Nascent Wars: How RARE enhancers regulate *Hoxb* gene transcripts

By

Zainab Afzal
M.S., University of the District of Columbia and Lombardi Comprehensive Cancer Center
Georgetown University, 2014
B.Sc., University of the District of Columbia, 2013
A.A.S.N., University of the District of Columbia, 2011

Submitted to the graduate degree program in Anatomy and Cell Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Committee Co-Chair/Mentor: Robb Krumlauf

_____
Committee Co-Chair: Julie Christianson

_____
Committee member: Kausik Si

_____
Committee member: Irfan Saadi

_____
Committee member: Paul Trainor

_____
Committee member: Chad Slawson

Date Defended: April 6, 2021

The dissertation committee for Zainab Afzal certifies that this is the approved version of the following dissertation:

Nascent Wars: How RARE enhancers regulate *Hoxb* gene transcripts

_____

Committee Co-Chair/Mentor: Robb Krumlauf

_____

Committee Co-Chair: Julie Christianson

_____

Graduate Director: Julie Christianson

Date Approved: May 11, 2021

Abstract

During development of an embryo, the anterior-posterior (A-P) body axis has to be properly defined, in order for body segments and organs to form in their correct locations. A family of transcription factors, the *Hox* genes, play an important role in determining the A-P body axis. In mammals, there are four *Hox* gene cluster (A-P) that are arranged on four different chromosomes. For proper A-P axis formation, the spatial and temporal domains of *Hox* genes within the clusters have to be precisely initiated and strictly maintained. Therefore, it is critical to understand the regulatory inputs that dictate the expression of *Hox* genes. It has been identified that *Hox* genes are responsive to morphogen gradients, e.g., retinoic acid (RA), along with cis-regulatory elements and long non-coding RNAs that are interspersed within the clusters.

In this study, I have looked at retinoic acid response elements (RAREs), which help to incorporate RA signaling to regulate the expression of *Hox* genes. I have optimized the single molecule fluorescent *in situ* hybridization technique (smFISH), to look at newly synthesized or nascent *Hoxb* transcripts in mouse tissue sections. I found that three RAREs – DE, B4U, and ENE while having individual inputs into regulating *Hoxb* genes, also appear to work together to ensure proper levels of nascent transcripts in the neural tube and adjacent somites. Furthermore, I see that these RARE have different inputs along the axial level of the embryo, such that DE plays a greater role anteriorly while B4U plays a greater role posteriorly. In the DE-B4U double mutants, I observe that antagonism from the individual mutants is neutralized as levels of nascent transcripts in the double mutant appear similar to wildtype. Results from the triple DE-B4U-ENE mutants highlight that these RAREs are critical for nascent transcription of *Hox* genes, as in the triple mutant we observe a very low number of nascent transcripts. These results demonstrate

how multiple RAREs integrate response to RA to fine-tune the transcript levels of *Hox* genes. My study is a starting point towards understanding the *in vivo* transcriptional dynamics of *Hox* genes in mouse embryos. This study is a step towards better understanding the enhancer-promoter interactions required for regulating genes that ensure proper embryonic development.

Acknowledgements

I want to thank friends and family for their undying support, kindness, and constant check-ins. You guys have made KC home, and the PhD journey full of happy memories. You have made this PhD journey truly memorable, and I am both grateful and indebted to you all. Especially want to thank all my friends especially Viraj, Vivek, Younshim, Khyati, Tseday and Habib for all the fun times, lots of discussions on science and otherwise, and for the countless times we have whined, cried, and jumped for joy about our projects and life events. Thank you for listening to my talks over and over and reading my work over the years and thank you for always being there through life's ups and downs. I want to thank my little sis Nooro and my cuz bestie Safiyah for always being there. You both are my adorable gems, and I am so happy you are my family. I really want to thank my mom and dad. I love you both dearly and want to thank you for always letting me chase after my dreams and supporting my decisions. I wouldn't be where I am without your love and guidance.

I want to thank Robb for his mentorship, guidance, and kindness over the years. When projects worked or didn't work, you allowed me to experiment and pursue different ideas. Thank you for supporting me in my experimental adventures, and also for helping me to focus when I started doing a lot of different things :) I also want to thank all my current and past lab members. Especially want to thank Alice, Hugo, Nishal, Narendra, Bony, and Christof. You guys are amazing, and I am super glad to have met you as lab mates and as friends and sub-mentors. Thank you Thank you for sitting through countless practice talks and giving feedback on several drafts of my work. I also want to thank Carrie for helping keep everything functional in the lab. I

also want to thank Leanne for guidance and feedback and thank Lisa for all the administrative support.

I would like to thank all my committee members (Kausik Si, Paul Trainor, Irfaan Saadi, Julie Christianson, and Chad Slawson) for their questions, guidance, continued encouragement over the years. Thank you for always giving me constructive feedback on my work. Thank you, especially Julie, for always making sure things were going well every semester and lending a listening and comforting space. I really want to thank the Anatomy department for constant support over the years. I admit that at times I have dreaded Thursday afternoons, but I have had a lot of good times in the many journal club meetings. I also want to thank IGPBS program that helped me choose the best lab and allowed me to happily get started on my PhD journey; Thank you Dr. Werle and Dr. Abmayr for all the encouragement and advices.

I want to thank all the core facilities in Stowers Institute, especially the microscopy core. It has been the most amazing time collaborating with you all. You all taught me all about imaging and analysis. I can vividly remember not following more than half the conversations you all had in the beginning, and because of all the kind and thorough explanations I follow a whole lot more now :) Also, thank you to the LASF and LAT team (esp. Heidi, Kathleen, Haylee, Andrea, Tim, and Michael), molecular biology, and genome engineering (esp. MaryEllen and Kyle), Histology (esp. Karen, Nancy, and Yongfu), Aquatics (esp. Darin, Adam, Kathie, and Angela), and Ariel in Bioinformatics. Thank you for always humoring my questions, giving me lots of guidance and helping me with different experiments. I appreciate it immensely. Thank you thank you.

Table of Contents

Abstract iii

Acknowledgements v

Table of Contents vii

List of Figures xi

List of Tables xii

Chapter 1: Introduction 1

1.1 Historical perspective of *Hox* genes – Why are they called the *Hox* genes? 3

1.2 Conserved *Hox* genes 5

1.3 Role of *Hox* genes in development 8

1.3.1 Role of Hox genes in the neuroectoderm 9

1.3.2 Role of Hox genes in the mesoderm 15

1.3.3 Hox genes and development of limbs, reproductive and other organs 18

1.3.4 Loss-of-function and gain-of-function studies for Hox genes 22

1.3.5 Human mutations and diseases linked to Hox genes 26

1.4 Regulatory inputs for *Hox* genes 27

1.4.1 Morphogen gradients 27

1.4.2 Non-coding RNAs 31

1.4.3 Cis-regulatory elements 35

1.5 Transcriptional control of *Hox* genes 37

1.5.1 Hox genes as transcription factors 37

1.5.2 How enhancers regulate transcription 38

1.5.3 RAREs in the Hoxb cluster 42

Chapter 2: Materials and Methods 44

2.1 Materials 45

2.1.1 Mouse lines 45

2.1.2 Fixing embryos 45

2.1.3 Small molecule Florescent in situ hybridization (smFISH) 46

2.1 Methods 46

2.2.1 Genotyping of animals 46

2.2.1.1 Genotyping by Transnetyx *46*
2.2.1.2 Genotyping for RARE mutants *47*
2.2.1.3 Genotyping for Hobbit lines *48*
2.2.2 Staging of embryos 49
2.2.3 Collection of embryo samples 49
2.2.3.1 Embryo collection for smFISH *50*
2.2.3.2 Embryo collection for RNA sequencing *50*
2.2.4 Histological preparation of embryos *50*
2.2.5 Whole mount *in situ* hybridization with DIG labelled probes 51
2.2.6 smFISH optimized protocol 54
2.2.6.1 Probe design *54*
2.2.6.2 Probe synthesis *54*
2.2.6.3 Optimized smFISH protocol *55*
2.2.6.4 Imaging with Nikon spinning disk *58*
2.2.6.5 Deep learning (DL) processing on images *59*
2.2.6.6 Distance measurement methods *59*
2.2.6.7 Measurements post-DL on images *60*
2.2.6.8 Data Analysis on measurements made post DL *61*
2.2.6.9 Statistical analysis of data *63*
2.2.6.10 Visual representation of images *63*
2.2.7 RNA seq sample processing 64
2.2.8 RACE (3' and 5') to determine *Hobbit* start and end site 65
2.2.9 *Hobbit* CRISPR experiments 65
2.2.9.1 Identification of guides *65*
2.2.9.2 Deletions strategies *66*
2.2.9.3 Homology Directed Repair (HDR) strategies *67*
2.2.10 Microinjections and electroporation's of *Hobbit* CRISPR 68
Chapter 3: Characterization of nascent *Hoxb* transcripts in wildtype embryos 69
3.1 Background 70
3.2 *Hoxb* cluster is an ideal genomic locus to study transcriptional regulation 70

3.3 Single molecule Fluorescent *in situ* hybridization optimized for *in vivo* detection of newly synthesized RNA molecules of coding and non-coding *Hoxb* transcripts 71
3.4 Transcription of *Hoxb* genes is primarily independently regulated 78
3.5 Independent regulation of *Hoxb* genes is true at multiple axial levels in the embryo 80
3.6 Trends of nascent *Hoxb* transcripts can recapture known patterns of processed *Hoxb* transcripts along the embryonic axis 83
3.6.1 Single cell resolution with smFISH can detect dorsal-ventral expression differences 84
3.6.2 Sum DAPI intensity can be used to estimate cell counts for regions of interest 85
3.7 Non-coding RNAs within the *Hoxb* cluster are also differentially regulated like the coding transcripts 87
3.8 Trends of expression for *Hoxb* transcripts in the adjacent somites are similar to trends in the neural tube 89
3.9 Each *Hoxb* gene has its own pattern of expression in different tissues along the embryonic axis 90
3.10 Levels of *Hoxb* nascent transcripts correlate with levels of processed transcripts 91
3.11 Differential gene expression in wildtype mouse embryos in response to RA induction .. 92
3.11.1 RA induction has different effect on nascent transcripts of adjacent Hoxb genes..... 93
3.11.2 RA induction does not appear to effect levels of processed transcripts 93
3.12 Discussion 96
Chapter 4: Characterization of nascent *Hoxb* transcripts in RARE mutants 98
4.1 RARE mutants differentially regulate *Hoxb* coding and non-coding transcripts 99
4.2 DE RARE differentially effects anterior and posterior *Hoxb* genes in the tail 102
4.3 B4U RARE mediates RA response to predominantly regulate *Hoxb* genes in the tail 103
4.3.1 RA mediated response is altered in the B4U mutant 104
4.3.2 B4U element potentially a repressor or is in competition with adjacent RAREs 106
4.3.3 Hobbit lncRNA regulates Hoxb genes downstream of B4U element 108
4.4 Intermediate transcriptional activation of *Hoxb* transcripts by ENE RARE 110
4.5 Competition between enhancers for regulating *Hoxb* nascent transcripts 111
4.5.1 Compound RARE mutant demonstrates balance or competition between elements.. 112
4.5.2 Triple mutant highlights how critical RAREs are for nascent transcription 113
4.6 Co-activation of *Hoxb* transcripts is actively not favored 113
4.7 RAREs differentially regulate *Hoxb* transcripts along the embryonic axis 115

4.8 Transcriptional activation by RAREs could be mediated though dynamic changes in enhancer-promoter distances 119
4.9 Discussion 122
Chapter 5: Discussion 125
5.1 Different nascent transcriptional dynamics observed for single RARE mutants 127
5.2 Enhancer RNAs may be playing a role towards dynamic nascent transcription 128
5.3 Competition or balance observed between activity of critical RAREs 129
5.4 Transcriptional regulation of *Hox* genes 130
References 136

List of Figures

Figure 1-1 Sketches of the diverse life forms in nature 2
Figure 1-2 Hox genes clusters and their spatio-temporal expression 4
Figure 1-3 Overlapping domains of *Hox* gene expression in hindbrain and spinal cord of a developing mouse embryo 6
Figure 1-4. Development time for the embryonic ectoderm from which the nervous system also develops 10
Figure 1-5. *Hox* expression domains and the opposing RA and FGF gradients that define expression boundaries 11
Figure 1-6. Neural tube at the posterior (tail end) of the embryo. NT- neural tube, S- somites... 13
Figure 1-7. Expression domains of the Hox gene clusters in the neural tube..... 15
Figure 1-8. Mesoderm and its derivatives. NT-neural tube, NC- neural crest cells 17
Figure 1-9. Mechanism of RAR/RXRs binding to RAREs to activate transcription of genes..... 36
Figure 1-10 Potential transcriptional models for the *Hoxb* cluster 41
Figure 1-11. RAREs present in the middle of the *Hoxb* cluster 43
Figure 3-1 Tracking *Hoxb* gene expression in the neural tube along the embryonic axis..... 70
Figure 3-2 Schematic of *Hoxb* genomic locus 71
Figure 3-3 smFISH on fixed mouse embryo cryo-sections 72
Figure 3-4 Visualizing nascent transcripts in mouse embryo sections 73
Figure 3-5 Depiction of how Deep learning algorithm detects nascent transcripts 74
Figure 3-6 Max projections to visualize nascent transcripts..... 75
Figure 3-7 Max projections to visualize nascent transcripts in grayscale 76
Figure 3-8 Heatmaps of nascent transcript intensity..... 77
Figure 3-9 Graphs of *Hoxb* nascent transcripts..... 78
Figure 3-10 Probabilities of co-localization of *Hoxb* nascent transcripts..... 80
Figure 3-11 Nascent transcript profiles over neural tube of the tail 81
Figure 3-12 Nascent transcript profiles over neural tube of the body section 82
Figure 3-13 Expression of *Hoxb* genes in neural tube over whole embryo..... 84
Figure 3-14 Dorsal-ventral expression difference for *Hoxb* genes..... 84
Figure 3-15 StarDist analysis using fiji on wildtype body and tails sections 86
Figure 3-16 Counting of cells for each ROIs..... 87
Figure 3-17 Wildtype expression profile for exact nascent counts in neural tube of the tail 88
Figure 3-18 Wildtype expression profile for exact nascent counts in adjacent somites of the tail89
Figure 3-19 Wildtype expression profile in regions from the head and body sections 90
Figure 3-20 Graphs of average RPKM values of *Hoxb* transcripts 92
Figure 3-21 Relative expression of transcripts in wildtype and RA induced embryos 93
Figure 3-22 Differential gene expression in r4 region..... 94
Figure 3-23 Differential gene expression in trunk region..... 95
Figure 3-24 Differential gene expression in tail region 95

Figure 3-25 Venn diagram for differentially expressed genes 96
Figure 4-1 Schematic of Hoxb4-Hoxb5 intergenic region with RARE mutants 99
Figure 4-2 Visualization of nascent transcripts for wildtype and RARE mutants 100
Figure 4-3 Plots of exact nascent transcripts for wildtype and RARE mutants 101
Figure 4-4 Co-localization of nascent transcripts in DE RARE mutants 103
Figure 4-5 Transgenic assay for mouse B4U element in zebrafish 104
Figure 4-6 Heatmaps of differentially expressed genes for B4U RARE 105
Figure 4-7 Co-localization of nascent transcripts in B4U RARE mutants 107
Figure 4-8 Levels of nascent transcripts in the *Hobbit*-ko mutant 109
Figure 4-9 Co-localization of nascent transcripts in ENE RARE mutants 111
Figure 4-10 Plots of exact nascent transcripts for wildtype, double, and triple RARE mutants 112
Figure 4-11 Probabilities of random co-localizations 114
Figure 4-12 Probabilities of observed co-localizations 114
Figure 4-13 Nascent transcripts in the adjacent somites in the tail sections 116
Figure 4-14 Relative density of nascent transcripts in anterior body sections 117
Figure 4-15 Relative intensity of nascent transcripts for head region 118
Figure 4-16 Spot fitting triple co-localized transcripts 119
Figure 4-17 Spot fitting data to infer chromatin interactions 120
Figure 4-18 Hypothetical looping model to demonstrate distance changes 121
Figure 4-19 Transcriptional hub with simultaneous activation of hox genes 122
Figure 4-20 Dynamic interactions of RAREs with gene promoters 123
Figure 5-1 Dynamic interaction through which RAREs activate transcription of *Hoxb* genes .. 131
Figure 5-2 Nascent transcripts and transcriptional condensates 132
Figure 5-3 Transcriptional hub with multiple RAREs 134

List of Tables

Table 1-1 *Hox* gene mutations with focus on *Hoxb* cluster genes 23
Table 2-1 Processing of plasmids for probe synthesis 51
Table 2-2 SmFISH probes labelled in-house and optimized to use on mouse embryo sections .. 55
Table 2-3 Probe-set combinations used on alternate mouse embryo sections 57
Table 4-1 Differentially expressed genes for B4U mutant embryo 105
Table 4-2 GO term analysis for B4U RA induced embryos 106

Chapter 1: Introduction

In nature there exist a huge variety of living organisms (Figure 1-1). The beauty of different living forms is evident in the types of plants, animals, insects, and micro-organisms that have lived or are currently living in this world. Each species has its own footprint, displaying its own unique pattern. Some patterns are visibly discernible by eye, while some have to be observed at the microscopic level. How do all these different patterns in living organisms come about? Another fascinating aspect of the different body forms in biology, is the ability of a single cell to develop into a multicellular organism with different internal structures and organs systems. How are organ systems able to have their own specific cellular programming, especially when they all start with similar information as they develop from a single cell? One thread that unites both questions, the variety in body forms and the formation of different body segments/structures, are the existence of *Homeobox* or *Hox* genes (Reviewed in Swalla, 2006).

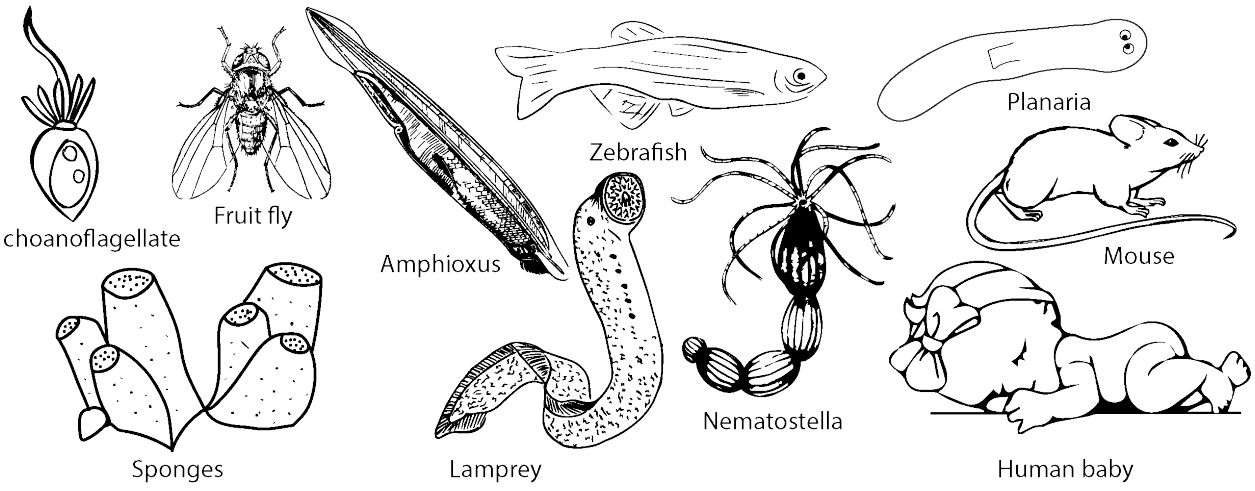


Figure 1-1 Sketches of the diverse life forms in nature

The variety of species depicted that either have Hox genes (e.g., fruit fly, amphioxus, lamprey, zebrafish, mouse, humans), don't have Hox genes (e.g., sponges), or have homeobox genes (TALE family) but not the *bona fide* Hox class genes (e.g., choanoflagellate).

1.1 Historical perspective of *Hox* genes – Why are they called the *Hox* genes?

The phenomenon of observing changes in body structures was described in 1790 by Geothe in the "Metamorphosis of plants" (Wolfgang von Goethe, 1952). However, metapmorphy had different meanings in different fields of biology, so to be more specific William Bateson about 127 years ago coined the term homeosis (Bateson, 1894). He defined homeosis as "something has been changed into the likeness of something else." It was later identified that homeotic mutations were changes happening to genes – to their expression levels or expression domains. The term homeosis has persisted to mean transformations of one body structure into another. When *Drosophila* was beginning to be established as a model system in the laboratory of Thomas H Morgan, they observed homeotic transformations in their mutants. In particular, Charles Bridges observed that the haltere of the *Drosophila* changed into another pair of wings in the Bithorax mutant (Bridges & Dobzhan, 1933); a segment identity change where the third segment changes into another second segment in the fly. Another homeotic mutant in the fly resulted in a pair of legs forming instead of antennae in the antennapedia mutant (Balkaschina, 1929).

Further investigations into the genes that were able to cause homeotic transformations led to the identification of a homeobox domain (McGinnis et al., 1984; Scott & Weiner, 1984) – a conserved 60 amino acid stretch that was present in all HOX proteins. Hence, genes that were able to cause homeotic transformations and contained the homeobox domain were called *Hox* genes. Bridges who in testing the hypothesis that genes underwent duplications, identified multiple *Hox* genes that are arranged in tandem and complemented each other in affecting the abdominal region of the fly – the Bithorax complex (Lindsley & Zimm, 1992). Other studies identified the Antennapedia complex of *Hox* genes and established that in *Drosophila* there

exists a split cluster of *Hox* genes consisting of two *Hox* complexes. Work by Ed Lewis showed that *Hox* genes within the cluster displayed collinearity, such that mutations in the 3' end of the cluster effected anterior regions of the embryo while mutations in the 5' end of the cluster effected the posterior regions (Lewis, 1978). Hence, there appeared to be a linear relationship between the order of the *Hox* genes in the genome and the region of the body plan they affect.

Through comparative analysis between *Drosophila* and other species, it was observed that not only are *Hox* genes clusters conserved across both invertebrate and vertebrate species, the spatio-temporal collinearity of the clusters is also maintained (Beeman et al., 1989; Boncinelli et al., 1988; Carrasco et al., 1984; Coen & Meyerowitz, 1991; Dolecki et al., 1986; Duboule & Dolle, 1989; Graham et al., 1989; Kelsh et al., 1993; Kenyon & Wang, 1991; Krumlauf et al., 1987). *Hox* gene paralogs are genes separated by a duplication event while orthologs are genes separated by species. The translated sequence of homeoboxes of anterior *Hox* genes in flies are more similar with their mammalian counterparts then their own posterior paralogs.

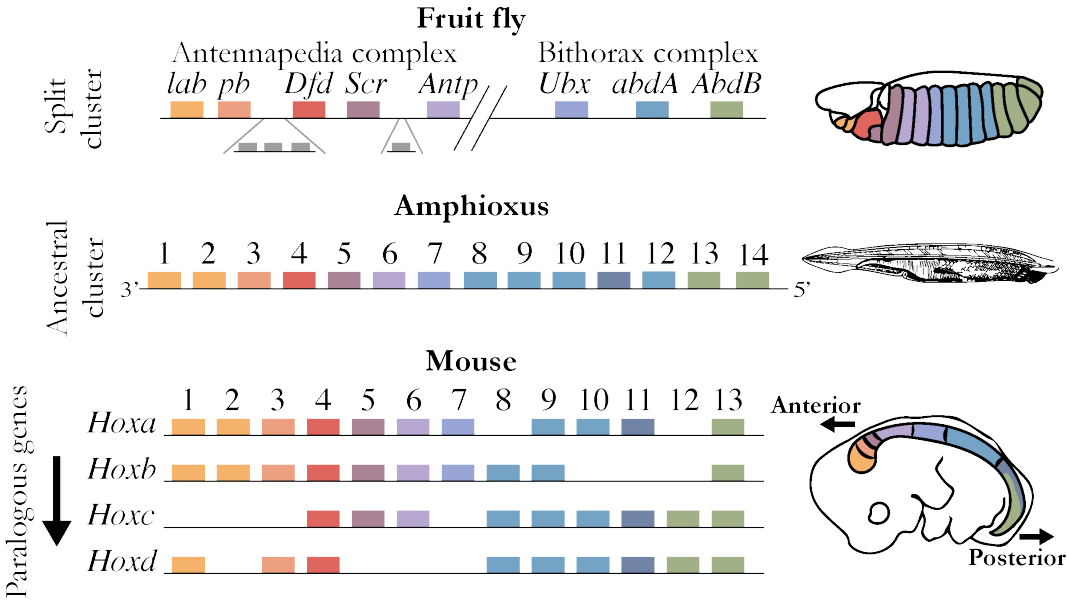


Figure 1-2 Hox genes clusters and their spatio-temporal expression

The *Hox* gene clusters of fruit fly and mouse compared to the amphioxus *Hox* cluster. The amphioxus cluster is thought to reflect the ancestral chordate *Hox* cluster. In mouse, the paralogous group members have similar expression domains (arrow to show members in same paralog group). The paralogous group members, and similar genes in terms of expression domains within all the embryos, are represented in the same color. The genes at the 3' end of the clusters are expressed early in development as well as anteriorly in the embryo, while genes at the 5' end are expressed later in development and posteriorly in the embryo. (Adapted from OpenStax, 2013)

In addition, regardless of species, genes located at the 'beginning' of the complex exert their 'function' in anterior domains while genes located at the end of the complexes regulate the identity of more posterior domains (Regulski et al., 1987b). In mammals in particular, there are four *Hox* gene clusters arranged in four distinct chromosomal locations within the genome (Figure 1-2). This comparison of *Hox* genes to mammals indicated that gene duplication events had resulted in four clusters as compared to the split one in *Drosophila*, and the presence of *Hox* clusters in both *Drosophila* and mouse indicated the presence of an ancestral *Hox* cluster that predated the separation of invertebrates and vertebrates. The *Hox* cluster in Amphioxus is thought to reflect the ancestral chordate (includes all species that at some point in their life cycle contain a spinal cord, notochord, etc.) cluster that contains 14 *Hox* genes.

1.2 Conserved *Hox* genes

The evolutionarily conserved *Hox* genes have been identified in almost all animal kingdom phyla (exceptions being Porifera which include sponges that have no *Hox* genes, and Ctenophora which include comb jellies where presence of *Hox* genes is debated) (de Rosa et al., 1999; Ryan et al., 2007). Currently, *Hox* genes have been characterized in a variety of invertebrate and vertebrate species. While in many species *Hox* genes appear in clusters, there are also some species where they are not clustered in the genome (Seo et al., 2004), and either exhibit collinearity or don't exhibit collinearity (Schiemann et al., 2017). In some species such as *Drosophila*, non-*Hox* genes can also be found within the cluster. However, in the four *Hox* clusters in mammals, which are organized on four different chromosomes, while they are no

intervening non-*Hox* genes present there are several non-coding RNAs within the clusters (De Kumar & Krumlauf, 2016). The gene numbers ranging from 1-13 exhibit spatio-temporal collinearity (Krumlauf, 1994) to exhibit a *Hox* code of expression (Kessel & Gruss, 1991). Genes within the clusters turn on sequentially in a space and time dependent manner from the 3' end to the 5' end (Gaunt, 2018) to form a nested pattern of expression, but with different anterior (and sometimes also posterior) domains within the embryo (Figure 1-3).

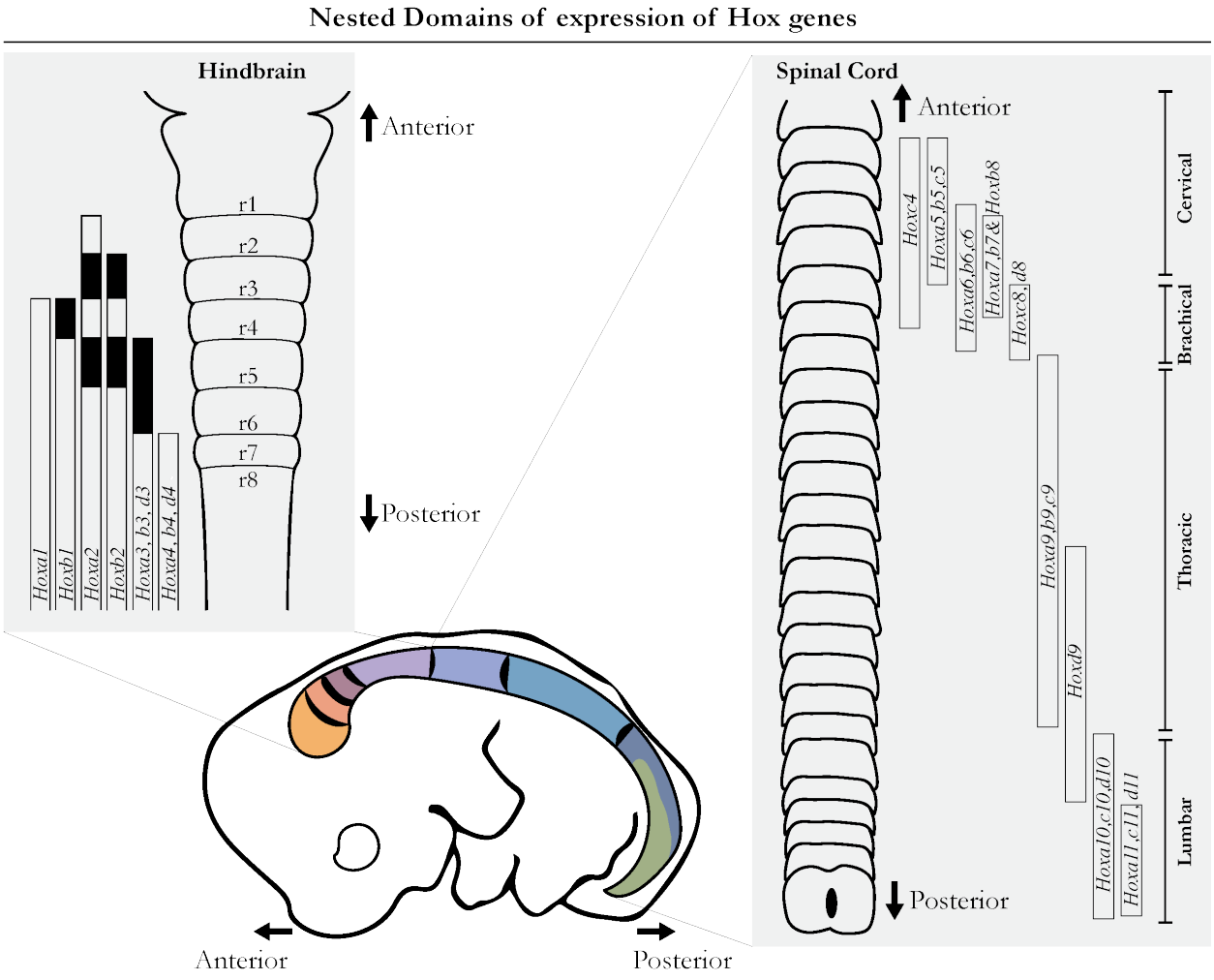


Figure 1-3 Overlapping domains of *Hox* gene expression in hindbrain and spinal cord of a developing mouse embryo

The hindbrain (left panel) and spinal cord (right panel) as viewed from the back (dorsal) side of the embryo. The nested *Hox* expression domains are shown with bars or rectangles against rhombomere boundaries - r1 to r8 in the

hindbrain, or against the spinal cord boundaries i.e., cervical, brachial, thoracic, and lumbar. In the hindbrain, the darker shade depicts higher expression levels, e.g., *Hoxb1* in r4. (Adapted from Gonçalves et al., 2020)

In a developing embryo, *Hox* genes are expressed in particular segments along the body axis, and often display sharp boundaries of expression rostrally (anteriorly or towards the front end of the embryo). Within the four *Hox* clusters, not all paralog group members are present, such that there are a total of 39 *Hox* genes in both humans and mice. Each member of the paralogous group - genes similar based upon their sequences and the relative position within the clusters, are often expressed in partly redundant domains in an embryo and at similar developmental time (Regulski et al., 1987a). This could be indicative of functional redundancy in the system to ensure proper formation of body structures. Usually, the most 3' *Hox* genes have more anterior boundaries than the posterior genes, and also the paralogous group members have similar expression boundaries in the hindbrain. However, there are a few exceptions in the expression of paralogous *Hox* group members in the hindbrain, which during development is segmented into 8 rhombomeres (r1-r8). The exceptions are 1) *Hoxa2* is the only *Hox* gene expressed in the r2 segment of the rhombencephalon, while its paralog *Hoxb2* is expressed posterior to the r2/r3 boundary. 2) *Hoxb1* while initially expressed up to the r3/r4 boundary, its expression gets restricted to r4, and at the stage of r4 restriction *Hoxb1*'s paralog *Hoxa1* is no longer expressed in the hindbrain. 3) *Hoxc4* and *Hoxd4* have different boundaries of expression also such that *Hoxd4* has an anterior rhombomere boundary of expression compared to *Hoxc4*. The boundary of *Hoxc4* appears to not be tied to any rhombomere segment but instead to the starting of the spinal cord. Also, while *Hox* gene paralog groups 1-4 are expressed in the spinal cord as well, their expression is generally much lower than in the hindbrain (Reviewed in Nolte & Krumlauf, 2006).

It was also observed that *Hox* genes exhibit "posterior prevalence" such that the posterior *Hox* genes repress the expression and function of the anterior *Hox* genes (Akam, 1987; Duboule &

Morata, 1994; Hafen et al., 1984; Harding et al., 1985) while paralogous group members may function together or similarly even though they are in different clusters. The paralogous genes have been shown to compensate for loss of one member (Wellik et al., 2002), but in some instances have evolved to have a unique function as the expression domains or regulatory domains of the paralogous group members differ (McClintock et al., 2002; Miguez et al., 2012; Singh et al., 2020). Over the years, a lot of work has been done to better understand the *Hox* gene clusters, but the exact logic behind why *Hox* genes are even present in a cluster (Hart et al., 1985; Levine et al., 1985), and how they are regulated to exhibit collinearity are areas still being actively studied. Collectively, *Hox* genes play a role not only initially in development by defining anterior-posterior body axis and conferring segmental identity to pattern tissues especially in bilaterians (Akam, 1987; Harding et al., 1985; Nolte et al., 2013; Patterson & Potter, 2004; Philippidou & Dasen, 2013; Roux & Zaffran, 2016), but also later in the adult life where they have been indicated to play roles in proper functioning of several organ systems (Chang et al., 2002; Du & Taylor, 2015; Neville et al., 2002; Qian et al., 2018; Rux & Wellik, 2017; Yamamoto et al., 2003).

1.3 Role of *Hox* genes in development

There are three stages of *Hox* gene expression that are critical for proper anterior-posterior (A-P) patterning of the embryo especially during the axial elongation (or growth of embryo posteriorly to anteriorly). It has been suggested that the amount of time cells spend in the primitive streak during gastrulation defines what *Hox* code they will eventually express (Reviewed in Deschamps & van Nes, 2005). This is because *Hox* gene transcription is first initiated in cells in the posterior part of the primitive streak. While these cells do not form future axial or paraxial tissues (neural tube or adjacent somites) in the embryo, they begin the process of sequential activation of *Hox*

genes towards the anterior part of the primitive streak or the node. It is from the primitive node that cells extend out and move anteriorly to form the axis of the embryo (Forlani et al., 2003). In the primitive streak, while *Hox* gene expression continues to spread up to and through the node outward and anteriorly, lineage tracing experiments have shown that these *Hox* gene expression domains are not final. Hence, the final patterns of expression are not are not fixed in the node. After generation of the new neuroectoderm and mesoderm, there are independent modifications made to the *Hox* gene expression in both the neural tissue and somites (Alexander et al., 2009; Forlani et al., 2003; Whiting et al., 1991). This final pattern of *Hox* genes has to be established and maintained for proper segmental identity along the A-P axis.

1.3.1 Role of *Hox* genes in the neuroectoderm

Formation of neuroectoderm is the first step in development of the nervous system (Figure 1-4). The nervous system develops from a thickened epithelial sheet – the neural plate. The neural plate extends anteriorly from the primitive streak and folds to form the neural tube which further segments into forebrain, midbrain, hindbrain, and spinal cord. The forebrain and midbrain do not express *Hox* genes, while the hindbrain and spinal cord do. These two *Hox* expressing regions form distinctly, anteriorly the hindbrain and posteriorly the spinal cord. The hindbrain develops from a portion of the epiblast (cell layer that forms one of the germ layers – the ectoderm) that is situated anterolaterally to the node during the late stage of the primitive streak (Forlani et al., 2003; Lawson et al., 1991). While the spinal cord is made when the embryonic axis elongates from cells produced by the stem cell region in the primitive node. Since the neural tube elongates anteriorly with cells proliferating at the posterior end, the embryo is more developmentally advanced at the anterior (or rostral) end than at posterior (or caudal) end.

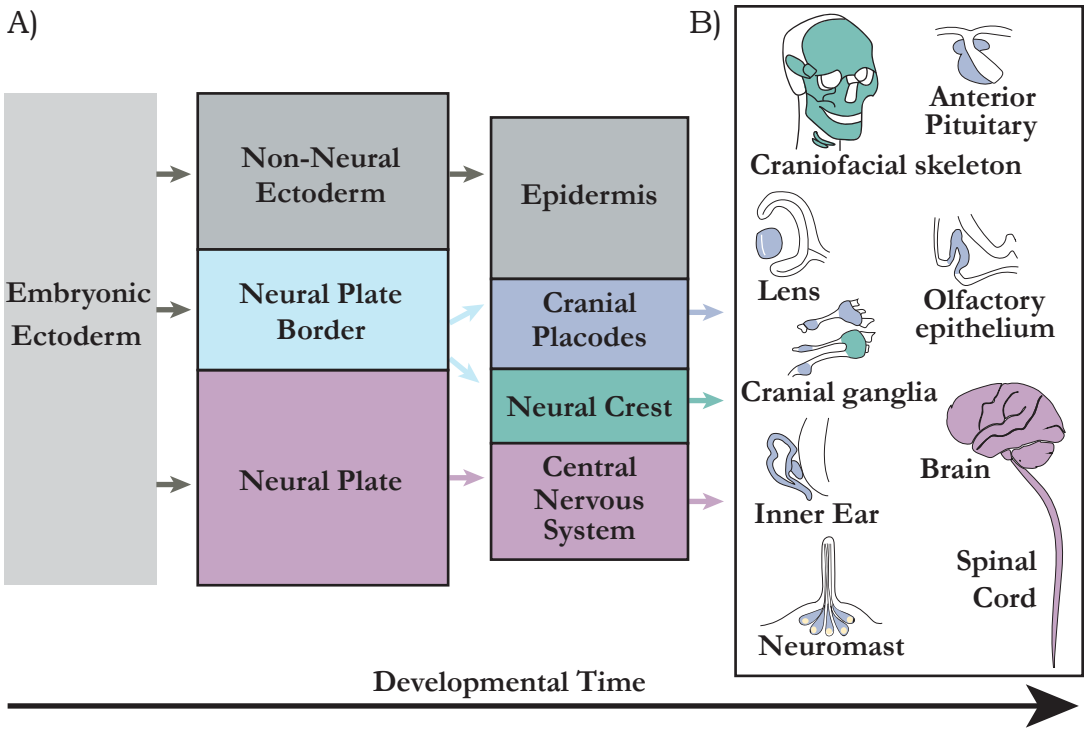


Figure 1-4. Development time for the embryonic ectoderm from which the nervous system also develops

Developmental time trajectory for the embryonic ectoderm and its derivates (A). The tissue/organ structures on the right are color coded to depict their developmental origins from the ectoderm derivates (B). (Adapted from Dubey et al., 2018)

The **hindbrain** is contiguous with the midbrain and connects to the spinal cord. It is a vital portion of the brain responsible for receiving sensory information (such as taste, touch, hearing, and balance), integration of sensory and motor signals to dictate movement (i.e. locomotion and posture), and enables continuity in timed events (e.g. breathing and swallowing). This happens through cranial nerves, reticulospinal neurons, and the pacemaker-like neuronal circuits originating from or present within the hindbrain. During development, strict boundaries of the rhombomere segments of the hindbrain are maintained. The establishment of strict rhombomere boundaries due to the Hox code, results in specific neuronal circuits within the hindbrain and the emergence of specific types of neural crest cells (which in turn give rise to specific structures of the body) from the individual segments. *Hox* gene expression in the hindbrain is fine-tuned by

two opposing signaling gradients from each end (Retinoic acid – RA and Fibroblast growth factors – FGFs), and these genes in turn are major players in defining the gene regulatory network that plays a crucial role in the patterning of neural circuits during development (Reviewed in Parker & Krumlauf, 2017).

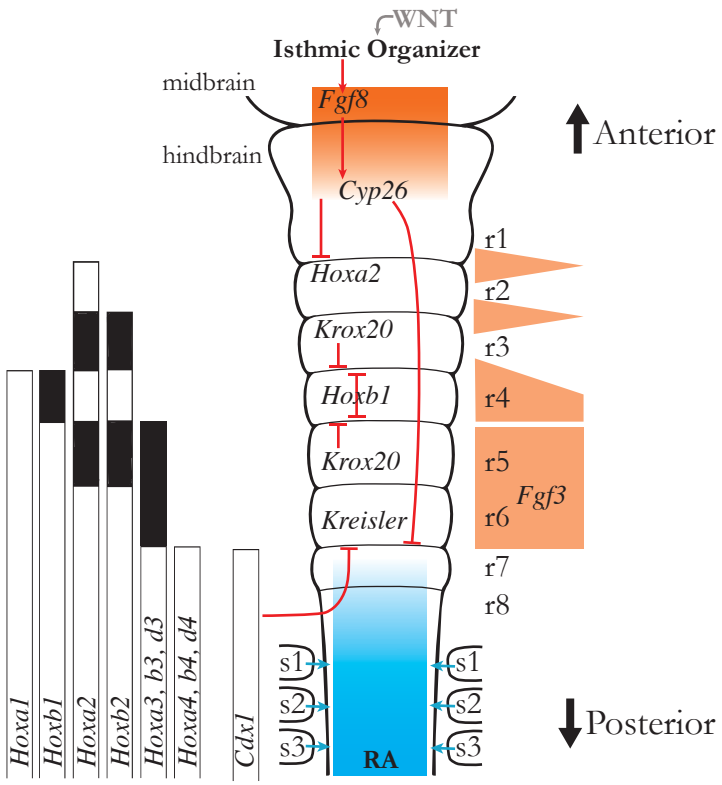


Figure 1-5. *Hox* expression domains and the opposing RA and FGF gradients that define expression boundaries

Morphogen gradients and signaling during the segmental patterning of the hindbrain. During gastrulation WNT induces formation of the Isthmic organizer which becomes the source of FGF8. There is an anterior to posterior gradient of *Fgf8* which induces *Cyp26*. This inhibits *Hox* genes from being expressed in r1. CYP26 enyzmes also metabolize RA. The metabolism of RA combined with its release from cervical somites, creates a gradient of high RA in the posterior hindbrain to low RA in the anterior hindbrain. RA sensitivity of *Hox* genes decreases going from 3' to 5' end of the cluster. Another morphogen gradient in the hindbrain is of *Fgf3* which with development gets restricted to r5 and r6. Domains of expression and signaling from other genes (*Cdx1*, *Krox20*, *Kreisler*) are also shown which directly effect to restrict expression of *Hox* genes. While KROX20 inhibits *Hoxb1*, HOXB1 inhibits *Krox20*, which results in a strict expression of *Hoxb1* in r4. (Adapted from Mahmood et al., 1996; Parker & Krumlauf, 2017).

In the hindbrain, paralogous groups 1-4 are expressed and exhibit nested domains of expression that coincide with rhombomere segments (Figure 1-5). Paralog group 1 members have anterior

boundary of expression in r3/r4, paralog group 3 members in r4/r5, and paralog group 4 members in r6/r7. There are no *Hox* genes expressed in r1, and with some exceptions most genes exhibit collinearity. The genes appear to exhibit a two-rhombomere periodicity such that while *Hoxb1* is expressed up to the r3 boundary, *Hoxb3* is expressed up to r5 boundary. Also, within each paralog group, members have different timings of expression. Cis-regulatory elements that work both locally and globally regulate *Hox* gene patterns by activating and maintaining *Hox* genes and/or by incorporating morphogen gradients into the dynamic expression profiles of *Hox* genes (Reviewed in Nolte & Krumlauf, 2006).

The **spinal cord** develops from the posterior part of the embryo (Figure 1-6). As gastrulation proceeds, the descendants of stem cells contribute to the elongation of the neural tube and mesoderm (Diez del Corral & Storey, 2004; Dubrulle & Pourquie, 2004; Mathis et al., 2001). Opposing morphogen gradients ensure that the developing spinal cord is extending at the same time as the adjacent somites (Diez del Corral & Storey, 2004). The regulation of *Hox* genes is initiated from this tail end of the developing embryo. (Deschamps & van Nes, 2005).

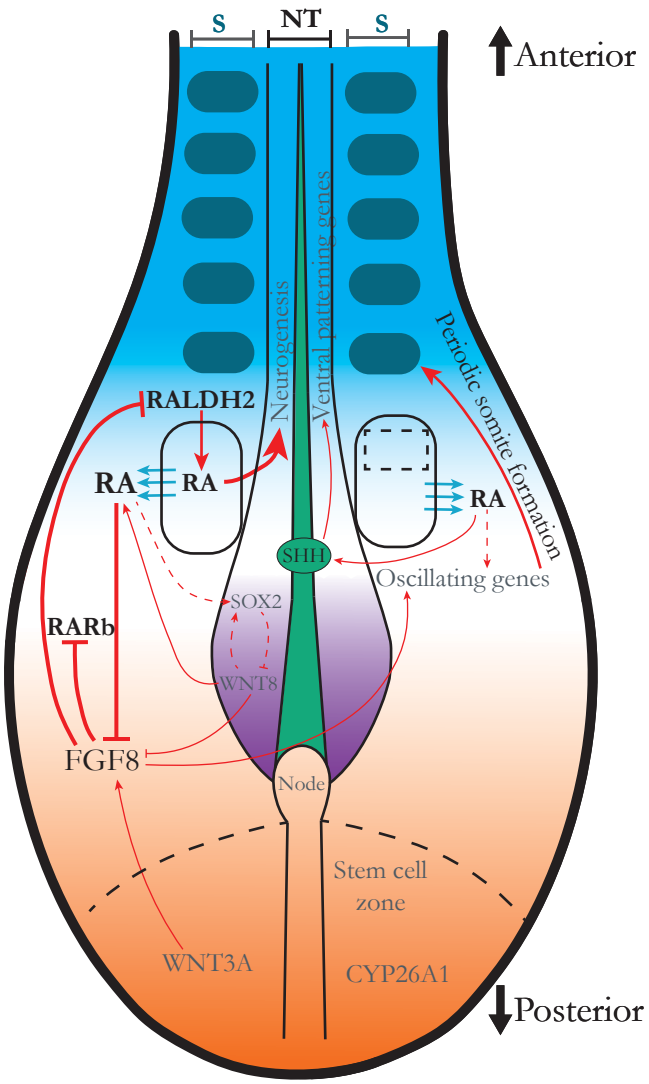


Figure 1-6. Neural tube at the posterior (tail end) of the embryo. NT- neural tube, S- somites

Schematic overview of how development proceeds from the tail (caudal end) of the embryo. Interaction of signaling molecules, such as RA and FGF which effect *Hox* gene expression are depicted. Antagonistic gradients of RA (blue) and FGF (orange) help coordinate the elongation of neural tube with synchronous deposition of adjacent somites. RA is made in the somites by RALDH2 and gets released. RA inhibits *Fgf8* and itself is repressed by FGF8. FGF is induced in cells in the stem cell zone, and its levels progressively decrease in descendants of the stem cells. FGF8 along with RA also inhibits retinoic acid receptor b (RARb). While RA promotes differentiation, FGF maintains the stem cell zone. (Adapted from Niederreither & Dolle, 2008; Patel et al., 2013; Rhinn & Dolle, 2012).

The spinal cord can be divided into segments based on the vertebra that will surround it and upon the neuronal projections that extend out in a symmetrical pattern. Going from anterior to posterior, the spinal cord is divided into cervical, brachial, thoracic, lumbar, and sacral domains. While there is no visible segmentation of the spinal cord, from chick neural tube there is some

evidence of indirect segmentation from the adjacent somites (Bronner-Fraser, 2000; Ensini et al., 1998; Lim et al., 1991; Phelan & Hollyday, 1990; Stern et al., 1991). Also, the symmetrical pattern of axonal projections from the spinal cord may also hint towards some intrinsic segmentation. Anterior *Hox* genes (paralogous groups 1-4) are expressed in the spinal cord, but at much lower levels compared to their expression in the hindbrain. The *Hox* paralog groups 5-9 have anterior domains of expression that map to the cervical region while those of paralogous group 10-13 to the lumbar region (Figure 1-3).

Along with anterior domains, many of the *Hox* genes also have posterior domains of expression. The exact patterns of expression for some members (e.g. *Hoxc4*) can be dynamic with cyclic changes in higher or lower expression taking place over time along the spinal cord (Geada et al., 1992). Some *Hox* genes (e.g. *Hoxb5*, *Hoxb6*, and *Hoxb8*) while initially have anterior borders in the spinal cord, exhibit an expansion of their expression into the hindbrain by 12.5dpc (Oosterveen et al., 2004). There is some discrepancy in the exact expression domains of the *Hox* genes in the spinal cord which can be attributed to the dynamic nature of the expression profiles and to the presence of splice variants - each of which can have different domains, as well as embryo staging across different studies. While most of these expression boundaries are based on mRNA levels, there have been some studies looking into protein levels (Liu et al., 2001). For *Hoxc* genes it has been observed that the mRNA levels don't necessarily overlap with protein levels, and this further complicates our understanding of posterior prevalence exhibited by *Hox* genes. (Reviewed in Nolte & Krumlauf, 2006).

In addition to A-P differences in the spinal cord which are visible with different vertebrae and symmetrical axonal patterns, there are also dorsal-ventral (D-V) differences which are evident from the types of neuronal projections that protrude from the dorsal or ventral side. *Hox* genes

exhibit distinct dorsal-ventral patterns of expression which dictate the development of different classes of neurons (Figure 1-7). It was observed that while initially *Hoxb* cluster genes have relatively equal dorsal-ventral expression patterns, by 11.5dpc the expression intensifies in the dorsal region of the spinal cord which correlates with formation of commissural and sensory neurons (Reviewed in Krumlauf, 2016). *Hoxb* genes therefore help to organize the dorsal sensory region in the developing spinal cord. Complementary to *Hoxb* genes are the *Hoxc* genes which exhibit stronger ventral expression profiles, and along with their A-P domains specify which motor neurons will form in the ventral region of the spinal cord (Dasen et al., 2003). The domains of *Hox* genes dictate specific neuronal formation, how exactly *Hox* genes are regulated to ensure such patterns still need to be studied.

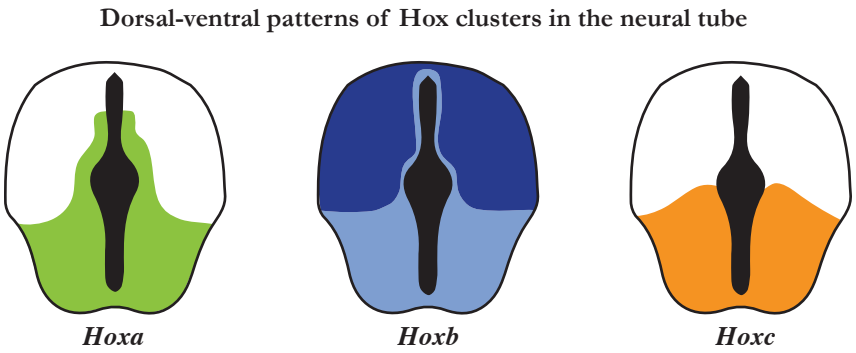


Figure 1-7. Expression domains of the Hox gene clusters in the neural tube

The expression profiles for the *Hox* clusters in the neural tube are depicted in different colors. The complementary expression profiles of *Hoxb* and *Hoxc* gene clusters dictate the different neuronal types emerging from the spinal cord; high dorsal expression of *Hoxb* corelates with spinal neurons, while ventral expression of *Hoxc* corelates with motor neurons. Reprinted with permission from Elsevier. Chapter Thirty-Four Hox Genes and the Hindbrain A Study in Segments. Curr Top Dev Biol, 116, 581-596. https://doi.org/10.1016/bs.ctdb.2015.12.011.

1.3.2 Role of *Hox* genes in the mesoderm

Cells that initially leave the primitive streak at 7.5dpc express 3' *Hox* genes and the second batch of cells leaving expresses both 3' and 5' genes. This expression of *Hox* genes is not the final expression for descendants of either the axial or paraxial cells, but they do lay the groundwork

for *Hox* gene expression for the mesoderm and derivatives (Figure 1-8). The patterns of expression are refined by regulatory inputs (morphogen gradients of RA, FGFs, WNTs, NOTCH, and YAP) during 9.5dpc and up to their established forms at 12.5dpc. The paraxial mesodermal cells receive positional information as they travel through the pre-somatic mesoderm, after they come from the stem cell zone (Nicolas et al., 1996) - a region posterior to the node (Figure 1-6). *Hox* gene expression especially in the paraxial mesoderm lays the groundwork for proper somite formation. Somite formation on either side of the spinal cord is a result of an oscillatory clock mechanism that responds to morphogen gradients (Hubaud & Pourquié, 2014; Hubaud et al., 2017; Oates et al., 2012). Proper expression of *Hox* genes in response to these morphogens ends up ensuring that the left-right symmetry as well as the correct A-P segments are formed as the embryonic axis elongates (Cordes et al., 2004; Galceran et al., 2004; Hofmann et al., 2004; Ikeya & Takada, 2001; Kessel & Gruss, 1991; Lohnes et al., 1993; Moreno & Kintner, 2004; Partanen et al., 1998; Zákány et al., 2001).

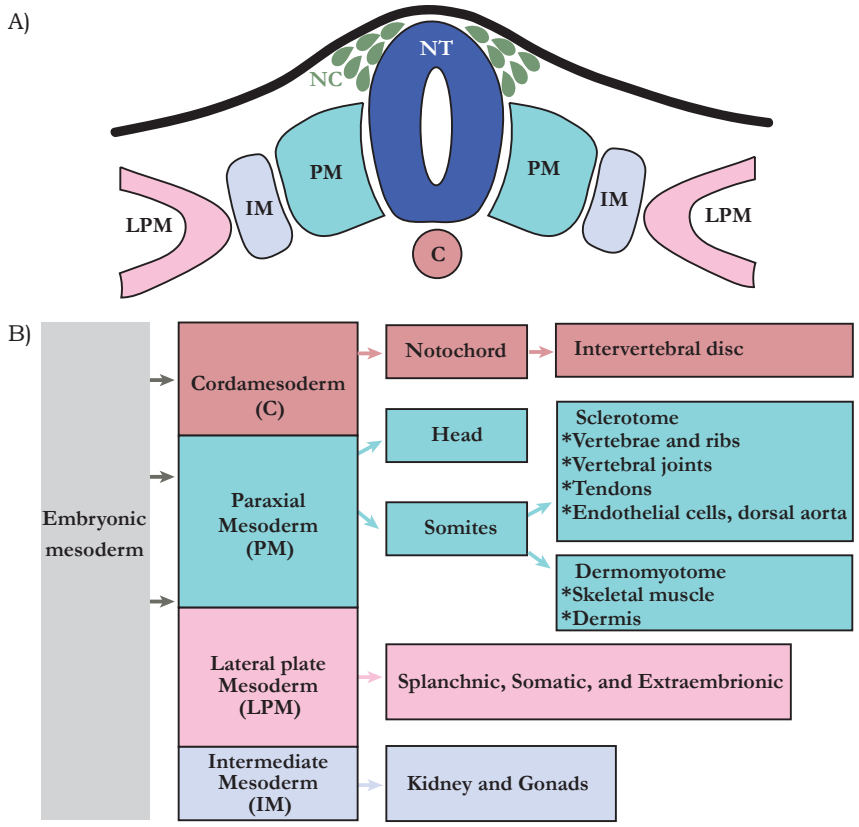


Figure 1-8. Mesoderm and its derivatives. NT-neural tube, NC- neural crest cells

Schematic of a transverse section that depicts neural tube and mesoderm derivatives (A); Cordamesoderm (C), paraxial mesoderm (PM), lateral plate mesoderm (LPM), and intermediate mesoderm (IM) are color coded to show their localization (top image) and what their derivatives will be over developmental time (B). The tissue/organ structures that are formed in the body are color coded to depict their developmental origins from the mesoderm. (Adapted from Tani et al., 2020)

Different patterns of *Hox* genes have been linked with the type of vertebra that will form. The A-P axis of *Hox* gene expressions are important because 1) seemingly similar looking vertebra such as cervical C3-C5 have a unique *Hox* code of expression and 2) the cervical C6-C7 which look different also have a distinct expression of multiple *Hox* genes (Kessel & Gruss, 1991). Somites are important developmental structures that will grow into majority of the axial skeleton. The first 4.5 somites will fuse to form the head (occipital skull). The rest of the somites interestingly will re-segment, such that posterior half of one somite will fuse to the anterior half of the next somite, and together will develop into one axial vertebra. Furthermore, the anterior limits of *Hox*

gene expression are offset in somites compared to the neural tube (Reviewed in Wellik, 2007). What role this offset expression between the adjacent structures (somites and neural tube) has in determining specifics of A-P axis still needs to be studied.

1.*3*.3 *Hox* genes and development of limbs, reproductive and other organs

Hox genes have majorly been shown to be involved in patterning of the nervous system and the axial skeleton, but they are also important in development of the reproductive organs, limb formation, and in the proper formation of various other tissues and organs throughout the developing embryo. *Hox* genes play distinct roles in the formation of limbs. Evolutionary analysis by comparisons with ancient bony fish have led to the hypotheses of 1) fin to limb transitions and 2) the positioning of the vertebrate limbs (Nakamura et al., 2016; Shubin et al., 1997, 2009). In the vertebrate developing forelimb and hindlimb buds, paralogous groups 9-13 of *Hoxa* and *Hoxd* clusters in particular exhibit temporal and spatial collinearity (Dollé & Duboule, 1989; Haack & Gruss, 1993; Yokouchi et al., 1991). At 10.5-11.5dpc, *Hoxa* cluster genes display expression boundaries perpendicular to the body axis, and genes in the *Hoxd* cluster display expression boundaries that are tilted towards the posterior side of the limb bud. While *Hoxc* cluster genes are also expressed in the limb bud, they do not appear to exhibit any collinearity (Nelson et al., 1996; Peterson et al., 1994). Early experiments on limb development were done in chick embryos, and it was observed that grafting of the zone of polarizing activity, or a bead implantation of RA, or the ectopic expression of SHH, all lead to a mirror image duplication of the digits in the limbs. It was observed that this mirror image result was due to an ectopic activation of *Hoxd* cluster genes (Izpisua-Belmonte et al., 1991; Nohno et al., 1991).

Studies in chick lead to models being proposed where specific *Hox* gene paralogs had to be expressed in certain domains for normal forelimb and hindlimb formations, but deletion of *Hox*

genes in mouse did not yield equivalent digit homeotic transformations (Dollé et al., 1993). In mice, the *Hoxd13* genes effected cartilage growth and delayed ossification after birth. *Hoxa9* and *Hoxd9* effected humerus but had no effect on femur. *Hoxa11*, *Hoxd11*, and *Hoxd12* effected only forelimb while *Hoxa10* solely effected the femur (Reviewed in Favier & Dolle, 1997). These phenotypes showed redundancy in the system and compensatory mechanisms by which *Hox* genes differentially effected development of the forelimbs and hindlimbs. This redundancy of function in limb development was not only among paralogous group members but also observed for members of the non-paralogous groups 9-13 (Davis et al., 1995; Favier et al., 1996).

Posterior *Hox* genes have also been observed to be expressed in collinear patterns in the reproductive organs of both males and females (Dollé, Izpisùa-Belmonte, et al., 1991). The *Hoxa10* mutants studied were viable but hypofertile (Rijli et al., 1995; Satokata et al., 1995). In these mutant males, the testes had different degrees of cryptorchidism (abnormal descent of the testes), and histologically the males had defects in spermiogenesis. Further, both *Hoxa10* and *Hoxa11* male mutants had defects where morphology of the ductus deferens (also called vas deferens and transports mature sperm into urethra) appeared like the epididymis (a very convoluted duct behind the testis) (Benson et al., 1996; Hsieh-Li et al., 1995). Similar anterior transformations of reproductive organs were also seen in *Hoxa10* mutant females. In these females, the proximal part of the uterus appeared to transform into a structure similar to the oviduct (Benson et al., 1996). In terms of fertility, these mutant females were either sterile or had very small litters. Studies showed that embryos failed to implant or got reabsorbed (Satokata et al., 1995). Misexpression of *Hox* genes can also result in posterior transformations. When *Hoxa11* gene was replaced with *Hoxa13* gene, the uterus transformed to appear like the posterior cervix and vagina (Benson et al., 1996; Satokata et al., 1995; Warot et al., 1997).

The female reproductive organs (oviduct, uterus, cervix, and upper vagina) develop from the Mullerian duct; ovaries develop majorly from the genital ridge (somatic precursor) and partially from the mesonephros. *Hoxa9-Hoxa13* are expressed in the Mullerian duct in some redundant overlapping patterns (Goodman, 2002; Taylor, 2000; Taylor et al., 1997). *Hox* expression in the reproductive tract is not restricted to development but is also present in the adult tissue. *Hoxa10* is expressed in the luminal and glandular epithelium of the uterus at 0.5 day of gestation, and in 3.5 to 5.5 days is expressed in the stroma and developing decidua. It has been postulated that the lack of these expression profiles in the *Hoxa10* mutants could be contributing to the infertility of the female mutants (Satokata et al., 1995). *Hoxa10* and *Hoxa11* are also expressed in the proliferative and secretory phase of the uterus (Taylor et al., 1998; Taylor et al., 1997). It is now believed that *Hox* genes in adults play a role in the differentiation and cellular proliferation of the uterus during the menstrual cycle and pregnancy (Reviewed in Du & Taylor, 2015). During pregnancy (and even in mammary development after birth), paralogous group 9 genes have been shown to be expressed in the mammary glands, coinciding with their development. Further proof for role of *Hox* genes became evident when hypoplasia of mammary glands was observed after pregnancy in the *Hox* mutants (Chen & Capecchi, 1999).

In addition, *Hox* genes also exhibit spatial temporal collinearity in expression in the skin during fetal development. Specifically, *Hoxb* cluster genes have been shown to be involved in fetal as well as adult skin in mice (Bieberich et al., 1991; Detmer et al., 1993; Mathews et al., 1993; Rieger et al., 1994). Further studies have also implicated *HOX* genes playing a role in human skin (Stelnicki et al., 1998). In the skin, hair follicles which undergo phases of regression and regeneration during the whole lifespan have also been studied. It has been shown that *Hoxc13* plays a role in hair follicle development. Studies have shown that *Hoxc13* target genes such as

hair-specific keratins and keratin-associated proteins are important for hair growth (Reviewed in Awgulewitsch, 2003). In mutant *Hoxc13* animals, which exhibit alopecia, it was identified that the problem was not in the morphogenesis but in the differentiation of hair follicles (Godwin & Capecchi, 1998).

Hox genes also play a role in development of kidneys (Davis et al., 1995). They observed that while single *Hox* gene mutants didn't have any visible abnormalities, compound *Hoxa11* and *Hoxd11* mutants, in addition to skeletal defects, had severe hypoplasia of the kidneys. *Hox* genes are also functionally important in the developing gut (Dollé, Izpisúa-Belmonte, et al., 1991; Kondo et al., 1996). In the developing gut, the colinear expression of *Hoxa* and *Hoxb* cluster genes starts from the duodenum and for *Hoxc* genes from the jejunum. *Hoxd* genes are expressed in the posterior portions of the midgut, in the ilium, and cecum. In the developing gut therefore, it appears like the different clusters exhibit a different code in the different regions (foregut, midgut, and hindgut). *Hox* genes in the gut play a role in the mesenchymal specification initially, and then also in the regional identity through differentiation as a result of epithelial-mesenchymal interactions (Kawazoe et al., 2002). *Hox* genes exhibit collinearity in the developing adrenal gland, such that 3' *Hox* genes are expressed in the outer gland and the 5' *Hox* genes are expressed in the inner region of the gland (Neville et al., 2002). A study has also looked at *HOX* expression in adult tissues and identified that many *HOX* genes are differentially expressed in human adult tissues (Yamamoto et al., 2003). Furthermore, in the adult skeleton, *Hox* genes are highly expressed in the mesenchymal stem cells and hence involved in the healing process during fractures (Rux & Wellik, 2017). These are not the only stem cells *Hox* genes are expressed in. The *Hoxb* cluster genes have been shown to play a role in maintaining normal hematopoietic stem cells and to inhibit leukemogenesis (Qian et al., 2018). Thus, *Hox* genes are

fundamentally important especially during development but also in adults for proper functioning of various tissues and organs.

1.3.4 Loss-of-function and gain-of-function studies for *Hox* genes

The spatial expression patterns of *Hox* genes ensure proper segmental identity and formation of body axis and tissues during development. Similarly, the temporal activation of *Hox* genes is functionally important because experimental conditions resulting in premature or delayed *Hox* gene activation have been shown to have a phenotypic effect, which is true even in instances where the final patterns of *Hox* genes are restored (Gerard et al., 1997; Juan & Ruddle, 2003; Kondo & Duboule, 1999; Zakany et al., 1997). It has been observed that sometimes deletion of the whole cluster has less of an effect than deletion of multiple paralogous group members, even though all members of paralogous groups may or may not have the same function or expression domains (Reviewed in Maconochie et al., 1996). Animal models have been generated where individual or complete cluster deletions have been made. Complete deletion of the *Hoxa* cluster is embryonically lethal after 11-12.5dpc because of placental insufficiency and premature stenosis of the umbilical artery (Scotti & Kmita, 2012; Shaut et al., 2008; Warot et al., 1997). Complete deletion of the *Hoxb* cluster results in circulation defects with mice developing severe edemas and hemorrhages in internal jugular veins by 13.5-15.5dpc (Medina-Martinez et al., 2000). Compound deletion of both *Hoxa* and *Hoxb* cluster is lethal by 9.5dpc which has been attributed to defects in development of the heart (Nolte et al., 2013; Soshnikova et al., 2013). Deletion of the *Hoxc* cluster mice live to birth without any extreme skeletal or internal organ defects. Death after birth has been attributed to some defects in the respiratory system (Suemori & Noguchi, 2000). *Hoxd* cluster genes have been shown to express in two waves during

development and specific cluster deletions have been linked to limb deformations (Zakany et al., 2004).

While instances of compensation among *Hox* genes have sometimes resulted in no visible phenotype being observed, several mutations in groups 3-11 *of Hox* genes have been implicated in specific phenotypic defects especially in the axial skeleton. Also, as described before *Hox* genes especially play an important role in the development of the nervous system. Further details of phenotype defects for *Hox* gene cluster are reviewed in detail by other groups (Maconochie et al., 1996; Philippidou & Dasen, 2013; Quinonez & Innis, 2014). Compilation of defects in neural and skeletal with loss-of-function and gain-of-function studies, and with a particular focus on *Hoxb* genes, as it pertains to our study, is shown in Table1-1.

Table 1-1 *Hox* gene mutations with focus on *Hoxb* cluster genes

Mutant	**Phenotype**	**References**
Hoxb1 loss-of-function	r4 homeotically transformed to r2/r3, motor neurons VII acquire identity of V, loss of motor nucleus and nerve as consequence of problems in motor neurons VII nucleus migration and axonal pathfinding, contralateral vestibuloacoustic afferents specified incorrectly, serotonergic neurons ectopically generated and visceral sensory neurons in r4 lost, auditory circuits formed incorrectly, lateral vestibulospinal tract projections lost	(Chen et al., 2012; Di Bonito et al., 2013; Gaufo et al., 2004; Gavalas et al., 2003; Goddard et al., 1996; Pattyn et al., 2003; Studer et al., 1996)
Hoxb1 gain-of-function	overexpression in r2 causes motor axon projections in r2 to start resembling those in r4	(Bell et al., 1999)

Hoxa1/*Hoxb1*	absence of r4 and r5, cranial nerves VII to XI have patterning defects, lack of the 2nd branchial arch in *Hoxa1*/*Hoxb1* double mutants and defects in the formation of the 2nd and 3rd branchial pouches, neural crest cells not able to migrate properly	(Gavalas et al., 1998; Rossel & Capecchi, 1999; Studer et al., 1998)
Hoxb2	reduction in VII nucleus along with slight transformation to identity of V, serotonergic neuron ectopically generated in r4, oligodendrocyte production reduced in r4, auditory circuit formed incorrectly, sternum malformations and C2 anteriorly transformed to C1	(Barrow & Capecchi, 1996; Davenne et al., 1999; Di Bonito et al., 2013; Gavalas et al., 2003; Miguez et al., 2012; Pattyn et al., 2003; Sham et al., 1993)
Hoxa2/*Hoxb2*	inter-rhombomeric boundaries from r1 to r4 lost, loss of *Evx1*+ interneurons in both r2 and r3, dorsoventral neurons specified incorrectly	(Davenne et al., 1999)
Hoxa3/*Hoxb3*	lack of VI nucleus, visceral sensory neurons in r5 missing, slightly abnormal formations of C1 and C2	(Gaufo et al., 2003; Gaufo et al., 2004; Manley & Capecchi, 1997)

Mutant	Phenotype	References
Hoxb4	ventral body wall formation abnormal, C2 slightly anteriorly transformed to C1 (ventral tubercle incorrectly located, neural arch wider, arch atlas anteriorized), sternum is split, anemia as result of reduced spleen and bone marrow cellularity	(Brun et al., 2004; Manley et al., 2001; Ramirez-Solis et al., 1993)
Hoxb2/*Hoxb4*	abnormally formed body wall and sternum	(Barrow & Capecchi, 1996; Manley et al., 2001; Ramirez-Solis et al., 1993)
Hoxb5	forelimbs with respect to axial skeleton are anteriorized which causes shoulder girdle to move rostrally, ectopic brachial plexus, C6 and C7 formed incorrectly, some lung abnormalities	(Boucherat et al., 2013; D.E. Rancourt et al., 1995)
Hoxb5 (neural crest cell specific deletion)	apoptosis of neural crest cells, sympathetic and dorsal root ganglia defects, reduced pigmentation, abnormalities in the enteric nervous system	(Kam et al., 2014)Kam et al 2014.
Hoxb6	Loss of first rib, second rib bifid, defects in somitogenesis and abnormal anterior-posterior somite patterning, intercostal nerve innervation defects	(Casaca et al., 2016; D.E. Rancourt et al., 1995)
Hoxb5/*Hoxb6*	cervicothoracic vertebrae from C6 to T1 homeotically transformed to anterior identities	(D. E. Rancourt et al., 1995)
Hoxb7	twelve percent of mutants have defects in first and second rib, sternum segmented incorrectly	(Chen et al., 1998)
Hoxa7/*Hoxb7*	greater penetrance and higher expressivity in abnormalities of first and second rib	(Chen et al., 1998)

Mutant	Phenotype	References
Hoxb8	second spinal ganglion deteriorated, defects in neural distribution in the dorsal horn, mutants either normal or exhibit first and second rib fusion, abnormal sternum and T2 spinous process, T3 spinous process and C7 vertebrae anteriorly transformed, morphological defects in C2 spinal ganglion, compulsively excessive grooming	(Greer & Capecchi, 2002; Holstege et al., 2008; van den Akker et al., 2001; van den Akker et al., 1999)
Hoxb9	first and second ribs fusion at the sternum attachment point, defects in segmentation of sternum, extra rib attached to sternum	(Chen & Capecchi, 1997)
Hoxa9/*Hoxb9*/*Hoxd9*	first, second, and third ribs fused (100% penetrance of mutants), abnormal humerus, radius, and ulna, mammary glands exhibit hypoplasia during pregnancy and post parturition	(Chen & Capecchi, 1997, 1999)
Hoxb13	spinal cord extended at the tail side, increased number of Doral root ganglions, abnormal innervation of tail sensory neurons	(Economides et al., 2003)

1.3.5 Human mutations and diseases linked to *Hox* genes

Hox genes are evolutionary conserved, and hence it is not surprising that they also play a significant role in humans, where there are also a total 39 *HOX* genes in 4 clusters which exhibit spatio-temporal collinearity. There are 10 different mutations in *HOX* genes that have been linked to developmental diseases in humans (Alasti et al., 2008; Bosley et al., 2008; Garcia-Barceló et al., 2008; Lin et al., 2012; Sayli et al., 1995; Shrimpton et al., 2004; Thompson & Nguyen, 2000; Tischfield et al., 2005) (Also see reviewed in Goodman, 2002; Quinonez & Innis, 2014). For example, *HOXB1* mutations have been linked to congenital facial palsy, hearing loss, strabismus, midface retrusion and an upturned nose (Webb et al., 2012). In addition to developmental defects, *HOX* genes in humans have also been linked to metabolic diseases

(Reviewed in Procino & Cillo, 2013) and different cancers, where instances of their abnormal expression has been a predictor of poor prognosis (Sun et al., 2018). For example, *HOXB13* has been linked to prostate and colorectal cancers (Akbari et al., 2013; Ewing et al., 2012) with conflicting reports about its link to breast cancer (Akbari et al., 2012; Alanee et al., 2012). Further, the hexapeptide motif in *HOXB9* has been shown to be involved in gastric cancer (Chang et al., 2015). Also, missense mutations in *HOXD4* were identified in children with acute lymphoid malignancy and were accompanied either with or without skeletal anomalies (van Scherpenzeel Thim et al., 2005). In certain leukemias, *HOX* genes have been identified to be fused with NUP98 (a nucleoporin gene) and have disrupted transcriptional activity (Reviewed in Krumlauf & Ahn, 2013). Therefore, with respect to cancers, it is not clear whether the increased expression of *HOX* genes or just their altered functional activity is playing into malignancy or poorer prognosis (Reviewed in Luo et al., 2019).

1.4 Regulatory inputs for *Hox* genes

Hox genes play a critical role in development as well as in adult tissues. They have also been identified to play roles in etiology of diseases. Therefore, many in the field have tried to understand how exactly *Hox* genes are able to exhibit such precise spatial and temporal expression domains, and what exactly is regulating their expression patterns. Research into both the transcriptional regulation of *Hox* transcripts as well as translational control of HOX proteins has been studied (Reviewed in Brend et al., 2003; Kondrashov et al., 2011; Nolte et al., 2019; Parker & Krumlauf, 2017; Rezsohazy et al., 2015).

1.4.1 Morphogen gradients

Over the years, some regulatory inputs that dictate or modulate the expression domains, especially in the neuroectoderm and mesoderm, have been identified. These include major

signaling pathways such as Retinoic acid (RA), fibroblast growth factor (FGF), and wingless-related integration site (WNT) (Figures 1-5 and 1-6). Collinearity of *Hox* genes has also been observed in response to these signaling molecules in cell culture experiments (Diez del Corral et al., 2003; Papalopulu et al., 1991; Simeone et al., 1990; Simeone et al., 1991). In an embryo, the final expression domains and collinearity along the A-P axis exhibited by the *Hox* genes are related to their ability to respond to the dynamic and opposite morphogen gradients (Mallo & Alonso, 2013; Nolte et al., 2019; Parker & Krumlauf, 2017).

During gastrulation and early somitogenesis, crosstalk between signaling factors and other transcription factors (e.g., OTX2, GBX1/2, CDX1, and CYP26) help to define boundaries for expression of *Hox* genes. This crosstalk enables segmentation borders for the midbrain, hindbrain, and spinal cord. Initially, the spatial distinction between anterior vs posterior regions occurs in response to opposite RA and FGF/WNT gradients. **WNT** signaling in the primitive streak modulates the expansion and the anterior spreading of the *Hox* genes during gastrulation (Ciruna & Rossant, 2001; Forlani et al., 2003). *Wnt3a* expression gets restricted to the posterior epiblast during early gastrulation, where along with FGF it acts as a key player in ensuring segmental fate to paraxial mesoderm. Mutants for *Wnt3a* and *fgfr1* cause anterior transformation of vertebra along the A-P axis by posterior restriction of *Hox* gene expression domains (Ikeya & Takada, 2001; Partanen et al., 1998). Also, during gastrulation WNT triggers the secondary signaling center – isthmic organizer (IO) in the midbrain. The IO defines the *Otx2* and *Gbx1/2* expression that sets the anterior limit for the hindbrain at the midbrain-hindbrain boundary (Kiecker & Lumsden, 2012; Liu et al., 1999; Millet et al., 1999; Rhinn et al., 2005; Wurst & Bally-Cuif, 2001). In the posterior hindbrain, *Cdx* genes are expressed, and they specify spinal cord fate by promoting posterior embryonic development (Deschamps & van Nes, 2005; Metzis

et al., 2018; Young et al., 2009). Mutations in the *Wnt* or *Notch* signaling pathway effect *Hox* gene expression. It has been shown that WNT which controls pre-somatic mesoderm (PSM) maturation directly cross regulates the dynamic expression of segmentation genes in the *Notch* pathway. Hence, mutations in the *Notch* pathway during somitogenesis also effect the expression of *Hox* genes in the PSM which in turn effects the vertebral patterning (Cordes et al., 2004; Galceran et al., 2004; Hofmann et al., 2004; Zákány et al., 2001).

During late gastrulation, ***Fgf*** gene transcription is restricted to the stem cell region of the node. As cells divide the levels of *fgf8* transcript and FGF8 protein levels in the descendants decrease, forming a caudorostral gradient - high in posterior tail and decreasing anteriorly towards head (Dubrulle & Pourquie, 2004). High levels of FGF along with CYP26 inhibits RA, which helps to maintain stemness in the posterior region of the embryo. FGF8 inhibits RALDH2 enzyme (which synthesizes RA in the somites), and CYP26 clears RA by metabolizing it. In the anterior head region of the embryo, IO is the source of FGF signaling. FGF in the hindbrain is essential for segmentation to form rhombomeres. Anteriorly, high levels of FGF along with CYP26 inhibit RA that is diffusing into the hindbrain from the cervical somites. This helps to create a gradient of RA, and this inhibits expression of *Hox* genes in anterior neural regions and restricts *Hoxa2* expression to the r1/r2 boundary (Parker & Krumlauf, 2017). Another source of FGFs in the hindbrain is *Fgf3*. In mouse embryos *Fgf3* has dynamic expression patterns which at 8.5dpc is high, in the region that will become r4/r5, and decreases by 9.5dpc. In 9.5dpc, *fgf3* is expressed at slightly lower levels in r5 and r6, with even lower levels in the r4 and at the boundaries of r1/r2 and r2/r3. *Fgf3* is involved in otic development and expressed in the second brachial arch which helps to form several muscles and skeletal bones of the face (Mahmood et al., 1996).

Critical to the expression of *Hox* genes along the A-P axis of the developing embryo is **RA**. In the posterior part of early post-implantation embryos, RA was detected (Hogan et al., 1992), but it has been hard to ascertain a specific role of RA in regulating 3' *Hox* genes in this early developmental stage (Niederreither et al., 1999). Beginning at late gastrulation, RA is made in the pre-somatic mesoderm and the somites (Begemann et al., 2001; Molotkova et al., 2005). RA is a product of retinol (vitamin A) which gets processed by RDH and RALDH2 enzymes. While RA has not been linked to the initial activation of *Hox* genes in the primitive streak, it does play a role in regulating mesodermal segmentation (Moreno & Kintner, 2004). As RA gets synthesized, it diffuses into the neuroepithelium of the hindbrain. RA regulates its own expression by upregulating the expression of CYP26 enzymes which clears RA anteriorly (Gould et al., 1998; Hernandez et al., 2007; Sirbu et al., 2005; White et al., 2007; White & Schilling, 2008). While RA is actively synthesized in newly formed somites, as the embryo elongates and anterior somites begin to differentiate, the expression of *Raldh2* is downregulated. In the hindbrain, all this results in a gradient of high RA posteriorly to low RA anteriorly (Deschamps & van Nes, 2005; Gavalas & Krumlauf, 2000; Schilling et al., 2016; Shimozono et al., 2013). While the gradient of RA decreases in the hindbrain, the sensitivity of *Hox* genes to RA increases; the 3' *Hox* genes are more sensitive to RA than the 5' *Hox* genes. This creates unique *Hox* expressions and gives individual rhombomeres their segmental identity (Dupe & Lumsden, 2001; Gavalas & Krumlauf, 2000; Gould et al., 1998). The RA gradient is interpreted by *Hox* and target genes due to the presence of Retinoic Acid Response elements (RAREs). They are direct repeats on the genome to which Retinoic Acid Receptors (RARs/RXRs) can bind to modulate transcription of genes (Chen & Evans, 1995; Glass & Rosenfeld, 2000; Horlein et al., 1995; Lavinsky et al., 1998; Loudig et al., 2005; Nolte et al., 2019; Tumpel et al., 2009). RA and

FGF have antagonistic effect on cells. In the young developing spinal cord, while RA from the mesoderm stimulates maturation and differentiation of cells, the FGF from ectoderm and mesoderm in the node maintains stemness by preventing differentiation (Reviewed in Deschamps & van Nes, 2005). Together the antagonistic signaling from RA and FGF ensure that, as the embryonic axis grows, the neural maturation occurs together with generation of new somites (Diez del Corral & Storey, 2004).

1.4.2 Non-coding RNAs

Lewis in 1978 had postulated that many cis-regulatory regions in the *Drosophila* bithorax complex were regulatory RNAs, and since then his claim has turned out to be true. Within the four mammalian *Hox* clusters, as mentioned before, there exist several non-coding RNAs. These include microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). Based on several studies, these non-coding RNAs may act as another layer of regulation to control the expression of *Hox* genes at either the transcriptional or translational level (Leppek et al., 2020; Parker & Krumlauf, 2017; Rezsohazy et al., 2015).

MiRNAs are short non-coding RNAs approximately 22 nucleotides in length. Mature miRNAs are made from longer pri-RNA that gets transcribed by RNA polymerase II. Pri-miRNA is cleaved in the nucleus by DGCR8 and DROSHA to make pre-miRNA- a stem-loop structured RNA. This cleaved form (pre-miRNA) gets exported into the cytoplasm, where DICER cuts the stem-loop to make a linearized double stranded RNA (Winter et al., 2009). One of the strands gets loaded onto Argonaute proteins to form the RNA inducing silencing complex (RISC) which can target multiple messenger RNAs (mRNAs). miRNAs target genes mostly by binding to the 3’ untranslated region (UTR). Targeting takes place by pairing of the miRNAs seed sequences (2-7 nucleotides) to the target mRNAs of genes. Since miRNAs can bind even if all base pairs

don’t match and in mRNAs there are plenty of 6-8 mers, a miRNA can target and regulate multiple mRNAs (Agarwal et al., 2015). From the first discovery of miRNAs in c-elegans (Lee et al., 1993), multiple evolutionary conserved and some species-specific miRNAs have been identified. In mammals, miRNAs are involved in development where they can coordinate developmental timing, regulate cell fate, maintain homeostasis, and can have general or tissue specific roles (Reviewed in DeVeale et al., 2021). miRNAs have also been implicated in several diseases e.g., in different cancers (Reviewed in Bhaskaran & Mohan, 2014; Nguyen et al., 2020; Qin et al., 2019; Shea et al., 2016; Srivastava et al., 2015). In the mammalian *Hox* clusters there are present multiple miRNAs; miRNAs conserved between mouse and humans *Hox* clusters include miR196a/b, miR10a/b, and miR615. The miR196 members are present within the *Hoxa*, *Hoxb*, and *Hoxc* clusters and they can regulate the expression of *Hox* genes both in cis (effecting same cluster from which it is transcribed) and trans (effecting a cluster other than the one from which it is transcribed). The miR10 members have also been implicated to play roles in cis and trans to effect *Hox* genes directly or through feedforward regulatory loops by effecting signaling pathways such as WNTs, FGFs, and NOTCH. The expression of these miR10 members has been observed to correlate with both the expression and response to RA of adjacent *Hox* genes (Reviewed in De Kumar & Krumlauf, 2016).

LncRNAs have been shown to be involved in control of diverse biological functions in development and disease (Briggs et al., 2015; Du et al., 2016; Goff et al., 2015; Nguyen & Carninci, 2016; Spadaro et al., 2015; Younger & Rinn, 2014). LncRNAs are generally characterized as non-coding RNAs greater than 200 nucleotides, having a 5’ Cap, may or not be polyadenylated, and lacking an open reading frame. In the genome, lncRNAs can be transcribed in sense, anti-sense, from intronic regions, or from overlapping regions with respect to genes

(Kashi et al., 2016; Mattick & Rinn, 2015; Melé et al., 2017; Schein et al., 2016; Watanabe et al., 2015). They can be transcribed by both RNA polymerase II (Pol II) and RNA polymerase III (Pol III) and may also contain multiple splice variants. LncRNAs can function in both cis and trans to regulate mRNAs and miRNAs. While lncRNAs may share some transcriptional properties with mRNAs, they are distinguishable based on their chromatin environments and transcriptional regulation (Reviewed in Melé et al., 2017). Often times lncRNAs are not evolutionarily conserved, but some lncRNAs that are conserved among mammals and some even outside mammals have been identified (Hezroni et al., 2015; Sarropoulos et al., 2019). While many lncRNAs are transcribed, their relative levels of expression are often low, and they can be difficult to study. Hence, there is uncertainty about whether all lncRNAs are functional on their own, or if they are just transcriptional noise and are transcribed just to enable processing of adjacent genes by being run-through transcripts. Nonetheless, there has been some evidence of lncRNAs functioning in the nuclei as well as in the cytoplasm to regulate genes (Carlevaro-Fita et al., 2016; Melé et al., 2017; Zuckerman et al., 2020). LncRNAs can regulate chromatin either through binding chromatin modifying proteins (Chu et al., 2011; Isoda et al., 2017) or by directly effecting DNA through triple helices or R-loops to silence (Schmitz et al., 2010) or activate (Grote et al., 2013; Mondal et al., 2015) the genome. In the nucleus, lncRNAs have been linked to the assembly and function of nuclear condensates (Clemson et al., 2009; West et al., 2016). In the cytoplasm, lncRNAs have been identified to be localized to specific organelles such as exosomes and mitochondria (Fatima & Nawaz, 2017; Zhao et al., 2018). Particularly mitochondrial lncRNAs appear to be transcribed either from nuclear DNA or mitochondrial DNA to play a role in mitochondrial functions and enable crosstalk between mitochondria and nuclei (Reviewed in Statello et al., 2021).

In mammals, about 40% of the lncRNAs are expressed in the mammalian brain (Briggs et al., 2015). In the central nervous system, lncRNAs have been implicated in differentiation and regeneration after injury of neurons (He et al., 2017; Perry et al., 2018; Sauvageau et al., 2013). *Hox* genes, that also are critical for segmental identify of neurons along the A-P axis, have embedded within and adjacent to their clusters several lncRNAs. In the *Hox* clusters, the lncRNAs transcribed may act in cis to regulate *Hox* genes or trans to regulate genes outside the cluster or even non-*Hox* genes (Reviewed in De Kumar & Krumlauf, 2016). Upstream to *Hoxa1* is an ~16kb region that transcribes several isoforms of *Heater* lncRNA. *Heater* may be linked to *Hoxa1*'s response to retinoids (De Kumar et al., 2015). Deletion of some *Heater* isoforms has resulted in increased levels of *Hoxa1* in uninduced ES cells (Maamar et al., 2013). Other examples of lncRNAs, within the *Hoxa* cluster, that have been identified to have functional roles include *Hottairm1*, *HoxA2-AS2*, *HoxA11-AS*, *Hottip* (Hsieh-Li et al., 1995; Kim, Song, et al., 2013; Zhang et al., 2014; Zhao et al., 2013). The *Hoxb* cluster contains several lncRNAs including *Hobbit* and *HoxBlinc*. *HoxBlinc* has been implicated to have a role in hematopoiesis by mediating *Hoxb* genes by working through the Notch signaling pathway (Deng et al., 2016). *Hobbit* is unique in that it gets induced by RA similarly to adjacent *Hoxb* genes in ES cells, and in embryos its expression seems to depend on shared RAREs (Ahn et al., 2014; De Kumar et al., 2015). Other lncRNAs identified with functional roles include *Hotair* (Dasen, 2013; Rinn et al., 2007; Schorderet & Duboule, 2011), *Hotdog* and *Tog* (Delpretti et al., 2013) from the *Hoxd* cluster. LncRNA within the *Hox* clusters appear to be more conserved and also be affected by similar cis-regulatory elements as *Hox* genes. Thus, it is important to understand the regulatory interplay between lncRNAs and enhancer elements, in order to better understand how *Hox* gene expression may be fine-tuned or maintained.

1.4.3 Cis-regulatory elements

Experimental evidence has shown that *Hox* genes can functionally compensate for each other. Swapping and replacement of homeodomains have also put into question the degree of specificity within HOX proteins. These experiments point towards the need to better understand, in a context dependent manner, how *Ho*x genes are regulated and what the total HOX protein levels are that ensure segmental identity. Cis-regulatory elements interspersed within and around the clusters have emerged as important players that regulate transcription of *Hox* genes. They modulate *Hox* gene expression by directly incorporating the opposite morphogen gradients along the body axis (Krumlauf, 2018). Specifically, to regulate nested *Hox* expressions in the hindbrain, several cis-regulatory elements or Retinoic acid response elements (RAREs), have been identified that enable *Hox* genes to respond to RA (Reviewed in Alexander et al., 2009; Parker & Krumlauf, 2017; Tumpel et al., 2009). In an embryo, in response to RA, the paralogous group 1 genes are the first to get activated (Marshall et al., 1992; Marshall et al., 1994; Murphy & Hill, 1991; Wilkinson et al., 1987).

During development, *Hoxb1* gets restricted to r4 in the hindbrain. The initial activation of *Hoxb1* occurs in response to RA through 3' RAREs (Dupe et al., 1997; Marshall et al., 1994; Studer et al., 1998). *Hoxb1* restriction occurs through repression in adjacent rhombomeres, r3 and r5, though an element that is responsive to both RA and KROX20 (Studer et al., 1994). In r6, *Hoxb1* is repressed by posterior genes, *Hoxa3* and *Hoxb3* (Gaufo et al., 2003). In addition to activation and restriction of *Hoxb1*, the expression in r4 is maintained through a 5' *Hox* response element which incorporates *Hoxb1*'s autoregulation and cross-regulation by *Hoxb2* and *Hoxa1* (Davenne et al., 1999; Pattyn et al., 2003; Popperl et al., 1995; Studer et al., 1998). How RAREs enable response to RA is through binding of RAR/RXR receptor and/or through change in binding

partners of retinoid receptors to RA (Reviewed in Cunningham & Duester, 2015). There are two different ways through which RA induces transcription, examples of which are seen with *Hoxa1* and *Hoxb1* (Figure 1-10). In case of *Hoxa1*, the 3' RAREs already have retinoid receptors bound along with the repressor Ncor. When RA binds to RARs, a conformational change occurs to the receptors and the repressor is removed and replaced with an activator. Through this relief of repression, *Hoxa1* gets activated in response to RA. In the case of *Hoxb1*, initially there are no retinoid receptors bound to the 3' RARE. In presence of RA, the RAR/RXRs are recruited to the RAREs and *Hoxb1* is activated for transcription. In this way, RA is able to de-novo activate *Hoxb1* (De Kumar et al., 2015) (Also see reviewed in Nolte & Krumlauf, 2019; Parker & Krumlauf, 2017).

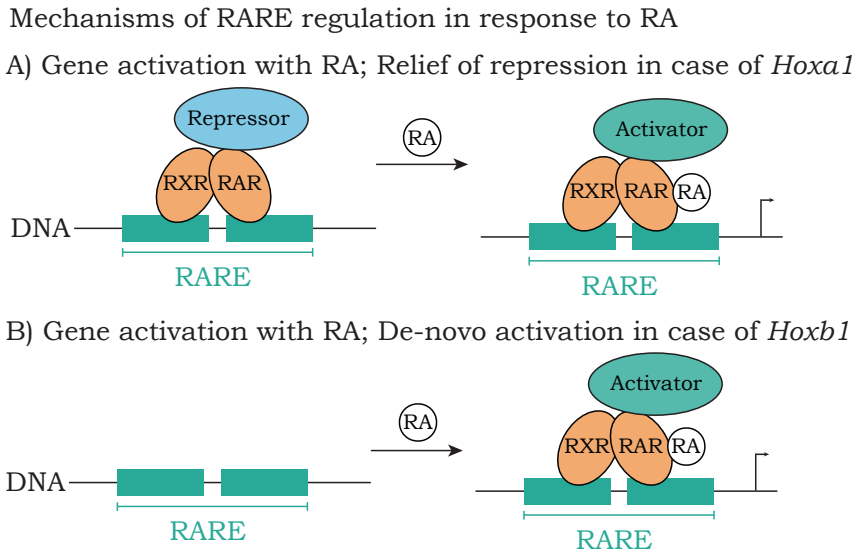


Figure 1-9. Mechanism of RAR/RXRs binding to RAREs to activate transcription of genes

Schematic of how retinoid acid receptors (RARs and RXRs) heterodimerize and bind to specific retinoic acid response elements (RAREs) on the DNA to regulate transcription. In the case of *Hoxa1* (A), RA has to bind to RARs to activate transcription by relieving repression from bound repressors. In the case of *Hoxb1* (B), RA bound to RARs enables them to bind to RAREs to activate transcription. (Adapted from Cunningham & Duester, 2015)

1.5 Transcriptional control of *Hox* genes

1.5.1 *Hox* genes as transcription factors

Hox genes contain homeobox sequences that were found to be highly conserved across species. They were proposed to have transcriptional activity, based on the similarity of the homeobox sequence to the mating-type transcriptional regulatory proteins in budding yeast (Klar, 1987) (Hagen et al., 1993) and to the helix-turn-helix transcriptional regulators in bacteria (Anderson et al., 1981; McKay & Steitz, 1981). Initial structural studies suggested that HOX proteins would have similar DNA binding properties (Gehring et al., 1990; Jabet et al., 1999; Kissinger et al., 1990). Then it was identified that *Hox* genes indeed bound to the DNA, but also that different homeodomain sequences preferred to bind to different regions (Desplan et al., 1985; Levine & Hoey, 1988). *In vivo* evidence for homeodomains having transcriptional activity came from the analyses of *fushi tarazu* (FTZ). A gene was identified that contained an auto-regulatory region to which FTZ could bind and regulate its feedback interactions (Schier & Gehring, 1992). Currently, we know that *Hox* genes encode for transcription factors, which bind to enhancers elements in order to either activate or repress promoters. The *Hox* transcription factors bind to DNA through their homeodomain, which is a helix-turn-helix motif that binds to the major groove of DNA (Gehring et al., 1994). Through this binding *Hox* genes can auto and/or cross-regulate their own expression and also that of their target genes. Due to the similarity in the homeodomains, the *Hox* paralogous group members have similar bindings *in vitro*. However, the members have distinct functions in vivo and can exhibit specific binding preferences in different cellular contexts (Crocker et al., 2015; Slattery et al., 2011). This specificity in binding could partly be attributed to specific domains of *Hox* expression, but also comes from the cis-regulatory activity and differences in the HOX protein complexes formed with co-factors, such

as the TALE proteins, PBX and MEIS (Reviewed in Mann et al., 2009; Merabet & Mann, 2016; Zandvakili & Gebelein, 2016; Zeiske et al., 2018). How exactly do all the inputs regulate and confer tissue specific activity for *Hox* genes, are still major questions in the field.

1.5.2 How enhancers regulate transcription

Transcription of genes, i.e., the act of copying a DNA sequence into an RNA transcript by RNA polymerase, is the first step towards the expression of a gene. Essentially, transcription can be considered to be under the control of promoters, enhancers, and boundary elements present within the genome. Promoters can be of different types and in time either be active, poised, or in an inactive state, but are mostly located near Transcription start sites (TSS) (Reviewed in Haberle & Stark, 2018). Enhancers and boundary elements can be found within genes, but mostly are observed to be present in non-coding DNA regions. Enhancers can be present at distances as small as 1kb to distances as far as 1000kb away from genes (Reviewed in Marsman & Horsfield, 2012). In response to signaling cues, enhancers by interacting with promoters, can activate transcription independently of their orientation to genes (Sanyal et al., 2012). Enhancer elements contain motifs, like a specific DNA signature, to which particular transcription factors bind in a tissue specific manner to regulate gene expression (Reviewed in Furlong & Levine, 2018; Long et al., 2016; Schoenfelder & Fraser, 2019). Transcription factor occupancy (both the affinity to bind and the spacing between factors), motif identification and composition (Avsec et al., 2021), histone modifications, and hence the enhancer 'grammar' are active fields of study to better understand how gene expression is modulated.

Advances in imaging and genome wide chromatin capture techniques (such as Hi-C) have shown that our genomes are arranged into Topologically Associated Domains (TADs) (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). These TADs bring enhancers and promoters near

each other, and some studies have shown that enhancers are more likely to regulate promoters within the TAD than outside. Boundary elements at the edges of the TADs, contain binding sites for insulator proteins such as CTCF, and this may help prevent improper or untimely enhancer-promoter interactions which are corelated with gene expression levels (Guo et al., 2015; Lupiáñez et al., 2015; Narendra et al., 2015). Whether TADs form before or after chromatin looping is still debated, and how exactly TADs are formed is not known. Currently, the loop-extrusion model aided by cohesin is considered the most fitting model (Furlong & Levine, 2018). Disruption of some TAD boundaries in cis has resulted in significant phenotypes (Symmons et al., 2014). Interestingly however, mutants of both CTCF (Gambetta & Furlong, 2018; Nora et al., 2012) and cohesin (Rao et al., 2017), while disrupting TADs, don't appear to completely disrupt genome compartmentalization or the resulting gene expression (de Wit et al., 2015). This points towards maybe there are elements in addition to CTCF and cohesin that organize the 3D genome to restrict enhancer-promoter interactions, and/or it could also be that TADs do not restrict interaction of all enhancers (Schwarzer et al., 2017). Thus, the question of how exactly enhancers, dynamically acting either locally or globally, gain specificity for regulating single or multiple promoters needs to be studied further.

In eukaryotes, it has been observed that transcription occurs in bursts – on and off cycles of transcription (Bartman et al., 2016; Fukaya et al., 2016). Over the years, several models for how enhancers may contact promoters to regulate transcriptional bursts have been proposed. These include tracking (Kong et al., 1997), linking (Morcillo et al., 1997), and looping (Deng et al., 2012; Dunn et al., 1984). The sheer number of enhancers, insulators, and transcriptional factor binding sites make it difficult to imagine tracking and linking as means by which enhancers quickly find the correct promoter to regulate. In addition, cohesin and CTCF are ubiquitously

expressed and can interact with different factors to confer tissue specificity. These points make looping, provided it is as dynamic as the enhancer-promoter interactions, the most likely model for transcriptional regulation. Furlong and Levine propose that gene activation is a two-step process where first enhancers are primed, i.e., brought into close proximity with promoters and with presence of paused Pol II are in a ready-to-go state (Ghavi-Helm et al., 2014; Shao & Zeitlinger, 2017). Then secondly, in response to some cue, change in topology takes place and gene transcription gets activated. In the field, condensates of transcriptional hubs have also been proposed as a model by which transcriptional events are regulated (Hnisz et al., 2017). According to this model, enhancer and promoter do not need to be in physical contact but can be in proximity of anywhere between 100-300nm (Lim et al., 2018). In these condensates, the presence of multiple mediator complexes, pre-initiation complexes, and Pol II could act as bridges that connect enhancers to promoters (Cho et al., 2018).

In terms of the *Hoxb* cluster, it can be envisioned that if the run-through transcriptional model was true, then Pol II would go along the *Hoxb* cluster, and sequentially turn on all the genes within the cluster. Since there can be multiple Pol IIs going through and transcribing the cluster, at a given timepoint all the genes would appear to be transcriptionally on (Figure 1-10, panel A). If this model is true, one would also expect that adjacent genes or non-coding RNAs would be turned on or off in synchrony. However, if this run-through model is not true and there are dynamic interactions between various enhancer elements and promoters of genes within the cluster, then most likely the cluster is configured in some form of a loop. A loop allows for quick movement between various cis-regulatory elements, allowing for variations in enhancer-promoter interactions. Thus, in the looping model, one can envision that the *Hoxb* cluster, in different cellular contexts especially along the embryonic axis, has different enhancer-promoter

contacts (Figure 1-10, panel B). These contact differences would allow for genes to be turned on and off differentially under independent regulatory mechanism. In this scenario then *Hoxb* genes would not be on at the same time as they would transcriptionally be activated by different RAREs or different combination of enhancer-promoter contacts.

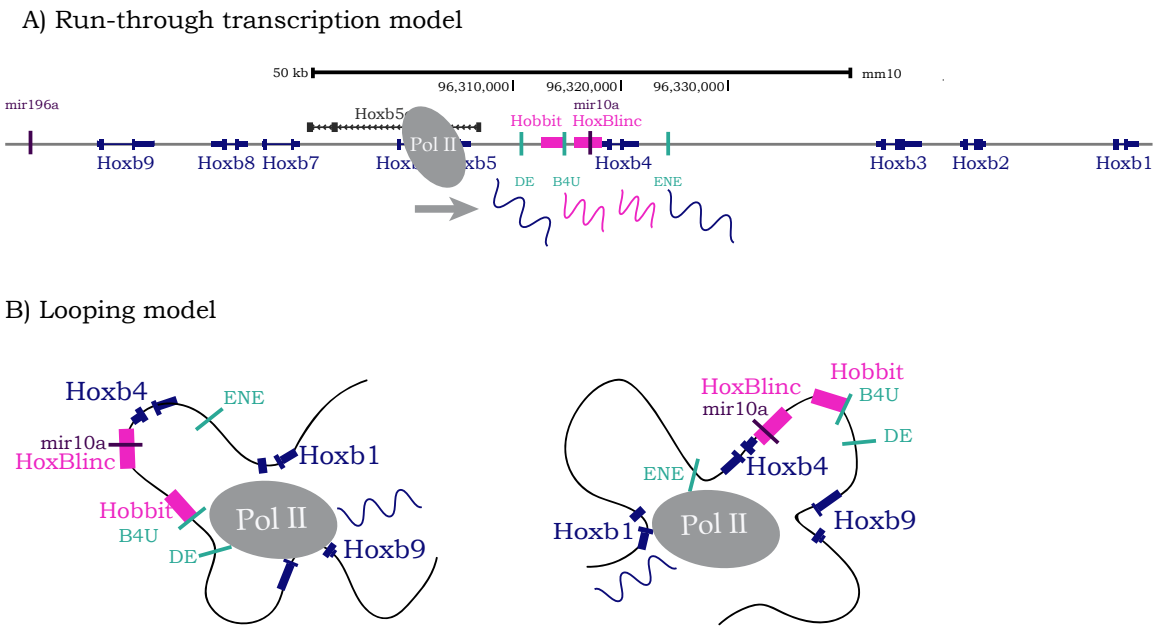


Figure 1-10 Potential transcriptional models for the *Hoxb* cluster

Potential models how genes within the *Hoxb* cluster are transcriptionally activated. A) shows the run-through model where genes get transcribed one after another as if Pol II is sliding along a track. B) shows the looping model where enhancer-promoter contacts are dynamically changing, and Pol II only transcribes upon activation cue. Hence, there is differential activation of genes within the cluster.

In light of recent evidence of condensates regulating nascent transcription (Henninger et al., 2021), it might well be that dynamic looping within the transcriptional condensates are regulating gene transcription, and synthesized transcripts themselves are modulating further transcription. Other recent studies, have also shown that multiple cis-regulatory modules may be present within a transcriptional hub, competing or working together towards transcriptional states of genes (Espinola et al., 2021). Intriguingly, this group also found that contacts between cis-regulatory modules were indistinguishable in neuroectoderm, mesoderm, and dorsal ectoderm,

and that these contacts appeared early in development before the formation of TADs and expression of genes. Both these studies raise questions about whether these phenomena, of condensate formation and dissolution with feedback from RNA molecules and the stability of enhancer-enhancer or enhancer-promoters contacts, are globally true for all transcriptional activities or if they highlight specific examples among a variety of scenarios that might exist to regulate gene expression in a developing embryo.

To broaden our understanding of transcriptional control, the *Hoxb* cluster, serves as an ideal genomic locus to study how multiple enhancers may work to regulate transcription of multiple *Hox* coding and non-coding genes. Important questions like how enhancers are shared between the different *Hox* genes, and how they may make stable interactions to activate all genes simultaneously within the cluster or dynamically activate a specific subset of genes, will help to get a better picture for how genes are regulated by enhancer elements.

1.5.3 RAREs in the *Hoxb* cluster

In the *Hoxb* cluster there are several RAREs that have been identified to act as enhancer elements; RAREs 3' of Hoxb1 and RAREs in the intergenic region between *Hoxb4-b5* and flanking *Hoxb4* (Ahn et al., 2014; Nolte et al., 2013; Nolte & Krumlauf, 2019). The RAREs in the middle of the *Hoxb* cluster are all DR5, meaning they are a unit of two direct repeats with a 5bp spacer region (Reviewed in Nolte et al., 2019). B4U and ENE, that are on either side of *Hoxb4*, have equivalent RAREs present in other *Hox* clusters. However, DE is specific only to the *Hoxb* cluster (Ahn et al., 2014; Gould et al., 1998; Oosterveen et al., 2003; Sharpe et al., 1998). Through enhancer assays and BAC clones, the activity of these enhancers and their regulation in cis has previously been studied (Ahn et al., 2014). It was observed that the DE enhancer was able to locally and globally regulate the whole *Hoxb* cluster in the neural tube and

in hematopoiesis (Ahn et al., 2014; Qian et al., 2018). It was also observed that in addition to regulating *Hoxb* coding genes, the RA induction of *Hobbit* in the DE mutant was also reduced (De Kumar et al., 2015). The B4U is unique in its positioning compared to DE and ENE RAREs, such that it is positioned at the end of a lncRNA, *Hobbit*, making it an enhancer from which a lncRNA is transcribed. In the *Hoxb* BAC reporters, it was shown that at 9.5dpc the ENE mutated embryos had low *Hoxb4* expression which was restored to normal by 12.5dpc. While *Hoxb9* rostral expansion is gone in the double DE-ENE BAC mutant, no change in the neural tube is observed in the single mutants (Ahn et al., 2014). These RAREs or RA responsive enhancers in the middle of the cluster have been shown to be functional, and through this thesis, I have looked at the endogenous role of these RAREs in the mouse embryo and elucidated their potential mechanisms of action in regulating transcription of *Hoxb* cluster coding and non-coding genes.

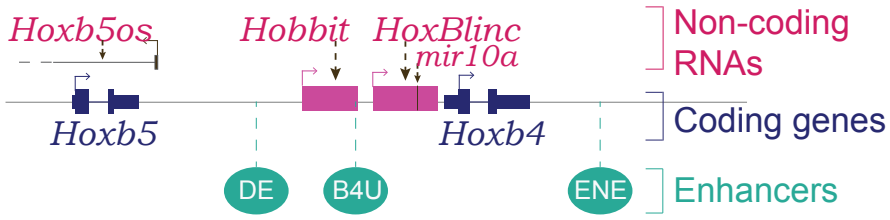


Figure 1-11. RAREs present in the middle of the *Hoxb* cluster

Schematic of the *Hoxb4-Hoxb5* intergenic region. The relative positions of long non-coding RNAs in pink and the enhancer elements (RAREs) in green are shown with respect to coding *Hoxb4-b5* in blue.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Mouse lines

For wildtype animals, Stowers F1 strain was used. F1 are a cross between CBA/CaJ x C57Bl/10J and are maintained by Laboratory Animal Services (LAS) at Stowers Institute for Medical Research (SIMR).

The RARE enhancer mutants used for the project were generated in the Krumlauf lab by Christof Nolte and Youngwook Ahn using a CRISPR approach. Guides were targeted to RAREs and the genomic sequence was either replaced with an EcoRI site (B4U mutant and DE/B4U mutant) or by Dox sequence which was floxed out (DE mutant).

Hobbit-KO and *Hobbit* eGFP lines were generated by LAS at SIMR using CRISPR guide mixtures prepared by Genome Engineering at SIMR. For *Hobbit*-KO line, guides targeting a region 250bp upstream and a region 250bp downstream from *Hobbit* start site were used. For *Hobbit*-eGFP line, guides targeting *Hobbit* start site and 1kb downstream of start site were used along with single stranded DNA (ssDNA) of 100bp homology for the whole region. The homology arms were designed such that 1kb of *Hobbit* would be substituted for ~1kb of eGFP.

2.1.2 Fixing embryos

For fixing embryos, paraformaldehyde 16% vial (Alfa Aesar, Catalog #43368) was used and diluted to 4% with 1X Tissue culture grade PBS (Life technologies, Catalog #20012050). For dehydrating and rehydrating embryos pure methanol diluted in PBS was used (VWR Catalog #BDH1135-1LP), and embryos were put through gradients of solutions (25% -75% methanol in PBS). Details on histology sectioning in section 2.2.4.

2.1.3 Small molecule Florescent *in situ* hybridization (smFISH)

Designed coding or non-coding gene probes were ordered from LGC Biosearch technologies. Unlabeled probes were ordered and labelled in-house with Alexa Florophores.

For smFISH the buffers used were ordered from LGC Biosearch technologies; Hybridization buffer (Catalog #SMF-HB1-10), Buffer A (Catalog #SMF-WA1-60), and Buffer B (SMF-WB1-20). For preparing working solutions of hybridization buffer and wash buffer A, formamide either new vial or sealed for use less than 3 months was used (Amresco Deionized formamide; VWR Catalog #97062_008). For permeabilization of embryo sections, Proteinase K was used (stock of 20mg/ml solution, Fisher Scientific Quality Biological Inc Catalog #E195_5ml). For DAPI a 1:5000 dilution was used in a prepared solution A (Thermofisher 1mg/ml stock, Catalog #62248). For overnight Hyb and final mounting, 0.13-0.16mm long coverslips were used. For mounting slides for imaging, Prolong Gold was use (Invitrogen Catalog #P36930).

2.1 Methods

2.2.1 Genotyping of animals

To genotype the animals, tail samples were collected after weaning the animals. The tails were either processed internally and genotyped (details in section 2.2.1.2) or directly sent for processing to Transnetyx (details 2.2.1.1)

2.2.1.1 Genotyping by Transnetyx

Currently, for each mutant animal, Transnetyx assays have been set up which detects for presence of wildtype or mutant DNA sequence. The assays are designed such that they detect the presence of sequences and can identify if a change in sequence has been made. The link to the website is https://www.transnetyx.com/

2.2.1.2 Genotyping for RARE mutants

Internal genotyping of the mutant tails was initially done by extracting DNA from the tails. Quick crude DNA extraction was done for these tail samples. Tails were dissolved in 300µl of NaOH and kept at 100ºC for 1 hour. To each sample 15µl of Tris-HCl was added. 1µl of sample was then used for the PCR reaction. For PCR, Dreamtaq ready mix polymerase was used to set up a 20µl reaction. The annealing temperature of 57ºC for 30 seconds for 30 cycles was used with 10 minutes extension at 72ºC. After PCR was done, each reaction was digested with specific restriction enzyme for each mutant. The animals' genotype was determined by visualization of bands from each reaction on an agarose gel. The following PCR primers and consequent restriction enzyme were used for the mutant samples.

DE mutant

Forward primer – CDN 854 = aaggaggcgcaaatgagttg

Reverse primer – CDN 855 = gtactggccaacaggttcca

Restriction enzyme – NotI

B4U mutant

Forward primer – CDN 856 = tgaatcaaatcctgctctgtc

Reverse primer – CDN 857 = ggagctcgctcctgtccac

Restriction enzyme – EcoRI

ENE mutant

Forward primer – CDN 858 = tctgggggccttacttctcag

Reverse primer – CDN 859 = agctaggaaccctgaagaggtg

Restriction enzyme – EcoRI

2.2.1.3 Genotyping for *Hobbit* lines

To generate *Hobbit* mouse mutants, CRISPR-Cas9 technology was used. GuideRNA target sites were designed using CCTop target predictor tool (Stemmer et al., 2015). The target sites were selected by evaluating the predicted on-target efficiency score and the off-target potential (Labuhn et al., 2018). Selected guideRNA sequences were ordered as an Alt-R crRNA from IDT. To generate the EGFP insertion, guideRNA target sites were selected at 50bp upstream and 750bp downstream of the *Hobbit* lncRNA transcription start site. A donor plasmid containing the genomic sequence with EGFP in place of the N-term 750bp *Hobbit* sequence was used as a template to produce long single stranded DNA according to the published Easi-CRISPR method (Quadros et al., 2017). To generate the 500bp deletion, two guideRNA target sites were selected at 250bp up and down stream of the *Hobbit* transcription start site.

Each full length guideRNA was formed by hybridizing the specific crRNA with the universal tracrRNA. Ribonucleoprotein (RNP) complexes were prepared by hybridizing Cas9 HiFi v3 enzyme (IDT) and the full length guideRNA at room temperature for 10 mins. For mouse embryo injections, 10ng/ul of each guideRNA and 10ng/ul of Cas9 were combined with 10ng/ul of the purified long single stranded DNA donor.

G0 Mice were screened for the expected mutations by an ear clip. Mutation detection was done by first lysing the mouse ear clips in Epicenter QuickExtract DNA Extraction Solution. For the *Hobbit*-eGFP samples PCR was performed to amplify the specific genomic location at both guideRNA target sites as well as the junctions of the eGFP insertion. With the *Hobbit* deletion samples PCR was performed to amplify the specific genomic location at both guide sites and for the presence of the deletion between the guide sites. A second round of amplification incorporated sample-specific dual barcodes. All amplicons were pooled and size-selected using

ProNex Size-Selective Purification System (Promega). Cleaned pools were quantified on a Qubit Fluorometer and ran on an Agilent Bioanalyzer to check sizing and purity. Purified pools were run on an Illumina MiSeq 2x250 flow cell. Resulting sequence data was demultiplexed and read pairs were joined. On-target indel frequency and expected mutations were analyzed using CRIS.py (Connelly & Pruett-Miller, 2019). In addition, for the *Hobbit* deletion samples for the F1 generation and later the PCR amplicons were run on a Caliper Labchip GX using the DNA 5K kit to detect positive deletion samples.

2.2.2 Staging of embryos

For each embryo collected, the number of somites were counted to ensure same developmental age was used for all experiments. Embryos are considered 9.5dpc if they have anywhere from 21 to 29 somite pairs. For wildtype and mutant RAREs, embryos containing 21-24 somite pairs were used for all experiments.

2.2.3 Collection of embryo samples

For collecting wildtype embryos, Stowers F1 animals were paired and pregnant females upon plug checking were marked for embryo collection at 9.5dpc. For enhancer mutants, mouse lines were bred to homozygosity. The homozygous animals were then crossed, and pregnant females upon plug checking were marked for embryo collection at 9.5dpc. The pregnant females were euthanized according to IACUC protocol number 2019-064 (detailed protocol is available at eprotocol.stowers.org). The uterine tissue containing embryo sacs were extracted from the abdominal cavity and kept in cold PBS solution. The embryos were taken out of the sacs, the outer membrane removed, and they were kept in cold PBS until further processed as needed for specific procedures. If a heterozygous animal cross was made, then for each embryo, its yolk sac was also saved to determine genotype.

2.2.3.1 Embryo collection for smFISH

After collecting 9.5dpc embryos in cold PBS. The embryos were kept in 4% PFA solution overnight at 4ºC. The embryos were then dehydrated using methanol gradients of 25%, 50%, 75% and 100%. The embryos were stored in -20ºC for at least one day before histological processing.

2.2.3.2 Embryo collection for RNA sequencing

The embryos were collected and then segmented into parts. A Head section (cut above brachial arches), trunk section (the section left after head and tail cuts), and a tail section (pre-somatic mesoderm plus 3 somite level). The brachial arches and heart samples for each embryo were also collected. All samples were stored snap frozen and kept at -80ºC.

2.2.4 Histological preparation of embryos

For smFISH RNAase free conditions were used. The embryos were kept in methanol and rehydrated slowly with PBS. The embryos were cut into three sections to aid with correct orientations; head (cut made after brachial arches), mid (cut made after hindlimb bud), and tail sections. After this the samples were dehydrated using a 15% solution of sucrose made in DEPC treated PBS. Once, the samples were saturated, they were embedded into OCT (VWR #25608-930) using the Histochill unit (SP Scientific Products # HC80). Using a Thermo Scientific Microm Cryostar NX70 cryostat, transverse sections at 10microns were placed on Sure Bond charged slides (Avantik #SL6332-1); Alternate sections were placed onto a single slide. The slides were stored at -80ºC if not going to be used within the same week or kept at -20ºC if going to be used within the same week.

2.2.5 Whole mount *in situ* hybridization with DIG labelled probes

Probe making from plasmid constructs for *Hoxb1*, *Hoxb4*, *Hoxb9*, and *Hobbit*.

1) 5µg of DNA was linearized by cutting with appropriate restriction enzymes (Table 2-1) to generate anti-sense probes for each gene.

Table 2-1 Processing of plasmids for probe synthesis

Probe	Clone Number	Restriction enzyme	RNA polymerase
Hoxb1	866	ClaI	T7
Hoxb4	650	BamHI	T3
Hoxb9	803	EcoRI	T7
Hobbit	3-4	SacI	T7

2) RNA polymerase reaction was performed on 1µg of linearized DNA using either T7 or T3 polymerase with the 10X DIG labelling mix from Roche (Catalog #11277073910)
3) Probe integrity was checked on gel before precipitating probe with ammonium acetate. Probe was dissolved in 100µL of formamide and stored at -70^{o}C.
4) For all probes 1:500 dilution was used during hybridization

In situ

<u>Day 1</u>

1) Rehydrated sample (mouse embryos stored in methanol at -20^{o}C)
    a. Series of 75%, 50%, 25% methanol in PBT series
        i. 75%: 7.5mL methanol 2.5mL DD H_2O
        ii. 50%: 5mL methanol 5mL DD H_2O
        iii. 25%: 2.5mL methanol 7.5mL DD H_2O

iv. 100% PBT

b. Washed embryos for each step by aspirating out media with thin tip pipette; left some solution inside, and placed on rocker for 5 mins for each step at room temperature

2) Washed twice with PBT at room temperature for 5 mins

3) Proteinase K treatment (5µL in 10mL from a 20mg/ml stock)

i. For mouse embryos incubated in 10µg/mL proteinase K in PBT (time depending on embryonic stage)

9.5dpc : 10 mins.

10.5dpc : 15 mins

4) Washed 10 minutes in 2mg/ml glycine in PBT (**made fresh each time**)

5) Washed twice with PBT at room temperature for 5 mins

6) Postfixed with 4% paraformaldehyde and 0.2% glutaraldehyde in PBT for 20 minutes at room temperature (kept samples on rocker)

a. 25% stock of glutaraldehyde 200µL

b. 16% Paraformaldehyde 1mL

c. PBT 23.8 mL (to make 25mL)

7) Washed twice with PBT at room temperature for 5 mins

8) For slow equilibration with hybridization solution (can be warmed if gets cloudy)

a. Washed 10 minutes in a 1:1 mixture of hybridization solution/PBT

b. Washed 10 minutes in hybridization solution

9) Incubated at 70°C in hybridization solution for at least 1 hour

10) Replaced hybridization solution with a fresh solution containing the RNA probes (1µg/ml concentration); incubated at 70°C overnight

a. Up until adding separate probes, all the embryos can be kept in the same container
b. After adding individual probe solutions, used Eppendorf tubes and used 200µL solution mix
c. The head of the embryo was pierced with 21gauge needle before adding probe to prevent otic probe trapping.

Day 2

1) Pre-warmed solution I to 70°C
    a. Washed embryos 3 times for 30 minutes each at 70°C with prewarmed solution
2) Pre-warm solution III to 65°C
    a. Washed embryos 3 times for 30 minutes each at 70°C with prewarmed solution
3) Washed 3 times with fresh TBST for 5 minutes at room temperature (at this point DEPC treated or not doesn't make a difference)
4) Blocked embryos by incubating on rocker for 60-90 minutes in blocking solution
    a. To make blocking solution added 0.02mg or 2µg of Boerhinger Mannheim blocking reagent and added it to 18ml of PBT. Kept it at 65°C for an hour or two. In order to completely dissolve it made sure to shake/mix it after 15/20 minutes.
    b. After powder was dissolved added 2mL of inactivated sheep serum
5) Incubated with anti-dig AP antibody (1:2000 concentration; antibody to blocking reagent)
6) Add 200µL antibody to each sample tube; left overnight at shaker in 4°C room

Day3

1) Washed 3 times for 5 minutes each with TBST at room temperature
2) Washed 5 times for 1 hour (60 or 90 minutes) in TBST at room temperature

3) Washed overnight in TBST at 4ºC over gentle rocker

Day 4

1) Washed 3 times in NTMT for at least 10 minutes each at room temperature (over rocker)
2) Added BCIP/NBT reaction mix (10µL/mL of NTMT solution) and covered tubes with aluminum foil.
3) Kept solution until reaction appears to be done (typically after 6 hours started seeing reaction and left them over night at 4ºC before post fix and imaging).

2.2.6 smFISH optimized protocol

The smFISH existing protocol (Raj & Tyagi, 2010) was optimized and adopted to use for mouse embryo sections.

2.2.6.1 Probe design

For probe design genomic sequences were obtained from Genome Browser (genome.uscs.edu). The multiple 20-22 bp intronic and exonic probes (for *Hoxb1*, *Hoxb4*, *Hoxb5*, *Hobbit*, and *HoxBlinc*) and the multiple 29bp intronic probe (for *Hoxb9*) were designed using the Stellaris Probe generator from LGC Biosearch technologies (https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer).

2.2.6.2 Probe synthesis

For smFISH the probe-set sequences were either bought from the Design Ready version offered by Stellaris or were designed using the Stellaris probe designer tool. For probe labeling, unlabeled probe sets carrying a C-term TEG-Amino tag, were purchased from Biosearch Technologies and labelled as previously described (De Kumar et al., 2017). Probes (5 nmol) were fluorescently labeled overnight in 0.1 M sodium tetraborate pH 9 at 4 °C. Labeling

occurred with AlexaFluor-488, AlexaFluor-568 or AlexaFluor-647. Two units of amine reactive succinimidyl ester Decapacks (ThermoFisher) were used for each reaction and following quenching labeled probes were purified using Reverse-Phase HPLC. Probes were separated using an Ettan LC (GE Healthcare) using a 4.6 × 250-mm, 5-µm, C18-EMS end-capped Kinetex column (Phenomenex). Mobile phase A was 0.1 M ammonium acetate (EMD) pH7 and mobile phase B was acetonitrile (Millipore). A linear gradient of 5% B to 100% (vol/vol) B was run over the course of 20 min at 1 mL/min. Peaks were monitored at 280 nm for probe and at 488, 568 or 647 nm, depending on the dye. Dual positive peaks were collected by hand and concentrated by spin vac. Samples were then resuspended in DEPC water to a final volume of 100µL.

Table 2-2 SmFISH probes labelled in-house and optimized to use on mouse embryo sections

	Hoxb4	***Hoxb4*** **intron**	***Hoxb5***	***Hoxb5*** **intron**	***Hobbit***	***HoxBlinc***
Probes labelled with fluorophores	555	555	555	647	647	555
	488	488	488	555	488	488
	Atto488	594	Atto488		Atto488	
	594	647	594		594	

* The highlighted probe sets are the ones that have been validated to work.

Hoxb1 probe was ordered pre-labelled with Quazi-570.

2.2.6.3 Optimized smFISH protocol

Day 1

1) For smFISH on the sectioned 9.5dpc mouse embryos (that have already been fixed with 4% PFA overnight, and dehydrated with methanol series, and then cryo-sectioned).

a. The slides were allowed to thaw to room temperature by keeping them at room temperature for ~10mins.
b. The samples were permeabilized with 10ug/ml concentration of proteinase K in PBS (re-fixing not done as sample embryos were already fixed before they were sectioned). I use slide holder containers to put sample slides on top of the solution; I use the blue microscope slide mailers (Amazon, Catalog #B00X6L1NM4), as they can hold up to 1ml of solution in slide holder spot. To prepare 10mL of permeabilization solution, use 10mL of PBS and add 1µL of proteinase K solution. [Other containers I have used, appear to distort the sections when I close the containers]

During this step, I prepare working solution for wash buffer A. For making 10mL solution, I use 7ml of nuclease free water, 2ml of concentrated wash buffer A, and 1ml of formamide.

c. The slide was washed with buffer A for 5mins by replacing solution in the blue slide mailer [The samples can be left longer in solution, but I have not crossed more then 15mins]

While samples are being washed, I prepare the hybridization buffer with probes. For making 1ml of Hyb solution, take 900µL of hybridization buffer and 100µL of formamide. Per sample slide I added about 100µL of Hyb with probes. I use a 1:100 dilution for probes; each probe is ~5nmol stock concentration dissolved in 100µL of nuclease free water. Hence, in 100µL of Hyb I add 1µL of smFISH probe.

d. For each slide, I added 100µL Hyb solution with probes. I add solution in a droplet manner over the sections, and then cover the slide with a coverslip. I then keep the

slides in a wet chamber. First, I heat the slides at 65ºC for 10mins, and then leave them for an overnight incubation at 30ºC (~16hrs).

Up to 3 different probes can be added to the Hyb solution (same volume added for each probe and using the 1:100 dilution factor for each ~5nmol probe) and 4 colors can be imaged simultaneously with DAPI being the fourth color). With embryo sections placed alternatively on to slides, for the approximately the same axial level I can have smFISH nascent transcript data for up to 6 probes. The major combinations I have tested on the wildtype and all mutant embryos are:

Table 2-3 Probe-set combinations used on alternate mouse embryo sections

	Combination 1	Combination 2
Probe 1	*Hoxb1* – 570	*Hobbit* – 647
Probe 2	*Hoxb4* – atto 488	*HoxBlinc* – 555
Probe 3	*Hoxb9* – 647	*Hoxb4* – atto 488

(gene probe combinations written out by genomic locations)

Day 2

1) The coverslips were carefully removed. If the slides appeared little dry, I soaked in wash buffer A and then removed coverslip. After Hyb the embryos may start to look a little distorted around the edges if they have been kept much longer than 16hrs and the slide is drying.
2) DAPI was added to wash buffer A (1:5000 ratio), and slides kept on DAPI solution. After keeping the slides with DAPI solution at room temp for 10mins, slides were moved to 4ºC for ~2-4 hours.

3) Slides were washed with Buffer B, by keeping slides on solution at 4°C for 4 hours (I have also washed it at room temperature for 15mins, but with some probes that cluster I prefer to wash longer in 4°C as it appears to help the signal).
4) The slides were mounted with prolong gold by taking 5µL volumes and putting drops along the center of the slide. Then the slides were covered with a 0.13-0.16mm long coverslip and kept for drying at room temperature.

Samples were then kept at 4°C until imaging. Slides used to be good for imaging up to even a year after mounting, but now the signal seems to fade away after 4 weeks. [I have replaced and tested with all new solutions and fixing reagents, but cannot identify where this difference is coming from]

2.2.6.4 Imaging with Nikon spinning disk

Tissue sections on the slides were imaged using the Nikon spinning disk using the Hamatsu scMos camera. The images were captured using 60X wide angle oil objective. The sections to image were identified by focusing with the 10X objective with a 1.45 NA. The images were obtained at 100% laser power for far red 633, red 561, green 488, and DAPI 405 nm lasers. For each channel the following filters were used: DAPI, ET455/50m, green, ET525/36m, red, ET605/70m, and far red, ET700/75m. A job Tiler was used to make final capture of the images and if tiles were taken, the image was stitched later in image J. The order of experiments was Lambda (z-series), so each color was images in z before moving to next color. For obtaining a z-stack through the tissue section, each slice imaged was 0.3µm apart and a range of 20 steps was taken to be imaged. Images were obtained in the order of 633 (for the 647 probe), 561 (for the 555 probe), 488 (for the atto 488 probe), and 405 (for DAPI). On a slide the whole tissue section

was imaged, and for each slide a single row was imaged all the way across for each genotype of embryo. All images were stitched using imageJ before further processing.

2.2.6.5 Deep learning (DL) processing on images

RNAFish spots were segmented using DeepFiji (Nuckolls et al., 2020). In brief, a small subset of spots from different probes were annotated in Fiji and used to train a 2D Unet (Ronneberger et al., 2015) model. This model was used to infer spot probabilities from full image sets. After obtaining the results from DeepFiji, 3D image sets were thresholded, and 3D segmented in order to apply a size filter, and then projected to 2D. Combinations of these projections were examined to find cases where spots had overlap in multiple channels. Original spots and overlapped spots were then reduced to a single pixel and blurred to provide heat maps. Spot locations were also saved for future analysis. Relevant code for preparing image files for DeepFiji, and the post-processing afterwards can be found at https://github.com/jouyun/smc-macros.

2.2.6.6 Distance measurement methods

Distances were measured between smFISH spots via Gaussian fitting with color correction utilizing custom plugins written for the open source image processing program ImageJ. These tools are available at https://research.stowers.org/imagejplugins. Color correction parameters were determined by fitting tetraspeck beads (Thermo Fisher) to Gaussian functions and finding the translation, rotation, and scaling parameters that minimize the distances between peaks in different channels. Those same transformations were applied to smFISH images collected with the same parameters. After color correction, images were pre-processed via Gaussian blurring with a standard deviation of 1 pixel and rolling ball background subtraction with a ball radius of 15 pixels. Peaks were then found in the far-red channel (due to it's low background) by a max

not mask (spot finding Fiji plugin) approach. Briefly, the maximum voxel in the image is repeatedly found and then a spheroid around that voxel is masked with zeros. For this analysis the spheroid diameter was 20 pixels in xy and 6 slices in z. This procedure is repeated until no other voxels are found with and intensity above 5% of the image maximum.

Once positions were found, they were refined by non-linear least squares 3D Gaussian fitting of 20 x 20 pixel stack of 6 slices surrounding the potential position in all 3 channels. Best fit peak positions were constrained to be within two pixels distance in xy and 2 slices in z of the maximum position found above. For each pair of probes, 2D histograms of spot amplitude vs. spot distance were created (see Figure 4-16) and the area surrounding the main peak of this histogram extending from distances of 0 to distances of 1 micron was gated and plotted in D (Figure 4-16). Finally that histogram was fit to a one dimensional Gaussian, again by non-linear least squares to obtain the peak position of the distance distribution.

2.2.6.7 Measurements post-DL on images

Images were processed using imageJ. Using max projection from the DAPI channel, the region of interests (ROIs) were marked in each tissue section. ROIs for neural tube in forebrain, midbrain, and hindbrain, brachial arches, and adjacent somites were outlined by hand in the DAPI channel and saved. The saved ROIs were then used to obtain measurements from the intensity plot images for each channel.

The intensity plot image files contain 7 channels, outlined below:

1 – Intensity of nascent spots in Channel 1

2 – Intensity of nascent spots in Channel 2

3 – Intensity of nascent spots in Channel 3

4 – Intensity of the co-localized nascent spots in Channels 1 and 2

5 – Intensity of the co-localized nascent spots in Channels 2 and 3

6 – Intensity of the co-localized nascent spots in Channels 1 and 3

7 – Intensity of the co-localized nascent spots in Channels 1, 2, and 3

Measurements of min, max, and average intensity were obtained from the intensity plot files for each ROI in the tissue section. For each tissue section, a background measurement was also taken. This background average intensity was subtracted from the average intensity for each channel for each ROI. The average background intensity was subtracted from the mean intensity from each channel, and the background subtracted values of intensity were then plotted for visualization of data. For each ROI, a Stowers filter table spatial ROI plugin was used to also extract the exact count of nascent spots and their co-localizations from the DL processing's excel file output for every nascent spot over whole tissue.

2.2.6.8 Data Analysis on measurements made post DL

For graphical representation of data either the exact nascent counts or the relative intensity of nascent spots (measurements of background subtracted mean), for specific ROIs over the whole embryo for *Hoxb1*, *Hoxb4*, and *Hoxb9* were plotted. The mouse embryo was divided into head, mid, and tail sections, and R was used to plot graphs of for specific ROIs in wildtype (wt) and RARE mutants (DE, B4U, and ENE) samples for each body segment.

2.2.6.8.1 Processing Head samples

For head samples, I averaged 5-7 sections which contained neural tube from the forebrain, midbrain, and hindbrain regions. In graphs, the hindbrain regions become obvious as they have expression of *Hox* genes, while in wildtype forebrain and midbrain there are no *Hox* nascent transcripts. I also averaged 3-5 brachial arch samples from the head samples.

2.2.6.8.2 Processing Mid samples

For mid samples, I averaged 6-8 neural tube, adjacent somites, and developing gut. These sections were imaged such that each consecutive section is a posterior region in the same embryo. The size of the neural tube and adjacent somites decreases with posterior sections.

2.2.6.8.3 Processing Tail samples:

For tail sections, I averaged measurement values from neural tube and adjacent somite of 6-8 tail sections. These sections were imaged such that each consecutive section is a posterior region in the same embryo. This was done for each experimental condition – wildtype or RARE mutant.

2.2.6.8.4 Plots

There were several different plots that were for each coding and non-coding gene i.e. *Hoxb1*, *Hoxb4*, *Hoxb9*, *Hobbit*, *HoxBlinc*, these include:

1) Comparisons for relative expression of these genes in wildtype and RARE mutants

2) Comparisons for relative expression of the co-localization of these genes in wildtype and RARE mutants

3) Comparisons of the exact number of nascent spots for these genes in wildtype and RARE mutants

4) Comparisons of the exact number of co-localized nascent spots for these genes in wildtype and RARE mutants

2.2.6.8.5 Cell counts for each ROI

In the neural tube, the cells are densely packed together which makes it very hard to automate their counting. Conventional algorithms have a hard time separating out cells and even manually it is difficult to go through each z-slice and count. To counter this problem. A sum projection file was generated using the DAPI channel for each tissue section. A small region where there is a lower density of cells was outlined, cells in that region were counted, and the integrated DAPI intensity per nuclei for this region was made. A ratio of integrated DAPI intensity/cell was used to roughly estimate the number of cells in the ROIs marked in each section. (Integrated DAPI Intensity of each ROI divided by the ratio of Integrated DAPI intensity/cell, to equal the number of cells present within the ROI).

2.2.6.9 Statistical analysis of data

The nascent transcripts datasets were assessed for normality, and it was observed that the samples are bell-shaped. Hence, the non-parametric test Mann-Whitney (also called Wilcoxon rank-sum) test was performed, and all RARE mutant samples were compared against wildtype for each probe set. Stars over plots were used to denote significance.

2.2.6.10 Visual representation of images

There are different ways in which the data was presented for visualizations using Image J.

1) The original image of nascent spots for each channel was false colored and overlaid onto a DAPI image of the section. This image was useful to get an estimate for how much

background might be in the image and importantly against validating how accurately the DL marked nascent spots.

2) Seeing the original image over the whole section was hard to see unless image was zoomed in. So, over a section to see which region of the embryo contains more nascent spots, the intensity plots were used with ROIs drawn from DAPI channel of the image to mark regions over the embryo. These appeared like heatmaps of intensity plots over the whole tissue sections.
3) To see nascent spots over the section, another way to visualize the spots was by overlaying the marked nascent spots by DL over the DAPI image. Since the DL nascent spots are a binary output (spot present or spot absent), these images were cleaner to see than the original images.
4) To be able to see nascent spots detected over the whole tissue section without having to zoom into image, the nascent spots localization file was used and a spot file made. This file could be edited in image J, such that the sizes of the individual spots could be increased. Spot size for each channel was increased consistently across all samples, and this file was overlaid onto DAPI. To generate a spot file for this visual representation, Chris Wood's python program was used.

2.2.7 RNA seq sample processing

Bulk RNA isolation was done for different cut fragments of the embryo; the R4 region, the tail, and trunk fragment. RNA quantification and quality control were performed using an Agilent 2100 Bioanalyzer. Library preparation was done using KAPA HTP kit using stock adapters by Bioo Scientific Nextflex DNA barcodes. Samples were polyA-selected using TruSeq poly A protocol for the cDNA preparation and library construction. Library quality was checked using

an Agilent 2100 Bioanalyzer. Samples were divided into two pools of 20 libraries each and loaded onto 6 lanes of an Illumina HiSeq flow cell to sequence 50 bp single reads for a total of 20-30X genomic coverage. The RNA sequenced reads were aligned to mm10 using Tophat 2.1.1 (Kim, Pertea, et al., 2013). Downstream analysis was performed in R 3.3.2 using EdgeR quasi-likelihood pipeline 1.4.1 (Chen et al., 2016) for differential expression analysis with default settings. Differentially expressed genes between samples were called with adjusted $p < 0.05$.

2.2.8 RACE (3' and 5') to determine *Hobbit* start and end site

Thermofisher kits (5' RACE Catalog #18374-058 and 3' RACE Catalog #18373-019) were used to identify the 5' end and 3' end of the *Hobbit* lncRNA. After running gel in the final step, the gel bands of sizes similar to previously estimated were extracted and purified. Blunt-end cloning was done for the purified sample and several colonies for picked, mini-prepped, and sent for sequencing. Sequencing results identified were used to determine the starting and ending regions. The starting site was 177bp upstream and ending site was 46bp downstream of previous estimate. Hence the new co-ordinates for *Hobbit* (KR231012 on UCSC genome browser) with the mm10 genome assembly are ch11: 96,312,352-96,314,750.

2.2.9 *Hobbit* CRISPR experiments

2.2.9.1 Identification of guides

Initially to identify guides to target specific *Hobbit* regions, the *Hobbit* sequence was input into the CRISPR.MIT.EDU (site has now been shut down). Guides were ordered based on their targeting score and minimum number of off-target effect score.

2.2.9.2 Deletions strategies

Many different deletion strategies were tried with different combinations of guide(s) and reagents. Some deletions were targeted towards deletion of the whole 2.7kb *Hobbit* region, some for deletion of first 1kb, some for the second 1kb, and some towards the last kb. Several different combinations of the guides were used. The strategies that were tried include:

1) Single guide px330 plasmid which would also encode for the Cas9 protein and tracer. To target a deletion, two or four of these constructs (two different guides to target same region) were used together for either electroporation or microinjection.
2) Dual guide px330 plasmid which encode for 2 guides and also encode for the Cas9 protein and tracer. One or two constructs were thus set up for use. Two constructs were used with a similar idea of using multiple guides to target the same region.
3) Guides were ordered as oligos, and an in-vitro reaction mixture with the Cas9 protein and tracer was set up before use.

None of the above strategies resulted in the expected deletion. One combination with px330 single guides resulted in a 15bp deletion, but those animals eventually struggled to breed, and the line was eventually culled.

4) Final strategy employed to target inactivation of *Hobbit* lncRNA was to design guides against the promoter and initial region of *Hobbit*. Guides were designed 250bp upstream and 250bp downstream of *Hobbit* start site. The purpose being to delete *Hobbit* promoter and have no functional *Hobbit* since potential start site is also deleted. This resulted in a total of ~500bp *Hobbit* deletion. Guides and mixture were prepared by Genome engineering group. *Hobbit*-KO line was generated with this strategy. We have 7 founders, 6 heterozygous animals and 1 homozygous animal positive for deletion.

2.2.9.3 Homology Directed Repair (HDR) strategies

The idea was to replace a portion of *Hobbit* such that *Hobbit* lncRNA becomes inactive, but the spatial resolution of the genomic loci remains the same. For this a strategy to replace *Hobbit* with a fluorescent protein was decided in hopes to inactivate *Hobbit*, preserve spatial resolution, and also potentially track *Hobbit*'s early expression.

1) Initially a 3kb homology arm was made with eYFP from Christofs eYFP construct. The final 3kb construct contained a minimal promoter and a poly-A tail and was thus not used.
2) A new 3kb of homology arm was designed (guidance from Narendra) and a construct containing that was made. Long 60-70bp PCR primers were ordered from IDT and three different regions were amplified;

   1) 1kb of *Hobbit* upstream region with overhang for puc19 vector and eGFP

   2) eGFP sequence with overhangs for *Hobbit* 1kb upstream and *Hobbit* region 1kb downstream 3) *Hobbit* region 1kb downstream (region after the 1kb that will be targeted for replacement)

   The 3kb arm was amplified by PCR and purified or an endoplasma free midi-prep was performed on the construct and it linearized. Either of these were mixed in with the guides and Cas9 protein mixture. The prepared solutions were then co-injected into single or two-cell stage embryos. The PCR purified sample with guides mixture which was injected into a 2-cell stage embryo generated an eGFP positive mouse line. However, in addition to the line being mosaic, it was eventually found that the whole 3kb homology arm had gotten inserted into a different genomic location than the targeted *Hobbit* loci. The animals of this like developed severe malocclusions, and the line was culled.

3) A new strategy was employed that used multiple single stranded DNA (ssDNA) of 100bp length to span the 3kb homology arm (made in #2 above). Two guides were used to target *Hobbit* start site and 1kb downstream region. This strategy of using ssDNA and in-vitro preparation of cas9 protein complex resulted in a positive eGFP replacement for *Hobbit* region. These constructs were microinjected into 1-cell stage embryos.

2.2.10 Microinjections and electroporation's of *Hobbit* CRISPR

Single and double cell microinjections were performed by Tim and Andrea in LASF core for the *Hobbit* deletion with eGFP insertion line. Electroporations for *Hobbit* promoter plus start site deletion were done by Michael and Tim in LASF done using NEPA21 electroporater as per company recommendations.

Chapter 3: Characterization of nascent *Hoxb* transcripts in wildtype embryos

3.1 Background

Hox genes have to be expressed in correct spatio-temporal domains for proper embryonic development. In order to understand how the *Hox* genes are regulated, I am looking at their transcriptional dynamics. Specifically, I am looking at the newly synthesized or nascent RNAs for the *Hoxb* genes in the neural tube of the 9.5dpc mouse embryo. Within the 9.5dpc embryo, I can get an estimate of development time through the neural tube. The neural tube towards the tail (caudal) side is less developed and at an early stage, while the neural tube towards the head (rostral) side is more developed and at a late stage. The gradient of RA is also higher towards the head and decreases going towards the tail side, while the sensitivity of *Hox* genes is higher for anterior genes and decreases for posterior genes.

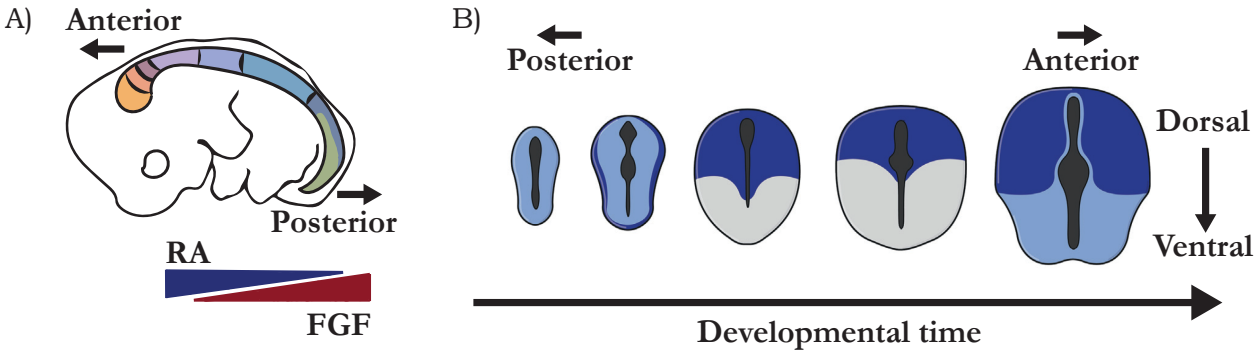


Figure 3-1 Tracking *Hoxb* gene expression in the neural tube along the embryonic axis

Panel (A) shows mouse embryo with gross trends of RA and FGF. Panel (B) shows schematic of transverse section through the neural tube along the embryonic axis with general trends of expression for the *Hoxb* genes. Darker blue represents higher expression levels. As development proceeds the anterior neural tube which is made earlier in development is more advanced than the neural tube at the posterior tail of the embryo(Adapted from Krumlauf, 2016). Panel B reprinted with permission from Elsevier. Chapter Thirty-Four Hox Genes and the Hindbrain A Study in Segments. Curr Top Dev Biol, 116, 581-596. https://doi.org/10.1016/bs.ctdb.2015.12.011.

3.2 *Hoxb* cluster is an ideal genomic locus to study transcriptional regulation

I am studying the *Hoxb* genomic locus to understand the regulatory interplay between enhancer elements and non-coding RNAs. I want to decipher how these cis-regulatory elements potentially

ensure spatio-temporal timing for the transcription of coding *Hoxb* genes. The *Hoxb* cluster is a prime locus to address such a question because interspersed within its coding genes are RARE (Retinoic Acid Response Elements) or enhancers and long non-coding RNAs (lncRNAs). Figure 3-2 (A) shows a schematic of the *Hoxb* cluster. Of particular interest is the intergenic region between *Hoxb4* and *Hoxb5*. Between these two coding genes, lie two important lncRNAs – the novel *Hobbit* and previously studied *HoxBlinc* (Deng et al., 2016). Also, in this cluster, as Figure 3-2 panel B) shows, are present three RAREs of interest- DE, B4U, and ENE. It isn't clear exactly, if the lncRNAs, *Hobbit* and *HoxBlinc* have direct inputs into regulating *Hoxb* genes or if they do this under the control of the RAREs. Hence, the *Hoxb* cluster, which contains along with coding genes both RAREs and non-coding RNAs, is a primed genomic locus to understand the regulatory interplay that dictates transcription of *Hoxb* genes.

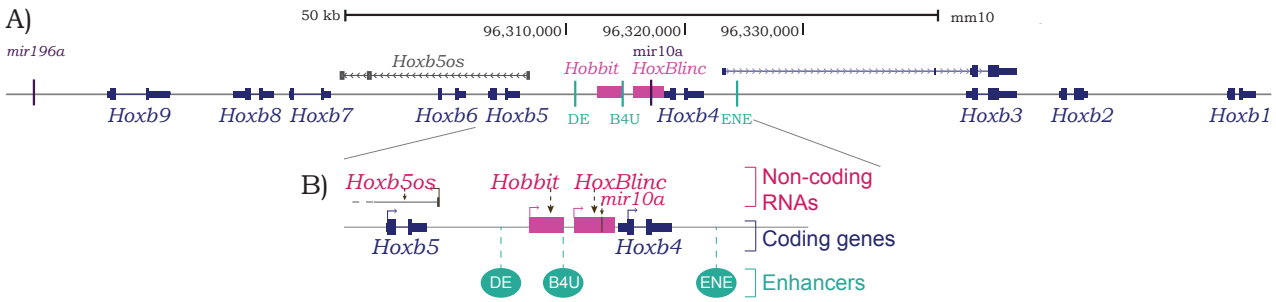


Figure 3-2 Schematic of *Hoxb* genomic locus

Panel (A) shows a complete *Hoxb* genomic locus from the mm10 genome assembly ucsc.genomebrowser.edu. The inset (panel B) shows the intergenic region between *Hoxb4* and *Hoxb5*, depicting the long non-coding RNAs and enhancer elements (RAREs) that are present in the middle of the cluster.

3.3 Single molecule Fluorescent *in situ* hybridization optimized for *in vivo* detection of newly synthesized RNA molecules of coding and non-coding *Hoxb* transcripts

To understand the coordinated transcription of *Hox* genes, the single molecule FISH (smFISH) technique was optimized to use for mouse embryonic sections. This optimized technique allows for taking snapshots in developmental time through the neural tube across the whole embryo.

The smFISH (Figure 3-3) was performed on fixed cryo-sections of 9.5dpc wild type F1 mice (in-house hybrid between CBA/Ca/J x C57BL/10 background). While data for expression of newly synthesized or nascent transcripts over the whole section is obtained because the whole section was probed and imaged, the primary focus of this study is on the developing neural tube through the whole embryo. The developing neural tube is a precursor to the future central nervous system, the brain and spinal cord. While the main focus is the neural tube, in order to draw some comparisons, in instances some other regions, e.g., the adjacent somite, from various sections of the embryo have also been processed.

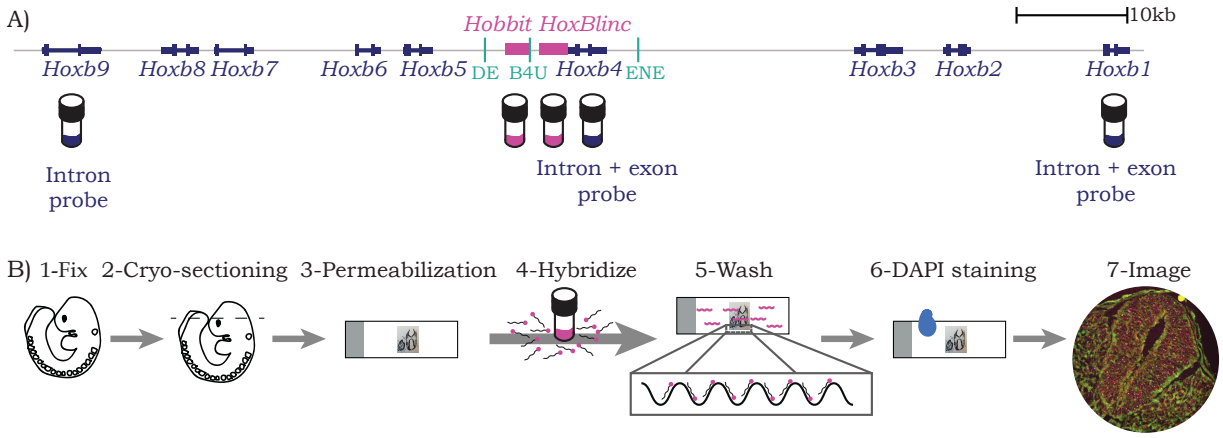


Figure 3-3 smFISH on fixed mouse embryo cryo-sections

Panel (A) shows the *Hoxb* locus with probes that were used along the cluster. Panel (B) shows a schematic for the single molecule fluorescent *in situ* hybridization protocol that was optimized from (Raj & Tyagi, 2010) to detect nascent transcripts in mouse embryo cryo-sections.

Probes were designed against the ends and middle of the *Hoxb* cluster, and up to 3 probes can be imaged simultaneously. The probe sets used were either a mix of intron and exon or intron-only and ranged from 22-26 bp in length (Figure 3-3 A). To observe more than 3 genes within the same relative embryo region, alternate cryo-sections of the embryo were probed to get datasets for up to 6 genes across a single embryo. Figure 3-4 shows one z-slice image through an alternate section of the tail; A) shows DAPI, B) show segments of the neural tube with nascent

transcripts for *Hoxb1*, *Hoxb4*, and *Hoxb9* in panel C). Zoomed insets to see nascent transcripts are shown in D).

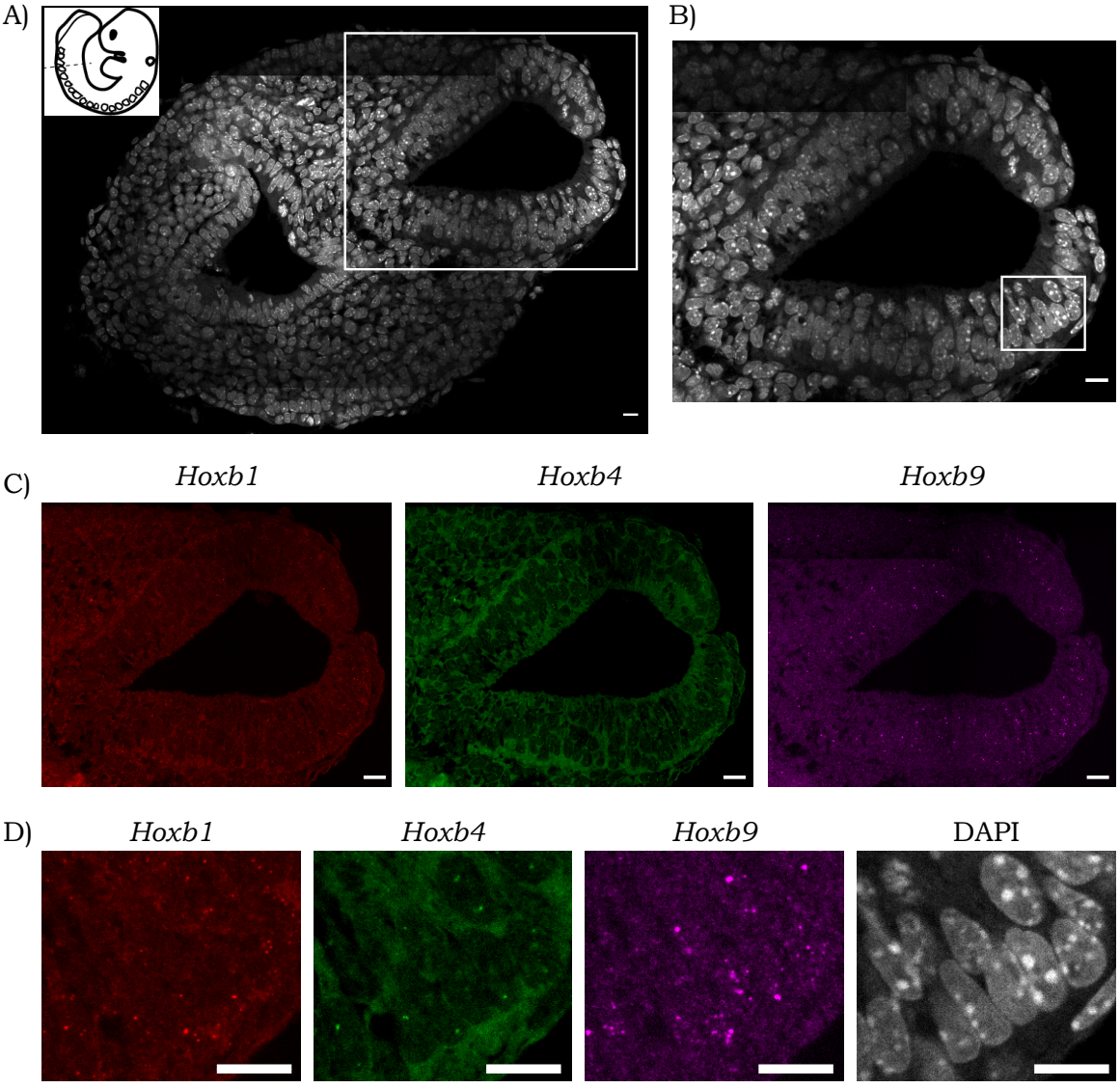


Figure 3-4 Visualizing nascent transcripts in mouse embryo sections

Panel (A) shows image of a tail section, (B) is the zoomed image of the neural tube to show (C) nascent transcripts detected for *Hoxb1*, *Hoxb4*, and *Hoxb9* over whole neural tube in the section. (D) Shows zoomed in region over the neural tube to see nascent transcripts detected at the single cell level. Scale bar - 10microns.

To quantify data in a high throughput and unbiased manner, a Deep learning (DL) algorithm was applied to the probed tissue section images (Figure 3-5). The algorithm was trained on a few

images with a specific probe, and the trained algorithm was able to then be used on all probes across different tissues for both wildtype and mutant embryos.

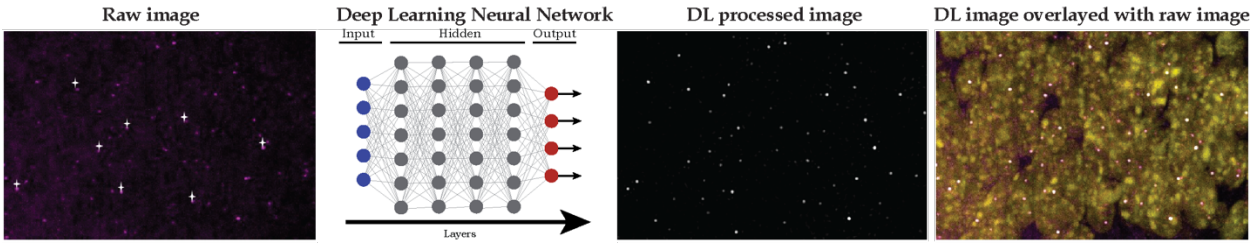


Figure 3-5 Depiction of how Deep learning algorithm detects nascent transcripts

Schematic showing how Nascent spots from raw image files are marked in an unbiased and high throughput manner using Deep Learning. The most right panel shows an overlay of the spots detected by DL over the raw image. Mostly white spots indicate the effectivenss of the DL algorithm being able to detect the nascent spots. DL neural network image adapted from Nosratabadi et al., 2020.

Max projections for each z-slice of the image was run through the DL algorithm to detect and mark spots. The intron probes usually have very little background, and the marked spots look almost like the binary DL output spots (*Hoxb9*). The exon plus intron mixed probe of *Hoxb4*, while marking nascent transcripts, also marks the single transcripts present in the nucleus. Hence, the nuclei outline in the *Hoxb4* images is seen due to cytoplasmic signal. The single transcripts, however, are difficult to quantify because it is very hard to distinguish between autofluorescence, unlabeled probes, and actual single transcripts. In theory they can be spectrally isolated, but it would still be hard to distinguish between bound and unbound probe, and along with that it would significantly slow down imaging and the high-throughput pipeline of smFISH. Hence, analysis was restricted solely to nascent transcripts quantified by DL in the nuclei of cells. A visual readout for the DL processed image compared to max projections of the raw images through the neural tube are shown in Figure 3-6. The figure shows neural tube segments from the tail sections (same tissue section as in Figure 3-4) with zoomed inset to see nascent transcripts in individual cells.

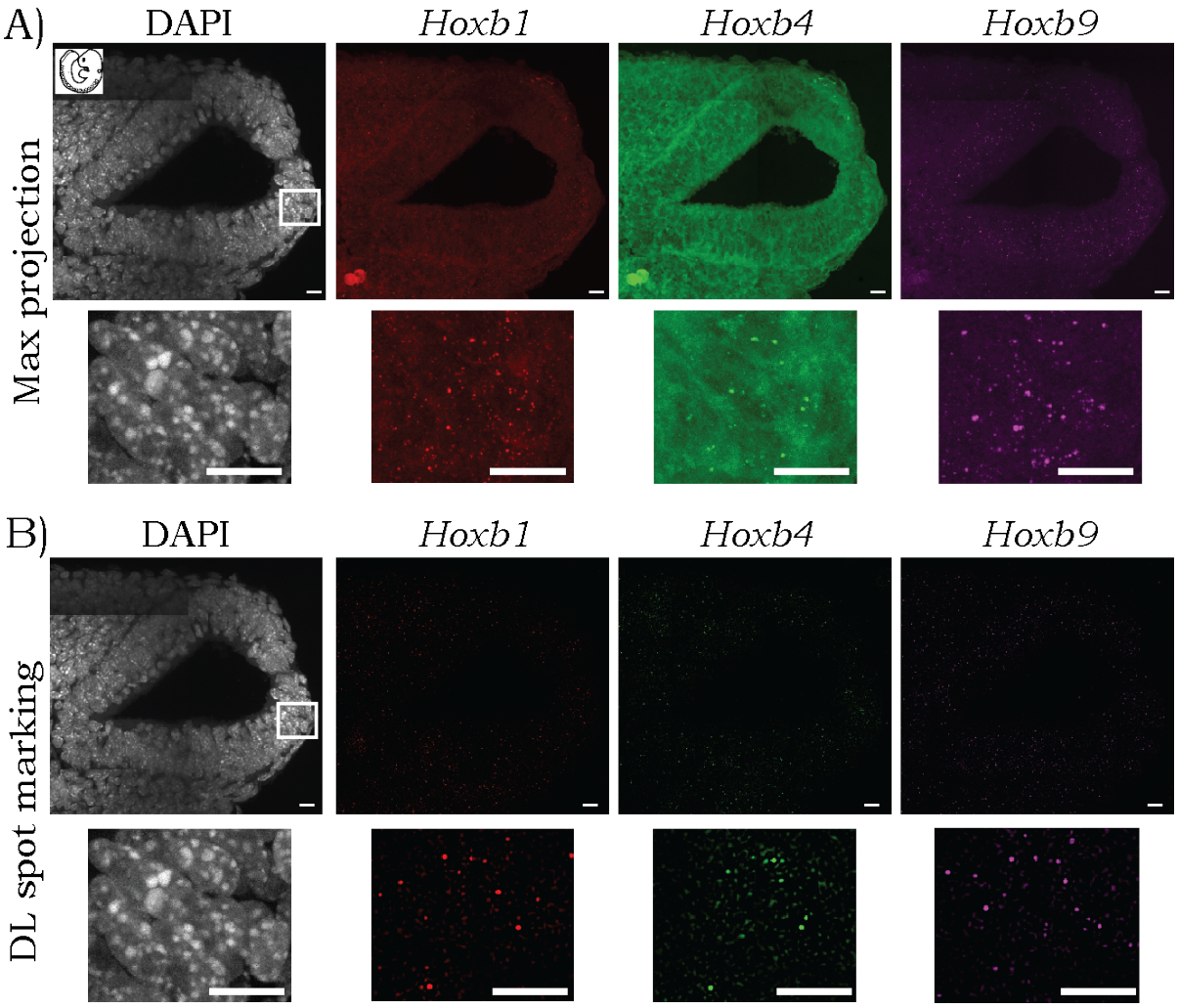


Figure 3-6 Max projections to visualize nascent transcripts

A) Max projections to visualize nascent transcripts are shown for raw image files with zoomed in insets at the bottom. B) Shows the corresponding DL processed image with zommed in insets to show the detected spots. Scale bar - 10microns.

In figure 3-6 and 3-7, panels (A) show the max projection of raw images while panels (B) show the same neural tube segment but with the DL markings of nascent transcripts over the whole region. Figure 3-7 is the same Figure as 3-6, with the difference being that 3-7 is images shown in grayscale.

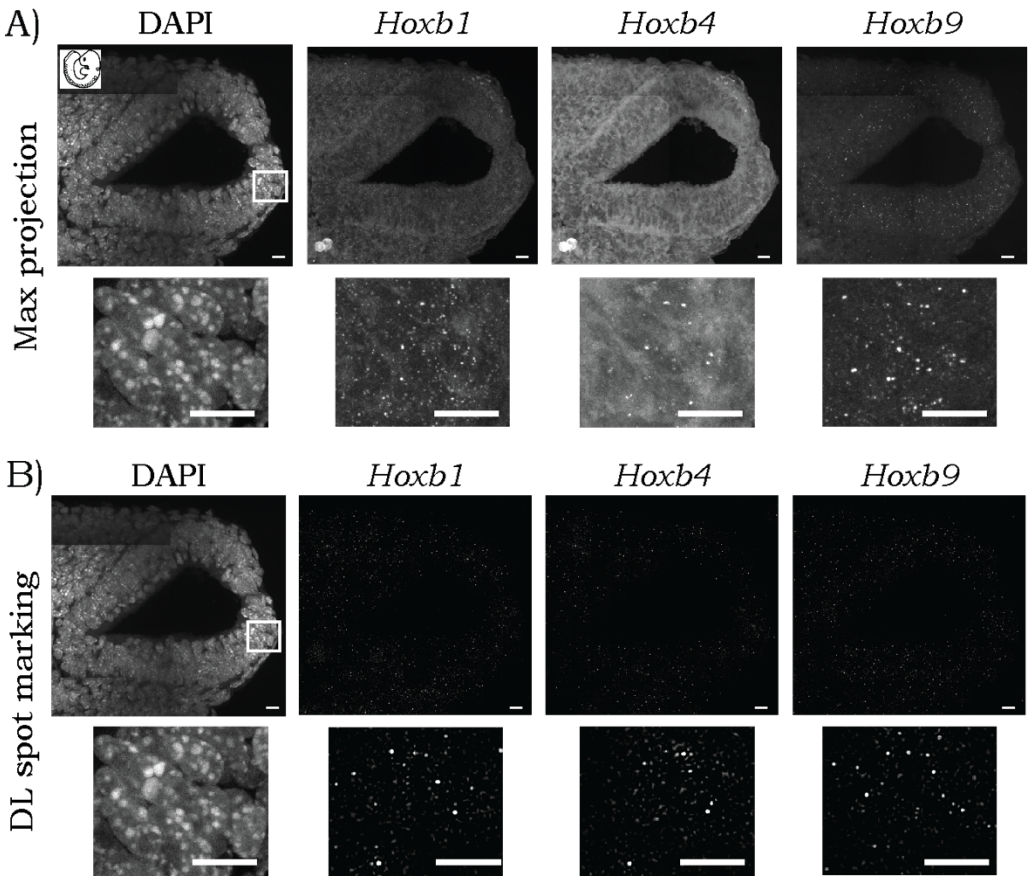


Figure 3-7 Max projections to visualize nascent transcripts in grayscale

Visualization in grayscale for the same image as Figure 3-6. Scale bar - 10microns.

The DL output, for each image processed, consists of a file with spots marked, a file with all co-ordinates/locations for the spots identified on the image, and the intensity plots for which regions have high or low density of nascent transcript spots. The density plots are a sort of heatmap over the image and they allow for quickly being able to identify regions of high or low numbers of nascent transcripts over the whole processed section. For example, in the tail section of the embryo, *Hoxb9* and *Hoxb4* would be expected to have higher expression in the tail than *Hoxb1*, and that can quickly be seen in the heatmaps. Just by looking at spots over DAPI, it is hard to get an estimate of regions where there are cells having more nascent transcripts. However, through the density heatmaps, regions of dense nascent transcript expression can quickly be gauged.

Figure 3-8 show two different LUT's (pseudo color image Look Up Tables) for visualization of *Hoxb1*, *Hoxb4*, and *Hoxb9* transcripts. Panel A) shows grayscale while panel B) shows "fire" LUT with high density of nascent transcripts depicted in yellow to red while low density of nascent transcripts depicted in blue.

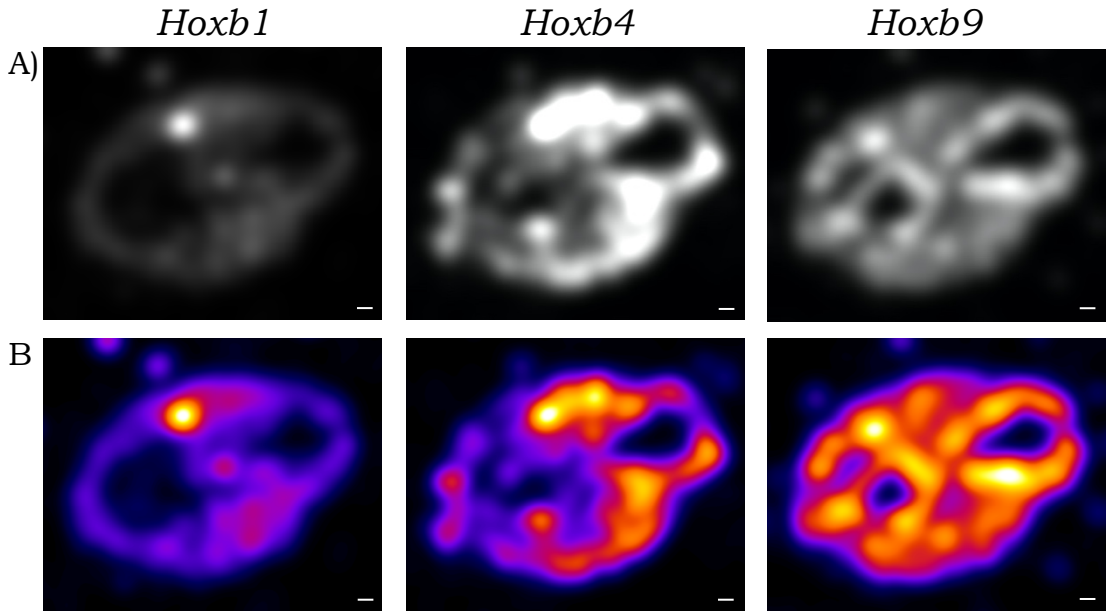
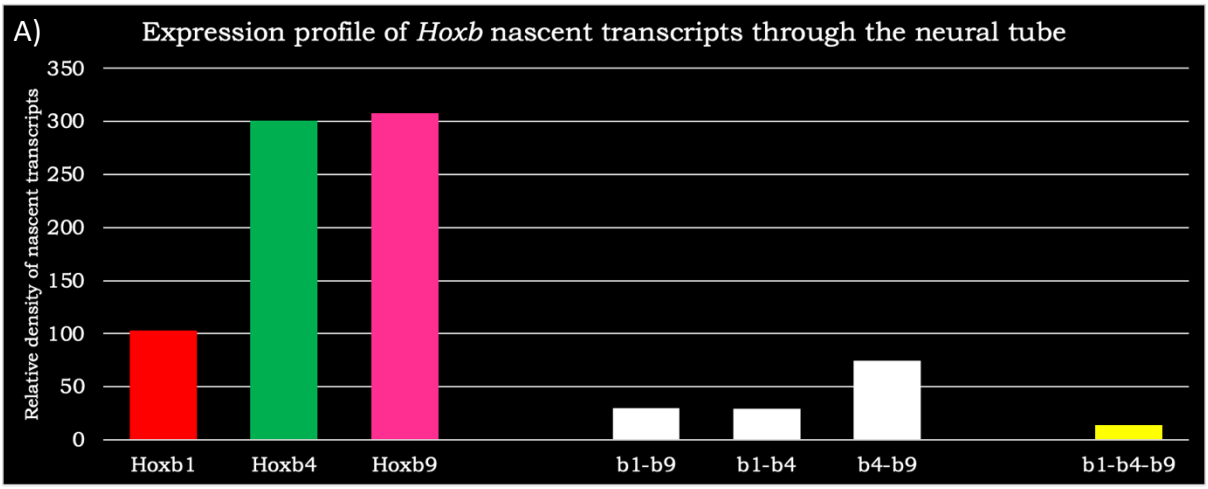


Figure 3-8 Heatmaps of nascent transcript intensity

DL output of intensity heatmaps for *Hoxb1*, *Hoxb4*, and *Hoxb9* (grayscale LUT on top) and the corresponding fire LUT (bottom) for better visualization of regions with high number of nascent transcripts. Scale bar - 20microns.

In addition to heatmaps over the embryo, through DL processing the exact spot locations of nascent transcripts over the embryo can be obtained. This allows for extraction of the exact number of nascent transcripts for any user defined regions of interest (ROIs). Then, either the exact number or the relative densities of nascent transcripts can be used to plot graphs for expression of nascent transcripts over the whole embryo. Figure 3-9 shows graphs with expression profiles for coding *Hoxb* transcripts. Both panel A) with relative densities of nascent transcripts and panel B) with exact count of nascent transcripts can be used to visualize the trends of expression for nascent *Hoxb* gene transcripts.



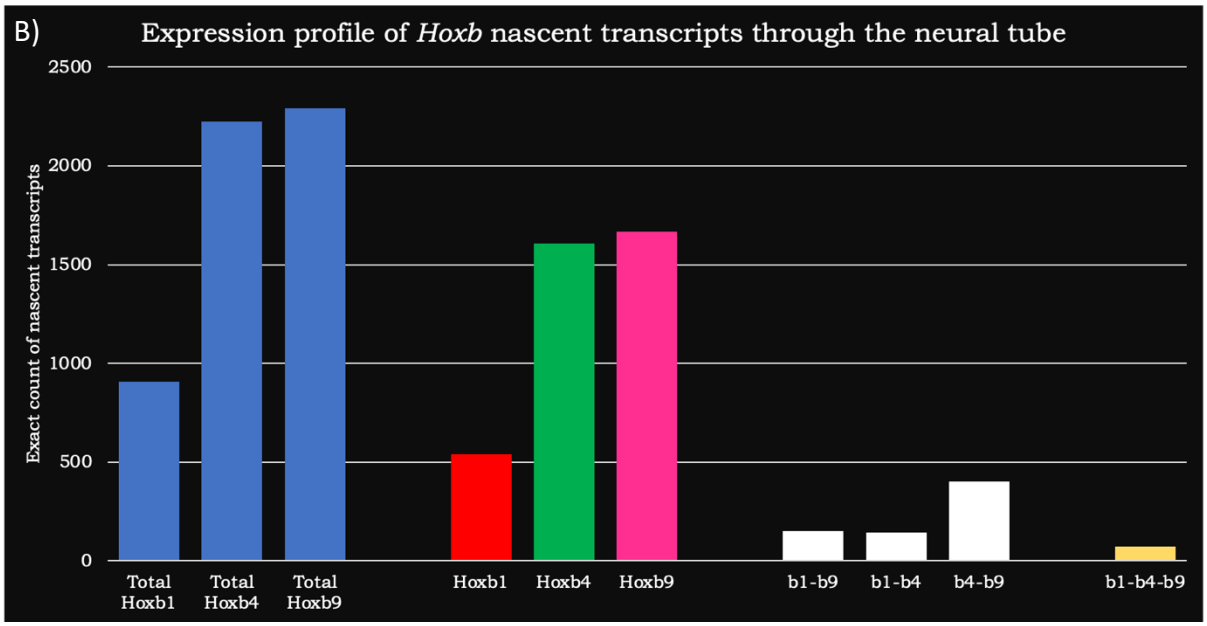


Figure 3-9 Graphs of *Hoxb* nascent transcripts

Graph in (A) shows relative densities for ROI's marked over the heatmap intensity plots, while graph in (B) is the corresponding ROIs showing exact nascent transcript counts from spots marked by DL (n=1 tail section).

3.4 Transcription of *Hoxb* genes is primarily independently regulated

The graph of exact number nascent transcripts for *Hoxb* genes through the neural tube in the tail segment that were plotted (Figure 3-9, B), depict the total number of transcripts for each gene along with a breakdown of the total, the single and co-localized transcripts. The graphs show that levels of individual *Hoxb* transcripts (the singles i.e., *Hoxb1*, *Hoxb4*, and *Hoxb9*) are much higher than the co-localized *Hoxb* transcripts (the doubles i.e., *b1-b9*, *b1-b4*, *b4-b9*, and the triple *b1-b4-b9*). Co-localized transcripts means that the two genes are being transcribed at the same

time. Hence, evidence of co-localized nascent transcripts is indicative of instances where there is some common regulatory control that is in play to turn on multiple *Hoxb* genes at the same time. For the *Hoxb* cluster, it appears that while there is some common regulatory mechanism to turn some genes on simultaneously, the individual *Hoxb* genes are primarily turned on one at a time. This is because a much higher proportion of single transcripts compared to co-localized transcripts is observed. In terms of co-localized transcripts, there are more co-localized transcripts for *Hoxb4* and *Hoxb9* in the tail. This could be explained by either that *Hoxb4* and *Hoxb9* are posterior genes and must be activated in the posterior/tail segment of the developing embryo and are responding to the same stimuli. Or it could be that the highest number of co-localizations for these transcripts is because *Hoxb4* and *Hoxb9* have a greater number of nascent transcripts to begin with in the tail. Hence, the probability of them being co-localized is higher. Furthermore, the expression of *Hoxb9* is the highest in the tail, and it is possible that *Hoxb9* by some mechanism dictates expression of anterior genes in the tail similar to *Hoxc9* (Dasen & Jessell, 2009). The observed probabilities of co-localization of two or more *Hoxb* transcripts were also compared to the random chance of transcripts getting co-localized in a confined space (equivalent to the neural tube defined region). The observed probability of co-localization is less than the probability of random chance of co-localization for the transcripts (Figure 3-10). The question then arises, if the observed probability being lower than random probability, is indicative of the fact that there are actually regulatory mechanisms that are actively repressing two *Hoxb* genes from being co-expressed together. Furthermore, the fact that there are more single transcripts for each *Hoxb* gene, that do not co-localize with any other *Hoxb* transcript, shows that the ends and center of the *Hoxb* cluster primarily have independent regulatory mechanisms that define their transcription.

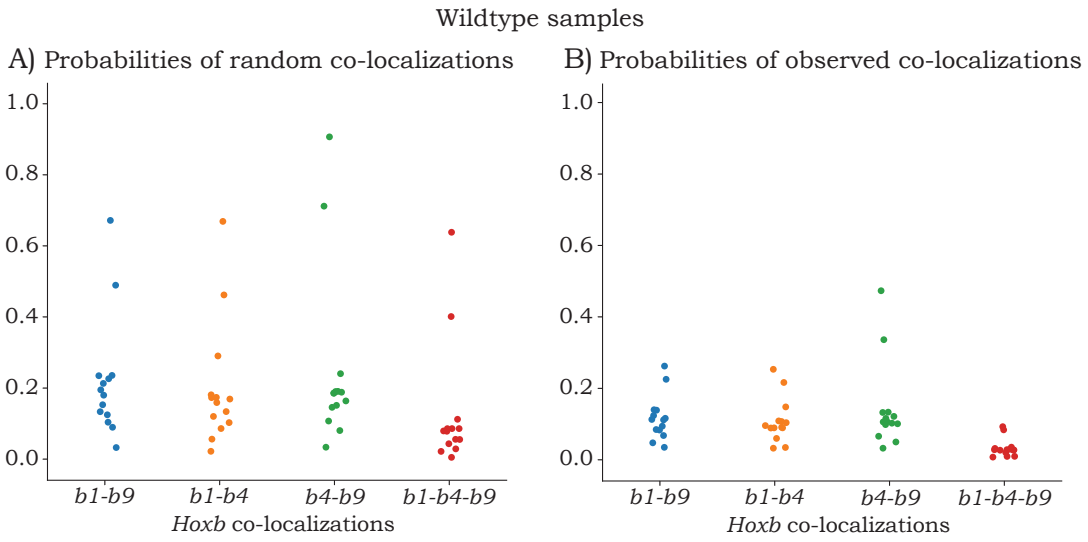


Figure 3-10 Probabilities of co-localization of *Hoxb* nascent transcripts

Graph in (A) shows the probability of random chance of two or three *Hoxb* nascent transcripts getting colocalized. Graph in (B) shows the probabilities of observed colocalization (n=10-14 tail sections from single embryo).

3.5 Independent regulation of *Hoxb* genes is true at multiple axial levels in the embryo

The neural tube in the tail of the embryo shows that the expression profile of single nascent transcripts is higher than the co-localized transcripts. Figure 3-11 shows the trends of nascent transcripts, but in addition the image also depicts cell insets (Figure 3-11, D) that have single transcripts, double co-localizations, or triple co-localization of all three *Hoxb* genes for the same tail section as in Figure 3-8. As the tail has a very high expression of nascent transcripts for the *Hoxb* genes, Figure 3-11 shows a single z-slice through a 10micron tail section with transcripts overlaid on to the DAPI channel to visualize nascent transcripts over nuclei. However, the graphs (Figure 3-11, C) are the total number of nascent transcripts for the whole z-stack through the neural tube.

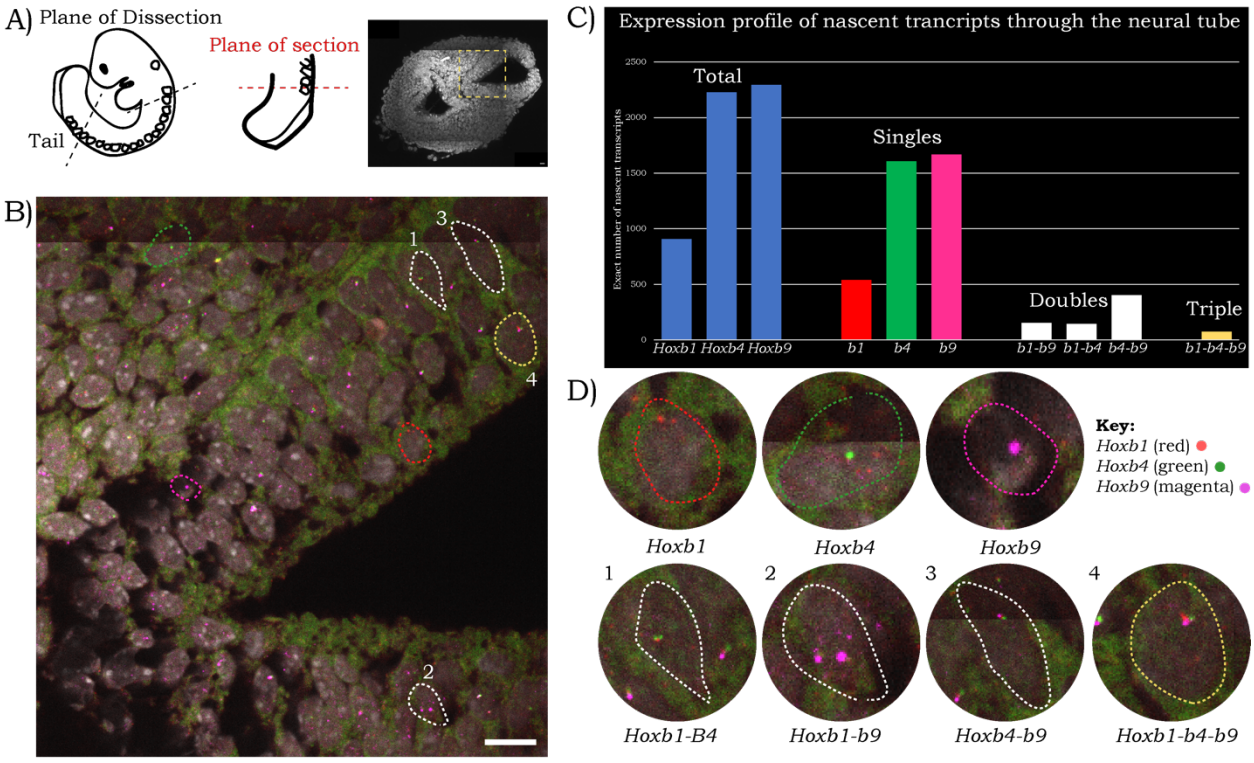


Figure 3-11 Nascent transcript profiles over neural tube of the tail

Panel (A) shows which plane of section from the tail the image is from. In (B) the zoomed transverse section is highlighting the corresponding insets of individual cells (in D) to show cells expressing the nascent transcripts. Total number of nascent transcripts in the neural tube were used to plot graphs of expression for single, double, and triple transcripts (shown in C, n=1 tail section). Scale bar - 10microns.

To observe if trends of *Hoxb* gene expression were similar or different along the axial levels of the embryo, anterior neural tube of the embryo was also imaged and analyzed. A similar visualization (to Figure 3-11) for the expression profile of *Hoxb* genes in the neural tube of anterior body section of the embryo is shown in Figure 3-12. In this region also, all three coding *Hoxb* transcripts are expressed, and hence their trends of single expression or of co-localized transcripts can be observed. In this anterior neural tube, the level of single transcripts for each *Hoxb* gene again is much higher than the co-localized transcripts. This indicates that the even at this anterior axial level, the *Hoxb* genes are largely independently turned on or transcriptionally activated. Further, in this anterior segment of the neural tube, the expression of *Hoxb4* is the

highest. The number of transcripts co-localized the most are those of *Hoxb1* and *Hoxb4* followed by those of *Hoxb4* and *Hoxb9*. The reason *Hoxb4* is more co-localized with the other two genes could either be because it has the highest number of transcripts or it could be that genes in this mid-section of the body are responding to similar regulatory control as *Hoxb4*. A point to note is that the triple co-localizations of *Hoxb1*, *b4*, and *b9* always appear to be much lower than single or any other double co-localizations at anterior body or tail neural tube sections. This would indicate that genes in the cluster are differentially regulated and have distinct regulatory mechanisms towards activating transcription which ensures that all *Hoxb* genes do not get turned on at the same time.

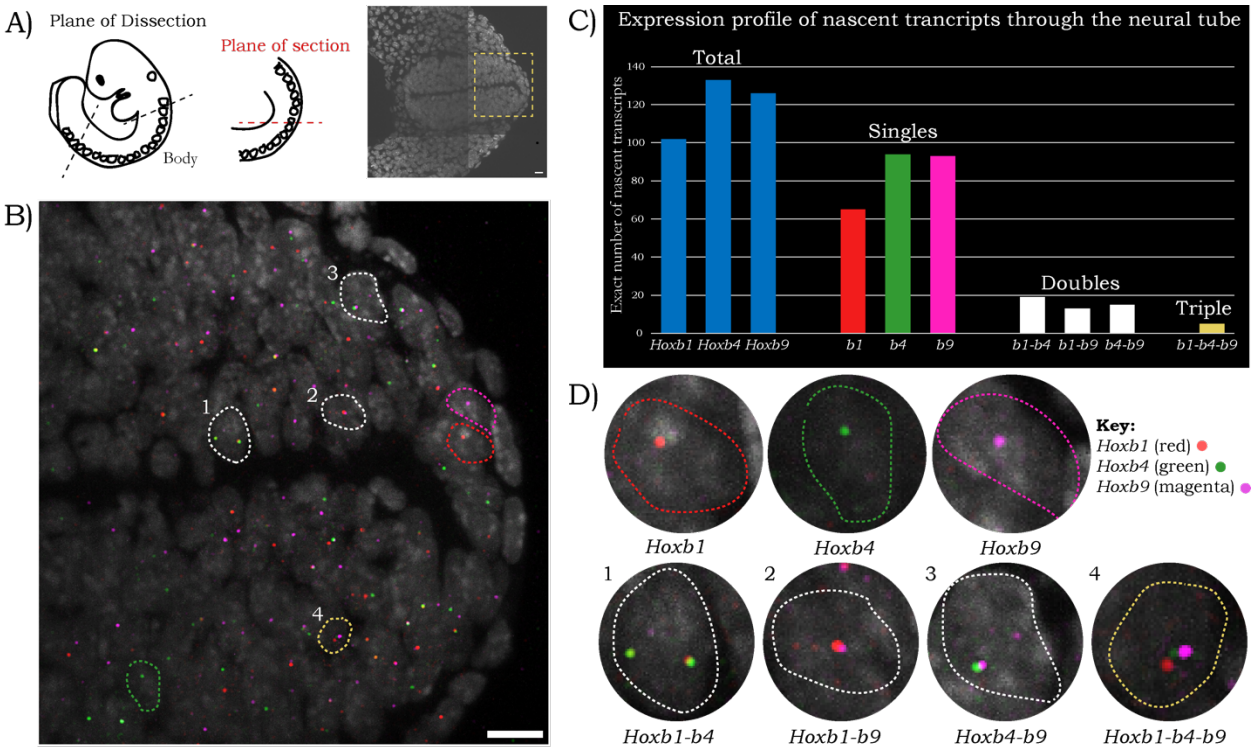


Figure 3-12 Nascent transcript profiles over neural tube of the body section

Panel (A) shows which plane of section the transverse section (B) was imaged. Total number of nascent transcripts in the neural tube were used to plot graphs of expression for single, double, and triple transcripts (C, n=1 tail section). Corresponding insets of cells to show cells expressing the nascent transcripts are also shown (D). Scale bar - 10microns.

3.6 Trends of nascent *Hoxb* transcripts can recapture known patterns of processed *Hoxb* transcripts along the embryonic axis

For wildtype embryos, the whole embryo was imaged and neural tube along the whole embryonic axis was DL processed and analyzed. The relative densities of nascent transcripts in the neural tube along the embryo were plotted (Figure 3-11). It was observed that the expression pattern of nascent transcripts for each coding *Hoxb* gene corroborated with the known expression patterns of processed transcripts of these genes through *in situs* analysis with DIG labelled probes (Medina-Martinez & Ramirez-Solis, 2003; Sharpe et al., 1998; Trainor & Krumlauf, 2000). There are no nascent transcripts observed for *Hoxb4* and *Hoxb9* in the head region, while at specific segments of the head *Hoxb1* nascent transcripts can be detected. This is potentially rhombomere 4 or r4 (segment of the developing hindbrain), which expresses *Hoxb1* like a stripe at r4 during this 9.5dpc developmental stage. Going posteriorly in the embryo, in the mid body sections, initially the nascent transcripts of *Hoxb4* begin to be detected, followed by *Hoxb9* and *Hoxb1*. In mid sections, expression of *Hoxb4* is typically the highest. In the tail sections, all three coding genes are actively transcribed with expression for *Hoxb9* being the highest. In the pre-somatic mesoderm (or edge of the tail), *Hoxb1* is also highly expressed, and similar observations were made for nascent transcript levels being high for *Hoxb1* in the tail. In Figure 3-13, panel A) shows a schematic of the mouse embryo with dotted lines to illustrate roughly the segment of the embryo from which the plotted data in panel B) is coming from. The optimized smFISH technique can capture nascent transcripts of *Hoxb* genes in embryonic sections, but also correlates well with what is known in the literature about the expression patterns of processed *Hoxb* transcripts.

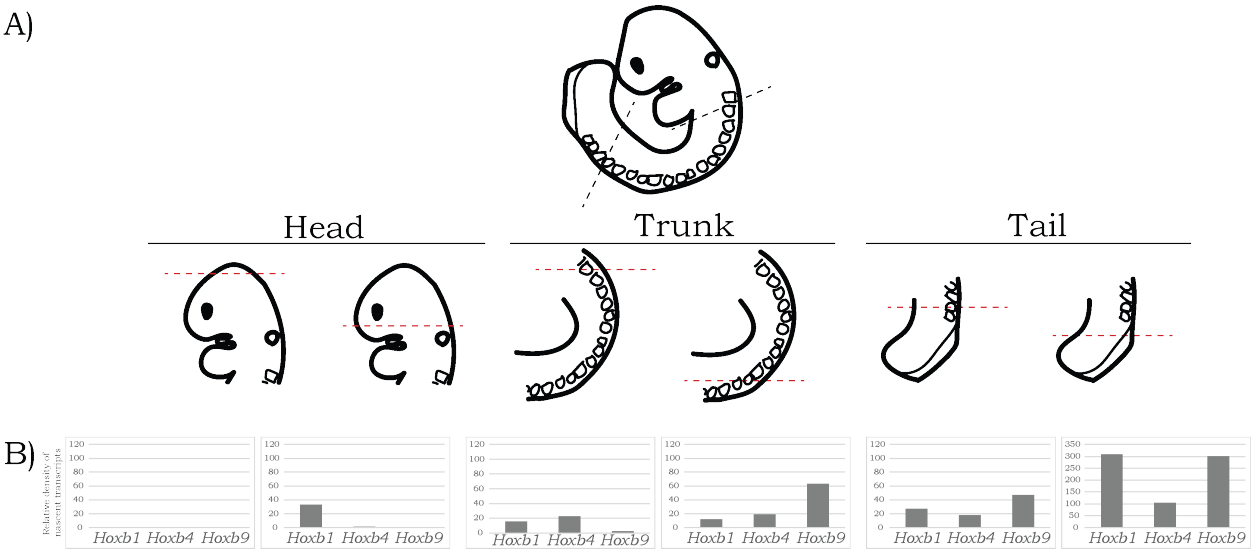


Figure 3-13 Expression of *Hoxb* genes in neural tube over whole embryo

Panel (A) shows sections through the whole embryo to depict changes in relative levels of nascent transcripts in the neural tube along the whole embryonic axis. The graphs (B) are plotted for relative intensity of nascent transcripts from neural tube ROI for the corresponding plane of section (for each graph n=1 tissue section, and all sections are from a single embryo).

3.6.1 Single cell resolution with smFISH can detect dorsal-ventral expression differences

In specific mid body sections near the tail, *Hoxb1* expression appears to be higher in the dorsal side of the embryo. *Hoxb4* expression appears to be the same dorsally and ventrally. While *Hoxb9* expression is highest among others ventrally, its expression is slightly higher in the dorsal side of the neural tube (Figure 3-14, panel B).

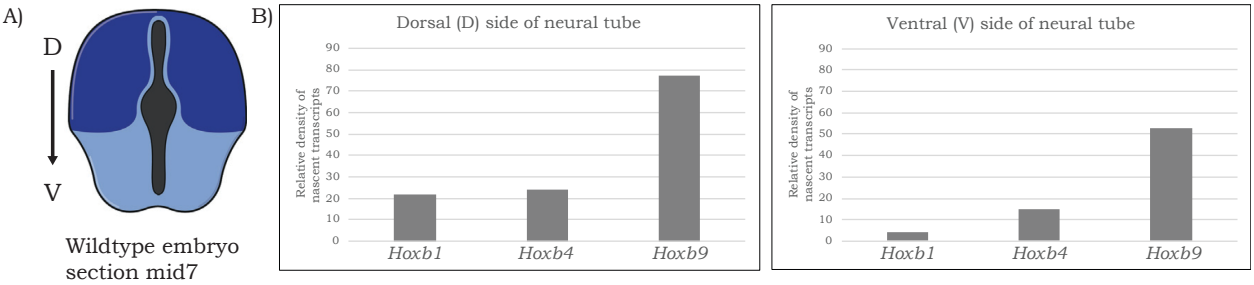


Figure 3-14 Dorsal-ventral expression difference for *Hoxb* genes

Panel (A) shows schematic of transverse section through the neural tube to indicate dorsal and ventral sides. The graphs in (B) show relative intensity of nascent transcripts for *Hoxb1*, *b4*, *b9* in the dorsal and ventral sides of the neural tube (n=1 tissue section from single embryo, ventral and dorsal ROIs marked on the same section by drawing line through approximately half the neural tube).

In the developing neural tube sensory neurons and motor neurons project from dorsal and ventral sides, respectively. Different activation of *Hoxb* genes could be indicative of the distinct dorsal-ventral identities that *Hoxb* genes confer to the developing neural tube.

3.6.2 Sum DAPI intensity can be used to estimate cell counts for regions of interest

Going anteriorly to posteriorly along the axis of the embryo, the number of cells within the neural tube change because the neural tube is at different stages of development. For this reason, when looking across the whole embryo, the relative intensities of nascent transcripts for each region are depicted, knowing that expression high or low indicates high or low number of cells that are actively expressing nascent transcripts. This assumption is possible because for the most part, each cell in the datasets typically expresses one nascent transcript.

However, to compare expression of nascent transcripts in similar regions across different embryos, there is a need to account for cell numbers within the marked ROI's. Best possible way would be to have a method to outline cells. However, counting cells especially densely packed cells as is the case in the neural tube is particularly challenging. Regular cell segmentation or deep learning algorithms don't work to effectively outline packed cells. The best cell segmentation method that recently came out – StarDist (Schmidt et al., 2018), is also not able to segment our cells from the neural tube. In some regions of the tissue StarDist algorithm was able to separate some cells, but in most regions it is clumping 2 or more cells together or separating bright regions within cells as a single cells (Figure 3-15). I have used the pre-trained model of the plugin, so maybe re-training after manual cell marking on our specific needs might give us a better output.

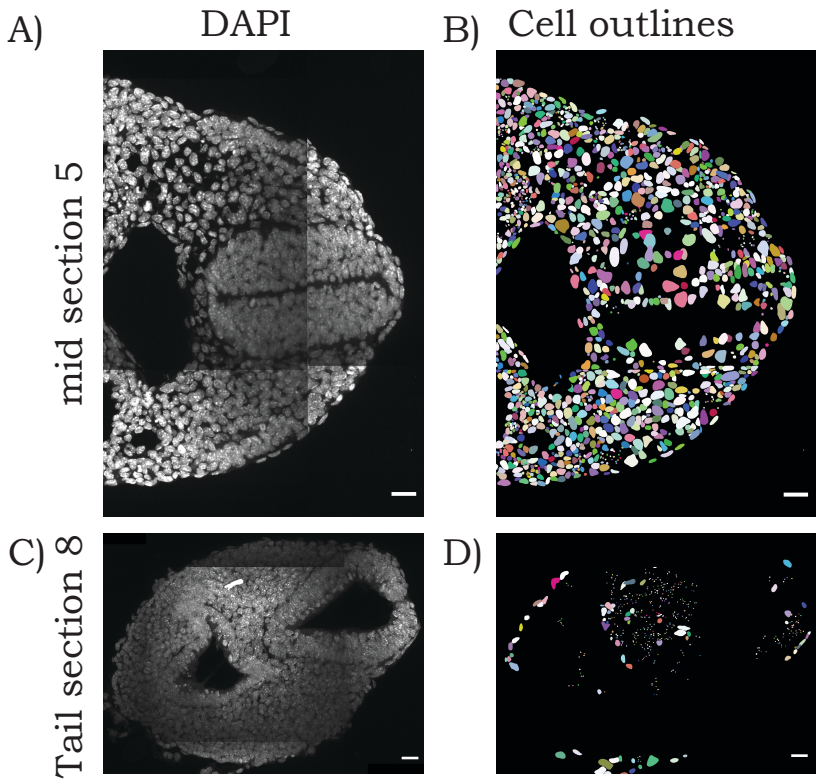


Figure 3-15 StarDist analysis using fiji on wildtype body and tails sections

Panel (A & C) shows DAPI images from body and tail sections (n=1 for each segment) on which StarDist plugin was used to try to outline cells. In the body section where cells are larger the plug does better (B). However, it fails to detect most cells from the neural tube. In the tail (D) where cells are smaller, it does poorly; it clumps multiple cells, marks bright regions within cells as individual cells, and cannot detect of the densely packed neural tube cells. Scale bar – 20microns.

Ideally, use of a nuclear laminin or a cell membrane marker could help outline the cells, and algorithms can be used to detect the outline and count cells. However, to maintain a high throughput imaging and DL pipeline, spectral unmixing is avoided and the maximum four fluorophores, are simultaneously imaged for smFISH probes and DAPI for each section. To circumvent the issue of not having labelled or outlined cells, a sum projection of the DAPI channel that labels nuclei is taken, and it is used to get an approximation of the number of cells present in a specific ROI. Figure 3-16 shows how a ratio of Integrated DAPI intensity/cell count is estimated (1) and how the ratio is divided by the Integrated Density for a specific ROI to estimate the cell count within that ROI. Manual checking of this method was done several times,

and the estimation by this method (Figure 3-16) is similar to the manual counting of cells within a region.

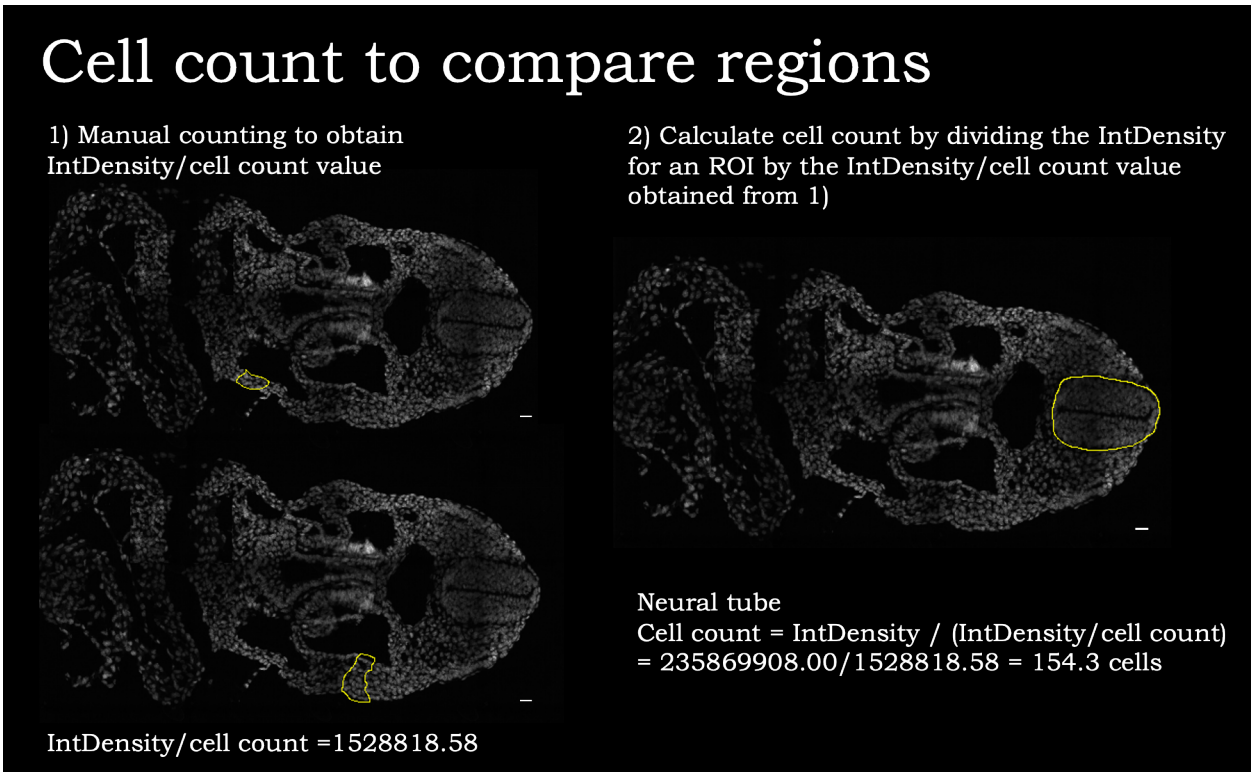


Figure 3-16 Counting of cells for each ROIs

Panel (1 & 2) shows steps for how cell counts for each section are estimated. Section in figure is from mid section 5. A ratio of Integrated DAPI Density/cell count is calculated (1). This ratio value is used to calculate the number of cells in the ROI for each section (2). Scale bar - 20microns.

3.7 Non-coding RNAs within the *Hoxb* cluster are also differentially regulated like the coding transcripts

Similar to coding *Hoxb* genes, the expression profiles of *Hobbit* and *HoxBlinc* present within the cluster were also tracked in the tail of the 9.5dpc embryos (Figure 3-17). Where possible, alternate sections from the embryo were taken to probe with the non-coding probes in order to compare averages of expression for the tail region. Figure 3-17 shows non-coding RNA expression compared to coding *Hoxb* genes. Typically, in the tail, the level of nascent transcripts

for *Hobbit* are higher than all the other transcripts (Figure 3-17, panel B). In terms of co-localizations, *Hobbit* and *HoxBlinc* have very different levels of co-localization with the coding *Hoxb4* transcripts (Figure 3-17, panel C). Even though directly adjacent and could potentially be under the same regulatory control, *HoxBlinc* has lower levels of colocalization with *Hoxb4* compared to *Hobbit*. In the genome, the order from 5' to 3' end is *Hobbit*, *HoxBlinc*, and then *Hoxb4*. Hence, while *Hobbit* and *HoxBlinc* and *HoxBlinc* and *Hoxb4* are adjacent to each other, the expression profiles of these non-coding RNAs indicates that these lncRNAs are not products of run-through transcription which would sequentially activate them. Instead, these lncRNAs are dynamically turning on or off with respect of *Hoxb4* and are potentially under differential regulatory control.

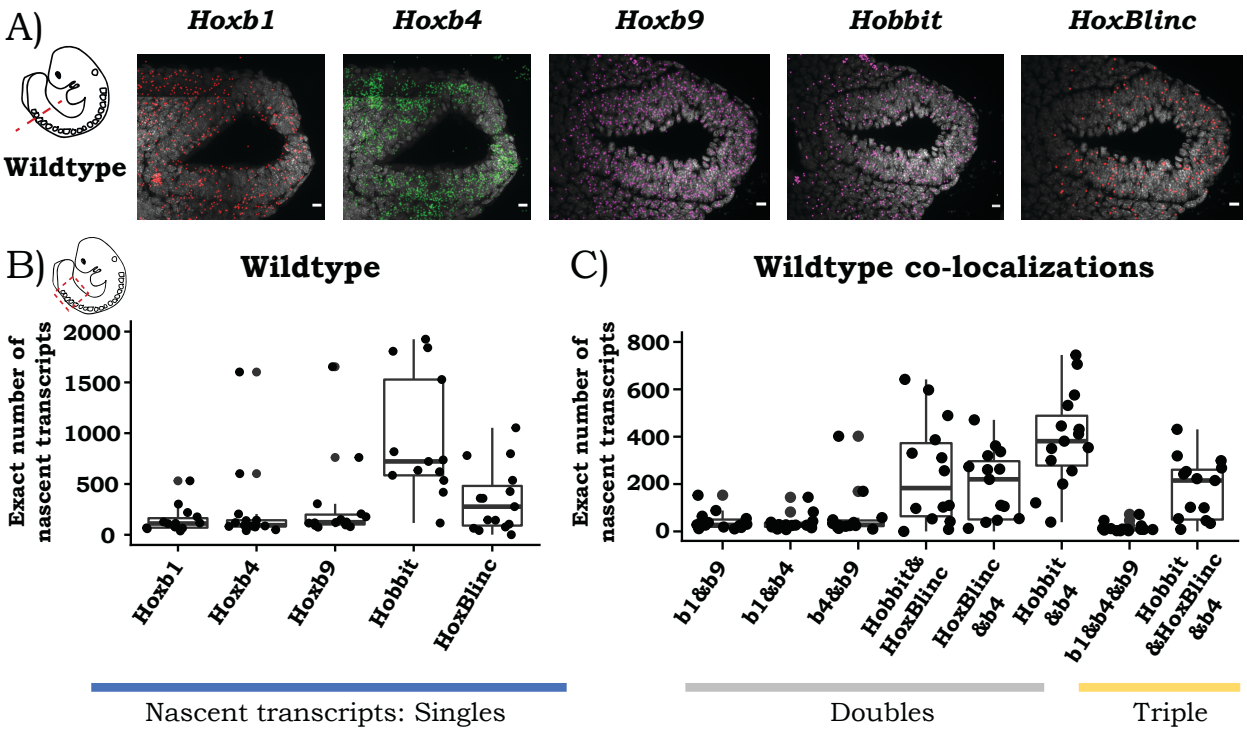


Figure 3-17 Wildtype expression profile for exact nascent counts in neural tube of the tail

Boxplots for average wildtype levels of nascent transcripts are depicted. Panel (A) shows images that have been processed with spot fitting on python to increase size of DL processed nascent spots. The graphs show exact nascent counts for single transcripts (B) and the double and triple co-localization of transcripts (C). Box plots show average

values and the box spans first to third quartiles with whiskers that extend from the smallest to largest values, n=9 -12 tail sections from one or two embryos in case of non-coding RNAs. Statistical significance calculated using non-parametric Wilcoxon rank-sum test. Scale bar - 10microns.

3.8 Trends of expression for *Hoxb* transcripts in the adjacent somites are similar to trends in the neural tube

Adjacent to the neural tube are mesodermal structures called somites. The expression of coding and non-coding *Hoxb* transcripts for the adjacent somites were also quantified. On average for the tail sections, the levels of transcripts are almost half as that in the neural tube. However, the trends of expression and co-localizations are very similar to the neural tube (Figure 3-18, A-B). This suggests signaling patterns or regulatory mechanisms that dictate transcription of *Hoxb* genes are very similar in these near adjacent tissues along the embryonic axis.

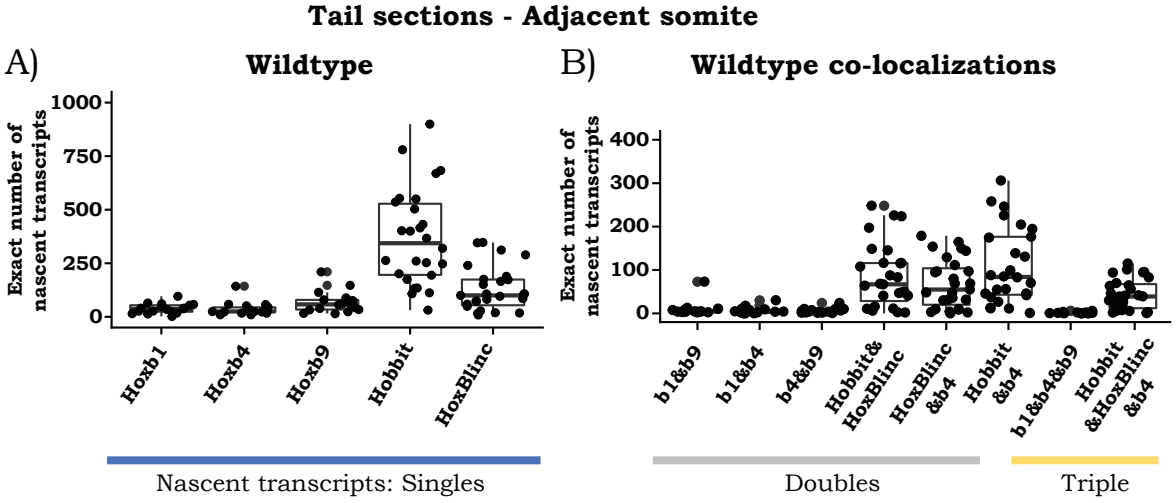


Figure 3-18 Wildtype expression profile for exact nascent counts in adjacent somites of the tail

Boxplots show average wildtype levels of nascent transcripts in the adjacent somites. The plots show exact nascent counts for single transcripts (A) and the double and triple co-localization of transcripts (B). Box plots show average values and the box spans first to third quartiles with whiskers that extend from the smallest to largest values, n=7-9 tissue sections for up to 2 adjacent somites marked for each section. Statistical significance calculated using non-parametric Wilcoxon rank-sum test.

3.9 Each *Hoxb* gene has its own pattern of expression in different tissues along the embryonic axis

For wildtype embryos, a thorough analysis of *Hoxb* nascent transcripts along the whole embryonic axis was done (Figure 3-19, A & C). Several ROIs were marked, and the nascent transcripts in each were measured. In the head, the neural tube of the forebrain and roughly the hindbrain region were measured along with brachial arches (Figure 3-19, B). In the mid body sections, the neural tube, adjacent somites, and gut ROIs were measured (Figure 3-19, C). The boxplots of nascent transcripts averaged for each ROI show different trends of expression in the different tissues (B & D). These boxplots are averages for each section, and depict the trends observed when comparing transcripts along axial plane for the whole embryo in Figure 3-13. Each *Hoxb* gene has its own pattern of expression along the embryonic axis, while there are instances of similar expression profiles observed in some ROIs.

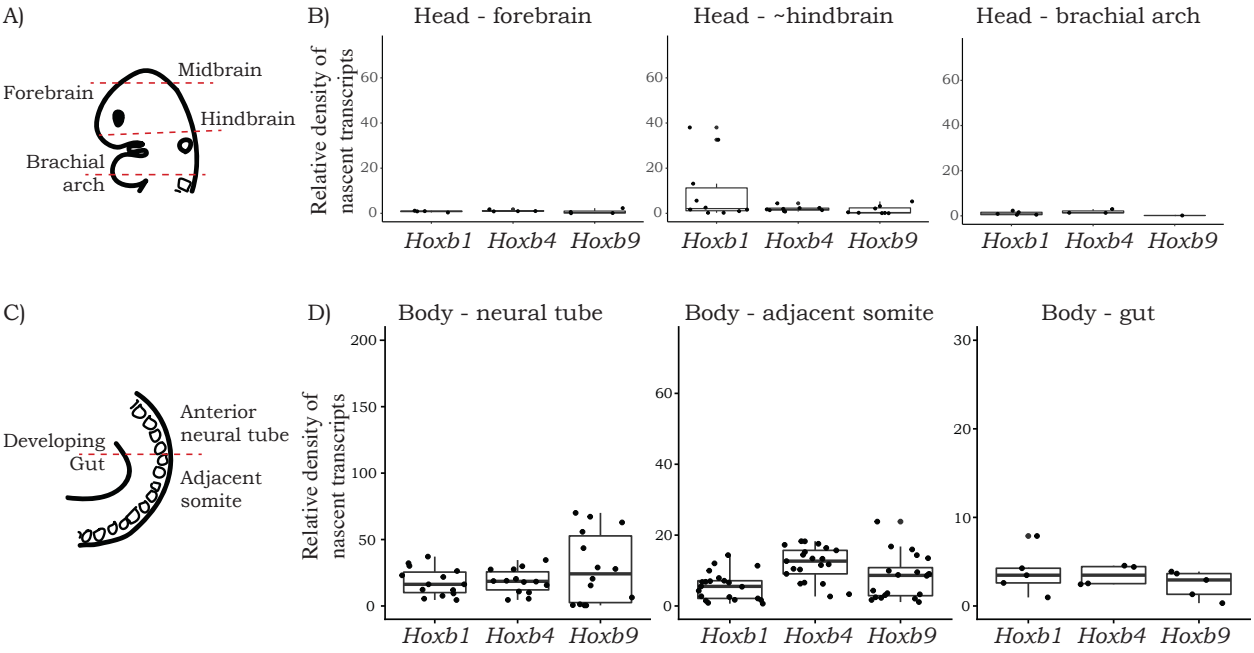


Figure 3-19 Wildtype expression profile in regions from the head and body sections

The boxplots show the relative density of nascent transcripts in the different tissues from regions of the head and mid sections (A & C). In the head, neural tube regions have been divided into forebrain and roughly location of the hindbrain and along with brachial arches the average nascent transcripts are plotted (B). From the body sections, the neural tube, adjacent somites, and gut regions were analyzed and plotted (D). Box plots show average values and the box spans first to third quartiles with whiskers that extend from the smallest to largest values, n=3-9 tissue sections averaged for each corresponding region. Statistical significance calculated using non-parametric Wilcoxon rank-sum test.

3.10 Levels of *Hoxb* nascent transcripts correlate with levels of processed transcripts

With smFISH snapshots in time for the number of nascent transcripts for *Hoxb* genes that are being made has been captured. In an attempt to better understand the transcriptional rate of *Hoxb* transcripts, these nascent transcripts levels can be correlated with the mRNA levels of transcripts. To achieve this, bulk RNA seq analysis of wildtype 9.5dpc embryos was performed. Three distinct regions through the embryo were sequenced; the r4 region where *Hoxb* expression is present in a stripe (crude cut of region below the otic vesicle and above the brachial arches), the trunk region (crude cut of embryo from somite 1 to roughly 10/11), and the tail region (crude cut at 4 somites level above pre-somatic mesoderm). Bulk RNA was extracted from these regions and sequenced. The reads were processed and further analyzed.

Figure 3-20 shows the expression of coding *Hoxb* genes and the non-coding RNAs that were observed with the RNA-seq analysis. It was observed that the trends of mRNA expression for the coding genes do corelate with their relative intensity of nascent transcription. In the r4 sample, expression of only *Hoxb1* is observed, while in trunk and tail sections there are higher levels of posterior *Hoxb4* and *Hoxb9. Hoxb1* levels increase in expression again in the tail. The number of nascent transcripts for non-coding RNAs *Hobbit* and *HoxBlinc* while comparable to those of coding genes by smFISH analysis, appear to have much lower levels of expression with the bulk RNAseq. This discrepancy between nascent transcripts by smFISH and bulk RNA levels by RNA-seq potentially indicates a higher turnover of these non-coding RNA transcripts - their transcriptional activation and RNA degradation.

One important caveat in this comparison is that RNAseq is bulk with tissues including the neural tube and its adjacent mesodermal structures, while the smFISH is focused on specific regions such as the neural tube. These differences in non-coding RNAs could hence either be that 1) they have a high turnover, so at any given time their expression through RNA seq will be low and their nascent levels are comparable to coding genes. Or 2) they are expressed highly in specific tissues (such as neural tube), but not expressed or very lowly expressed in other tissues. This could result in the average expression being very low when comparing bulk data.

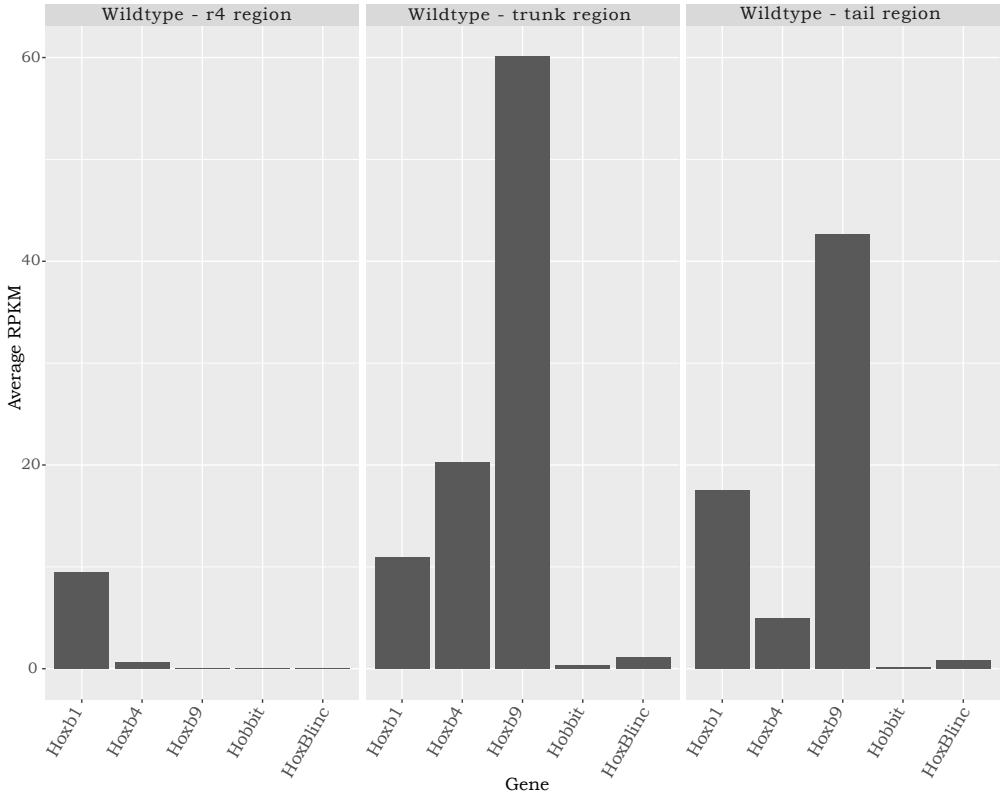


Figure 3-20 Graphs of average RPKM values of *Hoxb* transcripts

The graphs show the average RPKM values of coding and non-coding *Hoxb* transcripts in three different regions of the embryo; the R4 (hindbrain), the trunk, and the tail (n=3 replicates, each replicate a separate single embryo).

3.11 Differential gene expression in wildtype mouse embryos in response to RA induction

We can induce embryos with RA, by giving RA gavage to pregnant females, to track how *Hox* gene expression changes in response to increased levels of RA.

3.11.1 RA induction has different effect on nascent transcripts of adjacent *Hoxb* genes

As the three RARE of interest are located within *Hoxb4-b5* intergenic region or flanking *Hoxb4*, I focused on looking at transcript levels for coding *Hoxb4*, *Hoxb5* along with non-coding RNA *Hobbit* by smFISH. While the number of coding *Hoxb5* nascent transcripts increases in response to RA induction, there isn't much difference in the density of nascent transcripts for *Hoxb4*. What is striking to see is that the nascent intensity of *Hobbit* lncRNA significantly reduces in this section of the neural tube in response to RA induction. This change in *Hobbit* is intriguing because we have previously seen (De Kumar et al., 2015) that *Hobbit* expression increases in response to RA, similar to coding *Hoxb* coding genes during RA differentiation of ES cells. Also, that *Hobbit* nascent transcripts typically increase in the RARE mutants. Hence, this decrease with RA gavage points towards complex activation and repression of *Hobbit*, maybe in an axial level specific manner, that is mediated through RA signaling.

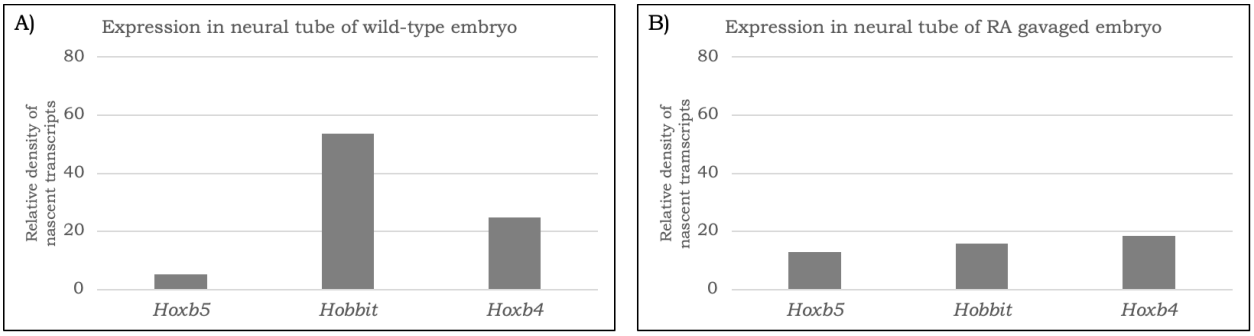


Figure 3-21 Relative expression of transcripts in wildtype and RA induced embryos

Graphs show expression of nascent transcripts for genes in the middle of the *Hoxb* cluster for wildtype uninduced (A) and RA induced samples (B). The graphs of relative density of nascent transcripts are plotted for the coding *Hoxb4* and *Hoxb5*, along with non-coding *Hobbit*.

3.11.2 RA induction does not appear to effect levels of processed transcripts

To better understand transcriptional dynamics in response to RA induction, the r4, trunk, and tail tissue samples were also analyzed by RNA sequencing. While the nascent transcript numbers of

Hox coding genes changed in response to RA, there was mostly no significant difference in the mRNA levels of the coding genes upon induction. While many genes remain same between the uninduced and RA induced conditions, there are some genes differentially expressed between wildtype and RA induced samples. They are depicted by tissue sections below.

In the **r4 region** of the embryo (Figure 3-22), genes that are upregulated for *Hoxb* cluster include *Hoxb4* and *Hoxb5*, and an RA receptor *RARb*. There is also an upregulation of non-coding RNAs such as *Hotairm1* and *Halr* which are present in different *Hox* clusters.

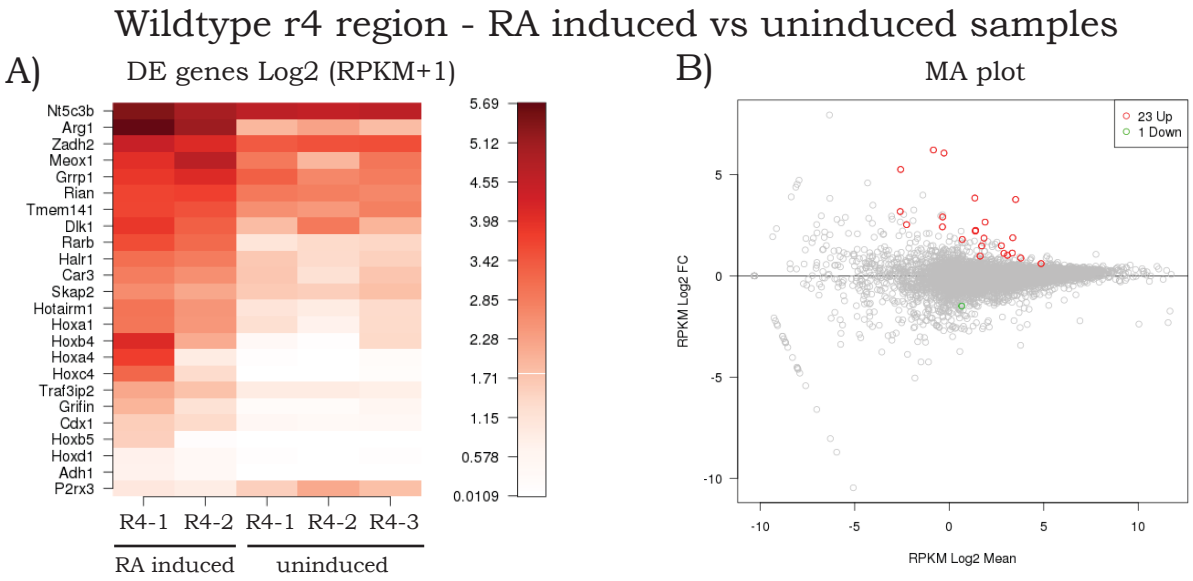


Figure 3-22 Differential gene expression in r4 region

Panel (A) shows a heatmap of differentially expressed (DE) genes in the R4 region between RA induced and wildtype embryos (n=2-3 replicates, each replicate a single embryo). A corresponding MA plot (B) is shown to depict relative levels of genes that change relative to the ones that stay the same. Red circles indicate genes that are differentially upregulated while green circles indicate genes differentially downregulated. Differentially expressed genes were calculate through edgeR analysis, and having a Log2 fold change along with p-value < 0.05.

In the **trunk region** of the embryo (Figure 3-23), there are only five differentially expressed genes. There appear to be no *Hox* genes that are differentially expressed in response to RA induction in this region.

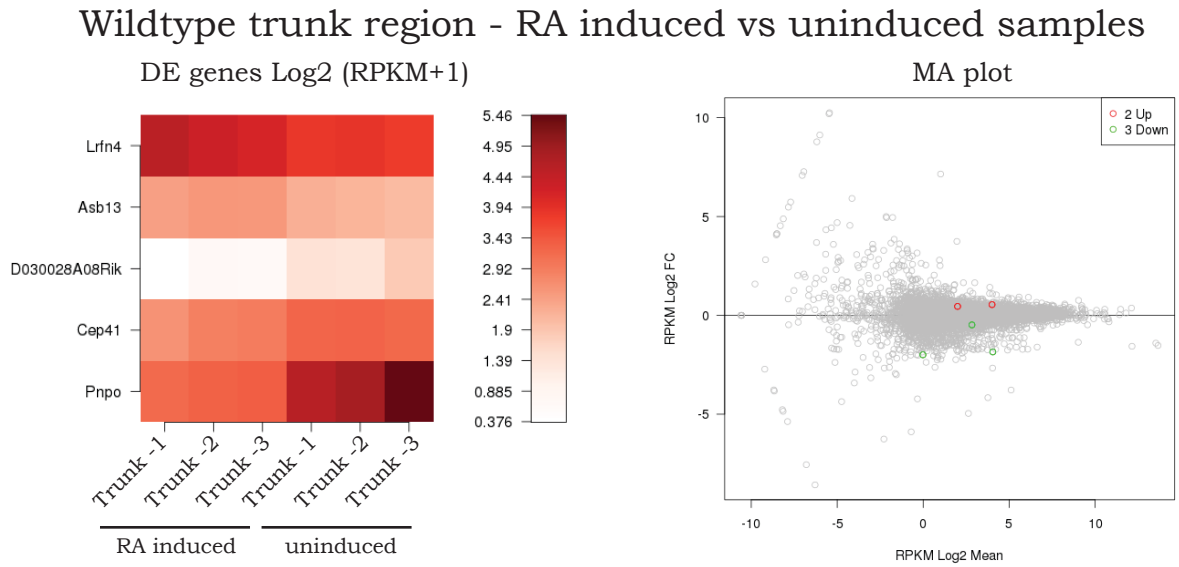


Figure 3-23 Differential gene expression in trunk region

Panel (A) shows a heatmap of differentially expressed (DE) genes in the Trunk region between RA induced and wildtype embryos (n=3 replicates, each replicate a single embryo). A corresponding MA plot (B) is shown to depict relative levels of genes that change relative to the ones that stay the same. Red circles indicate genes that are differentially upregulated while green circles indicate genes differentially downregulated. Differentially expressed genes were calculate through edgeR analysis, and having a Log2 fold change along with p-value < 0.05.

In tail region of the embryo (Figure 3-24), there is again like the trunk no significant differential expression of the coding *Hoxb* genes. But similar to the r4 region, there is an upregulation of *Cdx1* and *RARb* in the RA induced samples.

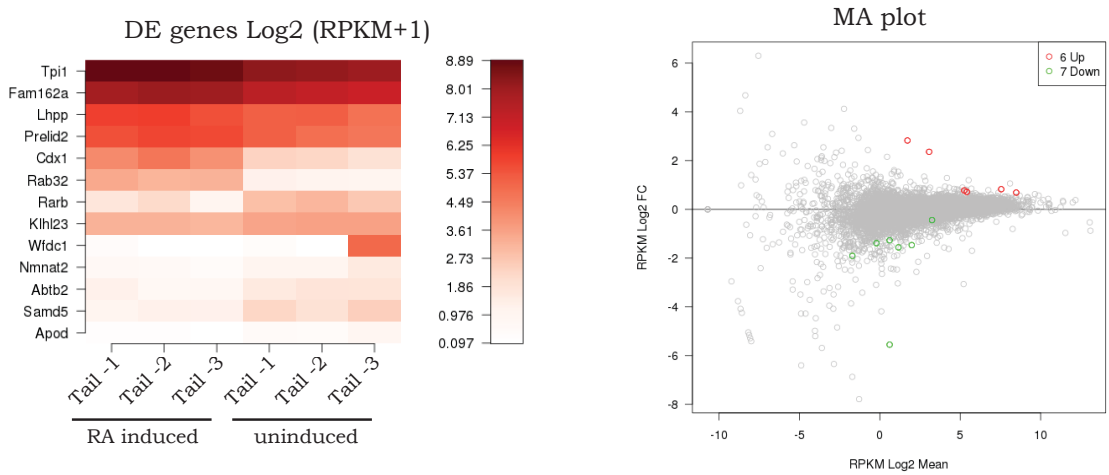


Figure 3-24 Differential gene expression in tail region

Panel (A) shows a heatmap of differentially expressed (DE) genes in the Tail region between RA induced and wildtype embryos (n=3 replicates, each replicate a single embryo). A corresponding MA plot (B) is shown to depict relative levels of genes that change relative to the ones that stay the same. Red circles indicate genes that are differentially upregulated while green circles indicate genes differentially downregulated. Differentially expressed genes were calculate through edgeR analysis, and having a Log2 fold change along with p-value < 0.05.

A Venn diagram was made to compare genes that are changing across the whole embryo in response to RA. The major changes as previously shown are in the R4 region. There are 2 genes (*Cdx1* and *RARb*) that are commonly differentially expressed for r4 and tail regions, while all differentially expressed genes differ slightly for each region.

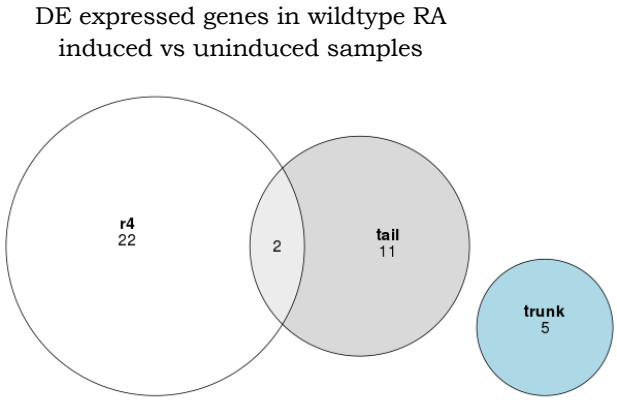


Figure 3-25 Venn diagram for differentially expressed genes

Venn diagram made from differentially expressed genes for each region – r4, trunk, and tail. Then these genes were compared in each region to see which genes are similarly having altered expression in these regions in response to RA induction.

Overall, there isn't much difference observed with RA induction in these tissue segments. This indicates that there are either redundant mechanism that ensure mRNA levels of coding *Hox* genes remains same despite changes in transcriptional bursting (evidenced by increased nascent transcripts), or that bulk RNA seq which includes tissue adjacent to neural tube is not able to detect subtle changes occurring at the mRNA levels for coding *Hox* genes in the neural tube.

3.12 Discussion

In this chapter, I demonstrated that smFISH technique can be optimized to detect nascent transcripts for *Hoxb* genes *in vivo* in the developing neural tube of an embryo. To analyze nascent transcripts in an unbiased and high throughput manner, a DL approach was applied to

count nascent transcripts over the whole tissue sections. As a result, expression levels of *Hoxb* nascent transcripts in wildtype embryos was established. The main focus was on the neural tube of the developing embryo. It was observed that with smFISH expression of nascent transcripts along the axial levels in the embryo correlated with known expression profiles of processed transcripts for coding *Hoxb* genes.

Further, with the single cell resolution obtained with smFISH, some dorsal-ventral expression patterns in the neural were also detected in specific mid body sections. Detailed analysis for coding *Hoxb* nascent transcripts revealed that each gene in the cluster appears to be independently regulated and under different regulatory control, evidenced by a significantly reduced number of co-localized transcripts compared to single transcripts. These trends of expression for coding genes in the neural tube also appear to be true in the adjacent somites, suggesting that loop conformations or enhancer-promoter interactions are similar in both tissues. The non-coding RNAs within the cluster are also differentially expressed in the neural tube, and similar to coding *Hoxb* genes also have distinct expression profiles along the embryonic axis.

To get an idea of transcriptional dynamics within the cluster, bulk RNA-seq was also performed for three regions in the embryo – the r4, trunk, and tail. The levels of nascent transcripts in these regions for the coding genes correlated well with levels of processed transcripts through RNA-seq. However, while the levels of nascent non-coding RNA transcripts were comparable to coding genes in smFISH analysis, their levels of expression in bulk RNA-seq are much lower in comparison to coding genes. These indicate that there are different rates of transcription and/or turnover of transcripts for coding and non-coding RNAs within the cluster, further highlighting that genes within the cluster are under different regulatory control to activate transcription.

Chapter 4: Characterization of nascent *Hoxb* transcripts in RARE mutants

4.1 RARE mutants differentially regulate *Hoxb* coding and non-coding transcripts

Characterization of the *Hoxb* nascent transcripts in wildtype samples revealed that each gene likely has independent mechanism that activate their transcription. To understand how RAREs incorporate RA signaling to transcriptionally activate *Hoxb* genes, simultaneously or independently, the nascent transcript levels and co-localization of the *Hoxb* coding and non-coding transcripts in the RARE mutants were compared against wildtype. In order to achieve this for each RARE, the DR5 binding sites of the enhancers were mutated to restriction enzyme sites (e.g., ecoRI in B4U RARE mutant) that would prevent proper binding of Retinoic Acid Receptors (RARs) (Figure 4-1).

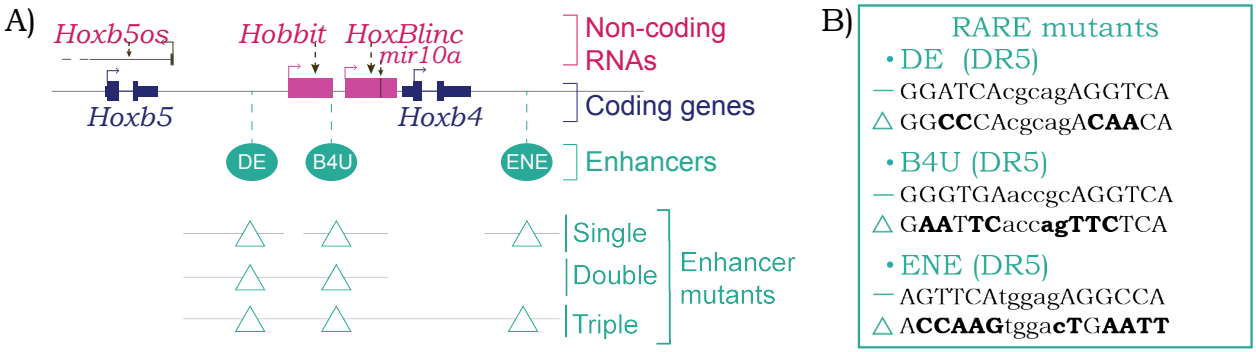


Figure 4-1 Schematic of Hoxb4-Hoxb5 intergenic region with RARE mutants

(A) Shows coding *Hoxb* genes in blue, non-coding RNAs in pink, and RARE enhancers in green. The intergenic region of *Hoxb4* and *b5* is drawn to scale in the genome. The single, double, and triple mutants are depicted with a triangle; there is no deletion in the genomic region, but just a mutation that inactivates the enhancer sequences (B).

In these mutants, no obvious visible morphological phenotypic defects nor any significant behavioral changes were observed. However, changes in expression patterns and in response to RA were noted. From previous studies it is known that proper expression, especially the rostral expansion, of *Hoxb* genes is affected by these cis-regulatory elements (Ahn et al., 2014). In the *Hoxb* BAC reporters, it has been shown that at 9.5dpc the ENE mutated embryos had low *Hoxb4* expression which was restored to normal by 12.5dpc. While *Hoxb9* rostral expansion is gone in the double DE-ENE BAC mutant. There was no change in neural tube observed in single

mutants by *in situ* analysis. Furthermore, a previous study has shown that levels of *Hoxb* processed transcripts in the DE RARE mutant are significantly reduced in hematopoietic stem cells (Qian et al., 2018). These studies highlight the intricacies that might exist in how these RAREs regulate *Hoxb* transcripts during development. For this study, the mutant embryos were stage matched to wildtype embryos, such that all were 9.5dpc embryos and all embryos also had 24 somites. This was to ensure that any transcriptional changes in body segments was not due to size differences associated with differences in developmental age. Similar to the analysis for wildtype embryos, the nascent transcripts for the tail region of the embryo were quantified by DL algorithm, and then the nascent transcripts for representative tail sections were spot enhanced for each mutant embryo for better visualization (Figure 4-2, A-D).

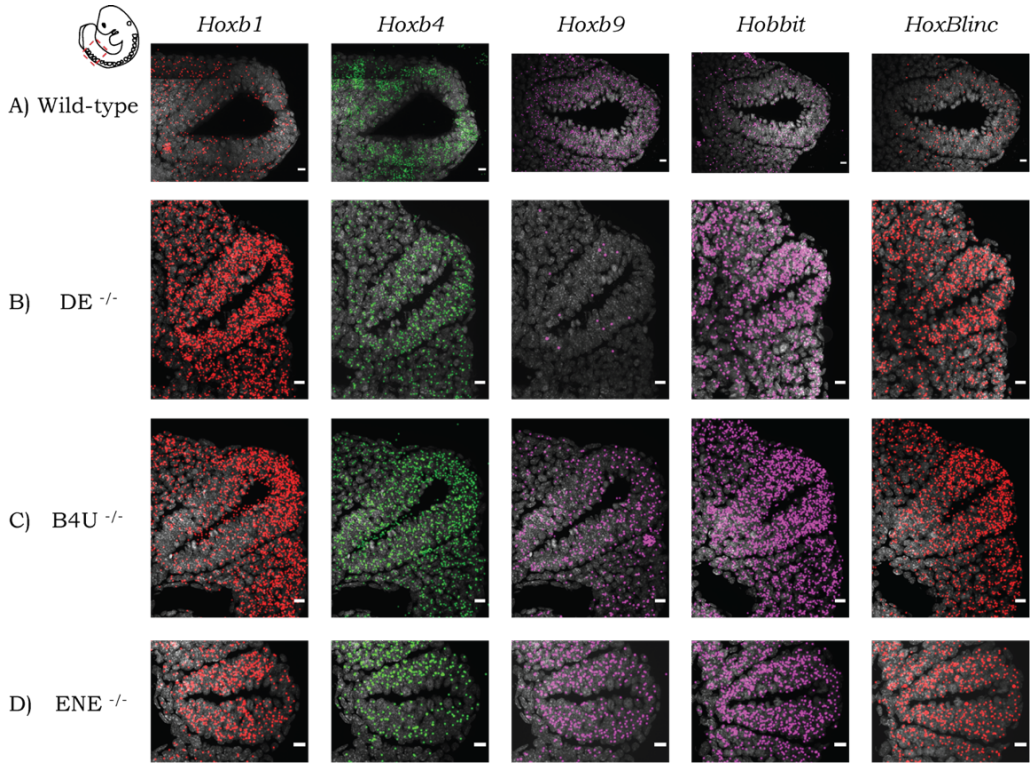


Figure 4-2 Visualization of nascent transcripts for wildtype and RARE mutants

Panels (A-D) show a visualization for the levels of nascent transcripts in wildtype and RARE mutants. Images have been processed with spot fitting on python to increase size of DL processed nascent spots for better visualization. Scale bar – 10 microns.

In all the RARE mutants, for most of the probes (except of *Hoxb9* in the DE mutant), the levels of nascent transcripts increase significant in the tail sections. It is also possible that individual RARE or competition to activate between the RAREs results in some repression, and thus with a mutation we see an increase in the total number of nascent transcripts. These observations also support the idea that if there is some competition between these RAREs, when one RARE is mutated or non-functional, the others can step up to drive the increase in nascent transcription.

To assess changes for the mutant embryos, the levels of exact number nascent transcripts were also averaged for the whole tail region of the embryo and plotted to compare against wildtype embryos (Figure 4-3, A-D). Significant changes in expression levels of many transcripts were observed, suggesting that at a single axial level these RAREs have a different effect on regulating *Hoxb* transcripts. Response of the individual RAREs is discussed in the following sections.

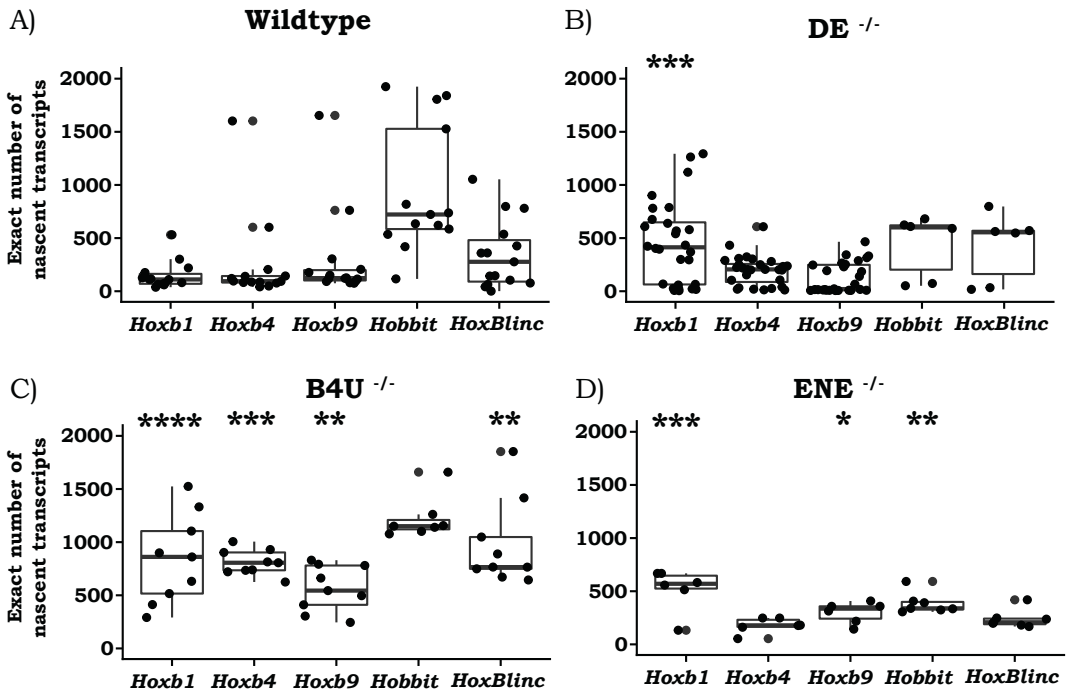


Figure 4-3 Plots of exact nascent transcripts for wildtype and RARE mutants

The boxplots show an average for the tail section for exact levels of nascent transcripts for wildtype (A) and RARE mutants (B-D). Box plots show average values and the box spans first to third quartiles with whiskers that extend from the smallest to largest values. Statistically significant boxplots are depicted by stars, n=7-10 tissue sections from one or two embryo in each condition averaged for each corresponding region. Statistical significance calculated using non-parametric Wilcoxon rank-sum test.

4.2 DE RARE differentially effects anterior and posterior *Hoxb* genes in the tail

In the DE RARE mutant, the levels of nascent transcripts significantly increase in the case of anterior *Hoxb* genes (*Hoxb1* and *Hoxb4*) and lncRNAs *Hobbit* and *HoxBlinc*, while it stays the same or significantly decreases for the posterior *Hoxb9* gene (Figure 4-2, panel B). At specific sections in the tail region, there is almost no expression of *Hoxb9*, while in some sections there is normal expression. This indicates that there is either some repression on the anterior *Hoxb* genes which is released in the DE RARE mutant, or that there is some processing defect or degradation happening for the processed anterior *Hoxb* transcripts. The increase in nascent transcripts thus, could be a compensatory feedback to maintain the levels of total transcripts.

With smFISH analysis, the range in the number of nascent transcripts in the tail sections of the DE mutants significantly increases for the coding *Hoxb1* as well as the lncRNAs *Hobbit* and *HoxBlinc*. This increase in the range of nascent transcripts indicates how multiple or different types of secondary mechanism might be getting turned on in the DE mutant at different axial levels. To maintain normal processed RNA and protein expression for *Hoxb* genes, these mechanisms could be distinct along different axial levels in the embryo. Furthermore, while most nascent transcripts are expressed as single transcripts (indicative of independent regulation of genes in the cluster), there are also slight changes in the co-localizations of the transcripts that are co-regulated in the DE mutant embryos compared to wildtype (Figure 4-4, panel B). There is a general increase in co-localization for all coding transcripts, with the highest increase for the anterior *Hoxb1* and *Hoxb4* transcripts. The co-localization between the lncRNA *Hobbit* and

Hoxb4 slightly decreases compared to wildtype, while other co-localizations for the lncRNAs *Hobbit* and *HoxBlinc* with coding *Hoxb4* increase significantly.

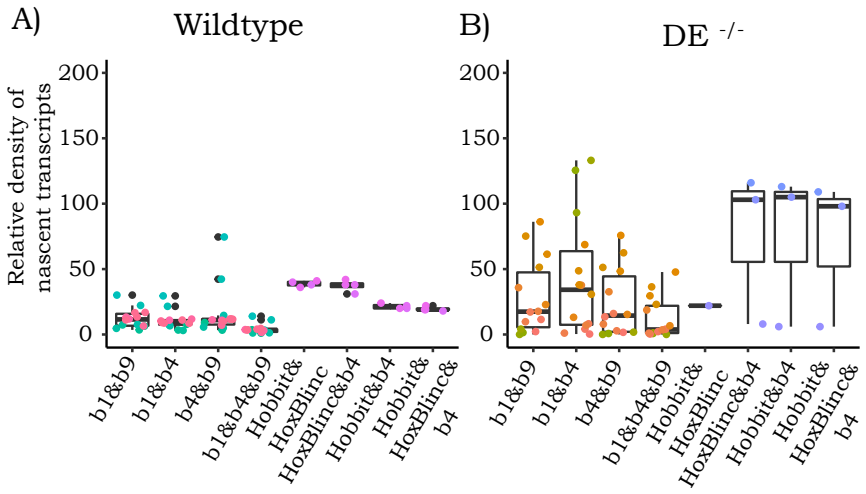


Figure 4-4 Co-localization of nascent transcripts in DE RARE mutants

The boxplots show an average co-localization for the tail sections for coding and non-coding *Hoxb* transcripts. The relative intensity of co-localized nascent transcripts is plotted for wildtype (A) and DE RARE mutants (B).

While some changes in co-localized transcripts could be attributed to the changes in the total number of nascent transcripts. However, the increase in *Hobbit* and *HoxBlinc* transcripts co-localizing with *Hoxb4,* especially when the levels of these non-coding RNAs have slightly decreased compared to wildtype, hints at independent mechanisms regulating the individual coding and non-coding *Hoxb* transcripts within the cluster.

4.3 B4U RARE mediates RA response to predominantly regulate *Hoxb* genes in the tail

The B4U RARE is unique in its positioning compared to DE and ENE RAREs, such that it is positioned at the end of the lncRNA, *Hobbit*. Enhancer activity of this previously uncharacterized RARE was observed through transgenic assays in zebrafish (Figure 4-5). The B4U element appears to be conserved across the different *Hox* clusters as well as conserved across species (Nolte et al., 2019). Like the DE and ENE RARE mutants, B4U mutant embryos

can also achieve homozygosity, and do not appear to have any morphological or significant behavioral phenotypes (while all RAREs do appear to start having breeding issues after a generation under homozygosity).

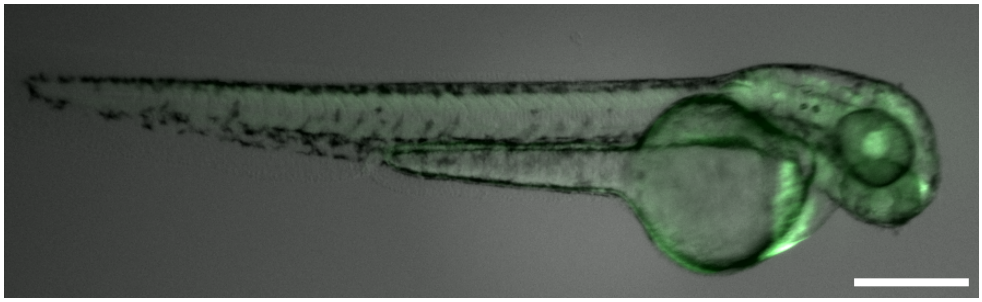


Figure 4-5 Transgenic assay for mouse B4U element in zebrafish

The cloned mouse B4U element was added to Hulk vector to check for enhancer activity of B4U element. The B4U-hulk vector consistently showed expression in and around the hindbrain, pharyngeal arch, and along the notochord regions. Scale bar - 1mm.

4.3.1 RA mediated response is altered in the B4U mutant

To characterize gene expression changes in the B4U mutant compared to wildtype, an RNA-seq was performed. RNA-seq analysis revealed that there are 330 genes that are differentially expressed when comparing tail segments of B4U mutants against wildtype (Table 4-1, B). In the B4U mutants the RA response is observed to be altered, as a higher number of genes are differentially expressed when comparing B4U embryos induced with RA compared to wildtype embryos induced with RA (447 genes differentially expressed; Table 4-1, D). Response to RA being affected is also evident when comparing B4U RA induced embryos with B4U uninduced mutant embryos (438 genes differentially expressed; Table 4-1, C) against wildtype embryos induced with RA and not induced (13 genes differentially expressed; Table 4-1, A). Heatmap to show the changes is shown in Figure 4-6. All these changes in differential gene expressions point towards RA response being altered in the B4U mutants. This is especially evident as there is hardly any difference in wildtype embryos induced with RA, while there are significant

differences in B4U embryos upon RA induction. This effect also seems to predominantly be in the tail, as in the r4 and trunk regions there isn't as much difference observed.

Table 4-1 Differentially expressed (DE) genes for B4U mutant embryo

	Contrast	DE FDR	DE Total	DE Upregulated	DE Downregulated
A)	Wildtype RA induced compared to wildtype uninduced	0.05	13	6	7
B)	B4U uninduced compared to wildtype uninduced	0.05	330	114	216
C)	B4U RA induced compared to B4U uninduced	0.05	438	360	78
D)	B4U RA induced compared to wildtype RA induced	0.05	447	363	84

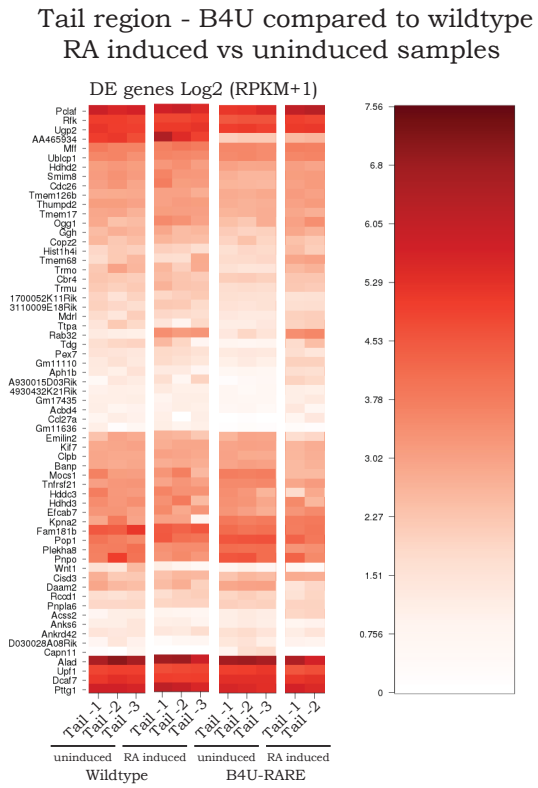


Figure 4-6 Heatmaps of differentially expressed genes for B4U RARE

The heatmap compares wildtype (uninduced and RA induced) samples against B4U mutant (uninduced and RA induced sample).

The greatest number of differentially expressed genes were observed when comparing B4U mutants induced with RA against the wildtype embryos induced with RA. To see what processes where affected which altered the RA mediated response in the mutants, a GO term analysis was performed. The GO-enrichment terms are shown in Table 4-2 for these B4U embryos. Processes like Retinol metabolism, nervous system development, and tube formation highlight the critical role B4U RARE may be playing in transcriptional regulation of the *Hoxb* genes which are linked to these processes.

Table 4-2 GO term analysis for B4U RA induced embryos

GO-term
Apical plasma membrane
Protein homodimerization activity
Endocytic vesicle
Endocytic vesicle
Retinol metabolic process
Tube formation
Regulation of protein stability
Negative regulation of nervous system development
Negative regulation of cell cycle process

4.3.2 B4U element potentially a repressor or is in competition with adjacent RAREs

At a single axial level, i.e., in the tail, for the B4U mutant embryos, it is observed that the increased levels of nascent transcripts is the highest in this mutant compared to other RAREs for both coding and non-coding RNA transcripts. B4U mutant appears to have the greatest impact on anterior *Hoxb1* gene, followed by its most adjacent coding gene *Hoxb4* (Figure 4-3, panel C). The range in expression changes for *Hoxb1* along the axial levels in the tail also increases in the B4U mutant. The range is not as high as in the DE, but it is higher than wildtype and it is also

higher when compared against ENE mutant embryos. For *Hoxb9*, in the B4U mutant we see an opposite effect compared to DE mutant. While in the DE mutant at the same axial levels the expression of *Hoxb9* is completely lost or stays the same, in the B4U mutant the expression of *Hoxb9* nascent transcripts significantly increases. This opposite effect for *Hoxb9* points towards an antagonistic effect of DE and B4U RAREs on this coding gene.

For the non-coding RNAs, *Hobbit* and *HoxBlinc*, there are different trends of expression in the mutants compared to wildtype. Like the coding *Hoxb* genes, the levels of nascent transcripts are highest in the B4U mutant. While in wildtype *HoxBlinc* has higher levels of nascent transcripts, the opposite is true in the B4U mutant where *Hobbit* expression is higher than all coding and *HoxBlinc* lncRNA levels. Since B4U is actually present within the *Hobbit* lncRNA (in the 3' end), it is probable that B4U is repressing *Hobbit*, or that B4U plays a part in some autoregulatory control for *Hobbit*'s expression, or that B4U and *Hobbit* work in the same regulatory pathway to regulate *Hoxb* gene transcripts.

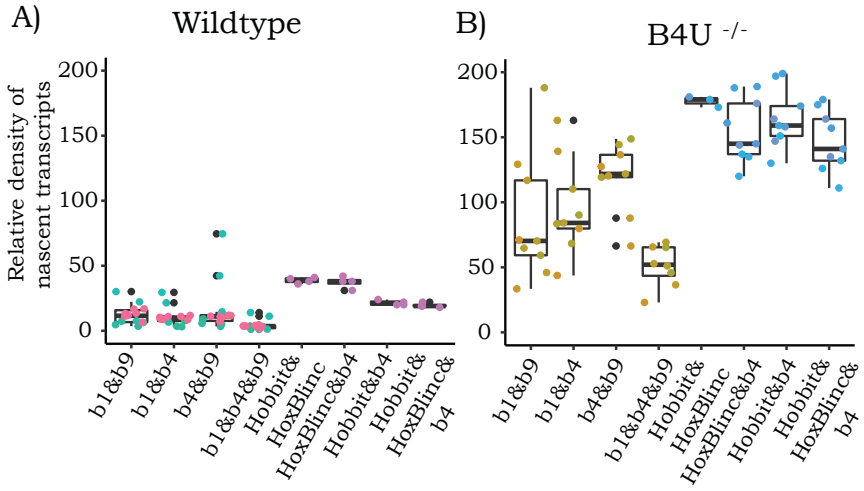


Figure 4-7 Co-localization of nascent transcripts in B4U RARE mutants

The boxplots show an average co-localization for the tail sections for coding and non-coding H*oxb* transcripts. The relative intensity of co-localized nascent transcripts is plotted for wildtype (A) and B4U RARE mutants (B).

Further, upon looking at co-localization of nascent transcripts in the B4U mutant, one can appreciate that they have significantly increased (Figure 4-7, panel B). While in the DE mutant, the anterior *Hoxb* genes were more co-localized, we observe that in the B4U mutant the posterior *Hoxb4* and *Hoxb9* genes have higher co-localizations. Furthermore, while *Hobbit* and *HoxBlinc* had the lowest co-localization in the DE mutant, they appear to have the highest co-regulation in the B4U mutant. Also, while in wildtype embryos *HoxBlinc* seemed to have higher co-localization with *Hoxb4* compared to *Hobbit* and *Hoxb4*, in the B4U mutant it is the opposite with *Hobbit* and *Hoxb4* having higher co-localized transcripts. These changing trends of co-localization ensure us that there exist independent mechanisms that regulate transcription of *Hobbit* lncRNA and *HoxBlinc* lncRNA, and that these lncRNAs are not run-through transcripts being made. Especially since the lncRNA co-localizations with *Hoxb4* change in opposite directions in the two RARE mutants. In the B4U mutant while levels of nascent transcripts significantly increase for anterior *Hoxb* genes, the co-localized or co-regulated transcripts increase for posterior Hoxb genes. This fact highlights how multiple RAREs have specific inputs for activating individual *Hoxb* genes, but also have combinatory inputs for co-regulating *Hoxb* genes.

4.3.3 *Hobbit* lncRNA regulates *Hoxb* genes downstream of B4U element

To decipher the role *Hobbit* lncRNA plays in the transcriptional control mediated by B4U, a knockout mutant was generated. B4U is at the 3' end of the *Hobbit* lncRNA, and a knockout mutant for the promoter and transcriptional start site of *Hobbit* was generated. This mutant has a deletion 250bps upstream and 250bp downstream from the start site of *Hobbit* which was mapped by 5' RACE, and the B4U RARE is maintained at the 3' end of the mutant. While changes in trends of nascent transcripts are like in B4U mutant in the *Hobbit*-ko mutant, for both

the neural tube and the adjacent somite, the actual increase in nascent transcripts is not as high as it was observed in the B4U mutant. The levels of nascent transcripts for anterior *Hoxb1* and *Hoxb4* seem to be impacted more than the posterior *Hoxb9* in *Hobbit*-ko mutant, particularly in the neural tube (Figure 4-8, panel B,D). While there is a slight decrease and the range of *Hobbit* transcripts is lower in the *Hobbit*-ko mutant, there is some expression. This expression could be attributed to the presence of multiple start sites for *Hobbit* lncRNA and needs to be further characterized. *HoxBlinc* appears to significantly increase in the *Hobbit*-ko, further validating that these two non-coding RNAs that have opposite trends, for levels of nascent transcripts, are under different transcriptional control. Overall, the *Hobbit*-ko data, which shows similar trends, just to a lower degree than observed in the B4U mutant, points at *Hobbit* response being downstream of B4U RARE and/or that *Hobbit* is being regulated itself by B4U RARE. Hence, *Hobbit*-ko on its own alone doesn't have the same heightened response as seen in the B4U mutant.

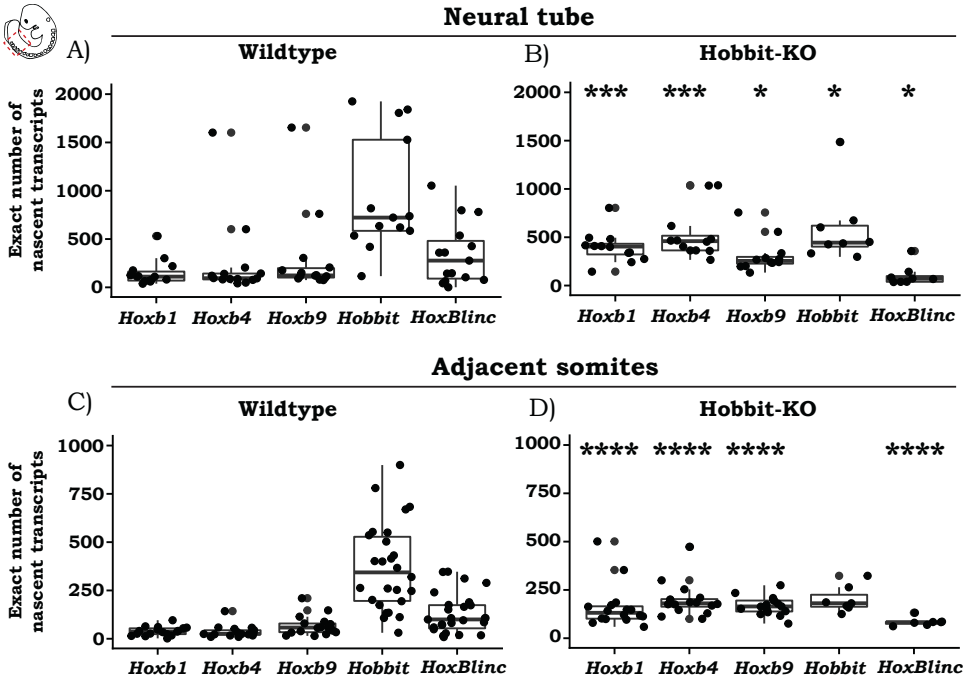


Figure 4-8 Levels of nascent transcripts in the *Hobbit*-ko mutant

The boxplots show an average for the tail section for exact levels of nascent transcripts for wildtype (A,C) and *Hobbit*-ko mutants (B,D) in neural tube and adjacent somites. Box plots show average values and the box spans first to third quartiles with whiskers that extend from the smallest to largest values. Statistically significant boxplots are depicted by stars, n=7-10 tissue sections from one or two embryo in each condition averaged for each corresponding condition or region. Statistical significance calculated using non-parametric Wilcoxon rank-sum test.

4.4 Intermediate transcriptional activation of *Hoxb* transcripts by ENE RARE

Levels of nascent transcripts for coding and non-coding genes in the ENE mutant were observed to be halfway between levels seen in DE and B4U mutants (Figure 4-3, panel D). There is an increase in nascent transcripts again, but the highest change is for the anterior *Hoxb1* followed by the posterior *Hoxb9*. The slight increase in *Hoxb4* nascent transcripts that is observed might not be sufficient to maintain expression of processed *Hoxb4* RNA, as has previously been observed through *in situs* analysis (Ahn et al., 2014) at this embryonic stage. For *Hoxb1*, the range of expression changes in the axial levels in the tail of the embryo is not as wide as it was with the other two RARE mutants. Also, for the lncRNAs there is an increase in nascent transcripts, but it is not as high as with other RAREs. The trend of *Hobbit* nascent transcript levels being higher than *HoxBlinc* is similar to what was observed with the B4U mutant. It appears that changes in the ENE mutant partly lean towards DE and partly towards B4U mutants, with the exception of *Hoxb1* nascent transcripts. The increase in range of expression for *Hoxb1* nascent transcripts along the axis of tail sections, which is observed for both DE and B4U mutants, is not seen with the ENE mutants. This raises the question if ENE RARE is involved in conferring axial level changes to the *Hoxb* transcripts. Looking at co-localizations of our coding genes, it is observed that the anterior genes and posterior *Hoxb* genes are almost equally co-localized. This is also true for localization of the lncRNAs with *Hoxb4*. The anterior *Hoxb1* and posterior *Hoxb9* have the highest co-localization in the ENE mutant as is the co-localization of *Hobbit* and *HoxBlinc*, like what was seen with the B4U mutants (Figure 4-9, panel B). The average co-localizations for coding genes are higher in the ENE mutant compared for DE, but of lncRNAs with *Hoxb4* are

lower than those observed with DE. Comparing trends of exact number of nascent transcripts and transcript co-localization observed in the ENE mutant to the other RAREs, it appears that ENE plays a role intermediate between the other two RAREs. ENE either helps to buffer competition or antagonism between DE and B4U, or that ENE itself also mildly competes with the other two RAREs to activate *Hoxb* genes, and hence the effects observed in this mutant are halfway between those observed in DE and B4U RARE mutants.

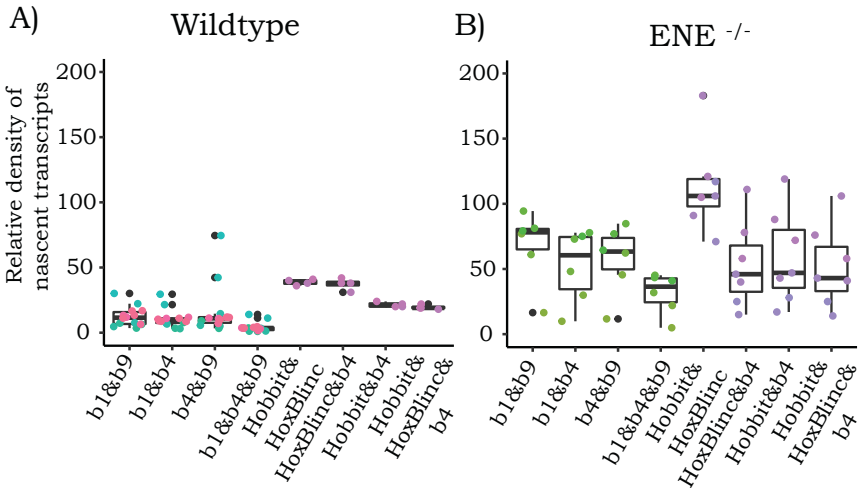


Figure 4-9 Co-localization of nascent transcripts in ENE RARE mutants

The boxplots show an average co-localization for the tail sections for coding and non-coding *Hoxb* transcripts. The relative intensity of co-localized nascent transcripts is plotted for wildtype (A) and ENE RARE mutants (B).

4.5 Competition between enhancers for regulating *Hoxb* nascent transcripts

The individual RARE mutants showed that each RARE has its own input towards how it regulates the initial transcription of the *Hoxb* cluster coding and non-coding genes. To further understand whether these RAREs work together or independently, double and triple RARE mutants were generated in the lab, and nascent transcripts for the *Hoxb* genes were analyzed.

4.5.1 Compound RARE mutant demonstrates balance or competition between elements

In order to decipher whether there exist any joint inputs of two RAREs, DE and B4U, to regulate *Hoxb* genes, smFISH analysis was done for the double DE-B4U RARE mutant. The levels of *Hoxb* nascent transcripts, that increased in both single RARE mutants, not only did not increase but went towards or lower than normal wildtype levels (Figure 4-10, panel B). This return to wildtype levels, at this axial level in the tail, hints at some antagonistic role of DE and B4U RAREs towards maintaining a steady state expression of *Hoxb* genes during development.

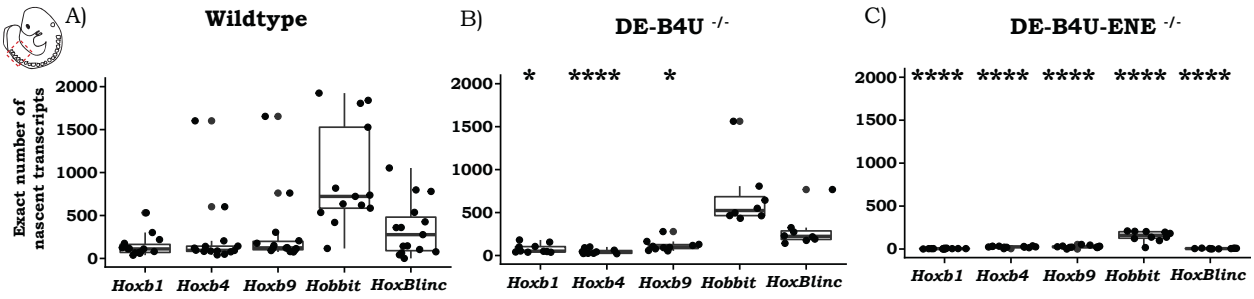


Figure 4-10 Plots of exact nascent transcripts for wildtype, double, and triple RARE mutants

The boxplots show an average for the tail section for exact levels of nascent transcripts for wildtype (A) and RARE mutants (B-C). Box plots show average values and the box spans first to third quartiles with whiskers that extend from the smallest to largest values. Statistically significant boxplots are depicted by stars, n=7-10 tissue sections from one or two embryo in each condition averaged for each corresponding condition. Statistical significance calculated using non-parametric Wilcoxon rank-sum test.

It is important to note that, while levels trends towards normal wildtype, the range of expression levels along the axis of the embryo is definitely impacted in the DE-B4U double mutant. This is evidenced from the reduced levels of nascent transcripts in the highest expressing sections along the embryonic axis compared to wildtype. However, the average expression of nascent transcripts in the tail are very similar to wildtype. The greatest impact to the normal differences in axial expression is to the posterior *Hoxb4* an *Hoxb9* genes. Levels of *Hobbit* and *HoxBlinc* in the double mutant also return to wildtype levels, and their range of expression is also decreased. The DE-B4U double RARE mutant analysis, with regards to exact levels and co-regulated

nascent transcripts, shows that while DE and B4U RAREs have some antagonism they are also working together to fine tune and ensure proper expression of the *Hoxb* genes.

4.5.2 Triple mutant highlights how critical RAREs are for nascent transcription

Previous work in the lab and current smFISH analysis have shown that ENE also has important inputs towards regulating *Hoxb* transcripts. Therefore, to understand if all these RAREs are critical elements or of if they are just redundant elements in the system that activate *Hoxb* genes, triple mutant embryos were generated. Levels of nascent transcripts in the triple DE-ENE-B4U RARE mutants were analyzed (Figure 4-10, panel C). The triple mutants like other RARE mutants also reach homozygosity, and like other mutants the animals also show no significant morphological or behavioral changes. However, the levels of nascent transcripts in the triple mutants were observed to be very low or close to zero for all *Hoxb* transcripts. The exception to this was the *Hobbit* lncRNA which had levels equivalent to wildtype. The patterns of low expression and therefore the non-existence of co-regulatory trends in the triple RARE mutants, underscores that these three RAREs are critical to the initial nascent transcription of the coding genes. However, the mutant animals surviving homozygosity and having no morphological phenotype, suggests that these RAREs are not the sole mechanism dictating or maintaining the expression of *Hoxb* genes. Other secondary mechanisms must be getting triggered in these triple mutants to ensure activation of *Hoxb* genes to ensure proper development.

4.6 Co-activation of *Hoxb* transcripts is actively not favored

In all the mutant conditions, where nascent transcripts were observed through smFISH analysis, the number of single transcripts were significantly higher than the co-localized or co-regulated transcripts. These observed probabilities of co-localizations (Figure 4-11) were compared to probabilities of co-localization by random chance (Figure 4-12). It was seen that in all the cases

the observed probabilities were significantly lower than those by random chance.

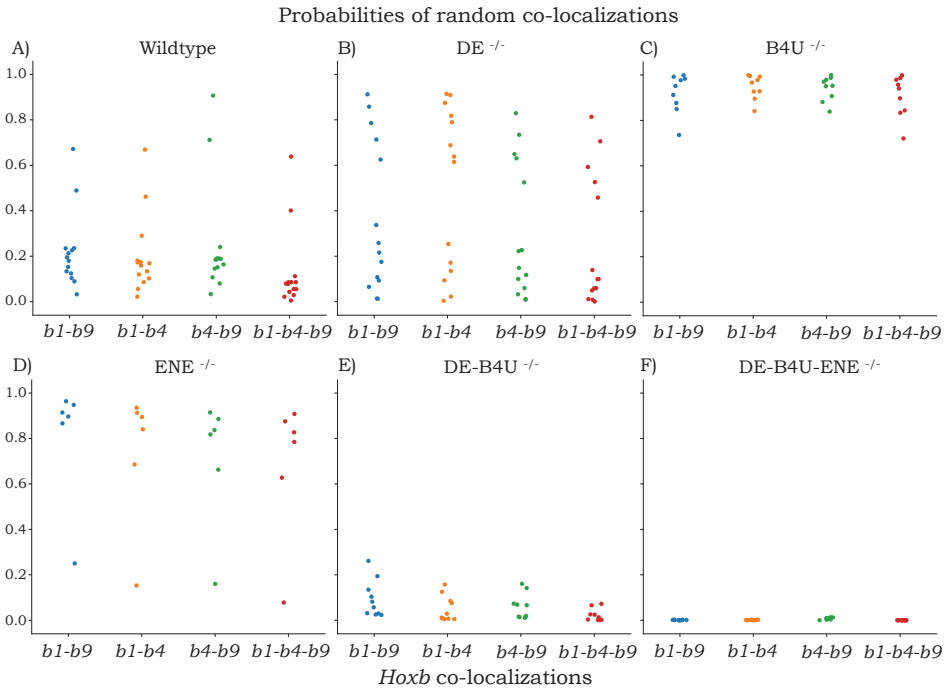


Figure 4-11 Probabilities of random co-localizations

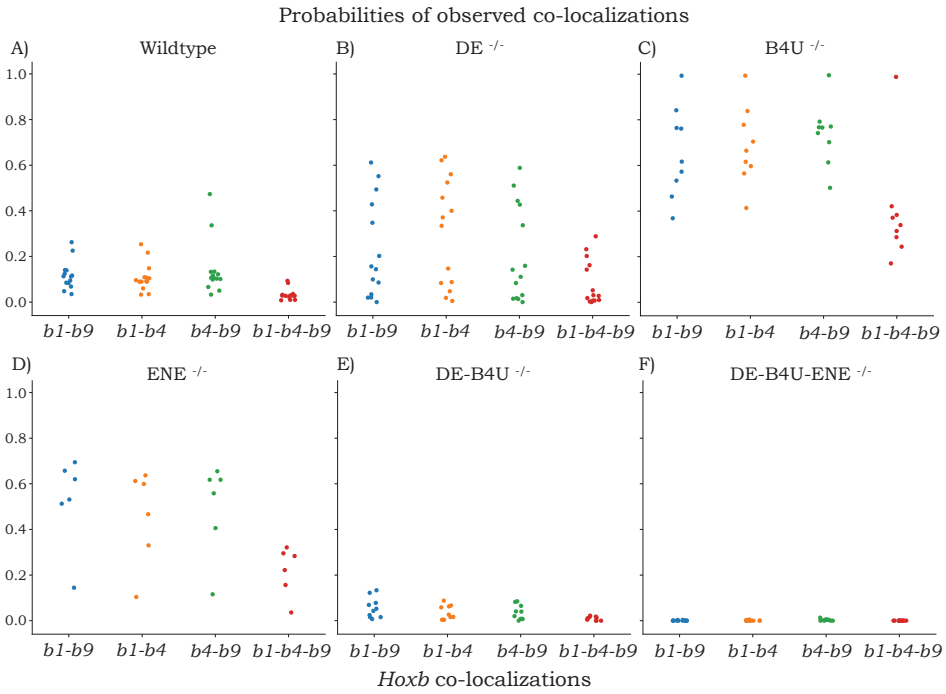


Figure 4-12 Probabilities of observed co-localizations

The probabilities of random co-localization were calculated for wildtype (A) and mutants (B-F) based on the number of single transcripts expressed. The observed co-localizations (Figure 4-12) were lower than calculated for random co-localization (Figure 4-11).

This indicates that there are active mechanisms that are ensuring that *Hoxb* transcripts are not co-localized in the neural tube of the developing tail in mouse embryos. This raises further questions on why co-activation of *Hoxb* genes is actively not favored, and what are the functional consequences of increased co-activation going to be on cellular identities that are being dictated by *Hoxb* genes during development.

4.7 RAREs differentially regulate *Hoxb* transcripts along the embryonic axis

The bulk of the analysis in this study was done in the neural tube of the tail segment of the embryo. However, as whole sections were probed and imaged, to draw comparative conclusions, data for other regions in the sections which were collected in the process were also further processed for analysis. In the somites adjacent to the neural tube in the tail region, it was observed that while levels of expression are half that of the neural tube, the trends of expression changes in all the mutants are very similar to the ones seen in the neural tube (Figure 4-13).

Comparing the neural tube in the tail segment to the neural tube in an anterior axial plane of the embryo, the mid body sections, it was observed that RARE mutants may have different regulatory preferences (Figure 4-14, panel A). Generally, compared to the tail, the levels of nascent transcripts for *Hoxb* genes are much lower anteriorly. Nonetheless, it was observed that while in the tail there was a significant increase in the levels of *Hoxb* nascent transcripts in the B4U mutants, the nascent transcript levels appeared to stay the same as wildtype in the anterior neural tube. Indicating that while B4U RARE played a significant role posteriorly in the tail, it doesn't appear to have as much of a role anteriorly. It could still be competing with DE RARE or acting as a secondary redundant RARE to activate *Hoxb* transcripts, because levels of nascent

transcripts anteriorly in the double mutant are very close to the wildtype levels seen in the tail segments.

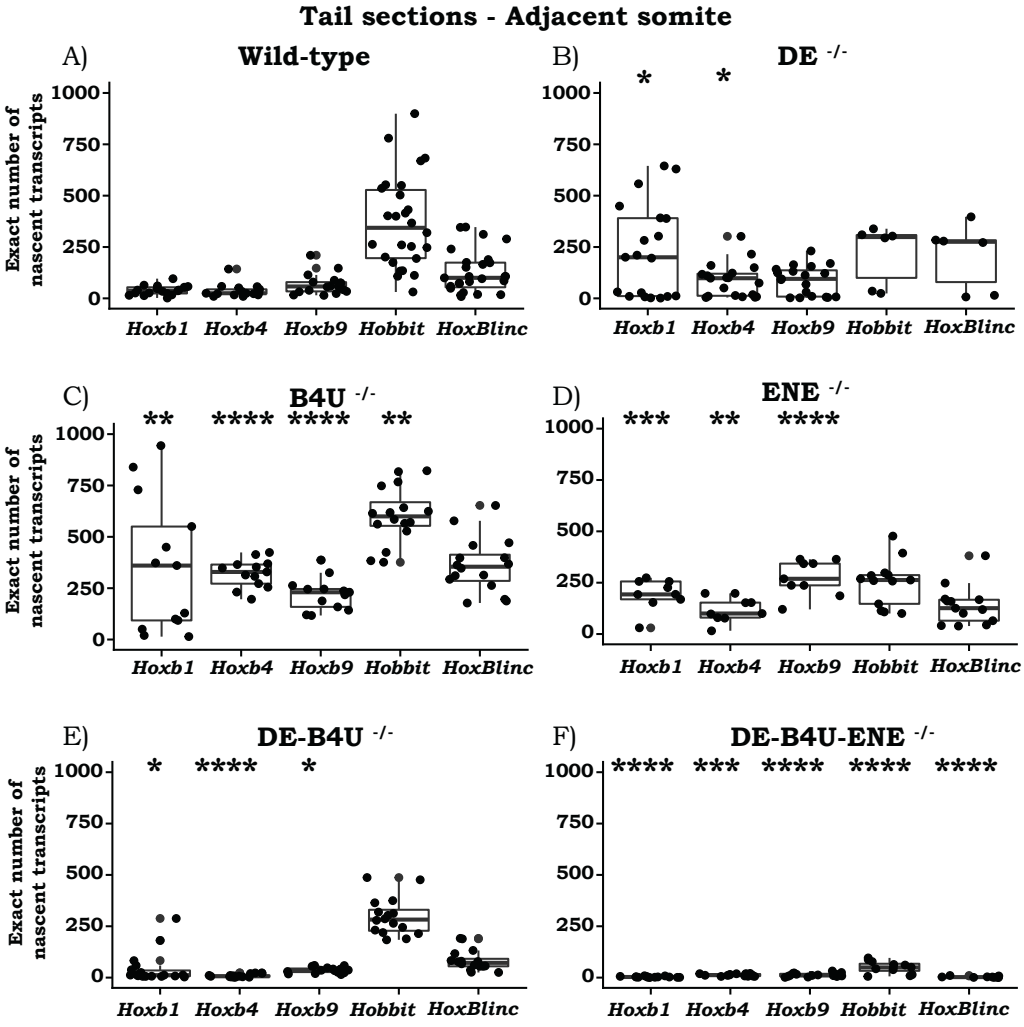


Figure 4-13 Nascent transcripts in the adjacent somites in the tail sections

The boxplots show an average for the tail section for exact levels of nascent transcripts for wildtype (A) and RARE mutants (B-F) in the adjacent somite of the tail sections. Box plots show average values and the box spans first to third quartiles with whiskers that extend from the smallest to largest values. Statistically significant boxplots are depicted by stars, n=7-10 tissue sections from one or two embryo in each condition averaged for each corresponding condition. Statistical significance calculated using non-parametric Wilcoxon rank-sum test.

Further, in this anterior axial plane of the embryo the DE RARE appears to have a significantly greater role compared to its role posteriorly in the tail. The differential changes in the levels of anterior and posterior *Hoxb* gene transcripts in the DE mutant are even stronger at this axial

plane of the embryo. This suggests that while DE plays a greater role anteriorly in the developing embryo, B4U plays a greater role posteriorly. These trends of expression changes at this axis appear to be true for both the neural tube and adjacent somites (Figure 4-14, panels A-B). In the double mutant, while levels appear like wildtype, the levels of *Hoxb4* nascent transcripts appear to be significantly reducing. This points at the need for both DE and B4U RAREs to transcriptionally activate *Hoxb4*. To fully tease out the critical impact of these RAREs, further experiments and analysis need to be done on how non-coding RNAs are getting impacted at this axial plane, and whether in the triple mutant the *Hoxb* transcripts are also severely reduced at this axial plane.

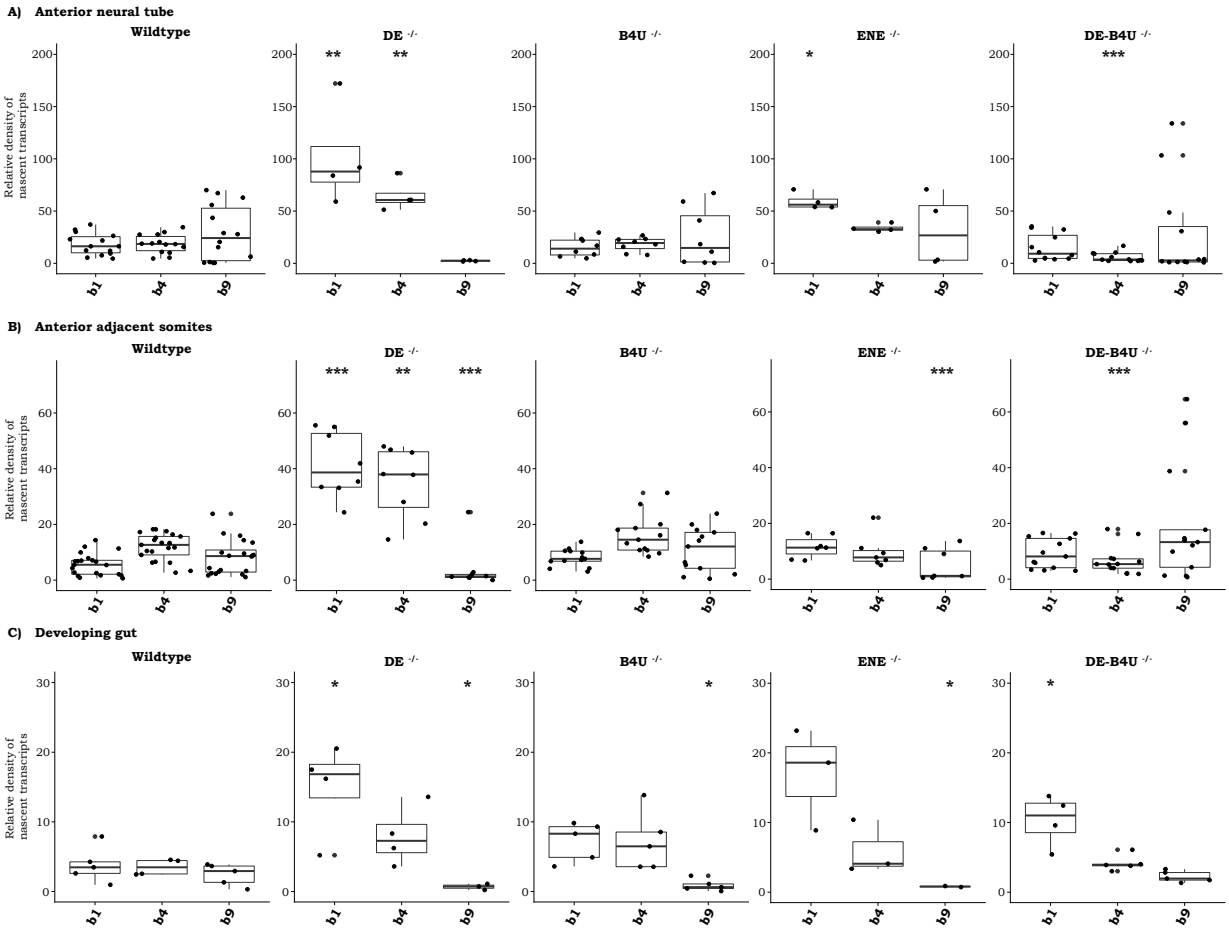


Figure 4-14 Relative density of nascent transcripts in anterior body sections

The boxplots show an average for the relative density of nascent transcripts for wildtype and RARE mutants in the anterior body section for neural tube (A), adjacent somite (B), and developing gut (C). The levels of transcripts are much lower at this axial level compared to at the tail. Statistically significant boxplots are depicted by stars, n=8-11 sections from one or two embryo in each condition averaged for each corresponding region. Statistical significance calculated using non-parametric Wilcoxon rank-sum test.

In the gut tissue of the anterior mid body sections, while expression was really low, there were slight changes seen in the levels of nascent transcripts for *Hoxb1* and *Hoxb9*. In the head region of the embryo, while not significant, there was observed some spurious expression of *Hoxb* nascent transcripts with both DE and B4U RARE mutants (Figure 4-15). This points at how these RAREs might be essential to not only activate nascent transcription, but also to maintain where the *Hoxb* transcripts are not transcriptionally activated.

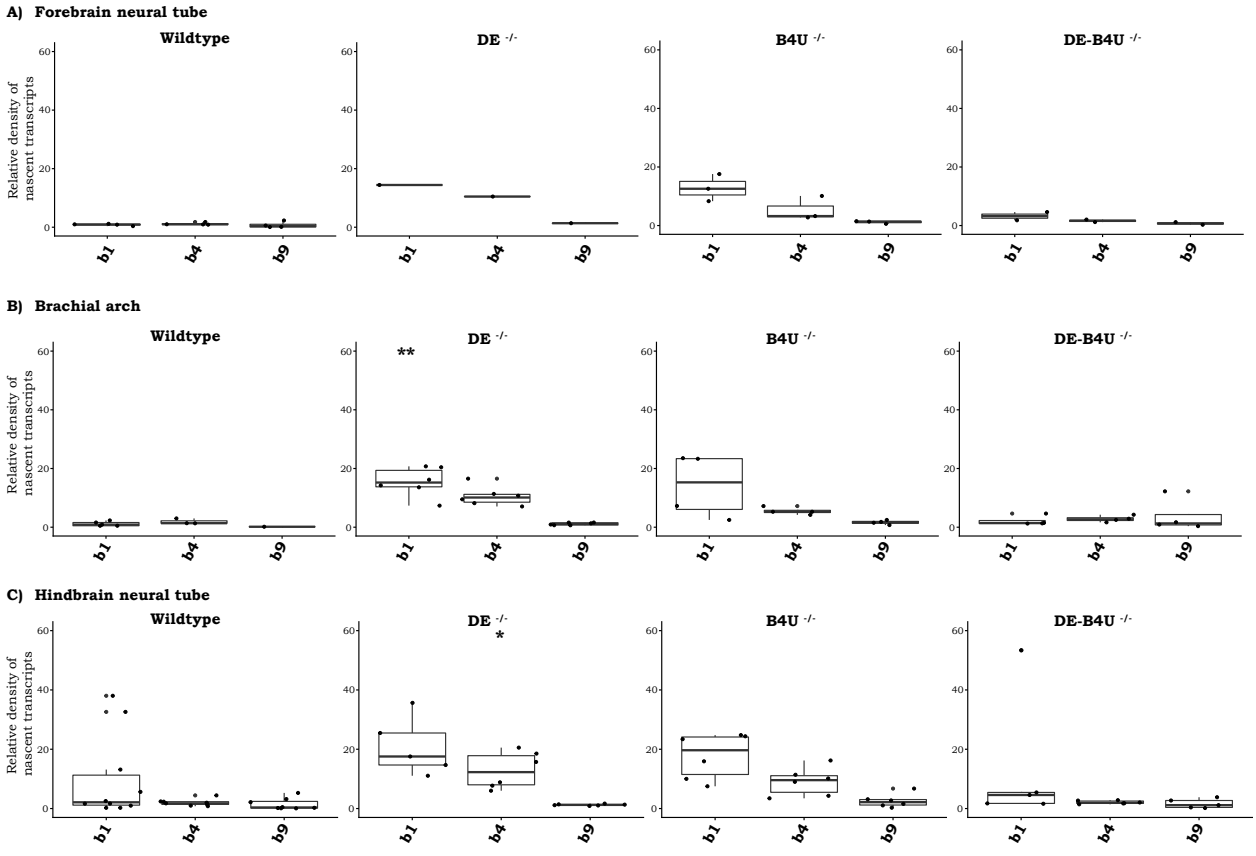


Figure 4-15 Relative intensity of nascent transcripts for head region

The boxplots show an average for the relative intensity of nascent transcripts for wildtype and RARE mutants in the anterior head region (A-C). The neural tube in these sections is divided into forebrain and approximate region of hindbrain. Along with that levels of nascent transcripts are measured for brachial arches. Statistically significant

boxplots are depicted by stars, n=3-5 sections from one or two embryo in each condition averaged for each corresponding region. Statistical significance calculated using non-parametric Wilcoxon rank-sum test.

4.8 Transcriptional activation by RAREs could be mediated though dynamic changes in enhancer-promoter distances

Hoxb nascent transcripts are regulated differentially by RAREs. There appear to be both individual inputs from RAREs to transcriptionally activate single *Hoxb* transcripts and combinatory inputs from RAREs to ensure co-regulation of multiple *Hoxb* genes is not favored. To assess if chromatin changes play a role in transcriptional activation of *Hoxb* genes, wildtype data for co-localized transcripts was spot fitted to infer distances between the gene promoters (Figure 4-16). These distances can then be used to infer chromatin interactions present between enhancers and the genes promoters they activate.

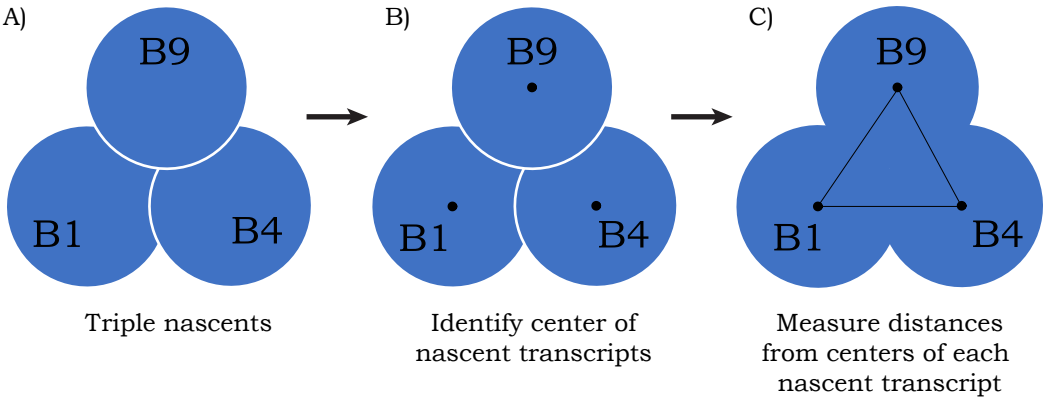


Figure 4-16 Spot fitting triple co-localized transcripts

Schematic to depict how spot fitting analysis was performed on triple co-localized transcripts (A). The center of each transcript was identified (B) and pairwise comparisons were made to get distances for each of the three *Hoxb* transcripts (C).

To shed light on whether RAREs may be functioning by changing chromatin interactions, spot fitting data from co-localized transcripts was also generated for the triple co-localized *Hoxb* transcripts in the individual RARE mutants and compared against wildtype (Figure 4-17, panel A). Spot fitting analysis was performed for each sample by identifying the center of one nascent

transcript (among the triple co-localized) and then measuring the distance between two co-localized spots. Using this pairwise approach, the distances between each of the three co-localized *Hoxb* transcripts was made. This data points towards a dynamic nature of enhancer-promoter contacts, because a range of distances between nascent transcripts of co-localized spots was observed (Figure 4-17 B). The peak distance for each pairwise comparison between *Hoxb* nascent transcripts was plotted for wildtype and each mutant condition (Figure 4-17 C).

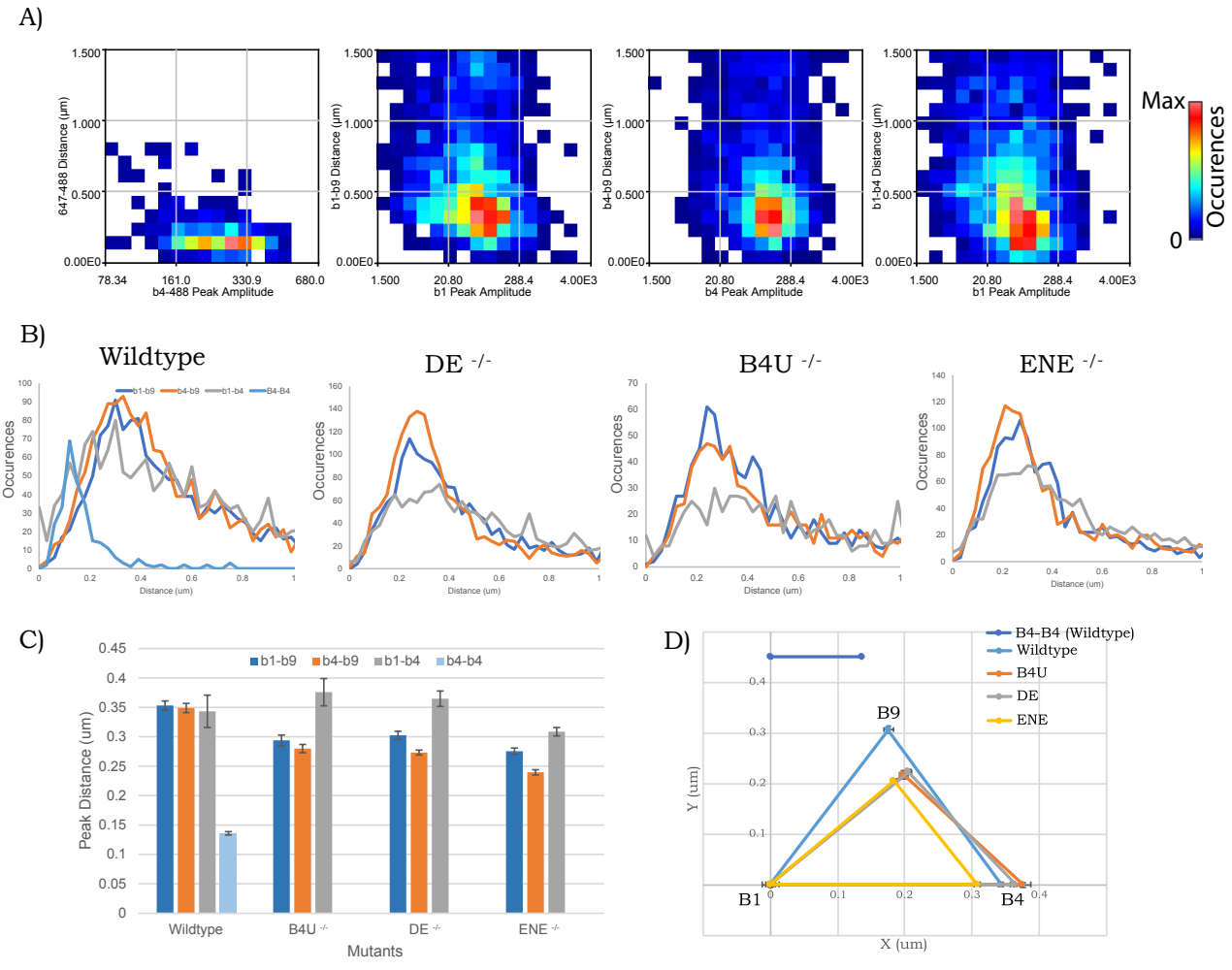


Figure 4-17 Spot fitting data to infer chromatin interactions

Spot fitting data generated by looking at co-localized transcripts in wildtype and RARE mutants. The B4-B4 bar is an internal control with an intron and exon probe to depict the threshold of imaging resolution; distance changes below this threshold cannot be accurately distinguished. Detailed methodology for generation of plots is in the methods.

For better visualization of the changes in distances that were occurring in the mutants a triangle plot was generated (Figure 4-17 D). It was observed that the distances between *Hoxb1-b9* and between *Hoxb4-b9* are decreasing, while there is no significant difference in the distance between *Hoxb1-b4*. Assuming that the *Hoxb* cluster is forming a loop, from this data it can be proposed that the chromatin is getting compacted in the mutants. Taking DE RARE as an example to show a schematic for how *Hoxb* genes could be arranged in a loop, it can be hypothesized that an active RARE maintains an optimal or required distance between promoters to ensure proper transcriptional activity (Figure 4-18, panel A)

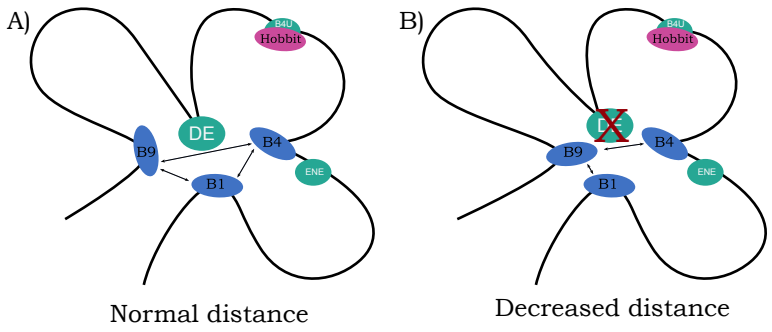


Figure 4-18 Hypothetical looping model to demonstrate distance changes

Schematic to demonstrate how chromatin may be getting compacted in the RARE mutants. DE is shown as an example to show how in the mutants the *Hoxb1* and *Hoxb9* distance and *Hoxb4* and *Hoxb9* distance are decreasing while the *Hoxb1* and *Hoxb4* distance is staying relatively unchanged (B).

A decrease in distance between the assumed promoter region for *Hoxb* genes could be indicative of the *Hoxb* cluster loop getting smaller, which is seen for all the RARE mutants including DE shown in the schematic (Figure 4-18, panel B). This change in loop compaction, or enhancer-promoter distances, could potentially be the reason for changes observed in the RARE mutants for the *Hoxb* nascent transcripts. The decreased distances with inactive RAREs or in the mutants could be hindering normal interactions required between RAREs and *Hoxb* gene promoters to regulate, maintain, or fine-tune the transcriptional activity of *Hoxb* transcripts.

4.9 Discussion

In this chapter, I have shown how the optimized smFISH technique can be applied to functionally characterize the RARE mutants. Changes occurring to the exact number of *Hoxb* nascent transcripts and to their co-localizations in the RARE mutants were compared against the characterized expression in wildtype embryos. The objective was to identify how RAREs transcriptionally activate *Hoxb* genes. One hypothesis was that the enhancer/RAREs exist in a transcriptional hub with all the *Hoxb* genes and can turn on all the *Hoxb* genes simultaneously (Figure 4-19). The second hypothesis was that there are dynamic interactions between the enhancer elements and promoters of individual Hox genes, such that RAREs turn on or activate a single *Hoxb* gene at a time (Figure 4-20, A). In case of these dynamic interactions, one could also imagine there exist conditions where an enhancer is beginning to turn on one gene while moving away from one gene, and thus instances where an enhancer is close enough to activate two or more genes could be captured, as evidenced from double or triple co-localized transcripts (Figure 4-20, B-C).

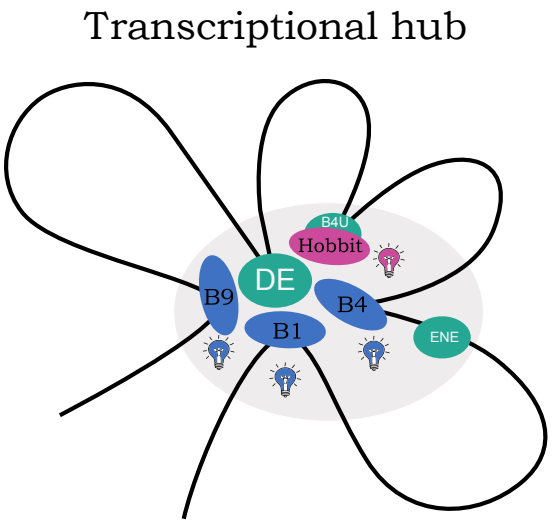


Figure 4-19 Transcriptional hub with simultaneous activation of hox genes

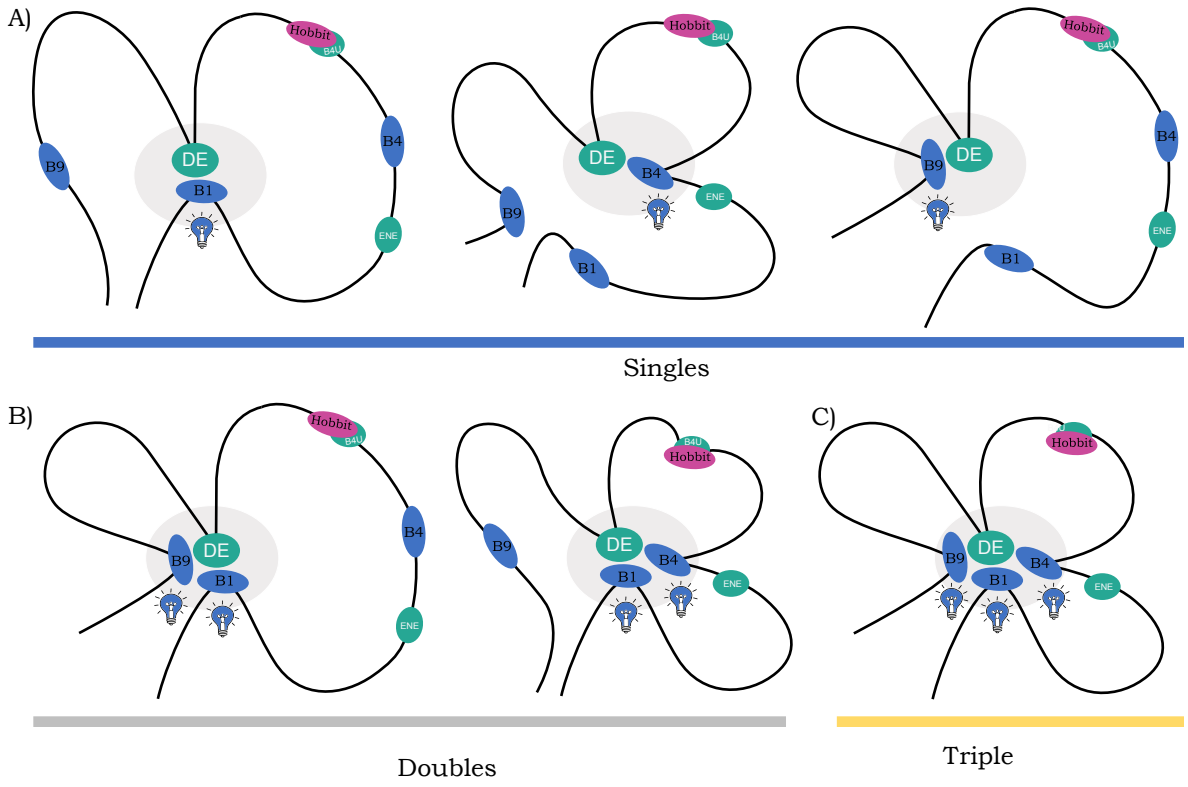


Figure 4-20 Dynamic interactions of RAREs with gene promoters

The looping models in 4-19 and 4-20 are inferred from spot fitting data from Figure 4-17. The RAREs-*Hoxb* gene promoter distances may be changing in the mutants, and that could be affecting the changed levels of nascent transcripts in the mutants.

Through the *Hoxb* nascent transcript analysis, it was observed that the second hypothesis of dynamic interactions is most probably how enhancers elements are functioning to transcriptionally activate *Hox* genes. Further, in single axial level (i.e. the tail region) the single enhancers appeared to show different effects of *Hoxb* nascent transcripts. This indicated how the RAREs do not all have the same impact on *Hoxb* transcription, which pointed towards competition, selectivity, or preference for specific types of enhancer interactions. It was also observed that enhancers played different roles along the embryonic axis, such that while DE seemed to play a greater role anteriorly in the embryo, B4U seemed to play a greater role posteriorly in the embryo. Furthermore, data from the double DE-B4U RARE mutants suggested there is some balance or competition that exists between the elements. While all the RARE

mutants are able to reach homozygosity without showing any obvious morphological or behavioral phenotype, the triple RARE mutants underscored the critical need of these RAREs in transcriptional activation of *Hoxb* transcripts. To infer how chromatin changes at the levels of enhancer-promoter contacts might be changing in the RARE mutants, a spot fitting analysis was performed for triple co-localized transcripts. This analysis allowed for making hypothetical looping models for the dynamics interactions that might exist between RAREs and *Hoxb* gene promoters. It appears that when enhancers are inactivated with mutations, they fail to maintain a specific distance between the *Hoxb* gene promoters, and as a result the distances between *Hoxb* gene promoters decreases. To validate this hypothetical looping model, the need to do a high-resolution chromatin capture experiment arises which can resolve within at least a 15kb region, and that is currently difficult to achieve.

Chapter 5: Discussion

I sought to understand the transcriptional regulation of coding and non-coding RNAs by cis-regulatory elements. The *Hoxb* cluster was chosen as a model to decipher the regulatory interplay between the cis-regulatory RAREs and non-coding RNAs in order to understand how expression of coding *Hoxb* genes was maintained within the neural tube. Previous observations by transgenics assays had shown that these RAREs have enhancer activity, and through use of BAC clones it was known that the rostral expansion of *Hoxb* genes was affected in the neural tube in response to RA. To understand what was happening at the single cell level to the transcription for *Hoxb* genes, CRISPR mutants were generated in the lab for three RAREs – DE, B4U, and ENE. Intriguingly, homozygous viable animals could be achieved with all the RARE mutants, suggesting that any initial changes to the expression of *Hoxb* genes was restored as the embryos developed. To track the dynamics of nascent transcription for both coding and non-coding RNAs within the *Hoxb* cluster, the smFISH technique was optimized which allowed for quantifying expression changes, which cannot yet be fully achieved with other *in situ* techniques, in fixed mouse embryos. A Deep Learning (DL) approach was applied to the images to count nascent transcripts in an unbiased and high throughput manner. The high throughput pipeline enabled establishment of normal *Hoxb* nascent transcriptional events. With focus on the neural tube, the nascent transcripts over whole axial plane of the embryo were observed. The images obtained capture developmental time through the neural tube and highlight how *Hoxb* genes could axially be responding to RA. To understand dynamics between the whole *Hoxb* cluster, analysis was focused on the developing end of the embryo i.e., the tail, where all three *Hoxb* transcripts of interest are actively transcribed.

5.1 Different nascent transcriptional dynamics observed for single RARE mutants

In the single RARE mutants, the levels of nascent transcripts increase. This increase did not appear to be attributed to a change in the number of cells within the neural tube of RARE mutants. While in wildtype usually one nascent transcript per nuclei is observed, in the RARE mutants two nascent transcripts per nuclei are often seen. This suggests that both alleles are being actively transcribed in the mutants. It could be that 1) the turning on and off of the alleles is disrupted or slowed down, hence nascent transcripts from both alleles in the mutants are captured. While in wildtype the dynamic on/off or the switch from one active allele to another is quick enough, allowing for only one nascent transcript at a time to be captured. Or it could be that 2) in wildtype the levels of processed transcripts are maintained, while in the mutants something is dysfunctional with the processed transcripts or they get degraded. In response to the problem or in effort to compensate for some disruption of processed transcripts, both alleles are kept on in the mutants and two nascent transcripts per cell are captured.

An ideal experiment to address this would be to count processed single transcripts, but with the current technique established in this study, the ability to count single transcripts isn't possible. Hence an estimate of transcriptional bursts or the turn-over of transcripts cannot be achieved. In the tissue sections imaged with exon probes, in theory processed single transcripts can be detected. However, there is a lot of cytoplasmic background which makes it harder to separate out single transcripts from background. If the spectra of the image could be separated out, computationally or with added imaging filters, in order to distinguish between cytoplasmic background, labeled probe, and unlabeled probe, it could be incorporated with the DL pipeline. This would increase the power of the study by adding the ability to calculate rates of transcription by comparing nascent transcripts to processed single transcripts. However, while

this is not currently possible, this study focused mostly on nascent transcripts in the nuclei and used as many intron-only probes as possible to reduce even capturing cytoplasmic signal in the images.

Analysis of nascent transcripts of *Hoxb* genes, point towards the dynamic nature of RAREs within the cluster. Different patterns of expression were observed for both coding and non-coding nascent transcripts in the various single and compound RARE mutants. In each of the single mutants, different trends of increase in nascent transcripts for the anterior and posterior *Hoxb* genes was seen suggesting these RAREs have individual inputs towards *Hoxb* gene regulation. In the future it would be interesting to have a methodology that captures the movements of these enhancers in response to transcriptional activation or repression of *Hoxb* genes. Ability to visualize dynamic movements of these RARE would help to visibly demonstrate if there is competition between enhancers for proximity to promoters or if there is a coordinated manner in which these RAREs function together along the embryonic axis to transcriptionally activate *Hoxb* genes.

5.2 Enhancer RNAs may be playing a role towards dynamic nascent transcription

There is increasing evidence that enhancer regions are transcriptionally active. While earlier it was thought that transcription at enhancers is background noise or the RNAs produced at enhancers are just byproducts of the transcriptional machinery, there are now several studies that have highlighted distinct roles for enhancer RNAs (Reviewed in Arnold et al., 2020). Therefore, it would be interesting, especially in terms of B4U RARE which is present at the 3' end of *Hobbit* lncRNA, to look at transcriptional roles lncRNAs may be playing in conjunction with RAREs to regulate *Hox* genes. LncRNAs can bind to other RNA, proteins, and even regions of DNA, and one could speculate if *Hobbit* within the *Hoxb* cluster can help maintain e.g., the

optimal B4U distance to *Hoxb* gene promoters by direct interaction or through its binding to other proteins. To test this, a ChiRP experiment (Chu et al., 2012), which could detect *Hobbit* lncRNA interactions with DNA regions and proteins can be performed. Comparisons of *Hobbit* interactions in wildtype, B4U mutant, *Hobbit* mutant (with and without scrambled binding sites in the B4U region) might give insights into interactions that exist. Further, it could be analyzed if *Hobbit* interactions effect enhancer-promoter distances and if changes in distances play a role in activating transcription of the *Hoxb* genes.

5.3 Competition or balance observed between activity of critical RAREs

The double mutant in this study hint towards antagonistic effects of DE and B4U RARE mutants because levels of expression for coding genes return to wildtype. The DE-B4U-ENE triple mutant indicates that the three RAREs are critical and must be working together to initiate nascent transcripts, since the levels of all coding nascent transcripts almost drops to zero. Despite this, these RAREs while essential are not the only means of regulating transcription of *Hoxb* genes. This is because, even though at 9.5dpc where there is a molecular phenotype of abnormal transcriptional activity for RARE mutants, the mutants develop to term and exhibit no obvious morphological phenotype. This suggested there are other mechanisms to correct for expression of *Hox* genes to ensure proper HOX proteins levels. However, it might be worthwhile in further studies to look in detail at morphological changes at the tissue or organ level along with detailed behavioral analysis (such as grooming or stress response) that might be present in these RARE mutants. This detailed analysis might reveal defects which could be traced to the disruption of *Hoxb* nascent transcription at 9.5dpc embryonic stage. Further, isolating secondary mechanisms that restore levels of *Hoxb* genes as the embryo develops might provide greater insight into mechanisms that dictate expression of *Hoxb* genes. Performing smFISH and RNA-seq analysis

on later developmental stages in wildtype and RARE mutants might help unveil these secondary mechanisms.

5.4 Transcriptional regulation of *Hox* genes

The arrangement of genes within the *Hox* clusters are correlated with the spatial temporal domains in which the *Hox* genes are activated during development (Dollé et al., 1989; Gaunt et al., 1988). Targeted deletions within the cluster and ectopic insertions of enhancers, have showcased the importance of organization for genes within the cluster (Kmita et al., 2000). However, they have also indicated that transcriptional activation of genes within the cluster is not a passive event (Ahn et al., 2014; Hunt & Krumlauf, 1991; Papalopulu et al., 1989). This raises the question of how exactly cis-regulatory elements, such as RAREs, are able to control the nascent transcriptions of coding and non-coding *Hoxb* genes in specific patterns along the embryonic axis? Analysis of *Hoxb* nascent transcripts in the RARE mutants revealed that *Hoxb* genes are most likely not in a transcriptional hub that would turn all genes on simultaneously. Rather there exist dynamic interactions between RAREs and *Hoxb* gene promoters that activate genes in a specific manner. However, to begin to understand the question of how exactly RAREs turn on *Hoxb* genes individually, in addition to looking at trends of nascent transcripts, one needs to also need to account for the 3D genome architecture and/or the Topologically associated domains (TADs), which may enable or restrict enhancer-promoter contacts. There is evidence that chromatin interactions are actively maintained within the *Hoxb* cluster, and they appear to corelate with presence of the non-canonical Structural Maintenance of Chromosomes (SMC) protein SMCHD1 (Jansz et al., 2018). It has previously also been observed that with activation of *Hoxb* genes, the inactive *Hox* genes remain within the chromosomal territories while active genes loop out (Chambeyron & Bickmore, 2004) (See also reviewed in Duboule & Deschamps,

2004). Hence, the specificity of regulation by RAREs, may greatly be due to the pre-specified chromosomal territories in which they lie as well as the dynamic looping changes occurring within the cluster.

With distance measurements of *Hoxb* nascent transcripts in this study, opposite to other studies, it was observed that co-localized nascent transcripts get closer in the RARE mutants, which could be inferred to the chromatin getting compacted in these mutants. This result is like approaching the problem from the other side, and it appears to be consistent with another paper from the Bickmore lab. They observe that when enhancers are active, the nascent transcript distances increase (Benabdallah et al., 2019). Hence from this study, it can be envisioned that there exist dynamic interactions of RAREs within a transcriptional hub such that at a given point a single RARE is activating nascent transcription of a different *Hoxb* gene (Figure 5-1). One can envision different hubs consisting of different RAREs and the different genes they are transcriptionally activating.

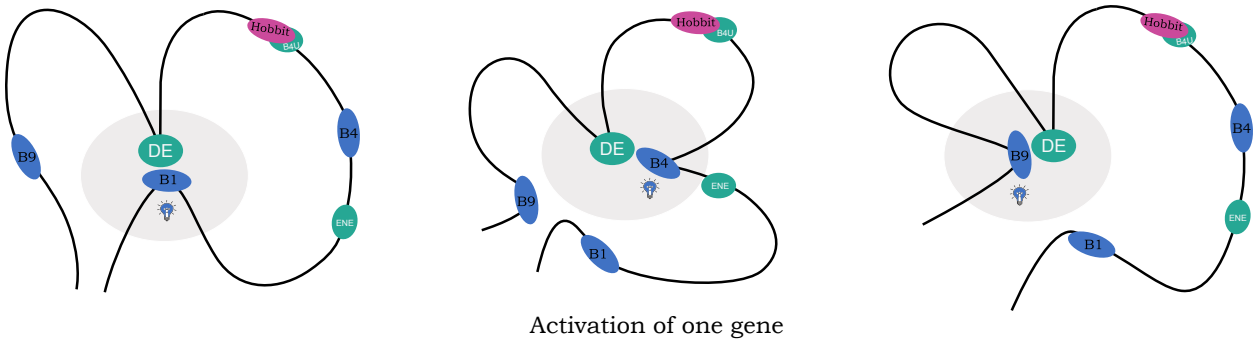


Figure 5-1 Dynamic interaction through which RAREs activate transcription of *Hoxb* genes

Hypothetical model of dynamic transcriptional hub where a single RARE is in close proximity to a single promoter and activates that *Hoxb* gene. In a specific region, e.g., in the neural tube, each cell at a given point in time has a different hub and hence activates a different Hox genes, evidenced from the great number of single transcripts that are observed in this study.

These hubs could also be getting feedback from processed *Hoxb* transcripts to decide upon the levels of nascent transcription, and this could be a potential mechanism for how these RAREs

can work differentially at a single axial level as well as along the embryonic axis. This mechanism seems plausible now in light of a recent paper that shows nascent transcripts stimulate condensate formation (Henninger et al., 2021), while bursts of RNA during elongation phase stimulate dissolution of condensates. Hence, it could be that nascent transcripts of *Hoxb* genes themselves get initiated by RAREs, and that these nascent *Hoxb* transcripts could be allowing for more nascent transcription in single mutants that still have two functional RAREs (Figure 5-2, panel A). However, when there are enough RNAs produced during elongation, these RNAs directly feedback into the activity of the RAREs to dissolve the condensate/hub formed, and stop nascent transcription (Figure 5-2, panel B). This mechanism could be the reason *Hoxb* expression probably gets normalized, and RARE mutant embryos end up developing normally.

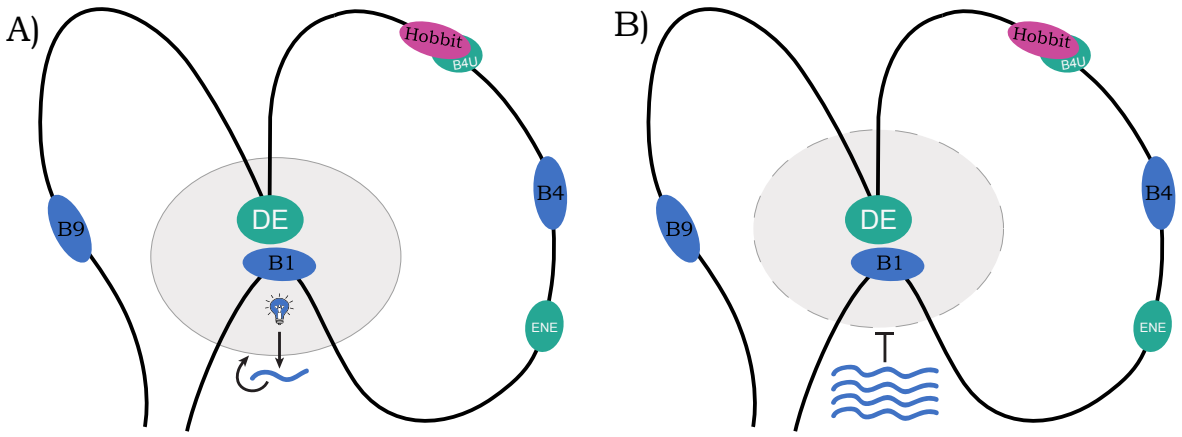


Figure 5-2 Nascent transcripts and transcriptional condensates

A) Nascent transcripts produced promote condensate formation (transcriptional hub) that allows for more nascent transcripts. B) RNAs formed during elongation from this process feedback and promote condensate dissolution.

Furthermore, what makes things more complicated is the fact that enhancer-promoter distances may or may not be linked to transcriptional states. Some groups have seen that decreasing enhancer-promoter distances or forcing loops to form increases transcription (Bartman et al., 2016; Chen et al., 2018; Larsson et al., 2019; Sanyal et al., 2012), indicating need for enhancer-

promoter proximity for active transcription. While other groups have observed the opposite where increased enhancer-promoter distances can also have active transcription, showing dissociation of enhancer-promoter contacts with active transcription (Alexander et al., 2019; Benabdallah et al., 2019). The value of enhancer-promoter distance, the timing of when loops form to create the distance, and whether or not loops or interactions are restricted between TADs or transcriptional condensates, all are emerging questions that still need to be further investigated. A recent paper by Espinola et. al., shows that multiple cis-regulatory modules come together to make transcriptional hubs early during development. They show that these hub formations is aided by the pioneer factor Zelda, and that the whole formation of a hub occurs before transcriptional activation (Espinola et al., 2021). Based on this paper, I can vision tissue specific or axial specific transcriptional hubs forming where all three RAREs are present in the hub, and the RAREs are either competing or working together towards transcriptional activation of *Hoxb* genes. Hence, at a given snapshot in time, axially in the embryo or along the dorsal-ventral axis through the neural tube, there are cell specific differences observed from the transcriptional output of the hub dependent on the active RARE(s). The nascent transcriptional trends could be emerging from dynamic shifts in activity of the RAREs within the hub, and/or the transcriptional readout could be the combinatory output of multiple interactions of the RAREs with the *Hoxb* gene promoters. As an example, Figure 5-3 shows a transcriptional hub with all three RAREs present, but only the DE element is active. DE can potentially activate all three Hoxb genes, but the other B4U and ENE elements interact with promoters of *Hoxb1* and *Hoxb4* and hinder DE from activating these two genes. Hence, DE element in this instance can only activate transcription for *Hoxb9*. This hypothetical mechanism could explain how in the DE mutant *Hoxb9* transcripts were significantly reduced. Also, if competition between elements

plays a role, this hypothetical model could also explain how with no competition from DE in the mutant the other functional RAREs could freely work resulting in more nascent transcripts for *Hoxb1* and *Hoxb4*. Testing competition and cooperativity of these RAREs will help in validating this hypothetical model.

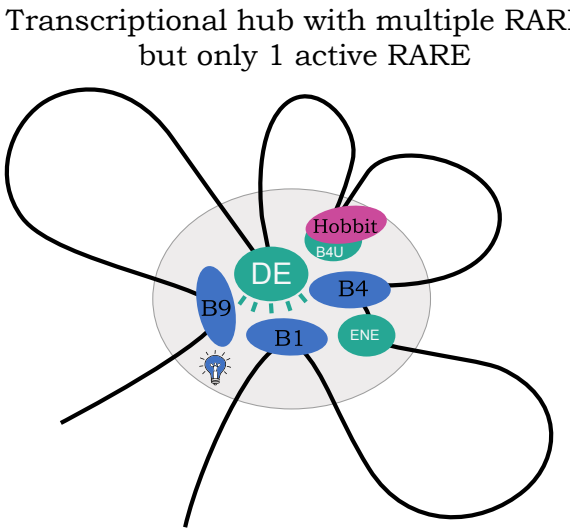


Figure 5-3 Transcriptional hub with multiple RAREs

Hypothetical model to show a hub which contains all the RARE elements, but at a given time one RARE is active. As an example, the DE is shown to be active. When active it blocks other RAREs and can activate transcription for all genes. However, the other RAREs interacting with promoters of *Hoxb1* and *Hoxb4,* can hold off their transcriptional until they are active. The net output of these interactions results in only *Hoxb9* being transcriptionally activated from this transcriptional hub.

Furthermore, recent papers have shown pioneer activity of posterior *Hox* genes (Bulajić et al., 2020; Desanlis et al., 2020). What is particularly interesting is that while Hox binding sites might be similar, HOX transcription factors have different affinities for compacted chromatin (Bulajić et al., 2020). Thus, in this study where analysis is focused on the tail, the posterior *Hoxb* genes are known to exert functional authority through posterior prevalence. It could be that there is a feedback mechanism by the posterior HOXB proteins themselves to regulate the RNA levels, which are detect through presence of nascent transcripts. Further studies need to be undertaken to see if indeed HOXB posterior as well as anterior HOXB transcription factors have some pioneer

activity to auto- or cross-regulate production of the nascent transcripts, especially in the different chromatin states or in the different active condensates, to confer tissue or cell specificity.

The complexity of observations in this study, such that individually these RAREs appear to be acting as repressive elements but together appear to be activators for regulating nascent transcription of the *Hoxb* cluster, has invoked several new questions that need further investigation. Future experiments where live imaging could be performed in cultured mouse embryos (Garcia et al., 2011) to capture transcriptional bursting dynamics. This could be done with a system like MS2/MCP (Lionnet et al., 2011), and it could be even more powerful if it was done in combination with llama tags to simultaneously detect protein localizations with nascent transcripts (Bothma et al., 2015; Bothma et al., 2018). In this way, a read out of processed *Hoxb* transcripts as well as levels and spatial localization of HOX proteins could be obtained. Further, if movements of both alleles could be tracked through live imaging, a better understanding of the mechanism behind the transcriptional trends that have observed, with smFISH snapshots of fixed tissue, could be achieved. Currently, these experiments have been difficult to undertake, but with all the new technologies emerging these experiments seem plausible in the near future.

References

Agarwal, V., Bell, G. W., Nam, J. W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, *4*. https://doi.org/10.7554/eLife.05005

Ahn, Y., Mullan, H. E., & Krumlauf, R. (2014). Long-range regulation by shared retinoic acid response elements modulates dynamic expression of posterior Hoxb genes in CNS development. *Dev Biol*, *388*(1), 134-144. https://doi.org/10.1016/j.ydbio.2014.01.027

Akam, M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. *Development*, *101*, 1-22.

Akbari, M. R., Anderson, L. N., Buchanan, D. D., Clendenning, M., Jenkins, M. A., Win, A. K., Hopper, J. L., Giles, G. G., Nam, R., Narod, S., Gallinger, S., & Cleary, S. P. (2013). Germline HOXB13 p.Gly84Glu mutation and risk of colorectal cancer. *Cancer Epidemiol*, *37*(4), 424-427. https://doi.org/10.1016/j.canep.2013.03.003

Akbari, M. R., Kluźniak, W., Rodin, R., Li, S., Wokołorczyk, D., Royer, R., Kashyap, A., Menkiszak, J., Lubinski, J., Narod, S. A., & Cybulski, C. (2012). The HOXB13 p.Gly84Glu mutation is not associated with the risk of breast cancer. *Breast Cancer Res Treat*, *136*(3), 907-909. https://doi.org/10.1007/s10549-012-2295-y

Alanee, S., Couch, F., & Offit, K. (2012). Association of a HOXB13 variant with breast cancer. *N Engl J Med*, *367*(5), 480-481. https://doi.org/10.1056/NEJMc1205138

Alasti, F., Sadeghi, A., Sanati, M. H., Farhadi, M., Stollar, E., Somers, T., & Van Camp, G. (2008). A mutation in HOXA2 is responsible for autosomal-recessive microtia in an Iranian family. *Am J Hum Genet*, *82*(4), 982-991. https://doi.org/10.1016/j.ajhg.2008.02.015

Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., & Weiner, O. D. (2019). Live-cell imaging reveals enhancer-dependent. *Elife*, *8*. https://doi.org/10.7554/eLife.41769

Alexander, T., Nolte, C., & Krumlauf, R. (2009). Hox genes and segmentation of the hindbrain and axial skeleton. *Annu Rev Cell Dev Biol*, *25*, 431-456. https://doi.org/10.1146/annurev.cellbio.042308.113423

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., & Matthews, B. W. (1981). Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. *Nature*, *290*(5809), 754-758. https://doi.org/10.1038/290754a0

Arnold, P. R., Wells, A. D., & Li, X. C. (2020). Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Frontiers in Cell and Developmental Biology*, *7*. https://doi.org/10.3389/fcell.2019.00377

Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., & Zeitlinger, J. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat Genet*. https://doi.org/10.1038/s41588-021-00782-6

Awgulewitsch, A. (2003). Hox in hair growth and development. *Naturwissenschaften*, *90*(5), 193-211. https://doi.org/10.1007/s00114-003-0417-4

Balkaschina, E. I. (1929). Ein Fall der Erbhomeosis (die Genovarition

"aristopedia") bei Drosophila melanogaster. In *Wilhelm Roux Arch Entwickl Mech Org* (Vol. 115, pp. 448-463).

Barrow, J., & Capecchi, M. (1996). Targeted disruption of the *Hoxb2* locus in mice interferes with expression of *Hoxb1* and *Hoxb4*. *Development*, *122*, 3817-3828.
Bartman, C. R., Hsu, S. C., Hsiung, C. C., Raj, A., & Blobel, G. A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell*, *62*(2), 237-247. https://doi.org/10.1016/j.molcel.2016.03.007
Bateson, W. (1894). *Materials for the study of variation*. MacMillan & Co.
Beeman, R. W., Stuart, J. J., Haas, M. S., & Denell, R. E. (1989). Genetic analysis of the homeotic gene complex (HOM-C) in the beetle Tribolium castaneum. *Dev Biol*, *133*(1), 196-209. https://doi.org/10.1016/0012-1606(89)90311-4
Begemann, G., Schilling, T. F., Rauch, G. J., Geisler, R., & Ingham, P. W. (2001). The zebrafish neckless mutation reveals a requirement for raldh2 in mesodermal signals that pattern the hindbrain. *Development*, *128*(16), 3081-3094.
Bell, E., Wingate, R. J., & Lumsden, A. (1999). Homeotic transformation of rhombomere identity after localized *Hoxb1* misexpression. *Science*, *284*, 21682171.
Benabdallah, N. S., Williamson, I., Illingworth, R. S., Kane, L., Boyle, S., Sengupta, D., Grimes, G. R., Therizols, P., & Bickmore, W. A. (2019). Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol Cell*, *76*(3), 473-484.e477. https://doi.org/10.1016/j.molcel.2019.07.038
Benson, G. V., Lim, H., Paria, B. C., Satokata, I., Dey, S. K., & Maas, R. L. (1996). Mechanisms of reduced fertility in Hoxa-10 mutant mice: uterine homeosis and loss of maternal Hoxa-10 expression. *Development*, *122*(9), 2687-2696.
Bhaskaran, M., & Mohan, M. (2014). MicroRNAs: history, biogenesis, and their evolving role in animal development and disease. *Vet Pathol*, *51*(4), 759-774. https://doi.org/10.1177/0300985813502820
Bieberich, C. J., Ruddle, F. H., & Stenn, K. S. (1991). Differential expression of the Hox 3.1 gene in adult mouse skin. *Ann N Y Acad Sci*, *642*, 346-353; discussion 353-344. https://doi.org/10.1111/j.1749-6632.1991.tb24400.x
Boncinelli, E., Somma, R., Acampora, D., Pannese, M., D'Esposito, M., Faiella, A., & Simeone, A. (1988). Organization of human homeobox genes. *Hum Reprod*, *3*(7), 880-886. https://doi.org/10.1093/oxfordjournals.humrep.a136802
Bosley, T. M., Alorainy, I. A., Salih, M. A., Aldhalaan, H. M., Abu-Amero, K. K., Oystreck, D. T., Tischfield, M. A., Engle, E. C., & Erickson, R. P. (2008). The clinical spectrum of homozygous HOXA1 mutations. *Am J Med Genet A*, *146A*(10), 1235-1240. https://doi.org/10.1002/ajmg.a.32262
Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T., & Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *Elife*, *4*. https://doi.org/10.7554/eLife.07956
Bothma, J. P., Norstad, M. R., Alamos, S., & Garcia, H. G. (2018). LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos. *Cell*, *173*(7), 1810-1822.e1816. https://doi.org/10.1016/j.cell.2018.03.069
Boucherat, O., Montaron, S., Berube-Simard, F. A., Aubin, J., Philippidou, P., Wellik, D. M., Dasen, J. S., & Jeannotte, L. (2013). Partial functional redundancy between Hoxa5 and Hoxb5 paralog genes during lung morphogenesis. *Am J Physiol Lung Cell Mol Physiol*, *304*(12), L817-830. https://doi.org/10.1152/ajplung.00006.2013

Brend, T., Gilthorpe, J., Summerbell, D., & Rigby, P. W. (2003). Multiple levels of transcriptional and post-transcriptional regulation are required to define the domain of Hoxb4 expression. *Development*, *130*, 2717-2728.

Bridges, C. B., & Dobzhan, T. (1933). The mutant "proboscipedia" in Drosophila melanogaster - a case of hereditary homeosis. In (Vol. 127: 575-590). Wilhelm Roux' Arch. Entwicklungsmech Org.

Briggs, J. A., Wolvetang, E. J., Mattick, J. S., Rinn, J. L., & Barry, G. (2015). Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. *Neuron*, *88*(5), 861-877. https://doi.org/10.1016/j.neuron.2015.09.045

Bronner-Fraser, M. (2000). Rostrocaudal differences within the somites confer segmental pattern to trunk neural crest migration. *Curr Top Dev Biol*, *47*, 279-296.

Brun, A. C., Björnsson, J. M., Magnusson, M., Larsson, N., Leveén, P., Ehinger, M., Nilsson, E., & Karlsson, S. (2004). Hoxb4-deficient mice undergo normal hematopoietic development but exhibit a mild proliferation defect in hematopoietic stem cells. *Blood*, *103*(11), 4126-4133. https://doi.org/10.1182/blood-2003-10-3557

Bulajić, M., Srivastava, D., Dasen, J. S., Wichterle, H., Mahony, S., & Mazzoni, E. O. (2020). Differential abilities to engage inaccessible chromatin diversify vertebrate Hox binding patterns. *Development*, *147*(22). https://doi.org/10.1242/dev.194761

Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L. A., & Johnson, R. (2016). Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA*, *22*(6), 867-882. https://doi.org/10.1261/rna.053561.115

Carrasco, A. E., McGinnis, W., Gehring, W. J., & De Robertis, E. M. (1984). Cloning of an X. laevis gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes. *Cell*, *37*, 409-414.

Casaca, A., Nóvoa, A., & Mallo, M. (2016). Hoxb6 can interfere with somitogenesis in the posterior embryo through a mechanism independent of its rib-promoting activity. *Development*, *143*(3), 437-448. https://doi.org/10.1242/dev.133074

Chambeyron, S., & Bickmore, W. A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*, *18*(10), 1119-1130. https://doi.org/10.1101/gad.292104

Chang, H. Y., Chi, J. T., Dudoit, S., Bondre, C., van de Rijn, M., Botstein, D., & Brown, P. O. (2002). Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc Natl Acad Sci U S A*, *99*(20), 12877-12882. https://doi.org/10.1073/pnas.162488599

Chang, Q., Zhang, L., He, C., Zhang, B., Zhang, J., Liu, B., Zeng, N., & Zhu, Z. (2015). HOXB9 induction of mesenchymal-to-epithelial transition in gastric carcinoma is negatively regulated by its hexapeptide motif. *Oncotarget*, *6*(40), 42838-42853. https://doi.org/10.18632/oncotarget.5814

Chen, F., & Capecchi, M. R. (1997). Targeted mutations in hoxa-9 and hoxb-9 reveal synergistic interactions. *Dev Biol*, *181*(2), 186-196. https://doi.org/10.1006/dbio.1996.8440

S0012-1606(96)98440-7 [pii]

Chen, F., & Capecchi, M. R. (1999). Paralogous mouse Hox genes, Hoxa9, Hoxb9, and Hoxd9, function together to control development of the mammary gland in response to pregnancy. *Proc Natl Acad Sci U S A*, *96*(2), 541-546. https://doi.org/10.1073/pnas.96.2.541

Chen, F., Greer, J., & Capecchi, M. R. (1998). Analysis of Hoxa7/Hoxb7 mutants suggests periodicity in the generation of the different sets of vertebrae. *Mech Dev*, *77*(1), 49-57. https://doi.org/S0925-4773(98)00126-9 [pii]

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B., & Gregor, T. (2018). Dynamic interplay between enhancer-promoter topology and gene activity. *Nat Genet*, *50*(9), 1296-1303. https://doi.org/10.1038/s41588-018-0175-z

Chen, J. D., & Evans, R. M. (1995). A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature*, *377*(6548), 454-457. https://doi.org/10.1038/377454a0

Chen, Y., Lun, A. T., & Smyth, G. K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res*, *5*, 1438. https://doi.org/10.12688/f1000research.8987.2

Chen, Y., Takano-Maruyama, M., Fritzsch, B., & Gaufo, G. O. (2012). Hoxb1 controls anteroposterior identity of vestibular projection neurons. *PLoS One*, *7*(4), e34762. https://doi.org/10.1371/journal.pone.0034762

Cho, W. K., Spille, J. H., Hecht, M., Lee, C., Li, C., Grube, V., & Cisse, I. I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*, *361*(6400), 412-415. https://doi.org/10.1126/science.aar4199

Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., & Chang, H. Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*, *44*(4), 667-678. https://doi.org/10.1016/j.molcel.2011.08.027

Chu, C., Quinn, J., & Chang, H. Y. (2012). Chromatin Isolation by RNA Purification (ChIRP). *Journal of Visualized Experiments*(61). https://doi.org/10.3791/3912

Ciruna, B., & Rossant, J. (2001). FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev Cell*, *1*(1), 37-49. https://doi.org/S1534-5807(01)00017-X [pii]

Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A., & Lawrence, J. B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell*, *33*(6), 717-726. https://doi.org/10.1016/j.molcel.2009.01.026

Coen, E. S., & Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature*, *353*(6339), 31-37. https://doi.org/10.1038/353031a0

Connelly, J. P., & Pruett-Miller, S. M. (2019). CRIS.py: A Versatile and High-throughput Analysis Program for CRISPR-based Genome Editing. *Sci Rep*, *9*(1), 4194. https://doi.org/10.1038/s41598-019-40896-w

Cordes, R., Schuster-Gossler, K., Serth, K., & Gossler, A. (2004). Specification of vertebral identity is coupled to Notch signalling and the segmentation clock. *Development*, *131*(6), 1221-1233. https://doi.org/10.1242/dev.01030

Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., Mann, R. S., & Stern, D. L. (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell*, *160*(1-2), 191-203. https://doi.org/10.1016/j.cell.2014.11.041

Cunningham, T. J., & Duester, G. (2015). Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat Rev Mol Cell Biol*, *16*(2), 110-123. https://doi.org/10.1038/nrm3932

Dasen, J. S. (2013). Long noncoding RNAs in development: solidifying the Lncs to Hox gene regulation. *Cell Rep*, *5*(1), 1-2. https://doi.org/10.1016/j.celrep.2013.09.033

Dasen, J. S., & Jessell, T. M. (2009). Hox networks and the origins of motor neuron diversity. *Curr Top Dev Biol*, *88*, 169-200. https://doi.org/10.1016/S0070-2153(09)88006-X

Dasen, J. S., Liu, J. P., & Jessell, T. M. (2003). Motor neuron columnar fate imposed by sequential phases of Hox-c activity. *Nature*, *425*(6961), 926-933. https://doi.org/10.1038/nature02051

Davenne, M., Maconochie, M. K., Neun, R., Pattyn, A., Chambon, P., Krumlauf, R., & Rijli, F. M. (1999). *Hoxa2* and *Hoxb2* control dorsoventral patterns of neuronal development in the rostral hindbrain. *Neuron*, *22*(4), 677-691.

Davis, A., Witte, D., Hsieh, L., Potter, S., & Capecchi, M. (1995). Absence of radius and ulna in mice lacking *Hoxa-11* and *Hoxd-11*. *Nature*, *375*, 791-795.

De Kumar, B., & Krumlauf, R. (2016). HOXs and lincRNAs: Two sides of the same coin. *Sci Adv*, *2*(1), e1501402. https://doi.org/10.1126/sciadv.1501402

De Kumar, B., Parker, H. J., Parrish, M. E., Lange, J. J., Slaughter, B. D., Unruh, J. R., Paulson, A., & Krumlauf, R. (2017). Dynamic regulation of Nanog and stem cell-signaling pathways by Hoxa1 during early neuro-ectodermal differentiation of ES cells. *Proc Natl Acad Sci U S A*, *114*(23), 5838-5845. https://doi.org/10.1073/pnas.1610612114

De Kumar, B., Parrish, M. E., Slaughter, B. D., Unruh, J. R., Gogol, M., Seidel, C., Paulson, A., Li, H., Gaudenz, K., Peak, A., McDowell, W., Fleharty, B., Ahn, Y., Lin, C., Smith, E., Shilatifard, A., & Krumlauf, R. (2015). Analysis of dynamic changes in retinoid-induced transcription and epigenetic profiles of murine Hox clusters in ES cells. *Genome Res*, *25*(8), 1229-1243. https://doi.org/10.1101/gr.184978.114

de Rosa, R., Grenier, J. K., Andreeva, T., Cook, C. E., Adoutte, A., Akam, M., Carroll, S. B., & Balavoine, G. (1999). Hox genes in brachiopods and priapulids and protostome evolution. *Nature*, *399*(6738), 772-776. https://doi.org/10.1038/21631

de Wit, E., Vos, E. S., Holwerda, S. J., Valdes-Quezada, C., Verstegen, M. J., Teunissen, H., Splinter, E., Wijchers, P. J., Krijger, P. H., & de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell*, *60*(4), 676-684. https://doi.org/10.1016/j.molcel.2015.09.023

Delpretti, S., Montavon, T., Leleu, M., Joye, E., Tzika, A., Milinkovitch, M., & Duboule, D. (2013). Multiple enhancers regulate Hoxd genes and the Hotdog LncRNA during cecum budding. *Cell Rep*, *5*(1), 137-150. https://doi.org/10.1016/j.celrep.2013.09.002

Deng, C., Li, Y., Zhou, L., Cho, J., Patel, B., Terada, N., Li, Y., Bungert, J., Qiu, Y., & Huang, S. (2016). HoxBlinc RNA Recruits Set1/MLL Complexes to Activate Hox Gene Expression Patterns and Mesoderm Lineage Development. *Cell Rep*, *14*(1), 103-114. https://doi.org/10.1016/j.celrep.2015.12.007

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A., & Blobel, G. A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, *149*(6), 1233-1244. https://doi.org/10.1016/j.cell.2012.03.051

Desanlis, I., Paul, R., & Kmita, M. (2020). Transcriptional Trajectories in Mouse Limb Buds Reveal the Transition from Anterior-Posterior to Proximal-Distal Patterning at Early Limb Bud Stage. *J Dev Biol*, *8*(4). https://doi.org/10.3390/jdb8040031

Deschamps, J., & van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development*, *132*(13), 2931-2942. https://doi.org/10.1242/dev.01897

Desplan, C., Theis, J., & O'Farrell, P. H. (1985). The Drosophila developmental gene, engrailed, encodes a sequence-specific DNA binding activity. *Nature*, *318*(6047), 630-635. https://doi.org/10.1038/318630a0

Detmer, K., Lawrence, H. J., & Largman, C. (1993). Expression of class I homeobox genes in fetal and adult murine skin. *J Invest Dermatol*, *101*(4), 517-522. https://doi.org/10.1111/1523-1747.ep12365890

DeVeale, B., Swindlehurst-Chan, J., & Blelloch, R. (2021). The roles of microRNAs in mouse development. *Nat Rev Genet*. https://doi.org/10.1038/s41576-020-00309-5

Di Bonito, M., Glover, J. C., & Studer, M. (2013). Hox genes and region-specific sensorimotor circuit formation in the hindbrain and spinal cord. *Dev Dyn*. https://doi.org/10.1002/dvdy.24055

Diez del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M., & Storey, K. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron*, *40*(1), 65-79.

Diez del Corral, R., & Storey, K. G. (2004). Opposing FGF and retinoid pathways: a signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. *Bioessays*, *26*(8), 857-869. https://doi.org/10.1002/bies.20080

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376-380. https://doi.org/10.1038/nature11082

Dolecki, G. J., Wannakrairoj, S., Lum, R., Wang, G., Riley, H. D., Carlos, R., Wang, A., & Humphreys, T. (1986). Stage-specific expression of a homeo box-containing gene in the non-segmented sea urchin embryo. *EMBO J*, *5*(5), 925-930.

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

Dollé, P., & Duboule, D. (1989). Two gene members of the murine *HOX 5* complex show regional and cell-type specific expression in developing limbs and gonads. *EMBO J.*, *8*, 1507-1515.

Dollé, P., Izpisùa-Belmonte, J. C., Brown, J. M., Tickle, C., & Duboule, D. (1991). *Hox-4* genes and the morphogenesis of mammalian genitalia. *Genes Dev.*, *5*, 1767-1776.

Dollé, P., Izpisùa-Belmonte, J. C., Falkenstein, H., Renucci, A., & Duboule, D. (1989). Co-ordinate expression of the murine *Hox-5* complex homeobox-containing genes during limb pattern formation. *Nature*, *342*(6251), 767-772.

Dollé, P., Izpisúa-Belmonte, J. C., Boncinelli, E., & Duboule, D. (1991). The Hox-4.8 gene is localized at the 5' extremity of the Hox-4 complex and is expressed in the most posterior parts of the body during development. *Mech Dev*, *36*(1-2), 3-13. https://doi.org/10.1016/0925-4773(91)90067-g

Du, H., & Taylor, H. S. (2015). The Role of Hox Genes in Female Reproductive Tract Development, Adult Function, and Fertility. *Cold Spring Harb Perspect Med*, *6*(1), a023002. https://doi.org/10.1101/cshperspect.a023002

Du, Z., Sun, T., Hacisuleyman, E., Fei, T., Wang, X., Brown, M., Rinn, J. L., Lee, M. G., Chen, Y., Kantoff, P. W., & Liu, X. S. (2016). Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nat Commun*, *7*, 10982. https://doi.org/10.1038/ncomms10982

Dubey, A., Rose, R. E., Jones, D. R., & Saint-Jeannet, J. P. (2018). Generating retinoic acid gradients by local degradation during craniofacial development: One cell's cue is another cell's poison. *Genesis*, *56*(2). https://doi.org/10.1002/dvg.23091
Duboule, D., & Deschamps, J. (2004). Colinearity loops out. *Dev Cell*, *6*, 738-740.
Duboule, D., & Dolle, P. (1989). The structural and functional organization of the murine *HOX* gene family resembles that of *Drosophila* homeotic genes. *EMBO J*, *8*(5), 1497-1505.
Duboule, D., & Morata, G. (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet*, *10*, 358-364.
Dubrulle, J., & Pourquie, O. (2004). fgf8 mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo. *Nature*, *427*(6973), 419-422. https://doi.org/10.1038/nature02216

nature02216 [pii]
Dunn, T. M., Hahn, S., Ogden, S., & Schleif, R. F. (1984). An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc Natl Acad Sci U S A*, *81*(16), 5017-5020. https://doi.org/10.1073/pnas.81.16.5017
Dupe, V., Davenne, M., Brocard, J., Dolle, P., Mark, M., Dierich, A., Chambon, P., & Rijli, F. M. (1997). In vivo functional analysis of the Hoxa-1 3' retinoic acid response element (3'RARE). *Development*, *124*(2), 399-410.
Dupe, V., & Lumsden, A. (2001). Hindbrain patterning involves graded responses to retinoic acid signalling. *Development*, *128*(12), 2199-2208.
Economides, K. D., Zeltser, L., & Capecchi, M. R. (2003). Hoxb13 mutations cause overgrowth of caudal spinal cord and tail vertebrae. *Dev Biol*, *256*, 317-330.
Ensini, M., Tsuchida, T., Belting, H.-G., & Jessell, T. (1998). The control of rostrocaudal pattern in the developing spinal cord: Specification of motor neuron subtype identity is initiated by signals from paraxial mesoderm. *Development*, *125*, 969-982.
Espinola, S. M., Götz, M., Bellec, M., Messina, O., Fiche, J.-B., Houbron, C., Dejean, M., Reim, I., Cardozo Gizzi, A. M., Lagha, M., & Nollmann, M. (2021). Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. *Nature Genetics*, *53*(4), 477-486. https://doi.org/10.1038/s41588-021-00816-z
Ewing, C. M., Ray, A. M., Lange, E. M., Zuhlke, K. A., Robbins, C. M., Tembe, W. D., Wiley, K. E., Isaacs, S. D., Johng, D., Wang, Y., Bizon, C., Yan, G., Gielzak, M., Partin, A. W., Shanmugam, V., Izatt, T., Sinari, S., Craig, D. W., Zheng, S. L., Walsh, P. C., Montie, J. E., Xu, J., Carpten, J. D., Isaacs, W. B., & Cooney, K. A. (2012). Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med*, *366*(2), 141-149. https://doi.org/10.1056/NEJMoa1110000
Fatima, F., & Nawaz, M. (2017). Vesiculated Long Non-Coding RNAs: Offshore Packages Deciphering Trans-Regulation between Cells, Cancer Progression and Resistance to Therapies. *Noncoding RNA*, *3*(1). https://doi.org/10.3390/ncrna3010010
Favier, B., & Dolle, P. (1997). Developmental functions of mammalian Hox genes. *Mol Hum Reprod*, *3*, 115-131.
Favier, B., Rijli, F. M., Fromental-Ramain, C., Fraulob, V., Chambon, P., & Dollé, P. (1996). Functional cooperation between the non-paralogous genes Hoxa-10 and Hoxd-11 in the developing forelimb and axial skeleton. *Development*, *122*, 449-460.

Forlani, S., Lawson, K. A., & Deschamps, J. (2003). Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. *Development*, *130*(16), 3807-3819.
Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell*, *166*(2), 358-368. https://doi.org/10.1016/j.cell.2016.05.025
Furlong, E. E. M., & Levine, M. (2018). Developmental enhancers and chromosome topology. *Science*, *361*(6409), 1341-1345. https://doi.org/10.1126/science.aau0320
Galceran, J., Sustmann, C., Hsu, S. C., Folberth, S., & Grosschedl, R. (2004). LEF1-mediated regulation of Delta-like1 links Wnt and Notch signaling in somitogenesis. *Genes Dev*, *18*(22), 2718-2723. https://doi.org/10.1101/gad.1249504
Gambetta, M. C., & Furlong, E. E. M. (2018). The Insulator Protein CTCF Is Required for Correct. *Genetics*, *210*(1), 129-136. https://doi.org/10.1534/genetics.118.301350
Garcia, M. D., Udan, R. S., Hadjantonakis, A. K., & Dickinson, M. E. (2011). Live imaging of mouse embryos. *Cold Spring Harb Protoc*, *2011*(4), pdb.top104. https://doi.org/10.1101/pdb.top104
Garcia-Barceló, M. M., Wong, K. K., Lui, V. C., Yuan, Z. W., So, M. T., Ngan, E. S., Miao, X. P., Chung, P. H., Khong, P. L., & Tam, P. K. (2008). Identification of a HOXD13 mutation in a VACTERL patient. *Am J Med Genet A*, *146A*(24), 3181-3185. https://doi.org/10.1002/ajmg.a.32426
Gaufo, G. O., Thomas, K. R., & Capecchi, M. R. (2003). Hox3 genes coordinate mechanisms of genetic suppression and activation in the generation of branchial and somatic motoneurons. *Development*, *130*, 5191-5201.
Gaufo, G. O., Wu, S., & Capecchi, M. R. (2004). Contribution of Hox genes to the diversity of the hindbrain sensory system. *Development*, *131*(6), 1259-1266. https://doi.org/10.1242/dev.01029
Gaunt, S. J. (2018). Hox cluster genes and collinearities throughout the tree of animal life. *Int J Dev Biol*, *62*(11-12), 673-683. https://doi.org/10.1387/ijdb.180162sg
Gaunt, S. J., Sharpe, P. T., & Duboule, D. (1988). Spatially restricted domains of homeo-gene transcripts in mouse embryos: relation to a segmented body plan. *Development*, *104 (Suppl.)*, 169-181.
Gavalas, A., & Krumlauf, R. (2000). Retinoid signalling and hindbrain patterning. *Curr Opin Genet Dev*, *10*(4), 380-386.
Gavalas, A., Ruhrberg, C., Livet, J., Henderson, C. E., & Krumlauf, R. (2003). Neuronal defects in the hindbrain of Hoxa1, Hoxb1 and Hoxb2 mutants reflect regulatory interactions among these Hox genes. *Development*, *130*(23), 5663-5679. https://doi.org/10.1242/dev.00802
Gavalas, A., Studer, M., Lumsden, A., Rijli, F. M., Krumlauf, R., & Chambon, P. (1998). Hoxa1 and Hoxb1 synergize in patterning the hindbrain, cranial nerves and second pharyngeal arch. *Development*, *125*(6), 1123-1136.
Geada, A. M. C., Gaunt, S. J., Azzawi, M., Shimeld, S. M., Pearce, J., & Sharpe, P. T. (1992). Sequence and embryonic expression of the murine Hox-3.5 gene. *Development*, *116*, 497-506.
Gehring, W., Muller, M., Affolter, M., Percival-Smith, A., Billeter, M., Qian, Y., Otting, G., & Wuthrich, K. (1990). The structure of the homeodomain and its functional implications. *TIG*, *6*(10), 323-329.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., & Wuthrich, K. (1994). Homeodomain-DNA recognition. *Cell*, *78*(2), 211-223.

Gerard, M., Zakany, J., & Duboule, D. (1997). Interspecies exchange of a Hoxd enhancer in vivo induces premature transcription and anterior shift of the sacrum. *Dev Biol*, *190*(1), 32-40.

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., & Furlong, E. E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*, *512*(7512), 96-100. https://doi.org/10.1038/nature13417

Glass, C. K., & Rosenfeld, M. G. (2000). The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev*, *14*(2), 121-141.

Goddard, J. M., Rossel, M., Manley, N. R., & Capecchi, M. R. (1996). Mice with targeted disruption of *Hoxb-1* fail to form the motor nucleus of the VIIth nerve. *Development*, *122*(10), 3217-3228.

Godwin, A. R., & Capecchi, M. R. (1998). Hoxc13 mutant mice lack external hair. *Genes Dev*, *12*(1), 11-20. https://doi.org/10.1101/gad.12.1.11

Goff, L. A., Groff, A. F., Sauvageau, M., Trayes-Gibson, Z., Sanchez-Gomez, D. B., Morse, M., Martin, R. D., Elcavage, L. E., Liapis, S. C., Gonzalez-Celeiro, M., Plana, O., Li, E., Gerhardinger, C., Tomassy, G. S., Arlotta, P., & Rinn, J. L. (2015). Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A*, *112*(22), 6855-6862. https://doi.org/10.1073/pnas.1411263112

Gonçalves, C. S., Le Boiteux, E., Arnaud, P., & Costa, B. M. (2020). HOX gene cluster (de)regulation in brain: from neurodevelopment to malignant glial tumours. *Cell Mol Life Sci*, *77*(19), 3797-3821. https://doi.org/10.1007/s00018-020-03508-9

Goodman, F. R. (2002). Limb malformations and the human HOX genes. *Am J Med Genet*, *112*(3), 256-265. https://doi.org/10.1002/ajmg.10776

Gould, A., Itasaki, N., & Krumlauf, R. (1998). Initiation of rhombomeric *Hoxb4* expression requires induction by somites and a retinoid pathway. *Neuron*, *21*(1), 39-51.

Graham, A., Papalopulu, N., & Krumlauf, R. (1989). The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell*, *57*(3), 367-378.

Greer, J. M., & Capecchi, M. R. (2002). Hoxb8 is required for normal grooming behavior in mice. *Neuron*, *33*, 23-34.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., & Herrmann, B. G. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell*, *24*(2), 206-214. https://doi.org/10.1016/j.devcel.2012.12.012

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M. Q., Ren, B., Krainer, A. R., Maniatis, T., & Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, *162*(4), 900-910. https://doi.org/10.1016/j.cell.2015.07.038

Haack, H., & Gruss, P. (1993). The establishment of murine Hox-1 expression domains during patterning of the limb. *Dev Biol*, *157*(2), 410-422. https://doi.org/10.1006/dbio.1993.1145

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*, *19*(10), 621-637. https://doi.org/10.1038/s41580-018-0028-8

Hafen, E., Levine, M., & Gehring, W. (1984). Regulation of *Antennapedia* transcript distribution by the bithorax complex of *Drosophila*. *Nature*, *307*, 287-289.
Hagen, D. C., Bruhn, L., Westby, C. A., & Sprague, G. F. (1993). Transcription of alpha-specific genes in Saccharomyces cerevisiae: DNA sequence requirements for activity of the coregulator alpha 1. *Mol Cell Biol*, *13*(11), 6866-6875. https://doi.org/10.1128/mcb.13.11.6866
Harding, K., Wedeen, C., McGinnis, W., & Levine, M. (1985). Spatially regulated expression of homeotic genes in *Drosophila*. *Science*, *229*, 1236-1242.
Hart, C., Awgulewitsch, A., Fainsod, A., McGinnis, W., & Ruddle, F. (1985). Homeobox gene complex on mouse chromosome 11: molecular cloning, expression in embryogenesis and homology to a human homeobox locus. *Cell*, *43*, 9-18.
He, D., Wang, J., Lu, Y., Deng, Y., Zhao, C., Xu, L., Chen, Y., Hu, Y. C., Zhou, W., & Lu, Q. R. (2017). lncRNA Functional Networks in Oligodendrocytes Reveal Stage-Specific Myelination Control by an lncOL1/Suz12 Complex in the CNS. *Neuron*, *93*(2), 362-378. https://doi.org/10.1016/j.neuron.2016.11.044
Henninger, J. E., Oksuz, O., Shrinivas, K., Sagi, I., LeRoy, G., Zheng, M. M., Andrews, J. O., Zamudio, A. V., Lazaris, C., Hannett, N. M., Lee, T. I., Sharp, P. A., Cissé, I. I., Chakraborty, A. K., & Young, R. A. (2021). RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell*, *184*(1), 207-225.e224. https://doi.org/10.1016/j.cell.2020.11.030
Hernandez, R. E., Putzke, A. P., Myers, J. P., Margaretha, L., & Moens, C. B. (2007). Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development. *Development*, *134*(1), 177-187. https://doi.org/10.1242/dev.02706
Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep*, *11*(7), 1110-1122. https://doi.org/10.1016/j.celrep.2015.04.023
Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell*, *169*(1), 13-23. https://doi.org/10.1016/j.cell.2017.02.007
Hofmann, M., Schuster-Gossler, K., Watabe-Rudolph, M., Aulehla, A., Herrmann, B. G., & Gossler, A. (2004). WNT signaling, in synergy with T/TBX6, controls Notch signaling by regulating Dll1 expression in the presomitic mesoderm of mouse embryos. *Genes Dev*, *18*(22), 2712-2717. https://doi.org/10.1101/gad.1248604
Hogan, B. L. M., Thaller, C., & Eichle, G. (1992). Evidence that Hensen's node is a site of retinoic acid synthesis. *Nature*, *359*, 237-241.
Holstege, J. C., de Graaff, W., Hossaini, M., Cano, S. C., Jaarsma, D., van den Akker, E., & Deschamps, J. (2008). Loss of Hoxb8 alters spinal dorsal laminae and sensory responses in mice. *Proc Natl Acad Sci U S A*, *105*(17), 6338-6343.
Horlein, A. J., Naar, A. M., Heinzel, T., Torchia, J., Gloss, B., Kurokawa, R., Ryan, A., Kamei, Y., Soderstrom, M., Glass, C. K., & et al. (1995). Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature*, *377*(6548), 397-404. https://doi.org/10.1038/377397a0
Hsieh-Li, H. M., Witte, D. P., Weinstein, M., Branford, W., Li, H., Small, K., & Potter, S. S. (1995). Hoxa 11 structure, extensive antisense transcription, and function in male and female fertility. *Development*, *121*(5), 1373-1385.

Hubaud, A., & Pourquié, O. (2014). Signalling dynamics in vertebrate segmentation. *Nat Rev Mol Cell Biol*, *15*(11), 709-721. https://doi.org/10.1038/nrm3891
Hubaud, A., Regev, I., Mahadevan, L., & Pourquié, O. (2017). Excitable Dynamics and Yap-Dependent Mechanical Cues Drive the Segmentation Clock. *Cell*, *171*(3), 668-682.e611. https://doi.org/10.1016/j.cell.2017.08.043
Hunt, P., & Krumlauf, R. (1991). Deciphering the Hox Code: Clues to patterning the branchial region of the head. *Cell*, *66*, 1075-1078.
Ikeya, M., & Takada, S. (2001). Wnt-3a is required for somite specification along the anteroposterior axis of the mouse embryo and for regulation of cdx-1 expression. *Mech Dev*, *103*(1-2), 27-33. https://doi.org/S0925477301003380 [pii]
Isoda, T., Moore, A. J., He, Z., Chandra, V., Aida, M., Denholtz, M., Piet van Hamburg, J., Fisch, K. M., Chang, A. N., Fahl, S. P., Wiest, D. L., & Murre, C. (2017). Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. *Cell*, *171*(1), 103-119.e118. https://doi.org/10.1016/j.cell.2017.09.001
Izpisua-Belmonte, J. C., Tickle, C., Dollé, P., Wolpert, L., & Duboule, D. (1991). Expression of the homeobox Hox-4 genes and the specification of position in chick wing development. *Nature*, *350*, 585-589.
Jabet, C., Gitti, R., Summers, M. F., & Wolberger, C. (1999). NMR studies of the pbx1 TALE homeodomain protein free in solution and bound to DNA: proposal for a mechanism of HoxB1-Pbx1-DNA complex assembly. *J Mol Biol*, *291*(3), 521-530. https://doi.org/10.1006/jmbi.1999.2983
Jansz, N., Keniry, A., Trussart, M., Bildsoe, H., Beck, T., Tonks, I. D., Mould, A. W., Hickey, P., Breslin, K., Iminitoff, M., Ritchie, M. E., McGlinn, E., Kay, G. F., Murphy, J. M., & Blewitt, M. E. (2018). Long-range chromatin interactions on the inactive X and at Hox clusters are regulated by the non-canonical SMC protein Smchd1. In. BioRxiv.
Juan, A. H., & Ruddle, F. H. (2003). Enhancer timing of Hox gene expression: deletion of the endogenous Hoxc8 early enhancer. *Development*, *130*(20), 4823-4834.
Kam, M. K., Cheung, M. C., Zhu, J. J., Cheng, W. W., Sat, E. W., Tam, P. K., & Lui, V. C. (2014). Perturbation of Hoxb5 signaling in vagal and trunk neural crest cells causes apoptosis and neurocristopathies in mice. *Cell Death Differ*, *21*(2), 278-289. https://doi.org/10.1038/cdd.2013.142
Kashi, K., Henderson, L., Bonetti, A., & Carninci, P. (2016). Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome. *Biochim Biophys Acta*, *1859*(1), 3-15. https://doi.org/10.1016/j.bbagrm.2015.10.010
Kawazoe, Y., Sekimoto, T., Araki, M., Takagi, K., Araki, K., & Yamamura, K. (2002). Region-specific gastrointestinal Hox code during murine embryonal gut development. *Dev Growth Differ*, *44*(1), 77-84. https://doi.org/10.1046/j.1440-169x.2002.00623.x
Kelsh, R., Dawson, I., & Akam, M. (1993). An analysis of abdominal-B expression in the locust Schistocerca gregaria. *Development*, *117*(1), 293-305.
Kenyon, C., & Wang, B. (1991). A cluster of Antennapedia-class homeobox genes in a nonsegmented animal. *Science*, *253*, 516-517.
Kessel, M., & Gruss, P. (1991). Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. *Cell*, *67*(1), 89-104.
Kiecker, C., & Lumsden, A. (2012). The role of organizers in patterning the nervous system. *Annu Rev Neurosci*, *35*, 347-367. https://doi.org/10.1146/annurev-neuro-062111-150543

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, *14*(4), R36. https://doi.org/10.1186/gb-2013-14-4-r36

Kim, D., Song, J., Han, J., Kim, Y., Chun, C. H., & Jin, E. J. (2013). Two non-coding RNAs, MicroRNA-101 and HOTTIP contribute cartilage integrity by epigenetic and homeotic regulation of integrin-alpha1. *Cell Signal*, *25*(12), 2878-2887. https://doi.org/10.1016/j.cellsig.2013.08.034

Kissinger, C., Liu, B., Martin-Blanco, E., Kronberg, T., & Pabo, C. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8A resolution: A framework for understanding homeodomain-DNA interactions. *Cell*, *63*(2), 579-590.

Klar, A. J. (1987). Determination of the yeast cell lineage. *Cell*, *49*(4), 433-435. https://doi.org/10.1016/0092-8674(87)90442-9

Kmita, M., van Der Hoeven, F., Zakany, J., Krumlauf, R., & Duboule, D. (2000). Mechanisms of Hox gene colinearity: transposition of the anterior Hoxb1 gene into the posterior HoxD complex. *Genes Dev*, *14*(2), 198-211.

Kondo, T., Dollé, P., Zakany, J., & Duboule, D. (1996). Function of posterior HoxD genes in the morphogenesis of the anal sphincter. *Development*, *122*(9), 2651-2659.

Kondo, T., & Duboule, D. (1999). Breaking colinearity in the mouse HoxD complex. *Cell*, *97*(3), 407-417.

Kondrashov, N., Pusic, A., Stumpf, C. R., Shimizu, K., Hsieh, A. C., Ishijima, J., Shiroishi, T., & Barna, M. (2011). Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell*, *145*(3), 383-397. https://doi.org/10.1016/j.cell.2011.03.028

Kong, S., Bohl, D., Li, C., & Tuan, D. (1997). Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Mol Cell Biol*, *17*(7), 3955-3965. https://doi.org/10.1128/mcb.17.7.3955

Krumlauf, R. (1994). *Hox* genes in vertebrate development. *Cell*, *78*(2), 191-201.

Krumlauf, R. (2016). Hox Genes and the Hindbrain: A Study in Segments. *Curr Top Dev Biol*, *116*, 581-596. https://doi.org/10.1016/bs.ctdb.2015.12.011

Krumlauf, R. (2018). Hox genes, clusters and collinearity. *Int J Dev Biol*, *62*(11-12), 659-663. https://doi.org/10.1387/ijdb.180330rr

Krumlauf, R., & Ahn, Y. (2013). Hox genes. In S. Maloy & K. Hughes (Eds.), *Brenner's Encyclopedia of Genetics (Second Edition)* (2nd ed., pp. 539-542). Academic Press. https://doi.org/http://dx.doi.org/10.1016/B978-0-12-374984-0.00742-7

Krumlauf, R., Holland, P. W., McVey, J. H., & Hogan, B. L. (1987). Developmental and spatial patterns of expression of the mouse homeobox gene, *Hox 2.1*. *Development*, *99*(4), 603-617.

Labuhn, M., Adams, F. F., Ng, M., Knoess, S., Schambach, A., Charpentier, E. M., Schwarzer, A., Mateo, J. L., Klusmann, J. H., & Heckl, D. (2018). Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic Acids Res*, *46*(3), 1375-1385. https://doi.org/10.1093/nar/gkx1268

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., Segerstolpe, Å., Rivera, C. M., Ren, B., & Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. *Nature*, *565*(7738), 251-254. https://doi.org/10.1038/s41586-018-0836-1

Lavinsky, R. M., Jepsen, K., Heinzel, T., Torchia, J., Mullen, T. M., Schiff, R., Del-Rio, A. L., Ricote, M., Ngo, S., Gemsch, J., Hilsenbeck, S. G., Osborne, C. K., Glass, C. K., Rosenfeld, M. G., & Rose, D. W. (1998). Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes. *Proc Natl Acad Sci U S A*, *95*(6), 2920-2925.

Lawson, K. A., Meneses, J. J., & Pedersen, R. A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development*, *113*(3), 891-911.

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75*(5), 843-854. https://doi.org/10.1016/0092-8674(93)90529-y

Leppek, K., Fujii, K., Quade, N., Susanto, T. T., Boehringer, D., Lenarčič, T., Xue, S., Genuth, N. R., Ban, N., & Barna, M. (2020). Gene- and Species-Specific Hox mRNA Translation by Ribosome Expansion Segments. *Mol Cell*, *80*(6), 980-995.e913. https://doi.org/10.1016/j.molcel.2020.10.023

Levine, M., Harding, K., Wedeen, C., Doyle, H., Hoey, T., & Radomska, H. (1985). Expression of the homeo box gene family in Drosophila. *Cold Spring Harb Symp Quant Biol*, *50*, 209-222. https://doi.org/10.1101/sqb.1985.050.01.027

Levine, M., & Hoey, T. (1988). Homeobox proteins as sequence-specific transcription factors. *Cell*, *55*(4), 537-540.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature*, *276*(5688), 565-570.

Lim, B., Heist, T., Levine, M., & Fukaya, T. (2018). Visualization of Transvection in Living Drosophila Embryos. *Mol Cell*, *70*(2), 287-296.e286. https://doi.org/10.1016/j.molcel.2018.02.029

Lim, T., Jaques, K., Stern, C., & Keynes, R. (1991). An evaluation of myelomeres and segmentation of the chick embryo spinal cord. *Development*, *113*, 227-238.

Lin, Z., Chen, Q., Shi, L., Lee, M., Giehl, K. A., Tang, Z., Wang, H., Zhang, J., Yin, J., Wu, L., Xiao, R., Liu, X., Dai, L., Zhu, X., Li, R., Betz, R. C., Zhang, X., & Yang, Y. (2012). Loss-of-function mutations in HOXC13 cause pure hair and nail ectodermal dysplasia. *Am J Hum Genet*, *91*(5), 906-911. https://doi.org/10.1016/j.ajhg.2012.08.029

Lindsley, D. L., & Zimm, G. G. (1992). The Genome of Drosophila Melanogaster. In (pp. 1160): Academic Press.

Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A. L., Chao, J. A., Park, H. Y., de Turris, V., Lopez-Jones, M., & Singer, R. H. (2011). A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat Methods*, *8*(2), 165-170. https://doi.org/10.1038/nmeth.1551

Liu, J. P., Laufer, E., & Jessell, T. M. (2001). Assigning the Positional Identity of Spinal Motor Neurons. Rostrocaudal Patterning of Hox-c Expression by FGFs, Gdf11, and Retinoids. *Neuron*, *32*(6), 997-1012.

Liu, P., Wakamiya, M., Shea, M. J., Albrecht, U., Behringer, R. R., & Bradley, A. (1999). Requirement for Wnt3 in vertebrate axis formation. *Nat Genet*, *22*(4), 361-365. https://doi.org/10.1038/11932

Lohnes, D., Kastner, P., Dierich, A., Mark, M., LeMeur, M., & Chambon, P. (1993). Function of the retinoic acid receptor gamma in the mouse. *Cell*, *73*, 643-658.

Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, *167*(5), 1170-1187. https://doi.org/10.1016/j.cell.2016.09.018

Loudig, O., Maclean, G. A., Dore, N. L., Luu, L., & Petkovich, M. (2005). Transcriptional co-operativity between distant retinoic acid response elements in regulation of Cyp26A1 inducibility. *Biochem J*, *392*(Pt 1), 241-248. https://doi.org/10.1042/BJ20050874

Luo, Z., Rhie, S. K., & Farnham, P. J. (2019). The Enigmatic HOX Genes: Can We Crack Their Code? *Cancers (Basel)*, *11*(3). https://doi.org/10.3390/cancers11030323

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., Visel, A., & Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, *161*(5), 1012-1025. https://doi.org/10.1016/j.cell.2015.04.004

Maamar, H., Cabili, M. N., Rinn, J., & Raj, A. (2013). linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. *Genes Dev*, *27*(11), 1260-1271. https://doi.org/10.1101/gad.217018.113

Maconochie, M., Nonchev, S., Morrison, A., & Krumlauf, R. (1996). Paralogous *Hox* genes: function and regulation. *Annual Review of Genetics*, *30*, 529-556. https://doi.org/10.1146/annurev.genet.30.1.529

Mahmood, R., Mason, I. J., & Morriss-Kay, G. M. (1996). Expression of Fgf-3 in relation to hindbrain segmentation, otic pit position and pharyngeal arch morphology in normal and retinoic acid-exposed mouse embryos. *Anat Embryol (Berl)*, *194*, 13-22.

Mallo, M., & Alonso, C. R. (2013). The regulation of Hox gene expression during animal development. *Development*, *140*(19), 3951-3963. https://doi.org/10.1242/dev.068346

Manley, N. R., Barrow, J. R., Zhang, T., & Capecchi, M. R. (2001). Hoxb2 and hoxb4 act together to specify ventral body wall formation. *Dev Biol*, *237*(1), 130-144.

Manley, N. R., & Capecchi, M. R. (1997). *Hox* group 3 paralogous genes act synergistically in the formation of somitic and neural crest-derived structures. *Dev Biol*, *192*(2), 274-288. https://doi.org/10.1006/dbio.1997.8765

Mann, R. S., Lelli, K. M., & Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. *Curr Top Dev Biol*, *88*, 63-101. https://doi.org/10.1016/S0070-2153(09)88003-4

Marshall, H., Nonchev, S., Sham, M. H., Muchamore, I., Lumsden, A., & Krumlauf, R. (1992). Retinoic acid alters hindbrain Hox code and induces transformation of rhombomeres 2/3 into a 4/5 identity. *Nature*, *360*(6406), 737-741. https://doi.org/10.1038/360737a0

Marshall, H., Studer, M., Popperl, H., Aparicio, S., Kuroiwa, A., Brenner, S., & Krumlauf, R. (1994). A conserved retinoic acid response element required for early expression of the homeobox gene *Hoxb-1*. *Nature*, *370*(6490), 567-571. https://doi.org/10.1038/370567a0

Marsman, J., & Horsfield, J. A. (2012). Long distance relationships: enhancer-promoter communication and dynamic gene transcription. *Biochim Biophys Acta*, *1819*(11-12), 1217-1227. https://doi.org/10.1016/j.bbagrm.2012.10.008

Mathews, C. H., Detmer, K., Lawrence, H. J., & Largman, C. (1993). Expression of the Hox 2.2 homeobox gene in murine embryonic epidermis. *Differentiation*, *52*(2), 177-184. https://doi.org/10.1111/j.1432-0436.1993.tb00628.x

Mathis, L., Kulesa, P. M., & Fraser, S. E. (2001). FGF receptor signalling is required to maintain neural progenitors during Hensen's node progression. *Nat Cell Biol*, *3*(6), 559-566.
Mattick, J. S., & Rinn, J. L. (2015). Discovery and annotation of long noncoding RNAs. *Nat Struct Mol Biol*, *22*(1), 5-7. https://doi.org/10.1038/nsmb.2942
McClintock, J. M., Kheirbek, M. A., & Prince, V. E. (2002). Knockdown of duplicated zebrafish hoxb1 genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. *Development*, *129*(10), 2339-2354.
McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., & Gehring, W. J. (1984). A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. *Cell*, *37*(2), 403-408.
McKay, D. B., & Steitz, T. A. (1981). Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA. *Nature*, *290*(5809), 744-749. https://doi.org/10.1038/290744a0
Medina-Martinez, O., Bradley, A., & Ramirez-Solis, R. (2000). A large targeted deletion of Hoxb1-Hoxb9 produces a series of single-segment anterior homeotic transformations. *Dev Biol*, *222*(1), 71-83. https://doi.org/S0012160600996830 [pii]
Medina-Martinez, O., & Ramirez-Solis, R. (2003). 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), 77-90.
Melé, M., Mattioli, K., Mallard, W., Shechner, D. M., Gerhardinger, C., & Rinn, J. L. (2017). Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res*, *27*(1), 27-37. https://doi.org/10.1101/gr.214205.116
Merabet, S., & Mann, R. S. (2016). To Be Specific or Not: The Critical Relationship Between Hox And TALE Proteins. *Trends Genet*, *32*(6), 334-347. https://doi.org/10.1016/j.tig.2016.03.004
Metzis, V., Steinhauser, S., Pakanavicius, E., Gouti, M., Stamataki, D., Ivanovitch, K., Watson, T., Rayon, T., Mousavy Gharavy, S. N., Lovell-Badge, R., Luscombe, N. M., & Briscoe, J. (2018). Nervous System Regionalization Entails Axial Allocation before Neural Differentiation. *Cell*, *175*(4), 1105-1118 e1117. https://doi.org/10.1016/j.cell.2018.09.040
Miguez, A., Ducret, S., Di Meglio, T., Parras, C., Hmidan, H., Haton, C., Sekizar, S., Mannioui, A., Vidal, M., Kerever, A., Nyabi, O., Haigh, J., Zalc, B., Rijli, F. M., & Thomas, J. L. (2012). Opposing roles for Hoxa2 and Hoxb2 in hindbrain oligodendrocyte patterning. *J Neurosci*, *32*(48), 17172-17185. https://doi.org/10.1523/JNEUROSCI.0885-12.2012
Millet, S., Campbell, K., Epstein, D. J., Losos, K., Harris, E., & Joyner, A. L. (1999). A role for Gbx2 in repression of Otx2 and positioning the mid/hindbrain organizer. *Nature*, *401*(6749), 161-164.
Molotkova, N., Molotkov, A., Sirbu, I. O., & Duester, G. (2005). Requirement of mesodermal retinoic acid generated by Raldh2 for posterior neural transformation. *Mech Dev*, *122*(2), 145-155. https://doi.org/10.1016/j.mod.2004.10.008
Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A. R., Hoberg, E., Moustakas, A., Gyllensten, U., Jones, S. J., Gustafsson, C. M., Sims, A. H., Westerlund, F., Gorab, E., & Kanduri, C. (2015). MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. *Nat Commun*, *6*, 7743. https://doi.org/10.1038/ncomms8743
Morcillo, P., Rosen, C., Baylies, M. K., & Dorsett, D. (1997). Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin

enhancer in Drosophila. *Genes Dev*, *11*(20), 2729-2740. https://doi.org/10.1101/gad.11.20.2729
Moreno, T. A., & Kintner, C. (2004). Regulation of segmental patterning by retinoic acid signaling during Xenopus somitogenesis. *Dev Cell*, *6*(2), 205-218. https://doi.org/10.1016/s1534-5807(04)00026-7
Murphy, P., & Hill, R. E. (1991). Expression of the mouse labial-like homeobox-containing genes, Hox 2.9 and Hox 1.6, during segmentation of the hindbrain. *Development*, *111*(1), 61-74.
Nakamura, T., Gehrke, A. R., Lemberg, J., Szymaszek, J., & Shubin, N. H. (2016). Digits and fin rays share common developmental histories. *Nature*, *537*(7619), 225-228. https://doi.org/10.1038/nature19322
Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., & Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science*, *347*(6225), 1017-1021. https://doi.org/10.1126/science.1262088
Nelson, C. E., Morgan, B. A., Burke, A. C., Laufer, E., DiMambro, E., Murtaugh, L. C., Gonzales, E., Tessarollo, L., Parada, L. F., & Tabin, C. (1996). Analysis of Hox gene expression in the chick limb bud. *Development*, *122*(5), 1449-1466.
Neville, S. E., Baigent, S. M., Bicknell, A. B., Lowry, P. J., & Gladwell, R. T. (2002). Hox gene expression in adult tissues with particular reference to the adrenal gland. *Endocr Res*, *28*(4), 669-673. https://doi.org/10.1081/erc-120016984
Nguyen, L., Schilling, D., Dobiasch, S., Raulefs, S., Santiago Franco, M., Buschmann, D., Pfaffl, M. W., Schmid, T. E., & Combs, S. E. (2020). The Emerging Role of miRNAs for the Radiation Treatment of Pancreatic Cancer. *Cancers (Basel)*, *12*(12). https://doi.org/10.3390/cancers12123703
Nguyen, Q., & Carninci, P. (2016). Expression Specificity of Disease-Associated lncRNAs: Toward Personalized Medicine. *Curr Top Microbiol Immunol*, *394*, 237-258. https://doi.org/10.1007/82_2015_464
Nicolas, J. F., Mathis, L., Bonnerot, C., & Saurin, W. (1996). Evidence in the mouse for self-renewing stem cells in the formation of a segmented longitudinal structure, the myotome. *Development*, *122*(9), 2933-2946.
Niederreither, K., & Dolle, P. (2008). Retinoic acid in development: towards an integrated view. *Nat Rev Genet*, *9*(7), 541-553. https://doi.org/10.1038/nrg2340
Niederreither, K., Subbarayan, V., Dolle, P., & Chambon, P. (1999). Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat Genet*, *21*(4), 444-448. https://doi.org/10.1038/7788
Nohno, T., Noji, S., Koyama, E., Ohyama, K., Myokai, F., Kuroiwa, A., Saito, T., & Tanaguchi, S. (1991). Involvement of the Chox-4 chicken homeobox genes in determination of anteroposterior axial polarity during limb development. *Cell*, *64*, 1197-1205.
Nolte, C., De Kumar, B., & Krumlauf, R. (2019). Hox genes: Downstream "effectors" of retinoic acid signaling in vertebrate embryogenesis. *Genesis*, *57*(7-8), e23306. https://doi.org/10.1002/dvg.23306
Nolte, C., Jinks, T., Wang, X., Martinez Pastor, M. T., & Krumlauf, R. (2013). Shadow enhancers flanking the HoxB cluster direct dynamic Hox expression in early heart and endoderm development. *Dev Biol*, *383*(1), 158-173. https://doi.org/10.1016/j.ydbio.2013.09.016

Nolte, C., & Krumlauf, R. (2006). Expression of Hox genes in the nervous system of vertebrates. In S. Papageorgiou (Ed.), *HOX Gene Expression* (pp. 14-41). Landes Bioscience and Springer.

Nolte, C., & Krumlauf, R. (2019). A conserved boundary element downstream of the HoxB governs regulatory interactions between the HoxB complex and the Skap1 locus in modulation of integrin signaling during mouse development. *Developmental Biology (San Diego, CA, United States)*, *DB_2018_360 in revision*.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Bluthgen, N., Dekker, J., & Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, *485*(7398), 381-385. https://doi.org/10.1038/nature11049

Nosratabadi, S., Mosavi, A., Duan, P., Ghamisi, P., Filip, F., Band, S., Reuter, U., Gama, J., & Gandomi, A. (2020). Data Science in Economics: Comprehensive Review of Advanced Machine Learning and Deep Learning Methods. *Mathematics*, *8*(10), 1799. https://doi.org/10.3390/math8101799

Nuckolls, N. L., Mok, A. C., Lange, J. J., Yi, K., Kandola, T. S., Hunn, A. M., McCroskey, S., Snyder, J. L., Bravo Núñez, M. A., McClain, M., McKinney, S. A., Wood, C., Halfmann, R., & Zanders, S. E. (2020). The. *Elife*, *9*. https://doi.org/10.7554/eLife.55694

Oates, A. C., Morelli, L. G., & Ares, S. (2012). Patterning embryos with oscillations: structure, function and dynamics of the vertebrate segmentation clock. *Development*, *139*(4), 625-639. https://doi.org/10.1242/dev.063735

Oosterveen, T., Meijlink, F., & Deschamps, J. (2004). Expression of retinaldehyde dehydrogenase II and sequential activation of 5' Hoxb genes in the mouse caudal hindbrain. *Gene Expr Patterns*, *4*, 243-247.

Oosterveen, T., Niederreither, K., Dolle, P., Chambon, P., Meijlink, F., & Deschamps, J. (2003). Retinoids regulate the anterior expression boundaries of 5' Hoxb genes in posterior hindbrain. *EMBO J*, *22*(2), 262-269. https://doi.org/10.1093/emboj/cdg029

OpenStax. (2013). Features of the animal kingdom: Figure 4. In: OpenStax CNX.

Papalopulu, N., Clarke, J. D., Bradley, L., Wilkinson, D., Krumlauf, R., & Holder, N. (1991). Retinoic acid causes abnormal development and segmental patterning of the anterior hindbrain in Xenopus embryos. *Development*, *113*(4), 1145-1158.

Papalopulu, N., Graham, A., Lorimer, J., Kenny, R., McVey, J., & Krumlauf, R. (1989). Structure, expression and evolutionary relationships of murine homeobox genes in the Hox-2 cluster. In S. de Laat, J. Bluemink, & C. Mummery (Eds.), *Cell to cell signals in mammalian development* (Vol. H26, pp. 11-21). Springer-Verlag.

Parker, H. J., & Krumlauf, R. (2017). Segmental arithmetic: summing up the *Hox* gene regulatory network for hindbrain development in chordates. *Wiley Interdiscip Rev Dev Biol*, *6*(6). https://doi.org/10.1002/wdev.286

Partanen, J., Schwartz, L., & Rossant, J. (1998). Opposite phenotypes of hypomorphic and Y766 phosphorylation site mutations reveal a function for Fgfr1 in anteroposterior patterning of mouse embryos. *Genes Dev*, *12*(15), 2332-2344.

Patel, N. S., Rhinn, M., Semprich, C. I., Halley, P. A., Dollé, P., Bickmore, W. A., & Storey, K. G. (2013). FGF signalling regulates chromatin organisation during neural differentiation via mechanisms that can be uncoupled from transcription. *PLoS Genet*, *9*(7), e1003614. https://doi.org/10.1371/journal.pgen.1003614

Patterson, L. T., & Potter, S. S. (2004). Atlas of Hox gene expression in the developing kidney. *Dev Dyn*, *229*(4), 771-779. https://doi.org/10.1002/dvdy.10474

Pattyn, A., Vallstedt, A., Dias, J. M., Samad, O. A., Krumlauf, R., Rijli, F. M., Brunet, J. F., & Ericson, J. (2003). Coordinated temporal and spatial control of motor neuron and serotonergic neuron generation from a common pool of CNS progenitors. *Genes & Development*, *17*(6), 729-737. https://doi.org/10.1101/gad.255803

Perry, R. B., Hezroni, H., Goldrich, M. J., & Ulitsky, I. (2018). Regulation of Neuroregeneration by Long Noncoding RNAs. *Mol Cell*, *72*(3), 553-567.e555. https://doi.org/10.1016/j.molcel.2018.09.021

Peterson, R. L., Papenbrock, T., Davda, M. M., & Awgulewitsch, A. (1994). The murine Hoxc cluster contains five neighboring AbdB-related Hox genes that show unique spatially coordinated expression in posterior embryonic subregions. *Mech Dev*, *47*(3), 253-260.

Phelan, K. A., & Hollyday, M. (1990). Axon guidance in muscleless chick wings: the role of muscle cells in motoneuronal pathway selection and muscle nerve formation. *J Neurosci*, *10*, 2699-2716.

Philippidou, P., & Dasen, J. S. (2013). Hox genes: choreographers in neural development, architects of circuit organization. *Neuron*, *80*(1), 12-34. https://doi.org/10.1016/j.neuron.2013.09.020

Popperl, H., Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., Mann, R. S., & Krumlauf, R. (1995). Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. *Cell*, *81*(7), 1031-1042.

Procino, A., & Cillo, C. (2013). The HOX genes network in metabolic diseases. *Cell Biol Int*, *37*(11), 1145-1148. https://doi.org/10.1002/cbin.10145

Qian, P., De Kumar, B., He, X. C., Nolte, C., Gogol, M., Ahn, Y., Chen, S., Li, Z., Xu, H., Perry, J. M., Hu, D., Tao, F., Zhao, M., Han, Y., Hall, K., Peak, A., Paulson, A., Zhao, C., Venkatraman, A., Box, A., Perera, A., Haug, J. S., Parmely, T., Li, H., Krumlauf, R., & Li, L. (2018). Retinoid-Sensitive Epigenetic Regulation of the Hoxb Cluster Maintains Normal Hematopoiesis and Inhibits Leukemogenesis. *Cell Stem Cell*, *22*(5), 740-754 e747. https://doi.org/10.1016/j.stem.2018.04.012

Qin, S., Shi, X., Wang, C., Jin, P., & Ma, F. (2019). Transcription Factor and miRNA Interplays Can Manifest the Survival of ccRCC Patients. *Cancers (Basel)*, *11*(11). https://doi.org/10.3390/cancers11111668

Quadros, R. M., Miura, H., Harms, D. W., Akatsuka, H., Sato, T., Aida, T., Redder, R., Richardson, G. P., Inagaki, Y., Sakai, D., Buckley, S. M., Seshacharyulu, P., Batra, S. K., Behlke, M. A., Zeiner, S. A., Jacobi, A. M., Izu, Y., Thoreson, W. B., Urness, L. D., Mansour, S. L., Ohtsuka, M., & Gurumurthy, C. B. (2017). Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biol*, *18*(1), 92. https://doi.org/10.1186/s13059-017-1220-4

Quinonez, S. C., & Innis, J. W. (2014). Human HOX gene disorders. *Mol Genet Metab*, *111*(1), 4-15. https://doi.org/10.1016/j.ymgme.2013.10.012

Raj, A., & Tyagi, S. (2010). Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. *Methods Enzymol*, *472*, 365-386. https://doi.org/10.1016/S0076-6879(10)72004-8

Ramirez-Solis, R., Zheng, H., Whiting, J., Krumlauf, R., & Bradley, A. (1993). Hoxb-4 (Hox-2.6) mutant mice show homeotic transformation of a cervical vertebra and defects in the closure of the sternal rudiments. *Cell*, *73*, 279-294.

Rancourt, D. E., Tsuzuki, T., & Capecchi, M. R. (1995). Genetic interaction between hoxb-5 and hoxb-6 is revealed by nonallelic noncomplementation. *Genes Dev*, *9*(1), 108-122. https://doi.org/10.1101/gad.9.1.108

Rancourt, D. E., Tsuzuki, T., & Capecchi, M. R. (1995). Genetic interaction between hoxb-5 and hoxb-6 is revealed by nonallelic noncomplementation. *Genes & Development*, *9*, 108-122.

Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D., Huang, X., Shamim, M. S., Shin, J., Turner, D., Ye, Z., Omer, A. D., Robinson, J. T., Schlick, T., Bernstein, B. E., Casellas, R., Lander, E. S., & Aiden, E. L. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell*, *171*(2), 305-320.e324. https://doi.org/10.1016/j.cell.2017.09.026

Regulski, M., McGinnis, N., Chadwick, R., & McGinnis, W. (1987a). Developmental and molecular analysis of Deformed: A homeotic gene controlling Drosophila head development. *EMBO J.*, *6*, 767-777.

Regulski, M., McGinnis, N., Chadwick, R., & McGinnis, W. (1987b). Developmental and molecular analysis of Deformed; a homeotic gene controlling Drosophila head development. *EMBO J*, *6*(3), 767-777.

Rezsohazy, R., Saurin, A. J., Maurel-Zaffran, C., & Graba, Y. (2015). Cellular and molecular insights into Hox protein action. *Development*, *142*(7), 1212-1227. https://doi.org/10.1242/dev.109785

Rhinn, M., & Dolle, P. (2012). Retinoic acid signalling during development. *Development*, *139*(5), 843-858. https://doi.org/10.1242/dev.065938

Rhinn, M., Lun, K., Luz, M., Werner, M., & Brand, M. (2005). Positioning of the midbrain-hindbrain boundary organizer through global posteriorization of the neuroectoderm mediated by Wnt8 signaling. *Development*, *132*(6), 1261-1272. https://doi.org/10.1242/dev.01685

Rieger, E., Bijl, J. J., van Oostveen, J. W., Soyer, H. P., Oudejans, C. B., Jiwa, N. M., Walboomers, J. M., & Meijer, C. J. (1994). Expression of the homeobox gene HOXC4 in keratinocytes of normal skin and epithelial skin tumors is correlated with differentiation. *J Invest Dermatol*, *103*(3), 341-346. https://doi.org/10.1111/1523-1747.ep12394888

Rijli, F. M., Matyas, R., Pellegrini, M., Dierich, A., Gruss, P., Dolle, P., & Chambon, P. (1995). Cryptorchidism and homeotic transformations of spinal nerves and vertebrae in Hoxa-10 mutant mice. *Proc Natl Acad Sci U S A*, *92*(18), 8185-8189.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., & Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, *129*(7), 1311-1323. https://doi.org/10.1016/j.cell.2007.05.022

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In.

Rossel, M., & Capecchi, M. R. (1999). Mice mutant for both *Hoxa1* and *Hoxb1* show extensive remodeling of the hindbrain and defects in craniofacial development. *Development*, *126*, 5027-5040.

Roux, M., & Zaffran, S. (2016). Hox Genes in Cardiovascular Development and Diseases. *J Dev Biol*, *4*(2). https://doi.org/10.3390/jdb4020014
Rux, D. R., & Wellik, D. M. (2017). Hox genes in the adult skeleton: Novel functions beyond embryonic development. *Dev Dyn*, *246*(4), 310-317. https://doi.org/10.1002/dvdy.24482
Ryan, J. F., Mazza, M. E., Pang, K., Matus, D. Q., Baxevanis, A. D., Martindale, M. Q., & Finnerty, J. R. (2007). Pre-bilaterian origins of the Hox cluster and the Hox code: evidence from the sea anemone, Nematostella vectensis. *PLoS One*, *2*(1), e153. https://doi.org/10.1371/journal.pone.0000153
Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature*, *489*(7414), 109-113. https://doi.org/10.1038/nature11279
Sarropoulos, I., Marin, R., Cardoso-Moreira, M., & Kaessmann, H. (2019). Developmental dynamics of lncRNAs across mammalian organs and species. *Nature*, *571*(7766), 510-514. https://doi.org/10.1038/s41586-019-1341-x
Satokata, I., Benson, G., & Maas, R. (1995). Sexually dimorphic sterility phenotypes in Hoxa10-deficient mice. *Nature*, *374*(6521), 460-463. https://doi.org/10.1038/374460a0
Sauvageau, M., Goff, L. A., Lodato, S., Bonev, B., Groff, A. F., Gerhardinger, C., Sanchez-Gomez, D. B., Hacisuleyman, E., Li, E., Spence, M., Liapis, S. C., Mallard, W., Morse, M., Swerdel, M. R., D'Ecclessis, M. F., Moore, J. C., Lai, V., Gong, G., Yancopoulos, G. D., Frendewey, D., Kellis, M., Hart, R. P., Valenzuela, D. M., Arlotta, P., & Rinn, J. L. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife*, *2*, e01749. https://doi.org/10.7554/eLife.01749
Sayli, B. S., Akarsu, A. N., Sayli, U., Akhan, O., Ceylaner, S., & Sarfarazi, M. (1995). A large Turkish kindred with syndactyly type II (synpolydactyly). 1. Field investigation, clinical and pedigree data. *J Med Genet*, *32*(6), 421-434. https://doi.org/10.1136/jmg.32.6.421
Schein, A., Zucchelli, S., Kauppinen, S., Gustincich, S., & Carninci, P. (2016). Identification of antisense long noncoding RNAs that function as SINEUPs in human cells. *Sci Rep*, *6*, 33605. https://doi.org/10.1038/srep33605
Schiemann, S. M., Martín-Durán, J. M., Børve, A., Vellutini, B. C., Passamaneck, Y. J., & Hejnol, A. (2017). Clustered brachiopod Hox genes are not expressed collinearly and are associated with lophotrochozoan novelties. *Proc Natl Acad Sci U S A*, *114*(10), E1913-E1922. https://doi.org/10.1073/pnas.1614501114
Schier, A. F., & Gehring, W. J. (1992). Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. *Nature*, *356*, 804-807.
Schilling, T. F., Sosnik, J., & Nie, Q. (2016). Visualizing retinoic acid morphogen gradients. *Methods Cell Biol*, *133*, 139-163. https://doi.org/10.1016/bs.mcb.2016.03.003
Schmidt, U., Weigert, M., Broaddus, C., & Myers, G. (2018). Cell Detection with Star-convex Polygons. In. (Reprinted from: arXiv:1806.03535v2).
Schmitz, K. M., Mayer, C., Postepska, A., & Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev*, *24*(20), 2264-2269. https://doi.org/10.1101/gad.590910
Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet*, *20*(8), 437-455. https://doi.org/10.1038/s41576-019-0128-0
Schorderet, P., & Duboule, D. (2011). Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet*, *7*(5), e1002071. https://doi.org/10.1371/journal.pgen.1002071

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N. A., Huber, W., Haering, C. H., Mirny, L., & Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, *551*(7678), 51-56. https://doi.org/10.1038/nature24281

Scott, M. P., & Weiner, A. J. (1984). 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*, 4115-4119.

Scotti, M., & Kmita, M. (2012). Recruitment of 5' Hoxa genes in the allantois is essential for proper extra-embryonic function in placental mammals. *Development*, *139*(4), 731-739. https://doi.org/10.1242/dev.075408

Seo, H. C., Edvardsen, R. B., Maeland, A. D., Bjordal, M., Jensen, M. F., Hansen, A., Flaat, M., Weissenbach, J., Lehrach, H., Wincker, P., Reinhardt, R., & Chourrout, D. (2004). Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. *Nature*, *431*, 67-71.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., & Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, *148*(3), 458-472. https://doi.org/10.1016/j.cell.2012.01.010

Sham, M. H., Vesque, C., Nonchev, S., Marshall, H., Frain, M., Das Gupta, R. D., Whiting, J., Wilkinson, D., Charnay, P., & Krumlauf, R. (1993). The zinc finger gene *Krox-20* regulates *Hoxb-2 (Hox2.8)* during hindbrain segmentation. *Cell*, *72*, 183-196.

Shao, W., & Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet*, *49*(7), 1045-1051. https://doi.org/10.1038/ng.3867

Sharpe, J., Nonchev, S., Gould, A., Whiting, J., & Krumlauf, R. (1998). Selectivity, sharing and competitive interactions in the regulation of *Hoxb* genes. *EMBO J*, *17*(6), 1788-1798. https://doi.org/10.1093/emboj/17.6.1788

Shaut, C. A., Keene, D. R., Sorensen, L. K., Li, D. Y., & Stadler, H. S. (2008). HOXA13 Is essential for placental vascular patterning and labyrinth endothelial specification. *PLoS Genet*, *4*(5), e1000073. https://doi.org/10.1371/journal.pgen.1000073

Shea, A., Harish, V., Afzal, Z., Chijioke, J., Kedir, H., Dusmatova, S., Roy, A., Ramalinga, M., Harris, B., Blancato, J., Verma, M., & Kumar, D. (2016). MicroRNAs in glioblastoma multiforme pathogenesis and therapeutics. *Cancer Med*, *5*(8), 1917-1946. https://doi.org/10.1002/cam4.775

Shimozono, S., Iimura, T., Kitaguchi, T., Higashijima, S., & Miyawaki, A. (2013). Visualization of an endogenous retinoic acid gradient across embryonic development. *Nature*, *496*(7445), 363-366. https://doi.org/10.1038/nature12037

Shrimpton, A. E., Levinsohn, E. M., Yozawitz, J. M., Packard, D. S., Cady, R. B., Middleton, F. A., Persico, A. M., & Hootnick, D. R. (2004). A HOX gene mutation in a family with isolated congenital vertical talus and Charcot-Marie-Tooth disease. *Am J Hum Genet*, *75*(1), 92-96. https://doi.org/10.1086/422015

Shubin, N., Tabin, C., & Carroll, S. (1997). Fossils, genes and the evolution of animal limbs. *Nature*, *388*(6643), 639-648. https://doi.org/10.1038/41710

Shubin, N., Tabin, C., & Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature*, *457*(7231), 818-823. https://doi.org/10.1038/nature07891

Simeone, A., Acampora, D., Arcioni, L., Andrews, P. W., Boncinelli, E., & Mavilio, F. (1990). Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. *Nature*, *346*, 763-766.

Simeone, A., Acampora, D., Nigro, V., Faiella, A., D'Esposito, M., Stornaiuolo, A., Mavilio, F., & Boncinelli, E. (1991). Differential regulation by retinoic acid of the homeobox genes of the four *HOX* loci in human embryonal carcinoma cells. *Mech Dev*, *33*(3), 215-227. https://doi.org/10.1016/0925-4773(91)90029-6

Singh, N. P., De Kumar, B., Paulson, A., Parrish, M. E., Zhang, Y., Florens, L., Conaway, J. W., Si, K., & Krumlauf, R. (2020). A six-amino-acid motif is a major determinant in functional evolution of HOX1 proteins. *Genes Dev*, *34*(23-24), 1680-1696. https://doi.org/10.1101/gad.342329.120

Sirbu, I. O., Gresh, L., Barra, J., & Duester, G. (2005). Shifting boundaries of retinoic acid activity control hindbrain segmental gene expression. *Development*, *132*(11), 2611-2622. https://doi.org/10.1242/dev.01845

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H. J., & Mann, R. S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell*, *147*(6), 1270-1282. https://doi.org/10.1016/j.cell.2011.10.053

Soshnikova, N., Dewaele, R., Janvier, P., Krumlauf, R., & Duboule, D. (2013). Duplications of hox gene clusters and the emergence of vertebrates. *Dev Biol*, *378*(2), 194-199. https://doi.org/10.1016/j.ydbio.2013.03.004

Spadaro, P. A., Flavell, C. R., Widagdo, J., Ratnu, V. S., Troup, M., Ragan, C., Mattick, J. S., & Bredy, T. W. (2015). Long Noncoding RNA-Directed Epigenetic Regulation of Gene Expression Is Associated With Anxiety-like Behavior in Mice. *Biol Psychiatry*, *78*(12), 848-859. https://doi.org/10.1016/j.biopsych.2015.02.004

Srivastava, A., Goldberger, H., Afzal, Z., Suy, S., Collins, S. P., & Kumar, D. (2015). Detection of circulatory microRNAs in prostate cancer. *Methods Mol Biol*, *1238*, 523-538. https://doi.org/10.1007/978-1-4939-1804-1_27

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol*, *22*(2), 96-118. https://doi.org/10.1038/s41580-020-00315-9

Stelnicki, E. J., Kömüves, L. G., Kwong, A. O., Holmes, D., Klein, P., Rozenfeld, S., Lawrence, H. J., Adzick, N. S., Harrison, M., & Largman, C. (1998). HOX homeobox genes exhibit spatial and temporal changes in expression during human skin development. *J Invest Dermatol*, *110*(2), 110-115. https://doi.org/10.1046/j.1523-1747.1998.00092.x

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., & Mateo, J. L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS One*, *10*(4), e0124633. https://doi.org/10.1371/journal.pone.0124633

Stern, C., Jaques, K., Lim, T., Fraser, S., & Keynes, R. (1991). Segmental lineage restrictions in the chick embryo spinal cord depend on the adjacent somites. *Development*, *113*, 239-244.

Studer, M., Gavalas, A., Marshall, H., Ariza-McNaughton, L., Rijli, F. M., Chambon, P., & Krumlauf, R. (1998). Genetic interactions between Hoxa1 and Hoxb1 reveal new roles in regulation of early hindbrain patterning. *Development*, *125*(6), 1025-1036.

Studer, M., Lumsden, A., Ariza-McNaughton, L., Bradley, A., & Krumlauf, R. (1996). Altered segmental identity and abnormal migration of motor neurons in mice lacking *Hoxb-1*. *Nature*, *384*(6610), 630-634. https://doi.org/10.1038/384630a0

Studer, M., Popperl, H., Marshall, H., Kuroiwa, A., & Krumlauf, R. (1994). Role of a conserved retinoic acid response element in rhombomere restriction of *Hoxb-1*. *Science*, *265*(5179), 1728-1732.

Suemori, H., & Noguchi, S. (2000). Hox C cluster genes are dispensable for overall body plan of mouse embryonic development. *Dev Biol*, *220*(2), 333-342.

Sun, Y., Zhou, B., Mao, F., Xu, J., Miao, H., Zou, Z., Phuc Khoa, L. T., Jang, Y., Cai, S., Witkin, M., Koche, R., Ge, K., Dressler, G. R., Levine, R. L., Armstrong, S. A., Dou, Y., & Hess, J. L. (2018). HOXA9 Reprograms the Enhancer Landscape to Promote Leukemogenesis. *Cancer Cell*, *34*(4), 643-658.e645. https://doi.org/10.1016/j.ccell.2018.08.018

Swalla, B. J. (2006). Building divergent body plans with similar genetic pathways. *Heredity (Edinb)*, *97*(3), 235-243. https://doi.org/10.1038/sj.hdy.6800872

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., & Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res*, *24*(3), 390-400. https://doi.org/10.1101/gr.163519.113

Tani, S., Chung, U. I., Ohba, S., & Hojo, H. (2020). Understanding paraxial mesoderm development and sclerotome specification for skeletal repair. *Exp Mol Med*, *52*(8), 1166-1177. https://doi.org/10.1038/s12276-020-0482-1

Taylor, H. S. (2000). The role of HOX genes in the development and function of the female reproductive tract. *Semin Reprod Med*, *18*(1), 81-89. https://doi.org/10.1055/s-2000-13478

Taylor, H. S., Arici, A., Olive, D., & Igarashi, P. (1998). HOXA10 is expressed in response to sex steroids at the time of implantation in the human endometrium. *J Clin Invest*, *101*(7), 1379-1384. https://doi.org/10.1172/JCI1057

Taylor, H. S., Vanden Heuvel, G. B., & Igarashi, P. (1997). A conserved Hox axis in the mouse and human female reproductive system: late establishment and persistent adult expression of the Hoxa cluster genes. *Biol Reprod*, *57*(6), 1338-1345. https://doi.org/10.1095/biolreprod57.6.1338

Thompson, A. A., & Nguyen, L. T. (2000). Amegakaryocytic thrombocytopenia and radio-ulnar synostosis are associated with HOXA11 mutation. *Nat Genet*, *26*(4), 397-398. https://doi.org/10.1038/82511

Tischfield, M. A., Bosley, T. M., Salih, M. A., Alorainy, I. A., Sener, E. C., Nester, M. J., Oystreck, D. T., Chan, W. M., Andrews, C., Erickson, R. P., & Engle, E. C. (2005). Homozygous HOXA1 mutations disrupt human brainstem, inner ear, cardiovascular and cognitive development. *Nat Genet*, *37*(10), 1035-1037. https://doi.org/10.1038/ng1636

Trainor, P., & Krumlauf, R. (2000). Plasticity in mouse neural crest cells reveals a new patterning role for cranial mesoderm. *Nature Cell Biology*, *2*(2), 96-102. https://doi.org/10.1038/35000051

Tumpel, S., Wiedemann, L. M., & Krumlauf, R. (2009). Hox genes and segmentation of the vertebrate hindbrain. *Curr Top Dev Biol*, *88*, 103-137. https://doi.org/10.1016/S0070-2153(09)88004-6

van den Akker, E., Fromental-Ramain, C., de Graaff, W., Le Mouellic, H., Brulet, P., Chambon, P., & Deschamps, J. (2001). Axial skeletal patterning in mice lacking all paralogous group 8 Hox genes. *Development*, *128*, 1911-1921.

van den Akker, E., Reijnen, M., Korving, J., Brouwer, A., Meijlink, F., & Deschamps, J. (1999). Targeted inactivation of Hoxb8 affects survival of a spinal ganglion and causes aberrant limb reflexes. *Mech Dev*, *89*(1-2), 103-114. https://doi.org/S0925-4773(99)00212-9 [pii]

van Scherpenzeel Thim, V., Remacle, S., Picard, J., Cornu, G., Gofflot, F., Rezsohazy, R., & Verellen-Dumoulin, C. (2005). Mutation analysis of the HOX paralogous 4-13 genes in children with acute lymphoid malignancies: identification of a novel germline mutation of HOXD4 leading to a partial loss-of-function. *Hum Mutat*, *25*(4), 384-395. https://doi.org/10.1002/humu.20155

Warot, X., Fromental-Ramain, C., Fraulob, V., Chambon, P., & Dollé, P. (1997). Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts. *Development*, *124*(23), 4781-4791.

Watanabe, T., Cheng, E. C., Zhong, M., & Lin, H. (2015). Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. *Genome Res*, *25*(3), 368-380. https://doi.org/10.1101/gr.180802.114

Webb, B. D., Shaaban, S., Gaspar, H., Cunha, L. F., Schubert, C. R., Hao, K., Robson, C. D., Chan, W. M., Andrews, C., MacKinnon, S., Oystreck, D. T., Hunter, D. G., Iacovelli, A. J., Ye, X., Camminady, A., Engle, E. C., & Jabs, E. W. (2012). HOXB1 founder mutation in humans recapitulates the phenotype of Hoxb1-/- mice. *Am J Hum Genet*, *91*(1), 171-179. https://doi.org/10.1016/j.ajhg.2012.05.018

Wellik, D. M. (2007). Hox patterning of the vertebrate axial skeleton. *Dev Dyn*, *236*(9), 2454-2463.

Wellik, D. M., Hawkes, P. J., & Capecchi, M. R. (2002). Hox11 paralogous genes are essential for metanephric kidney induction. *Genes Dev*, *16*(11), 1423-1432.

West, J. A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., Chujo, T., Yanaka, K., Kingston, R. E., Hirose, T., Bond, C., Fox, A., & Nakagawa, S. (2016). Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J Cell Biol*, *214*(7), 817-830. https://doi.org/10.1083/jcb.201601071

White, R. J., Nie, Q., Lander, A. D., & Schilling, T. F. (2007). Complex regulation of cyp26a1 creates a robust retinoic acid gradient in the zebrafish embryo. *PLoS Biol*, *5*(11), e304. https://doi.org/10.1371/journal.pbio.0050304

White, R. J., & Schilling, T. F. (2008). How degrading: Cyp26s in hindbrain development. *Dev Dyn*, *237*(10), 2775-2790. https://doi.org/10.1002/dvdy.21695

Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P. W., Stott, D., & Allemann, R. K. (1991). Multiple spatially specific enhancers are required to reconstruct the pattern of *Hox-2.6* gene expression. *Genes & Development*, *5*(11), 2048-2059.

Wilkinson, D., Bailes, J., & McMahon, A. (1987). Expression of the proto-oncogene int-1 is restricted to specific neural cells in the developing mouse embryo. *Cell*, *50*, 79-88.

Winter, J., Jung, S., Keller, S., Gregory, R. I., & Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol*, *11*(3), 228-234. https://doi.org/10.1038/ncb0309-228

Wolfgang von Goethe, J. (1952). The Metamorphosis of Plants. In G. L. Miller (Ed.). MIT Press.

Wurst, W., & Bally-Cuif, L. (2001). Neural plate patterning: upstream and downstream of the isthmic organizer. *Nat Rev Neurosci*, *2*(2), 99-108. https://doi.org/10.1038/35053516

Yamamoto, M., Takai, D., & Yamamoto, F. (2003). Comprehensive expression profiling of highly homologous 39 hox genes in 26 different human adult tissues by the modified systematic multiplex RT-pCR method reveals tissue-specific expression pattern that suggests an important role of chromosomal structure in the regulation of hox gene expression in adult tissues. *Gene Expr*, *11*(3-4), 199-210. https://doi.org/10.3727/000000003108749071

Yokouchi, Y., Sasaki, H., & Kuroiwa, A. (1991). Homeobox gene expression correlated with the bifurcation process of limb cartilage development. *Nature*, *353*, 443-445.

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

Younger, S. T., & Rinn, J. L. (2014). 'Lnc'-ing enhancers to MYC regulation. *Cell Res*, *24*(6), 643-644. https://doi.org/10.1038/cr.2014.54

Zakany, J., Gerard, M., Favier, B., & Duboule, D. (1997). Deletion of a *HoxD* enhancer induces transcriptional heterochrony leading to transposition of the sacrum. *EMBO J*, *16*(14), 4393-4402. https://doi.org/10.1093/emboj/16.14.4393

Zakany, J., Kmita, M., & Duboule, D. (2004). A dual role for Hox genes in limb anterior-posterior asymmetry. *Science*, *304*(5677), 1669-1672.

Zandvakili, A., & Gebelein, B. (2016). Mechanisms of Specificity for Hox Factor Activity. *J Dev Biol*, *4*(2). https://doi.org/10.3390/jdb4020016

Zeiske, T., Baburajendran, N., Kaczynska, A., Brasch, J., Palmer, A. G., 3rd, Shapiro, L., Honig, B., & Mann, R. S. (2018). Intrinsic DNA Shape Accounts for Affinity Differences between Hox-Cofactor Binding Sites. *Cell Rep*, *24*(9), 2221-2230. https://doi.org/10.1016/j.celrep.2018.07.100

Zhang, X., Weissman, S. M., & Newburger, P. E. (2014). Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells. *RNA Biol*, *11*(6), 777-787.

Zhao, H., Zhang, X., Frazao, J. B., Condino-Neto, A., & Newburger, P. E. (2013). HOX antisense lincRNA HOXA-AS2 is an apoptosis repressor in all trans retinoic acid treated NB4 promyelocytic leukemia cells. *J Cell Biochem*, *114*(10), 2375-2383. https://doi.org/10.1002/jcb.24586

Zhao, Y., Sun, L., Wang, R. R., Hu, J. F., & Cui, J. (2018). The effects of mitochondria-associated long noncoding RNAs in cancer mitochondria: New players in an old arena. *Crit Rev Oncol Hematol*, *131*, 76-82. https://doi.org/10.1016/j.critrevonc.2018.08.005

Zuckerman, B., Ron, M., Mikl, M., Segal, E., & Ulitsky, I. (2020). Gene Architecture and Sequence Composition Underpin Selective Dependency of Nuclear Export of Long RNAs on NXF1 and the TREX Complex. *Mol Cell*, *79*(2), 251-267.e256. https://doi.org/10.1016/j.molcel.2020.05.013

Zákány, J., Kmita, M., Alarcon, P., de la Pompa, J. L., & Duboule, D. (2001). Localized and transient transcription of Hox genes suggests a link between patterning and the segmentation clock. *Cell*, *106*(2), 207-217. https://doi.org/S0092-8674(01)00436-6 [pii]