Mémoire

Auteur : Lentjes, Colin
Promoteur(s) : Peers, Bernard
Faculté : Faculté des Sciences
Diplôme : Master en biochimie et biologie moléculaire et cellulaire, à finalité approfondie
Année académique : 2023-2024
URI/URL : http://hdl.handle.net/2268.2/21014

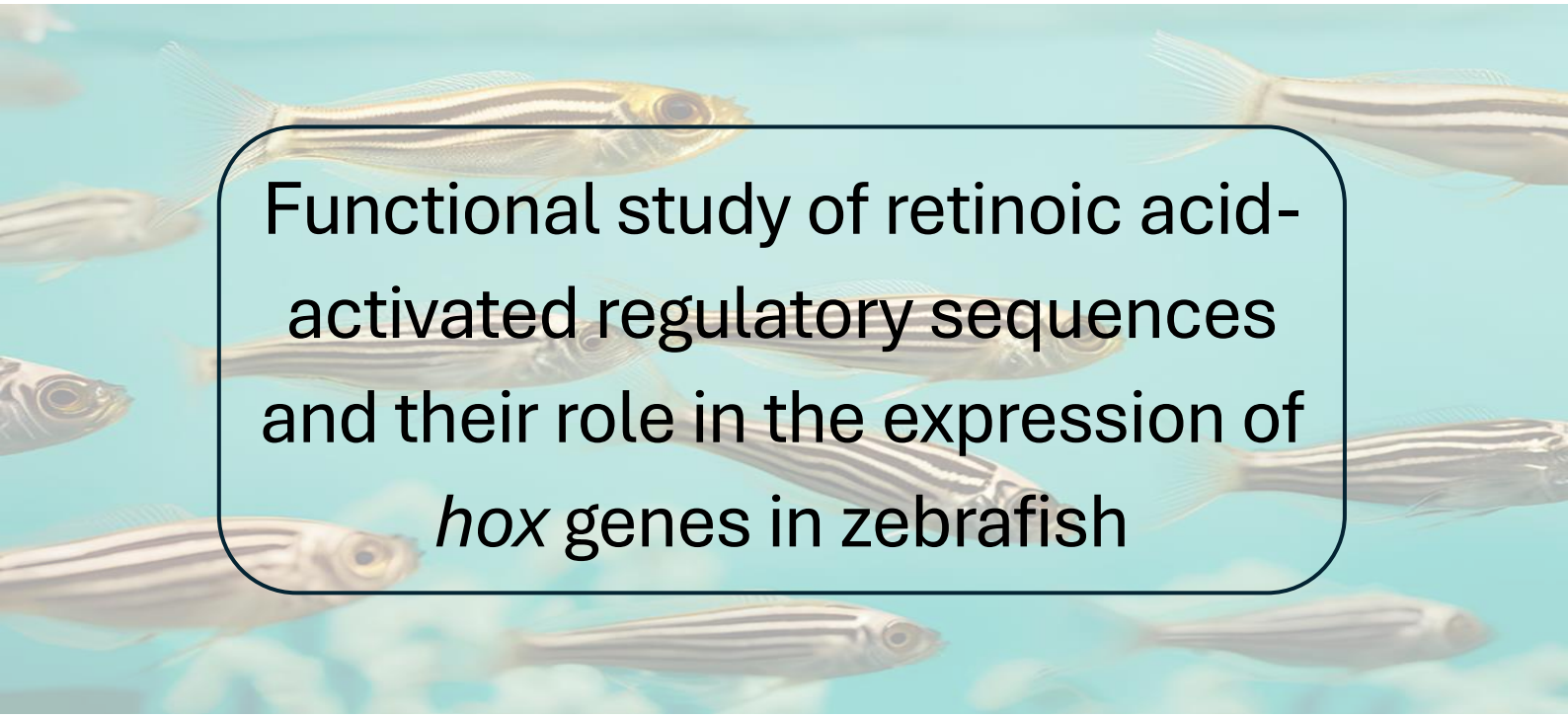
University of Liège

Department of Life Sciences

Faculty of Science

GIGA (Interdisciplinary Cluster of Applied Genoproteomics)

Laboratory of Zebrafish Development and Disease Models (ZDDM)



Functional study of retinoic acid-activated regulatory sequences and their role in the expression of *hox* genes in zebrafish

Master's thesis presented by

LENTJES Colin

With a view to obtaining a Master's degree in Biochemistry, Molecular and Cellular Biology, in depth (BBMC)

Promoter : PEERS Bernard

Supervisor : Désiront Caroline

September 2024

Acknowledgements

First, I would like to thank Drs. Marianne Voz and Bernard Peers for welcoming me into their laboratory to complete my master's thesis.

I would like to extend special thanks to my advisor, Dr. Bernard Peers, who has been consistently available to guide me and answer all my questions throughout this process. His dedication and support during the writing of this thesis have been crucial for the successful completion of this work.

I would also like to express my deep gratitude to my supervisor, Caroline Désiront, who generously shared her project, knowledge, and time to train me in the best possible way. I am thankful for her positivity and dedication to this work, as well as for her help in integrating me into the lab team. And, of course, I am also grateful for all her entertaining gossip.

I want to sincerely thank everyone in the lab—Chiara Goossens, Elisa Deom, Emma Comté, Lydie Flasse, Manon Dohet, Romane Vanhaeren, Colin Lolos, and Luca Koch. Your friendliness and support made it easy to fit in, and I really appreciated how you all made the workdays not just bearable, but actually enjoyable. The good times we had together, both in and out of the lab, really made a difference. Thanks for being such a great team to work with. I'm looking forward to continuing this adventure with you all !

I also want to thank Florent Dubois, who has been my "lab partner" throughout these six years of study. Without him, it would have been much more challenging to make it to the finish line.

Finally, I want to express my gratitude to my family—my parents and my sister—and of course Charlyne. They believed in me and supported me through the tougher times, especially during those challenging exam periods. Without their support, completing these studies wouldn't have been possible.

Table of Contents

List of Abbreviations

AF-2	Activation Function 2
ATAC-seq	Assay for Transposase-Accessible Chromatin with high-throughput sequencing
bp	Base pair
BCO1	Beta-Carotene Oxygenase 1
ChIP-seq	Chromatin Immunoprecipitation Sequencing
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CYP26	Cytochrome P450 26
DBD	DNA Binding Domain
DEAB	A Retinoic Acid Receptor Antagonist
DSB	Double-Strand Break
DNA	Deoxyribonucleic Acid
DHRS3	Dehydrogenase/Reductase (SDR Family) Member 3
Fgf8	Fibroblast Growth Factor 8
GFP	Green Fluorescent Protein
HDR	Homologous Recombination Repair
HAT	Histone Acetyltransferase
HDAC	Histone Deacetylase
HiC	High-Throughput Chromosome Conformation Capture
ISH	*In Situ* Hybridization
kb	Kilobase
LBD	Ligand Binding Domain
N-CoR	Nuclear Receptor Co-Repressor
NR	Nuclear Receptor
PTU	Phenylthiourea
RA	Retinoic Acid
RALDH	Retinaldehyde Dehydrogenase
RAR	Retinoic Acid Receptor
RARE	Retinoic Acid Response Element

RXR	Retinoid X Receptor
SMRT	Silencing Mediator of Retinoic Acid and Thyroid Hormone Receptors
TAD	Topologically Associating Domain
TALEN	Transcription Activator-Like Effector Nucleases
TIR	Terminal Inverted Repeats
Wnt	Wingless-related Integration Site
WT	Wild Type
ZFN	Zinc Finger Nucleases

1. Introduction

1.1. *Hox* genes

Understanding the genetic blueprint that orchestrates the development of complex multicellular organisms is a cornerstone of developmental biology. The complex form of multicellular organisms begins with proper embryonic development, which depends on a single-cell zygote possessing all necessary information for its ontogeny. An ancestral gene group, known as "*Hox*" genes, segment and shape the entire body of all metazoans. These genes encode transcription factors with a DNA-binding helix-loop-helix homeodomain, defined by a conserved 180 bp region (Hubert & Wellik, 2023; Ruddle et al., 1994). They were initially described in Drosophila in 1978 (Lewis, 1978) and have since been identified in numerous invertebrate species, as well as in all vertebrates. When their expression is disrupted or their coding sequences are mutated, it can lead to abnormal patterning, impaired cell differentiation, and diseases (Nolte et al., 2019).

Hox genes are ubiquitous among metazoans post-Porifera divergence and have been identified in Cnidaria (Burke et al., 1995). In bilaterians, these genes are organized into clusters, contrasting with the random genomic placement observed in Cnidaria (Duboule & Dollé, 1989). Invertebrates like Drosophila possess a single *Hox* cluster, while many vertebrates, due to whole-genome duplication events, have four clusters (*HoxA*, *HoxB*, *HoxC*, and *HoxD)* containing 39 paralogs gene in total. Fish have even more *Hox* clusters due to an additional genome duplication during teleost evolution. For example, the zebrafish genome contains seven clusters (*hoxaa*, *hoxab*, *hoxba*, *hoxbb*, *hoxca*, *hoxcb* and *hoxda)* with 48 *hox* genes, due to the deletion of some duplicated genes (Figure 1) (Amores et al., 1998; Izpisúa-Belmonte et al., 1991).

Hox proteins and their cofactors can act positively or negatively on gene transcription. Indeed, they regulate a diverse set of downstream target genes, orchestrating the activation and repression of genes and signalling pathways involved in specifying cellular identity. *Hox* genes definitely play a crucial role in regulating various aspects of gene regulatory networks that govern patterning and differentiation in vertebrate embryogenesis (Nolte et al., 2019).

1.1.1. Hox genes expression

The *Hox* clusters display a particular phenomenon : the spatial and temporal collinearity rule. Indeed, the *Hox* genes located in the 3' part of the cluster (*Hox1*, *Hox2*, …) are transcribed first and expressed in the anterior parts of the embryo, while the *Hox* genes located in the opposite side (*Hox13*, *Hox12*, …) are activated the last and expressed in the posterior part of the embryo. This phenomenon can be explained by an epigenetic control : before gastrulation, the *Hox* clusters are in an inaccessible chromatin conformation but at latter stages, the chromatin opens progressively on the 3' side of the cluster due presumably to the action of enhancers located far downstream from the *Hox* clusters. As gastrulation and somitogenesis begin first in the anterior part of embryos and continue progressively towards the posterior part, the *Hox1* and *Hox2* paralogues, the first to be expressed, appear in the anterior embryo region (Forlani et al., 2003). Subsequently, as the clusters become more accessible, expression of various *Hox* paralogues occurs, culminating in the expression of all *Hox* paralogues and ending with the *Hox13* paralogues expressed in the most posterior region of the embryo. This spatial and temporal collinearity is observed in all vertebrate *Hox* complexes and can bring an explanation to the nested patterns of *Hox* gene expression along the body axis (Figure 1, Nolte et al., 2019). Enhancer regions, whether located within or outside gene clusters, play a crucial role in activating *Hox* gene transcription, alongside chromatin conformation that can either facilitate or impede access to regulatory regions of target genes (Gould et al., 1997; Montavon et al., 2011; Spitz et al., 2001, 2003).

Studies have demonstrated that, at least for the *HoxA* cluster, sequential 3' to 5' activation is due to the presence of enhancers located far downstream from the cluster and which are activated by the Wnt signalling pathway (Neijts et al., 2016). *Hox* genes also respond to retinoic acid (RA) and genes located at the 3' end of the cluster exhibit greater responsiveness to RA compared to those at the 5' end. This differential responsiveness creates the distinct *Hox* expression pattern in the growing embryo, thus establishing the body axis of vertebrates (reviewed by Nolte et al., 2019).

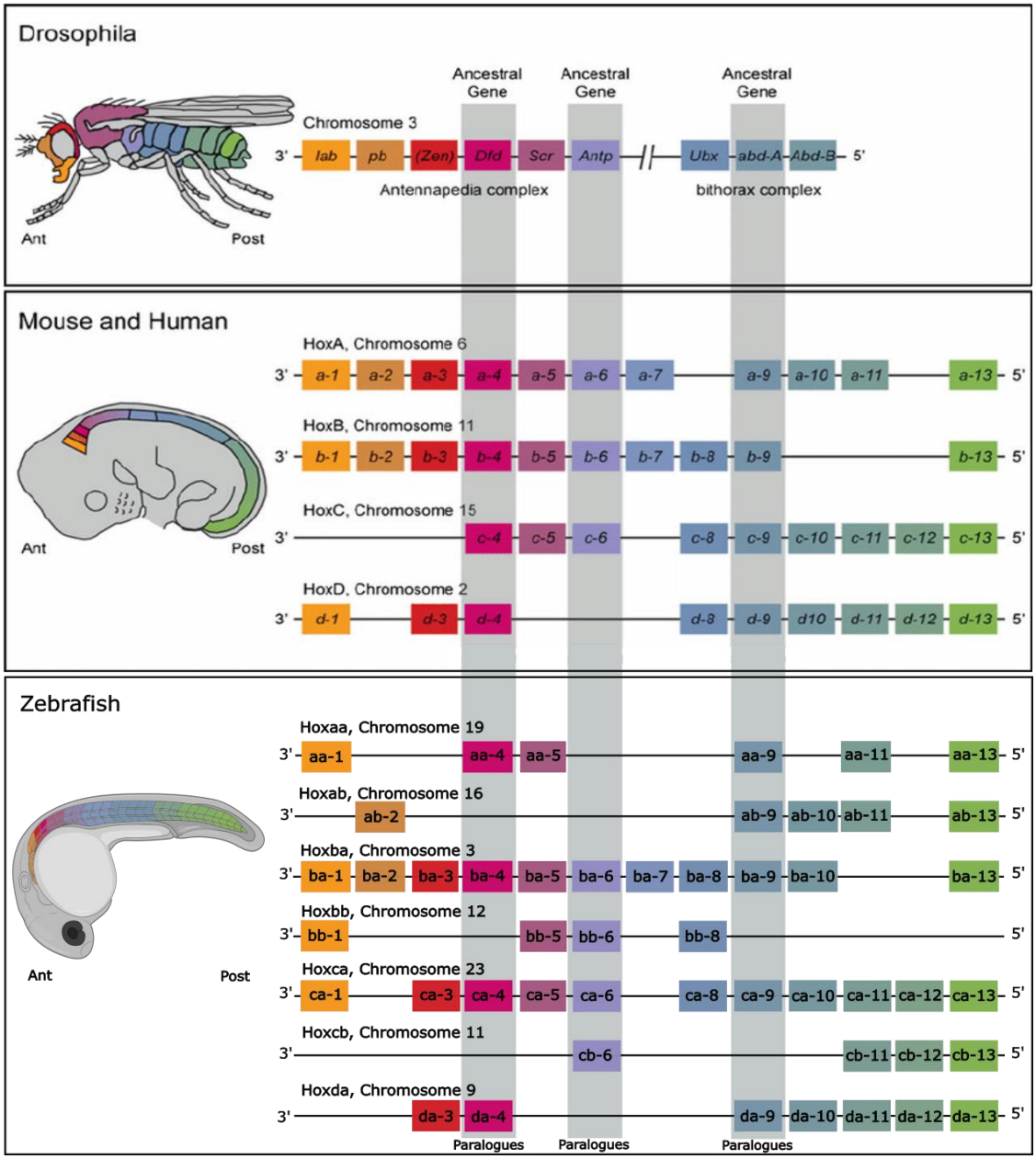


Figure 1 : Schematic Representation of *Hox* Gene Expression Patterns. The *Hox* gene clusters of mouse/human and zebrafish are compared with their paralogues in the ancestral cluster of Drosophila (the 3 grey bars highlighted 3 ancestral genes and their paralogues in mouse/human and the zebrafish). In mice and humans, 39 *Hox* genes are distributed across 4 different clusters on 4 chromosomes (6, 11, 15, and 2). In zebrafish, 48 *Hox* genes are found in 7 different clusters on 7 chromosomes (19, 16, 3, 12, 23, 11, and 9). The schematic illustrates the structural and temporal collinearity within the Hox clusters, where the lower numbered paralogues located at the anterior 3' end of the chromosome are expressed earlier and in more anterior regions of the developing embryo compared to the higher numbered paralogues located at the 5' end of the chromosome (see the correspondence between the gene colour and their expression domains along the embryonic axis) (Pang & Thompson, 2011).

1.2. Retinoic Acid

Vitamin A (retinol) is well-known for its role in the vision but is also crucial for good health and embryo development, as demonstrated in early 20th-century by experiments on rats fed a vitamin A-deficient diet (Osborne et al., 1913). In humans, this deficiency is prevalent in specific regions and lead to illness. Due to its significant health implications, extensive research has already been conducted to elucidate the vitamin A signalling pathway and its roles (Désiront, 2022).

Vitamin A is the precursor of an important signalling molecule called all-trans retinoic acid (RA). Indeed this molecule has been demonstrated to be an important morphogen for the regionalisation of the antero-posterior (A-P) axis and hindbrain patterning in vertebrate embryos (Diez Del Corral & Morales, 2014; Schubert et al., 2006). This liposoluble molecule is also known for its role in the formation of several organs such as the kidneys, liver and pancreas (Rhinn & Dollé, 2012; Stafford & Prince, 2002; Wingert et al., 2007). RA controls the expression of a large set of genes directly by binding to its receptor (RAR or the Retinoic Acid Receptor) acting as a transcription factor (Balmer & Blomhoff, 2002). Its presence is indispensable for the good development of embryos but in excess, a teratogenic effect has been observed in different species leading to problems in A-P axis patterning due to its regulation role of *Hox* genes (Avantaggiato et al., 1996; Durston et al., 1989; Schilling & Kinght, 2001). RA concentrations are strongly enzymatically regulated, permitting its synthesis from the maternal or egg-stored retinol and its degradation by subfamily of cytochrome P450 enzymes (reviewed by Rhinn & Dollé, 2012).

1.2.1. Retinoic acid metabolism

All-trans retinol (vitamin A) can't be synthesized de novo by animals and must be directly acquired from diet. Vitamin A can be taken up directly as retinol or under 2 possible precursor forms, retinyl-ester, present in animal products and carotenoids in plant products (Groff & Gropper, 2000). In vertebrate embryos, retinol is stocked in the yolk in oviparous species or acquired through placental circulation in placental species (Rhinn & Dollé, 2012).

Vitamin A is directly absorbed by intestinal cells while retinyl-ester need to be first hydrolysed. β-carotene can either be assimilated unchanged by enterocytes or converted to all-*trans*-retinaldehyde by β-carotene-15,15-monooxygenase-1 (*BCO1*) within these enterocytes (Di Masi et al., 2015). Retinol is then bound to the retinol binding protein and bring up mainly to the liver through circulation to be stored (Kanai et al., 1968).

During embryonic development, vitamin A undergoes two sequential reactions to be converted into retinoic acid. Initially, vitamin A (all-trans-retinol) is oxidised to retinaldehyde principally by the enzyme retinol dehydrogenase 10 (RDH10). The inverse reaction is carried out by a dehydrogenase

reductase (DHRS3) which controls the amount of RA produced. Retinaldehyde is then oxidised into RA by 3 different retinaldehyde dehydrogenase (RALDH1, RALDH2 (Aldh1a2) and RALDH3) acting at different moment of the development and in different tissues. RA can then be oxidised by cytochrome P450 family 26 (CYP26) enzymes and thus decreasing its concentration in different parts of the embryo. CYP26 are themselves controlled by AR, what formed a self-regulating loop. (Niederreither & Dollé, 2008). Species like the zebrafish, which can store carotenoids in the yolk, also utilize an alternative pathway for RA production. The production of retinaldehyde is ensured by the cleavage of β-carotene by β-carotene 15,15'-monooxygenase 1 (*BCO1*) (Niederreither & Dollé, 2008).

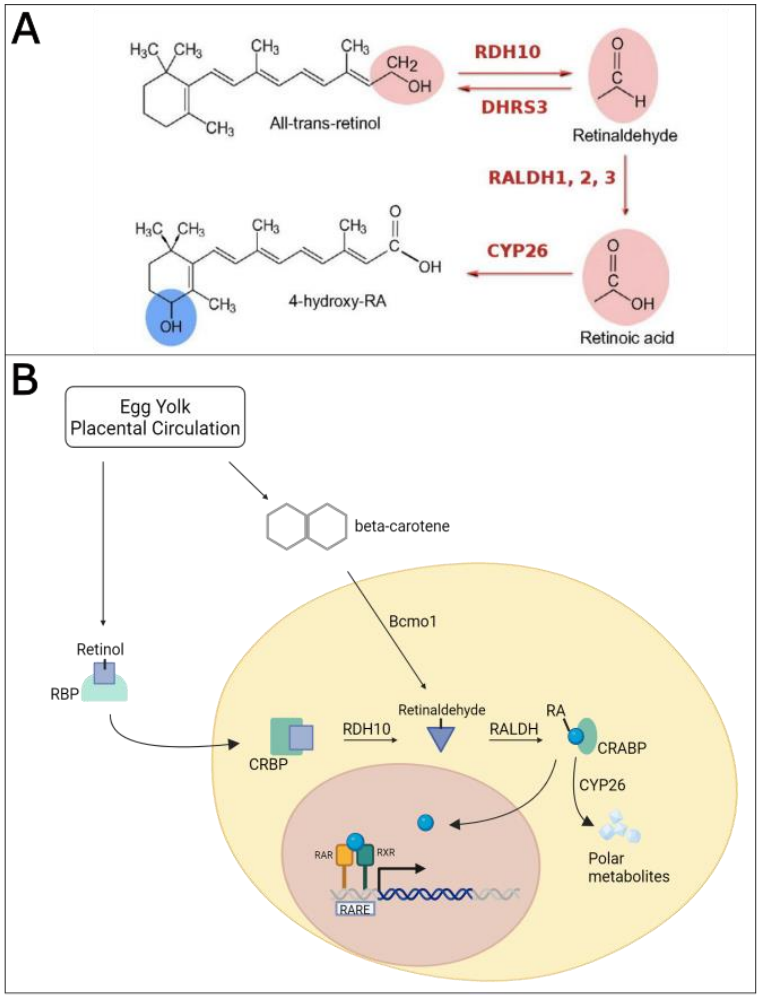


Figure 2 : Retinoic Acid pathway. A. Retinoic Acid metabolism. All-trans-retinol (Vitamin A) is oxidised by Retinol Dehydrogenase 10 (RDH10) into retinaldehyde. The reverse reaction is catalysed by Dehydrogenase/Reductase 3 (DHRS3). Retinaldehyde is subsequently converted into retinoic acid by Retinaldehyde Dehydrogenases (RALDH1, 2, 3). Retinoic acid is then degraded by oxidation into more polar metabolites by the cytochrome P450 Family 26 (CYP26) (Désiront, 2022). **B.** Schematic representation of the retinoic acid pathway. Retinoid sources can come as retinol from maternal circulation in placental species or as beta-carotene stored in the yolk sac of oviparous species. Retinol Binding Protein (RBP) transports retinol to the cell, where it is transferred intracellularly and bound to a Cellular Retinol Binding Protein (CRBP). Retinol is then converted into retinaldehyde by RDH10. Beta-carotene is cleaved by β-carotene 15,15'-monooxygenase 1 (BCO1) to produce retinaldehyde. This retinaldehyde is subsequently converted into retinoic acid (RA) by Retinaldehyde Dehydrogenases (RALDH). RA is then bound by a Cellular Retinoic Acid Binding Protein (CRABP) and either degraded by CYP26 or transported to the nucleus. In the nucleus, RA binds to its RAR receptor bound to a RARE sequence, acting as a transcription factor (Dollé, 2008).

This combination of enzymes creates RA gradients in the embryos, allowing it to fulfil its morphogen role. Retinoic acid is synthesized early in the gastrula stage embryo and forms a two-tailed gradient along the anterior-posterior axis (Figure 3). This gradient is established by differential expression of RA-synthesizing enzyme Raldh2 in the prospective paraxial mesoderm and the RA-metabolizing enzyme Cyp26a1 in the anterior and posterior extremities. RA acts as a posteriorizing agent by inducing expression of posterior genes like *Hox* genes. There is an increased concentration of RA in the mid-trunk where organs such as the pancreas and the liver form (Figure 3) (reviewed by Piersma et al., 2017). CYP26, DHSR3, and RALDH2 are directly regulated by RA, establishing a feedback loop to maintain RA levels.

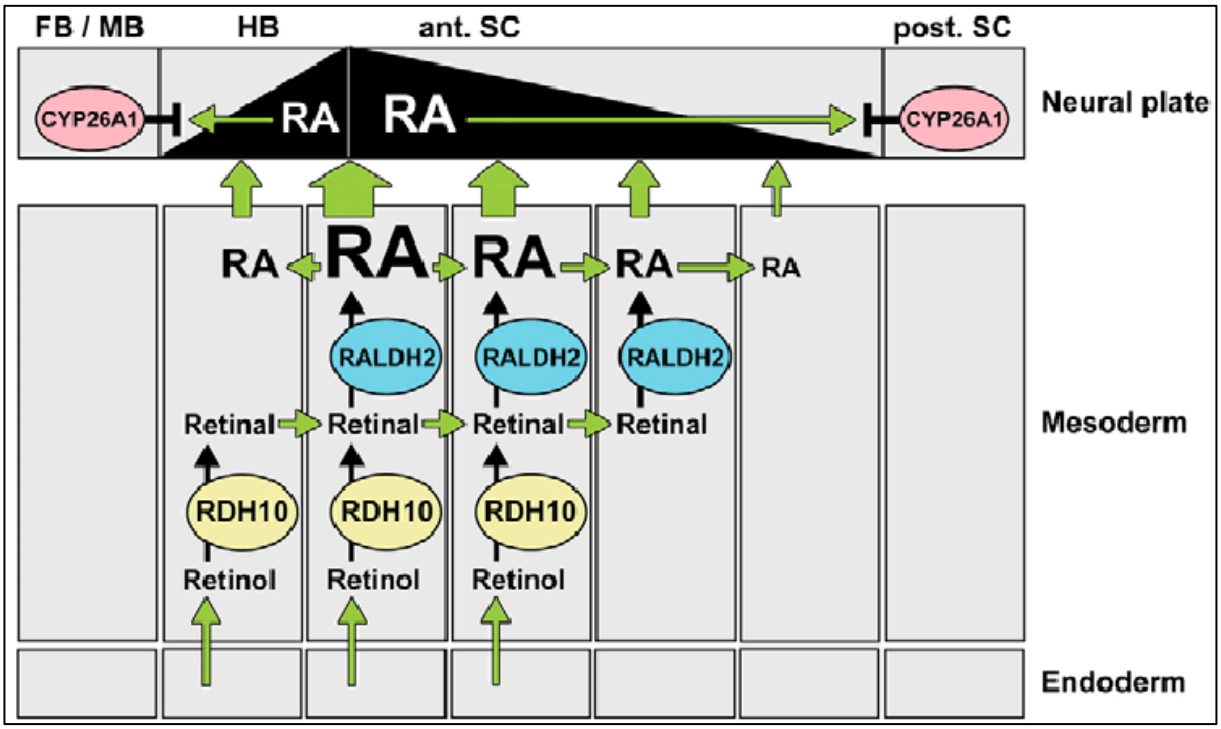


Figure 3 : Representation of the dynamic formation of retinoic acid (RA) gradients in the developing embryo. The coordinated expression and interaction of RDH10 and RALDH2 genes drive a posterior flow of retinal, resulting in the establishment of an initial RA gradient within the trunk mesoderm, peaking at the boundary between the hindbrain and spinal cord. Subsequent diffusion of RA leads to the formation of two gradients across the hindbrain and spinal cord. These gradients are further shaped by the enzymatic degradation of RA by CYP26A1 at both the anterior (ant.) and posterior (post.) ends of the neural plate. Key regions include the forebrain (FB), midbrain (MB), hindbrain (HB), and spinal cord (SC) (Strate et al., 2009).

1.2.2. Retinoic acid receptors

As a lipophilic molecule, retinoic acid diffuses into the cell nucleus and interacts with specific nuclear receptors (RAR) to induce transcription of various target genes. In metazoans, nuclear receptors (NRs) form a superfamily of receptors binding molecules like steroids, retinoids, and other lipophilic hormones. They act as transcription factors, regulated by their ligand, in various functions, notably in development. All NRs share a similar canonical structure with two well-conserved domains. The first one is the central DNA-binding domain (DBD) consisting of two zinc-finger motifs and the second one is the C-terminal ligand-binding domain (LBD). A nuclear localization signal is also present in these receptors and can be at different places (Petkovich & Chambon, 2022; Robinson-Rechavi et al., 2003).

Retinoic acid receptors (RARs) are NRs that contain five function-specific domains (A to F, from the N-terminal to the C-terminal end), as shown in Figure 4. Regions C and E are the two conserved regions, the DBD and the LBD, respectively. The other four regions are less conserved. The LBD is crucial for the specific binding of retinoic acid and contains the ligand-dependent activation function-2 (AF-2). It also plays a major role in dimerization by interacting with retinoid X receptors (RXRs) to form heterodimers. The A/B region is known as the N-terminal domain (NTD) and contains the AF-1 transcription activation domain. The role and function of F domain remain unknown for the moment (Rochette-Egly & Germain, 2009).

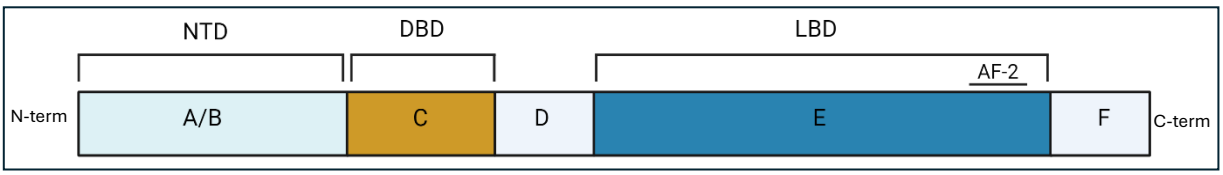


Figure 4 : Schematic representation of the five domains RAR protein. RARs exhibit a structure with five conserved domains labelled A to F. Region C includes the DNA-binding domain (DBD), while region E encompasses several domains such as the ligand-binding domain (LBD), the AF-2 domain and the dimerization domain. Additionally, it has phosphorylation sites. The N-terminal domain (NTD), comprising regions A and B, features a proline-rich motif with phosphorylation sites for Cdks and MAPKs (Rochette-Egly & Germain, 2009).

RARs enable all the diverse functions of RA. 3 receptors exist in mammals, RAR_{α} – RAR_{β} - RAR_{γ}, as well as 3 related retinoid-X-receptor ($RXR_{\alpha\text{-}\beta\text{-}\gamma}$) that make heterodimer with RAR. Splicing even increases the diversity of receptors with 2 isoforms for RAR_{α} and RAR_{γ} and 4 for RAR_{β} as well as 2 isoforms for each of the RXRs. During development, RARα is expressed ubiquitously throughout the embryo, while RARβ is specifically expressed in neurogenic and epithelial tissues, and RAR_{γ} is expressed in mesenchymal tissues and skin epithelium (reviewed by Petkovich & Chambon, 2022; Robinson-Rechavi et al., 2003). In zebrafish, some of these genes were lost during evolution, leaving only 2 isoforms of RAR_{α} and 2 isoforms of RAR_{γ}, with no remaining RAR_{β} (Bertrand et al., 2007).

RARs function as heterodimers, typically binding with one RXR partner to enhance their DNA binding capabilities and initiate RA signalling. This dimer is bound to specific DNA sequences, called Retinoic Acid Response Elements (RAREs) present in regulatory sequences of target genes, through the DBD. These RARE regions are composed of two direct repeats of the motif: 5'-RGKTCA-3' (R = A/G and K = G/T). These repeats are typically separated by 5 nucleotides (DR5) but can also be separated by 1 or 2 nucleotides (DR1 or DR2) (Bastien & Rochette-Egly, 2004; Moutier et al., 2012). The specific combinations of RAR/RXR heterodimers with repressor/activator complexes dictate the activation of RA signalling pathways, allowing for the regulation of distinct target genes that mediate the various effects of RA, including the modulation of axial and limb patterning during early embryo development, as well as organogenesis in later stages (Petkovich & Chambon, 2022).

RAR/RXR dimers adopt two conformations and follow the typical three-step mechanism of NRs: repression, derepression, and transcription activation (Figure 5). The first conformation is the inactive form in the absence of RA binding to the receptor. In this conformation, there is strong repression of target gene promoters. This repression occurs due to the binding of the dimeric receptor to a silencing mediator (SMRT) and Nuclear Receptor Corepressor (N-CoR1-2). This facilitates the recruitment of factors that encode histone deacetylase (HDAC) activity, leading to the closure of chromatin near the target promoter and preventing transcription (**Repression**). Other gene repression complexes, such as topoisomerase IIβ, polycomb group proteins, and calmodulin kinase, have also been shown to be recruited to unliganded RAR/RXR complexes (reviewed by Bastien & Rochette-Egly, 2004; Petkovich & Chambon, 2022; Robinson-Rechavi et al., 2003). The AF-2 activation domain, a conserved alpha-helical motif within the LBD, facilitates the release of N-CoR via conformational changes of the LBD following RA binding to the RAR/RXR dimer. Additionally, this domain enables the recruitment of coactivator proteins including some with histone acetyltransferase (HAT) activity promoting chromatin decompaction, facilitating the transcription of target genes (**Derepression**). This HAT complex dissociates to allow a second coactivator complex, such as the mediator complex, to interact with the transcription machinery and activate the transcription of these genes (**Transcription activation**). However, the activity of these complexes depends on the specific receptor subtype and the RARE domain they bind to (reviewed by Bastien & Rochette-Egly, 2004; Petkovich & Chambon, 2022; Robinson-Rechavi et al., 2003). Interestingly, studies on Fgf8 have shown that RARE can mediate gene repression rather than activation, contradicting the currently understood mechanism of action. The cause of this opposing effect, however, remains unknown (Kumar & Duester, 2014).

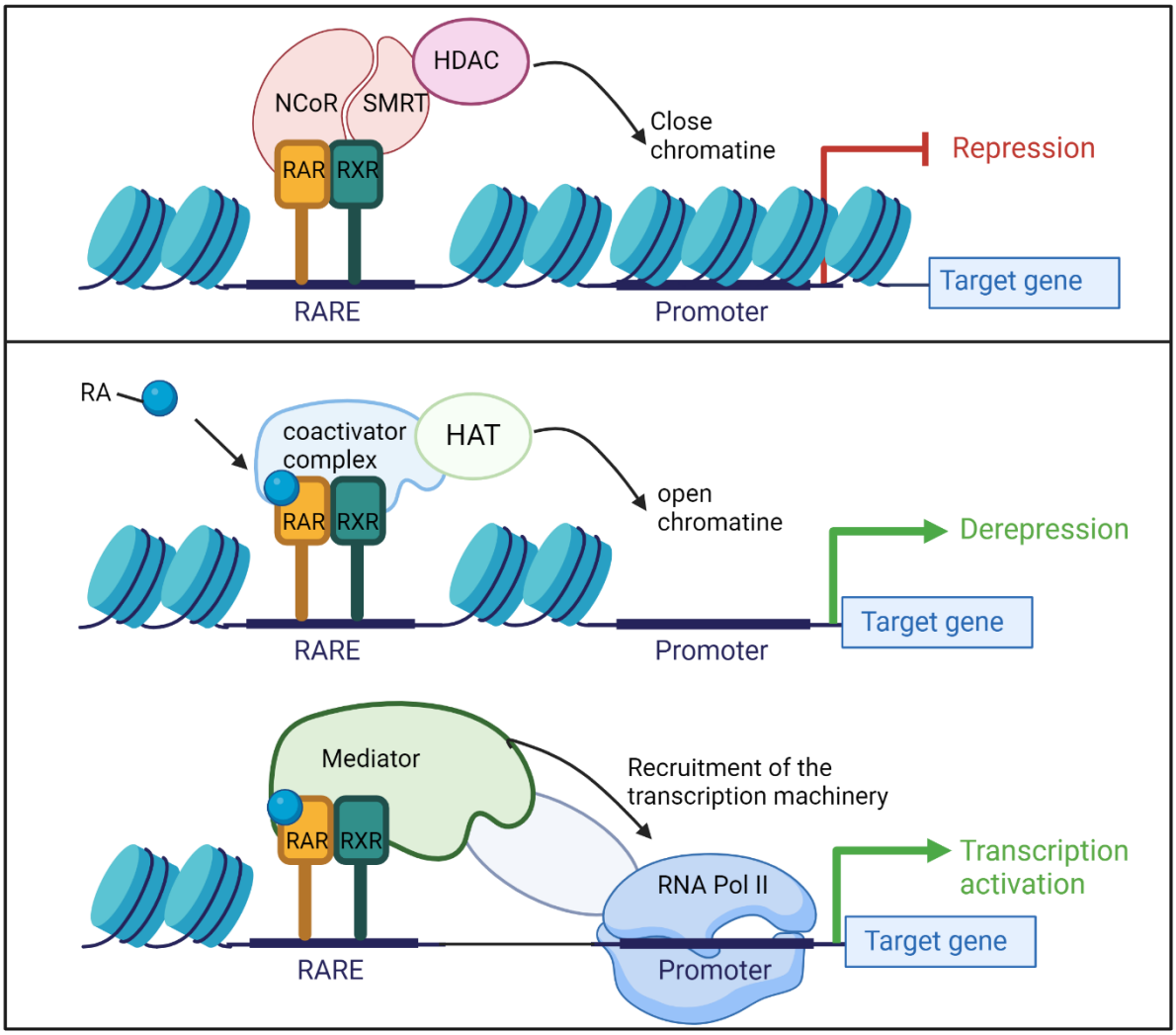


Figure 5 : Retinoic Acid receptor mechanism in 3 steps. A. In the absence of retinoic acid, the RAR/RXR receptor dimer binds to the Retinoic Acid Response Element (RARE), accompanied by a corepressor complex (NCoR/SMRT) and proteins containing histone deacetylase (HDAC) activity. This complex functions to repress transcription by condensing chromatin structure (REPRESSION) **B.** When RA binds to the RAR receptor, the corepressor dissociates, and a coactivator complex, including histone acetyltransferase (HAT) proteins, binds to the dimer, leading to chromatin opening (DEREPRESSION). **C.** In the final step, the coactivator complex dissociates, and a mediator complex binds to the receptor dimer, facilitating the recruitment of the transcriptional machinery (BioRender, s. d.).

1.3. Long-range gene regulation

It has long been established that *Hox* genes are direct targets of RA signalling through the RAREs sequences present in the clusters (Conlon & Rossant, 1992; reviewed by Nolte et al., 2019). While studies on RARE-based enhancers suggest they primarily initiate transcription of adjacent genes, the findings resulting from the ChIPseq experiment of Ana López-Pérez in my host laboratory suggest that RAREs could also regulate multiple *Hox* genes over long distances (>50 kb) as they can be found in long intergenic region (López-Pérez et al., 2021).

In contrast to prokaryotes, where a single regulatory element can control the expression of multiple genes involved in the same process and transcribed in the same operon, higher eukaryotes

possess multiple regulatory elements that can each regulate the same gene, often acting over long distances. Enhancers, such as RAREs, can be located at various positions relative to their target genes, including upstream, downstream, or within introns. They may also be located within introns of unrelated genes and act at long distance on neighbouring genes (reviewed by Schoenfelder & Fraser, 2019). To delimitate these long-distance interactions, the genome is organized in relatively independent domains named Topologically Associating Domain (TAD) where interactions are facilitated within a TAD, which are typically around 1000 kb in length, and generally not between two different TAD (Figure 6A). Long range interactions between enhancers and promoters are mediated by DNA loops within a TAD and can be detected by conformation capture technics such as HiC (reviewed by Cavalheiro et al., 2021; Dixon et al., 2012). Loops and TADs are cohesin-dependent structures, and within a TAD, multiple enhancer-promoter interactions may exist, underscoring their importance in gene regulatory mechanism (Figure 6B) (Razin et al., 2022; Schoenfelder & Fraser, 2019).

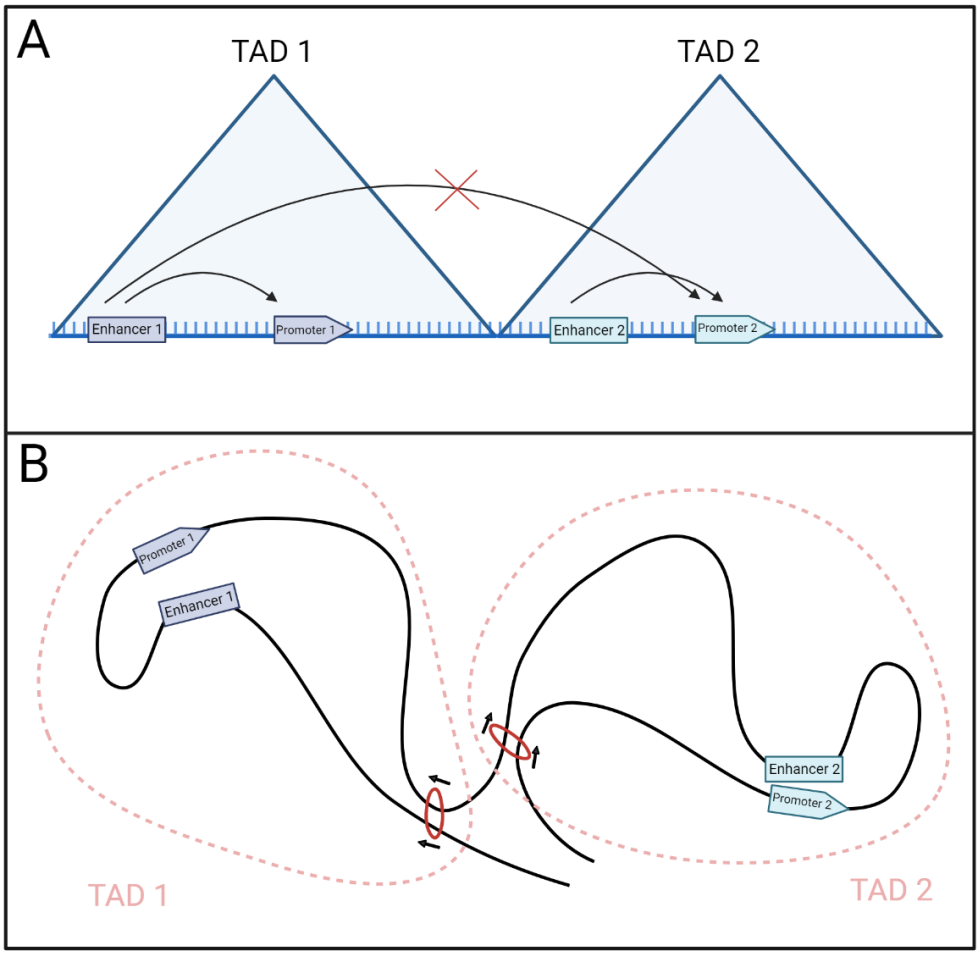


Figure 6 : Simplified Schematic Representation of TADs. (A) Schematic representation of a genomic region divided into 2 TADs represented by the light blue triangles. The active bound between an enhancer and a promoter is represented by the non-barred black arrows. The enhancer/promoter interaction are favourited inside a same TAD. **(B)** Schematic 3D organization of the genomic region comprising the 2 TADs. TAD formation is directed by a cohesin (red circle) and CCTF (CCCTC-binding factor) (black arrows) complex also permitting the formation of enhancer/promoter bounds (BioRender, s. d).

1.4. Previous findings

A previous study conducted in my host laboratory by Ana López-Pérez integrated RNA-seq, ChIP-seq, and ATAC-seq experiments, revealing the presence of four RAREs region (R1, R2, R3, and R4) downstream of the zebrafish *hoxbb* cluster, the R1 RARE being located just upstream from *skap1* gene and the 3 others within *skap1* (Figure 7A). Conservation of the R2 and R4 regulatory regions was indicated by the presence of similar RAR sites in mice by ChIP-seq (Chatagnon et al., 2015). Furthermore, the sequence of the second RARE of the R2 region (DR5-RAR-*skap1*) is the most conserved sequence between different species including human and mice, suggesting a critical role in gene regulation. While the R1-R4 are closed to *skap1*, this gene is not regulated by RA (Figure 7B) indicating that these RARE probably act at long distances on neighbouring genes. Furthermore, there is a conserved synteny between *HoxB* cluster and *Skap1* from cephalochordate to vertebrate, this suggest that regulatory elements near *Skap1*, like R1-R4, act at long distance on the neighbouring *HoxB* genes. Ana López-Pérez (Figure 7B) also demonstrated the RARE activity of the DR5-RAR-*skap1* RARE site. The results indicate an increase in GFP fluorescence upon addition of RA to the embryos. Conversely, a decrease in GFP fluorescence is observed when BMS493, an antagonist of RA for RAR binding, is introduced to the embryo culture medium. Furthermore, the activity of the DR5-*skap1* RARE site is very similar to the expression pattern of *hoxb1b* (López-Pérez et al., 2021).

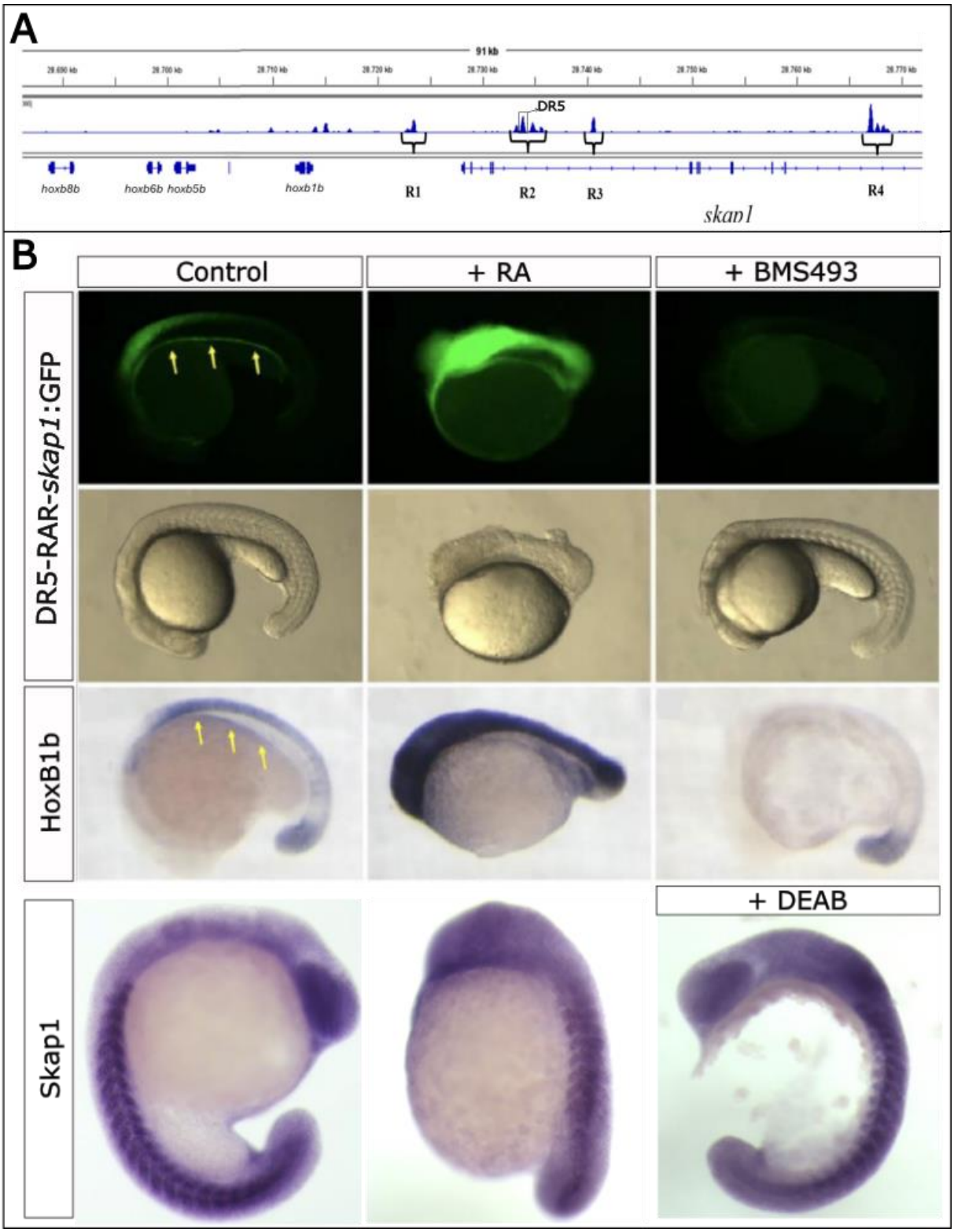


Figure 7 : Analysis of RARE around the *hoxbb* – Skap1 locus. A. Integrated data from ChIP-Seq analysis for the identification of RAR binding sites. Four RAR-binding regions (R1, R2, R3, and R4) were identified downstream of the Hoxbb cluster and upstream (R1) or within (R2, R3, and R4) the Skap1 gene (Désiront, 2022). **B.** Images of DR5-RARE:GFP transgenic embryos treated with DMSO (control), RA, or BMS493. Upper panels show GFP fluorescent expression, while lower panels depict embryo morphology. The last two lines present In Situ hybridization results for HoxB1b and Skap1 genes in embryos subjected to the same treatments as those in the R2 experiment, except for the third photograph in the last line, which shows an embryo treated with DEAB instead of BMS493 (López-Pérez et al., 2021).

During her master's thesis, Caroline Désiront analysed the activity of four regions : R1, R2, R3 and R4, containing each RARE sites, using transgenesis in zebrafish. She demonstrated that R2 and R4 sites confer strong expression to the GFP reporter gene in the presence of exogenous RA and these sites have no enhancer activity in the presence of DEAB, a pharmacological drug blocking the synthesis of RA by inhibiting aldh1a2 (Figure 8). The expression patterns of GFP closely resembled those of *hoxbb* genes. These results demonstrated the RARE activity of R2 and R4 elements and strongly suggested a role in *hoxbb* genes regulation (Désiront, 2022). Additionally, HiC analysis of adult zebrafish skeletal muscle permitted the identification of interactions between RARE elements (R1, R2, R3, and R4) and all promoters of the *hoxbb* gene cluster (Yang et al., 2020), suggesting a contribution of these RARE sequences to the initial activation of the upstream *hoxbb* cluster. Furthermore the activation of the *Hox* clusters is proposed to be controlled by enhancers located far downstream and allow progressive chromatin opening on the 3' side (Deschamps & Duboule, 2017; Noordermeer et al., 2011). This suggested that the R1-R4 elements could act in this chromatin opening. The goal of my master's thesis is to verify this hypothesis using zebrafish as a model organism.

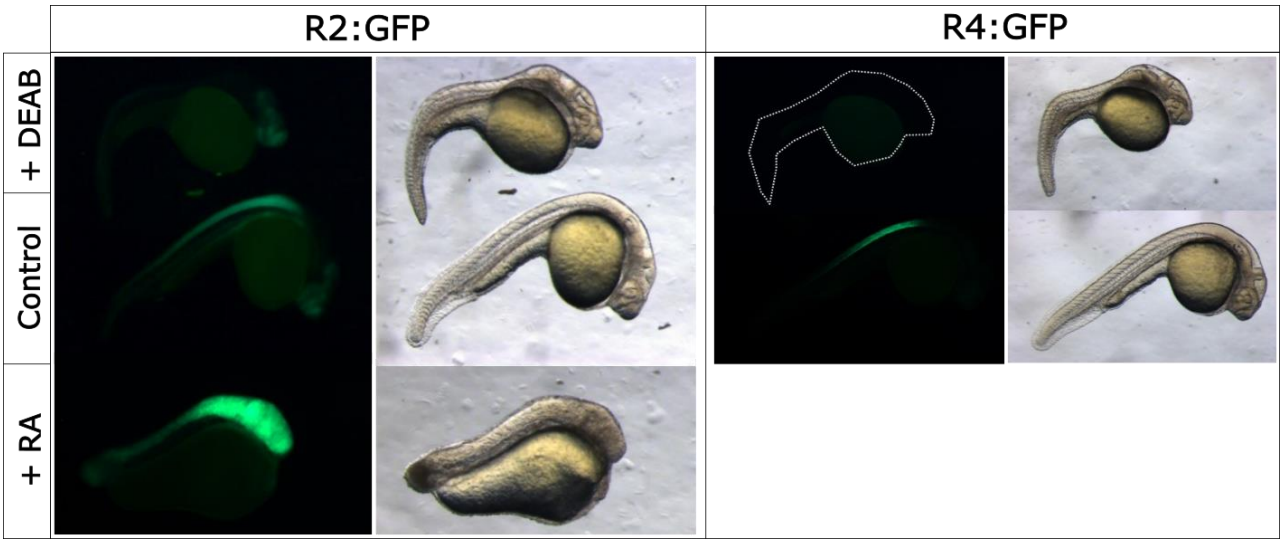


Figure 8 : Pictures of R2:GFP and R4:GFP transgenic lines treated with DMSO (control), RA and DEAB. For each transgenic, left panels show GFP fluorescent expression, while right panels show embryo morphology (Désiront, 2022).

1.5. The zebrafish as model animal

The zebrafish (*Danio rerio*) is an animal model which becomes more and more popular as it is a vertebrate offering many advantages compared to other models. It is easy to breed, allowing abundant embryo production. Its development is both rapid and external, and the embryos are transparent during early stages, facilitating direct observation of embryogenesis. Additionally, the zebrafish is easily genetically tractable with numerous available tools for developmental studies.

Techniques such as *in situ* hybridization and transgenesis are employed to study gene expression, CRISPR technology can be used to modify the genome and transcriptome, and protein levels can be altered in vivo by injecting morpholinos or mRNA (Hawkins & Wingert, 2023; Veldman & Lin, 2008). In addition to these qualities, the roles and pathways of RA are well conserved from zebrafish to humans, and the use of zebrafish as a model organism has proven successful in previous studies. Therefore, the zebrafish is an excellent model organism for this project (Hawkins & Wingert, 2023).

1.5.1. Transgenesis

To study the function of genes or regulatory elements, *in vivo* transgenesis is an interesting tool that can be easily used in zebrafish embryos due to the autonomously active Tol2 transposon discovered in the medaka fish (*Oryzias latipes*). It encodes a transposase protein capable of catalysing the transposition of a non-autonomous Tol2 construct, which includes a specific 200 bp region on the left side and a 150 bp region on the right side of the construct. These sequences also contain 12 bp terminal inverted repeats (TIRs) essential for the transposition of inserts up to 11 kb. This insert could include a non-coding sequence, such as a promoter or enhancer, followed by the GFP gene used as a reporter gene, to study the function of specific regulatory regions (Figure 9). The Tol2 technique is more efficient than plasmid injection, which has a germline transmission rate of only 5%, and is easier to manipulate than the pseudotyped retrovirus system. In zebrafish, the donor plasmid and synthetic RNA coding for the Tol2 transposase are microinjected into fertilized 1-cell stage eggs. The insert of interest is then stably integrated into the zebrafish genome with an integration frequency of 50 to 70% (Désiront, 2022; Kawakami, 2007). Furthermore, this system enables direct and straightforward visualization of the response of certain genes or regulatory sequences to specific drugs, thereby simplifying the study of their roles.

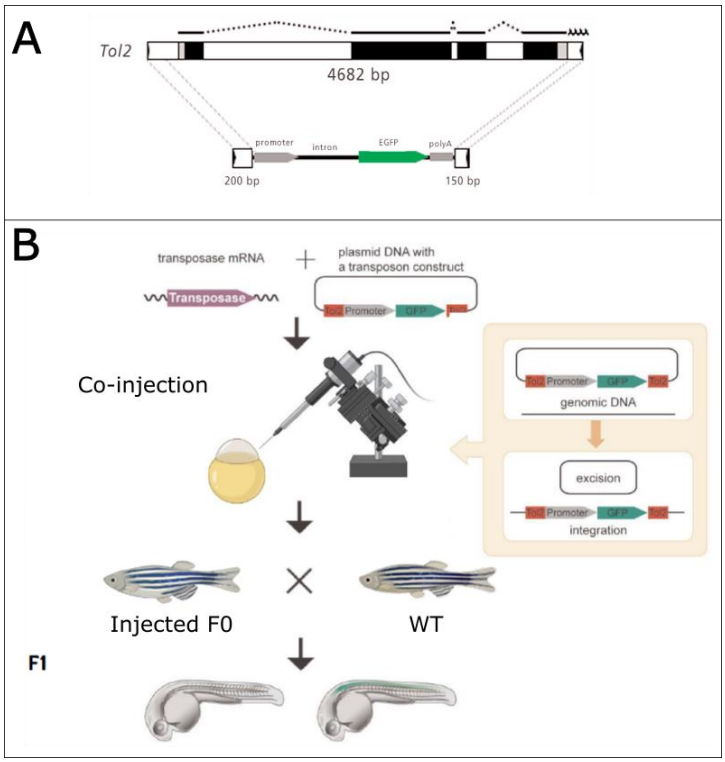


Figure 9 : Tol2 transposon principle. (A) Representation of the Tol2 transposon and the minimal Tol239 vector. The Tol2 element is shown at the top, with coding regions in black and non-coding regions in gray. The 12 base pair terminal inverted repeats (TIRs) are represented by the black arrows at the extremities. At the bottom, the minimal Tol2 vector, which includes the green fluorescent protein (GFP) expression cassette, contains 200 and 150 base pairs from the left and right ends of the transposon, including the TIR sequences. **(B)** Transgenesis with the Tol2 vector in zebrafish. Synthetic Tol2 transposase mRNA as well as a donor plasmid containing a promoter and GFP gene flanked by Tol2 sequences are co-injected into fertilized zebrafish eggs. The GFP construct is excised and integrated into the genome, and these insertions are passed to the F1 generation. Crossing wild-type fish with F0 injected fish results in heterozygous F1 offspring, yielding both transgenic and non-transgenic zebrafish (Désiront, 2022; Kawakami, 2007)

1.5.2. CRISPR

Mutagenesis is a powerful tool for studying the roles of genes or regulatory sequences. Initially, random mutagenesis was used, but this approach requires the subsequent genome localisation of the chosen mutation, which represent a difficult task. Targeted mutagenesis advanced significantly with the discovery of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system in 2012, based on a bacterial defence mechanism against phage infection, which enables the inactivation of 'intrusive' genes. It has replaced the zinc finger nuclease (ZFNs) and transcription activator-like effector nucleases (TALENs) systems due to its simplicity and efficiency in zebrafish. The most commonly used CRISPR tool is the CRISPR/Cas9 complex. The guide RNA (gRNA), composed of the CRISPR RNA (crRNA) and the trans-activating CRISPR RNA (tracrRNA), binds to the Cas9 endonuclease via the tracrRNA and is guided to the target DNA by the crRNA. Cas9 induces double-strand breaks (DSBs) in the target sequence, activating the cell's repair mechanisms (Figure 10) (Irfan et al., 2024). The non-homologous end joining (NHEJ) mechanism leads to random insertions or deletions in the targeted sequence, often resulting in gene knockout by changing the open reading frame. If two CRISPR/Cas9 complexes are targeting sequences of the same gene locus, a large deletion can be generated between the two targeted sequences. The homology-directed repair (HDR) mechanism enables precise gene editing by providing a template for the repair of the sequence (Irfan et al., 2024).

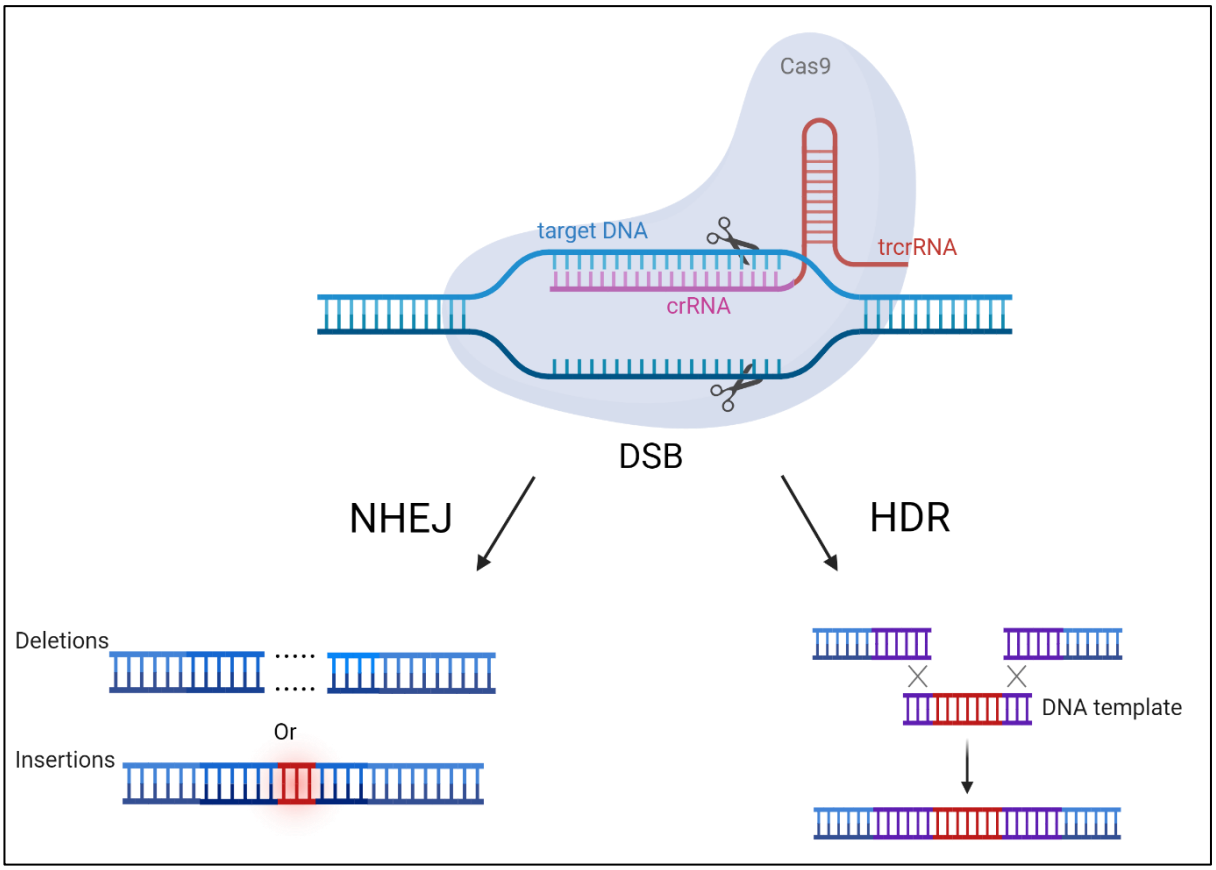


Figure 10 : CRISPR Cas9 experiment. Cas9 protein/mRNA and the specific guide RNA (gRNA, composed of crRNA and tracrRNA) are injected into embryos, inducing a double-stranded break (DSB) at the targeted sequence. This break can then be repaired by two different mechanisms: non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ often results in random insertions or deletions at the targeted site. In contrast, HDR enables precise modifications to the gene by co-injecting a template that serves as a model for the repair (BioRender, s. d).

2. Objectives

This project builds upon previous findings in my host laboratory and aims to further elucidate the role of the previously identified RARE regions within the *skap1* gene (R1 to R4, see Figure 6) on the *hoxbb* cluster. Indeed, as it was shown that the chromatin opening of the *HoxA* cluster is dependent of Wnt-regulated enhancer located far downstream the cluster, we can propose a similar hypothesis that the progressive activation of the *hoxbb* cluster could be controlled by the RA-dependant regulatory regions R1-R4.

This hypothesis will be tested by employing three distinct deletion mutants – specifically, an R2 mutant, an R1-R4 mutant, and R2-R4 mutant (Figure 11) – and by determining if the expression profile of the *hoxbb* genes is affected in these mutants by performing *in situ* hybridization.

The RARE activity of the R3 enhancer region will be assessed using transgenic fish, which will be cultured under RA and DEAB conditions, following the methodology previously employed by Caroline Désiront for R2 and R4 transgenics (Figure 8). Due to issues with the breeding of R1 transgenic fish, this test could not be conducted for the R1 region.

To further investigate the mechanisms of the R2 region, transgenic embryos were cultured under various DEAB conditions. Due to difficulties encountered with the breeding of R4 transgenic fish, this test could not be performed for the R4 region.

Finally, I also contribute to a ChIP-seq experiment in order to verify the interactions of retinoic acid receptors (RARs) to RARE regions in the zebrafish genome at the end of gastrulation. One cell-stage embryos have been injected with RAR_{α} mRNA and the ChIP-seq has been conducted on these embryos. The experiment did not yield any results; therefore, the results section will focus solely on the findings from the *in situ* observations and the experiments on transgenic fish.

3. Results

3.1. RAREs region control *hoxbb* cluster

3.1.1. General strategy used to investigate the role of the four RAREs regions.

To investigate the function of the 4 identified RARE regions (R1, R2, R3 and R4) on the expression of the *hoxbb* genes, the *hoxbb* expression patterns were compared at several developmental stage by *in situ* hybridization (ISH) experiments in embryos harbouring deletions for these regions and in wild type embryos (figure 11). To that end, fish heterozygous for each deletion were crossed together (incrosses) giving clutches of embryos, 25% being homozygous for the deletion mutation, 50 % being heterozygous and 25% being wild type (WT). Thus, the ISH performed on the whole clutch allows to compare directly the *hoxbb* expression level from the three genotypes. If a variation in the expression profile is observed between embryos of the same clutch, we next verify by genotyping each embryo that the expression level difference depends on the genotype. Nevertheless, in order to verify directly that the variation of expression profile is due to the difference of genotype, an ISH was also performed on a clutch of wild type embryos (obtained by crosses of wild type fish). In this control condition, all embryos are expected to display the same expression profile and level.

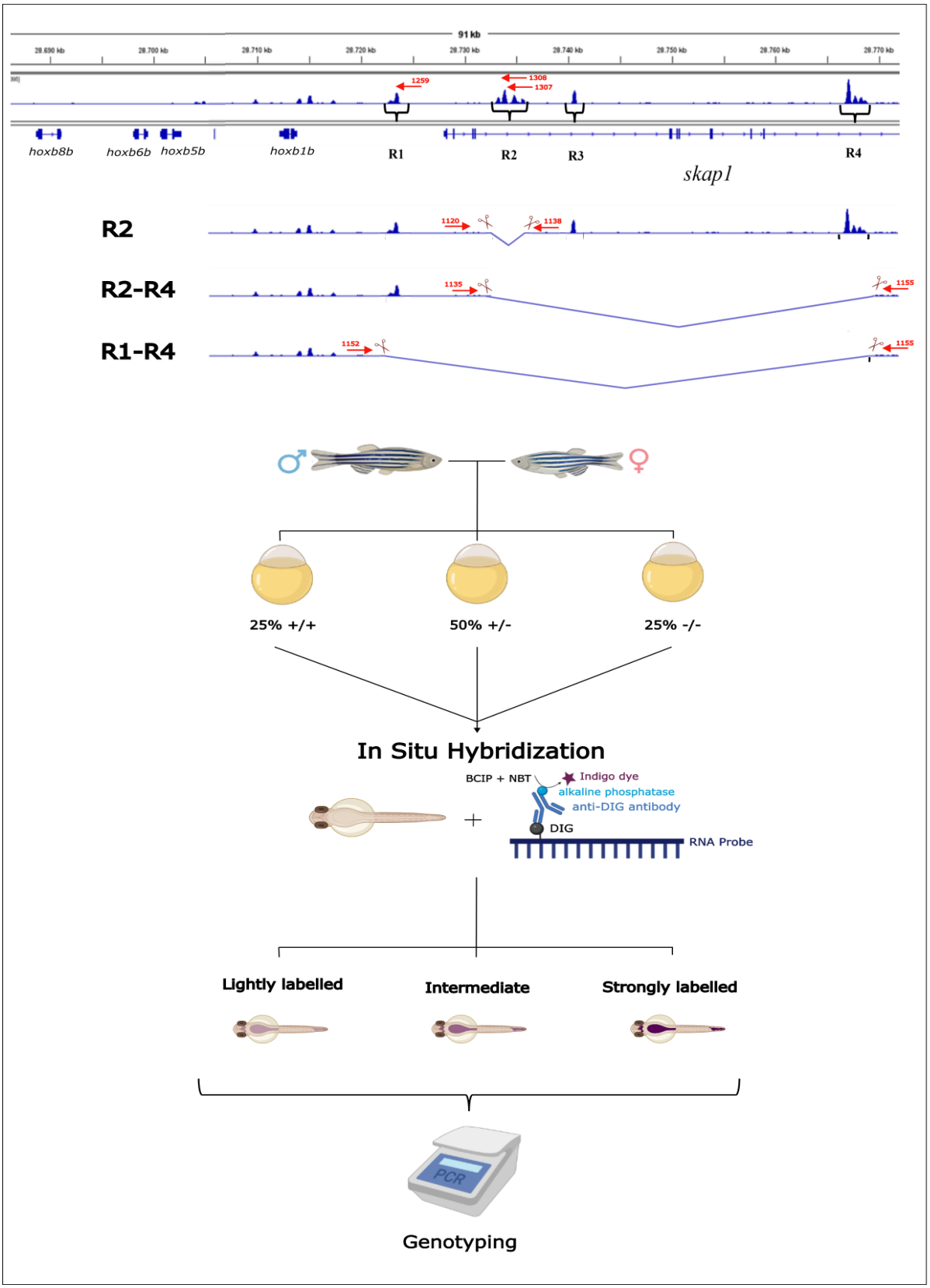


Figure 11 : Principle of the *in situ* hybridization experiment. Heterozygous mutants R2, R1-R4, and R2-R4 fish were incrossed. According to Mendelian laws, their laying contains 25% WT, 50% heterozygotes, and 25% homozygous mutant embryos. In situ hybridization was performed on embryos at different development stages (12-somite, 18-somite, 24 hpf and 48 hpf), using antisense RNA probes targeting the mRNA of *hoxb1b*, *hoxb5b*, *hoxb6b*, or *hoxb8b* in zebrafish. The stained embryos were then sorted based on their staining intensity (light, intermediate, or strong). Subsequently, the embryos were genotyped to determine if a dependence exists between expression pattern and the genotype.

3.1.2. Preparation of clutches of embryos.

We started the study by identification of adult fish harbouring the different deletions at the heterozygous state. These deletion mutants were previously generated in my host laboratory by CRISPR/Cas9 mutagenesis. We set up a genotyping strategy allowing the amplification of the deleted allele and of the wild type allele by PCR using three primers (primers presented as red arrows on figure 11). DNA was prepared from a piece of caudal fin taken from adult fish ("fin clips") and used for the PCR (see Material and Methods). These PCR enable to directly distinguish between wild-type (+/+), heterozygous mutants (+/-), and homozygous mutants (-/-) on gel. An example of the gel obtained with the deletion R2-R4 is shown on Figure 12, using the primers 1135, 1155, and 1308. The amplification of the wild type allele with primers 1135 and 1308 giving a fragment of 321 bp while the amplification of the deletion allele leading to a fragment of 214 bp. This strategy allowed the identification of adult fish heterozygous and homozygous fish for each deletion. These heterozygous fish were then crossed to obtained WT, heterozygous and homozygous embryos. WT fish were also crossed to only obtain WT embryos. Embryos were then fixed at different stages : 48 hours post fertilization (hpf), 24hpf, 18-somite (corresponding to 18 hpf) and 12-somite (15 hpf).

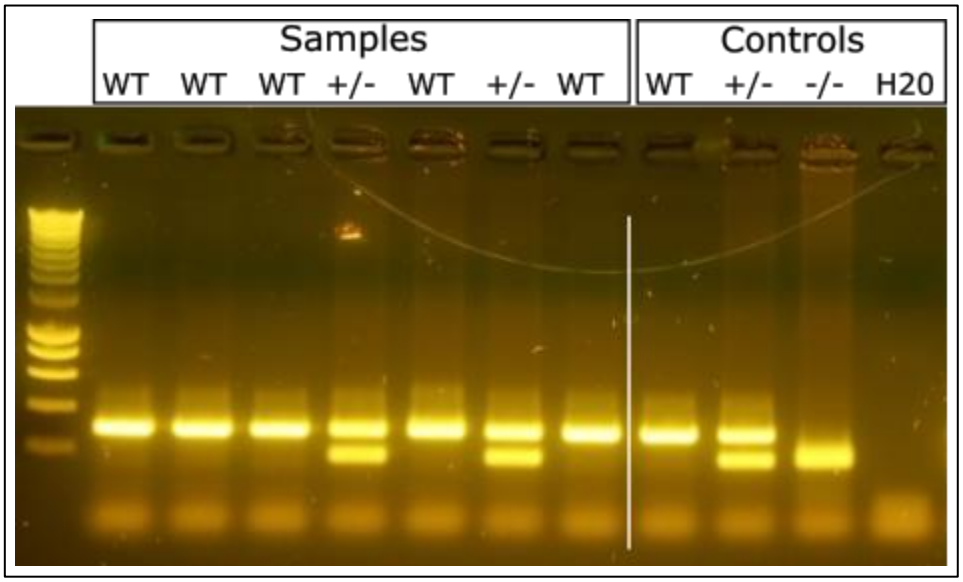


Figure 12 : Example of a gel following the Fin clips experiment. This gel was obtained after a fin clips experiment on R2-R4 mutant samples. The controls were performed using fish with known genotypes. The PCR, utilizing three primers, enables the detection of fish genotypes in a single experiment: wild type (WT), heterozygous (+/-), and homozygous (-/-).

3.1.3. *In situ* hybridization experiments (ISH) using the *hoxbb* probes.

The expression of the four genes presents in the zebrafish *hoxbb* cluster—*hoxb1b*, *hoxb5b*, *hoxb6b*, and *hoxb8b*—were tested by ISH for the 3 deletion mutants and compared to WT embryos, at the 4 developmental stages. Thus, we performed in total 64 different ISH experiments. These ISH were done using specific probes designed prior to my arrival at the laboratory. The intensity of the staining depends on the quantity and the place of probes bound in the embryo, reflecting the amount of target mRNA present. Following the experiment, photographs of the embryos were taken. For the embryos obtained from wild type fish, almost all embryos of the same clutch display the same labelling. In contrast, obvious labelling differences were observed among embryos obtained from R1-R4 and R2-R4 deletion mutants (example on Figure 13). The lightly labelled embryos consistently represented a smaller proportion than the strongly labelled ones, making up approximately one quarter of the total number of embryos.

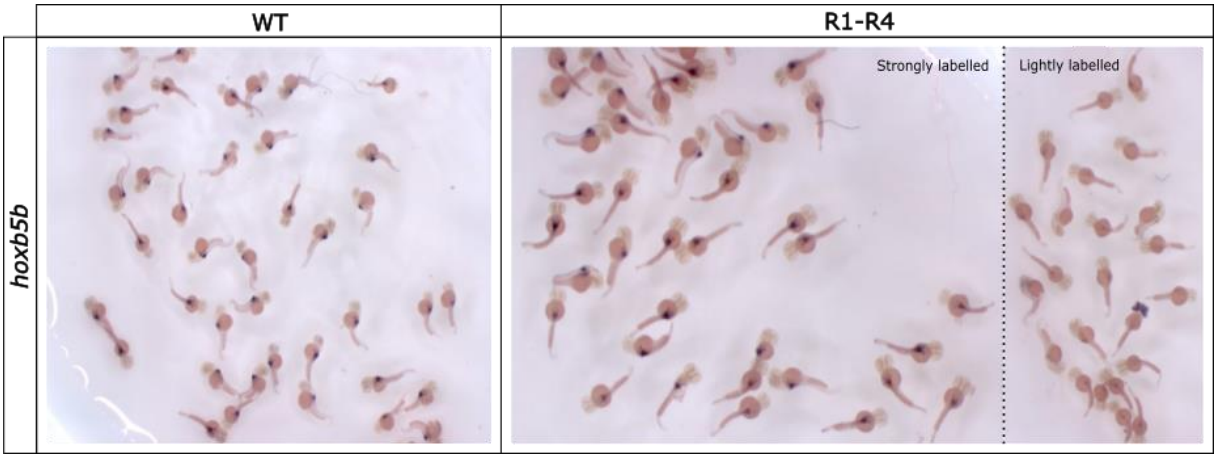


Figure 13 : *In situ* hybridization reveals uniform *hoxb5b* mRNA expression in WT embryos and variable expression in R1-R4 mutants at 48 hpf. R1-R4 embryos all originate from the same clutch and were in the same well during the experiment. They were sorted prior to being photographed

3.1.3.1. Analysis of R1-R4 mutant embryos at the 48 hpf developmental stage

Figure 14 show embryos at 48 hpf stained by the 4 probes and classified as strongly labelled (S) (resembling the WT embryos) and lightly labelled (L). Indeed, as shown in the images, there is a clear difference between embryos of the two categories.

In order to verify that the lower expression of *hoxbb* genes observed in the L category is due to the R1-R4 deletion, we genotyped embryos individually using the same 'triple' PCR with the three primers, as previously described for the fin clips. An example of genotyping results for such embryos is shown in Figure 14. The graphs shown in Figure 15 present a summary of the genotyping results for each *hoxbb* gene. These results showed a clear link between the intensity of the labelling and the genotype of the embryo. There is a perfect link between the lower expression with homozygous R1-R4

deletion mutant for *hoxb5b*. For the *hoxb6b* gene, all homozygous deletion mutant were classified in the "light labelling" category; still a few heterozygous mutants were also classified in this group. For *hoxb1b* and *hoxb8b*, we can notice that the link is far to be perfect : some embryos classified as "lightly labelled" were WT or heterozygous for the deletion, and inversely, a few embryos classified as "strongly labelled" were homozygous for the deletion. Nevertheless, when we perform a Fisher statistical test to determine if the difference of labelling is dependent of the genotype, the results demonstrate that it is the case for the 4 *hoxbb* genes as evidenced by all p-values below 0.05. Thus, at 48hpf the expression of the four *hoxbb* gene is under the control of the enhancer regions R1 to R4. However, we do not lose all expression, suggesting that these genes are also regulated by other enhancers.

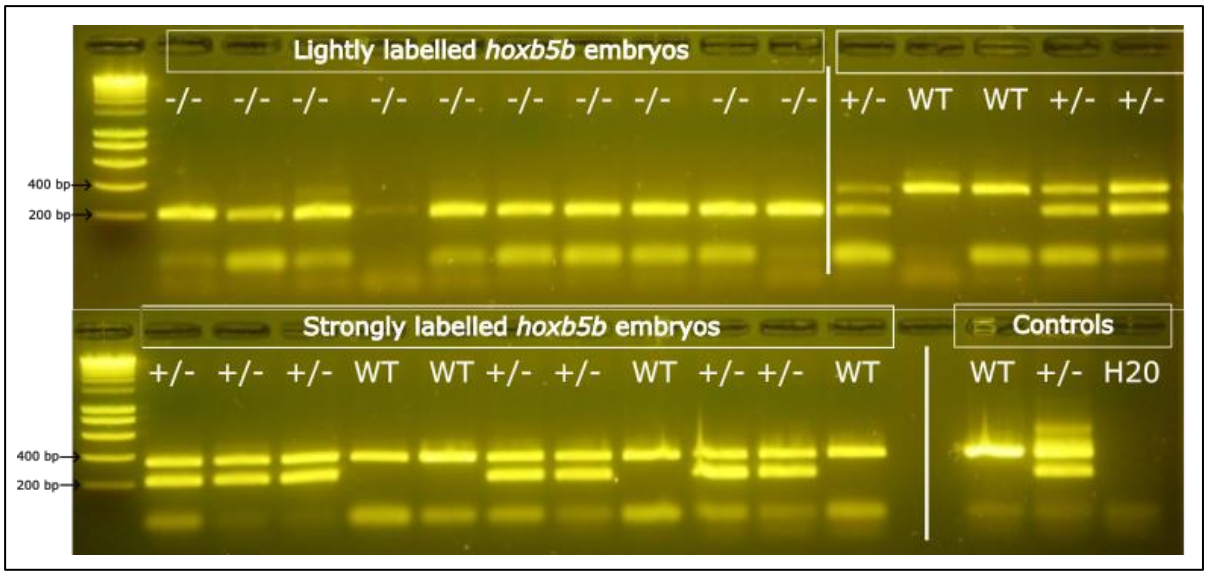


Figure 14 : **Example of a gel for genotyping following *in situ* hybridization.** Gel obtained after an *in situ* hybridization experiment labelling the *hoxb5b* mRNA, on embryos at the 48 hpf stage, resulting from the breeding of R1-R4 heterozygous mutants. Two categories of labelled embryos were genotyped: strongly labelled and lightly labelled embryos. The controls were performed using fish with known genotypes. The PCR, utilizing three primers, enables the detection of embryo genotypes in a single experiment: wild type (WT), heterozygous (+/-), and homozygous (-/-).

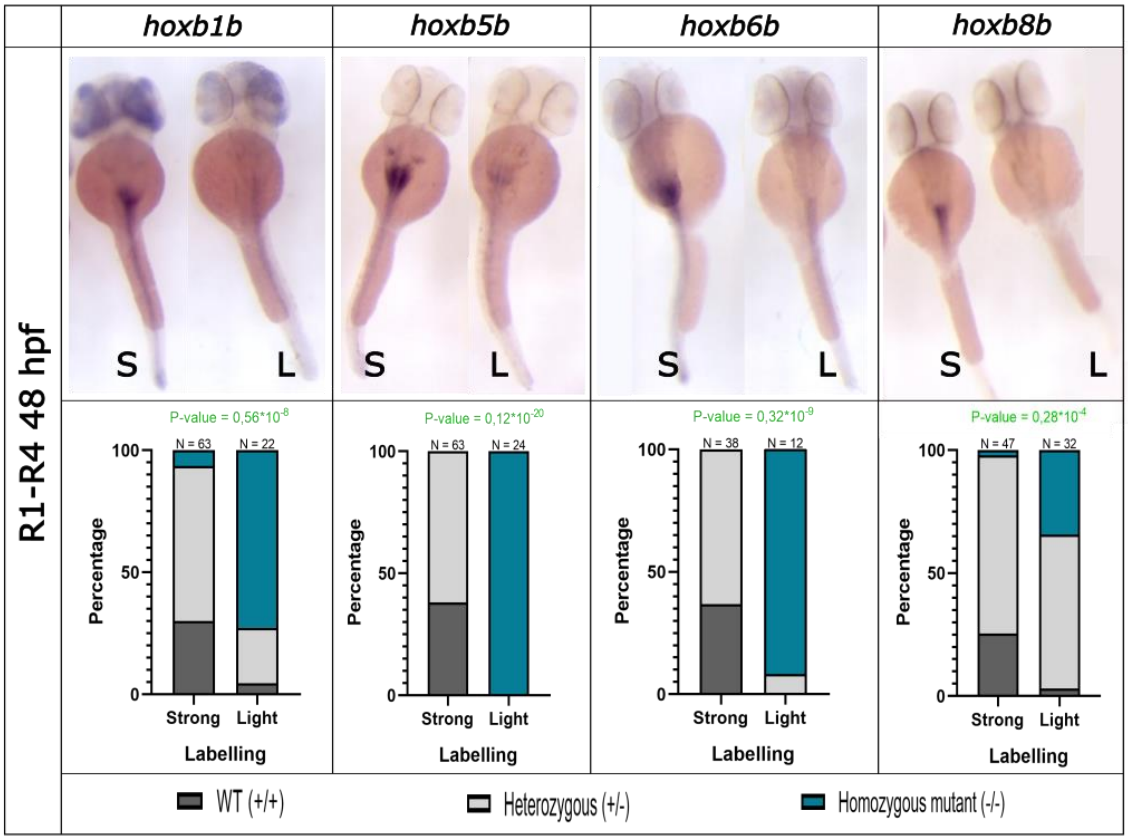


Figure 15 : Results of *in situ* hybridization for 48hpf embryos from R1-R4 heterozygous mutants. The upper panel shows images of embryos labelled with an indigo precipitate, displaying a strongly labelled embryo (S) and a lightly labelled embryo (L) for each gene. The bottom panel presents graphs from the genotyping of categorised embryos. The graphs display the percentage of each genotype in the different categories, with N representing the number of embryos genotyped in each category. The p-value, obtained from a Fisher's exact test, indicates whether there is a significant link between the genotype and the observed labelling. The p-value is shown in green if significant (< 0.05) and in red if not significant (> 0.05).

3.1.3.2. Analysis of R1-R4 mutant embryos at the 24 hpf developmental stage

At 24 hpf, no clear differences were observed for the *hoxb1b* gene in R1-R4 mutants. However, for the other three genes, three categories could be distinguished based on the labelling intensities of the embryos (strongly, lightly, or intermediately labelled; refer to the images in Figure 16.). For the *hoxb5b* gene, the proportion of homozygous mutants was the highest in the "lightly labelled" (L) category and was the smallest in the "strongly labelled" (S) category; however, the p-value obtained from Fisher's exact test is not significant; this may be due to the low the number of genotyped embryos. For the *hoxb6b* and *hoxb8b* genes, the embryos of the L category correspond mostly to homozygous mutants and the p-values obtained are significant, indicating the regulation of these genes by the R1-R4 region at 24 hpf.

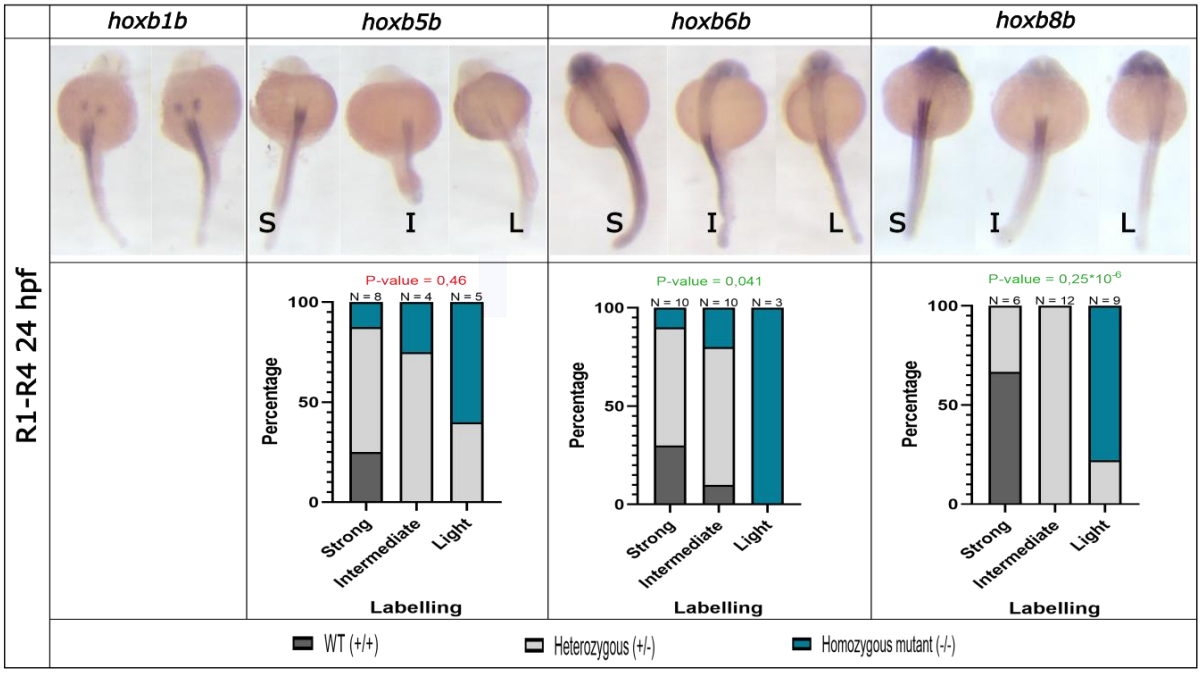


Figure 16 : Results of *in situ* hybridization for 24hpf embryos from R1-R4 heterozygous mutants. The upper panel shows images of embryos labelled with an indigo precipitate, displaying a strongly labelled embryo (S), an intermediate (I) and a lightly labelled embryo (L) for each gene (if differences in the labelling was observed). The bottom panel presents graphs from the genotyping of categorised embryos. The graphs display the percentage of each genotype in the different categories, with N representing the number of embryos genotyped in each category. The p-value, obtained from a Fisher's exact test, indicates whether there is a significant link between the genotype and the observed labelling. The p-value is shown in green if significant (< 0.05) and in red if not significant (> 0.05).

3.1.3.3. Analysis of R1-R4 mutant embryos at 18-somite developmental stage

At the 18-somite stage, two categories could be established for the *hoxb1b*, *hoxb6b*, and *hoxb8b* genes, and three categories for the *hoxb5b* gene (Figure 17). Clear differences were observed for *hoxb5b* and *hoxb6b*, while the differences for the other two genes were less pronounced, making embryo categorization more challenging. Nonetheless, all p-values indicate a significant association between genotype and embryo labelling. The results suggest that the regulation of all four genes by the R1-R4 region is evident at the 18-somite stage.

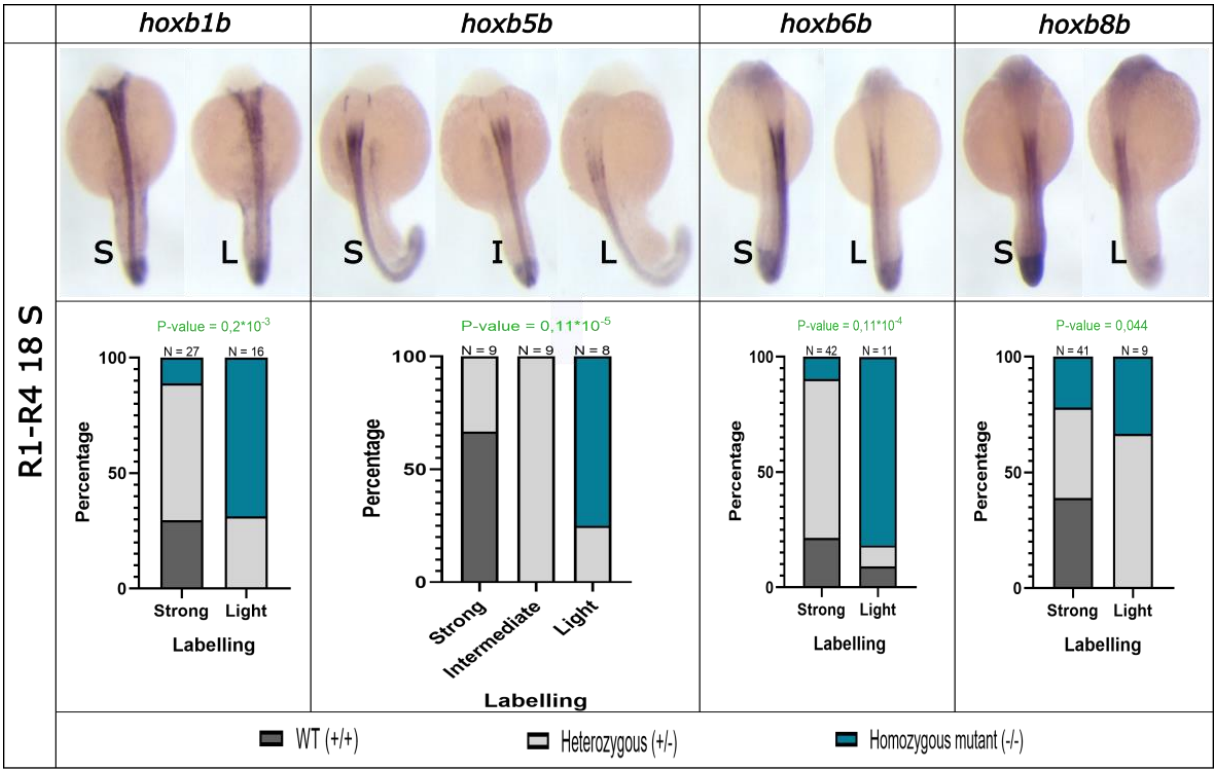


Figure 17 : Results of *in situ* hybridization embryos at the 18-somite stage from R1-R4 heterozygous mutants. The upper panel shows images of embryos labelled with an indigo precipitate, displaying a strongly labelled embryo (S), an intermediate (I) and a lightly labelled embryo (L) for each gene. The bottom panel presents graphs from the genotyping of categorised embryos. The graphs display the percentage of each genotype in the different categories, with N representing the number of embryos genotyped in each category. The p-value, obtained from a Fisher's exact test, indicates whether there is a significant link between the genotype and the observed labelling. The p-value is shown in green if significant (< 0.05) and in red if not significant (> 0.05).

3.1.3.4. Analysis of R1-R4 mutant embryos at 12-somite developmental stage

At the 12-somite stage (Figure 18), no obvious differences were observed for the *hoxb1b* gene. Two categories were established for the *hoxb5b* and *hoxb8b* genes; however, the labelling differences were not as distinct as those observed in later stages. Nonetheless, the p-value obtained was below 0.05, indicating a significant link between labelling differences and genotype even at the 12-somite stage. As depicted in Figure 18, two categories were also created for the *hoxb6b* gene, but this categorization was not based on the labelling intensity but on the expression pattern. Indeed, the *hoxb6b* expression was slightly more posterior in some embryos showing a longer area without staining just anteriorly (noted "Long" and illustrated in the image of the *hoxb6b* with black arrows). For some embryos, the labelling start point is closer to the forebrain compared to other embryos. Genotyping of five embryos from each category revealed that 100% of the embryos in the "long" category were homozygous mutants, demonstrating a clear link with the genotype and the absence of the R1-R4 region. This indicates a posterior shift in the expression profile of *hoxb6b* in R1-R4 mutants.

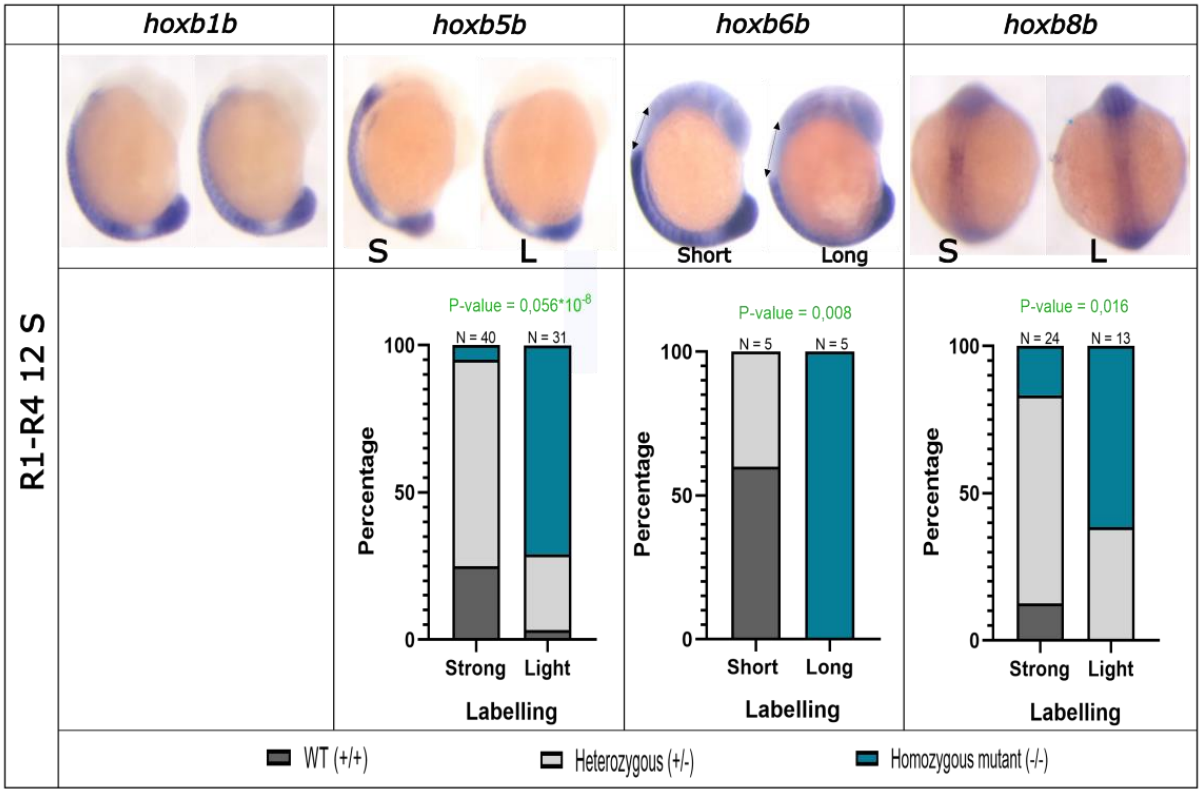


Figure 18 : Results of *in situ* hybridization of embryos at the 12-somite stage from R1-R4 heterozygous mutants. The upper panel shows images of embryos labelled with an indigo precipitate, displaying a strongly labelled embryo (S), and a lightly labelled embryo (L) for *hoxb5b* gene. For *hoxb6b* gene, the embryos were sorted between short and long labelling. The bottom panel presents graphs from the genotyping of categorised embryos. The graphs display the percentage of each genotype in the different categories, with N representing the number of embryos genotyped in each category. The p-value, obtained from a Fisher's exact test, indicates whether there is a significant link between the genotype and the observed labelling. The p-value is shown in green if significant (< 0.05) and in red if not significant (> 0.05).

The results obtained for the R1-R4 mutant strongly suggest that the four *hoxbb* genes within the cluster are regulated by the R1-R4 region. The differences in labelling were more pronounced at later stages (48 hpf), indicating that this enhancer region is active, or becomes more significant later in the development. It may play a greater role in maintaining their expression rather than in their initial activation.

3.1.3.5. Analysis of R2-R4 mutant embryos

The same experiment was conducted on embryos derived from the breeding of heterozygous R2-R4 mutants. The same developmental stages were analysed, and similar results were obtained, with a few exceptions. At 48 hpf (Figure 19), the results were consistent, and all p-values calculated were below 0.05, indicating a significant link between the observed labelling and the genotype of the embryos. This further indicates a regulative action of the R2-R4 region on all *hoxbb* genes in zebrafish at 48 hpf. It also suggests that the R1 region doesn't play significant role in the regulation of *hoxbb* genes.

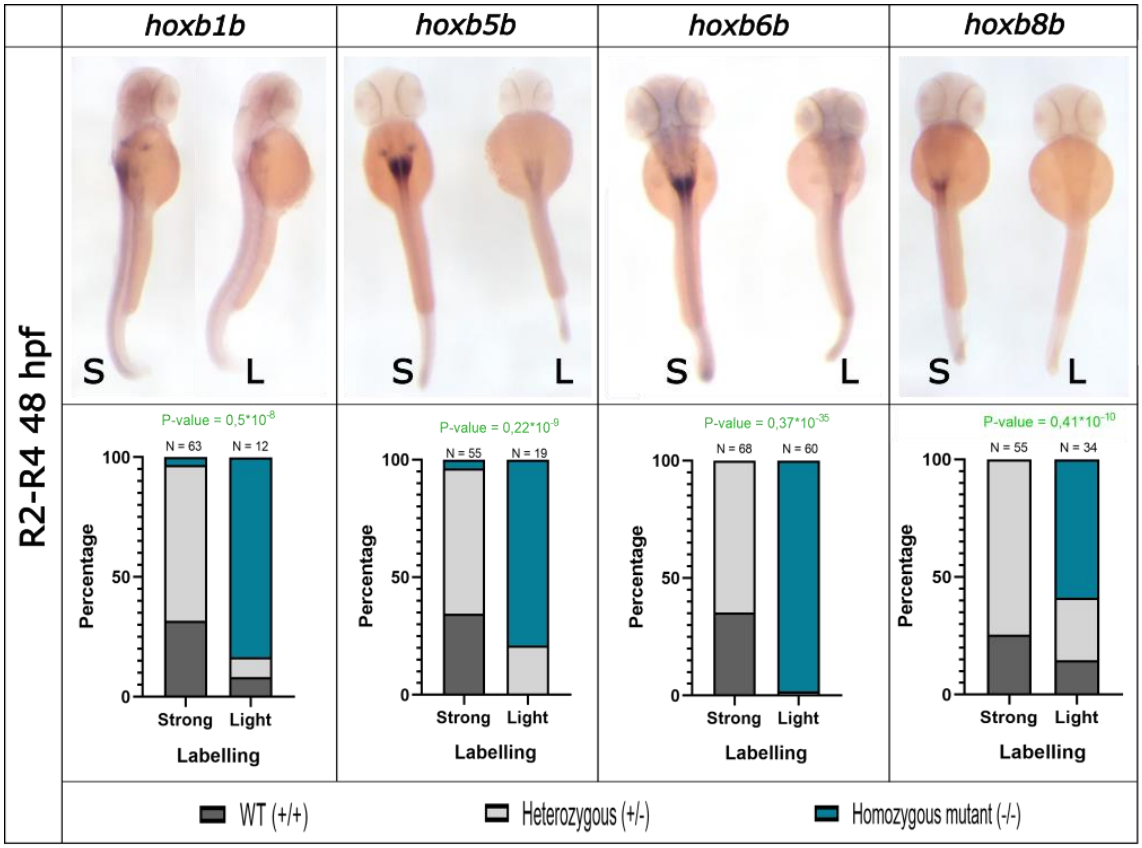


Figure 19 : Results of *in situ* hybridization for 48hpf embryos from R2-R4 heterozygous mutants. The upper panel shows images of embryos labelled with an indigo precipitate, displaying a strongly labelled embryo (S) and a lightly labelled embryo (L) for each gene. The bottom panel presents graphs from the genotyping of categorised embryos. The graphs display the percentage of each genotype in the different categories, with N representing the number of embryos genotyped in each category. The p-value, obtained from a Fisher's exact test, indicates whether there is a significant link between the genotype and the observed labelling. The p-value is shown in green if significant (< 0.05) and in red if not significant (> 0.05).

At 24 hpf (Figure 20), as with the R1-R4 mutant, no clear differences were observed for the *hoxb1b* gene. However, in contrast, the p-value calculated for the *hoxb5b* gene was significant, while those for the *hoxb6b* and *hoxb8b* genes were not. A trend consistent with an effect on *hoxb6b* gene is observed. Our expectations is noticeable and, if more embryos would have been genotyped, it is possible that the statistical test would have become significant.

At the 18-somite (Figure 21), the correlation of the labelling with the genotype was very good and highly significant for *hoxb5b*. However, for *hoxb6b*, the sorting of embryos was not statistically linked to the genotype. Similar observations were made at the 12-somite stage (Figure 22), as the sorting of *hoxb5b* labelling were statistically linked to the genotype. No difference was detected for *hoxb1b*, as with the R1-R4 mutant, nor for the *hoxb8b* gene. For *hoxb6b*, the same positional shift was observed as in R1-R4 mutants. Although the trend appears to be present, as no WT embryos were found in the "long" category, only heterozygous mutant embryos were found among the five genotyped in the "short" category. Here again, the small number of embryos genotyped could be a contributing factor.

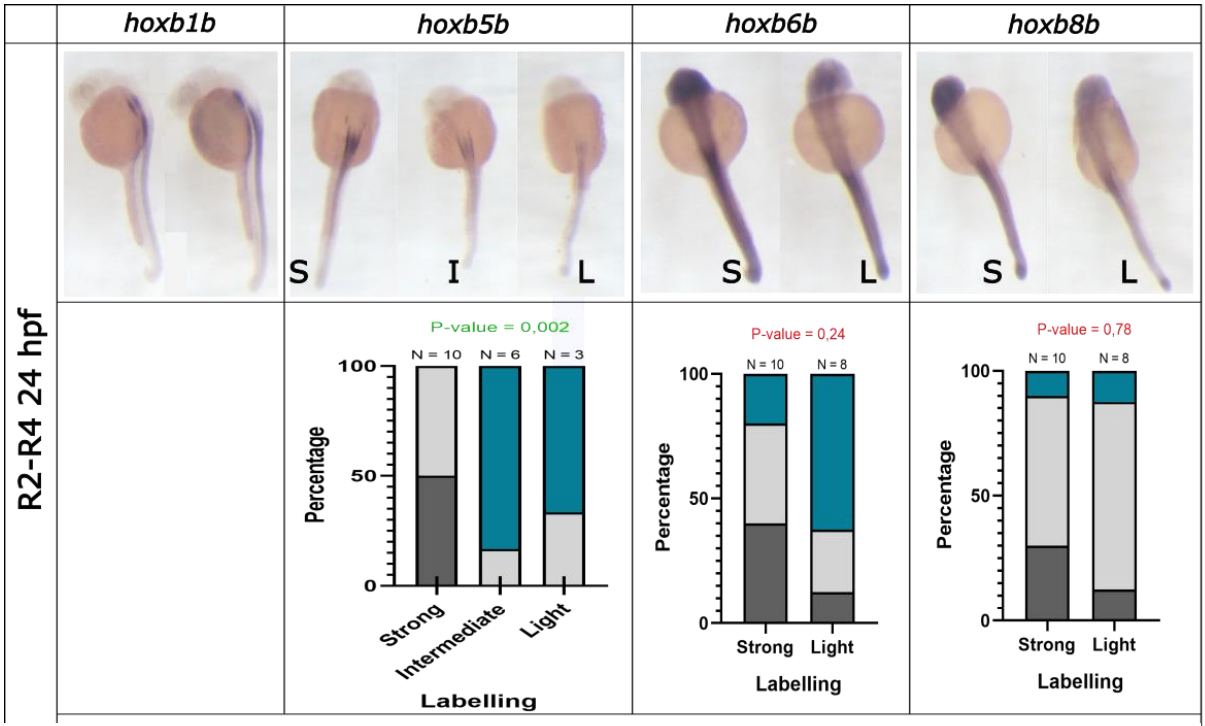


Figure 20 : Results of *in situ* hybridization for 24hpf embryos from R2-R4 heterozygous mutants. The upper panel shows images of embryos labelled with an indigo precipitate, displaying a strongly labelled embryo (S), an intermediate (I) and a lightly labelled embryo (L) for each gene (if differences in the labelling was observed). The bottom panel presents graphs from the genotyping of categorised embryos. The graphs display the percentage of each genotype in the different categories, with N representing the number of embryos genotyped in each category. The p-value, obtained from a Fisher's exact test, indicates whether there is a significant link between the genotype and the observed labelling. The p-value is shown in green if significant (< 0.05) and in red if not significant (> 0.05).

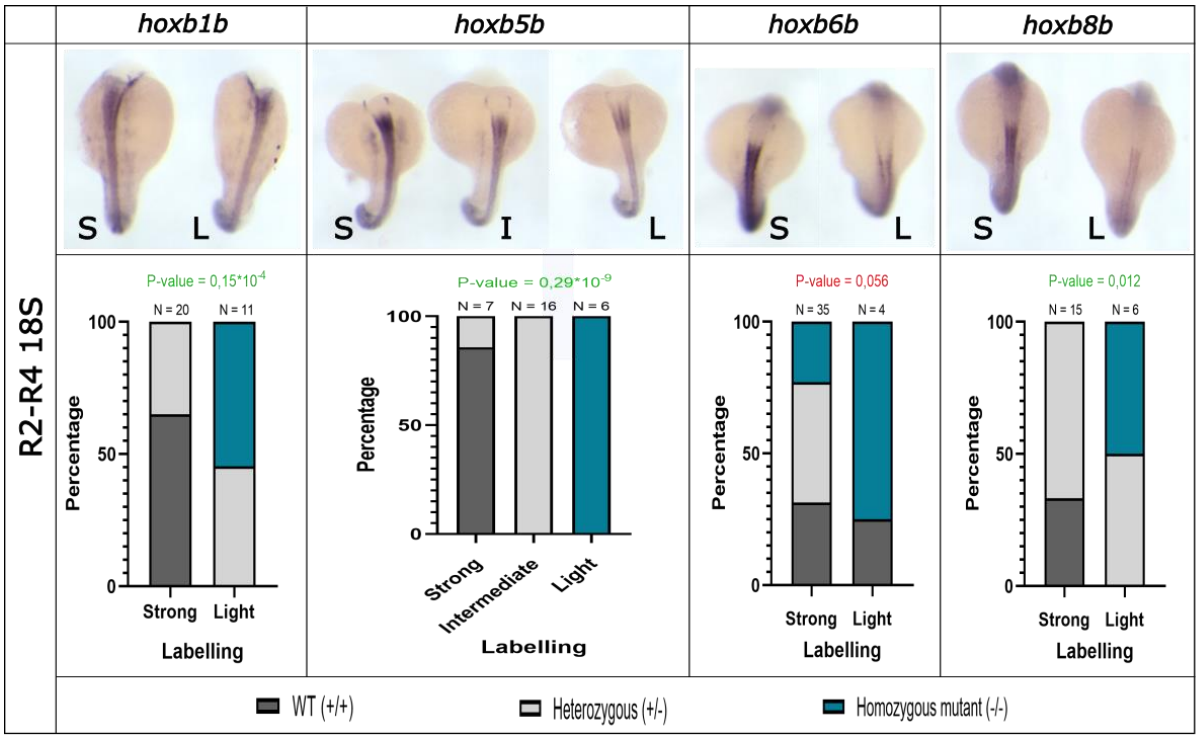


Figure 21 : Results of *in situ* hybridization embryos at the 18-somite stage from R2-R4 heterozygous mutants. The upper panel shows images of embryos labelled with an indigo precipitate, displaying a strongly labelled embryo (S), an intermediate (I) and a lightly labelled embryo (L) for each gene. The bottom panel presents graphs from the genotyping of categorised embryos. The graphs display the percentage of each genotype in the different categories, with N representing the number of embryos genotyped in each category. The p-value, obtained from a Fisher's exact test, indicates whether there is a significant link between the genotype and the observed labelling. The p-value is shown in green if significant (< 0.05) and in red if not significant (> 0.05).

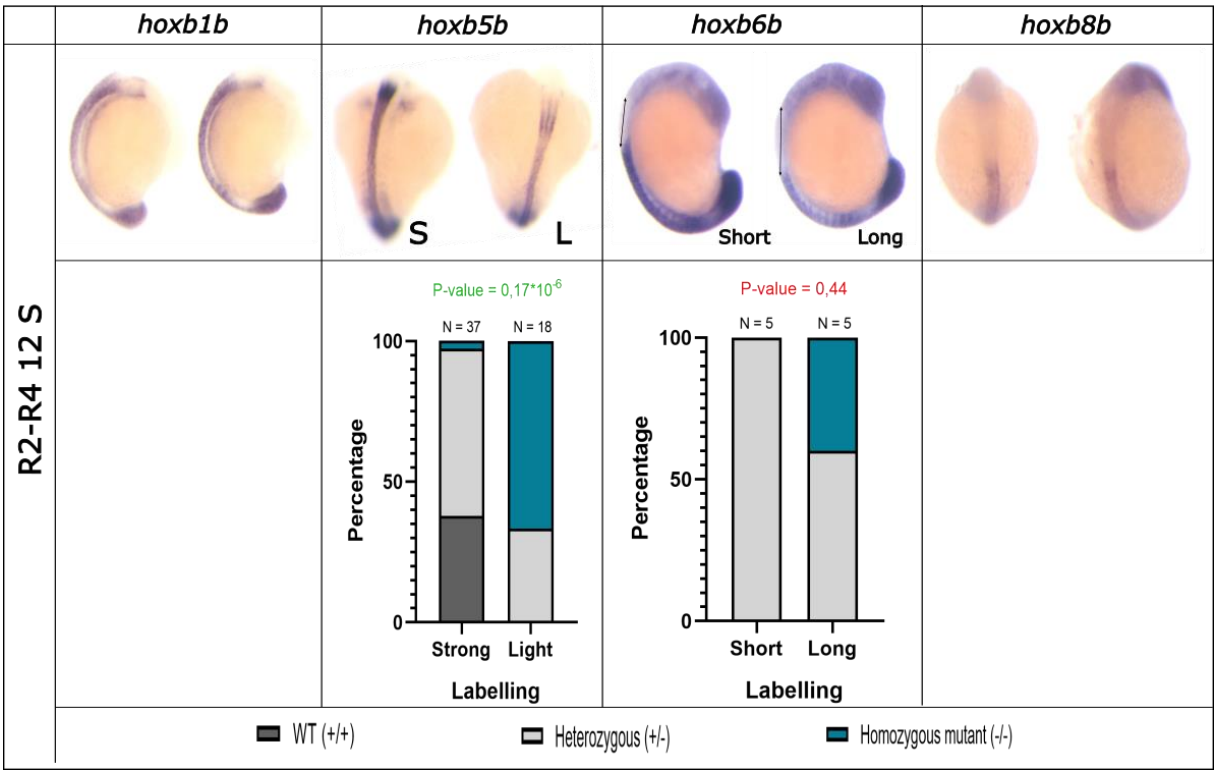


Figure 22 : Results of *in situ* hybridization of embryos at the 12-somite stage from R1-R4 heterozygous mutants. The upper panel shows images of embryos labelled with an indigo precipitate, displaying a strongly labelled embryo (S), and a lightly labelled embryo (L) for *hoxb5b* gene. For *hoxb6b* gene, the embryos were sorted between short and long labelling. The bottom panel presents graphs from the genotyping of categorised embryos. The graphs display the percentage of each genotype in the different categories, with N representing the number of embryos genotyped in each category. The p-value, obtained from a Fisher's exact test, indicates whether there is a significant link between the genotype and the observed labelling. The p-value is shown in green if significant (< 0.05) and in red if not significant (> 0.05).

The results obtained are largely consistent with those observed in the R1-R4 mutant. This consistency across these 2 different mutants implies that the regulatory influence of the R1 RARE region on the *hoxbb* genes may be limited or non-essential and indicate that the R2 to R4 regions are more important for the control of the *hoxbb* cluster.

3.1.3.6. Analysis of R2 mutant embryos

The same experiment was conducted at the four developmental stages for the R2 mutants. We anticipated a decrease in mRNA concentration in the embryos, similar to what was observed in the other two mutants, given that the R2 region is the most conserved. However, no labelling differences were detected following the ISH experiments (Figure 23), indicating that the R2 is not required for *hoxbb* expression. Several possibilities could explain this surprising result. The first is that the R2 region does not play an enhancer role in the regulation of the *hoxbb* cluster. A more plausible explanation is that when a single region is removed from the genome, the remaining three regions are sufficient to maintain the regulatory function. Additionally, as the enhancer activity of R4 is similar to R2 enhancer activity (Figure 8), it is possible that R4 compensates for the loss of R2. Further experiments will be necessary to confirm this hypothesis.

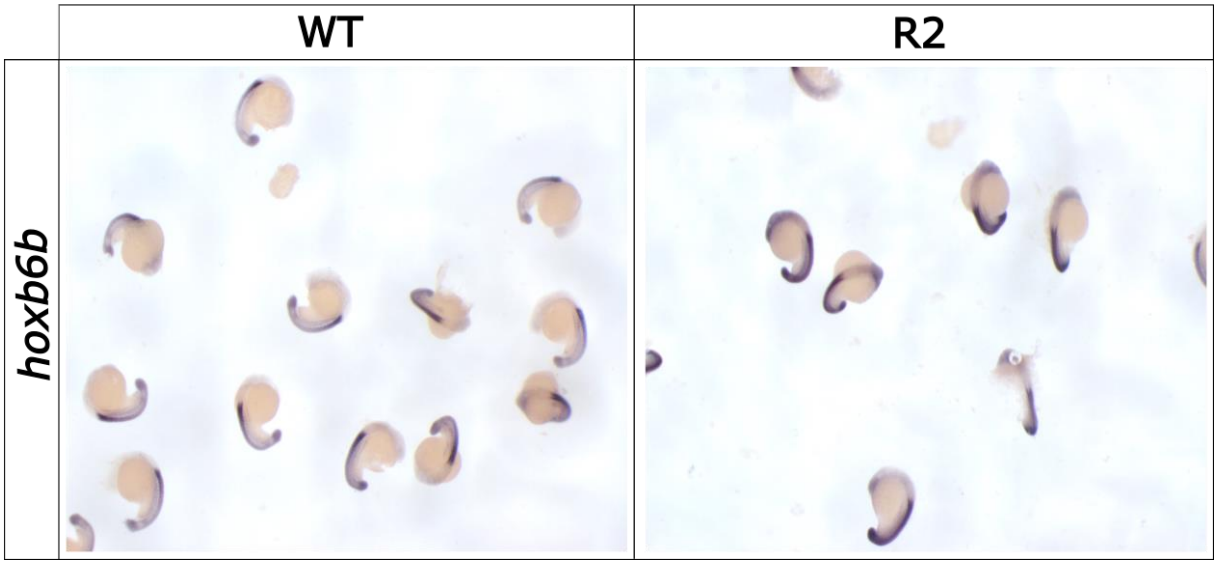


Figure 23 : Photographs of WT and R2 mutant embryos at the 18-somite stage after *in situ* hybridization for the hoxb6b gene. The R2 mutant embryos exhibit labelling patterns similar to those observed in WT embryos, and no variation was observed among the mutant embryos themselves.

3.2. R3 enhancer region is independent of retinoic acid regulation

Previous results from the host laboratory had shown that both the R2 and R4 regions were able to confer retinoic acid response when fused to the cfos promoter driving GFP expression (as presented in introduction, Figure 8). To determine whether the R3 region functions similarly, transgenic embryos harbouring a R3-cfos:GFP cassette were tested to determine whether they also confer RA responsiveness. Embryos from an outcross of R3 transgenic fish with wild-type fish were grown until the 50% epiboly stage (5.5 hpf). The embryos were then divided into three wells. The first well served as control with E3 medium and DMSO only. The second well contained E3 medium with RA (0.5 µM), and the third well contained E3 medium with DEAB, which is an endogenous RA synthesis inhibitor (4 µM). At 24hpf, photos of the embryos were taken with the fluorescent stereomicroscope. As shown on figure 24, the R3 transgenic embryos are not regulated by retinoic acid as the GFP fluorescence doesn't change in the different conditions, particularly when RA is added. In conclusion, the R3 region alone is not able to confer RA response on a cfos promoter.

Figure 24 : Pictures of R3:GFP transgenic embryos at 24 hpf treated with DMSO (control), DEAB and RA. The left picture shows embryo morphology and the right photograph shows GFP fluorescent expression.

3.3. R2 Enhancer Region Exhibits Activity in Later Developmental Stages

Our previous results indicated that the deletion of R2-R4 impacts *hox*bb expression mainly at late developmental stage (48hpf). Thus, we decided to determine the timing during which R2 is active in the context of the R2-cfos:GFP transgene. To that end, we have treated R2-cfos:GFP transgenic embryos at different developmental stage with DEAB to determine when the R2 activity is affected by the RA inhibitor. Embryos from outcrosses of R2 transgenic fish were grown until 50% epiboly. They were then divided into different wells of a multi-well plate, with each well containing embryos under different conditions (as shown in Figure 25). The control wells contained E3 medium with DMSO. Photographs of the embryos were taken using a fluorescence microscope at 24 and 48 hpf.

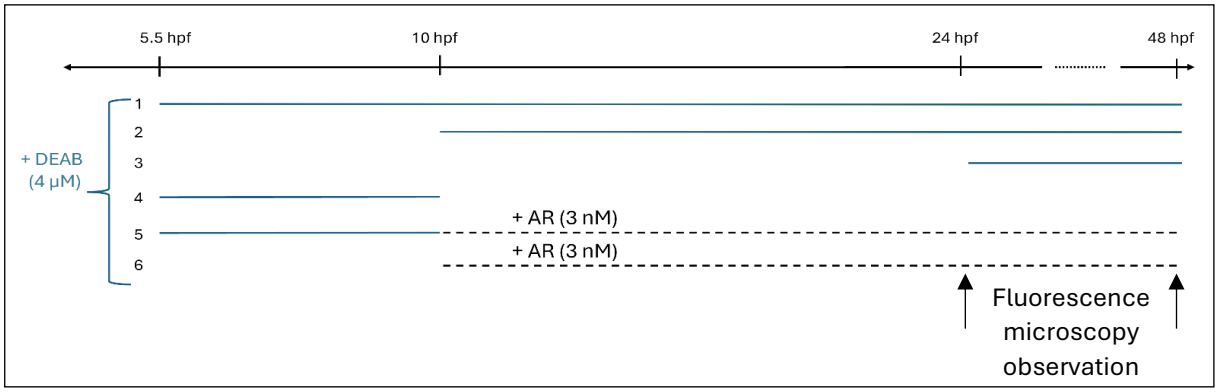


Figure 25 : Experimental Conditions for Transgenic Growth. Embryos were cultured under six different conditions with DEAB (1–6). The coloured lines indicate the time periods during which the drug was present in the culture medium. Absence of a line signifies that the embryos were grown in standard E3 medium with PTU to inhibit pigmentation. Control embryos were maintained in E3 medium with DMSO.

The results obtained at 24 hpf (Figure 26) show a complete extinction of GFP fluorescence when DEAB is added from 5.5 hpf to 24 hpf (condition 1) and a nearly complete extinction when added from 10 hpf to 24 hpf (condition 2). When DEAB is applied only between 5.5 and 10 hpf (condition 4), GFP fluorescence is still observed even if exogenous RA is applied or not from 10 hpf. All together, these data indicate that the R2 region is active mostly after 10 hpf, even if the synthesis of RA begin at early gastrulation (5.5 hpf).

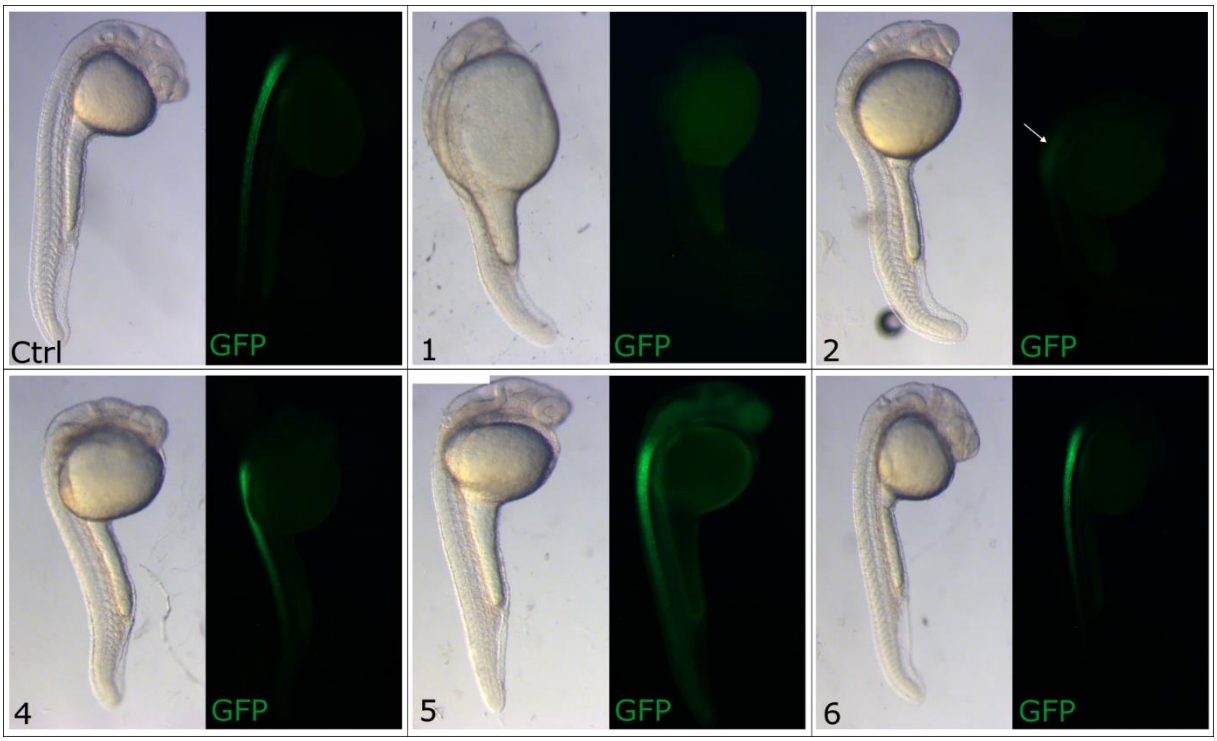


Figure 26 : Images of R2:GFP transgenic embryos at 24 hpf treated with DMSO (control), DEAB, and RA. The left image displays embryo morphology, while the right image shows GFP fluorescence. The number in the bottom left corner indicates the growth condition. White arrows highlight GFP fluorescence when it is faint.

At 48 hpf (Figure 27), GFP fluorescence is entirely absent in conditions 1 and 2, and almost entirely absent in condition 3, where DEAB was introduced into the medium at 24 hpf. The fluorescence seen in the other conditions indicate that the R2 region remains active between 24 and 48 hpf. However, the extent of the enhancer region's activity is diminished, as evidenced by the images, compared to the fluorescence observed at 24 hpf.

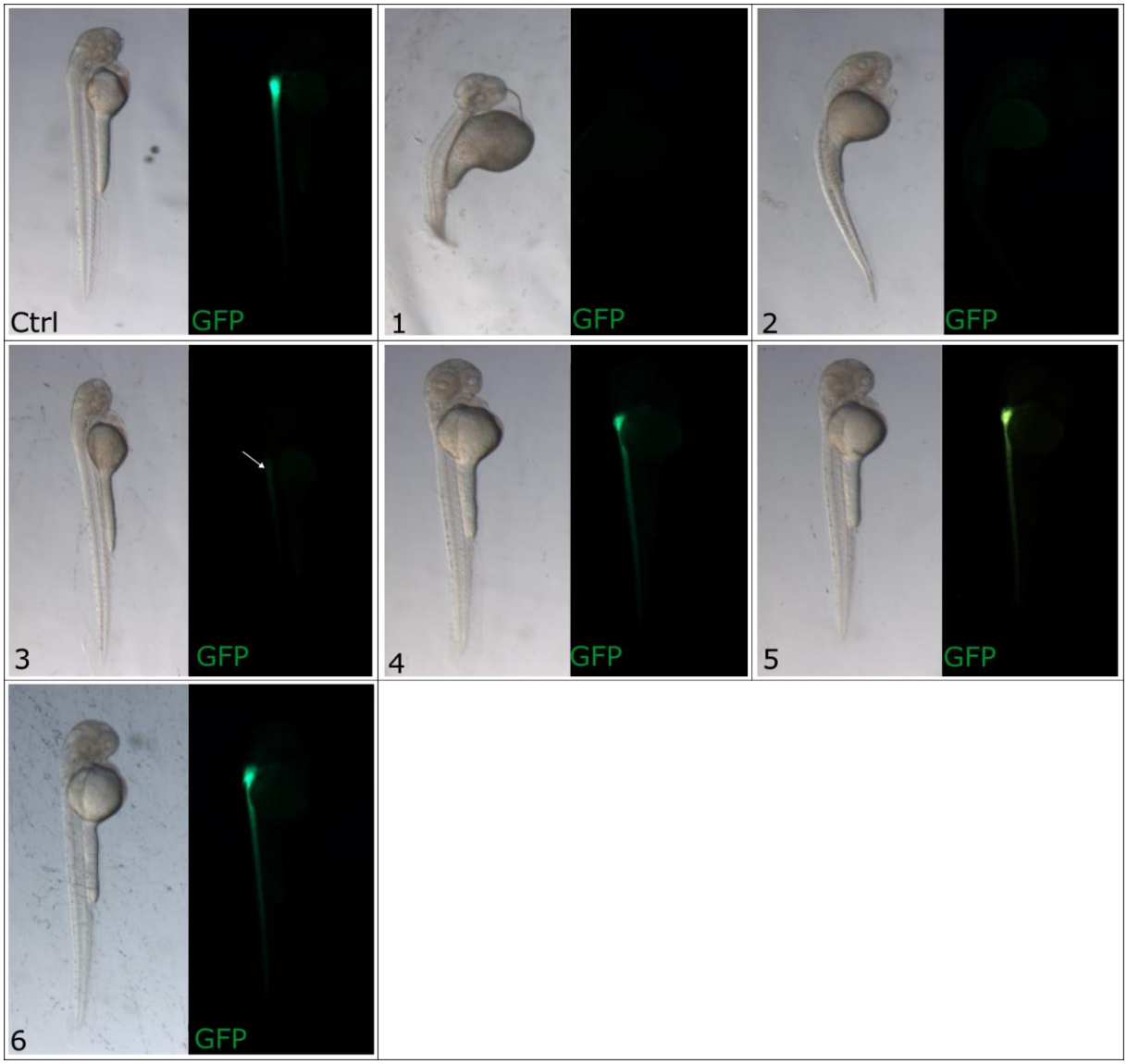


Figure 27 : Images of R2:GFP transgenic embryos at 48 hpf treated with DMSO (control), DEAB, and RA. The left image displays embryo morphology, while the right image shows GFP fluorescence. The number in the bottom left corner indicates the growth condition. White arrows highlight GFP fluorescence when it is faint.

These results indicate that the R2 enhancer is still active well after gastrulation (10 hpf). These observation on transgenic lines are thus consistent with findings obtained on deletion mutants, which revealed that the enhancer region R1-R4 have a stronger impact at relatively late developmental stages. This suggests that these enhancer regions may be more involved in maintaining the expression of *hoxbb* genes rather than in their initial activation during gastrulation.

4. Discussion

During this master's thesis, the objective was to investigate the function of four RAREs regions, R1 to R4, identified by a ChIP-seq analysis conducted previously by Ana Lopez in my host laboratory (López-Pérez et al., 2021). These regions are located far downstream from the *hoxbb* cluster, which includes the genes *hoxb1b*, *hoxb5b*, *hoxb6b*, and *hoxb8b*. We initially examined the expression patterns of these four genes through in situ hybridisation (ISH) in three different deletion mutants: R1-R4, R2-R4, and R2. The results from these *in situ* experiments indicate that both the R1-R4 region and the R2-R4 region regulate the expression of all *hoxbb* genes within the *hoxbb* cluster. In contrast, deletion of the R2 region alone—despite its high conservation among the four regions—did not significantly affect the expression of any of the four gene mRNAs. Additionally, we observed that the differences in labelling were more pronounced at a later stage, such as 48 hpf, compared to the 12-somite stage. However, at the 12-somite stage, a posterior shift in the expression profile of *hoxb6b* has been noticed in R1-R4 and R2-R4 mutants. To highlight this posterior shift more clearly, it will be interesting to perform other ISH but using 2 different probes at the same time : *hoxb6b* and another probe such as *myoD* (somite marker) or *krox20* (marker of rhombomeres 3 and 4) serving as landmark of the antero-posterior position. Such posterior shift of *hoxb6b* could be due to the gradual chromatin opening compared to wild-type embryos. In that case, the posterior shift should be also observed for the neighbouring *hoxb5b* and *hoxb8b*. This should be investigated in the future. Also, the possible delay in *hoxbb* chromatin opening imply that that the activation of these genes should be delayed. This hypothesis should be verified by analysing *hoxbb* expression at earlier stages.

The fact that the R2-R4 has a significant effect on *hoxbb* expression while R2 deletion had no obvious effect lead to the conclusion that either R3-R4 is the most important element or that a compensation occurs between R4 and R2. To address these 2 possibilities, other deletion mutants must be generated like deletion of R4 alone or together with R2. Embryos have been injected with two specific gRNAs and CRISPR/Cas9 to excise the R4 region (experiment not shown in this thesis). The same analysis will be conducted on these new mutant embryos. The primary objective of this experiment is to determine whether the results for R4 mutants will be the same as those of R2 mutants or if the removal of this RARE region will lead to a decrease in *hoxbb* gene mRNA levels. This experiment aims to establish whether a single region can independently influence the *hoxbb* cluster expression. If similar results are observed, the creation of a “R2 and R4” double mutant will be necessary to determine whether a combined action of the R2 and R4 regions affects *hoxbb* gene expression.

While we show in our study that the *hoxbb* expression were reduced upon R1-R4 and R2-R4 deletion, we can see that the correlation is not perfect in many cases. Indeed, some embryos sorted as "strongly labelled" were finally identified as harbouring the deletion at homozygous state. Several explanations can be proposed to explain this lack of perfect correlation. Firstly, it is possible that there is an actual heterogeneity among homozygous mutants, some presenting lower *hoxbb* expression while other have normal expression due to other regulatory regions around the cluster. This could explain why some homozygous deletion mutants survive until adulthood; such mutants may correspond to those expressing *hoxbb* genes correctly. Another explanation is that ISH experiment is not a quantitative experiment and the sorting of "strongly labelled" and "lightly labelled" is subjective. An accurate sorting is strongly dependant of the quality of the embryo labelling and some ISH experiments may give some background staining. In such case, an accurate sorting is not easy and will contain errors. Thus, it would be interesting to perform a more quantitative method to measure more precisely *hoxbb* expression. A reverse transcriptase quantitative PCR (RTqPCR) or a RNA-seq experiment could be good solution. However, such experiment would not be so easy to perform because these methods must be applied on individual embryo and furthermore a genotyping must also be performed on these embryos.

During her master's thesis, Caroline Désiront investigated the retinoic acid dependence of the two most conserved regions, R2 and R4 (Désiront, 2022). In this study, I demonstrated that the R3 region does not respond to retinoic acid, as evidenced by the lack of increase in GFP fluorescence upon RA addition to the medium. This does not mean that the R3 region has no role; it is possible that in the context of the *hoxbb* cluster this R3 element acts with another regulatory sequence that is missing in the R3-cfos:GFP transgene. Similar treatments need to be conducted on R1 transgenic embryos to assess its responsiveness to retinoic acid.

The final phase of this project involved investigating the time period during which the R2 region remains active. R2 transgenic embryos were cultured under various drug-medium conditions. The results indicate that the R2 region is active later in development, showing significant activity during somitogenesis and continuing to be active between 24 and 48 hpf. These findings are consistent with those observed in the deletion mutants. Similar experiments are still required to assess the activity of the R4 region in transgenic embryos.

It would be valuable to precisely determine the activation timing of the R2 region by conducting a kinetic analysis on these transgenic embryos. This could involve capturing images at various developmental stages using in situ hybridisation to detect the presence of GFP mRNA. Given that GFP fluorescence typically appears approximately three hours post-transcription, this approach would

provide insights into the precise timing of the enhancer region's activation. Such an experiment could offer a clearer understanding of the temporal dynamics of R2 region activation.

A ChIP-seq experiment was conducted with the intention of performing a HiChIP-seq experiment. Unfortunately, it did not work as intended. However, a recent article by Moreno-Oñate et al. (2024) on chromatin architecture through the induction of the AR signalling pathway has been published. This paper presents the results of a HiChIP experiment conducted on zebrafish, focusing on the same transcription factor, RAR. Therefore, it is no longer necessary to carry out these experiments ourselves, but we can analyse their results bioinformatically, particularly at the *hoxbb-skap1* locus, to assess the interactions between RARE elements R1, R2, R3, and R4 and the promoters of the hoxbb genes.

5. Material and Method

5.1. Zebrafish and embryos breeding

All the experiments were realised following European union and Belgian guidelines of the Animals welfare as well as the approval of the Uliege Ethics Committee (approval number : 2425).

To inhibit fungal growth, embryos were cultured in E3 medium (water, 5mM de NaCl ; 0.17 mM de KCl ; 0.33 mM de MgSO4 ; 0.33 CaCl2, supplemented with 0.01% methylene blue). Wild-type AB strain zebrafish were used as well as mutant and transgenic lines. All the mutant and transgenic lines used in this project were created in the laboratory by Désiront Caroline and Peers Bernard (Désiront, 2022). For embryos used for analyses, older than 24 hpf, phenylthiourea (PTU) was added to E3 medium at 24hpf to avoid pigmentation.

5.2. In situ hybridisation

Embryos at the appropriate developmental stage were fixed in 4% paraformaldehyde (PFA) overnight (O/N) at 4°C, followed by washing in 0.1% PBST and dechorionation. Subsequently, they underwent dehydration through sequential immersion in 25%, 50%, 75%, and 100% methanol and were stored at -20°C.

For *in situ* hybridisation, embryos were rehydrated in reverse order of methanol concentrations and subjected to Proteinase K treatment (concentration and duration depending on stage), PBST washing, 20-minute post-fixation in 4% PFA, and further PBST washing. Embryos were then incubated in HybMix+ for 3-5 hours at 65°C, followed by overnight incubation at 65°C in HybMix+ containing 5% Dextran Sulphate and 10-50 ng of probes (*hoxb1b*, *hoxb5b*, *hoxb6b*, *hoxb8b*). The antisense RNA probes contained digoxigenin (DIG)-labelled UTP. Subsequent steps included sequential immersion in 2X SSCT 25%, 50% ,75 and 100% (diluted in HybMix-), and then in PBST 25%, 50% ,75% and 100% (diluted in SSCT 0,125%). Embryos were then immersed in blocking buffer and then incubated overnight at 4°C with Blocking buffer containing anti-DIG antibodies bound to an alkaline phosphatase. This antibody recognises the probes containing the DIG-labelled UTP. They were then washed in PBST and immersed in freshly prepared staining buffer, and then in staining buffer with NBT and BCIP. Upon adding 5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt (BCIP) and nitro blue tetrazolium chloride (NBT), an indigo precipitate is formed through the reaction with alkaline phosphatase. The staining process was monitored under a microscope, and once satisfactory coloration appeared, the reaction was stopped by washing with PBST. Pictures of the embryos were taken using a binocular (Olympus, SZX10).

If staining differences were observed among the embryos, they were categorized based on their staining patterns and subsequently genotyped.

5.3. Fin clips and genotyping

Mutant fishes as well as embryos sorted after *in situ* hybridisation were genotyped. The piece of the caudal fin cutted from adult fish or single embryo were immerge in 50 µL of NaOH (50 mM), and the tubes were incubated at 95°C for 25 minutes, followed by a 10-minutes cooling period on ice. Finally, 5 µL of Tris-HCl (1M, pH 8) was added.

For the “triple” PCR amplification, 1 µL of the resultant mixture was combined with 0,3 µL of each of the 3 primers (IDT, USA), 5 µL of TERRA Buffer and 0,2 µM of TERRA polymerase (TaKaRa, Japan) and water to reach a final volume of 10 µL. The products obtained were run on a 2% agarose gel with SYBR Safe (Thermo Fisher Scientific, USA). Smart Ladder 200 – 10000 pb (Eurogentec, Belgium) was used as molecular weight ladder.

Table 1 : Primers sequences

Mutants	Primers sequence	Primer Name
R2	Fwd : 5’ - GTCATGCATACATCTTCTTGTTC - 3’	1120
	Rev1 : 5’ - AACATGCACACTTTGGCTGC - 3’	1138
	Rev2 : 5’ - CGTGAAGCGCCATTATTGAA - 3’	1307
R1-R4	Fwd : 5’ - CCCAGGTCTCTGAAGGGGTGT - 3’	1152
	Rev1 : 5’ - ATGAATCCATTTGGAGCAGTCT - 3’	1155
	Rev2 : 5’ - ATGCACAGCTGAAAGGGCTACA - 3’	1259
R2-R4	Fwd : 5’ - TTCGCATCATGTCGCTCTTTCA - 3’	1135
	Rev1 : 5’ - ATGAATCCATTTGGAGCAGTCT - 3’	1155
	Rev2 : 5’ - CAGCATGCAAGGATACCAGC - 3’	1308

5.4. Statistical analysis

Statistical analysis of the *in situ* results was performed using the R® program (R Foundation for Statistical Computing, Austria). Genotyped embryos were categorized into three groups based on labelling intensity (strong, intermediate, light). A contingency table was constructed to display the number of embryos of each genotype (WT, +/-, and -/-) within each category. Fisher's exact test was then applied to determine the p-values, assessing the association between labelling intensity and genotype. A p-value < 0.05 was considered statistically significant. Graphs were generated using GraphPad Prism® (GraphPad Software, USA).

5.5. Test of the Dependence of the R3 Region on Retinoic Acid

R3:GFP fish were outcrossed with WT fish. Resulting embryos were treated with Retinoic acid (0,5 µM) or DEAB (4 µM) from 5,5 hpf (Bud) to 24 hpf. Photos were taken at 24 hpf with a fluorescence binocular (Leica, M165FC).

5.6. Test of R2 region activation

As explain on Figure 25, R2:GFP embryos, resulting from outcrosses of R2:GFP fish, were treated with DEAB (4 µM) and/or RA (3nm) at different concentration and at different developmental stages. Photos were taken at 24 and 48 hpf with a fluorescence binocular (Leica, M165FC).

6. Bibliography

- Amores, A., Force, A., Yan, Y.-L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y.-L., Westerfield, M., Ekker, M., & Postlethwait, J. H. (1998). Zebrafish *hox* Clusters and Vertebrate Genome Evolution. *Science*, *282*(5394), 1711-1714. https://doi.org/10.1126/science.282.5394.1711

- Avantaggiato, V., Acampora, D., Tuorto, F., & Simeone, A. (1996). Retinoic Acid Induces Stage-Specific Repatterning of the Rostral Central Nervous System. *Developmental Biology*, *175*(2), 347-357. https://doi.org/10.1006/dbio.1996.0120

- Balmer, J. E., & Blomhoff, R. (2002). Gene expression regulation by retinoic acid. *Journal of Lipid Research*, *43*(11), 1773-1808. https://doi.org/10.1194/jlr.R100015-JLR200

- Bastien, J., & Rochette-Egly, C. (2004). Nuclear retinoid receptors and the transcription of retinoid-target genes. *Gene*, *328*, 1-16. https://doi.org/10.1016/j.gene.2003.12.005

- Bertrand, S., Thisse, B., Tavares, R., Sachs, L., Chaumot, A., Bardet, P.-L., Escrivà, H., Duffraisse, M., Marchand, O., Safi, R., Thisse, C., & Laudet, V. (2007). Unexpected Novel Relational Links Uncovered by Extensive Developmental Profiling of Nuclear Receptor Expression. *PLoS Genetics*, *3*(11), e188. https://doi.org/10.1371/journal.pgen.0030188

- Burke, A. C., Nelson, C. E., Morgan, B. A., & Tabin, C. (1995). *Hox* genes and the evolution of vertebrate axial morphology. *Development*, *121*(2), 333-346. https://doi.org/10.1242/dev.121.2.333

- Cavalheiro, G. R., Pollex, T., & Furlong, E. E. (2021). To loop or not to loop : What is the role of TADs in enhancer function and gene regulation? *Current Opinion in Genetics & Development*, *67*, 119-129. https://doi.org/10.1016/j.gde.2020.12.015

- Chatagnon, A., Veber, P., Morin, V., Bedo, J., Triqueneaux, G., Sémon, M., Laudet, V., d'Alché-Buc, F., & Benoit, G. (2015). RAR/RXR binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements. *Nucleic Acids Research*, *43*(10), 4833-4854. https://doi.org/10.1093/nar/gkv370

- Deschamps, J., & Duboule, D. (2017). Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes & Development*, *31*(14), 1406-1416. https://doi.org/10.1101/gad.303123.117

- Désiront, C. (2022). *Étude fonctionnelle des séquences régulatrices activées par l'acide rétinoïque chez le poisson-zèbre* [Mémoire]. ULiège.

- Di Masi, A., Leboffe, L., De Marinis, E., Pagano, F., Cicconi, L., Rochette-Egly, C., Lo-Coco, F., Ascenzi, P., & Nervi, C. (2015). Retinoic acid receptors : From molecular mechanisms to cancer therapy. *Molecular Aspects of Medicine*, *41*, 1-115. https://doi.org/10.1016/j.mam.2014.12.003

- Diez Del Corral, R., & Morales, A. (2014). Retinoic Acid Signaling during Early Spinal Cord Development. *Journal of Developmental Biology*, *2*(3), 174-197. https://doi.org/10.3390/jdb2030174

- 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
- Duboule, D., & Dollé, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. *The EMBO Journal*, *8*(5), 1497-1505. https://doi.org/10.1002/j.1460-2075.1989.tb03534.x
- Durston, A. J., Timmermans, J. P. M., Hage, W. J., Hendriks, H. F. J., De Vries, N. J., Heideveld, M., & Nieuwkoop, P. D. (1989). Retinoic acid causes an anteroposterior transformation in the developing central nervous system. *Nature*, *340*(6229), 140-144. https://doi.org/10.1038/340140a0
- 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. https://doi.org/10.1242/dev.00573
- Gould, A., Morrison, A., Sproat, G., White, R. A., & Krumlauf, R. (1997). Positive cross-regulation and enhancer sharing : Two mechanisms for specifying overlapping Hox expression patterns. *Genes & Development*, *11*(7), 900-913. https://doi.org/10.1101/gad.11.7.900
- Groff, J. L., & Gropper, S. A. S. (2000). *Advanced nutrition and human metabolism* (3rd ed). West/Wadsworth.
- Hawkins, M. R., & Wingert, R. A. (2023). Zebrafish as a Model to Study Retinoic Acid Signaling in Development and Disease. *Biomedicines*, *11*(4), 1180. https://doi.org/10.3390/biomedicines11041180
- Hubert, K. A., & Wellik, D. M. (2023). Hox genes in development and beyond. *Development*, *150*(1), dev192476. https://doi.org/10.1242/dev.192476
- Irfan, M., Majeed, H., Iftikhar, T., & Ravi, P. K. (2024). A review on molecular scissoring with CRISPR/Cas9 genome editing technology. *Toxicology Research*, *13*(4), tfae105. https://doi.org/10.1093/toxres/tfae105
- Izpisúa-Belmonte, J. C., Falkenstein, H., Dollé, P., Renucci, A., & Duboule, D. (1991). Murine genes related to the Drosophila AbdB homeotic genes are sequentially expressed during development of the posterior part of the body. *The EMBO Journal*, *10*(8), 2279-2289. https://doi.org/10.1002/j.1460-2075.1991.tb07764.x
- Kanai, M., Raz, A., & Goodman, D. S. (1968). Retinol-binding protein : The transport protein for vitamin A in human plasma. *Journal of Clinical Investigation*, *47*(9), 2025-2044. https://doi.org/10.1172/JCI105889
- Kawakami, K. (2007). Tol2 : A versatile gene transfer vector in vertebrates. *Genome Biology*, *8*(Suppl 1), S7. https://doi.org/10.1186/gb-2007-8-s1-s7
- Kumar, S., & Duester, G. (2014). Retinoic acid controls body axis extension by directly repressing *Fgf8* transcription. *Development*, *141*(15), 2972-2977. https://doi.org/10.1242/dev.112367
- Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature*, *276*(5688), 565-570. https://doi.org/10.1038/276565a0

- López-Pérez, A. R., Balwierz, P. J., Lenhard, B., Muller, F., Wardle, F. C., Manfroid, I., Voz, M. L., & Peers, B. (2021). Identification of downstream effectors of retinoic acid specifying the zebrafish pancreas by integrative genomics. *Scientific Reports*, *11*(1), 22717. https://doi.org/10.1038/s41598-021-02039-y

- Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., & Duboule, D. (2011). A Regulatory Archipelago Controls Hox Genes Transcription in Digits. *Cell*, *147*(5), 1132-1145. https://doi.org/10.1016/j.cell.2011.10.023

- Moutier, E., Ye, T., Choukrallah, M.-A., Urban, S., Osz, J., Chatagnon, A., Delacroix, L., Langer, D., Rochel, N., Moras, D., Benoit, G., & Davidson, I. (2012). Retinoic Acid Receptors Recognize the Mouse Genome through Binding Elements with Diverse Spacing and Topology. *Journal of Biological Chemistry*, *287*(31), 26328-26341. https://doi.org/10.1074/jbc.M112.361790

- Neijts, R., Amin, S., Van Rooijen, C., Tan, S., Creyghton, M. P., De Laat, W., & Deschamps, J. (2016). Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos. *Genes & Development*, *30*(17), 1937-1942. https://doi.org/10.1101/gad.285767.116

- Niederreither, K., & Dollé, P. (2008). Retinoic acid in development : Towards an integrated view. *Nature Reviews Genetics*, *9*(7), 541-553. https://doi.org/10.1038/nrg2340

- 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

- Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., & Duboule, D. (2011). The Dynamic Architecture of *Hox* Gene Clusters. *Science*, *334*(6053), 222-225. https://doi.org/10.1126/science.1207194

- Osborne, T. B., Mendel, L. B., Ferry, E. L., & Wakeman, A. J. (1913). THE RELATION OF GROWTH TO THE CHEMICAL CONSTITUENTS OF THE DIET. *Journal of Biological Chemistry*, *15*(2), 311-326. https://doi.org/10.1016/S0021-9258(18)88530-1

- Petkovich, M., & Chambon, P. (2022). Retinoic acid receptors at 35 years. *Journal of Molecular Endocrinology*, *69*(4), T13-T24. https://doi.org/10.1530/JME-22-0097

- Piersma, A. H., Hessel, E. V., & Staal, Y. C. (2017). Retinoic acid in developmental toxicology : Teratogen, morphogen and biomarker. *Reproductive Toxicology*, *72*, 53-61. https://doi.org/10.1016/j.reprotox.2017.05.014

- Razin, S. V., Zhegalova, I. V., & Kantidze, O. L. (2022). Domain Model of Eukaryotic Genome Organization : From DNA Loops Fixed on the Nuclear Matrix to TADs. *Biochemistry (Moscow)*, *87*(7), 667-680. https://doi.org/10.1134/S0006297922070082

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

- Robinson-Rechavi, M., Garcia, H. E., & Laudet, V. (2003). The nuclear receptor superfamily. *Journal of Cell Science*, *116*(4), 585-586. https://doi.org/10.1242/jcs.00247

- Rochette-Egly, C., & Germain, P. (2009). Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs). *Nuclear Receptor Signaling*, *7*(1), nrs.07005. https://doi.org/10.1621/nrs.07005

- Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murtha, M. T., & Pendleton, J. W. (1994). Evolution of Hox Genes. *Annual Review of Genetics*, *28*(1), 423-442. https://doi.org/10.1146/annurev.ge.28.120194.002231
- Schilling, T. F., & Kinght, R. D. (2001). Origins of anteroposterior patterning and *Hox* gene regulation during chordate evolution. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1414), 1599-1613. https://doi.org/10.1098/rstb.2001.0918
- Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics*, *20*(8), 437-455. https://doi.org/10.1038/s41576-019-0128-0
- Schubert, M., Holland, N. D., Laudet, V., & Holland, L. Z. (2006). A retinoic acid-Hox hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. *Developmental Biology*, *296*(1), 190-202. https://doi.org/10.1016/j.ydbio.2006.04.457
- Spitz, F., Gonzalez, F., & Duboule, D. (2003). A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the HoxD Cluster. *Cell*, *113*(3), 405-417. https://doi.org/10.1016/S0092-8674(03)00310-6
- Spitz, F., Gonzalez, F., Peichel, C., Vogt, T. F., Duboule, D., & Zákány, J. (2001). Large scale transgenic and cluster deletion analysis of the *HoxD* complex separate an ancestral regulatory module from evolutionary innovations. *Genes & Development*, *15*(17), 2209-2214. https://doi.org/10.1101/gad.205701
- Stafford, D., & Prince, V. E. (2002). Retinoic Acid Signaling Is Required for a Critical Early Step in Zebrafish Pancreatic Development. *Current Biology*, *12*(14), 1215-1220. https://doi.org/10.1016/S0960-9822(02)00929-6
- Strate, I., Min, T. H., Iliev, D., & Pera, E. M. (2009). Retinol dehydrogenase 10 is a feedback regulator of retinoic acid signalling during axis formation and patterning of the central nervous system. *Development*, *136*(3), 461-472. https://doi.org/10.1242/dev.024901
- Veldman, M. B., & Lin, S. (2008). Zebrafish as a Developmental Model Organism for Pediatric Research. *Pediatric Research*, *64*(5), 470-476. https://doi.org/10.1203/PDR.0b013e318186e609
- Wingert, R. A., Selleck, R., Yu, J., Song, H.-D., Chen, Z., Song, A., Zhou, Y., Thisse, B., Thisse, C., McMahon, A. P., & Davidson, A. J. (2007). The cdx Genes and Retinoic Acid Control the Positioning and Segmentation of the Zebrafish Pronephros. *PLoS Genetics*, *3*(10), e189. https://doi.org/10.1371/journal.pgen.0030189
- Yang, H., Luan, Y., Liu, T., Lee, H. J., Fang, L., Wang, Y., Wang, X., Zhang, B., Jin, Q., Ang, K. C., Xing, X., Wang, J., Xu, J., Song, F., Sriranga, I., Khunsriraksakul, C., Salameh, T., Li, D., Choudhary, M. N. K., … Yue, F. (2020). A map of cis-regulatory elements and 3D genome structures in zebrafish. *Nature*, *588*(7837), 337-343. https://doi.org/10.1038/s41586-020-2962-9

Master's Thesis presented by: Lentjes Colin
Supervisor : Peers Bernard
Laboratory of Zebrafish Development and Disease Models (ZDDM)
September 2024

Étude fonctionnelle des séquences régulatrices activées par l'acide rétinoïque et de leur rôle dans l'expression des gènes *hox* chez le poisson-zèbre

Ce mémoire examine le rôle fonctionnel de séquences régulatrices activées par l'acide rétinoïque et leur impact sur l'expression des gènes *hox* chez le poisson-zèbre (*Danio rerio*). Les gènes *Hox* sont essentiels pour le développement embryonnaire des vertébrés, en particulier pour la segmentation corporelle et la formation de structures le long de l'axe antéro-postérieur. L'acide rétinoïque, un métabolite de la vitamine A, joue un rôle clé dans la régulation de l'expression de ces gènes.

Dans cette étude, nous avons concentré nos investigations sur le rôle de quatre régions « RARE », R1 à R4, localisées loin en aval du complexe *hoxbb* (entre 10 et 60 kb) et dans le gène avoisinant *skap1*. Etant donné que le gène *skap1* n'est pas régulé par l'acide rétinoïque, ces régions R1-R4 étaient suspectées réguler à longue distance les gènes du complexe *hoxbb*. Nous avons vérifié cette hypothèse en analysant le profil d'expression des 4 gènes *hoxb1b*, *hoxb5b*, *hoxb6b*, et *hoxb8b* à divers stades de développement dans des mutants où les régions R1 à R4 ont été enlevées. Les résultats ont révélé que la délétion des séquences comprenant les éléments R2, R3 et R4 diminuait significativement l'expression des quatre gènes *hoxbb* testés. Le niveau d'expression des gènes *hox* était plus fortement diminué au stade tardif (48 hpf) qu'à des stades précoces (15 hpf). Cependant, nous avons remarqué que, au stade 15 hpf, le profil d'expression du gène *hoxb6b* était modifié, présentant une expression décalée vers la partie postérieure de l'embryon. Etonnamment, la délétion uniquement de la région R2 n'a pas modifié significativement l'expression des gènes *hoxbb*, bien que cette région soit la plus conservée parmi les quatre séquences RARE. Ceci suggère une redondance fonctionnelle potentielle ou une compensation notamment entre les régions R2 et R4.

Des expériences utilisant des lignées transgéniques, où les séquences R2 et R3 sont insérées en amont du promoteur cfos lié aux séquences codant pour la GFP, ont permis de montrer que la région R3 n'est pas suffisante pour conférer une régulation par l'acide rétinoïque, contrairement à la lignée R2-cfos:GFP. Des traitements DEAB sur la lignée R2 ont montré que cette région était plus active pendant la somitogenèse et restait active entre 24 et 48 heures après fécondation. Ces résultats étaient cohérents avec les observations issues des expériences de délétions des éléments RARE.

Bien que des études supplémentaires soient nécessaires pour une compréhension plus complète de l'action de ces régions, cette recherche a contribué à approfondir de manière significative notre connaissance des mécanismes moléculaires régissant la régulation des gènes *hoxbb* par l'acide rétinoïque chez le poisson-zèbre.

Master's Thesis presented by: Lentjes Colin
Supervisor : Peers Bernard
Laboratory of Zebrafish Development and Disease Models (ZDDM)
September 2024

Functional study of retinoic acid-activated regulatory sequences and their role in the expression of *hox* genes in zebrafish

This master's thesis examines the functional role of retinoic acid-responsive regulatory sequences and their impact on the expression of *hox* genes in zebrafish (Danio rerio). *Hox* genes are crucial for vertebrate embryonic development, particularly in body segmentation and the formation of structures along the anterior-posterior axis. Retinoic acid, a metabolite of vitamin A, plays a key role in regulating the expression of these genes.

In this study, we focused our investigations on the role of four "RARE" regions, R1 to R4, located far downstream of the *hoxbb* complex (between 10 and 60 kb) and within the neighbouring *skap1* gene. Since the *skap1* gene is not regulated by retinoic acid, these R1-R4 regions were suspected to regulate *hoxbb* genes over a long distance. We tested this hypothesis by analysing the expression profile of four hox genes—*hoxb1b*, *hoxb5b*, *hoxb6b*, and *hoxb8b*—at various developmental stages in mutants where the R1 to R4 regions were removed. The results revealed that deletion of sequences containing the R2, R3, and R4 elements significantly decreased the expression of the four tested *hoxbb* genes. The expression levels of *hox* genes were more significantly reduced at the late stage (48 hpf) than at early stages (15 hpf). However, we observed that at the 15 hpf stage, the expression profile of the *hoxb6b* gene was altered, showing a shift in expression towards the posterior part of the embryo. Surprisingly, the deletion of the R2 region alone did not significantly alter the expression of the *hoxbb* genes, although this region is the most conserved among the four RARE sequences. This suggests potential functional redundancy or compensation, particularly between the R2 and R4 regions.

Experiments using transgenic lines, where the R2 and R3 sequences were inserted upstream of the cfos promoter linked to GFP-coding sequences, demonstrated that the R3 region is not sufficient to confer retinoic acid regulation, in contrast to the R2-cfos line. DEAB treatments on the R2 line showed that this region was more active during somitogenesis and remained active between 24 and 48 hpf. These results were consistent with observations from the RARE element deletion experiments.

Although further studies are necessary for a more comprehensive understanding of the actions of these regions, this research has significantly deepened our knowledge of the molecular mechanisms governing the regulation of *hoxbb* genes by retinoic acid in zebrafish.