev. 5 (1995) 602–609, https://doi.org/10.1016/0959-437x(95)80029-8.
[40] G. Halder, P. Callaerts, W.J. Gehring, Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila, Science 267 (1995) 1788–1792, https://doi.org/10.1126/science.7892602.
[41] M.B. Hawkins, K. Henke, M.P. Harris, Latent developmental potential to form limb-like skeletal structures in zebrafish, in: Cell, 184, 2021, pp. 899–911, https://doi.org/10.1016/j.cell.2021.01.003.
[42] A. Hejnol, Acoelomorpha and Xenoturbellida, in: A. Wanninger (Ed.), Evolutionary Developmental Biology of Invertebrates 1, Springer, Vienna, 2015, pp. 203–214.
[43] J.C.-G. Hombría, M. García-Ferrés, C. Sánchez-Higueras, Anterior Hox genes and the process of cephalization, Front. Cell Dev. Biol. 9 (2021), 718175, https://doi.org/10.3389/fcell.2021.718175.
[44] C.L. Hughes, T.C. Kaufman, Hox genes and the evolution of the arthropod body plan, Evol. Dev. 4 (2002) 459–499.
[45] M. Jager, J. Murienne, C. Clabaut, J. Deutsch, H. Le Guyader, M. Manuel, Homology of arthropod anterior appendages revealed by Hox gene expression in a sea spider, Nature 441 (2006) 506–508, https://doi.org/10.1038/nature04591.
[46] W. Jakob, S. Sagasser, S. Dellaporta, P. Holland, K. Kuhn, B. Schierwater, The Trox-2 Hox/ParaHox gene of Trichoplax (Placozoa) marks an epithelial boundary, Dev. Genes Evol. 214 (2004) 170–175, https://doi.org/10.1007/s00427-004-0390-8.
[47] E. Jarvis Alberstat, K. Chung, D.A. Sun, S. Ray, N.H. Patel, Combinatorial interactions of Hox genes establish appendage diversity of the amphipod crustacean Parhyale hawaiensis, bioRxiv (2022), https://doi.org/10.1101/2022.03.25.485717.
[48] E. Jimenez-Guri, J. Paps, J. Garcia-Fernandez, E. Salo, Hox and ParaHox genes in Nemertodermatida, a basal bilaterian clade, Int. J. Dev. Biol. 50 (2006) 675–679, https://doi.org/10.1387/ijdb.062167ej.
[49] K. Juravel, L. Porras, S. Höhna, D. Pisani, G. Wörheide, Improved resolution of recalcitrant nodes in the animal phylogeny through the analysis of genome gene content and morphology, bioRxiv (2021), https://doi.org/10.1101/2021.11.19.469253.
[50] T. Kaufman, M.A. Seeger, G. Olsen, Molecular and genetic organization of the Antennapedia gene complex of Drosophila melanogaster, in: T.R.F. Wright (Ed.), Advances in genetics 27, Academic Press, San Diego, 1990, pp. 309–361.
[51] N.J. Kenny, K.W. Chan, W. Nong, Z. Qu, I. Maeso, H.Y. Yip, T.F. Chan, H.S. Kwan, P.W.H. Holland, K.H. Chu, J.H.L. Hui, Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs, Heredity 116 (2016) 190–199, https://doi.org/10.1038/hdy.2017.38.
[52] C. Larroux, B. Fahey, S.M. Degnan, M. Adamski, D.S. Rokhsar, B.M. Degnan, The NK homeobox gene cluster predates the origin of Hox genes, Curr. Biol. 17 (2007) 706–710, https://doi.org/10.1016/j.cub.2007.03.008.
[53] C.E. Laumer, H. Gruber-Vodicka, M.G. Hadfield, V.B. Pearse, A. Riesgo, J. C. Marioni, G. Giribet, Support for a clade of Placozoa and Cnidaria in genes with minimal compositional bias, ELife 7 (2018), e36278, https://doi.org/10.7554/eLife.36278.
[54] C.E. Laumer, R. Fernández, S. Lemer, D. Combosch, K.M. Kocot, A. Riesgo, S.C. S. Andrade, W. Sterrer, M.V. Sørensen, G. Giribet, Revisiting metazoan phylogeny with genomic sampling of all phyla, Proc. R. Soc. B Biol. Sci. 286 (2019), https://doi.org/10.1098/rspb.2019.0831.
[55] P.N. Lee, P. Callaerts, H.G. De Couet, M.Q. Martindale, Cephalopod Hox genes and the origin of morphological novelties, Nature 424 (2003) 1061–1065, https://doi.org/10.1038/nature01872.
[56] D. Lemons, W. McGinnis, Genomic evolution of Hox gene clusters, Science 313 (2006) 1918–1922, https://doi.org/10.1126/science.1132040.
[57] E.B. Lewis, A gene complex controlling segmentation in Drosophila, Nature 276 (1978) 565–570, https://doi.org/10.1038/276565a0.
[58] C. Liu, Y. Ren, Z. Li, Q. Hu, L. Yin, H. Wang, X. Qiao, Y. Zhang, L. Xing, Y. Xi, F. Jiang, S. Wang, C. Huang, B. Liu, H. Liu, F. Wan, W. Qian, W. Fan, Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic–terrestrial transition, Mol. Ecol. Resour. 21 (2020) 478–494, https://doi.org/10.1111/1755-0998.13261.
[59] R. Loker, J.E. Sanner, R.S. Mann, Cell-type-specific Hox regulatory strategies orchestrate tissue identity, Curr. Biol. 31 (2021) 4246–4255, https://doi.org/10.1016/j.cub.2021.07.030.
[60] M. Mallo, D.M. Wellik, J. Deschamps, Hox genes and regional patterning of the vertebrate body plan, Dev. Biol. 344 (2010) 7–15, https://doi.org/10.1016/j.ydbio.2010.04.024.
[61] R.S. Mann, G. Morata, The developmental and molecular biology of genes that subdivide the body of Drosophila, Annu. Rev. Cell Dev. Biol. 16 (2000) 243–271, doi: 1081-0706/00/1115-0243.
[62] A. Martin, J.M. Serano, E. Jarvis, H.S. Bruce, J. Wang, S. Ray, C.A. Barker, L. C. O'Connell, N.H. Patel, CRISPR/Cas9 mutagenesis reveals versatile roles of Hox genes in crustacean limb specification and evolution, Curr. Biol. 26 (2016) 14–26, https://doi.org/10.1016/j.cub.2015.11.021.
[63] W. McGinnis, R. Krumlauf, Homeobox genes and axial patterning, Cell 68 (1992) 283–302, https://doi.org/10.1016/0092-8674(92)90471-n.
[64] C. McQueen, M. Towers, Establishing the pattern of the vertebrate limb, Development 147 (2020) dev177956, https://doi.org/10.1242/dev.177956.
[65] H.K. Moghadam, M.M. Ferguson, R.G. Danzmann, Evolution of Hox clusters in Salmonidae: a comparative analysis between Atlantic salmon (Salmo salar) and rainbow trout (Oncorhynchus mykiss), J. Mol. Evol. 61 (2005) 636–649, https://doi.org/10.1007/s00239-004-0338-7.
[66] H.K. Moghadam, M.M. Ferguson, R.G. Danzmann, Evidence for Hox gene duplication in rainbow trout (Oncorhynchus mykiss): a tetraploid model species, J. Mol. Evol. 61 (2005) 804–818, https://doi.org/10.1007/s00239-004-0230-5.
[67] A.S. Monteiro, D.E. Ferrier, Hox genes are not always Colinear, Int. J. Biol. Sci. 2 (2006) 95–103, https://doi.org/10.7150/ijbs.2.95.
[68] C. Moreau, P. Caldarelli, D. Rocancourt, J. Roussel, N. Denans, O. Pourquie, J. Gros, Timed collinear activation of Hox genes during gastrulation controls the avian forelimb position, Curr. Biol. 29 (2019) 35–50, https://doi.org/10.1016/j.cub.2018.11.009.
[69] E. Moreno, M. Nadal, J. Baguñà, P. Martínez, Tracking the origins of the bilaterian Hox patterning system: insights from the acoel flatworm Symsagittifera roscoffensis, Evol. Dev. 11 (2009) 574–581, https://doi.org/10.1111/j.1525-142X.2009.00363.x.
[70] L.L. Moroz, K.M. Kocot, M.R. Citarella, S. Dosung, T.P. Norekian, I. S. Povolotskaya, A.P. Grigorenko, C. Dailey, E. Berezikov, K.M. Buckley, A. Ptitsyn, D. Reshetov, K. Mukherjee, T.P. Moroz, Y. Bobkova, F. Yu, V. V. Kapitonov, J. Jurka, Y.V. Bobkov, J.J. Swore, D.O. Girardo, A. Fodor, F. Gusev, R. Sanford, R. Bruders, E. Kittler, C.E. Mills, J.P. Rast, R. Derelle, V.V. Solovyev, F. A. Kondrashov, B.J. Swalla, J.V. Sweedler, E.I. Rogaev, K.M. Halanych, A. B. Kohn, The ctenophore genome and the evolutionary origins of neural systems, Nature 510 (2014) 109–114, https://doi.org/10.1038/nature13400.
[71] R.C. Neves, K.J.K. Sørensen, R.M. Kristensen, A. Wanninger, Cyclicphoran dwarf males break the rule: high complexity with low cell numbers, Biol. Bull. 217 (2009) 2–5.
[72] C. Nielsen, P. Martinez, Patterns of gene expression: homology or homocracy, Dev. Genes Evol. 213 (2003) 149–154, https://doi.org/10.1007/s00427-003-0301-4.
[73] R. Owen, Lectures on the comparative anatomy and physiology of the invertebrate animals: delivered at the Royal College of Surgeons, Longman, Brown, Green & Longmans, 1843.
[74] R.M. Pace, M. Grbić, L.M. Nagy, Composition and genomic organization of arthropod Hox clusters, Evodevo 7 (2016) 1–11.
[75] S. Papageorgiou, Hox gene collinearity: from A-P patterning to radially symmetric animals, Curr. Genom. 17 (2016) 444–449, https://doi.org/10.2174/1389202917666160616082436.
[76] H.J. Parker, M.E. Bronner, R. Krumlauf, The vertebrate Hox gene regulatory network for hindbrain segmentation: evolution and diversification: Coupling of a Hox gene regulatory network to hindbrain segmentation is an ancient trait originating at the base of vertebrates, Bioessays 38 (2016) 526–538, https://doi.org/10.1002/bies.201600010.
[77] C.C. Pastrana, M.B. DeBiasse, J.F. Ryan, Sponges lack ParaHox genes, Genome Biol. Evol. 11 (2019) 1250–1257, https://doi.org/10.1093/gbe/evz052.
[78] I.S. Peter, E.H. Davidson, Genomic Control Process: Development and Evolution, Academic Press, Boston, 2015.
[79] P. Philippidou, J.S. Dasen, Hox genes: choreographers in neural development, architects of circuit organization, Neuron 80 (2013) 12–34, https://doi.org/10.1016/j.neuron.2013.09.020.
[80] L. Pick, A. Heffer, Hox gene evolution: multiple mechanisms contributingto evolutionary novelties, Ann. N. Y. Acad. Sci. 1256 (2012) 15–32, https://doi.org/10.1111/j.1749-6632.2011.06385.x.
[81] M.C.C. de Pinna, Concepts and tests of homology in the cladistic paradigm, Cladistics 7 (1991) 367–394, https://doi.org/10.1111/j.1096-0031.1991.tb00045.x.

[82] D. Pisani, W. Pett, M. Dohrmann, R. Feuda, O. Rota-Stabelli, H. Philippe, N. Lartillot, G. Wörheide, Genomic data do not support comb jellies as the sister group to all other animals, Proc. Natl. Acad. Sci. U. S. A 112 (2015) 15402–15407, https://doi.org/10.1073/pnas.1518127112.
[83] O.M. Ramos, D. Barker, D.E. Ferrier, Ghost loci imply Hox and ParaHox existence in the last common ancestor of animals, Curr. Biol. 22 (2012) 1951–1956, https://doi.org/10.1016/j.cub.2012.08.023.
[84] A.K. Redmond, A. McLysaght, Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding, Nat. Commun. 12 (2021) 1783, https://doi.org/10.1038/s41467-021-22074-7.
[85] R. Robinson, Different paths, same structure: "developmental systems drift" at work, PLoS Biol. 9 (2011), e1001113, https://doi.org/10.1371/journal.pbio.1001113.
[86] J.F. Ryan, K. Pang, J.C. Mullikin, M.Q. Martindale, A.D. Baxevanis, NISC comparative sequencing program, the homeodomain complement of the ctenophore Mnemiopsis leidyi suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa, EvoDevo 1 (2010), https://doi.org/10.1186/2041-9139-1-9.
[87] D.A. Salamanca-Díaz, A.D. Calcino, A.L. de Oliveira, A. Wanninger, Non-collinear Hox gene expression in bivalves and the evolution of morphological novelties in mollusks, Sci. Rep. 11 (2021) 3575, https://doi.org/10.1038/s41598-021-82122-6.
[88] L. Samadi, G. Steiner, Involvement of Hox genes in shell morphogenesis in the encapsulated development of a top shell gastropod (Gibbula varia L.), Dev. Genes Evol. 219 (2009) 523–530, https://doi.org/10.1007/s00427-009-0308-6.
[89] L. Samadi, G. Steiner, Expression of Hox genes during the larval development of the snail, Gibbula varia (L.)-further evidence of non-colinearity in molluscs, Dev. Genes Evol. 220 (2010) 161–172, https://doi.org/10.1007/s00427-010-0338-0.
[90] S.M. Schiemann, J.M. Martín-Durán, A. Børve, B.C. Vellutini, Y.J. Passamaneck, A. Hejnol, Clustered brachiopod Hox genes are not expressed collinearly and are associated with lophotrochozoan novelties, Proc. Natl. Acad. Sci. 114 (2017) E1913–E1922, https://doi.org/10.1073/pnas.1614501114.
[91] B. Schierwater, M. Eitel, Placozoa, in: A. Wanninger (Ed.), Evolutionary Developmental Biology of Invertebrates 1, Springer, Vienna, 2015, pp. 107–114.
[92] B. Schierwater, D. de Jong, R. DeSalle, Placozoa and the evolution of Metazoa and intrasomatic cell differentiation, Int. J. Biochem. Cell Biol. 41 (2009) 370–379, https://doi.org/10.1016/j.biocel.2008.09.023.
[93] P.H. Schiffer, P. Natsidis, D.J. Leite, H. Robertson, F. Lapraz, F. Marlétaz, B. Fromm, L. Baudry, F. Simpson, E. Høye, A.-C. Zakrzewski, P. Kapli, K.J. Hoff, S. Mueller, M. Marbouty, H. Marlow, R.R. Copley, R. Koszul, P. Sarkies, M. J. Telford, The slow evolving genome of the xenacoelomorph worm Xenoturbella bocki, bioRxiv (2022), https://doi.org/10.1101/2022.06.24.497508.
[94] E.E. Schwager, P.P. Sharma, T. Clarke, D.J. Leite, T. Wierschin, M. Pechmann, Y. Akiyama-Oda, L. Esposito, J. Bechsgaard, T. Bilde, A.D. Buffry, H. Chao, H. Dinh, H. Doddapaneni, S. Dugan, C. Eibner, C.G. Extavour, P. Funch, J. Garb, L.B. Gonzalez, V.L. Gonzalez, S. Griffiths-Jones, Y. Han, C. Hayashi, M. Hilbran, D.S.T. Hughes, R. Janssen, S.L. Lee, I. Maeso, S.C. Murali, D.M. Muzny, R. Nunes da Fonseca, C.L.B. Paese, J. Qu, M. Ronshaugen, C. Schomburg, A. Schönauer, A. Stollewerk, M. Torres-Oliva, N. Turetzek, B. Vanthournout, J.H. Werren, C. Wolff, K.C. Worley, G. Bucher, R.A. Gibbs, J. Coddington, H. Oda, M. Stanke, N.A. Ayoub, N.-M. Prpic, J.-F. Flot, N. Posnien, S. Richards, A.P. McGregor, The house spider genome reveals an ancient whole-genome duplication during arachnid evolution, BMC Biol. 15 (2017) 62, https://doi.org/10.1186/s12915-017-0399-x.
[95] H.C. Seo, R.B. Edvardsen, A.D. Maeland, M. Bjordal, M.F. Jensen, A. Hansen, M. Flaat, J. Weissenbach, H. Lehrach, P. Wincker, R. Reinhardt, Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica, Nature 431 (2004) 67–71, https://doi.org/10.1038/nature02709.
[96] J.M. Serano, A. Martin, D.M. Liubicich, E. Jarvis, H.S. Bruce, K. La, W.E. Browne, J. Grimwood, N.H. Patel, Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiensis, Dev. Biol. 409 (2016) 297–309.
[97] P.P. Sharma, E.E. Schwager, C.G. Extavour, W.C. Wheeler, Hox gene duplications correlate with posterior heteronomy in scorpions, Proc. R. Soc. B 281 (2014) 20140661, https://doi.org/10.1098/rspb.2014.0661.
[98] O. Simakov, F. Marletaz, S.-J. Cho, E. Edsinger-Gonzales, P. Havlak, U. Hellsten, D.-H. Kuo, T. Larsson, J. Lv, D. Arendt, R. Savage, K. Osoegawa, P. De Jong, J. Grimwood, J.A. Chapman, H. Shapiro, A. Aerts, R.P. Otillar, A.Y. Terry, J. L. Boore, I.V. Grigoriev, D.R. Lindberg, E.C. Seaver, D.A. Weisblat, N.H. Putnam, D.S.O. Rokhsar, Insights into bilaterian evolution from three spiralian genomes, Nature 493 (2013) 526–531, https://doi.org/10.1038/nature11696.
[99] N.P. Singh, R. Krumlauf, Diversification and functional evolution of HOX proteins, Front. Cell Dev. Biol. 10 (2022), 798812, https://doi.org/10.3389/fcell.2022.798812.
[100] E. Sober, Reconstructing the Past: Parsimony, Evolution, and Inference, MIT Press, Cambridge (Mass), London, 1988.
[101] R.J. Sommer, Evolutionary changes of developmental mechanisms in the absence of cell lineage alterations during vulva formation in the Diplogastridae (Nematoda), 243-225, Development 124 (1997), https://doi.org/10.1242/dev.124.1.243.
[102] R.J. Sommer, Evolution of regulatory networks: nematode vulva induction as an example of developmental systems drift, in: O.S. Soyer (Ed.), Evolutionary Systems Biology, Advances in Experimental Medicine and Biology, 79–91, Springer, New York, 2012, p. 4, https://doi.org/10.1007/978-1-4614-3567-9.
[103] M.E. Svensson, Homology and homocracy revisited: gene expression patterns and hypotheses of homology, Dev. Genes Evol. 214 (2004) 418–421, https://doi.org/10.1007/s00427-004-0416-2.
[104] M. Thomas-Chollier, P. Martinez, Origin of metazoan patterning systems and the role of ANTP-class homeobox genes, eLS Evol. Divers. Life (2016), https://doi.org/10.1002/9780470015902.a0022852.pub2.
[105] J.R. True, E.S. Haag, Developmental system drift and flexibility in evolutionary trajectories, Evol. Dev. 3 (2001) 109–119, https://doi.org/10.1046/j.1525-142x.2001.003002109.x.
[106] G.P. Wagner, The developmental genetics of homology, Nat. Rev. Genet. 8 (2007) 473–479, https://doi.org/10.1038/nrg2099.
[107] G.P. Wagner, Homology in the age of developmental genomics, in: A. Wanninger (Ed.), Evolutionary Developmental Biology of Invertebrates 1, Springer, Vienna, 2015, pp. 25–44.
[108] G.P. Wagner, C. Amemiya, F. Ruddle, Hox cluster duplications and the opportunity for evolutionary novelties, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 14603–14606, https://doi.org/10.1073/pnas.2536656100.
[109] W. Wang, R.J. Sommer, Antagonism of LIN-17/Frizzled and LIN-18/Ryk in nematode vulva induction reveals evolutionary alterations in core developmental pathways, PLoS Biol. 9 (2011), e1001110, https://doi.org/10.1371/journal.pbio.
[110] A. Wanninger, Morphology is dead – long livemorphology! Integrating MorphoEvoDevo into molecular EvoDevo and phylogenomics, Front. Ecol. Evol. 3 (2015) 54, https://doi.org/10.3389/fevo.2015.00054.
[111] A. Wanninger, T. Wollesen, Mollusca, in: A. Wanninger (Ed.), Evolutionary Developmental Biology of Invertebrates 2, Springer, Vienna, 2015, pp. 103–153.
[112] A. Wanninger, T. Wollesen, The evolution of molluscs, Biol. Rev. 94 (2019) 102–115, https://doi.org/10.1111/brv.12439.
[113] N.V. Whelan, K.M. Kocot, L.L. Moroz, K.M. Halanych, Error, signal, and the placement of Ctenophora sister to all other animals, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 5773–5778, https://doi.org/10.1073/pnas.1503453112.
[114] N.V. Whelan, K.M. Kocot, T.P. Moroz, K. Mukherjee, P. Williams, G. Paulay, L. L. Moroz, K.M. Halanych, Ctenophore relationships and their placement as the sister group to all other animals, Nat. Ecol. Evol. 1 (2017) 1737–1746, https://doi.org/10.1038/s41559-017-0331-3.
[115] T. Wollesen, M. Scherholz, S.V. Rodríguez Monje, E. Redl, C. Todt, A. Wanninger, Brain regionalization genes are co-opted into shell field patterning in Mollusca, Sci. Rep. 7 (2017) 5486, https://doi.org/10.1038/s41598-017-05605-5.
[116] T. Zhao, A. Zwaenepoel, J.Y. Xue, S.M. Kao, Z. Li, M.E. Schranz, Y. Van de Peer, Whole-genome microsynteny-based phylogeny of angiosperms, in: Nat. Commun., 12, 2021, p. 3498, https://doi.org/10.1038/s41467-021-23665-0.

Seminars in Cell and Developmental Biology 152-153 (2024) 93–100

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Homeodomain complex formation and biomolecular condensates in Hox gene regulation

Joseph Salomone [a,b], Edward Farrow [a,b], Brian Gebelein [c,d,*]

[a] Graduate Program in Molecular and Developmental Biology, Cincinnati Children's Hospital Research Foundation, Cincinnati, OH 45229, USA
[b] Medical-Scientist Training Program, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA
[c] Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, MLC 7007, Cincinnati, OH 45229, USA
[d] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA

ARTICLE INFO

Keywords:
Hox
Transcription factor
Low-affinity binding sites
Biomolecular condensates
Short linear interaction motifs
Intrinsically disordered regions

ABSTRACT

Hox genes are a family of homeodomain transcription factors that regulate specialized morphological structures along the anterior-posterior axis of metazoans. Over the past few decades, researchers have focused on defining how Hox factors with similar in vitro DNA binding activities achieve sufficient target specificity to regulate distinct cell fates in vivo. In this review, we highlight how protein interactions with other transcription factors, many of which are also homeodomain proteins, result in the formation of transcription factor complexes with enhanced DNA binding specificity. These findings suggest that Hox-regulated enhancers utilize distinct combinations of homeodomain binding sites, many of which are low-affinity, to recruit specific Hox complexes. However, low-affinity sites can only yield reproducible responses with high transcription factor concentrations. To overcome this limitation, recent studies revealed how transcription factors, including Hox factors, use intrinsically disordered domains (IDRs) to form biomolecular condensates that increase protein concentrations. Moreover, Hox factors with altered IDRs have been associated with altered transcriptional activity and human disease states, demonstrating the importance of IDRs in mediating essential Hox output. Collectively, these studies highlight how Hox factors use their DNA binding domains, protein-protein interaction domains, and IDRs to form specific transcription factor complexes that yield accurate gene expression.

1. Introduction

The highly conserved Hox gene family has long fascinated geneticists, developmental biologists, and evolutionary biologists [1–3]. Early studies in *Drosophila melanogaster* revealed a series of linked genetic mutations that resulted in dramatic homeotic transformations of appendages and segments along the anterior-posterior (A-P) axis [4,5]. Subsequent molecular studies showed that each *Drosophila* Hox gene has a highly similar 180 nucleotide signature sequence called the homeobox that encodes a 60 amino acid homeodomain that binds DNA in a sequence-specific manner [6–8]. Hox genes are conserved across metazoans and often found clustered along the chromosome, albeit the number of Hox genes frequently differs between animal species due to either individual gene duplication or gene loss within a Hox cluster or the duplication/loss of an entire Hox gene cluster (Fig. 1A) [9,10]. Moreover, the expression of Hox genes along the A-P axis generally correlates with their order found within the cluster, and Hox genes have been categorized into distinct groups based on their role in anterior versus central versus posterior fate specification (Fig. 1A) [11]. Advances in gene targeting and genomic editing technologies led to an explosion in studies focused on altering Hox gene function in numerous different animal species [10,12,13]. Cumulatively, studies using model organisms as well as other species have consistently demonstrated how changes in Hox gene number and expression patterns along the developing A-P axis can instruct the formation of distinct cell fates and specialized morphological structures.

Given their central role in both the development and evolution of distinct morphological structures, a great deal of effort has been spent trying to understand how individual Hox genes accurately control the gene regulatory networks required for specific cell fates. However, biochemical studies revealed that in sharp contrast to their ability to direct distinct in vivo fates, the Hox homeodomain proteins largely bind

* Corresponding author at: Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, MLC 7007, Cincinnati, OH 45229, USA.
E-mail address: Brian.Gebelein@cchmc.org (B. Gebelein).

https://doi.org/10.1016/j.semcdb.2022.11.016
Received 30 July 2022; Received in revised form 21 October 2022; Accepted 30 November 2022
Available online 12 December 2022

the same AT-rich DNA sequences in vitro (Fig. 1B) [14–17]. This problem is further exacerbated by the fact that metazoans encode many additional non-Hox homeodomain transcription factors that bind highly similar AT-rich DNA sequences (Fig. 1B). For example, the human genome encodes approximately 200 homeodomain transcription factors, of which only 39 are the classic Hox genes that specify fates along the A-P axis [18,19]. Taken together, these findings present a paradox: How can a set of transcription factors with highly similar DNA binding activities achieve sufficient target gene specificity to direct the development of fundamentally different cell fates and morphological structures?

While many mechanisms are likely to contribute to Hox specificity, including post-translational modifications [20,21], binding to coactivator and corepressor proteins [20,22], and cellular chromatin status [23], this review focuses on the mechanisms that alter and enhance Hox DNA binding specificity. Over the past three decades, molecular studies have revealed that in addition to the conserved homeodomain that directly contacts DNA, Hox factors contain different combinations of short linear interaction motifs (SLiM) that bind other proteins [24–26] and intrinsically disordered regions (IDRs) that alter the biophysical properties of Hox proteins (Fig. 1C). Among the Hox SLiMs identified to date, several have been shown to specifically interact with additional transcription factors, especially other homeodomain proteins, and biochemical and genomic studies have revealed how homeodomain transcription factor complexes increase Hox DNA binding specificity [27]. While less is known about the role of IDRs in Hox function, recent work on IDRs has demonstrated a role in localizing proteins to nuclear sub-compartments [28,29], providing a new mechanism for how Hox proteins can be concentrated to increase binding to low-affinity binding sites within target gene *cis*-regulatory modules (CRMs). The distinct combinations of SLiMs and IDRs within Hox proteins provides key insight into how Hox proteins may gain sufficient regulatory specificity during development (Fig. 1C). Thus, this review will focus on the molecular mechanisms that target Hox proteins to specific DNA sequences during development.

2. Hox complex formation with the Pbx and Meis homeodomain proteins

While we still do not have a complete answer to how Hox transcription factors achieve accurate DNA target specificity, several key concepts have begun to emerge. Chief among these mechanisms is that Hox factors use SLiMs to form a variety of different transcription factor complexes on DNA with the Pbx (Extradenticle (Exd) in *Drosophila*) and Meis (Homothorax (Hth) in *Drosophila*) families of transcription factors [30]. Pbx and Meis are highly conserved members of the Three Amino Acid Loop (TALE) homeodomain family of proteins. Molecularly, the Pbx and Meis factors can interact with each other through highly conserved N-terminal domains, as well as with Hox factors to form larger transcription factor complexes on DNA. However, it is important to note, Pbx and Meis factors also have Hox-independent functions and can also interact with additional non-Hox transcription factors [31].

The role of the Pbx and Meis factors as Hox cofactors has been well described in the literature with several excellent reviews highlighting the many aspects of how Pbx and Meis impact Hox function [23,27,32,33]. Here, we focus on four mechanisms that reveal how Pbx and Meis enhance Hox transcription factor DNA binding specificity. First, the Pbx TALE homeodomain directly interacts with a short hexapeptide motif (typically containing a YPWM core) found N-terminal to the homeodomain of Hox factors [30]. This interaction, which is thought to be relatively weak off DNA, results in the formation of cooperative Pbx/Hox complexes on DNA through adjacent Pbx/Hox binding sites (Fig. 2A). Importantly, the added DNA sequence requirements needed to bind both Pbx and Hox factors as well as the direct protein-protein interactions enhance the DNA binding affinity and specificity relative to other homeodomain factors (Fig. 2A). For example, while Hox factors bind similar monomer DNA sites as many other homeodomain factors, such as Gsx2 (Fig. 1B), Gsx2 does not have a hexapeptide interaction motif, has not been shown to form complexes with Pbx, and does not enrich for Pbx/Hox binding sites in genomic assays [14–16,34]. In contrast, Hox factors have either a well-defined or atypical hexapeptide motif that functions as a SLiM to bind Pbx and thereby form

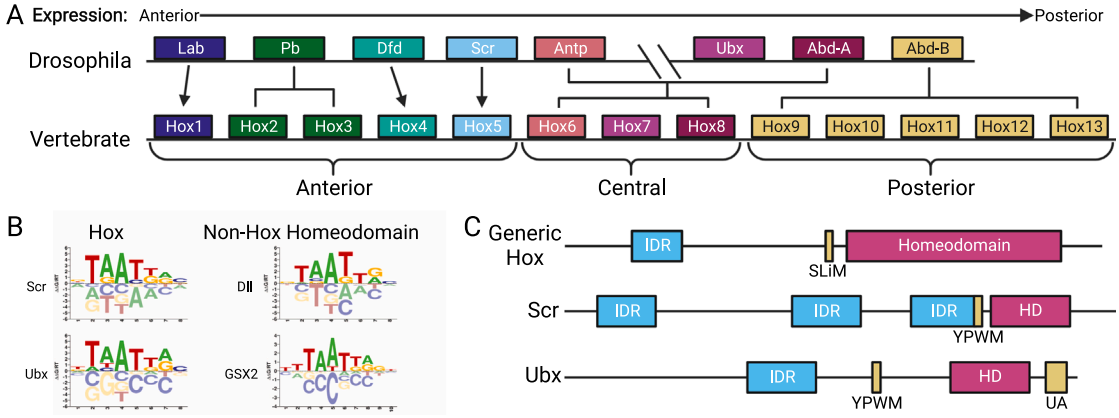


Fig. 1. : The clustered Hox genes encode conserved homeodomain transcription factors with characteristic protein regions across phyla. (A) Schematic showing the stereotypic chromosomal gene organization of Hox factors with the order within the cluster tied to their expression along the anterior-posterior axis. *Drosophila* contains a split chromosomal cluster on the same chromosome that separates the Hox genes into the Antennapedia cluster consisting of Lab, Pb, Dfd, Scr and Antp and the Bithorax cluster containing Ubx, Abd-A, and Abd-B. Vertebrates typically contain several Hox clusters that often lack a subset of the Hox genes due to gene loss. Here, we show a single vertebrate Hox cluster with a full complement of genes numbered from Hox1 to Hox13. Color coding denotes the approximate developmental relationship from ancestral Hox cluster, including duplication events of various genes, and the Hox genes can be broadly assigned into anterior, central, and posterior groups based on their expression and role in anterior-posterior patterning. (**B**) The clustered Hox genes and many non-Hox homeodomain proteins bind highly similar AT-rich monomer sequences in vitro. Representative PWMs of monomer Hox motifs for Scr and Ubx and the non-Hox homeodomain binding motifs for Distal-less (Dll) and GSX2 were taken from SELEX experiments uploaded to CisBP (http://cisbp.ccbr.utoronto.ca). (**C**) Schematics of a generic Hox protein and two *Drosophila* Hox factors with the homeodomain, short linear interaction motifs (SLiMs) and intrinsically disordered regions highlighted. Hox factors use various combinations of these sequences (i.e. Scr vs. Ubx) for specificity. IDRs in Scr and Ubx are annotated based on predicted regions on UniProt. *Lab = Labial; Pb = Proboscopedia; Dfd = Deformed; Scr; Sex Combs Reduced; Antp = Antennapedia; Ubx = Ultrabithorax; Abd-A = Abdominal-A; Abd-B = Abdominal-B; Dll = Distalless; GSX2 = GS Homeobox 2; IDR = Intrinsically Disordered Region; SLiM = Short Linear Interaction Motif; HD = Homeodomain; YPWM = hexapeptide motif; UA = the conserved UbdA motif.* Created with BioRender.com.

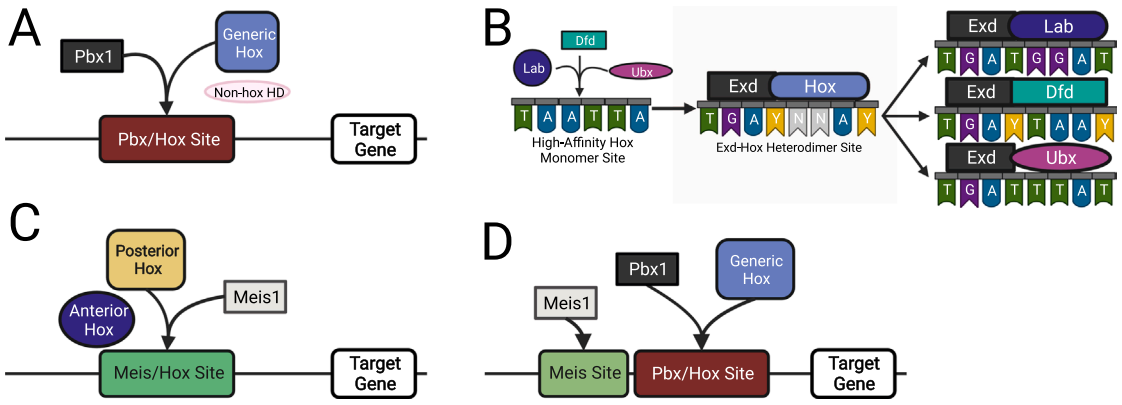


Fig. 2. Mechanisms of Hox factor DNA binding specificity in complex with the Pbx and Meis TALE homeodomain proteins. (**A**) All Hox factors can interact with Pbx/Exd to increase DNA binding sequence specificity and affinity. In contrast, most non-Hox homeodomains do not similarly interact with Pbx on DNA. (**B**) Hox factors largely bind similar monomer sites in vitro, but latent specificity differences are accentuated through interaction with Exd. Preferred Exd/Hox heterodimer sequences were defined by Slattery et al. [41] (**C**) Posterior Hox factors interact with Meis/Hth proteins to promote differential sequence recognition between anterior and posterior Hox factors. (**D**) The most common Hox multiplex organization uses nearby Meis binding site and adjacent Pbx/Hox heterodimer site sto increase binding specificity of Hox factors at target gene CRMs. Note, flexible configurations of these motifs are possible with distinct combinations of homeodomain binding sites potentially contributing to DNA binding specificity. *Pbx1 = PBX Homeobox 1; Exd = Extradenticle; Meis1= Meis Homeobox 1*. Created with BioRender.com.

heterodimers on DNA [30] (Fig. 2A). Highlighting the importance of the hexapeptide motif for Pbx/Hox genomic binding, a ChIP-seq study for the Sex combs reduced (Scr) Hox factor in *Drosophila* revealed that flies engineered with a Scr-YPWM mutation were homozygous lethal and only enriched for monomer Hox sites, whereas wild-type Scr strongly enriched for Exd/Hox dimer sequences [35]. Intriguingly, a subset of Hox factors contain more than one SLiM sequence that directly interacts with Pbx/Exd. For example, the Ultrabithorax (Ubx) and Abdominal-A (Abd-A) Hox factors contain both the generic hexapeptide sequence as well as a UbdA motif (UA motif, Fig. 1C) that can also mediate interactions with Exd [36–38]. Moreover, a subset of vertebrate Hox factors similarly contain multiple Pbx interacting motifs [39,40]. While current studies favor the idea that Hox factors with multiple Pbx/Exd interaction domains function redundantly to promote Hox heterodimer binding, many of these studies have relied upon ectopic expression assays and only assessed DNA binding to specific CRM sites. Hence, genomic studies using animals with specific Hox factor SLiM mutations in endogenous loci and unbiased SELEX-seq assays are needed to determine if the Hox factors use distinct Pbx/Exd interacting peptides to alter and/or diversify the sequence binding selectivity of the heterodimer complex on DNA.

Second, while all Hox factors can bind DNA as both monomers and with Pbx/Exd, heterodimer binding was found to accentuate differences in DNA binding site preferences between Hox factors (Fig. 2B). This concept, which is called latent specificity, was revealed when Slattery et al. performed a comprehensive unbiased DNA binding site selection assay (SELEX-seq) for eight *Drosophila* Hox factors as monomers and in combination with Exd [41]. Importantly, comparative analysis of the bound monomer versus Exd/Hox dimer sequences revealed that the differences in binding preferences between Hox factors increased when in complex with Exd relative to monomer binding alone (Fig. 2B). To better understand how Exd/Hox interactions invoke latent specificity, computational analysis based on DNA shape revealed that when in complex with Exd, the anterior Hox factors prefer DNA sites with narrower minor groove widths relative to the central/posterior Hox factors [41,42]. This computational prediction was supported by prior structural, biochemical, and genetic studies showing that Scr utilizes conserved amino acid residues in the N-terminal arm of its homeodomain to make specific contacts to DNA sequences with a narrow minor groove [43,44]. To formally test this idea, SELEX-seq and reporter assays using Scr proteins containing specific mutations in the N-terminal arm residues that contact the narrow minor groove demonstrated how paralog-specific residues are required for Exd/Scr binding to sequences with this DNA shape feature [42]. Moreover, selectively inserting these Scr residues into the N-terminal arm sequences of Antennapedia (Antp), a more posterior Hox factor that does not efficiently bind DNA with a narrow minor groove, was shown to be sufficient to confer this DNA binding site preference in vitro and target gene expression activity in vivo [42]. Thus, these studies demonstrate how Pbx/Hox heterodimer complexes can further differentiate Hox DNA binding activities relative to Hox monomer binding.

Third, the Meis/Hth TALE factors form complexes with a subset of Hox factors on DNA (Fig. 2C). While the molecular mechanisms and the Hox sequences underlying the interactions between Meis and Hox factors are not well understood, DNA binding assays revealed that Meis cooperatively binds adjacent Meis/Hox sites with vertebrate members of the posterior Hox9–13 class of factors [45]. Consistent with these findings, ChIP-seq studies for MEIS1 and HOXB13 in prostate cells revealed enrichment of adjacent MEIS and HOX sites at many genomic loci [46]. In contrast, the Hoxb4, Hoxb6, Hoxa7, and Hoxb8 anterior/central Hox factors failed to similarly bind such Meis/Hox sites in vitro, suggesting Meis factors selectively bind heterodimer DNA sites with only posterior Hox factors (Fig. 2C)[45]. In *Drosophila*, Hth was also shown to complex with a subset of Hox factors on adjacent binding sites to mediate both gene repression and activation. In particular, Hth cooperatively binds adjacent Hth/Hox DNA sites with Ubx and Abd-A to repress *Distal-less* (*Dll*) expression and leg development in abdominal segments [47,48]. In addition, Hth/Abd-A cooperatively bind and activate a CRM that ultimately promotes EGF signaling via similar adjacent Hth/Hox sites [49,50]. Interestingly, Ubx and Abd-A are members of the central Hox class (Fig. 1A), and the vertebrate central Hox factors (Hoxb6, Hoxa7, and Hoxb8) do not cooperatively bind with vertebrate Meis in DNA binding assays [45]. In addition, Antp, which is also a member of the *Drosophila* central Hox factors, failed to regulate these targets through Hth/Hox binding sites [49–51]. Hence, future studies are needed to determine which Hox factors are capable of cooperatively binding adjacent sites with Meis/Hth and to define the paralog-specific mechanisms underlying the formation of these complexes.

Fourth, Pbx, Meis, and Hox factors can form larger transcription factor complexes on DNA (Fig. 2D). Since Pbx and Meis interact via N-terminal domains separate from their C-terminal homeodomains [52], the Pbx and Meis binding sites can be found at variable distances and orientations relative to each other [27]. In complex with Hox factors, the most common trimeric protein complex contains adjacent Pbx/Hox sites

with a more distant Meis binding site [20]. Such complexes and motif enrichment profiles have been observed in both vertebrate and *Drosophila* studies [53–56]. Moreover, a SELEX-seq assay using *Drosophila* proteins demonstrated that each Hth binding site orientation was permissible relative to the Exd/Hox sites, although one orientation preferred a shorter spacer distance between the Hth and Exd/Hox sites than the other orientation [57]. Additional studies in *Drosophila* revealed that even larger Hox complexes can be observed as Hth/Hox and Exd/Hox sites separated by 7 bps can mediate the formation of functional tetrameric complexes [47,51]. Thus, these studies highlight how distinct combinations of Hox, Pbx/Exd, and Meis/Hth sites can be utilized to recruit larger homeodomain transcription factor complexes that mediate specific transcriptional outcomes.

2.1. Hox complex formation with other homeodomain transcription factors

While genetic, biochemical, and genomic studies strongly support the idea that the Pbx and Meis factors are widespread Hox co-factor proteins, studies have begun to uncover that additional transcription factors contribute to Hox DNA binding specificity. For example, the Engrailed (En) homeodomain protein was shown to cooperatively bind DNA with the Ubx and Abd-A Hox factors but not the Antp Hox factor [47]. Intriguingly, subsequent molecular studies revealed that Abd-A requires its UbdA motif to mediate interactions with En, suggesting that this motif may participate in binding both the Pbx and En homeodomain proteins [36]. In addition to En, recent ChIP-seq and biochemical pull-down studies have shown that the Distal-less (Dll) homeodomain factor may selectively function as a Hox co-factor protein with Scr during leg development. In particular, Feng at al found that Scr and Dll bind to many of the same genomic sequences containing two TAATTA sequences spaced 3 bp apart in the leg imaginal disc in the first thoracic segment (T1) [35]. Intriguingly, while Scr is expressed throughout the leg imaginal disc, the Exd/Hth factors and the Dll factors are expressed in largely mutually exclusive domains along the proximal-distal axis. Hence, these studies suggest that Scr forms distinct transcription factor complexes with different homeodomain factors to regulate distinct target genes during T1 leg development [35]. Lastly, additional bimolecular fluorescent complementation (BiFC) assays and mass-spectrometry studies have revealed that Hox factors can interact with many, perhaps even hundreds, of different transcription factors, including many homeodomain proteins [26,58,59]. While the mechanisms and impact of these potential interactions on DNA binding are largely unknown, these studies suggest that the formation of larger Hox transcription factor complexes with different partner proteins may be a critical, widespread mechanism underlying how Hox factors achieve target gene specificity.

2.2. The role of low-affinity binding sites in Hox specificity

The above studies support a model that Hox factors use SLiMs to form complexes with other homeodomain transcription factors to increase DNA binding specificity, suggesting that CRMs use different combinations of AT-rich DNA sequences to recruit specific Hox complexes. Intriguingly, studies on several known Hox-regulated CRMs have revealed an additional concept: Binding sites that yield Hox-specific output often contain clusters of sub-optimal or low-affinity sites (Fig. 3A). In fact, recent evidence suggests that low-affinity sites are generally important for proper CRM function, as they promote target specificity between related transcription factors, and increasing the binding affinity of such sites can result in CRM mis-regulation and developmental defects [60–67]. One of the best examples of these principles are the *shaven-baby* (*svb*) CRMs that mediate Ubx-specific activation using multiple low-affinity Exd/Hox sites to regulate gene expression required for trichome patterns in *Drosophila* [62,68]. Importantly, changing these sequences from low- to high-affinity resulted in the mis-regulation of *svb* CRM activity into additional segments in a Hox-dependent manner, supporting the idea that high-affinity sites are likely to be bound and regulated by many Hox factors, and thus are less specific to individual Hox factors (Fig. 3B) [62,69]. Consistent with this idea, a general computational approach called No Read Left Behind (NRLB) demonstrated a general inverse relationship between Hox/Exd binding affinity and Hox/Exd binding specificity [69]. Moreover, several other CRM studies similarly reveal how low-affinity sites can result in cell- and or segment-specific output, whereas high-affinity sites result in altered and/or expanded patterns of gene expression [50,70–72].

Relying upon low-affinity binding sites to mediate cell-specific gene regulatory outcomes that are essential for proper development comes with a caveat: DNA-transcription factor binding events are often transient with DNA binding and dissociation occurring over a relatively short time-period [73–75]. Hence, high transcription factor concentrations are required to reproducibly bind and regulate CRMs using low-affinity sites. However, transcription factors are frequently expressed at relatively low cellular levels, raising the question of how low-affinity CRM binding sites can reproducibly recruit the required transcription factors and cofactors to function efficiently? High-resolution imaging methods have begun to answer this question. For example, further investigation of the *svb* CRMs showed that Ubx is not uniformly distributed across the nucleus (Fig. 3C) [76]. Rather, Ubx is locally concentrated in specific nuclear subregions, and these regions of high Ubx concentration correspond to regions where the Hox cofactor Hth is concentrated. Tsai et al. further found that Ubx and Hth concentrations are high at the sites of active transcription of the *svb* locus [76]. In fact, *svb* CRMs on homologous chromosomes as well as

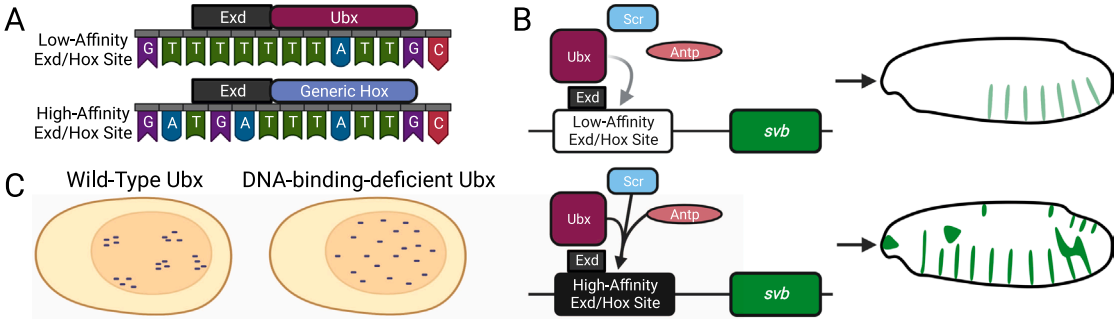


Fig. 3. The role of low-affinity sites and Hox transcriptional specificity. (**A**) Sequence differences between low- and high-affinity Exd/Hox sites promote the binding of specific versus pan Hox factors. Note, this example highlights a Ubx-specific low affinity Exd/Hox regulatory element whereas high affinity sites were found to be regulated by many Hox factors. Sequences were defined and tested by Rastogi et al. [69] (**B**) Presence of low-affinity Exd/Hox sites in vivo directs select expression of the svb target gene by Ubx. Converting the site into a high-affinity motif resulted in the mis-regulation of svb CRM activity due to more generalized Hox activation [62,69].(**C**) The wild-type Ubx proteins was found to be highly concentrated at transcriptional hubs in Drosophila nuclei whereas a DNA binding deficient Ubx protein was diffusely localized throughout the nucleus. *svb* = *shavenbaby* [76,77]. Created with BioRender.com.

transgenic *svb* CRMs reporters inserted on different chromosomes colocalized with one another at sites of high Ubx concentration [76]. High Ubx concentrations located in close proximity to the *svb* CRMs with clustered low-affinity sites may provide an avidity effect, whereby Ubx molecules that dissociate from one site would likely have an increased probability to bind other nearby low-affinity sites. Thus, CRMs with clustered sites are likely to play a key role in the formation of "transcriptional hubs" that yield cell-specific outputs.

Several lines of evidence suggest that the formation of these active transcriptional hubs on the *svb* CRMs are directly dependent on Ubx interactions with binding sites. First, analysis of flies expressing a DNA binding deficient Ubx protein shows a uniform nuclear distribution of the transcription factor in contrast to the regions of high Ubx and Hth concentration seen with the wild-type factor (Fig. 3C) [76]. Furthermore, when a redundant Ubx-regulated CRM from the *svb* locus was deleted, Ubx concentrations around the actively transcribed region were significantly decreased [77]. This resulted in overall decreased transcription from the *svb* locus, and increased sensitivity to stress with phenotypic defects appearing at increased temperatures [77]. Finally, insertion of another copy of the *svb* CRM onto a separate chromosome not only rescued the phenotypic effects of the deletion, but also resulted in increased local Ubx concentrations and transcription at the *svb* locus [77]. These data indicate that CRMs and gene loci on separate chromosomes can function cooperatively to promote formation of nuclear regions with increased transcription factor and cofactor concentration. These findings raise a new biological question: What are the mechanisms underlying how specific transcription factors are concentrated within specific nuclear sub-compartments?

3. Biomolecular condensates, IDRs, and transcriptional regulation

Increasingly, studies have revealed that the nucleus is highly heterogeneous, with transcription factors, such as the Hox factors, often found in discrete, concentrated nuclear sub-compartments (Fig. 4). Intriguingly, biophysical and cellular studies have led to the identification, characterization, and functional analysis of membraneless nuclear sub-compartments with particular emphasis on the notion that liquid-liquid phase separation can concentrate proteins including transcription factors and the transcriptional machinery and thereby define transcriptionally active chromatin [78–83]. The observation of these subdomains is intriguing because it suggests a rather intuitive model for transcriptional regulation, where nuclear sub-compartments enriched for transcription factors, the mediator complex, and RNA polymerase II can interact with specific chromatin regions (promoters and distal enhancers).

Though we focus on the implications of nuclear condensates on the regulation of transcription, and particularly on how they impact the function of Hox transcription factors, it should be noted that the identification of membraneless compartments that facilitate compartmentalization and concentration of components required for specific cellular functions is not new. The role of the nucleolus in ribosomal formation is a prominent example. In fact, these compartments have been implicated in a variety of nuclear processes including heterochromatin formation, DNA replication, DNA repair, transcriptional regulation, and RNA processing [83–87]. Before moving forward, it must also be addressed that just as there are many different types of sub-compartments found throughout the cell, there are multiple different and sometime conflicting terms used to describe them. Recent studies, especially those focused on transcriptionally active sub-compartments, have used the term biomolecular condensates; however, other terminology includes chromatin sub-compartments, membraneless organelles, nuclear bodies, transcriptional factories, and transcriptional hubs [81,83]. Some debate exists on the precise definition of these terms, for example some use the term condensate to generally refer to all such sub-compartments, while other groups hold that condensates imply the presence of liquid-liquid phase separation. For the purposes of this review, we will use the terminology defined by Sabari et al.: Namely, biomolecular condensates refer to a membraneless cellular compartment where specific biomolecules are concentrated and which are composed of higher-order assemblies of biomolecules held together by multiple, dynamic, weak interactions [83]. An overall key feature of these condensates is their capacity to alter biomolecule concentrations, such that biomolecules required for a particular process facilitate the formation of large multicomponent complexes by inclusion of necessary components and exclusion of others.

A general model has emerged from these studies in which a transcription factor's location within the nucleus is dictated by two types of domains; a highly structured DNA binding domain (DBD) and an often unstructured regulatory domain made of repetitive amino acid sequences referred to as intrinsically disordered regions (IDRs) [88,89]. In this model, the IDR functions to promote condensate formation, whereas the DBD directs condensate localization to specific genomic regions. Put another way, the sequence specificity of the DBD results in transcription factor binding to CRMs within the genome, whereas the IDRs from large numbers of transcription factor molecules and the transcriptional machinery weakly interact to form biomolecular condensates at specific genomic regions (Fig. 4) [88,90–92]. Lastly, histone modifications and

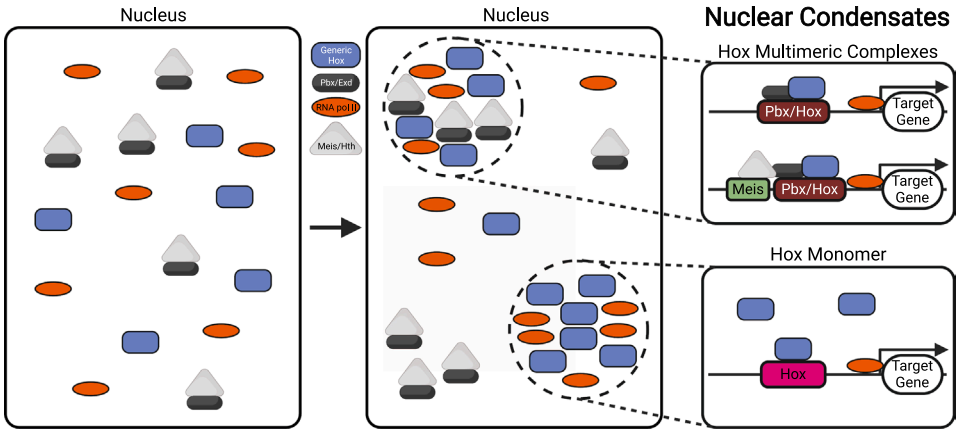


Fig. 4. The role of nuclear sub-compartments in transcriptional regulation. Schematic at left shows a graphic representation of a theoretical "homogenous" nucleus with proteins diffusely found throughout the nucleus. Schematic in middle shows a representation of a heterogenous nucleus based on high resolution imaging. The heterogenous nucleus contains distinct regions of concentrated proteins, often called biomolecular condensates, that typically include specific transcription factors and the transcriptional machinery (i.e. RNA polII, the mediator complex, etc). Close-up view of these nuclear condensates, which rely upon liquid-liquid phase separation of proteins due to weak protein-protein intermolecular interactions via IDRs, reveals a Hox multimeric condensate with Pbx/Exd and Meis/Hth and a Hox monomeric condensate. In this model, the high local protein concentrations increases the probability of Hox binding to CRMs, and thereby the activation of specific target genes within each nuclear sub-compartment. *RNA pol II = RNA polymerase II*; *Hth = Homothorax*. Created with BioRender.com.

the factors that "read" these marks also play key roles in condensate formation and specificity. For example, BRD4 interacts with acetylated nucleosomes and promotes formation of active condensates, while the factor HP1α binds nucleosomes with heterochromatin marks and drives formation of heterochromatin condensates [82,93].

The ability of transcription factors to form condensates via their IDRs has implications as a possible mechanism for achieving target specificity. As has been previously described, even high confidence information about a transcriptions factor's preferred DNA binding site is usually insufficient to predict in vivo targets. Further, related factors with nearly identical in vitro binding specificities often bind different targets when expressed in the same cell types. The in vivo and in vitro binding preferences of two yeast transcription factors that contain IDRs were analyzed, revealing the surprising contribution of the IDR to in vivo target selection [94]. The zinc finger transcription factor, Msn2, binds sites containing the 5′-AGGGG-3′ motif, which was confirmed in chromatin binding profiles. However, only a subset of promoters encoding that motif were bound by Msn2, and related transcription factors that bind the same core motif interact with a different subset of promoters. The importance of the DBD and the IDR of Msn2 were then tested by assaying the DNA binding profiles of a DBD-only and a DBD deficient protein. While the DBD-only construct bound regions enriched for the core AGGGG motif, the binding profile differed significantly from that of full-length Msn2. On the other hand, the DBD deficient construct localized to most of the same promoters as the full-length transcription factor, although enrichment of the AGGGG motif was no longer observed. Similar findings were also observed with the yeast transcription factor, Yap1 [94], highlighting the potential role of IDRs in targeting transcription factor proteins to specific nuclear sub-compartments.

3.1. *Hox transcription factors, IDRs, and condensate formation*

Many investigations of transcriptionally active biomolecular condensates focused on the key role played by IDRs in factors like MED1 and BRD4 in driving the formation of liquid-liquid phase separations and promoting transcriptional activation [82,88]. Transcription factor activation domains consisting of IDRs can take several different forms including acidic, Proline, Glutamine, or Serine/Threonine-rich regions [89,95,96]. The importance of IDR-containing activation domains for the function of Hox factors has been well established. For example, early functional domain mapping assays of the HOXD4 transcription factor revealed that in addition to its DNA binding domain, a proline-rich domain with characteristics of an IDR was necessary for transcriptional activation [97]. Interestingly, the Ubx Hox factor also contains IDRs that contribute to in vitro DNA binding affinity and its regulatory function [98,99]. However, to our knowledge, the ability of the Ubx IDRs to form biomolecular condensates and their role in regulating known Ubx target genes, such as *svb*, has yet to be tested.

Direct evidence that IDRs are not only essential for Hox gene function, but also contribute to biomolecular condensates and liquid-liquid phase separation comes from investigations into a set of inherited diseases in humans. These diseases comprise a group of more than 20 disorders that result from the expansion of short, repetitive DNA sequences encoding amino acid repeats, typically of alanine or glutamine [100–102]. 15 of these disease-associated repeats were found to be in nuclear proteins, most being transcription factors [100]. Of particular note, expansion of an alanine repeat in HOXD13 has been implicated in the hereditary limb malformation disorder, synpolydactyly [103]. More recent work using a synpolydactyly mouse model has shown that the polyalanine region in HOXD13 is part of an IDR, which can drive liquid-liquid phase separation, and the disease-associated expansion of alanine repeats enhanced phase separation and altered the composition of HOXD13 condensates [28]. More specifically, compared to wild-type HOXD13 condensates, those created by the HOXD13 variant with the expanded alanine repeats, less effectively recruited the mediator complex and RNA pol II. Thus, the disease-associated allele is less effective at activating transcription [28]. Consistent with these findings both wild-type and disease-associated HOXD13 proteins have very similar DNA binding profiles as defined by ChIP-seq assays, but their transcriptional profiles are clearly different [28]. Importantly, repeat expansions in other transcription factors implicated in disease, such as HOXA13, RUNX2, and TBP, were also shown to alter both phase separation and condensate composition [28]. Taken together, these findings suggest that pathological TFs with expanded IDRs share two key features: an enhanced capacity to undergo liquid-liquid phase separation and decreased target gene transcription.

Further evidence that altering Hox condensate formation contributes to human disease comes from investigation of cancer/tumorigenesis. It has been noted that many cancers have chromosomal translocations between genes that encode IDRs that are either associated with transcription factors or chromatin binding proteins. For example, a translocation that fuses the phenylalanine/glycine-rich IDR of a nucleoporin gene with one of several transcription factors has been implicated in leukemias [104–107]. Of note, a translocation that creates a fusion between the IDR from NUP98 with HOXA9 has been extensively studied in the context of leukemias [105,107]. The NUP98-HOXA9 fusion is not only required for leukemic transformation but also can drive the formation of liquid-liquid phase separations [29]. In fact, this NUP98-HOXA9 protein has increased chromatin occupancy compared to HOXA9 alone and produces a binding profile with broad characteristics suggestive of "super-enhancer" formation. Analysis of chromatin looping using Hi-C showed that the NUP98-HOXA9 fusion showed enrichment for key proto-oncogenes. Further experiments showed that a fusion protein created by replacing the nucleoporin IDR with an unrelated IDR could also drive liquid-liquid phase separated condensates and produce similar increases in DNA binding and target gene activation [29].

Despite a relatively small number of studies investigating biomolecular condensates and Hox transcription factors, there is clear emerging evidence that at least some Hox transcription factors can contribute to such condensates by promoting liquid-liquid phase separations via weak interactions between IDRs. Furthermore, the ability of Hox factors to participate in condensates has obvious implications for pathogenesis of human diseases.

4. Future directions

New evidence about the mechanisms by which Hox factors achieve specificity, regulate target genes, and both benefit from and participate in the formation of biomolecular condensates provides exciting new avenues for study, while also suggesting potential elegant solutions to long standing questions. It will be interesting to investigate whether regions of high Ubx concentration observed in *Drosophila* truly represent biomolecular condensates, whether they are mediated by IDRs within the Ubx protein or its cofactors, and to what extent these IDRs contribute to Hox target specificity. Further, it is interesting to consider that many Hox factors can function as both transcriptional activators and repressors. While this review focused mainly on the incorporation of Hox factors into condensates associated with activation and active transcription, it is worth noting that evidence of repressive condensates has also been described [108,109]. Investigation of Hox factors in active versus repressive condensate formation could provide further insight into how the factors achieve regulatory specificity on different targets.

Author contributions

The topic of this review was jointly conceived by J.S. and B.G. The initial draft of the manuscript was written by J.S. and B.G and edited by E.F. The figures were initially generated by E.F. and edited by B.G.

Funding

This work was supported by National Institutes of Health grants GM079428 and NS124660 to B.G. The funder had no role in the content and writing of this review.

Conflicts of Interest

None

Data availability

No data was used for the research described in the article.

Acknowledgments

We apologize to the researchers working on Hox factors that we could not include due to space requirements. We would like to thank Brittany Cain for her critical comments on the manuscript.

References

[1] W. McGinnis, R. Krumlauf, Homeobox genes and axial patterning, Cell 68 (2) (1992) 283–302.
[2] S.B. Carroll, Homeotic genes and the evolution of arthropods and chordates, Nature 376 (6540) (1995) 479–485.
[3] M.R. Capecchi, Hox genes and mammalian development, Cold Spring Harb. Symp. Quant. Biol. 62 (1997) 273–281.
[4] E.B. Lewis, A gene complex controlling segmentation in Drosophila, Nature 276 (5688) (1978) 565–570.
[5] E.B. Lewis, The 1991 Albert Lasker Medical Awards. Clusters of master control genes regulate the development of higher organisms, JAMA 267 (11) (1992) 1524–1531.
[6] W. McGinnis, et al., A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes, Nature 308 (5958) (1984) 428–433.
[7] M. Regulski, et al., Homeo box genes of the Antennapedia and bithorax complexes of Drosophila, Cell 43 (1) (1985) 71–80.
[8] W. McGinnis, et al., A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans, Cell 37 (2) (1984) 403–408.
[9] N. Hajirnis, R.K. Mishra, Homeotic genes: clustering, modularity, and diversity, Front Cell Dev. Biol. 9 (2021), 718308.
[10] N.P. Singh, R. Krumlauf, Diversification and functional evolution of HOX proteins, Front Cell Dev. Biol. 10 (2022), 798812.
[11] S.D. Hueber, et al., Improving Hox protein classification across the major model organisms, PLoS One 5 (5) (2010), e10820.
[12] Y. Tomoyasu, S.R. Wheeler, R.E. Denell, Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum, Nature 433 (7026) (2005) 643–647.
[13] D.R. Rux, D.M. Wellik, Hox genes in the adult skeleton: novel functions beyond embryonic development, Dev. Dyn. 246 (4) (2017) 310–317.
[14] M.B. Noyes, et al., Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites, Cell 133 (7) (2008) 1277–1289.
[15] M.F. Berger, et al., Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences, Cell 133 (7) (2008) 1266–1276.
[16] A. Jolma, et al., DNA-binding specificities of human transcription factors, Cell 152 (1–2) (2013) 327–339.
[17] K.R. Nitta, et al., Conservation of transcription factor binding specificities across 600 million years of bilateria evolution, Elife 4 (2015).
[18] T.R. Burglin, M. Affolter, Homeodomain proteins: an update, Chromosoma 125 (3) (2016) 497–521.
[19] S.A. Lambert, et al., The human transcription factors, Cell 172 (4) (2018) 650–665.
[20] A. Zandvakili, B. Gebelein, Mechanisms of specificity for hox factor activity, J. Dev. Biol. 4 (2) (2016).
[21] M. Primon, et al., Kinase Regulation of HOX Transcription Factors, Cancers (Basel) (4) (2019) **11**.
[22] R.S. Mann, K.M. Lelli, R. Joshi, Hox specificity unique roles for cofactors and collaborators, Curr. Top. Dev. Biol. 88 (2009) 63–101.
[23] B. Cain, B. Gebelein, Mechanisms underlying hox-mediated transcriptional outcomes, Front Cell Dev. Biol. 9 (2021), 787339.
[24] V. Neduva, R.B. Russell, Linear motifs: evolutionary interaction switches, FEBS Lett. 579 (15) (2005) 3342–3345.
[25] K. Van Roey, et al., Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation, Chem. Rev. 114 (13) (2014) 6733–6778.
[26] M. Baeza, et al., Inhibitory activities of short linear motifs underlie Hox interactome specificity in vivo, Elife 4 (2015).
[27] S. Merabet, R.S. Mann, To be specific or not: the critical relationship between Hox And TALE Proteins, Trends Genet 32 (6) (2016) 334–347.
[28] S. Basu, et al., Unblending of transcriptional condensates in human repeat expansion disease, Cell 181 (5) (2020), p. 1062-1079 e30.
[29] J.H. Ahn, et al., Phase separation drives aberrant chromatin looping and cancer development, Nature 595 (7868) (2021) 591–595.
[30] S. Merabet, et al., Classification of sequence signatures: a guide to Hox protein function, Bioessays 31 (5) (2009) 500–511.
[31] L. Selleri, V. Zappavigna, E. Ferretti, 'Building a perfect body': control of vertebrate organogenesis by PBX-dependent regulatory networks, Genes Dev. 33 (5–6) (2019) 258–275.
[32] C.B. Moens, L. Selleri, Hox cofactors in vertebrate development, Dev. Biol. 291 (2) (2006) 193–206.
[33] B. De Kumar, D.C. Darland, The Hox protein conundrum: The "specifics" of DNA binding for Hox proteins and their partners, Dev. Biol. 477 (2021) 284–292.
[34] J. Salomone, et al., Conserved Gsx2/Ind homeodomain monomer versus homodimer DNA binding defines regulatory outcomes in flies and mice, Genes Dev. 35 (1–2) (2021) 157–174.
[35] S. Feng, et al., Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors, Nat. Commun. 13 (1) (2022) 3808.
[36] K.M. Lelli, B. Noro, R.S. Mann, Variable motif utilization in homeotic selector (Hox)-cofactor complex formation controls specificity, Proc. Natl. Acad. Sci. USA 108 (52) (2011) 21122–21127.
[37] M. Saadaoui, et al., Selection of distinct Hox-Extradenticle interaction modes fine-tunes Hox protein activity, Proc. Natl. Acad. Sci. USA 108 (6) (2011) 2276–2281.
[38] N. Foos, et al., A flexible extension of the Drosophila ultrabithorax homeodomain defines a novel Hox/PBC interaction mode, Structure 23 (2) (2015) 270–279.
[39] A. Dard, et al., The human HOXA9 protein uses paralog-specific residues of the homeodomain to interact with TALE-class cofactors, Sci. Rep. 9 (1) (2019) 5664.
[40] A. Dard, et al., Human HOX Proteins Use Diverse and Context-Dependent Motifs to Interact with TALE Class Cofactors, Cell Rep. 22 (11) (2018) 3058–3071.
[41] M. Slattery, et al., Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins, Cell 147 (6) (2011) 1270–1282.
[42] N. Abe, et al., Deconvolving the recognition of DNA shape from sequence, Cell 161 (2) (2015) 307–318.
[43] R. Joshi, et al., Functional specificity of a Hox protein mediated by the recognition of minor groove structure, Cell 131 (3) (2007) 530–543.
[44] R. Joshi, L. Sun, R. Mann, Dissecting the functional specificities of two Hox proteins, Genes Dev. 24 (14) (2010) 1533–1545.
[45] W.F. Shen, et al., AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins, Mol. Cell Biol. 17 (11) (1997) 6448–6458.
[46] C. VanOpstall, et al., MEIS-mediated suppression of human prostate cancer growth and metastasis through HOXB13-dependent regulation of proteoglycans, Elife (2020) **9**.
[47] B. Gebelein, D.J. McKay, R.S. Mann, Direct integration of Hox and segmentation gene inputs during Drosophila development, Nature 431 (7009) (2004) 653–659.
[48] A. Zandvakili, et al., The cis-regulatory logic underlying abdominal Hox-mediated repression versus activation of regulatory elements in Drosophila, Dev. Biol. 445 (2) (2019) 226–236.
[49] D. Li-Kroeger, T.A. Cook, B. Gebelein, Integration of an abdominal Hox complex with Pax2 yields cell-specific EGF secretion from Drosophila sensory precursor cells, Development 139 (9) (2012) 1611–1619.
[50] D. Li-Kroeger, et al., Hox and senseless antagonism functions as a molecular switch to regulate EGF secretion in the Drosophila PNS, Dev. Cell 15 (2) (2008) 298–308.
[51] J.D. Uhl, A. Zandvakili, B. Gebelein, A Hox Transcription Factor Collective Binds a Highly Conserved Distal-less cis-Regulatory Module to Generate Robust Transcriptional Outcomes, PLoS Genet 12 (4) (2016), e1005981.
[52] G.E. Rieckhof, et al., Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein, Cell 91 (2) (1997) 171–183.
[53] B. De Kumar, et al., HOXA1 and TALE proteins display cross-regulatory interactions and form a combinatorial binding code on HOXA1 targets, Genome Res 27 (9) (2017) 1501–1512.
[54] H.D. Ryoo, et al., Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex, Development 126 (22) (1999) 5137–5148.
[55] Y. Jacobs, C.A. Schnabel, M.L. Cleary, Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity, Mol. Cell Biol. 19 (7) (1999) 5134–5142.
[56] B. Gebelein, et al., Specificity of Distalless repression and limb primordia development by abdominal Hox proteins, Dev. Cell 3 (4) (2002) 487–498.
[57] J.F. Kribelbauer, et al., Context-Dependent Gene Regulation by Homeodomain Transcription Factor Complexes Revealed by Shape-Readout Deficient Proteins, Mol. Cell 78 (1) (2020), p. 152-167 e11.
[58] J. Bischof, et al., Generation of a versatile BiFC ORFeome library for analyzing protein-protein interactions in live Drosophila, Elife 7 (2018).
[59] J. Carnesecchi, et al., Multi-level and lineage-specific interactomes of the Hox transcription factor Ubx contribute to its functional specificity, Nat. Commun. 11 (1) (2020) 1388.
[60] J. Crocker, N. Potter, A. Erives, Dynamic evolution of precise regulatory encodings creates the clustered site signature of enhancers, Nat. Commun. 1 (2010) 99.
[61] J. Crocker, E.P. Noon, D.L. Stern, The soft touch: low-affinity transcription factor binding sites in development and evolution, Curr. Top. Dev. Biol. 117 (2016) 455–469.

[62] J. Crocker, et al., Low affinity binding site clusters confer hox specificity and regulatory robustness, Cell 160 (1–2) (2015) 191–203.
[63] A.I. Ramos, S. Barolo, Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution, Philos. Trans. R. Soc. Lond. B Biol. Sci. 368 (1632) (2013), p. 20130018.
[64] J. Rister, et al., Single-base pair differences in a shared motif determine differential Rhodopsin expression, Science 350 (6265) (2015) 1258–1261.
[65] S. Rowan, et al., Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity, Genes Dev. 24 (10) (2010) 980–985.
[66] E.K. Farley, et al., Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers, Proc. Natl. Acad. Sci. USA 113 (23) (2016) 6508–6513.
[67] J.F. Kribelbauer, et al., Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes, Annu Rev. Cell Dev. Biol. 35 (2019) 357–379.
[68] T. Fuqua, et al., Dense and pleiotropic regulatory information in a developmental enhancer, Nature 587 (7833) (2020) 235–239.
[69] C. Rastogi, et al., Accurate and sensitive quantification of protein-DNA binding affinity, Proc. Natl. Acad. Sci. USA 115 (16) (2018) E3692–E3701.
[70] A. Zandvakili, et al., Degenerate Pax2 and Senseless binding motifs improve detection of low-affinity sites required for enhancer specificity, PLoS Genet 14 (4) (2018), e1007289.
[71] R.K. Delker, et al., Low affinity binding sites in an activating CRM mediate negative autoregulation of the Drosophila Hox gene Ultrabithorax, PLoS Genet 15 (10) (2019), e1008444.
[72] C. Sanchez-Higueras, et al., In vivo Hox binding specificity revealed by systematic changes to a single cis regulatory module, Nat. Commun. 10 (1) (2019) 3597.
[73] J. Chen, et al., Single-molecule dynamics of enhanceosome assembly in embryonic stem cells, Cell 156 (6) (2014) 1274–1285.
[74] I. Izeddin, et al., Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus, Elife 3 (2014).
[75] D. Normanno, et al., Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher, Nat. Commun. 6 (2015) 7357.
[76] A. Tsai, et al., Nuclear microenvironments modulate transcription from low-affinity enhancers, Elife 6 (2017).
[77] A. Tsai, M.R. Alves, J. Crocker, Multi-enhancer transcriptional hubs confer phenotypic robustness, Elife 8 (2019).
[78] W.K. Cho, et al., Mediator and RNA polymerase II clusters associate in transcription-dependent condensates, Science 361 (6400) (2018) 412–415.
[79] S. Chong, et al., Imaging dynamic and selective low-complexity domain interactions that control gene transcription, Science 361 (6400) (2018).
[80] A.J. Plys, R.E. Kingston, Dynamic condensates activate transcription, Science 361 (6400) (2018) 329–330.
[81] K. Rippe, Liquid-Liquid Phase Separation in Chromatin, Cold Spring Harb. Perspect. Biol. 14 (2) (2022).
[82] B.R. Sabari, et al., Coactivator condensation at super-enhancers links phase separation and gene control, Science 361 (6400) (2018).
[83] B.R. Sabari, A. Dall'Agnese, R.A. Young, Biomolecular Condensates in the Nucleus, Trends Biochem Sci. 45 (11) (2020) 961–977.
[84] Y.E. Guo, et al., Pol II phosphorylation regulates a switch between transcriptional and splicing condensates, Nature 572 (7770) (2019) 543–548.
[85] S. Kilic, et al., Phase separation of 53BP1 determines liquid-like behavior of DNA repair compartments, EMBO J. 38 (16) (2019), e101379.
[86] A.G. Larson, et al., Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin, Nature 547 (7662) (2017) 236–240.
[87] M.W. Parker, et al., A new class of disordered elements controls DNA replication through initiator self-assembly, Elife 8 (2019).
[88] A. Boija, et al., Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains, Cell 175 (7) (2018), p. 1842-1855 e16.
[89] P.B. Sigler, Transcriptional activation . Acid blobs and negative noodles, Nature 333 (6170) (1988) 210–212.
[90] R. Brent, M. Ptashne, A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor, Cell 43 (3 Pt 2) (1985) 729–736.
[91] L. Keegan, G. Gill, M. Ptashne, Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein, Science 231 (4739) (1986) 699–704.
[92] K. Shrinivas, et al., Enhancer features that drive formation of transcriptional condensates, Mol. Cell 75 (3) (2019), p. 549-561 e7.
[93] L. Wang, et al., Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism, Mol. Cell 76 (4) (2019), p. 646-659 e6.
[94] S. Brodsky, et al., Intrinsically disordered regions direct transcription factor in vivo binding specificity, Mol. Cell 79 (3) (2020), p. 459-471 e4.
[95] S.G. Roberts, Mechanisms of action of transcription activation and repression domains, Cell Mol. Life Sci. 57 (8–9) (2000) 1149–1160.
[96] L. Staby, et al., Eukaryotic transcription factors: paradigms of protein intrinsic disorder, Biochem J. 474 (15) (2017) 2509–2532.
[97] I. Rambaldi, E.N. Kovacs, M.S. Featherstone, A proline-rich transcriptional activation domain in murine HOXD-4 (HOX-4.2), Nucleic Acids Res 22 (3) (1994) 376–382.
[98] Y. Liu, K.S. Matthews, S.E. Bondos, Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila hox protein ultrabithorax, J. Biol. Chem. 283 (30) (2008) 20874–20887.
[99] S.E. Bondos, H.C. Hsiao, Roles for intrinsic disorder and fuzziness in generating context-specific function in Ultrabithorax, a Hox transcription factor, Adv. Exp. Med Biol. 725 (2012) 86–105.
[100] A.L. Darling, V.N. Uversky, Intrinsic disorder in proteins with pathogenic repeat expansions, Molecules 22 (12) (2017).
[101] A.R. La Spada, J.P. Taylor, Repeat expansion disease: progress and puzzles in disease pathogenesis, Nat. Rev. Genet 11 (4) (2010) 247–258.
[102] H.T. Orr, H.Y. Zoghbi, Trinucleotide repeat disorders, Annu Rev. Neurosci. 30 (2007) 575–621.
[103] Y. Muragaki, et al., Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13, Science 272 (5261) (1996) 548–551.
[104] A. Boija, I.A. Klein, R.A. Young, Biomolecular condensates and cancer, Cancer Cell 39 (2) (2021) 174–192.
[105] S.M. Gough, C.I. Slape, P.D. Aplan, NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights, Blood 118 (24) (2011) 6247–6257.
[106] D. Jankovic, et al., Leukemogenic mechanisms and targets of a NUP98/HHEX fusion in acute myeloid leukemia, Blood 111 (12) (2008) 5672–5682.
[107] A. Mendes, B. Fahrenkrog, NUP214 in Leukemia: it's more than transport, Cells 8 (1) (2019).
[108] R. Tatavosian, et al., Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation, J. Biol. Chem. 294 (5) (2019) 1451–1463.
[109] N. Treen, et al., Properties of repression condensates in living Ciona embryos, Nat. Commun. 12 (1) (2021) 1561.

Seminars in Cell and Developmental Biology 152-153 (2024) 4–15

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Evolution of the insect Hox gene cluster: Comparative analysis across 243 species

Peter O. Mulhair *, Peter W.H. Holland

Department of Biology, University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ, UK

ARTICLE INFO

Keywords:
Homeobox
Hexapoda
Gene cluster
Shx

ABSTRACT

The Hox gene cluster is an iconic example of evolutionary conservation between divergent animal lineages, providing evidence for ancient similarities in the genetic control of embryonic development. However, there are differences between taxa in gene order, gene number and genomic organisation implying conservation is not absolute. There are also examples of radical functional change of Hox genes; for example, the *ftz*, *zen* and *bcd* genes in insects play roles in segmentation, extraembryonic membrane formation and body polarity, rather than specification of anteroposterior position. There have been detailed descriptions of Hox genes and Hox gene clusters in several insect species, including important model systems, but a large-scale overview has been lacking. Here we extend these studies using the publicly-available complete genome sequences of 243 insect species from 13 orders. We show that the insect Hox cluster is characterised by large intergenic distances, consistently extreme in Odonata, Orthoptera, Hemiptera and Trichoptera, and always larger between the 'posterior' Hox genes. We find duplications of *ftz* and *zen* in many species and multiple independent cluster breaks, although certain modules of neighbouring genes are rarely broken apart suggesting some organisational constraints. As more high-quality genomes are obtained, a challenge will be to relate structural genomic changes to phenotypic change across insect phylogeny.

1. Introduction

Insects display an astounding range of developmental and morphological diversity. Comprising over half of all described animal species, insect diversity has been attributed to high rates of speciation and adaptive radiation in association with flowering plant diversification, underpinned by dynamic rates of gene and genome evolution. Together with the orders Protura, Diplura and Collembola, insects make up the Hexapoda, a clade within Arthropoda consisting of six-legged, mostly terrestrial species. Within Hexapoda there have been several major evolutionary transitions associated with novel phenotypic traits. The evolution of insect wings is one such event which resulted in diversification of body forms within the clade Pterygota [1]. A later event was the emergence of complete metamorphosis in the holometabolous insects, thought to have permitted rapid diversification. Indeed, the most diverse and speciose insect orders are found within the holometabolous pterygotes (Hymenoptera, Coleoptera, Diptera and Lepidoptera). While the insect body plan is generally well conserved, a myriad of morphological novelties have emerged through insect radiation, ranging from pronotal horns on some beetles, sucking mouthparts in Hemiptera and (most) Lepidoptera, stings in bees and wasps, and halteres in Diptera and Strepsiptera.

Changes in developmental processes underlie morphological diversity, and ultimately these developmental changes must be underpinned by inherited genetic changes. Identifying these genetic changes is one of the goals of evolutionary developmental biology (evo-devo) although this is a difficult task when the morphological transitions occurred tens or hundreds of millions of years ago. One place to start is with the genes shared between taxa, and with key roles in development: a set of genes sometimes called the developmental toolkit. The Hox genes are examples of such core developmental genes, encoding position along the anteroposterior axis of most animal embryos. Furthermore, Hox genes code for transcription factors that activate and repress cascades of downstream genes to sculpt the morphology appropriate to that position. Later in development, Hox genes also orchestrate cell differentiation decisions, primarily though not exclusively within their original embryonic expression domains [2]. Changes in the content, order and expression domains of these genes have been implicated in a huge

* Corresponding author.
E-mail addresses: peter.mulhair@biology.ox.ac.uk (P.O. Mulhair), peter.holland@biology.ox.ac.uk (P.W.H. Holland).

https://doi.org/10.1016/j.semcdb.2022.11.010
Received 29 July 2022; Received in revised form 28 November 2022; Accepted 30 November 2022
Available online 14 December 2022

range of morphological novelties in the arthropod body plan [2].

The insect Hox cluster is thought to have consisted ancestrally of 10 genes: *labial (lab)*, *proboscipedia (pb)*, *zerknüllt (zen)*, *Deformed (Dfd)*, *Sex combs reduced (Scr)*, *fushi tarazu (ftz)*, *Antennapedia (Antp)*, *Ultrabithorax (Ubx)*, *abdominal-A (abdA)*, and *Abdominal-B (AbdB)*, similar to that of the bilaterian ancestor [3,4]. Of these, *zen* and *ftz* have 'altered' roles, having switched from their ancestral roles in anteroposterior position specification to extraembryonic membrane patterning (*zen*) and segmentation (*ftz*). Until recently, we have lacked a thorough knowledge of the multiple evolutionary paths that have been taken from this ancestral state across the diversity of insect orders. Here, we provide an updated view of the evolution of the Hox gene cluster across the largest sample of insect genomes sampled to date. We focus on the evolution of Hox cluster organisation, and do not discuss recent work on Hox gene regulation or the changing downstream functions of Hox genes, such as co-option of Hox genes to accessory roles in different orders (e.g. [5–9]). Furthermore, we focus on protein-coding loci within Hox clusters, and do not cover non-coding RNAs since these cannot be predicted reliably from genome sequences. There is certainly good evidence for antisense lncRNAs produced from within the Hox cluster of insects and other arthropods, as well as annelids, but comparative data are sparse (for review see [10]). We do not report new sequence data here: these analyses are based on publicly-available complete genome sequences, interpreted in the light of previous analyses.

2. Hox genes in a new era of insect genomics

In the pre-genomic era of molecular biology, from the early 1980s to around 2005, Hox gene clusters were analysed either by painstaking positional cloning of mutants or by cross- hybridisation of probes to genomic libraries followed by laborious genomic walking and clone-by-clone sequencing, sometimes coupled with in situ hybridization to chromosome spreads [11,12]. These heroic efforts were limited to a few species, but they started to generate a picture of comparative stasis of insect Hox gene clusters. For example, in the fruitfly *Drosophila melanogaster*, the insect in which Hox genes were first studied, it was clear that the *zen* gene had undergone tandem duplication to give three genes: *zen*, *z2* and *bcd*. Otherwise, there were no examples of Hox gene duplication in that evolutionary lineage. Similarly, in the red flour beetle *Tribolium castaneum*, there is a *zen* duplication to give two genes, but otherwise the cluster is unaltered [13,14]. It was, of course, clear from the earliest days that splitting of the cluster was possible, as evidenced by the cluster split in *Drosophila melanogaster* with a 9.6 Mb gap between *Ubx* and *Antp*. It was also clear that intergenic distances could be very large, as in *Drosophila melanogaster* and the locust *Schistocerca gregaria* [11]. Nonetheless, these are relatively minor changes compared to whole Hox gene cluster duplication in vertebrates and extensive posterior Hox gene duplication in vertebrates and amphioxus [15,16].

From the mid-2000s onwards, complete genome sequencing began to be applied to single species, or small sets of related species, and the evolutionary picture was refined. First, it became clear that splitting of the cluster was not a unique event in *Drosophila melanogaster*, as there have been independent splits at different positions in some other *Drosophila* species [17]. Second, lability of the *zen* gene, in terms of propensity to duplication, was reinforced by a remarkable discovery in the genome of the silk moth *Bombyx mori* [18]. This analysis revealed extensive tandem gene duplication, generating at least 13 copies of *zen*: one locus with an amino acid sequence similar to the ancestral *zen* gene, and 12 that have diverged extensively and given the name Shx (Special homeobox) genes. Later analysis of a refined genome assembly suggested the number may be even greater: *zen* plus 15 Shx loci, although not all can encode functional proteins [19]. Genome sequences of two butterflies (*Heliconius melpomene* and the Monarch *Danaus plexippus*) revealed presence of four Shx genes [20,21], as did low coverage genome skims of the Comma butterfly *Polygonia c-album*, Speckled Wood butterfly *Pararge aegeria*, Scarlet Tiger moth *Callimorpha dominula* and Horse Chestnut leaf-miner *Cameraria ohridella* [19].

The above historical perspective is one of a gradually unfolding picture emerging in a piecemeal manner as each additional genome is sequenced or analysed. But the landscape is now changing rapidly. In 2018, the Earth BioGenome Project was announced, as a bold vision to determine the complete genome sequence of all living eukaryote species [22]. This vision has galvanised action from over 40 affiliated projects, each attempting to determine high quality genome sequences at scale [23]. Among these, the project that has generated the largest number of high quality insect genomes to date is the Wellcome Trust-funded 'Darwin Tree of Life' (DToL), focussed on species living in Britain and Ireland [24]. Since 2019, the DToL project has generated 381 complete genome assemblies with over 2000 more species in the genome sequencing pipeline (data as of July 2022: https://portal.darwintreeoflife.org/tracking). These genome assemblies have been determined using long-read DNA sequencing technology (primarily PacBio HiFi) and scaffolding to chromosome-level using Hi-C. As such, they surpass in quality the large majority of genome assemblies previously available. In particular, the large contig sizes scaffolded to chromosomes provides opportunity to determine gene order and distances. Importantly, all data from the DToL project are released openly.

In reviewing the evolution of Hox gene clusters, we consider that the landscape of the field has changed so remarkably in the past two years that we cannot draw conclusions solely from previously published analyses. Instead, we supplement previous findings with analyses of the openly released data from DToL and other genome sequencing projects. We do not report new experimental data here, but rather draw new conclusions from available data. We use these data to summarise patterns of Hox gene duplication and the changes to genomic organisation across insects, using genomic data from 243 species representing 13 insect orders, plus one order of non-insect hexapod as an outgroup (Table 1). We show that insects continue to be an important model for understanding Hox gene evolution and, with the development of further methods and models for genetic manipulation from a phylogenetically diverse set of orders, will be vital for progress in the field of evolutionary developmental biology [25].

3. Insect Hox gene clusters

3.1. Gene loss in insect Hox clusters

There are no clear examples of Hox gene loss within insects, at least for the 'canonical' Hox genes that play roles in specifying anteroposterior position. All eight of the expected canonical Hox genes - *pb*, *lab*, *Dfd*, *Scr*, *Antp*, *Ubx*, *abdA*, *AbdB* - are present in all insects (Fig. 1). This contrasts to some other arthropod lineages (see [26]). For example, within Crustacea the *abdA* gene is proposed to be missing in three barnacles that have been studied (*Elminius modestus*, *Trypetesa lampas* and *Sacculina carcini*) and within Chelicerata the same gene has not been found in two mites (*Archegozetes longisetosus* and *Tetranychus urticae*) and a pycnogonid (*Endeis spinosa*), although not all these surveys were based on high quality genome assemblies (see [26]).

The *ftz* gene, which evolved a role in segmental patterning in insects rather than specification of position, seems to be absent in the genome of the stick insect *Timema cristina* (order Phasmatodea; [27]; assembly tcristinae_2.1). However, since this is a finding from analysis of a single genome assembly, verification is needed. The other Hox gene with a changed function in insects, the paralogy group 3 gene *zen*, is present in most insects. Interestingly, *zen* appears to be lost completely from the genomes of two related flies, *Epicampocera succincta* and *Thecocarcelia acutangulata*, which are both within the dipteran family Tachinidae. Similar loss of *zen* may have occurred within some Chelicerata, where this gene is reported absent from the genomes of the mites *Tetranychus urticae* [28] and *Metaseiulus occidentalis* [29]. Other cases of gene loss affect more recent duplicates. For example, *zen* has undergone tandem duplication in several lineages of insects and in some cases there has

Table 1
Order, species and genomes used in this study.

Order	Species	Genome
Trichoptera	Limnephilus lunatus	GCA_917563855.1_iiLimLuna2.1_genomic
Trichoptera	Limnephilus marmoratus	GCA_917880885.1_iiLimMarm1.1_genomic
Trichoptera	Limnephilus rhombicus	GCA_929108145.1_iiLimRhom1.1_genomic
Trichoptera	Glyphotaelius pellucidus	GCA_936435175.1_iiGlyPell1.1_genomic
Trichoptera	Eubasilissa regina	GCA_022840565_Eubasilissa_regina
Phasmatodea	Timema cristinae	GCA_002926335_ipTimCris1
Plecoptera	Nemoura dubitans	GCA_921293005.1_ipNemDubi1.1_genomic
Plecoptera	Nemurella pictetii	GCA_921293315.1_ipNemPict2.1_genomic
Plecoptera	Brachyptera putata	GCA_907164805.1_ipBraPut3m.1_genomic
Coleoptera	Pyrochroa serraticornis	GCA_905333025.1_icPyrSerr1.1_genomic
Coleoptera	Rhagonycha fulva	GCA_905340355.1_icRhaFulv1.1_genomic
Coleoptera	Coccinella septempunctata	GCA_907165205.1_icCocSept1.1_genomic
Coleoptera	Malachius bipustulatus	GCA_910589415.1_icMalBipu1.1_genomic
Coleoptera	Adalia bipunctata	GCA_910592335.1_icAdaBipu1.1_genomic
Coleoptera	Ocypus olens	GCA_910593695.1_icOcyOlen1.1_genomic
Coleoptera	Cantharis rustica	GCA_911387805.1_icCanRust1.1_genomic
Coleoptera	Harmonia axyridis	GCA_914767665.1_icHarAxyr1.1_genomic
Coleoptera	Apoderus coryli	GCA_911728435.1_icApoCory1.1_genomic
Coleoptera	Pterostichus madidus	GCA_911728475.1_icPteMadi1.1_genomic
Coleoptera	Agrypnus murinus	GCA_929113105.1_icAgrMuri1.1_genomic
Coleoptera	Podabrus alpinus	GCA_932274525.1_icPodAlpi1.1_genomic
Coleoptera	Philonthus cognatus	GCA_932526585.1_icPhiCogn1.1_genomic
Coleoptera	Leistus spinibarbis	GCA_933228885.1_icLeiSpin1.1_genomic
Coleoptera	Polydrusus cervinus	GCA_935413205.1_icPolCerv1.1_genomic
Coleoptera	Melolontha melolontha	GCA_935421215.1_icMelMelo1.1_genomic
Coleoptera	Rutpela maculata	GCA_936432065.1_icLepMacu1.1_genomic
Coleoptera	Halyzia sedecimguttata	GCA_937662695.1_icHalSede1.1_genomic
Coleoptera	Ophonus ardosiacus	GCA_943142095.1_icOphArdo1.1_genomic
Lepidoptera	Micropterix aruncella	GCA_944548615.1_ilMicArun2.1_genomic
Lepidoptera	Autographa gamma	GCA_905146925.1_ilAutGamm1.1_genomic
Lepidoptera	Laspeyria flexula	GCA_905147015.1_ilLasFlex1.1_genomic
Lepidoptera	Inachis io	GCA_905147045.1_ilAglIoxx1.1_genomic
Lepidoptera	Pieris brassicae	GCA_905147105.1_ilPieBrab1.1_genomic
Lepidoptera	Blastobasis lacticolella	GCA_905147135.1_ilBlaLact1.1_genomic
Lepidoptera	Nymphalis urticae	GCA_905147175.1_ilAglUrti1.1_genomic
Lepidoptera	Euproctis similis	GCA_905147225.1_ilEupSimi1.1_genomic
Lepidoptera	Erynnis tages	GCA_905147235.1_ilEryTage1.1_genomic
Lepidoptera	Hypena proboscidalis	GCA_905147285.1_ilHypProb1.1_genomic
Lepidoptera	Mythimna impura	GCA_905147345.1_ilMytImpu1.1_genomic
Lepidoptera	Apotomis turbidana	GCA_905147355.1_ilApoTurb1.1_genomic
Lepidoptera	Aricia agestis	GCA_905147365.1_ilAriAges1.1_genomic
Lepidoptera	Hylaea fasciaria	GCA_905147375.1_ilHylFasc1.1_genomic
Lepidoptera	Limenitis camilla	GCA_905147385.1_ilLimCami1.1_genomic
Lepidoptera	Xestia xanthographa	GCA_905147715.1_ilXesXant1.1_genomic
Lepidoptera	Phlogophora meticulosa	GCA_905147745.1_ilPhlMeti2.1_genomic
Lepidoptera	Thyatira batis	GCA_905147785.1_ilThyBati1.1_genomic
Lepidoptera	Pieris rapae	GCA_905147795.1_ilPieRapa1.1_genomic
Lepidoptera	Phalera bucephala	GCA_905147815.1_ilPhaBuce1.1_genomic
Lepidoptera	Endotricha flammealis	GCA_905163395.1_ilEndFlam1.1_genomic
Lepidoptera	Noctua fimbriata	GCA_905163415.1_ilNocFimb1.1_genomic
Lepidoptera	Mamestra brassicae	GCA_905163435.1_ilMamBras1.1_genomic
Lepidoptera	Pararge aegeria	GCA_905163445.1_ilParAegt1.1_genomic
Lepidoptera	Craniophora ligustri	GCA_905163465.1_ilCraLigu1.1_genomic
Lepidoptera	Cosmia trapezina	GCA_905163495.1_ilCosTrap1.1_genomic
Lepidoptera	Lymantria monacha	GCA_905163515.1_ilLymMona1.1_genomic
Lepidoptera	Notocelia uddmanniana	GCA_905163555.1_ilNotUddm1.1_genomic
Lepidoptera	Celastrina argiolus	GCA_905187575.1_ilCelArgi3.1_genomic
Lepidoptera	Cyaniris semiargus	GCA_905187585.1_ilCyaSemi1.1_genomic
Lepidoptera	Colias croceus	GCA_905220415.1_ilColCroc2.1_genomic
Lepidoptera	Amphipyra tragopoginis	GCA_905220435.1_ilAmpTrag2.1_genomic
Lepidoptera	Deilephila porcellus	GCA_905220455.1_ilDeiPorc1.1_genomic
Lepidoptera	Ennomos fuscantarius	GCA_905220475.1_ilEnnFusc2.1_genomic
Lepidoptera	Laothoe populi	GCA_905220505.1_ilLaoPopu1.1_genomic
Lepidoptera	Lysandra coridon	GCA_905220515.1_ilLysCori1.1_genomic
Lepidoptera	Mellicta athalia	GCA_905220545.1_ilMelAtha1.1_genomic
Lepidoptera	Melitaea cinxia	GCA_905220565.1_ilMelCinx1.1_genomic
Lepidoptera	Nymphalis polychloros	GCA_905220585.1_ilNymPoly1.1_genomic
Lepidoptera	Spilosoma lubricipeda	GCA_905220595.1_ilSpiLubr1.1_genomic
Lepidoptera	Tinea trinotella	GCA_905220615.1_ilTinTrin1.1_genomic
Lepidoptera	Boloria selene	GCA_905231865.2_ilBolSele5.2_genomic
Lepidoptera	Pieris napi	GCA_905231885.1_ilPieNapi4.1_genomic
Lepidoptera	Vanessa atalanta	GCA_905147765.1_ilVanAtal1.1_genomic
Lepidoptera	Notodonta dromedarius	GCA_905147325.1_ilNotDrom1.1_genomic
Lepidoptera	Vanessa cardui	GCA_905220365.1_ilVanCard2.1_genomic
Lepidoptera	Hecatera dysodea	GCA_905332915.1_ilHecDyso1.1_genomic
Lepidoptera	Mimas tiliae	GCA_905332985.1_ilMimTili1.1_genomic
Lepidoptera	Lycaena phlaeas	GCA_905333005.1_ilLycPhla1.1_genomic
Lepidoptera	Lysandra bellargus	GCA_905333045.1_ilLysBell1.1_genomic
Lepidoptera	Maniola jurtina	GCA_905333055.1_ilManJurt1.1_genomic
Lepidoptera	Pheosia tremula	GCA_905333125.1_ilPheTrem1.1_genomic
Lepidoptera	Abrostola tripartita	GCA_905340225.1_ilAbrTrip1.1_genomic
Lepidoptera	Noctua pronuba	GCA_905220335.1_ilNocPron1.1_genomic
Lepidoptera	Atethmia centrago	GCA_905333075.2_ilAteCent1.2_genomic
Lepidoptera	Glaucopsyche alexis	GCA_905404095.1_ilGlaAlex1.1_genomic
Lepidoptera	Pheosia gnoma	GCA_905404115.1_ilPheGnom1.1_genomic
Lepidoptera	Hesperia comma	GCA_905404135.1_ilHesComm1.1_genomic
Lepidoptera	Biston betularia	GCA_905404145.1_ilBisBetu1.1_genomic
Lepidoptera	Plebejus argus	GCA_905404155.1_ilPleArgu1.1_genomic
Lepidoptera	Anthocharis cardamines	GCA_905404175.1_ilAntCard3.1_genomic
Lepidoptera	Fabriciana adippe	GCA_905404265.1_ilFabAdip1.1_genomic
Lepidoptera	Hedya salicella	GCA_905404275.1_ilHedSali1.1_genomic
Lepidoptera	Erannis defoliaria	GCA_905404285.1_ilEraDefo1.1_genomic
Lepidoptera	Ochlodes sylvanus	GCA_905404295.1_ilOchSylv3.1_genomic
Lepidoptera	Leptidea sinapis	GCA_905404315.1_ilLepSina1.1_genomic
Lepidoptera	Autographa pulchrina	GCA_905475315.1_ilAutPulc1.1_genomic
Lepidoptera	Clostera curtula	GCA_905475355.1_ilCloCurt1.1_genomic
Lepidoptera	Schrankia costaestrigalis	GCA_905475405.1_ilSchCost1.1_genomic
Lepidoptera	Ochropleura plecta	GCA_905475445.1_ilOchPlec1.1_genomic
Lepidoptera	Zeuzera pyrina	GCA_907165235.1_ilZeuPyri1.1_genomic
Lepidoptera	Habrosyne pyritoides	GCA_907165245.1_ilHabPyri1.1_genomic
Lepidoptera	Zygaena filipendulae	GCA_907165275.1_ilZygFili1.1_genomic
Lepidoptera	Crocallis elinguaria	GCA_907269065.1_ilCroElin1.1_genomic
Lepidoptera	Idaea aversata	GCA_907269075.1_ilIdaAver1.1_genomic
Lepidoptera	Blastobasis adustella	GCA_907269095.1_ilBlaAdus2.1_genomic
Lepidoptera	Mythimna ferrago	GCA_910589285.1_ilMytFerr1.1_genomic
Lepidoptera	Noctua janthe	GCA_910589295.1_ilNocJant1.1_genomic
Lepidoptera	Bembecia ichneumoniformis	GCA_910589475.1_ilBemIchn1.1_genomic
Lepidoptera	Ennomos quercinarius	GCA_910589525.1_ilEnnQuei1.1_genomic
Lepidoptera	Carcina quercana	GCA_910589575.1_ilCarQuer1.1_genomic

(*continued on next page*)

Table 1 (*continued*)

Order	Species	Genome
Lepidoptera	Chrysoteuchia culmella	GCA_910589605.1_ilChrCulm1.1_genomic
Lepidoptera	Tinea semifulvella	GCA_910589645.1_ilTinSemi1.1_genomic
Lepidoptera	Acronicta aceris	GCA_910591435.1_ilAcrAcer1.1_genomic
Lepidoptera	Cydia splendana	GCA_910591565.1_ilCydSple1.1_genomic
Lepidoptera	Ypsolopha scabrella	GCA_910592155.1_ilYpsScab1.1_genomic
Lepidoptera	Amphipyra berbera	GCA_910594945.1_ilAmpBerb1.1_genomic
Lepidoptera	Parapoynx stratiotata	GCA_910589355.1_ilParStra1.1_genomic
Lepidoptera	Pyrgus malvae	GCA_911387765.1_ilPyrMalv3.1_genomic
Lepidoptera	Thymelicus sylvestris	GCA_911387775.1_ilThySylv1.1_genomic
Lepidoptera	Apamea monoglypha	GCA_911387795.1_ilApaMono1.1_genomic
Lepidoptera	Neomicropteryx cornuta	GCA_020383195.1_ilNeoCorn1.1_genomic
Lepidoptera	Hemaris fuciformis	GCA_907164795.1_ilHemFuc1.1_genomic
Lepidoptera	Papilio machaon	GCA_912999745.1_ilPapMach1.1_genomic
Lepidoptera	Sesia apiformis	GCA_914767545.1_ilSesApif2.1_genomic
Lepidoptera	Hydraecia micacea	GCA_914767645.1_ilHydMica1.1_genomic
Lepidoptera	Ptilodon capucinus	GCA_914767695.1_ilPtiCapc1.1_genomic
Lepidoptera	Agrochola circellaris	GCA_914767755.1_ilAgrCirc1.1_genomic
Lepidoptera	Eupsilia transversa	GCA_914767815.1_ilEupTran1.1_genomic
Lepidoptera	Agriopis aurantiaria	GCA_914767915.1_ilAgrAura1.1_genomic
Lepidoptera	Eilema depressum	GCA_914767945.1_ilEilDepe1.1_genomic
Lepidoptera	Eilema sororculum	GCA_914829495.1_ilEilSoro1.1_genomic
Lepidoptera	Spilarctia lutea	GCA_916048165.1_ilSpiLutu1.1_genomic
Lepidoptera	Griposia aprilina	GCA_916610205.1_ilGriApri1.1_genomic
Lepidoptera	Omphaloscelis lunosa	GCA_916610215.1_ilOmpLuno1.1_genomic
Lepidoptera	Mesoligia furuncula	GCA_916614155.1_ilMesFuru1.1_genomic
Lepidoptera	Xestia c-nigrum	GCA_916618015.1_ilXesCnig1.1_genomic
Lepidoptera	Emmelina monodactyla	GCA_916618145.1_ilEmmMono1.1_genomic
Lepidoptera	Agrochola macilenta	GCA_916701695.1_ilAgrMaci1.1_genomic
Lepidoptera	Orgyia antiqua	GCA_916999025.1_ilOrgAnti1.1_genomic
Lepidoptera	Erebia ligea	GCA_917051295.1_ilEreLige1.1_genomic
Lepidoptera	Dryobotodes eremita	GCA_917490735.1_ilDryErem1.1_genomic
Lepidoptera	Selenia dentaria	GCA_917880725.1_ilSelDent1.1_genomic
Lepidoptera	Synanthedon vespiformis	GCA_918317495.1_ilSynVesp1.1_genomic
Lepidoptera	Notodonta ziczac	GCA_918843915.1_ilNotZicz1.1_genomic
Lepidoptera	Eulithis prunata	GCA_918843925.1_ilEulPrun1.1_genomic
Lepidoptera	Philereme vetulata	GCA_918857605.1_ilPhiVetu1.1_genomic
Lepidoptera	Melanargia galathea	GCA_920104075.1_ilMelGala2.1_genomic
Lepidoptera	Furcula furcula	GCA_911728495.1_ilFurFurc1.1_genomic
Lepidoptera	Peribatodes rhomboidaria	GCA_911728515.1_ilPerRhom1.1_genomic
Lepidoptera	Pammene fasciana	GCA_911728535.1_ilPamFasc1.1_genomic
Lepidoptera	Aporia crataegi	GCA_912999735.1_ilApoCrat1.1_genomic
Lepidoptera	Hydriomena furcata	GCA_912999785.1_ilHydFurc1.1_genomic
Lepidoptera	Campaea margaritaria	GCA_912999815.1_ilCamMarg1.1_genomic
Thysanoptera	Thrips palmi	GCF_012932325_itThrPalm1
Collembola	Folsomia candida	GCF_002217175_hcFolCand1
Hymenoptera	Bombus hortorum	GCA_905332935.1_iyBomHort1.1_genomic
Hymenoptera	Bombus pascuorum	GCA_905332965.1_iyBomPasc1.1_genomic
Hymenoptera	Bombus campestris	GCA_905333015.1_iyBomCamp1.1_genomic
Hymenoptera	Vespula germanica	GCA_905340365.1_iyVesGerm1.1_genomic
Hymenoptera	Vespula vulgaris	GCA_905475345.1_iyVesVulg1.1_genomic
Hymenoptera	Nomada fabriciana	GCA_907165295.1_iyNomFabr1.1_genomic
Hymenoptera	Vespa crabro	GCA_910589235.1_iyVesCrab1.1_genomic
Hymenoptera	Cerceris rybyensis	GCA_910591515.1_iyCerRyby1.1_genomic
Hymenoptera	Nysson spinosus	GCA_910591585.1_iyNysSpin1.1_genomic
Hymenoptera	Ectemnius continuus	GCA_910591665.1_iyEctCont1.1_genomic

Table 1 (*continued*)

Order	Species	Genome
Hymenoptera	Bombus terrestris	GCA_910591885.1_iyBomTerr1.1_genomic
Hymenoptera	Andrena haemorrhoa	GCA_910592295.1_iyAndHaem1.1_genomic
Hymenoptera	Ectemnius lituratus	GCA_910593735.1_iyEctLitu1.1_genomic
Hymenoptera	Dolichovespula media	GCA_911387685.1_iyDolMedi1.1_genomic
Hymenoptera	Bombus hypnorum	GCA_911387925.1_iyBomHypn1.1_genomic
Hymenoptera	Dolichovespula saxonica	GCA_911387935.1_iyDolSaxo1.1_genomic
Hymenoptera	Osmia bicornis	GCA_907164935.1_iyOsmBic2.1_genomic
Hymenoptera	Vespa velutina	GCA_912470025.1_iyVesVel2.1_genomic
Hymenoptera	Seladonia tumulorum	GCA_913789895.1_iySelTumu1.1_genomic
Hymenoptera	Sphecodes monilicornis	GCA_913789915.1_iySphMoni1.1_genomic
Hymenoptera	Tenthredo notha	GCA_914767705.1_iyTenNoth1.1_genomic
Hymenoptera	Anoplius nigerrimus	GCA_914767735.1_iyAnoNige1.1_genomic
Hymenoptera	Ancistrocerus nigricornis	GCA_916049575.1_iyAncNigr1.1_genomic
Hymenoptera	Macropis europaea	GCA_916610135.1_iyMacEuro1.1_genomic
Hymenoptera	Lasioglossum morio	GCA_916610235.1_iyLasMori1.1_genomic
Hymenoptera	Lasioglossum lativentre	GCA_916610255.1_iyLasLatv2.1_genomic
Hymenoptera	Athalia rosae	GCA_917208135.1_iyAthRosa1.1_genomic
Hymenoptera	Mimumesa dahlbomi	GCA_917499265.1_iyMimDahl1.1_genomic
Hymenoptera	Ichneumon xanthorius	GCA_917499995.1_iyIchXant1.1_genomic
Hymenoptera	Dolichovespula sylvestris	GCA_918808275.1_iyDolSylv1.1_genomic
Hymenoptera	Bombus sylvestris	GCA_911622165.1_iyBomSyle1.1_genomic
Hymenoptera	Andrena dorsata	GCA_929108735.1_iyAndDors1.1_genomic
Hymenoptera	Andrena minutula	GCA_929113495.1_iyAndMinu1.1_genomic
Hymenoptera	Bombus pratorum	GCA_930367275.1_iyBomPrat1.1_genomic
Diptera	Scaeva pyrastri	GCA_905146935.1_idScaPyra1.1_genomic
Diptera	Syritta pipiens	GCA_905187475.1_idSyrPipi1.1_genomic
Diptera	Tachina fera	GCA_905220375.1_idTacFera2.1_genomic
Diptera	Xylota sylvarum	GCA_905220385.1_idXylSylv2.1_genomic
Diptera	Eristalis tenax	GCA_905231855.1_idEriTena2.1_genomic
Diptera	Volucella inanis	GCA_907269105.1_idVolInan1.1_genomic
Diptera	Eristalis pertinax	GCA_907269125.1_idEriPert2.1_genomic
Diptera	Bibio marci	GCA_910594885.1_idBibMarc1.1_genomic
Diptera	Xanthogramma pedissequum	GCA_910595825.1_idXanPedi1.1_genomic
Diptera	Chrysotoxum bicinctum	GCA_911387755.1_idChrBici1.1_genomic
Diptera	Melanostoma mellinum	GCA_914767635.1_idMelMell2.1_genomic
Diptera	Coremacera marginata	GCA_914767935.1_idCorMarg1.1_genomic
Diptera	Thecocarcelia acutangulata	GCA_914767995.1_idTheAcut1.1_genomic
Diptera	Bellardia pandia	GCA_916048285.1_idBelPand1.1_genomic
Diptera	Platycheirus albimanus	GCA_916050605.1_idPlaAlba1.1_genomic
Diptera	Cheilosia vulpina	GCA_916610125.1_idCheVulp2.1_genomic
Diptera	Eristalis arbustorum	GCA_916610145.1_idEriArbu1.1_genomic
Diptera	Gymnosoma rotundatum	GCA_916610165.1_idGymRotn1.1_genomic
Diptera	Criorhina berberina	GCA_917880715.1_idCriBerb1.1_genomic
Diptera	Eupeodes latifasciatus	GCA_920104205.1_idEupLati1.1_genomic
Diptera	Clusia tigrina	GCA_920105625.1_idCluTigr1.1_genomic
Diptera	Sicus ferrugineus	GCA_922984085.1_idSicFerr1.1_genomic
Diptera	Sarcophaga caerulescens	GCA_927399465.1_idSarCaer1.1_genomic
Diptera	Volucella inflata	GCA_928272305.1_idVolInfl1.1_genomic
Diptera	Epistrophe grossulariae	GCA_929447395.1_idEpiGros1.1_genomic
Diptera	Myathropa florea	GCA_930367185.1_idMyaFlor2.1_genomic
Diptera		GCA_930367215.1_idPolAngu1.1_genomic

(*continued on next page*)

Table 1 (*continued*)

Order	Species	Genome
	Pollenia angustigena	
Diptera	Sarcophaga rosellei	GCA_930367235.1_idSarRose1.1_genomic
Diptera	Sarcophaga variegata	GCA_932273835.1_idSarVari1.1_genomic
Diptera	Leucozona laternaria	GCA_932273885.1_idLeuLate1.1_genomic
Diptera	Protocalliphora azurea	GCA_932274085.1_idProAzur1.1_genomic
Diptera	Nephrotoma flavescens	GCA_932526605.1_idNepFlae1.1_genomic
Diptera	Epicampocera succincta	GCA_932526305.1_idEpiSucc1.1_genomic
Diptera	Bombylius major	GCA_932526495.1_idBomMajo1.1_genomic
Diptera	Rhingia campestris	GCA_932526625.1_idRhiCamp1.1_genomic
Diptera	Stomorhina lunata	GCA_933228675.1_idStoLuna1.1_genomic
Diptera	Machimus atricapillus	GCA_933228815.1_idMacAtri3.1_genomic
Diptera	Cheilosia pagana	GCA_936431705.1_idChePaga1.1_genomic
Diptera	Nowickia ferox	GCA_936439885.1_idNowFero1.1_genomic
Diptera	Sarcophaga subvicina	GCA_936449025.1_idSarSubv1.1_genomic
Diptera	Thecophora atra	GCA_937620795.1_idTheAtra2.1_genomic
Diptera	Cistogaster globosa	GCA_937654795.1_idCisGlob1.1_genomic
Diptera	Bombylius discolor	GCA_939192795.1_idBomDisc1.1_genomic
Diptera	Phyto melanocephala	GCA_941918925.1_idPhyMeln1.1_genomic
Diptera	Calliphora vomitoria	GCA_942486065.1_idCalVomi1.1_genomic
Odonata	Ischnura elegans	GCA_921293095.1_ioIscEleg1.1_genomic
Odonata	Platycnemis pennipes	GCA_933228895.1_ioPlaPenn1.1_genomic
Odonata	Pantala flavescens	GCA_020796165_Panflav1_CAAS_Pfla_1.0
Psocodea	Liposcelis brunnea	GCA_023512825_ipLipBrun1
Orthoptera	Schistocerca piceifrons	GCA_021461385_ioSchPice1
Orthoptera	Schistocerca gregaria	GCA_023897955_ioSchGreg1
Orthoptera	Schistocerca americana	GCA_021461395_ioSchAmer1
Hemiptera	Aelia acuminata	GCA_911387785.1_ihAelAcum1.1_genomic
Hemiptera	Acanthosoma haemorrhoidale	GCA_930367205.1_ihAcaHaem1.1_genomic
Neuroptera	Chrysoperla carnea	GCA_905475395.1_inChrCarn1.1_genomic
Neuroptera	Chrysopa pallens	GCA_020423425_inChrPall1

been secondary loss of derived *zen* duplicates (see Section 3.4). This includes a shared loss of the *zen*-derived *ShxD* gene in all Lycaenidae butterflies.

3.2. Splits, rearrangements and inversions in the insect Hox cluster

Even before the molecular cloning of Hox genes, it was clear that the mutant loci giving homeotic phenotypes in *Drosophila melanogaster* were located in two distinct complexes on chromosome 3: the ANT-C and the BX-C [30,31]. Cloning revealed the ANT-C contains from *lab* to *Antp* (of the ancestral gene order), whereas BX-C contains the genes from *Ubx* to *AbdB*, with a 9.6 Mb gap between them. Splits have also occurred, at different positions, in other *Drosophila* species [17]. In the mosquito *Anopheles gambiae*, the cluster is not split. The clear implication is that an unbroken Hox cluster is ancestral for this clade of Diptera and by implication (assuming a split cluster cannot reform into a complete cluster) ancestral for all insects, as it is for the Bilateria. In surveying the structure of Hox clusters across insects, therefore, we are not asking whether a split cluster is ancestral. Instead, we can ask whether there are particular intergenic regions where splits are more frequent evolutionarily, and conversely whether particular sets of Hox genes always stay together in evolution. Here we define intergenic regions as the genomic content between the homeobox sequences of the Hox genes, as current data do not allow us to identify the ends of every transcription unit.

Examination of 243 insect genomes reveals Hox cluster splits in many species (Fig. 1). For example, these include splits in the Hox clusters of *Platycnemis pennipes* (Odonata), *Aelia acuminata* (Hemiptera), *Chrysoperla carnea* (Neuroptera), *Coremacera marginata* (Diptera), *Limnephilus marmoratus* (Trichoptera) and in *all* Lepidoptera. In some cases, these splits lead to dramatic expansion in the overall size of the Hox cluster. For example, the *lab*, *pb* and *zen* genes in *Platycnemis pennipes* are located ~84 Mb from the rest of the cluster. Similarly, splits between *Scr* and *Dfd* and between *pb* and *lab* in *Aelia acuminata*, resulted in genomic distances of ~33 Mb and ~21 Mb between these genes, respectively. In *Chrysoperla carnea* the split occurs between *Scr* and *Dfd* and results in a distance of ~67 Mb, and in *Coremacera marginata* a distance of ~66 Mb separates *zen* and *pb*. The cluster split in Lepidoptera lies between *lab* and the rest of the cluster, and is present in every lepidopteran species analysed (124 species) including two representatives of the most basal family, Micropterygidae (*Neomicropteryx cornuta* and *Micropterix aruncella*).

The *lab* gene is found distal to the 'posterior' end of the cluster in most Lepidoptera (represented in Fig. 1 by the Silver-Y moth *Autographa gamma*). This repositioning is clearly a secondary event following 'escape' of the Hox gene from tight linkage to other Hox genes, since in *Neomicropteryx cornuta* (in the basal family Micropterygidae) the split has occurred but the repositioning has not. The finding that the *lab* gene is also distant from the rest of the Hox cluster in Trichoptera (caddisfly) genomes suggests this split probably occurred prior to the common ancestor of Lepidoptera+Trichoptera (Amphiesmenoptera). Interestingly, relocation of *lab* to a different chromosome was also found in two mosquito species, *Aedes aegypti* and *Culex quinquefasciatus* [32].

Similar cases of translocation or inversion have occurred in Odonata, Thysanoptera and Trichoptera, after splitting of the Hox cluster. This is the case for *zen*, *pb* and *lab* split from the rest of the cluster in *Platycnemis pennipes* (Odonata), *Thrips palmi* (Thysanoptera) and *Limnephilus marmoratus* (Trichoptera). Other inversions of genes occur unrelated to splits, for example, *Dfd* is inverted in *Coremacera marginata* and *Nephrotoma flavescens* (Diptera) and *Neomicropteryx cornuta* and *Autographa gamma* (Lepidoptera) (Fig. 1).

While splits have occurred frequently in insect evolution, the overall genomic order of Hox genes in insects is comparable to that seen for the homologous genes in vertebrates. This represents the colinear correspondence between gene order and the body position where each gene is expressed and functional during early embryonic development, for those Hox genes that still play this role (Fig. 2). Although we do not find clear cases of shuffling this order when the genes are together in a single intact cluster, there are cases of rearrangement caused by cluster breakage, in some cases involving inversions. Interestingly, these changes are almost always associated with paralogy groups 1–4: *lab, pb, zen* and *Dfd* (Fig. 2). The rearrangements found affect these four Hox genes in different ways. In each case of gene, or gene block rearrangement, there is a link between splits in the Hox cluster and subsequent rearrangement events within insect orders.

In some insect orders, there are rearrangements in all species sampled; for example, five Trichoptera (caddisfly) species have *pb*, *zen* and *Dfd* in derived positions. In all four species in the family Limnephilidae, *pb*, *zen* and *Dfd* are located at the 'posterior' (*AbdB*) end of the cluster, with an inversion in two species, *Limnephilus marmoratus* and *Limnephilus rhombicus*. In *Eubasilissa regina* (family Phryganeidae), *pb*, *zen* and *Dfd* are found outside the cluster, upstream of *lab*. Similarly, in all Lepidoptera (butterflies and moths; 124 species) *lab* is found away from the rest of the Hox cluster. The rearrangement of *lab* away from the rest of the Hox cluster was noted previously in the silk moth *Bombyx mori* [18,33]; the higher quality genome assemblies now available confirm that the *lab* gene is usually located at the 'posterior' end of the lepidopteran Hox cluster, separated by a large distance (from 1.4 Mb in *Tinea semifulvella* to 24 Mb in *Phalera bucephala*) containing numerous non-Hox genes.

In two of the three Odonata (dragonfly and damselfly) species

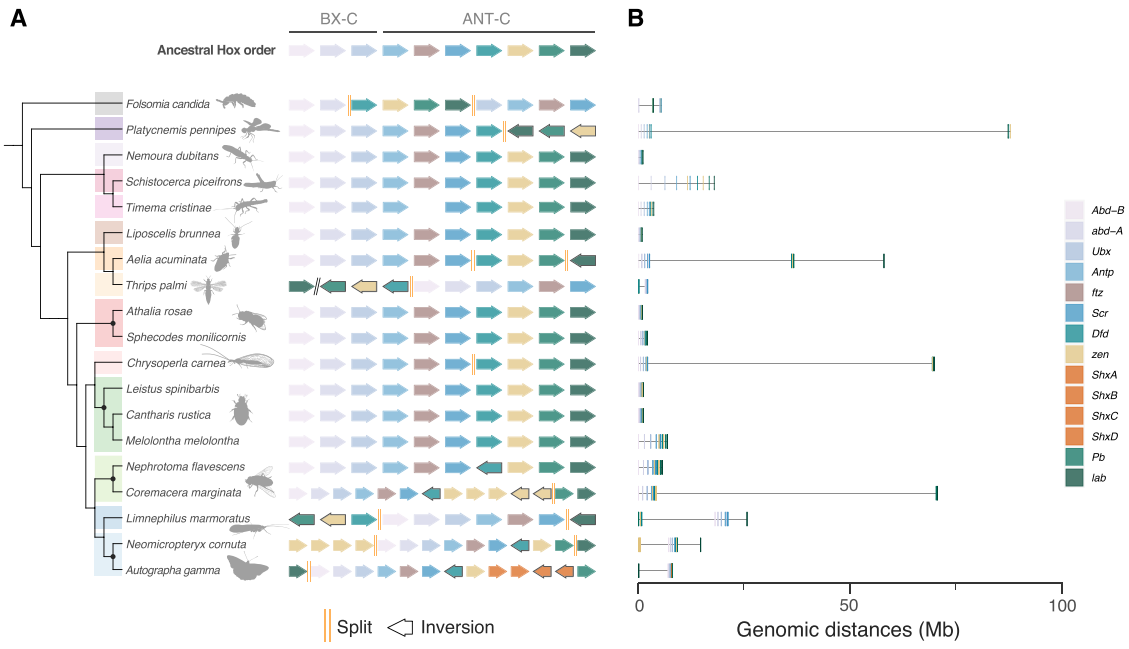


Fig. 1. Genomic organisation and gene orientation across insect Hox clusters (A) Left shows the phylogeny for subset of species analysed. Hexapod orders, from top to bottom are: Collembola (grey), Odonata (dark purple), Plecoptera (light purple), Orthoptera (dark pink), Phasmatodea (light pink), Psocodea (brown), Hemiptera (dark orange), Thysanoptera (light orange), Hymenoptera (dark red), Neuroptera (light red), Coleoptera (dark green), Diptera (light green), Trichoptera (dark blue), and Lepidoptera (light blue).Dots on the phylogeny represent hexapod orders for which data are shown from more than one species. Right shows the order and transcriptional orientation of Hox genes (coloured as per the legend) in each species. Splits within the Hox cluster are denoted by double orange lines, inversions are annotated with a black border around the gene. Slanted double black lines represent translocation to a separate scaffold. (B) Structure of the Hox cluster per species shown using actual genomic distances. Each line represents a Hox gene as it occurs in the genome, coloured as per the legend. Genomic distances are shown in Megabases.

analysed, *lab*, *pb* and *zen* are rearranged, but *Dfd* is in its ancestral position in the cluster. In the white-legged damselfly *Platycnemis pennipes* there has been an inversion that switched the order and transcriptional orientation of these genes as a block, and in the blue-tailed damselfly *Ischnura elegans* there has been an inversion plus a translocation to the other end of the cluster. In Plecoptera (stoneflies), Thysanoptera (thrips) and Neuroptera (lacewings and allies) various rearrangements are found. In certain species in these groups *lab*, *pb*, *zen* and *Dfd* have all been translocated to the 'posterior' end of the cluster, nearer to *AbdB,* with a subsequent inversion of this gene cassette in the plecopteran species *Nemurella pictetii*. In the only thysanopteran species in our dataset (*Thrips palmi*), an additional rearrangement resulted in *lab* positioned on a separate scaffold to the rest of the cluster and *pb*, *zen* and *Dfd* translocated to the posterior end. In Hemiptera, *lab* is located at the 'posterior' end of the cluster in *Acanthosoma haemorrhoidale*, although larger rearrangements affecting *lab*, *pb*, *zen* and *Dfd* in another hemipteran species (*Diaphorina citri*) has been observed [34]. Diptera displays the largest number of rearrangements, with at least five different rearrangement events occurring across the tree, resulting in translocations of one or more of the *lab*, *pb*, *zen* and *Dfd* genes to the opposite end of the gene cluster. In Coleoptera, three species show translocation of *zen* copies outside of the Hox cluster, resulting from independent lineage-specific events.

3.3. Hox cluster size across insects

While splits and rearrangements in the Hox cluster occur frequently across insects, there are certain genes which have rarely been split apart in the insect genomes studied to date. For example, the three genes found in the Bithorax complex of *Drosophila* (*AbdB*, *abdA* and *Ubx*) are found in the same order, in all 243 insect genomes studied, although we note that a cluster split between *Ubx* and *abdA* occurred in a clade of *Drosophila* [17]. Within the set of genes corresponding to the ANT-C of *Drosophila* (Figs. 1–2), the genes *Antp*, *ftz* and *Scr* are most conserved in their organisation and orientation. To our knowledge, there are no known cases of split between these genes, indicating there may be a selective pressure to maintain their linkage. Indeed, overall there are relatively few cluster splits between *AbdB* and *Scr*.

When the intergenic distances between each pair of genes (measured as the distance between homeobox sequences of the Hox genes) are compared between insect orders, a very intriguing pattern emerges (Fig. 3A). Excluding the first three Hox genes located at the 'anterior' end of the cluster (*lab*, *pb* and *zen*), which underwent significant rearrangements in many different species, we see that the intergenic distances between the next four genes (*Antp*, *ftz*, *Scr* and *Dfd*) are consistently small. These four 'tightly linked' genes are all orthologues of the ANT-C genes of *Drosophila melanogaster*. In contrast, the intergenic distances between the three orthologues of the BX-C genes (*AbdB*, *abdA* and *Ubx*) are consistently larger. This trend is most easily seen when the distances are compared within an insect order, and is seen regardless of whether the insect order has more or less 'relaxed' organisation of the Hox cluster (Fig. 3A). This may imply that there is a deep and fundamental difference between Hox gene organisation between ANT-C and BX-C genes, dating to long before the homeotic complex split in *Drosophila*. Interestingly, the intergenic distance between *Ubx* and *Antp* in most insects (the position of the BX-C/ANT-C split in *Drosophila melanogaster*) falls into the range of the BX-C intergenic distances, even in gene clusters that are not split.

The relative conservation in gene order and organisation from *AbdB* to *Scr* across all orders provides a useful opportunity to compare the evolution of the Hox cluster size across insects. The size of this conserved core region of the Hox cluster ranges from 0.57 Mb (Common Plume moth *Emmelina monodactyla*; Lepidoptera) to 5.8 Mb (*Tachina fera*; Diptera). These genomic distances are much larger than the same region in vertebrates where whole Hox clusters are only ~0.1 Mb [35,36]. Odonata, Hemiptera and Trichoptera have consistently large sizes for

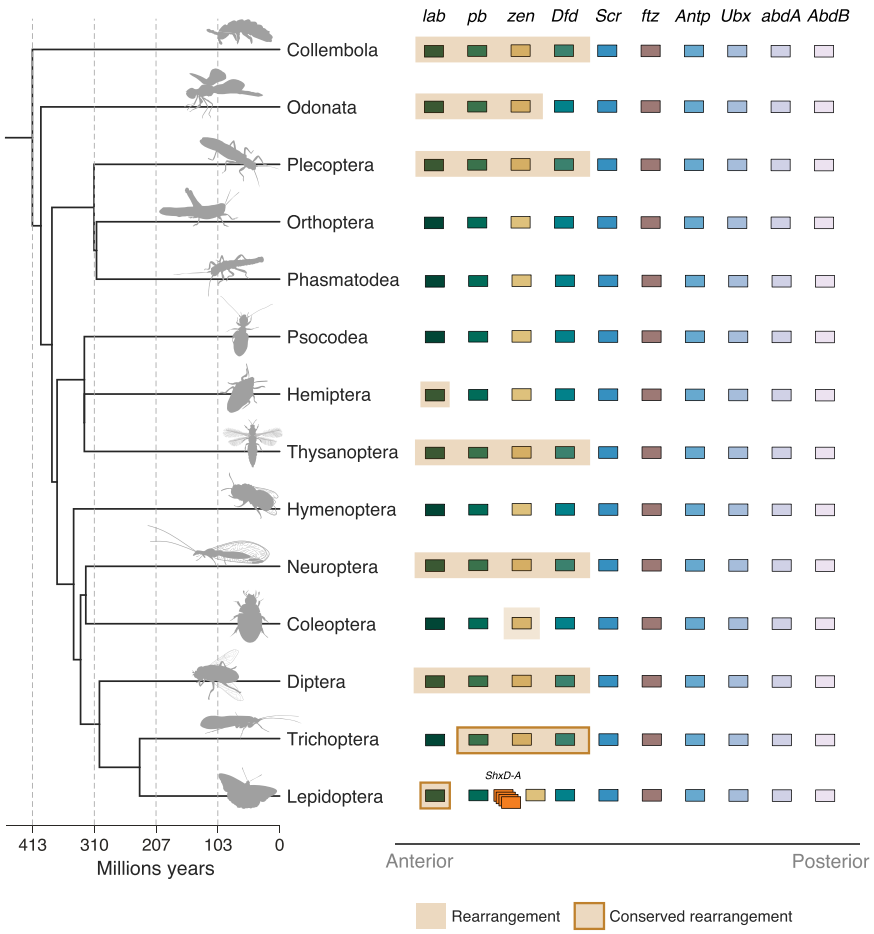


Fig. 2. Hox genes prone to rearrangement in insect Hox clusters Left shows a time calibrated species tree of insect Orders analysed in this study. Right shows the composition of the Hox gene cluster, in their ancestral order. Shaded orange regions represent genes that have undergone rearrangement from the ancestral order of Hox genes. Those conserved across species sampled have a border around the box. Splits in the Hox cluster are not depicted.

this core cluster, reflecting large intergenic distances, while Coleoptera and Diptera each show great variation in cluster size across the order (Fig. 3B). For most insect orders, the size of the core part of the Hox cluster correlates with genome size (Fig. 3C). However, Diptera and Lepidoptera both show low correlation values ($r = 0.37$ and 0.27, respectively), suggesting that there are other factors driving the size of the Hox cluster other than genome size in these groups of insects.

The size of the Hox cluster in *Schistocerca piceifrons* (Orthoptera; assembly iqSchPice1.1) is significantly expanded with larger distances between all genes, relative to most other insects, suggesting relaxation of the constraints acting on the whole cluster (Fig. 1). When we compare cluster size across other *Schistocerca* species (*Schistocerca americana*; iqSchAmer2.1 and *Schistocerca gregaria*; iqSchGreg1.1), the total size of the Hox cluster ranges from 16 Mb to 17.8 Mb, and the 'core' Hox cluster size (*AbdB* to *Scr*) ranges from 10.8 Mb to 12.2 Mb, significantly larger than any other insect species analysed in this study (Fig. 3B). This contrasts with earlier (pre-genomic) analysis in *Schistocerca gregaria*, where the total cluster size was determined using chromosomal in situ hybridization and estimated to be at least 700Kb in length, and no longer that 2 Mb in total [11]. Although linkage in the Hox cluster in this genus has relaxed significantly, there are no rearrangements found in the order of the Hox genes within the genome.

3.4. Tandem duplication of insect Hox genes: Zerknüllt and fushi tarazu

Tandem duplication within a Hox gene cluster is rare, with some of the clearest examples being the initial expansion of the Hox cluster in early bilaterian evolution [37,38] and expansions at the 'posterior' end of the cluster in vertebrates, amphioxus and echinoderms [15,39]. In analysing publicly-available insect genomes, we find only two cases of putative tandem duplication of a canonical Hox gene (Fig. 4): two copies of *Dfd* present in *Acronicta aceris* (Sycamore Moth) and two copies of *pb* present in *Micropterix aruncella*. While genomic position and gene trees provide support for these putative Hox duplicates as real events, since they are present in a single species each of the findings needs further verification. Indeed, it is expected that tandem duplications of canonical Hox genes would be deleterious since they could disrupt the spatial regulation of these genes, and thereby disrupt anteroposterior body patterning. In contrast, the two Hox genes that have derived roles, *ftz* and *zen*, might be expected to have less constraint against duplication. This is because the *zen* gene lost its ancestral homeotic function in early insect evolution, acquiring a novel role in the formation of extraembryonic tissues, while *ftz* has a new role in segmentation.

As noted in Section 3.1, a putative loss of *ftz* is observed in one insect, and conversely there are two *ftz* copies in *Spilarctia lutea* (Buff Ermine Moth), as well as two closely related wasps: *Vespula germanica* and *Vespula vulgaris* (Fig. 4). Finding the duplication in related species gives stronger support to this observation. Duplications of *zen* in insects have been known about and intensively studied for many years. First, there is a well-studied duplication of *zen* in *Tribolium castaneum*, which gave *zen* and *zen2* [13]. Recent work has shown that this duplication is shared by three closely related *Tribolium* species, and that the gene products interact in a negative feedback loop that may confer precision of temporal expression [40]. Second, it was shown over 20 years ago that the dipteran *bcd* gene is a derived tandem duplicate of *zen* [41]. This duplication was followed by extensive sequence divergence in the locus

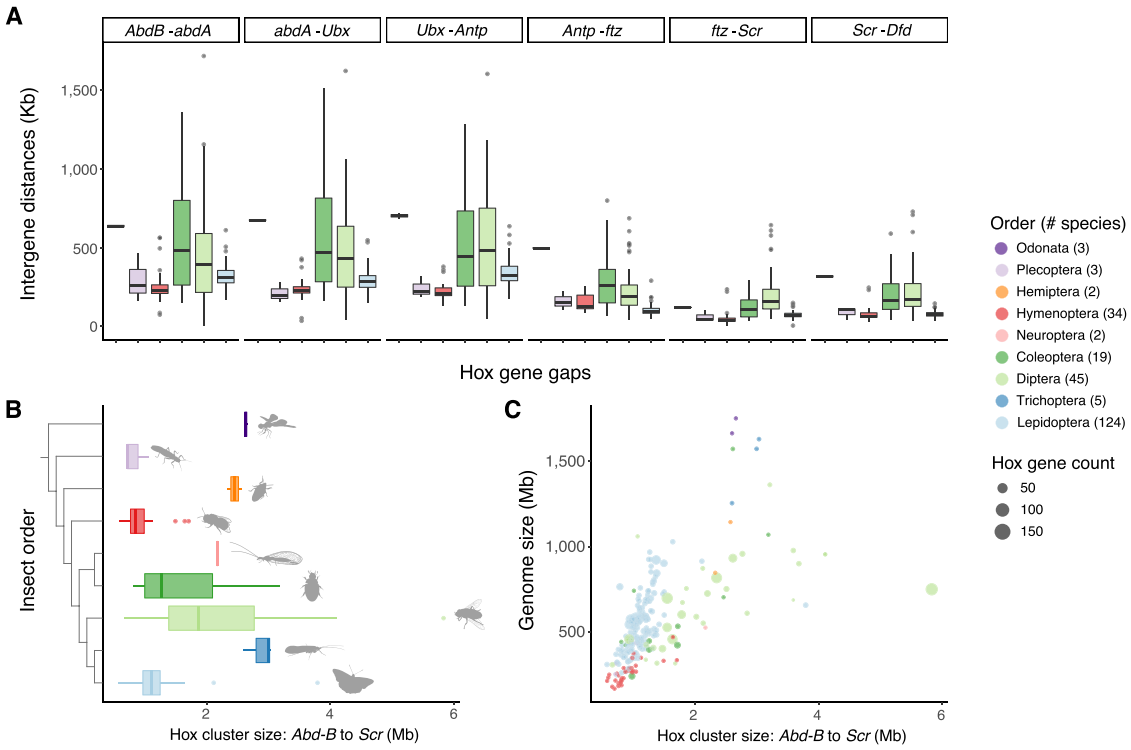


Fig. 3. Hox cluster size across insects. (A) Distribution of intergenic regions per Hox gene in different insect orders. Only shows orders where the Hox genes are conserved in the 'normal' order. (B) Distribution of core Hox cluster (AbdB to Scr) size for each order, ordered by phylogeny. (C) Correlation between core Hox cluster size and genome size. Each dot represents a species, which are colour by order as per the figure legend. Dots are sized by the total number of Hox genes present in each species. In (A) and (B) boxplots have a rectangle between the 25th and 75th percentiles of the range, with the median as a dark line, whiskers reach the largest and smallest values within 1.5x Interquartile range, and outliers are points beyond 1.5 × Interquartile range (plotted using geom_boxplot in ggplot2).

which became *bcd*, in a classic case of 'asymmetric sequence divergence' where one daughter gene undergoes far more sequence change than the other [42,43]. Key amino acid changes in the homeodomain include a mutation from glutamine to lysine at position 50 (Q50K) and a switch from methionine to arginine at position 54 (M54R); these substitutions contributed to changing downstream targets and altered biological role [44,45]. Third, a further duplication of *zen* within the *Drosophila* genus produced *zen2*. Fourth, the most dramatic cases of *zen* duplication have been reported in Lepidoptera. In ditrysian lepidopterans, the *zen* gene duplicated to give four additional fast-evolving copies named *ShxA*, *ShxB*, *ShxC* and *ShxD* [19]. These highly derived Shx genes are expressed in the developing serosa, not the embryo itself, and may pre-pattern this extraembryonic tissue, as judged by the striking pattern of maternal RNA localisation [19]. These genes duplicated even further in *Bombyx mori* resulting in at least 12, possibly 15, tandem Shx gene copies [18,19]. Furthermore, recent work has shown that the extreme duplication in *Bombyx* is not unique: at least 18 other lineages of Lepidoptera have highly expanded sets of Shx genes, in some cases reaching over 100 homeobox copies (Fig. 4). There has also been occasional loss of specific Shx genes; for example, *ShxD* was lost in butterflies of the family Lycaenidae ('blues' and their allies) and fritillary butterflies of the genus *Melitaea* [46]).

With the availability of many high-quality insect genomes, it is now possible to ask if there are additional cases of *zen* duplication, in addition to those mentioned above (Fig. 4). Within the coleopteran species for which genomes are available, multiple duplications of *zen* occur in the Cucujiformia infraorder and range from 2 copies in *Polydrusus cervinus* (weevil) to 17 copies in *Harmonia axyridis* (harlequin ladybird) and 19 copies in *Pyrochroa serraticornis* (cardinal beetle). This is in addition to the well-studied duplication in *Tribolium*. In Diptera, copy number of *zen* ranges from a single copy in the early diverging lineages to 118 in *Tachina fera* and 93 in *Sarcophaga variegata*. Even within a family of flies for which there is a large number of species sampled, Syrphidae, there is significant variation in *zen* copy number between these related species (Fig. 4). The number of copies and the branching patterns within the gene tree (Fig. 4) suggest that large tandem duplication events occurred multiple times independently in this lineage.

It is striking that duplication of *zen*, and its progenitors (Shx in Lepidoptera) occur only in the highly speciose orders Diptera, Lepidoptera and Coleoptera, within the holometabolous insects. As described above, *zen* lost its homeotic function early on in insect evolution, and in many insect species is involved in development of extraembryonic membranes. In insects these membranes consist of two distinct layers: the amnion and serosa (these form a single epithelium known as the amnioserosa in higher flies) [47,48]. The amnion is the inner membrane which surrounds the ventral side of the developing embryo, while the serosa is an outer membrane which lies just inside the chorion and envelops the embryo, amnion and yolk [49–51]. This structure is hypothesised to be involved in a wide range of functions unrelated to development of body form, such as a general protective role including structural stability, water regulation and desiccation resistance [52–54], and innate immune response [55–57]. Interestingly, while this dual structure is present in most pterygote insects, derived hymenopterans (Apocrita) usually lack an amnion, or have a temporary amniotic-like structure which covers the yolk [58,59]. It is intriguing to consider whether the highly dynamic copy numbers of *zen* along with its functions in the extraembryonic tissues may have played a role in facilitating speciation and adaptation to diverse habitats in Diptera, Lepidoptera and Coleoptera. Indeed, it is striking that copy number variation of *zen* is particularly variable in highly speciose families such as Syrphidae and Coccinellidae (Fig. 4). However, whether these large expansions in gene number are functional, or even expressed during early development, requires further analyses. Furthermore, the neutral theory posits that increases in copy number may not always have

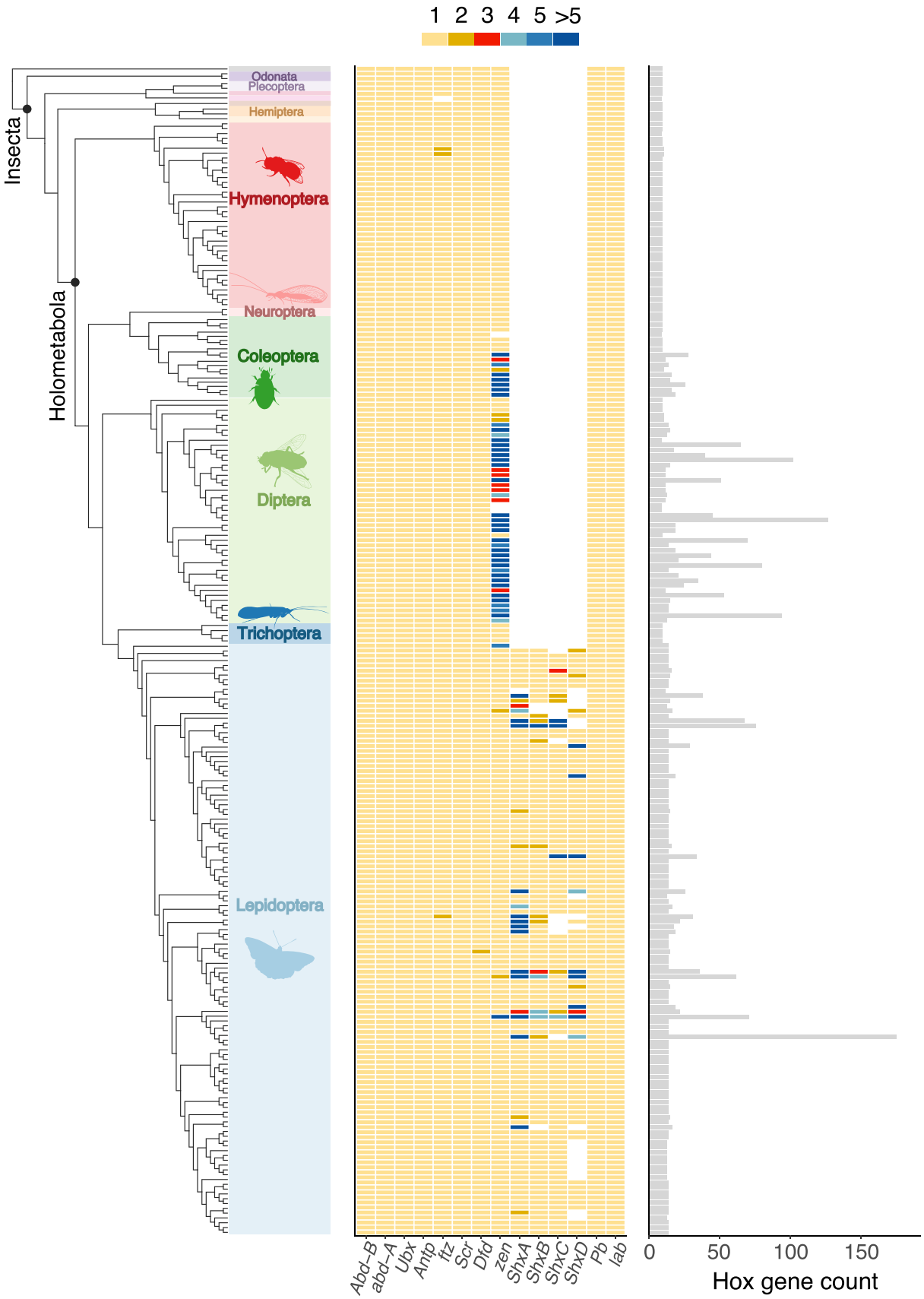


Fig. 4. Copy number of Hox genes across insects. Left shows the phylogeny of all species analysed, with block colours signifying the Order to which they belong. Heatmap shows the copy number of all Hox genes per species. Barchart of the right shows the total number of Hox genes annotated per species.

adaptive significance, and may instead result from mutational processes within the genome, affected by intragenomic variation in copying fidelity and the effects of transposable element accumulation.

4. Conclusions

We are entering a new era of genomics, as new technologies are facilitating the imminent sequencing and assembly of thousands of eukaryotic species. At the time of writing, there are more than 200 high quality, complete insect genomes available for analysis, and although this number is expected to rise very rapidly, now is an excellent time to pause and take stock of the lessons that can be learned. This is an opportune time for two reasons. First, the available high quality genomes span a wide phylogenetic diversity of insects, including representatives of at least 13 orders (Odonata, Plecoptera, Orthoptera, Phasmatodea, Psocodea, Hemiptera, Thysanoptera, Hymenoptera, Neuroptera, Coleoptera, Diptera, Trichoptera, Lepidoptera). Second, within some taxa (notably Lepidoptera and Diptera) 'deep dives' have been undertaken, yielding genomes from closely related species, thereby permitting insights into genomic change on shorter time frames. Here we have used these data, in combination with previously published analyses, to compare Hox gene cluster organisation across insects. We have searched for patterns of evolutionary conservation or general trends across insects, examples of convergent evolution, and anomalies.

First, we examine gene loss and conclude that canonical Hox genes have not been lost in insect evolution: *pb*, *lab*, *Dfd*, *Scr*, *Antp*, *Ubx*, *abdA*, *AbdB* are present in all insects studied. The two 'non-canonical' Hox genes, *zen* and *ftz*, are lost rarely. We find two closely related insect species putatively lacking *zen*, possibly a shared loss inherited from a common ancestor, and one example of a putative loss of *ftz*. The rarity of these losses highlights that further verification is needed. However, the finding that canonical Hox genes are never lost in insects has a biological implication. We suggest that each Hox gene has remained indispensable through insect radiation because segment number and tagmatization, has remained consistent, giving no opportunity for gene redundancy and loss.

Second, we find many independent cases of splitting of the insect Hox gene cluster, in an analogous fashion to the separation of ANT-C and BX-C in *Drosophila melanogaster*. Although these splits can occur in several different places in the Hox cluster, they are most commonly seen affecting the first four paralogy groups (PG1 to PG4): *lab*, *pb*, *zen*, and *Dfd*. There are cases where just *lab* (PG1) is split away (Lepidoptera and Trichoptera), one dipteran in which *lab* plus *pb* are separated away, Odonata with *lab*, *pb* and *zen* split away, and many insects with a split between *Dfd* and *Scr* (separating PG1 to PG4 from the rest). We do not find cases of complete 'atomisation' of the Hox cluster, as seen in larvacean chordates and predatory mites for example. It would be interesting to compare patterns of Hox cluster breakage and rearrangement with the overall genome-wide recombination and inversion rates for each taxa, to test if Hox cluster rearrangements reflect general genomic properties. From the pattern of splitting observed, we suggest that insect Hox genes are not generally regulated as a whole cluster, but there are selective pressures acting to prevent many rearrangements. These selective pressures could include shared regulation of neighbouring games, interdigitated control (enhancers for one Hox gene located beyond the neighbouring gene) or simply a high density of regulatory elements. We suggest these constraints are lowest around paralogy groups 1–4. We speculate that shared and interdigitated control may have evolved around 'posterior' insect Hox genes to fine-tune expression within overlapping domains in the abdomen.

Third, it has long been known that insect Hox gene clusters have much larger intergenic distances than in vertebrates. We find that intergenic distances in the Hox cluster vary greatly across insects, with particularly large genomic distances in Orthoptera, Odonata, Hemiptera and Trichoptera, and highly variable intergenic distances in Coleoptera and Diptera. Intergenic lengths correlate with genome size in most, but not all, insect orders. We note a striking and puzzling trend in intergenic distance within insect Hox clusters: the distances between 'posterior' genes are consistently greater than distances between each pair of 'central' or 'anterior' genes. Specifically, intergenic distances from *AbdB* to *Antp* are greater than intergenic distances across the rest of the cluster. We do not know the biological basis for this observation. Counterintuitively, the region with the largest intergenic distances is also the region least prone to genomic rearrangement in evolution. We suggest that fundamental mechanisms of gene regulation may be different at the two ends of the insect Hox gene cluster.

Fourth, we examine gene duplication and conclude that insect Hox genes are rarely duplicated, with the exception of *zen*. We do find putative cases of *Dfd* duplication and *pb* duplication, but these are seen in single genomes and require further verification. A *ftz* duplication is seen in genome assemblies for two wasps and can be treated as more definitive. The *zen* gene, in contrast, has undergone tandem duplication many times independently, undergoing dramatic copy number expansion in some insect lineages. The most striking examples of *zen* duplication are seen in genomes from the highly speciose orders, Coleoptera, Diptera and Lepidoptera, where over 100 *zen*-derived homeobox sequences can be present in some species. It is unclear why such dramatic copy number changes have occurred, and indeed whether retention of extra genes was selectively advantageous through subfunctionalization, neofunctionalization or simply dosage effects. The fact that *zen* genes play roles in extraembryonic patterning, rather than position-specific cell fate in the embryo, may underpin why tandem duplications are not instantly deleterious, but this does not seem to explain the preponderance of *zen* gene arrays observed. Further work is required to determine if the locus is particularly prone to unequal crossover at meiosis, and therefore a hotspot of mutation, and/or whether duplicated *zen* genes were repeatedly recruited to novel roles in extraembryonic membrane patterning as insects adapted to their multitude of ecological niches.

Declarations of interest

None.

Acknowledgements

We acknowledge the contributions of the many laboratories and research groups that have made insect genome sequences openly available for analysis, including the Darwin Tree of Life project coordinated by the Wellcome Sanger Institute. This research was funded by the Wellcome Trust Darwin Tree of Life Discretionary Award [218328] and the John Fell OUP Research Fund. For the purpose of open access, the authors have applied a CC BY public copyright licence.

References

[1] R. Dudley, The Biomechanics of Insect Flight: Form, Function, Evolution, Princeton University Press, 2002.
[2] C.L. Hughes, T.C. Kaufman, Hox genes and the evolution of the arthropod body plan, Evol. Dev. 4 (2002) 459–499.
[3] R. de Rosa, J.K. Grenier, T. Andreeva, C.E. Cook, A. Adoutte, M. Akam, S.B. Carroll, G. Balavoine, Hox genes in brachiopods and priapulids and protostome evolution, Nature 399 (1999) 772–776.
[4] C.E. Cook, M.L. Smith, M.J. Telford, A. Bastianello, M. Akam, Hox genes and the phylogeny of the arthropods, Curr. Biol. 11 (2001) 759–763.
[5] A.J.J. Crumière, A. Khila, Hox genes mediate the escalation of sexually antagonistic traits in water striders, Biol. Lett. 15 (2019), 20180720.
[6] L. Tian, S.R. Rahman, B.D. Ezray, L. Franzini, J.P. Strange, P. Lhomme, H.M. Hines, A homeotic shift late in development drives mimetic color variation in a bumble bee, Proc. Natl. Acad. Sci. USA 116 (2019) 11857–11865.
[7] A.M. Rafiqi, A. Rajakumar, E. Abouheif, Origin and elaboration of a major evolutionary transition in individuality, Nature 585 (2020) 239–244.
[8] B.-P. Liu, G. Ding, Y. Miao, B.-Z. Hua, The Hox gene Abdominal-B regulates the appendage development during the embryogenesis of scorpionflies, Insect Mol. Biol. (2022), https://doi.org/10.1111/imb.12790.
[9] S.N. Murugesan, H. Connahs, Y. Matsuoka, M. Das Gupta, G.J.L. Tiong, M. Huq, V. Gowri, S. Monroe, K.D. Deem, T. Werner, Y. Tomoyasu, A. Monteiro, Butterfly eyespots evolved via cooption of an ancestral gene-regulatory network that also

patterns antennae, legs, and wings, Proc. Natl. Acad. Sci. USA 119 (2022), https://doi.org/10.1073/pnas.2108661119.
[10] E.L. Novikova, M.A. Kulakova, There and back again: hox clusters use both DNA strands, J. Dev. Biol. 9 (2021), https://doi.org/10.3390/jdb9030028.
[11] D.E. Ferrier, M. Akam, Organization of the Hox gene cluster in the grasshopper, Schistocerca gregaria, Proc. Natl. Acad. Sci. USA 93 (1996) 13024–13029.
[12] Holland, The fall and rise of evolutionary developmental biology, Syst. Assoc. Spec. Vol. (2004). ⟨https://books.google.com/books?hl=en&lr=&id=qnUhL9z9QqcC&oi=fnd&pg=PA261&dq=The+Fall+and+Rise+of+Evolutionary+Developmental+Biology&ots=ss4mShxI2l&sig=lOQvpEl35SP065UFXJF2k4x2lqs⟩.
[13] S.J. Brown, J.P. Fellers, T.D. Shippy, E.A. Richardson, M. Maxwell, J.J. Stuart, R. E. Denell, Sequence of the Tribolium castaneum homeotic complex: the region corresponding to the Drosophila melanogaster antennapedia complex, Genetics 160 (2002) 1067–1074.
[14] M. van der Zee, N. Berns, S. Roth, Distinct functions of the tribolium zerknüllt genes in serosa specification and dorsal closure, Curr. Biol. 15 (2005) 624–636.
[15] D.E. Ferrier, C. Minguillón, P.W. Holland, J. Garcia-Fernàndez, The amphioxus Hox cluster: deuterostome posterior flexibility and Hox14, Evol. Dev. 2 (2000) 284–293.
[16] J. Pascual-Anaya, S. D'Aniello, S. Kuratani, J. Garcia-Fernàndez, Evolution of Hox gene clusters in deuterostomes, BMC Dev. Biol. 13 (2013) 26.
[17] B. Negre, A. Ruiz, HOM-C evolution in Drosophila: is there a need for Hox gene clustering? Trends Genet. 23 (2007) 55–59.
[18] C.-L. Chai, Z. Zhang, F.-F. Huang, X.-Y. Wang, Q.-Y. Yu, B.-B. Liu, T. Tian, Q.-Y. Xia, C. Lu, Z.-H. Xiang, A genomewide survey of homeobox genes and identification of novel structure of the Hox cluster in the silkworm, Bombyx mori, Insect Biochem. Mol. Biol. 38 (2008) 1111–1120.
[19] L. Ferguson, F. Marlétaz, J.-M. Carter, W.R. Taylor, M. Gibbs, C.J. Breuker, P.W. H. Holland, Ancient expansion of the hox cluster in lepidoptera generated four homeobox genes implicated in extra-embryonic tissue formation, PLOS Genet. 10 (2014), e1004698.
[20] S. Zhan, C. Merlin, J.L. Boore, S.M. Reppert, The monarch butterfly genome yields insights into long-distance migration, Cell 147 (2011) 1171–1185.
[21] Heliconius Genome Consortium, Butterfly genome reveals promiscuous exchange of mimicry adaptations among species, Nature 487 (2012) 94–98.
[22] H.A. Lewin, G.E. Robinson, W.J. Kress, W.J. Baker, J. Coddington, K.A. Crandall, R. Durbin, S.V. Edwards, F. Forest, M.T.P. Gilbert, M.M. Goldstein, I.V. Grigoriev, K.J. Hackett, D. Haussler, E.D. Jarvis, W.E. Johnson, A. Patrinos, S. Richards, J. C. Castilla-Rubio, M.-A. van Sluys, P.S. Soltis, X. Xu, H. Yang, G. Zhang, Earth biogenome project: sequencing life for the future of life, Proc. Natl. Acad. Sci. USA 115 (2018) 4325–4333.
[23] H.A. Lewin, S. Richards, E. Lieberman Aiden, M.L. Allende, J.M. Archibald, M. Bálint, K.B. Barker, B. Baumgartner, K. Belov, G. Bertorelle, M.L. Blaxter, J. Cai, N.D. Caperello, K. Carlson, J.C. Castilla-Rubio, S.-M. Chaw, L. Chen, A.K. Childers, J.A. Coddington, D.A. Conde, M. Corominas, K.A. Crandall, A.J. Crawford, F. DiPalma, R. Durbin, T.E. Ebenezer, S.V. Edwards, O. Fedrigo, P. Flicek, G. Formenti, R.A. Gibbs, M.T.P. Gilbert, M.M. Goldstein, J.M. Graves, H.T. Greely, I.V. Grigoriev, K.J. Hackett, N. Hall, D. Haussler, K.M. Helgen, C.J. Hogg, S. Isobe, K.S. Jakobsen, A. Janke, E.D. Jarvis, W.E. Johnson, S.J.M. Jones, E.K. Karlsson, P. J. Kersey, J.-H. Kim, W.J. Kress, S. Kuraku, M.K.N. Lawniczak, J.H. Leebens-Mack, X. Li, K. Lindblad-Toh, X. Liu, J.V. Lopez, T. Marques-Bonet, S. Mazard, J.A. K. Mazet, C.J. Mazzoni, E.W. Myers, R.J. O'Neill, S. Paez, H. Park, G.E. Robinson, C. Roquet, O.A. Ryder, J.S.M. Sabir, H.B. Shaffer, T.M. Shank, J.S. Sherkow, P. S. Soltis, B. Tang, L. Tedersoo, M. Uliano-Silva, K. Wang, X. Wei, R. Wetzer, J. L. Wilson, X. Xu, H. Yang, A.D. Yoder, G. Zhang, The earth biogenome project 2020: starting the clock, Proc. Natl. Acad. Sci. USA 119 (2022), https://doi.org/10.1073/pnas.2115635118.
[24] The Darwin Tree of Life Project Consortium, Sequence locally, think globally: the Darwin tree of life project, Proc. Natl. Acad. Sci. USA (2022). ⟨https://www.pnas.org/content/pnas/119/4/e2115642118⟩. accessed March 25, 2022.
[25] A. Heffer, L. Pick, Conservation and variation in Hox genes: how insect models pioneered the evo-devo field, Annu. Rev. Entomol. 58 (2013) 161–179.
[26] R.M. Pace, M. Grbić, L.M. Nagy, Composition and genomic organization of arthropod Hox clusters, Evodevo 7 (2016) 11.
[27] V. Soria-Carrasco, Z. Gompert, A.A. Comeault, T.E. Farkas, T.L. Parchman, J. S. Johnston, C.A. Buerkle, J.L. Feder, J. Bast, T. Schwander, S.P. Egan, B.J. Crespi, P. Nosil, Stick insect genomes reveal natural selection's role in parallel speciation, Science 344 (2014) 738–742.
[28] M. Grbić, T. Van Leeuwen, R.M. Clark, S. Rombauts, P. Rouzé, V. Grbić, E. J. Osborne, W. Dermauw, P.C.T. Ngoc, F. Ortego, P. Hernández-Crespo, I. Diaz, M. Martinez, M. Navajas, É. Sucena, S. Magalhães, L. Nagy, R.M. Pace, S. Djuranović, G. Smagghe, M. Iga, O. Christiaens, J.A. Veenstra, J. Ewer, R. M. Villalobos, J.L. Hutter, S.D. Hudson, M. Velez, S.V. Yi, J. Zeng, A. Pires-daSilva, F. Roch, M. Cazaux, M. Navarro, V. Zhurov, G. Acevedo, A. Bjelica, J.A. Fawcett, E. Bonnet, C. Martens, G. Baele, L. Wissler, A. Sanchez-Rodriguez, L. Tirry, C. Blais, K. Demeestere, S.R. Henz, T.R. Gregory, J. Mathieu, L. Verdon, L. Farinelli, J. Schmutz, E. Lindquist, R. Feyereisen, Y. Van de Peer, The genome of Tetranychus urticae reveals herbivorous pest adaptations, Nature 479 (2011) 487–492.
[29] M.A. Hoy, R.M. Waterhouse, K. Wu, A.S. Estep, P. Ioannidis, W.J. Palmer, A. F. Pomerantz, F.A. Simão, J. Thomas, F.M. Jiggins, T.D. Murphy, E.J. Pritham, H. M. Robertson, E.M. Zdobnov, R.A. Gibbs, S. Richards, Genome sequencing of the phytoseiid predatory mite metaseiulus occidentalis reveals completely atomized Hox genes and superdynamic intron evolution, Genome Biol. Evol. 8 (2016) 1762–1775.
[30] E.B. Lewis, A gene complex controlling segmentation in Drosophila, Nature 276 (1978) 565–570.
[31] T.C. Kaufman, R. Lewis, B. Wakimoto, Cytogenetic analysis of chromosome 3 in drosophila melanogaster: the homoeotic gene complex in polytene chromosome interval 84a-B, Genetics 94 (1980) 115–133.
[32] B.J. Matthews, O. Dudchenko, S.B. Kingan, S. Koren, I. Antoshechkin, J. E. Crawford, W.J. Glassford, M. Herre, S.N. Redmond, N.H. Rose, G.D. Weedall, Y. Wu, S.S. Batra, C.A. Brito-Sierra, S.D. Buckingham, C.L. Campbell, S. Chan, E. Cox, B.R. Evans, T. Fansiri, I. Filipović, A. Fontaine, A. Gloria-Soria, R. Hall, V. S. Joardar, A.K. Jones, R.G.G. Kay, V.K. Kodali, J. Lee, G.J. Lycett, S.N. Mitchell, J. Muehling, M.R. Murphy, A.D. Omer, F.A. Partridge, P. Peluso, A.P. Aiden, V. Ramasamy, G. Rašić, S. Roy, K. Saavedra-Rodriguez, S. Sharan, A. Sharma, M. L. Smith, J. Turner, A.M. Weakley, Z. Zhao, O.S. Akbari, W.C. Black 4th, H. Cao, A. C. Darby, C.A. Hill, J.S. Johnston, T.D. Murphy, A.S. Raikhel, D.B. Sattelle, I. V. Sharakhov, B.J. White, L. Zhao, E.L. Aiden, R.S. Mann, L. Lambrechts, J. R. Powell, M.V. Sharakhova, Z. Tu, H.M. Robertson, C.S. McBride, A.R. Hastie, J. Korlach, D.E. Neafsey, A.M. Phillippy, L.B. Vosshall, Improved reference genome of Aedes aegypti informs arbovirus vector control, Nature 563 (2018) 501–507.
[33] Y. Yasukochi, L.A. Ashakumary, C. Wu, A. Yoshido, J. Nohata, K. Mita, K. Sahara, Organization of the Hox gene cluster of the silkworm, Bombyx mori: a split of the Hox cluster in a non-Drosophila insect, Dev. Genes Evol. 214 (2004) 606–614.
[34] T.D. Shippy, P.S. Hosmani, M. Flores-Gonzalez, L.A. Mueller, W.B. Hunter, S. J. Brown, T. D'Elia, S. Saha, Annotation of Hox cluster and Hox cofactor genes in the Asian citrus psyllid, Diaphorina citri, reveals novel features, Gigabyte 2022 (2022) 1–18.
[35] D. Duboule, The rise and fall of Hox gene clusters, Development 134 (2007) 2549–2560.
[36] F. Darbellay, C. Bochaton, L. Lopez-Delisle, B. Mascrez, P. Tschopp, S. Delpretti, J. Zakany, D. Duboule, The constrained architecture of mammalian *Hox* gene clusters, Proc. Natl. Acad. Sci. USA 116 (2019) 13424–13433.
[37] P.W.H. Holland, Evolution of homeobox genes, Wiley Interdiscip. Rev. Dev. Biol. 2 (2013) 31–45.
[38] W. Nong, J. Cao, Y. Li, Z. Qu, J. Sun, T. Swale, H.Y. Yip, P.Y. Qian, J.-W. Qiu, H. S. Kwan, W. Bendena, S. Tobe, T.F. Chan, K.Y. Yip, K.H. Chu, S.M. Ngai, K.Y. Tsim, P.W.H. Holland, J.H.L. Hui, Jellyfish genomes reveal distinct homeobox gene clusters and conservation of small RNA processing, Nat. Commun. 11 (2020) 3051.
[39] R. Szabó, D.E.K. Ferrier, Two more Posterior Hox genes and Hox cluster dispersal in echinoderms, BMC Evol. Biol. 18 (2018) 203.
[40] D. Gurska, I.M. Vargas Jentzsch, K.A. Panfilio, Unexpected mutual regulation underlies paralogue functional diversification and promotes epithelial tissue maturation in Tribolium, Commun. Biol. 3 (2020) 552.
[41] M. Stauber, H. Jäckle, U. Schmidt-Ott, The anterior determinant *bicoid* of *Drosophila* is a derived *Hox* class 3 gene, Proc. Natl. Acad. Sci. USA 96 (1999) 3786–3789.
[42] P.W.H. Holland, F. Marlétaz, I. Maeso, T.L. Dunwell, J. Paps, New genes from old: asymmetric divergence of gene duplicates and the evolution of development, Philos. Trans. R. Soc. Lond. B Biol. Sci. 372 (2017), https://doi.org/10.1098/rstb.2015.0480.
[43] R. Bao, S.E. Dia, H.A. Issa, D. Alhusein, M. Friedrich, Comparative evidence of an exceptional impact of gene duplication on the developmental evolution of drosophila and the higher diptera, Front. Ecol. Evol. 6 (2018), https://doi.org/10.3389/fevo.2018.00063.
[44] Q. Liu, P. Onal, R.R. Datta, J.M. Rogers, U. Schmidt-Ott, M.L. Bulyk, S. Small, J. W. Thornton, Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development, eLife 7 (2018), https://doi.org/10.7554/eLife.34594.
[45] P. Onal, H.I. Gunasinghe, K.Y. Umezawa, M. Zheng, J. Ling, L. Azeez, A. Dalmeus, T. Tazin, S. Small, Suboptimal intermediates underlie evolution of the bicoid homeodomain, Mol. Biol. Evol. 38 (2021) 2179–2190.
[46] V. Ahola, R. Lehtonen, P. Somervuo, L. Salmela, P. Koskinen, P. Rastas, N. Välimäki, L. Paulin, J. Kvist, N. Wahlberg, J. Tanskanen, E.A. Hornett, L. C. Ferguson, S. Luo, Z. Cao, M.A. de Jong, A. Duplouy, O.-P. Smolander, H. Vogel, R.C. McCoy, K. Qian, W.S. Chong, Q. Zhang, F. Ahmad, J.K. Haukka, A. Joshi, J. Salojärvi, C.W. Wheat, E. Grosse-Wilde, D. Hughes, R. Katainen, E. Pitkänen, J. Ylinen, R.M. Waterhouse, M. Turunen, A. Vähärautio, S.P. Ojanen, A. H. Schulman, M. Taipale, D. Lawson, E. Ukkonen, V. Mäkinen, M.R. Goldsmith, L. Holm, P. Auvinen, M.J. Frilander, I. Hanski, The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera, Nat. Commun. 5 (2014) 4737.
[47] U. Schmidt-Ott, The amnioserosa is an apomorphic character of cyclorrhaphan flies, Dev. Genes Evol. 210 (2000) 373–376.
[48] U. Schmidt-Ott, C.W. Kwan, How two extraembryonic epithelia became one: serosa and amnion features and functions of Drosophila's amnioserosa, Philos. Trans. R. Soc. Lond. B Biol. Sci. 377 (2022), 20210265.
[49] K.A. Panfilio, Extraembryonic development in insects and the acrobatics of blastokinesis, Dev. Biol. 313 (2008) 471–491.
[50] U. Schmidt-Ott, C.W. Kwan, Morphogenetic functions of extraembryonic membranes in insects, Curr. Opin. Insect Sci. 13 (2016) 86–92.
[51] K.A. Panfilio, S.M. Chuva de Sousa Lopes, The extended analogy of extraembryonic development in insects and amniotes, Philos. Trans. R. Soc. Lond. B Biol. Sci. 377 (2022), 20210268.
[52] Y. Goltsev, G.L. Rezende, K. Vranizan, G. Lanzaro, D. Valle, M. Levine, Developmental and evolutionary basis for drought tolerance of the Anopheles gambiae embryo, Dev. Biol. 330 (2009) 462–470.
[53] C.G.C. Jacobs, G.L. Rezende, G.E.M. Lamers, M. van der Zee, The extraembryonic serosa protects the insect egg against desiccation, Proc. Biol. Sci. 280 (2013), 20131082.
[54] L.C. Farnesi, R.F.S. Menna-Barreto, A.J. Martins, D. Valle, G.L. Rezende, Physical features and chitin content of eggs from the mosquito vectors Aedes aegypti,

Anopheles aquasalis and Culex quinquefasciatus: connection with distinct levels of resistance to desiccation, J. Insect Physiol. 83 (2015) 43–52.
[55] M.J. Gorman, P. Kankanala, M.R. Kanost, Bacterial challenge stimulates innate immune responses in extra-embryonic tissues of tobacco hornworm eggs, Insect Mol. Biol. 13 (2004) 19–24.
[56] C.G.C. Jacobs, H.P. Spaink, M. van der Zee, The extraembryonic serosa is a frontier epithelium providing the insect egg with a full-range innate immune response, eLife 3 (2014), https://doi.org/10.7554/eLife.04111.
[57] C.G.C. Jacobs, R. van der Hulst, Y.-T. Chen, R.P. Williamson, S. Roth, M. van der Zee, Immune function of the serosa in hemimetabolous insect eggs, Philos. Trans. R. Soc. Lond. B Biol. Sci. 377 (2022), 20210266.
[58] A.L. Bull, Stages of living embryos in the jewel wasp Mormoniella (nasonia) vitripennis (walker) (hymenoptera: pteromalidae), Int. J. Insect Morphol. Embryol. 11 (1982) 1–23.
[59] R. Fleig, K. Sander, Honeybee morphogenesis: embryonic cell movements that shape the larval body, Development 103 (1988) 525–534.

Seminars in Cell and Developmental Biology 152-153 (2024) 44–57

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Establishing and maintaining *Hox* profiles during spinal cord development

Alexander Miller *, Jeremy S. Dasen *

NYU Neuroscience Institute and Developmental Genetics Programs, Department of Neuroscience and Physiology, NYU School of Medicine, New York, NY 10016, USA

ARTICLE INFO

Keywords:
Hox gene
Neural development
Spinal cord
Transcription factor
Motor neuron

ABSTRACT

The chromosomally-arrayed *Hox* gene family plays central roles in embryonic patterning and the specification of cell identities throughout the animal kingdom. In vertebrates, the relatively large number of *Hox* genes and pervasive expression throughout the body has hindered understanding of their biological roles during differentiation. Studies on the subtype diversification of spinal motor neurons (MNs) have provided a tractable system to explore the function of *Hox* genes during differentiation, and have provided an entry point to explore how neuronal fate determinants contribute to motor circuit assembly. Recent work, using both in vitro and in vivo models of MN subtype differentiation, have revealed how patterning morphogens and regulation of chromatin structure determine cell-type specific programs of gene expression. These studies have not only shed light on basic mechanisms of rostrocaudal patterning in vertebrates, but also have illuminated mechanistic principles of gene regulation that likely operate in the development and maintenance of terminal fates in other systems.

1. Introduction

Our ability to coordinately move and interact with the environment relies on the activity of neural circuits within the spinal cord. A key step in the assembly of motor circuits is the establishment of synaptic connections between spinal motor neurons (MNs) and their peripheral targets. The ability of mammalian nervous systems to coordinate movement depends on the generation of dozens of anatomically and functionally distinct MN subtypes. In vertebrates, spinal MN subtypes are housed in motor columns longitudinally arrayed along the rostrocaudal axis. Each motor column is further subdivided into motor pools, clustered groups of MNs that target a single muscle. Motor columns and pools are positionally-defined along the rostrocaudal axis, and their location is largely stereotyped between animals of the same species. Over the past decade significant progress has been made in define the genetic programs that determine the molecular profiles and connectivity of diverse spinal MN subtypes [31,150,122].

The selective and regulated expression of transcription factors (TFs) is integral during MN differentiation, with cell-type specific TF activity driving expression of effector genes that confer MN molecular identities and synaptic specificity. Central to spinal MN diversification is the large family of TFs encoded by chromosomally arrayed *Homeobox* (*Hox*) genes. *Hox* genes determine key features of MN fates including subtype-specific molecular profiles, somatotopic organization, and postsynaptic specificity [17,32,127]. *Hox* genes are conserved in all metazoans, and their activity confers the rostrocaudal positional identity of neural and non-neuronal tissues in all organisms that have been examined [49,65,157].

In addition to MNs, *Hox* genes function in the diversification of multiple classes of spinal neurons, operating both during early development and after terminal differentiation. Recent studies have shown *Hox* genes are involved in the subtype diversification of proprioceptive sensory neurons, spinal projection neurons, and locally-connected spinal interneurons [152,144,3]. After differentiation, *Hox* genes are also required to maintain expression of genes that define terminal fates [21]. Understanding the mechanisms by which *Hox* genes are regulated and maintained could provide insights into how spinal circuits are assembled during development and contribute to mature functional features of MNs.

This review outlines the mechanisms of *Hox* gene regulation during spinal cord development, focusing on the process of MN subtype differentiation. We will categorize these regulatory mechanisms in relation to the relative stages of neural differentiation. Namely, the early silencing of *Hox* expression preceding gastrulation, the establishment and further refinement of *Hox* expression domains during axis extension, and the postmitotic maintenance of *Hox* patterns. We outline the known and emerging mechanisms of *Hox* gene regulation, focusing on the regulatory mechanism operating within the developing spinal cord. In doing so, we hope to summarize the current knowledge on the regulation of *Hox* genes during neural development and provide a background

* Corresponding authors.
E-mail addresses: alexander.miller@nyulangone.org (A. Miller), jeremy.dasen@nyulangone.org (J.S. Dasen).

https://doi.org/10.1016/j.semcdb.2023.03.014
Received 22 November 2022; Received in revised form 18 January 2023; Accepted 30 March 2023
Available online 5 April 2023

for further study.

2. *Hox* gene organization and the subtype diversification of spinal MNs

Each *Hox* gene is characterized by a shared ~180 bp sequence element, encoding a 60 amino acid DNA-binding motif termed the Homeodomain (HD) [106,142]. Possessing a HD allows Hox proteins to associate with targets and directly regulate gene expression. Comparative genetic studies have found that *Hox* genes are conserved in vertebrates and other Animalia, from protostomes to humans [18,49,53,106,125,15]. Analyses of *Hox* gene organization has led to the understanding that *Hox* clusters were generated during repeated cycles of tandem gene duplications, combined with slow acquisition of new *Hox* regulatory

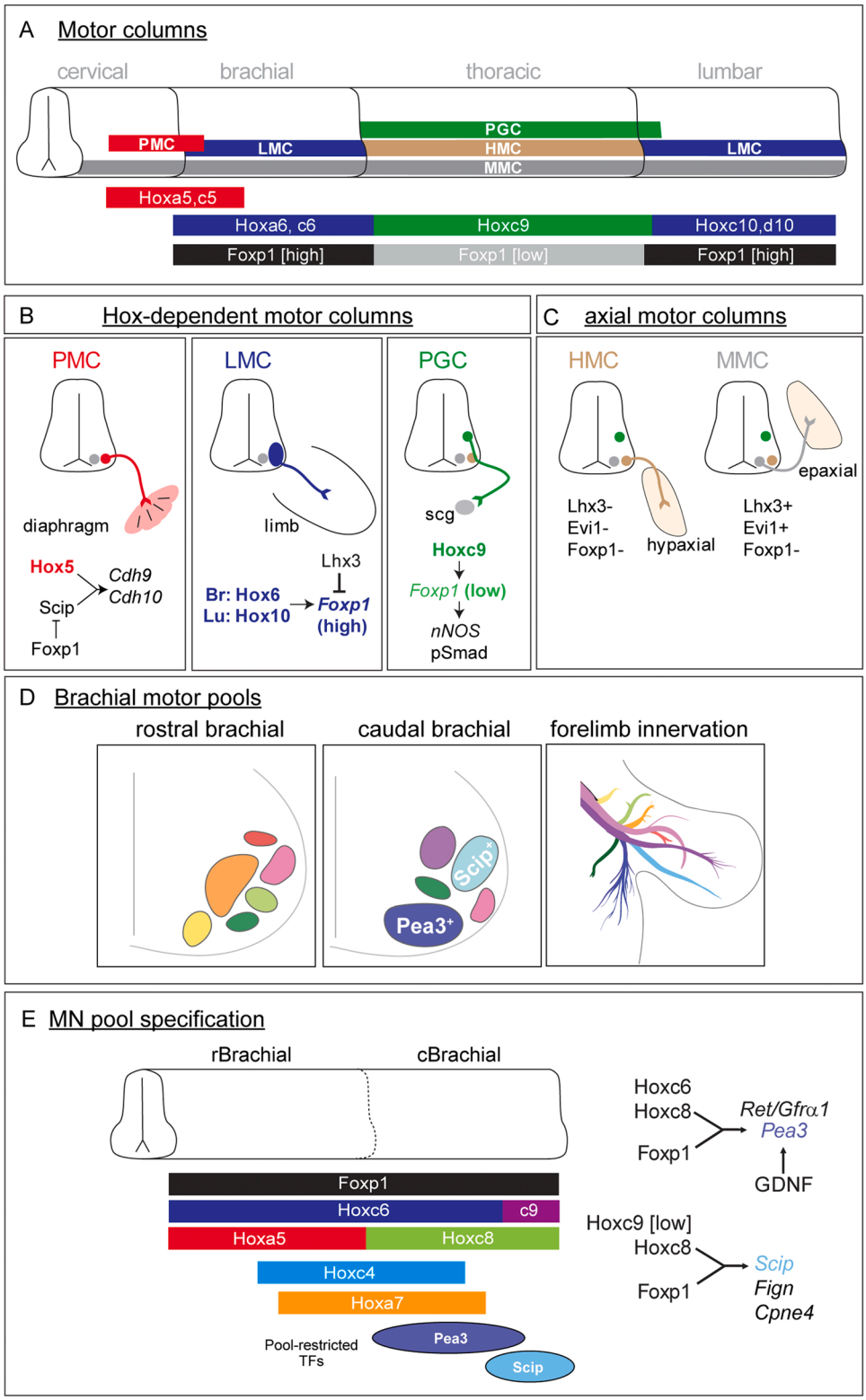


Fig. 1. *Hox* gene profiles and function during spinal MN development. (A) Organization of spinal motor columns at cervical, brachial, thoracic, and lumbar levels in mice. *Hox* genes expressed by segmentally restricted motor columns are shown. Phrenic motor column (PMC) neurons express *Hox5* genes (*Hoxa5* and *Hoxc5*), brachial and lumbar lateral motor column (LMC) neuron express *Hox6* and *Hox10* genes, respectively. *Hoxc9* is expressed by thoracic preganglionic column (PGC) and hypaxial motor column (HMC) neurons. Medial motor column (MMC) neurons are present at all segmental levels. Foxp1 is expressed at high levels in LMC neurons and reduced levels in PGC neurons. **(B)** *Hox* regulatory interactions involved in motor column specification. Downstream targets of Hox proteins in motor columns are shown. Peripheral targets of motor columns are also indicated. **(C)** Axial motor column markers and innervation pattern. MMC neurons express Lhx3 and Evi1, while no definitive embryonic TF markers for HMC neuron are currently known. **(D)** Organization of motor pools and innervation pattern of mouse forelimb at e12.5 (based on [20]). Motor pools expressing the TFs Pea3 (also known as Etv4) and Scip (Pou3f1) are shown, and each color-coded pool corresponds to a specific axonal trajectory in the limb. **(E)** Hox expression in brachial LMC neurons and regulatory interactions involved in motor pool specification. Hoxc6 and Hoxc8 function in conjunction with Foxp1 to promote expression of *Ret* and *GFRα1*, which are expressed in Pea3^{+} motor pools. A subset of Scip^{+} MNs express *Fign* and *Cpne4*, and target forelimb digit muscles [107].

functions [54,37,81].

2.1. Vertebrate Hox clusters and early neural expression pattern

In tetrapods, *Hox* genes reside within four chromosomally-arrayed *Hox* clusters termed *HoxA, HoxB, HoxC,* and *HoxD*. Due to an additional round of genome duplication, zebrafish, salmon, pufferfish and medaka, have an average of 7 clusters [164]. The ordering of individual *Hox* genes within a cluster is linked to the timing of their induction and spatial domains of expression. This correlation is termed temporal and spatial collinearity, with early- and rostrally-expressed *Hox* genes positioned 3' in the cluster, while late- and caudal-expressed *Hox* genes existing 5' [53,71]. A prevailing model is that *Hox* gene clustering arose from the necessity to sequentially activate individual genes along the chromosome, reflecting the progressive opening of repressive chromatin structure [52,147]. Interestingly, in some species where *Hox* genes lack a clustered organization, such as in the larvacean tunicate *Oikopleura dioica*, some aspects of temporal and spatial collinearity are retained [53,125].

The function of *Hox* genes during embryonic development has been intensely studied in both vertebrates and invertebrates, with mutation of individual *Hox* genes often having a profound impact on the formation of spatially resolved structures [127,103,102]. The majority of *Hox* genes are expressed within the developing CNS, and the activity of over a dozen *Hox* genes are involved in the diversification and synaptic specificity of spinal MN subtypes [127]. *Hox1-Hox4* paralogs are expressed in the developing hindbrain, where their activity patterns the development of transiently-segmented structures called rhombomeres [123]. In regions of the neural tube that eventually give rise to the spinal cord, *Hox4-Hox13* genes are expressed.

2.2. Hox function in spinal MN diversification and connectivity

In the spinal cord, *Hox* genes determine the subtype identity and connectivity of segmentally-restricted MN subtypes [31]. Here we provide a brief summary of the known functions of Hox proteins during two key steps in MN diversification – the formation of motor columns and motor pools. Motor columns are longitudinally arrayed groups of MNs that project their axons to a specific region (e.g. limb or axial muscle) (Fig. 1 A,B). Within each motor column, MNs further segregate into motor pools, each pool innervating a single muscle. Motor columns and pools are generated within defined rostrocaudal positions, and multiple columns and pools can occupy a single segment. While most MNs rely on *Hox* function in tetrapods, MNs targeting axial muscle (HMC and MMC neurons) do not appear to rely on Hox function for their specification (Fig. 1C).

Hox proteins promote MN diversification by regulating expression of other fate determinants and can act in concert with additional TF classes, including Lim-, Mnx-, and Pbx homeodomain factors. A key direct target of Hox proteins in MNs is the transcription factor *Foxp1*, which is induced at high levels by multiple Hox proteins expressed by limb-innervating LMC neurons, and at reduced levels by *Hoxc9* in thoracic preganglionic column (PGC) neurons (Fig. 1A,B) [73,75,86,33]. Mutation of *Foxp1* in mice leads to a loss of molecular signatures of Hox-dependent MN columnar and pool subtypes, and reversion of MN identities to the more ancestral axial MN fate [35,139]. Subsequent to its induction by Hox proteins, Foxp1 also acts act in conjunction with Hox proteins in LMC neurons to promote motor pool fates (Fig. 1E).

In forelimb LMC neurons, multiple genes in the *Hox4-Hox8* paralog groups are involved in specifying motor pools. For example, *Hoxc8* is essential for establishing the molecular identities and connectivity of MN pools targeting distal forelimb muscles. *Hoxc8* contributes to motor pool fates by regulating expression of *Ret* and *GFRα* surface receptor genes that are essential for proper forelimb innervation (Fig. 1D, E) [20]. Individual *Hox* genes and paralog groups also play more restricted roles in MN diversification. For example, two *Hox5* paralogs (*Hoxa5* and *Hoxc5*) are essential for the development of phrenic MNs targeting respiratory muscle [128]. Recent studies indicate Hox5 proteins regulate expression of cadherin proteins involved in phrenic MN clustering and connectivity to premotor respiratory networks (Fig. 1B) [156]. As described later, loss of *Hox* function also can also lead to MN fate transformations through derepression of *Hox* genes normally expressed in adjacent segments.

Hox proteins also act in conjunction with more broadly expressed cofactors to promote MN fate specification. Members of the Pbx family are well known cofactors for Hox proteins, which enhance the affinity and specificity of Hox proteins to target sites [9]. After deletion of *Pbx* genes (*Pbx1* and *Pbx3*) from MNs in mice, Hox-dependent subtype features are lost [62]. *Pbx* mutants are therefore similar to *Foxp1* mutants, but appear to affect a broader range of MN subtypes, including phrenic MNs. Pbx TFs also have Hox-independent functions that promote the organization of axial MNs, as *Pbx* mutants are characterized by an erosion in axial MN molecular signatures and somatotopic organization [62].

3. Establishing presumptive *Hox* boundaries in neural progenitors

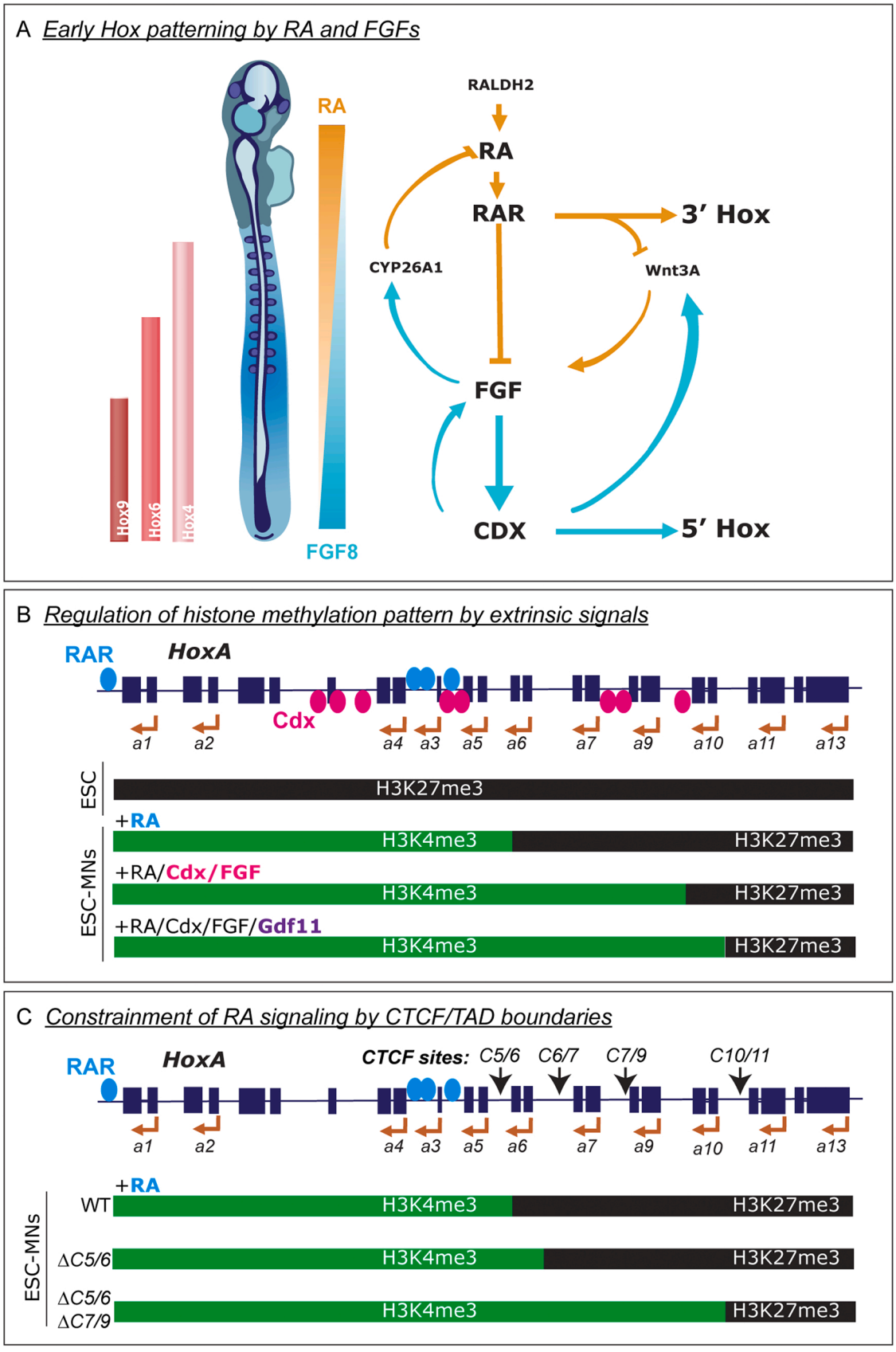
In order to specify MN fates, *Hox* expression must be precisely regulated during development. Immediately following gastrulation, *Hox1-Hox3* paralogs are expressed in the mesoderm and weakly expressed in the epiblast at the caudal section of the primitive streak [36,170]. As the embryo lengthens, expression of these early *Hox* genes extends rostrally, eventually reaching their definitive boundaries in the neuroectoderm, mesoderm, and endoderm. The initial induction of *Hox* genes, and therefore the axial patterning of the embryo, appears to occur prior to neurogenesis [108]. The exact rostral limit of *Hox* expression is more rostral for 3' genes than 5'. This rostral extension of the *Hox* domains results in an overlapping pattern of expression which is further refined, to produce demarcated *Hox* expression domains (Fig. 2A). In the following subsections, we will discuss how signaling from extrinsic morphogens, regulation of chromatin structure, and intrinsic factors establish presumptive *Hox* domains in the neural tube during axial elongation.

3.1. Regulation of axial positional identities by morphogens during neurogenesis

In neural progenitors, retinoic acid (RA) signaling functions to promote the expression of *Hox1-Hox5* paralogs, while fibroblast growth factor (FGF) in conjunction with Wnt signaling initiates caudal *Hox* expression. The synthesis of RA during development is regulated by Retinaldehyde Dehydrogenase 2 (RALDH2), which is expressed in trunk somites and presomitic mesoderm [115]. In addition, RA is degraded by Cytochrome P450 26 proteins (CYP26A1/B1/C1) expressed in the tailbud, creating a rostral-caudal gradient of RA activity along the neural tube [8]. *Fgf8* is transcribed in the tail bud and generates a caudal-rostral gradient through mRNA decay [44]. FGF signaling also restricts RA synthesis by activating *CYP26A1* expression [116], preventing RA signaling in regions of high FGF [11,41]. In rostral regions, RA inhibits the expression of *FGF8* through binding of retinoic acid receptors (RARs) at regulatory sequence upstream of the gene [83,41]. RA and FGF therefore sets up rostrocaudal patterning through reciprocal cross-regulatory interactions [39], providing the necessary inductive signals to establish the early pattern of *Hox* expressions (Fig. 2A).

3.2. Retinoic acid and rostral patterning of Hox expression in spinal progenitors

The initial rostral expansion of *Hox* expression in the neural tube is driven by the activity of RA, which is essential for many developmental processes including rostrocaudal patterning, somitogenesis, and neural development [67,72]. Following synthesis of RA and diffusion to



(caption on next page)

Fig. 2. Regulation of *Hox* expression and chromatin structure by extrinsic cues. **(A)** During vertebrate axial elongation, *Hox* genes are sequentially induced by RA and FGF. In neural progenitors RA functions to regulate *Hox1-Hox5* genes in rostral segments while FGF regulates *Hox6-Hox9* genes in caudal regions. Stage 12 chick embryo shown. Regulatory interactions between RA and FGF signaling pathways are shown on the right. **(B)** Model for the effects of morphogens on distribution of histone marks associated with active (H3K4me3) and repressed (H3K27me3) *Hox* genes in ESC-MNs. In general, the presence of H4K4me3 is associated with *Hox* expression, while H3K27me3 with *Hox* repression. Signaling through RA and FGF depletes H3K27me3 from *Hox* genes and establishes presumptive rostrocaudal expression boundaries. Genes in the *HoxA* cluster are shown. RA clears H3K27me3 marks from *Hox1-Hox5* genes through RAR-RXR recruitment. Binding of CDX following FGF treatment results in the clearance of H3K27me3 from *Hox6-Hox9* genes. Gdf11 is likely required to clear H3K27me3 from more caudal *Hox10* genes. Approximate binding sites for RAR and CDX are shown. **(C)** Role of CTCF in insulating the effects of RA in H3K27me3 clearance. CTCF sites in the *HoxA* cluster are shown. Mutation in CTCF binding sites results in an extension of H3K4me3 signal at more 5' ends of *HoxA* cluster after RA treatment in ESC-MNs, and a concomitant spreading of activating H3K4me3 marks. Changes in chromatin marks after CTCF site mutation results in ectopic expression of progressively more 5' *HoxA* genes.

neighboring cells, RA acts as a ligand for RARs which heterodimerize with Retinoic X Receptors (RXRs), both of which are widely expressed in many tissues [40,76,101]. In the presence of RA, RAR-RXR heterodimers interact with co-activator complexes at Retinoic Acid Response Elements (RAREs) to activate gene expression. These RAREs are exemplified by direct repeats of the hexameric "(A/G)G(G/T)TCA", often found in close proximity to promoter regions of target genes [87,101]. Once activated by RA and bound to RAREs, RAR/RXRs can initiate transcription through the interactions with general coactivators and RNA Polymerase II machinery.

Genes within *Hox* clusters are differentially susceptible to RA concentration, with genes such as *Hoxa1* rapidly induced following RA induction and more caudal *Hox* genes responding slower. RAREs have been found proximal to *Hox1-Hox5* genes [117,105,118]. They primarily function to promote the expression of nearby *Hox* genes, but in the developing heart and gut RAREs flanking the *HoxB* cluster have been found to distally regulate *Hox* expression over longer genomic distances [119,134].

The mechanisms of gene activation following RA induction is varied. During hindbrain patterning, RARs associated with the promoter of *Hoxa1* and in the absence of RA pause RNA PolII and prevent transcription [50]. Activation of RARs by RA results in the unpausing of RNA PolII and the fast release of repression. Contrasting this, RARs are only recruited to the RARE proximal to *Hoxb1* in the presence of RA, thereby resulting in a slower onset of transcription compared to *Hoxa1* [97]. Preliminary evidence suggests that regions flanking the core RARE sequence dictate enhancer strength, and thus the differential sensitivity of *Hox* groups to RA concentrations.

3.3. Roles of RA and TAD boundaries in regulating Hox expression in MN progenitors

The induction of *Hox1-Hox5* genes by RA in ESC-derived MNs (ESC-MNs) is associated with the clearance of Polycomb-associated repressive H3K27me3 histone marks and presence of Trithorax complex-associated activation histone marks (H3K4me3) (Fig. 2B) [105,112]. The extent of clearance of PRC repressive marks is constrained by the position of binding sites for CTCF (CCCTC-binding factor). CTCF is an evolutionarily conserved DNA binding protein that localizes to borders between topologically-associated domains (TADs), mega-base regions of localized chromatin interactions [42]. Although the precise functions of TADs in regulating gene expression are unclear, CTCF binding at TAD boundaries appears to act as an insulator that prevents interactions between enhancer and repressor elements in neighboring TADs [6]. Studies in the developing mouse limb bud support an essential role for CTCF function and its binding sites in regulating the pattern of *Hox* expression [137,148].

Studies in ESC-MNs have examined the role of CTCF sites in the regulation of *Hox* expression during neural patterning. Deletion of a CTCF binding site between *Hoxa5* and *Hoxa6* (C5|6) results in the ectopic expression of *Hoxa7* following RA induction. Combined deletion of the *Hoxa5|6* site and a CTCF site located between *Hoxa7* and *Hoxa9* (C7|C9) leads to the loss of the C5|6 TAD boundary and ectopic expression of *Hoxa7-Hoxa10* genes (Fig. 2 C) [112,111]. Similar deletion of CTCF sites from the *HoxC* cluster result in ectopic expression of *Hoxc6* and *Foxp1* in ESC-derived MNs [111].

Studies performed in vivo also support a role for CTCF sites in *Hox* gene regulation. Mutation of CTCF binding sites in the *HoxA* and *HoxC* clusters leads to homeotic transformations of axial skeletal elements, confirming that CTCF is required for *Hox*-dependent body patterning [111]. A participating cofactor in the CTCF complex, Myc-Associated Z-Finger (MAZ) has also been found to directly interact with CTCF in ESC-MNs, and is required for the demarcation of active-inactive compartments of the *HoxA* cluster [121]. These results indicate that transcriptional activation of *Hox* genes following RA treatment is constrained by the activity of CTCF and MAZ, preventing large-scale derepression of the *Hox* clusters.

3.4. FGFs induce caudal Hox genes in spinal progenitors via Cdx proteins

In addition to the regulation of *Hox1-Hox5* paralog groups by RA, FGFs provide a caudalizing signal for patterning *Hox* expression in the neural tube. FGFs are known to be involved in many aspects of early vertebrate development, including endoderm formation, gastrulation, and neural induction [10]. Early evidence supporting a role for FGF in caudalization of *Hox* patterning came from studies in *Xenopus*, where FGFs were shown to upregulate caudal *Hox* gene expression in a dose-dependent manner [88]. Higher concentrations of FGF results in the expression of progressively 5' *Hox* genes [133,79,29]. Furthermore, expression of a dominant-negative FGFR results in the reduction of *Hoxb9* expression and defects in caudal CNS development [57]. In mice, hypomorphic mutants of FGFR1 results in caudal shifts in expression of *Hoxd4* and *Hoxb9* [124].

In explants of chick neural progenitors, addition of FGF8 is sufficient to promote expression of *Hoxc6, Hoxc8,* and *Hoxc9* in MNs in a dose dependent manner [99], indicating that *Hox* genes are sensitive to graded level of FGF. Similarly, adding FGFs to cultured chicken embryos results in a rostral shift of *Hoxb9* expression [7]. Elevation of FGF8 signaling in vivo extends the rostral limits of *Hoxc9* expression into brachial-level MN progenitors, which in turn inhibits the expression of *Hoxc6* [33]. The transformation of *Hox* pattern after FGF elevation also leads to a switch in MNs fates in brachial segments, causing a loss of the LMC marker *Raldh2* and a gain of the PGC marker *Bmp5* [33]. Thus, elevation of FGF8 expression results in the transformation of brachial spinal MNs into a thoracic identity.

The induction of caudal *Hox* paralogs by FGF is mediated through the activity of Cdx homeodomain proteins. FGF signaling promotes expression of *Cdx* genes in chick and *Xenopus* through cooperation with Wnt and MAP kinase pathways [7,78,120,171]. Overexpression of an activated form of the *Xenopus* homolog of *Cdx, XcadVP16,* results in an upregulation of *Hoxb9* and rostral extension of its expression domain [7]. By contrast, expression of a dominant negative allele of *Xcad3* abrogates activation of *Hoxb9* following FGF treatment, showing the role of the *Cdx* family downstream of FGF [68].

Studies in ESC-MNs indicate that *Cdx2* directly binds at caudal *Hox* genes and its binding is associated with the removal of H3K27me3 repressive marks from *Hox6-Hox9* genes in the presence of FGF (Fig. 2B) [113,105]. Cdx/FGF signaling also decompacts *Hox* clusters and enables the deposition of the activating H3K27ac signal [105,113,114]. Complete removal of H3K27me3, and thus full activation of more caudal *Hox*

paralogs, likely requires additional patterning signals. These signals include *Gdf11*, which has been shown to be necessary for activation of *Hox10* genes chick explants, mouse embryos, and human ESC-MNs [143, 98,110]. This work indicates that extrinsic signals and *Cdx* proteins regulate the chromatin landscape of *Hox* clusters, removing inhibitory marks and making *Hox* targets available for transcriptional activation.

In addition to the binding of Cdx and RAR within *Hox* clusters, distal regulatory control regions, and topological DNA organization have been implicated in the regulation *Hox* expression [30]. Defining the specific contributions of local and long-range regulatory elements in regulating *Hox* expression and chromatin structure has been challenging to disentangle. To tease apart the functional requirement for specific genomic elements, a synthetic construct of the rat *HoxA* cluster has recently been developed for fine control of *cis*-regulatory elements within the cluster [129]. Consistent with previous studies, deletion of RAREs within the 3'-region of the *HoxA* cluster prevented removal of H3K27me3 marks after RA treatment, and *Hox1-Hox5* failed to be activated. By contrast, the absence of distal enhancers had no effect on chromatin remodeling in response to RA, but were required for high levels of *Hox* gene expression. These studies support a model in which internal elements within the *Hox* clusters, such as local RAR, Cdx, and CTCF binding sites, are necessary to establish appropriate patterns of *Hox* gene expression, while distal enhancers are required for robust levels of gene expression.

4. Hox cross-repressive interactions and establishment of postmitotic MN positional identities

Following the morphogen-induced patterning of *Hox* expression in neural progenitors, there is initially considerable overlap between the domains of *Hox* expression in caudal segments of the neural tube. In order to correctly specify neuronal subtype identities, *Hox* gene expression is further restricted along the rostrocaudal axis and within a single segment. This pattern of refinement is driven through direct cross-regulatory interactions between Hox proteins and *Hox* genes and appears to occur in postmitotic neurons.

Pairs of *Hox* genes display mutually-exclusive patterns of expression in MNs along the rostrocaudal axis, similar to the boundaries of TF expression established in spinal progenitors along the dorsoventral axis [33,34,13]. For example, the boundary between brachial and thoracic segments is established through cross-repressive interactions between *Hoxc6* and *Hoxc9,* which demarcates the positional boundary between limb-innervating LMC and thoracic PGC neurons (Fig. 3A). Ectopic postmitotic expression of *Hoxc9* in brachial segments can inhibit expression of *Hoxc6* and other brachial-expressed *Hox* genes, while mutation in *Hoxc9* in mice results in derepression of brachial *Hox* genes in thoracic segments [75]. Depression of brachial *Hox* genes in *Hoxc9* mutants leads to an extension of forelimb-innervating MN subtypes into thoracic segments. Interestingly, in vertebrates which normally lack brachial limb MN subtypes, such as in limbless snakes, the pattern of *Hoxc9* extends into rostral spinal cord and is associated with an absence of *Hox4-Hox8* paralog expression by MNs [73]. Although the precise

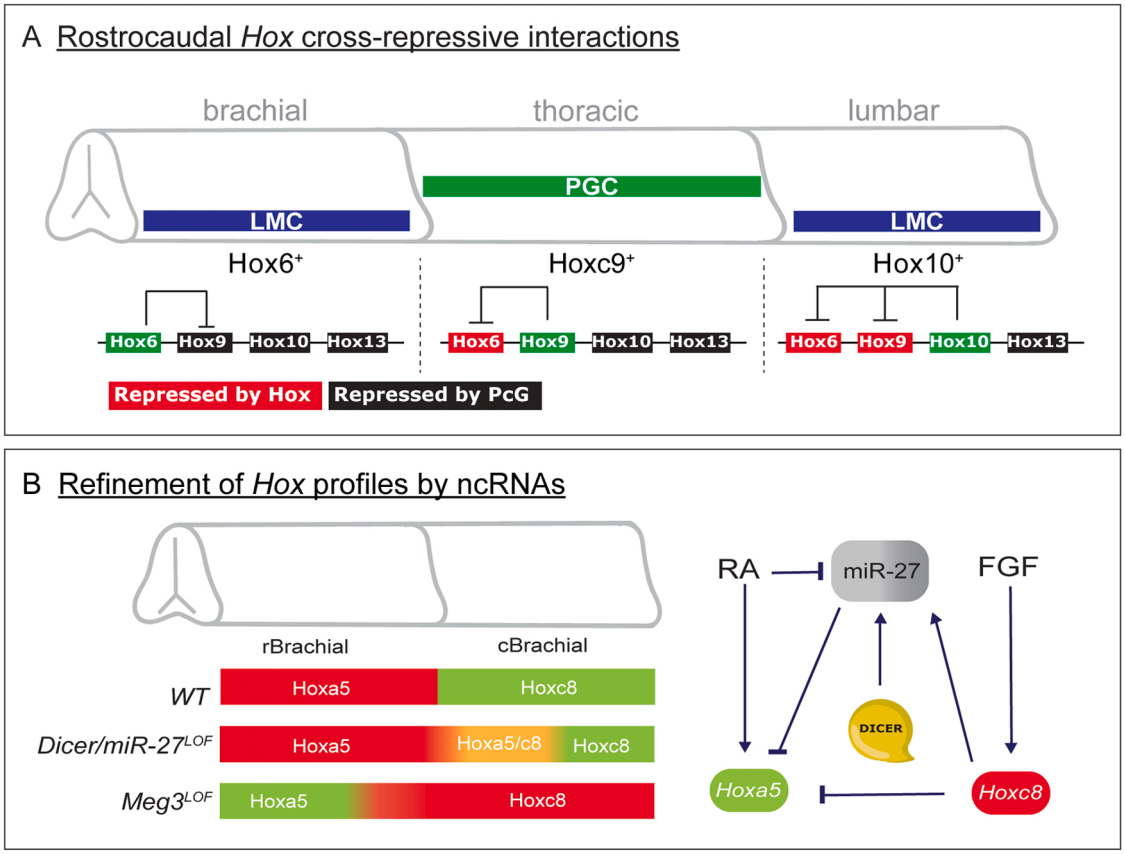


Fig. 3. Establishment and refinement of *Hox* boundaries in postmitotic MNs. (A) *Hox* boundaries along the rostrocaudal axis are established through cross-repression and PcG-mediated silencing. For simplicity only *Hox6*, *Hox9*, *Hox10*, and *Hox13* paralogs are shown. Limb-innervating LMC neurons and thoracic PGC neurons are shown. Cross-repressive interactions appear to be conserved among paralogs (e.g. Hoxa9, Hoxb6, Hoxc9, and Hoxd9 all repress *Hoxc6*). At thoracic levels, Hox9 proteins repress brachial *Hox6* genes, while more caudal *Hox10* and *Hox13* genes are silenced by PcG proteins. **(B)** The activity of miRNAs fine-tunes expression of *Hoxa5* and *Hoxc8* at the boundary between rostral (r) and caudal (c) brachial segments. Depletion of miR-27 or Dicer function leads to erroneous expression of *Hoxa5* in caudal Hoxc8^{+} brachial segments. Loss of the lncRNA *meg3* in MNs results in the expansion of *Hoxc8* into rostral brachial segments and reduced Hoxa5 expression.

mechanisms that determine the expanded domain of *Hoxc9* in snakes is unclear, these observations suggest that changes in *Hox* gene expression contributed to the evolution of spinal MN organization. This idea is further supported by studies in the little skate *Leucoraja erinacea*, which lack the *Hoxc9* gene, and display a caudally expanded domain of fin-innervating LMC neurons [74].

Cross-repressive interactions can also occur within a single segmental level of the spinal cord. Cross-repressive interactions between *Hox4-Hox8* genes determine identity of motor pools within the brachial LMC [34,20]. However, not all intrasegmental Hox cross-repressive interactions give rise to mutually exclusive expression of *Hox* genes. Interactions between Hoxc6, Hoxc8, and Hoxc9 allow for the specification of a subtype of LMC neuron that innervate forelimb digits (Fig. 1E) [107]. These MNs lack the canonical limb MN determinant Hoxc6, and express low levels of Hoxc9 and Hoxc8, possibly reflecting differential weighting in the repressive activity of Hoxc9 towards the *Hoxc8* and *Hoxc6* genes.

Studies in ESC-MNs indicate that cross-repressive activities are mediated through direct interactions of Hox proteins on *Hox* genes. Chip-seq studies in ESC-MNs show that Hoxc9 binds to genomic regions near *Hox4-Hox8* paralogs [75,14]. Misexpression of more caudally-expressed *Hox* genes such as *Hoxc10* or *Hoxc13* results in the repression of a larger group *Hox* genes. For example, misexpression of Hoxc13 in chick represses *Hox4-Hox10* paralogs, and binds at additional 5'-sites within *Hox* clusters [14,140]. The extent to which *Hox* genes are repressed therefore appears to correlate with cluster position, with proteins encoded by 5'-*Hox* genes repressing larger numbers of 3'-*Hox* genes. This ability of caudal *Hox* genes to repress more rostral *Hox* genes is reminiscent of a phenomena first described in *Drosophila* termed posterior dominance or phenotype repression [43].

5. Post-transcriptional regulation of *Hox* genes by non-coding RNAs

The pattern of *Hox* expression in MNs can also be fine-tuned by non-coding RNAs that post-transcriptionally regulate specific *Hox* genes. A major class of ncRNAs, long non-coding RNAs (lncRNAs), have been implicated in *Hox* gene regulation. One of the first lncRNAs reported to affect *Hox* expression in vertebrates is *HOTAIR,* transcribed from the *HoxC* cluster [135]. Following deletion of the *HOTAIR* in fibroblasts, *HoxD* cluster paralogs are derepressed [95]. *HOTAIR* has been shown to associate with PRC2, suggesting this lncRNA functions to recruit repressors to *Hox* loci [165]. Despite this, mouse mutants of *HOTAIR* display no detectable change in embryonic *HoxD* expression [1], and *HOTAIR*-mediated repression can occur independently of PRC2 activity [132], bringing into question the significance of *HOTAIR* for developmental patterning.

A lncRNA produced from the *Dlk1-Dio3* locus, *meg3,* is highly expressed in postmitotic brachial MNs [168]. *Meg3* associates with the Jarid2/PRC2 complex to inhibit the expression of caudal *Hox* genes [168]. Knockdown of *meg3* on ESC-MNs results in a reduction in PRC2-deposited H3K27me3 mark and an upregulation of *Hox8-Hox13* paralogs. In mice lacking *meg3* activity, there is derepression of *Hoxc8* in rostral-brachial segments, and an expansion in the motor pools normally located in caudal-brachial segments (Fig. 3B). Interestingly, more caudal *Hox* genes, such as *Hox9* and *Hox10* genes, are not affected by *meg3* depletion [168], suggesting segment-restricted functions for lncRNAs in refining *Hox* patterns in MNs.

A second class of ncRNAs, microRNAs, are 22-bp single stranded RNA molecules which function to post-transcriptionally regulate gene expression. In *Drosophila,* several miRNAs that inhibit *Hox* expression have been identified, including *miR-iab-4–5p,* mapped upstream of the *Abd-A* gene in the *BX-C* cluster [153]. Expression of miRNAs is often restricted to the same spatial domains as their coding *Hox* counterparts [2]. *miR-iab-4–5p* has been found to be functionally conserved in vertebrates [153], and its overexpression in *Drosophila* phenotypically mirrors loss of *Ubx* [138]. This pattern of *Hox*-repression by miRNA is consistent with other miRNA sequences found associated to the *BX-C* cluster in *Drosophila,* such as *miR-iab-9–5b*, mapped upstream of the *Abd-B* locus. Due to their close proximity to the coding sequence of the *Hox* homologs, it is thought that these miRNAs are regulated in tandem with their gene targets [162]. *miR-iab-4–5p* and *miR-iab-9–5p* both inhibit the activity of rostrally-expressing *Ubx, Abd-A,* and *Antp*. This raises the possibility that the generation of miRNAs in caudal regions facilitates the posterior dominance and refinement of *Hox* domains [146].

The posttranscriptional regulation of *Hox* genes by microRNAs is conserved in mammals. In mammalian cell lines overexpressing *miR-196* (analogous to *iab-4*), a miRNA associated with the *HOXB8* locus, show repression of the *HOXB8* transcript [167]. In ESC-MNs, loss of the miRNA synthesizing enzyme *Dicer* results in precocious and noisy expression of Hoxa5 protein and an erosion of the normal Hoxa5/Hoxc8 boundary within brachial segments, indicating miRNAs are involved in the tempering of *Hox* expression (Fig. 3B) [93]. Depleting the function of *miR-27* in both mESCs and chick embryos leads to ectopic Hoxa5 expression in the *Hoxc8* expression domain, mirroring the disrupted *Hoxa5-Hoxc8* boundary observed in *Dicer* mutants.

6. Polycomb group proteins and the regulation *Hox* expression

Due to their integral role in neuronal subtype diversification, the timing of *Hox* gene expression must be tightly regulated in both space and time. Prior to gastrulation, as well as in pluripotent embryonic stem cells (ESCs), *Hox* clusters are compacted and transcriptionally silent (Fig. 4A) [12,109]. The early silencing of *Hox* genes, and the eventual maintenance of *Hox* repression, is regulated by histone-associated proteins in the Polycomb group (PcG) family. In *Drosophila*, mutation of the *Polycomb* (*Pc*) PcG protein results in misexpression of *Bithorax* genes and homeotic transformation of body pattern, providing evidence that PcG proteins function to silence *Hox* expression [92]. Further characterization of PcG proteins in *Drosophila*, mouse, and other model systems solidified an evolutionary conserved role in the developmental regulation of *Hox* expression [56,90].

PcG proteins associate to form two histone-modifying Polycomb Repressive Complexes (PRCs), PRC1 and PRC2. The activity of these complexes allows for the binding, methylation, and compaction of target histones, resulting the transcriptional silencing of genomic loci [100, 145]. In vertebrates, PRC2 can be recruited to target histones by Polycomb-Like (PCL) proteins, often in the presence of unmethylated CpG islands [82,94]. Once associated with a locus, Ezh1/2 constituents of PRC2 methylate histone H3 at lysine 27 (H3K27me1/2/3). Mono- and bi-methylation of H3K27 promote further methyltransferase activity through a positive-feedback loop driven by the core PRC2 subunits Eed and Suz12 [25]. The H3K27me3 mark recruits PRC1 to a target locus, through interaction with the Cbx protein constituent of PRC1. Once bound, Ring1A/B proteins ubiquitinate Histone H2A at lysine119 (H2K119ub), while PHC polymerization contributes to spreading PRC1, thereby increasing the presence of the complex at a locus [55,69,70]. The activities of the PRCs lead to the compaction of associated chromatin, the inhibition of RNA Pol II transcriptional initiation/elongation, and the transcriptional silencing of resident genes [151].

6.1. Early silencing of Hox loci by PcG proteins

PRC activities are essential for the controlled silencing of *Hox* clusters, until the correct timing of induction during development. Depletion of PRC2 activity in ESCs results in the loss of H3K27me3 marks in *Hox* clusters and ectopic *Hox* gene expression. Removal of PRC1 function during these early stages results in a similar decondensation of the *Hox* clusters, and loss of transcriptional silencing [47,159]. Mouse mutants of core PRC1 and PRC2 components are embryonic lethal, often stalling development during early gastrulation [48,158]. Mutation of

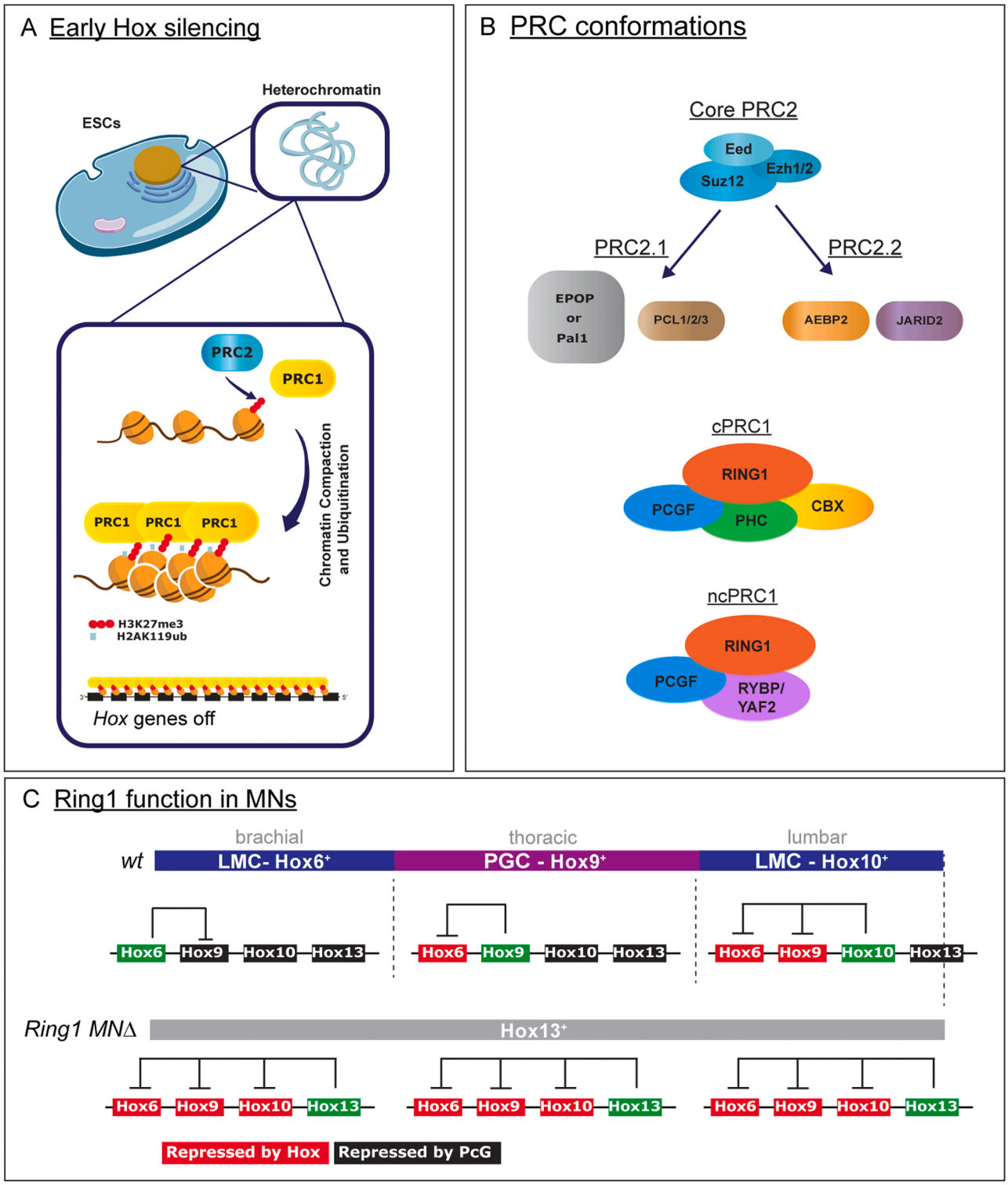


Fig. 4. Diversity and function of PRCs during MN differentiation. (A) In early embryos and mouse embryonic stem cells, *Hox* clusters are kept transcriptionally silent through PRC1 and PRC2, which modify histones and compact chromatin. **(B)** Alternate PRC2 and PRC1 subunit compositions. PRC2 can be subdivided into PRC2.1 and PRC2.2. Canonical (c) PRC1 contains PHC and CBX proteins while non-canonical (nc) PRC1 contains RYBP or YAF2. **(C)** Function of Ring1 during MN differentiation. In the absence of *Ring1* function (*Ring1A* and *Ring1B*), *Hox13* genes are ectopically expressed in rostral spinal segments, leading the repression of *Hox4-Hox10* gene expression. For simplicity only *Hox6*, *Hox9*, *Hox10*, and *Hox13* paralogs are shown.

sub-stoichiometric PRC components display later developmental defects and homeotic transformations, suggesting a cell-type specificity or temporal necessity of PRC constituents [19,23,58,61,70,69,94]. Core PRC subunits are therefore broadly required during early development, while specific interacting cofactors appear to enable cell-type specific control of gene expression in a variety of contexts.

Following induction of *Hox* expression in neural progenitors, and refinement of rostrocaudal boundaries through cross-repression, the patterns of *Hox* expression are maintained in newly postmitotic neurons. As Hox proteins regulate key target effectors in postmitotic neurons, the

profiles of *Hox* expression must be preserved through the transition from progenitors to terminal fates. Studies in *Drosophila* indicate the inheritance of positional identities ("epigenetic memory") are facilitated through PRC activities [24,26,28]. Below we discuss the mechanisms of PRC-mediated repression in vertebrates, their various accessory subunits, and the known functions of PRCs in the regulation of *Hox* gene expression in the CNS.

6.2. Mechanisms of PRC2 function in Hox regulation

The core of PRC2 consists of three subunits: Enhancer of Zeste 2 (Ezh2 or its paralog Ezh1), Suppressor of zeste 12 (Suz12), and Embryonic ectoderm development (Eed) [100]. Ezh2 is the main catalytic constituent of PRC2, while its Ezh1 paralog has weaker activity and more restricted expression. Ezh2 functions as a histone methyl transferase (HMTase) to catalyze the methylation of H3K27, allowing for recruitment PRC1 to genomic loci. Both Suz12 and Eed positively regulate the HMTase activity of Ezh1/2. Suz12 serves a structural function, stabilizing the PRC2 complex and maintaining HMTase activity [126,66]. Eed promotes Ezh1/2 activity through a positive-feedback loop, recognizing and binding to H3K27me3 signals and allosterically activating the HMTase activity of PRC2 [104].

Additional cofactors can associate with PRC2 and modulate its activity (Fig. 4B). PRC2.1 contains Polycomb-like proteins 1–3 (PCL1–3), which are thought to aid in the initial recruitment of PRC2 to loci [16,4]. Loss of PCL proteins in *Drosophila* and vertebrates results in homeotic transformations and *Hox* misexpression [45,96]. PRC2.2 is defined by the presence of Jumonji, AT rich interactive domain 2 (JARID2) and adipocyte enhancer-binding protein 2 (AEBP2). JARID2 and AEBP2 both work to recruit PRC2 to chromatin and associate with PRC1-deposited H2AK119ub, facilitating cross-talk between the PRC1 and PRC2 [89,27]. Using separation-of-function mutants of SUZ12 in induced pluripotent stem cells (iPSCs), it has been found PRC2.2 associates with target genes at lower levels than PRC2.1, and enrichment of relative PRC2.2 abundance is paired with an upregulation of *Hox* target genes [169]. The exact functional relationship between PRC2.1 and PRC2.2 conformations is currently unknown, but they appear to act in both synergistic and independent fashions to deposit H3K27 methyl marks on target genes, and are both required for efficient silencing of *Hox* genes [63].

The role of PRC2 in the establishment and maintenance of neuronal identities has been explored in vertebrates through deletion of core subunit-encoding genes. Removal of PRC2 function can lead to defects in the temporal transition from neurogenesis to gliogeneis, neuronal subtype specification programs, and repression of *Hox* genes [38,64,161, 166,141]. In spinal MNs, PRC2 function has been examined through deletion of *Eed* from neural progenitors and through combined deletion of both *Ezh1/2* genes in MNs. Neural-specific deletion of these core PRC2 constituents in mice does not have a noticeable effect on embryonic *Hox* pattern or MN subtype diversification programs [140]. The absence of an observable phenotype in PRC2 mutants may reflect residual H3K27me3 carried over from earlier stages, or a more pronounced reliance on PRC1 for maintaining postmitotic *Hox* expression.

6.3. PRC1 is essential to maintain Hox pattern in MNs

PRC1 conformations are also structurally and functionally diverse (Fig. 4B). PRC1 conformations can be broadly classified as either canonical (cPRC1) or noncanonical PRC1 (ncPRC1), distinguished by their specific constituents. The core proteins of PRC1, which are shared in cPRC1 and ncPRC1, include Ring1A/B and one of six Polycomb Group ring fingers 1–6 (PCGF1–6). Ring1A/B is an E3 ubiquitin-ligase, which deposits H2AK119ub on target histones. There is uncertainty regarding the function of Ring1 ubiquitination activity for *Hox* gene regulation. Early silencing of *Hox* clusters appears independent of Ring1 ubiquitin-ligase activity, as enzymatically inactive mutants of Ring1B are capable of rescuing *Hox* silencing in *Ring1B*-null backgrounds [47,155]. Recent work has suggested that H2AK119ub is required in ESCs and neural progenitors to temporally repress neuronal genes, but is dispensable for long-term silencing of *Hox* targets [155]. Other reports show that the ubiquitin-ligase activity of Ring1B is required for the maintenance of ESC identity [46].

In canonical PRC1, Ring1A/B associates with a PCGF, one of three Polyhomeotic-like (PHC1–3) proteins, and a CBX protein (CBX2/4/6–8). CBX proteins contain a chromodomain which recognizes H3K27me3 marks and is involved in the recruitment of PRC1 to target loci [77]. CBX proteins have also been found to possess chromatin-compaction activities and are necessary for *Hox* repression [60,91]. Once bound to H3K27me3, polymerization of PRC1 subunits results in the spreading of PRC1 occupancy across a locus, the formation of discrete genomic domains of high interaction, and the generation of phase-separated PcG conglomerates (termed PcG bodies) [69,70,136,163]. Formation of PcG bodies involves polymerization of PHC proteins via its SAM-domain as well as through Cbx proteins [70,131]. The formation of tightly-compacted loci are inaccessible to activating transcriptional machinery.

Due to compaction, genomic loci associated with PRCs become physically clustered and tightly associated. The linkage of genomically distant sites allows for increased coordinated regulation of the resident target genes. Through chromatin conformation analyses, it has been shown that sites enclosed within PcG bodies preferentially interact in discrete domains [84,154,5]. In mice, these domains demarcate active vs. inactive genomic regions, and following RA-differentiation they separate rostral and caudal-expressing *Hox* groups into two associated populations which are differentially regulated [22].

Non-canonical PRC1 configurations contain RYBP (Ring1 and YY1 binding protein) or YAF2 (YY1 associated factor 2). RYBP can compete for the same binding pocket on Ring1B, and displace CBX from Ring1 [160]. The presence of RYBP in ncPRC1 stimulates the E3 ubiquitin-ligase activity of Ring1A/B [51], and the majority of H2AK119ub genomic signals are thought to be deposited by ncPRC1 [85]. Despite this, the significance of H2AK119ub deposition and ncPRC1 activity for the regulation of *Hox* transcription during embryonic development is currently unknown. The function of ncPRC1 in specification of spinal MN subtypes also appears to be minimal. While global *Rybp* mutant mice are embryonic lethal [130], combined conditional deletion of *Rybp* and *Yaf2* from MN progenitors show no obvious changes in *Hox* expression or MN subtype differentiation programs at embryonic stages [140].

In spinal MNs, cPRC1 plays a critical role in maintaining *Hox* patterns and ensuring the silencing of inappropriate gene programs. Depletion of the PCGF protein Bmi1 in chick embryos leads to ectopic expression of *Hoxc9* in brachial segments and a switch in the fate of brachial LMC neurons to a thoracic PGC fate [58]. In mice, mutation of *Ring1A/B* from MNs results in a widespread derepression of dozens of fate determinants, including caudal *Hox* genes [140]. Derepression of caudal *Hox* genes, in particular *Hox13* paralogs, leads to the suppression of *Hox4-Hox10* expression in MNs (through cross-repression), and a loss of all Hox-dependent columnar and pool subtypes (Fig. 4C). This indicates that in postmitotic MNs, cPRC1 contributes to the maintained repression caudal *Hox* genes. The degree to which postmitotic MNs rely on PRC2 function, and the specific contribution of PRC1 subunits CBX and PHC in local compaction and silencing of *Hox* genes, remains to be fully explored.

7. Future directions and remaining questions

Spatial and temporal control of *Hox* expression is necessary for body patterning and cellular differentiation. Work over the last decades has provided a wealth of information on the dynamic regulation of *Hox* expression during neural development, from early progenitor stages to maintenance in postmitotic neurons. Recent studies have also raised

important questions concerning how and when patterning signals operate to regulate *Hox* expression. A classic model for neural development is the "activation-transformation" model in which neural progenitors are initially specified with a rostral (brain) identity, which lack *Hox* gene expression, and are later transformed to a caudal (hindbrain/spinal cord) fate [149]. In vitro and in vivo studies of regions of condensed and open chromatin in neural progenitors indicate that opening of *Hox* regions precedes neural induction, suggesting that early patterning signals may act initially on progenitor cells prior to neural fate specification [108]. Further studies in vitro support a model in which the timing of *Hox* expression is governed by extrinsic patterning cues, as opposed to an intrinsic "Hox clock"-based timer [110]. The precise relationships between the timing of *Hox* induction, regulation of local chromatin structure, and mechanisms through which morphogens regulate *Hox* pattern remain to be determined.

Early programming of *Hox* profiles in stem cell populations prior to neurogenesis may serve to coordinate positional identities among neural and non-neural tissues. Patterning morphogens such as RA and FGF could also act at later stages to control more unique *Hox* codes in neural tissues. After rostrocaudal positional identities are established, *Hox* gene expression is further shaped along the dorsoventral axis. For example, early studies indicated that while *Hox* genes are initially uniformly expressed in neural progenitors, at later postmitotic stages *HoxB* genes become restricted to the dorsal spinal cord, while *HoxC* genes localize to ventral types, including MNs [59,99]. These later changes could reflect further modification to *Hox* clusters, mediated through developmental regulation of chromatin-modifying proteins. The abundance of sub-stoichiometric PRC components are dynamic during progressive stages of neural differentiation [80], with PRC2 proteins becoming markedly downregulating during NPC differentiation, while PRC1 remains bound to H3K27me3 loci. The differential function of PRC conformations at various stages of neural differentiation has proved challenging to decipher, due to the wide variety of interacting cofactors and the embryonic lethality of core PRC mutants. With the advent and improvement of genomic analysis methods, our ability to observe the topographical organization and compaction of *Hox* clusters has significantly improved. Correctly utilizing these techniques to study *Hox* cluster compaction, PcG body polymerization, and TF binding will be integral to fully resolve the dynamic regulation of *Hox* genes during nervous system development.

Declarations of Competing Interest

None.

Acknowledgements

This work was supported by NIH NICHD grant T32 HD007520 to A. M and NIH NINDS grant R35 NS116858 to J.D.

References

[1] A.R. Amândio, A. Necsulea, E. Joye, B. Mascrez, D. Duboule, Hotair Is dispensible for mouse development, PLoS Genet. 12 (12) (2016), e1006232, https://doi.org/10.1371/journal.pgen.1006232. PMID: 27977683; PMCID: PMC5157951.

[2] E. Bae, V.C. Calhoun, M. Levine, E.B. Lewis, R.A. Drewell, Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the Drosophila bithorax complex, Proc. Natl. Acad. Sci. USA 99 (2002) 16847–16852.

[3] M. Baek, V. Menon, T.M. Jessell, A.W. Hantman, J.S. Dasen, Molecular logic of spinocerebellar tract neuron diversity and connectivity, Cell Rep. 27 (9) (2019) 2620–2635.e4, https://doi.org/10.1016/j.celrep.2019.04.113. PMID: 31141687; PMCID: PMC6555431.

[4] C. Ballaré, M. Lange, A. Lapinaite, et al., Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity, Nat. Struct. Mol. Biol. 19 (2012) 1257–1265, https://doi.org/10.1038/nsmb.2434.

[5] F. Bantignies, V. Roure, I. Comet, B. Leblanc, B. Schuettengruber, J. Bonnet, V. Tixier, A. Mas, G. Cavalli, Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila, Cell 144 (2) (2011) 214–226, https://doi.org/10.1016/j.cell.2010.12.026. PMID: 21241892.

[6] P.J. Batut, X.Y. Bing, Z. Sisco, J. Raimundo, M. Levo, M.S. Levine, Genome organization controls transcriptional dynamics during development, Science 375 (6580) (2022) 566–570, https://doi.org/10.1126/science.abi7178. Epub 2022 Feb 3. PMID: 35113722.

[7] S. Bel-Vialar, N. Itasaki, R. Krumlauf, H.O.X. Initiating, expression: in the early chick neural tube differential sensitive to FGF and RA signaling subdivides the HOXB genes into two distinct groups", Development (2002).

[8] S. Bernheim, S.M. Meilhac, Mesoderm patterning by a dynamic gradient of retinoic acid signalling, Oct 12, Philos. Trans. R. Soc. Lond. B Biol. Sci. 375 (1809) (2020) 20190556, https://doi.org/10.1098/rstb.2019.0556. Epub 2020 Aug 24. PMID: 32829679; PMCID: PMC7482219..

[9] N. Bobola, C.G. Sagerström, TALE transcription factors: cofactors no more, Semin Cell Dev. Biol. S1084–9521 (22) (2022), https://doi.org/10.1016/j.semcdb.2022.11.015. Epub ahead of print. PMID: 36509674.

[10] R.T. Böttcher, C. Niehrs, Fibroblast growth factor signaling during early vertebrate development, Endocr. Rev. 26 (1) (2005) 63–77, https://doi.org/10.1210/er.2003-0040.

[11] A.M. Boulet, M.R. Capecchi, Signaling by FGF4 and FGF8 is required for axial elongation of the mouse embryo, Dev. Biol. 371 (2) (2012) 235–245, https://doi.org/10.1016/j.ydbio.2012.08.017. Epub 2012 Aug 30. PMID: 22954964; PMCID: PMC3481862.

[12] L.A. Boyer, K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S. S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D. K. Gifford, R.A. Young, R. Jaenisch, Polycomb complexes repress developmental regulators in murine embryonic stem cells, Nature 441 (7091) (2006) 349–353, https://doi.org/10.1038/nature04733. Epub 2006 Apr 19. PMID: 16625203.

[13] J. Briscoe, A. Pierani, T.M. Jessell, J. Ericson, A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube, Cell 101 (4) (2000) 435–445, https://doi.org/10.1016/s0092-8674(00)80853-3.

[14] M. Bulajić, D. Srivastava, J.S. Dasen, H. Wichterle, S. Mahony, E.O. Mazzoni, Differential abilities to engage inaccessible chromatin diversify vertebrate Hox binding patterns, Development 147 (22) (2020) dev194761, https://doi.org/10.1242/dev.194761. PMID: 33028607; PMCID: PMC7710020.

[15] T.R. Bürglin, M. Affolter, Homeodomain proteins: an update, Chromosoma 125 (3) (2016) 497–521, https://doi.org/10.1007/s00412-015-0543-8. PMID: 26464018; PMCID: PMC4901127.

[16] L. Cai, S.B. Rothbart, R. Lu, B. Xu, W.Y. Chen, A. Tripathy, S. Rockowitz, D. Zheng, D.J. Patel, C.D. Allis, B.D. Strahl, J. Song, G.G. Wang, An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting, Mol. Cell. 49 (3) (2013) 571–582, https://doi.org/10.1016/j.molcel.2012.11.026. Epub 2012 Dec 27. PMID: 23273982; PMCID: PMC3570589.

[17] E.M. Carpenter, Hox genes and spinal cord development, Dev. Neurosci. 24 (1) (2002) 24–34, https://doi.org/10.1159/000064943.

[18] Andrés E. Carrasco, William McGinnis, Walter J. Gehring, Eddy M. De Robertis, Cloning of an X. laevis gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes, Cell Volume 37 (Issue 2) (1984).

[19] J.E. Castelli-Gair, J.L. Micol, A. García-Bellido, Transvection in the Drosophila Ultrabithorax gene: a Cbx1 mutant allele induces ectopic expression of a normal allele in trans, Genetics 126 (1) (1990) 177–184, https://doi.org/10.1093/genetics/126.1.177. PMID: 2121595; PMCID: PMC1204122.

[20] C. Catela, M.M. Shin, D.H. Lee, J.P. Liu, J.S. Dasen, Hox proteins coordinate motor neuron differentiation and connectivity programs through ret/gfrα genes, Cell Rep. 14 (8) (2016) 1901–1915, https://doi.org/10.1016/j.celrep.2016.01.067. Epub 2016 Feb 18. PMID: 26904955; PMCID: PMC4775310.

[21] C. Catela, Y. Chen, Y. Weng, K. Wen, P. Kratsios, Control of spinal motor neuron terminal differentiation through sustained Hoxc8 gene activity, Elife 11 (2022), e70766, https://doi.org/10.7554/eLife.70766. PMID: 35315772; PMCID: PMC8940177.

[22] S. Chambeyron, W.A. Bickmore, Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription, Genes Dev. 18 (10) (2004) 1119–1130, https://doi.org/10.1101/gad.292104. PMID: 15155579; PMCID: PMC415637.

[23] K. Chawengsaksophak, R. James, V. Hammond, et al., Homeosis and intestinal tumours in *Cdx2* mutant mice, Nature 386 (1997) 84–87, https://doi.org/10.1038/386084a0.

[24] F. Ciabrelli, F. Comoglio, S. Fellous, B. Bonev, M. Ninova, Q. Szabo, A. Xuéreb, C. Klopp, A. Aravin, R. Paro, F. Bantignies, G. Cavalli, Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila, Nat. Genet. 49 (6) (2017) 876–886, https://doi.org/10.1038/ng.3848. Epub 2017 Apr 24. PMID: 28436983; PMCID: PMC5484582.

[25] C. Ciferri, G.C. Lander, A. Maiolica, F. Herzog, R. Aebersold, E. Nogales, Molecular architecture of human polycomb repressive complex 2, Elife 1 (2012), e00005, https://doi.org/10.7554/eLife.00005. . PMID: 23110252; PMCID: PMC3482686.

[26] R.T. Coleman, G. Struhl, Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene, Science 356 (6333) (2017) eaai8236, https://doi.org/10.1126/science.aai8236. Epub 2017 Mar 16. PMID: 28302795; PMCID: PMC5595140.

[27] S. Cooper, A. Grijzenhout, E. Underwood, K. Ancelin, T. Zhang, T.B. Nesterova, B. Anil-Kirmizitas, A. Bassett, S.M. Kooistra, K. Agger, K. Helin, E. Heard, N. Brockdorff, Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2, Nat. Commun. 7 (2016) 13661, https://doi.org/10.1038/ncomms13661. PMID: 27892467; PMCID: PMC5133711.

[28] M. Corley, K.L. Kroll, The roles and regulation of Polycomb complexes in neural development, Cell Tissue Res. 359 (1) (2015) 65–85, https://doi.org/10.1007/s00441-014-2011-9. Epub 2014 Nov 1. PMID: 25367430; PMCID: PMC4286515.
[29] W.G. Cox, A. Hemmati-Brivanlou, Caudalization of neural fate by tissue recombination and bFGF, Development 121 (12) (1995) 4349–4358, https://doi.org/10.1242/dev.121.12.4349.
[30] F. Darbellay, D. Duboule, Topological domains, metagenes, and the emergence of pleiotropic regulations at Hox loci, Curr. Top. Dev. Biol. 116 (2016) 299–314, https://doi.org/10.1016/bs.ctdb.2015.11.022. Epub 2016 Jan 27.
[31] J.S. Dasen, Establishing the molecular and functional diversity of spinal motoneurons, Adv. Neurobiol. 28 (2022) 3–44, https://doi.org/10.1007/978-3-031-07167-6_1.
[32] J.S. Dasen, T.M. Jessell, Hox networks and the origins of motor neuron diversity, Curr. Top. Dev. Biol. 88 (2009) 169–200, https://doi.org/10.1016/S0070-2153(09)88006-X.
[33] J.S. Dasen, J.P. Liu, T.M. Jessell, Motor neuron columnar fate imposed by sequential phases of Hox-c activity, Oct 30, Nature 425 (6961) (2003) 926–933, https://doi.org/10.1038/nature02051.
[34] J.S. Dasen, B.C. Tice, S. Brenner-Morton, T.M. Jessell, A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity, Sep 4, Cell 123 (3) (2005) 477–491, https://doi.org/10.1016/j.cell.2005.09.009.
[35] J.S. Dasen, A. De Camilli, B. Wang, P.W. Tucker, T.M. Jessell, Hox repertoires for motor neuron diversity and connectivity gated by a single accessory factor, FoxP1, Jul 25, Cell 134 (2) (2008) 304–316, https://doi.org/10.1016/j.cell.2008.06.019.
[36] J. Deschamps, M. Wijgerde, Two phases in the establishment of HOX expression domains, Dev. Biol. 156 (2) (1993) 473–480, https://doi.org/10.1006/dbio.1993.1093.
[37] J. Deschamps, D. Duboule, Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock, Genes Dev. 31 (14) (2017) 1406–1416, https://doi.org/10.1101/gad.303123.117. PMID: 28860158; PMCID: PMC5588924.
[38] T. Di Meglio, C.F. Kratochwil, N. Vilain, A. Loche, A. Vitobello, K. Yonehara, S. M. Hrycaj, B. Roska, A.H. Peters, A. Eichmann, D. Wellik, S. Ducret, F.M. Rijli, Ezh2 orchestrates topographic migration and connectivity of mouse precerebellar neurons, Science 339 (6116) (2013) 204–207, https://doi.org/10.1126/science.1229326. PMID: 23307742; PMCID: PMC4824054.
[39] R. Diez del Corral, K.G. Storey, Opposing FGF and retinoid pathways: a signaling switch that controls differentiation and patterning onset in the extending vertebrate axis, Bioessays (2004).
[40] R. Diez del Corral, D.N. Breitkreuz, K.G. Storey, Onset of neuronal differentiation is regulated by paraxial mesoderm and requires attenuation of FGF signalling, Development 129 (2002) 1681–1691.
[41] R. Diez del Corral, I. Olivera-Martinez, A. Goriely, E. Gale, M. Maden, K. Storey, Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension, Sep 25, Neuron 40 (1) (2003) 65–79, https://doi.org/10.1016/s0896-6273(03)00565-8.
[42] J.R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions, Nature 485 (7398) (2012) 376–380, https://doi.org/10.1038/nature11082. PMID: 22495300; PMCID: PMC3356448.
[43] D. Duboule, G. Morata, Colinearity and functional hierarchy among genes of the homeotic complexes, Trends Genet 10 (1994) 358–364.
[44] J. Dubrulle, O. Pourquié, *fgf8* mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo, Nature 427 (2004) 419–422 (doi:10.1038/nature02216).
[45] I.M. Duncan, Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster, Genetics 102 (1) (1982) 49–70, https://doi.org/10.1093/genetics/102.1.49. PMID: 6813190; PMCID: PMC1201924.
[46] M. Endoh, T.A. Endo, T. Endoh, K. Isono, J. Sharif, O. Ohara, T. Toyoda, T. Ito, R. Eskeland, W.A. Bickmore, M. Vidal, B.E. Bernstein, H. Koseki, Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity, PLoS Genet. 8 (7) (2012), e1002774, https://doi.org/10.1371/journal.pgen.1002774. Epub 2012 Jul 26. PMID: 22844243; PMCID: PMC3405999.
[47] R. Eskeland, M. Leeb, G.R. Grimes, C. Kress, S. Boyle, D. Sproul, N. Gilbert, Y. Fan, A.I. Skoultchi, A. Wutz, W.A. Bickmore, Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination, Mol. Cell. 38 (3) (2010) 452–464, https://doi.org/10.1016/j.molcel.2010.02.032. PMID: 20471950; PMCID: PMC3132514.
[48] C. Faust, K.A. Lawson, N.J. Schork, B. Thiel, T. Magnuson, The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo, Development 125 (22) (1998) 4495–4506, https://doi.org/10.1242/dev.125.22.4495.
[49] D.E. Ferrier, P.W. Holland, Ancient origin of the Hox gene cluster, Nat. Rev. Genet 2 (1) (2001) 33–38, https://doi.org/10.1038/35047605.
[50] B. Gaertner, J. Johnston, K. Chen, N. Wallaschek, A. Paulson, A.S. Garruss, J. Zeitlinger, Poised RNA polymerase II changes over developmental time and prepares genes for future expression, Cell Rep. 2 (6) (2012) 1670–1683, https://doi.org/10.1016/j.celrep.2012.11.024.
[51] Z. Gao, J. Zhang, R. Bonasio, F. Strino, A. Sawai, F. Parisi, Y. Kluger, D. Reinberg, PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes, Mol. Cell. 45 (3) (2012) 344–356, https://doi.org/10.1016/j.molcel.2012.01.002. PMID: 22325352; PMCID: PMC3293217.
[52] S.J. Gaunt, A.L. Gaunt, Possible rules for the ancestral origin of Hox gene collinearity, J. Theor. Biol. 410 (2016) 1–8, https://doi.org/10.1016/j.jtbi.2016.09.009. Epub 2016 Sep 10.
[53] Stephen Gaunt, Hox cluster genes and collinearities throughout the tree of animal life, Int. J. Dev. Biol. 62 (2018) 673–683, https://doi.org/10.1387/ijdb.180162sg.
[54] W.J. Gehring, U. Kloter, H. Suga, Evolution of the Hox gene complex from an evolutionary ground state, Curr. Top. Dev. Biol. 88 (2009) 35–61, https://doi.org/10.1016/S0070-2153(09)88002-2.
[55] Z. Geng, Z. Gao, Mammalian PRC1 complexes: compositional complexity and diverse molecular mechanisms, Int J. Mol. Sci. 21 (22) (2020) 8594, https://doi.org/10.3390/ijms21228594. PMID: 33202645; PMCID: PMC7697839.
[56] C. Gentile, M. Kmita, Polycomb repressive complexes in hox gene regulation: silencing and beyond: the functional dynamics of polycomb repressive complexes in Hox gene regulation, Bioessays 42 (10) (2020), e1900249, https://doi.org/10.1002/bies.201900249. Epub 2020 Aug 2. PMID: 32743818.
[57] S.F. Godsave, A.J. Durston, Neural induction and patterning in embryos deficient in FGF signaling, Int J. Dev. Biol. (1997).
[58] M.G. Golden, J.S. Dasen, Polycomb repressive complex 1 activities determine the columnar organization of motor neurons, Genes Dev. 26 (19) (2012) 2236–2250, https://doi.org/10.1101/gad.199133.112. PMID: 23028147; PMCID: PMC3465743.
[59] A. Graham, M. Maden, R. Krumlauf, The murine Hox-2 genes display dynamic dorsoventral patterns of expression during central nervous system development, Development 112 (1) (1991) 255–264, https://doi.org/10.1242/dev.112.1.255.
[60] D.J. Grau, B.A. Chapman, J.D. Garlick, M. Borowsky, N.J. Francis, R.E. Kingston, Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge, Genes Dev. 25 (20) (2011) 2210–2221, https://doi.org/10.1101/gad.17288211. PMID: 22012622; PMCID: PMC3205590.
[61] A. Grijzenhout, J. Godwin, H. Koseki, M.R. Gdula, D. Szumska, J.F. McGouran, S. Bhattacharya, B.M. Kessler, N. Brockdorff, S. Cooper, Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs, Development 143 (15) (2016) 2716–2723, https://doi.org/10.1242/dev.123935. Epub 2016 Jun 17. PMID: 27317809; PMCID: PMC5004903.
[62] O. Hanley, R. Zewdu, L.J. Cohen, H. Jung, J. Lacombe, P. Philippidou, D.H. Lee, L. Selleri, J.S. Dasen, Parallel Pbx-dependent pathways govern the coalescence and fate of motor columns, Neuron 91 (5) (2016) 1005–1020, https://doi.org/10.1016/j.neuron.2016.07.043. Epub 2016 Aug 25. PMID: 27568519; PMCID: PMC5017921.
[63] E. Healy, M. Mucha, E. Glancy, D.J. Fitzpatrick, E. Conway, H.K. Neikes, C. Monger, G. Van Mierlo, M.P. Baltissen, Y. Koseki, M. Vermeulen, H. Koseki, A. P. Bracken, PRC2.1 and PRC2.2 synergize to coordinate H3K27 trimethylation, Mol. Cell. 76 (3) (2019) 437–452.e6, https://doi.org/10.1016/j.molcel.2019.08.012. Epub 2019 Sep 11.
[64] Y. Hirabayashi, N. Suzki, M. Tsuboi, T.A. Endo, T. Toyoda, J. Shinga, H. Koseki, M. Vidal, Y. Gotoh, Polycomb limits the neurogenic competence of neural precursor cells to promote astrogenic fate transition, Sep 10, Neuron 63 (5) (2009) 600–613, https://doi.org/10.1016/j.neuron.2009.08.021.
[65] O. Hobert, Homeobox genes and the specification of neuronal identity, Nat. Rev. Neurosci. 22 (10) (2021) 627–636, https://doi.org/10.1038/s41583-021-00497-x. Epub 2021 Aug 26. PMID: 34446866.
[66] J.W. Hojfeldt, A. Laugesen, B.M. Willumsen, H. Damhofer, L. Hedehus, A. Tvardovskiy, F. Mohammad, O.N. Jensen, K. Helin, Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2, Nat. Struct. Mol. Biol. 25 (2018) 225–232.
[67] T. Iimura, N. Denans, O. Pourquié, Establishment of Hox vertebral identities in the embryonic spine precursors, Curr. Top. Dev. Biol. 88 (2009) 201–234, https://doi.org/10.1016/S0070-2153(09)88007-1. PMID: 19651306; PMCID: PMC3523337.
[68] H.V. Isaacs, M.E. Pownall, J.M. Slack, Regulation of Hox gene expression and posterior development by the *Xenopus* caudal homologue Xcad3, EMBO J. 17 (1998) 3413–3427, https://doi.org/10.1093/emboj/17.12.3413.
[69] K. Isono, Y. Fujimura, J. Shinga, M. Yamaki, J. O-Wang, Y. Takihara, Y. Murahashi, Y. Takada, Y. Mizutani-Koseki, H. Koseki, Mammalian polyhomeotic homologues Phc2 and Phc1 act in synergy to mediate polycomb repression of Hox genes, Mol. Cell Biol. 25 (15) (2005) 6694–6706, https://doi.org/10.1128/MCB.25.15.6694-6706.2005. Erratum in: Mol Cell Biol. 2014 Jul;34 (14):2771. PMID: 16024804; PMCID: PMC1190356.
[70] K. Isono, T.A. Endo, M. Ku, D. Yamada, R. Suzuki, J. Sharif, T. Ishikura, T. Toyoda, B.E. Bernstein, H. Koseki, SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing, Dev. Cell 26 (6) (2013) 565–577, https://doi.org/10.1016/j.devcel.2013.08.016.
[71] J.-C. Izpisda-Belmonte, H. Falkenstein, P. Doll, A. Bnucci, D. Duboule, Murine genes related to the Dmmphilu AbdB homeotic gene are sequentially expressed during development of the posterior part of the body, EMBO J. (1991), 102279-102289.
[72] A. Janesick, S.C. Wu, B. Blumberg, Retinoic acid signaling and neuronal differentiation, Cell Mol. Life Sci. 72 (8) (2015) 1559–1576, https://doi.org/10.1007/s00018-014-1815-9. Epub 2015 Jan 6. PMID: 25558812.
[73] H. Jung, E.O. Mazzoni, N. Soshnikova, O. Hanley, B. Venkatesh, D. Duboule, J. S. Dasen, Evolving Hox activity profiles govern diversity in locomotor systems, Dev. Cell. 29 (2) (2014) 171–187, https://doi.org/10.1016/j.devcel.2014.03.008. Epub 2014 Apr 17. PMID: 24746670; PMCID: PMC4024207.
[74] H. Jung, M. Baek, K.P. D'Elia, C. Boisvert, P.D. Currie, B.H. Tay, B. Venkatesh, S. M. Brown, A. Heguy, D. Schoppik, J.S. Dasen, The ancient origins of neural

substrates for land walking, Cell 172 (4) (2018) 667–682.e15, https://doi.org/10.1016/j.cell.2018.01.013. PMID: 29425489; PMCID: PMC5808577.
[75] H. Jung, J. Lacombe, E.O. Mazzoni, K.F. Liem Jr, J. Grinstein, S. Mahony, D. Mukhopadhyay, D.K. Gifford, R.A. Young, K.V. Anderson, H. Wichterle, J. S. Dasen, Global control of motor neuron topography mediated by the repressive actions of a single hox gene, Neuron 67 (5) (2010) 781–796, https://doi.org/10.1016/j.neuron.2010.08.008. PMID: 20826310; PMCID: PMC2955411.
[76] P. Kastner, et al., Genetic evidence that the retinoid signal is transduced by heterodimeric RXR/RAR functional units during mouse development, Development 124 (1997) 313–326.
[77] L. Kaustov, H. Ouyang, M. Amaya, A. Lemak, N. Nady, S. Duan, G.A. Wasney, Z. Li, M. Vedadi, M. Schapira, J. Min, C.H. Arrowsmith, Recognition and specificity determinants of the human cbx chromodomains, J. Biol. Chem. 286 (1) (2011) 521–529, https://doi.org/10.1074/jbc.M110.191411. Epub 2010 Nov 3. PMID: 21047797; PMCID: PMC3013012.
[78] I.D. Keenan, R.M. Sharrard, H.V. Isaacs, FGF signal transduction and the regulation of Cdx gene expression, Dev. Biol. 299 (2) (2006) 478–488, https://doi.org/10.1016/j.ydbio.2006.08.040. Epub 2006 Aug 24. PMID: 16982047.
[79] M. Kengaku, H. Okamoto, Basic fibroblast growth factor induces differentiation of neural tube and neural crest lineages of cultured ectoderm cells from Xenopus gastrula (Dec), Development 119 (4) (1993) 1067–1078, https://doi.org/10.1242/dev.119.4.1067.
[80] S.L. Kloet, M.M. Makowski, H.I. Baymaz, L. van Voorthuijsen, I.D. Karemaker, A. Santanach, P.W.T.C. Jansen, L. Di Croce, M. Vermeulen, The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation, Nat. Struct. Mol. Biol. 23 (7) (2016) 682–690, https://doi.org/10.1038/nsmb.3248. PMID: 27294783; PMCID: PMC4939079.
[81] R. Krumlauf, Hox genes, clusters and collinearity, Int. J. Dev. Biol. 62 (11–12) (2018) 659–663, https://doi.org/10.1387/ijdb.180330rr.
[82] M. Ku, R.P. Koche, E. Rheinbay, E.M. Mendenhall, M. Endoh, T.S. Mikkelsen, A. Presser, C. Nusbaum, X. Xie, A.S. Chi, M. Adli, S. Kasif, L.M. Ptaszek, C. A. Cowan, E.S. Lander, H. Koseki, B.E. Bernstein, Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains, PLoS Genet. 4 (10) (2008), e1000242, https://doi.org/10.1371/journal.pgen.1000242. Epub 2008 Oct 31. PMID: 18974828; PMCID: PMC2567431.
[83] S. Kumar, G. Duester, Retinoic acid controls body axis extension by directly repressing Fgf8 transcription, Development 141 (15) (2014) 2972–2977, https://doi.org/10.1242/dev.112367. PMID: 25053430; PMCID: PMC4197666.
[84] S. Kundu, F. Ji, H. Sunwoo, G. Jain, J.T. Lee, R.I. Sadreyev, J. Dekker, R. E. Kingston, Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation, Mol. Cell. 65 (3) (2017) 432–446.e5, https://doi.org/10.1016/j.molcel.2017.01.009. Erratum in: Mol Cell. 2018 Jul 5; 71(1):191. PMID: 28157505; PMCID: PMC5421375.
[85] L. Tavares E. Dimitrova D. Oxley J. Webster R. Poot J. Demmers K. Bezstarosti S. Taylor H. Ura H. Koide A. Wutz M. Vidal S. Elderkin N. Brockdorff RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell. 2012 Feb 17;148(4):664–678. doi: 10.1016/j.cell.2011.12.029. Epub 2012 Feb 9. Erratum in: Cell. 2012 Jun 22;149(7): 1647–8. PMID: 22325148; PMCID: PMC3281992.
[86] J. Lacombe, O. Hanley, H. Jung, P. Philippidou, G. Surmeli, J. Grinstein, J. S. Dasen, Genetic and functional modularity of Hox activities in the specification of limb-innervating motor neurons, PLoS Genet. 9 (1) (2013), e1003184, https://doi.org/10.1371/journal.pgen.1003184. Epub 2013 Jan 24. PMID: 23359544; PMCID: PMC3554521.
[87] S. Lalevée, Y.N. Anno, A. Chatagnon, E. Samarut, O. Poch, V. Laudet, G. Benoit, O. Lecompte, C. Rochette-Egly, Genome-wide in silico identification of new conserved and functional retinoic acid receptor response elements (direct repeats separated by 5 bp), J. Biol. Chem. 286 (38) (2011) 33322–33334, https://doi.org/10.1074/jbc.M111.263681. Epub 2011 Jul 29. PMID: 21803772; PMCID: PMC3190930.
[88] T.M. Lamb, R.M. Harland, Fibroblast growth factor is a direct neural inducer, which combined with noggin generates anterior-posterior neural pattern, Development (1995).
[89] D. Landeira, S. Sauer, R. Poot, M. Dvorkina, L. Mazzarella, H.F. Jorgensen, C. F. Pereira, M. Leleu, F.M. Piccolo, M. Spivakov, et al., Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators, Nat. Cell Biol. 12 (2010) 618–624.
[90] C. Lanzuolo, V. Orlando, Memories from the polycomb group proteins, Annu Rev. Genet. 46 (2012) 561–589, https://doi.org/10.1146/annurev-genet-110711-155603.
[91] M.S. Lau, M.G. Schwartz, S. Kundu, A.J. Savol, P.I. Wang, S.K. Marr, D.J. Grau, P. Schorderet, R.I. Sadreyev, C.J. Tabin, R.E. Kingston, Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning, Science 355 (6329) (2017) 1081–1084, https://doi.org/10.1126/science.aah5403.
[92] E.B. Lewis, A gene complex controlling segmentation in Drosophila, Dec 7, Nature 276 (5688) (1978) 565–570, https://doi.org/10.1038/276565a0.
[93] C.J. Li, T. Hong, Y.T. Tung, Y.P. Yen, H.C. Hsu, Y.L. Lu, M. Chang, Q. Nie, J. A. Chen, MicroRNA filters Hox temporal transcription noise to confer boundary formation in the spinal cord, Nat. Commun. 8 (2017) 14685, https://doi.org/10.1038/ncomms14685. PMID: 28337978; PMCID: PMC5376671.
[94] H. Li, R. Liefke, J. Jiang, J.V. Kurland, W. Tian, P. Deng, W. Zhang, Q. He, D. J. Patel, M.L. Bulyk, Y. Shi, Z. Wang, Polycomb-like proteins link the PRC2 complex to CpG islands, Nature 549 (7671) (2017) 287–291, https://doi.org/10.1038/nature23881. Epub 2017 Sep 6. PMID: 28869966; PMCID: PMC5937281.
[95] L. Li, B. Liu, O.L. Wapinski, M.C. Tsai, K. Qu, J. Zhang, J.C. Carlson, M. Lin, F. Fang, R.A. Gupta, J.A. Helms, H.Y. Chang, Targeted disruption of Hotair leads to homeotic transformation and gene derepression, Cell Rep. 5 (1) (2013) 3–12, https://doi.org/10.1016/j.celrep.2013.09.003. Epub 2013 Sep 26. PMID: 24075995; PMCID: PMC4038295.
[96] X. Li, K. Isono, D. Yamada, T.A. Endo, M. Endoh, J. Shinga, Y. Mizutani-Koseki, A. P. Otte, M. Casanova, H. Kitamura, T. Kamijo, J. Sharif, O. Ohara, T. Toyada, B. E. Bernstein, N. Brockdorff, H. Koseki, Mammalian polycomb-like Pcl2/Mtf2 is a novel regulatory component of PRC2 that can differentially modulate polycomb activity both at the Hox gene cluster and at Cdkn2a genes, Mol. Cell Biol. (2) (2011) 351–364, https://doi.org/10.1128/MCB.00259-10. Epub 2010 Nov 8. Erratum in: Mol Cell Biol. 2014 Jul;34(14):2773. PMID: 21059868; PMCID: PMC3019975.
[97] C. Lin, A.S. Garrett, B. De Kumar, E.R. Smith, M. Gogol, C. Seidel, A. Shilatifard, Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC), Genes Dev. 25 (14) (2011) 1486–1498, https://doi.org/10.1101/gad.2059211.
[98] J.P. Liu, The function of growth/differentiation factor 11 (Gdf11) in rostrocaudal patterning of the developing spinal cord, Development 133 (15) (2006) 2865–2874, https://doi.org/10.1242/dev.02478. Epub 2006 Jun 21. PMID: 16790475.
[99] J.P. Liu, E. Laufer, T.M. Jessell, Assigning the positional identity of spinal motor neurons: rostrocaudal patterning of Hox-c expression by FGFs, Gdf11, and retinoids, Neuron 32 (2001) 997–1012.
[100] V. Loubiere, A.M. Martinez, Cavalli G. cell fate and developmental regulation dynamics by polycomb proteins and 3D genome architecture, Bioessays 41 (3) (2019), e1800222, https://doi.org/10.1002/bies.201800222. Epub 2019 Feb 22. PMID: 30793782.
[101] S. Mahony, E.O. Mazzoni, S. McCuine, et al., Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis, Genome Biol. 12 (2011) R2, https://doi.org/10.1186/gb-2011-12-1-r2.
[102] M. Mallo, C.R. Alonso, The regulation of Hox gene expression during animal development (Oct), Development 140 (19) (2013) 3951–3963, https://doi.org/10.1242/dev.068346.
[103] M. Mallo, D.M. Wellik, J. Deschamps, Hox genes and regional patterning of the vertebrate body plan, Dev. Biol. 344 (1) (2010) 7–15, https://doi.org/10.1016/j.ydbio.2010.04.024. Epub 2010 May 7. PMID: 20435029; PMCID: PMC2909379.
[104] R. Margueron, N. Justin, K. Ohno, M.L. Sharpe, J. Son, W.J. Drury 3rd, P. Voigt, S. R. Martin, W.R. Taylor, V. De Marco, V. Pirrotta, D. Reinberg, S.J. Gamblin, Role of the polycomb protein EED in the propagation of repressive histone marks, Nature 461 (7265) (2009) 762–767, https://doi.org/10.1038/nature08398. PMID: 19767730; PMCID: PMC3772642.
[105] E.O. Mazzoni, S. Mahony, M. Peljto, T. Patel, S.R. Thornton, S. McCuine, C. Reeder, L.A. Boyer, R.A. Young, D.K. Gifford, H. Wichterle, Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals, Nat. Neurosci. 16 (9) (2013) 1191–1198, https://doi.org/10.1038/nn.3490. Epub 2013 Aug 18. PMID: 23955559; PMCID: PMC3799941.
[106] W. McGinnis, M. Levine, E. Hafen, et al., A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes, Nature 308 (1984) 428–433, https://doi.org/10.1038/308428a0.
[107] A.I. Mendelsohn, J.S. Dasen, T.M. Jessell, Divergent Hox coding and evasion of retinoid signaling specifies motor neurons innervating digit muscles, Neuron 93 (4) (2017) 792–805.e4, https://doi.org/10.1016/j.neuron.2017.01.017. Epub 2017 Fe**b 9. PMID: 28190640; PMCID:** PMC5325683.
[108] V. Metzis, S. Steinhauser, E. Pakanavicius, M. Gouti, D. Stamataki, K. Ivanovitch, T. Watson, T. Rayon, S.N. Mousavy Gharavy, R. Lovell-Badge, N.M. Luscombe, J. Briscoe, Nervous system regionalization entails axial allocation before neural differentiation, Cell 175 (4) (2018) 1105–1118.e17, https://doi.org/10.1016/j.cell.2018.09.040. Epub 2018 Oct 18. PMID: 30343898; PMCID: PMC6218657.
[109] T. Montavon, N. Soshnikova, Hox gene regulation and timing in embryogenesis, Semin Cell Dev. Biol. 34 (2014) 76–84, https://doi.org/10.1016/j.semcdb.2014.06.005. Epub 2014 Jun 12. PMID: 24930771.
[110] V. Mouilleau, C. Vaslin, R. Robert, S. Gribaudo, N. Nicolas, M. Jarrige, A. Terray, L. Lesueur, M.W. Mathis, G. Croft, M. Daynac, V. Rouiller-Fabre, H. Wichterle, V. Ribes, C. Martinat, S. Nedelec, Dynamic extrinsic pacing of the HOX clock in human axial progenitors controls motor neuron subtype specification, Development 148 (6) (2021) dev194514, https://doi.org/10.1242/dev.194514. PMID: 33782043; PMCID: PMC8034877.
[111] V. Narendra, M. Bulajić, J. Dekker, E.O. Mazzoni, D. Reinberg, CTCF-mediated topological boundaries during development foster appropriate gene regulation, Genes Dev. 30 (24) (2016) 2657–2662, https://doi.org/10.1101/gad.288324.116. Erratum in: Genes Dev. 2017 Aug 15;31(16):1714. PMID: 28087711; PMCID: PMC5238725.
[112] V. Narendra, P.P. Rocha, D. An, R. Raviram, J.A. Skok, E.O. Mazzoni, D. Reinberg, CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation, Science 347 (6225) (2015) 1017–1021, https://doi.org/10.1126/science.1262088. PMID: 25722416; PMCID: PMC4428148.Oh et al., "Retinoic Acid and CTCF Play Key Roles in Inducing the Collinear Expression of the Hoxa Cluster.".
[113] R. Neijts, S. Amin, C. van Rooijen, J. Deschamps, Cdx is crucial for the timing mechanism driving colinear Hox activation and defines a trunk segment in the Hox cluster topology, Dev. Biol. 422 (2) (2017) 146–154, https://doi.org/10.1016/j.ydbio.2016.12.024. Epub 2016 Dec 29. PMID: 28041967.
[114] R. Neijts, S. Amin, C. van Rooijen, S. Tan, M.P. Creyghton, W. de Laat, J. Deschamps, Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos, Genes Dev. 30 (17) (2016) 1937–1942, https://doi.

org/10.1101/gad.285767.116. Epub 2016 Sep 15. PMID: 27633012; PMCID: PMC5066237.
[115] K. Niederreither, P. McCaffery, U.C. Dräger, P. Chambon, P. Dollé, Restricted expression and retinoic acid-induced downregulation of the retinaldehyde dehydrogenase type 2 (RALDH-2) gene during mouse development (Feb), Mech. Dev. 62 (1) (1997) 67–78, https://doi.org/10.1016/s0925-4773(96)00653-3.
[116] K. Niederreither, S. Abu-Abed, B. Schuhbaur, M. Petkovich, P. Chambon, P. Dollé, Genetic evidence that oxidative derivatives of retinoic acid are not involved in retinoid signaling during mouse development, Nat. Genet. 31 (1) (2002) 84–88, https://doi.org/10.1038/ng876. Epub 2002 Apr 15. PMID: 11953746.
[117] C. Nolte, R. Krumlauf, Expression of Hox genes in the nervous system of vertebrates, 2000-. Madame Curie Bioscience Database [Internet]. Austin (TX), Landes Bioscience, 2013, 2000-, ⟨https://www.ncbi.nlm.nih.gov/books/NBK6519/⟩.
[118] C. Nolte, B. De Kumar, R. Krumlauf, *Hox* genes: Downstream "effectors" of retinoic acid signaling in vertebrate embryogenesis, Genesis 57 (2019), e23306, https://doi.org/10.1002/dvg.23306.
[119] C. Nolte, T. Jinks, X. Wang, M.T. Martinez Pastor, R. Krumlauf, Shadow enhancers flanking the HoxB cluster direct dynamic Hox expression in early heart and endoderm development, Dev. Biol. 383 (1) (2013) 158–173, https://doi.org/10.1016/j.ydbio.2013.09.016. Epub 2013 Sep 18. PMID: 24055171.
[120] J.L. Northrop, D. Kimelman, Dorsal-ventral differences in Xcad-3 expression in response to FGF-mediated induction in Xenopus (Feb), Dev. Biol. 161 (2) (1994) 490–503, https://doi.org/10.1006/dbio.1994.1047.
[121] H. Ortabozkoyun, P.Y. Huang, H. Cho, V. Narendra, G. LeRoy, E. Gonzalez-Buendia, J.A. Skok, A. Tsirigos, E.O. Mazzoni, D. Reinberg, CRISPR and biochemical screens identify MAZ as a cofactor in CTCF-mediated insulation at Hox clusters, Nat. Genet. 54 (2) (2022) 202–212, https://doi.org/10.1038/s41588-021-01008-5. Epub 2022 Feb 10. PMID: 35145304; PMCID:.
[122] P.J. Osseward 2nd, S.L. Pfaff, Cell type and circuit modules in the spinal cord, Curr. Opin. Neurobiol. 56 (2019) 175–184, https://doi.org/10.1016/j.conb.2019.03.003. Epub 2019 Apr 5. PMID: 30954861; PMCID: PMC8559966.
[123] H.J. Parker, R. Krumlauf, A Hox gene regulatory network for hindbrain segmentation, Curr. Top. Dev. Biol. 139 (2020) 169–203, https://doi.org/10.1016/bs.ctdb.2020.03.001. Epub 2020 Apr 9. PMID: 32450960.
[124] J. Partanen, L. Schwartz, J. Rossant, Opposite phenotypes of hypomorphic and Y766 phosphorylation site mutations reveal a function for fgfr1 in anteroposterior patterning of mouse embryos, Genes Dev. (1998).
[125] J. Pascual-Anaya, S. D'Aniello, S. Kuratani, et al., Evolution of *Hox* gene clusters in deuterostomes, BMC Dev. Biol. 13 (2013) 26, https://doi.org/10.1186/1471-213X-13-26.
[126] D. Pasini, A.P. Bracken, M.R. Jensen, E.L. Denchi, K. Helin, Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity, EMBO J. 23 (2004) 4061–4071, https://doi.org/10.1038/sj.emboj.7600402.
[127] P. Philippidou, J.S. Dasen, Hox genes: choreographers in neural development, architects of circuit organization, Neuron 80 (1) (2013) 12–34, https://doi.org/10.1016/j.neuron.2013.09.020. Epub 2013 Oct 2. PMID: 24094100; PMCID: PMC3835187.
[128] P. Philippidou, C.M. Walsh, J. Aubin, L. Jeannotte, J.S. Dasen, Sustained Hox5 gene activity is required for respiratory motor neuron development, Nat. Neurosci. 15 (12) (2012) 1636–1644, https://doi.org/10.1038/nn.3242. Epub 2012 Oct 28. PMID: 23103965; PMCID: PMC3676175.
[129] S. Pinglay, M. Bulajić, D.P. Rahe, E. Huang, R. Brosh, N.E. Mamrak, B.R. King, S. German, J.A. Cadley, L. Rieber, N. Easo, T. Lionnet, S. Mahony, M.T. Maurano, L.J. Holt, E.O. Mazzoni, J.D. Boeke, Synthetic regulatory reconstitution reveals principles of mammalian Hox cluster regulation, Science 377 (6601) (2022) eabk2820, https://doi.org/10.1126/science.abk2820.
[130] M.K. Pirity, J. Locker, N. Schreiber-Agus, Rybp/DEDAF is required for early postimplantation and for central nervous system development, Mol. Cell Biol. 25 (16) (2005) 7193–7202, https://doi.org/10.1128/MCB.25.16.7193-7202.2005. PMID: 16055728; PMCID: PMC1190233.
[131] A. Piunti, A. Shilatifard, Author Correction: the roles of Polycomb repressive complexes in mammalian development and cancer, Nat. Rev. Mol. Cell Biol. 23 (6) (2022) 444, https://doi.org/10.1038/s41580-022-00495-6. Erratum for: Nat Rev Mol Cell Biol. 2021 May;22(5):326-345. PMID: 35534714.
[132] M. Portoso, R. Ragazzini, Z. Brencic, A. Moiani, A. Michaud, I. Vassilev, M. Wassef, N. Servant, B. Sargueil, R. Margueron, PRC2 is dispensable for HOTAIR-mediated transcriptional repression, EMBO J. 36 (2017) 981–994.
[133] M. Pownall, A. Tucker, J. Slack, H. Isaacs, eFGF, *Xcad3* and Hox genes form a molecular pathway that establishes the anteroposterior axis in *Xenopus*, Development (1996).
[134] P. Qian, B. De Kumar, X.C. He, C. Nolte, M. Gogol, Y. Ahn, S. Chen, Z. Li, H. Xu, J. M. Perry, D. Hu, F. Tao, M. Zhao, Y. Han, K. Hall, A. Peak, A. Paulson, C. Zhao, A. Venkatraman, A. Box, A. Perera, J.S. Haug, T. Parmely, H. Li, R. Krumlauf, L. Li, Retinoid-sensitive epigenetic regulation of the hoxb cluster maintains normal hematopoiesis and inhibits leukemogenesis, Cell. Stem Cell. 22 (5) (2018) 740–754.e7, https://doi.org/10.1016/j.stem.2018.04.012.
[135] J.L. Rinn, M. Kertesz, J.K. Wang, S.L. Squazzo, X. Xu, S.A. Brugmann, L. H. Goodnough, J.A. Helms, P.J. Farnham, E. Segal, H.Y. Chang, Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs, Cell 129 (7) (2007) 1311–1323, https://doi.org/10.1016/j.cell.2007.05.022. PMID: 17604720; PMCID: PMC2084369.
[136] A.K. Robinson, B.Z. Leal, L.V. Chadwell, R. Wang, U. Ilangovan, Y. Kaur, S. E. Junco, V. Schirf, P.A. Osmulski, M. Gaczynska, A.P. Hinck, B. Demeler, D. G. McEwen, C.A. Kim, The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain, J. Biol. Chem. 287 (12) (2012) 8702–8713, https://doi.org/10.1074/jbc.M111.336115. Epub 2012 Jan 24. PMID: 22275371; PMCID: PMC3308824.
[137] E. Rodríguez-Carballo, L. Lopez-Delisle, A. Willemin, L. Beccari, S. Gitto, B. Mascrez, D. Duboule, Chromatin topology and the timing of enhancer function at the *HoxD* locus, Proc. Natl. Acad. Sci. USA 117 (49) (2020) 31231–31241, https://doi.org/10.1073/pnas.2015083117. Epub 2020 Nov 23. PMID: 33229569; PMCID: PMC7733857.
[138] M. Ronshaugen, F. Biemar, J. Piel, M. Levine, E.C. Lai, The Drosophilamicro RNA iab-4 causes a dominant homeotic transformation of halteresto wings, Genes Dev. 19 (2005) 2947–2952.
[139] D.L. Rousso, Z.B. Gaber, D. Wellik, E.E. Morrisey, B.G. Novitch, Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons, Neuron 59 (2) (2008) 226–240, https://doi.org/10.1016/j.neuron.2008.06.025. PMID: 18667151; PMCID: PMC2547125.
[140] A. Sawai, S. Pfennig, M. Bulajić, A. Miller, A. Khodadadi-Jamayran, E.O. Mazzoni, J.S. Dasen, PRC1 sustains the integrity of neural fate in the absence of PRC2 function, Elife 11 (2022), e72769, https://doi.org/10.7554/eLife.72769.
[141] M. von Schimmelmann, P.A. Feinberg, J.M. Sullivan, S.M. Ku, A. Badimon, M. K. Duff, Z. Wang, A. Lachmann, S. Dewell, A. Ma'ayan, M.H. Han, A. Tarakhovsky, A. Schaefer, Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration, Nat. Neurosci. 19 (10) (2016) 1321–1330, https://doi.org/10.1038/nn.4360. PMID: 27526204; PMCID: PMC5088783.
[142] M.P. Scott, A.J. Weiner, Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila, Proc. Natl. Acad. Sci. USA 81 (13) (1984) 4115–4119, https://doi.org/10.1073/pnas.81.13.4115. PMID: 6330741; PMCID: PMC345379.
[143] Y. Shi, J.P. Liu, Gdf11 facilitates temporal progression of neurogenesis in the developing spinal cord, J. Neurosci. 31 (3) (2011) 883–893, https://doi.org/10.1523/JNEUROSCI.2394-10.2011. PMID: 21248112; PMCID: PMC4112097.
[144] M.M. Shin, C. Catela, J. Dasen, Intrinsic control of neuronal diversity and synaptic specificity in a proprioceptive circuit, Elife 9 (2020), e56374, https://doi.org/10.7554/eLife.56374. PMID: 32808924; PMCID: PMC7467731.
[145] J.A. Simon, R.E. Kingston, Mechanisms of polycomb gene silencing: knowns and unknowns, Nat. Rev. Mol. Cell Biol. 10 (10) (2009) 697–708, https://doi.org/10.1038/nrm2763. Epub 2009 Sep 9. PMID: 19738629.
[146] N.P. Singh, R.K. Mishra, A double-edged sword to force posterior dominance of Hox genes (Nov), Bioessays 30 (11–12) (2008) 1058–1061, https://doi.org/10.1002/bies.20847.
[147] N. Soshnikova, D. Duboule, Epigenetic temporal control of mouse Hox genes in vivo, Science 324 (5932) (2009) 1320–1323, https://doi.org/10.1126/science.1171468.
[148] N. Soshnikova, T. Montavon, M. Leleu, N. Galjart, D. Duboule, Functional analysis of CTCF during mammalian limb development, Dec 14, Dev. Cell 19 (6) (2010) 819–830, https://doi.org/10.1016/j.devcel.2010.11.009.
[149] C.D. Stern, Initial patterning of the central nervous system: how many organizers? (Feb), Nat. Rev. Neurosci. 2 (2) (2001) 92–98, https://doi.org/10.1038/35053563.
[150] N. Stifani, Motor neurons and the generation of spinal motor neuron diversity, Front Cell Neurosci. 8 (2014) 293, https://doi.org/10.3389/fncel.2014.00293. PMID: 25346659; PMCID: PMC4191298.
[151] J.K. Stock, S. Giadrossi, M. Casanova, E. Brookes, M. Vidal, H. Koseki, N. Brockdorff, A.G. Fisher, A. Pombo, Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells, Nat. Cell Biol. 9 (12) (2007) 1428–1435, https://doi.org/10.1038/ncb1663. Epub 2007 Nov 25. PMID: 18037880.
[152] L.B. Sweeney, J.B. Bikoff, M.I. Gabitto, S. Brenner-Morton, M. Baek, J.H. Yang, E. G. Tabak, J.S. Dasen, C.R. Kintner, T.M. Jessell, Origin and segmental diversity of spinal inhibitory interneurons, Neuron 97 (2) (2018) 341–355.e3, https://doi.org/10.1016/j.neuron.2017.12.029. Epub 2018 Jan 4. PMID: 29307712; PMCID: PMC5880537.
[153] A. Tanzer, C.T. Amemiya, C.B. Kim, P.F. Stadler, Evolution of microRNAs located within Hox gene clusters, J. Exp. Zool. B Mol. DevEvol 304 (2005) 75–85.
[154] B. Tolhuis, M. Blom, R.M. Kerkhoven, L. Pagie, H. Teunissen, M. Nieuwland, M. Simonis, W. de Laat, M. van Lohuizen, B. van Steensel, Interactions among Polycomb domains are guided by chromosome architecture, PLoS Genet. (2011), e1001343, https://doi.org/10.1371/journal.pgen.1001343.
[155] M. Tsuboi, Y. Kishi, W. Yokozeki, H. Koseki, Y. Hirabayashi, Y. Gotoh, Ubiquitination-independent repression of PRC1 targets during neuronal fate restriction in the developing mouse neocortex, Dev. Cell. 47 (6) (2018) 758–772. e5, https://doi.org/10.1016/j.devcel.2018.11.018.
[156] A.N. Vagnozzi, K. Garg, C. Dewitz, M.T. Moore, J.M. Cregg, L. Jeannotte, N. Zampieri, L.T. Landmesser, P. Philippidou, Phrenic-specific transcriptional programs shape respiratory motor output, Elife 9 (2020), e52859, https://doi.org/10.7554/eLife.52859.
[157] L. Venkatasubramanian, R.S. Mann, The development and assembly of the Drosophila adult ventral nerve cord, Curr. Opin. Neurobiol. 56 (2019) 135–143, https://doi.org/10.1016/j.conb.2019.01.013. Epub 2019 Feb 28. PMID: 30826502; PMCID: PMC6551290.
[158] J.W. Voncken, B.A. Roelen, M. Roefs, S. de Vries, E. Verhoeven, S. Marino, J. Deschamps, M. van Lohuizen, Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition, Proc. Natl. Acad. Sci. USA 100 (5) (2003) 2468–2473, https://doi.org/10.1073/pnas.0434312100. Epub 2003 Feb 14. PMID: 12589020; PMCID: PMC151364.

[159] H. Wang, L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R.S. Jones, Y. Zhang, Role of histone H2A ubiquitination in Polycomb silencing, Nature 431 (7010) (2004) 873–878, https://doi.org/10.1038/nature02985. Epub 2004 Sep 22.

[160] R. Wang, A.B. Taylor, B.Z. Leal, L.V. Chadwell, U. Ilangovan, A.K. Robinson, V. Schirf, P.J. Hart, E.M. Lafer, B. Demeler, A.P. Hinck, D.G. McEwen, C.A. Kim, Polycomb group targeting through different binding partners of RING1B C-terminal domain, Structure 18 (8) (2010) 966–975, https://doi.org/10.1016/j.str.2010.04.013. PMID: 20696397; PMCID: PMC4445678.

[161] W. Wang, H. Cho, D. Kim, Y. Park, J.H. Moon, S.J. Lim, S.M. Yoon, M. McCane, S. A. Aicher, S. Kim, B. Emery, J.W. Lee, S. Lee, Y. Park, S.K. Lee, PRC2 acts as a critical timer that drives oligodendrocyte fate over astrocyte identity by repressing the notch pathway, Cell Rep. 32 (11) (2020), 108147, https://doi.org/10.1016/j.celrep.2020.108147. PMID: 32937136; PMCID: PMC8070886.

[162] Wang X., Gu J., Zhang M.Q., Li Y. 2008. Identification of phylogenetically conserved microRNA cis-regulatory elements across 12 Drosophila species. Bioinformatics 24:165 –171Simon G. Sprecher, Martin Müller, Lars Kammermeier, David F.B. Miller, Thomas C. Kaufman, Heinrich Reichert, Frank Hirth. Hox gene cross-regulatory interactions in the embryonic brain of Drosophila. Mechanisms of Development. Volume 121, Issue 6. 2004.

[163] A. Wani, A. Boettiger, P. Schorderet, et al., Chromatin topology is coupled to Polycomb group protein subnuclear organization, Nat. Commun. 7 (2016) 10291, https://doi.org/10.1038/ncomms10291.

[164] J.M. Woltering, A.J. Durston, The zebrafish hoxDb cluster has been reduced to a single microRNA (Jun), Nat. Genet. 38 (6) (2006) 601–602, https://doi.org/10.1038/ng0606-601.

[165] L. Wu, P. Murat, D. Matak-Vinkovic, A. Murrell, S. Balasubramanian, Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins, Biochemistry 52 (52) (2013) 9519–9527, https://doi.org/10.1021/bi401085h. Epub 2013 Dec 17. PMID: 24320048; PMCID: PMC3964825.

[166] B. Yaghmaeian Salmani, I. Monedero Cobeta, J. Rakar, S. Bauer, J.R. Curt, A. Starkenberg, S. Thor, Evolutionarily conserved anterior expansion of the central nervous system promoted by a common PcG-Hox program, Apr 5, Development 145 (7) (2018) dev160747, https://doi.org/10.1242/dev.160747.

[167] S. Yekta, I.H. Shih, D.P. Bartel, MicroRNA-directed cleavage of HOXB8 mRNA, Science 304 (5670) (2004) 594–596, https://doi.org/10.1126/science.1097434. PMID: 15105502.

[168] Y.P. Yen, W.F. Hsieh, Y.Y. Tsai, Y.L. Lu, E.S. Liau, H.C. Hsu, Y.C. Chen, T.C. Liu, M. Chang, J. Li, S.P. Lin, J.H. Hung, J.A. Chen, Dlk1-Dio3 locus-derived lncRNAs perpetuate postmitotic motor neuron cell fate and subtype identity, Elife (2018), e38080, https://doi.org/10.7554/eLife.38080.

[169] D.T. Youmans, A.R. Gooding, R.D. Dowell, T.R. Cech, Competition between PRC2.1 and 2.2 subcomplexes regulates PRC2 chromatin occupancy in human stem cells, Mol. Cell (2021) 488–501.e9, https://doi.org/10.1016/j.molcel.2020.11.044.

[170] T. Young, J.Hox Deschamps, Cdx, and anteroposterior patterning in the mouse embryo, Curr. Top. Dev. Biol. 88 (2009) 235–255, https://doi.org/10.1016/S0070-2153(09)88008-3.

[171] T. Young, J.E. Rowland, C. van de Ven, M. Bialecka, A. Novoa, M. Carapuco, J. van Nes, W. de Graaff, I. Duluc, J.N. Freund, F. Beck, M. Mallo, J. Deschamps, Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos (Oct), Dev. Cell 17 (4) (2009) 516–526, https://doi.org/10.1016/j.devcel.2009.08.010.

Seminars in Cell and Developmental Biology 152-153 (2024) 35–43

Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Hox function and specificity – A tissue centric view

Pedro B. Pinto [a,b], Katrin Domsch [a], Ingrid Lohmann [a,*]

[a] Heidelberg University, Centre for Organismal Studies (COS) Heidelberg, Department of Developmental Biology and Cell Networks - Cluster of Excellence, Heidelberg, Germany

[b] Institut de Génomique Fonctionnelle de Lyon, Univ Lyon, CNRS UMR 5242, École Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, Lyon, France

ARTICLE INFO

Keywords:
Hox
Specificity
Repression
Activation
Ectoderm
Mesoderm
Fat body
Polycomb
BiFC
BioID
Interactomes
Enhancer
Vvl
Svb
AP-2
Sal
Dll
Tin
Grh

ABSTRACT

Since their discovery, the *Hox* genes, with their incredible power to reprogram the identity of complete body regions, a phenomenon called homeosis, have captured the fascination of many biologists. Recent research has provided new insights into the function of Hox proteins in different germ layers and the mechanisms they employ to control tissue morphogenesis. We focus in this review on the ectoderm and mesoderm to highlight new findings and discuss them with regards to established concepts of Hox target gene regulation. Furthermore, we highlight the molecular mechanisms involved the transcriptional repression of specific groups of Hox target genes, and summarize the role of Hox mediated gene silencing in tissue development. Finally, we reflect on recent findings identifying a large number of tissue-specific Hox interactor partners, which open up new avenues and directions towards a better understanding of Hox function and specificity in different tissues.

1. Introduction

An important characteristic of animals is the development of specialized morphological features along the body plan necessary for their development and survival. How such diversity of morphologies is generated with such incredible precision, has been a fundamental question in Biology and the subject of many studies. The discovery that mutations in one group of genes, the *Hox* genes, induced homeotic transformations in which one structure is transformed to resemble, in form and shape, a homologous structure present in the body was fundamental to our understanding of how morphological diversity can be generated [1–3]. These studies identified *Hox* genes as master regulators of development and showed that this group of genes controls the specification and differentiation of different cell and tissue types along the anterior-posterior (AP) axis in bilaterian animals [4]. On the molecular level, Hox proteins are transcription factors (TFs) and, thus instruct the development of different morphological features by controlling specific morphogenetic programs [5,6]. Hence, an immense body of work focused on the characterization of these genetic networks. However, to understand how Hox TFs regulate gene expression in a spatial and temporal manner so as to deploy specific programs, it is important to understand how these proteins specifically regulate their target genes. This is indeed essential, since Hox TFs share the same DNA-binding domain, the homeodomain (HD) [7,8]. In a clear contrast to their highly specific function in vivo, the Hox TFs recognize similar binding sequences in vitro [7,8]. This so-called Hox paradox is the major challenge to understand how Hox proteins play such diverse and specific roles in development. Consequently, much of the work in the Hox field is focused on identifying the molecular mechanisms used by Hox TFs to regulate precise gene expression. Hox specificity is even more of a challenge when we consider that in addition to having different Hox TFs with different specificities, a single Hox protein also acts highly specific:

* Corresponding author.
E-mail address: ingrid.lohmann@cos.uni-heidelberg.de (I. Lohmann).

https://doi.org/10.1016/j.semcdb.2022.11.011
Received 30 August 2022; Received in revised form 11 November 2022; Accepted 30 November 2022
Available online 12 December 2022

it controls the specification and differentiation of very different cell and tissue types present in each segment where it is expressed. Thus, despite the immense effort and the ground-breaking results [9], many questions are still unsolved including the following: Do Hox TFs employ similar molecular mechanisms in different tissues? What are the different mechanisms employed by Hox TFs that provide their unique specificity? Are different mechanisms used to regulate general and tissue specific target genes? Do activator and repressor functions of Hox TFs share common molecular mechanisms? Which interaction partners/co-factors assist Hox proteins in target gene regulation in the different tissues?

In this review, we focus on the mesoderm and the ectoderm, and highlight findings of previous and recent publications, which influenced our view on how Hox proteins control the development of these two germ layers in the fruit fly *Drosophila*. We will put these findings in

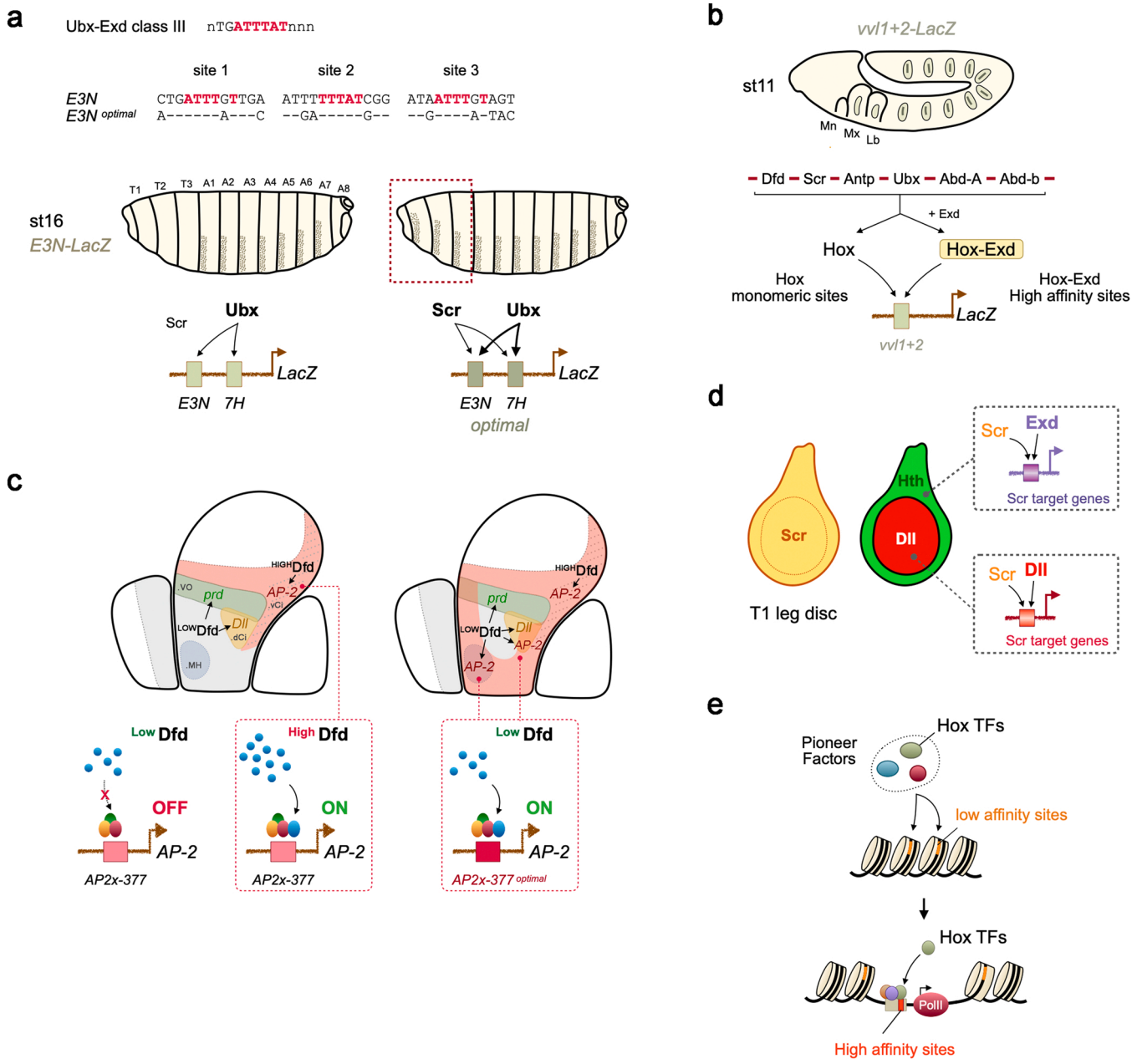


Fig. 1. Hox TFs rely on different molecular mechanisms to achieve specificity in the ectoderm. (a) In stage 16 embryos, two enhancers, *E3N* and *7H*, control the expression of *shavenbaby* (*svb*) via Ubx-Exd low-affinity sites. The *E3N* enhancer is shown as an example. The Ubx-Exd low-affinity sites lead to a specific activation of the enhancer by Ubx but not but not by anterior Hox proteins like Scr. Conversion of these sites into high-affinity Ubx-Exd sites ($E3N^{optimal}$) results in loss of regional specificity, as the $E3N^{optimal}$ enhancer is activated also by anterior Hox proteins. (b) Expression of *ventral veinless* (*vvl*) in the ectoderm of stage 11 embryos is controlled by the *vvl1+2* enhancer, which contains Hox-Exd high-affinity sites and is controlled by all Hox proteins in the different segments. In addition, this enhancer uses Hox monomeric binding sites as additional input. (c) *AP-2* is activated in posterior ectodermal border cells of the maxillary segment (Mx) by high Dfd levels (^{HIGH}Dfd) via the maxillary *AP2x-377* enhancer. In the rest of the maxillary segment, Dfd is expressed at lower concentrations (^{LOW}Dfd) and unable to activate *AP2x-377*. Optimization of Dfd-Exd sites in *AP2x-377* ($AP2x\text{-}377^{optimal}$) results in an increase in Dfd-Exd binding affinity. This results in the activation of $AP2x\text{-}377^{optimal}$ in anterior and ventral ectodermal cells of the maxillary segment. Mn: Mandibular segment; Mx: Maxillary segment; Lb: Labial segment. (d) In the first thoracic leg disc (T1), Scr interacts with different homeodomain TFs leading to defined target gene activations. In the Hth expression region, Scr interacts with Exd, whereas in the in the central area Scr interacts with Dll to activate a different set of genes. (e) Hox TFs may interact alone or with other factors to promote the opening of the chromatin using low-affinity binding, whereas binding to high-affinity binding sites could induce transcription of target genes.

relation to known mechanisms employed by Hox proteins and will discuss their general relevance with regards to segmental as well as cellular diversity. And finally, we will pose several avenues for further investigation necessary to bridge critical gaps in our understanding of how Hox proteins control specific and generic aspects of development and homeostasis.

2. Hox function in the ectoderm: A story of specificity

Several studies have shown that Hox proteins interact with many chromatin regions in the genome [9–14], indicating that the number of directly Hox-controlled target genes is high. However, only a very selected number of enhancers that are under the direct control of Hox TFs have been studied in detail [4], with many of them driving expression in the ectoderm. Thus, much of our knowledge on the molecular aspects of Hox specificity derives from Hox enhancer studies in this tissue.

2.1. Specificity: The same cofactor – Variable sequence binding affinities

In recent years, a number of studies have broadened our understanding of Hox specificity. These studies have shown that the precise control of target genes seems to rely on the suboptimal organization of *cis*-regulatory regions with the result being a trade-off between enhancer activity and specificity. In *Drosophila*, the study of the Ubx-regulated enhancers of the *shavenbaby* (*svb*) gene is one of the best examples that illustrate how Hox specificity can be achieved [15]. *svb* is the master control gene for trichome development [16], which develop in abdominal segments only. The work of Crocker and colleagues (2015) showed that the Hox TF Ubx, in complex with its cofactor Extradenticle (Exd) [4,17,18], which is termed Pbx in vertebrates, uses clusters of low-affinity binding sites in the *E3N* and *7H* enhancers to control the specific activation of *svb* [15] (Fig. 1a). Conversion of these sites to Ubx-Exd canonical/high-affinity binding sites, although increasing the activity of these regulatory regions (Fig. 1a), resulted in the loss of cell specificity due to the activation of these enhancers by Sex combs reduced (Scr)-Exd complexes in the more anterior region of the *Drosophila* embryo [15]. Thus, this work suggests that low-affinity sites are critical when individual Hox proteins are required to control specific targets while Hox high-affinity binding sites are present when Hox specificity is secondary, as for example in the case of target genes that are activated by all Hox proteins. The study of the *Drosophila ventral veinless* (*vvl*) enhancer *vvl1+2* supports this specificity-affinity trade-off model [19]. This enhancer is a *cis*-regulatory module responsible for the early activation of the *vvl* gene in a segmentally repeated pattern of patches that extend from the maxillary to the ninth abdominal segment on the lateral ectoderm. And consistent with the hypothesis put forward by Crocker and colleagues [15], this enhancer contains high affinity Hox-Exd sites and it is activated across the AP axis by different Hox TFs [19] (Fig. 1b). Importantly, this enhancer does not only depend on Hox-Exd activity but also on Hox monomeric binding events, showing that different mechanisms contribute to Hox target gene regulation and specificity. Intriguingly, such an Exd-independent Hox target gene regulation relying only on Hox monomer sites has already been demonstrated 20 years ago [20]. Galant and colleagues (2002) analysed haltere development, a modified flight appendage in the third thoracic segment in Diptera under the control of the Hox TF Ubx [21] and found that repression of the target gene *spalt* (*sal*) was under the control of Ubx via the binding to a specific haltere *cis*-regulatory element. Importantly, at that time repression was shown to be independent of cofactors (Exd and Hth) and to require only Ubx and its interaction with multiple monomeric Ubx binding sites. However, with recent findings indicating that many more proteins function as Hox cofactors [17,22–26], it seems not unlikely that these monomeric Hox sites are recognized by Hox proteins interacting with so far unknown cofactors.

The existence of Hox-Exd low-affinity binding sites have contributed to our understanding of Hox specificity by providing a model that explains how Hox-Exd complexes activate their target genes in a region-specific manner while avoiding activation by other Hox-Exd complexes. It is also in agreement with the general view in the transcription field that low-affinity binding sites rather than high-affinity ones play important regulatory roles in gene expression. However, a detailed study of the previously identified *AP-2* enhancer [11] suggests that this generalization might be too simplistic. The *AP-2* enhancer directs the expression of the TF encoding gene *AP-2* in a specific domain of the maxillary segment, under the control of the anterior Hox protein Dfd (Fig. 1c). An analysis of the *AP-2* enhancer using the *No Reads Left Behind* (*NRLB*) algorithm [27] employed in previous studies to identify Hox-Exd binding sites, revealed that the enhancer lacks canonical Dfd-Exd sites and contains instead several predicted low-affinity/non-canonical sites. However, in contrast to prediction, Pinto et al. [30] showed that the predicted relative affinity of these sites did not reflect their experimentally determined affinities. By conducting electrophoretic mobility shit assays, they determined the equilibrium dissociation constants (K_D) for the interaction of Dfd-Exd complexes with both high affinity canonical Dfd-Exd binding sites and *AP-2* non-canonical sites. The results showed that Dfd-Exd complexes bound the *AP-2* non-canonical Dfd-Exd sites with high affinity. But nonetheless the Dfd-Exd sites are under tight constraints: converting them into Dfd-Exd canonical binding sites resulted in a 2–2.5-fold increase in affinity that resulted in ectopic activation of the enhancer in other parts of the maxillary segment. Overall, Pinto et al. [30] showed that Dfd-Exd complexes bind strongly to non-canonical Dfd-Exd sites with the cell-specific activation of the *AP-2* enhancer resulting from a balance between the affinity of Dfd-Exd binding sites and the levels of Dfd protein present throughout the maxillary segment (Fig. 1c). More importantly, the authors identified that this configuration is crucial for Dfd function in the maxillary segment, as it allows Dfd to control and coordinate the morphogenesis of the different maxillary structures (Fig. 1c). These findings are consistent with recent data by Paul et al. [28], which showed that the Hox dosage (and not necessarily the Hox identity) is critical for driving segment-specific morphogenesis [28]. Furthermore, although Dfd-Exd and Scr-Exd complexes have been shown to bind the same Hox/Exd binding sequences with similar affinities [29], activation of the *AP-2* enhancer was restricted to the maxillary segment and was not induced in segments controlled by other Hox proteins [30]. In sum, this study showed that Dfd-Exd high-affinity sites can act highly specific in Hox target gene regulation, which challenges the view based on posterior Hox protein studies that low-affinity Hox-Exd sites are the rule to ensure Hox protein specificity.

2.2. Specificity: Different cofactors – Different binding sequences

The study of the posterior *svb* enhancer [15] and the anterior *AP-2* enhancer [30] raises the question whether anterior and posterior Hox-regulated enhancers generally rely on different mechanisms. A very recent study adds more information to this question. Feng and colleagues (2022) studied the development of thoracic legs in *Drosophila*, which are morphologically similar yet slightly distinct due to different Hox input: the adult morphology of first thoracic segment (T1) leg is controlled by the anterior Hox protein Scr, while the posterior Hox TF Ubx controls the third thoracic segment (T3) leg morphology [31]. By comparing the genome-wide DNA binding of the two Hox proteins in their respective cellular contexts, the authors found that about 8% of the binding events were different (despite the transcriptomes of the two tissues being very similar), suggesting that these two factors differentially interact with some enhancers while the majority of regulatory elements are bound by both proteins. Importantly, the authors found that interaction of Scr with its cofactor Exd explains many of the Scr specific binding events in T1 legs, and, consistent with the anterior *AP-2* enhancer [30] and in contrast to the posterior *svb* enhancer [15], the identified Scr-Exd sites were of high affinity [24] (Fig. 1d). Intriguingly,

binding specificity of Scr in T1 legs is not only mediated by interaction with the known cofactor Exd, but Feng et al. [24] found that Scr interacts with another HD TF, Distal-less (Dll), which enables Scr to bind to a different subset of loci in a different region of the leg primordium not controlled by Exd. Importantly, binding of Scr with the different cofactors to DNA is quite distinct, as cooperativity in binding was stronger between Scr and Exd in comparison to Scr and Dll. In addition, binding sequences were found to be different, consisting of overlapping Scr and Hox half-sites in the case of Scr and Exd, in contrast to two HD binding motifs separated by a spacer for Scr and Dll [24] (Fig. 1d). There are two important lessons to be learned from this study. First, there are other Hox cofactors than Exd, which has long been suggested and now the large number of additional Hox interactors identified in large scale screens [22,23,25] can be tested in a systematic manner. And second, it might be indeed that anterior and posterior Hox proteins use different mechanisms when it comes to binding specificity with anterior Hox proteins recognizing binding sites with a broad affinity spectrum while posterior Hox proteins rely more on low-affinity sites. How can this be explained? Although speculative, it is possible that during evolution the appearance of new Hox TFs may have required the development of innovative molecular mechanisms including the recognition of a wider range of binding sequences to ensure a specific function for these TFs. These innovations may have certainly continued throughout evolution as Hox proteins were recruited to the development of new structures, as for example, during the process of cephalisation. As reviewed in Hombria et al. [32], the expression of anterior *Hox* genes in anterior segments predate their involvement in the development of the head of vertebrates and arthropods. It is likely that the anterior Hox proteins have acquired new regulatory features to allow them to contribute to the development of head-specific structures and functions. Thus, it will be particularly interesting to compare Hox binding preferences in metazoan animals like *Hydra* or *Nematostella*, which have a much simpler organized head region, to the ones of more complex organisms like *Drosophila*.

2.3. Low- and high-affinity Hox-Exd sites: A matter of pioneering?

The studies of Crocker et al. [15], Feng and colleagues [24] and Pinto and colleagues [30] provide important insights into Hox specificity. Adding to the characterization of Hox-Exd sites from SELEX experiments [29], these studies reveal that the diversity of Hox (and Hox-Exd) binding sites is larger than initially expected, with Hox TFs recognizing multiple binding sites with different affinities. It also shows that Hox binding site affinities as well as Hox TF levels are key features in determining Hox specificity and critically control the specific and coordinated expression of Hox targets. Nonetheless, many questions remain open, and one of them is whether differences in Hox binding site affinities could play a role in other contexts. And this might be indeed the case! The study by Porcelli et al. [14] showed that *Drosophila* Hox proteins, in complex with Exd and Hth, are able to bind less accessible chromatin *in cellulo* [14], suggesting that Hox TFs could function as pioneer TFs. In addition, Desanlis et al. (2020) [33] provided compelling evidence that some Hox TFs act as pioneer factors in vivo to activate regional developmental programs [33]. Intriguingly, the analysis by Porcelli et al. [14] indicated that such less accessible chromatin regions are bound by Hox proteins via high-affinity binding sites, leading to a model whereby Hox TFs bind inaccessible chromatin through affinity-based competition with nucleosomes [14]. A recent study by Loker et al. [34] supports this observation and sheds some additional light on this topic. By profiling gene expression, chromatin accessibility and TF binding in *Drosophila* wing and haltere imaginal disks, the authors found that the Hox TF Ubx, which controls haltere development [35], can increase chromatin accessibility, supporting the notion that Hox (with their cofactors Exd and Hth) have pioneering function. Interestingly, this study shows that Ubx also closes chromatin regions in the haltere (relative to the wing) suggestive of anti-pioneering activity. As in the case of Porcelli et al. [14], the canonical high-affinity Ubx-Exd binding sequence was the most significantly enriched DNA motif in regions that change chromatin accessibility in the haltere. Does this mean that Hox proteins generally interact with high-affinity sites to change chromatin accessibility? This is so far not completely resolved, in particular when considering a recent study by Meers et al. [36], which provides some additional and important mechanistic insights into how TFs engage with different chromatin configurations and how TF binding site strengths contribute to these interactions. Instead of correlating ATAC- and ChIP-seq profiles to map TF binding to accessible and less accessible chromatin regions, the authors used CUT&RUN, a chromatin profiling method that uses MNase to specifically liberate DNA fragments bound by a target protein [37,38]. The advantage of this method is that it preserves information about the size of sequenced fragments [39], which is used to predict TF binding to different chromatin configurations, with small fragments of < 120 bp representing direct TF contacts with DNA while fragments of > 150 bp are indicative of nucleosomal binding [38]. Using CUT&RUN, Meers et al. [36] profiled chromatin interactions of several TFs during human embryonic stem cell (hESC) differentiation and found that in their context individual TFs access their targets either via direct DNA binding, as for example during chromatin remodeling or DNA unwrapping from the nucleosome, or via nucleosome binding, with TFs binding DNA wrapped around nucleosomes. They propose that nucleosome binding does not occur exclusively as a pioneer mechanism to uniquely access inactive chromatin but rather occurs in concert with other TF binding events providing TFs with the opportunity to sample diverse chromatin configurations, both accessible and occluded. Moreover, TFs are likely to engage low-affinity binding sites via a combination of direct DNA binding and nucleosomal interactions with the latter stabilizing low affinity interactions while high-affinity sites are bound more effectively through direct DNA binding. Thus, it could be that regions pioneered by TFs harbor a combination of low- and high-affinity TF binding sites to ensure chromatin opening and effective TF binding to regions made more accessible. In future, it will be highly relevant to apply the same strategy as Meers et al. [36] to other cellular contexts and to functionally test the identified Hox DNA motifs, low- and high-affinity sites, in regions changing their chromatin accessibility to ultimately resolve the underlying molecular cues of Hox pioneering function.

3. Hox function in the mesoderm: A focus on repression

As part of their function as master regulators of development, Hox TFs coordinate the regulation of target genes in order to deploy specific morphogenetic programs and to specify and maintain the identity of cell lineages. The majority of studies have focused on identifying and characterizing *cis*-regulatory regions activated by Hox TFs. Thus, most of the information regarding the regulation of Hox *cis*-regulatory regions concerns the role of Hox TFs in transcriptional activation. In comparison, only a small number of *cis*-regulatory elements have been identified, which are repressed by Hox proteins. This includes the repression of the *sal* gene in halteres, which is Exd- and Hth-independent and critically depends on multiple Hox binding sites with which Ubx interacts as monomer [40]. This study was very important, as most of the work on Hox regulated target genes focused at that time (and still today) largely on Hox-Exd protein complexes and their binding sites. Galant et al. [40] demonstrated that other modes of Hox target gene regulation exist already 20 years ago, and that clusters of Hox monomer binding sites are sufficient to repress a Hox target [40]. However, repression can also involve Hox-Exd complexes, which is the case for the *DMX-R* enhancer. This element is important for the repression of *Distal-less* (*Dll*) in the embryonic abdomen. Silencing of *Dll* requires the Hox TFs Ubx and Abd-A, which bind together with the cofactors Exd and Hth in a cooperative manner to *DMR-X* [41]. However, this repression does not only require Hox-Exd but two additional proteins, the HD TF Engrailed (En) and the Forkhead TF Sloppy paired 1 (Slp1), highlighting that factors other than Exd and Hth can function as Hox cofactors.

We will focus now on the mesoderm where recent studies provided important insights into how Hox proteins maintain the repression of a large number of genes to ensure the development of this germ layer.

3.1. *Alternative fate repression: The Polycomb-Hox axis*

By analysing the genome-wide chromatin binding of Ubx in the embryonic mesoderm, Domsch et al. [12] showed that most of the chromatin interactions they found were in the vicinity of transcriptionally inactive genes [12]. In line with a repressive function, interference with Ubx in the mesoderm demonstrated that a substantial fraction of these genes was indeed de-repressed in the absence of Ubx. Domsch et al. [12] further showed that the repressive activity of Ubx is achieved by interacting with Pleiohomeotic (Pho), a DNA-binding component of the Polycomb group (PcG) complex, and therefore, possibly recruiting the PcG complex, which sets repressive chromatin marks found at Ubx-Pho co-bound sites [12]. An interesting observation that resulted from this study concerned the nature of genes silenced by Ubx in the mesoderm: many of them are normally active in other tissues (Fig. 2a). This means that one of the functions of Hox proteins in tissue development is the restriction of cellular and temporal plasticity, which is critical for stably maintaining cell fates. This is absolutely essential for an organism to function properly, and Hox proteins are ideally suited to fulfill this generic function, since Hox TFs are expressed in most/all cells during the life-time of an organism. The study of Domsch et al. [12] shows that they do so by stabilizing the Polycomb complex, which is known to establish, together with the counteracting Trithorax group proteins, an epigenetic cellular memory by faithfully maintaining transcription states determined early in embryogenesis after cell fates have been specified [42]. Many studies highlight that the PcG complex is majorly involved in repressing most alternative genetic programs in any given cell type, while excluding the subset that is required in that cell type [43–45]. This is highly reminiscent to what Hox proteins do according to Domsch et al. [12], suggesting that the Hox-PcG interaction is generally used to imprint a permanent memory on cells to maintain their specific identity by repressing the expression of alternative fate genes. And in line with this, the study by Domsch et al. [12] indicates that Ubx controls the repression of alternative fate genes in the embryonic nervous system, which includes mesodermal genes, while neuronal genes are repressed in the mesoderm [12]. One fundamental question arising from these findings is whether interaction with the Polycomb complex is a general mechanism used by Hox proteins to maintain transcriptional repression. There are several evidences supporting this hypothesis. Previous profiling of mesodermal chromatin interactions of Pleiohomeotic (Pho), a DNA binding component of the Polycomb complex thought to recruit the complex to specific regions of the genome [46,47], revealed a large number of loci bound by Pho in this tissue during embryogenesis [48]. Importantly, many of these sites were co-bound by Ubx in the mesoderm and characterized by repressive H3K27me3 histone marks [12]. Decreasing Ubx levels in the mesoderm resulted in a loss of Pho chromatin interaction at about one quarter of the Ubx-Pho co-bound sites [12], suggesting interdependency of the two factors in repressing many target genes. This finding was the basis for another study, which showed that Ubx represses the early mesodermal master regulator *twist* (*twi*) by interacting with the NK-HD TF Tinman (Tin) at the *twi* promoter, which results in the recruitment of Pho (and the Polycomb complex) [49]. Thus, the Hox-Polycomb interaction is not only involved in the repression of alternative fate genes but also in the timely inactivation of early specification genes, which is critical for cell differentiation (Fig. 2b). Based on these finding it will be interesting to tackle the following questions: Does the Polycomb – Hox interaction play a general role in gene silencing, for example also in Ubx-repression of *sal* in the haltere and *Dll* in the abdomen? What about the repression of anterior *Hox* gene expression by posterior Hox proteins, the so-called posterior suppression [50,51]? Interestingly, it has been shown by Garaulet and colleagues that the Hox TF Ultrabithorax (Ubx) repressed its own transcription in a PcG-dependent manner [52]. Thus, in future, it will be very important to tackle whether these repressive (and many other) events are mediated by the Hox-Polycomb interaction axis not only in the mesoderm but also in other tissues.

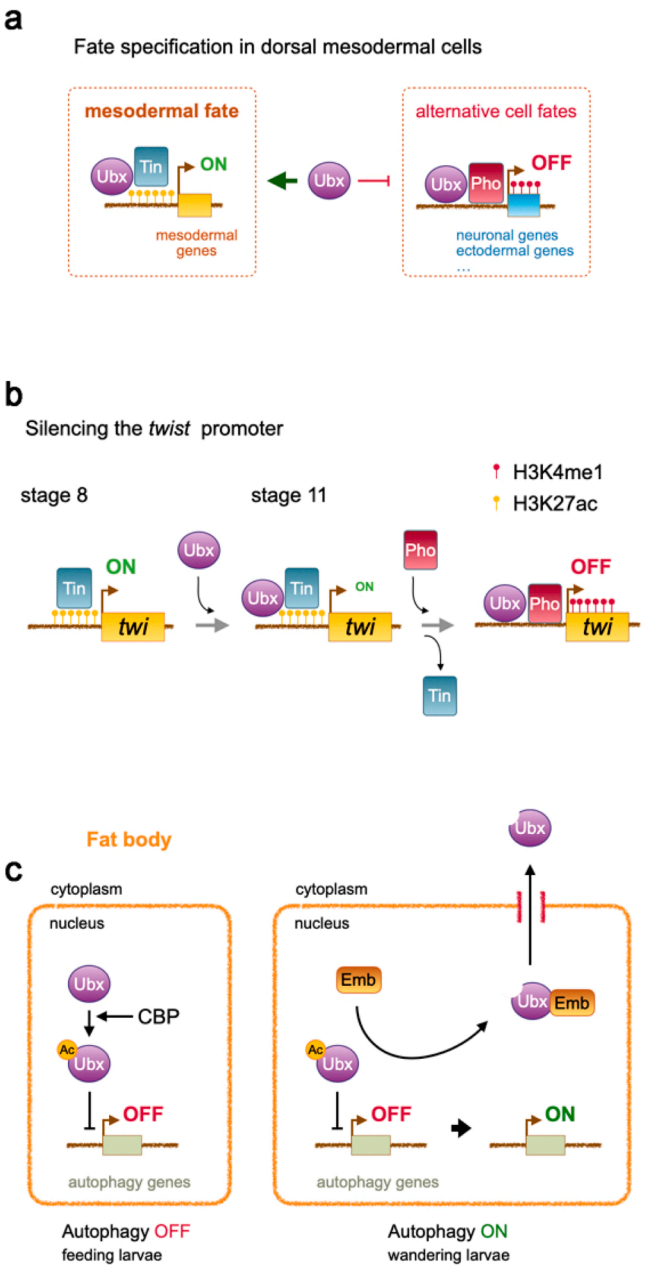


Fig. 2. Hox transcriptional repressive activity in the mesoderm. (a) In dorsal mesodermal cells, Ubx activates mesoderm-specific genes via the interaction with tissue specific TF (ex.: Tinman (Tin)). In addition, Ubx represses genes normally expressed in alternative cell types in collaboration with the Polycomb complex protein Pleiohomeotic (Pho). (b) Ubx plays a role in the repression of early mesoderm specification genes like *twist* (*twi*). In stage 8 embryos, *twi* is activated by the mesoderm-specific TF Tin. As soon as Ubx starts to be expressed (at stage 11), Ubx interacts with *twi* promoter, leading to displacement of Tin and recruitment of Pho followed by silencing of this region and repression of *twi*. (c) In feeding *Drosophila* larvae, Hox proteins including Ubx repress autophagy related genes in the fat body thereby contributing to the maintenance of this tissue. In wandering larvae, Hox/Ubx protein is actively removed from the nucleus thereby releasing the repression of autophagy genes resulting in the destruction of this tissue.

3.2. Repression of autophagy genes: To leave or not to leave the nucleus

Repression by Hox proteins plays also a very important role in a different cellular context, the *Drosophila* fat body, a mesoderm-derived tissue that controls the use and storage of energy to meet the energy demands of the fly during developmental stages [53]. During metamorphosis, a transition phase between the larval and adult stage in holometabolous insects as *Drosophila*, many tissues including the fat body are extensively remodeled or in some cases even re-built to form the adult structures. This degradation and recycling of cellular materials that allows for tissue remodeling is achieved via the induction of autophagy, which is mediated by the different Hox proteins in the fat body [54]. The mechanims used to control autophagy in this tissue is remarkable for various reasons. Contrary to expectation Hox TFs do not initiate autophagy at the onset of fat body degradation by activating autophagy genes, but rather by acting as potent repressors of these genes throughout larval development (Fig. 2c) with repression being released when autophagy is initiated [54]. This release of Hox-mediated repression is achieved by the inactivation of *Hox* gene expression in the fat body [54] and more importantly by the active removal of Hox proteins from fat body nuclei (Fig. 2c) when the tissue is ready to enter autophagy [55]. How are the Hox proteins forced out of the nuclei? The work of Duffraisse et al. [55] shows that Hox proteins contain an unconventional nuclear export signal (NES) which overlaps with a highly conserved motif, the hexpeptide (HX) or W-containing motif [55]. This motif is better known for its function in mediating the interaction with Exd [18]. However, in the context of the fat body, this motif mediates the interaction with the major exportin protein Embargoed (Emb), also known as CRM1, thereby inducing the active export of Hox proteins from fat body nuclei, leading to a release of the Hox-mediated autophagy repression. What prevents the interaction of Hox proteins with Emb during larval stages when autophagy is silenced? Intriguingly, the results of Duffraisse and colleagues [55] suggest that Hox nuclear export is controlled by the acetyltransferase CBP/p300, which seems to interfere with the interaction of Hox and Emb by modifying the W-containing motif via acetylation/deacetylation mechanisms [55]. There are two important lessons to be learned from the fat body example: first, Hox transcriptional activity is not only controlled by the interaction with different co-factors but also by the removal of the protein from the nucleus, the major place of TF action (Fig. 2c). This is a rather elegant mechanism, as the shuttling of regulatory proteins in and out of the nucleus allows for a fast and reversible adjustment of whole gene expression programs to environmental and developmental cues. And second, transcriptional repression could be a major mechanism generally used by Hox proteins to control the development of individual lineages. What is so far not resolved is how the repression of autophagy genes is mediated. Does this also involve the interaction of Hox proteins with the PcG complex? Although unclear, a very recent report shows that Polycomb group proteins play a role in the repression of autophagy in other cellular contexts. Puri and colleagues (2022) found that during the differentiation of mouse embryonic stem cells (mESCs) autophagy is induced and identified the Polycomb enzyme EZH2 as an important regulator in this process [56]. It is known that *Hox* gene expression is activated in mESCs only when these cells start to differentiate [57], raising the intriguing possibility that the temporal activation of autophagy at the onset of mESC differentiation relies on the same mechanism, the interaction of Hox and PcG proteins. Thus, it will be interesting to study the Hox-Polycomb interactions in additional developmental contexts where Hox proteins play a role.

4. Tissue specificity of Hox action: A global view

We have discussed so far prominent examples of Hox action in the ectoderm and mesoderm/fat body, that shows that Hox proteins, despite being expressed in different cells and tissues types, perform highly specific functions. How can this be? How can Hox proteins act so specifically in individual cell and tissue types?

Several determinants dictate the output of TFs, which includes amongst others context-dependent interaction partners [58,59] and post-translational modifications [60–62] as well as the cell-specific chromatin architecture defined during development [63]. We will focus in this section on Hox interaction partners, as most of our understanding on Hox regulatory specificity stems from the analysis of such proteins, in particular the Three Amino acids Loop Extension (TALE) family of HD-containing TFs, which includes the *Drosophila* Exd and the vertebrate Pbx1–4 proteins [20]. These proteins cooperatively bind DNA with Hox TFs thereby increasing their binding specificity [20–23]. However, although TALE TFs are important for Hox function, they can only partially explain how Hox TFs can function in a context/tissue-specific manner in vivo, in particular as they are expressed in many different cell types themselves [27]. In order to identify proteins interacting with Hox proteins in a tissue-specific manner, two recent large-scale approaches have been performed, which greatly increased our understanding of Hox functional specificity.

4.1. BiFC: Large-scale probing of Hox-protein interactions using pre-selected candidates

To identify tissue-specific Hox cofactors, one approach tested the ability of pre-selected TFs to interact with the Hox proteins Ubx and Abd-A in two different tissues, the ectoderm and the mesoderm, using Bimolecular Fluorescence Complementation (BiFC). This technique relies on the property of monomeric fluorophores to be reconstituted from two sub-fragments upon spatial proximity [64], and when combined with the UAS-GAL4 system [65] enables the easy and non-invasive visualization of protein interactions in specific cell and tissue types in vivo. Bischof and colleagues tested 260 (of the approximately 1000) TFs in *Drosophila* embryos and found that about two-third of the interactions were common to Ubx and Abd-A [23]. This result is not unexpected, as these two Hox TFs control many identical developmental processes. And in line, these common interactions were found to occur mostly in the same tissues. In contrast, only few interactions were shown to be more specific for one or the other Hox TF. For example, Spalt major (Salm), a TF important for oenocyte specification [66], was found to be enriched among the Abd-A (but not Ubx) interactome, and in line, Abd-A (but not Ubx) has been shown to play a role in this process [67,68]. Another important finding of this study is the enrichment of Zinc finger and HD TFs among the Ubx and Abd-A interaction partners. This result is highly interesting in the light of recent studies, showing that cell-specific combinations of HD TFs, so-called homeo-codes, control the specification and differentiation of individual cell (sub)types, for example in the nervous system in different animals [69,70]. Intriguingly, Reilly et al. [71] showed that within the developing nervous system of the nematode *C. elegans* unique combinations of HD TFs are expressed in individual neurons and specify their identity [71], while Velten et al. [70] demonstrated that even within one neuronal subtype, the motoneurons, cell specific expression of multiple HD TFs defines the properties and positions of differentiated motoneurons along the major body axes of *Drosophila* embryos [70]. Thus, Hox TFs might achieve specificity in target gene regulation controlling cell specific features by interacting in each cell with combinations of other HD TFs (besides Exd/Pbx). In future, it will be interesting to test the Hox cofactor activity of all BiFC identified candidates to unravel the complexity of interaction partners conferring specificity to Hox proteins.

4.2. BioID: An unbiased proteomics approach to identify Hox interactors

One of the advantages but also problems of the BiFC approach is that it uses pre-selected candidates to test an interaction with a protein of interest. However, in order to reveal the regulatory complexes that drive the multi-faceted outputs of TFs in the animal, unbiased methods are required to identify stable and transient TF interaction networks in vivo.

Proximity-labeling of proteins coupled with mass spectrometry (MS) offers a systematic analysis of spatially restricted proteomes, providing a comprehensive understanding of cellular functions in different contexts [28–32]. Carnesecchi and colleagues established proximity-dependent biotin identification (BioID) in combination with the UAS-GAL4 system [65] to capture tissue-specific interactomes in vivo. They used the Hox TF Ubx as their model and identified Ubx interaction partners in the mesoderm and the nervous system. Their results were remarkable for various reasons. They first revealed that Ubx interactomes in the different tissue were largely non-overlapping, meaning that Ubx interacts with very different proteins in the different tissues. Contrary to expectations, they found that Ubx interacts mostly with ubiquitously expressed proteins playing roles in general transcriptional regulation, such as chromatin remodeling proteins or RNA processing factors, in a tissue-specific manner. Only very few of the interaction partners were expressed in tissue-restricted manner and only some of them were TFs. This included in particular Tinman, a HD TF specifically expressed in the mesoderm [72], and the CP2 TF Grainy head (Grh), which is active in the ectoderm and nervous system but not in the mesoderm [73,74]. Genetic analyses showed that, in vivo, all the identified interactions acted tissue-specifically. These results suggest that Hox proteins control gene expression programs at multiple regulatory levels and not only by recognizing in combination with other DNA binding proteins distinct sequence codes written in enhancers and promoters. In line with these findings, a very recent study showed that Ubx modulates transcriptomes by regulating co-transcriptional splicing, which is suggested to be mediated by a dynamic interplay between Ubx and PolII on chromatin [75].

One question arising from this behavior is how interaction specificity between Ubx and other proteins is achieved. It is known that Ubx protein, like many other TFs, harbors intrinsically disordered domains that are important for selecting interacting partners [63–65]. Thus, the few tissue-restricted Ubx interactors identified in this study, like Tin or Grh, could be responsible for Ubx's differential interaction potential by binding to these intrinsically disordered domains. They could enforce tissue-specific protein conformations that would allow for specific interaction with only a subset of the many Ubx interactors. In addition, these interactions could be responsible for the differential gene expression in the two tissues. In line with this assumption, comparison of genome-wide binding profiles of Tin [76] and Grh [77] to Ubx chromatin interactions [12] identified substantial co-binding of Ubx-Tin and Ubx-Grh to non-overlapping regions in the different tissues. Importantly, Carnesecchi and colleagues identified Tin to function as a specific Hox cofactor, as Ubx and Tin were found to interact on the well-described *dpp674* enhancer to combinatorically activate *decapentaplegic* (*dpp*) expression in the visceral mesoderm [25]. In the same line, it has been shown just recently that Grh acts as a Hox cofactor in controlling apoptosis of specific neuroblasts in the CNS [78]. What makes the interaction of Hox proteins with Tin and Grh particularly interesting is their ability to function as pioneer TFs. This has been suggested for Tin in heart and dorsal vascular musculature formation [79], while Grh has been shown to pioneer and open epithelial enhancers in *Drosophila* [80]. Thus, both tissue-restricted factors could interact with Ubx (and other Hox proteins) to control regional and lineage-specific gene expression programs by changing the chromatin landscape.

Another question arising from this study is whether Hox proteins interact indeed with only very few other TFs in a certain tissue. Very likely not and besides technical reasons the most reasonable explanation for the identification of only a few TFs is that most of the Hox - TF interactions critical for Hox output occur only in very few or single cells and can thus not be captured with such an approach. These interactions, however, might be identified by BiFC, which relies on the overexpression of the tagged interaction partners [22,23]. Thus, targeted BioID is a valuable and powerful method, as it captures dynamic, weak and specific interactions in vivo in an unbiased manner, and is ideally complemented with other approaches like BiFC. Intriguingly, a very recent study has just shown that the interaction of a TF with a no-DNA binding proteins can change the DNA binding specificity of TF [81]. Thus, it seems likely that some of the BioID identified factors could even play a role in changing the binding preference of Hox proteins even if these factors do not interact with DNA themselves and have so far not been shown to control gene expression at the transcriptional level.

Although these large scale-studies have provided valuable data of Hox TFs interactomes and contributed to a deeper understanding of Hox specificity, BiFC and BioID approaches rely on the overexpression of the proteins of interest. Thus, the possibility of false positives is intrinsic to these methods, requiring validation through additional experiments. Therefore, in the future, it will be interesting to combine these two methodologies to CRISPR to characterize Hox interactomes at the endogenous level in several cellular/tissue contexts during different stages of development.

5. Conclusions and perspective

Hox proteins are very important regulators which control the morphological diversity along the AP axis by acting highly specifically in all the different tissue types present in the segments they are expressed in. So far, many studies focused on understanding how different Hox proteins control different regional identities without considering that different mechanisms might be at work in different cell and tissue types. We have summarized in this review recent findings on how Hox proteins control target gene expression in different tissues, with a focus on the ectoderm and the mesoderm. Studies in the ectoderm show that Hox TF binding affinity in coordination with Hox expression levels are two key features driving Hox specificity in this tissue. The question is whether similar mechanisms are used in other tissues, like the mesoderm. This seems likely however, a recent study challenges this view. By analysing genomic chromatin interaction of Ubx in the mesoderm and nervous system, Folkendt et al. [82] found that the classical Ubx/Hox motif was only present at Ubx bound sites in the nervous system while a novel non-canonical Hox motif was enriched among the regions bound by Ubx in the mesoderm [82]. Such differences in binding preferences could be due to Hox TFs interacting with many different proteins that change their binding preferences in a tissue-specific manner. Based on recent single-cell transcriptome data and large-scale interaction screens this will include a multitude of HD TFs (and very likely other classes of TFs), which will act as Hox cofactors, in a similar fashion as Exd/Pbx and Hth/Meis. Furthermore, we hypothesize that non-DNA binding proteins could cooperate with Hox proteins to change their binding specificity, which would expand the repertoire of binding sequences even more. In sum, we assume that Hox target gene regulation relies on a fine-tuned interaction of Hox proteins with many functionally different proteins, which will assemble, depending on their expression levels, in a cell-specific manner on enhancers thereby driving precise target gene regulation using a multitude of different sequences. Resolving this complexity will require multi-modal single cell profiling strategies to capture all levels of regulation, which will reveal similarities and differences in Hox target gene regulation in different tissues and will provide more insights into the Hox specificity problem in the in vivo chromatin context. And these studies might indeed reveal that Hox proteins follow as few "regulatory rules" as possible to be able to function in so many different lineages and developmental stages, as formulated by Pearson and colleagues almost 20 years ago [4].

Conflict of interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgments

We would like to emphasize that as a result of the vast amount of published information as well as due to practical limitations (restriction on the number of citations) we tried to present the most recent and relevant data that would support our description of the tissue-specific function and specificity of Hox TFs. Therefore, we would like to apologize to all whose work was not cited due to space limitations.

References

[1] J. Garcia-Fernàndez, The genesis and evolution of homeobox gene clusters, Nat. Rev. Genet. 6 (2005) 881–892.
[2] E.B. Lewis, A gene complex controlling segmentation in Drosophila, Nature 276 (1978) 565–570.
[3] T.C. Kaufman, M.A. Seeger, G. Olsen, Molecular and genetic organization of the antennapedia gene complex of drosophila melanogaster, Adv. Genet 27 (1990) 309–362.
[4] J.C. Pearson, D. Lemons, W. McGinnis, Modulating Hox gene functions during animal body patterning, Nat. Rev. Genet. 6 (2005) 893–904.
[5] J.C. Hombría, B. Lovegrove, Beyond homeosis – HOX function in morphogenesis and organogenesis, Differentiation 71 (2003) 461–476.
[6] E.Hox Sánchez-Herrero, Targets and cellular functions, Science 2013 (2013), 738257.
[7] M.B. Noyes, et al., Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites, Cell 133 (2008) 1277–1289.
[8] M. Affolter, M. Slattery, R.S. Mann, A lexicon for homeodomain-DNA recognition, Cell 133 (2008) 1133–1135.
[9] D. Shlyueva, A.C.A. Meireles-Filho, M. Pagani, A. Stark, Genome-wide ultrabithorax binding analysis reveals highly targeted genomic loci at developmental regulators and a potential connection to polycomb-mediated regulation, PLOS One 11 (2016), e0161997.
[10] M. Slattery, L. Ma, N. Négre, K.P. White, R.S. Mann, Genome-wide tissue-specific occupancy of the hox protein ultrabithorax and hox cofactor homothorax in drosophila, PLOS One 6 (2011), e14686.
[11] S. Sorge, et al., The cis-regulatory code of Hox function in Drosophila, EMBO J. 31 (2012) 3323–3333.
[12] K. Domsch, et al., The Hox transcription factor Ubx stabilizes lineage commitment by suppressing cellular plasticity in Drosophila, eLife 8 (2019), e42675.
[13] P. Agrawal, F. Habib, R. Yelagandula, L.S. Shashidhara, Genome-level identification of targets of Hox protein Ultrabithorax in Drosophila: novel mechanisms for target selection, Sci. Rep. 1 (2011) 205.
[14] D. Porcelli, B. Fischer, S. Russell, R. White, Chromatin accessibility plays a key role in selective targeting of Hox proteins, Genome Biol. 20 (2019) 115.
[15] J. Crocker, et al., Low affinity binding site clusters confer hox specificity and regulatory robustness, Cell 160 (2015) 191–203.
[16] F. Payre, A. Vincent, S. Carreno, Ovo/svb integrates Wingless and DER pathways to control epidermis differentiation, Nature 400 (1999) 271–275.
[17] R.S. Mann, M. Affolter, Hox proteins meet more partners, Curr. Opin. Genet. Dev. 8 (1998) 423–429.
[18] R.S. Mann, S.-K. Chan, Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins, Trends Genet. 12 (1996) 258–262.
[19] C. Sánchez-Higueras, et al., In vivo Hox binding specificity revealed by systematic changes to a single cis regulatory module, Nat. Commun. 10 (2019) 3597.
[20] C.M. Walsh, S.B. Carroll, Collaboration between Smads and a Hox protein in target gene repression, Development 134 (2007) 3585–3592.
[21] F. Roch, M. Akam, Ultrabithorax and the control of cell morphology in Drosophila halteres, Development 127 (2000) 97–107.
[22] B. Hudry, S. Viala, Y. Graba, S. Merabet, Visualization of protein interactions in living Drosophila embryos by the bimolecular fluorescence complementation assay, Bmc Biol. 9 (2011) (5–5).
[23] J. Bischof, et al., Generation of a versatile BiFC ORFeome library for analyzing protein–protein interactions in live Drosophila, eLife 7 (2018), e38853.
[24] S. Feng, et al., Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors, Nat. Commun. 13 (2022) 3808.
[25] J. Carnesecchi, et al., Multi-level and lineage-specific interactomes of the Hox transcription factor Ubx contribute to its functional specificity, Nat. Commun. 11 (2020) 1388.
[26] J. Carnesecchi, P.B. Pinto, I. Lohmann, Hox transcription factors: an overview of multi-step regulators of gene expression, Int. J. Dev. Biol. 62 (2018) 723–732.
[27] C. Rastogi, et al., Accurate and sensitive quantification of protein-DNA binding affinity, Proc. Natl. Acad. Sci. USA 115 (2018) E3692–E3701.
[28] R. Paul, et al., Hox dosage contributes to flight appendage morphology in Drosophila, Nat. Commun. 12 (2021) 2892.
[29] M. Slattery, et al., Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins, Cell 147 (2011) 1270–1282.
[30] P.B. Pinto, et al., Specificity of the Hox member deformed is determined by transcription factor levels and binding site affinities, Nat. Commun. 13 (2022) 5037.
[31] G. Struhl, Genes controlling segmental specification in the Drosophila thorax, Proc. Natl. Acad. Sci. USA 79 (1982) 7380–7384.
[32] J.C.-G. Hombría, M. García-Ferrés, C. Sánchez-Higueras, Anterior Hox genes and the process of cephalization, Front. Cell Dev. Biol. 9 (2021), 718175.
[33] I. Desanlis, et al., HOX13-dependent chromatin accessibility underlies the transition towards the digit development program, Nat. Commun. 11 (2020) 2491.
[34] R. Loker, J.E. Sanner, R.S. Mann, Cell-type-specific Hox regulatory strategies orchestrate tissue identity, Curr. Biol. 31 (2021) 4246–4255 (e4).
[35] S.D. Weatherbee, G. Halder, J. Kim, A. Hudson, S. Carroll, Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the Drosophila haltere, Gene Dev. 12 (1998) 1474–1482.
[36] M.P. Meers, D.H. Janssens, S. Henikoff, Pioneer factor-nucleosome binding events during differentiation are motif encoded, Mol. Cell 75 (2019) 562–575.
[37] P.J. Skene, J.G. Henikoff, S. Henikoff, Targeted in situ genome-wide profiling with high efficiency for low cell numbers, Nat. Protoc. 13 (2018) 1006–1019.
[38] P.J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites, eLife 6 (2017), e21856.
[39] N.A. Kent, S. Adams, A. Moorhouse, K. Paszkiewicz, Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing, Nucleic Acids Res. 39 (2011) (e26–e26).
[40] R. Galant, C.M. Walsh, S.B. Carroll, Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites, Development 129 (2002) 3115–3126.
[41] B. Gebelein, D.J. McKay, R.S. Mann, Direct integration of Hox and segmentation gene inputs during Drosophila development, Nature 431 (2004) 653–659.
[42] L. Ringrose, R. Paro, Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins, Annu. Rev. Genet. 38 (2004) 413–443.
[43] Y.B. Schwartz, V. Pirrotta, Polycomb silencing mechanisms and the management of genomic programmes, Nat. Rev. Genet. 8 (2007) 9–22.
[44] Y.B. Schwartz, T.G. Kahn, G.I. Dellino, V. Pirrotta, Polycomb silencing mechanisms in Drosophila, Cold Spring Harb. Symp. 69 (2004) 301–308.
[45] B. Schuettengruber, H.-M. Bourbon, L.D. Croce, G. Cavalli, Genome regulation by polycomb and trithorax: 70 years and counting, Cell 171 (2017) 34–57.
[46] C. Fritsch, J.L. Brown, J.A. Kassis, J. Muller, The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene, Development 126 (1999) 3905–3913.
[47] J.L. Brown, C. Fritsch, J. Mueller, J.A. Kassis, The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing, Development 130 (2003) 285–294.
[48] J. Erceg, et al., Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements, Gene Dev. 31 (2017) 590–602.
[49] K. Domsch, J. Schröder, M. Janeschik, C. Schaub, I. Lohmann, The Hox transcription factor Ubx ensures somatic myogenesis by suppressing the mesodermal master regulator twist, Cell Rep. 34 (2021), 108577.
[50] A. Gonzalez-Reyes, N. Urquia, W.J. Gehring, G. Struhl, G. Morata, Are cross-regulatory interactions between homoeotic genes functionally significant? Nature 344 (1990) 78–80.
[51] A. González-Reyes, G. Morata, The developmental effect of overexpressing a Ubx product in Drosophila embryos is dependent on its interactions with other homeotic products, Cell 61 (1990) 515–522.
[52] D.L. Garaulet, D. Foronda, M. Calleja, E. Sánchez-Herrero, Polycomb-dependent Ultrabithorax Hox gene silencing induced by high Ultrabithorax levels in Drosophila, Development 135 (2008) 3219–3228.
[53] E. Parra-Peralbo, A. Talamillo, R. Barrio, Origin and development of the adipose tissue, a key organ in physiology and disease, Front. Cell Dev. Biol. 9 (2021), 786129.
[54] A. Banreti, B. Hudry, M. Sass, A.J. Saurin, Y. Graba, Hox proteins mediate developmental and environmental control of autophagy, Dev. Cell 28 (2014) 56–69.
[55] M. Duffraisse, et al., Role of a versatile peptide motif controlling Hox nuclear export and autophagy in the Drosophila fat body, J. Cell Sci. 133 (2020), jcs241943.
[56] D. Puri, A. Kelkar, G.S. Bhaskar, D. Subramanyam, Autophagy-cell death balance is maintained by Polycomb-mediated regulation during stem cell differentiation, Biorxiv (2022), 486206, https://doi.org/10.1101/2022.03.29.486206.
[57] S.B. Bahrami, M. Veiseh, A.A. Dunn, N.J. Boudreau, Temporal changes in Hox gene expression accompany endothelial cell differentiation of embryonic stem cells, Cell Adhes. Migr. 5 (2011) 133–141.
[58] Z. Ji, A.D. Sharrocks, Changing partners: transcription factors form different complexes on and off chromatin, Mol. Syst. Biol. 11 (2015) 782.
[59] N. Bobola, S. Merabet, Homeodomain proteins in action: similar DNA binding preferences, highly variable connectivity, Curr. Opin. Genet. Dev. 43 (2017) 1–8.
[60] A. Draime, L. Bridoux, Y. Graba, R. Rezsohazy, Post-translational modifications of HOX proteins, an underestimated issue, Int. J. Dev. Biol. 62 (2018) 733–744.
[61] M. Reichlmeir, L. Elias, D. Schulte, Posttranslational modifications in conserved transcription factors: a survey of the TALE-homeodomain superclass in human and mouse, Front. Cell Dev. Biol. 9 (2021), 648765.
[62] T.M. Filtz, W.K. Vogel, M. Leid, Regulation of transcription factor activity by interconnected post-translational modifications, Trends Pharmacol. Sci. 35 (2014) 76–85.
[63] L. Isbel, R.S. Grand, D. Schübeler, Generating specificity in genome regulation through transcription factor sensitivity to chromatin, Nat. Rev. Genet. (2022) 1–13, https://doi.org/10.1038/s41576-022-00512-6.
[64] T.K. Kerppola, Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells, Nat. Protoc. 1 (2006) 1278–1286.
[65] A.H. Brand, N. Perrimon, Targeted gene expression as a means of altering cell fates and generating dominant phenotypes, Development 118 (1993) 401–415.

[66] P.R. Elstob, V. Brodu, A. Gould, P. spalt-dependent switching between two cell fates that are induced by the Drosophila EGF receptor, Development 128 (2001) 723–732.
[67] G. Wang, L. Gutzwiller, D. Li-Kroeger, B. Gebelein, A Hox complex activates and potentiates the Epidermal Growth Factor signaling pathway to specify Drosophila oenocytes, PLOS Genet. 13 (2017), e1006910.
[68] L.M. Gutzwiller, et al., Proneural and abdominal Hox inputs synergize to promote sensory organ formation in the Drosophila abdomen, Dev. Biol. 348 (2010) 231–243.
[69] O. Hobert, Homeobox genes and the specification of neuronal identity, Nat. Rev. Neurosci. 22 (2021) 627–636.
[70] J. Velten, et al., Single-cell RNA sequencing of motoneurons identifies regulators of synaptic wiring in Drosophila embryos, Mol. Syst. Biol. 18 (2022), e10255.
[71] M.B. Reilly, C. Cros, E. Varol, E. Yemini, O. Hobert, Unique homeobox codes delineate all the neuron classes of C. elegans, Nature 584 (2020) 595–601.
[72] N. Azpiazu, M. Frasch, tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila, Gene Dev. 7 (1993) 1325–1340.
[73] K. Venkatesan, H.R. McManus, C.C. Mello, T.F. Smith, U. Hansen, Functional conservation between members of an ancient duplicated transcription factor family, LSF/Grainyhead, Nucleic Acids Res. 31 (2003) 4304–4316.
[74] M. Narasimha, A. Uv, A. Krejci, N.H. Brown, S.J. Bray, Grainy head promotes expression of septate junction proteins and influences epithelial morphogenesis, J. Cell Sci. 121 (2008) 747–752.
[75] J. Carnesecchi, et al., The Hox transcription factor Ultrabithorax binds RNA and regulates co-transcriptional splicing through an interplay with RNA polymerase II, Nucleic Acids Res. 50 (2021) 763–783.
[76] H. Jin, et al., Genome-wide screens for in vivo tinman binding sites identify cardiac enhancers with diverse functional architectures, PLOS Genet. 9 (2013), e1003195.
[77] M. Nevil, E.R. Bondra, K.N. Schulz, T. Kaplan, M.M. Harrison, Stable binding of the conserved transcription factor grainy head to its target genes throughout drosophila melanogaster development, Genetics 205 (2017) 605–620.
[78] R. Sipani, R. Joshi, Hox genes collaborate with helix-loop-helix factor Grainyhead to promote neuroblast apoptosis along the anterior-posterior axis of the Drosophila larval central nervous system, Genetics (2022), https://doi.org/10.1093/genetics/iyac101.
[79] B.W. Busser, et al., Contribution of distinct homeodomain DNA binding specificities to drosophila embryonic mesodermal cell-specific gene expression programs, PLOS One 8 (2013), e69385.
[80] J. Jacobs, et al., The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes, Nat. Genet. 50 (2018) 1011–1020.
[81] P. Rieu, et al., The F-box cofactor UFO redirects the LEAFY floral regulator to novel cis-elements, Biorxiv (2022), https://doi.org/10.1101/2022.06.14.495942.
[82] L. Folkendt, I. Lohmann, K. Domsch, An evolutionary perspective on hox binding site preferences in two different tissues, J. Dev. Biol. 9 (2021) 57.