Characterization of the Hoxd4 neural enhancer

Ler, Ser Yeng

2012

Ler, S. Y. (2012). Characterization of the Hoxd4 neural enhancer. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/50718>

<https://doi.org/10.32657/10356/50718>

NANYANG
TECHNOLOGICAL
UNIVERSITY

Characterization of the *Hoxd4* neural enhancer

A thesis submitted to

Nanyang Technological University

In partial fulfilment of the requirement for
the degree of Doctor of Philosophy of Science

by

Ler Ser Yeng

Dec 2011

© Ler Ser Yeng, 2011

For my grandmum.

Table of contents

Abstract.....	4
Acknowledgement.....	6
List of abbreviations.....	7
Chapter 1 Introduction.....	11
Chapter 2 Materials and methods.....	58
Chapter 3 Study of <i>Hoxd4/hoxd4a</i> neural enhancer in P19 cells.....	72
Chapter 4 Recapitulation of <i>Hoxd4/hoxd4a</i> expression in zebrafish embryos.....	83
Chapter 5 Study of <i>hoxd4a</i> neural enhancer in zebrafish embryos.....	140
Chapter 6 The role of Kreisler/valentino in regulating <i>Hoxd4/hoxd4a</i> neural expression.....	156
Chapter 7 Conclusion/Future work.....	169
References.....	174

Abstract

Hox genes encode homeodomain-containing transcription factors, required for correct patterning of the anterior-posterior axis, genital tubercle and limbs, organogenesis, morphogenesis, hematopoiesis, hair growth and development, and mammary gland maturation. They are highly conserved throughout the animal kingdom and their expression is tightly regulated.

The Featherstone lab aims to understand how *Hox* expression is controlled through the study of the vertebrate *Hoxd4* orthologs. In mammals and birds, there are 39 *Hox* genes. These genes are distributed over four clusters, named A-D, and they are subdivided into thirteen paralog groups. *Hoxd4* is the fourth paralog in the *HoxD* cluster. *Hoxd4* expression has an anterior expression border between segments (rhombomeres (r)) 6/7 of the developing hindbrain and at cervical vertebra (C) 1 of the axial skeleton. Our lab is interested in understanding *Hoxd4* expression in the developing hindbrain. In this regard, a 3' neural enhancer in murine *Hoxd4* has been mapped and shown to be required for setting the r6/7anterior expression border in the developing hindbrain.

In zebrafish, there are 49 *hox* genes, distributed over seven clusters, denoted Aa, Ab, Ba, Bb, Ca, Cb and D. A 3' neural enhancer has also been mapped in orthologous zebrafish *hoxd4a*. Previous sequence comparison between mouse and zebrafish *Hoxd4/hoxd4a* neural enhancers revealed eight blocks of homology, termed sites A to G and a retinoic acid response element (RARE). Sites D to F and the RARE have been previously characterized in the murine 3' neural enhancer. It is now of interest to characterize the remaining sites A, B, C and G.

A GFP-encoding transcriptional reporter based on the murine presumptive *Hoxd4* promoter and murine 3' neural enhancer was used to assess enhancer function in neurally differentiating P19 cells. Results were not conclusive, possibly because the presumptive promoter region used to drive transgene expression may not actually harbour promoter activity.

An unsuccessful attempt to assess enhancer function in zebrafish embryos was made using a Tol2 transposon-based vector with a GFP-encoding transcriptional reporter

including the murine *Hoxd4* 5' upstream region spanning from -1 to -5398 bp with respect to the ATG start codon, and murine 3' neural enhancer. However, r6/7 expression was recapitulated in the developing hindbrain of zebrafish embryos at 24 hpf when I tested a Tol2 transposon-based vector with a GFP-encoding transcriptional reporter bearing the zebrafish *hoxd4a* 5' upstream region spanning from -1 to -5235 bp with respect to the ATG start codon, and zebrafish 3' neural enhancer. Further investigation into the role of the 3' neural enhancer showed that it was not responsible for setting the r6/7 boundary. Therefore, I investigated whether an unexpected and additional enhancer may lie in the 5' upstream region. My mapping studies led to the identification of a 5' neural enhancer with r6/7 activity located between -355 and -538 bp, the position of the *hoxd4a* P3 promoter between -1 and -200 bp and the speculative proposition for a repressor region between -3469 and -3368 bp with respect to the ATG start codon.

Though the 3' neural enhancer was not responsible for setting the r6/7 boundary, the evolutionary conservation of the sites indicates some importance. This was borne out by mutational analyses. Expression was altered in embryos injected with constructs containing mutated i) site B, ii) sites A and B, and iii) sites A, B and C in the 3' neural enhancer. From these results, sites A, B and G are speculated to be evolutionarily important to the activity of the 3' neural enhancer. The r6/7 expression observed with construct containing mutated sites A, B, C and G may imply that sites D, E, F and RARE are adequate for the function of 3' neural enhancer.

Genetic evidence suggests that the Maf family transcription factor Kreisler (Kr)/MafB (Valentino (Val) in zebrafish) is a negative regulator of *Hoxd4* in the mouse embryonic hindbrain. *In situ* hybridization of zebrafish *val* morphants with *hoxd4a* and *krox20* RNA probes showed the anteriorization of the r6/7 boundary to r5/6. Sequence alignment revealed that conserved sites F and G show weak homology to the consensus maf recognition sequence (TGCTGANT/CCNGNN), suggesting that these elements might mediate the repressive effects of Val. Preliminary results from electrophoretic mobility shift assays (EMSA) and coinjection of the *val* morpholino and the transgene containing the 5' and 3' neural enhancers show that Kr does not interact with the neural enhancer(s), nor does it bind to sites F and G. These results suggest that Val binds sequences elsewhere beyond the 3' neural enhancer.

Acknowledgement

I thank Prof Mark Featherstone (Nanyang Technological University, Singapore) for his guidance and supervision. Special thanks to Dr Cyrus Papan for his advice on microscopy, Geraldine Lee for her advice on flow cytometry, as well as Dr Shen Hui and Dr Ragini Rai Khandelia for their advice on molecular biology. I will also like to thank Assoc/Prof Andrew Waskiewicz (University of Alberta, Canada) for his supervision during my four-month exchange programme in Canada.

List of abbreviations

1. **abd-A**: abdominal-A
2. **Abd-B**: Abdominal-B
3. **ANT-C**: Antennapedia complex
4. **Antp**: Antennapedia
5. **AP**: Activating Protein
6. **ARE**: autoregulatory element
7. **BCIP**: 5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt
8. **BD**: Becton-Dickinson
9. **bp**: base pair
10. **BLAST**: Basic Local Alignment Search Tool
11. **BMP**: Bone Morphogenetic Protein
12. **Bra**: brachyury
13. **bZIP**: basic leucine zipper
14. **BX-C**: Bithorax complex
15. **C**: cervical vertebra
16. **CBP**: cAMP response element binding protein
17. **CDK4**: cyclin-dependent kinase 4
18. **C.elegans**: *Caenorhabditis elegans*
19. **ChIP**: **Ch**romatin **Immu**no**P**recipitation
20. **Cyp26**: Cytochrome P450, family 26
21. **Dfd**: Deformed
22. **dpc**: days post-coitum
23. **dpi**: dots per inch
24. **dpf**: days post-fertilization
25. **DR**: direct repeat
26. **EGTA**: ethylene glycol tetraacetic acid
27. **ELCR**: early limb control region
28. **EMSA**: electrophoretic mobility shift assay
29. **Eng**: Engrailed
30. **ERE**: estrogen response element
31. **eve**: even-skipped
32. **Evx-2**: Even-skipped homeobox-2

- 33. **Exd**: Extradenticle
- 34. **Eya**: eyes absent
- 35. **FACS**: Fluorescence Activated Cell Sorting
- 36. **FSC**: Forward Scatter
- 37. **fgf**: fibroblast-growth factor
- 38. **FGFR1**: fibroblast growth factor receptor-1
- 39. **FLI1**: Friend leukemia virus integration 1
- 40. **ftz**: fushi tarazu
- 41. **GATA-1**: GATA binding protein 1
- 42. **GCR**: global control region
- 43. **Gdf11**: Growth differentiation factor 11
- 44. **GFP**: green fluorescent protein
- 45. **GLI**: glioblastoma-associated oncogene
- 46. **GST**: glutathione-S-transferase
- 47. **HAND2**: Heart and neural crest derivatives expressed 2
- 48. **HMG**: High mobility group
- 49. **HOM-C**: homeotic complex
- 50. **HSC**: hematopoietic stem cell
- 51. **hpf**: hours post-fertilization
- 52. **Hth**: homothorax
- 53. **hyb**: hybridization
- 54. **iab**: infraabdominal
- 55. **IL**: interleukin
- 56. **IPTG**: Isopropyl β -D-1-thiogalactopyranoside
- 57. **KAP**: keratin-associated proteins
- 58. **Kr**: Kreisler
- 59. **lab**: labial
- 60. **LnP**: lunapark
- 61. **Lzr**: Lazarus
- 62. **MEIS**: murine ecotropic integration site
- 63. **mRNA**: messenger RNA
- 64. **miRNA**: microRNA
- 65. **MLL**: mixed-lineage leukaemia
- 66. **NBT**: Nitro blue tetrazolium chloride

67. **NE**: neural enhancer
68. **NFY**: Nuclear Factor Y
69. **NRL**: neural retina leucine zipper
70. **OCT**: octamer-binding protein
71. **OD** : optical density
72. **PAX**: Paired box
73. **pb**: proboscipedia
74. **PBC**: pre-B-cell
75. **PBS**: Phosphate-Buffered Saline
76. **PBX**: pre-B-cell leukaemia transcription factor
77. **PcG**: polycomb group
78. **PCR**: polymerase chain reaction
79. **PREP**: Pbx regulating protein
80. **PVDF**: Polyvinylidene fluoride
81. **RA**: Retinoic acid
82. **RARE**: retinoic acid response element
83. **r**: rhombomere
84. **RB** : Rohon-Beard
85. **Scr**: Sex comb reduced
86. **SDS-PAGE**: Sodium Dodecyl Sulfate-PolyAcrylamide Gel
Electrophoresis
87. **shh**: sonic hedgehog
88. **six**: sine oculis
89. **SOX**: Sex-determining Region Y-type high-mobility-group box
90. **SSC**: saline-sodium citrate
91. **SSC**: side scatter
92. **svp**: seven up
93. **TALE**: Three Amino acid Loop Extension
94. **TBST**: Tris-Buffered Saline Tween-20
95. **TGF**: transforming growth factor
96. **TNT**: transcription/translation
97. **trxG**: trithorax
98. **Ubx**: Ultrabithorax
99. **UTR**: untranslated region

- 100. **Val:** Valentino
- 101. **vhnf1:** variant hepatocyte nuclear factor 1
- 102. **WDR5:** tryptophan-aspartic acid repeat domain 5
- 103. **Wnt:** Wingless-Int
- 104. **YY1:** Yin Yang-1

Chapter 1 Introduction

1.1 Hox genes

Hox genes were first characterized in *Drosophila melanogaster* by E.B. Lewis in 1978. *Drosophila* has altogether eight *Hox* genes, arranged into two sub-clusters, located on the third chromosome. The Antennapedia complex (ANT-C) contains the labial (*lab*), proboscipedia (*pb*), Deformed (*Dfd*), Sex comb reduced (*Scr*), and Antennapedia (*Antp*) genes; and the Bithorax complex (BX-C) contains the Ultrabithorax (*Ubx*), abdominal-A (*abd-A*), and Abdominal-B (*Abd-B*) genes. Together, these complexes are referred to as the homeotic complex (HOM-C) (**Fig. 1-1A**).

Early in evolution, the HOM-C was duplicated twice, resulting in the four clusters seen in mammals and birds. Thirty-nine genes are distributed over these four clusters, A-D, and they are further subdivided into thirteen paralog groups (**Fig. 1-1B**). Due to gene loss, no single cluster retains all thirteen. Mammalian paralog groups 1, 2, 4, 5, 6, 7 and 8 are related to *lab*, *pb*, *Dfd*, *Scr*, *Antp*, *Ubx* and *abd-A* respectively. Paralog groups 9-13 are more closely related to the *Abd-B* gene of the BX-C. No *Drosophila* counterpart is found to be related to group 3. Further whole genome duplication occurred in teleost fish; hence zebrafish has seven clusters with 49 *Hox* genes (**Fig. 1-1C**).

1.2 Hox structure

Mammalian *Hox* genes are relatively small. They generally contain two coding exons. The 180 base pair (bp) homeobox is always found within the second exon (**Fig. 1-2**). *Hoxa1* has, in addition, a shorter transcript that lacks the homeobox. This is due to an alternative splicing event that removes 203 bp within exon 1, hence resulting in a frameshift and a premature termination (LaRosa and Gudas, 1988).

The homeobox encodes an evolutionarily highly conserved 60 residue helix-turn-helix protein structure called the homeodomain. It consists of three α -helices with a variable smaller fourth helix and an N-terminal arm (Qian et al., 1989). The second and third helices form a helix-turn-helix motif that recognizes and binds to the major groove of DNA. Residues 47, 50, 51 and 54 of the third α -helix contact bps 3 through 6 of the recognition sequence. Residues 50, 51 and 54 are glutamine, asparagine and

Figure 1-1: The evolution of Hox clusters. A) The HOM-C is made up of two sub-clusters – ANT-C and BX-C. The ANT-C contains *lab*, *pb*, *Dfd*, *Scr*, *Antp* and the BX-C contains *Ubx*, *abd-A* and *Abd-B*. Early in evolution, the HOM-C was duplicated twice, resulting in the four clusters seen in mouse as shown in B). B) Mammalian paralog groups 1, 2, 4, 5, 6, 7 and 8 are related to *lab*, *pb*, *Dfd*, *Scr*, *Antp*, *Ubx* and *abd-A* respectively. Paralog groups 9-13 are more closely related to the *Abd-B* gene of the BX-C. No *Drosophila* counterpart is found to be related to group 3. C) Further whole genome duplication occurred in teleost fish; hence zebrafish has seven clusters (Prince and Pickett, 2002).

Figure 1-2: Structure of *Hox* gene. The *Hox* genes generally contain two exons and the homeobox is always found in the second exon (Lappin et al., 2006).

methionine respectively (LaRonde-LeBlanc and Wolberger, 2003). Except Pb which has valine as the 47th residue, others have isoleucine (Cribbs et al., 1992) .

Specific residues of the N-terminal arm make additional base-pair contacts with the minor groove. Differences in residues 3 and 5 of the N-terminal arm determine the binding site preferences (**Fig. 1-3**) (Phelan et al., 1994) .

1.3 HOX and their cofactors

HOX proteins are known to interact with the pre-B-cell (PBC) and murine ecotropic integration site (MEIS) classes of Three Amino acid Loop Extension (TALE) homeodomain proteins, Sex-determining Region Y-type high-mobility-group box (SOX) and POU domain-containing Octamer-binding proteins (OCT), to enhance their DNA-binding specificity. The PBC class comprises *Drosophila* Extradenticle (Exd), mammalian pre-B-cell leukaemia transcription factor (Pbx) 1, Pbx2, Pbx3 and Pbx4, zebrafish Lazarus (Lzr) and *Caenorhabditis elegans* (*C.elegans*) homeobox-20 (Ceh-20). The MEIS class comprises members of the MEIS and Pbx regulating protein (PREP). Members of MEIS are the three murine MEIS1, MEIS2 and MEIS3, *Drosophila* homothorax (Hth) and Ceh-25. Members of PREP are the mammalian PREP1 and PREP2.

HOX proteins from mammalian paralog groups 1 to 8, equivalent to all *except Abd-B* in flies, contain a conserved YPWM motif N-terminal to the homeodomain (**Fig. 1-4**). This is crucial for interaction with the PBC class of homeodomain proteins. HOX proteins in paralogs 9 and 10 have ANW instead. These conserved motifs are collectively called the Pbx-interacting motifs. The extreme N-terminus is somewhat conserved among HOX proteins too.

Only HOX proteins of paralog groups 1-10 are able to interact with PBX proteins because HOX proteins of the other three paralog groups lack a Pbx-interacting motif. PBX binds to the 5' TGAT half-site of the PBX recognition site (TGATTNATGG) and the HOX Pbx-interacting motif in the N-terminal region upstream of the homeodomain makes contact to a hydrophobic pocket in the PBX homeodomain (**Fig.1-5**) (Lufkin, 2003).

HOX proteins of paralog groups 9-13 bind cooperatively with MEIS to the sequence 5' TGACAGTTTACGAC 3'. MEIS binds to the 5' TGAC half site and the HOX proteins bind to TTACG. Besides heterodimerization with HOX proteins, MEIS/PREP

proteins and PBX proteins can also heterodimerize. Three complexes were described both in the presence and absence of DNA (**Fig. 1-6**) (Lufkin, 2003).

Data shows that SOX/OCT heterodimers bind to the b1-autoregulatory element (ARE) in combination with HOXA1/PBC to mediate retinoic acid (RA) *in vivo* response of the b1-ARE (Di Rocco et al., 2001) .

Figure 1-3: The homeodomain bound to DNA. The homeodomain consists of three α -helices (I, II and III) with a variable smaller fourth helix (not shown). The second and third helices form a helix-turn-helix motif that recognizes and binds to the major groove of DNA. Residues 47, 50, 51 and 54 of the third α -helix contact bps 3 through 6 of the recognition sequence (Nolte, 2002).

Figure 1-4: Structure of HOX protein. HOX proteins from paralog groups 1 to 8, equivalent to all except Abd-B in flies, contain a conserved YPWM motif at the N-terminal site of the homeodomain (Lappin et al., 2006).

Figure 1-5: PBX.HOX dimer. PBX binds to the 5' TGAT half-site of the PBX recognition site (TGATTNATGG) and the HOX Pbx interacting motif in the N-terminal region upstream of the homeodomain makes contact to a hydrophobic pocket in the PBX homeodomain (Lufkin, 2003).

Figure 1-6: MEIS/PREP proteins can heterodimerize with PBX in the presence and absence of DNA. A: PBX.MEIS dimer binds to TGATTGACAG. B: PBX.HOX.MEIS trimer. MEIS binds to PBX without contacting DNA. C: PBX.HOX.MEIS trimer. MEIS binds at CTGTCA upstream of PBX.HOX complex (Lufkin, 2003).

1.4 Hox functions

Hox genes are required for correct patterning of the anterior-posterior axis, organogenesis, hematopoiesis and hair growth and development during embryogenesis. They are also expressed in adult tissues.

1.4.1 Correct patterning of the anterior-posterior axis

In embryonic development, *Hox* genes are homeotic selector genes that are responsible for the specification of segment identity. They are regulated by the gap genes and pair rule genes (**Fig. 1-7**). Gap genes are zygotic genes that act downstream of maternal effect genes and divide the embryo into broad territories along the anterior-posterior axis. In *Drosophila*, these genes include zygotic *hunchback*, *krüppel*, *giant*, *tailless* and *knirps*. Hunchback and Krüppel proteins are known to inhibit *abdA* and *AbdB* expression in the head and thorax. Certain levels of Hunchback protein activate *Ubx* in the middle of the embryo. Krüppel activates *Antp* transcription (Struhl et al., 1992; Erfurth et al., 2008). Their boundaries of expression are then confined to the parasegments by *even-skipped* (*eve*) and *fushi tarazu* (*ftz*) (pair rule genes which act downstream of gap genes) (Gilbert, 2000).

Since all genes are transcribed from 5' to 3', we can assign a 3' end and a 5' end to each *Hox* cluster. The 3' genes are expressed more anteriorly and earlier and respond more rapidly to RA than the 5' genes (**Fig. 1-8**). *Hox* genes share enhancers and cross-regulate to reinforce the overlapping expression domains (Gould et al., 1997). Different combinations of Hox proteins generated by the nested domains of expression determine a cell's identity. Misexpression of these genes during embryonic development results in homeotic transformations or malformations. The term "homeosis" was coined by Dr William Bateson in 1894 which means that one body part is developed into the likeness of another.

Figure 1-7: Flow chart of how the anterior-posterior genes are regulated. Maternal effect genes regulate gap genes, which in turn regulate pair-rule genes. Pair rule genes regulate segment polarity genes and *Hox* genes. *Hox* genes can also be regulated by gap genes. Examples of the genes are in colour (Sanson, 2001).

Figure 1-8: The mammalian *Hox* complex. The 3' genes are expressed more anteriorly (spatial) and earlier (temporal) and respond more rapidly to RA than the 5' genes (Nolte, 2002).

In many cases, gain of function mutations result in formation of more posterior structures and loss of function mutations result in formation of more anterior structures. Synpolydactyly (fusions and malformations in fingers and toes) and hand-foot-genital syndrome (big toes, short thumbs and urogenital defects) are the first two malformations shown to be caused by mutations in human *Hox* genes, namely *HOXD13* and *HOXA13*, respectively (Goodman, 2002).

1.4.2 Organogenesis

Organogenesis is the process whereby the internal organs are formed from the three germ layers, namely the ectoderm, mesoderm and endoderm. *Hox* genes were found to be involved in the development of mammary gland, heart, kidney, prostate, gut and lung, limb formation and genital tubercle formation.

1.4.2.1 Mammary gland development

The mammary glands do not develop much during embryogenesis and their development is independent of hormones. Currently, only a few genes have been found to be involved. *Hoxb9* and *Hoxd9* have been shown to be expressed in the mammary mesenchyme surrounding the ductal epithelium (Lewis, 2000).

1.4.2.2 Heart development

Hox genes are involved in establishing the correct anterior-posterior patterning of the heart tube. The heart tube is divided into aorta and heart. The aorta is further divided into anterior and posterior aorta. This is well-studied in the *Drosophila* equivalent of the heart, known as the dorsal vessel. Since the involvement of *Hox* genes in anterior-posterior patterning of the body axis is similar in both *Drosophila* and vertebrates, similarities in heart development are likely to be observed in both *Drosophila* and vertebrates.

In *Drosophila*, no *Hox* gene is expressed in anterior aorta. *Antp*, *abdA*, *AbdB* and *Ultrabithorax* are expressed in posterior aorta. In the heart, only *abdA* is expressed. During the differentiation of posterior tube cardiomyocytes, *abdA* in the heart activates the gene encoding the anion transporter Na⁺-Driven Anion Exchanger 1 that is repressed by *Ubx* in the posterior aorta (Perrin et al., 2004). *abdA* is coexpressed with *Seven up* (*svp*). Upon sonic hedgehog (*shh*) signaling from the dorsal ectoderm, six

pairs of *syp*-expressing cells differentiate into functional inlet-outlet valves for hemolymph, known as ostia (Lovato et al., 2002).

1.4.2.3 Kidney development

The lowest vertebrates form a phylogenetically primitive kidney known as the pronephros. More advanced classes like fish and amphibia build upon the pronephros to form a more advanced kidney called the mesonephros. The amniote classes (reptiles, mammals and birds) extend this further to third stage called the metanephric kidney that persists into adult life.

The more 3' *Hox* genes are expressed more anteriorly, in the ureteric bud. *Hox11* genes are involved in controlling early metanephric induction. Hox11 proteins physically interact with the Pax2 and Eyes absent 1 (*Eya1*) proteins to upregulate the inducing growth factor, *Glial cell line-derived neurotrophic factor* (*Gdnf*), required for ureteric bud formation, and *sine oculis 2* (*Six2*), required for progenitor cell renewal (Brodbeck and Englert, 2004; Gong et al., 2007).

1.4.2.4 Prostate development

The prostate is divided into four lobes- anterior, dorsal, lateral and ventral. Studies in mice show that *Hox* expression in the prostate is androgen-dependent. *Hox10*, *Hox11* and *Hox13* paralog groups, except *Hoxc13*, are involved in prostate development in a collinear fashion. Apparently, *Hox10* paralogs are involved in patterning the anterior prostate, *Hox11* paralogs are involved in the dorsolateral prostate and *Hox13* paralogs are involved in the ventral prostate (Weelik DM, unpublished).

Hoxa13 and *Hoxd13* are involved in outgrowth and branching morphogenesis of the prostate. *Hoxb13* function is redundant to *Hoxd13* (Economides and Capecchi, 2003).

1.4.2.5 Gut development

The gut tube develops from the endoderm. It divides into three parts: foregut, midgut and hindgut. The foregut gives rise to the esophagus, stomach and the first and second part of the duodendum. The midgut gives rise to the third part of the duodendum, jejunum, ileum and the first 2/3 of the large intestine. The hindgut gives rise to the remaining 1/3 of the large intestine and the rectum.

Several *Hox* genes have been found to be involved in gut development. In mouse at 12.5 days post-coitum (dpc), all *HoxA* and *HoxB* genes from paralog groups 4-9 have been found to be expressed in the stomach. Colinearity is observed in the midgut whereby the more 3' genes are expressed more rostrally and the more 5' genes are expressed more caudally. *HoxA* and *HoxB* cluster gene expression starts rostrally from the duodenum, followed by *HoxC* cluster genes from the jejunum and finally *HoxD* cluster genes in the caudal part of the midgut, ileum and cecum. *HoxD* cluster gene expression continues in the hindgut, together with the *Abd-B* family genes (Kawazoe et al., 2002). The budding of the cecum involves the regulation of the Fibroblast-growth factor 10 (*Fgf10*) signaling by the upstream anterior *Hox* genes. *Hoxd12* and *Hoxd13* have been found to be expressed only near the anus (Zacchetti et al., 2007).

1.4.2.6 Lung development

The ventral-lateral foregut endoderm gives rise to lung and trachea. Lung buds form before the branching morphogenesis.

In mouse, lung development starts at 9.5 dpc. *Hoxb3*, *Hoxb4* and *Hoxb5* are highly expressed in the region where lung buds forms. Later in development from 10.5-14.5 dpc, *Hoxb3* and *Hoxb4* are expressed in the mesenchyme of the trachea, bronchi, and peripheral lung while *Hoxb2* and *Hoxb5* are expressed in the mesenchyme of the peripheral lung buds (Maeda et al., 2007). *Hoxa1-5*, *Hoxb3*, *Hoxb4*, *Hoxb6*, *Hoxb7* and *Hoxb8* have been reported to be expressing in the branching region. Opposing effect of RA and transforming growth factor β (TGF β) specifies the expression pattern of *Hoxa2*, *Hoxa4* and *Hoxa5* in the proximal mesenchyme (Cardoso et al., 1996; Mollard and Dziadek, 1997; Packer et al., 2000). *Hoxd4* is also expressed in the developing lung and its expression is lost after 12.5 dpc (Gaunt et al., 1989).

Hoxa5, *Hoxb4* and *Hoxb6* are involved in branching morphogenesis and alveogenesis (Mandeville et al., 2006; Volpe et al., 2008).

1.4.2.7 Limb formation

Hox genes are involved in positioning the limb fields, regulating *Shh* expression and forming the digits (Papageorgiou, 2007). It has been suggested that *Hoxb5*, *Hoxc6*,

Hoxc8 and *Hox9* paralogs such as *Hoxb9*, expressed in the lateral plate mesoderm, work together in positioning the presumptive limbs. Three phases of expression have been described after the prelimb stage with respect to the *HoxA* and *HoxD* clusters (Nelson et al., 1996). These are more studied with the *HoxD* cluster.

During phase one, *Hoxd9* and *Hoxd10* are expressed throughout the lateral plate mesoderm. During phase two, *Hoxd1-d9* genes are uniformly expressed throughout the emerging bud as their promoters respond equally to the early limb control region (ELCR), located on the 3' side of the *HoxD* cluster (**Fig. 1-9**). However, the expressions of *Hoxd10-d13* genes, further away from ELCR, are progressively restricted to the more posterior cells. The accumulation of *Hoxd10-d13* transcripts in the posterior region relieves the repression of glioblastoma-associated oncogene family zinc finger 3 (GLI3) protein on *Shh* and *Heart and neural crest derivatives expressed 2* (*HAND2*) in phase three. *Shh* and *HAND2* activate each other and limit the accumulation of GLI3 in the posterior region. Under the control of the global control region (GCR), *Shh* promotes 5' most *Hoxd* gene transcription in the presumptive digits. GCR is located on the 5' side upstream of the 5' most *Hoxd* genes and contains the digit enhancer and some neural enhancers (Zakany et al., 2004). The digit enhancer acts on *Hoxd13*, *lunapark* (*Lnp*) and *Even-skipped homeobox-2* (*Evx2*) (Tarchini and Duboule, 2006) while the neural enhancers act on *Evx2* and *Lnp* (Spitz et al., 2003). Prox enhancer element, located upstream of *Lnp*, works together with GCR in the expression of posterior *Hoxd* genes (Gonzalez et al., 2007).

Figure 1-9: Genomic organization of *HoxD* cluster. ELCR lies downstream of *Hoxd1*. *Lnp* and *Evx2* lie upstream of *Hoxd13*. *Lnp* and *Evx2* are transcribed in the reverse direction from the *HoxD* genes. Prox, an enhancer element, lies in *Lnp-Evx2* intergenic region, while GCR lies downstream of *Lnp*.

1.4.2.8 Genital tubercle formation

Genital tubercle is a body of tissue that gives rise to penis in males and clitoris in females. *HoxD* genes are found to be expressed in the genital tubercle. Since it is the last structure to develop along the proximal-distal axis, all the genes in the *HoxD* cluster are already expressed. The 5' most genes are expressed more abundantly than the 3' most genes in the mesenchyme.

Before birth, the *HoxD* expression patterns in males and females are the same. After birth, *Hoxd13* transcripts concentrate in the proximal erectile tissue and around the ossifying area of the penis in males but are weak and diffuse in the anlage of the clitoris in females. Depending on the presence or absence of androgen hormones, different structures are formed in males and females (Dolle et al., 1991). It has been found later that *Hoxc11* is also expressed in the genital tubercle (Hostikka and Capecchi, 1998).

Genital tubercle development is similar to that of limbs along the proximal-distal axis except that the *Hox* gene expression domains are more distinct in the limbs than the genital tubercle, and sequential activation of the 3' most genes to the 5' most genes is observed in the limbs but only a sequential decrease of the 3' to 5' genes is observed in the genital tubercle. In the mouse *Ulnaless* mutation, *Hoxd13* expression was reduced in the genital bud; *Hoxd11*, *Hoxd12* and *Hoxd13* expressions were reduced in the distal limbs; and *Hoxd12* and *Hoxd13* were misexpressed in the proximal limbs. This suggests that *Hoxd* expression in the genital bud and limb bud is controlled by the same cis-regulatory element in *HoxD* cluster. Hence, there is possibly a phylogentic relationship between limb formation and genital tubercle formation (Peichel et al., 1997).

1.4.3 Hematopoiesis

Hematopoiesis refers to the blood formation, which starts in the yolk sac, then the liver and finally in the bone marrow of the developing embryo. In the adult, it occurs in the marrow and lymphatic tissues. All blood cells arise from pluripotent stem cells that commit into either the myeloid (erythrocytes, platelets, monocytes, neutrophils, basophils and eosinophils) or the lymphoid lineage (B and T cells).

Most *Hox* genes of the A, B and C clusters are expressed early in hematopoiesis and downregulated during differentiation and maturation. *Hoxa9* is the most important gene in

early hematopoiesis because mice knockout of *Hoxa9* suffer from the most severe phenotypes. Its function overlaps with *Hoxb3* and *Hoxb4*. That is why loss of *Hoxb3* and *Hoxb4* showed only mild phenotypes, suggesting that they are dispensable for normal hematopoietic stem cell (HSC) function (Argiropoulos and Humphries, 2007).

Knockouts of *Hoxa5*, *Hoxa7*, *Hoxa10*, *Hoxb3*, *Hoxb6* and *Hoxc8* possess phenotypes ranging from inability to self-renew HSCs to inability to differentiate into myeloid or B lymphoid cells (Argiropoulos and Humphries, 2007). GATA binding protein 1 (GATA-1), Friend leukemia virus integration 1 (FLI1) and HOXA11 were found to bind to *HOXA10* to repress HOXA10 expression in megakaryocytic cells (Gosiengfiao et al., 2007) which otherwise blocked erythroid and megakaryocytic development (Magnusson et al., 2007).

1.4.4 Adult functions

Hox expression in adult tissues is less studied than in embryonic development. Like in embryonic development, *Hox* genes are widely expressed in adult human body and colinearity is observed.

1.4.4.1 Hair growth and development

Hoxc13 is the first *Hox* gene found to be involved in hair development. *Hoxc13* is expressed only in the epidermal part of the follicle. *Hoxc13* mutant mice have normal hair follicles but the hair fibres are so brittle that they break off at the surface of the skin. *Hoxc13* was found to downregulate hair-specific keratins and keratin-associated proteins (KAPs) genes, most of which are located in the novel KAP gene cluster mapped to a region in mouse chromosome 16 that is syntenic to human chromosome 21q22.11 (Godwin and Capecchi, 1998; Tkatchenko et al., 2001). Electrophoretic mobility shift assays showed that HOXC13 binds exclusively to the proximal promoter regions via the TAAT and TTAT core motifs (Jave-Suarez et al., 2002).

Other *Hox* genes found to be active in hair follicles by *in situ* studies are *Hoxa4*, *Hoxa5*, *Hoxa7*, *Hoxb4*, *Hoxb7*, *Hoxc4*, *Hoxc8*, *Hoxc9*, *Hoxc10*, *Hoxc11*, *Hoxc12*, *Hoxc13*, *Hoxd9*, *Hoxd11* and *Hoxd13* (Awgulewitsch, 2003). In hair development, the temporal colinearity of gene activation is reversed. The 5' genes are expressed earlier than the 3' genes, e.g. *Hoxc12* is expressed later than *Hoxc13*. It is also expressed only in the epidermal part of the follicle but more distally than *Hoxc13* (Shang et al., 2002).

1.4.4.2 Mammary gland

Mammary gland development continues in adulthood. *Hox* gene expression in mammary gland is linked with the sex steroids. In response to estrogen, estrogen receptors α and β bind to consensus estrogen response element (ERE) on *HOXA10* which then regulates mammary gland differentiation (Martin et al., 2007) .

Paralogous *Hox9* genes (*Hoxa9*, *Hoxb9*, *Hoxc9* and *Hoxd9*) are necessary for the proper mammary gland development during pregnancy and childbirth. *Hoxa9*, *Hoxb9*, *Hoxd9* triple mutant females are unable to undergo mammary gland maturation in preparation for lactation (Chen and Capecchi, 1999). *Hoxc6* which is expressed during puberty and into adulthood is downregulated during pregnancy (Friedmann et al., 1994).

Hoxa5 is expressed only in the mammary stroma at specific stages of mammary gland development. *Hoxa5*^{-/-} female mice have lactation problems and this suggests its function in regulating the growth and differentiation of mammary alveolar epithelium (Garin et al., 2006).

1.4.4.3 Prostate

Hoxb13 is the only *Hox* gene that continues to be expressed during adulthood in mice. Its expression gradates from the highest in the ventral lobe to the lowest in the dorsal lobe. Its expression is androgen-independent so it is not affected by castration (McMullin et al., 2009).

1.4.4.4 Behaviour

Loss of function mutation of *Hoxb8* in the microglia has been implicated to be the cause of excessive grooming and hair removal. Transplantation of the wild type bone marrow into *Hoxb8* mutant mice rescues their grooming dysfunction but not their spinal cord defects. Deletion of *Hoxb8* in the hematopoietic system results in excessive grooming but does not result in spinal cord defects (Chen et al., 2010).

1.5 Hox enhancers

An enhancer is defined as a DNA element that transcription factors bind to increase the rate of gene transcription. It can be located thousands of base pairs away from the gene it

regulates. It can be found upstream, downstream or within the intron of the gene, and in any orientation.

Enhancer timing of *Hox* gene expression is important for *Hox* function. Deletion of the early enhancer of the murine *Hoxc8* delays but does not eliminate *Hoxc8* expression. However, this is sufficient to result in many axial skeletal defects as observed in *Hoxc8*-null mutation (Juan and Ruddle, 2003).

Due to the close proximity of the vertebrate *Hox* genes, enhancers found in the intergenic regions can be shared by adjacent genes, interacted selectively by certain promoters and competed for by different promoters. These three mechanisms were displayed in the analysis of the *Hoxb4-Hoxb6* region. **Enhancer sharing:** *Hoxb4* and *Hoxb5* share an r6/7 neural enhancer and an intergenic s7/8 somite enhancer. *Hoxb5* and *Hoxb6* share a limb enhancer that lies upstream of *Hoxb6*. Moreover, it has been shown that enhancer sharing occurs within the same cell. **Enhancer selectivity:** The limb enhancer in Region D, upstream of *Hoxb4*, interacts selectively with the *Hoxb4* promoter and not the *Hoxb5* promoter. The same applies to the neural enhancer in the same region. However, the neural enhancer has been found to interact with the *Hoxb5* promoter at a later stage of expression. The neural and mesodermal enhancers that lie upstream of *Hoxb6* act only on the *Hoxb6* promoter and not the *Hoxb5* promoter. **Promoter competition:** The neural enhancer in Region E, upstream of *Hoxb4*, is able to direct expression from both *Hoxb4* and *Hoxb5* promoters when they are tested individually. However, *Hoxb4* promoter out-competes *Hoxb5* promoter when they are tested together. Insertion of the region D balances out the competition and the neural enhancer is able to interact with *Hoxb5* promoter (Sharpe et al., 1998).

In *Drosophila*, *Hox* genes are further apart. A 255 bp promoter tethering element, located 5' of the Abd-B transcriptional start site, has been shown to tether the *infraabdominal* (*iab*) 5 enhancer to the Abd-B promoter in transgenic embryo assay. This enables the enhancer to bypass the insulators and to act on the promoter that is located at 55 kb away from the enhancer (Akbari et al., 2008; Ho et al., 2011).

Autoregulation and cross-regulation are also involved in *Hox* expression. *Hoxb1* r4 enhancer is the most extensively studied *Hox* responsive element. *Hoxb1* expression was

first initiated by retinoids via the 3' RARE (Langston and Gudas, 1992; Ogura and Evans, 1995). Expression in the r4 region is restricted by the negative regulation from the 5' RARE (Studer et al., 1994) and the binding of Hoxb3 (negative cross-regulation) to the S3 site of the *Hoxb1* locus (Wong et al., 2011). Multiple Hox/cofactor binding sites are found on this enhancer. Hoxb1 (autoregulation), Hoxa1 (para-regulation), members of the Pbx and Meis family of Hox cofactors, Sox and Oct bind cooperatively to the ARE during the initial activation of *Hoxb1*. At the later stage of *Hoxb1* expression, when *Hoxa1* expression is lost, Hoxb1 auto-regulates (Popperl et al., 1995; Studer et al., 1998; Di Rocco et al., 2001; Ferretti et al., 2005).

An enhanceosome refers to a functional unit that consists of an enhancer, its transcription factors and their interacting cofactors. High mobility group 1 (HMG-1) protein, an architectural DNA-binding protein, has been shown to interact with HOXB1, B3, C6, D3, D8, D9, D10 and D11 in the absence of DNA. In the presence of a 100 bp HOXD9 ARE found in the upstream promoter region, HMG-1 binds to the N terminus/helix 1 of the HOXD9 homeodomain and allows HOXD9 to bind faster to DNA. HOXD8 competes with HMG1 and represses HOXD9 activity (Zappavigna et al., 1996).

1.6 Hox regulation during embryonic development

Hox gene expression is regulated at the initiation, establishment and maintenance stages. During the initiation stage, transcription factors and RNA polymerase II bind in proximity of the *Hox* clusters to enable chromatin opening. During the establishment stage, *Hox* genes are sequentially activated and their expression domains are set. During the maintenance stage, *Hox* gene expression domains are maintained.

1.6.1 Initiation of *Hox* gene expression

Hox genes are first expressed far from the organizer or its equivalent during gastrulation. Initiation of *Hox* gene expression involves interactions between numerous factors and signaling pathways.

1.6.1.1 RA

RA, a derivative of Vitamin A, was shown to induce the 3'→5' sequential expression of *Hox* genes (Deschamps et al., 1999). The 3' *Hox* genes are more sensitive to RA than the 5' *Hox* genes. Excess RA has teratogenic effects in embryos. Embryos are posteriorized

and possess extra digits and limbs (Cusic and Dagg, 1985; Kessel and Gruss, 1991). When there is too little or no RA, rhombomeres are partially transformed or the posterior hindbrain and anterior spinal cord are completely lost (Papalopulu et al., 1991).

Free RA exists *in vivo* in two major forms, all-*trans* and 9-*cis*. All-*trans*-RA acts as ligand to retinoic acid receptors (RARs) and retinoid X receptors (RXRs) while 9-*cis*-RA can only act on RXRs. In mammals, three genes encode RARs (RAR α , - β and - γ) and three more encode RXRs (RXR α , - β and - γ). It has been reported that all-*trans*-RA binds with little affinity to RXR α and must be isomerized to 9-*cis*-RA to bind to RXR α with greater affinity (Levin et al., 1992).

RXR forms dimers with either RXR or RAR. Binding of all-*trans*-RA to RAR in RXR-RAR complex activates transcription and this activation is enhanced by the binding of 9-*cis*-RA to RXR (Minucci et al., 1997). RAR-RXR recognizes RAREs and RXR-RXR recognizes retinoid X response elements (RXREs).

The typical RARE is composed of two direct repeats (DRs) of the consensus sequence (PuGG/TTCA), separated by a 2 or 5 nucleotide spacer. RAREs have been described in *Hoxa1* (Langston and Gudas, 1992), *Hoxa4* (Packer et al., 1998), *Hoxb1* (Ogura and Evans, 1995), *Hoxb4* (Gould et al., 1998), *Hoxb5*, *Hoxb6*, *Hoxb8* (Oosterveen et al., 2003), *Hoxd4* (Popperl and Featherstone, 1993) and human *HOXA7* (Kim et al., 2002).

RAR/RXR heterodimers bind co-operatively to the RAREs and recruit coactivators bearing histone acetylase activity eg cAMP response element binding protein (CBP)/p300. In the absence of RA, corepressors are recruited instead (Bastien and Rochette-Egly, 2004).

It has been shown that the posterior RA-responsive genes are not exposed longer to RA (Maves and Kimmel, 2005; Hernandez et al., 2007). In Maves and Kimmel, 2005, they attributed the expression of posterior RA-responsive genes to the increasing RA source. The gene is expressed when the RA threshold is reached. However, in Hernandez et al., 2007, they attributed the expression to the “sink” (Cytochrome P450, family 26 (Cyp26) enzymes). They hypothesized that the dynamic expression of Cyp26 enzymes delimits the gene expression boundaries.

1.6.1.2 Fgf

Fgfs are heparin-binding proteins. The 5' *Hox* genes are more sensitive to Fgf than the 3' *Hox* genes. It has been implicated that fibroblast growth factor receptor-1 (FGFR1) plays a role in determining the expression domains of the *Hox* genes. Mutation of tyrosine 766 autophosphorylation site of FGFR1 to phenylalanine resulted in subtle expression changes in *HOXD4* and *HOXB5*. *HOXD4* expression was posteriorized by one somite while *HOXB5* expression did not extend as posteriorly as its wild type expression (Partanen et al., 1998).

Fgf also prevents RA-induced maturation and differentiation of cells in the young spinal cord. In this way, neurons mature when new somites are formed. Fgf signaling is enhanced by the TGF β family member, Growth differentiation factor 11 (Gdf11), at a later stage resulting in the induction of 5' *Hoxc* gene expression at thoracolumbar levels (Deschamps and van Nes, 2005).

1.6.1.3 Upstream regulators

In *Xenopus*, it has been found that the initial *Hox* expression domain is defined by the overlap of the mesodermal transcription factor brachyury (Xbra) and the secreted growth factor BMP-4. Xbra limits it in the animal-vegetal direction while BMP-4 limits it in the organizer/non-organizer direction (Wacker et al., 2004). Wingless-Int 8 (*Xenopus*) (XWnt8) directly initiates the expression of *hoxa1*, *hoxb1* and *hoxd1* via *Tcf/Lef* transcription factors in the canonical Wnt signaling pathway (In der Rieden et al., 2010).

1.6.1.4 Long intergenic noncoding RNA

A long intergenic noncoding RNA, transcribed from the 5' tip of *HOXA* locus, was identified to activate several 5' *HOXA* genes. This transcript, known as *HOTTIP*, was brought close to its target genes by chromosomal looping. Binding of *HOTTIP* RNA to tryptophan-aspartic acid repeat domain 5/ mixed-lineage leukemia (WDR5/MLL) complex promotes histone H3 lysine 4 trimethylation and ultimately gene transcription (**Fig. 1-10**) (Wang et al., 2011).

1.6.2 Establishment of *Hox* gene expression

Products of *caudal*-related (*Cdx*) genes from the *ParaHox* gene family bind to Cdx-binding sites within *Hox* clusters and directly regulate *Hox* genes. They are believed to

play a role in Wnt and Fgf signaling pathways due to the similarities in the posterior body truncations of *Cdx* mutants, loss-of-function *Wnt3a* and *Fgfr1* mutants (Deschamps and van Nes, 2005). It has been shown in that *Cdx1a* and *Cdx4* control the tissue responsiveness of Fgf and RA signals to prevent ectopic expression of posterior hindbrain and anterior spinal cord genes (Shimizu et al., 2006).

RA was shown in zebrafish to induce the expression of a member of the *maf* family transcription factor called *val* via the activation of *variant hepatocyte nuclear factor 1* (*vhnf1*). *vhnf1* cooperates with Fgf to activate *val* expression. Fgfs also induce and/or maintain the expression of *Krox20* (Marin and Charnay, 2000; Hernandez et al., 2004). *Kr* and *Krox20* work together to regulate *Hoxb3* enhancer activity and restricting its expression to r5, while the *Hoxa3* enhancer activity does not require *Krox20* (Manzanares et al., 2002).

Activating Protein-2 (AP-2) binding sites were found in *Hoxa2*. AP-2 was shown to negatively regulate *Hoxa2* expression, hence restricting its expression in the neural crest cells that subsequently give rise to the second branchial arch and the more posterior regions (Maconochie et al., 1999).

Figure 1-10: *HOTTIP* RNA binds to WDR5/MLL complex, and promotes histone H3 lysine 4 trimethylation and ultimately gene transcription (Wang et al., 2011).

1.6.2.1 The Maf proteins

The Maf proteins are a subset of the basic leucine zipper (bZIP) proteins. They are so named because of their structural similarity to the product of the *v-maf* oncogene, which is identified as the transduced transforming component of avian musculoaponeurotic fibrosarcoma virus, AS42 (Nishizawa et al., 1989). The bZIP structural motif consists of a basic DNA-binding domain and a leucine zipper dimerization domain.

Maf proteins are divided into two subgroups: the large Maf proteins (236-370 amino acids) and the small Maf proteins (149-162 amino acids). The large Maf proteins which include c-Maf, MafB and neural retina leucine zipper (NRL) possess an acidic transactivation domain that is not found in small Maf proteins which include MafK, MafF and MafG. Except NRL, which is expressed only in retina and neurons, the other Mafs are expressed in a wide variety of tissues (Blank and Andrews, 1997).

In particular, MafB is essential for r5 and r6 development. Its homologues were found in rats, mice, zebrafish and humans. They are known as Maf-1 in rats, Kr in mice, val in zebrafish and MafB in humans. It has been best studied in mice and zebrafish.

In mice, *kr* expression starts at 8.0 dpc with a band in the caudal hindbrain. This is before the rhombomere boundaries are established and this coincides with the refinement of *Hox* gene expression. At 8.5 dpc, its expression peaks in r5 and r6, with a sharp rostral border at r4/5 boundary and a more diffuse caudal border in the vicinity of r6/7. After 9.5 dpc, expression rapidly decreases first in r5 and subsequently in r6 (Cordes and Barsh, 1994).

The zebrafish *val* expression has a similar onset as mouse *kr* expression, at 10 hours post-fertilization (hpf) at the end of gastrulation. However, unlike *kr*, its expression persists throughout r5 and r6 until the 20 somite stage (19 hpf), which is equivalent to 9.75 dpc in mice (Moens et al., 1998).

kr/kr homozygous mutants frequently toss their heads and run in a circling manner. They are also deaf and cannot swim. This suggests inner ear defects. Morphological examination of the *kr* hindbrain reveals that the otic vesicle, found normally adjacent to r5 and 6, is displaced laterally and develops into a cyst-like structure rather than the vestibular apparatus and cochlea. Furthermore, the posterior hindbrain cadual to r1-r3

appears unusually smooth, forming a single large, unsegmented bulge. Similar hindbrain abnormality was found in *val*⁻ embryos (Moens et al., 1998).

Thus far, it was found that the *kr* mutation affects the expression of *Hoxa3*, *Hoxb1*, *Hoxb3*, *Hoxb4*, *Hoxd4*, *Krox-20*, fibroblast growth factor-3 (*fgf3*) and cellular retinoic acid binding protein I (*CRABP-I*). In homozygous *kr* null hindbrain, the expression boundary of *Hoxa3* at r4/5 appeared less sharp and expression was less intense in r5. The expression domain of *Hoxb1* was expanded, resulting in the posteriorization of the caudal boundary from r4/5 to r5/6. Expression boundaries of *Hoxb3* and *Hoxb4* were anteriorized from r4/5 and r6/7 to r3/4 and r5/6 respectively. *Hoxd4* expression was anteriorized by two rhombomeres. *Krox-20* expression in what should be r5 and *fgf3* expression in r5 and r6 are lost. *CRABP-I* expression faded in r4 instead of r6 (Frohman et al., 1993; McKay et al., 1994).

val mutation affects the expression of *hoxb1*, *hoxa2*, *hoxb2*, *hoxb3*, *hoxd3*, and *hoxb4* later in development. Expression of *hoxb1* and *hoxb4* expanded gradually across rX (between r4 and r7, where no boundaries are formed). The expression of *hoxb2* in r5 increased and the expression of *hoxa2* was not upregulated in r5. The anterior limits of *hoxb3* and *hoxd3* are less clear (Prince et al., 1998). Hence, these findings suggest a repressive role of MafB.

Studies have shown that Kr/Val is responsible for *Hoxa3* and *Hoxb3* expression in r5 and r6 early in vertebrate evolution, but later in evolution, it cooperates with Krox20 to activate *Hoxb3* enhancer, restricting its activity to r5 (Manzanares et al., 1999). Like *Hox* genes, *MafB/kr/val* is also subjected to environmental cues. Fgf signals from r4 cooperate with *vhnf1* to activate RA-induced *val* expression. Thus, *val* acts downstream of *vhnf1* to repress r4-like cell surface properties in r5 and r6 (Hernandez et al., 2004).

Besides functions in hindbrain, MafB is also involved in megakaryocyte differentiation (Sevinsky et al., 2004), central respiratory control (Blanchi et al., 2003), embryonic development of epidermis and hair follicles (Ogata et al., 2004), renal development (Moriguchi et al., 2006) and islet beta cell maturation (Artner et al., 2007).

1.6.3 Maintenance of *Hox* gene expression

Polycomb group (PcG) and trithorax group (trxG) proteins act by maintaining previously established transcriptional states of the *Hox* genes so as to maintain the positional identity of *Hox* genes along the antero-posterior axis. PcG protein complex is recruited by transcriptionally repressive histone, methylated at lysine 9 and 27, while trxG protein complex is recruited by transcriptionally active histone, acetylated at lysine 9 and dimethylated at lysine 4 (Deschamps and van Nes, 2005). Both proteins compete to bind to the Polycomb Response Element of the *Hox* genes. The PcG proteins prevent rostral *Hox* genes from expressing outside their normal expression domains while the trxG proteins ensure the correct *Hox* expression within their respective domains (Deschamps et al., 1999).

An additional tier of regulation is by the stalled RNA polymerase (pol) II which has been observed in *lab*, *Antp*, *Ubx* and *Abd-B* of early *Drosophila* embryo (Zeitlinger et al., 2007; Chopra et al., 2009). This implies that the gene silencing is not entirely due to the PcG proteins. It has been proposed that the paused Pol II prevents additional Pol II complexes from binding and it is important for the rapid gene induction and activation at later stages of embryogenesis.

Hox proteins can also auto/crossregulate *Hox* gene expression. HOXB4 and its paralogs, together with PBX, bind to the *Hox*-responsive element of *Hoxb4* late enhancer to maintain *Hoxb4* expression in the rostral-most domain. HOXB5 can also bind to the autoregulatory element (ARE) of *Hoxb4* to potentiate *Hoxb4* expression in their overlapping region (Deschamps et al., 1999).

Hox genes have been observed to be regulated post-transcriptionally by microRNAs (miRNAs). miRNAs are ~22-nucleotide non-coding RNAs that bind to partially complementary messenger RNAs (mRNAs) of target genes to downregulate their gene expression.

Two conserved miRNA loci, miR-10 and miR-196, are found at two positions in the *Hox* clusters, between paralogs IV and V and between paralogs IX and X respectively. Only the *HoxD* cluster lacks a miR-196 member (**Fig. 1-11**).

These genes are transcribed in the same orientation and observe spatial colinearity like the *Hox* genes. They are more highly expressed in the neural tube than in the trunk mesoderm. They are believed to fine-tune *Hox* gene expression. In fact, miR-196 was discovered to repress *Hoxb8* expression (Deschamps and van Nes, 2005; Yekta et al., 2008).

Hoxb4 was proposed to be regulated by miR10a because both miR10a and *Hoxb4* have similar expression patterns (Brend et al., 2003). *Hoxb4* transcripts were detected throughout the neural tube starting from the r6/7 boundary to the tail bud of the embryo but its protein was only detected in the posterior hindbrain. Hence, binding of miR10a to the 3' untranslated region (UTR) of *Hoxb4* destabilizes its transcripts in the posterior somatic domain, restricting its expression to somites 7 to 13.

Alternative polyadenylation was observed in the twelve *Drosophila* species with known genomic sequences, resulting in mRNA transcripts of different lengths in different set of tissues. As a result, new cis-regulatory sequences are added or RNA secondary structures are altered. Hence, their 3' UTR serve as very different target sites for miRNA (Patraquim et al., 2011).

Figure 1-11: Predicted miRNA-repression of *Hox* genes. Blue and green lines indicate repression by miR-10 and miR-196, respectively. Two conserved miRNA loci, miR-10 and miR-196, are found at two positions in the *Hox* clusters, between paralogs IV and V and between paralogs IX and X respectively. Only the *HoxD* cluster lacks a miR-196 member (Yekta et al., 2008).

1.7 Hox regulation during adulthood

Similar regulatory mechanisms of *Hox* gene expression observed during embryogenesis are also observed during adulthood. Endocrine regulation of *Hox* gene expression is characterized in the female reproductive tract. Unlike the embryonic situation, RA was not shown to regulate *Hox* gene expression during adulthood.

1.7.1 Endocrine regulation of Hox gene expression in female reproductive tract

HOXA10 is a mediator of endometrial functional differentiation and decidualization. Its expression is upregulated by Vitamin D metabolite (1, 25-dihydroxycholecalciferol), estradiol, progesterone and testosterone.

1, 25-dihydroxycholecalciferol, bound to Vitamin D receptor, binds to the Vitamin D regulatory element located at -385 to -343 bp upstream of the *HOXA10* transcription start site. Estradiol-bound estrogen receptors ER α and ER β bind to two putative EREs at 5' regulatory region of *HOXA10* (Daftary and Taylor, 2006).

1.8 Hox4 paralogs

Hoxd4 is the fourth paralog in the *HoxD* cluster. Paralog group IV genes are homologous to *Drosophila Dfd*. *Hoxa4*, *Hoxb4* and *Hoxd4* have similar expression domains in the developing hindbrain, delimiting rostrally at r6/7 (**Fig 1-12A**). However, *Hoxa4* expression delimits at the r6/7 boundary at later time points (after 11.5 dpc) than *Hoxb4* and *Hoxd4* (after 10.5 dpc). They are also expressed in the neural crest of the 4th branchial arch. *Hoxc4* is the most disparate member of the paralog group IV genes because its anterior boundary is more posterior than the others in the rostral spinal cord and neural crest.

Their axial limits in the somitic mesoderm are all slightly different. *Hoxb4* and *Hoxd4* have the most anterior limit at C1. However, *Hoxb4* expression at C1 is negligible. *Hoxc4* has the most posterior limit at C4 while *Hoxa4* extends to C2 (**Fig.1-12B**) (Maconochie et al., 1996) .

Hoxa4, *Hoxb4* and *Hoxd4* differ in abundance within the tissues and stages. *Hoxd4* become less abundant from 8.5 to 12.5 dpc in the mesodermal derivatives of the lung and stomach and the prevertebral column in mouse while *Hoxa4* and *Hoxb4* become more

abundant over the same period of time. At 12 dpc, *Hoxa4* and *Hoxb4* are found within mesodermal derivatives of the trachea, lung and gut while *Hoxd4* is found in the mesenchymal component of the gonad. *Hoxb4* and *Hoxd4* extend anteriorly to the thyroid duct but *Hoxa4* does not (Gaunt et al., 1989).

Hoxa4, *Hoxb4* and *Hoxd4* work cooperatively in embryonic patterning. Compound *Hoxa4*, *Hoxb4* and *Hoxd4* mutants show more complete homeotic transformations than single mutants. C2 through C5 are transformed to the C1 identity (Horan et al., 1995).

Figure 1-12A: Schematic drawing of paralog group IV gene expression in the developing hindbrain (Maconochie et al., 1996). *Hoxa4*, *Hoxb4* and *Hoxd4* have similar expression domains in the developing hindbrain, delimiting rostrally at r6/7. *Hoxc4* has the most posterior expression in the spinal cord. B: Schematic drawing of paralog group IV gene expression in the axial skeleton. *Hoxb4* and *Hoxd4* have the most anterior limit at C1. *Hoxa4* has anterior boundary at C2 and *Hoxc4* has the anterior boundary at C4. Blue shading indicates highest level of expression and grey shading indicates relatively lower level of expression. mb = midbrain, sc = spinal cord.

1.9 Transcriptional regulation of *Hox4* paralogs

1.9.1 *Hoxa4*

Hoxa4 expression is autoregulated and regulated by RA. Four potential *Hoxa4* binding sites are found in *Hoxa4* promoter and intron. Three potential RAREs are found upstream of the coding region at -2.9, -1.8 and -1.4 kb. The -2.9 RARE is identical to a *Hoxd4* RARE (Packer et al., 1998). It has been shown that Activating protein (AP)-2 binds near the RARE at -2.9 kb and interacts with it. Autoregulation of *Hoxa4* expression is required for the maintenance of the transient RA-induced expression (Doerksen et al., 1996; Packer et al., 1998).

1.9.2 *Hoxb4*

Hoxb4 expression is regulated by regulatory elements found in Regions A-E (Whiting et al., 1991; Sharpe et al., 1998; Brend et al., 2003) (**Fig. 1-13**), and ventralizing Bone Morphogenetic Protein (BMP)-4 signals in response to RA or somite signals. *Hoxb4* tends to be expressed more dorsally in the neural tube so BMP4 upregulates *Hoxb4* expression in the ventral neural tube to ensure even expression on both sides of the tube (Amirthalingam et al., 2009).

Region A consists of an early neural enhancer and a late neural enhancer. DR5 RARE is found in the early neural enhancer. The early neural enhancer is induced by somites and RA. RA induction on early neural enhancer is rapid and transient. The early neural enhancer initiates *Hoxb4* expression but is unable to maintain it. The maintenance of the expression is mediated by the late neural enhancer, previously known as CR 3 and activated one day later in mice. It is shared by both *Hoxb3* and *Hoxb4* in maintaining their expression in the posterior hindbrain. It is responsive to RA and precisely marks the r6/7 boundary of *Hoxb4* expression (Gould et al., 1997).

Region B, found in the 3' UTR, contains regulatory elements responsible for maintaining the anterior boundary in the paraxial mesoderm, established by the intronic enhancer found in Region C, throughout the embryonic development. Sequences within the *Hoxb4* promoter are also required in maintaining the expression established by Region C (Brend et al., 2003). Nuclear Factor Y (NFY) (transcriptional activator) and Yin Yang-1 (YY1) (transcriptional repressor) are found to bind to a short conserved region within Region C known as CR1 (Gilthorpe et al., 2002). CR1 is conserved between mouse and *Fugu Hoxb4*

(Aparicio et al., 1995) . Region D consists of a limb enhancer and a neural enhancer. The limb enhancer regulates only *Hoxb4* limb expression but the neural enhancer is shared by both *Hoxb4* and *Hoxb5*. It directs expression up to an s7/8 anterior boundary. The neural enhancer in region E is also shared by both *Hoxb5* and *Hoxb4* but *Hoxb4* and *Hoxb5* compete to interact with the neural enhancer (Sharpe et al., 1998).

1.9.3 *Hoxc4*

Transcriptional regulation of *Hoxc4* is not well-studied. *Hoxc4* has been observed to be involved in lymphocyte activation and proliferation. Its expression is upregulated by the binding of estrogen receptors to 3 evolutionarily conserved and cooperative EREs on *Hoxc4* promoter, and synergizing with CD154 or lipopolysaccharide and interleukin (IL)-4 signaling. This hence induces activation-induced cytosine deaminase induction, immunoglobulin class switch DNA recombination, and somatic hypermutation (Mai et al., 2010).

Figure 1-13: Schematic diagram showing *Hoxb4* regulatory regions. Region A is a 3'-flanking enhancer and consists of an early neural enhancer and a late neural enhancer. Region B is found in the 3' UTR. Region C consists of an intronic enhancer. CR1 in Region C is found to be necessary for the enhancer activity. Region D consists of a limb enhancer and a neural enhancer. Region E consists of a neural enhancer. Black box indicates homeobox. NE = neural enhancer

1.9.4 Hoxd4

Transcriptional regulation of human *HOXD4* and mouse *Hoxd4* is fairly well-studied. Transcription of human *HOXD4* is driven by two promoters, P1 and P2 (**Fig. 1-14**). Depending on the choice of the two transcription start sites and two polyadenylation sites, there are four transcripts of different lengths. These transcripts are variably expressed in different locations. Two transcripts of 5.4 and 2.8 kb, transcribed from P1, are expressed in spinal cord and limb buds. Two transcripts of 4.2 and 1.4 kb, transcribed from P2, are expressed in the hindbrain and spinal cord. 4.2 and 1.4 kb transcripts were shown to be induced by low concentrations of RA (10 to 100 nM) while 5.4 and 2.8 kb transcripts were induced by higher concentrations of RA (0.1 to 10 μ M). This reinforces the hypothesis that less RA is required to activate more anterior structures (Moroni et al., 1993).

Human *HOXD4* expression is regulated by four enhancers – one 5' neural enhancer, one 3' neural enhancer and two mesodermal 3' enhancers. The 5' neural enhancer is responsible for more posterior *HOXD4* expression in the developing hindbrain. The 3' neural enhancer is flanked by two mesodermal 3' enhancers and it is responsible for the delimitation of expression at r6/7 in the developing hindbrain. Of the two partially overlapping subfragments of the minimal 700 bp region of the 3' neural enhancer tested, the 3' 468 bp region was found to direct neural-restricted expression while the 5' 405 bp region was found to be required for the setting of the r6/7 boundary. Both neural enhancers contain a DR5 type RARE and are therefore responsive to RA. The mesoderm 3' enhancers are responsible for delimitation of expression at somites (s) 5/6 of the paraxial mesoderm. (Morrison et al., 1996; Morrison et al., 1997).

Murine *Hoxd4* expression is believed to be driven by two promoters – P2 and P3. P2 is 5.2 kb upstream of the coding region. P2 is active in a more posterior expression (Folberg et al., 1997). The location of the promoter active in the r6/7 expression, denoted by P3, is still unknown. P1 which was previously believed to be the promoter active in the r6/7 expression is not a true promoter (Phua et al., 2011).

Hoxd4 expression is regulated by a 5' ARE (Popperl and Featherstone, 1992), a 5' mesodermal enhancer, a 3' neural enhancer (Zhang et al., 1997) and a 3' mesodermal enhancer. The 5' mesodermal enhancer is found 5' to codon 43. The 3' neural enhancer

Figure 1-14: Diagram of human genomic *HOXD4* locus (adapted from (Moroni et al., 1993). Open boxes: coding region, black box: homeobox. *HOXD4* is driven by two promoters P1 and P2. $(A)^n$ is the poly adenylation site.

and 3' mesodermal enhancer are found downstream of exon 2 of *Hoxd4* (Morrison et al., 1997).

It was observed that HOXD4 and PBX1 were recruited to the ARE to positively regulate *Hoxd4* expression in differentiating P19 cells (Kobrossy et al., 2006). The 5' mesodermal enhancer sets an anterior boundary one somite posterior to wild type and needs to work with the 3' mesodermal enhancer to set the right anterior boundary in mesoderm. A RARE is found in the 5' mesodermal enhancer to make the chromatin accessible to other trans-acting factors (Zhang et al., 1997).

The *Hoxd4* 3' neural enhancer is located ~ 1.6 kb downstream of exon 2. It has been shown that its function is conserved in mouse and zebrafish. Sequence comparison between mouse and zebrafish *Hoxd4/hoxd4a* neural enhancers revealed eight blocks of homology. They are sites A-G and the RARE (**Fig. 1-15**) (Nolte et al., 2003). A comparison between *Hoxd4* and *hoxd4a* transcription units is shown in **Fig. 1-16**. Note that for completeness and to assist the reader in later sections of the thesis, this figure includes elements that were only brought to light through the work reported here. This includes the element designated as "5' neural enhancer".

Figure 1-15: Sequence comparison between mouse and zebrafish *Hoxd4/hoxd4a* 3' neural enhancers. Eight blocks of homology (sites A-G and RARE) are found.

Figure 1-16: Comparison between mouse *Hoxd4* and zebrafish *hoxd4a* transcription units (not drawn to scale). ARE: autoregulatory element; CR1/2/3: conserved region 1/2/3; ME = mesodermal enhancer and R = repressor element; miR-10b: microRNA-10b; NE = neural enhancer; P1/2/3: promoter 1/2/3; RARE: retinoic acid response element.

Sites D-F and the RARE were previously characterized in the Featherstone lab. They were found to be required for *Hoxd4* enhancer function. The RARE is composed of two DRs separated by a 5 nucleotide spacer. RAR and RXR heterodimerize and bind cooperatively to the RARE. The RARE is important for the initiation and maintenance of *Hoxd4* expression. Mutation of the RARE delayed the onset of *Hoxd4* expression from 8.5 dpc to 9.0 dpc in the mouse embryonic hindbrain. Neural expression was dramatically impaired, such that at 9.5 dpc, only a small patch on the dorsal side of the neural tube was observed (**Fig. 1-17**) (Zhang et al., 2000).

Simultaneous deletion of sites D-F abolished *Hoxd4* expression in the mouse embryonic hindbrain and individual mutations caused posteriorization of the anterior border (**Fig. 1-18**). Site F mutants showed the slightest posteriorization (**Fig. 1-18E and I**). Sites D and E mutants showed similar extents of posteriorization but site E mutants showed more pronounced posteriorization in the ventral aspect (**Fig. 1-18C, D, G and H**) (Zhang et al., 2000; Nolte et al., 2006).

The spacing between sites D and E, and between the RARE and site D is evolutionarily conserved, suggesting that transcription factors bound adjacently to these sites might undergo stereospecifically constrained interactions. Introduction of 5 bp between the RARE and site D, corresponding to one-half turn of the DNA helix, abolished enhancer activity (**Fig. 1-18K and P**). Expression was rescued by introducing 11 bp which completed a full DNA turn (**Fig. 1-18L and Q**). There was no effect on enhancer activity when 5 bp were inserted between site C and the RARE (**Fig. 1-18J and O**) and mutants were anteriorized when 5 bp were inserted between sites E and F (**Fig. 1-18N and S**). Mutants with 5 bp insertion between sites D and E were posteriorized just like the site D mutants (**Fig. 1-18M and R**). This suggests that retinoid receptors interact stereospecifically with the factors bound to at least site D. In fact, it was found that the Paired Box gene 6 (PAX6) paired domain binds to sites D and E (Nolte et al., 2006). PAX6 is important for setting the anterior border and strength of *Hoxd4* expression. In zebrafish, knockdown of both paralogs of pax6 (*pax6a* and *pax6b*) reduces and anteriorizes *hoxd4a* expression in the zebrafish hindbrain (**Fig. 1-19**). Since PAX6 is known to be involved in dorsal-ventral patterning, the above observation shows a possible interaction between antero-posterior patterning and dorso-ventral patterning (Nolte et al., 2006).

Figure 1-17: Transgenic mouse embryos with mutation of RARE. A: At 9.0 dpc, *Hoxd4* expression was first observed. B: At 9.5 dpc, only a small patch on the dorsal side of the neural tube was observed (Zhang et al., 2000).

Figure 1-18A: Cartoon of E9.5 mouse embryo. The boundaries of rhombomeres are indicated by dashed lines and *Hoxd4* expression is shown with solid fill. B-S: Top panels show lateral view of transgenic embryos and bottom panels show the flatmount view of transgenic embryos. These embryos correspond to the construct number and pictorial shown above each pair of photographs. The white solid line on the right side of embryo indicates the posterior edge of the otic vesicle. #2 and #3 show similar extents of posteriorization but #3 shows more pronounced posteriorization in the ventral aspect. #4 shows slight posteriorization of the anterior boundary. #5 shows no effect on enhancer activity when 5 bp are inserted before the RARE. The enhancer activity is abolished when 5 bp are inserted between RARE and site D (#6) and the activity is rescued when 11bp are inserted (#7). 5bp insertion between sites D and E posteriorizes the anterior boundary (#8). The anterior boundary is anteriorized when 5bp are inserted between sites E and F (#9) (Nolte et al., 2006).

Figure 1-19: Knockdown of *pax6a* and *pax6b* anteriorizes and reduces *hoxd4a* expression in the zebrafish hindbrain. A: Dorsal view of wild type 14-somite embryo probed for the expression of *engrailed (eng) 2b*, *krox20a* and *hoxd4a*. B: Dorsal view of *pax6a+b* 14-somite morphant probed for the expression of *eng2b*, *krox20a* and *hoxd4a* (Nolte et al., 2006) .

1.9.5 Comparison of the transcriptional regulation of group 4 paralogs

Studies have shown that the arrangement of *cis*-regulatory regions between *Hoxa4*, *Hoxb4* and *Hoxd4* is conserved (**Fig. 1-20**) and these genes are responsive to RA around 8.5-9.0 dpc in mouse. However, *Hoxb4* and *Hoxd4* are more closely-related to each other in the neural regulation and RA response than *Hoxa4*. Both neural enhancers involved in setting the r6/7 anterior boundary are located in the 3' flanking region and their r6/7 boundaries are set at 10.5 dpc. In response to RA at 8.5 dpc, their neural expression was shifted anteriorly by one rhombomere; and at 9.5 dpc, it shifted further up to r4. Though *Hoxb4* and *Hoxd4* neural enhancers are similar in function, they do not share any blocks of homology (Morrison et al., 1997).

On the other hand, the *Hoxa4* r6/7 enhancer is found in the 5' flanking region and its neural expression extends to r6/7 only after 11.5 dpc. Both the r6/7 enhancer and the 3' neural enhancer are responsive to RA. RA induction on the r6/7 enhancer promotes ectopic expression in the neural tube at 10 dpc and in the hindbrain at 10.5 dpc. There was no anterior shift in expression but the r6/7 expression was observed prematurely (Packer et al., 1998). RA induction on the 3' neural enhancer only promotes ectopic expression in the ventral neural tube at 10.5 dpc (Morrison et al., 1997).

Their mesodermal regulation is less studied because it is complex and variable between paralogs. It has been found that the intronic enhancer in *Hoxb4* acts similarly (Whiting et al., 1991) as the 5' and 3' mesodermal enhancers in *Hoxa4* (Behringer et al., 1993; Morrison et al., 1997) and the 3' mesodermal enhancers in *Hoxd4* (Morrison et al., 1996) in determining the anterior boundary of the somitic expression.

Figure 1-20: Alignment of *Hoxa4*, *Hoxb4* and *Hoxd4* and their regulatory regions (Morrison et al., 1997) . Yellow triangle= neural regulatory element. Exons are represented by blue rectangles. Not drawn to scale. Both *Hoxb4* and *Hoxd4* have their r6/7 neural enhancers in the 3' flanking region, which are responsive to RA. *Hoxa4* r6/7 neural enhancer is in the 5' flanking region and is also responsive to RA (Packer et al., 1998).

Rationale

Since the activity and position of the r6/7 neural enhancer are conserved in paralogous *Hoxb4* and *Hoxd4*, studying either enhancer in detail will reveal common and diverged regulatory mechanisms. Such studies are also likely to reveal mechanisms applicable to *Hox* gene regulation generally. Therefore, the Featherstone lab aims to understand how *Hox* expression is controlled through the study of the vertebrate *Hoxd4* 3' neural enhancer.

Sequence comparison between mouse and zebrafish *Hoxd4/hoxd4a* 3' neural enhancers revealed eight blocks of homology, denoted as sites A-G and RARE. Sites D, E and F and the RARE were previously characterized in the Featherstone lab (Zhang et al., 2000; Nolte et al., 2003; Nolte et al., 2006). In previous studies on sites D, E, F and RARE, the rabbit beta-globin minimal promoter (Nolte et al., 2003) and the entire *Hoxd4* transcription unit (Nolte et al., 2006) were used to drive the beta-galactosidase reporter gene in transgenic mouse embryos.

It has been proposed that *Hoxd4* transcription is driven by two promoters, P2 located 5.2 kb upstream from the coding region and P1 located 1.1 kb upstream from the coding region (Folberg et al., 1997). P1 was proposed to be responsible for the r6/7 expression in the developing hindbrain. Since zebrafish *hoxd4a* neural enhancer has been tested to be fully functional in mouse embryos and has been shown to direct expression to r6/7 boundary in the developing hindbrain, mouse *Hoxd4* neural enhancer and promoters can therefore be tested in zebrafish embryos (Nolte et al., 2003). Therefore we were interested in recapitulating *Hoxd4/hoxd4a* neural expression in zebrafish embryos using *Tol2* transposons and to use derivatives of these constructs for subsequent studies.

At the beginning of this work, sites A, B, C and G had not been tested for their contributions to enhancer function. Therefore, we were interested in testing them in both neurally differentiating P19 cells and zebrafish embryos. P19 cells are suitable for studying *Hoxd4* expression because they are pluripotent stem cells of murine teratocarcinoma that differentiate along a neural cell lineage upon RA induction and aggregation. Endogenous *Hoxd4* expression is activated only during the differentiation process. Zebrafish was chosen as the model organism because it is a vertebrate displaying

conserved *Hox* function and allows for rapid and cost-effective analysis of enhancer function.

Genetic evidence suggests that a member of the maf family transcription factor (Kr in the mouse and Val in zebrafish) is a negative regulator of murine *Hoxd4* and zebrafish *hoxb4a*, *respectively*, in the embryonic hindbrain. Such repression occurs in rhombomeres 5 and 6, thereby helping to set the r6/7 anterior expression border of these *Hox* genes. Sequence alignment shows that conserved sites F and G bear weak homology to the consensus maf recognition sequence (TGCTGANT/CCNGNN) (Yang and Cvekl, 2007) (**Fig. 1-21**). Therefore, it was of interest to determine whether *val* represses *hoxd4a* and whether such repression might act through the *hoxd4a* neural enhancer at the proposed sites.

Figure 1-21: Sequence alignment of conserved sites F and G of the 3' *Hoxd4* neural enhancer (top strand) with consensus maf recognition site (bottom strand). *Hoxd4* sites F and G are highlighted in purple.

Hypotheses

1. Evolutionarily conserved sites A-C and G are important in *Hoxd4* 3' neural enhancer function.
2. Val represses expression of *hoxd4a* in r5 and r6 by binding to sites F and G in the neural enhancer and interacting with PAX6 bound to sites D and E (Nolte et al., 2006).

Specific Aims

1. Establish transgenic reporters that recapitulate *Hoxd4/hoxd4a* expression in the hindbrain of zebrafish embryos. Transposons containing the *Hoxd4/hoxd4a* 3' neural enhancers, putative promoter regions and GFP reporter genes were injected into the zebrafish embryos at one-cell stage and the embryos were observed at 24 hpf for GFP expression.
2. Characterize conserved sites A-C and G of *Hoxd4/hoxd4a* 3' neural enhancer. Mutations were made to the conserved sites. Constructs containing the mutated *Hoxd4/hoxd4a* 3' neural enhancers, putative promoter regions and GFP reporter genes were tested in neurally differentiating P19 cells and zebrafish embryos.
3. Determine whether Kr/Val/MafB represses the hindbrain expression of *hoxd4a* and, if so, binds to sites F and G of the 3' neural enhancer. *val* is knocked down by injecting *val* morpholino in zebrafish embryos at 1-cell stage. Electrophoretic mobility shift assay was performed to find out whether MafB binds to the 3' neural enhancer.

Chapter 2 Materials and Methods

2.1 General chemicals and reagents:

Restriction enzymes and buffers used in cloning were obtained from New England Biolabs, Inc (Beverly, MA, USA) and Fermentas, Inc. (Hanover, MD, USA). Polymerase chain reaction (PCR) reagents, T4 DNA ligase (Cat. No: EL0011), DNA ladders (Cat. No: SM0321), PageRuler Plus Prestained Protein Ladder (Cat. No: SM1811) were from Fermentas, Inc. (Hanover, MD, USA). Gateway® BP Clonase® II (Cat no.: 11789-020) and Gateway® LR Clonase® II Plus (Cat. no.: 12538-120) enzymes were obtained from Invitrogen, Carlsbad, CA, USA. BP Clonase® II enzyme was used for the generation of entry clones, involving the cloning of DNA fragments into pDONR™ vectors, while LR Clonase® II Plus enzyme was used to catalyze the *in-vitro* recombination of the entry clones with the destination vector, pDESTTol2pA (<http://chien.neuro.utah.edu/tol2kitwiki/index.php?title=PDestTol2pA&printable=yes>).

2.2 Study of *Hoxd4/hoxd4a* neural enhancer in P19 cells

2.2.1 Plasmid construction

Mouse transgenic constructs: A 1457 bp fragment containing the 570 bp CL region of the mouse neural enhancer and -780 to +140 bp of *Hoxd4* P1 promoter (Kobrossy et al., 2006; Folberg et al., 1997; Zhang et al., 2000) was amplified with (sense) 5' TGATAAGCTTGATCCTTGGTGGTAGTAT 3' and (antisense) 5' ATTTACCGGTGTACAATTTACACAGGC 3' and cloned into HindIII and AgeI sites of the pAcGFP1-1 promoterless vector (a gift from Dr Li Hoi Yeung, NTU, Singapore). This plasmid is called construct 1. Mutations of the mouse neural enhancer were first checked with MatInspector (<http://www.genomatix.de/en/produkte/genomatix-software-suite.html>) for unintentional introduction of new transcription factor binding sites before generating by PCR overlap extension (Ho et al., 1989) using construct 1 as the DNA template. The PCR primers used to generate a mutation in site A were (sense) 5' CCCATCACAACTCTAGCTGAAAAGGTATTTTAC 3', (antisense) 5' GCTAGAGTTGTGATGGGCGTCAGAAAAAAG 3'; a mutation in site B (sense) 5' AAACAGTCGGCTGTGAAGTTCGGCTAAATTACAG 3', (antisense) 5' TCACAGCCGACTGTTTTCAGCTTCTCAACG 3'; a mutation in site C (sense) 5' TTTTACTCTGCCCGGGCTAAATTACAGAATCAG 3', (antisense)

5'GCCCCGGGCAGAGTAAAATACCTTTTCAGC 3'; a mutation in site G (sense)
 5'AGGCAGTTAGAGACAACCCTTAACCTTTGGG 3', (antisense) 5'
 GTTGTCTCTAACTGCCTCCAGGCTGAAATAC 3'. Compound mutation of sites A, B,
 C was generated using the plasmid containing the neural enhancer with mutation in site A
 (construct 1mutA) as the template for PCR overlap extension and primers (sense) 5'
 GAAAACAGTCGGCTGTCTGCCCCGGGCTAAATTACAGAATC 3' and (antisense) 5'
 AGCCCCGGGCAGACAGCCGACTGTTTTTCAGCTAGAGTTGTGA 3'.

This plasmid was then used as the template to generate compound mutation of sites A, B,
 C and G (construct 1mutABCG) using the same pair of primers for generating mutation in
 site G. Compound mutation of sites A-G except RARE (construct 1allmut) was generated
 using construct 1mutABCG as the template and primers (sense) 5'
 CAACCCGCGGATTTCGCCCCGCTGTACCGCGGCCTGGAGGCAGTTAG 3' and
 (antisense) 5' CCAGGCCGCGGTACAGCGGGGCGAATCCGCGGGTTGTCCTCTGG
 3'. Enhancers with i) mutated RARE and ii) insertion between sites E and F were
 subcloned from i) pSHCRmL (Zhang et al., 2000) and ii) pSH C-E5CF-L (Nolte et al.,
 2006) into pAcGFP1-1 and the final constructs are known as construct 1mutRARE and
 construct 1EFspace respectively. Plasmids were extracted using the QIAGEN Plasmid
 Midi Kit (Qiagen Inc, Valencia, CA, USA) and all clones were confirmed by sequencing.

Zebrafish transgenic constructs: An orthologous fragment of the *hoxd4a* enhancer and
 P1 region as the mouse counterpart were cloned into HindIII and AgeI sites of the
 pAcGFP1-1 promoterless vector. Plasmid was extracted using the QIAGEN Plasmid Midi
 Kit (Qiagen Inc, Valencia, CA, USA) and confirmed by sequencing.

2.2.2 Cell culture, transfection and differentiation:

Plasmids containing the mouse/zebrafish *Hoxd4/hoxd4a* neural enhancers and promoters
 were stably transfected into P19 mouse embryonic carcinoma cells by the calcium
 phosphate co-precipitation method as described in Sambrook and Russell, 2001. These
 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Cat. no: 12100-046,
 Gibco-Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (Cat. no: SH
 30109.03, Hyclone, Logan, UT, USA), 1 mM sodium pyruvate (Cat. no: 11360-070,
 Gibco-Invitrogen, Carlsbad, CA) and 1% penicillin-streptomycin (Cat. no: 15140-122,
 Gibco-Invitrogen, Carlsbad, CA). Cells were seeded at a density of 3×10^5 cells/10-cm
 diameter plate. Twenty-two hours later, the medium was changed and 2 h after that, the

cells were transfected with 15 µg of DNA containing 9 µg of plasmid of interest, 1 µg of a phosphoglycerate kinase 1 promoter-driven puromycin-resistance vector and 5 µg of pCAB-B17 (McBurney et al., 1994) as the carrier DNA. At 16 h post-transfection, the medium was removed and the cells were washed with 5 mM ethylene glycol tetraacetic acid (EGTA) in phosphate-buffered saline (PBS) and then with PBS. Beginning at 40 h post-transfection, cells were selected with 2 µg of puromycin/ml for at least 10 days. Then cells were aggregated in 10 cm bacterial petri plates at a density of 10^5 cells/ml and treated with 0.3 µM RA (Cat. no: R2625, Sigma Aldrich, St. Louis, MO) for 72 h, with subculturing after 48 h in fresh medium with RA. The fluorescence intensity was quantified using Becton-Dickinson (BD) Fluorescence Activated Cell Sorting (FACS) Calibur on day 3. The FACS results were analyzed with BD FACSDiva Software. Cell images were taken with a Carl Zeiss inverted microscope Axio Observer A1 using 10 x objective.

2.3 Recapitulation of *Hoxd4/hoxd4a* expression in zebrafish embryos

2.3.1 Plasmid construction

Mouse transgenic constructs: CL region of the mouse neural enhancer was amplified with 5' GGGGACAACCTTTGTATAGAAAAGTTGTGATCCTTGGTGGTAGTAT 3' (sense) and 5' GGGGACTGCTTTTTTTGTACAACTTGAGTCCAAAAGGTTAAGGGTT 3' (anti-sense) and cloned into pDONRTMP4-P1R (a gift from Dr Andrew Waskiewicz, University of Alberta, Canada) using Gateway® BP Clonase® II enzyme mix (Cat no.: 11789-020, Invitrogen, Carlsbad, CA, USA). P1 promoter was amplified with 5' GGGGACAAGTTTGTACAAAAAAGCAGGCTCGAACTTTTGTACTCTTC 3' (sense) and 5' GGGGACCACTTTGTACAAGAAAGCTGGGTTGTACAATTTACACCAGGC 3' (anti-sense) and cloned into pDONRTM221 (Invitrogen, Carlsbad, CA, USA) using Gateway® BP Clonase® II enzyme mix (Cat no.: 11789-020, Invitrogen, Carlsbad, CA, USA). These two plasmids are recombined with p3E-EGFPpA (<http://chien.neuro.utah.edu/tol2kitwiki/index.php/P3E-EGFPpA>) into pDESTTol2pA destination vector (<http://chien.neuro.utah.edu/tol2kitwiki/index.php?title=PDestTol2pA&printable=yes>) using LR Clonase® II Plus enzyme (Cat. no.: 12538-120, Invitrogen, Carlsbad, CA, USA) and the final construct was named construct 10.

P2-ATG promoter region was amplified with 5' TGAAAAGATGCAAGTTGAAAAGCC 3' (sense) and 5' TAATTTTCAGAAGGTAGAAAATACTAATTTTTC 3' (anti-sense) and cloned into pCR®8/GW/TOPO® (Invitrogen, Carlsbad, CA, USA). This plasmid was recombined with the plasmid containing the enhancer region, p3E-EGFPpA and pDESTTol2pA and the final construct was named construct 8.

P2-P1 promoter region was amplified with 5' TGAAAAGATGCAAGTTGAAAAGCC 3' (sense) and 5' TGTACAATTTTACCASGGCAAAG 3' (anti-sense) and cloned into pCR®8/GW/TOPO® (Invitrogen, Carlsbad, CA, USA). This plasmid was recombined with the plasmid containing the enhancer region, p3E-EGFPpA and pDESTTol2pA and the final construct was named construct 9.

P2 promoter region was amplified with 5' TGAAAAGATGCAAGTTGAAAAGCC 3' (sense) and 5' TAGAAAGATGCAAGTTGAAAAGCC 3' (anti-sense) and cloned into pCR®8/GW/TOPO® (Invitrogen, Carlsbad, CA, USA). This plasmid was recombined with the plasmid containing the enhancer region, p3E-EGFPpA and pDESTTol2pA and the final construct was named construct 11.

Zebrafish transgenic constructs: Zebrafish orthologous region of 570 bp mouse CL region was amplified with 5' GGGGACAACCTTTGTATAGAAAAGTTGAAGCTTTATGTTGTGAACG 3' (sense) and 5' GGGGACTGCTTTTTTTGTACAACTTGAAGCTTAAAAACCTCGAC 3' (antisense) and cloned into pDONRTMP4-P1R. P1 promoter was amplified with 5' GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGCGTCAGTGTTAATCA 3' (sense) and 5' GGGGACCACTTTGTACAAGAAAGCTGGGTATGACTAATTGATCATATA 3' (antisense) and cloned into pDONRTM221. These two plasmids are recombined with p3E-EGFPpA into pDESTTol2pA and the final construct was named construct 4.

P2-ATG promoter region was amplified with 5' GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTGAGTGCAAGTGAATGG 3' (sense) and 5' TAATTTTCAGAAGGTAGAAAATACTAATTTTTC 3' (anti-sense) and cloned into pDONRTM221. This plasmid (known as pDONRzfP2-ATG) was recombined with the plasmid containing the 570 bp enhancer region, p3E-EGFPpA and pDESTTol2pA and the final construct was named construct 2.

P2-P1 promoter was amplified with 5' TCTGAGTGCAGTGAATGGGAATTG 3' (sense) and 5' ATGACTAATTGATCATATATAAATATTTTACAG 3' (antisense) and cloned into pCR®8/GW/TOPO® (Invitrogen, Carlsbad, CA, USA). This plasmid was recombined with the plasmid containing the 570 bp enhancer region, p3E-EGFPpA and pDESTTol2pA and the final construct was named construct 3.

P2 promoter region was amplified with 5' TCTGAGTGCAGTGAATGGGAATTG 3' (sense) and 5' ACCCAAGCTTACCTCCAGAAG 3' (anti-sense) and cloned into pCR®8/GW/TOPO® (Invitrogen, Carlsbad, CA, USA). This plasmid was recombined with the plasmid containing the 570 bp enhancer region, p3E-EGFPpA and pDESTTol2pA and the final construct was named construct 5.

PostP1-ATG promoter region was amplified with 5' GGGGACAAGTTTGTACAAAAAAGCAGGCTGTGCTTTTAATTCTTTAC 3' (sense) and 5' TAATTTTCAGAAGGTAGAAAATACTAATTTTTC 3' (anti-sense) and cloned into pDONRTM221. This plasmid (known as pDONRzfpostP1-ATG) was recombined with the plasmid containing the 570 bp enhancer region, p3E-EGFPpA and pDESTTol2pA and the final construct was named construct 6.

PostP1-200ATG promoter region was amplified with 5' GAGATAATGTGAAGTTTGTAAAAATCACTACTAGG 3' (sense) and 5' TAATTTTCAGAAGGTAGAAAATACTAATTTTTC 3' (anti-sense) and cloned in pCR®8/GW/TOPO®. This plasmid (known as pTOPOzfpostP1-200ATG) was recombined with the plasmid containing the 570 bp enhancer region, p3E-EGFPpA and pDESTTol2pA and the final construct was named construct 7.

To generate a 750 bp zebrafish enhancer region, 5' GGGGACAACCTTTGTATAGAAAAGTTGTGCCTGAAGTTCTTGAAAAG 3' (sense) and 5' GGGGACTGCTTTTTTTGTACAACTTGAACTTACACGCCATGGCAG 3' (anti-sense) were used in PCR amplification. This enhancer fragment was cloned into pDONRTMP4-P1R and the resulting plasmid was recombined with pTOPOzfpostP1-200ATG, p3EGFPpA and pDESTTol2pA. The final construct was named construct 7NE750bp.

For the negative controls, in place of the enhancer fragment, 5'GGGGACAACCTTTGTATAGAAAAGTTGTGCTTTTACAGATGC 3' (sense) and 5' GGGGACTGCTTTTTTTGTACAACTTGGTGCAATTGTTTTGTC 3' (anti-sense) were used to amplify 460 bp of the luciferase gene from a pXP2 derivative vector (AF093682). The fragment was cloned into pDONRTMP4-P1R. The construct (pDONR5'luc) was recombined with p3E-EGFPpA, pDONRzfp2-ATG/pDONRzfp2postP1-ATG/pTOPOzfp2postP1-200ATG and pDESTTol2pA. The final constructs are known as construct 2ΔNE, construct 6ΔNE and construct 7ΔNE respectively depending on the promoter region.

Deletion mutants: Three conserved regions in the postP1-ATG region were identified using VISTA browser (<http://pipeline.lbl.gov/cgi-bin/gateway2>) and NCBI BLAST (Basic Local Alignment Search Tool) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Since they are conserved, it is highly possible that one of them harbours an evolutionarily conserved r6/7 enhancer. These regions were known as CR1, CR2 and CR3. CR1 and CR3 were chosen for analysis. To generate a CR1 deletion, 5' CGGCACCTGGCAACACAGATG 3' (sense) and 5' TAATTTTCAGAAGGTAGAAAATACTAATTTTTC 3' (anti-sense) were used to amplify the region from construct 6 and the fragment was cloned into pCR®8/GW/TOPO®. The resulting construct was known as pTOPOzfp2postP1-ATGΔCR1. It was then recombined with pDONR5'luc, p3E-EGFPpA and pDESTTol2pA. The final construct was named construct 6ΔCR1ΔNE. As for CR3 deletion, the zfp2postP1-ATGΔCR3 region was first checked with MatInspector (<http://www.genomatix.de/en/produkte/genomatix-software-suite.html>) for unintentional introduction of new transcription factor binding sites. The downstream region from CR3 was amplified using 5' ACTTGG ATCCTTTACAAAAAATGTTATACACTC 3' (sense) and 5' TAATTTTCAGAAGGTAGAAAATACTAATTTTTC 3' (anti-sense). To facilitate better annealing during the PCR overlap extension in the generation of the final postP1-ATGΔCR3 region, it was further amplified with 5' ATGAGAAGCGCGGATCCTTTAC 3' (sense) and 5' TAATTTTCAGAAGGTAGAAAATACTAATTTTTC 3' (anti-sense). The upstream region was amplified with 5' GTGCTTTTAATTCTTTACAACCAAG 3' (sense) and 5' GTAAAGGAT CCGCGCTTCTCATCC 3' (anti-sense). The two fragments were mixed and amplified with only the primers at the far ends, 5' GTGCTTTTAATTCTTTACAACCAAG 3' (sense) and 5' TAATTTTCAGAAGGTAGAAAATACTAATTTTTC 3' (anti-sense). This final fragment

was cloned into pCR®8/GW/TOPO®. The resulting construct was known as pTOPOzfpostP1-ATGΔCR3. It was then recombined with pDONR5'Luc, p3E-EGFPpA and pDESTTol2pA. The final construct was named construct 6ΔCR3ΔNE.

All plasmids were isolated from DH5alpha/ TOP10 competent cells with QIAGEN Plasmid Midi Kit (Qiagen Inc, Valencia, CA, USA). All pDESTzfP2-ATG (construct 2) and construct 6 derivative constructs were checked by restriction digestion with XhoI and BglIII and sequencing. All construct 7 derivative constructs were checked by sequencing.

2.3.2 Injection into zebrafish embryos

Both mouse and zebrafish constructs (20 ng/μl) were injected at the one-cell stage with 70 ng/μl Tol2 corresponding mRNA transposase. Tol2 corresponding transposase mRNA was synthesized by cleaving with NotI and transcribing with Sp6 polymerase using Ambion mMessage Machine kit (Austin, Texas, United States). Images of embryos were taken with Carl Zeiss stereomicroscope Luminar V12 using an 80 x objective. The fluorescent images were taken at an exposure time of 8 s and all brightfield and fluorescent images were enhanced to a resolution of 300 dots per inch (dpi) with Adobe Photoshop.

2.4 Study of *hoxd4a* 3' neural enhancer in zebrafish embryos

2.4.1 Plasmid construction

Zebrafish constructs: Mutations of the conserved sites within the 570 bp 3' neural enhancer were first checked with MatInspector (<http://www.genomatix.de/en/produkte/genomatix-software-suite.html>) for unintentional introduction of new transcription factor binding sites before generating by PCR overlap extension. Fragments containing the mutated conserved sites in various derivatives of construct 2 were cloned into pDONRTMP4-P1R. These enhancer constructs were then recombined with pDONRzfP2-ATG, p3EGFPpA and pDESTTol2pA. In the case of construct 7mutRARE and construct 7mutABCG, the constructs containing the mutated enhancer fragment were recombined with pTOPOzfpostP1-200ATG instead of pDONRzfP2-ATG.

The mutation in site A of the enhancer fragment in construct 2mutA was generated using 5' CATCTAGCGACTCAGCATGAAAGGGTATTTTTACAG 3' (sense) and 5'

TGCTGAGTCGCTAGATGCTTCAGTGTGAAG 3' (antisense) with the wild type enhancer fragment as the template.

The mutation in site B of the enhancer fragment in construct 2mutB was generated using 5' GCAACTGTCCCATAGAATACAGAAGTTCAACAAAAGC 3' (sense) and 5' GTATTCTATGGGACAGTTGCTCTCAACGTGAATG 3' (antisense) with the wild type enhancer fragment as the template.

The mutation in site C of the enhancer fragment in construct 2mutC was generated using 5' ACACTTCAAGAACAAAAGCCTTGCGGGC 3' (sense) and 5' AGGCTTTTGTTCCTTGAAGTGTAATAATACCCTTTCATGCTC 3' (antisense) with the wild type enhancer fragment as the template.

The mutation in site G of the enhancer fragment in construct 2mutG was generated using 5' AGCTACTCTGTAGACACTCGTGACCCGATG 3' (sense) and 5' AGTGTCTACAGAGTAGCTCGCGGCTGAAAAG 3' (antisense) with the wild type enhancer fragment as the template.

The mutation in sites A and B of the enhancer fragment in construct 2mutAB was generated using 5' ATCTAGCGACTCAGCAACTGTCCCATAG 3' (sense) and 5' TGCTGAGTCGCTAGATGCTTCAGTGTGAAG 3' (antisense) with the enhancer fragment mutated in site B as the template.

The mutation in sites A and C of the enhancer fragment in construct 2mutAC was generated using 5' ACACTTCAAGAACAAAAGCCTTGCGGGC 3' (sense) and 5' AGGCTTTTGTTCCTTGAAGTGTAATAATACCCTTTCATGCTC 3' (anti-sense) with the enhancer fragment mutated in site A as the template.

The mutation in sites B and C of the enhancer fragment in construct 2mutBC was generated using 5' ACACTTCAAGAACAAAAGCCTTGCGGGC 3' (sense) and 5' CTTTTGTTCCTTGAAGTGTAATTCTATGGGACAGTTGC 3' (antisense) with the enhancer fragment mutated in site B as the template.

The mutation in sites A, B and C of the enhancer fragment in construct 2mutABC was generated using 5' ATCTAGCGACTCAGCAACTGTCCCATAG 3' (sense) and 5' CTTTGTCTTGAAGTGTATTCTATGGGACAGTTGC 3' (antisense) with the enhancer fragment mutated in sites B and C as the template.

The mutation in sites A, B, C and G of the enhancer fragment in construct 2mutABCG was generated with 5' AGCTACTCTGTAGACACTCGTGACCCGATG 3' (sense) and 5' AGTGTCTACAGAGTAGCTCGCGGCTGAAAAG 3' (anti-sense) using the enhancer fragment mutated in sites A, B and C as the template.

The mutation in RARE of the enhancer fragment in construct 2mutRARE was generated using 5' GCATCGCGATCCAGGGTACCCAGTTTCTGTTGCATTAC 3' (sense) and 5' CTGGGTACCCTGGATCGCGATGCCCGCAAGGCTTTTG 3' (antisense) with the wild type enhancer fragment as the template.

All plasmids were isolated from DH5alpha/ TOP10 competent cells with QIAGEN Plasmid Midi Kit (Qiagen Inc, Valencia, CA, USA). All construct 2 derivative constructs were checked by restriction digestion with XhoI and BglII and sequencing. All construct 7 derivative constructs were checked by sequencing.

2.4.2 Injection into zebrafish embryos

Zebrafish constructs (20 ng/μl) were injected at the one-cell stage with 70 ng/μl Tol2 corresponding mRNA transposase. Tol2 corresponding transposase mRNA was synthesized by cleaving with NotI and transcribing with Sp6 polymerase using Ambion mMessage Machine kit (Austin, Texas, United States). Images of embryos were taken with Carl Zeiss stereomicroscope Luminar V12 using an 80 x objective. The fluorescent images were taken at an exposure time of 8 s and all brightfield and fluorescent images were enhanced to a resolution of 300 dots per inch (dpi) with Adobe Photoshop.

2.5 The role of Kr/Val in regulating *Hoxd4/hoxd4a* neural expression

2.5.1 *val* knockdown

***val* morpholinos:** Morpholino oligos are short chains of about 25 nucleotide subunits. Each subunit is comprised of a nucleic acid base, a morpholine ring and a non-ionic phosphorodiamidate intersubunit linkage. Translational-blocking morpholino binds

complementary to the mRNA and sterically blocks the progression of the ribosomal initiation complex from the 5' cap to the start codon. The control morpholino (CCTCTTACCTCAGTTACAATTTATA) targets a human beta-globin intron mutation that causes beta-thalassemia and has not been reported to generate phenotypes in any other known test system.

Translation-blocking AUG-targeting *val* morpholino (GCGCTCATCGTGCCCTGTTCTTCA) was designed, titrated and injected into each embryo at the 1-cell stage. 4 ng - 5 ng of morpholino per embryo was found to be enough to knock down *val* expression without causing any deformity and lethality. To verify that the knockdown was specific, a 5'UTR-targeting *val* morpholino (TTGGGAAAACCAGACTTTGGCGTTT) was used independently. All morpholinos were tagged with 3' fluorescein to monitor successful injection.

To test whether Val binds to the neural enhancer(s), 4 ng of translation-blocking AUG-targeting untagged *val* morpholino was co-injected with 20 ng/μl construct 2 and 70 ng/μl transposase in zebrafish embryos at one-cell stage.

2.5.2 In situ hybridization of zebrafish embryos:

Embryos were fixed in 4% paraformaldehyde/PBS overnight at 4°C. Then they were washed 5 times every 5 min at room temperature in PBST (PBS/0.1% Tween-20). After the second wash, they were dechorionated. The 8-somite, 10-somite and 14-somite stages embryos were treated with 10 μg/ml Proteinase K for 1 min while the 18-somite stage and 24 hpf embryos were treated for 3 min. They were refixed in 4% paraformaldehyde/10X PBS for 20 min at room temperature, rinsed 5 x 5 min in PBST and prehybridized in 500 μl hybridization (hyb) mix with tRNA for 1 h at 65 °C. Then the hyb mix was removed and the warmed hyb mix with *hoxd4a* probe (1:200) was added to the embryos. In the case of double in-situ hybridization with *krox20* and *hoxd4a* probes, 43.75 ng of *hoxd4a* probe and 6.25 ng of *krox20* probe were used in 200 μl of hyb mix. They were hybridized at 65°C overnight. On the next day, the hyb mix was removed and the embryos were washed in the following solutions:

1. 66% hyb mix (no tRNA), 33% 2x saline-sodium citrate (SSC) for 5 min at 65°C
2. 33% hyb mix (no tRNA), 66% 2x SSC for 5 min at 65 °C

3. 2x SSC for 5 min at 65 °C
4. 1 x 20 min in 0.2x SSC + 0.1% Tween-20 at 65 °C
5. 2 x 20 min in 0.1x SSC + 0.1% Tween-20 at 65 °C
6. 66% 0.2x SSC, 33% PBST for 5 min at room temperature
7. 33% 0.2x SSC, 66% PBST for 5 min at room temperature
8. PBST for 5 min at room temperature

The wash was removed and the embryos were incubated in blocking solution (PBST plus 2% sheep serum, 2 mg/ml bovine serum albumin) for 1 h at room temperature and then in 1:5000 anti-digoxigenin-AP antibody: blocking solution (Cat. no: 1 1093274 910, Roche Diagnostics Cooperation, Indianapolis, IN) for 2 h at room temperature/ overnight at 4 °C. The antibody: blocking solution was removed and the embryos were washed 5 x 15 min in PBST. The embryos were then washed 4 x 5 min in colouration buffer and incubated in 500 µl of nitro-blue tetrazolium: colouration buffer mix (5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt (BCIP) Cat. no: 1383221 and Nitro blue tetrazolium chloride (NBT), Cat no: 11383213, Roche Diagnostics Cooperation, Indianapolis, IN) in the dark for 1 h at 37°C. Finally the reaction was stopped by washing 2 x in stop solution (PBST pH5.5). Images were taken with Carl Zeiss stereomicroscope Luminar V12 using 60 x objective and enhanced with Adobe Photoshop.

2.5.3 Confirmation of the efficacy of ATG-targetting *val* morpholino

The coding region of *val* gene was amplified from zebrafish cDNA with 5' ATGAGCGCCGATCTCGCC 3' (sense) and 5' TCACATGAAGAACTCTGGGGAGG 3' (anti-sense) and cloned into pCR4TOPO (Invitrogen, Carlsbad, CA, USA). The plasmid is known as pTOPOval(full)T3. Its protein was synthesized with TNT®T7/T3 Coupled Reticulocyte Lysate System (Promega, Heidelberg, Germany, Cat no: L5010) using T3 polymerase. One µg of pTOPOval(full)T3 was co-incubated with 1 µg of *val* translation-blocking morpholino in the reaction. pET15bMEIS1A was used as a negative control and MEIS1A protein was synthesized with TNT®T7 Quick Coupled Transcription/Translation System, Trial Size (Promega, Heidelberg, Germany, Cat no: 1171) and 1 µg of plasmid was co-incubated with 1 µg of *val* translation-blocking morpholino in T7-driven reaction. Eight µl of each TNT reaction was assayed by Western blot. Val protein was detected with Rabbit Anti-MafB Polyclonal Antibody, Unconjugated (Cat. no: ab56242, Abcam,

Cambridge, MA, USA) and goat anti-rabbit horseradish peroxidase. MEIS1A protein was detected with 1: 5000 mouse Anti-Meis 1 / 2 / 3 monoclonal antibody, Unconjugated, Clone 9.2.7 (Cat. no: 05-779, Millipore, Bedford, MA, UK) and 1:10 000 mouse secondary antibody conjugated with horseradish peroxidase.

2.5.4 Cloning, expression and purification of MafB

Human *MafB* was amplified from pCMVTag2CWTmafB containing 1005 bp Flag-tagged human *MafB* coding region (NM_005461) (gift of Dr Natalie Ahn, University of Colorado, USA) with 5' CAA CAG CTG CCC ACT AGC 3' (sense) and 5' TCA CAG AAA GAA CTC GGG AG 3' (antisense) and cloned into pCR®8/GW/TOPO® Entry vector (Invitrogen, Carlsbad, CA, USA). Then a LR recombination reaction with LR clonase II Plus enzyme (Invitrogen, Carlsbad, CA, USA, Cat. no.12538-120) was performed between this plasmid and pDEST15 (obtained from Dr Eugene Makeyev, NTU, Singapore). The final construct named pDESTmafBDN was transformed into BL21 cells. A colony was picked and inoculated in Luria-Bertani broth with 100 µg/ml ampicillin overnight. The next day, an inoculum was added to 6 equivalent flasks of 500 ml Luria-Bertani broth with 100 µg/ml ampicillin at a starting optical density (O.D.) of less than 0.01. At OD ~ 0.6, the cells were induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown at 37 °C for 4 h. The culture was pelleted and snap-frozen in liquid nitrogen. 1.28g of pellet was solubilized with a buffer containing 50 mM Tris and 400mM sodium chloride at pH 7.5, 1 mM phenylmethanesulfonylfluoride and 0.8 mM dithiothreitol. The solution was sonicated using pulse (2 s on and 1 s off) for 10 min at 30% amplitude and centrifuged at 10, 000 x g for 35 min at 4 °C. The supernatant was filtered with a 0.45 µm syringe filter. The filtrate was purified using the glutathione-S-transferase (GST) purification module (GE Healthcare Bio-sciences, Piscataway, NJ, USA, 27-4570-01). The concentration of MafB was quantitated by Bradford assay (Bio-Rad, Hercules, California). The purified MafB protein was loaded onto a 15% denaturing polyacrylamide gel which was stained with Comassie blue or transferred to a polyvinylidene fluoride (PVDF) membrane for Western blotting. The uninduced and untransformed BL21 cells were also grown for 4 h after reaching an OD ~0.6 at 37 °C.

i) Coomassie blue staining

The gel was stained with 0.1% Coomassie Brilliant Blue G-250 (AppliChem, Darmstadt, Germany), 40% methanol and 10% acetic acid for 10 min with gentle shaking at 25 rpm on a Labnet GyroTwister Shaker and quickly rinsed with destaining solution 1 (10% acetic acid, 25% isopropanol). Then it was destained overnight in destaining solution 2 (10% acetic acid, 7.5% methanol).

ii) Western blot

Proteins were boiled at 100 °C for 8 min, vortexed, quick-spun and cooled on ice before loading for SDS-PAGE (Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis). The gel was pre-run at 60 V for 10 min and run at 150 V for 70 min. Subsequently, it was washed for 10 min in cold transfer buffer before transfer onto polyvinylidene fluoride (PVDF) membrane (GE Healthcare Bio-sciences, Piscataway, NJ, USA) at 100 V for 1 h. The membrane was rinsed in Tris-Buffered Saline Tween-20 (TBST) for 5 min and incubated overnight in 5% skimmed milk at 4 °C. The membrane was then incubated for 1 h in 1: 5000 rabbit anti-GST (a gift from Dr Peter Cheung, NTU, Singapore): 5% skimmed milk at room temperature, washed 6 times for 5 minutes each with TBST, incubated for 1 h in 1:10,000 goat anti-rabbit horseradish peroxidase: 5% skimmed milk at room temperature. Finally the blot was washed 4 times for 10 minutes each in TBST. Equal amounts of luminol reagent and peroxidase reagent (Cat. No: WBKLS0500 Millipore, Bedford, MA, UK) were mixed and incubated for 1 min on the blot.

2.5.5 Electrophoretic mobility shift assay analysis:

The sequences of oligonucleotides used for EMSA (one strand) were: i) probe containing murine *Hoxd4* sites F and G 5' GCTGTATTTTCAGCCTGGAGGACTGACCTCTAAACCCT 3' and ii) consensus MafB probe 5' TAAGATTGCTGACTCAGCATTTGGGCTTTCC 3'. They were annealed to their 3' biotin-labeled complementary strand with 1 x Oligo Annealing Buffer (10 mM Tris, 1 mM EDTA, 50 mM NaCl, pH 8.0). The reaction mixture was annealed with a thermocycler (95 °C for 5 min, 70 cycles of -1 °C/cycle for 1 min and hold at 4 °C). The interactions of MafB protein and control BL21 bacterial proteins with the probes were analyzed with LightShift Chemiluminescent EMSA Kit (Pierce, Rockford, IL, USA) and blotted on positively charged Biodyne® B Nylon Membrane (Cat no.: 77016) (Pierce,

Rockford, IL, USA) according to instructions in kit. The binding reaction consisted of 0.25 µg protein, 50 ng/µl poly (dI.dC), 2.5% glycerol, 0.05% NP-40, 5 mM magnesium chloride and 20 fmol biotin-labelled probe. In the case involving cold competitor probe, 4 pmol unlabelled probe was used.

Chapter 3: Study of *Hoxd4/hoxd4a* neural enhancer in P19 cells

Sites D, E and F and the RARE of the murine *Hoxd4* 3' neural enhancer were previously characterized in the Featherstone lab (Zhang et al., 2000; Nolte, 2002; Nolte et al., 2006). Sites A, B, C and G have not been tested for their contributions to enhancer function. A preliminary study of the *Hoxd4/hoxd4a* neural enhancer was performed in P19 cells and three trials were conducted. P19 cells are pluripotent stem cells derived from a murine teratocarcinoma that differentiate along a neural cell lineage upon RA induction and aggregation. Endogenous *Hoxd4* expression is activated only during the differentiation process. The *Hoxd4* P1 presumptive promoter region was proposed to work in conjunction with the 3' neural enhancer to give r6/7 expression in the developing hindbrain, and I therefore used the P1 region to supply promoter activity in the constructs described below (Folberg et al., 1997).

To test the contributions of sites A, B, C, and G to enhancer function, these sites were mutated singly or in combination. Mouse enhancer derivatives and the P1 region (-1125 to -1928 bp) (Folberg et al., 1997), cloned into the pAcGFP1-1 promoterless GFP vector (**Fig. 3-1**), were stably transfected into undifferentiated P19 cells. Cells were then selected with 2 μ g/ml of puromycin for at least 10 days. Colonies that survived the puromycin selection were assumed to have taken up the transfected plasmid of interest, puromycin-resistance vector and pCAB-B17, which promotes stability of integration in the cells (McBurney et al., 1994). These cells were then aggregated and treated with 0.3 μ M RA for 72 h in bacterial Petri dishes. The presence of GFP expression indicated functional enhancer activity. Since enhancers have been shown to increase the probability of turning on the transcriptional active state and not the level of expression per cell (Walters et al., 1995), the effects of mutations were compared with the wild type neural enhancer based on the number of GFP-expressing cells.

The promoterless pAcGFP1-1 was used as the negative control in this study. It served as an appropriate control because it gave no expression. A green fluorescent protein (GFP)-expressing plasmid, pXJ-GFP (Zhao et al., 2000), was used as a positive control. Deletion of sites D-G (Zhang et al., 2000) was shown to abolish *Hoxd4* enhancer activity in transgenic mouse embryos (c.f. construct 1mutDEF). Insertion of 5 bp between sites E and

Figure 3-1: Schematic diagram of constructs tested. Each of the elements of the murine *Hoxd4* 3' neural enhancer contained within the smaller 570 bp fragment are shown in relation to the P1 presumptive promoter region used to drive GFP expression. Mutations are indicated by rose shading and an "X". Construct names are given at the right. Construct 1EFspace bears an insertion of 5 cytosine residues between sites E and F.

F (Nolte et al., 2006) was shown to have very intense anteriorized *LacZ* expression in transgenic mouse embryos (c.f. construct 1EFspace). Observations in the mouse embryos were tested for correlations in terms of expression in differentiating P19 cells (construct 1mutDEF and construct 1EFspace).

Hoxd4 is not expressed in undifferentiated P19 cells and its expression is induced as early as day 1 of RA treatment and peaks between days 3 and 4 (Rastegar et al., 2004). This has been independently confirmed in the Featherstone lab at NTU (Phua et al., 2011). Therefore, it is robust enough to perform analysis of the effects of the mutations on *Hoxd4* neural enhancer on day 3.

GFP expression was observed in cells transfected with all constructs except construct 1mutRARE at varying expression levels (**Fig. 3-2**). FACS, a specialized type of flow cytometry that sorts cells based upon the specific light scattering and fluorescent characteristics of each cell, was used to quantitate GFP expression. One fluorescent event refers to one fluorescing cell. Hence, the greater the probability that the transcriptional state is turned on, the greater the number of fluorescent events.

A sampling of 10,000 cells was analyzed for each construct. To minimize technical errors affecting the flow cytometer readings, cells were analyzed thrice in the first trial and six times in the second and third trials (**Tables 3-1, 3-2 and 3-3**). To minimize the difference in the extent of RA differentiation, a master mix of RA culture medium was used in the same trial.

Viable cells were gated based on the physical characteristics of the untransfected P19 cells. In general, viable cells have higher forward scatter (FSC) and lower side scatter (SSC) than dead cells. FSC measures the cell size while SSC measures cell granularity. To eliminate background fluorescence, all autofluorescing events on FL1-H (height) vs FSC-H plot were excluded from the analysis. FL1 is the channel for green fluorescence. For consistency, the same gates for SSC-H vs FSC-H plot and FL1-H vs FSC-H plot were used for transfected P19 cells.

Figure 3-2: Images of transfected cells on day 3 taken with 10 x objective. **A:** construct 1 (containing wild type mouse *Hoxd4* neural enhancer). **B:** construct 1mutA. **C:** construct 1mutB. **D:** construct 1mutC. **E:** construct 1mutG. **F:** construct 1mutABC. **G:** construct 1mutABCG. **H:** construct 1mutDEF. **I:** construct 1allmut. **J:** construct 1EFspace.

	No of fluorescent events			
	1 st reading	2 nd reading	3 rd reading	Average
WT	64	65	85	71
mutA	17	9	10	12
mutB	176	172	154	167
mutC	43	48	30	40
mutG	208	226	255	230
mutABC	142	176	177	165
mutABCG	12	10	13	12
mutDEF	33	34	23	30
RARE	0	0	0	0
Allmut	154	167	180	167
EFspace	1	0	0	0

Table 3-1: Number of fluorescent events for all constructs in Trial 1. WT: construct 1.

	No of fluorescent events						
	1st	2nd	3rd	4th	5th	6th	Average
WT	306	272	299	280	309	281	291
mutA	669	596	644	628	634	623	632
mutB	505	402	474	476	461	450	475
mutC	567	565	536	548	523	589	555
mutG	398	371	360	391	396	394	385
mutABC	558	566	556	540	531	558	552
mutABCG	52	38	53	55	37	41	46
mutDEF	307	281	282	300	307	321	300
RARE	0	0	0	0	0	0	0
Allmut	413	387	337	345	377	369	371
EFspace	2	5	4	2	7	3	4

Table 3-2: Number of fluorescent events for all constructs in Trial 2. WT: construct 1.

	No of fluorescent events						
	1st	2nd	3rd	4th	5th	6th	Average
WT	32	14	28	29	16	13	22
mutA	140	124	94	79	113	90	107
mutB	106	127	103	109	103	86	106
mutC	73	82	79	61	75	82	75
mutG	45	31	25	37	39	34	35
mutABC	182	184	162	178	145	157	168
mutABCG	11	3	10	7	4	9	7
mutDEF	7	3	6	6	6	6	6
RARE	0	0	0	0	0	0	0
Allmut	86	100	115	109	91	90	99
EFspace	0	0	0	0	0	0	0

Table 3-3: Number of fluorescent events for all constructs in Trial 3. WT: construct 1.

The number of fluorescent events in construct 1 is compared pairwise with the number of fluorescent events for individual mutants. The two-tailed paired student's t-test was used to determine the probability that the two populations have the same mean, with the assumption that the cells differentiated well and to the same extent. The significance level was set at 5%. If the p-value is less than 0.05, the two populations are considered to be significantly different and the observed upregulation or downregulation due to the mutated sites is implied to be significant as well (**Tables 3-4 and 3-5**). Data in **Table 3-5** are plotted into a graph as shown in **Fig. 3-3**.

Individual mutations in sites A and C (constructs 1mutA and 1mutC) show upregulation of the enhancer in two trials and downregulation of the enhancer in one trial. Compound mutation in sites D, E and F (construct 1mutDEF) shows downregulation of the enhancer in two trials and no change in one trial. Therefore, for these three mutations, I am unable to draw any conclusion. Enhancers with individual mutations in sites B and G, compound mutation in sites A, B and C (constructs 1mutB, 1mutG and 1mutABC) and mutations in sites A-G (construct 1allmut) were statistically found to be stronger than wild type mouse *Hoxd4* neural enhancer (construct 1). Enhancers with compound mutations in sites A, B, C and G (construct 1mutABCG) and 5 bp insertion between sites E and F (construct 1EFspace) were statistically found to be weaker than wild type mouse *Hoxd4* neural enhancer. It is interesting to note that the impaired activity of construct 1EFspace in P19 cells appears to contrast with the strong anteriorized expression observed for the same mutation in transgenic mouse embryos (Nolte et al., 2006).

A construct containing the homologous zebrafish sequences corresponding to the *Hoxd4* enhancer and P1 regions was also tested in neurally differentiating P19 cells. No GFP expression was observed. This may imply some species differences. In fact, later in this thesis I will show that the contribution of zebrafish *hoxd4a* regulatory elements to neural expression is very different to that observed in mouse.

	Trial 1	Trial 2	Trial 3
	WT		
mutB	0.02	$< 10^{-5}$	$< 10^{-4}$
mutG	$< 10^{-2}$	$< 10^{-3}$	0.02
mutABC	0.01	$< 10^{-5}$	$< 10^{-5}$
mutABCG	0.01	$< 10^{-6}$	$< 10^{-2}$
RARE	0.01	$< 10^{-6}$	$< 10^{-2}$
Allmut	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-4}$
EFspace	0.01	$< 10^{-6}$	$< 10^{-2}$

Table 3-4 shows the p-value of the comparison between the wild type enhancer and the mutant enhancers using two-tailed paired student's t-test. Values are all rounded off to 2 decimal places. WT: construct 1.

	Trial 1	Trial 2	Trial 3	Average	SD
mutB	2.4	1.6	4.8	2.9	1.7
mutG	3.2	1.3	1.6	2.0	1.0
mutABC	2.3	1.9	7.6	4.0	3.2
mutABCG	0.2	0.2	0.3	0.2	0.1
RARE	0.0	0.0	0.0	0.0	0.0
Allmut	2.3	1.3	4.5	2.7	1.6
EFspace	~ 0.0	~ 0.0	0.0	~ 0.0	~ 0.0

Table 3-5: Fold difference in the number of fluorescent events of the mutants in comparison with wild type (construct 1). Values are rounded off to 1 decimal place. SD = standard deviation.

Figure 3-3: The number of GFP-positive cells was determined for differentiating P19 cells (day 3) stably transfected with mutant constructs. The plotted values were obtained by dividing the number of GFP-positive obtained under the same conditions with the wild type construct. Values are the average of three trials. Error bars give standard deviation.

Discussion

The P19 cell line originates from a murine teratocarcinoma so it may not be a true representative of the mouse embryo (McBurney, 1993). Nonetheless, results from this study may provide preliminary indications of how mutations of different conserved sites may affect the *Hoxd4/hoxd4a* expression in zebrafish embryos.

An increase in expression in cells transfected with construct 1allmut, implies that only the RARE is an important positive regulatory element for *Hoxd4* neural enhancer activity in the context of P19 cells. This derives from the undetectable expression level in cells transfected with construct 1RAREmut.

In mouse embryos, RA treatment results in anteriorized expression of posterior *Hox* genes and posteriorized body plan, meaning the anterior structures are greatly reduced or missing and supplanted by more posterior structures. Hence, RA-treated P19 cells may correspond to cells found in more posterior structures of the embryo. In support of this suggestion, previous studies have shown that RA-treated P19 cells express miR-10b, a microRNA that is embedded in primary *Hoxd4* P2 transcripts and expressed considerably posterior to r6/7 (Phua et al., 2011) and many posterior *Hox* genes (Reese and Ramos-Valle, 2007). Therefore, the number of fluorescent events observed in this experiment may correlate to the likelihood of the gene expression in the posterior structures of the zebrafish embryos (**Chapter 4**).

If RA-treated P19 cells correspond to posterior embryonic central nervous system, the increased expression following single mutation of sites B and G (construct 1mutB and construct 1 mutG) and compound mutations of sites A-C (construct 1mutABC) and sites A-G (construct 1allmut) may be due to a release of constraints on “posterior” expression. It has been shown that insertion of 5 bp between sites E and F (Nolte et al., 2006) results in very intense anteriorized *LacZ* expression in transgenic mouse embryos but a decreased expression, implying “posterior” expression, in cells transfected with construct 1EFspace was observed in this study. This may be because only 920 bp of the *Hoxd4* P1 region is present in these constructs, while a much longer promoter region was used in the constructs tested in mouse embryos. Therefore, it is likely that the full set of transcriptional machinery was unable to bind to the short promoter region to reproduce the

observations in mouse embryos. More trials should be carried out to draw more confident conclusions and a longer promoter region should be used.

Nevertheless, we have to exercise extra caution in interpreting these results because a “posterior” environment seems to be invoked in these RA-treated P19 cells and the “posterior” character of these cells is influenced by the length of exposure time to RA (Rastegar et al., 2004; Phua et al., 2011). Therefore, more reliable results will be obtained from *in vivo* study by studying the mouse enhancer in mouse embryos and the zebrafish enhancer in zebrafish embryos.

Chapter 4: Recapitulation of *Hoxd4/hoxd4a* expression in zebrafish embryos

4.1 Examination of *hoxd4a* expression pattern by *in situ* hybridization

To independently confirm the *hoxd4a* expression pattern, and specifically to verify the position of its anterior boundary at different developmental stages, zebrafish embryos were hybridized *in situ* with a *hoxd4a* probe at 8-somite, 10-somite, 14-somite and 18-somite stages, and 24 hpf (**Fig. 4-1**). A *krox20* probe was also used at 24 hpf (1 day post-fertilization (dpf)). At 24 hpf, *hoxd4a* is expressed from the r6/7 boundary of the posterior hindbrain to the posterior spinal cord, in the pectoral fin buds and in the neural crest-derived pharyngeal arch mesoderm beneath and posterior to the otic vesicle (**Fig. 4-2**). It has also been observed to be expressed in the epibranchial placode (ventral to the otic vesicle and dorsal to the posterior pharyngeal pouches) (Punnamoottil et al., 2008).

Figure 4-1: *In situ* hybridization with *hoxd4a* probe. A: 8-somite stage. B: 10-somite stage. C: 14-somite stage. D: 18-somite stage. E: 24 hpf. Arrowhead indicates *hoxd4a* anterior expression boundary.

Figure 4-2: *In situ* hybridization of 24 hpf embryos with *hoxd4a* and *krox20* probe. *krox20* is expressed in r3 and r5. The otic vesicle (not shown) lies adjacent to r5. Embryos were imaged at 80x magnification. A: lateral view (image was enhanced using Adobe Photoshop). B: dorsal view. C: deyolked.

4.2 Recapitulation of zebrafish *hoxd4a* expression in transgenic zebrafish embryos

The neural enhancer was cloned into a 5' donor vector (pDONRP4-P1R), the promoter regions were cloned into a middle donor vector (pDONR221) and an EGFP reporter gene was cloned into a 3' donor vector (pDONRP2R-P3). These three vectors were then recombined with pDESTTol2pA destination vector (**Fig. 4-3**). Each transposon was tested at least twice to ensure consistency of results. All transposons mentioned here and in later chapters were cloned using MultiSite Gateway Technology (Invitrogen, Carlsbad, CA, USA).

Transposons containing the *hoxd4a* 3' neural enhancer regions and putative promoter regions homologous to mouse promoter regions (Folberg et al., 1997) were coinjected with transposase into zebrafish embryos at 1-cell stage (**Figs. 4-4 and 4-5**). Derivatives made in the background of construct X are named "construct Xn", where "X" is the construct number and "n" is a description of the derivative. The 570 bp *hoxd4a* 3' neural enhancer has been shown to exhibit neural enhancer activity in transgenic mouse embryos identical to the murine counterpart including an ability to set the r6/7 boundary in the hindbrain (Nolte et al., 2003).

To orientate readers to the anatomy of zebrafish embryos, close-up bright field images of rhombomeres in the developing hindbrain, with labels, at 18 hpf (Moens and Prince, 2002), 24 hpf and 48 hpf are shown in **Figs. 4-6A-C** respectively. The otic vesicle lies next to r5 and at 24 hpf, the distance between the r6/7 boundary and the inner posterior limit of the otic vesicle is about the length of the inner diameter of the otic vesicle. At 48 hpf, the r6/7 boundary is slightly anterior to the posterior end of the otic vesicle. Rhombomeres disappear at later stages. Rhombomeres are not visible at 72 hpf (**Fig. 4-6D**). Images of 24 hpf, 48 hpf and 72 hpf zebrafish hindbrain were taken with a Zeiss Observer A1 inverted microscope using a 32 x objective.

Figure 4-3: Cloning scheme of Gateway Technology. The neural enhancer and promoter region were PCR amplified with the addition of *att* sites and recombined with BP clonase into *pDONRP4-P1R* and *pDONR221* respectively. These entry clones were then recombined with *p3E-EGFP* (3' entry clone with *EGFP* reporter gene) in the presence of LR clonase into *pDESTTol2pA* destination vector.

Figure 4-4: Schematic representation of the putative promoter regions (not drawn to scale). The homeobox is given by the black rectangle. The exons are coloured in pink. NE = neural enhancer. The 3' neural enhancer contains conserved sites A-G and the RARE. P3 promoter lies in the postP1-ATG region.

Figure 4-5: Schematic diagram of zebrafish constructs tested (not drawn to scale). NE = neural enhancer. The size of 3'NE in all constructs except construct 7NE750bp is 570 bp. The size of 3'NE in construct 7NE750bp is 750 bp.

Figure 4-6: Lateral view of zebrafish hindbrain. Anterior is to the left and posterior is to the right. ov = otic vesicle. Except for A, all images were taken with 32 x objective from Carl Zeiss Axio Observer A1 inverted microscope. A: At 18 hpf (Moens and Prince, 2002). The otic vesicle is next to r5. The series of bulges are the rhombomeres along the anterior-posterior axis of the hindbrain. B: At 24 hpf. C: At 48 hpf. D: At 72 hpf.

4.2.1 Recapitulation of *hoxd4a* expression at 24 hpf (1 dpf)

It has been proposed that *Hoxd4* transcription is driven by two promoters, P2 located 5.2 kb upstream from the coding region and P1 located 1.1 kb upstream from the coding region (**Fig. 4-4**). P1 was proposed to be active in an anterior domain reaching the *Hoxd4* r6/7 expression border (Folberg et al., 1997), and since the mouse P1 region was able to drive GFP expression in pAcGFP1-1 derivative constructs in differentiating P19 cells, as described in **Chapter 3**, I was interested to find out whether the zebrafish homologous P1 region is responsible for driving *hoxd4a* expression. Hence, the zebrafish homologous P1 region was cloned together with the EGFP reporter gene and *hoxd4a* 3' neural enhancer into pDESTTol2pA (**Fig. 4-5**, construct 4) and the transposon was injected into zebrafish embryos at the 1-cell stage. No expression was observed in any injected embryos.

P2 was proposed to drive a more posterior expression in the central nervous system (Folberg et al., 1997). Since the zebrafish P1 homologous region was unable to drive *hoxd4a* r6/7 expression, I was interested to determine whether the zebrafish region homologous to the murine P2 region could have replaced the P1 region to drive *hoxd4a* r6/7 expression. On average, 10.6% of injected embryos with construct 5 (**Fig. 4-5**) expressed GFP. As it is difficult to analyze all the injected live embryos at 24 hpf, 13 embryos were randomly selected, photographed and analyzed. In all 13 embryos examined, the zebrafish P2 region was associated with neural expression that was more posterior than the endogenous *hoxd4a* gene (**Figs. 4-7C and 4-8C**).

Since constructs containing P2 or P1 were unable to establish r6/7 expression, a longer promoter region of ~ 5 kb was tested. Because an unknown downstream promoter may be required for r6/7 expression of *hoxd4a*, I designated this promoter as P3. This ~ 5 kb 5' region spans a contiguous stretch of genomic DNA and consists of the P2 promoter, P1 region and a presumptive P3 promoter region that ends at the nucleotide immediately preceding the *hoxd4a* ATG start codon (**Fig. 4-5**). This construct is known as construct 2. Using this construct, 68.7 % of injected embryos expressed GFP. Twelve embryos were randomly selected, photographed and analyzed at 24 hpf. 11/12 expressing embryos recapitulated *hoxd4a* expression by delimiting at the r6/7 boundary (**Figs. 4-7A and 4-8A**), while the expression of the remaining embryo appeared slightly posteriorized. Unlike the endogenous gene, this expression did not extend fully posteriorly along the spinal cord.

Severely crippled expression was observed with a construct containing the P2-P1 promoter region but lacking -1 to -2103 (construct 3) (**Fig. 4-5**). No expression was observed in all 133 embryos in a first trial and 35/36 embryos injected with construct 3 in a second trial. Only a single embryo from the second trial had spotty expression (**Figs. 4-7B and 4-8B**). These results indicate that expression of the *hoxd4a* transgene with a correct r6/7 boundary in the hindbrain is dependent on sequences located between -1 and -2103 relative to the ATG. I suggest that this region, designated PostP1-ATG, harbours an unknown P3 promoter, and could also include neural enhancer elements.

To determine whether the PostP1-ATG region can direct r6/7 expression independently of other 5' regions, embryos were injected with construct 6 lacking P1 and P2 regions (**Fig. 4-5**). On average, 81.9 % of injected embryos expressed GFP. Eleven embryos were randomly selected, photographed and analyzed at 24 hpf. 10/11 embryos recapitulated *hoxd4a* expression by delimiting at r6/7 boundary (**Figs. 4-7D and 4-8D**) while the expression of 1/11 embryos appeared slightly posteriorized. Just like embryos injected with construct 2, this expression did not extend fully posteriorly along the spinal cord. These results demonstrate that the PostP1-ATG region harbours r6/7 expression independently of the P1 and P2 regions.

Figure 4-7: Lateral view of embryos injected with constructs 2, 3, 5 and 6, observed at 24 hpf. White arrowhead points to r6/7 boundary. Red arrowhead points to the GFP expression limit. A: Embryo injected with construct 2. It shows r6/7 expression. B: Embryo injected with construct 3. It shows spotty expression. C: Embryo injected with construct 5. It shows posteriorized expression. D: Embryo injected with construct 6. It shows r6/7 expression.

Figure 4-8: Dorsal view of embryos injected with constructs 2, 3, 5 and 6, observed at 24 hpf. Red arrowhead points to the GFP expression limit. A: Embryo injected with construct 2. It shows r6/7 expression. B: Embryo injected with construct 3. It shows spotty expression. C: Embryo injected with construct 5. It shows posteriorized expression. D: Embryo injected with construct 6. It shows r6/7 expression. Note: Embryos shown in Figures 4-7C and 4-8C are different.

All the constructs described above carry the 3' neural enhancer, which I expected to be critical for r6/7 activity for reasons previously mentioned. To verify that the r6/7 expression observed was due to the 570 bp enhancer, it was removed from constructs 2 and 6 and we expected the r6/7 expression to be abolished in its absence. To our surprise, GFP expression in embryos injected with constructs 2 Δ NE and 6 Δ NE (**Fig. 4-5**) was not abolished. On average, 49.1 % of embryos injected with construct 2 Δ NE expressed GFP. Twelve embryos were randomly selected, photographed and analyzed at 24 hpf. 11/12 embryos recapitulated *hoxd4a* expression by delimiting at r6/7 boundary (**Figs. 4-9A and 4-10A**) while the expression of 1/12 embryos appeared posteriorized. On average, 40.3 % of embryos injected with construct 6 Δ NE expressed GFP. Eleven embryos were randomly selected, photographed and analyzed at 24 hpf. 8/11 embryos recapitulated *hoxd4a* expression by delimiting at r6/7 boundary (**Figs. 4-9B and 4-10B**) while the expression of 3/11 embryos appeared anteriorized. In both cases, r6/7 expression did not extend fully posteriorly along the spinal cord. The presence of r6/7 expression in the absence of the 570 bp 3' *hoxd4a* neural enhancer suggests the presence of a 5' *hoxd4a* neural enhancer upstream of *hoxd4a* coding exons capable to setting the correct r6/7 boundary. Such activity 5' to a vertebrate 4th group paralogs has not been previously described.

Figure 4-9: Lateral view of embryos injected with constructs 2 Δ NE and 6 Δ NE, observed at 24 hpf. White arrow points to r6/7 boundary. A: Embryo injected with construct 2 Δ NE. It shows r6/7 expression. B: Embryo injected with construct 6 Δ NE. It shows r6/7 expression.

Figure 4-10: Dorsal view of embryos injected with constructs 2ΔNE and 6ΔNE, observed at 24 hpf. A: Embryo injected with construct 2ΔNE. It shows r6/7 expression. B: Embryo injected with construct 6ΔNE. It shows r6/7 expression.

Sequence comparison of the first 500 bp upstream of zebrafish *hoxd4a* ATG start site with mouse *Hoxd4* region with NCBI blast shows homology from -112 relative to the ATG. This suggests that a P3 promoter and/or other regulatory elements required for r6/7 expression may lie in this region. To test this, a 200 bp region upstream of zebrafish *hoxd4a* ATG start codon was cloned into pDESTTol2pA destination vector with EGFP reporter gene. This construct was designated construct 7 Δ NE (**Fig. 4-5**). None of the injected embryos expressed GFP in three independent trials. Insertion of the 570 bp neural enhancer, construct 7 (**Fig. 4-5**) restored partial neural expression. On average, 38.1 % of the injected embryos expressed GFP. Of these, 17 embryos were randomly selected, photographed and analyzed at 24 hpf. All of them had faint expression that neither demarcated clearly at r6/7 boundary nor extended fully posteriorly (**Figs. 4-11A and 4-12A**). As their expression was weak, the images were taken with a more sensitive monochrome camera instead of a colour camera with the same exposure time (8 s) as the other images shown elsewhere in the results section. The fluorescent expression shown in white was given a green colouration.

Next, we posed the question whether it is possible that the 570 bp region of the 3' *hoxd4a* neural enhancer might lack important flanking elements required to fully recapitulate *hoxd4a* expression in transgenic zebrafish embryos, since a 700-bp *Hoxd4* 3' neural enhancer has been previously shown to set the r6/7 expression boundary in transgenic mouse embryos (Morrison et al., 1997; Zhang et al., 2000) . To clarify this doubt, a comparable 750 bp region of the 3' *hoxd4a* neural enhancer was cloned into the pDESTTol2pA destination vector with the EGFP reporter gene and the 200 bp promoter region. This construct is known as construct 7NE750bp (**Fig. 4-5**). On average, 50.4 % of embryos expressed GFP. 18 embryos were randomly selected, photographed and analyzed at 24 hpf. These embryos (**Figs. 4-11B and 4-12B**) had hindbrain expression fainter than those injected with construct 7 and which did not demarcate clearly at r6/7. In general, it is interesting to note that the expression was stronger in the tail regions in embryos injected with construct 7NE750bp than those injected with construct 7 (**Figs. 4-13A and 4-13B**).

Figure 4-11: Lateral view of embryos injected with constructs 7 and 7NE750bp, observed at 24 hpf. White arrow points to r6/7 boundary A: Embryo injected with construct 7. It shows faint hindbrain expression that does not demarcate clearly at r6/7 boundary. B: Embryo injected with construct 7NE750bp. It shows hindbrain expression that also does not demarcate clearly at r6/7 boundary.

Figure 4-12: Dorsal view of embryos injected with constructs 7 and 7NE750bp, observed at 24 hpf. A: Embryo injected with construct 7. It shows faint hindbrain expression that does not demarcate clearly at r6/7 boundary. B: Embryo injected with construct 7NE750bp. It shows hindbrain expression that also does not demarcate clearly at r6/7 boundary.

Figure 4-13: Tail regions of embryos injected with constructs 7 and 7NE750bp, observed at 24 hpf. Images were taken with a 58 x objective. A: Tail region of embryo injected with construct 7. Note: This embryo is different from the embryo shown in Figures 4-11A and 4-12A. B: Tail region of embryo injected with construct 7NE750bp. It shows stronger expression than that of embryo injected with construct 7.

Table 4-1 summarizes the expression patterns described above and average expression level is rated from 1 (lowest) to 5 (highest). These scores for the average hindbrain expression level serve as a rough guide as to how strongly GFP was expressed in different contexts and they are not meant to be quantitative. Unpaired t-test was performed to compare the difference in activity of these constructs based on the percentage of expressing embryos, with the significance level set at $p < 0.05$. The enhancer strength is correlated to the percentage of expressing embryos relative to injected embryos. Each experiment was repeated at least thrice and only percentages that were consistent in at least three trials were computed in the t-test. This is to reduce the possibility that the percentage was influenced by integration efficiency.

The probability that construct 2 has stronger activity than 2 Δ NE by chance is 0.02 (< 0.05) (**Table 4-2**). The probability that construct 6 has stronger activity than 6 Δ NE by chance is less than 0.01 (< 0.05) (**Table 4-3**). These suggest that the 3' neural enhancer plays a complementary role to the 5' neural enhancer in *hoxd4a* expression. This is supported by the hindbrain expression observed when only the 3' neural enhancer was cloned with the 200 bp promoter region (construct 7). The probability that construct 6 has stronger activity than construct 2 by chance is 0.01 (< 0.05) (**Table 4-4**). This suggests the presence of a repressor element in P2-P1 region. The probability that construct 6 Δ NE has stronger activity than construct 2 Δ NE is 0.44 (> 0.05) (**Table 4-5**). This shows that the both constructs have the same activity. The probability that construct 7NE750bp has stronger activity than construct 7 by chance is 0.05 (**Table 4-6**). This is a borderline case so this suggests that the additional 180 bp fragment may not be necessary for better *hoxd4a* neural enhancer activity.

constructs	% of expressing embryos relative to injected	Expression pattern	Ratio of embryos with this expression pattern	Average hindbrain expression level of analyzed embryos (1 = lowest and 5 = highest)
construct 2	68.7 %	r6/7	11/12	3.1
construct 2 Δ NE	49.4 %	r6/7	11/12	3.3
construct 3	2.8 %	Spotty	1/36	N.A
construct 4	0 %	N.A.	N.A.	N.A
construct 5	10.6 %	Posteriorized	13/13	1.9
construct 6	81.9 %	r6/7	10/11	3.5
construct 6 Δ NE	40.3 %	r6/7	8/11	3.4
construct 7	38.1 %	Faint hindbrain expression	17/17	1.5
construct 7 Δ NE	0 %	N.A.	N.A.	N.A.
construct 7NE750bp	50.4 %	Very faint hindbrain expression	18/18	0.3

Table 4-1: Summary of expression patterns.

construct	Percentage of expressing embryos				p-value
	Trial 1	Trial 2	Trial 3	Average	
construct 2	71.4	66.6	68.0	68.7	0.02
construct 2ΔNE	43.2	55.0	50.0	49.4	

Table 4-2: construct 2 vs construct 2 Δ NE. Since the p-value is less than 0.05, this shows that construct 2 has a stronger activity than construct 2 Δ NE. The p-value is rounded off to 2 decimal places.

construct	Percentage of expressing embryos				p-value
	Trial 1	Trial 2	Trial 3	Average	
construct 6	89.3	100.0	100.0	96.4	< 0.01
construct 6ΔNE	50.0	53.9	54.2	52.7	

Table 4-3: construct 6 vs construct 6 Δ NE. Since the p-value is less than 0.05, this shows that construct 6 has a stronger activity than construct 6 Δ NE. The p-value is rounded off to 2 decimal places.

construct	Percentage of expressing embryos				p-value
	Trial 1	Trial 2	Trial 3	Average	
construct 2	71.4	66.6	68.0	68.7	0.01
construct 6	89.3	100.0	100.0	96.4	

Table 4-4: construct 2 vs construct 6. Since the p-value is less than 0.05, this shows that construct 6 is stronger than construct 2. The p-value is rounded off to 2 decimal places.

construct	Percentage of expressing embryos				p-value
	Trial 1	Trial 2	Trial 3	Average	
construct 2ΔNE	43.2	55.0	50.0	49.4	0.44
construct 6ΔNE	50.0	53.9	54.2	52.7	

Table 4-5: construct 2 Δ NE vs construct 6 Δ NE. Since the p-value is more than 0.05, this shows that both constructs have the same activity. The p-value is rounded off to 2 decimal places.

construct	Percentage of expressing embryos					p-value
	Trial 1	Trial 2	Trial 3	Trial 4	Average	
construct 7	39.3	35.3	29.0	31.1	33.7	0.05
construct 7NE750bp	44.4	50.0	48.8	58.3	50.4	

Table 4-6: construct 7 vs construct 7NE750bp. Since p-value is equal to 0.05, this suggests that construct 7 and construct 7NE750bp may have the same activity. The p-value is rounded off to 2 decimal places.

4.2.2 Recapitulation of *hoxd4a* expression from 2 dpf (48 hpf) to 5 dpf

To determine the extent to which my transgenes could recapitulate *hoxd4a* expression, transgenic embryos with r6/7 expression (constructs 2, 2 Δ NE, 6 and 6 Δ NE) at 24 hpf were observed for 5 days and compared with the images from Punnamoottil et al., 2008. As for the transgenic embryos with posteriorized expression (construct 5), they were observed for 3 days to see if the P2 promoter region plays a later role in r6/7 expression. As far as possible, the same embryos observed at 24 hpf were observed for subsequent days. However, as it is difficult to position the embryos in the right view, these embryos were sometimes destroyed accidentally and I was unable to observe them on subsequent days. Hence, other embryos of the same age were photographed and kept for observation on subsequent days. All fluorescent images were taken with an exposure time of 8 s unless stated.

4.2.2.1 P2 promoter region does not play a role in *hoxd4a* r6/7 expression

Embryos injected with construct 5 were analyzed at 48 hpf and 72 hpf. Expression remained posteriorized in 9/9 embryos analyzed at 48 hpf (**Figs. 4-14A and 4-15A**) and in 14/14 embryos analyzed at 72 hpf (**Figs. 4-14B and 4-15B**). This suggests that the P2 promoter region does not play a role in *hoxd4a* r6/7 expression.

Figure 4-14: Lateral view of embryos injected with construct 5, observed at 48 hpf and 72 hpf. White arrowhead points to r6/7 boundary. Red arrowhead points to the GFP expression limit. A: At 48 hpf. B: At 72 hpf. Expression remained posteriorized at 48 hpf and 72 hpf.

Figure 4-15: Dorsal view of embryos injected with construct 5, observed at 48 hpf and 72 hpf. White dashed line marks the r6/7 boundary. Red arrowhead points to the GFP expression limit. A: At 48 hpf. Note: Embryos shown in Figures 4-14A and 4-15A are different. B: At 72 hpf. Expression remained posteriorized at 48 hpf and 72 hpf.

4.2.2.2 *hoxd4a* expression was recapitulated in transgenic embryos injected with constructs 2, 2ΔNE, 6 and 6ΔNE from 2 dpf to 5 dpf

Transgenic embryos with r6/7 expression (constructs 2, 2ΔNE, 6 and 6ΔNE) were observed for 5 days. At 2 dpf, *hoxd4a* transgenic expression in 7/10 embryos injected with construct 2 continued to delimit at r6/7 boundary (**Figs. 4-16A and 4-17A**), directly below the otic vesicle while it looked slightly posteriorized in 3/10 embryos (**Figs. 4-16B and 4-17B**). Besides hindbrain expression, expression was also observed in Rohon-Beard cells and muscle fibres in 6/10 embryos. Rohon-Beard (RB) cells are large embryonic mechanosensory neurons, found in the dorsal spinal cord. Each RB neuron extends three axons- two longitudinally within the dorsal spinal cord and one peripherally, innervating the skin (Solnica-Krezel, 2002). Muscle fibres appear as streak-like structures around the spinal cord. In the absence of the 3' neural enhancer (construct 2ΔNE), 8/12 analyzed embryos showed slight posteriorized expression (**Figs. 4-18A and 4-19A**) while only 4/12 analyzed embryos had r6/7 expression. Expression in Rohon-Beard cells and muscle fibres was observed in 2/12 embryos.

12/14 analyzed embryos injected with construct 6 had r6/7 expression (**Figs. 4-18B and 4-19B**) and 2/14 embryo showed slight posteriorized expression. Expression in Rohon-Beard cells and muscle fibres was observed in 6/14 embryos. In the absence of the 3' neural enhancer (construct 6ΔNE), 10/14 embryos had r6/7 expression (**Figs. 4-18C and 4-19C**) while 4/14 embryos showed slight posteriorized expression. Expression in Rohon-Beard cells and muscle fibres was observed in 1/14 embryos. Based on the scores, it seems that without the 3' neural enhancer, expression in Rohon-Beard cells and muscle fibres is hardly observed. None of the 2 dpf transgenic embryos injected with constructs 2, 2ΔNE, 6 and 6ΔNE was found to recapitulate *hoxd4a* expression in the branchial arches and vagal ganglion.

48 hpf

Figure 4-16: Lateral view of embryos injected with construct 2, observed at 48 hpf. r6/7 boundary is shown with a white arrowhead. Red arrowhead points to the GFP expression limit. A: Expression in this embryo continued to delimit at r6/7 boundary. B: Slight posteriorized expression was observed in this embryo.

Figure 4-17: Dorsal view of embryos injected with construct 2, observed at 48 hpf. r6/7 boundary is marked with white dashed lines. Red arrowhead points to the GFP expression limit. A: Expression in this embryo continued to delimit at r6/7 boundary. B: Slight posteriorized expression was observed in this embryo.

Figure 4-18: Lateral view of embryos injected with constructs 2ΔNE, 6 and 6ΔNE, observed at 48 hpf. r6/7 boundary is shown with a white arrowhead. Red arrowhead points to the GFP expression limit. A: Embryo injected with construct 2ΔNE. Slight posteriorization was observed. B: Embryo injected with construct 6. Expression continued to delimit at r6/7 boundary. C: Embryo injected with construct 6ΔNE. Expression continued to delimit at r6/7 boundary. Image was taken at 3 sec because the GFP signal at 8 sec was very strong.

Figure 4-19: Dorsal view of embryos injected with construct 2 Δ NE, 6 and 6 Δ NE, observed at 48 hpf. r6/7 boundary is marked with white dashed lines. Red arrowhead points to the GFP expression limit. A: Embryo injected with construct 2 Δ NE. Slight posteriorization was observed. Image was taken at 500 ms because the GFP signal at 8 sec was very strong. Note: Embryos shown in Figures 4-18A and 4-19A are different. B: Embryo injected with construct 6. Expression continued to delimit at r6/7 boundary. Note: Embryos shown in Figures 4-18B and 4-19B are different. C: Embryo injected with construct 6 Δ NE. Expression continued to delimit at r6/7 boundary.

At 3 dpf, *hoxd4a* transgenic expression (**Figs. 4-20 to 4-23**) was recapitulated in the posterior hindbrain in all embryos injected with constructs 2 (**Figs. 4-20A and 4-21A**), 2 Δ NE (**Figs. 4-20B and 4-21B**), 6 (**Figs. 4-22A and 4-23A**) and 6 Δ NE (**Figs. 4-22B and 4-23B**). Expression in the heart wall was recapitulated in embryos injected with constructs 2 Δ NE and 6 Δ NE; expression in the pectoral fins was recapitulated in embryos injected with constructs 2, 2 Δ NE and 6; and expression in the vagal ganglion and nerve was recapitulated in embryos injected with construct 2 Δ NE.

At 4 dpf, *hoxd4a* transgenic expression (**Figs. 4-24 to 4-27**) continued to be recapitulated in the posterior hindbrain in all embryos injected with constructs 2 (**Figs. 4-24A and 4-25A**), 2 Δ NE (**Figs. 4-24B and 4-25B**), 6 (**Figs. 4-26A and 4-27A**) and 6 Δ NE (**Figs. 4-26B and 4-27B**). Expression in the heart wall and pectoral fins was recapitulated in embryos injected with constructs 2, 2 Δ NE and 6; expression in the vagal ganglion was recapitulated in constructs 2 Δ NE and 6; expression in the cerebellum was recapitulated in embryos injected with construct 2; and expression in the medioventral hindbrain was recapitulated in embryos injected with construct 6 Δ NE.

At 5 dpf, *hoxd4a* transgenic expression (**Figs. 4-28 to 4-31**) continued to be recapitulated in all embryos injected with constructs 2 (**Figs. 4-28A and 4-29A**), 2 Δ NE (**Figs. 4-28B and 4-29B**), 6 (**Figs. 4-30A and 4-31A**) and 6 Δ NE (**Figs. 4-30B and 4-31B**). Expression in fine fibres innervating the pectoral fins and pharynx was recapitulated in embryos injected with constructs 2, 2 Δ NE and 6 Δ NE. Expression in the nerve innervating the pectoral fins was recapitulated in embryos injected with construct 6.

3 dpf

Figure 4-20: Lateral view of embryos injected with constructs 2 and 2ΔNE, observed at 72 hpf (3 dpf). White arrowhead points to the delimitation of *hoxd4a* expression. A: Embryo injected with construct 2. Expression was observed in the posterior hindbrain and branchial arches. B: Embryo injected with construct 2ΔNE. Expression was observed in the heart wall, posterior hindbrain and vagal ganglion.

Figure 4-21: Dorsal view of embryos injected with constructs 2 and 2ΔNE, observed at 3 dpf. Delimitation of *hoxd4a* expression is marked with white dashed lines. A: Embryo injected with construct 2. Expression was observed in the pectoral fins and posterior hindbrain. B: Embryo injected with construct 2ΔNE. Expression was observed in the pectoral fins and posterior hindbrain.

Figure 4-22: Lateral view of embryos injected with constructs 6 and 6 Δ NE, observed at 3 dpf. White arrowhead points to the delimitation of *hoxd4a* expression. A: Embryo injected with construct 6. Image was taken at 3 sec because the GFP signal at 8 sec was strong. Expression was observed in the posterior hindbrain. B: Embryo injected with construct 6 Δ NE. Expression was observed in the heart wall and posterior hindbrain.

Figure 4-23: Dorsal view of embryos injected with construct 6 and 6ΔNE, observed at 3 dpf. Delimitation of *hoxd4a* expression is marked with white dashed lines. A: Embryo injected with construct 6. Image was taken at 3 sec because the GFP signal at 8 sec was strong. Expression was observed in the pectoral fins and posterior hindbrain. B: Embryo injected with construct 6ΔNE. Expression was observed in the posterior hindbrain.

4 dpf

Figure 4-24: Lateral view of embryos injected with constructs 2 and 2 Δ NE, observed at 4 dpf. A: Embryo injected with construct 2. Expression was observed in the posterior hindbrain, heart and cerebellum. B: Embryos injected with construct 2 Δ NE. Image was taken at 3 sec because GFP signal at 8 sec was strong. Expression was observed in the posterior hindbrain, vagal ganglion and heart wall.

Figure 4-25: Dorsal view of embryos injected with constructs 2 and 2 Δ NE, observed at 4 dpf. A: Embryo injected with construct 2. Expression was observed in the posterior hindbrain, pectoral fins and cerebellum. B: Embryo injected with construct 2 Δ NE. Image was taken at 3 sec because GFP signal at 8 sec was strong. Expression was observed in the pectoral fins and posterior hindbrain.

Figure 4-26: Lateral view of embryos injected with constructs 6 and 6 Δ NE, observed at 4 dpf. A: Embryo injected with construct 6. Image was taken at 3 sec because the GFP signal at 8 sec was strong. Expression was observed in the heart wall, posterior hindbrain and vagal ganglion. **B:** Embryo injected with construct 6 Δ NE. Expression was observed in the posterior hindbrain.

Figure 4-27: Dorsal view of embryos injected with constructs 6 and 6 Δ NE, observed at 4 dpf. A: Embryo injected with construct 6. Image was taken at 3 sec because the GFP signal at 8 sec was strong. Expression was observed in the pectoral fins and posterior hindbrain. B: Embryo injected with construct 6 Δ NE. Expression was observed in the pectoral fins and medioventral and posterior hindbrain.

5 dpf

Figure 4-28: Lateral view of embryos injected with constructs 2 and 2 Δ NE, observed at 5 dpf. A: Embryo injected with construct 2. Expression was observed in the hindbrain and fine fibres innervating the fins (shown with red arrow) and pharynx. B: Embryo injected with construct 2 Δ NE.

Figure 4-29: Dorsal view of embryos injected with constructs 2 and 2 Δ NE, observed at 5 dpf. A: Embryo injected with construct 2. Expression was observed in the pectoral fins and hindbrain. B: Embryo injected with construct 2 Δ NE. Image was taken at 3 sec because GFP signal at 8 sec was strong. Expression was observed in the pectoral fins and hindbrain.

Figure 4-30: Lateral view of embryos injected with constructs 6 and 6 Δ NE, observed at 5 dpf. A: Embryo injected with construct 6. Expression was observed in the hindbrain and nerve innervating the fin (shown with the red arrow). B: Embryo injected with construct 6 Δ NE. Expression was observed in the hindbrain and fine fibres innervating the fins (shown with the red arrow) and pharynx.

Figure 4-31: Dorsal view of embryos injected with constructs 6 and 6 Δ NE, observed at 5 dpf. A: Embryo injected with construct 6. Expression was observed in the hindbrain. B: Embryos injected with construct 6 Δ NE. Expression was observed in the hindbrain and pectoral fins.

4.3 Identification of zebrafish *hoxd4a* 5' neural enhancer

Failure of constructs 2 Δ NE and 6 Δ NE to abrogate *hoxd4a* r6/7 expression in zebrafish embryos suggests the presence of a 5' *hoxd4a* neural enhancer upstream of the gene. Analysis of the PostP1-ATG region in various fish species and vertebrates using VISTA browser (not shown) and NCBI blast (**Fig.4-32**) revealed three conserved regions, denoted as CR1-3 (**Fig. 4-33**). CR3 contains two RARE half-sites with slight deviation in the second half-site while CR1 and CR2 contain only one RARE half site. In CR2, the first base of RARE half-site sequence is not conserved. Since CR3 is the only region that contains both RARE half-sites, we hypothesized that the 5'neural enhancer lies within this region. To test this hypothesis, I deleted CR3 from the 6 Δ NE background (construct 6 Δ NE Δ CR3) and examined expression in transgenic zebrafish embryos. Correct r6/7 expression was virtually eliminated, though weak and aberrant expression was noted. On average, 54.5% of embryos injected with construct 6 Δ NE Δ CR3 expressed GFP. Ten embryos were randomly selected, photographed and analyzed at 24 hpf. In all embryos analyzed, there was some expression in the tail and ectopically in the eye (**Fig. 4-34**). Hindbrain expression was faint and unspecific (**Figs. 4-35A and C**). This implies that CR3 contributes activity to a 5' neural enhancer.

To determine whether the other conserved regions might also contribute to 5' neural enhancer activity, a construct with CR1 deleted (construct 6 Δ NE Δ CR1) was tested. On average, 51.5 % of embryos injected with construct 6 Δ NE Δ CR1 expressed GFP. Fourteen embryos were randomly selected, photographed and analyzed at 24 hpf. Of these, 11/14 embryos recapitulated *hoxd4a* expression by delimiting at r6/7, while 3/14 embryos showed slight anteriorization (**Figs. 4-35B and D**).

Unpaired t-test was performed to find out whether there is any difference in the enhancer strength between constructs 6 Δ NE and 6 Δ NE Δ CR1 based on the percentages of expressing embryos, with the significance level set at 5%. The probability that construct 6 Δ NE has stronger activity than construct 6 Δ NE Δ CR1 by chance is 0.73 (> 0.05) (**Table 4-7**). This shows that constructs 6 Δ NE and 6 Δ NE Δ CR1 have the same activity and the 5' neural enhancer does not lie within CR1. Hence, removal of CR1 does not abolish *hoxd4a* expression.

Query	1500	TGGAACGATGAGACTCCAGTAACTTCTGTG	TGACCTCTACA	TGACAAAGACGGTCTTCAT	1559
Sbjct	46338	TGAAATGATGAGATTCTTGTAACTTTGGAT	TGACC	CCTTCTTAACAAAGACAGTCCCCAT	46397
Query	1560	CAATTGCTGCGAGGCTTTATGCGTCTCT	ttttgtgccctgga	-ccggtttt-tatcttttg	1617
Sbjct	46398	CAATTGCTCCCCAGGCTTTATTCCTCTTTT	TGAGCCAGGGAGACGTTTGTATCTTTTG		46457
Query	1618	tattgttttgttttt	ACCAGTCAAAATGCCTGGGCTGGGTTTGATTTAGCTTAGTTGAAT	1677	
Sbjct	46458	T-TTG-----TTTATCTGTCAGACTCACTGGCCGTAGTTTGATTTAGTATTGTAGAAC		46509	
Query	1678	ATTAA	1682		
Sbjct	46510	AATAA	46514		

Query	1506	GATGAGACTCCAGTAACCTCTGTG	TCGACCT	CTACA	TCACAA	AGACGGTCTTCATCAATTG	1565	
Sbjct	92650	GATGAGATTTCGTGTAACCTTTGGAT	TCGACG	CCTTCT	TCACAA	AGACAGTGTCATCAATTG	92709	
Query	1566	TCTGCAGGCTTTATGCGTCTC	ttttt	gtgc	cctggac	cgtttt-tatc	ttttgt	1618
Sbjct	92710	TCCTCACTCATTATTCG-CTGTCTTTATG	CCAA	GGGGC	TTTTG	TATC	TTTTGT	92762

Query	1506	GATGAGACTCCAGTAACCTCTGTG	FGACCT	CTACA	FGACAA	AGACGGTCTTCATCAATTG	1565
Sbjct	169334	GATGAGATTCTGTGTAACCTTTGGAT	FGACC	CCTCTT	FGACAA	AGACAGCTTCATCAATTG	169393
Query	1566	TCTGCAGGCTTTATGCGCTCTC	tttttgtg	gccctgg			1600
Sbjct	169394	TCCTCAGTCTTTCTTC	ACTGCTTT	TATGCCATGG			169427

Query	1506	GATGAGACTCCAGTAACCTCTGTG	TCACATGACAAAGACGGTCTTCATCAATTG	1565
Sbjct	166229	GATGAGATTCTGTGTAACTTTTCGATGACG	CCTTCTTGACAAAGACAGTGTCATCAATTG	166170
Query	1566	TCTGCAGGCTTTTATGCGCTCTC	tttttggtgcc-ctggacggtttttatcctttgtattgt	1623
Sbjct	166169	TCCTCACTCATTATTCG-CTGCTTTTATGCCGAAGGGC	TTTTGTATCTTTGT-TTGT	166113

Query	1	TGAAATGATGAGATTCTTTGTAACTTTTTGGAT TGACCC CTTCT TAACAA AGACAGTCCCCAT	60
Sbjct	98128	TGAAGAGATGAGAGTCATGTAAC TTTTGGAT TGACC CATCT TGACA AAGACAGTGTC CAT	98187
Query	61	CAATTGTC CCCCAGGCTTTATTC CCCTC TTTT GAGCCAGGGAGACGTTTGTATCTTTTG	120
Sbjct	98188	CAC TTGTCCTCAGTCATTCTT-CCCTGC TTTTATGCCAAAG-GGCTTTTGTGTCTTTTG	98245
Query	121	TT	122
Sbjct	98246	TT	98247

Salmo salar to *Danio rerio*

```

Query 1      TGAAATGATGAGATTCTTGTAACCTTTGGATGACCCCTTCTTAACAAAGACAGTCCCCAT 60
           || || ||||| || ||||| || ||||| || || ||||| || || ||
Sbjct 87933  TGGAACGATGAGACTCCAGTAACTTCTGTGTGACCTCTACATGACAAAGACGGTCTTCAT 87874

Query 61     CAATTGTCCCCAGGCTTTATTCCCCTCTTTTGTAGCCAGGGAGACGTTTGTATCTTTTG 120
           ||||| ||||| || ||||| || ||||| || || ||||| |||||
Sbjct 87873  CAATTGTCTGCAGGCTTTATGCGTCTCTTTTGTGCCCTGGA-CCGTTT-TATCTTTTG 87816

Query 121    T-TTG-----TTTATCTGTCAGACTCACTGGCCGTAGTTTGATTAGTATTGTAGAAC 172
           | ||| ||||| ||||| || ||||| || ||||| || || |||
Sbjct 87815  TATTGTTTGTTTTACCAGTCAAAATGCCCTGGGCTGGGTTTGATTAGCTTAGTTGAAT 87756

Query 173    AATAA 177
           | |||
Sbjct 87755  ATTAA 87751

```

Salmo salar to *Gallus gallus*

```

Query 10     GAGATTCTTGTAACCTTTGGATGACCCCTTCTTAACAAAGACAGTCCCCATCAATTGT 67
           || ||||| ||||| ||||| || || ||||| || || |||||
Sbjct 177    GAAATTCTGGTAACCTTTGGATGACCTCTGCGAGACAAAGACAGGACAGATCTATTGT 234

```

Salmo salar to *Homo sapiens*

```

Query 11     AGATTCTTGTAACCTTTGGATGACCCCTTCTTAACAAA 49
           ||||| ||||| ||||| ||||| || ||
Sbjct 4654   AGATTCTGGTAACCTTTGGGTGACCCGTATGAAGACAA 4692

```

Figure 4-32: Sequence alignment of PostP1-ATG region using NCBI blast. Conserved region (CR3) was identified using NCBI blast. Bases that align with the RARE half-sites are highlighted in green.

CR1

GTGCTTTTAATTCTTTACAACCAAGTAAATGTTCTAACTCAATTTATGTTCTAACTTAACTTTATAAGTTACGCAAGCTGTTTTAAGTTA
GTTTAACATAATATAAGTTCAATGGACTCATAAAGCTCATGGGGTACTTTTAAAAAAAAAATTTTCCAAATTTTCGAGTGCCTTG
GACTGATGTTTGCTGTTTATTCTGATGAATACCTTACACAAAGAGCACCGACACATTATTTAAAGCTACCCCTGAGCACTTTTTGGTT
GATTTTGGATTTTGGAAATAAGGTTAATTTGATTCAAGTTAAAATTTAAGACAACCCGGATGTTTTTGTACAGTGACAGGCAATATA
AATTGCGTATATATGTTGGAAGACGGCACCTGGCAACACAGATGACCATTAAATTAAGAAAATCAACTGATATTCATATGCTGATTATG
CATATTATTTTAATAACTTCTGTTGTGATGCTTTTTGCTGCGCTTTCTTAGTGAACATTGACTTTTTATTGCTGTCGATGGTTACTTGA

CR2

TGGACGTTCAACTAGGCCTATTGAACATTTCCATCTCACAACGTTCTTCACACTAAAGAAAATGCTTGTTTGGCGTAATGCAATGTT

CR3

TGGTCAATTATAACAAACCCAAAAATTGGGTAATTTCTTACTAATTTAATTTAAATAACATTTTAAACCAAAACGTTTGGGTTTGCTA
TATTTGACCCAACGTTGGAATACAACAACCCAGAAATTTAATTTAATTTAAATTCAAACGAATGTGTGGAATGTTCTTCACATTTAAAG
AACGTGTTCTATTAAGAGCTTTGAACTGAATGGAATCCCAAAAGGGTTCTTATTCTACACTGTTAACAAATCCTGGTTGCCTTAATTTT
TTTTATCAAATTAACCTTAAGAGTCCAAAAATCCTAAATCAAACAAAAAGTGCTCAGATGTAGCTTAACTAATGCTAACTCTGGGTAA
GGGATTTAATAGAATAAACTACAAAAATCAGTCCAAGGTACTCGAAAAAATTTGATTTTTTTTAAAGTAACCTTATGAGCTTTATGAGT
CCATTGAACCTTATATTATGTCAAACCTGACTTAAACCGCTTGCGTAGCTACAAAGTTAGGTTAGAACATGATTAACCTTAGTTTATAAAA
ATGAAAACACATGATGTCATGACTAACTGATCATATAATTTCTACAGTGTCGAATATTACAATCAAACCCCATTTTTACACTTTTTAT
AAAGGTCTCCTGAACATTTTAACATTTTAAATACCAATCCTACAGTTTAGTTCTTCTTAAAGACGCAGGGAGCAGTAATTAACATTA
GTAATATTTTTGTATGAATGTTAACTCACTTTGCTCCCTGGTCTTACCTTATGAGCTGCCATACGATTATTTCTTTTATGTGCTGTTTG
ATTATTGATTGGCACGGCGCATTGATGGCGAACCAGGGCGCCCTATACAGCGAGCTGAAAGGATGAGAAGCGCTGGAACGATGAG

CR3

ACTCCAGTAACTTCTGTGTGACCTCTACATGACAAAGACGGTCTTCATCAATTGCTGCAAGGCTTTATGCGTCTCTTTTTGTGCCCTG

GACCGTTTTTATCTTTTGATTGTTTTGTTTTACCAGTCAAAATGCCTGGGCTGGGTTTGATTTAGCTTAGTTGAATATTAAATTTACAA
AAAATGTTATACACTCACACACACATGCTTGTTATGTGGCATGAGAACATGTTTTTATGATAGCATCTAATGTGTAATAGAGATCCT
CCCATATGATGATGAACAATTGGCTATTGCATGAGTCCCAGCACTGCTTTCCTCCTTGCTCTTTCTTTTATATGGGTATTTAATCAAT
AAAATTTTATATTGTGTTATTTCCATCCTAATTGAGAGATAATGTGAAGTTGTAAAAATCACTACTAGGTGTGCACAGCTTCTCTCG
GTTGATGAAGTCCCTCCACCTCGATTGGCCAAGCTGGTCACATGGTAGGCAACTTTATTAGTTGACAGCAAGTAGGAGGGCTTTAT
GGAGGGAGGAAAAAAGACAACCTCGAGAAAAATTAGTATTTCTACCTTCAGAAATTA

Figure 4-33: PostP1-region (from -2103 to -1). Conserved regions are underlined and highlighted in yellow. CR1 is from -2103 to -1723. CR2 is from -1594 to -1467. CR3 is from -602 to -420. RARE half-sites are shown in red. Consensus RARE sequence is (A/G)G(T/G)TCA.

Figure 4-34: Embryo injected with construct 6ΔNEΔCR3, observed at 24 hpf (lateral view). White arrow points to r6/7 boundary. Some expression was observed in the tail and eye regions. Brightfield and fluorescent images at 50 x magnification were taken with a monochrome camera with an exposure time of 4 ms and 8 s respectively.

Figure 4-35: A, B: Lateral view of embryos injected with constructs 6ΔNEΔCR1 and 6ΔNEΔCR3, observed at 24 hpf. White arrow points to r6/7 boundary. A: Embryo injected with construct 6ΔNEΔCR3. Brightfield and fluorescent images at 80 x magnification were taken with a monochrome camera with an exposure time of 4 msec and 8 sec respectively. B: Embryo injected with construct 6ΔNEΔCR1. Brightfield and fluorescent images were taken with a colour camera with an exposure time of 4 msec and 8 sec respectively. C, D: Dorsal view of embryos injected with constructs

6 Δ NE Δ CR1 and 6 Δ NE Δ CR3, observed at 24 hpf. C: Embryo injected with construct 6 Δ NE Δ CR3. Brightfield and fluorescent images were taken with a monochrome camera with an exposure time of 4 msec and 8 sec respectively. D: Embryo injected with construct 6 Δ NE Δ CR1. Brightfield and fluorescent images were taken with a colour camera with an exposure time of 4 msec and 8 sec respectively.

construct	Percentage of expressing embryos				p-value
	Trial 1	Trial 2	Trial 3	Average	
construct 6 Δ NE	50.0	53.9	54.2	52.7	0.73
construct 6 Δ NE Δ CR1	54.7	54.2	45.5	51.5	

Table 4-7: construct 6 Δ NE vs construct 6 Δ NE Δ CR1. Since the p-value is more than 0.05, this shows that both constructs have the same activity. The p-value is rounded off to 2 decimal places.

4.4 Activity of murine *Hoxd4* regulatory elements in transgenic zebrafish embryos

It has been proposed that murine *Hoxd4* transcription is driven by two promoters, P2 located 5.2 kb upstream from the coding region and P1 located 1.1 kb upstream from the coding region. P1 was proposed to be used in anterior regions up to the r6/7 expression boundary (Folberg et al., 1997). Given our observation that the equivalent P1 region of zebrafish *hoxd4a* lacks promoter activity, and given that the ability of the murine P1 promoter has been called into question (Phua et al, 2011), it was of interest to test the activity of the murine P1 region in zebrafish embryos.

The transposon containing the murine *Hoxd4* 3' neural enhancer and P1 region (**Fig. 4-36**, construct 10) was injected into zebrafish embryos at the 1-cell stage. No expression was observed. Hence, it is possible that there is another promoter downstream of the P1 region that drives r6/7 expression, just as in zebrafish. This presumptive downstream promoter is named P3. To locate the P3 promoter, transposons containing the *Hoxd4* neural enhancer region and putative promoter regions were coinjected with transposase into the zebrafish embryos at the 1-cell stage (**Fig. 4-36**).

The P2 region was shown to drive a more posterior *Hoxd4* expression (Folberg et al., 1997). The same expression pattern was recapitulated in 8/8 zebrafish embryos analyzed at 24 hpf (construct 11) (**Figs. 4-37A and 4-38A**). In comparison to embryos injected with construct 5 (**Figs. 4-7C and 4-8C**), the expression in these embryos looked less posteriorized. The expression remained posteriorized at 48 hpf in 18/18 embryos analyzed (3/18 showed slight posteriorization) (**Figs. 4-37B and 4-38B**) and at 72 hpf in 7/7 embryos analyzed (**Figs. 4-37C and 4-38C**).

No expression was observed in embryos injected with constructs 8 (with P2-ATG region) and 9 (with P2-P1 region) (**Fig 4-36**). This implies that the mouse *Hoxd4* and zebrafish *hoxd4a* transcriptional regulations are different. Zebrafish P2-ATG region was able to drive r6/7 *hoxd4a* expression but not mouse P2-ATG region.

Figure 4-36: Schematic diagram of mouse constructs tested (not drawn to scale). NE = neural enhancer.

Figure 4-37: Lateral view of embryo injected with construct 11, observed at 24 hpf, 48 hpf and 72 hpf. White arrowhead points to anterior limit of *endogenous hoxd4a* expression. Red arrowhead points to anterior limit of expression driven by the mouse P2 promoter. A: 24 hpf. B: 48 hpf. C: 72 hpf.

Figure 4-38: Dorsal view of embryo injected with construct 11, observed at 24 hpf, 48 hpf and 72 hpf. Red arrowhead points to anterior limit of expression driven by mouse P2 promoter. A: 24 hpf. B: 48 hpf. C: 72 hpf.

4.5 Discussion

Transcription regulation of zebrafish *hoxd4a* is different from that of murine *Hoxd4* in terms of both enhancer and promoter functions. A 570 bp enhancer 3' to murine *Hoxd4* directs transgene expression in the mouse embryo with an anterior boundary at r6/7, thus faithfully recapitulating this aspect of *Hoxd4* expression. The 570 bp zebrafish sequence that I chose to test for *hoxd4a* 3' neural enhancer function spans a 150 bp region that is highly similar to the murine *Hoxd4* 3' neural enhancer, encompassing sites A through G and the RARE. Unexpectedly, however, I showed in the context of a zebrafish *hoxd4a* 200 bp region (construct 7) (where P3 promoter lies), that the zebrafish *hoxd4a* 3' neural enhancer could not direct r6/7 expression in transgenic zebrafish embryos. This result was all the more surprising given that this same zebrafish element had been shown to set the correct r6/7 boundary when tested with a rabbit beta-globin promoter in transgenic murine embryos, just like the homologous murine enhancer (Nolte et al., 2003). A possible explanation for the discrepancy in the activity of the zebrafish *hoxd4a* 3' neural enhancer tested with a zebrafish *hoxd4a* promoter region *versus* a beta-globin promoter lies with the specificity of promoter-enhancer interactions. It is possible that with the beta-globin promoter, the binding of activators and coactivators to the zebrafish *hoxd4a* 3' neural enhancer fortuitously provokes a favourable local environment (such as might be mediated by DNA bending) that provides a productive stereospecific interface with the basal transcription machinery, thereby resulting in r6/7 expression. Such favourable and fortuitous interactions may be blocked in the endogenous context. Non-exclusively, the transcription factors bound to the neural enhancer in the two species may play a determining or contributing role.

Embryos injected with construct 7, containing the 570 bp neural enhancer and the 200 bp region upstream of the ATG start site, had faint expression that did not demarcate clearly at r6/7 boundary. Extending this enhancer region to 750 bp (construct 7NE750bp) likewise failed to set r6/7 expression. This result is supported by the observations of another group who tested a 989 bp fragment 3' to zebrafish *hoxd4a* (Punnamoottil et al., 2010). This 989 bp fragment entirely included the 570 bp 3' neural enhancer fragment that I tested, and all but 70 bp of the extended 750 bp region (construct 7NE750bp). In their study, GFP expression faded from r8 anteriorly to r7 and the anterior boundary was not clearly delimited.

The r6/7 expression observed in zebrafish embryos injected with constructs 2, 2ΔNE, 6, and 6ΔNE suggests that a 5' neural enhancer in zebrafish is responsible for the setting of the r6/7 boundary. The location of this 5' neural enhancer was identified by mutational analysis. Deletion of a conserved region, CR3, from the postP1-ATG region (construct 6ΔNEΔCR3) abolished the neural expression in the injected embryos. Though it will be useful to test the activity of CR3 independently, it poses a challenge to determine the 5' and 3' ends that encompass full activity. A 5' neural enhancer, which is RA-responsive and responsible for neural tube-specific expression, has been reported in human *HOXD4* (Morrison et al., 1996). Although a strong 5' mesodermal enhancer has been reported in mouse, no neural activity has been associated with a broad 5' region (P2 to downstream of the ATG) (Zhang et al., 1997). By contrast, two 5' neural enhancers have been reported in paralogous mouse *Hoxb4* – one in a fragment termed Region D and another in Region E (**Fig. 1-13**) (Sharpe et al., 1998).

The zebrafish *hoxd4a* promoter region also functions differently from the mouse *Hoxd4* promoter region. The zebrafish P2-ATG region (construct 2) was able to drive GFP expression but not mouse P2-ATG region (construct 8). The zebrafish P2 region (construct 5) drove a more posterior expression than the mouse P2 region at 24 hpf (construct 11). However, neither zebrafish nor mouse P2 plays a role in driving r6/7 expression at later stages. It is possible that the mouse promoter regions did not function properly in zebrafish embryos due to the unfavourable stereospecific interface provided by the activators and coactivators. Therefore, it will be necessary to test these regions in murine embryos. At the same time, it will be useful to test the zebrafish promoter regions in murine embryos to check if there is a difference in activity in murine and zebrafish embryos. If there is a difference in activity, this shows that there is a need to test the regulatory regions in their endogenous contexts.

Three possible fates were proposed in the “duplication–degeneration–complementation” model for new paralogs- a) nonfunctionalization (one copy is lost and one copy is left), b) neofunctionalization (one copy acquires new function while the other copy preserves its ancestral function), and c) subfunctionalization (both copies lost part of the ancestral function and they have to work together to achieve full functionality) (Force et al., 1999). The presence of the 5' NE in zebrafish *hoxd4a* and the difference in the promoter structures of zebrafish *hoxd4a* and mouse *Hoxd4* are probably due to subfunctionalization

and possibly also neofunctionalization. Zebrafish *hox* genes have evolved rapidly from the mammalian orthologs under some selective pressures.

With the assumption that all transposons were efficiently integrated into the zebrafish embryos at the same rate, statistical comparison of the percentage of expressing embryos in different promoter contexts using unpaired t-test provides some insights into the organization of the *hoxd4a* transcription unit. A higher percentage of expressing embryos was observed with constructs 2 and 6 than constructs 2 Δ NE and 6 Δ NE suggests that the 3' neural enhancer plays a complementary role to the 5' neural enhancer in *hoxd4a* expression. Therefore, when this 3' neural enhancer was cloned with the 200 bp promoter region (construct 7), hindbrain expression was observed, though it did not delimit at r6/7. The functionality of construct 7 also implies that a promoter (P3) lies in the 200 bp region upstream of ATG.

A higher percentage of expressing embryos observed with construct 6 than construct 2 suggests the presence of a repressor element in P2-P1 region. Sequence comparison of the postP2-preP1 region with various fish species and vertebrates using NCBI blast (**Fig- 4-39**) reveals a conserved region from -3469 bp to -3368 bp, upstream of the ATG site (**Fig. 4-40**). This is the same region where the 5' RARE lies. A 340 bp fragment that included this region has been tested in Punnamootil's lab and has been found to have no reporter expression and hence no *cis*-regulatory activity (Punnamootil et al., 2010). This implies that the conserved region cannot be an enhancer element even though the RARE is present within the region. It has been observed in *Hoxb1* that an RARE can repress *Hox* gene expression (Studer et al., 1994). I thus propose that there may be a repressor element in the postP2-preP1 region possibly dependent on the 5' RARE previously mapped to the region (Poepperl 1993). This possibly explains why a higher percentage of expressing embryos was observed with construct 6 (with PostP1-ATG region) than construct 2 (with P2-ATG region). This hypothesis, however, would be at odds with the function of the paralogous 5' RARE in *Hoxa4* which is responsive to RA and activates *Hoxa4* expression (Packer et al., 1998).

As the above analysis was based on only one study and the activities of the constructs were estimated from the less widely used method - percentage of embryos, the conclusions drawn are weak and not very convincing. The experiments should be repeated

with the analysis of more embryos and stable lines of the transgenic fish should be established. This is to ensure that the outcome of the experiments is not affected by the quality and purity of the injected transposons, the integration efficiency and DNA conformation. Though the assessment of the activity of cis-regulatory elements by the determination of GFP intensity is widely used, it has its own limitations. Firstly, GFP intensity is an average measurement of the GFP expression and it may vary from cell to cell. Secondly, it may be overestimated if more than one copy of GFP gene is integrated into the genome. Therefore, transgene copy number in each embryo should be determined by real-time PCR, in order to compare reliably the expression levels. Thirdly, enhancers increase the probability and not the level of gene expression (Walters et al., 1995).

At 48 hpf (2 dpf), endogenous *hoxd4a* expression has been observed in the hindbrain, branchial arches 3-7 and vagal ganglion. Its hindbrain expression delimits slightly anterior to the posterior end of the otic vesicle (r6/7 boundary c.f. **Fig. 4-6C**). From 3 dpf to 5 dpf, expression has also been observed in the cerebellum (Punnamoottil et al., 2008).

The zebrafish enhancer-trap line CLGY846 was used to visualize and identify the neurons expressing *hoxd4a* due to its *hoxd4a*-like expression pattern (Punnamoottil et al., 2008). In 2 dpf embryos and 3 dpf larvae, they observed *hoxd4a* expression in the efferent neurons of the vagal nucleus. In 3 dpf larvae, they also observed its expression in the posterior and medioventral hindbrain, fins and heart wall. In 5 dpf larvae, *hoxd4a* was shown to be expressed in vagal fibres exiting the hindbrain, vagal ganglion cells and fine fibres innervating the pharynx.

Transgenic embryos injected with constructs 2, 2ΔNE, 6 and 6ΔNE were observed from 2 dpf to 5 dpf. *hoxd4a* expression was not fully recapitulated, probably because the regions tested were not sufficient for full activity. None of the 2 dpf transgenic embryos exhibited expression in the branchial arches and vagal ganglion. Expression in the cerebellum was not observed in 3 dpf and 5 dpf transgenic embryos. Expression in the medioventral hindbrain was also not observed in 3 dpf transgenic embryos.

Danio rerio to *Homo sapiens*

```

Query 1424      TGACTCTCTAATTCTTTTGTTCAG---TAACTCG---GCTTTGACCCGCCTGAACAAGTC 1477
||| | | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 27223141  TGAATATTTAACGCTTTTGTTCATCAATAACTCGTTGGCTTTGACCTGTCTGAACAAGTC 27223200

Query 1478      --GCAATTCAAGGTGAAACACAGGTCACGCTGTCTAACAAA 1516
||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 27223201  GAGCAAT--AAGGTGAAATGCAGGTCACAGCGTCTAACAAA 27223239

```

Danio rerio to *Mus musculus*

```

Query 1424      TGACTCTCTAATTCTTTTGTTCAG---TAACTCG---GCTTTGACCCGCCTGAACAAGTC 1477
||| | | ||| ||||| ||||| ||||| | ||||| ||||| ||||| |||||
Sbjct 15604046  TGAATATTTAACGCTTTTGTTCATCAATAACTTGTGGCTTTGACCTGTCTGAACAAGTC 15604105

Query 1478      --GCAATTCAAGGTGAAACACAGGTCACGCTGTCTAACAAA 1516
||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 15604106  GAGCAAT--AAGGTGAAATGCAGGTCACAATGTCTAACAAA 15604144

```

Danio rerio to *Takifugu rubripes*

```

Query 1424      TGACTCTCTAATTCTTTTGTTCAGTAACTCGGCTTTGACCCGCCTGAACAAGTCGCAATT 1483
||| | | ||||| ||||| || | || ||||| ||||| ||||| | |||
Sbjct 167708    TGACTTTATCATTCCTTTGTGTGAGAACTCTGGGCTTTGACCCGCTCTGAAGGACGAGCAGGA 167767

Query 1484      CAAGGTGAAACACAGGTCACGCTGTCTAACAAA-ATAGATAAG 1525
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 167768    GAAGGTGAAAGACAGGTCAGGCTGTCTAGCAAATATGGAAAAG 167810

```

Danio rerio to *Gasterosteus aculeatus*

```

Query 1418      AAATGATGACTCTCTAATTCTTTTGTTCAGTAACTCGGCTTTGACCCGCCTGAACAAGTC 1477
||| | ||||| | | ||||| ||||| | ||||| ||||| ||||| ||||| |||||
Sbjct 168068    AAACGCTGACTTTATCGTTCCTTTGTTTGAATCTCGGCTTTGACCCGCTCTGAAGGACAC 168009

Query 1478      GCAATTCAAGGTGAAACACAGGTCACGCTGTCTAACAAAA 1517
|| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 168008    GCTGGGGAAGGTGAAAGACAGGTCAGGCTGTCTAACTAAA 167969

```

Danio rerio to *Astatotilapia burtoni*

```

Query 1417      GAAATGATGACTCTCTAATTCTTTTGTTCAGTAACTCGGCTTTGACCCGCCTGAACAAGT 1476
||| | ||||| | | ||||| ||||| | ||||| ||||| ||||| ||||| |||||
Sbjct 90384    GAAACGCTGACTTTATCATTCCTTTGTGTGGGGAACCTCGGCTTTGACCCGTCACAAGGACC 90443

Query 1477      CGCAATTCAAGGTGAAACACAGGTCACGCTGTCTAACAAA 1516
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 90444    GGCAGGGTCAGGTGAAAGGCAGGTCAGGCTGTCTAATAAA 90483

```

Danio rerio to *Salmo salar*

```

Query 1455      GCTTTGACCCGCCTGAACAAGTCGCAATTCAAGGTGAAACACAGGTCACGCTGTCTAACA 1514
||||||| ||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 203020    GCTTTGACCCGCTCTGAACAAGTCTCAGGGCAAGGTGA-----AGGTCAAGCTGTCTAACA 203074

Query 1515      AA 1516
||
Sbjct 203075    AA 203076

```

Figure 4-39: Sequence alignment of postP2-preP1 region with various fish species and vertebrates using NCBI Blast.

GTGCAGAATTAATGTGACAAAACGAAGTAGTAGTCTATTTATTCATTATTAACCTGAAACCTGTAAGTACACAGTTTCAATTA
 TCTAGAACATCTATGTTAACTGCTTTGGCACAATCTACATTGTAAAAGCGCTGTAGAAATAAAGATGAATTGAATTGAAGTAT
 ACACGCACTTCCCTGGATTATAAGTGAACCTTAATAAGCAGGTTTTTCGGGGTTATATTGACAAGCTCTCTAGCGCCCCCAGGCA
 AAAGAGTCATTTCTGTCGTTATTTATACCTTTATTTTAATAATTATATTTAAATTTAAATAAACACTGTCTGCTTAGAAAGCCATGA
 TAATTAATACCTGTATTAAATGCTTTGGATTAGAATAGCAATTAATTTATGTTTCATATTATAAAGGTAAATACAAACATAATAAT
 TTAGCCACTGCTTTAGCCAAACATTATATCAGCTGAATGACGATTGCGTGCTATTTAATTATTGAGTTATAACTGTGAAGTCTAT
 AAACAATAAAAAATATGTGTATTTTTAAATGTTTAAATATTAGTAAACACTTGAATGGGGAAATAACTATATTTTCATCGTTCTAT
 AACTTATTTTCTTTTATATAGATAACTTTTTGTATAAGTATAAAATATCGTACTAAACAATAACATATCGATTAAATAACAATGA
 CCTATCTAAACCTAGAAATCTACAAGCGTATACATTTCCAACCTAGTTAACTGGAAAATTTAATCATTTTCACCCCTATATGCACG
 GATACGACTTATATAAAGGTGCATGCACAAATATATAAATTAATAAATCAACATAAGGTCGTTTTATTTAAGCGTAACTTAGAGTTT
 AATAGAAATGATGACTCTCTAATTCTTTGTTTCAGTAACTCGGCTTTGACCCGCCTGAACAAGTCGCAATTCAAGGTGAACAC
 AGGTCAAGCTGTCTAACAAATAGATAAGATTCACATGTCTTATTTAAATAGAAAGAAAAATATTAAGTGGGTACATGTGATA
 ATAAAAACGCACACAAAAATCTAATATTATTTATCTTATATAAAGAGCGAAAGGCCGCCAGAAATTATAAGCCTTTGCGTC
 TTTTAAATTTACTATTTATTTGTATGCTAAAATGCTATAATCACTATTAAGTAGTAAACATCACGCATAATGTTGTTATACTTT
 GCAAATTATTGAATGCAAAATTTAAATATGTTGATGTGTTTCTCGTGCGCTTTGCTGATATTTTAAATCATGTTAAATACTCCT
 GAGAAAAAAATGAGG

Figure 4-40: Post P2-pre P1 region (from -4886 to -3031). 5' RARE is shown in red. The conserved region (from -3469 to -3368) is underlined.

Chapter 5: Study of *hoxd4a* 3' neural enhancer in zebrafish embryos

5.1 Characterization of sites A-C and G of the zebrafish *hoxd4a* 3' neural enhancer

As noted previously, sequence comparison between mouse and zebrafish *Hoxd4/hoxd4a* 3' neural enhancers revealed eight blocks of homology, denoted as sites A-G and a retinoic acid response element (RARE). The mouse *Hoxd4* 3' neural enhancer has been shown to be responsible for the setting of r6/7 boundary of *Hoxd4* expression and sites D-F and the RARE were previously characterized in mouse embryos (Zhang et al., 2000; Nolte et al., 2003; Nolte et al., 2006). In **Chapter 4**, it is suggested that the zebrafish *hoxd4a* 3' neural enhancer plays only a complementary role to the 5' neural enhancer which actually sets r6/7 delimited expression (c.f. **Figs. 4-22 and 4-23**). Nonetheless, the conservation of sites A through G and the RARE between mouse and zebrafish strongly suggests that they play an important role in both species. (Note: The following analysis was performed before I had discovered the existence of the zebrafish 5' neural enhancer. Thus, the analysis was performed in the presence of this latter element which almost certainly masked some of the effects of the mutations under study.)

To test the biological relevance of sites A, B, C and G, these sites were mutated singly, doubly and compoundly. Different variations of construct 2 (**Fig. 5-1**) were injected into zebrafish embryos at 1-cell stage. All mutations were made in the background of construct 2 (**Fig. 4-5**), and so are named “construct 2n”, where “n” is a description of the mutation. To ensure consistency in the integration of the mutant constructs, the same preparation of transposase mRNA was coinjected. All brightfield images were taken at 4 ms and all GFP images were taken at 8 s for consistency.

Nine expressing embryos injected with construct 2mutA were randomly selected and analyzed. 6/9 embryos showed r6/7 expression (**Figs. 5-2A and 5-3A**) while 3/9 embryos showed anteriorized expression. Ten expressing embryos injected with construct 2mutB were randomly selected and analyzed. 1/10 embryos showed r6/7 expression while 9/10 embryos showed anteriorized expression (**Figs. 5-2B and 5-3B**). Eleven expressing embryos injected with construct 2mutC were randomly selected and analyzed. Of these, 8/11 embryos showed r6/7 expression (**Figs. 5-2C and 5-3C**). 3/11 embryos showed anteriorized expression. Fifteen expressing embryos injected with construct 2mutG were

Figure 5-1: Schematic diagram of different variations of construct 2 tested (not drawn to scale). These constructs contain the P2-ATG promoter region (a) P2 promoter, b) P1 region, and c) region post-P1 and pre-ATG site).

randomly selected and analyzed. Of these, 12/15 embryos showed r6/7 expression (**Figs. 5-2D and 5-3D**), 1/15 showed anteriorized expression and 2/15 embryos showed posteriorized expression. Fifteen expressing embryos injected with construct 2mutAC were randomly selected and analyzed. Of these, 8/15 embryos showed r6/7 expression (**Figs. 5-4A and 5-5A**) while 7/15 embryos showed anteriorized expression (**Figs. 5-4B and 5-5B**). Fourteen expressing embryos injected with construct 2mutAB were randomly selected and analyzed. Of these, 2/14 embryos showed r6/7 expression while 12/14 embryos showed anteriorized expression (**Figs. 5-6A and 5-7A**). Eleven expressing embryos injected with construct 2mutBC were randomly selected and analyzed. 9/11 embryos showed r6/7 expression (**Figs. 5-6B and 5-7B**) while 2/11 embryos showed anteriorized expression. Nine expressing embryos injected with construct 2mutABC were randomly selected and analyzed. Of these, 1/9 embryos showed r6/7 expression while 8/9 embryos showed anteriorized expression (**Figs. 5-6C and 5-7C**). Eight expressing embryos injected with construct 2mutABCG were randomly selected and analyzed. Of these, 7/8 embryos showed r6/7 expression (**Figs. 5-8A and 5-9A**) while 1/8 embryos showed slight anteriorized expression. Ten expressing embryos injected with construct 2mutRARE were randomly selected and analyzed. Of these, 7/10 embryos showed r6/7 expression (**Figs. 5-8C and 5-9C**) while 3/10 embryos showed anteriorized expression. In general, we observed that embryos injected with mutants that resulted in anteriorized expression (i.e. constructs 2mutB, 2mutAB and 2mutABC) appeared to have stronger expression than embryos injected with wild type construct (i.e. construct 2).

Mutation of the RARE in the mouse 3' neural enhancer abolished expression in E8.5 mouse embryos (Zhang et al., 2000). However, I did not make a similar observation when I mutated the *hoxd4a* RARE in the zebrafish 3' neural enhancer (**Figs. 5-8C and 5-9C**), quite possibly due to the presence of the 5' neural enhancer. To check the importance of the RARE in the zebrafish 3' neural enhancer, embryos were injected with construct 7mutRARE, which is driven by the 200 bp promoter region. Eight embryos were randomly selected and analyzed. Images were taken with a monochrome camera with the same exposure time as the other embryos described earlier. No specific hindbrain expression was observed in these embryos (**Figs. 5-8D and 5-9D**) but some expression was observed in the tail region (**Fig. 5-10A**). Since the GFP-expressing embryos injected with construct 7 (with the presence of RARE) had faint specific hindbrain expression, this suggests that the RARE is responsible for the hindbrain expression. Hindbrain expression

was not abolished in embryos injected with construct 7mutABCG. Eight embryos were randomly selected and analyzed. Hindbrain expression was similar to those embryos injected with construct 7. It faintly delimited around the r6/7 boundary in 6/8 embryos (**Figs. 5-8B and 5-9B**) and failed to delimit in 2/8 embryos. The expression in the tail region was stronger in 4/8 embryos analyzed (**Fig. 5-10B**).

Table 5-1 summarizes the expression patterns and expression level observed with the various mutants. The expression level is rated from 1 (lowest) to 5 (highest). The scores for the average expression level serve as a rough guide to the strength of GFP expression and are not meant to be quantitative. Embryos injected with constructs 2mutAC and 2mutABC have the strongest expression while embryos injected with constructs 2mutG and 2mutBC have the weakest expression.

Figure 5-2: Lateral view of expressing embryos injected with constructs 2mutA, 2mutB, 2mutC and 2mutG, observed at 24 hpf. White arrow points to r6/7 boundary. A: Embryo injected with construct 2mutA. Embryo shows r6/7 expression. B: Embryo injected with construct 2mutB. Embryo shows anteriorized expression. C: Embryo injected with construct 2mutC. Embryo shows r6/7 expression. D: Embryo injected with construct 2mutG. Embryo shows r6/7 expression.

Figure 5-3: Dorsal view of embryos injected with constructs 2mutA, 2mutB, 2mutC and 2mutG, observed at 24 hpf. A: Embryo injected with construct 2mutA. Embryo shows r6/7 expression. B: Embryo injected with construct 2mutB. Embryo shows anteriorized expression. Note: the embryo is different from the one shown in Figure 5-2B. C: Embryo injected with construct 2mutC. Embryo shows r6/7 expression. D: Embryo injected with construct 2mutG. Embryo shows r6/7 expression. Note: the embryo is different from the one shown in Figure 5-2D.

Figure 5-4: Lateral view of embryos injected with construct 2mutAC, observed at 24 hpf. White arrow points to r6/7 boundary. A: Embryo shows r6/7 expression. B: Embryo shows anteriorized expression.

Figure 5-5: Dorsal view of embryos injected with construct 2mutAC, observed at 24 hpf. A: Embryo shows r6/7 expression. Note: the embryo is different from the one shown in Figure 5-4A. B: Embryo shows anteriorized expression. Note: the embryo is different from the one shown in Figure 5-4B.

Figure 5-6: Lateral view of embryos injected with constructs 2mutAB, 2mutBC and 2mutABC, observed at 24 hpf. White arrow points to r6/7 boundary. A: Embryo injected with construct 2mutAB. Embryo shows anteriorized expression. B: Embryo injected with construct 2mutBC. Embryo shows r6/7 expression. C: Embryo injected with construct 2mutABC. Embryo shows anteriorized expression.

Figure 5-7: Dorsal view of embryos injected with constructs 2mutAB and 2mutBC, observed at 24hpf.
A: Embryo injected with construct 2mutAB. Embryo shows anteriorized expression. **B:** Embryo injected with construct 2mutBC. Embryo shows *r6/7* expression. **C:** Embryo injected with construct 2mutABC. Embryo shows anteriorized expression. Note: the embryo is different from the one shown in Figure 5-6C.

Figure 5-8: Lateral view of embryos injected with constructs 2mutABCG, 7mutABCG, 2mutRARE and 7mutRARE, observed at 24 hpf. White arrow points to r6/7 boundary A: Embryo injected with construct 2mutABCG. Embryo shows r6/7 expression. B: Embryo injected with construct 7mutABCG. Expression faintly delimits around the r6/7 boundary. C: Embryo injected with construct 2mutRARE. Embryo shows r6/7 expression. D: Embryo injected with construct 7mutRARE. No specific hindbrain expression was observed.

Figure 5-9: Dorsal view of embryos injected with construct 2mutABCG, 7mutABCG, 2mutRARE and 7mutRARE, observed at 24 hpf. A: Embryo injected with construct 2mutABCG. Embryo shows r6/7 expression. Note: the embryo is different from the one shown in Figure 5-8A. B: Embryo injected with construct 7mutABCG. Expression faintly delimits around the r6/7 boundary. C: Embryo injected with construct 2mutRARE. Embryo shows r6/7 expression. Note: the embryo is different from the one shown in Figure 5-8C. D: Embryo injected with construct 7mutRARE. No specific hindbrain expression was observed.

Figure 5-10: Lateral view of embryos injected with constructs 7mutABCG and 7mutRARE, observed at 24 hpf. A: Embryo injected with construct 7 mutABCG. Image was taken with a 60 x objective. Strong expression was observed in the tail region. B: Embryo injected with construct 7mutRARE. Image was taken with 48 x objective. Expression was observed in the tail region.

Constructs	Embryos analyzed	No. of embryos with r6/7 expression	No. of embryos with altered hindbrain expression	Expression level					
				+	++	+++	++++	+++++	Ave
Construct 2	12	11	Posteriorized: 1 (slightly)	0	3	6	2	1	3.1
Construct 2mutA	9	6	Anteriorized: 3	0	0	7	1	1	3.3
Construct 2mutB	10	1	Anteriorized: 9	0	1	3	2	4	3.9
Construct 2mutC	11	8	Anteriorized: 3	1	3	5	2	0	2.7
Construct 2mutG	15	12	Anteriorized: 1 Posteriorized: 2	0	4	8	3	0	2.9
Construct 2mutAB	14	2	Anteriorized: 12	0	0	3	3	8	4.4
Construct 2mutAC	15	8	Anteriorized: 7	0	0	0	3	12	4.8
Construct 2mutBC	11	9	Anteriorized: 2	0	3	6	2	0	2.9
Construct 2mutABC	9	1	Anteriorized: 8	0	0	1	1	7	4.7
Construct 2mutABCG	9	8	Anteriorized: 1 (slightly)	0	0	2	1	6	4.4
Construct 2mutRARE	10	7	Anteriorized: 3	0	1	3	1	5	4.0
Construct 7mutRARE	8	0	0	NA	NA	NA	NA	NA	NA
Construct 7mutABCG	8	6	No delimitation: 2	0	0	6	2	0	3.3

Table 5-1: Summary of hindbrain expression patterns and expression levels of injected embryos.

5.2 Discussion

In the P19 study (**chapter 3**), which involved the study of the murine *Hoxd4* neural enhancer, I speculated that *Hoxd4* expression would be anteriorized *in vivo* if individual sites A, B, C and G and compound sites A-C were to be mutated and the expression would be posteriorized if compound sites A-C and G were to be mutated. However, this was not observed when similar mutations were made in the zebrafish *hoxd4a* elements and tested *in vivo*. In the *in vivo* study in zebrafish embryos, anteriorized *hoxd4a* expression was generally observed in embryos injected with constructs 2mutB and 2mutABC, and r6/7 wild type expression was generally observed in embryos injected with constructs 2mutA, 2mutC, 2mutG and 2mutABCG. Perhaps the only consistent result from both the P19 study and *in vivo* study in zebrafish embryos is that mutation of the RARE (i.e. constructs 1mutRARE and 7mutRARE) abolished *Hoxd4/hoxd4a* expression. Along with the lab's previous demonstration that this same RARE is strictly required for expression in the mouse embryo, this suggests that the RARE is important for the function of the 3' neural enhancer regardless of the model system or species. Since the promoter region used for the P19 study does not contain a true promoter, this perhaps explains why most results obtained from the P19 study vary from those obtained from the *in vivo* study in zebrafish embryos. Perhaps more important is my discovery that the major activity directing r6/7 hindbrain expression lies not 3' to gene, as in murine *Hoxd4*, but rather 5' to *hoxd4a* in a discrete region just upstream of the coding region. This suggests a radical divergence in assignment of regulatory functions among enhancer regions in the two species.

Except for the less clear delimitation at the r6/7 boundary observed with constructs 7 (**Figs. 4-22 and 4-23**) and 7mutABCG (**Figs. 5-8B, 5-9B and 5-10A**), the hindbrain expression patterns observed with constructs 2 (**Figs. 4-12 and 4-13**) and 2mutABCG (**Figs. 5-8A and 5-9A**) are similar to those observed with constructs 7 and 7mutABCG in zebrafish embryos. Hence, this suggests that all observations on the mutants in the context of the P2-ATG region are reliable, even in the presence of the 5' neural enhancer. Alteration of the expression pattern in embryos injected with some mutants (i.e. constructs 2mutB, 2mutAB, 2mutAC and 2mutABC) suggests that the conserved sites play a role in the function of the *hoxd4a* 3' neural enhancer during its interaction with the 5' neural enhancer, in the delimitation of the r6/7 boundary. Expression beyond the r6/7 border observed with constructs 2mutB, 2mutAB and 2mutABC suggests that a repressor binds to site B. The r6/7 expression observed with construct 2mutABCG may imply that sites D, E,

F and RARE are adequate for the function of the 3' neural enhancer and an activator possibly binds to site G to counteract the repression effect at site B. Though sites D, E, F and RARE have been previously tested in mouse embryos (Nolte et al., 2003; Nolte et al., 2006), it will be nice to confirm their importance in zebrafish embryos. In support of these arguments, embryos injected with construct 2mutBC should have anteriorized expression and embryos injected with construct 2mutAC should have r6/7 expression. However, I observed r6/7 expression in embryos injected with construct 2mutBC and anteriorized expression in 50% of the expressing embryos injected with construct 2mutAC. This implies that a repressor binds to site A and works in place of the repressor at site B when it is absent. Further experimentation is required to confirm the importance of sites A, B and G.

It has been shown in other papers that *hoxb4a* expression is naturally mosaic and GFP expression under the control of the endogenous *hoxb4a* promoter is asymmetric across the midline (Hadrys et al., 2006; Ma et al., 2009). The neuroepithelial midline is established by the symmetrical deposition of daughter cells on each side of the midline between 14 hpf and 18 hpf. The daughter cells then extend their cell bodies until they traverse the full width of the neuroepithelium from apical to basal surfaces (Tawk et al., 2007). For asymmetrical expression across the midline to happen, there are three possibilities- 1) GFP expression is turned on in the mother cell before mid-line crossing and then turned off in one or both daughter cells, 2) GFP expression is turned on in one of the two daughter cells after mid-line crossing, or 3) late integration of the injected transgene. It has been observed in the figures shown in **Chapter 4** that *hoxd4a* expression is mosaic too. Therefore, it will be of interest in our future work to find out whether hypothesis 1, 2 or 3 is correct. If hypothesis 3 is ruled out, we can then proceed to find out the effects of the mutations and RA (due to the presence of RARE) on the asymmetric expression in daughter cells. This will be studied by tracking the GFP expression under the control of *hoxd4a* endogenous promoter in the zebrafish embryos injected with construct 2 and its derivatives, starting from 14 hpf.

Chapter 6: The role of Kr/Val in regulating *Hoxd4/hoxd4a* neural expression

6.1 Knockdown of *val* anteriorizes *hoxd4a* expression

Studies have shown that *kr* mutation affects *Hoxd4* expression in the mouse (McKay et al., 1994). Loss of *val* has been shown to result in inappropriately anteriorized *hoxb4* expression in zebrafish (Prince et al., 1998) but a similar response has not been shown for *hoxd4a*. Since *hoxb4* and *hoxd4a* belong to the same paralog group, they are highly related. Therefore, it is possible that *val* also affects *hoxd4a* expression. To determine whether *val* played a role in setting the anterior boundary of *hoxd4a* expression in the embryonic hindbrain of zebrafish, I employed translation-blocking morpholinos directed against the *val* mRNA start codon. Knockdown of *val* expression using 4 ng of translation-blocking AUG-targeting *val* morpholino per embryo anteriorized the r6/7 boundary by almost one rhombomere (**Fig. 6-1A**), leaving a thin indistinct gap between the *krox20* r5 expression domain and that of *hoxd4a*. By contrast, injection of control morpholino did not interfere with the establishment of r6 territory devoid of *hoxd4a* expression.

To check the efficiency of the morpholino knockdown, proteins were extracted from ~100 control and ~ 100 *val* morphants to perform a Western blot with anti-MafB antibody. The Val protein was not detected in the extracts of either the control or *val* morphants, perhaps because *val* is only expressed in a small region of the zebrafish embryo (r5 and r6).

Subsequently, the efficacy of the *val* morpholino to block translation of *val* specifically was analyzed using an *in vitro* transcription/translation (TNT) assay. Eight µl of TNT reaction was loaded into each lane. The *val* translation-blocking morpholino specifically blocked Val protein synthesis (**Fig. 6-1B**) but not that of a MEIS1A control (**Fig. 6-1C**). The expected sizes of Val and MEIS1A are ~ 39 kDa and ~ 43 kDa respectively. Hence, the morpholino efficiently blocks Val protein synthesis.

To verify that *val* knockdown is specific, a 5'UTR-targeting *val* morpholino was used. The same phenotype was observed in 10/10 embryos injected with 4 ng of morpholino injected per embryo (**Fig. 6-1D**).

Figure 6-1A: Images of 24 hpf *val* morphants (right) in comparison with 24 hpf control morphants (left) at 60x magnification. All had a dosage of 4 ng of morpholinos. These embryos were probed with 6.25 ng of *krox20* probe and 43.75 ng of *hoxd4a* probe. The gap between the *krox20* r5 band and *hoxd4a* r6/7 boundary is less distinct compared to the control morphants. **B:** Western blot of TNT of Val protein, exposed for 10 sec. Eight µl of TNT reaction was loaded into each lane on a 15% gel. Val protein synthesis is affected by *val* morpholino. Lane 1: negative control (TNT of water). Lane 2: TNT of 1 µg of plasmid containing *val* sequence. Translated Val protein is indicated by the white arrow. Lane 3: TNT of 1 µg of plasmid containing *val* sequence and 1 µg of *val* translation-blocking morpholino. **C:** Western blot of TNT of MEIS1A protein, exposed for 5 sec. Eight µl of TNT reaction was loaded into each lane on a 15% gel. MEIS1A protein synthesis is not affected by *val* morpholino. Lane 1: TNT of 1 µg of plasmid containing *MEIS1A* sequence. Lane 2: TNT of 1 µg of plasmid containing *MEIS1A* sequence and 1 µg of *val* translation-blocking morpholino. Translated MEIS1A protein is indicated by the black arrow. **D:** Dorsal view of embryos injected with 5' UTR-targeting *val* morpholino at 60 x magnification. Like the embryos injected with AUG-targeting *val* morpholino (Fig. 6-1A), the gap between the *krox20* r5 band and *hoxd4a* r6/7 boundary is less distinct, compared to the gap between the r3 and r5 band.

6.2 MafB does not bind to sites F and G of the 3' neural enhancer

Sequence alignment (**Fig. 1-21**) has shown that conserved sites F and G of the 3' neural enhancer show weak homology to the consensus maf recognition sequence (TGCTGANT/CCNGNN) (Yang and Cvekl, 2007). Given my demonstration that *val* knockdown anteriorizes *hoxd4a* expression, I hypothesized that *val* binds to *hoxd4a* neural enhancer(s).

Alignment of murine Kr, human MafB and zebrafish Val protein sequences using CLUSTALW shows that Kr is 97% homologous to MafB and 69% homologous to Val; MafB is 67% homologous to Val. The DNA binding domain of Kr is 99% homologous to MafB and 81% homologous to Val; the MafB DNA binding domain is 80% homologous to that of Val (**Fig. 6-2**). Hence, due to their high homology, MafB can be used in place of Kr to determine its binding to sites F and G.

Kreisler	MAAELSMGQELPTSPPLAMEYVNDFFLLKFVDVKKEPLGRAERP-GRPCTRLQPAGSVSSTP	59
MafB	MAAELSMGPGLPTSPPLAMEYVNDFFLLKFDVKKEPLGRAERP-GRPCTRLQPAGSVSSTP	59
valentino	MSADLAMGPGLPTSPPLALEYVNDFFLDMKFEVKEMAHAHRANIRQCNRQLPPQS SVSSTP *:.:*: * *****:*****:*:****:. . :*. * *. **** *	60
Kreisler	LSTPCSSVPSPSFSPTEPKTHLEDLYWMAS--NYQQMNPEALNLT PEDAVEALIGS---	114
MafB	LSTPCSSVPSPSFSPTEQKTHLEDLYWMAS--NYQQMNPEALNLT PEDAVEALIGS---	114
valentino	ISTPCSSVPSPSFSPTEQKNHLEELYWMPSGAYPQQIDPQTLSLTPEDAVEALIGATAH :***** ** *:***:****.* **:~::~~.*****:	120
Kreisler	-HPVP---QPLQS--FDGFSAHHHHHHHHHP-----HPHHGYPGAGVTH--DDL G--	156
MafB	-HPVP---QPLQS--FDSFRGAHHHHHHHHHP-----HPHHAYPGAGVAH--DEL G--	156
valentino	GHPPPP HVQQQLQAGAFDGYRGAHHHHGHAAQQQQQQPQH HHHHQHYGAIPHH PDDLSGH ** * * **: ~::~~.***** * : * ** : ..:~* **:.	180
Kreisler	----QHAHPHHHHHHHQASPFPSSAASPAQQLPTSHPGPGPHATAAATAAGNG--SVE	208
MafB	----PHAHPHHHHHHHQASPFPSSAASPAQQLPTSHPGPGPHATASAATAAGNG--SVE	208
valentino	PGAHGHH PHHHHHHHHSQDPDSPSPTSPEQLHRRHHHHHHHPGHGHPGGQH HGVGGLNVE *.* *****. * ..:::** * * **.. . * * ..*	240
Kreisler	DRFSDDQLVSMSVRELNRHLRGFTKDEVIRLKQKRRTLKNRGYA QSCRYKRVQ QKHLEN	268
MafB	DRFSDDQLVSMSVRELNRHLRGFTKDEVIRLKQKRRTLKNRGYA QSCRYKRVQ QKHLEN	268
valentino	DRFSDDQLVTMSVRELNRHLRGFTKDEVIRLKQKRRTLKNRGYA QSCRKRVQ QKHLEN *****:*****:*****:*****:*****:***** *	300
Kreisler	EKTQLIQQEVLKQEVSR LARERDAYKV KCEKLANS-GFREAGSTD SPSSPEFFL	323
MafB	EKTQLIQQEVLKQEVSR LARERDAYKV KCEKLANS-GFREAGSTD SPSSPEFFL	323
valentino	EKTQLINQVEVLKQEI NRLARERDAYKLKCEKLTGANGFREAGSTDNP SSP EFM *****:*****:*****:*****:.....: *****:*****:	356

I expressed MafB as a GST fusion protein in *E. coli* and purified it on a GST column. The expected GST-tagged protein size is ~ 38.4 kDa. The MafB band is the dominant band. Some fine bands of lower molecular weights were observed in the gel stained with Comassie blue and may have been residual bacterial proteins because these were also observed in the uninduced, untransformed BL21 sonicate in lane 2 (**Fig. 6-3A**). The identity of GST-tagged human MafB was confirmed with rabbit anti-GST primary antibody and goat anti-rabbit secondary antibody (**Fig. 6-3B**). The blot has a few bands of smaller size, probably due to a minute quantity of degraded or truncated MafB protein which is not detected by Comassie blue staining.

The annealing efficiency of the 34-mer consensus and 38-mer sites F-G probes was assessed through ethidium bromide staining of a 15% nondenaturing polyacrylamide gel (**Fig. 6-4**). I compared the annealing efficiencies in which both strands were unlabeled vs. conditions in which one of the two strands was biotin labeled. The higher molecular weight band represents annealed double-stranded oligo and the lower molecular weight band represents unannealed single-stranded oligo. All probes were efficiently annealed.

0.25 µg of MafB protein was used for EMSA using chemiluminescent detection. BL21 cell lysate was used as a negative control to eliminate false positive results because the MafB protein used was not 100% pure. Consensus MafB probe and sites F-G probe bound weakly to BL21 cell lysate (lanes 2 and 7). A shift was observed in lane 3 indicating the binding of MafB to the consensus MafB probe. This shift disappeared when competed with unlabelled consensus MafB probe (lane 4). The shift did not disappear when competed with unlabelled consensus sites F-G probe (lane 5), implying that MafB protein does not bind to sites F and G. This result was confirmed in reactions of MafB protein and sites F-G probe which failed to yield a shifted complex (lane 8) (**Fig. 6-5**).

Figure 6-3 A: lane 1: PageRuler Plus Prestained Protein Ladder (5 µl) ; lane 2: BL21 sonicate (5 µl) ; lane 3: eluate (purified MafB) (5 µl). B: Western blot of 5 µl of purified MafB protein.

Figure 6-4: Assessment of oligo annealing efficiency through ethidium bromide staining of a 15% nondenaturing polyacrylamide gel. Lane 1: 100 bp ladder, lane 2: single-stranded consensus probe (10 pmol), lane 3: biotin-labelled double-stranded consensus probe (10 pmol), lane 4: unlabelled double-stranded consensus probe (10 pmol), lane 5: single-stranded sites F-G probe (10 pmol), lane 6: biotin-labelled double-stranded sites F-G probe (10 pmol), lane 7: unlabelled double-stranded sites F-G probe (10 pmol) and lane 8: 100 bp ladder.

Figure 6-5: Use of EMSA to assess the binding of MafB protein to sites F and G of the *Hoxd4* neural enhancer. BL21 cell lysate refers to control bacterial cell lysate obtained by induction of BL21 *E.coli* lacking the MafB expression vector was used as a negative control. Cold competitors refer to unlabelled probes. White arrowhead points to the shifted complex due to the interaction between MafB protein and consensus MafB probe.

6.3 Val does not act through the neural enhancer(s)

Since MafB does not bind to sites F and G of the 3' neural enhancer, I asked whether Val acts through other *hoxd4a* neural enhancers. If so, then a transgene driven by *hoxd4a* neural enhancers should show anteriorized expression in response to *val* knockdown, just as I have shown for the endogenous gene. Construct 2 was co-injected with 4 ng AUG-blocking *val* morpholino into zebrafish embryos at the 1-cell stage. Its expression pattern was compared with two controls, i) embryos injected with construct 2 alone (c.f. **Chapter 4 Figs. 4-12 and 4-13**) and ii) embryos injected with construct 2 and control morpholino at 24 hpf (**Figs. 6-6A and 6-7A**). Twelve embryos injected with construct 2 and control morpholino were randomly selected, photographed and analyzed at 24 hpf. Of these, 10/12 embryos have GFP expression delimiting at the r6/7 boundary, while 1/12 embryos has a slightly posteriorized expression and 1/12 embryos has a slightly anteriorized expression. Eleven embryos injected with construct 2 and *val* morpholino were randomly selected, photographed and analyzed at 24 hpf. 10/11 embryos have GFP expression delimiting at the r6/7 boundary (**Figs. 6-6B and 6-7B**) while 1/11 embryos has a slightly posteriorized expression. These results suggest that the expression of construct 2 is not influenced by *val* knockdown.

To confirm that *val* was efficiently knocked down in the embryos injected with construct 2 and *val* morpholino, they were fixed and hybridized with *krox20* and *hoxd4a* probes. Their expression pattern (**Fig. 6-8**) recapitulated the observation with the *val* morphants shown in **Fig. 6-1A**. Together, these results suggest that Val does not bind to either of the *hoxd4a* neural enhancers.

Figure 6-6: Embryos injected with A) construct 2 and control morpholino and B) construct 2 and *val* morpholino (lateral view at 80 x magnification). White arrowhead points to r6/7 boundary. GFP expression delimits at r6/7 boundary in both cases.

Figure 6-7: Embryos injected with A) construct 2 and control morpholino and B) construct 2 and *val* morpholino (dorsal view at 80 x magnification). A: GFP expression delimits at r6/7 boundary. Note: Embryos shown in Figures 6-6A and 6-7A are different. B: GFP expression delimits at r6/7 boundary. Note: Embryos shown in Figures 6-6B and 6-7B are different.

Figure 6-8: *In situ* hybridization of embryo injected with construct 2 and *val* morpholino, using *krox20* and *hoxd4a* probes. Image was taken at 60 x magnification. *val* was successfully knocked down in embryos injected with construct 2 and *val* morpholino.

6.4 Discussion

Different combinations of Hox proteins generated by the nested domains of expression determine a cell's identity. Misexpression of these genes during embryonic development results in homeotic transformations or malformations. Different muscles of the head are controlled by a unique set of motoneurons, given rise by each rhombomere in the developing hindbrain. Rhombomere identity is important for controlling our facial expressions and mouth functions (Sanes et al., 2006).

Hoxd4 expression delimits at r6/7 boundary of the developing hindbrain. R6 and 7 give rise to the neurons of glossopharyngeal nerve, which controls swallowing (Sanes et al., 2006). Studies have shown that *kr* mutation anteriorizes *Hoxd4* expression in the mouse (McKay et al., 1994) and loss of Val results in inappropriately anteriorized *hoxb4* expression in zebrafish (Prince et al., 1998). I showed here that knockdown of *val* results in anteriorization of *hoxd4a* expression by almost one rhombomere. This implies that Val is a negative regulator of *hoxd4a* in the zebrafish embryonic hindbrain, establishing a broadly conserved role for Kr/Val in the negative regulation of 4th group paralogs.

Sequence alignment (**Fig. 1-21**) has shown that conserved sites F and G of the 3' neural enhancer showed weak homology to the consensus maf recognition sequence (Yang and Cvekl, 2007). However, I was unable to obtain results supporting the binding of Val to the 5' or 3' neural enhancers, either by EMSA or in the co-injection of construct 2 and *val* morpholino. Since only 570 bp of the 3' neural enhancer was used in the test, it is possible that Val binds to some other region that is upstream/downstream of the 570 bp fragment tested. To check whether Kr/Val binds to the neural enhancer(s), the best method is to use chromatin immunoprecipitation (ChIP). In this method, proteins are cross-linked to DNA in cell lysate. The DNA-protein complexes are sheared by sonication and the DNA fragments associated with Kr/Val are immunoprecipitated with Kr/Val antibody. The associated DNA fragments are purified and analyzed by sequencing. I attempted to use this method but due to lack of efficient commercially available Kr/Val/MafB antibody, I was unable to pull down the DNA-protein complexes. Therefore, there is a need to produce this antibody to test whether Kr/Val really binds to the neural enhancer(s). Identification of transcription factors that bind to the neural enhancer(s) helps us understand how *Hoxd4/hoxd4a* expression is regulated.

Chapter 7 Conclusion/Future work

A model of the *hox4a* transcription unit is given in **Fig. 7-1**. Previous studies have shown that murine *Hoxb4* and *Hoxd4*, being in the same paralog group, are closely related to each other in their neural regulation and both bear 3' neural enhancers involved in setting the r6/7 anterior boundary (Morrison et al., 1997). However, I showed in this thesis that this relationship holds only in the mouse context. In fact, for zebrafish *hoxd4a*, there is a 5' neural enhancer that is responsible for driving r6/7 expression, while the 3' neural enhancer instead plays a complementary role to the 5' neural enhancer (c.f **Chapter 4**, a comparison between mouse *Hoxd4* and zebrafish *hoxd4a* transcription units is shown in **Fig. 7-2**). Although previous studies from the Featherstone lab indicate that no 5' neural enhancer is present in mouse *Hoxd4* (Zhang et al., 1997), a 5' neural enhancer has been reported in human *HOXD4* (Morrison et al., 1996) and two 5' neural enhancers have been reported in mouse *Hoxb4* (Sharpe et al., 1998). Thus, interspecies comparisons reveal a strong tendency for conservation of neural regulatory mechanisms, but with some clear divergence within zebrafish *hoxd4a* as shown here. The regulatory elements that I have identified (the 5' neural enhancer that lies between -420 and -602 and the 5' repressor element that lies between -3368 and -3469) require further verification and characterization as noted below. An eventual definition of the contribution of individual transcription factors to the function of *hoxd4a* enhancer elements is rendered more complicated by the fact that the zebrafish genome is more complex than the mouse genome due to an ancestral whole genome duplication event.

To gain confidence about the presence of these regulatory elements, we need to repeat the experiments described in **Chapter 4** and observe more zebrafish embryos. To test the 5' repressor function that is possibly dependent on the 5' RARE, we can mutate the RARE or delete the entire element to check if there is an increase in *hoxd4a* expression. To test the 5' neural enhancer function, we can isolate CR3 (where the 5' neural enhancer lies) and test its activity independently in *in vitro* reporter assay using cell culture and *in vivo* study by microinjecting transposons containing these elements into zebrafish embryos (the methods described and used earlier in this thesis). However, it poses a challenge to

Figure 7-1: Model of *hoxd4a* transcription unit (not drawn to scale). Blue boxes: regulatory regions, white boxes: non-coding exonic sequence; pink boxes; coding sequence; black box: homeobox. *hoxd4a* neural expression is driven by two promoters – P2 and P3 and regulated by two neural enhancers – 5' NE and 3' NE and a repressor – R.

Figure 7-2: Comparison between mouse *Hoxd4* and zebrafish *hoxd4a* transcription units (not drawn to scale). ARE: autoregulatory element; CR1/2/3: conserved region 1/2/3; ME = mesodermal enhancer and R = repressor element; miR-10b: microRNA-10b; NE = neural enhancer; P1/2/3: promoter 1/2/3; RARE: retinoic acid response element.

determine the 5' and 3' ends that encompass the full activity. To counter this, we need to make various deletion constructs to determine the minimal region required for the regulatory activities. Moreover, since we do not know whether these elements require the presence of some other elements to function, there is always a possibility of misinterpreting the results. A complementary method – locating the transcription factors binding to the *hoxd4a* transcription unit to determine the regulatory elements – has been employed in some transcription regulation labs. However, this poses a challenge too because the transcription factors may be involved in various processes and we need to know the transcription factors that bind to the regulatory elements first. To identify the transcription factors bound to these enhancers, we can do a preliminary study in zebrafish embryos by knocking down the genes that have been found to play a role in *Hoxd4/hoxd4a/HOXD4* regulation and doing an *in situ* hybridization with *hoxd4a* probe. This approach has already been used successfully by the Featherstone and Postlethwait labs to confirm the role of Pax6 in the regulation of *hoxd4a* neural expression (Nolte et al., 2006). After this, we can perform ChIP experiments to find out whether these transcription factors bind and where they bind to the *hoxd4a* transcription unit. This is only feasible if we have the antibodies for the transcription factors in question. The dissection of enhancer function at both the gene level and the protein level should ensure that our results are accurate.

It had been proposed that mouse *Hoxd4* transcription was driven by two promoters, P2 located 5.2 kb upstream from the coding region and P1 located 1.1 kb upstream from the coding region. P1 was proposed to be used in anterior regions up to the r6/7 expression boundary (Folberg et al., 1997). However, it has been shown subsequently that murine P1 is not a true promoter (Phua et al., 2011) and I showed here in the *in vivo* study in zebrafish embryos that zebrafish P1 does not have promoter activity either. In fact, the promoter used in anterior regions up to the r6/7 expression boundary (P3) lies in the 200 bp region upstream of the ATG start site (c.f. **Chapter 4**). To confirm this, we can perform rapid amplification of cDNA ends with the *hoxd4a* transcripts in order to map their 5' ends of *hoxd4a* transcripts precisely.

Previous sequence comparison between mouse and zebrafish *Hoxd4/hoxd4a* 3' neural enhancers revealed eight blocks of homology, denoted sites A-G and RARE. Therefore, it has been hypothesized that these evolutionarily conserved sites are important in *Hoxd4/*

hoxd4a 3' neural enhancer function. Sites D, E and F and the RARE were previously characterized in the Featherstone lab (Zhang et al., 2000; Nolte et al., 2003; Nolte et al., 2006). Though I have shown that the zebrafish 3' neural enhancer plays a complementary role to 5' neural enhancer, the altered expression patterns observed in the context of the P2-ATG region (constructs 2mutB, 2mutAB, 2mutAC and 2mutABC) show that these conserved sites are important for the r6/7 expression (c.f. **Chapter 5**). I herein propose that repressors bind to sites A and B. The activity of the repressor that binds to site A is revealed when binding of the repressor to site B is blocked. Giving careful consideration to my data (namely, anteriorized expression dependent on site G; compare construct 2mutABC to construct 2mutABCG), I further propose that an activator binds to site G. To identify the regulatory factors binding to sites A, B and G, we can employ reverse ChIP which involves incubating a DNA probe corresponding to a specific chromatin locus, isolating the DNA fragment together with the associated proteins and identifying these proteins using mass spectrometer (Dejardin and Kingston, 2009; Mittler et al., 2009; Rusk, 2009). However, for this method to work, we have to optimize the condition such that the signal-to-noise ratio is high.

It has been shown in other papers that *hoxb4* expression is naturally mosaic and YFP expression under the control of the endogenous *hoxb4* promoter is asymmetric across the neuroepithelial midline (Hadrys et al., 2006; Ma et al., 2009). I observed here that GFP expression under the control of *hoxd4a* cis-acting elements is also mosaic (c.f. **Chapters 4 and 5**). Therefore, asymmetric Hox expression is not artifactual but a true biological process whose mechanism must be able to act through the regulatory elements present in my *hoxd4a* transgene constructs. Since *hoxb4* and *hoxd4a* are paralogous and display mosaic expression, this may bring about an interesting situation in which neuronal daughter cells after mid-line crossing may express (a) both *hoxb4* and *hoxd4a*, (b and c) one or the other of these genes, or (d) neither Hox gene. Such differential *Hox* expression could delegate specific neurons to specific tasks within rhombomeres 7 and 8, such as may be useful for the formation of particular r7/8-derived structures (e.g. the inferior olive, vagus nerve) or specific functions mediated by these rhombomeres including innervation of the pectoral fin bud, pacemaker oscillations and respiration (Bass and Baker, 1997; Punnamoottil et al., 2008).

For asymmetrical expression across the midline to happen, there are two possibilities- 1) GFP expression is turned on in the mother cell before mid-line crossing and then turned off in one or both daughter cells, or 2) GFP expression is turned on in one of the two daughter cells after mid-line crossing. Hence, it will be of interest in our future work to find out whether hypothesis 1 or 2 is correct, as well as the effects of the mutations and retinoic acid (due to the presence of RARE) on the asymmetric expression in daughter cells. This will be studied by tracking GFP expression under the control of the *hoxd4a* endogenous promoter in zebrafish embryos injected with construct 2 and its derivatives, starting from 14 hpf, using the same constructs as described in **Chapters 4 and 5**.

Studies have shown that *kr* mutation affects *Hoxd4* expression in the mouse (McKay et al., 1994) and loss of *val* results in inappropriately anteriorized *hoxb4* expression in zebrafish (Prince et al., 1998). I showed here that *val* likewise represses *hoxd4a* expression in zebrafish embryos. Though preliminary observations in electrophoretic mobility shift assay and coinjection of the *val* morpholino with construct 2 were unable to provide evidence that *val* acts directly upon the neural enhancer(s) (c.f. **Chapter 6**), it is possible that *val* binds to some other region that is upstream/downstream of the 570 bp 3' neural enhancer fragment tested. Therefore, further investigation using ChIP has to be performed.

References

- Akbari, O. S., Bae, E., Johnsen, H., Villaluz, A., Wong, D. and Drewell, R. A. (2008) 'A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the *Drosophila* embryo', *Development* 135(1): 123-31.
- Amirthalingam, G. S., Howard, S., Alvarez, S., de Lera, A. R. and Itasaki, N. (2009) 'Regulation of Hoxb4 induction after neurulation by somite signal and neural competence', *BMC developmental biology* 9: 17.
- Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R. and Brenner, S. (1995) 'Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*', *Proceedings of the National Academy of Sciences of the United States of America* 92(5): 1684-8.
- Argiropoulos, B. and Humphries, R. K. (2007) 'Hox genes in hematopoiesis and leukemogenesis', *Oncogene* 26(47): 6766-76.
- Artner, I., Blanchi, B., Raum, J. C., Guo, M., Kaneko, T., Cordes, S., Sieweke, M. and Stein, R. (2007) 'MafB is required for islet beta cell maturation', *Proceedings of the National Academy of Sciences of the United States of America* 104(10): 3853-8.
- Awgulewitsch, A. (2003) 'Hox in hair growth and development', *Die Naturwissenschaften* 90(5): 193-211.
- Bass, A. H. and Baker, R. (1997) 'Phenotypic specification of hindbrain rhombomeres and the origins of rhythmic circuits in vertebrates', *Brain, behavior and evolution* 50 Suppl 1: 3-16.
- Bastien, J. and Rochette-Egly, C. (2004) 'Nuclear retinoid receptors and the transcription of retinoid-target genes', *Gene* 328: 1-16.
- Behringer, R. R., Crotty, D. A., Tennyson, V. M., Brinster, R. L., Palmiter, R. D. and Wolgemuth, D. J. (1993) 'Sequences 5' of the homeobox of the Hox-1.4 gene direct

tissue-specific expression of lacZ during mouse development', *Development* 117(3): 823-33.

Blanchi, B., Kelly, L. M., Viemari, J. C., Lafon, I., Burnet, H., Bevendut, M., Tillmanns, S., Daniel, L., Graf, T., Hilaire, G. et al. (2003) 'MafB deficiency causes defective respiratory rhythmogenesis and fatal central apnea at birth', *Nature neuroscience* 6(10): 1091-100.

Blank, V. and Andrews, N. C. (1997) 'The Maf transcription factors: regulators of differentiation', *Trends in Biochemical Sciences* 22(11): 437-441.

Brend, T., Gilthorpe, J., Summerbell, D. and Rigby, P. W. (2003) 'Multiple levels of transcriptional and post-transcriptional regulation are required to define the domain of Hoxb4 expression', *Development* 130(12): 2717-28.

Brodbeck, S. and Englert, C. (2004) 'Genetic determination of nephrogenesis: the Pax/Eya/Six gene network', *Pediatr Nephrol* 19(3): 249-55.

Cardoso, W. V., Mitsialis, S. A., Brody, J. S. and Williams, M. C. (1996) 'Retinoic acid alters the expression of pattern-related genes in the developing rat lung', *Dev Dyn* 207(1): 47-59.

Chen, F. and Capecchi, M. R. (1999) 'Paralogous mouse Hox genes, Hoxa9, Hoxb9, and Hoxd9, function together to control development of the mammary gland in response to pregnancy', *Proceedings of the National Academy of Sciences of the United States of America* 96(2): 541-6.

Chen, S. K., Tvrdik, P., Peden, E., Cho, S., Wu, S., Spangrude, G. and Capecchi, M. R. (2010) 'Hematopoietic origin of pathological grooming in Hoxb8 mutant mice', *Cell* 141(5): 775-85.

Chopra, V. S., Hong, J. W. and Levine, M. (2009) 'Regulation of Hox Gene Activity by Transcriptional Elongation in Drosophila', *Curr Biol*.

Cordes, S. P. and Barsh, G. S. (1994) 'The Mouse Segmentation Gene Kr Encodes a Novel Basic Domain Leucine-Zipper Transcription Factor', *Cell* 79(6): 1025-1034.

Cribbs, D. L., Pultz, M. A., Johnson, D., Mazzulla, M. and Kaufman, T. C. (1992) 'Structural Complexity and Evolutionary Conservation of the Drosophila Homeotic Gene Proboscipedia', *Embo Journal* 11(4): 1437-1449.

Cusic, A. M. and Dagg, C. P. (1985) 'Spontaneous and retinoic acid-induced postaxial polydactyly in mice', *Teratology* 31(1): 49-59.

Daftary, G. S. and Taylor, H. S. (2006) 'Endocrine regulation of HOX genes', *Endocr Rev* 27(4): 331-55.

Dejardin, J. and Kingston, R. E. (2009) 'Purification of proteins associated with specific genomic Loci', *Cell* 136(1): 175-86.

Deschamps, J., van den Akker, E., Forlani, S., De Graaff, W., Oosterveen, T., Roelen, B. and Roelfsema, J. (1999) 'Initiation, establishment and maintenance of Hox gene expression patterns in the mouse', *International Journal of Developmental Biology* 43(7): 635-50.

Deschamps, J. and van Nes, J. (2005) 'Developmental regulation of the Hox genes during axial morphogenesis in the mouse', *Development* 132(13): 2931-42.

Di Rocco, G., Gavalas, A., Popperl, H., Krumlauf, R., Mavilio, F. and Zappavigna, V. (2001) 'The recruitment of SOX/OCT complexes and the differential activity of HOXA1 and HOXB1 modulate the Hoxb1 auto-regulatory enhancer function', *J Biol Chem* 276(23): 20506-15.

Doerksen, L. F., Bhattacharya, A., Kannan, P., Pratt, D. and Tainsky, M. A. (1996) 'Functional interaction between a RARE and an AP-2 binding site in the regulation of the human HOX A4 gene promoter', *Nucleic Acids Res* 24(14): 2849-56.

Dolle, P., Izpisua-Belmonte, J. C., Brown, J. M., Tickle, C. and Duboule, D. (1991) 'HOX-4 genes and the morphogenesis of mammalian genitalia', *Genes & development* 5(10): 1767-7.

Economides, K. D. and Capecchi, M. R. (2003) 'Hoxb13 is required for normal differentiation and secretory function of the ventral prostate', *Development* 130(10): 2061-9.

Erfurth, F. E., Popovic, R., Grembecka, J., Cierpicki, T., Theisler, C., Xia, Z. B., Stuart, T., Diaz, M. O., Bushweller, J. H. and Zeleznik-Le, N. J. (2008) 'MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression', *Proceedings of the National Academy of Sciences of the United States of America* 105(21): 7517-22.

Ferretti, E., Cambronero, F., Tumpel, S., Longobardi, E., Wiedemann, L. M., Blasi, F. and Krumlauf, R. (2005) 'Hoxb1 enhancer and control of rhombomere 4 expression: complex interplay between PREP1-PBX1-HOXB1 binding sites', *Molecular and Cellular Biology* 25(19): 8541-52.

Folberg, A., Kovacs, E. N. and Featherstone, M. S. (1997) 'Characterization and retinoic acid responsiveness of the murine Hoxd4 transcription unit', *J Biol Chem* 272(46): 29151-7.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. and Postlethwait, J. (1999) 'Preservation of duplicate genes by complementary, degenerative mutations', *Genetics* 151(4): 1531-45.

Friedmann, Y., Daniel, C. A., Strickland, P. and Daniel, C. W. (1994) 'Hox genes in normal and neoplastic mouse mammary gland', *Cancer Res* 54(22): 5981-5.

Frohman, M. A., Martin, G. R., Cordes, S. P., Halamek, L. P. and Barsh, G. S. (1993) 'Altered Rhombomere-Specific Gene-Expression and Hyoid Bone Differentiation in the Mouse Segmentation Mutant, Kreisler (Kr)', *Development* 117(3): 925-936.

Garin, E., Lemieux, M., Coulombe, Y., Robinson, G. W. and Jeannotte, L. (2006) 'Stromal Hoxa5 function controls the growth and differentiation of mammary alveolar epithelium', *Dev Dyn* 235(7): 1858-71.

Gaunt, S. J., Krumlauf, R. and Duboule, D. (1989) 'Mouse homeo-genes within a subfamily, Hox-1.4, -2.6 and -5.1, display similar anteroposterior domains of expression in the embryo, but show stage- and tissue-dependent differences in their regulation', *Development* 107(1): 131-41.

Gilbert, S. F. (2000) *Developmental biology*. Sunderland, Mass.: Sinauer Associates.

Gilthorpe, J., Vandromme, M., Brend, T., Gutman, A., Summerbell, D., Totty, N. and Rigby, P. W. (2002) 'Spatially specific expression of Hoxb4 is dependent on the ubiquitous transcription factor NFY', *Development* 129(16): 3887-99.

Godwin, A. R. and Capecchi, M. R. (1998) 'Hoxc13 mutant mice lack external hair', *Genes Dev* 12(1): 11-20.

Gong, K. Q., Yallowitz, A. R., Sun, H., Dressler, G. R. and Wellik, D. M. (2007) 'A Hox-Eya-Pax complex regulates early kidney developmental gene expression', *Molecular and Cellular Biology* 27(21): 7661-8.

Gonzalez, F., Duboule, D. and Spitz, F. (2007) 'Transgenic analysis of Hoxd gene regulation during digit development', *Developmental biology* 306(2): 847-59.

Goodman, F. R. (2002) 'Limb malformations and the human HOX genes', *American Journal of Medical Genetics* 112(3): 256-65.

Gosiengfiao, Y., Horvat, R. and Thompson, A. (2007) 'Transcription factors GATA-1 and Fli-1 regulate human HOXA10 expression in megakaryocytic cells', *DNA and cell biology* 26(8): 577-87.

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

Gould, A., Morrison, A., Sproat, G., White, R. A. and Krumlauf, R. (1997) 'Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns', *Genes & development* 11(7): 900-13.

Hadrys, T., Punnamoottil, B., Pieper, M., Kikuta, H., Pezeron, G., Becker, T. S., Prince, V., Baker, R. and Rinkwitz, S. (2006) 'Conserved co-regulation and promoter sharing of *hoxb3a* and *hoxb4a* in zebrafish', *Developmental biology* 297(1): 26-43.

Hernandez, R. E., Putzke, A. P., Myers, J. P., Margaretha, L. and Moens, C. B. (2007) 'Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development', *Development* 134(1): 177-87.

Hernandez, R. E., Rikhof, H. A., Bachmann, R. and Moens, C. B. (2004) '*vhnf1* integrates global RA patterning and local FGF signals to direct posterior hindbrain development in zebrafish', *Development* 131(18): 4511-20.

Ho, M. C., Schiller, B. J., Akbari, O. S., Bae, E. and Drewell, R. A. (2011) 'Disruption of the abdominal-B promoter tethering element results in a loss of long-range enhancer-directed Hox gene expression in *Drosophila*', *PLoS One* 6(1): e16283.

Horan, G. S., Kovacs, E. N., Behringer, R. R. and Featherstone, M. S. (1995) 'Mutations in paralogous Hox genes result in overlapping homeotic transformations of the axial skeleton: evidence for unique and redundant function', *Developmental biology* 169(1): 359-72.

Hostikka, S. L. and Capecchi, M. R. (1998) 'The mouse *Hoxc11* gene: genomic structure and expression pattern', *Mech Dev* 70(1-2): 133-45.

In der Rieden, P. M., Vilaspasa, F. L. and Durston, A. J. (2010) '*Xwnt8* directly initiates expression of labial Hox genes', *Dev Dyn* 239(1): 126-39.

Jave-Suarez, L. F., Winter, H., Langbein, L., Rogers, M. A. and Schweizer, J. (2002) '*HOXC13* is involved in the regulation of human hair keratin gene expression', *J Biol Chem* 277(5): 3718-26.

Juan, A. H. and Ruddle, F. H. (2003) 'Enhancer timing of Hox gene expression: deletion of the endogenous Hoxc8 early enhancer', *Development* 130(20): 4823-34.

Kawazoe, Y., Sekimoto, T., Araki, M., Takagi, K., Araki, K. and Yamamura, K. (2002) 'Region-specific gastrointestinal Hox code during murine embryonal gut development', *Dev Growth Differ* 44(1): 77-84.

Kessel, M. and Gruss, P. (1991) 'Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid', *Cell* 67(1): 89-104.

Kim, M. H., Shin, J. S., Park, S., Hur, M. W., Lee, M. O., Park, H. and Lee, C. S. (2002) 'Retinoic acid response element in HOXA-7 regulatory region affects the rate, not the formation of anterior boundary expression', *International Journal of Developmental Biology* 46(3): 325-8.

Kobrossy, L., Rastegar, M. and Featherstone, M. (2006) 'Interplay between chromatin and trans-acting factors regulating the Hoxd4 promoter during neural differentiation', *J Biol Chem* 281(36): 25926-39.

Langston, A. W. and Gudas, L. J. (1992) 'Identification of a retinoic acid responsive enhancer 3' of the murine homeobox gene Hox-1.6', *Mech Dev* 38(3): 217-27.

Lappin, T. R., Grier, D. G., Thompson, A. and Halliday, H. L. (2006) 'HOX genes: seductive science, mysterious mechanisms', *Ulster Med J* 75(1): 23-31.

LaRonde-LeBlanc, N. A. and Wolberger, C. (2003) 'Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior', *Genes & development* 17(16): 2060-72.

LaRosa, G. J. and Gudas, L. J. (1988) 'Early retinoic acid-induced F9 teratocarcinoma stem cell gene ERA-1: alternate splicing creates transcripts for a homeobox-containing protein and one lacking the homeobox', *Molecular and Cellular Biology* 8(9): 3906-17.

Levin, A. A., Sturzenbecker, L. J., Kazmer, S., Bosakowski, T., Huselton, C., Allenby, G., Speck, J., Kratzeisen, C., Rosenberger, M., Lovey, A. et al. (1992) '9-cis retinoic acid stereoisomer binds and activates the nuclear receptor RXR alpha', *Nature* 355(6358): 359-61.

Lewis, M. T. (2000) 'Homeobox genes in mammary gland development and neoplasia', *Breast Cancer Res* 2(3): 158-69.

Lovato, T. L., Nguyen, T. P., Molina, M. R. and Cripps, R. M. (2002) 'The Hox gene abdominal-A specifies heart cell fate in the Drosophila dorsal vessel', *Development* 129(21): 5019-27.

Lufkin, T. (2003) *Murine homeobox gene control of embryonic patterning and organogenesis*, Amsterdam ; Boston ; Oxford: Elsevier Science.

Ma, L. H., Punnamoottil, B., Rinkwitz, S. and Baker, R. (2009) 'Mosaic hoxb4a neuronal pleiotropism in zebrafish caudal hindbrain', *PLoS One* 4(6): e5944.

Maconochie, M., Krishnamurthy, R., Nonchev, S., Meier, P., Manzanares, M., Mitchell, P. J. and Krumlauf, R. (1999) 'Regulation of Hoxa2 in cranial neural crest cells involves members of the AP-2 family', *Development* 126(7): 1483-94.

Maconochie, M., Nonchev, S., Morrison, A. and Krumlauf, R. (1996) 'Paralogous Hox genes: function and regulation', *Annual review of genetics* 30: 529-56.

Maeda, Y., Dave, V. and Whitsett, J. A. (2007) 'Transcriptional control of lung morphogenesis', *Physiol Rev* 87(1): 219-44.

Magnusson, M., Brun, A. C., Miyake, N., Larsson, J., Ehinger, M., Bjornsson, J. M., Wutz, A., Sigvardsson, M. and Karlsson, S. (2007) 'HOXA10 is a critical regulator for hematopoietic stem cells and erythroid/megakaryocyte development', *Blood* 109(9): 3687-96.

Mai, T., Zan, H., Zhang, J., Hawkins, J. S., Xu, Z. and Casali, P. (2010) 'Estrogen receptors bind to and activate the HOXC4/HoxC4 promoter to potentiate HoxC4-mediated activation-induced cytosine deaminase induction, immunoglobulin class switch DNA recombination, and somatic hypermutation', *The Journal of biological chemistry* 285(48): 37797-810.

Mandeville, I., Aubin, J., LeBlanc, M., Lalancette-Hebert, M., Janelle, M. F., Tremblay, G. M. and Jeannotte, L. (2006) 'Impact of the loss of Hoxa5 function on lung alveogenesis', *Am J Pathol* 169(4): 1312-27.

Manzanares, M., Nardelli, J., Gilardi-Hebenstreit, P., Marshall, H., Giudicelli, F., Martinez-Pastor, M. T., Krumlauf, R. and Charnay, P. (2002) 'Krox20 and kreisler co-operate in the transcriptional control of segmental expression of Hoxb3 in the developing hindbrain', *The EMBO journal* 21(3): 365-76.

Marin, F. and Charnay, P. (2000) 'Hindbrain patterning: FGFs regulate Krox20 and mafB/kr expression in the otic/preotic region', *Development* 127(22): 4925-35.

Martin, R., Taylor, M. B., Krikun, G., Lockwood, C., Akbas, G. E. and Taylor, H. S. (2007) 'Differential cell-specific modulation of HOXA10 by estrogen and specificity protein 1 response elements', *J Clin Endocrinol Metab* 92(5): 1920-6.

Maves, L. and Kimmel, C. B. (2005) 'Dynamic and sequential patterning of the zebrafish posterior hindbrain by retinoic acid', *Developmental biology* 285(2): 593-605.

McBurney, M. W. (1993) 'P19 embryonal carcinoma cells', *International Journal of Developmental Biology* 37(1): 135-40.

McBurney, M. W., Fournier, S., Jardine, K. and Sutherland, L. (1994) 'Intragenic regions of the murine Pgk-1 locus enhance integration of transfected DNAs into genomes of embryonal carcinoma cells', *Somatic cell and molecular genetics* 20(6): 515-28.

McKay, I. J., Muchamore, I., Krumlauf, R., Maden, M., Lumsden, A. and Lewis, J. (1994) 'The kreisler mouse: a hindbrain segmentation mutant that lacks two rhombomeres', *Development* 120(8): 2199-211.

McMullin, R. P., Mutton, L. N. and Bieberich, C. J. (2009) 'Hoxb13 regulatory elements mediate transgene expression during prostate organogenesis and carcinogenesis', *Dev Dyn* 238(3): 664-72.

Minucci, S., Leid, M., Toyama, R., Saint-Jeannet, J. P., Peterson, V. J., Horn, V., Ishmael, J. E., Bhattacharyya, N., Dey, A., Dawid, I. B. et al. (1997) 'Retinoid X receptor (RXR) within the RXR-retinoic acid receptor heterodimer binds its ligand and enhances retinoid-dependent gene expression', *Molecular and Cellular Biology* 17(2): 644-55.

Mittler, G., Butter, F. and Mann, M. (2009) 'A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements', *Genome Research* 19(2): 284-293.

Moens, C. B., Cordes, S. P., Giorgianni, M. W., Barsh, G. S. and Kimmel, C. B. (1998) 'Equivalence in the genetic control of hindbrain segmentation in fish and mouse', *Development* 125(3): 381-391.

Moens, C. B. and Prince, V. E. (2002) 'Constructing the hindbrain: insights from the zebrafish', *Dev Dyn* 224(1): 1-17.

Mollard, R. and Dziadek, M. (1997) 'Homeobox genes from clusters A and B demonstrate characteristics of temporal colinearity and differential restrictions in spatial expression domains in the branching mouse lung', *International Journal of Developmental Biology* 41(5): 655-66.

Moriguchi, T., Hamada, M., Morito, N., Terunuma, T., Hasegawa, K., Zhang, C., Yokomizo, T., Esaki, R., Kuroda, E., Yoh, K. et al. (2006) 'MafB is essential for renal development and F4/80 expression in macrophages', *Molecular and Cellular Biology* 26(15): 5715-27.

Moroni, M. C., Vigano, M. A. and Mavilio, F. (1993) 'Regulation of the human HOXD4 gene by retinoids', *Mechanisms of development* 44(2-3): 139-54.

Morrison, A., Ariza-McNaughton, L., Gould, A., Featherstone, M. and Krumlauf, R. (1997) 'HOXD4 and regulation of the group 4 paralog genes', *Development* 124(16): 3135-46.

Morrison, A., Moroni, M. C., Ariza-McNaughton, L., Krumlauf, R. and Mavilio, F. (1996) 'In vitro and transgenic analysis of a human HOXD4 retinoid-responsive enhancer', *Development* 122(6): 1895-907.

Nelson, C. E., Morgan, B. A., Burke, A. C., Laufer, E., DiMambro, E., Murtaugh, L. C., Gonzales, E., Tessarollo, L., Parada, L. F. and Tabin, C. (1996) 'Analysis of Hox gene expression in the chick limb bud', *Development* 122(5): 1449-66.

Nishizawa, M., Kataoka, K., Goto, N., Fujiwara, K. T. and Kawai, S. (1989) 'V-Maf, a Viral Oncogene That Encodes a Leucine Zipper Motif - (Avian Retrovirus Transformation DNA-Binding Protein)', *Proceedings of the National Academy of Sciences of the United States of America* 86(20): 7711-7715.

Nolte, C., Amores, A., Nagy Kovacs, E., Postlethwait, J. and Featherstone, M. (2003) 'The role of a retinoic acid response element in establishing the anterior neural expression border of Hoxd4 transgenes', *Mech Dev* 120(3): 325-35.

Nolte, C., Rastegar, M., Amores, A., Bouchard, M., Grote, D., Maas, R., Kovacs, E. N., Postlethwait, J., Rambaldi, I., Rowan, S. et al. (2006) 'Stereospecificity and PAX6 function direct Hoxd4 neural enhancer activity along the antero-posterior axis', *Developmental biology* 299(2): 582-93.

Nolte, C. D. (2002) The role of a retinoic acid response element in setting the anterior border of Hoxd4 expression in the developing hindbrain.

Ogata, A., Shimizu, T., Abe, R., Shimizu, H. and Sakai, M. (2004) 'Expression of c-maf and mafB genes in the skin during rat embryonic development', *Acta histochemica* 106(1): 65-7.

Ogura, T. and Evans, R. M. (1995) 'A Retinoic Acid-Triggered Cascade of Hoxb1 Gene Activation', *Proceedings of the National Academy of Sciences of the United States of America* 92(2): 387-391.

Oosterveen, T., Niederreither, K., Dolle, P., Chambon, P., Meijlink, F. and Deschamps, J. (2003) 'Retinoids regulate the anterior expression boundaries of 5' Hoxb genes in posterior hindbrain', *The EMBO journal* 22(2): 262-9.

Packer, A. I., Crotty, D. A., Elwell, V. A. and Wolgemuth, D. J. (1998) 'Expression of the murine Hoxa4 gene requires both autoregulation and a conserved retinoic acid response element', *Development* 125(11): 1991-8.

Packer, A. I., Mailutha, K. G., Ambrozewicz, L. A. and Wolgemuth, D. J. (2000) 'Regulation of the Hoxa4 and Hoxa5 genes in the embryonic mouse lung by retinoic acid and TGFbeta1: implications for lung development and patterning', *Dev Dyn* 217(1): 62-74.

Papageorgiou, S. (2007) *HOX gene expression*, Austin,Texas New York: Landes Bioscience/Eurekah.com ; Springer Science + Business Media.

Papalopulu, N., Clarke, J. D. W., Bradley, L., Wilkinson, D., Krumlauf, R. and Holder, N. (1991) 'Retinoic Acid Causes Abnormal-Development and Segmental Patterning of the Anterior Hindbrain in Xenopus Embryos', *Development* 113(4): 1145-&.

Partanen, J., Schwartz, L. and Rossant, J. (1998) 'Opposite phenotypes of hypomorphic and Y766 phosphorylation site mutations reveal a function for Fgfr1 in anteroposterior patterning of mouse embryos', *Genes & development* 12(15): 2332-44.

Patraquim, P., Warnefors, M. and Alonso, C. R. (2011) 'Evolution of Hox post-transcriptional regulation by alternative polyadenylation and microRNA modulation within twelve Drosophila genomes', *Molecular biology and evolution*.

Peichel, C. L., Prabhakaran, B. and Vogt, T. F. (1997) 'The mouse Ulnaless mutation deregulates posterior HoxD gene expression and alters appendicular patterning', *Development* 124(18): 3481-92.

Perrin, L., Monier, B., Ponzielli, R., Astier, M. and Semeriva, M. (2004) 'Drosophila cardiac tube organogenesis requires multiple phases of Hox activity', *Developmental biology* 272(2): 419-31.

Phelan, M. L., Sadoul, R. and Featherstone, M. S. (1994) 'Functional Differences between Hox Proteins Conferred by 2 Residues in the Homeodomain N-Terminal Arm', *Molecular and Cellular Biology* 14(8): 5066-5075.

Phua, S. L., Sivakamasundari, V., Shao, Y., Cai, X., Zhang, L. F., Lufkin, T. and Featherstone, M. (2011) 'Nuclear Accumulation of an Uncapped RNA Produced by Drosha Cleavage of a Transcript Encoding miR-10b and HOXD4', *PLoS One* 6(10): e25689.

Popperl, H., Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., Mann, R. S. and Krumlauf, R. (1995) 'Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx', *Cell* 81(7): 1031-42.

Popperl, H. and Featherstone, M. S. (1992) 'An autoregulatory element of the murine Hox-4.2 gene', *The EMBO journal* 11(10): 3673-80.

Popperl, H. and Featherstone, M. S. (1993) 'Identification of a retinoic acid response element upstream of the murine Hox-4.2 gene', *Molecular and Cellular Biology* 13(1): 257-65.

Prince, V. E., Moens, C. B., Kimmel, C. B. and Ho, R. K. (1998) 'Zebrafish hox genes: expression in the hindbrain region of wild-type and mutants of the segmentation gene, valentino', *Development* 125(3): 393-406.

Prince, V. E. and Pickett, F. B. (2002) 'Splitting pairs: the diverging fates of duplicated genes', *Nat Rev Genet* 3(11): 827-37.

Punnamoottil, B., Herrmann, C., Pascual-Anaya, J., D'Aniello, S., Garcia-Fernandez, J., Akalin, A., Becker, T. S. and Rinkwitz, S. (2010) 'Cis-regulatory characterization of sequence conservation surrounding the Hox4 genes', *Developmental biology* 340(2): 269-82.

Punnamoottil, B., Kikuta, H., Pezeron, G., Erceg, J., Becker, T. S. and Rinkwitz, S. (2008) 'Enhancer detection in zebrafish permits the identification of neuronal subtypes that express Hox4 paralogs', *Dev Dyn* 237(8): 2195-208.

Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J. and Wuthrich, K. (1989) 'The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors', *Cell* 59(3): 573-80.

Rastegar, M., Kobrossy, L., Kovacs, E. N., Rambaldi, I. and Featherstone, M. (2004) 'Sequential histone modifications at Hoxd4 regulatory regions distinguish anterior from posterior embryonic compartments', *Molecular and Cellular Biology* 24(18): 8090-103.

Reese, D. H. and Ramos-Valle, M. (2007) 'A high-throughput method for monitoring changes in homeobox gene expression', *Biochemical and biophysical research communications* 357(4): 882-888.

Rusk, N. (2009) 'Reverse ChIP', *Nature Methods* 6(3): 187-187.

Sanes, D. H., Reh, T. A. and Harris, W. A. (2006) *Development of the nervous system*, Amsterdam ; Boston: Elsevier Academic Press.

Sanson, B. (2001) 'Generating patterns from fields of cells. Examples from *Drosophila* segmentation', *EMBO reports* 2(12): 1083-8.

Sevinsky, J. R., Whalen, A. M. and Ahn, N. G. (2004) 'Extracellular signal-regulated kinase induces the megakaryocyte GPIIb/CD41 gene through MafB/Kreisler', *Molecular and Cellular Biology* 24(10): 4534-45.

Shang, L., Pruetz, N. D. and Awgulewitsch, A. (2002) 'Hoxc12 expression pattern in developing and cycling murine hair follicles', *Mech Dev* 113(2): 207-10.

Sharpe, J., Nonchev, S., Gould, A., Whiting, J. and Krumlauf, R. (1998) 'Selectivity, sharing and competitive interactions in the regulation of Hoxb genes', *The EMBO journal* 17(6): 1788-98.

Shimizu, T., Bae, Y. K. and Hibi, M. (2006) 'Cdx-Hox code controls competence for responding to Fgfs and retinoic acid in zebrafish neural tissue', *Development* 133(23): 4709-19.

Solnica-Krezel, L. (2002) *Pattern formation in zebrafish*, Berlin ; London: Springer.

Spitz, F., Gonzalez, F. and Duboule, D. (2003) 'A global control region defines a chromosomal regulatory landscape containing the HoxD cluster', *Cell* 113(3): 405-17.

Struhl, G., Johnston, P. and Lawrence, P. A. (1992) 'Control of *Drosophila* body pattern by the hunchback morphogen gradient', *Cell* 69(2): 237-249.

Studer, M., Gavalas, A., Marshall, H., Ariza-McNaughton, L., Rijli, F. M., Chambon, P. and Krumlauf, R. (1998) 'Genetic interactions between Hoxa1 and Hoxb1 reveal new roles in regulation of early hindbrain patterning', *Development* 125(6): 1025-36.

Studer, M., Popperl, H., Marshall, H., Kuroiwa, A. and Krumlauf, R. (1994) 'Role of a conserved retinoic acid response element in rhombomere restriction of Hoxb-1', *Science* 265(5179): 1728-32.

Tarchini, B. and Duboule, D. (2006) 'Control of Hoxd genes' collinearity during early limb development', *Developmental cell* 10(1): 93-103.

Tawk, M., Araya, C., Lyons, D. A., Reugels, A. M., Girdler, G. C., Bayley, P. R., Hyde, D. R., Tada, M. and Clarke, J. D. (2007) 'A mirror-symmetric cell division that orchestrates neuroepithelial morphogenesis', *Nature* 446(7137): 797-800.

Tkatchenko, A. V., Visconti, R. P., Shang, L., Papenbrock, T., Pruett, N. D., Ito, T., Ogawa, M. and Awgulewitsch, A. (2001) 'Overexpression of Hoxc13 in differentiating keratinocytes results in downregulation of a novel hair keratin gene cluster and alopecia', *Development* 128(9): 1547-58.

Volpe, M. V., Wang, K. T., Nielsen, H. C. and Chinoy, M. R. (2008) 'Unique spatial and cellular expression patterns of Hoxa5, Hoxb4, and Hoxb6 proteins in normal developing murine lung are modified in pulmonary hypoplasia', *Birth Defects Res A Clin Mol Teratol* 82(8): 571-84.

Wacker, S. A., McNulty, C. L. and Durston, A. J. (2004) 'The initiation of Hox gene expression in *Xenopus laevis* is controlled by Brachyury and BMP-4', *Developmental biology* 266(1): 123-37.

Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M. and Martin, D. I. (1995) 'Enhancers increase the probability but not the level of gene expression', *Proceedings of the National Academy of Sciences of the United States of America* 92(15): 7125-9.

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B. R., Protacio, A., Flynn, R. A., Gupta, R. A. et al. (2011) 'A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression', *Nature* 472(7341): 120-124.

Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P. W., Stott, D. and Allemann, R. K. (1991) 'Multiple spatially specific enhancers are required to

reconstruct the pattern of Hox-2.6 gene expression', *Genes & development* 5(11): 2048-59.

Wong, E. Y., Wang, X. A., Mak, S. S., Sae-Pang, J. J., Ling, K. W., Fritzsche, B. and Sham, M. H. (2011) 'Hoxb3 negatively regulates Hoxb1 expression in mouse hindbrain patterning', *Developmental biology* 352(2): 382-92.

Yang, Y. and Cvekl, A. (2007) 'Large Maf Transcription Factors: Cousins of AP-1 Proteins and Important Regulators of Cellular Differentiation', *Einstein J Biol Med* 23(1): 2-11.

Yekta, S., Tabin, C. J. and Bartel, D. P. (2008) 'MicroRNAs in the Hox network: an apparent link to posterior prevalence', *Nat Rev Genet* 9(10): 789-96.

Zacchetti, G., Duboule, D. and Zakany, J. (2007) 'Hox gene function in vertebrate gut morphogenesis: the case of the caecum', *Development* 134(22): 3967-73.

Zakany, J., Kmita, M. and Duboule, D. (2004) 'A dual role for Hox genes in limb anterior-posterior asymmetry', *Science* 304(5677): 1669-72.

Zappavigna, V., Falciola, L., Helmer-Citterich, M., Mavilio, F. and Bianchi, M. E. (1996) 'HMG1 interacts with HOX proteins and enhances their DNA binding and transcriptional activation', *The EMBO journal* 15(18): 4981-91.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J. W., Nechaev, S., Adelman, K., Levine, M. and Young, R. A. (2007) 'RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo', *Nat Genet* 39(12): 1512-6.

Zhang, F., Nagy Kovacs, E. and Featherstone, M. S. (2000) 'Murine *hoxd4* expression in the CNS requires multiple elements including a retinoic acid response element', *Mech Dev* 96(1): 79-89.

Zhang, F., Popperl, H., Morrison, A., Kovacs, E. N., Prideaux, V., Schwarz, L., Krumlauf, R., Rossant, J. and Featherstone, M. S. (1997) 'Elements both 5' and 3' to

the murine Hoxd4 gene establish anterior borders of expression in mesoderm and neurectoderm', *Mech Dev* 67(1): 49-58.

Zhao, Z. S., Manser, E., Loo, T. H. and Lim, L. (2000) 'Coupling of PAK-interacting exchange factor PIX to GIT1 promotes focal complex disassembly', *Molecular and Cellular Biology* 20(17): 6354-63.