University of Alberta

Regulation of Vertebrate *Ladybird* Genes

by

Christopher Michael Lukowski

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Molecular Biology and Genetics

Department of Biological Sciences

Spring 2010
Edmonton, Alberta

Examining Committee

Dr. Andrew Waskiewicz, Department of Biological Sciences

Dr. Dave Pilgrim, Department of Biological Sciences

Dr. William (Ted) Allison, Department of Biological Sciences

Dr. Alan Underhill, Department of Medical Genetics

Abstract

Development of the vertebrate central nervous system is a complex process that relies on the accurate spatiotemporal distribution of signaling centers during embryogenesis. These signals provide cells with positional information, which is integrated via transcription factors and gene regulatory elements to generate a specific downstream gene expression profile that confers specific cellular functions. It is of interest to determine how cells acquire their unique spatiotemporal gene expression patterns. The wide variety of expression profiles established along the dorsoventral axis of the neural tube provides a great system to address this question. Recent advances in zebrafish transgenic technology, along with the phenomenon of a fish-specific genome duplication event, have been exploited here to provide an efficient way of identifying and characterizing gene regulatory elements. An identified neuronal-specific enhancer near the *ladybird* locus has been incorporated into a transgenic zebrafish strain driving fluorescent reporter protein expression in a subset of dorsal interneurons.

Preface

This little fishy has forebrain, and this little fishy has none. This little fishy has one eye and this little fishy has none....and after banging it's head against the Petri dish for hours this little fishy went WEE WEE WEE all the way home!!!

– Jen Becker

Acknowledgements

This accomplishment was inspired by many people in my life, particularly those whom I met in my senior undergraduate year and early graduate career. I am thankful to my supervisor, Andrew, for providing me an environment to develop and excel at many skills, along with encouragement and enthusiasm for my success. James MacLagan and Monica Engstrom are two individuals whom had I never met, probably would not have entered into a graduate program in molecular biology and genetics. Their sense of humor and passion for lab work was truly contagious which made me want to be in the lab all the time. Gary Ritzel, aka "God" to the inquisitive student, provided an answer to everything regarding both theoretical and practical molecular biology.

Although my supervisors and mentors have been an invaluable source of motivation and encouragement, the feeling I get from teaching and helping students understand a concept or technique is what made my years as a graduate student the best years of my life. It didn't matter how many students complained for extra marks, or were unsatisfied with a graded assignment or report, I knew sooner or later I would see a smile on a puzzled student's face when they "got it". Being a teaching assistant for the course I enjoyed so much, genetics 420, made my masters – Dave Pilgrim is an incredible instructor and the fun times with motivated and enthusiastic students made those two, four-month terms as good as it got. Thanks to my amazing girlfriend, Jen – without your support and encouragement I wouldn't have finished this thesis… I love you ♥

Table of Contents

List of Tables:

List of Figures:

List of Abbreviations

AP, anteroposterior
BCIP, 5-bromo-4-chloro-3-indoyl-phosphate
bHLH, basic helix-loop-helix
BLAST, Basic local alignment search tool
BLAT, BLAST-like alignment tool
Bmp, Bone morphogenetic protein
ChIP, chromatin immuniprecipitation
CNS, central nervous system
DMSO, dimethylsulfoxide
dpf, days post-fertilization
DV, dorsoventral
FACS, fluorescence activated cell sorting
GW, Gateway
HAT, histone acetyltransferase
HDAC, histone deacetylase
hpf, hours post-fertilization
INT, iodonitro-tetrazolium
LG, linkage group
MeOH, methonal
MESAB, ethyl-m-aminobenzoate methanesulfonate
MHB, mid-hindbrain boundary
MHO, mid-hindbrain organizer
NBT, nitroblue-tetrazolium
NT, nucleotide
OCT, Optimal cutting temperature
ORF, open reading frame
PBS, phosphate-buffered saline
PFA, paraformaldehyde
proK, proteinase K
PTU, phenylthiourea
RA, retinoic acid
RE, restriction endonuclease
RNAP, RNA polymerase
RT, reverse-transcriptase
Shh, sonic hedgehog
SSC, standard saline citrate
TALE, three-amino-acid-loop extension
TF, transcription factor
TFBS, transcription factor binding sequence
UTR, untranslated region
ZIRC, zebrafish information resource center

Chapter 1: Introduction to Vertebrate Neural Development

Progression of a single cell zygote into a complex adult requires the spatiotemporal coordination of numerous signaling pathways. Cells at specific positions integrate this signaling information and develop accordingly through modulation of their gene expression profile. When gene products involved in determining cell fate are aberrantly expressed, a range of developmental defects can result, including abnormal limb development (Clark *et al.*, 2001), brain malformations (Totori-Donati *et al.*, 2005) or cranial dysinnervation defects (van der Zwaag, 2006). Understanding the mechanisms that generate these specific gene expression patterns will elucidate the means by which these abnormalities manifest and offer potential therapeutic techniques.

The central nervous system provides a great model to examine changes in gene expression and cell fate as the signaling mechanisms generating neuronal diversity have been well characterized. The anteroposterior and dorsoventral axes are two major embryonic axes in which a variety of signaling centers have been identified. One particular group of neurons occupying an intermediate position in the dorsoventral axis of the neural tube develops independently of these signaling mechanisms. This group relies on the Ladybird homeodomain transcription factor Lbx1 and the defects observed from the loss of Lbx1 contradict to the classical model of dorsoventral patterning. Identifying the mechanisms driving *Lbx1* expression in the central nervous system will help explain the signal-independent development of these intermediate neurons and further our understanding of the genes required in neural tube patterning during embryonic development.

1) Vertebrate Neural Induction

The vertebrate central nervous system is specified early in development during the cellular reorganization events of gastrulation, where cells of a flat ectodermal sheet ingress and involute to create multiple germ layers. The tissue responsible for neural induction was first identified in amphibians as the dorsal lip of the blastopore (Spemann and Mangold, 1924), which is a group of cells that collectively are able to 'organize' neural ectoderm for the surrounding non-neural ectoderm (Spemann, 1938). This organizer region has been under meticulous examination over the past three decades in order to identify the molecular mechanisms capable of generating such an important cell fate decision. A striking observation was that transplantation of a donor organizer region to a host embryo not only creates neural tissue, but induces a new and nearly complete body axis that is derived almost entirely from host cells. Analogous organizer regions have been identified in a wide variety of species including the node in mammals (Beddington, 1994; Boettger *et al.*, 2001) and the shield in teleost fish (Shih and Fraser, 1996).

The gastrula organizer resides on the presumptive dorsal side of the embryo and mediates neural specification through the antagonism of epidermal fate signaling molecules that emanate from the ventral side (De Robertis and Kurodo, 2004; Stern, 2005; De Robertis, 2009; Rogers *et al.*, 2009). Bone morphogenetic proteins (Bmp) released from the ventral side act to promote

epidermal specification, while the organizer region releases a suite of Bmp antagonists preventing their interaction with cognate receptors, which promotes neural specification. The major inhibitors of Bmp signaling have been identified as Chordin (Sasai *et al.*, 1994; Sasai *et al.*, 1995; Oelgeschläger *et al.*, 2003), Noggin (Lamb *et al.*, 1993; Zimmerman *et al.*, 1996) and Follistatin (Hemmati-Brivanlou *et al.*, 1994), proteins whose actions are highly conserved from arthropods to vertebrates (Holley *et al.*, 1995; Ferguson, 1996). Thus, the process of neural specification through antagonism of Bmp signaling seems to be an evolutionarily conserved mechanism during embryonic development (De Robertis, 2008).

The cell movements during gastrulation, combined with the activity of the organizer, yield a flat sheet of neural ectoderm, also known as the neural plate. Understanding the genetic mechanisms that sub-divide the neural plate along the anteroposterior (AP) and dorsoventral (DV) axes will provide potential treatment opportunities for neuronal diseases. Patterning molecules work by imposing regional restrictions on the potential fate of a neuronal cell through the activation and/or repression of neural fate determinant genes. As additional factors restrict cell lineage potential further, neuronal differentiation proceeds and morphological changes consistent with future cellular functions occur. Dorsoventral patterning of the central nervous system is facilitated by the process of neurulation (Colas and Schoenwolf, 2001; Lowery and Sive, 2004; Harrington *et al.*, 2009), where the neural plate converges towards the midline and folds up into a hollow neural tube.

The neural tube is found along the dorsal side of the embryo in vertebrates and is situated between two crucial tissues. The underlying notochord specifies ventral neural fates and the overlying dorsal ectoderm specifies dorsal fates. Interestingly, it was observed that the notochord was the only tissue which was derived from the cells of a transplanted organizer (Spemann and Mangold, 1924), suggesting a vital role in providing instructions for surrounding tissues during early development.

2) Anteroposterior Patterning of the Vertebrate Neural Tube

The vertebrate neural tube is initially sub-divided along the anteroposterior axis into two main compartments, the presumptive brain and the spinal cord. The brain is partitioned into the forebrain (prosencephalon), midbrain (mesencephalon) and hindbrain (rhombencephalon) while the spinal cord retains uniformity along the entire trunk of the embryo (Kiecker and Lumsden, 2005). Numerous genes and signaling mechanisms play a prominent role in the development of these structures (Figure 1-1). The forebrain is a highly complex structure that is specified by signals from the rostral notochord, also known as the pre-chordal plate (Pera and Kessel, 1997), and the anterior ectoderm (Shimamura and Rubenstein, 1997; Houart *et al.*, 1998). The division between the midbrain and hindbrain is generated by the actions of a secondary signaling center, the isthmic organizer, which delineates the mid-hindbrain boundary (MHB; Rhinn and Brand, 2001). The hindbrain and spinal cord are patterned along the AP axis by the highly conserved Hox protein family (Akin and Nazarali, 2005; Hueber

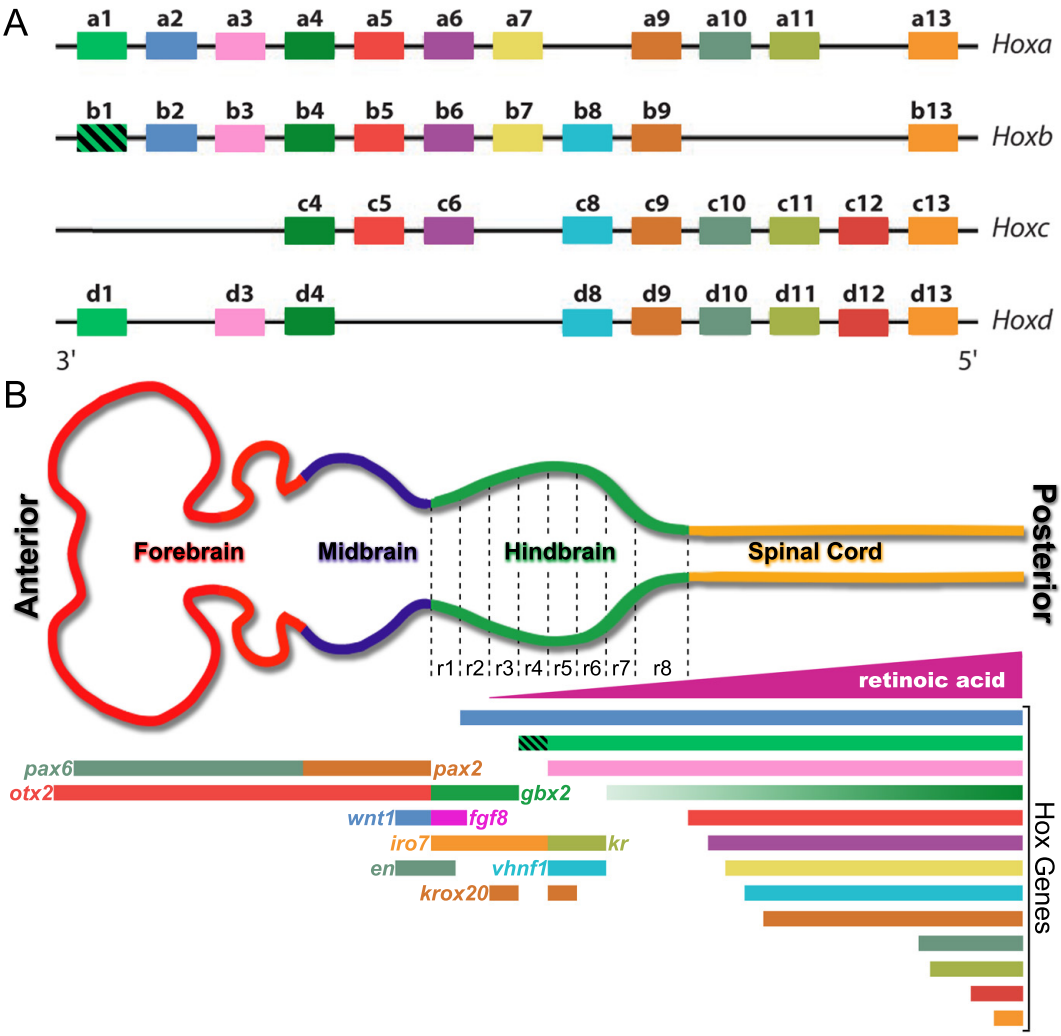

Figure 1-1. **An Overview of Anteroposterior Patterning in the Vertebrate Embryo**. *Hox* genes exhibit colinearity, where genes situated along a 3' to 5' direction (A) from each cluster (*Hox- a*, *b*, *c*, *d*) are expressed in an anterior to posterior fashion along the embryo length (B; below). A retinoic acid gradient and a variety of other genes that are important in the formation of specific neural tissues are also shown. The expression pattern of each gene and the approximate spatial arrangement along the anteroposterior axis of the embryo is indicated below the schematic by a colored box. The various divisions of the neural tube have been indicated with anterior to the left and posterior to the right. Modified from Alexander *et al.*, 2009

and Lohmann, 2008; Narita and Rijli, 2009; Tumpel *et al.*, 2009; Wellik, 2009). *Hox* gene expression is highly over-lapping (Figure 1-1) and there are many functional redundancies stemming from a series of genome duplications that took place during the evolution of the vertebrate lineage.

The forebrain and midbrain are very distinct from the rest of the vertebrate central nervous system in their morphological complexity and the developmental pathways from which they are derived. The major divisions of the rostral neural tube are the telencephalon, diencephalon and the eyes, which comprise the forebrain, and the mesencephalon. The forebrain is marked by expression of a Paired homeodomain transcription factor, *Pax6* (Manuel and Price, 2005) while the midbrain is marked by *Pax2* expression (Okafuji *et al.*, 1999; Nakamura, 2001). The region of the neural tube that develops into the forebrain and midbrain is defined by a homeodomain transcription factor, Otx2, which, through heterologous rescue experiments, has been shown to be functionally equivalent to the *Drosophila* orthodenticle protein (Acampora *et al.*, 1998; Leuzinger *et al.*, 1998). The forebrain is further sub-divided into functional compartments through distinct patterns of gene expression that arise from a series of complex genetic interactions (Schuurmans and Guillemot, 2002). In the adult these telencephalic domains serve numerous functions such as touch and smell sensation, intelligence, memory, and a variety of motor functions.

The regional distinction made between midbrain and hindbrain territory is accomplished through the mutual antagonism of Otx2/*Otx2* and Gbx2/*Gbx2*

(Gastrulation brain homeobox 2), which specifies the position of the mid-hindbrain organizer (MHO) along the AP axis (Prakash and Wurst, 2004; Hidalgo-Sánchez *et al.*, 2005). The reduction or loss of anterior *Otx2* expression results in a rostral shift of the MHO with a concomitant loss of midbrain tissue and an expansion of hindbrain tissue. Expansion of *Otx2* expression posteriorly through heterologous transgenes results in expansion of the midbrain and loss of the rostral hindbrain. Thus, the Otx2-Gbx2 boundary defines the position of the MHO (Reichert, 2002; Simeone *et al.*, 2002) and seems to be the most upstream event in separating the midbrain from hindbrain. The downstream signalling mechanisms involve a wingless-related protein, Wnt1, which becomes restricted to the rostral side of the MHO, and fibroblast growth factor 8 (Fgf8) which is on the caudal side of the MHO (Wurst and Bally-Cuif, 2001).

Wnt signaling has other functions in anteroposterior patterning of the vertebrate neural tube as it is strongly repressed in anterior regions and activated in posterior regions (Yamaguchi, 2001). Ectopic activation of Wnt signaling in anterior regions through de-repression or ectopic expression leads to the loss of forebrain-derived tissues, most notably the lack of eyes (Pöpperl *et al.*, 1997; Kim *et al.*, 2000). At the transcriptional level, Wnt repression is accomplished directly by Six3 (Lagutin *et al.*, 2003), a oculus homeobox protein whose expression is tightly controlled by several transcription factors (Lengler and Graw, 2001). Similar to how Bmp signaling molecules are inhibited by organizer-derived antagonists, anterior Wnt signaling is antagonized by proteins such as secreted

Frizzled-related protein, Dickkopf and Cerebrus (Kawano and Kypta, 2003; Niehrs, 2006). The combination of these precisely controlled regulatory mechanisms permits forebrain specification within a small anterior domain and it is necessary to understand the potential effects of minor shifts or changes gene expression has on embryonic development and human disease.

In the initial stages of embryonic neural development, the hindbrain appears contiguous with the rest of the neural tube but it quickly becomes compartmentalized into morphologically distinct metameric units known as rhombomeres. Each rhombomere is defined by a unique combinatorial code of gene expression, which consists mainly of homeotic (*Hox*) or other homeodomain containing transcription factors. Homeodomain proteins have been studied tremendously since it was observed that disruption of *Hox* gene expression in *Drosophila* resulted in 'homeotic' transformations where one body segment took on characteristics of a different body segment (Lewis, 1978). The study of *Hox* genes and their involvement in anteroposterior patterning during embryonic development expanded significantly upon the discovery of the conservation between *Hox* genes in *Drosophila* and mammals (Graham *et al.*, 1989). Unfortunately, the study of *Hox* genes in vertebrates has been limited by multiple genome duplications for example, which results in a high amount of functional redundancy between paralogs and often requires simultaneous disruption of several genes in order to produce a phenotype.

Molecular mapping and further genetic studies in *Drosophila* by Lewis and others identified *Hox* genes in a chromosomal cluster exhibiting co-linearity (Kaufman *et al.*, 1980), where their expression patterns along the anteroposterior axis corresponded to their chromosomal position. The principle of co-linearity in the *Hox* cluster, where 5' situated genes are expressed more posteriorly and 3' genes are expressed more anteriorly (Figure 1-1, B), has been maintained throughout evolution (Kmita, 2003; Lemons and McGinnis, 2006). In vertebrates perturbation of *Hox* expression results in the loss or alteration of rhombomere identity or vertebral morphology (Boncinelli *et al.*, 1989; Duboule and Dolle, 1989) instead of whole body segment transformations like in *Drosophila*. Introduction of ectopic *Hox* gene expression may be an attractive treatment for limb deficiencies caught early in development of families with heritable conditions. Recently, zebrafish has emerged as a great model system to examine the effects of disrupting *Hox* gene expression through the manipulation of retinoic acid signaling (Hernandez *et al.*, 2004; Hernandez *et al.*, 2007).

Retinoic acid (RA) signaling is a major component in determining the anteroposterior expression pattern of *Hox* genes, which is directed through retinoid receptors RAR and RXR. The upstream regulatory regions of several early *Hox* genes such as *HoxA1* and *HoxB1* contain identified retinoic acid response elements (Marshall *et al.*, 1994; Frasch *et al.*, 1995). RA signaling is highly conserved amongst vertebrates but a wide variety of basal metazoans such as Cnidarians and Protostomes posses components of the RA signaling pathway

(Campo-Paysaa *et al.*, 2008). Phylogenetic analysis suggests the RA pathway has undergone significant changes in certain lineages. Once retinoic acid signaling has established the initial position of early *Hox* gene expression, downstream signaling mechanisms and transcription factors are turned on in a tightly controlled manner (Alexander *et al.*, 2009), which drives the compartmentalization observed within the hindbrain. A key feature involved in rhombomere formation is cell sorting – a process mediated by the bi-directional signaling of membrane bound ephrin ligands and their cognate Eph receptors (Xu *et al.*, 2000; Cooke and Moens, 2002). Initially, cells of specific rhombomeres do not form a sharp boundary but instead are interspersed with cells of adjacent rhombomeres and cell sorting facilitates the grouping of similar cells into these compartments.

The property of co-linear expression exhibited by *Hox* genes and their ability to drive formation of specific tissues has dubbed this transcription factor group as selectors of segmental cell fate. This concept was quickly re-evaluated upon the discovery of a relatively simple Hox protein consensus DNA binding sequence of TAATT (Berger *et al.*, 2008; Noyes *et al.*, 2008). However, very few homeodomain containing transcription factors work alone, notably including the Hox proteins. They impose their effects on transcription as heteromeric protein complexes with a well known group of Hox cofactors, which are the three amino acid loop extension (TALE) class. A hydrophobic pocket is created by the TALE domain to accommodate interaction with a tryptophan residue from a highly

conserved hexapeptide motif found in Hox proteins (Chang *et al.*, 1995; Knoepfler and Kamps, 1995; Jabet *et al.*, 1999). This class includes extradenticle (Exd) and homothorax (Hth) from *Drosophila* and their vertebrate homologs pre-B-cell leukemia homeobox (Pbx) and myeloid ecotropic viral integration site (Meis), respectively. Although TALE class proteins have Hox-independent functions, they mainly function to provide a cooperative and increased DNA binding specificity for Hox proteins. Identifying the genomic sequences to which these and other heteromeric protein complexes bind will determine the role Hox cofactors play in modulating binding specificity.

Hox cofactors such as Pbx have been shown to be critical in the anteroposterior patterning of the vertebrate hindbrain (Waskiewicz *et al.*, 2002). The loss of functional Pbx proteins results in abolishment of all specific rhombomere characteristics and the entire hindbrain exhibits a uniform character, which is also known as the hindbrain 'ground state'. Meis on the other hand appears to be important for stabilization of Pbx (Waskiewicz *et al.*, 2001) and its transport into the nucleus (Rieckhof *et al.*, 1997), such that the ubiquitous gene expression pattern observed for *Pbx* (Waskiewicz *et al.*, 2002) does not accurately reflect the regions where its encoded protein is functional. Pbx, Meis and Hox have been shown to interact through specific protein domains (Jacobs *et al.*, 1999) and these domains exist in other proteins such as Engrailed (Peltenburg and Murre, 1996; Erickson *et al.*, 2007), which adds to the complex regulatory properties of Hox proteins. A microarray study examining the changes of gene

expression in response to loss of Pbx in zebrafish (Waskiewicz, unpublished data) has identified additional genes whose protein products potentially act during embryonic patterning, including a homeodomain containing transcription factor similar to the ladybird proteins found in *Drosophila*.

Posterior to the vertebrate hindbrain is the spinal cord, which is positioned along the dorsal side of the trunk of the embryo. The spinal cord is an important conduit for neuronal communication between the brain and the rest of the body (Colon-Ramos, 2009; Dasen, 2009). Information from the skin, gut and internal organs is relayed along ascending axons to various parts of the brain and communicated back via descending axons to neurons that generate a motor response. *Hox* genes play an important role in posterior patterning although many *Hox* genes exhibit overlapping expression patterns and perturbations of these genes often results in very subtle effects such as altered vertebral morphology. Disrupting the *HoxD* paralogs alters limb morphology and from their expression patterns it seems that co-linearity is maintained along the proximodistal axis of limbs, much like the AP axis (Kmita *et al.*, 2002). In addition to posterior *Hox* genes, Wnt signaling is critical for the proper patterning of posterior trunk segments or somites, in the developing embryo (Martin and Kimelman, 2008). Elucidating the necessary components, such as Hox proteins, for downstream gene expression is crucial for understanding human diseases such as cancer, where occasionally a developmentally regulated gene is ectopically activated.

3) Dorsoventral Patterning of the Vertebrate Neural Tube

Dorsoventral patterning of the neural tube is accomplished by two morphogenetic gradients positioned above and below the neural tube (Figure 1-2). These gradients originate from the roof plate and floor plate, which are the dorsal- and ventral-most cells of the neural tube, respectively. Both of these signaling centers span nearly the entire anteroposterior axis of the embryo. Their signaling has best been described in the spinal cord where changes in neuronal patterning are more easily observed. Cells of the floor plate and roof plate are part of the neural tube but do not exhibit traditional neuronal cell characteristics. The floor plate, for example, exhibits characteristics of glial cells, which primarily function to provide a scaffold for axonogenesis (Lane *et al.*, 2004). Indeed, many ascending and descending axons that run parallel to the spinal column are positioned at the ventral side of the spinal cord (Goulding, 2009). The roof plate forms later in development after neurulation is complete but shares a common precursor with neural crest cells (Echelard *et al.*, 1994), which has made it more difficult to examine the factors necessary for roof plate specification.

The floor plate is specified by the notochord, a mesodermal tissue that lies underneath the neural plate which arises from the embryonic organizer (Harland and Gerhart, 1997). The notochord induces formation of the floor plate in the ventral-most cell population of the overlying neurectoderm via the release of Sonic hedgehog (Shh), a diffusible signaling molecule (Strähle *et al.*, 2004). The

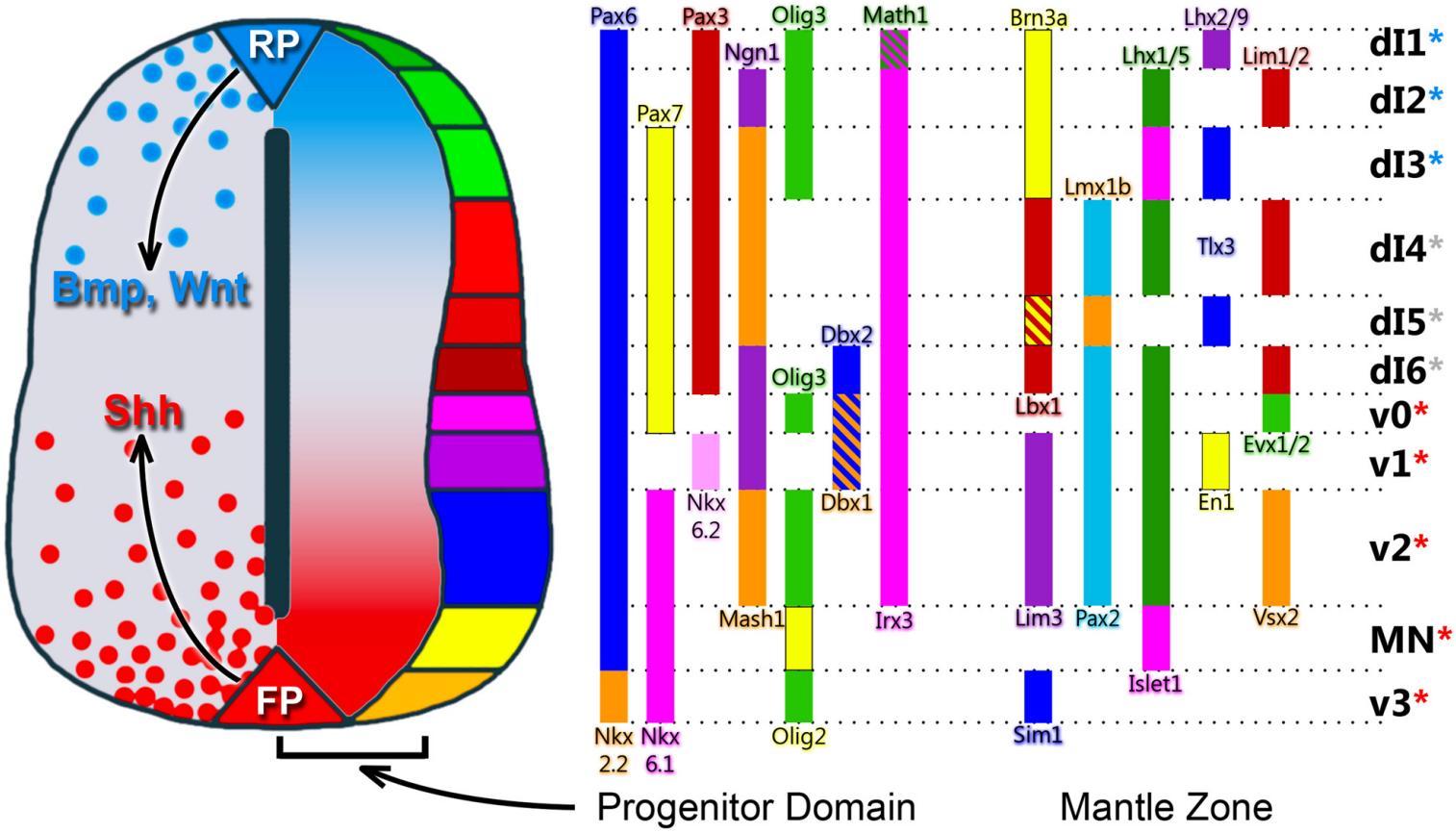

Figure 1-2. **Overview of Dorsoventral Patterning in the Vertebrate Neural Tube**. A schematic cross-section through the spinal cord depicts the patterning mechanisms that create neuronal diversity along the dorsoventral axis. Bmp/Wnt (blue) and Sonic Hedgehog (red) are released from the roof (RP) and floor (FP) plates, respectively, and expose progenitor cells to a morphogenetic gradient. Factors involved in neuronal specification within the progenitor domain are shown while those involved in differentiation are expressed in the mantle zone as highlighted by the neuronal classes along the lateral edge of the schematic. An colored asterisk denotes dependence on morphogens.

floor plate in turn acts as a signaling center, by also releasing Shh, which sets up a morphogenetic gradient that patterns the ventral half of the neural tube (Dessaud *et al.*, 2008). The ventral neural tube contains a wide variety of neuron types, including several classes of interneurons, and most notably, the motor neurons (Litingtung and Chiang, 2000).

It is known that Sonic hedgehog plays a critical role in ventral patterning of the neural tube as Shh knockout mice display a complete loss of ventral neuron development (Litingtung and Chiang, 2000). There is evidence that Shh acts in a concentration dependent manner as *in vitro* exposure to high levels of Shh induces expression of markers typically observed in cells adjacent to the floor plate (Briscoe *et al.*, 1999). Exposure to low levels of Shh results in expression of genes commonly found near the middle of the neural tube. The prominent role of Shh is to act as both a short and long range signaling molecule, which is accomplished by unique post-translational modification events. These include auto-cleavage (Bumcrot *et al.*, 1995), palmitoylation at the amino terminus (Buglino and Resh, 2008) and cholesterol modification at the carboxy terminus (Huang *et al.*, 2007). Cholesterol modification affects the functional range of Shh as removal of the cholesterol modification site increases diffusion to more distant tissues and affects neuronal specification. Conversely, palmitoylation of Shh is critical for its function rather than its functional range (Chen *et al.*, 2004).

Development of the roof plate is not as well understood, cells that adopt a roof plate identity arise from a multipotent progenitor that also gives rise to neural

crest and dorsal interneurons (Lee and Jessel, 1999). This tight association of a common progenitor which generates distinct cell types has made the characterization of the molecular mechanisms responsible for roof plate specification difficult. It is known that the ectodermal cells at the lateral edges of the neural plate are crucial for formation of the roof plate through Bmp signaling (Chizhikov and Millen, 2004a). These lateral regions of non-neural ectoderm eventually overlie the neural tube after neurulation is complete and release morphogens much like the floor plate does. However, several other factors have been shown to be necessary for establishment and maintenance of roof plate identity and the signaling mechanisms that originate from the roof plate are more complex than the activities of sonic hedgehog observed in the floor plate.

The roof plate releases a combination of well characterized morphogens including Bmp-like ligands, Gdf7 (Lee *et al.*, 1998) and Bmp4 (Liem *et al.*, 1997), and two Wnt-like ligands, Wnt1 and Wnt3a (Muroyama *et al.*, 2002; Ulloa and Marti, 2009). This combination of morphogens has both similar and different effects to that observed in Shh-dependent patterning of the ventral neural tube. These molecules function in both a concentration-dependent manner as well as a cell-specific manner where only certain dorsal cells are capable of responding to certain ligands (Lee and Jessel, 1999). Evidence from chick embryo manipulation and mouse knockouts suggest that a transcription factor, Lmx1a, is necessary and sufficient for the specification of roof plate identity (Chizhikov and Millen, 2004b). Loss of Lmx1a in the *dreher* mutant mouse line results in a complete lack

of the roof plate and the expression of Bmp- and Wnt-like dorsal patterning molecules is absent. Cells still exist at these positions but their specification is aberrant. Interestingly, eliminating either Gdf7, Bmp4, Wnt1 or Wnt3a alone does not affect roof plate formation and only mildly effects the development of dorsal neurons, mainly their abundance rather than their specification (Caspary and Anderson, 2003; Chizhikov and Millen, 2005). However, the most striking observation in roof plate-dependent patterning is observed in transgenic mice expressing diphtheria toxin A under control of endogenous Gdf7 regulatory elements, which essentially ablates the entire roof plate. Although there is still transient expression of signaling molecules prior to toxin-induced cell death, loss of the roof plate in this manner results in complete loss of many dorsal neuron types in the spinal cord and an expansion of more intermediate spinal cord neurons (Lee *et al.*, 2000).

The morphogenetic gradients set up by the roof plate and floor plate create a great diversity of cell types in the neural tube, which allows for the neuronal complexity observed in the vertebrate central nervous system. It is essential we understand the characteristics and functional properties of each neuronal subtype and the specific code of transcription factors that controls their development so specific neuron types may be cultured *in vitro* and transplanted into candidate host. To date, the mouse has provided much of this data regarding a transcription factor code (Briscoe *et al.*, 1999; Wilson and Maden, 2005), although comparative studies to other model vertebrates like zebrafish (Blader and Strähle,

2000; Downes *et al.*, 2002; Lewis and Eisen, 2003) and chicken (Wilson *et al.*, 2004; Alvarez-Medina *et al.*, 2008) are now receiving more attention. The attractiveness of zebrafish as a vertebrate model organism is a driving force behind such comparative studies in the hope that mechanisms controlling neuronal specification and diversity are highly conserved between zebrafish and mammals. The limited supply of transgenic strains and poor resolution of colorimetric mRNA *in situ* hybridization are currently major limitations to the analysis of the transcription factor code governing central nervous system development in zebrafish.

4) Transcription Factor Code of the Vertebrate Neural Tube

The signaling molecules from the roof plate and floor plate, most notably sonic hedgehog (Dessaud *et al.*, 2008; Nishi *et al.*, 2009) and bone morphogenetic proteins (Caspery and Anderson, 2003; Helms and Johnson, 2003), induce the differential expression of many target genes (Figure 1-2). Positional information is conveyed to neuronal progenitor cells through a concentration gradient established by each morphogen – the further away a cell is from the signal source, the less receptor activation occurs. Early during DV patterning, expression of several proneural transcription factors is initiated in a broad domain in response to the early morphogenetic gradient (Briscoe *et al.*, 2000). These transcription factors are expressed in neuronal progenitors (Figure 1-2) that are still in a multipotent state and are therefore termed progenitor domain transcription factors.

The neuronal progenitors reside in the inner regions of the neural tube defined as the ventricular zone, which is area of active neuronal cell division.

While in the progenitor domain, opposing, mutually repressive transcription factor networks act to progressively restrict neuronal cell fate by altering the gene expression profile within a specific subset of neurons (Dessaud *et al.*, 2008). Eventually, the overlapping expression domains and mutual repression activities of these transcription factors signals the process of neuronal differentiation to begin. As neuronal differentiation begins, cells exit the cell cycle and migrate to the lateral edges of the neural tube where they acquire a new code of expressed transcription factors (Caspery and Anderson, 2003). These new transcription factors drive both cell cycle exit and the acquisition of characteristics consistent with future neuronal cell functions. The domain occupied by differentiating neurons is known as the mantle zone, which is at the periphery of the neural tube. In the mouse, additional paths of neuronal migration are also observed, which results in a significant reorganization of the dorsal neuron types (Gross *et al.*, 2002; Müller *et al.*, 2002) making comparative studies to chicken and zebrafish difficult.

Development of ventral neural fates in the spinal cord is supported with far more experimental evidence when compared to development of dorsal neural fates. Disruption of a single gene, *sonic hedgehog*, reveals that the signaling mechanisms are less complex and therefore changes are more easily visualized. The discoveries of ventral neural tube patterning have laid the framework for

studying the development of dorsal neural fates. Recently, however, it has been shown that the signaling pathway from even a single morphogen, Shh, can be greatly elaborated by post-translational modification, the presence of a variety of receptor types and feedback loops (Marti and Bovolenta, 2002; Dessaud *et al.*, 2008). The complex activities of Shh are due to a number of different cellular responses which are now comparable to the complexities observed in the development of the dorsal neural tube. The cellular effectors of Shh signaling are the Gli protein family of zinc finger transcription factors, which exhibit a variety of context-dependent gene regulatory functions (Ruiz I Altaba, 1999; Persson *et al.*, 2002; Jacob and Briscoe, 2003).

Gli proteins are converted between activator and repressor forms based on two main criteria: the amount of Shh signaling and the type of Gli protein (Ruel and Thérond, 2009). Three Gli proteins exist in vertebrates: Gli1 functions only as a transcriptional activator while Gli2 and Gli3 can function as either a transcriptional activator or repressor based upon the level of transduced Shh signal (Persson *et al.*, 2002; Meyer and Roelink, 2003; Ruiz i Altaba *et al.*, 2007). The *Gli* homolog in *Drosophila*, *cubitus interruptus* (*ci*), also functions in the same way, suggesting this is an evolutionarily conserved signaling mechanism (Müller and Basler, 2000) where Gli/Ci transcription factors are the key sensors of the Shh gradient. The ventral most cells of the neural tube experience the highest levels of Shh signaling and thus have high levels of activator Gli proteins, while cells most distal in the Shh gradient exhibit high levels of repressor Gli

proteins. The balance between activator and repressor forms, in conjunction with early expressed proneural proteins, determines the neuronal fates at specific ventral positions. Once the early pattern is established, steady state levels of a complex gene regulatory network are reached (Lupo *et al.*, 2006; Nishi *et al.*, 2009).

The information passed on by the Gli code results in distinct gene expression patterns of homeodomain transcription factors at specific dorsoventral positions of the neural tube (Briscoe *et al.*, 2000; Bai *et al.*, 2004; Stamataki *et al.*, 2005). These Shh-responsive genes fall into two classes: class I genes are repressed by Shh signaling and class II genes require Shh signaling for their expression. In order to generate the neuronal classes in the ventral neural tube, each gene responds differently, such that some genes are repressed by very little Shh signaling and other genes require high levels of Shh signaling to be turned off. The same is also true for genes that are activated by Shh signaling, making some genes expressed only adjacent to the floor plate and some throughout the ventral half of the neural tube (Ruiz I Altaba *et al.*, 2003). Over time, sharp boundaries of expression are formed between pairs of class I and class II Shh-responsive genes through mutual antagonism, which defines the borders between each specific neuronal class.

There are four classes of ventral interneurons (v0-v3), as well as groups of motor neurons (MN), which are positioned (dorsal to ventral) v0, v1, v2, MN and v3 in the progenitor domain (Figure 1-2). The mutually antagonistic pairs of class

I and class II Shh-responsive targets that define these distinct progenitor boundaries have been well characterized. Dbx1 and Nkx6.2 define the v0-v1 border, Dbx2 and Nkx6.1 define the v1-v2 border, Irx3 and Olig2 define the v2-MN border and Pax6 and Nkx2.2 define the MN-v3 border (Briscoe *et al.*, 2000; Vallstedt *et al.*, 2001; Ulloa and Briscoe, 2007; Dessaud *et al.*, 2008). Ectopic expression or knockouts of any one of these progenitor domain transcription factors results in the loss of one or several specific neuronal classes and the concomitant expansion of an adjacent neuronal class. As these ventral neurons begin differentiation and exit the cell cycle, a new code of transcription factors is established: *Lhx1/5* and *Pax2* are expressed in v0-v2 neurons, *Evx1/2* in v0 neurons, *En1* in v1 neurons, *Vsx2* in v2 neurons, *Isl1/2* in motor neurons and *Sim1* in the ventral-most v3 neurons (Wilson and Maden, 2005; Yang *et al.*, 2006). The post-mitotic transcription factors typically do not form mutually antagonistic interactions with each other but instead regulate the acquisition of specific cellular characteristics.

The knowledge obtained from studying Shh-dependent patterning of the ventral neural tube has led to many similar discoveries in the dorsal neural tube, although some key differences exist. It has been clearly demonstrated that Bmp signaling occurs as a morphogenetic gradient (Liem *et al.*, 1997; Timmer *et al.*, 2002; Chizhikov and Millen, 2005) much like Shh signaling does. In addition to Bmp proteins, Wnt-like proteins are also important for patterning and cell proliferation in the dorsal neural tube (Megason and McMahon, 2002; Muroyama

et al., 2002; Chesnutt *et al.*, 2004). Analyses of mouse knockouts as well as ectopic expression in chick have revealed that dorsal neural tube patterning is more complex than a simple gradient model. The six classes of dorsal interneurons (dI1-dI6) that have been characterized (Figure 1-2) in the vertebrate neural tube (Caspery and Anderson, 2003; Helms and Johnson, 2003) fall into two groups – one that is dependent upon roof plate signals (class A) and another which still develops in the absence of a roof plate (class B). The class B neurons are typically referred to as intermediate neurons as they are positioned in the middle of the DV axis.

The development of dorsal neurons is similar to that of ventral neurons, where a progenitor domain exists in the interior of the neural tube and as differentiation occurs cells migrate to the mantle zone at the periphery. However, a late neuronal migration event occurs where dorsal born dI1-dI3 neuronal populations migrate ventrally and intermediate dI4-dI5 populations migrate laterally. A second group of late developing class B neurons (dI4L) migrate dorsally to occupy the entire dorsal half of the neural tube (Gross *et al.*, 2002; Müller *et al.*, 2002), which results in a significant reorganization of the dorsoventral positioning of neurons. Mutually antagonistic relationships also form in the dorsal neural tube, but between transcription factors of the basic helix-loop-helix (bHLH) protein family (Gowan *et al.*, 2001). Evidence suggests bHLH and homeodomain transcription factors work together to pattern dorsal neural tube progenitors. The Atonal homolog 1 (Ath1) specifies dI1 neurons, Neurogenin 1

and 2 (Ngn1/2) promote the dI2 and dI6 fate and the Achaete-Scute homolog 1 (Ash1) is important for dI3-dI5 neuron development (Timmer *et al.*, 2002; Caspery and Anderson, 2003; Helms and Johnson, 2003; Nakada *et al.*, 2004). The expression of these transcription factors is heavily dependent upon Bmp signaling from the roof plate.

Neurons of the dorsal neural tube also express a new set of transcription factors upon the onset of differentiation. However, there is a great amount of variability between their expression domains – some exhibit an alternating periodicity such as Lim domain proteins 1 and 2 (Lim1/2) and Lim homeobox proteins 1 and 5 (Lhx1/5) in dI2, dI4 and dI6 neurons (Gowan *et al.*, 2001). Others are responsible for the development of a specific class of neurons, such as the ladybird homeodomain 1 (Lbx1) protein, which is expressed in dI4-dI6 neurons (Gross *et al.*, 2002; Muller *et al.*, 2002). Other important neuronal differentiation factors include a variety of homeodomain-containing proteins, *Brn3a* in dI1-dI3 and dI5 interneurons, *Tlx3* in dI3 and dI5 neurons, *Pax2* in dI4 neurons, *Dbx2* in dI6 neurons, *Islet1* in dI3 neurons, *Lhx2/9* in dI1 neurons and a basic helix-loop-helix protein *Olig3* in dI1-dI3 neurons. Several of these factors are also expressed in ventral interneurons such as *Olig3* in v0, v2 and v3 neurons (Takebayashi *et al.*, 2002), *Dbx2* in v0-v1 neurons, *Pax2* and *Lhx1/5* in v0-v2 and *Islet1* in motor neurons, suggesting there are common features between dorsal and ventral neurons despite the differences in their initial patterning.

Evidence that challenges the morphogenetic gradient theory of dorsal neural tube patterning is the characterization of Lbx1-deficient mice, where an intermediate class of spinal cord neurons acquires characteristics of more dorsal neurons. It is thought that $Lbx1^{+}$ neurons (dI4-dI6) develop independently of morphogens released from the roof plate since mice lacking a roof plate posses dI4-like neurons positioned at the dorsal edge of the neural tube (Müller *et al.*, 2005). However, in $Lbx1^{GFP/GFP}$ mice (Gross *et al.*, 2002) or $Lbx1^{LacZ/LacZ}$ mice (Müller *et al.*, 2002), gene expression patterns indicative of dI4 and dI5 neuron specification was lost and replaced by a transcription factor code similar to that of dI2 and dI3 dorsal interneurons. The dI6 neurons appear to be respecified also but show molecular markers of both dI4 and dI6 neurons. These results argue against a simple morphogenetic gradient in the dorsal neural tube as neurons positioned at different distances from the morphogen source develop similarly.

It is of interest to determine whether Lbx1 can block the effects of roof plate-derived morphogens on the development of more intermediate (dI4-dI6) interneurons. It is known that the bHLH transcription factor Olig3 controls the development of dorsal dI1-dI3 interneurons and relies on Bmp and Wnt signaling from the roof plate for its expression in the ventricular zone (Müller *et al.*, 2005; Zechner *et al.*, 2007). The main function of Olig3 is to repress Lbx1 expression and thereby potentially permit morphogens to act upon the dI1-dI3 neuronal populations (Ding *et al.*, 2005; Müller *et al.*, 2005). The loss of Olig3 results in a dorsal expansion of Lbx1 expression and consequently, more intermediate

neuronal populations, which suggests that the mechanism distinguishing dorsal dI1-dI3 neurons and intermediate dI4-dI6 neurons is controlled by the mutual repression of Olig3 and Lbx1. Lbx1 expression is initiated once neurons transition to a post-mitotic character and exit the ventricular zone while Olig3 is only expressed in the progenitor domain. The entire dorsal half of the neural tube in $Lbx1^{-/-}$ $Olig3^{-/-}$ double mutant mice is occupied by class A (dI1-dI3) neurons (Müller *et al.*, 2005), suggesting Lbx1 functions to inhibit dI1-dI3 neuronal characteristics in intermediate neuronal populations. These observations also imply that Olig3 is not required for the interpretation of roof plate signals but instead acts to prevent Lbx1 expression in class A dorsal interneurons.

Mouse transgenic strains bearing gene replacements of *Lbx1* have provided much insight into the functions of Lbx1. Studies have shown that Lbx1 forms a mutually antagonistic relationship with Tlx3 in determining the neurotransmitter phenotype of specific dorsal interneurons of the spinal cord (Cheng *et al.*, 2005). Lbx1 also selects for a somatosensory (external sensation) function over a viscerosensory (internal organ sensation) function (Gross *et al.*, 2002; Müller *et al.* 2002; Sieber *et al.*, 2007). However, the regulatory input that drives Lbx1 expression in intermediate neurons while excluding expression in dorsal or ventral neurons remains unclear. Mouse transgenics have also been taken advantage of to elucidate genomic regulatory elements driving gene expression but the precise region controlling the dorsoventral expression domain of Lbx1 has not been examined. Zebrafish transgenic technology provides a much

quicker and consistently reliable assessment of potential regulatory function of genomic sequences (Kawakami, 2005; Fisher *et al.*, 2006). The presence of duplicated *Lbx1* orthologs in zebrafish and, hypothetically, their regulatory elements, further advances the attractiveness of using zebrafish to identify the genomic region(s) responsible for controlling *Lbx1* expression. Once the regulatory elements of Lbx1 have been identified, determining the transcription factor input should become more straightforward and will give clues as to why *Lbx1* is only expressed in intermediate dorsal interneurons.

5) Using Zebrafish to Understand Dorsal Interneuron Development

The teleost fish *Danio rerio* (zebrafish) has been receiving more and more attention as an outstanding model organism for studying vertebrate development. With the advances in transgenic technology (Jessen *et al.*, 1998; Kawakami, 2007), gene knockdown or ectopic expression (Robu *et al.*, 2007; Yuan and Sun, 2009) and fluorescent mRNA *in situ* hybridization (Clay and Ramakrishnan, 2005; Brend and Holley, 2009) much work has been done to validate to use of zebrafish for studying vertebrate development and disease. The molecular mechanisms of axial patterning are highly conserved between zebrafish and mammals (Schier and Talbot, 2005), including neural tube formation (Clarke, 2009), *Hox*-dependent anteroposterior patterning (Moens and Prince, 2002) and development of the central nervous system (Blader and Stähle, 2000; Lewis and Eisen, 2003). Bmp signaling has been shown to be critical for the development of

the dorsal neural tube (Nguyen *et al.*, 2000) and Gli proteins are also necessary to interpret the Shh gradient established by the notochord and floor plate (Tyurina *et al.*, 2005; Vanderlaan *et al.*, 2005).

Despite the significant similarity between teleost and mammalian embryonic development, very little is known about the molecular and functional similarities of neurons in the spinal cord, especially dorsal neurons. The expression pattern of *olig* genes (Tiso *et al.*, 2009) and other markers of the dorsoventral axis such as *nkx2.2*, *vsx1/2*, *islet1/2*, *irx3*, *eng1b*, *nkx6.1*, *lhx3* and neurotransmitter markers (Park *et al.*, 2004; Batista *et al.*, 2008) appear to be very similar to those observed in the mouse, but the corresponding neuronal function has not been compared in depth. The major limitation for comparative studies between zebrafish and mouse has been the methodology used to distinguish neuronal sub-types. In the mouse, specific combinatorial codes of expressed transcription factors and other neuronal-specific genes have been used to sub-divide the dorsoventral axis of the neural tube into molecular domains. This molecular segregation has led to the characterization of neuronal function and morphology, whereas in zebrafish, neurons have been primarily characterized and labeled by their morphology and axonal projections as opposed to the genes they express (Bernhardt *et al.*, 1990; Hale *et al.*, 2001; Downes *et al.*, 2002; Higashijima *et al.*, 2004; McLean and Fetcho, 2007).

Additionally, some neuronal classes exist as 1-2 cells along the length of a trunk segment in zebrafish, which normally contain 5 cells (Henry *et al.*, 2002),

thus it is difficult to compile dorsoventral gene expression data. In order to definitively compare and validate zebrafish as a model for vertebrate neuronal specification and function we need to understand the molecular mechanisms that create neuronal morphology, function and axonal trajectories. This will be facilitated by the creation of transgenic zebrafish where one can follow specific neuronal classes in live embryos and examine their phenotypes with subsequent fluorescent labeling of molecular markers. Compound transgenic strains will also be able to further subdivide broad neuronal domains into more distinct classes.

To assess the regulatory input on genes expressed in specific dorsoventral domains within the vertebrate neural tube, the zebrafish *ladybird* homeodomain-encoding genes have been examined. Three *ladybird*-like genes were identified and their mRNA expression patterns markedly resemble murine *Lbx1*. Comparative genomics and the hypothesized genome duplication event in teleost fish identified several highly conserved sequence blocks surrounding *ladybird* loci. These conserved genomic sequences were tested for enhancer activity with a Tol2 plasmid-based transgenic assay where cloned sequences were placed upstream of a fluorescent reporter gene and a minimal promoter. Transgenic constructs were injected into zebrafish at the 1-cell stage and embryos were examined up to 2 weeks post-fertilization. The transgene expression patterns generated from some of the cloned genomic regions were remarkably similar to that of the endogenous *ladybird* locus. Additionally, the corresponding region from the human genome was also able to drive reporter expression in a similar

pattern. Constructs with robust enhancer activity were used to create stable transgenic strains of zebrafish where a fluorescent reporter protein marks a subset of dorsally-located neurons in the hindbrain and spinal cord. These enhancers were further broken down into fragments or engineered with small deletions to determine the location of specific regulatory elements. It was determined that some regions specifically conveyed expression in the hindbrain while others established spinal cord expression.

6) References

Acampora D, Avantaggiato V, Tuorto F, Barone P, Reichert H, Finkelstein R and Simeone A, **1998**. Murine Otx1 and Drosophila otd genes share conserved genetic functions required in invertebrate and vertebrate brain development. *Development*. 125:1691-1702.

Akin ZN and Nazarali AJ, **2005**. Hox genes and their candidate downstream targets in the developing central nervous system. *Cell Mol Neurobiol*. 25:697-741.

Alexander T, Nolte C and Krumlauf R, **2009**. Hox genes and segmentation of the hindbrain and axial skeleton. *Annu Rev Cell Dev Biol*. 25:431-456.

Alvarez-Medina R, Cayuso J, Okubo T, Takada S and Marti E, **2008**. Wnt canonical pathway restricts graded Shh/Gli patterning activity through the regulation of Gli3 expression. *Development*. 135:237-247.

Bai CB, Stephen D and Joyner AL, **2004**. All mouse ventral spinal cord patterning by hedgehog is Gli dependent and involves an activator function of Gli3. *Dev Cell*. 6:103-115.

Beddington RS, **1994**. Induction of a second neural axis by the mouse node. *Development*. 120:613-620.

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Pena-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML and Hughes TR, **2008**. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*. 133:1266-1276.

Blader P and Strahle U, **2000**. Zebrafish developmental genetics and central nervous system development. *Hum Mol Genet*. 9:945-951.

Boettger T, Knoetgen H, Wittler L and Kessel M, **2001**. The avian organizer. *Int J Dev Biol*. 45:281-287.

Boncinelli E, Acampora D, Pannese M, D'Esposito M, Somma R, Gaudino G, Stornaiuolo A, Cafiero M, Faiella A and Simeone A, **1989**. Organization of human class I homeobox genes. *Genome*. 31:745-756.

Brend T and Holley SA, **2009**. Zebrafish whole mount high-resolution double fluorescent in situ hybridization. *J Vis Exp*. 25.

Briscoe J, Pierani A, Jessell TM and Ericson J, **2000**. A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. *Cell*. 101:435-445.

Briscoe J, Sussel L, Serup P, Hartigan-O'Connor D, Jessell TM, Rubenstein JL and Ericson J, **1999**. Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. *Nature*. 398:622-627.

Buglino JA and Resh MD, **2008**. Hhat is a palmitoylacyltransferase with specificity for N-palmitoylation of Sonic Hedgehog. *J Biol Chem*. 283:22076-22088.

Bumcrot DA, Takada R and McMahon AP, **1995**. Proteolytic processing yields two secreted forms of sonic hedgehog. *Mol Cell Biol*. 15:2294-2303.

Campo-Paysaa F, Marletaz F, Laudet V and Schubert M, **2008**. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. *Genesis*. 46:640-656.

Caspary T and Anderson KV, **2003**. Patterning cell types in the dorsal spinal cord: what the mouse mutants say. *Nat Rev Neurosci*. 4:289-297.

Chang CP, Shen WF, Rozenfeld S, Lawrence HJ, Largman C and Cleary ML, **1995**. Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev*. 9:663-674.

Chen MH, Li YJ, Kawakami T, Xu SM and Chuang PT, **2004**. Palmitoylation is required for the production of a soluble multimeric Hedgehog protein complex and long-range signaling in vertebrates. *Genes Dev*. 18:641-659.

Chesnutt C, Burrus LW, Brown AM and Niswander L, **2004**. Coordinate regulation of neural tube patterning and proliferation by TGFbeta and WNT activity. *Dev Biol*. 274:334-347.

Chizhikov VV and Millen KJ, **2004a**. Mechanisms of roof plate formation in the vertebrate CNS. *Nat Rev Neurosci*. 5:808-812.

Chizhikov VV and Millen KJ, **2004b**. Control of roof plate formation by Lmx1a in the developing spinal cord. *Development*. 131:2693-2705.

Chizhikov VV and Millen KJ, **2005**. Roof plate-dependent patterning of the vertebrate dorsal central nervous system. *Dev Biol*. 277:287-295.

Clark RM, Marker PC, Roessler E, Dutra A, Schimenti JC, Muenke M and Kingsley DM, **2001**. Reciprocal mouse and human limb phenotypes caused by gain- and loss-of-function mutations affecting Lmbr1. *Genetics*. 159:715-726.

Clarke J, **2009**. Role of polarized cell divisions in zebrafish neural tube formation. *Curr Opin Neurobiol*. 19:134-138.

Clay H and Ramakrishnan L, **2005**. Multiplex fluorescent in situ hybridization in zebrafish embryos using tyramide signal amplification. *Zebrafish*. 2:105-111.

Colas JF and Schoenwolf GC, **2001**. Towards a cellular and molecular understanding of neurulation. *Dev Dyn*. 221:117-145.

Colon-Ramos DA, **2009**. Synapse formation in developing neural circuits. *Curr Top Dev Biol*. 87:53-79.

Cooke JE and Moens CB, **2002**. Boundary formation in the hindbrain: Eph only it were simple. *Trends Neurosci*. 25:260-267.

Dasen JS, **2009**. Transcriptional networks in the early development of sensory-motor circuits. *Curr Top Dev Biol*. 87:119-148.

De Robertis EM, **2008**. Evo-devo: variations on ancestral themes. *Cell*. 132:185-195.

De Robertis EM, **2009**. Spemann's organizer and the self-regulation of embryonic fields. *Mech Dev*. doi:10.1016/j mod.2009.08.004.

De Robertis EM and Kuroda H, **2004**. Dorsal-ventral patterning and neural induction in Xenopus embryos. *Annu Rev Cell Dev Biol*. 20:285-308.

Dessaud E, McMahon AP and Briscoe J, **2008**. Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development*. 135:2489-2503.

Ding L, Takebayashi H, Watanabe K, Ohtsuki T, Tanaka KF, Nabeshima Y, Chisaka O, Ikenaka K and Ono K, **2005**. Short-term lineage analysis of dorsally derived Olig3 cells in the developing spinal cord. *Dev Dyn*. 234:622-632.

Downes GB, Waterbury JA and Granato M, **2002**. Rapid in vivo labeling of identified zebrafish neurons. *Genesis*. 34:196-202.

Duboule D and Dolle P, **1989**. The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. *EMBO J*. 8:1497-1505.

Echelard Y, Vassileva G and McMahon AP, **1994**. Cis-acting regulatory sequences governing Wnt-1 expression in the developing mouse CNS. *Development*. 120:2213-2224.

Erickson T, Scholpp S, Brand M, Moens CB and Waskiewicz AJ, **2007**. Pbx proteins cooperate with Engrailed to pattern the midbrain-hindbrain and diencephalic-mesencephalic boundaries. *Dev Biol*. 301:504-517.

Ferguson EL, **1996**. Conservation of dorsal-ventral patterning in arthropods and chordates. *Curr Opin Genet Dev*. 6:424-431.

Frasch M, Chen X and Lufkin T, **1995**. Evolutionary-conserved enhancers direct region-specific expression of the murine Hoxa-1 and Hoxa-2 loci in both mice and Drosophila. *Development*. 121:957-974.

Goulding M, **2009**. Circuits controlling vertebrate locomotion: moving in a new direction. *Nat Rev Neurosci*. 10:507-518.

Gowan K, Helms AW, Hunsaker TL, Collisson T, Ebert PJ, Odom R and Johnson JE, **2001**. Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons. *Neuron*. 31:219-232.

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

Gross MK, Dottori M and Goulding M, **2002**. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron*. 34:535-549.

Harland R and Gerhart J, **1997**. Formation and function of Spemann's organizer. *Annu Rev Cell Dev Biol*. 13:611-667.

Harrington MJ, Hong E and Brewster R, **2009**. Comparative analysis of neurulation: first impressions do not count. *Mol Reprod Dev*. 76:954-965.

Helms AW and Johnson JE, **2003**. Specification of dorsal spinal cord interneurons. *Curr Opin Neurobiol*. 13:42-49.

Hemmati-Brivanlou A, Kelly OG and Melton DA, **1994**. Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neuralizing activity. *Cell*. 77:283-295.

Henry CA, Urban MK, Dill KK, Merlie JP, Page MF, Kimmel CB and Amacher SL, **2002**. Two linked hairy/Enhancer of split-related zebrafish genes, her1 and her7, function together to refine alternating somite boundaries. *Development*. 129:3693-3704.

Hernandez RE, Putzke AP, Myers JP, Margaretha L and Moens CB, **2007**. Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development. *Development*. 134:177-187.

Hernandez RE, Rikhof HA, Bachmann R and Moens CB, **2004**. vhnf1 integrates global RA patterning and local FGF signals to direct posterior hindbrain development in zebrafish. *Development*. 131:4511-4520.

Hidalgo-Sanchez M, Millet S, Bloch-Gallego E and Alvarado-Mallart RM, **2005**. Specification of the meso-isthmo-cerebellar region: the Otx2/Gbx2 boundary. *Brain Res Brain Res Rev*. 49:134-149.

Holley SA, Jackson PD, Sasai Y, Lu B, De Robertis EM, Hoffmann FM and Ferguson EL, **1995**. A conserved system for dorsal-ventral patterning in insects and vertebrates involving sog and chordin. *Nature*. 376:249-253.

Houart C, Westerfield M and Wilson SW, **1998**. A small population of anterior cells patterns the forebrain during zebrafish gastrulation. *Nature*. 391:788-792.

Huang X, Litingtung Y and Chiang C, **2007**. Region-specific requirement for cholesterol modification of sonic hedgehog in patterning the telencephalon and spinal cord. *Development*. 134:2095-2105.

Hueber SD and Lohmann I, **2008**. Shaping segments: Hox gene function in the genomic age. *Bioessays*. 30:965-979.

Jabet C, Gitti R, Summers MF and Wolberger C, **1999**. NMR studies of the pbx1 TALE homeodomain protein free in solution and bound to DNA: proposal for a mechanism of HoxB1-Pbx1-DNA complex assembly. *J Mol Biol*. 291:521-530.

Jacob J and Briscoe J, **2003**. Gli proteins and the control of spinal-cord patterning. *EMBO Rep*. 4:761-765.

Jacobs Y, Schnabel CA and Cleary ML, **1999**. Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity. *Mol Cell Biol*. 19:5134-5142.

Kaufman TC, Lewis R and Wakimoto B, **1980**. Cytogenetic Analysis of Chromosome 3 in DROSOPHILA MELANOGASTER: The Homoeotic Gene Complex in Polytene Chromosome Interval 84a-B. *Genetics*. 94:115-133.

Kawano Y and Kypta R, **2003**. Secreted antagonists of the Wnt signalling pathway. *J Cell Sci*. 116:2627-2634.

Kiecker C and Lumsden A, **2005**. Compartments and their boundaries in vertebrate brain development. *Nat Rev Neurosci*. 6:553-564.

Kim CH, Oda T, Itoh M, Jiang D, Artinger KB, Chandrasekharappa SC, Driever W and Chitnis AB, **2000**. Repressor activity of Headless/Tcf3 is essential for vertebrate head formation. *Nature*. 407:913-916.

Kmita M and Duboule D, **2003**. Organizing axes in time and space; 25 years of colinear tinkering. *Science*. 301:331-333.

Kmita M, Fraudeau N, Herault Y and Duboule D, **2002**. Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs. *Nature*. 420:145-150.

Knoepfler PS and Kamps MP, **1995**. The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. *Mol Cell Biol*. 15:5811-5819.

Lagutin OV, Zhu CC, Kobayashi D, Topczewski J, Shimamura K, Puelles L, Russell HR, McKinnon PJ, Solnica-Krezel L and Oliver G, **2003**. Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes Dev*. 17:368-379.

Lamb TM, Knecht AK, Smith WC, Stachel SE, Economides AN, Stahl N, Yancopolous GD and Harland RM, **1993**. Neural induction by the secreted polypeptide noggin. *Science*. 262:713-718.

Lane S, McDermott K, Dockery P and Fraher J, **2004**. The developing cervical spinal ventral commissure of the rat: a highly controlled axon-glial system. *J Neurocytol*. 33:489-501.

Lee KJ, Dietrich P and Jessell TM, **2000**. Genetic ablation reveals that the roof plate is essential for dorsal interneuron specification. *Nature*. 403:734-740.

Lee KJ and Jessell TM, **1999**. The specification of dorsal cell fates in the vertebrate central nervous system. *Annu Rev Neurosci*. 22:261-294.

Lee KJ, Mendelsohn M and Jessell TM, **1998**. Neuronal patterning by BMPs: a requirement for GDF7 in the generation of a discrete class of commissural interneurons in the mouse spinal cord. *Genes Dev*. 12:3394-3407.

Lemons D and McGinnis W, **2006**. Genomic evolution of Hox gene clusters. *Science*. 313:1918-1922.

Lengler J and Graw J, **2001**. Regulation of the human SIX3 gene promoter. *Biochem Biophys Res Commun*. 287:372-376.

Leuzinger S, Hirth F, Gerlich D, Acampora D, Simeone A, Gehring WJ, Finkelstein R, Furukubo-Tokunaga K and Reichert H, **1998**. Equivalence of the fly orthodenticle gene and the human OTX genes in embryonic brain development of Drosophila. *Development*. 125:1703-1710.

Lewis EB, **1978**. A gene complex controlling segmentation in Drosophila. *Nature*. 276:565-570.

Lewis KE and Eisen JS, **2003**. From cells to circuits: development of the zebrafish spinal cord. *Prog Neurobiol*. 69:419-449.

Liem KF, Jr., Tremml G and Jessell TM, **1997**. A role for the roof plate and its resident TGFbeta-related proteins in neuronal patterning in the dorsal spinal cord. *Cell*. 91:127-138.

Litingtung Y and Chiang C, **2000**. Control of Shh activity and signaling in the neural tube. *Dev Dyn*. 219:143-154.

Lowery LA and Sive H, **2004**. Strategies of vertebrate neurulation and a re-evaluation of teleost neural tube formation. *Mech Dev*. 121:1189-1197.

Lupo G, Harris WA and Lewis KE, **2006**. Mechanisms of ventral patterning in the vertebrate nervous system. *Nat Rev Neurosci*. 7:103-114.

Manuel M and Price DJ, **2005**. Role of Pax6 in forebrain regionalization. *Brain Res Bull*. 66:387-393.

Marshall H, Studer M, Popperl H, Aparicio S, Kuroiwa A, Brenner S and Krumlauf R, **1994**. A conserved retinoic acid response element required for early expression of the homeobox gene Hoxb-1. *Nature*. 370:567-571.

Marti E and Bovolenta P, **2002**. Sonic hedgehog in CNS development: one signal, multiple outputs. *Trends Neurosci*. 25:89-96.

Martin BL and Kimelman D, **2009**. Wnt signaling and the evolution of embryonic posterior development. *Curr Biol*. 19:R215-219.

McGinnis W, Hart CP, Gehring WJ and Ruddle FH, **1984**. Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of Drosophila. *Cell*. 38:675-680.

Megason SG and McMahon AP, **2002**. A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. *Development*. 129:2087-2098.

Meyer NP and Roelink H, **2003**. The amino-terminal region of Gli3 antagonizes the Shh response and acts in dorsoventral fate specification in the developing spinal cord. *Dev Biol*. 257:343-355.

Moens CB and Prince VE, **2002**. Constructing the hindbrain: insights from the zebrafish. *Dev Dyn*. 224:1-17.

Muller B and Basler K, **2000**. The repressor and activator forms of Cubitus interruptus control Hedgehog target genes through common generic gli-binding sites. *Development*. 127:2999-3007.

Muller T, Anlag K, Wildner H, Britsch S, Treier M and Birchmeier C, **2005**. The bHLH factor Olig3 coordinates the specification of dorsal neurons in the spinal cord. *Genes Dev*. 19:733-743.

Muller T, Brohmann H, Pierani A, Heppenstall PA, Lewin GR, Jessell TM and Birchmeier C, **2002**. The homeodomain factor lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron*. 34:551-562.

Muroyama Y, Fujihara M, Ikeya M, Kondoh H and Takada S, **2002**. Wnt signaling plays an essential role in neuronal specification of the dorsal spinal cord. *Genes Dev*. 16:548-553.

Nakada Y, Hunsaker TL, Henke RM and Johnson JE, **2004**. Distinct domains within Mash1 and Math1 are required for function in neuronal differentiation versus neuronal cell-type specification. *Development*. 131:1319-1330.

Nakamura H, **2001**. Regionalization of the optic tectum: combinations of gene expression that define the tectum. *Trends Neurosci*. 24:32-39.

Narita Y and Rijli FM, **2009**. Hox genes in neural patterning and circuit formation in the mouse hindbrain. *Curr Top Dev Biol*. 88:139-167.

Nguyen VH, Trout J, Connors SA, Andermann P, Weinberg E and Mullins MC, **2000**. Dorsal and intermediate neuronal cell types of the spinal cord are established by a BMP signaling pathway. *Development*. 127:1209-1220.

Niehrs C, **2006**. Function and biological roles of the Dickkopf family of Wnt modulators. *Oncogene*. 25:7469-7481.

Nishi Y, Ji H, Wong WH, McMahon AP and Vokes SA, **2009**. Modeling the spatio-temporal network that drives patterning in the vertebrate central nervous system. *Biochim Biophys Acta*. 1789:299-305.

Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH and Wolfe SA, **2008**. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell*. 133:1277-1289.

Oelgeschlager M, Kuroda H, Reversade B and De Robertis EM, **2003**. Chordin is required for the Spemann organizer transplantation phenomenon in Xenopus embryos. *Dev Cell*. 4:219-230.

Okafuji T, Funahashi J and Nakamura H, **1999**. Roles of Pax-2 in initiation of the chick tectal development. *Brain Res Dev Brain Res*. 116:41-49.

Park HC, Shin J and Appel B, **2004**. Spatial and temporal regulation of ventral spinal cord precursor specification by Hedgehog signaling. *Development*. 131:5959-5969.

Peltenburg LT and Murre C, **1996**. Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J*. 15:3385-3393.

Pera EM and Kessel M, **1997**. Patterning of the chick forebrain anlage by the prechordal plate. *Development*. 124:4153-4162.

Persson M, Stamataki D, te Welscher P, Andersson E, Bose J, Ruther U, Ericson J and Briscoe J, **2002**. Dorsal-ventral patterning of the spinal cord requires Gli3 transcriptional repressor activity. *Genes Dev*. 16:2865-2878.

Popperl H, Schmidt C, Wilson V, Hume CR, Dodd J, Krumlauf R and Beddington RS, **1997**. Misexpression of Cwnt8C in the mouse induces an ectopic embryonic axis and causes a truncation of the anterior neuroectoderm. *Development*. 124:2997-3005.

Prakash N and Wurst W, **2004**. Specification of midbrain territory. *Cell Tissue Res*. 318:5-14.

Reichert H, **2002**. Conserved genetic mechanisms for embryonic brain patterning. *Int J Dev Biol*. 46:81-87.

Rhinn M and Brand M, **2001**. The midbrain--hindbrain boundary organizer. *Curr Opin Neurobiol*. 11:34-42.

Rieckhof GE, Casares F, Ryoo HD, Abu-Shaar M and Mann RS, **1997**. Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. *Cell*. 91:171-183.

Robu ME, Larson JD, Nasevicius A, Beiraghi S, Brenner C, Farber SA and Ekker SC, **2007**. p53 activation by knockdown technologies. *PLoS Genet*. 3:e78.

Rogers CD, Moody SA and Casey ES, **2009**. Neural induction and factors that stabilize a neural fate. *Birth Defects Res C Embryo Today*. 87:249-262.

Ruel L and Therond PP, **2009**. Variations in Hedgehog signaling: divergence and perpetuation in Sufu regulation of Gli. *Genes Dev*. 23:1843-1848.

Ruiz i Altaba A, **1999**. Gli proteins encode context-dependent positive and negative functions: implications for development and disease. *Development*. 126:3205-3216.

Ruiz i Altaba A, Mas C and Stecca B, **2007**. The Gli code: an information nexus regulating cell fate, stemness and cancer. *Trends Cell Biol*. 17:438-447.

Ruiz i Altaba A, Nguyen V and Palma V, **2003**. The emergent design of the neural tube: prepattern, SHH morphogen and GLI code. *Curr Opin Genet Dev*. 13:513-521.

Sasai Y, Lu B, Steinbeisser H and De Robertis EM, **1995**. Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in Xenopus. *Nature*. 377:757.

Sasai Y, Lu B, Steinbeisser H, Geissert D, Gont LK and De Robertis EM, **1994**. Xenopus chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell*. 79:779-790.

Schier AF and Talbot WS, **2005**. Molecular genetics of axis formation in zebrafish. *Annu Rev Genet*. 39:561-613.

Schuurmans C and Guillemot F, **2002**. Molecular mechanisms underlying cell fate specification in the developing telencephalon. *Curr Opin Neurobiol*. 12:26-34.

Shih J and Fraser SE, **1996**. Characterizing the zebrafish organizer: microsurgical analysis at the early-shield stage. *Development*. 122:1313-1322.

Shimamura K and Rubenstein JL, **1997**. Inductive interactions direct early regionalization of the mouse forebrain. *Development*. 124:2709-2718.

Simeone A, Puelles E and Acampora D, **2002**. The Otx family. *Curr Opin Genet Dev*. 12:409-415.

Spemann H, **1938**. Embryonic Development and Induction. *Yale University Press, New Haven, Conn*.

Spemann H and Mangold H, **1924**. Induction of embryonic primordial by implantation of organizers from a different species. *Roux's Arch Entw Mech*. 100:599-638.

Stamataki D, Ulloa F, Tsoni SV, Mynett A and Briscoe J, **2005**. A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. *Genes Dev*. 19:626-641.

Stern CD, **2005**. Neural induction: old problem, new findings, yet more questions. *Development*. 132:2007-2021.

Strahle U, Lam CS, Ertzer R and Rastegar S, **2004**. Vertebrate floor-plate specification: variations on common themes. *Trends Genet*. 20:155-162.

Takebayashi H, Ohtsuki T, Uchida T, Kawamoto S, Okubo K, Ikenaka K, Takeichi M, Chisaka O and Nabeshima Y, **2002**. Non-overlapping expression of Olig3 and Olig2 in the embryonic neural tube. *Mech Dev*. 113:169-174.

Timmer JR, Wang C and Niswander L, **2002**. BMP signaling patterns the dorsal and intermediate neural tube via regulation of homeobox and helix-loop-helix transcription factors. *Development*. 129:2459-2472.

Tiso N, Filippi A, Benato F, Negrisolo E, Modena N, Vaccari E, Driever W and Argenton F, **2009**. Differential expression and regulation of olig genes in zebrafish. *J Comp Neurol*. 515:378-396.

Totori-Donati P, Rossi A and Biancheri R, **2005**. Brain Malformations. *Pediatric Neuroradiology: Brain, Head, Neck and Spine*. 1:92-95.

Tumpel S, Wiedemann LM and Krumlauf R, **2009**. Hox genes and segmentation of the vertebrate hindbrain. *Curr Top Dev Biol*. 88:103-137.

Tyurina OV, Guner B, Popova E, Feng J, Schier AF, Kohtz JD and Karlstrom RO, **2005**. Zebrafish Gli3 functions as both an activator and a repressor in Hedgehog signaling. *Dev Biol*. 277:537-556.

Ulloa F and Briscoe J, **2007**. Morphogens and the control of cell proliferation and patterning in the spinal cord. *Cell Cycle*. 6:2640-2649.

Ulloa F and Marti E, **2009**. Wnt won the war: Antagonistic role of Wnt over Shh controls dorso-ventral patterning of the vertebrate neural tube. *Dev Dyn*. 239:69-76.

Vallstedt A, Muhr J, Pattyn A, Pierani A, Mendelsohn M, Sander M, Jessell TM and Ericson J, **2001**. Different levels of repressor activity assign redundant and specific roles to Nkx6 genes in motor neuron and interneuron specification. *Neuron*. 31:743-755.

van der Zwaag A, **2006**. Congenital Cranial Dysinnervation Disorders. Nijmagen.

Vanderlaan G, Tyurina OV, Karlstrom RO and Chandrasekhar A, **2005**. Gli function is essential for motor neuron induction in zebrafish. *Dev Biol*. 282:550-570.

Waskiewicz AJ, Rikhof HA, Hernandez RE and Moens CB, **2001**. Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning. *Development*. 128:4139-4151.

Waskiewicz AJ, Rikhof HA and Moens CB, **2002**. Eliminating zebrafish pbx proteins reveals a hindbrain ground state. *Dev Cell*. 3:723-733.

Wellik DM, **2009**. Hox genes and vertebrate axial pattern. *Curr Top Dev Biol*. 88:257-278.

Wilson L, Gale E, Chambers D and Maden M, **2004**. Retinoic acid and the control of dorsoventral patterning in the avian spinal cord. *Dev Biol*. 269:433-446.

Wilson L and Maden M, **2005**. The mechanisms of dorsoventral patterning in the vertebrate neural tube. *Dev Biol*. 282:1-13.

Wurst W and Bally-Cuif L, **2001**. Neural plate patterning: upstream and downstream of the isthmic organizer. *Nat Rev Neurosci*. 2:99-108.

Xu Q, Mellitzer G and Wilkinson DG, **2000**. Roles of Eph receptors and ephrins in segmental patterning. *Philos Trans R Soc Lond B Biol Sci*. 355:993-1002.

Yamaguchi TP, **2001**. Heads or tails: Wnts and anterior-posterior patterning. *Curr Biol*. 11:R713-724.

Yang X, Tomita T, Wines-Samuelson M, Beglopoulos V, Tansey MG, Kopan R and Shen J, **2006**. Notch1 signaling influences v2 interneuron and motor neuron development in the spinal cord. *Dev Neurosci*. 28:102-117.

Yuan S and Sun Z, **2009**. Microinjection of mRNA and morpholino antisense oligonucleotides in zebrafish embryos. *J Vis Exp*.

Zechner D, Muller T, Wende H, Walther I, Taketo MM, Crenshaw EB, 3rd, Treier M, Birchmeier W and Birchmeier C, **2007**. Bmp and Wnt/beta-catenin signals control expression of the transcription factor Olig3 and the specification of spinal cord neurons. *Dev Biol*. 303:181-190.

Zimmerman LB, De Jesus-Escobar JM and Harland RM, **1996**. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell*. 86:599-606.

1) Adult Zebrafish and Embryo Collection

Zebrafish were bred and maintained according to standard specifications and animal care guidelines. Embryos were raised in fish water (reverse osmosis water with Crystal Sea Marine salt mix to conductivity of 750 ± 100 µsiemens) or embryo media (15 mM NaCl, 500 nM KCl, 1 mM $CaCl_2$, 150 nM KH_2PO_4, 50 nM Na_2HPO_4, 1 mM $MgSO_4$, 715 nM $NaHCO_3$) at 28.5°C. Approximately 18 hours post-fertilization (hpf), embryo media containing 0.0045% (w/v) phenylthiourea (PTU) was substituted and subsequently replenished every 24 hours to inhibit pigment formation at later developmental stages. Embryos were staged using hours post-fertilization and morphological features such as somite number, yolk extension, and body axis curvature as guidelines (Kimmel *et al.*, 1995). For embryos older than 24 hpf, chorions were manually removed with fine tip (Dumont #5) forceps or enzymatically through digestion with 10-15 µg/mL pronase E (from *Streptomyces griseus*; Sigma) in embryo media for 1-2 hours at 28.5°C. Embryos were fixed in 4% (w/v) paraformaldehyde (PFA) in phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM NaH_2PO_4, 1.75 mM KH_2PO_4, pH 7.4) at room temperature for 4 hours or overnight at 4°C. Embryos were dehydrated in 100% methanol (MeOH) at room temperature for ~1 hour and rinsed again with 100% MeOH prior to storage at negative 20°C for >24 hrs. Embryos expressing fluorescent proteins were fixed for 2 hours at room temperature and stored in the dark to preserve fluorescence.

2) Cloning Zebrafish Ladybird Genes

To identify zebrafish *Ladybird*-like genes, Ensembl (Birney *et al.*, 2006; Hubbard *et al.*, 2007; Flicek *et al.*, 2008) and NCBI (Jenuth *et al.*, 2000; Pruitt *et al.*, 2005; Baxevanis, 2006; Pruitt *et al.*, 2007; Wolfsberg, 2007; Baxevanis, 2008; Johnson *et al.*, 2008) databases were used for Basic Local Alignment Search Tool (BLAST; Altschul *et al.*, 1990) analysis using mammalian *Ladybird homeobox-1* (*Lbx1*) mRNA and protein sequences. To distinguish between *Lbx* paralogs, *Lbx1* and *Lbx2*, and to increase the likelihood of identifying true *Lbx*-like genes, mRNA and amino acid sequences outside of the predicted homeodomain region were initially used in BLAST analysis. Three distinct *Lbx1*-like predicted open reading frames (ORF) were identified in zebrafish on linkage groups 1, 13 and 14. Forward primers were targeted to sequences upstream of the predicted start codon and reverse primers targeted to sequences downstream of the predicted stop codon of each of the three putative zebrafish *lbx* genes (Table 2-1). Total RNA was isolated from embryonic stages between 10 hpf and 72 hpf with TRIzol (Invitrogen) and used as a template in SuperScript III RT/PlatinumTaq One-Step reverse transcriptase-PCR (RT-PCR; Invitrogen). PCR products of each zebrafish *ladybird*-like gene were run on an agarose gel and products of expected size were isolated using a QIAquick gel extraction kit (Qiagen) and cloned into pCR 4-TOPO (Invitrogen). cDNA clones were transformed into Top10 chemically competent *E. coli* (Invitrogen) and both strands of isolated plasmids (Qiagen) were sequenced with DYEnamic ET (Amersham) on automated sequencers.

Table 2-1: Primers Designed to Clone Zebrafish *Ladybird* Genes, Generate mRNA *in situ* Probes and Identify Enhancer Elements

Cloning Zebrafish *Ladybird* Genes, *in situ* Probes and mRNA Overexpression

	Forward (5'-3')	Reverse (5'-3')
Lbx1a ORF (Clone/Sequence)	TTTTTCGGAGAACTGCAAGGAC	TGTGTGGGCAATCGTGCTTC
Lbx1b ORF (Clone/Sequence)	CAACATGCTGCTGATCGGCTAC	CACAACCTTGCCCTTGCATTG
Lbx2 ORF (Clone/Sequence)	GCTGTTTTGAGGCTCCTTGAGAC	AGTTCTTCCGTGCAAAAAGTCG
Lbx1a (PCR* *in situ* probe)	CGGTGGGCAATGTACCTTTTG	CTGAATACGACTCACTATAGGGGCCACAATGTTAAAAGGCACTCC
Lbx1b (*in situ* probe)	ATGGAGGACATGGAAGACGCTC	CACAACCTTGCCCTTGCATTG
Lbx2 (*in situ* probe)	GCTGTTTTGAGGCTCCTTGAGAC	AGGCGTTGATGTGTTCACGACC
Lbx1a ORF (T7TS expression)	CAGTAGATCTGCCACCATGACAAACCATCAAAGTCGTC	TGACACTAGTAATCAGACATCGACGTCTATTTCC
Lbx1b ORF (T7TS expression)	CAGTAGATCTGCCACCATGACCTCCAAAGAAGACGCGAAAG	TGACACTAGTAATCAGTCATCTACATCAATTTCCTCGTC
Lbx2 ORF (T7TS expression)	CAGTAGATCTGCCACCATGACCTCCAGCTCTAAAGACATG	TGACACTAGTCTGCGTTTAATCGTCCACCTC

Cloning Putative Regulatory Regions of Zebrafish *Ladybird* Genes

	Forward (5'-3')	Reverse (5'-3')
Lbx1b (promoter)	ACCTCGAGCAGTAGCTACAAACTTGCTAGTGTC	CTGGATCCGCTTTTGAAAATACGCTACAG
Lbx2 (promoter)	ACCTCGAGTTGTTGTCCAATAACTGATGCG	ACGGATCCTCAGTCCCAAGAAACTATTTCG
Lbx1a -US (enhancer)	TGGCTGCATCCTAAAACTACATCCAGTGTC	CTCCGAAAAACTGCCTAACAACAATGGC
Lbx1a -DS (enhancer)	AGAGCTTAACAAACCTGTTCACGACGG	CACCTGCTACTGACAGAATGCTTAGTG
Lbx1b -US (enhancer)	CTACCTGACTCCACAACACAATTAGCG	CCGTAGAATTAGATGTAGCCGATCAGCAGC
Lbx1b -DS-A (enhancer)	TGTAGCCCACTTTTATTCCCTGCACC	GCTCTAATTCCATCCATAGTGTGACTCCAG
Lbx1b -DS-B (enhancer)	TGAAGCGGCTGTAATAGCAGGATCAGAGC	CTTGTAGGTGAGCGATAAAAGGGGTTTCC
Lbx2 -US (enhancer)	ACGTCATGCTTTACTCAAGGCTGTG	CGAATAGAAGTCACCTAGCTCGCTC
Lbx2 -DS (enhancer)	GTCCAAAACAATGCAAGGGCAAGGTTGTG	CGATCTGCTCACAACAGACACAAATGGGC
Human *LBX1* (enhancer)	AAGAATGGGAAGAGAGGAGAAG	CGCTTCATTGGGATTCATG

Intended primer use is indicated in parentheses and restriction sites designed into the primers are underlined. For putative regulatory regions, US and DS denote conserved sequence upstream and downstream of the coding region, respectively. Primers designed to clone promoters have XhoI (ACTAGT) and BamHI (GGATCC) restriction sites (underlined) so that the entire Gateway cassette (GW) and minimal *cFos* promoter can be replaced in pTol2-GW:*cFos:EGFP* vector. *Since sub-cloning of the insert into pCR 4-TOPO was unsuccessful, the reverse primer for the *lbx1a* probe has a T7 Polymerase binding site (underlined) designed into the sequence so that purified PCR product could be used as a template for probe synthesis.

3) Zebrafish *Ladybird*-like Gene Sequence Analysis

Sequenced open reading frames corresponding to each zebrafish *ladybird* gene were conceptually translated and compared to known vertebrate Lbx protein sequences. Detailed analysis of genomic information available for other fish species, *Tetraodon nigroviridis* (*Tetraodon*), *Gasterosteus aculeatus* (Stickleback), *Takifugu rubripes* (Fugu), and *Oryzias latipes* (Medaka), was conducted to determine orthology of the three *lbx1*-like genes (currently designated *lbx1a, lbx1b* and *lbx2*) identified in zebrafish. Alignments were done using the ClustalW alignment algorithm (Thompson *et al.*, 2002) within MacVector v7.2.2 software. All three *ladybird* genes identified in fish species share significant identity and similarity to vertebrate *Lbx1* though considerably less sequence conservation to another *Ladybird* gene in mammals, *Lbx2*. A detailed linkage map of genes neighbouring each of the three zebrafish *lbx* genes was constructed and compared to the genomic location of human *LBX1* on chromosome 10 to determine synteny. The evolutionary relationship of vertebrate *Lbx* loci was assessed through phylogenetic analysis of Lbx1 and Lbx2 protein sequences and comparative genomics of nearby intergenic sequences.

4) mRNA *in situ* Hybridization

To determine the spatiotemporal expression pattern of genes during zebrafish embryonic development, antisense RNA probes were targeted to mRNA sequences of the gene of interest. When designing probes, regions of low identity

such as untranslated regions (UTR) or rare coding sequences were chosen to reduce the chance of probe cross-reactivity, and probe length was generally >600 bases. To design 3' UTR-specific probes for incomplete transcripts, genomic sequence downstream of the predicted stop codon was examined to locate putative transcriptional termination sequences (AATAAA) to determine the potential extent of 3' UTR sequences.

For zebrafish *lbx1a* specifically, a 592 NT probe targeting the last 208 NT of coding sequence, which is downstream of the predicted homeodomain, and 384 NT of presumptive 3' UTR was designed. The same strategy was applied to design a 523 NT probe to zebrafish *lbx1b*. Complete transcript information was available for zebrafish *lbx2*, allowing a 613 NT probe targeting the 5' UTR sequence and coding sequence upstream of the homeodomain to be designed. Probe sequences were RT-PCR-amplified (Invitrogen) using gene-specific primers (Table 2-1) from RNA isolated from various embryonic stages. Products of expected size were gel-extracted (Qiagen), cloned into pCR 4-TOPO, and sequenced as previously described. Plasmid preparations (Qiagen) of clones containing confirmed probe sequences were linearized with restriction endonucleases (RE) predicted to cut downstream of antisense sequences, purified by standard phenol-chloroform extraction, and used as template in probe synthesis. Labelled antisense RNA probes were synthesized from *in vitro* transcription reactions (Roche) containing digoxigenin- or fluorescein-conjugated UTP (Roche) using T7 or T3 RNA polymerases (Roche), and purified by lithium

chloride precipitation or post reaction clean up columns (Sigma). In the case of *lbx1a*, a T7 RNA polymerase binding site was added to the reverse primer, which provided purified PCR product to serve as template for antisense probe synthesis reactions.

Embryos in 100% MeOH were rehydrated through a series of MeOH:PBST (PBS, 0.1% Tween 20) washes (3:1, 1:1, 1:3) and then rinsed 4 times for 5 minutes in PBST. Embryos that had been fixed with intact chorions were manually de-chorionated with forceps (Dumont #5) during the previous PBST rinses. Embryo permeablization was done with proK (10 µg/mL proteinase K in PBST), which varied in time depending on embryonic stage as tissue thickness changes significantly from 10 hpf to 72 hpf.

Stage	Time
<12 hpf	none
12-14 hpf	1 min
15-17 hpf	2 min
18-22 hpf	3 min

Stage	Time
23-29 hpf	5 min
30-36 hpf	7.5 min
37-48 hpf	10 min
49-54 hpf	15 min

Stage	Time
55-64 hpf	20 min
65-74 hpf	30 min
75-96 hpf	45 min
>96 hpf	60 min

ProK solution was removed and embryos were re-fixed in 4% PFA for 20 minutes at room temperature. Embryos were rinsed 4 times for 5 minutes in PBST and then pre-hybridized at 65°C in hybridization solution (50% formamide, 5xSSC (standard saline citrate; 1x is 150 mM NaCl, 15 mM $Na_3C_3H_5O(CO_2)_3$), 50 µg/mL heparin, 500 µg/mL tRNA and 0.1% Tween 20, pH to 6.0 with citric acid) for at least 60 minutes with occasional mixing. Approximately 200 ng of

antisense probe was hybridized overnight at 65°C in 200-400 µL hyb-solution. In some cases, multiple RNA probes were hybridized to embryos at the same time. After hybridization, hyb-solution containing probe was carefully removed and stored at -20°C for reuse. Embryos were rinsed in hyb solution:2xSSC (2:1, 1:2, 0:1) at 65°C for 5 minutes each to remove the majority of remaining unbound probe. Three 20 minute stringency washes, once in 0.2xSSC, 0.1% Tween 20 and twice in 0.1xSSC, 0.1% Tween 20, were done at 65°C with occasional mixing to remove probe bound non-specifically.

Embryos were prepared for blocking with 5 minute rinses at room temperature in 0.2xSSC:PBST (2:1, 1:2, 0:1) with gentle agitation. Blocking was done at room temperature for at least one hour in blocking solution (PBST with 2% (v/v) sheep serum and 2 mg/mL bovine serum albumin) with gentle agitation. Incubation with alkaline phosphatase conjugated anti-digoxigenin antibody (1:5000; Roche) was done overnight at 4°C in blocking solution with constant agitation. Excess antibody was rinsed out with 5 successive 15 minute PBST washes at room temperature and prior to colouration embryos were equilibrated with 4 successive 5 minute rinses in colour buffer (100 mM Tris-HCl pH 9.5, 50 mM $MgCl_2$, 100 mM NaCl, 0.1% Tween 20). Detection of alkaline phosphatase activity was done at room temperature in the dark, with colour buffer containing 450 µg/mL nitroblue-tetrazolium (NBT) and 150 µg/mL 5-bromo-4-chloro-3-indoyl-phosphate (BCIP). The colouration reaction was examined every 30 minutes until adequate precipitate was visible, which ranged from 45 minutes to

overnight, depending on the probe. Colouration was stopped with 3 successive 5 minute rinses in ddH_2O, 0.1% Tween 20 followed by an overnight wash in MeOH, 0.1% Tween 20 to clear embryo discolouration caused by the *in situ* protocol. Prior to photography, embryos were re-hydrated with MeOH:PBST (3:1, 1:1, 1:3) and rinsed twice in PBST.

Detection of fluorescein labeled probes using a different coloured precipitate was done by stopping the NBT-colouration reaction with 4 successive 5 minute rinses in ddH_2O, 0.1% Tween 20 followed by a 10 minute incubation at room temperature in 0.1M glycine pH 2.2 to inactivate alkaline phosphatases. Embryos were rinsed 4 times for 5 minutes in PBST, re-blocked for at least one hour, and incubated overnight with alkaline phosphatase conjugated anti-fluorescein antibody (1:10000; Roche) at 4°C. Excess antibody was removed by 5 successive 15 minute PBST washes with constant agitation. Embryos were once again equilibrated in colour buffer with 4 successive 5 minute washes and alkaline phosphatase activity was detected with colour buffer containing 165 μg/mL iodo-nitrotetrazolium (INT) and 150 μg/mL BCIP in the dark. Deposition of red precipitate was stopped with 4 successive 5 minute rinses in ddH_2O, 0.1% Tween 20 followed by storage of embryos in 4% PFA at 4°C.

5) *Ladybird* mRNA and Morpholino Injection

Coding sequences for each zebrafish *ladybird* gene were PCR-amplified with Phusion high fidelity DNA polymerase (New England Biolabs) from

template cDNA that was synthesized (Invitrogen) from RNA isolated at various embryonic stages. Forward primers were designed with BglII sites at the 5' end and an optimal ribosome binding sequence (GCCACC; Kozak and Shatkin, 1979; Kozak, 1987) directly upstream of the ATG initiation codon, while reverse primers had SpeI sites at the 5' end (Table 2-1). PCR products and T7TS expression plasmid were digested with BglII and SpeI overnight at 37°C, gel purified (Qiagen), and ligated (Invitrogen) together according to manufacturers specifications for sticky-end ligations. One Shot Top10 chemically competent *E. Coli* (Invitrogen) were transformed with ligation mixtures and plasmid preps (Qiagen) were isolated and sequenced. Plasmids from positive clones were linearized with BamHI and m^7G-capped mRNA was synthesized using a mMessage Machine T7 *in vitro* transcription kit (Ambion). *Lbx1* mRNA was purified using YM-50 columns (Microcon) and analyzed for concentration and integrity. Zebrafish embryos were microinjected with approximately 15 ng mRNA at the one-cell stage and raised in embryo media containing 1% penicillin-streptomycin. A proportion of embryos were fixed at 24 hpf for *in situ* hybridization while the rest were monitored for developmental defects. Live embryos were anaesthetized in 0.5 mM MESAB (ethyl-m-aminobenzoate methanesulfonate; Sigma) in embryo media and photographed.

Translation blocking antisense morpholinos were designed by Gene Tools (www.gene-tools.com) to each zebrafish *lbx* transcript; *lbx1a* (5'-GAT GGT TGT CAT TCC AGC CTT TGT C-3'), *lbx1b* (5'-CTT TCG CGT CTT CTT TGG AGG

TCA T-3') and *lbx2* (5'-TTT AGA GCT GGA GGT CAT CTC AGT C-3'). Morpholinos were kept at a stock concentration of 10 mg/mL in Danieau buffer at -20°C while working concentrations (1-3 mg/mL) were stored at 4°C. Prior to injection, morpholinos were heated to 65°C for 10 minutes and loaded into 1.2 mm glass needles. One-cell zebrafish embryos were injected with 1-10 ng of morpholino and monitored until 3 dpf for any developmental defects. Combinations of morpholinos were also injected to assess any functional redundancy of duplicated *ladybird* genes. p53 morpholino was co-injected at a concentration of 2 mg/mL as it has been reported to prevent non-specific cell death associated with the microinjection procedure itself (Robu *et al.*, 2007).

6) Identifying Enhancers that Drive *Ladybird* Expression

Genomic sequence available from numerous vertebrates was compared using the BLAST-Like Alignment Tool (BLAT; Kent, 2002) at http://genome.ucsc.edu/, which identified several highly conserved sequence blocks neighbouring the *lbx1* locus. Thorough sequence analysis of these regions yielded a much more comprehensive map of conservation between vertebrates. Conserved sequence blocks that could potentially act as enhancer elements at zebrafish *lbx1* loci were amplified via PCR (Table 2-2) with Phusion DNA polymerase (New England Biolabs) from previously isolated zebrafish genomic DNA and cloned into pCR8/GW/TOPO (Invitrogen). To establish potential enhancer activity, confirmed clones were recombined into the well characterized

Table 2-2: Primers Used for Deletion Analysis of the Zebrafish *Lbx1a* Enhancer

***Lbx1a* Enhancer Fragment Analysis**

	Sequence (5'-3')
Lbx1a-DS-F1	AGAGCTTAACAAACCTGTTCACGACGG
Lbx1a-DS-F179	CTTTTACTCGCCAATAGCGTGACTTGG
Lbx1a-DS-F316	GTTTTGGTCCAGTTTCTCCAG
Lbx1a-DS-F457	CCTGCGGTTCACCATACTGAAAC
Lbx1a-DS-R206	GCCAAGTCACGCTATTGGCGAG
Lbx1a-DS-R484	CACTTGTTTCAGTATGGTGAACCGCAGG
Lbx1a-DS-R635	GGAGAGACTCCAATGTATGGC
Lbx1a-DS-R680	CGAGATGAAAAGAGCGGACATTC
Lbx1a-DS-R1067	CACCTGCTACTGACAGAATGCTTAGTG

***Lbx1a* Enhancer Deletion Analysis**

	Sequence (5'-3')
Lbx1a-DS-R243	ATCGGGATCCCGCCCAGAAAGGAAAAACACACG
Lbx1a-DS-R371	ATCGGGATCCCTGCTTCACCCGTGAACAAAAAGG
Lbx1a-DS-R469	ATCGGGATCCGGTGAACCGCAGGAGTCCAGCC
Lbx1a-DS-R521	ATCGGGATCCCCTCCAAACGTTTGCGTCTTTTTTATGTG
Lbx1a-DS-R691	ATCGGGATCCAGCTAAAAAGACGAGATGAAAAGAGCGG
Lbx1a-DS-F271	ATCGAGATCTGGCGCTGGGGAGAAAAAGGAG
Lbx1a-DS-F413	ATCGAGATCTCATCCACCCTGATGTGACAAGTGAC
Lbx1a-DS-F571	ATCGAGATCTGCGTTAACCCCTCTTTTAAACTGTTTAACG
Lbx1a-DS-F620	ATCGAGATCTCATTGGAGTCTCTCCTCGAGAAAGC
Lbx1a-DS-F728	ATCGAGATCTGCTCTCTCTTTTTTTGTGAGGCACTG

Combinations of various forward (F) and reverse (R) primers were used to generate a series of fragments or small deletions of the identified downstream (DS) *lbx1a* enhancer. Numbers correspond to the position along the 1067 bp *lbx1a* enhancer element to which the 5' end of the primer is designed to bind. BamHI (GGATCC) and BglII (AGATCT) restriction sites (underlined) allow for deletions of small conserved regions within the enhancer through ligation of compatible GATC overhangs.

pTol2-GW:*cFos*:*EGFP* (Figure 2-1; Fisher *et al.*, 2006a; Fisher *et al.*, 2006b; Kawakami, 2007) or pTol2-GW:*cFos*:*dTomato* (Figure 2-2), with LR clonase II (Invitrogen). The pTol2-GW:*cFos*:*dTomato* variant was a kind gift from Dr. Brad Magor in which the Gateway (GW) cassette is in opposite orientation to the pTol2-GW:*cFos*:*EGFP* vector, and coding sequence for a dimeric red fluorescent protein (dTomato; Shaner *et al.*, 2004) has replaced the EGFP coding sequence. The recombination reactions were transformed into OmniMAX 2 T1-phage resistant *E. Coli* (Invitrogen) and positive clones were confirmed by sequencing using Tol2-vector specific primers; Forward: 5'-TAG CAG GAA ACG TGA GCA GAG ACT CC-3'; Reverse for the EGFP vector: 5'-ATG AAC TTC AGG GTC AGC TTG CCG TAG G-3'; or Reverse for the dTomato vector: 5'-AAC TCT TTG ATG ACC TCC TCG CCC TTG C-3'. Wild type strain AB zebrafish embryos were microinjected at the one-cell stage with 2-3 nL of 25 nM purified transgenic construct plasmid, and 25 nM of *in vitro*-synthesized *Tol2*-mRNA (Ambion) to facilitate genomic integration of each construct. Transgenic constructs yielded a wide variety of reporter expression patterns with variable intensities and inserts initially tested showed no variability whether they were inserted in the 5' to 3' or 3' to 5' direction relative to the *cFos* promoter. All future recombination reactions were performed so that putative enhancers were oriented in the 5' to 3' direction relative to the *cFos* promoter within the Tol2 vector. Constructs with a low level of transient expression were injected at an increased dose of 100 nM to discriminate between genomic regions that are unable to drive

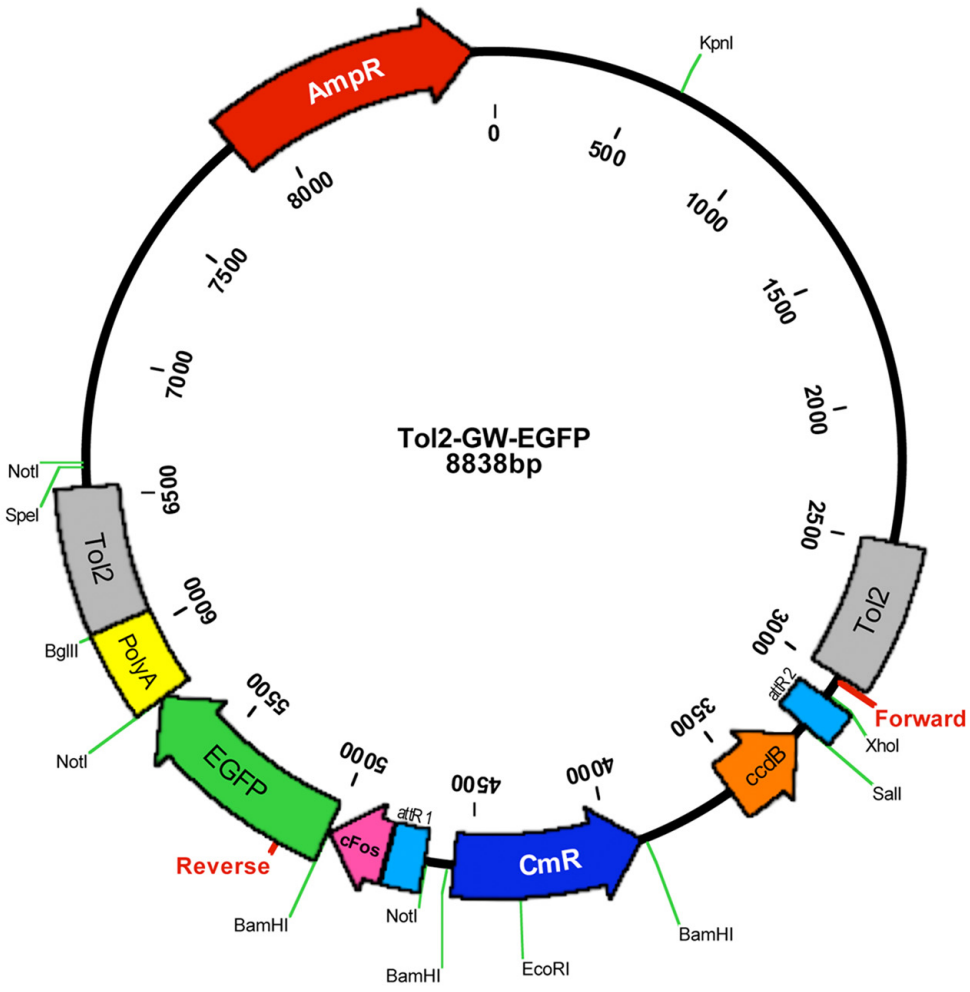

Figure 2-1. **Map of pTol2-GW:*cFos:EGFP***. Tol2 transposase binding sites (grey) flank a construct containing a gateway cassette (5'-attR2-ccdB-CmR-attR1-3'), which allows for insertion of cloned sequences, a minimal cFos promoter (pink) and coding sequence for EGFP (green) fused to a poly-A tail (yellow). Notable restriction sites are indicated and pTol2-GW sequencing primers (see Table 2-1) are labeled in red. Vector sequence was annotated in MacVector v7.2.2 and displayed as shown. AmpR, ampicillin resistance gene; CmR, chloramphenicol resistance gene; ccdB, host lethality gene; attR1 and attR2, LR clonase recombination enzyme binding sites.

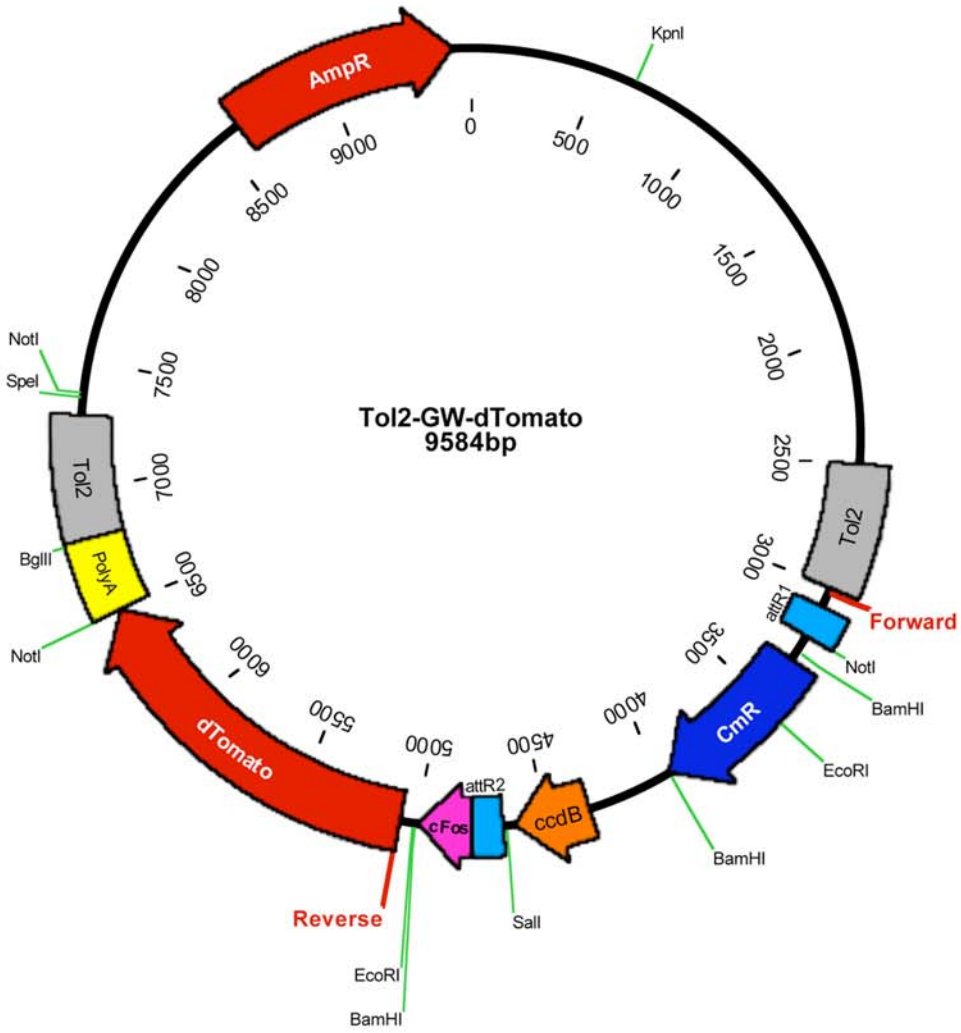

Figure 2-2. **Map of pTol2-GW:*cFos:dTomato***. The construct has Tol2 transposase binding sites (grey) flanking a gateway cassette (5'-attR1-CmR-ccdB-attR2-3') which allows for insertion of cloned sequences, a minimal *cFos* promoter (pink) and coding sequence for *dTomato* (red) fused to a poly-A tail (yellow). Notable restriction sites are shown and pTol2-GW sequencing primers (see Table 2-1) are labeled in red. Vector sequence was annotated in MacVector v7.2.2 and displayed as shown. AmpR, ampicillin resistance gene; CmR, chloramphenicol resistance gene; ccdB, host lethality gene; attR1 and attR2, LR clonase recombination enzyme binding sites.

expression in a tissue-specific manner from those driving expression in a very restricted domain. To corroborate the transient expression patterns observed from individual constructs, injected embryos were raised to adulthood and screened for stable transgenic progeny.

mRNA *in situ* hybridization allowed for detection of reporter *EGFP* or *dTomato* transcripts in tissues that had lower than detectable levels of fluorescent protein in live embryos, which provided a comparison of transgene expression to endogenous *ladybird* expression. This technique was also used to identify transgenic adults bearing construct insertions near promiscuous enhancers that were able to act on the reporter construct. Stable zebrafish lines with either *EGFP* or *dTomato* transgene expression driven by putative enhancer elements were generated by selecting stable transgenic adults yielding a Mendelian inheritance pattern of transgene expression when crossed to non-transgenic zebrafish. The corresponding genomic sequences downstream of the human *LBX1* locus were also cloned into pTol2-GW:*cFos*:*dTomato* and examined for the ability to respond to zebrafish transcription factors and drive transgene expression during early zebrafish development. Stable transgenic progeny were examined by live fluorescence microscopy and mRNA *in situ* hybridization for transgene expression.

Promoter sequences of zebrafish *ladybird* genes that showed significant conservation to other vertebrates were also cloned into the Tol2-GW:*cFos:EGFP* vector. In this case, XhoI and BamHI sites were designed into the forward and

reverse primers (Table 2-1), respectively, which allows for the replacement of the entire gateway cassette and minimal cFos promoter with the desired cloned sequences. Putative promoter PCR products and purified Tol2-GW:*cFos*:*EGFP* plasmid were digested with XhoI and BamHI overnight at 37°C and gel purified (Qiagen). Cloned promoter inserts were ligated to linear Tol2-*EGFP* vector overnight at room temperature and transformed into One Shot Top10 chemically competent *E. coli* (Invitrogen). Positive clones were sequenced to confirm the presence of putative *ladybird* promoter sequences and plasmid was injected into 1-cell zebrafish embryos with Tol2-mRNA to assess transient expression.

7) Cryo-sectioning of Transgenic Embryos

Tg(*lbx1a*:*dTomato*) fish were crossed to Tg(*isl1*:*GFP*) fish (Higashijima *et al.*, 2000) to generate embryos expressing 2 different fluorescent reporters. Embryos were raised to 84 hpf and fixed in 4% PFA, 5% sucrose overnight at 4°C with constant agitation. For all future steps embryos were kept in the dark as much as possible. Fixed embryos were rinsed three times with 5% sucrose in 1xPBS for 20 minutes, then with 12.5% sucrose in 1xPBS for 20 minutes and finally with 20% sucrose in 1xPBS overnight at 4°C to cryo-protect tissues. Embryos were equilibrated in embedding media comprised of a 2:1 mixture of 20% sucrose in 1xPBS and Optimal Cutting Temperature (OCT) embedding solution for 30 minutes at room temperature with constant agitation. Embryos were positioned with anterior down in sectioning moulds and cooled on a ~1.5 cm

thick steel plate positioned over top of dry ice. Moulds were allowed to freeze for at least 30 minutes prior to sectioning. Frozen samples were removed from the moulds and mounted on a Leica Cryostat0074 sectioning apparatus to generate 10 μm sections that were adhered to Superfrost Plus Glass slides and stored at 4°C.

8) Whole Mount Immunofluorescence

Fixed Tg(*lbx1a:EGFP*) embryos in 100% MeOH were rehydrated through a series of MeOH:PBST (3:1, 1:1, 1:3) washes and then rinsed in PBST 4 times for 5 minutes. 10 μg/mL proteinase K in PBST was added for 7.5 minutes (32 hpf embryos) to 15 minutes (60 hpf embryos) to permeablize tissues. Embryos were re-fixed in 4% PFA 1xPBS for 20 minutes at room temperature, rinsed in PBST 4 times for 5 minutes and blocked in 2.5% goat serum, 2.5 mg/mL BSA and 0.1% Triton X-100 in PBST at room temperature in the dark for 60-120 minutes with constant agitation. Primary antibodies obtained from the Zebrafish International Resource Center (ZIRC), mouse anti-Zrf1 (1:2) or mouse anti-Zn12 (1:500), were applied in blocking solution overnight at 4°C. Excess antibody was removed with 5 – 30 minute room temperature washes in PBST with 1% DMSO and 0.1% Triton X-100. Embryos were blocked again and incubated with rabbit polyclonal anti-GFP (1:500) antibody (Invitrogen) overnight at 4°C. After removing excess anti-GFP antibody with 5 – 30 minute PBST washes at room temperature, goat anti-rabbit-alexa488 (Invitrogen) and goat anti-mouse-alexa568 (Invitrogen) secondary antibodies (1:1000) were applied in blocking solution overnight at 4°C

in the dark. Excess secondary antibody was removed with several PBST washes and embryos were stored in PBST at 4°C until documentation.

9) Identifying Critical Regions of the *lbx1a* Enhancer

To narrow down sequence blocks within the 1067 bp *lbx1a* enhancer responsible for driving expression in the hindbrain and spinal cord, both fragment- and deletion-based approaches were used. Enhancer fragments generated by PCR-amplification from full-length enhancer template using specific primers (Table 2-2) were purified and cloned into pCR8/GW/TOPO (Invitrogen). Confirmed inserts were recombined into pTol2-GW:*cFos:EGFP* as previously mentioned and injected into 1-cell zebrafish embryos with *Tol2*-mRNA at a concentration of 25-50 nM. Transient reporter expression was assessed during early zebrafish development from numerous regions and fragments of the 1067 bp *lbx1a* enhancer.

Deletions of short, highly conserved sequence blocks were accomplished by ligating a fragment upstream of the desired deletion to a fragment downstream of the desired deletion. Ligation was possible through the incorporation of a BamHI (GGATCC) restriction site onto the 5' end of the reverse primer used to generate the fragment upstream of the deletion and a BglII (AGATCT) restriction site onto the 5' end of the forward primer used to generate the fragment downstream of the deletion (Table 2-2). This procedure effectively replaced a highly conserved sequence block with a hybrid BamHI-BglII (GGATCT)

restriction site. Ligation reactions were used as template for PCR amplification of the deletion fragment using the same primers that amplify the full length 1067 bp *lbx1a* enhancer. Purified deletion fragments were cloned into pCR8/GW/TOPO (Invitrogen), confirmed by sequencing and recombined into the transgenic pTol2-GW:*cFos:EGFP* vector. Transient expression analysis of deletion constructs was achieved by injecting purified plasmids with *Tol2*-mRNA at 25 nM into 1-cell zebrafish embryos and observing GFP by live fluorescence microscopy.

10) Photography

Live transgenic embryos were anaesthetized in 0.5 mM MESAB in embryo media and embedded in 2% (w/v) methylcellulose that was over-laid with anaesthetic. Embryos were visualized on a Leica MZ 16 F fluorescent stereomicroscope and imaged using a Leica DFC 420 C camera with Leica Application Suite v3.1.0 software. Whole embryos that had gone through mRNA *in situ* hybridizations were photographed in 2% methylcellulose on a Zeiss Discovery.V8 light stereomicroscope using a Q-Imaging MP5-RTV camera with Q-Capture Pro 6.0 software. Alternatively, embryos were deyolked in 1% methylcellulose using 0.2 mm diameter insect pins or fine-tip forceps (Dumont #5) and rinsed 3 times for 10 minutes in PBST prior to flat mounting. For flat mounting, embryos were equilibrated through a series of glycerol washes (30%, 50%, 70%) in 1xPBS and oriented on glass slides with a cover slip over-top. Cryo-sections and flat mounted embryos were photographed with an AxioCam

HRm camera on a Zeiss Imager.Z1 compound microscope using AxioVision v4.4 software. Photograph modification and composite figure preparation was done using Adobe Photoshop CS3/CS4 and ImageJ v1.40g.

11) References

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ, **1990**. Basic local alignment search tool. *J Mol Biol*. 215:403-410.

Baxevanis AD, **2006**. Searching the NCBI databases using Entrez. *Curr Protoc Hum Genet*. Chapter 6:Unit 6 10.

Baxevanis AD, **2008**. Searching NCBI databases using Entrez. *Curr Protoc Bioinformatics*. Chapter 1:Unit 1 3.

Birney E, Andrews D, Caccamo M, Chen Y, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM, Flicek P, Graf S, Hammond M, Herrero J, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Kokocinski F, Kulesha E, London D, Longden I, Melsopp C, Meidl P, Overduin B, Parker A, Proctor G, Prlic A, Rae M, Rios D, Redmond S, Schuster M, Sealy I, Searle S, Severin J, Slater G, Smedley D, Smith J, Stabenau A, Stalker J, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C and Hubbard TJ, **2006**. Ensembl 2006. *Nucleic Acids Res*. 34:D556-561.

Fisher S, Grice EA, Vinton RM, Bessling SL and McCallion AS, **2006a**. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science*. 312:276-279.

Fisher S, Grice EA, Vinton RM, Bessling SL, Urasaki A, Kawakami K and McCallion AS, **2006b**. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat Protoc*. 1:1297-1305.

Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Eyre T, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Holland R, Howe KL, Howe K, Johnson N, Jenkinson A, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Megy K, Meidl P, Overduin B, Parker A, Pritchard B, Prlic A, Rice S, Rios D, Schuster M, Sealy I, Slater G, Smedley D, Spudich G, Trevanion S, Vilella AJ, Vogel J, White S, Wood M, Birney E, Cox T, Curwen V, Durbin R, Fernandez-Suarez XM, Herrero J, Hubbard TJ, Kasprzyk A, Proctor G, Smith J, Ureta-Vidal A and Searle S, **2008**. Ensembl 2008. *Nucleic Acids Res*. 36:D707-714.

Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K, Johnson N, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Melsopp C, Megy K, Meidl P, Ouverdin B, Parker A, Prlic A, Rice S, Rios D, Schuster M, Sealy I, Severin J, Slater G, Smedley D, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wood M, Cox T, Curwen V, Durbin R, Fernandez-Suarez XM, Flicek P, Kasprzyk A, Proctor G, Searle S, Smith J, Ureta-Vidal A and Birney E, **2007**. Ensembl 2007. *Nucleic Acids Res*. 35:D610-617.

Jenuth JP, **2000**. The NCBI. Publicly available tools and resources on the Web. *Methods Mol Biol*. 132:301-312.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S and Madden TL, **2008**. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 36:W5-9.

Kawakami K, **2007**. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol*. 8 Suppl 1:S7.

Kent WJ, **2002**. BLAT--the BLAST-like alignment tool. *Genome Res*. 12:656-664.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B and Schilling TF, **1995**. Stages of embryonic development of the zebrafish. *Dev Dyn*. 203:253-310.

Kozak M, **1987**. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*. 15:8125-8148.

Kozak M and Shatkin AJ, **1979**. Characterization of translational initiation regions from eukaryotic messenger RNAs. *Methods Enzymol*. 60:360-375.

Pruitt KD, Tatusova T and Maglott DR, **2005**. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 33:D501-504.

Pruitt KD, Tatusova T and Maglott DR, **2007**. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 35:D61-65.

Robu ME, Larson JD, Nasevicius A, Beiraghi S, Brenner C, Farber SA and Ekker SC, **2007**. p53 activation by knockdown technologies. *PLoS Genet*. 3:e78.

Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE and Tsien RY, **2004**. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol*. 22:1567-1572.

Thompson JD, Gibson TJ and Higgins DG, **2002**. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. Chapter 2:Unit 2 3.

Wolfsberg TG, **2007**. Using the NCBI Map Viewer to browse genomic sequence data. *Curr Protoc Bioinformatics*. Chapter 1:Unit 1 5.

Chapter 3: Identifying Zebrafish *Ladybird* Homeodomain Encoding Genes

1) Introduction

a) Molecular Characteristics of *Ladybird*

The *ladybird* (*lb*) class of homeodomain transcription factor genes was first identified in *Drosophila melanogaster* through its sequence similarity to and chromosomal position within the NK-homeobox cluster in region 93D-E of chromosome 3R (Kim and Nirenberg, 1989; Jagla *et al.*, 1993). *Drosophila ladybird* genes occur as a tandem duplication within the NK-homeobox gene cluster (Jagla *et al.*, 1997b) yielding predicted proteins of 479 amino acids (*ladybird early*; *lbe*) and 346, 372 and 411 amino acids (*ladybird late*; *lbl*). Ladybird homeodomain sequences are most closely related to T-cell leukemia homeobox (Tlx) proteins (93Bal in *Drosophila*) and these two protein families share specific amino acid differences within the homeodomain when compared to other homeodomain proteins (Kennedy *et al*., 1991; Raju *et al.*, 1993). Mammalian *Ladybird* genes (*Lbx*) were cloned shortly after their identification in *Drosophila* showing these genes are highly conserved in both chromosomal position and transcript sequence (Jagla *et al.*, 1995). Ladybird homeodomain sequences differ from all other homeobox genes at several key residues in helix 1 and helix 3, which are thought to be important for DNA sequence recognition (Jagla *et al.*, 1993).

Mammalian *Ladybird* genes possess a simple chromosomal structure of two exons spanned by an intron approximately 650 bp in length (Jagla *et al.*, 1995). There is considerable sequence conservation within the intron and

untranslated regions (UTR) between mice and humans, ranging from 65% to 82% identity. Vertebrate Lbx1 protein sizes range from 213 amino acids in chicken to 318 amino acids in amphioxus; however, many vertebrates have an *Lbx1* gene that encodes a protein roughly 280 amino acids in length. The homeodomain is encoded by the 5' end of the second exon and a unique feature shared amongst only Ladybird and Hox11 protein classes is a threonine at position 47 of the homeodomain instead of valine or isoleucine, which has a critical role in DNA recognition (Gehring, 1987). *In situ* hybridization on metaphase chromosomes from mouse and human cells position the *Lbx1* locus adjacent to the *Hox11* (*Tlx*) locus which is a featured shared by *ladybird* and *93Bal* in *Drosophila* (Jagla *et al.*, 1995).

b) Expression Patterns of Identified *Ladybird* Genes

Expression of *Drosophila ladybird* occurs early in embryonic development and transcripts can also be detected in 3rd instar larvae and pupae (Jagla *et al.*, 1993). The onset of expression occurs at embryonic stage 8 in clusters of the central nervous system (CNS), mesoderm and epidermis, which are reiterated amongst the repeated body segments of the early *Drosophila* embryo (Jagla *et al.*, 1993; Jagla *et al.*, 1994; Jagla *et al.*, 1997b; Urbach *et al.*, 2003). Epidermal *ladybird* expression overlaps with *wingless* (*wg*) expression directly adjacent to *engrailed*-expressing cells at parasegmental boundaries (Jagla *et al.*, 1997b). During germ band elongation *ladybird* transcripts can be detected in the

anal plate primordium, neuroblasts of the CNS, heart precursors and specific muscle precursor cells that develop into the segmental border muscle and lateral adult muscle (Jagla *et al.*, 1997a; Jagla *et al.*, 1997b; Jagla *et al.*, 1998; Han *et al.*, 2002; De Graeve *et al.*, 2004). As embryonic development proceeds, *Drosophila ladybird* transcripts become restricted to clusters of the central and peripheral nervous system (Jagla *et al.*, 1997b; De Graeve *et al.*, 2004). In 3rd instar larvae, *lbe* exhibits dynamic expression within myoblasts of the leg imaginal discs (Maqbool *et al.*, 2006)

The majority of mammalian *Lbx1* expression occurs in a very short window, between stages E9 and E12 in mouse and at eight weeks in the human fetus (Jagla *et al.*, 1995). *Lbx1* transcripts are restricted to the developing central nervous system (Jagla *et al.*, 1995; Gross *et al.*, 2002; Muller *et al.*, 2002) and various muscle types including the tongue, diaphragm, shoulders and limbs (Jagla *et al.*, 1995; Schäfer and Braun, 1999; Brohmann *et al.*, 2000; Gross *et al.*, 2000; Uchiyama *et al.*, 2000). The rostrocaudal boundary of mouse *Lbx1* expression occurs within the rostral metencephalon just posterior to the mid-hindbrain boundary, and expression within the neural tube extends posteriorly in the spinal cord (Jagla *et al.*, 1995). Along the dorsoventral axis, neural tube expression is only seen within the mantle zone, where neural progenitors differentiate, in the dorsal alar plate of the hindbrain and the dorsal spinal cord (Jagla *et al.*, 1995; Gross *et al.*, 2000; Muller *et al.* 2002). Several transgenic mouse lines have assisted in determining *ladybird* expression patterns in discrete regions of the

embryo which were not detected by standard *in situ* experiments (Schäfer and Braun, 1999; Gross *et al.*, 2000). These strains carry a combination gene fusion and deletion where most of the *Lbx1* coding sequence has been replaced by either an in frame *lacZ* (Schäfer and Braun, 1999; Brohmann *et al.*, 2000) or *EGFP* (Gross *et al.*, 2000) reporter construct, however the EGFP transgene is expressed at low levels and requires detection with an EGFP antibody.

In situ hybridization for *ladybird* mRNA in chick (Dietrich *et al.*, 1998; Schubert *et al.*, 2001; Kanamoto *et al.*, 2006) and *Xenopus* (Martin and Harland, 2006) embryos confirms expression in similar regions of the hindbrain, spinal cord and muscle progenitors of the limb bud. Strong expression in both mouse (Schäfer and Braun, 1999; Brohmann *et al.*, 2000; Gross *et al.*, 2000) and chicken (Dietrich *et al.*, 1998) is observed in paraxial mesoderm at levels of both fore- and hind-limbs, as well as the dorsal and ventral regions of the developing limb buds. *Lbx1* marks myogenic precursors that delaminate from the ventrolateral dermomyotome and are capable of long range migration into the limb bud, muscle of the tongue and diaphragm (Dietrich *et al.*, 1998; Schäfer and Braun, 1999; Gross *et al.*, 2000). It has also been observed that *Ladybird* expression is present in regenerating muscle stem cells of the adult mouse (Watanabe *et al.*, 2007).

c) Function of Ladybird in Muscle and Nervous System Development

The functions of Ladybird proteins have been particularly well studied with regards to both muscle and nervous system development. In *Drosophila*,

ladybird has been shown to positively regulate *wingless* expression in the dorsal epidermis, and if *ladybird* is introduced ectopically, can activate *wingless* expression in cells that do not normally express the well known signalling molecule (Jagla *et al.*, 1997b). This *ladybird-wingless* loop is also present within cardiac progenitors such that ectopic *ladybird* expression greatly expands the number of cardiac and pericardial progenitors (Jagla *et al.*, 1997a; Han *et al.*, 2002). Additionally, *ladybird* forms a mutual repression loop with *even-skipped* (*eve*) in the development of cardiac mesoderm (Han *et al.*, 2002). Somatic muscle progenitors within segmental units are also affected when *ladybird* expression is perturbed in *Drosophila* (Jagla *et al.*, 1998). Specific neural progenitors giving rise to both early and late neuronal sub-types also require ladybird for their specification (De Graeve *et al.*, 2004). More recently, *ladybird* has been shown to be important for the development and function of adult leg muscles (Maqbool *et al.*, 2006), again with ties to the *wingless* pathway.

Vertebrate functions of Ladybird are very similar to *Drosophila*. In chick (Dietrich *et al.*, 1998), mouse (Brohmann *et al.*, 2000; Gross *et al.*, 2000) and frog (Martin and Harland, 2006) Lbx1 contributes to the development of hypaxial musculature leading to dorsal and ventral muscle masses in fore- and hind-limbs. This is accomplished through the delayed differentiation of delaminated myoblasts as well as the proposed ability to make migrating myogenic precursors capable of responding to lateral cues from the developing limb bud. In the vertebrate neural tube, Lbx1 has a distinct role in distinguishing two major neural

subtypes of the dorsal spinal cord (Müller *et al.*, 2002) which gives rise to association interneurons that relay sensory information (Gross *et al.*, 2002). Neuronal fates within the hindbrain are also determined by Lbx1, where sensory neurons expressing *Lbx1* take on a somatosensory fate over a viscerosensory fate (Sieber *et al.*, 2007). Vertebrate Ladybird also plays a role in promoting a basal GABAergic neurotransmitter phenotype, where in conjunction with Tlx3 proteins, creates an opposing switch between GABAergic and glutamatergic neurotransmitter fate (Cheng *et al.*, 2005).

It is evident that Ladybird proteins play a highly conserved underlying role in both myogenesis and neurogenesis, and that complexity is built upon this basal role to generate varying developmental programs of muscle tissues and neuron types. Ladybird proteins have been shown to be critical in appendicular muscle development (Neyt *et al.*, 2000) by controlling lateral migration into the developing limb, and the specification of a subset of dorsal spinal cord neurons that relay sensory information. More specialized muscle types that require *ladybird* are the heart, diaphragm and tongue. In *Drosophila* it has been shown that a major signalling pathway controlled, at least in part, by ladybird is the *wingless* signalling pathway, of which ladybird is a positive regulator. It would be of interest to examine the interplay between Wnt signalling and *Lbx1* expression in vertebrates for comparison to what is known in *Drosophila* development.

2) Results and Discussion

a) Bioinformatics of Vertebrate *Ladybird* Genes

To identify true *ladybird*-like genes within the zebrafish (*Danio rerio*) genome, human Ladybird homeobox 1 (*LBX1*) sequences were used in BLAST (Basic Local Alignment Search Tool) analysis for similar genomic, predicted mRNA, and predicted protein sequences in the zebrafish genome. Since identified mammalian Lbx1 sequences are >90% identical to each other and >70% identical to *Xenopus* Lbx1, using human LBX1 coding and protein sequences as the query sequence would likely identify putative ladybird encoding genes in the zebrafish genome. Significant changes between the March 2006 (Zv6) and July 2007 (Zv7) zebrafish genome assemblies prompted the use of both in data searches and analysis. BLAST searches that yielded the most conclusive results occurred when homeodomain sequences were omitted from the query sequence, as it eliminated false-positives of homeodomain-containing proteins that had no other similarity to Ladybird. Using either protein or nucleic acid sequences in BLAST searches provided a similar set of results, thus substantiating the presence of putative *ladybird*-like genes in the zebrafish genome.

The zebrafish genome was found to encode three identifiable *ladybird*-like genes on linkage groups (LG) 1, 13 and 14. The predicted amino acid sequence of these putative *lbx* genes ranges from 58% to 71% identity with human LBX1 along the entire protein length. Zebrafish *lbx* genes encode proteins that are slightly shorter than the 281 amino acid long human LBX1, and are currently

named *lbx1a* (265 aa; LG 1), *lbx1b* (269 aa; LG 13) and *lbx2* (257 aa; LG 14). To verify each predicted ladybird gene is expressed during zebrafish development, total RNA was isolated from various embryonic stages and used as template in reverse transcriptase PCR (Table 2-1). Products of expected size were gel purified, cloned and sequenced to corroborate the predicted position of exon-intron boundaries and the putative coding region. Zebrafish *ladybird* genes exhibit the same intron-exon boundaries as mammalian *Ladybird*. Sequencing results of all three zebrafish ladybird transcripts agreed with database predictions, with few polymorphisms, revealing that these genes are all expressed during embryonic development and encode a Ladybird-like protein. From human *LBX1* information, a similar gene structure of two exons is observed in all three zebrafish *ladybird* genes. Full length transcripts for zebrafish *ladybird* genes were not recovered and therefore sequence information for untranslated regions (UTR) is limited.

The amino-terminal half of Lbx1, which contains a predicted engrailed homology domain (Tolkunova *et al.*, 1998; Copley, 2005) and an uncharacterized conserved domain, is encoded by the first exon. The second exon encodes the homeodomain, which is followed by a long, poorly conserved region and an acidic carboxy-terminus that is rich in both glutamic and aspartic acid residues. Alignment of predicted zebrafish ladybird protein sequences generated from conceptual translation of cloned mRNAs to known vertebrate Lbx1 sequences clearly demonstrates the high degree of conservation among this protein family (Figure 3-1). The central 125 amino acids of the protein, which contains the

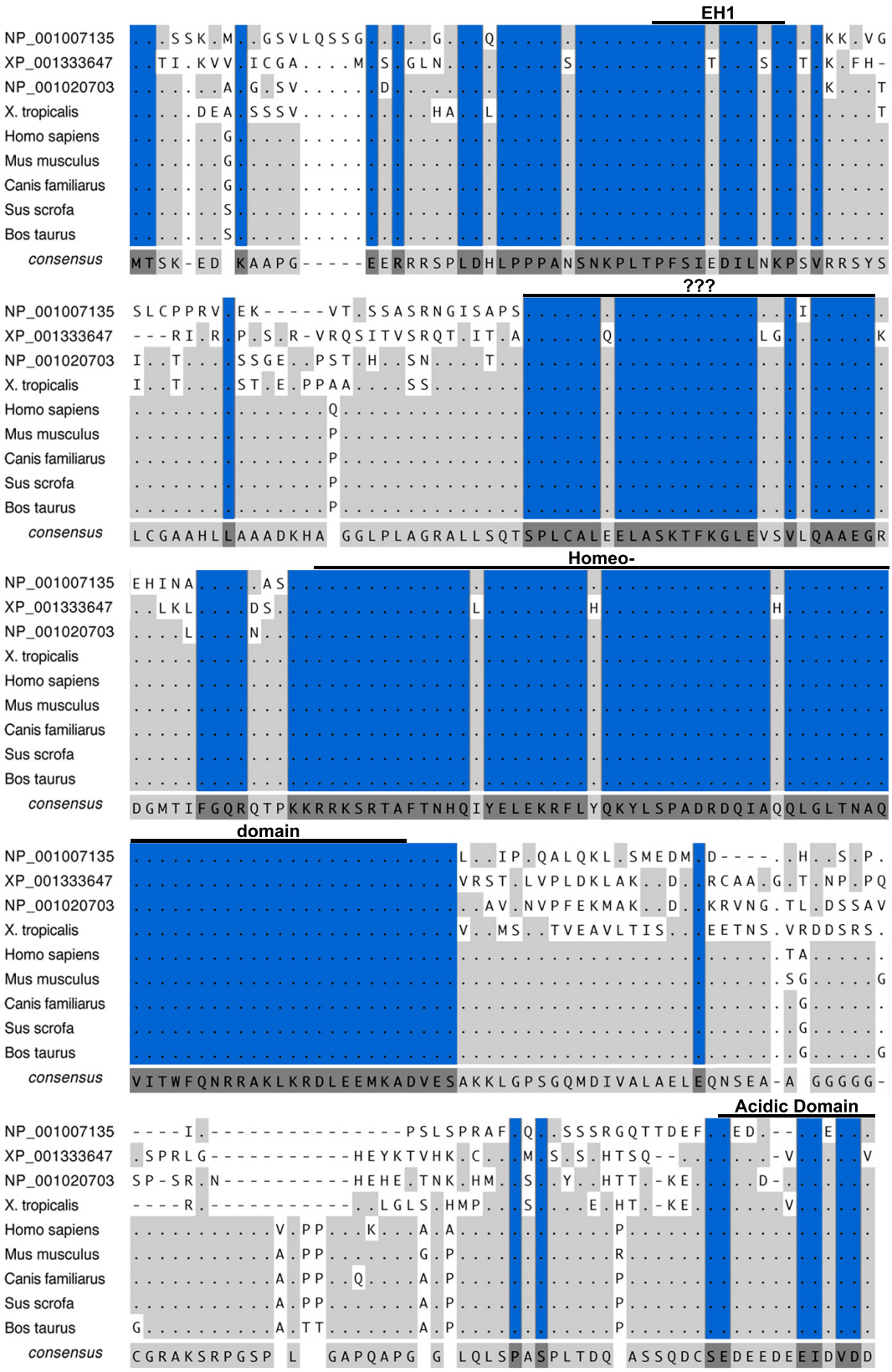
EH1
???
Homeo-
domain
Acidic Domain
NP_001007135
XP_001333647
NP_001020703
X. tropicalis
Homo sapiens
Mus musculus
Canis familiarus
Sus scrofa
Bos taurus
consensus

Figure 3-1. **Identification of three zebrafish *Lbx1*-like homeodomain encoding genes**. The genomic sequence from the March 2006 (Zv6) and July 2007 (Zv7) zebrafish assemblies were queried using the BLAST algorithm with human *LBX1* sequences. Predicted open reading frames NP_001007135 (*lbx2*), XP_001333647 (*lbx1a*) and NP_001020703 (*lbx1b*) were identified with a high score and aligned using the ClustalW algorithm to several known Lbx1 proteins from frog (NP_001072559), human (NP_006553), mouse (NP_034821), dog (XP_543979), pig (XP_001926254) and cow (XP_614245). Sequence highlighted in dark blue exhibits 100% identity between the Lbx proteins shown while those highlighted in light grey are commonly conserved only between mammalian Lbx1 proteins. Three highly conserved domains are indicated with black bars above the alignment while a fourth (???) does not match any database predictions. Dashes indicate a sequence gap and periods represent a match the consensus.

homeodomain and another conserved region, are >90% identical between zebrafish and mammals. The amino-terminal half shows significantly more conservation than the carboxy-terminal half as there is a poorly conserved putative linker region connecting the homeodomain to the acidic carboxy-terminus. This observation gives an initial insight into the function of Ladybird proteins, as acidic domains are thought to act as transcriptional activation domains (Li and Botchan, 1993; Kasamatsu *et al.*, 2004; Huang *et al.*, 2005).

Mammals possess another gene encoding an Lbx protein family member, *Lbx2*, the sequence of which differs significantly from Lbx1 both in the amino terminal half and within the homeodomain. These key differences make Lbx1 and Lbx2 sequences easily distinguishable from one another and it was of interest to determine if any zebrafish ladybird protein sequences resembled mammalian Lbx2. Alignment of predicted zebrafish Lbx proteins to known mammalian Lbx1 and Lbx2 amino acid sequences reveals that the zebrafish genome encodes three Lbx1-like proteins (Figure 3-2), despite the naming discrepancies encountered from current databases. This analysis does not preclude any zebrafish ladybird proteins from functioning in an analogous manner to mammalian Lbx2, however from database searches, it appears that identifiable Lbx2 proteins are only present in marsupial and placental mammal species. Thus the divergence of this second *Ladybird* gene may have accompanied the evolution of higher mammals and it is not surprising that zebrafish do not share this feature.

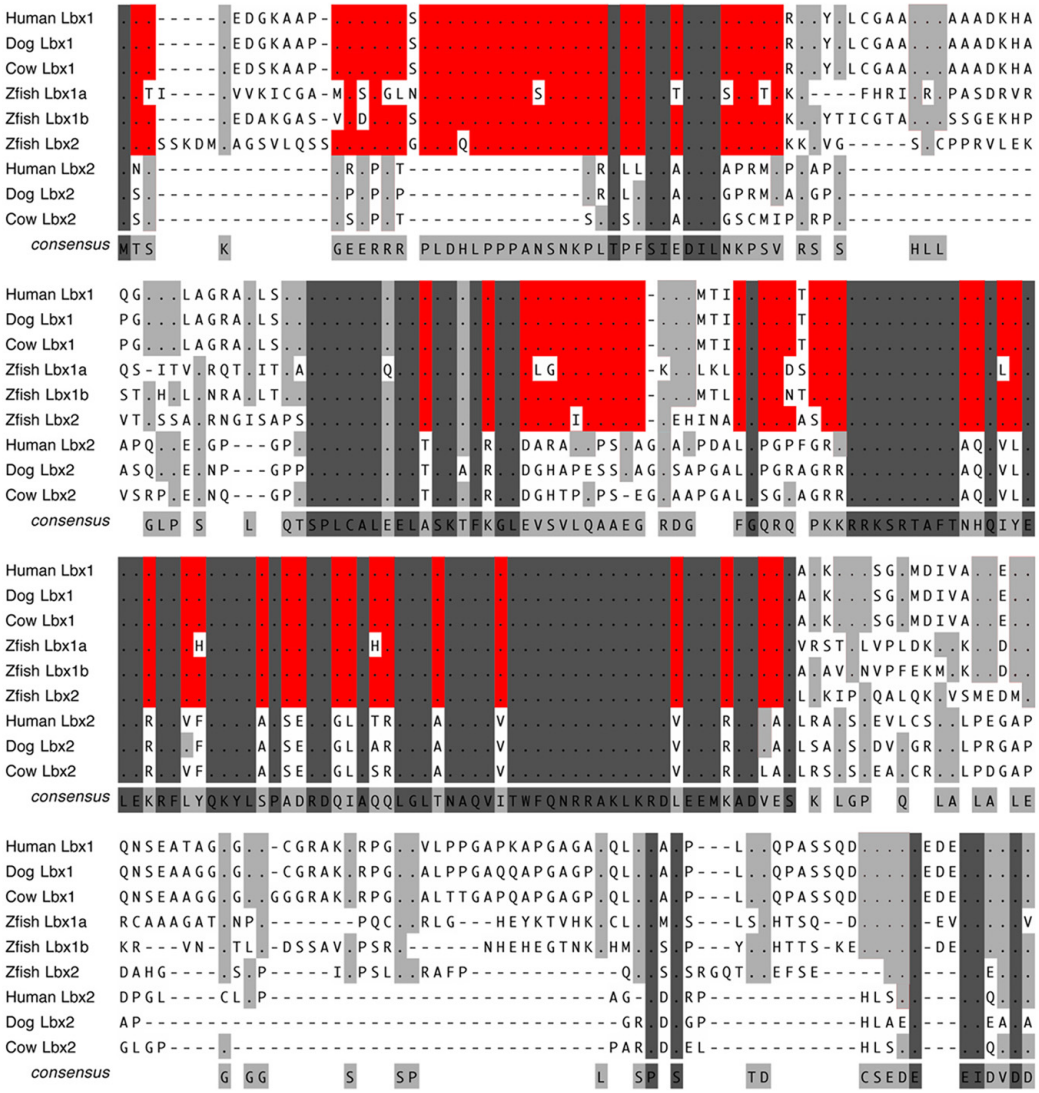

Figure 3-2. **Zebrafish Ladybird proteins are more similar to mammalian Lbx1**. Amino acid sequence alignment of identified zebrafish Ladybird proteins to mammalian homologs does not agree with current naming of zebrafish Lbx proteins. Zebrafish Lbx2 is more similar in sequence to mammalian Lbx1 as shown by sequences highlighted in red, which are conserved between the zebrafish Lbx-like proteins and mammalian Lbx1, but not Lbx2. Sequences with 100% identity between all Lbx proteins shown are in dark grey and partial (>50%) conservation is shown in light grey. Residues indicated by a period match the consensus and dashes indicate a gap in the sequence.

There is strong evidence that the teleost fish lineage underwent a whole genome duplication event (Hoegg *et al.*, 2004; Meyer and Van de Peer, 2005), leading to the possibility of multiple Ladybird-encoding genes, which might account for the naming discrepancies between databases. Although it is unclear as to why zebrafish retains three *ladybird*-like genes predicted to encode a protein that closely resembles known Lbx1 proteins, there is some data on the evolution and maintenance of *ladybird* genes in vertebrate species (Wotton *et al.*, 2008). It is hypothesized that through 2 rounds of genome duplication (Sidow *et al.*, 1996), ancestral vertebrates potentially had 4 *ladybird* genes, however, prior to the split between lobe- and ray-finned fishes, loci were lost such that only *Lbx1* and *Lbx2* existed (Wotton *et al.*, 2008). Genetic maps show *ladybird* genes are tightly linked within a gene cluster reminiscent of the NK-homeobox cluster found in *Drosophila* and that tetrapod species do possess the chromosomal remnants of four *ladybird* loci. If two *Lbx* loci existed prior to the genome duplication of teleost fish, zebrafish is expected to contain two *lbx1*-like loci and two *lbx2*-like loci and not the observed three *lbx1*-like loci.

To determine if the presence of three *lbx1*-like loci was unique to zebrafish, other available teleost fish genomes were analyzed using the BLAST algorithm with zebrafish and mammalian *Ladybird* sequences. Genomes of *Oryzias latipes* (Medaka), *Gasterosteus aculeatus* (Stickleback), *Tetraodon nigroviridis* and *Takifugu rubripes* (Fugu) all yielded similar results to zebrafish – three *Lbx1*-like predicted loci (Figure 3-3; Table 3-1), although the predicted

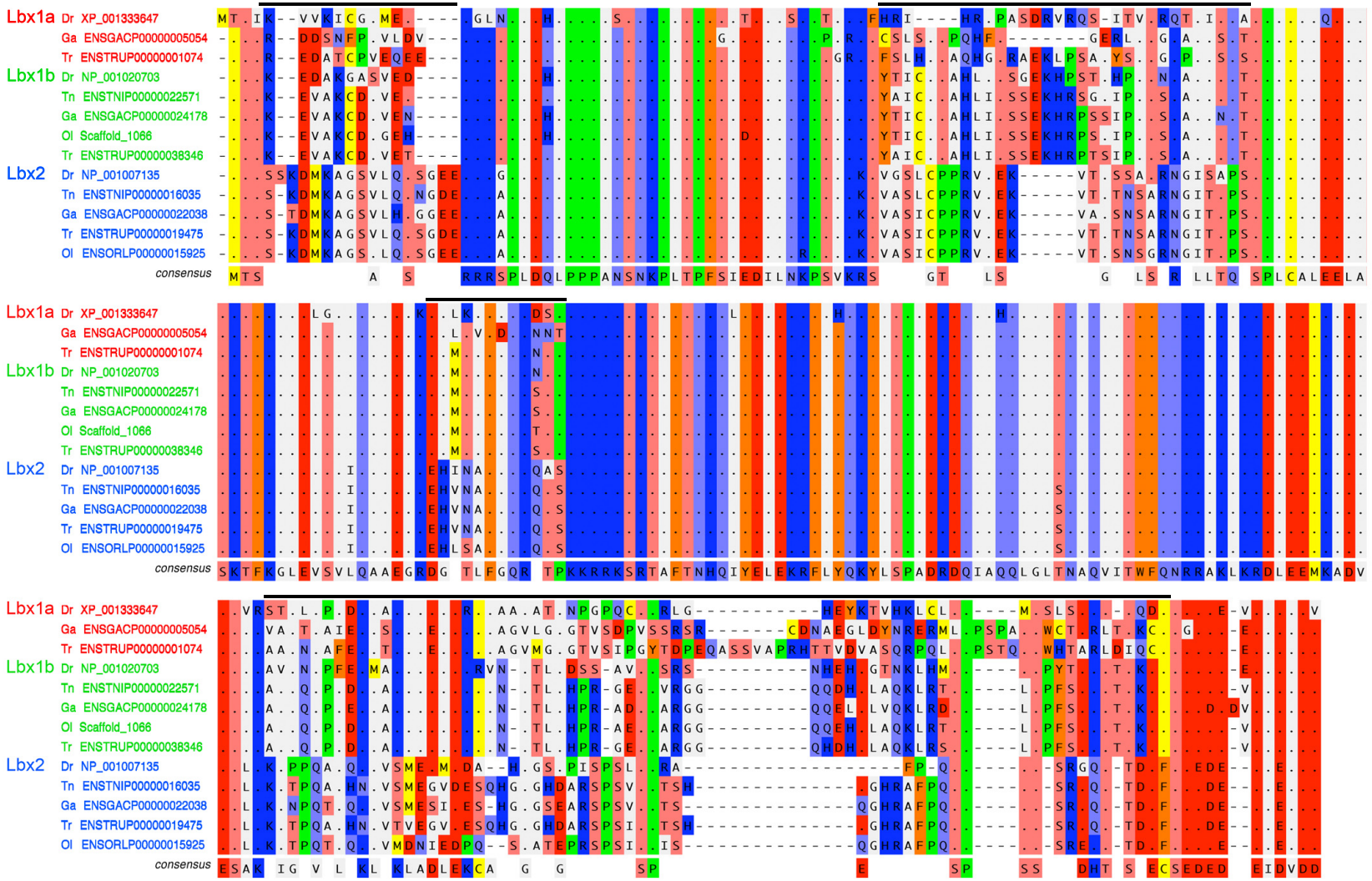
Lbx1a Dr XP_001333647
Ga ENSGACP00000005054
Tr ENSTRUP00000001074
Lbx1b Dr NP_001020703
Tn ENSTNIP00000022571
Ga ENSGACP00000024178
Ol Scaffold_1066
Tr ENSTRUP00000038346
Lbx2 Dr NP_001007135
Tn ENSTNIP00000016035
Ga ENSGACP00000022038
Tr ENSTRUP00000019475
Ol ENSORLP00000015925
consensus

Figure 3-3. **All three predicted zebrafish Ladybird proteins have orthologs in other teleost fish**. Genomes from *Tetraodon* (*Tetraodon nigroviridis*; Tn), Stickleback (*Gasterosteus aculeatus*; Ga), Fugu (*Takifugu rubripes*; Tr), and Medaka (*Oryzias latipes*; Ol) were BLAST with zebrafish (*Danio rerio*; Dr) Lbx1a (red), Lbx1b (green) and Lbx2 (blue) coding and protein sequences. Homologues were identified for all three different zebrafish Lbx proteins and accession numbers are shown. An incomplete Lbx1a sequence was identified in Tetraodon, however, the predicted genomic location of Lbx1a coding sequences in Medaka were not available for sequence analysis. Regions over-lined with black bars exhibit conservation within each Lbx protein type amongst teleost fish. Residues indicated by a period match the consensus and dashes indicate a gap in the sequence.

Table 3-1 : Predicted *Ladybird* -encoding Genes in Teleost Fish

Organism	Gene	Linkage Group	Protein Accession
Zebrafish	*lbx1a*	1	XP_001333647
	lbx1b	13	NP_001020703
	lbx2	14	NP_001007135
Stickleback	*lbx1a*	Group VI	ENSGACP00000005054
	lbx1b	Group IX	ENSGACP00000024178
	lbx2	Group IV	ENSGACP00000022038
Medaka	*lbx1b*	1	ENSORLP00000015925
	lbx2	scaffold_1066	n/a
Takifugu	*lbx1a*	scaffold_52	ENSTRUP00000001074
	lbx1b	scaffold_62	ENSTRUP00000038346
	lbx2	scaffold_70	ENSTRUP00000019475
Tetraodon	*lbx1a**	Un-mapped	ENSTNIP00000006408
	lbx1b	18	ENSTNIP00000022571
	lbx2	20	ENSTNIP00000016035

Zebrafish Lbx protein coding sequences, mRNA and predicted amino acid sequences were used to BLAST other fish genomes. From available genomic sequence, homologs of all three zebrafish Ladybird-like genes were found in other teleost fish, with the exception of *lbx1a* in Medaka. * A partial *lbx1a* coding sequence was identified in Tetraodon and the predicted location of *lbx1a* in the Medaka genome does not have available sequence data. The same *ladybird* gene designation has been demonstrated by Wotton *et al.* , 2008.

location of *lbx1a* in Medaka has not been sequenced yet. Each ladybird protein contains several variable regions that are not conserved between vertebrates, but do exhibit significant conservation within teleost fish, which allows for easy determination of orthology (Figure 3-3). As database information is incomplete, only a partial *lbx1a* coding region was found in *Tetraodon* and a sequence gap in the Medaka genome occurs in the region where the *lbx1a* locus is expected to reside.

Zebrafish Ladybird proteins have diverged from the other four teleost fish species mentioned previously and on average show 10% less identity (Table 3-2) to other fish species for each ladybird protein. For example, zebrafish Lbx1a is 57% and 55% identical to Stickleback and Fugu Lbx1a, respectively, while Stickleback and Fugu Lbx1a exhibit 73% identity between each other. Additionally, both zebrafish Lbx1b and Lbx2 are on average 84% identical to the corresponding protein in other teleost fish species, however, excluding zebrafish ladybird proteins from amino acid sequence comparisons, Lbx1b and Lbx2 are on average 93% identical between the four other teleost fish species mentioned. These results are expected as *Tetraodon*, Fugu, Stickleback and Medaka are evolutionarily more related to each other than to zebrafish (Froschauer *et al.*, 2006). Conservation of Lbx1 between mammalian species is significantly high, typically >90 % identity and the conservation between mammals and the sole Ladybird protein in *Xenopus* is still >70% identity (Table 3-2). An interesting observation of mammalian Lbx2 protein comparisons is that amino acid identity

Table 3-2. Amino Acid Sequence Identity Between Ladybird Proteins

	Human Lbx1	Monkey Lbx1	Mouse Lbx1	X. tropicalis Lbx1	X. laevis Lbx1	Zebrafish Lbx1a	Stickleback Lbx1a	Takifugu Lbx1a	Zebrafish Lbx1b	Stickleback Lbx1b	Takifugu Lbx1b	Tetraodon Lbx1b	Medaka Lbx1b	Zebrafish Lbx2	Stickleback Lbx2	Takifugu Lbx2	Tetraodon Lbx2	Medaka Lbx2	Human Lbx2	Monkey Lbx2	Mouse Lbx2
Human Lbx1		98	96	71	72	58	64	64	71	69	69	70	70	58	58	58	57	58	35	35	33
Monkey Lbx1			97	71	72	58	64	64	71	69	69	70	70	59	59	58	58	58	35	35	33
Mouse Lbx1				71	72	57	64	65	71	69	69	69	70	58	58	57	57	57	35	35	33
X. tropicalis Lbx1					93	58	64	64	74	72	71	72	72	60	61	61	60	60	37	37	35
X. laevis Lbx1						58	64	64	72	72	71	72	72	60	61	61	60	59	37	37	35
Zebrafish Lbx1a							57	55	62	69	66	66	66	55	54	54	53	55	36	35	33
Stickleback Lbx1a								73	68	67	68	68	67	57	57	57	57	56	35	34	32
Takifugu Lbx1a									68	66	67	67	67	56	56	55	54	55	33	32	31
Zebrafish Lbx1b										84	84	84	85	61	61	61	60	60	35	35	34
Stickleback Lbx1b											94	93	95	59	60	60	59	59	36	36	33
Takifugu Lbx1b												97	95	60	61	60	60	60	36	36	34
Tetraodon Lbx1b													95	60	61	61	60	61	36	36	34
Medaka Lbx1b														59	60	60	59	60	36	36	34
Zebrafish Lbx2															85	84	84	83	38	39	36
Stickleback Lbx2																92	92	89	38	38	37
Takifugu Lbx2																	97	88	38	38	36
Tetraodon Lbx2																		87	37	39	36
Medaka Lbx2																			38	38	37
Human Lbx2																				93	73
Monkey Lbx2																					74
Mouse Lbx2																					

Results of pairwise comparison between vertebrate Ladybird protein amino acid sequences. Red, orange, yellow, green and blue colored boxes indicate >85%, 76-85%, 66-75%, 56-65% and <56% amino acid sequence identity, respectively. All Ladybird proteins found in teleost fish exhibit much higher identity to mammalian and amphibian Lbx1, and not Lbx2. Fish Lbx1b shows the highest degree of conservation, with an average of 70% identity to mammalian Lbx1, while Lbx1a and Lbx2 from fish average 58% identity to mammalian Lbx1. Within fish species alone, Lbx1b and Lbx2 show a high degree of conservation, averaging 91% and 88% identity, respectively, while Lbx1a appears to be accumulating sequence changes more rapidly. Complete Lbx1a amino acid sequences were not identified in Tetraodon and Medaka.

is typically a lot lower than for Lbx1 proteins between the same two species, suggesting rapid sequence divergence for Lbx2 proteins (Wotton *et al.*, 2008).

The most striking distinction between Lbx1 and Lbx2 is in the homeodomain region, where nearly 35% of the residues are different between mammalian Lbx1 and Lbx2. All three ladybird proteins in teleost fish have homeodomain sequences that are identical to mammalian Lbx1 except for three residues in zebrafish Lbx1a and one residue in all fish Lbx2 proteins (Figure 3-4). Furthermore, *Drosophila melanogaster* ladybird early and ladybird late homeodomain sequences display 87% similarity to mammalian Lbx1. These results point to the consideration that the divergence of a second Lbx protein is specific to birds and mammals. While the amino acids shown to be critical in DNA sequence recognition in the amino terminal arm and helix 3 of the homeodomain (Berger *et al.*, 2008) are the same between Lbx1 and Lbx2 proteins, helices 1 and 2, in addition to other regions within the protein, are significantly different, which could modulate DNA binding specificity through interactions with other proteins (Chang *et al.*, 1996; Hassan *et al.*, 1997).

Phylogenetic analysis of ladybird amino acid sequences provides an overall perspective of ladybird protein evolution in vertebrates (Figure 3-5). With the lower sequence conservation mentioned previously, teleost fish Lbx1a sequences show a higher degree of divergence and do not fall onto a clear branch of the phylogeny. It is worthy to note that only partial Lbx1a sequences have been identified in *Tetraodon* and the Medaka genome has a sequence gap at the

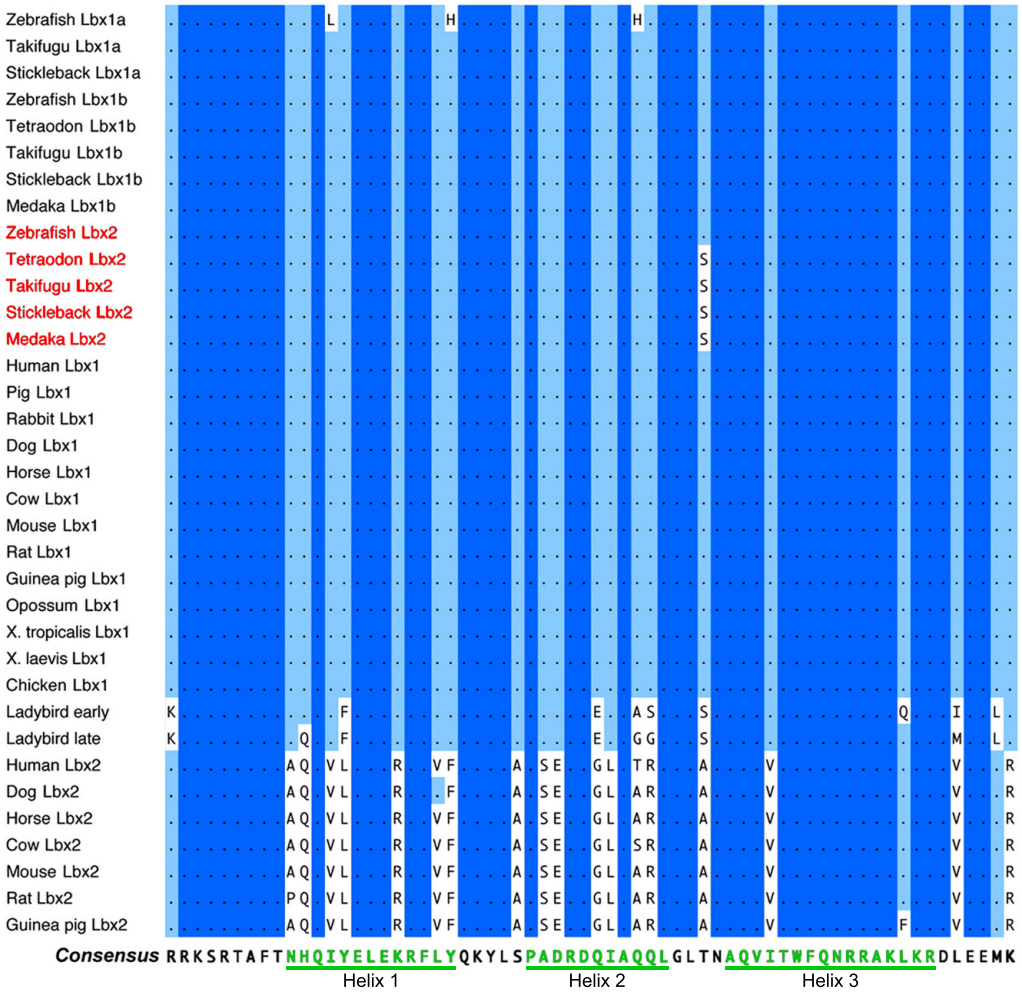

Figure 3-4. **Homeodomain sequences of Lbx2 have diverged from Lbx1**. Amino acid alignment of Ladybird homeodomain regions from a variety of species, including *Drosophila melanogaster* Ladybird early and Ladybird late. Sequences in dark blue are conserved between all Ladybird proteins shown while those in light blue are typically conserved in only Lbx1 family members. Note that in teleost fish, homeodomain sequences of proteins named Lbx2 (red) more closely resemble Lbx1 homeodomain sequences. Predicted homeodomain helices are indicated in green in the consensus sequence shown below and a period in the alignment indicates a match to the consensus.

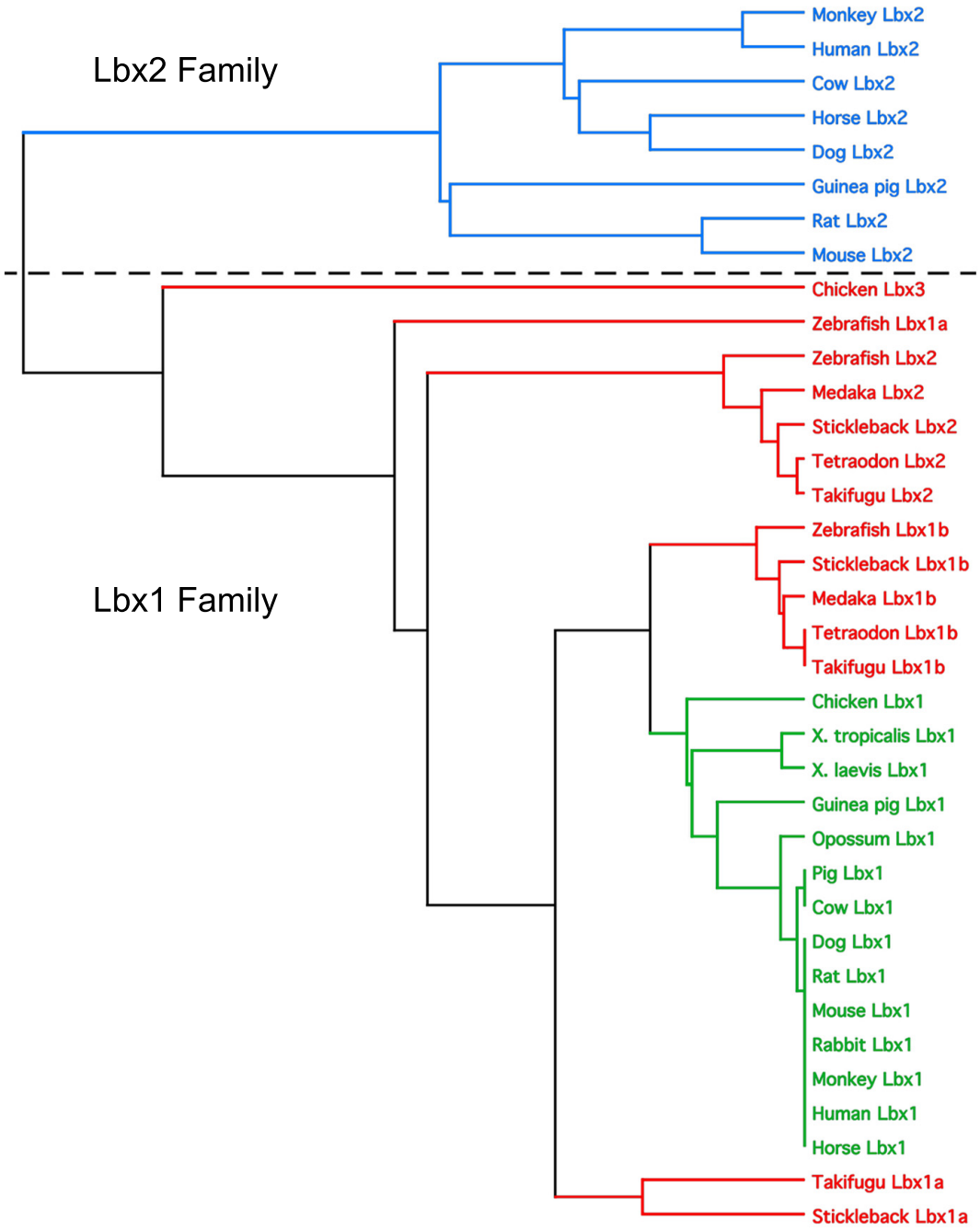

Figure 3-5. **Phylogenetic analysis of vertebrate Lbx proteins**. Shown is the best tree constructed from ClustalW sequence alignment analysis using UPGMA (Michener and Sokal, 1957) with poisson-correction and sequence gaps were ignored. Lbx proteins group into 2 distinct families, however, fish and bird species have Lbx proteins that exhibit clear differences (red) from Lbx1 (green) and Lbx2 (blue), but fall onto the Lbx1 family branch. Fish Lbx1b is most similar to Lbx1 in other vertebrates but Lbx1a and Lbx2 form discrete out-groups from the Lbx1 cluster.

predicted location of the *lbx1a* locus; two pieces of data that would make the phylogeny of *lbx1a* genes more clear. Of all the predicted teleost fish Lbx proteins, Lbx1b is grouped closest to other vertebrate Lbx1 proteins (Figure 3-5), with an average of 70% identity to human LBX1. Teleost fish Lbx2 proteins form a tight, separate branch within the Lbx1-like protein family; however, are still very distant to the non-fish Lbx1 cluster. Another interesting feature of Ladybird protein phylogeny is the increased branch length exhibited by mammalian Lbx2 proteins when compared to Lbx1 proteins, which suggests more rapid sequence divergence of Lbx2.

The differences between teleost fish Lbx proteins to other known vertebrate Ladybird proteins, and the naming discrepancies between various databases was clarified by a detailed synteny analysis of paralogous *lbx* genes in fish and their homology to mammalian *Lbx1* or *Lbx2*. Due to the high level of amino acid sequence conservation and position on the phylogenetic tree, it was expected that teleost fish *lbx1b* would reside in a chromosomal environment similar to mammalian Lbx1. However, it was unclear which mammalian *ladybird* gene would show synteny to teleost *lbx1a* or *lbx2*. A comprehensive genomic map of human chromosome 10, where *LBX1* resides, reveals that both *lbx1a* and *lbx1b* from zebrafish demonstrate synteny to human *LBX1* (Figure 3-6). Numerous genes and gene clusters within 3 Mbp on either side of human *LBX1* are dispersed throughout zebrafish LG 1, where *lbx1a* resides, and LG 13, where *lbx1b* resides. There are several genes in proximity of zebrafish *lbx2* on LG 14 which are

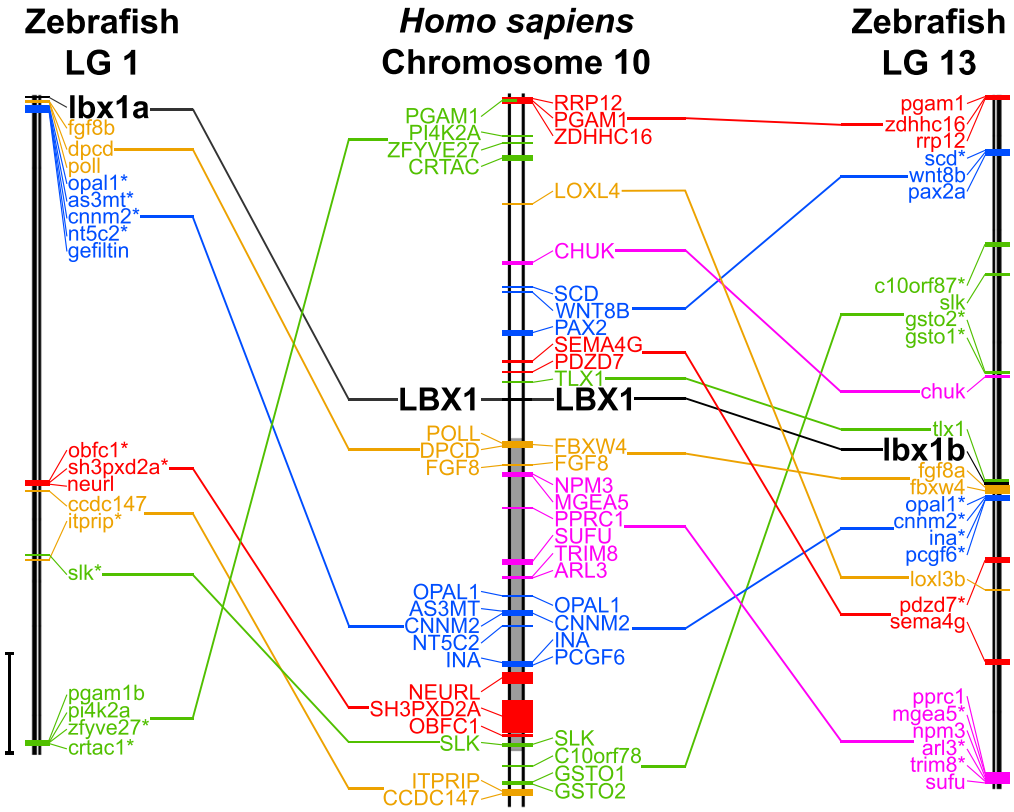

Figure 3-6. **Linkage and synteny analysis of zebrafish *lbx1a* and *lbx1b***. Synteny maps for zebrafish *lbx1a*, *lbx1b* and human *LBX1* were drawn from the genomic information available at http://www.ncbi.nlm.nih.gov and http://www.ensembl.org. Syntenic genes and gene clusters between zebrafish linkage group (LG) 1 and 13 and human chromosome 10 are matched by color. Zebrafish genes marked with an asterisk do not have a formal name but sequence alignments show a high degree of homology to the corresponding human homologue. Maps and lengths spanned by genes or gene clusters are drawn approximately to scale; scale bar represents 5 Mb for zebrafish linkage groups and 1.5 Mb on human chromosome 10. The side of *lbx1a* from LG 1 not shown exhibits significant synteny to human chromosome 13 and 3. The length shaded in grey denotes region 10q24.32.

syntenic to human *LBX2* on chromosome 2, however, are also syntenic to human chromosome 5, which contains the chromosomal remnants of a second ancestral vertebrate *Lbx2* locus. Unfortunately, the number of genes and gene clusters surrounding zebrafish *lbx2* are low and not nearly as definitive as the synteny observed for genes and gene clusters surrounding zebrafish *lbx1a* and *lbx1b* loci. As fish *lbx1b* is genetically more similar to mammalian *Lbx1*, it is not surprising that a higher percentage of genes neighbouring human *LBX1* are also found on linkage group 13 in zebrafish, near *lbx1b*. Many of the genes at the proximal end of region 10q24.32 in the immediate vicinity of human *LBX1* can be found in the immediate vicinity of either *lbx1a* or *lbx1b*, however, numerous gene clusters that are syntenic to this region of human chromosome 10 are located several, sometimes tens of Mbp away in zebrafish. Intrachromosomal recombination such as inversions, which are usually far more prevalent than interchromosomal recombination (Kaplan and Hudson, 1987), is a likely explanation for the observation that many genes found near *LBX1* still reside on the same chromosome as *lbx1a* or *lbx1b*.

Strictly from sequence information and chromosomal data, the chromosomal location of *lbx1b* from teleost fish resembles mammalian *Lbx1* the most, and *lbx1a* is a paralog. Teleost fish *lbx2* still remains in question, as the coding sequence resembles *Lbx1*, however, the chromosomal environment in the immediate vicinity of zebrafish *lbx2* suggests it might be the ortholog of mammalian *Lbx2*. It is possible a gene conversion (Benovoy and Drouin, 2009)

event took place in fish whereby *lbx1*-like coding sequences were used to repair the true *lbx2* gene. One line of evidence to support this theory is an analysis of the mammalian *Lbx2* locus, where no surrounding genomic sequence is conserved to teleost fish *lbx2* loci. The proposed gene conversion would place *lbx1*-like coding sequences in a chromosomal environment syntenic to mammalian *Lbx2*, thus giving teleost fish three *lbx1*-like genes of which only two are similar and syntenic to mammalian *Lbx1*.

b) Expression Patterns of Zebrafish *Ladybird* Genes

A powerful tool to assess gene expression during early development is mRNA *in situ* hybridization, which has many advantages in zebrafish as embryos are easily manipulated, plentiful from adult crosses, and optically transparent throughout early developmental stages. In general, a labelled probe with sequence complementary to the mRNA target is hybridized at high stringency and then detected by colorimetric (Oxtoby and Jowett, 1993; Thisse and Thisse, 2008) or more recently, fluorometric (Clay and Ramakrishnan, 2005; Brend and Holley, 2009) reactions. A challenge with designing probes is to target as much unique transcript sequence as possible to avoid probe cross-reactivity to other transcripts, such as those with highly conserved homeodomains. To achieve maximal probe signal it is important to examine the adenine content of the mRNA target as most probe labels, such as digoxigenin and fluorescein, are conjugated to deoxyuridine triphosphate (dUTP). With nearly complete transcript sequence information

available, it was possible to produce reasonably unique probes to each of the three zebrafish *lbx* transcripts. For both *lbx1a* and *lbx1b*, probe sequence was derived from the coding region downstream of the homeodomain and the presumptive 3'UTR sequences. Putative 3'UTR sequences were determined by scanning the downstream genomic sequence for transcription termination signals AATAAA. Complete transcript information was available for *lbx2* from databases and therefore more unique probe sequences could be generated to target the 5'UTR and coding sequence prior to the homeodomain.

Expression of all three zebrafish *ladybird* genes begins early in development at the end of gastrulation (epiboly), 9-10 hours post-fertilization (hpf). It persists well up to 5 days post-fertilization (dpf) where embryo thickness and tissue density limits the detection of gene transcripts by whole mount mRNA *in situ* hybridization. Transcripts of zebrafish *lbx1a* (Figure 3-7) and *lbx1b* (Figure 3-8) are detected in very similar regions of the embryo, namely the dorsal neural tube and pectoral fin muscle. During early somitogenesis stages from 10 – 14 hpf, expression is detected in the hindbrain and later, in the dorsal spinal cord, as well as a characteristic double-striped pattern indicative of pectoral fin muscle. However, both the pattern within the neural tube and the onset of pectoral fin muscle expression, are very different between *lbx1a* and *lbx1b*. While at later stages, *lbx1a* and *lbx1b* are expressed in the same areas in roughly the same pattern, *lbx2* (Figure 3-9) expression is considerably more dynamic. *Lbx2* expression ranges from pre-somitic and lateral plate mesoderm at the end of

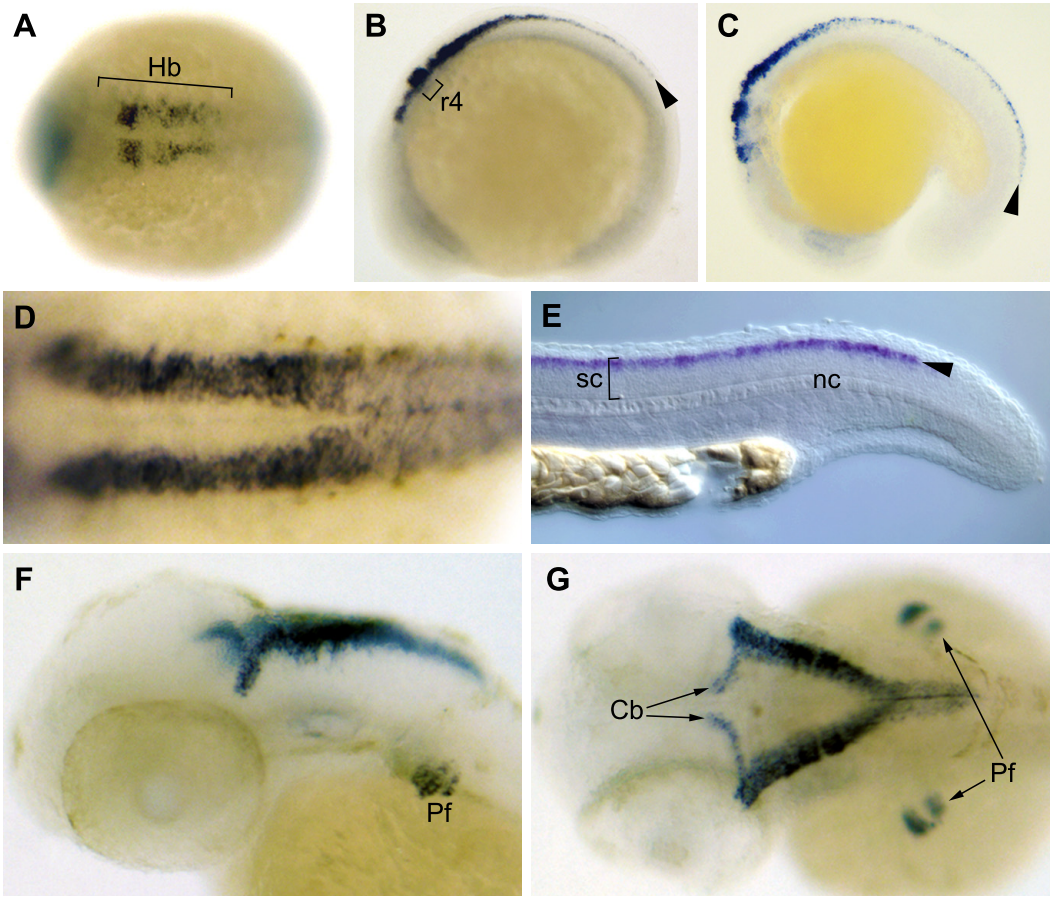

Figure 3-7. **Expression of zebrafish *lbx1a* during early embryo development**. With mRNA *in situ* hybridization, *lbx1a* transcripts are first detected in the hindbrain (Hb) during early somitogenesis stages (4S; A) with rhombomere-specific differences. This variation persists and is easily seen at 10S (B) and 18S (C) where rhombomere 4 (r4) exhibits a much broader expression domain along the DV axis. Transcripts are also detected in the dorsal spinal cord (sc; arrowhead) and at 24 hpf (D, E) the variability between rhombomeres has subsided. Later in development, at 48hpf (F and G), hindbrain expression is much more uniform and *lbx1a* is expressed in pectoral fin muscle (Pf) and the cerebellum (Cb). All panels have anterior to the left with dorsal (A, D, G) or lateral (B, C, E, F) views and panel E is a flat mount image using DIC optics.

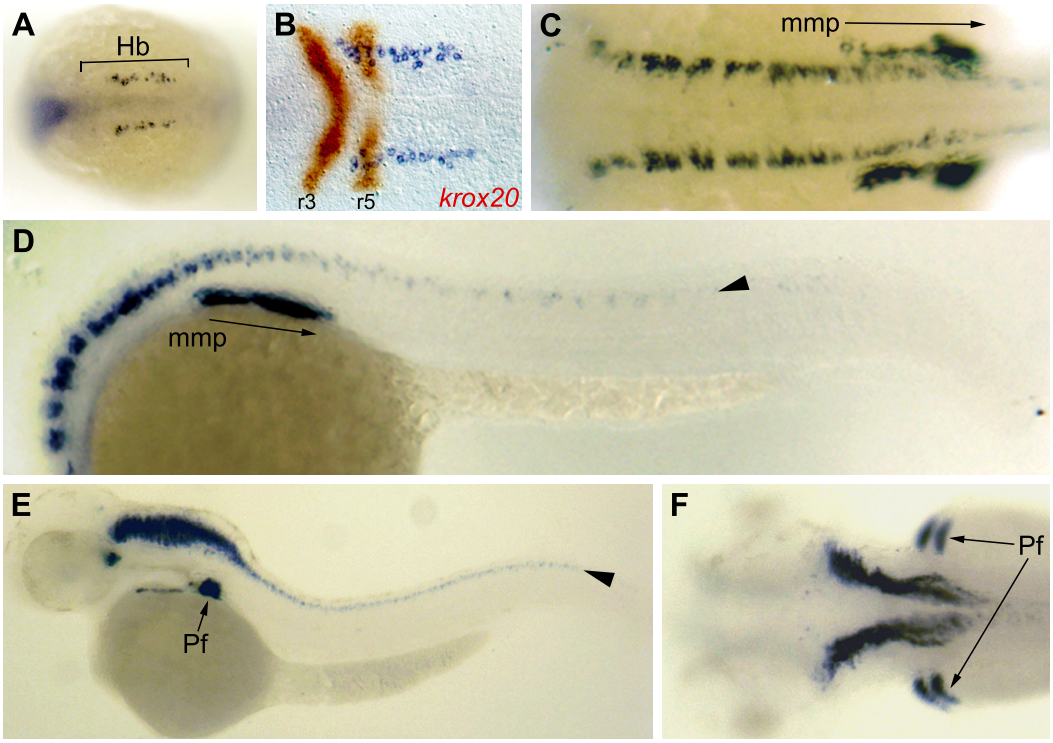

Figure 3-8. **Expression of *lbx1b* during early zebrafish development**. Shortly after gastrulation has completed (4S; A), *lbx1b* transcripts are detected in discrete bilateral stripes in the hindbrain (Hb) as shown with 2-color *in situ* hybridization (B) for *krox20* (red), which marks rhombomeres (r) 3 and 5. At 24 hpf (C) expression is seen in presumptive migratory muscle precursors (mmp) of the pectoral fin that migrate (arrow in C and D) to their final position in the pectoral fins. At 48 hpf (E) the majority of *lbx1b* expression is in the hindbrain and dorsal spinal cord (arrowhead), as well as pectoral fin (Pf) muscle, which is clearly seen at 72 hpf (F). All panels have anterior to the left with dorsal (A-C, F) or lateral (D, E) views. The embryo in panel B was deyolked and flat mounted.

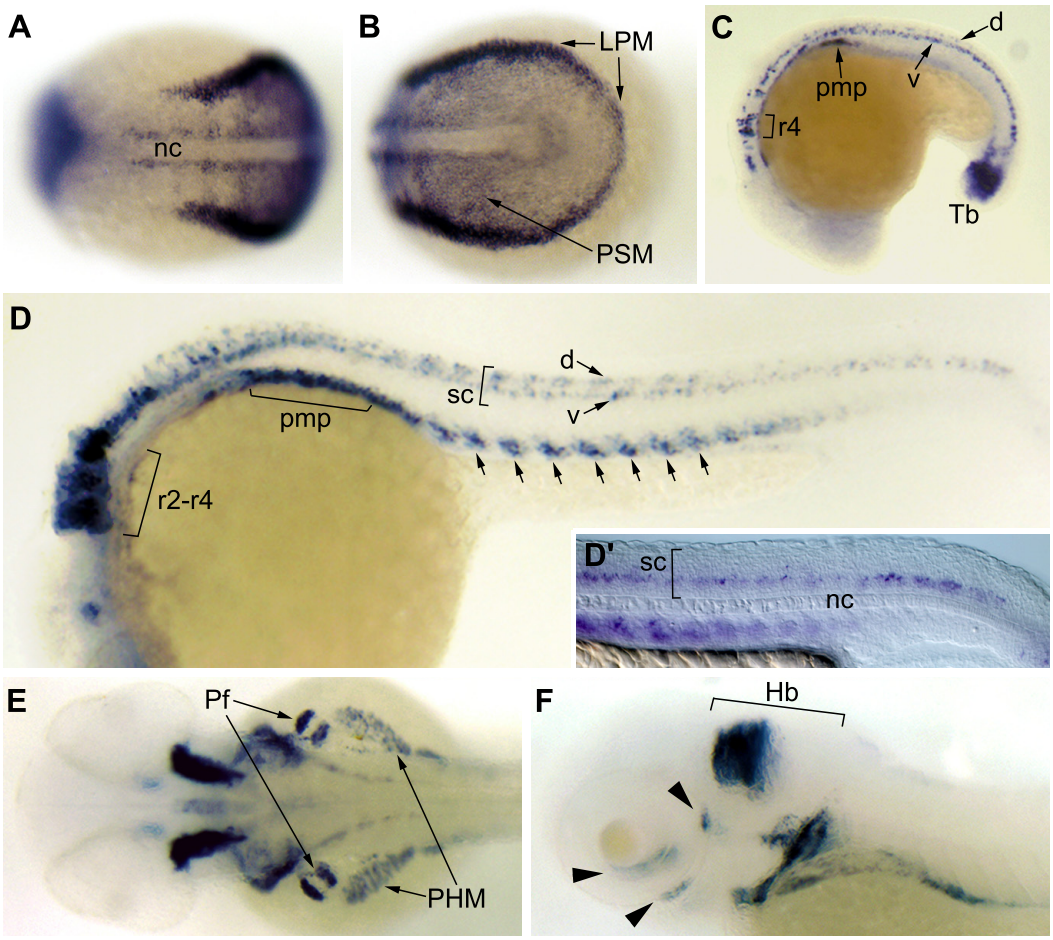

Figure 3-9. **Zebrafish *lbx2* expression during early embryo development**. At early embryonic stages (4S, A and B) expression of *lbx2* is seen in pre-somitic (PSM) and lateral plate (LPM) mesoderm as well as putative spinal cord neurons adjacent to the notochord (nc). At 18S (C) rhombomere 4 (r4) of the hindbrain and the tailbud (Tb) show high levels of *lbx2* expression. Additionally, transcripts are detected in both ventral (v) and dorsal (d) neurons within the dorsal half of the spinal cord (sc) during somitogenesis. The ventral myotome exhibits clear expression between somites (arrows) at 24 hpf (D) and in presumptive muscle precursors (pmp) along the lateral trunk. At the beginning of hatching stages (48 hpf; E and F), *lbx2* is restricted to the rostral hindbrain and several clusters of putative muscle around the eye (arrowheads), pectoral fin (Pf) muscle and presumptive posterior hypaxial muscle (PHM). All panels have anterior to the left with dorsal (A, B and E) or lateral (C, D and E) views. Panel D' shows high magnification of the tail using DIC optics to visualize the plane of expression in the spinal cord.

gastrulation to the tailbud and a small subset of dorsal neurons during mid-somitogenesis and progresses to various fin muscle groups in pharyngula (24 – 48 hpf) and hatching (48 – 72 hpf) stages (Kimmel *et al.*, 1995).

More detailed analysis of *lbx1a* expression reveals a very distinct variation in intensity between rhombomeres of the early segmented hindbrain. Rhombomere 4 (r4) exhibits the highest levels observed, r2 and r6-8 exhibit a discernable intermediate expression level, and r1, r3 and r5 have barely detectable levels of *lbx1a* transcript (Figure 3-7, A-C). A very thin layer of spinal cord neurons, which appear to reside at the dorsal edge, also express *lbx1a* starting shortly after somitogenesis begins at 10 hpf (Figure 3-7, B, C and E). The characteristic double-stripe expression pattern of *lbx1a* in pectoral fin muscle (Figure 3-7, G) is apparent at approximately 36 hpf and persists to 5 dpf. The variable expression between rhombomeres of the hindbrain becomes uniform at approximately 24 hpf (Figure 3-7, D) and remains this way past 48 hpf (Figure 3-7, F). At stages where hindbrain expression of *lbx1a* is more uniform, the caudal cerebellum, which arises from the rostral-most region of the hindbrain, referred to as r0, also expresses *lbx1a* (Figure 3-7, F). The pattern of *lbx1a* expression in zebrafish is fairly tissue specific being in the dorsal neural tube posterior to the midbrain and in pectoral fin musculature, however, the dynamic expression observed early in the hindbrain poses an intriguing avenue of research which will be addressed in chapter 4.

Detection of *lbx1b* follows much the same pattern as *lbx1a* with key differences in hindbrain and pectoral fin muscle. Hindbrain expression is restricted to 2 discrete bilateral stripes positioned approximately midway between the lateral edge of the hindbrain and the midline at stages where the neural plate is still flat (Figure 3-8, A). Shortly after gastrulation is complete, a small number of neurons spanning r4 to the presumptive caudal end of the hindbrain (Figure 3-8, B) express *lbx1b* and this develops into a series of neuronal clusters along these fine mediolateral stripes throughout the entire hindbrain (Figure 3-8, C) and dorsal spinal cord (Figure 3-8, D and E). The significant difference in pectoral fin muscle expression between *lbx1a* and *lbx1b* is that the onset of *lbx1b* expression occurs much earlier. At 24 hpf (Figure 3-8, C) a population of cells migrating posteriorly along the ventrolateral edge of the trunk (Figure 3-8, C and D) end up in the position where pectoral muscles are definitively identified by two mediolateral stripes of expression (Figure 3-8, F), suggesting these cell populations are pectoral fin progenitors.

The dynamic nature of *lbx2* expression in zebrafish suggests a multifaceted role in early development, which is not compensated for by the two additional *lbx* genes mentioned previously. Prior to the end of gastrulation, *lbx2* transcripts are detected in the lateral plate mesoderm and pre-somitic mesoderm (Figure 3-9, A and B). While expression is even more refined in the hindbrain, by mid-somitogenesis stages (Figure 3-9, C), the pattern of *lbx2* is remarkably similar to the *lbx1* genes, where levels are highest in r4, intermediate in r2, r6-8

and very low in r1, r3 and r5. High levels are also detected in the tailbud (Figure 3-9, C), a region that generates additional trunk segments, or somites, through a coordinated action of growth factors (Özbudak and Pourquié, 2008). Another interesting feature of *lbx2* expression is in the spinal cord; a very thin layer at the dorsal edge (Figure 3-9, C and D) is detectable, similar in position to that of *lbx1a* and *lbx1b*, as well as another layer of spinal cord neurons positioned in the middle of the dorsoventral axis. By 24 hpf rhombomeres 2-4 exhibit very high levels of *lbx2* transcripts and a visible stream-like pattern of cells laterally adjacent to the caudal hindbrain exists, which is reminiscent of *lbx1b* expression in presumptive migrating muscle precursors of the pectoral fin (Figure 3-9, D). At 48 hpf several fin muscle groups are positive for *lbx2* (Figure 3-7, D), including the pectoral fin muscles and posterior hypaxial muscle as well as a series of tiny clusters surrounding the eye. While *lbx2* is present along the entire length of the hindbrain at earlier embryonic stages, beyond 36 hpf there is clear restriction to the rostral hindbrain (Figure 3-9, E and F).

Ladybird was identified as a potential early patterning gene in zebrafish through its apparent decreased expression in embryos lacking pre-B-cell leukemia homeodomain (Pbx) proteins by microarray analysis (Waskiewicz, unpublished data). Microarray data was corroborated by mRNA *in situ* hybridization for *ladybird* genes (Figure 3-10) in compound *pbx* knockdown/mutant zebrafish embryos with a phenotype that is indicative of Pbx protein deficiency. Early Pbx protein deficiency was accomplished by injecting morpholinos targeting *pbx2* and

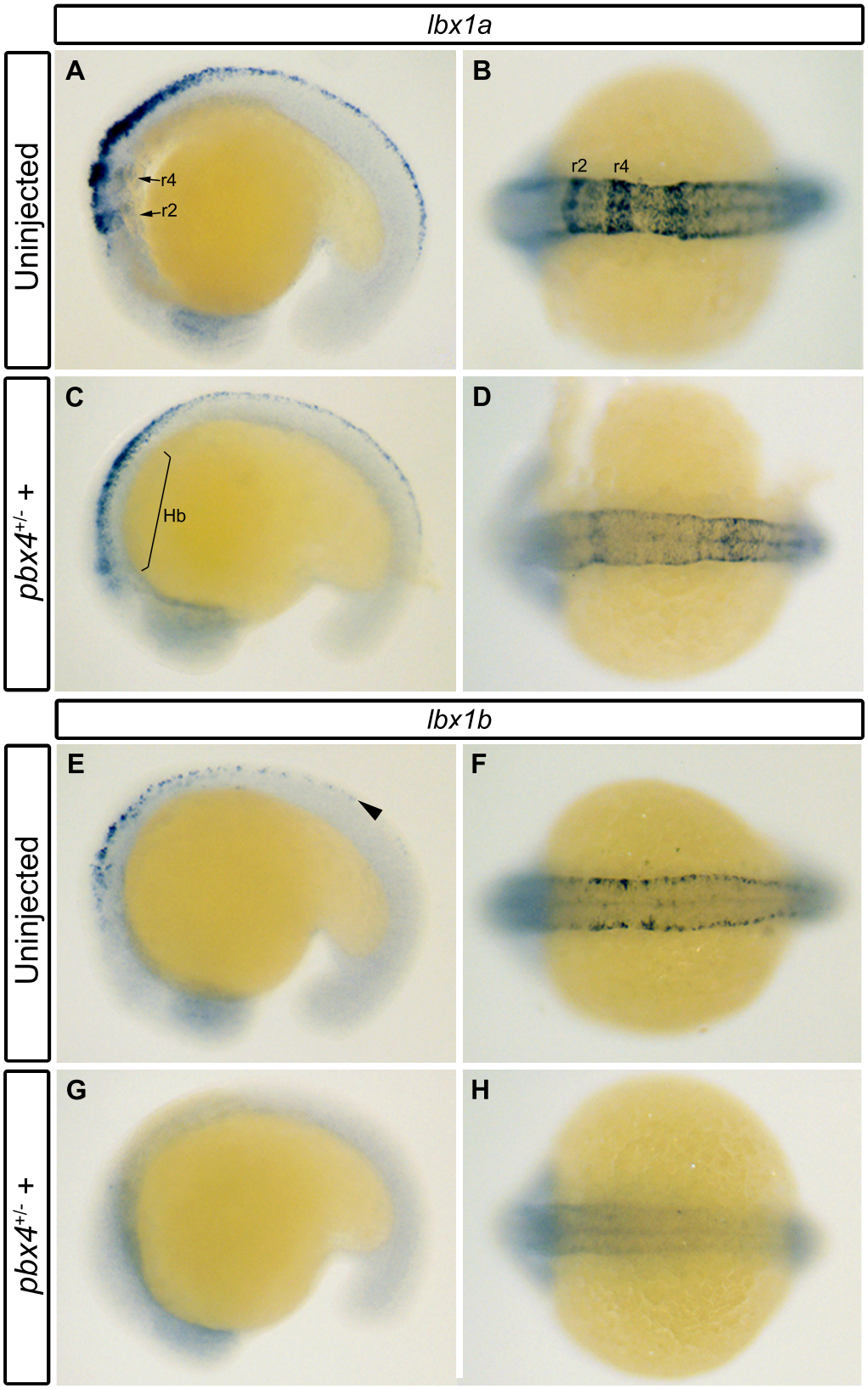
lbx1a
Uninjected
A
r4
r2
B
r2
r4
pbx4+/- +
C
Hb
D
lbx1b
Uninjected
E
F
pbx4+/- +
G
H

Figure 3-10. **Expression of *lbx1a* and *lbx1b* in *pbx*-MO injected *pbx4*$^{+/-}$ embryos**. Genotyped *pbx4*$^{+/-}$ (*Lzr*) adult fish were crossed and one-cell embryos were injected with a morpholino (MO) cocktail to reduce the levels of early Pbx proteins. At 18S (A-H), Pbx-knockdown (E-H) embryos can be visually sorted by a distinct morphological phenotype when compared to WT uninjected embryos (A-D), and analyzed for *in situ* hybridization for *lbx1a* (A, B, E, F) or *lbx1b* (C, D, G, H) mRNA. Expression of *lbx1b* in the dorsal neural tube (arrowhead) is completely dependent on Pbx proteins but *lbx1a* is still expressed, however, loses the distinct differences between certain rhombomeres (r) in the hindbrain (Hb). All panels have anterior to the left with dorsal (B, D, F, H) or lateral (A, C, E, G) views.

pbx4 into embryos derived from genotyped *pbx4*$^{+/-}$ adult crosses. At 18 hpf there is quite a different response between *lbx1a* and *lbx1b* to the loss of Pbx proteins, as *lbx1a* is still detected in the hindbrain and spinal cord (Figure 3-10, C and D), while *lbx1b* appears to be undetectable (Figure 3-10, G and H). This difference can be explained by how the hindbrain develops in the absence of Pbx; all of the rhombomeres exhibit a developmental ground state program (Waskiewicz *et al.*, 2002) and take on the identity of r1. It is known that *lbx1a* is expressed in the cerebellum at later stages, which is derived from r0, and transcripts can clearly be seen in r1 at low levels in wild type embryos (Figure 3-7, C, D). Therefore, with the conversion of all hindbrain segments to an r1-like character in the absence of Pbx protein function, it is expected that a low level of *lbx1a* expression would appear throughout the hindbrain. Another observation furthering the theory of a hindbrain ground state is that all rhombomere-specific differences in expression intensity of *lbx1a* are lost in Pbx-deficient embryos (Figure 3-10, A-D), suggesting either a uniform developmental timeline or a uniform character.

Furthermore, at 18 hpf, all visible expression of *lbx1b* is lost in zebrafish embryos lacking Pbx proteins (Figure 3-10, E-H). The phenomenon that all rhombomeres of the hindbrain take on a ground state with similar character to r1 potentially explains this observation. At 72 hpf the cerebellum is clearly distinct and genes expressed in this tissue appear as bilaterally symmetric curves, as seen in the case of *lbx1a* (Figure 3-7, G). Conversely, *lbx1b* is barely detectable in the cerebellum at 72 hpf, suggesting it is not expressed within r0 and expression in r1

is only detected after 24 hpf. Therefore, since rhombomeres in a Pbx-deficient hindbrain take on an r1-like character it is not surprising that *lbx1b* is not expressed in early Pbx-deficient embryos. The unusual aspect of *lbx1b* expression in a Pbx-deficient embryo is its absence in the spinal cord as it suggests zebrafish *ladybird* genes are regulated differently, despite being expressed in the same tissues. One explanation for this observation is the refined expression of *lbx1b* within the hindbrain and spinal cord when compared to *lbx1a*. At early embryonic stages, it is clear that *lbx1a* is more broadly expressed within these tissues (Figure 3-7) in contrast to *lbx1b* (Figure 3-8) and conceivably *lbx1b* is regulated in a spinal cord neuron type which requires Pbx while *lbx1a* regulation is Pbx-independent in the spinal cord.

The embryonic expression pattern of zebrafish *ladybird* genes follows the pattern observed in other vertebrates – expression in limb muscle and the dorsal neural tube. The results presented here add to the theory that Ladybird homeodomain proteins are an ancient class of transcription factors important for the development of both muscle and neural tissue. Several unknown clusters of *ladybird* expressing cells do exist in the zebrafish, around the eye and in the vicinity of the caudal pharyngeal arches. A comparison to the gene expression pattern observed in craniofacial muscles of zebrafish (Lin *et al*., 2006) suggests these unknown *lbx1*$^{+}$ clusters may in fact be muscles involved in movement of the eye and jaw.

c) Functional Analysis of Zebrafish *Ladybird* Genes

Zebrafish also provide the opportunity to assess gene function through knockdown of mRNA translation by the binding of a synthetic, chemically-modified oligonucleotide to the translation initiation site. These very stable morpholino oligonucleotides have very minimal side-effects; however, as cell division occurs and more transcripts accumulate, the pool of unbound morpholinos diminishes. Morpholino micro-injection is done at the 1-cell stage and can lead to non-specific developmental defects such as increased cell death and necrosis, which sometimes mask true gene-knockdown effects. Injection of each *lbx* morpholino alone yielded >75 % embryonic lethality prior to 24 hpf. Fortunately, these defects can be alleviated by co-injection of morpholino targeting the pro-apoptotic gene, *p53* (Langheinrich *et al.*, 2002; Robu *et al.*, 2007). Morpholinos were designed to bind to the translation start site of each zebrafish *ladybird* gene by Gene Tools, LLC and injected at a variety of doses (1-10 ng) in combination with 2 ng *p53* morpholino. It was expected that embryos may develop swimming abnormalities due to mis-formed pectoral fins or to be unresponsive to gentle touch from abnormal sensory connections of the dorsal spinal cord. No consistent phenotypic abnormalities were observed during early embryo development and embryos did not display any obvious behavioural defects. These results could be explained by gene product redundancies and/or incomplete knockdown by morpholinos, or the presence of defects that are too subtle to detect by observing the phenotype or behaviour of live embryos.

Ectopic expression of *ladybird* genes in zebrafish was accomplished through injection of *in vitro* synthesized mRNA into 1-cell embryos. Zebrafish *ladybird* coding sequences were cloned into the pTST7 expression plasmid, which contains a multiple cloning site flanked by the *Xenopus* β-globin untranslated regions. The β-globin UTRs provide transcript stability which allows for persistent translation of the injected mRNA through numerous cell divisions. The disadvantage of injecting mRNA at the 1-cell stage is that its activity is generally instantaneous and can sometimes produce confusing results. Introducing a gene product at the 1-cell stage, which normally functions later in development can over-ride complete developmental programs thereby preventing the embryo from reaching stages at which a discernable phenotype can be observed. The observed defects from injecting *lbx1a*, *lbx1b* or *lbx2* mRNA were very similar, suggesting a parallel function, however, penetrance was an issue for *lbx1a* and *lbx1b* mRNA.

The most consistent results were obtained from injection of *lbx2* mRNA, where a very specific tissue, the forebrain, was severely affected. Early in somitogenesis the eye fields begin to separate from the forebrain and it is at this stage the first signs of a developmental defect from *lbx* mRNA injection are visible. Either one or both eye fields are absent, suggesting ectopic ladybird protein is preventing anterior tissues from being specified correctly. These embryos survive quite well in the first week of development and their behaviour is seemingly normal, however, they lack eyes and some forebrain tissue (Figure 3-11). Other than forebrain and eye tissue, it is striking how phenotypically normal

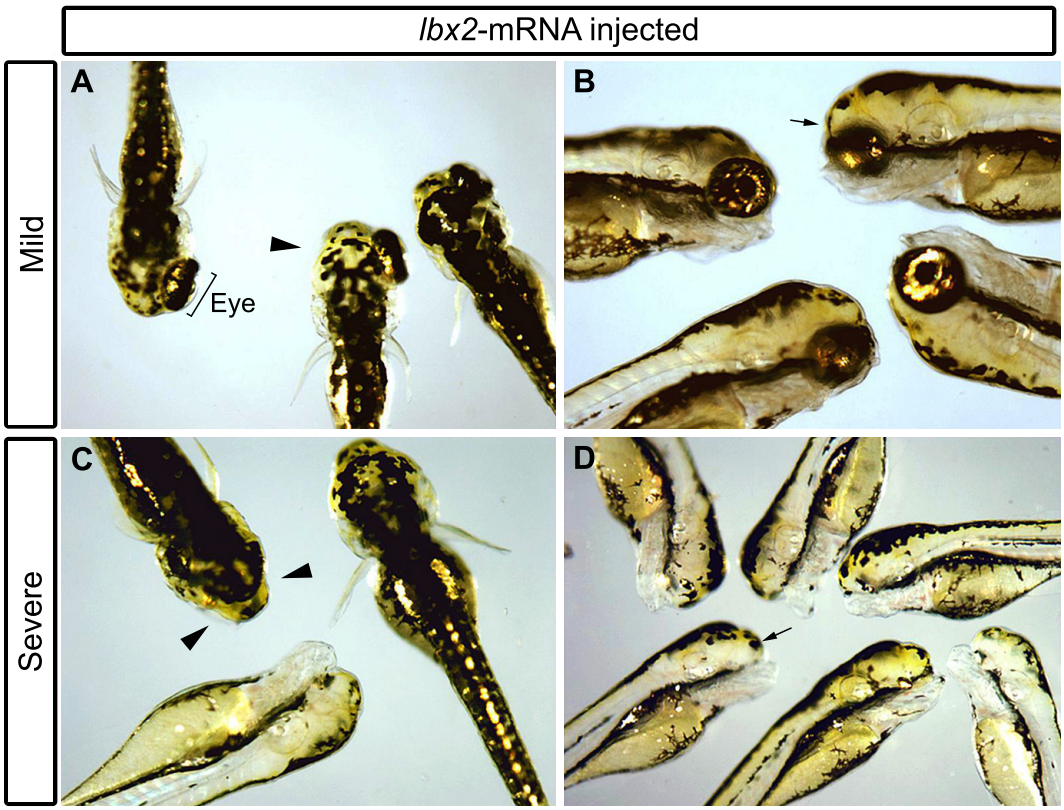

Figure 3-11. **Ectopic expression of *lbx2* inhibits anterior tissue development**. Wild type embryos were injected at the one-cell stage with *in vitro* transcribed 5'-capped mRNA encoding zebrafish Lbx2 and observed for severity of developmental defects. At 72 hpf (A-D) embryos show a near wild type body morphology but lack one (A, B) or both (C, D) eyes (arrowheads), as well as forebrain tissue (arrows). Eye morphology of single-eye embryos (B) appears normal and all embryos are able to swim and survive up to 2 weeks. Similar results were observed by injecting *lbx1a* or *lbx1b* mRNA, however, at lower penetrance than that of *lbx2* mRNA.

these embryos appear; they swim and respond to physical stimuli much like a wild type embryo does. In some cases, a single eye does develop properly and is located in its proper position (Figure 3-11, A and B) and in the most severe cases both eyes, as well as forebrain tissue, which normally protrudes over the jaw, are missing (Figure 3-11, C, D). These phenotypes are reminiscent of zebrafish mutations characterized in the *headless*/*tcf3* (Kim *et al*., 2000) and *masterblind*/*axin1* (Heisenberg *et al.*, 2001) genes, where repression of wingless (Wnt) target genes is lost and anterior tissue development is severely compromised leading to embryos lacking eyes and forebrain tissue.

To understand the mechanism behind the loss of anterior tissues in *lbx* mRNA injected zebrafish embryos, eye and forebrain markers, *pax6a* and *pax2a*, were examined by mRNA *in situ* hybridization between 18 and 24 hpf. Expression of *pax2a* in the optic stalk is lost in *lbx2*-mRNA injected embryos and the position of the mid-hindbrain boundary (MHB) is shifted forward significantly (Figure 3-12, A-D). The distance between the MHB and otic vesicle remains unchanged and the trunk of the embryo appears normal, suggesting that ectopic *lbx* mRNA only affects anterior tissues. In uninjected embryos the eye fields become clearly separated from the forebrain at approximately 14 hpf, yet by 20 hpf, *lbx2*-mRNA injected embryos fail to show any visible signs of eye development nor do they show molecular markers (Figure 3-12, C and D). In the case of single eyed-embryos, the expression pattern of *pax6a* explains the observed phenotype fairly well. *Pax6a* is expressed in parts of the telencephalon,

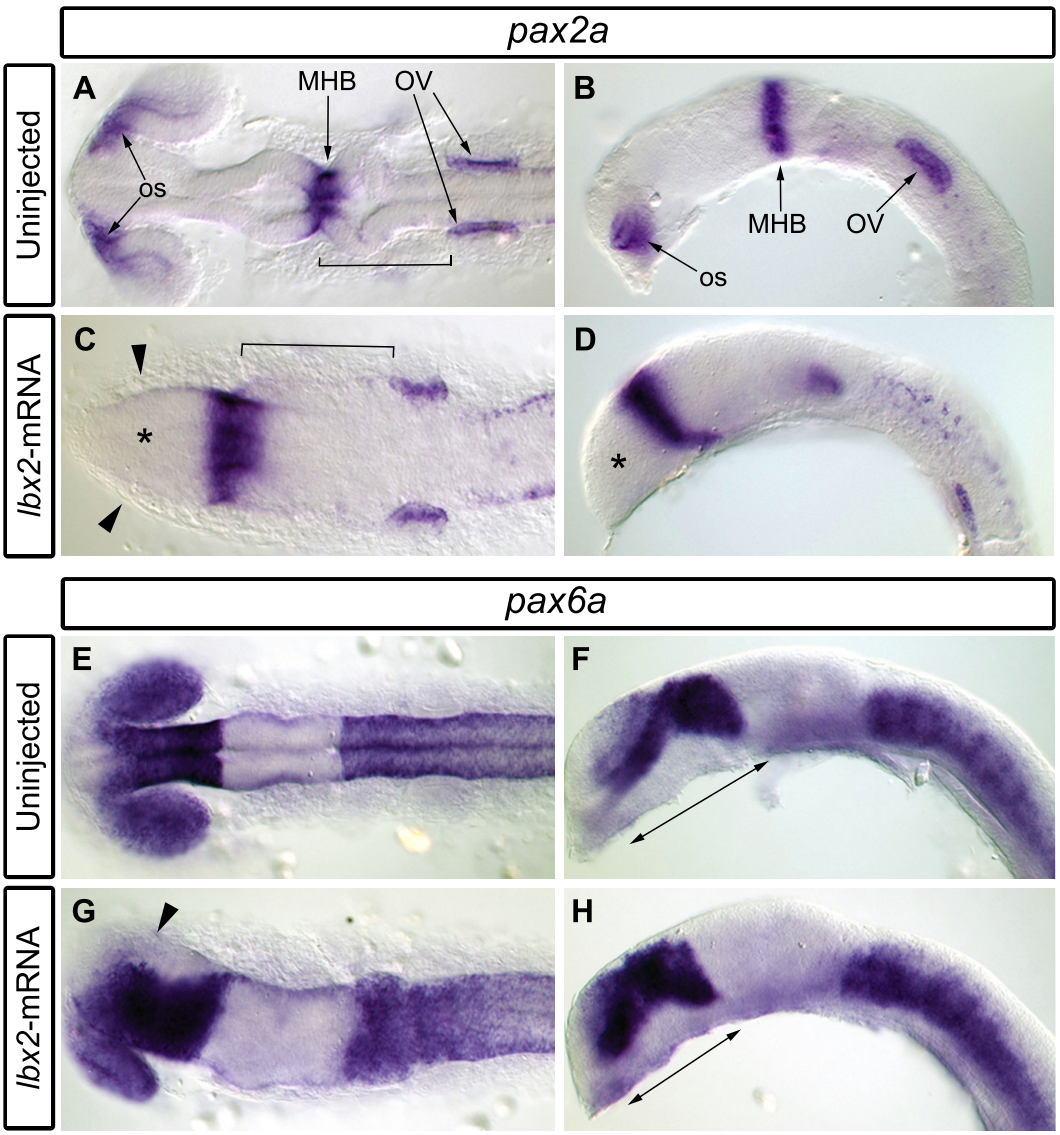

Figure 3-12. **Expression of *pax2a* and *pax6a* in *lbx2*-mRNA injected embryos**. Forebrain and eye markers, *pax2a* (A-D) and *pax6a* (E-H) were examined by *in situ* hybridization at 24 hpf (A-D) and 18 hpf (E-H) in uninjected (A, B, E, F) and *lbx2*-mRNA injected (C, D, G, H) embryos to determine the cause of the developmental defects seen in Figure 3-11. Absent eye fields (arrowhead) and loss of optic stalk (os) expression are visible early in development and in more severe cases (C, D), a significant proportion of anterior tissue (asterisk) is missing. The distance (brackets) between the mid-hindbrain boundary (MHB) and the otic vesicle (OV) appears unchanged, however, the anterior domain (double-arrow) of *pax6a* expression appears slightly shorter (F, H). All panels show anterior to the left with dorsal (A, C, E, G) or lateral (B, D, F, H) views.

diencephalon, eye, hindbrain and spinal cord (Figure, 3-12, E and F). However, in embryos with ectopically-expressed *ladybird* genes, a single eye field is absent (Figure 3-12, G) and the eye which is present appears to be in a normal location lateral to neural tissue.

These observations suggest that ladybird can override specific anterior developmental programs to prevent forebrain and eye development. The phenotype from ectopic *lbx* expression in early zebrafish embryos is similar to zebrafish *wnt* pathway mutants, specifically a mutation in the *tcf3* locus named *headless* (Kim *et al.*, 2000). The *headless* phenotype arises from the activation of *wnt* signalling in anterior tissues through the loss of a *wnt* repressor, Tcf3, which can be rescued by injection of mRNA encoding an anterior determinant gene, Six3 (Lagutin *et al.*, 2003). A similar phenotype arises when a mutation in *axin1* disrupts the glycogen synthase kinase 3 (gsk3) binding domain and leads to perturbation in Tcf-dependent transcription (Heisenberg *et al.*, 2001). Other studies have shown that ectopic *wnt* mRNA is capable of inhibiting the formation of anterior neural tissues in zebrafish (Kelly *et al.*, 1995). As *ladybird* has been shown to be a positive regulator of *wnt* signalling (Jagla *et al.*, 1997b), it is possible the observed phenotype in zebrafish with ectopic *ladybird* genes is due to ectopic *wnt* pathway activation.

3) References

Benovoy D and Drouin G, **2009**. Ectopic gene conversions in the human genome. *Genomics*. 93:27-32.

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Pena-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML and Hughes TR, **2008**. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*. 133:1266-1276.

Brend T and Holley SA, **2009**. Zebrafish whole mount high-resolution double fluorescent in situ hybridization. *J Vis Exp*. 25.

Brohmann H, Jagla K and Birchmeier C, **2000**. The role of Lbx1 in migration of muscle precursor cells. *Development*. 127:437-445.

Chang CP, Brocchieri L, Shen WF, Largman C and Cleary ML, **1996**. Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. *Mol Cell Biol*. 16:1734-1745.

Cheng L, Samad OA, Xu Y, Mizuguchi R, Luo P, Shirasawa S, Goulding M and Ma Q, **2005**. Lbx1 and Tlx3 are opposing switches in determining GABAergic versus glutamatergic transmitter phenotypes. *Nat Neurosci*. 8:1510-1515.

Clay H and Ramakrishnan L, **2005**. Multiplex fluorescent in situ hybridization in zebrafish embryos using tyramide signal amplification. *Zebrafish*. 2:105-111.

Copley RR, **2005**. The EH1 motif in metazoan transcription factors. *BMC Genomics*. 6:169.

De Graeve F, Jagla T, Daponte JP, Rickert C, Dastugue B, Urban J and Jagla K, **2004**. The ladybird homeobox genes are essential for the specification of a subpopulation of neural cells. *Dev Biol*. 270:122-134.

Dietrich S, Schubert FR, Healy C, Sharpe PT and Lumsden A, **1998**. Specification of the hypaxial musculature. *Development*. 125:2235-2249.

Froschauer A, Braasch I and Volff JN, **2006**. Fish Genomes, Comparative Genomics and Vertebrate Evolution. *Current Genomics*. 7:43-57.

Gehring WJ, **1987**. Homeo boxes in the study of development. *Science*. 236:1245-1252.

Gross MK, Dottori M and Goulding M, **2002**. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron*. 34:535-549.

Gross MK, Moran-Rivard L, Velasquez T, Nakatsu MN, Jagla K and Goulding M, **2000**. Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. *Development*. 127:413-424.

Han Z, Fujioka M, Su M, Liu M, Jaynes JB and Bodmer R, **2002**. Transcriptional integration of competence modulated by mutual repression generates cell-type specificity within the cardiogenic mesoderm. *Dev Biol*. 252:225-240.

Hassan B, Li L, Bremer KA, Chang W, Pinsonneault J and Vaessin H, **1997**. Prospero is a panneural transcription factor that modulates homeodomain protein activity. *Proc Natl Acad Sci U S A*. 94:10991-10996.

Heisenberg CP, Houart C, Take-Uchi M, Rauch GJ, Young N, Coutinho P, Masai I, Caneparo L, Concha ML, Geisler R, Dale TC, Wilson SW and Stemple DL, **2001**. A mutation in the Gsk3-binding domain of zebrafish Masterblind/Axin1 leads to a fate transformation of telencephalon and eyes to diencephalon. *Genes Dev*. 15:1427-1434.

Hoegg S, Brinkmann H, Taylor JS and Meyer A, **2004**. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J Mol Evol*. 59:190-203.

Huang H, Rastegar M, Bodner C, Goh SL, Rambaldi I and Featherstone M, **2005**. MEIS C termini harbor transcriptional activation domains that respond to cell signaling. *J Biol Chem*. 280:10119-10127.

Jagla K, Dolle P, Mattei MG, Jagla T, Schuhbaur B, Dretzen G, Bellard F and Bellard M, **1995**. Mouse Lbx1 and human LBX1 define a novel mammalian homeobox gene family related to the Drosophila lady bird genes. *Mech Dev*. 53:345-356.

Jagla K, Frasch M, Jagla T, Dretzen G, Bellard F and Bellard M, **1997a**. ladybird, a new component of the cardiogenic pathway in Drosophila required for diversification of heart precursors. *Development*. 124:3471-3479.

Jagla K, Georgel P, Bellard F, Dretzen G and Bellard M, **1993**. A novel homeobox nkch4 gene from the Drosophila 93E region. *Gene*. 127:165-171.

Jagla K, Jagla T, Heitzler P, Dretzen G, Bellard F and Bellard M, **1997b**. ladybird, a tandem of homeobox genes that maintain late wingless expression in terminal and dorsal epidermis of the Drosophila embryo. *Development*. 124:91-100.

Jagla K, Stanceva I, Dretzen G, Bellard F and Bellard M, **1994**. A distinct class of homeodomain proteins is encoded by two sequentially expressed Drosophila genes from the 93D/E cluster. *Nucleic Acids Res*. 22:1202-1207.

Kanamoto T, Terada K, Yoshikawa H and Furukawa T, **2006**. Cloning and expression pattern of lbx3, a novel chick homeobox gene. *Gene Expr Patterns*. 6:241-246.

Kaplan NL and Hudson RR, **1987**. On the divergence of genes in multigene families. *Theor Popul Biol*. 31:178-194.

Kasamatsu S, Sato A, Yamamoto T, Keng VW, Yoshida H, Yamazaki Y, Shimoda M, Miyazaki J and Noguchi T, **2004**. Identification of the transactivating region of the homeodomain protein, hex. *J Biochem*. 135:217-223.

Kelly GM, Greenstein P, Erezyilmaz DF and Moon RT, **1995**. Zebrafish wnt8 and wnt8b share a common activity but are involved in distinct developmental pathways. *Development*. 121:1787-1799.

Kennedy MA, Gonzalez-Sarmiento R, Kees UR, Lampert F, Dear N, Boehm T and Rabbitts TH, **1991**. HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24. *Proc Natl Acad Sci U S A*. 88:8900-8904.

Kim CH, Oda T, Itoh M, Jiang D, Artinger KB, Chandrasekharappa SC, Driever W and Chitnis AB, **2000**. Repressor activity of Headless/Tcf3 is essential for vertebrate head formation. *Nature*. 407:913-916.

Kim Y and Nirenberg M, **1989**. Drosophila NK-homeobox genes. *Proc Natl Acad Sci U S A*. 86:7716-7720.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B and Schilling TF, **1995**. Stages of embryonic development of the zebrafish. *Dev Dyn*. 203:253-310.

Lagutin OV, Zhu CC, Kobayashi D, Topczewski J, Shimamura K, Puelles L, Russell HR, McKinnon PJ, Solnica-Krezel L and Oliver G, **2003**. Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes Dev*. 17:368-379.

Langheinrich U, Hennen E, Stott G and Vacun G, **2002**. Zebrafish as a model organism for the identification and characterization of drugs and genes affecting p53 signaling. *Curr Biol*. 12:2023-2028.

Li R and Botchan MR, **1993**. The acidic transcriptional activation domains of VP16 and p53 bind the cellular replication protein A and stimulate in vitro BPV-1 DNA replication. *Cell*. 73:1207-1221.

Maqbool T, Soler C, Jagla T, Daczewska M, Lodha N, Palliyil S, VijayRaghavan K and Jagla K, **2006**. Shaping leg muscles in Drosophila: role of ladybird, a conserved regulator of appendicular myogenesis. *PLoS ONE*. 1:e122.

Martin BL and Harland RM, **2006**. A novel role for lbx1 in Xenopus hypaxial myogenesis. *Development*. 133:195-208.

Meyer A and Van de Peer Y, **2005**. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays*. 27:937-945.

Michener CD and Sokal RR, **1957**. A quantitative approach to a problem in classification. *Evolution*. 11:130-162.

Muller T, Brohmann H, Pierani A, Heppenstall PA, Lewin GR, Jessell TM and Birchmeier C, **2002**. The homeodomain factor lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron*. 34:551-562.

Oxtoby E and Jowett T, **1993**. Cloning of the zebrafish krox-20 gene (krx-20) and its expression during hindbrain development. *Nucleic Acids Res*. 21:1087-1095.

Ozbudak EM and Pourquie O, **2008**. The vertebrate segmentation clock: the tip of the iceberg. *Curr Opin Genet Dev*. 18:317-323.

Raju K, Tang S, Dube ID, Kamel-Reid S, Bryce DM and Breitman ML, **1993**. Characterization and developmental expression of Tlx-1, the murine homolog of HOX11. *Mech Dev*. 44:51-64.

Robu ME, Larson JD, Nasevicius A, Beiraghi S, Brenner C, Farber SA and Ekker SC, **2007**. p53 activation by knockdown technologies. *PLoS Genet*. 3:e78.

Schafer K and Braun T, **1999**. Early specification of limb muscle precursor cells by the homeobox gene Lbx1h. *Nat Genet*. 23:213-216.

Schubert FR, Dietrich S, Mootoosamy RC, Chapman SC and Lumsden A, **2001**. Lbx1 marks a subset of interneurons in chick hindbrain and spinal cord. *Mech Dev*. 101:181-185.

Sidow A, **1996**. Genome duplications in the evolution of early vertebrates. *Curr Opin Genet Dev*. 6:715-722.

Sieber MA, Storm R, Martinez-de-la-Torre M, Muller T, Wende H, Reuter K, Vasyutina E and Birchmeier C, **2007**. Lbx1 acts as a selector gene in the fate determination of somatosensory and viscerosensory relay neurons in the hindbrain. *J Neurosci*. 27:4902-4909.

Thisse C and Thisse B, **2008**. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc*. 3:59-69.

Tolkunova EN, Fujioka M, Kobayashi M, Deka D and Jaynes JB, **1998**. Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. *Mol Cell Biol*. 18:2804-2814.

Uchiyama K, Ishikawa A and Hanaoka K, **2000**. Expression of lbx1 involved in the hypaxial musculature formation of the mouse embryo. *J Exp Zool*. 286:270-279.

Urbach R and Technau GM, **2003**. Molecular markers for identified neuroblasts in the developing brain of Drosophila. *Development*. 130:3621-3637.

Waskiewicz AJ, Rikhof HA and Moens CB, **2002**. Eliminating zebrafish pbx proteins reveals a hindbrain ground state. *Dev Cell*. 3:723-733.

Watanabe S, Kondo S, Hayasaka M and Hanaoka K, **2007**. Functional analysis of homeodomain-containing transcription factor Lbx1 in satellite cells of mouse skeletal muscle. *J Cell Sci*. 120:4178-4187.

Wotton KR, Weierud FK, Dietrich S and Lewis KE, **2008**. Comparative genomics of Lbx loci reveals conservation of identical Lbx ohnologs in bony vertebrates. *BMC Evol Biol*. 8:171.

Chapter 4: Identification of a Neural-Specific Enhancer at the *Lbx1* Locus

1) Introduction

a) General mechanisms of Transcriptional Regulation

Transcription factors (TF) elicit their effects by binding DNA at specific sequences (Mitchell and Tjian, 1989) leading to the subsequent modulation of the chromatin to either enhance or inhibit the recruitment of the RNA polymerase (RNAP) complex (Ptashne and Gann, 1997). Several sequence domains have been identified in transcription factors which include DNA binding, transactivation, and occasionally, ligand binding domains (Latchman, 1997). Numerous co-factors are required for chromatin remodelling and subsequent RNAP complex formation; nonetheless, three key factors play a vital role in determining whether transcription of a gene can occur. The first is chromatin state – activity of histone acetyltransferases (HAT) yields open chromatin that is accessible to TFs and activity of histone deacetylases (HDAC) yields closed chromatin, which is inaccessible to most TFs (Anderson *et al.*, 2001; Narlikar *et al.*, 2002; Li *et al.*, 2007; Bártová *et al.*, 2008; Choi and Howe, 2009). The second is the presence or absence of transcriptional repressors, which usually function by occluding RNAP complex formation at the transcriptional start site. Lastly, if no repressors are present, transcriptional activators need to be recruited to the promoter to facilitate RNAP complex formation as gene expression usually does not occur by chance.

Transcription factors encompass the largest protein group known with over 2600 members identified in humans (Babu *et al.*, 2004) having a wide

variety of DNA binding domains. Identifying the genomic regions responsible for recruiting specific TFs serves two major purposes in the field of developmental biology: the first is to reveal the molecular mechanisms underlying gene regulation and the second is to aid the creation of reporter strains. Expression of developmentally important genes (Lobe, 1992) is usually transient which makes it difficult to follow specific cell populations at later stages. With stable reporters such as fluorescent proteins (Chudakov *et al.*, 2005) or β-galactosidase (β-gal), one can trace specific cell lineages well beyond the stages at which the endogenous gene expression is normally deactivated. Secondly, transcriptional regulatory elements can be used to identify the transcription factors that bind there and the exact sequences to which they bind. This permits the elucidation of a gene regulatory network and provides clues as to why specific genes are expressed in the spatiotemporal pattern observed. TF binding sites usually lie within a certain proximity to the gene they regulate and many approaches have been devised accordingly, to examine the genomic sequence surrounding coding sequences with the goal of locating TF binding sites.

b) Factors Influencing Expression of *Ladybird* during Development

Little evidence exists regarding the nature and source of transcriptional regulation at the *ladybird* locus. The embryonic expression pattern of *ladybird* demarcates neural and muscle tissue therefore it is of interest to determine the regulatory mechanisms that create tissue-specific *Lbx1* expression. Based on the

expression patterns of *Lbx1* in vertebrates, regulation of *ladybird* likely involves two components: one that controls mesodermal expression and another that controls neural expression. Direct regulation of *Ladybird* expression has not been demonstrated, although Pax3 (Mennerich *et al.*, 1998), in conjunction with Six1 and Six4 (Grifone *et al.*, 2005), seem to be upstream factors for *Lbx1* expression in migrating limb muscle precursors. Tlx3 negatively regulates *Lbx1* expression during neurotransmitter fate selection in the dorsal spinal cord of mice (Cheng *et al.*, 2005). The differences in expression between rhombomeres, fin muscle and the fine dorsoventral planes within the spinal cord suggest *ladybird* genes are capable of responding to several different embryonic axial patterning mechanisms.

Characterizing the genomic sequences that are responsible for generating such a complex expression pattern would facilitate the identification of transcription factors responsible for *Lbx1* expression. Transgenic zebrafish reporter strains from which one can follow *Lbx1*$^{+}$ neural and muscle progenitor cells in live embryos would also be useful. Both β-gal (Schäfer and Braun, 1999; Brohmann *et al.*, 2000) and EGFP (Gross *et al.*, 2000) reporter mice exist for the *Lbx1* locus; however, neural reporter gene expression appears to be significantly weaker than in muscle. Efficient methods of zebrafish transgenesis have been well documented (Fisher *et al.*, 2006) which exploit the Tol2-transposon system originally discovered in the Medaka fish (Koga *et al.*, 1996). The synteny observed between zebrafish *lbx1a*, *lbx1b* and mammalian *Lbx1*, along with the

similarities in expression pattern, facilitate the identification of evolutionarily conserved enhancer elements that regulate *Lbx1*. Since the mechanisms of early embryonic development are well conserved between vertebrates it is expected that regulatory elements of developmental genes are also conserved.

c) Bioinformatics Tools to Identify Highly Conserved Regulatory Elements

Large-scale sequencing projects such as the mammalian genome project by the Broad Institute (McCarthy, 2005; https://www.broad.harvard.edu/) and the Sanger Institute (http://www.sanger.ac.uk/) have made comparative genomics more straightforward. DNA that was once viewed as "junk" DNA without function (Ohno, 1972) appears to now have a conserved function. Conserved regions serve multiple purposes: (i) transcribed sequences that are not translated into polypeptides (Shabalina *et al.*, 2004), (ii) untranslated regions (UTR) of RNA transcripts or (iii) transcriptional regulatory elements such as enhancers (Woolfe *et al.*, 2005; Woolfe *et al.*, 2007). Comparison between mammalian genomes and vertebrates such as *Xenopus* (Evans, 2008), chicken (Internation Chicken Genome Sequencing and Consortium, 2004) and fish (Barbazuk *et al.*, 2000; Gilligan *et al.*, 2002; Shin *et al.*, 2005), has led to the identification of many highly conserved gene regulatory elements.

With the increasing efficiency and speed of genome sequencing observed, bioinformatics approaches are becoming more appealing over classical methods such as DNase sensitivity (Crawford *et al.*, 2004; Follows *et al.*, 2006; Boyle *et*

al., 2008) to detect transcription factor binding sites (TFBS). More comprehensive analysis using chromatin immunoprecipitation (ChIP) combined with mass sequencing (Visel *et al.*, 2009) or microarrays (Ren *et al.*, 2000; Won *et al.*, 2009) have also been used to identify potential TFBS, however, these methods remain costly. Functional assessment of these conserved regions has demonstrated that many genes important during early embryonic development have very complex regulatory mechanisms. Genes with an apparently simple expression pattern, such as the pan-neural pattern exhibited by *Sox2* in chicken, has been shown to involve multiple enhancer elements (Uchikawa *et al.*, 2004). Furthermore, enhancer activity near the *Sall1* locus in mammals has been demonstrated from seven different conserved regions (Pennacchio *et al.*, 2006).

Several types of bioinformatics software tools exist that are designed to locate highly conserved non-coding genomic sequences between species. It is thought that conserved non-coding DNA contains the specific sequences necessary to recruit transcription factors (TF) which can modulate gene expression. These algorithms work by comparing small blocks, typically 20 nucleotides (NT) of a query sequence at a time, against an entire genome of sequence. The output consists of an overall level of conservation, usually a histogram, exhibited between species which have available genomic sequence. The major drawback of these tools is that identification of very short highly conserved sequences is often missed as the algorithm parameters are optimized for quickness of the alignments. Transcription factor binding sites are typically

less than 10 bp, therefore, if surrounding sequence is not also highly conserved, these regions often remain undetected. Fortunately, many TFBS are accompanied by conserved sequences that create a specific topographical environment of genomic DNA or chromatin (Rubstov *et al.*, 2006; Parker *et al.*, 2009), or additional neighbouring TFBS.

The two most successful, publicly available comparative genomics software programs for visualizing conserved non-coding genomic sequence are the VISTA family of tools (Frazer *et al.*, 2004; Ratnere and Dubchak, 2009; http://www-gsd.lbl.gov/vista/) and the BLAT algorithm available from the UCSC Genome browser (Kent, 2002; Karolchik *et al.*, 2007; http://genome.ucsc.edu/). Both of these tools yield a graphical representation of conservation between species of the user's choice. Additionally, comprehensive, step by step tutorials exist for the UCSC Genome browser (Zweig *et al.*, 2008; Pevsner *et al.*, 2009). While the VISTA family of comparative genomics tools is equally as efficient at detecting highly conserved non-coding sequences with tissue-specific enhancer activity (Miller *et al.*, 2007a), the results presented here were obtained through the use of the UCSC Genome browser. The recent addition of a conservation track of 28 different vertebrate genomes (Miller *et al.*, 2007b) has made identification of evolutionarily conserved intergenic sequence even more straightforward. Several assays to determine potential enhancer activity of these conserved regions exist, mainly through the use of transgenic reporter constructs, such as the Tol2 transposon (Koga *et al.*, 1996), which has now been heavily exploited in zebrafish

(Kawakami *et al.*, 1998; Kawakami *et al.*, 2004b; Fisher *et al.*, 2006; Kawakami, 2007).

d) The Tol2 Transposable Element as a Tool to Test Enhancer Activity

The discovery of a mobile, autonomous, transposable element in the Medaka fish (Koga *et al.*, 1996) was the most significant advancement for zebrafish transgenic technology to date. The Tol2 element belongs to the inverted-terminal-repeat class of transposons, meaning transposition occurs in a cut and paste fashion leaving behind a short, 8 bp, duplication. Interestingly, the Tol2 transposable element was identified in a commercial Medaka strain bearing an albino phenotype, which is due to the disruption of a *tyrosinase* gene by Tol2 (Koga *et al.*, 1996). Furthermore, identification of the minimal cis-sequence required for transposition by Tol2 transposase to a mere 150 bp on one side and 200 bp on the other (Urasaki *et al.*, 2006), has provided a vector backbone where genes and regulatory elements can be inserted into.

A striking feature of the Tol2 transposon is its applicability to several different vertebrate model organisms (reviewed in Kawakami, 2007) such as *Xenopus* (Kawakami *et al.* 2004a; Hamlet *et al.*, 2006), chicken (Sato *et al.*, 2007), zebrafish (Kamakami *et al.*, 2004b) and mouse (Kawakami and Noda, 2004), as well as *Drosophila* (Urasaki *et al.*, 2008). Zebrafish has benefitted significantly from the Tol2 system, as the short generation time and ease of micro-injection yields a high rate of germline transposition and robust stable

transgenic progeny (Kawakami, 2005). Several different Tol2 vectors have been created thus far, each having a specific purpose. For testing enhancer activity, pTol2-GW:*cFos*:*EGFP* is a suitable vector (Fisher *et al.*, 2006a; Fisher *et al.*, 2006b). A Gateway (GW) cassette, where desired sequences can be inserted, is upstream of a minimal *cFos* promoter, which is linked to a fluorescent reporter. Transient transgene expression is observed by co-injecting purified transgenic construct DNA with *in vitro* transcribed mRNA encoding a functional Tol2 transposase into a 1-cell zebrafish (Kawakami, 2004).

2) Results and Discussion

a) Highly Conserved Non-coding Genomic Sequences at *Lbx1* Loci

Contiguous genomic sequence surrounding the *Lbx1* locus is available for many vertebrates, which allows identification of putative regulatory elements based on sequence conservation using the BLAT algorithm (Kent, 2002) available at http://genome.ucsc.edu/ (Figure 4-1, A). In humans, the *LBX1* locus resides on chromosome 10 with 100 kbp and 120 kbp of intervening genomic sequence between *TLX1* and *BTRC*, respectively. Similarly, on murine chromosome 19 there is 80 kbp of intergenic sequence between *Tlx1* and *Lbx1* and 120 kbp between *Brtc* and *Lbx1*. Therefore, it is likely that any regulatory elements of the *Lbx1* locus reside within this ~200 kbp stretch of genomic DNA. The expected position of such regulatory elements is unknown. They could be long range distal enhancers such as at the *Pax6* (Kleinjan *et al.*, 2006), *Sox8* (Guth *et al.*, 2009) and

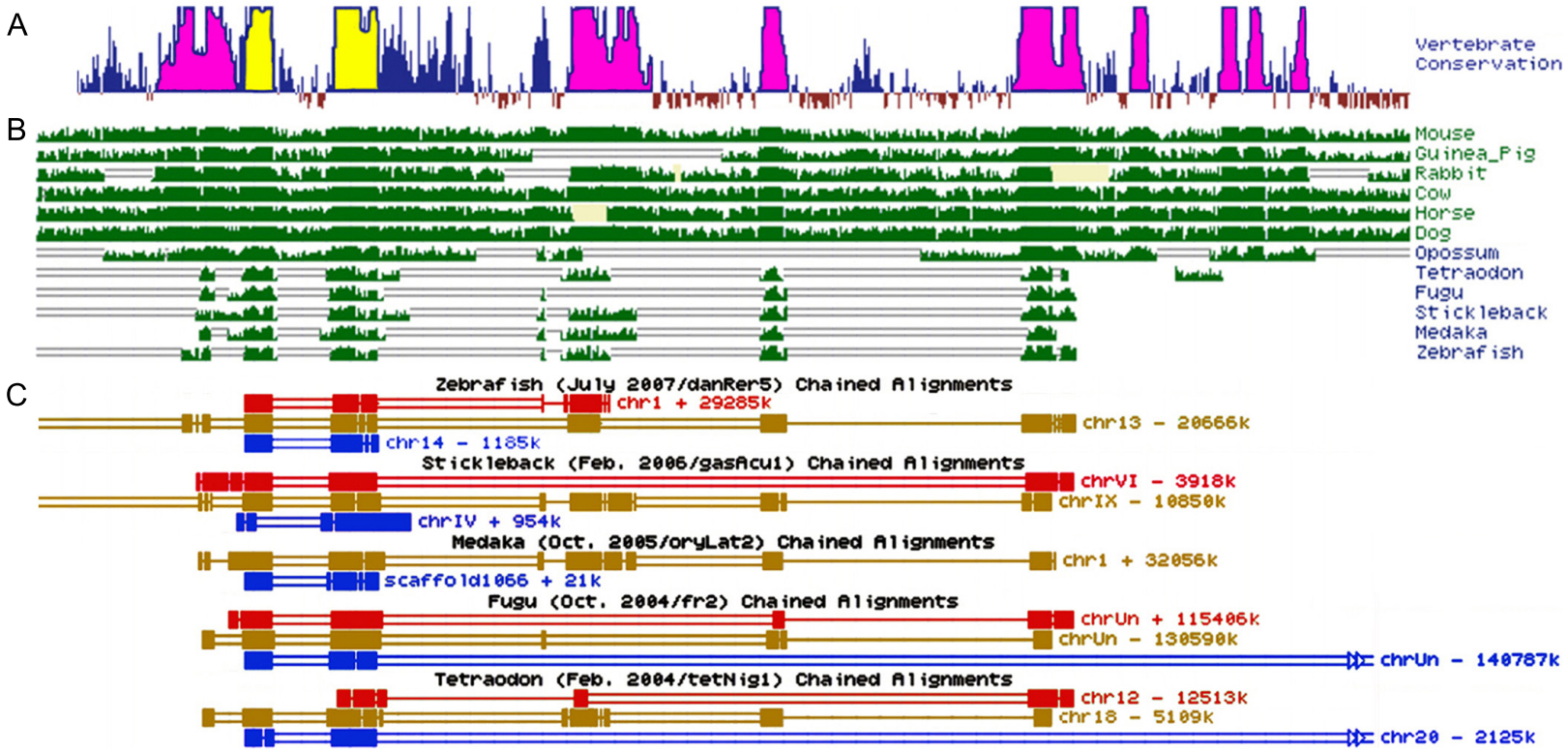

Figure 4-1. **Genomic conservation surrounding the vertebrate *Lbx1* locus**. Shown is the modified output of the BLAT algorithm at the UCSC genome browser (http://genome.ucsc.edu/) from ~15 kb of genomic sequence surrounding the Human *LBX1* locus which reveals highly conserved sequence blocks as a histogram (A). Regions filled in yellow indicate the two exons of *LBX1* and regions filled in magenta are putative regulatory regions due to their high level of conservation. Pairwise conservation histograms (B; green) of human sequence to various vertebrate species suggests these regions have been maintained throughout evolution. Fish species also possess these conserved genomic regions in similar position and distance to *ladybird* coding regions. Genomic sequence chains of fish *ladybird* genes (C) in red, brown and blue represent *lbx1a*, *lbx1b* and *lbx2*, respectively. Conserved regions appear as blocks in the chain.

Gata2 (Khandekar *et al.*, 2004) loci, or proximal enhancers as seen at *NeuroD2* (Lin *et al.*, 2004), *Fgf3* (Powles *et al.*, 2004) and *Msx1* (Miller *et al.*, 2007) loci.

A pairwise sequence comparison between humans and other vertebrates (Figure 4-1, B) surrounding the *Lbx1* locus demonstrates the conservation of several blocks of proximal non-coding sequences. These highly conserved intergenic regions are also present at teleost fish *lbx1a* and *lbx1b* loci at a similar distance from the start of the coding regions (Figure 4-1, C). All teleost fish *lbx2* loci display weak intergenic sequence conservation surrounding the *Lbx1* locus, therefore, attention was focused on *lbx1a* and *lbx1b*. BLAT analysis indicated that *lbx1* loci in zebrafish contain a different set of conserved regions compared to that of other teleost fish. To examine this observation further, 10 kbp of genomic sequence surrounding the *lbx1* loci in zebrafish was compared to that of *Oryzias latipes* (Medaka), *Gasterosteus aculeatus* (Stickleback), *Tetraodon nigroviridis* (*Tetraodon*) and *Takifugu rubripes* (Fugu) in relation to the conservation exhibited to mammals (Figure 4-2, B).

The conserved sequence surrounding both *lbx1a* and *lbx1b* loci in teleost fish, when taken together, closely resembles the pattern of conserved sequence amongst vertebrates (Figure 4-2, A and B). However, there is a distinct difference between zebrafish and other teleost fish examined; both zebrafish *lbx1a* and *lbx1b* loci contain a highly conserved region downstream of the presumptive 3'UTR, whereas only the *lbx1b* locus from other teleost fish shares this region (Figure 4-2, C). Another region further downstream at only the *lbx1b* locus in Zebrafish is

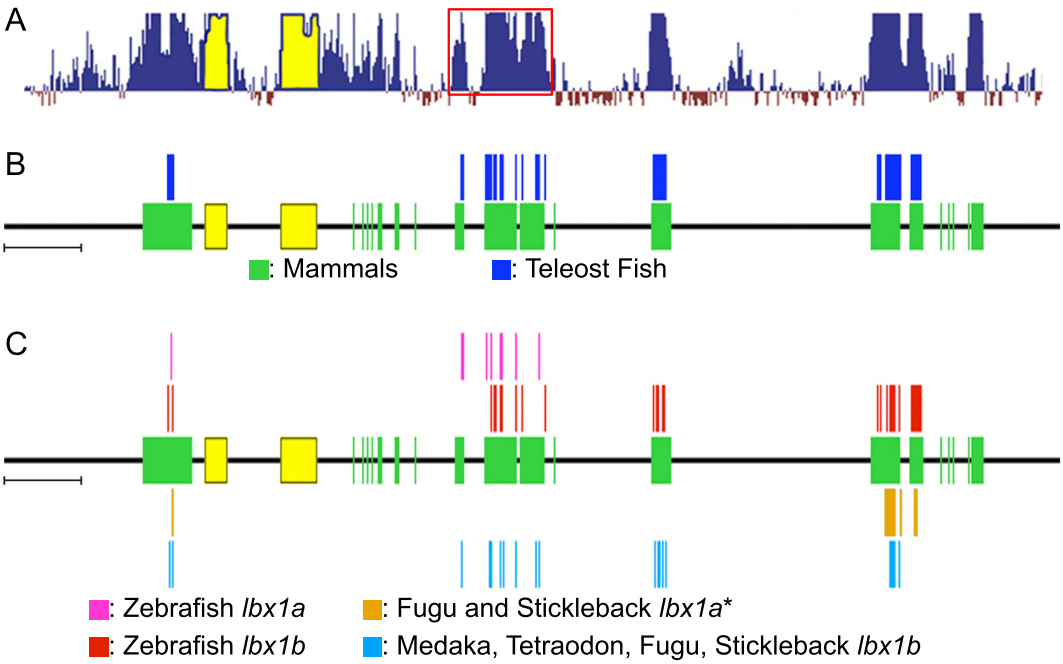

Figure 4-2. **Genomic conservation of the human *LBX1* locus to mammals and fish**. Conservation histogram of non-coding (navy) regions near the human *LBX1* locus to vertebrates (A) or to mammals and teleost fish (B). Yellow regions indicate *LBX1* coding sequences and the region boxed in red denotes the neural enhancer element to be described later. Since fish species have undergone a genome duplication event and retain two *lbx1* genes that are syntenic to mammalian *Lbx1*, putative regulatory elements would also have been duplicated. The conservation to mammalian *Lbx1* observed at zebrafish *lbx1a* and *lbx1b* loci is significantly different to that of other teleost fish (C). Scale bars represent 1 kb of genomic sequence and highly conserved (>90%) non-coding sequence blocks are shown as colored boxes above or below the human *LBX1* locus map with legends shown below. *Genomic information for Medaka and Tetraodon *lbx1a* was not available.

present at both *lbx1a* and *lbx1b* loci from the four other teleost fish mentioned. Since all modern teleost fish are thought to be the descendants of an ancestral fish-specific genome duplication event (Meyer and Van de Peer, 2005; Froschauer *et al.*, 2006), their subsequent divergence into more than 23 500 different species is potentially due to the differential loss of paralogous genes and regulatory elements. As more regulatory elements are characterized using zebrafish as a model, it would be interesting to examine the genomes of Medaka, Stickleback, *Tetraodon* and Fugu to distinguish whether differential maintenance of conserved non-coding sequence occurs throughout the genome, or if this phenomenon is unique to the *ladybird* locus.

Due to their proposed evolutionary relationship (Froschauer *et al.*, 2006), it is expected that Medaka, Stickleback, *Tetraodon* and Fugu would show significantly more conservation surrounding the *lbx1a* and *lbx1b* loci to each other than to zebrafish. Closer examination of non-coding sequence surrounding the *lbx1* loci in teleost fish confirms this prediction but also reveals a striking difference between zebrafish, a member of the *Ostariophysi* superorder, and the *Acanthopterygii* superorder, of which Medaka, Stickleback, Tetraodon and Fugu are members (Figure 4-3). Very little conservation near the *lbx1a* locus exists between zebrafish and other teleost fish *lbx1a* loci. Between Medaka, Stickleback, *Tetraodon* and Fugu, the *lbx1a* locus also shows little conservation, except for one downstream region that appears to be conserved among vertebrate *Lbx1* loci. The fish *lbx1b* locus exhibits quite a different pattern of conservation between

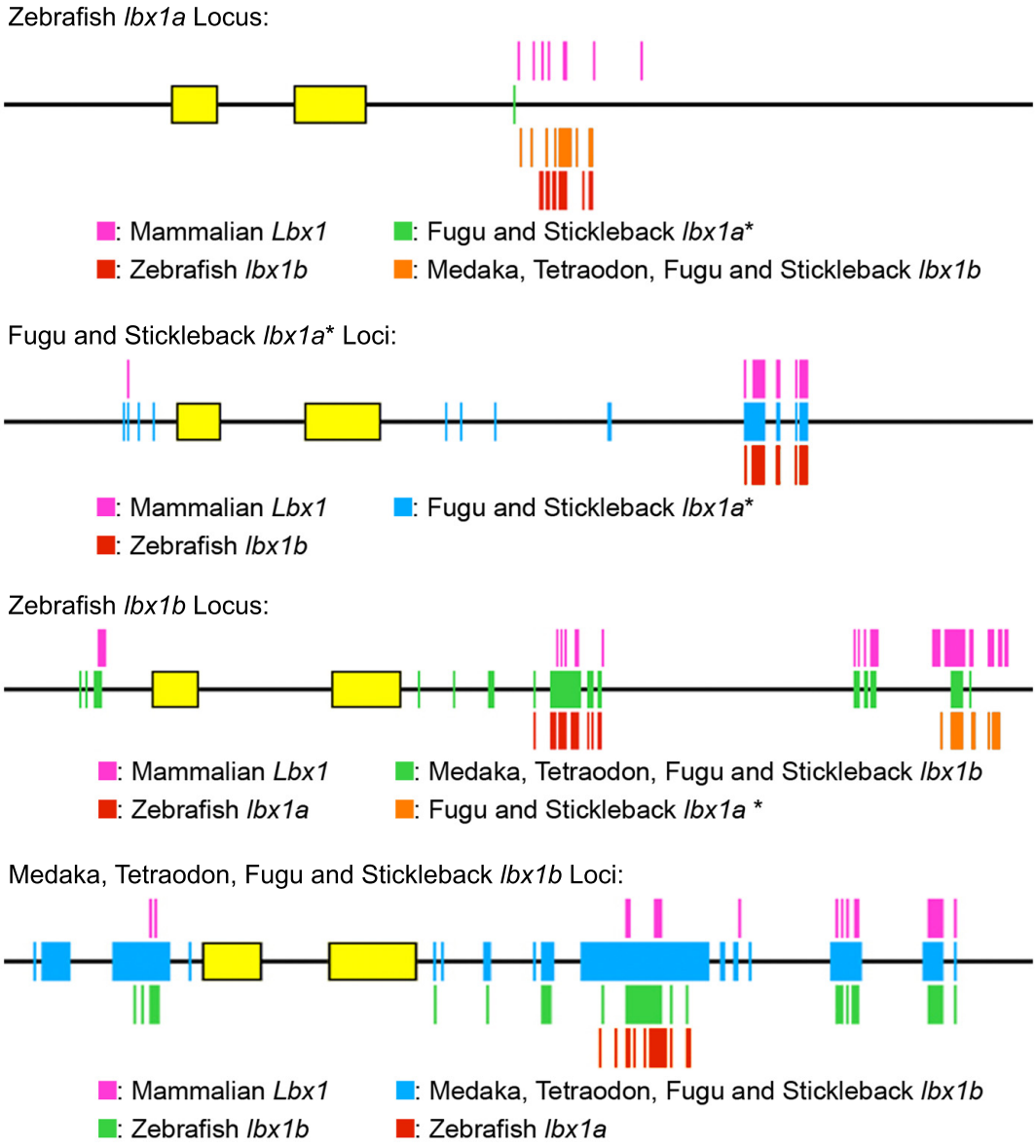

Figure 4-3. **Genomic conservation of *lbx1a* and *lbx1b* loci in teleost fish**. Detailed sequence comparison of all *lbx1* loci in fish species demonstrates a clear difference in the genomic sequences that were maintained after the genome duplication. Highly conserved (>90%) regions among other vertebrates are indeed conserved in fish species, however, the locus at which these corresponding sequences resides at is different. Each locus or set of loci is denoted by a black bar and colored blocks indicate sequence conservation according to the legend shown below. Yellow regions indicate the 2 exons of *ladybird* genes. Zebrafish *lbx1a* and *lbx1b* retain a different set of conserved genomic sequences when compared to the same locus in other fish species. *Genomic information for Medaka and Tetraodon *lbx1a* was not available.

zebrafish and the *Acanthopterygii* members. In Medaka, Stickleback, *Tetraodon* and Fugu, significantly more conservation is seen upstream of the start codon as well as the sequence surrounding a highly conserved region approximately 750 bp downstream of the stop codon. This pattern of high conservation surrounding the *Acanthopterygii lbx1b* locus and seemingly low sequence conservation when compared to the zebrafish is not just unique to the *lbx1* loci. A similar observation was made at the *Hox* clusters of *Acanthopterygii* species (Hoegg *et al.*, 2007), where the intervening sequence is far more conserved amongst that superorder, than when compared to zebrafish.

b) Examining the Regulatory Capacity of Putative Enhancer Elements

To determine if the conserved regions surrounding the zebrafish *lbx1* loci influence its expression pattern, genomic sequences surrounding zebrafish *lbx1a* and *lbx1b* coding regions were amplified by PCR and sub-cloned into the pTol2-GW:*cFos*:*EGFP* vector. Because several of the cloned regions occur downstream of *ladybird* coding sequences, a cloning strategy was devised to test putative enhancers in both 5' to 3' and 3' to 5' directions relative to the minimal *cFos* promoter. Purified constructs were confirmed by sequencing and injected into 1-cell zebrafish embryos along with *in vitro* transcribed *Tol2-transposase* mRNA from the pCSTP vector (Kawakami *et al.*, 2004b). Embryos were observed for transient transgene expression over the first 5 days of development by live fluorescence microscopy. Through multiple rounds of injections, an optimal

volume of 2-3 nL of a solution containing 25 nM of both transgenic construct and *Tol2* mRNA was found to yield high levels of transgene expression with minimal embryonic abnormalities or lethality. The drawback to this method is that an excess amount of plasmid is injected into 1-cell embryos and numerous insertions occur in random chromosomal locations. This results in 'position effects' in which the transgene is regulated by enhancers near the integration site causing ectopic expression of the reporter.

A region approximately 750 bp downstream of the zebrafish *lbx1a* stop codon (boxed in Figure 4-2, A), which is also present at Zebrafish, Medaka, Stickleback, *Tetraodon* and Fugu *lbx1b* loci, showed consistent and robust expression in a pattern similar to endogenous zebrafish *lbx1a* (Figure 4-4, A). The 1067 bp putative enhancer element (Figure 4-5) is able to drive EGFP throughout the dorsal hindbrain and spinal cord. To corroborate this result, the fragment was also able to drive dTomato (Figure 4-4, B) from the pTol2-GW:*cFos*:*dTomato* vector in a similar pattern. Each reporter construct was tested with the *lbx1a* enhancer in both orientations and they all yielded a similar set of results, therefore, all subsequent enhancer analyses were done with cloned regions inserted upstream of the *cFos* promoter in the 5' to 3' direction relative to endogenous coding sequences. These two transgenic zebrafish strains are currently designated Tg(*lbx1a*:*EGFP*) and Tg(*lbx1a*:*dTomato*).

Embryos with high levels of transient transgene expression were raised to adulthood as the F1 generation. To create stable transgenic zebrafish lines

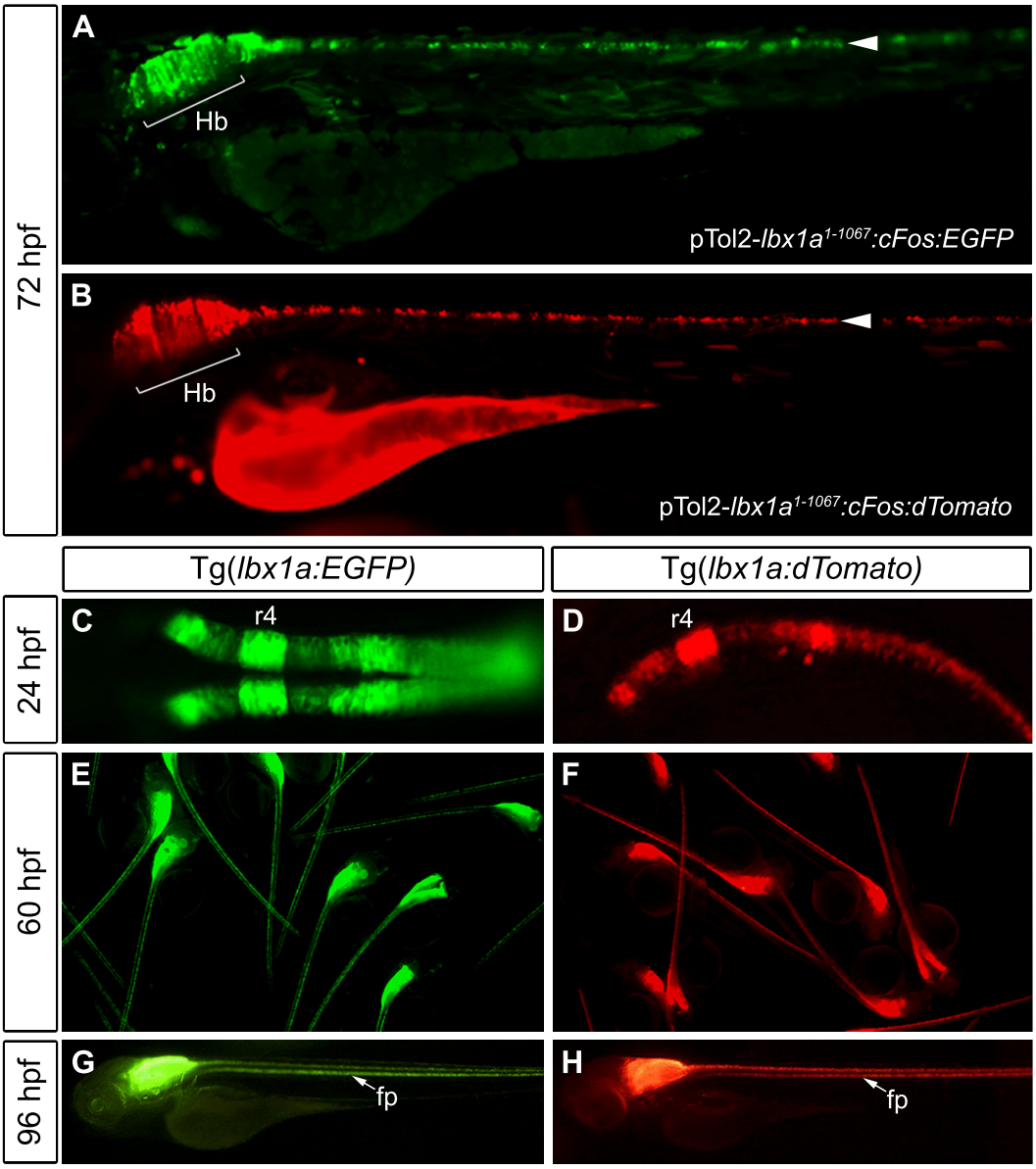

Figure 4-4. **Identification of a neural enhancer element at the zebrafish *lbx1a* locus**. Conserved genomic regions surrounding zebrafish *lbx* loci were cloned into pTol2-GW:*cFos:EGFP* or pTol2-GW:*cFos:dTomato*. Confirmed constructs were injected into one-cell zebrafish embryos with *Tol2*-mRNA and observed for fluorescent reporter expression (A, B). A 1067 bp element downstream of *lbx1a* coding sequences was able to drive reporter expression in the hindbrain (Hb) and dorsal spinal cord (arrowhead), a pattern very similar to endogenous *lbx1a* expression. Embryos were grown to adulthood and stable transgene expression (C-H) exhibited a remarkable resemblance to *lbx1a*, specifically the variation between rhombomeres (r) of the hindbrain (C, D). Floor plate (fp) expression is also driven by the enhancer. Panels A-D, G and H have anterior to the left with lateral (A, B, D, G, H) or dorsal (C) views. E and F show a collection of stable transgenic progeny from a single identified founder adult.

AGAGCTTAACAAACCTGTTCACGACGGCCAAGGGTTTCATACTTTTTTTT

AAAATTACTACTGCCTTGATTTCTTTACTCAAAAAGAAAAAAGGAAGAAA 100

AACATTCTTGCCCCTTTAACGCTATTGTTTTTTGCTCTCTGCAGTAATAA

AGAGCCTTTGGCTGGCTTCACTCTTTCTTTTACTCGCCAATAGCGTGACT 200

TGGCTTTTGACCCAACAACGTGTGTTTTTCCTTTCTGGGCGAAATTTATG

TGAGAAACACAAATCTGGGGCGCTGGGGAGAAAAAGGAGAAGAAAGACAC 300

AAAAAAACGCCAAGTTTTGGTCCAGTTTCTCCAGGGCTATATAATCCTTT

TTGTTCACGGGTGAAGCAGTAATTTCCAGCGTAAAGCATGAAAGCTGAGG 400

ACAAATGAGCCATCCACCCTGATGTGACAAGTGACAATGGCGCATGGCTG

GACTCCTGCGGTTCACCATACTGAAACAAGTGTGTCCAGCCACATAAAAA 500

AGACGCAAACGTTTGGAGGCCATATGGTTTCTGAAATTTTCACACAATTT

CAGACAATTTGATGGATTGCGTTAACCCCTCTTTTAAACTGTTTAACGCA 600

GCGCGAACAAAGGCCATACATTGGAGTCTCTCCTCGAGAAAGCATCTGTA

CCCATGAATGTCCGCTCTTTTCATCTCGTCTTTTTAGCTTGGCGTTTTGA 700

ACAAGTCAATGAATTACAATGAAACGCTCTCTCTTTTTTGTGAGGCACTG

ACTTTCTTCCCCCTTTTGGTGATTTTAAATGATTTGTTAGACTTGTGTGG 800

GGGACTTTTTTTGGTTAGAGAATACGTTCTCATTTTGTTGGAGTTTTTTT

TTCCATGTTCAGGAGGGGGACTCGTTAAGCGTTCAAAGTCCCTCATGTCA 900

ACTTTACTAAAAGGGGCTCTTTTGTTTAATGCCTAATTCCTTCGAATCTT

TACCTTAAGACTGCTTTACTTCAGGAGATTAATAACGGCAAAAACTAAAC 1000

AATGTGCTAAAGAGCGTGTGAGCTTTAAAAAAAGAAGACACACTAAGCAT

TCTGTCAGTAGCAGGTG 1067

Figure 4-5. **Sequence of the zebrafish *lbx1a* neural enhancer**. The full length 1067 bp fragment was cloned with primers shown in blue and sequenced. Sequences in magenta denote >90% conservation to mammals while those underlined in green indicate >90% conservation to the corresponding region at the *lbx1b* locus in other teleost fish as this enhancer fragment is not present at their *lbx1a* loci. The region boxed in red shows the Δ522-570 deletion construct which exhibited little to no transient reoprter gene expression (described later in chapter 4). Base pair positions are indicated on the right.

expressing either *EGFP* or *dTomato* under control of the 1067 bp enhancer element at the *lbx1a* locus, embryos from adults yielding extremely low frequencies of transgenic progeny were selected as founder adults. These embryos were raised to adulthood and again crossed to wild type strain AB zebrafish, and clutches that exhibited a Mendelian inheritance pattern of transgene expression were kept as founder adults. Once homozygous embryos were obtained, strains yielding both consistent and early onset of transgene expression were established for both *EGFP* and *dTomato* transgenic reporters. Each reporter strain did not exhibit any new transgene expression after 3 dpf suggesting the putative enhancer element isolated contains no additional activities.

The main criteria for selecting stable founder embryos was the ability of the transgenic construct to mimic very early expression of *lbx1a*, which implies the chromosomal location of the transgenic construct does not interfere with its expression and that transcription factors are promptly recruited to the transgene. One such line for each transgene was established where the dynamic expression of endogenous *lbx1a* within rhombomeres of the hindbrain was completely recapitulated by the transgene (Figure 4-4, C and D). This dynamic pattern becomes uniform at later embryonic stages (Figure 4-4, E and F), much like endogenous gene expression. The stability of the fluorescent protein provides a prolonged pattern up to 10 dpf, which is well beyond the limitations of mRNA *in situ* hybridization for endogenous transcripts (Figure 4-4, G and H).

An interesting characteristic of the 1067 bp *lbx1a* enhancer is that it only drives transgene expression in neural tissues that express *lbx1a* and not in pectoral fin muscle. This suggests that two independent enhancers may be responsible for neural- and mesodermal-specific expression separately although the latter has not yet been identified. The downstream nature and close proximity to coding sequences of this neural enhancer in all species examined indicates its genomic position may be critical to its function, and provides a rationale as to why mouse transgenic lines have yielded very poor responses in neural tissues. The EGFP (Gross *et al.*, 2000) and first β-gal (Schäfer and Braun, 1999) transgenic lines created displace this enhancer region approximately 4.5 kbp further downstream from the transcriptional start site. The second β-gal reporter strain for the mouse *Lbx1* locus (Brohmann *et al.*, 2000) results in a partial deletion of this region. An unexpected ability of the *lbx1a* neural enhancer yielded transgene expression in the floor plate, which is a region of the neural tube that never expresses endogenous *lbx1a*. Transgene expression in the floor plate is likely not a consequence of neighbouring enhancers nor an artefact of the transgenic construct sequences, such as the *cFos* promoter, as several different putative enhancers that were cloned did not produce this expression pattern. Additionally, both the *EGFP* and *dTomato* transgenes were expressed in the floor plate in nearly all first generation stable transgenic progeny, which implies sequences within the 1067 bp *lbx1a* enhancer element can be positively regulated in the floor plate and that repressor elements also exist at the *lbx1a* locus. Intriguingly, a similar observation

of floor plate expression was made for transgenic homozygous *Lbx1*$^{GFP/GFP}$ transgenic mice (Gross *et al.*, 2002).

c) Characterizing the Enhancer Element Identified at the *Lbx1* Locus

At early embryonic stages, reporter expression appeared to be more robust in Tg(*lbx1a*:*dTomato*) embryos in comparison to Tg(*lbx1a*:*EGFP*) embryos, and were therefore chosen for subsequent analysis. To get a better understanding of the precise timing of transgene expression, mRNA *in situ* hybridization for *dTomato* transcripts was performed at a variety of embryonic stages and compared to endogenous *lbx1a*. This procedure is far more sensitive than live fluorescence imaging as dTomato requires time to accumulate and mature into its dimeric form in order to be visible. Detection of *dTomato* transcripts occurs as early as 11 hpf (4 somites) and by 14 hpf (10 somites) the pattern resembles that of endogenous *lbx1a* transcripts (Figure 4-6, A and C). From a lateral view, rhombomeres 4 and 7 have the highest levels of *dTomato* expression, followed by low levels of expression in r2, r6 and r8 and barely detectable levels in r1, r3 and r5. By 36 hpf, it is remarkable how similar transgene expression is to zebrafish *lbx1a* expression in the neural tube (Figure 4-6, B and D). The cerebellum, hindbrain and spinal cord all exhibit nearly identical expression patterns between *lbx1a* and *dTomato*. At 60 hpf, both dorsal and lateral views of the hindbrain (Figure 4-6, E – H) clearly show the incredible similarity in expression patterns between the two genes.

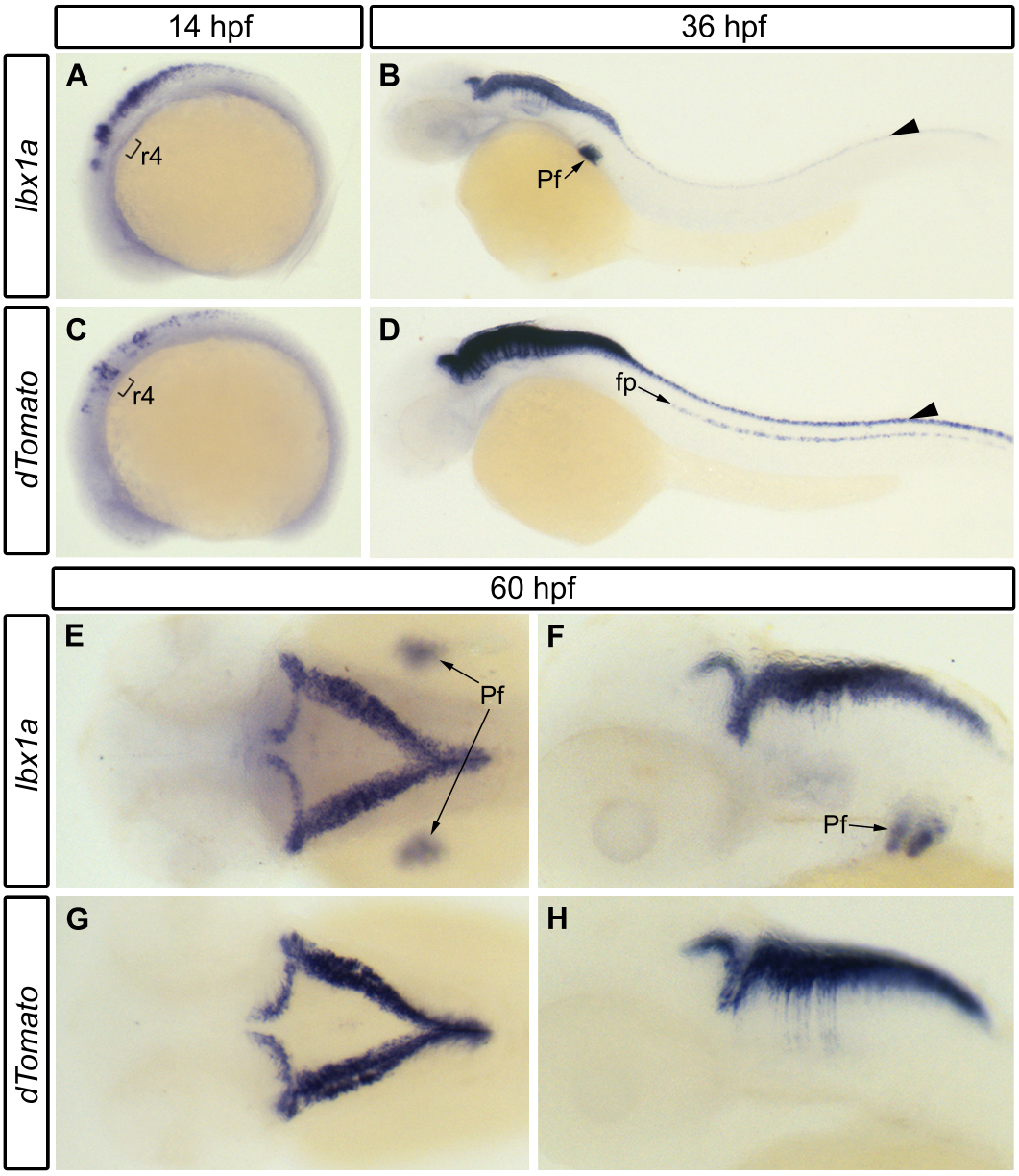

Figure 4-6. **Transgene expression driven by the 1067 bp *lbx1a* neural enhancer**. Stable Tg(*lbx1a:dTomato*) embryos were analyzed for transgene (C, D, G, H) expression by mRNA *in situ* hybridization and compared to endogenous *lbx1a* (A, B, E, F). At early somitogenesis stages (14 hpf; A, C), *dTomato* expression is slightly delayed with respect to endogenous *lbx1a*, however, still follows the specific differences between each rhombomere (r) within the hindbrain. By 36 hpf (B, D), transgene expression is more robust within the hindbrain and spinal cord (arrowhead) although it is clearly not expressed in pectoral fin muscle (Pf). The enhancer is also able to drive transgene expression in the floor plate (fp). At 60 hpf (E-H) it is evident that the cloned enhancer region drives transgene expression in a pattern nearly identical to endogenous *lbx1a*. All panels have anterior to the left with lateral (A-D, F, H) or dorsal views (E, G).

With a positive result for enhancer activity of a highly conserved region downstream of *Lbx1* coding sequences, it was of interest to examine this region at both zebrafish *lbx1b* and human *LBX1* loci. The corresponding genomic region downstream of zebrafish *lbx1b* was cloned into pTol2-GW:*cFos*:*dTomato* and assayed for transient transgene expression. The enhancer fragment cloned is 742 bp (Figure 4-7), a slightly smaller size to avoid repetitive elements surrounding this region at the *lbx1b* locus. While transgene expression driven by this region was significantly less intense than observed for the *lbx1a* enhancer, fluorescent protein was visible in much the same pattern – the dorsal hindbrain and spinal cord. Embryos were raised to adulthood and the methodology applied previously was used to generate a stable founder strain designated Tg(*lbx1b*:*dTomato*), with a putative single insertion. Detection of the fluorescent reporter by live microscopy was difficult, presumably due to low expression levels, however, the increased sensitivity of mRNA *in situ* hybridization allowed *dTomato* transcripts to be easily detected.

Tg(*lbx1b*:*dTomato*) embryos display a distinct expression pattern which is similar to endogenous *lbx1b*. However, this transgene also display characteristics of the dynamic pattern generated by the *lbx1a* neural enhancer in the hindbrain. Like the enhancer region cloned from the *lbx1a* locus, the *lbx1b* enhancer drives transgene expression only in the neural tube and not in muscle cells. Transgene expression in the hindbrain driven by the *lbx1b* enhancer is unexpectedly similar to the dynamic pattern of *lbx1a*, as endogenous *lbx1b* never exhibits any variation

TGTAGCCCACTTTTATTCCCTGCACCAACACTCTTTAAACATGCATCGCG

TGTGTGTTTTGGATTGGACATTTTACGCTTTTAATTTTCAGAATGGCATA 100

ACATGTTTTTGTGTTTCAATATCCACAATGGATGCGTGTGAGAAACAGAA

ATCTGGATTGATGAATGCGCGCATAAAGGCTGGGCAAAAACGCAGGGTTT 200

TTGGCCCGTTTCTCTGCGGTTGAAAGACGCTATTTCCCTGCGAAGCGAAG

GGAGGCAGTCATTTCTAGCGTAAGGCGTGAAAACTGGGGACAAATGAGCA 300

GCCTCCCCTGATGTGACAAGCTACAATGAAGCATGCTCTGCCTCCCGCGG

CTCTCAATACGGAGACAAGCGTGTCCCACCGAACAAAAAGGACACAAACG 400

TTTGGAGGCCATATGGTTTTTGAATTTTCCTCACAATTTTCAGACAATTT

GATGGATCGAGCTGGCCCTCTTTTTAAAGTGTTTCGCTCGGTTTTCACAA 500

AGGCAATAATGCGCACAGGGGCGAGGGCATCTGGGCCTATTAATGTCCCC

TCTATTCTCTCTTTTCACTCGGCTGTTTTAAACAAGTCTATGAATTACAA 600

TGATACTGACTCCTCTTGTTAAGAGGTTGTCTTCTGCCTTTATTTGTTTG

AGTGTTTGGCTTAAAATAATAAAGTTAACGTTTTGGCTACAAGTAAAAAG 700

CTGAGCATATGGCTGGAGTCACACTATGGATGGAATTAGAGC

Figure 4-7. **Sequence of the zebrafish *lbx1b* neural enhancer**. The full length 742 bp fragment was cloned with primers shown in blue and sequenced. Sequences highlighted in magenta indicate >90% conservation to mammals while those underlined in green indicate >90% conservation to the *lbx1b* loci in other teleost fish. Base pair positions are shown on the right.

between rhombomeres. This pattern appears later in development just prior to 24 hpf (Figure 4-8, A and B), which means the chromosomal insertion site may inhibit transgene expression or the enhancer itself is very weak. Rhombomeres 4 and 7 exhibit the highest levels observed, r8 exhibits intermediate levels of expression and the remaining hindbrain segments express very low levels of *dTomato*. The *lbx1b* enhancer also restricts transgene expression within a narrow mediolateral plane, sometimes to only a few cells, in comparison to *lbx1a* enhancer-driven expression, which occurs throughout the mediolateral domain of the hindbrain. With only a few cells expressing the transgene in some rhombomeres, it makes it difficult to discern intermediate and low levels of expression. From all the embryonic stages examined of both transgenic reporter strains, the cloned *lbx1b* neural enhancer regulates gene expression in a similar mediolateral and dorsoventral patterns to endogenous *lbx1b* but also with the dynamic hindbrain pattern exhibited by *lbx1a*.

Another unexpected pattern of expression driven by the *lbx1b* neural enhancer is the persistence of the variation between rhombomeres. The pattern follows endogenous *lbx1a*, however, after 24 hpf, *lbx1a* expression becomes uniform throughout the hindbrain whereas transgene expression driven by the *lbx1b* enhancer remains variable to 72 hpf (Figure 4-8, C – H) and later. Rhombomeres 4, 7 and 8 are clearly marked with increased expression levels and several cells within the cerebellum show *dTomato* transgene expression. It is evident that the genomic region downstream of the zebrafish *lbx1b* locus is able to

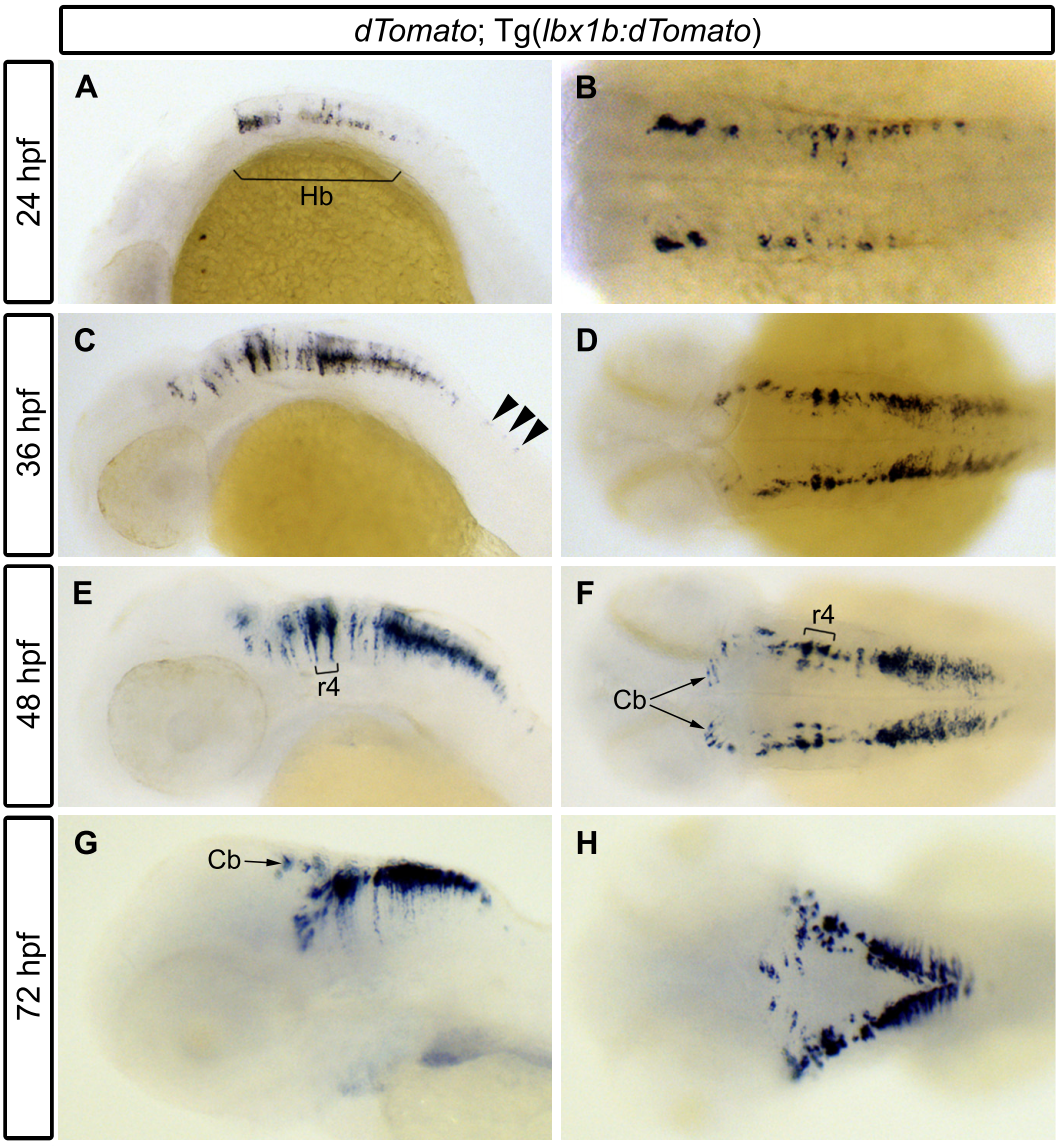

Figure 4-8. **Transgene expression driven by the 742 bp *lbx1b* neural enhancer**. Transgenic Tg(*lbx1b*:*dTomato*) embryos were analyzed by mRNA *in situ* hybridization for expression of the *dTomato* transgene. Transcripts are detected in the hindbrain (Hb) at 24 hpf (A, B) with remarkable similarity to the early pattern generated by the *lbx1a* enhancer. Rhombomere 4 (r4) and more posterior hindbrain segments have high expression levels while the rest of the hindbrain, including the cerebellum (Cb) and dorsal spinal cord (arrowheads), have low expression levels. A major difference is that this pattern from the *lbx1b* enhancer persists at 36 hpf (C, D), 48 hpf (E, F) and 72 hpf (G, H) whereas *lbx1a*:*dTomato* expression appears more uniform at stages later than 24 hpf (see Figure 4-5). Panels show anterior to the left with lateral (A, C, E, G) or dorsal (B, D, F, H) views.

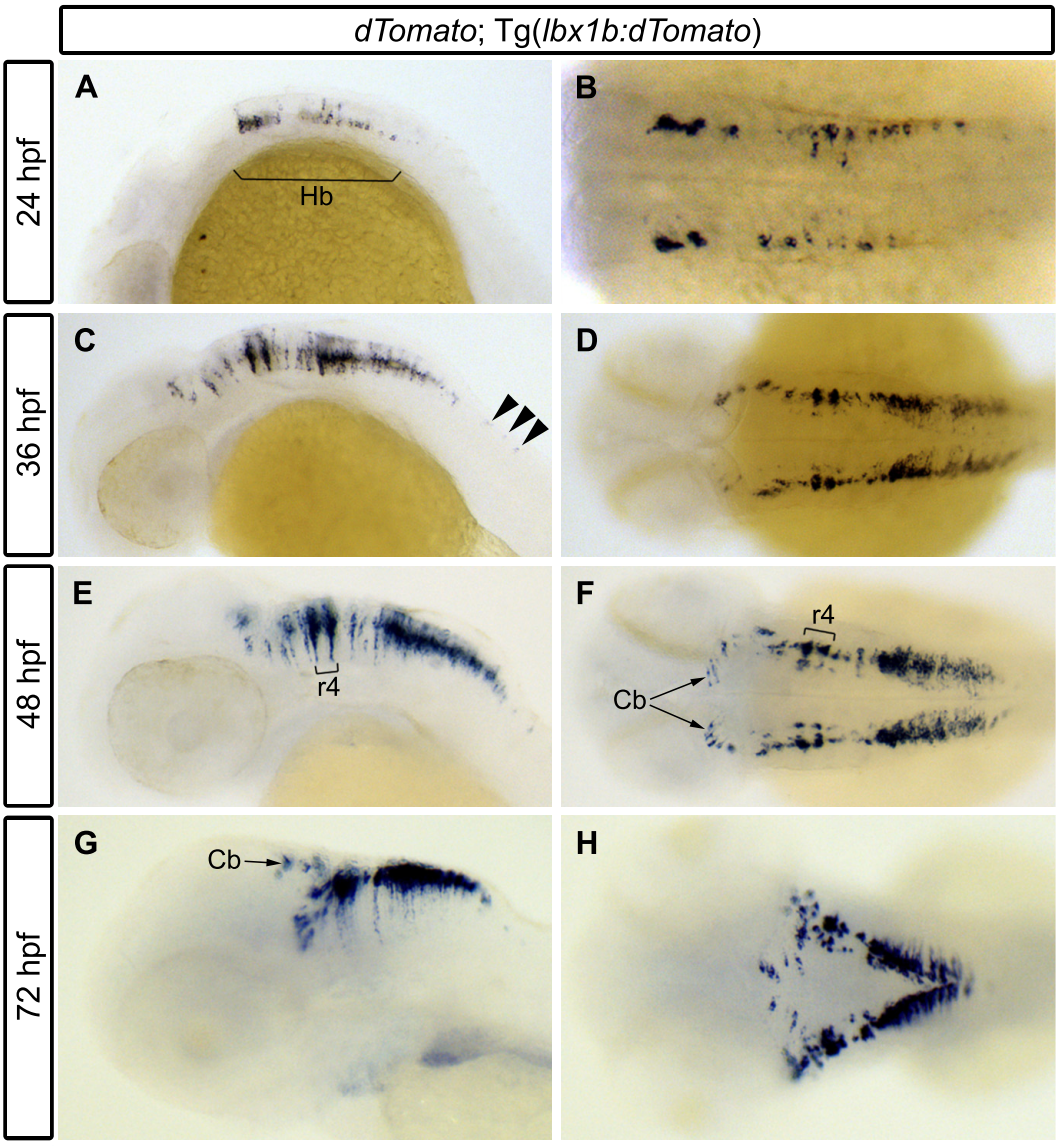

Figure 4-8. **Transgene expression driven by the 742 bp *lbx1b* neural enhancer**. Transgenic Tg(*lbx1b*:*dTomato*) embryos were analyzed by mRNA *in situ* hybridization for expression of the *dTomato* transgene. Transcripts are detected in the hindbrain (Hb) at 24 hpf (A, B) with remarkable similarity to the early pattern generated by the *lbx1a* enhancer. Rhombomere 4 (r4) and more posterior hindbrain segments have high expression levels while the rest of the hindbrain, including the cerebellum (Cb) and dorsal spinal cord (arrowheads), have low expression levels. A major difference is that this pattern from the *lbx1b* enhancer persists at 36 hpf (C, D), 48 hpf (E, F) and 72 hpf (G, H) whereas *lbx1a*:*dTomato* expression appears more uniform at stages later than 24 hpf (see Figure 4-5). Panels show anterior to the left with lateral (A, C, E, G) or dorsal (B, D, F, H) views.

drive neural tube expression but the pattern observed takes on characteristics of both endogenous *lbx1a* and *lbx1b* expression patterns. The variability in expression between rhombomeres, which is a characteristic of *lbx1a*, is generated by the *lbx1b* enhancer and persists well past the stage at which *lbx1a* hindbrain expression becomes uniform. Additionally, expression along the mediolateral axis is restricted for endogenous *lbx1b* and this pattern is evident from the *lbx1b* enhancer. These results suggest that the highly conserved genomic sequence downstream of the *lbx1* coding region responds to positional cues in the embryo that mediate anteroposterior and mediolateral axis development in the hindbrain as well as dorsoventral axis specification of the neural tube. It is also likely that the other conserved regions surrounding the *lbx1b* locus contribute to modulating this neural enhancer driven expression pattern in order to attain the pattern exhibited by endogenous *lbx1b*.

The majority of sequence within the two neural enhancers identified at zebrafish *lbx1* loci is also conserved in other vertebrates. To determine if zebrafish proteins are able to regulate expression via this enhancer across species, the corresponding region of the human *LBX1* locus was cloned and assayed for transient transgene expression in zebrafish embryos. Several highly repetitive sequences are interspersed within this region at the human *LBX1* locus which are not present at zebrafish *lbx1a* and *lbx1b* loci, making PCR amplification challenging. Only a 489 bp fragment encompassing the central region of the enhancer was amplified from the human *LBX1* locus and tested for regulatory

capacity. Interestingly, this region was able to drive expression in the zebrafish hindbrain, albeit at very low levels, but not the spinal cord (Figure 4-9). *In situ* hybridization for *dTomato* transcripts in Tg(*LBX1*:*dTomato*) stable transgenic embryos yielded an unexpectedly high number of different patterns. This is likely a result of other enhancer elements in the vicinity of the construct insertion site, although, a distinct consistency of hindbrain expression does appear in the stable transgenic embryos recovered.

A highly conserved region downstream of the vertebrate *Ladybird* coding sequence that is able to drive transgene expression in a distinct pattern within neural tissues has been identified. The pattern generated by the cloned enhancer at the zebrafish *lbx1a* locus is nearly identical to endogenous *lbx1a* as confirmed by two-color mRNA *in situ* hybridization (Figure 4-10, A and B). The *dTomato* transgene also overlaps *lbx1b* expression in the hindbrain (Figure 4-10, C – F). Although when compared to zebrafish *lbx2* (Figure 4-10, G and H), the *lbx1a* neural enhancer displays clear differences in the hindbrain. The corresponding enhancer cloned from the zebrafish *lbx1b* locus generates a pattern reminiscent of both *lbx1a* and *lbx1b* and additional regulatory elements outside of the cloned region likely exist at the zebrafish *lbx1b* locus.

The pTol2-GW:*cFos* vectors are generally used to identify positive regulatory elements and therefore, putative inhibitory elements at the *lbx1b* locus remain unidentified. Indeed there are highly conserved regions at the *lbx1b* locus that did not show consistent transgene expression when placed in a transgenic

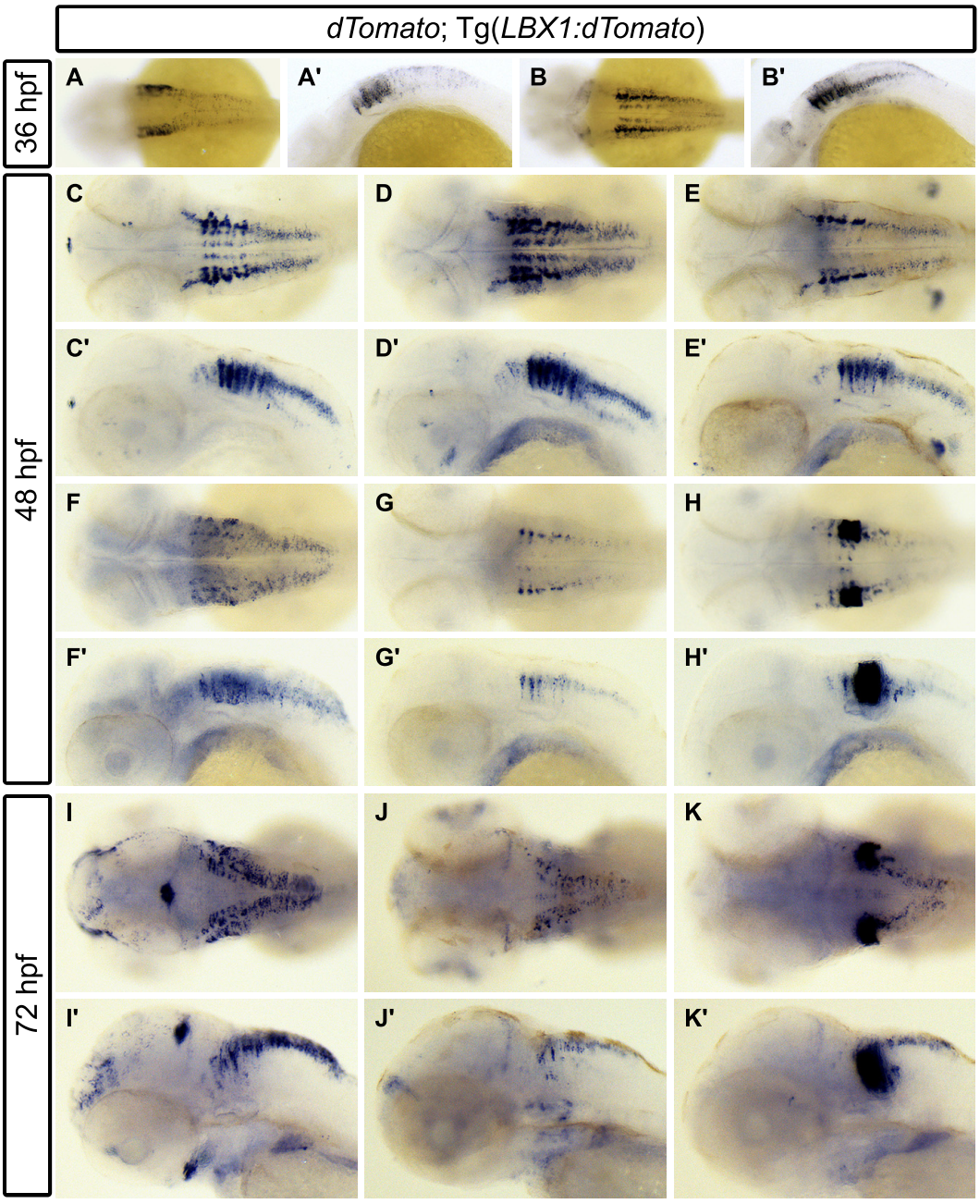

Figure 4-9. **Transgene expression driven by the 489 bp human LBX1 enhancer**. A wide variety of transgene expression patterns are observed in progeny from transgenic Tg(*LBX1*:*dTomato*) adult zebrafish. However, these stable transgenic embryos exhibit detectable levels of *dTomato* transcripts within the hindbrain prior to 36 hpf (A, B) in a pattern similar to those driven by the zebrafish *lbx1a* or *lbx1b* enhancers. Transgene expression is maintained throughout 48 hpf (C-H) and 72 hpf (I-K) in the dorsal hindbrain. Embryos are shown with dorsal (A-K) and corresponding lateral (A'-K') views of the same embryo.

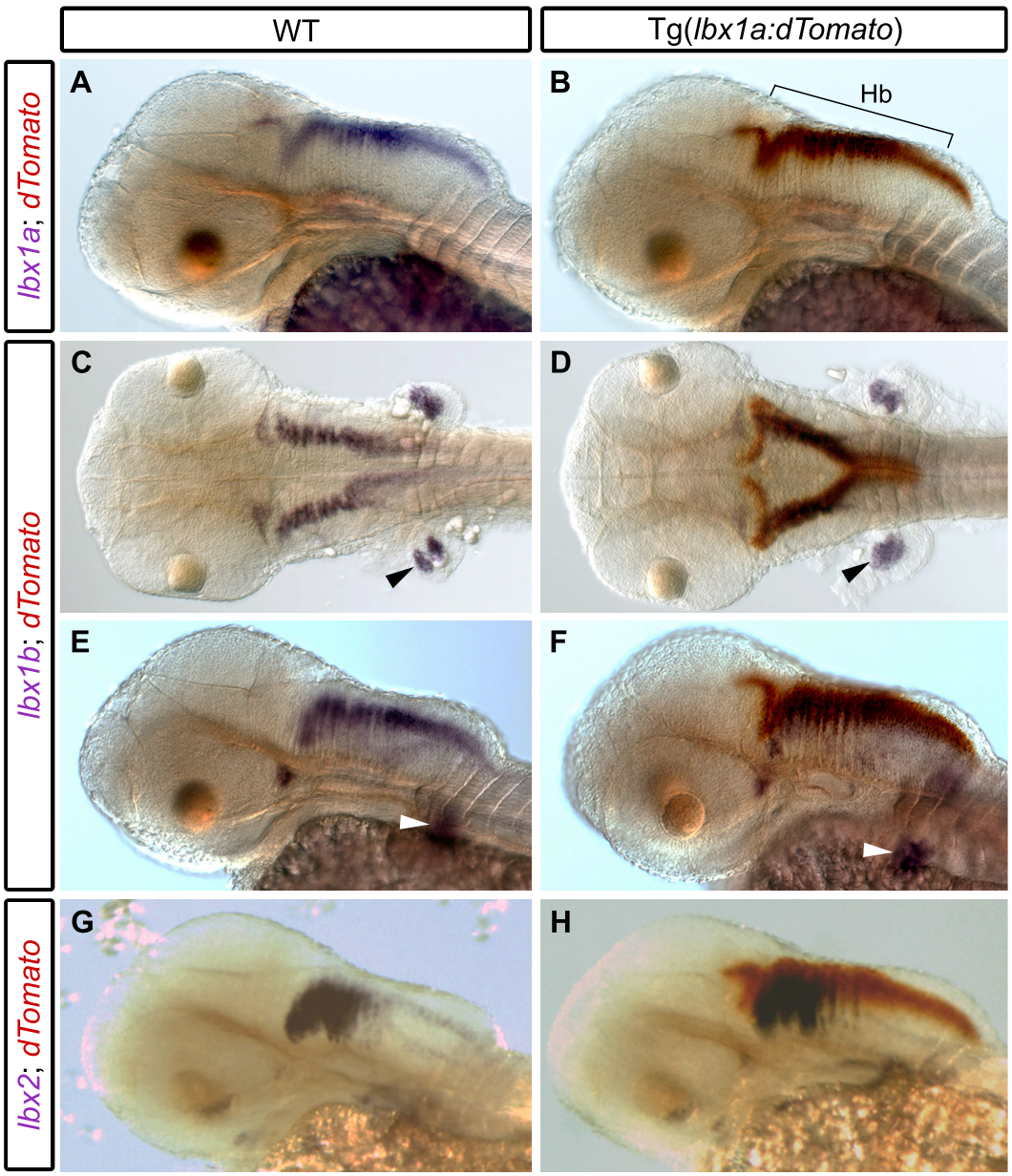

Figure 4-10. **Overlapping expression between endogenous zebrafish *lbx* genes and the *dTomato* reporter in Tg(*lbx1a:dTomato*) transgenic zebrafish**. Once stable Tg(*lbx1a*:*dTomato*) transgenic progeny were isolated and maintained, two-color *in situ* hybridization was performed at 56 hpf to validate reporter expression in comparison to endogenous *lbx1a*, and to compare its expression to other zebrafish *lbx* genes. DIG-labeled (purple) probes to *lbx1a* (A, B), *lbx1b* (C-F) and *lbx2* (G, H) transcripts were detected in combination with a fluorescein-labeled (red) *dTomato* probe (B, D, F, H). It is clear in the hindbrain (Hb) that *dTomato* transgene expression overlaps endogenous *lbx1a* and *lbx1b*, however, the cloned enhancer lacks the ability to drive pectoral fin expression (arrowheads). Panels A-F are merged DIC Z-stack images and G, H are single plane brightfield images, with lateral (A, B, E-H) or dorsal (C, D) views.

construct alone. It is possible that in combination with positive regulatory elements such as the neural enhancer identified, they may recapitulate the expression pattern observed by endogenous *lbx1b*. It also shows that neural expression of *lbx1a* and *lbx1b* can be separated from mesodermal expression. Work in *Drosophila* suggests that the dual functionality of Ladybird in muscle and neurons arose concomitantly (Jagla *et al.*, 1998; De Graeve *et al.*, 2004), therefore, it would be interesting to study *ladybird* expression in other invertebrates to determine if function was co-opted from one tissue to the other.

To further examine the dorsoventral domain of expression driven by the *lbx1a* neural enhancer, transgenic Tg(*lbx1a*:*dTomato*) embryos were crossed to Tg(*isl1*:*GFP*) fish (Higashijima *et al.*, 2000), a strain expressing GFP in facial motor neurons of the hindbrain and the ventral motor neurons of the spinal cord. Cryo-sections of 84 hpf homozygous transgenic embryos were prepared as this stage when near peak levels of both fluorescent reporters is observed (Figure 4-11). Sections reveal that between rhombomeres 1 and 7, axons of the dorsally-located dTomato^{+} neurons project ventrally and occasionally cross the midline in the ventral hindbrain. In the caudal hindbrain (r8) and the rostral spinal cord, axonal projections extend towards the lateral edge of the neural tube. In the spinal cord (Figure 4-12), axons are projected towards the ventral region of the spinal cord along the lateral edges. Transgene expression in the floor plate occurs in two adjacent cells just beneath the spinal cord motor neurons marked by GFP, indicating they are lateral floor plate cells and not medial floor plate.

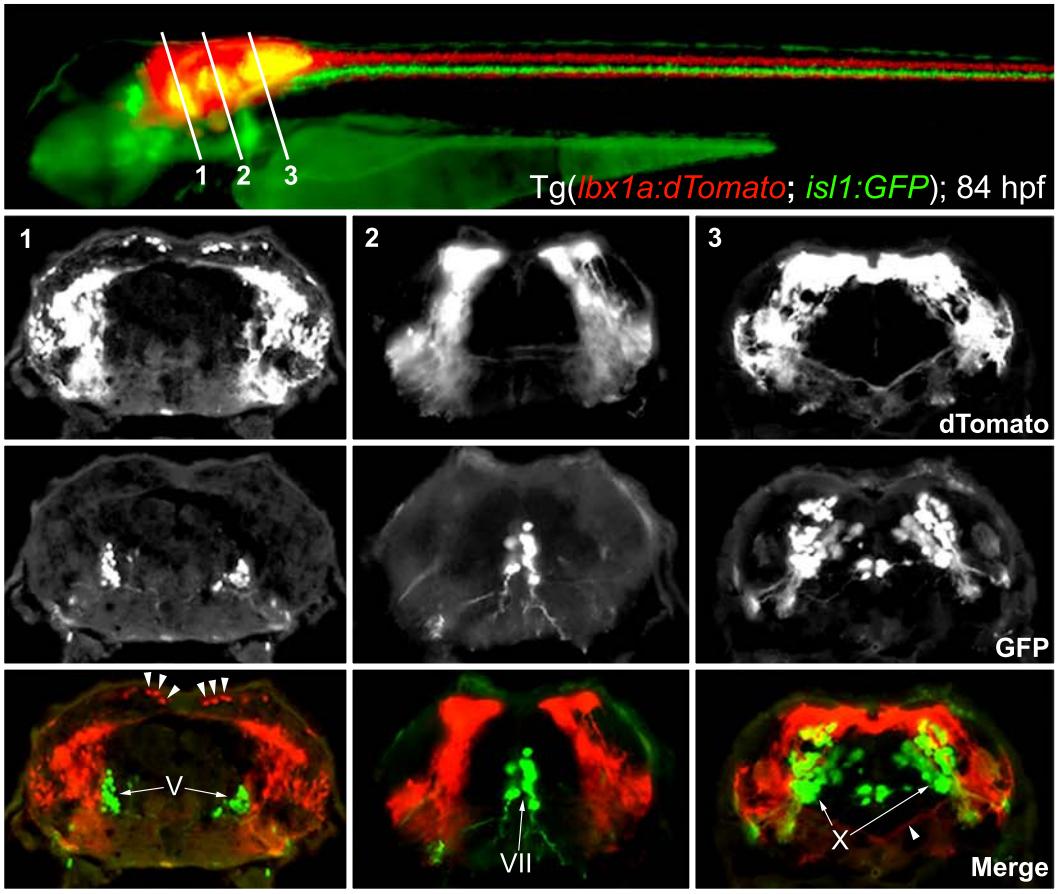

Figure 4-11. **The *lbx1a* enhancer drives transgene expression in the dorsal hindbrain**. Tg(*lbx1a:dTomato*) adult fish were crossed with Tg(*isl1:GFP*) adult fish to generate embryos expressing both trasngenes. GFP expression in the hindbrain marks trigeminal (V in r2 and r3), facial (VII in r4-r6) and vagal (X in r8) motor neurons. Section layout is anterior (left) to posterior (right) and approximate positions are in r2 (1), r4 (2), and rostral r8 (3). The *lbx1a* enhancer drives dTomato in the dorsal hindbrain with extensive ventral axonal projections in the hindbrain. Fluorescent dTomato protein is also seen in the cerebellum (arrowheads in 1) and commisural axons (arrowhead in 3).

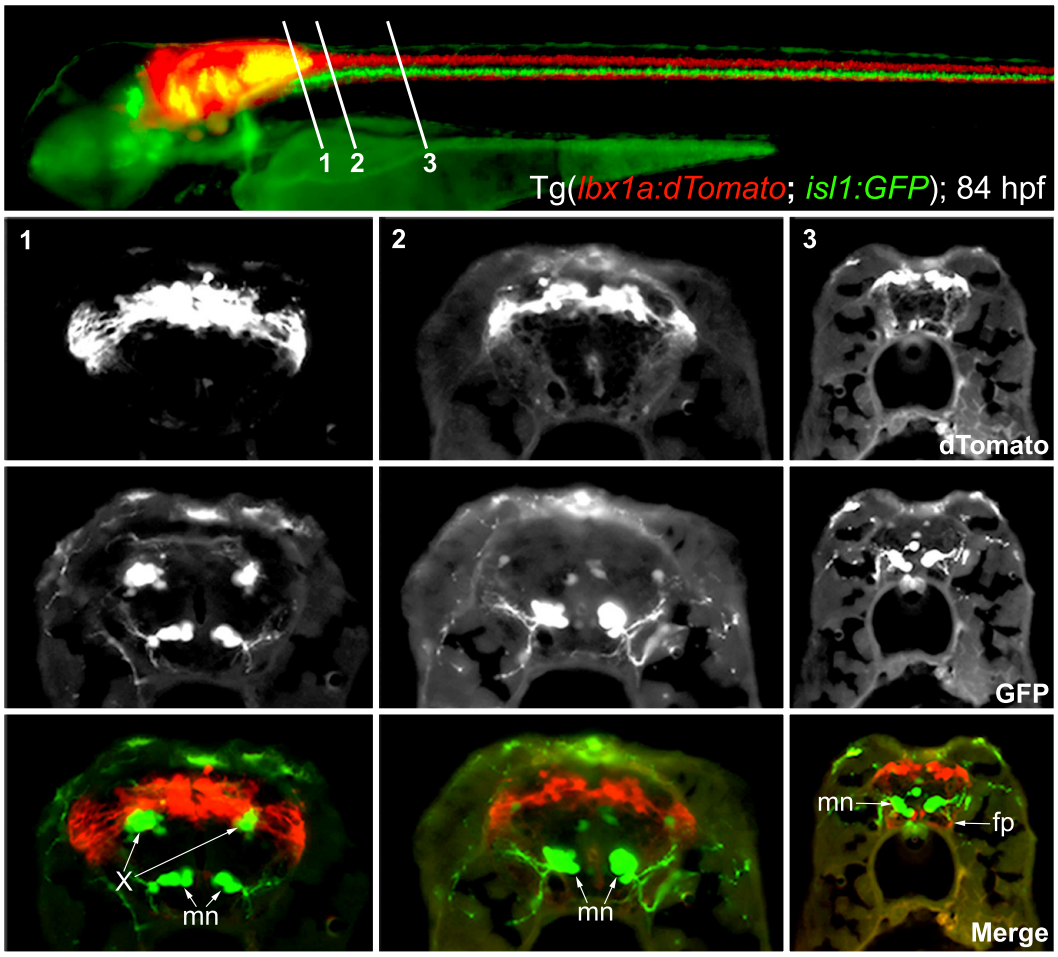

Figure 4-12. **The *lbx1a* enhancer drives transgene expression in the dorsal spinal cord**. Tg(*lbx1a:dTomato*) adult fish were crossed with Tg(*isl1:GFP*) adult fish to produce embryos expressing both trasngenes. GFP expression marks vagal (X in r8) motor neurons and spinal cord motor neurons (mn). Sections are shown with anterior (left) to posterior (right) and approximate positions are in caudal r8 (1) rostral spinal cord (2) and spinal cord (3). The *lbx1a* enhancer regulates dTomato expression in the dorsal neural tube with extensive ventrolateral projections in the caudal hindbrain (1), rostral spinal cord (2) and ventral projections in the spinal cord (3). Fluorescent dTomato protein is also seen in the medial floor plate (fp) as 2 adjacent cells directly underneath the GFP positive spinal cord motor neurons.

The mRNA *in situ* expression data and analysis of the *lbx1a* enhancer transgenic strain indicates that *lbx1a* may be expressed in the dorsal-most region of the zebrafish spinal cord. Another class of neurons in the dorsal spinal cord are the Rohon Beard (Rb) sensory neurons (Rossi *et al.*, 2009), which detect mechanosensory stimuli from the skin. Immunofluorescence with monoclonal Zn-12 antibody (Metcalfe *et al.*, 1990) reveals the dorsoventral plane that dTomato^{+} neurons occupy is just beneath the Rb sensory neurons (Figure 4-13, A). At later stages, Zrf-1 antibody (Marcus and Easter, 1995) which marks the entire zebrafish spinal cord, reveals these neurons have remained at that position, just beneath the dorsal edge of the spinal cord (Figure 4-13, B), whereas by stage E12.5 in mice, a population of Lbx1^{+} neurons has migrated to the very top of the dorsal horn in the spinal cord (Muller *et al.*, 2002; Gross *et al.*, 2002).

d) Deletion Analysis of the Zebrafish *Lbx1a* Neural Enhancer

With such an array of conserved regions present in the identified zebrafish *lbx1a* neural enhancer, the use of TFBS prediction software, such as PROMO (Messeguer *et al.*, 2002) or TRANSFAC (Matys *et al.*, 2003), is problematic as hundreds of predicted binding sites are anticipated. In order to narrow down the genomic regions responsible for driving transgene expression in the dorsal hindbrain and spinal cord, several constructs encompassing different regions of the *lbx1a* enhancer were cloned into the pTol2-GW:*cFos*:*EGFP* vector and assayed for transient *EGFP* expression. Two approaches were used: the first

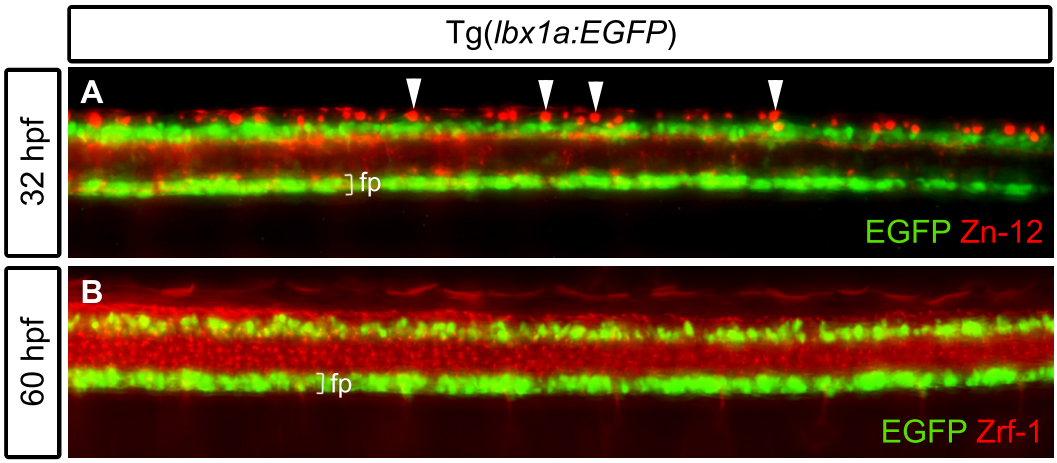

Figure 4-13. **The dorsal position of EGFP^{+} spinal cord neurons in Tg(*lbx1a:EGFP*) embryos**. Immunofluorescence with Zn-12 or Zrf-1 antibodies shows that the *lbx1a* neural enhancer drives transgene expression just beneath the dorsal edge of the spinal cord. At 32 hpf, Zn-12 (A) is a clear marker of Rohon-Beard (Rb) sensory neurons (arrowheads), which exist at the dorsal edge of the spinal cord. At 60 hpf Zrf-1 (B), which is a glial marker, spans the entire dorsoventral axis of the spinal cord. Transgene expression is also driven in the medial floor plate (fp). Both panels are a close-up lateral view of the spinal cord with anterior to the left and dorsal up.

involved cloning smaller fragments of the 1067 bp region while the second involved making short deletions of highly conserved sequences within the 1067 bp (Figure 4-14). All amplified regions were inserted into the transgenic vector in a 5'-3' direction and injected embryos were followed over the first week of development for the presence of fluorescent protein. Each construct is designated by numbers corresponding to the sequence present from the total 1067 bp (1-1067) length or of the short deletion generated. The sequence spanned by each fragment was determined by the overall level of conservation within that region and the availability of sequences capable of acting as suitable PCR primers. Expression of each construct was assessed by at least two independent sets of injections and those that exhibited questionable expression were tested with a variety of vector concentrations and in the pTol-GW:*cFos*:*dTomato* vector as well.

The most important criteria for assessing enhancer fragment or deletion construct activity was the consistency of observed expression in injected embryos over the first week of development. Based on this, three classes of transgene expression patterns were observed from the various constructs made. Some recapitulated the pattern observed for the full length enhancer while others were visible in only specific regions of the neural tube, and many constructs yielded completely inconsistent expression patterns. All of the transgenic constructs created were compared to transient expression of the full length *lbx1a* neural enhancer and an empty vector control (Figure 4-15, A and B). Fragments

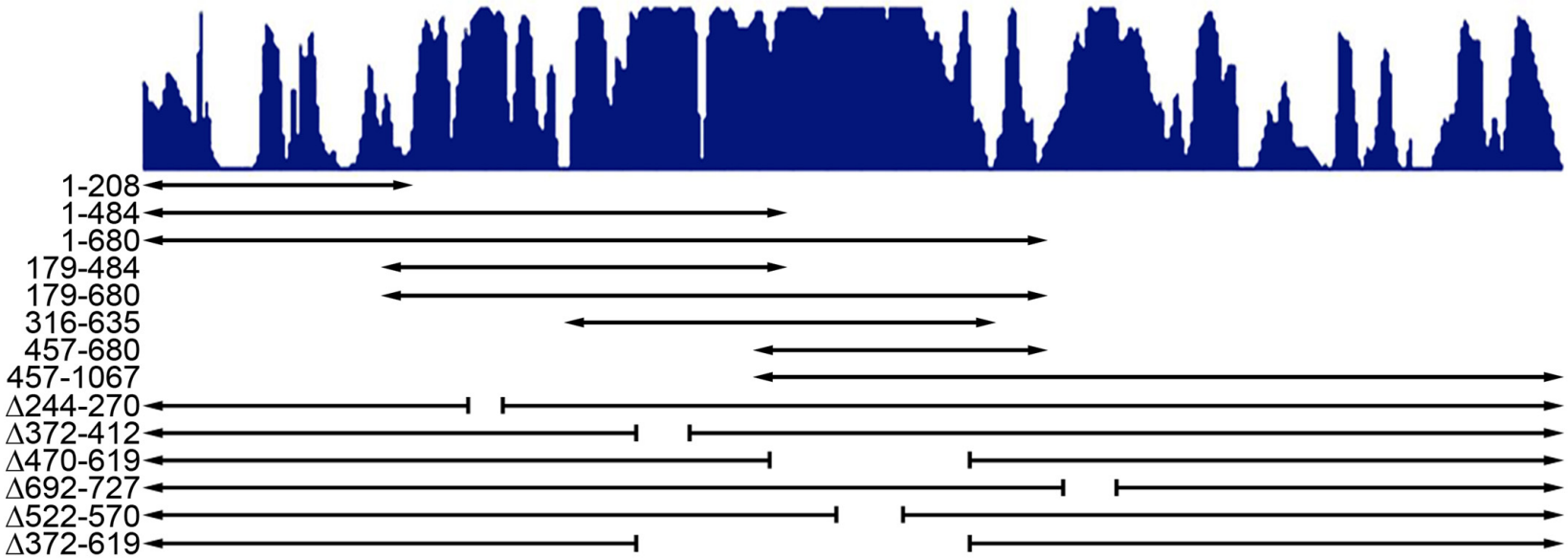

Figure 4-14. **Deletion analysis of the 1067 bp zebrafish *lbx1a* neural enhancer**. Shown is the conservation histogram of the zebrafish *lbx1a* neural enhancer region to the corresponding region at *Takifugu* and *Tetraodon lbx1b* loci, as this region is not present at other teleost fish *lbx1a* loci. In zebrafish, the enhancer lies 727 bp downstream of the stop codon. Fragments (double arrow) of the enhancer were generated through simple PCR while deletion constructs (interrupted double arrow) were made by ligating two seperate PCR products using compatible restriction enzyme sites designed into the primers. Each construct was subsequently cloned into pCR8/GW/TOPO and recombined into the pTol2-GW:*cFos:EGFP* transgenic vector for transient expression analysis via injection with *Tol2*-mRNA into 1-cell zebrafish embryos.

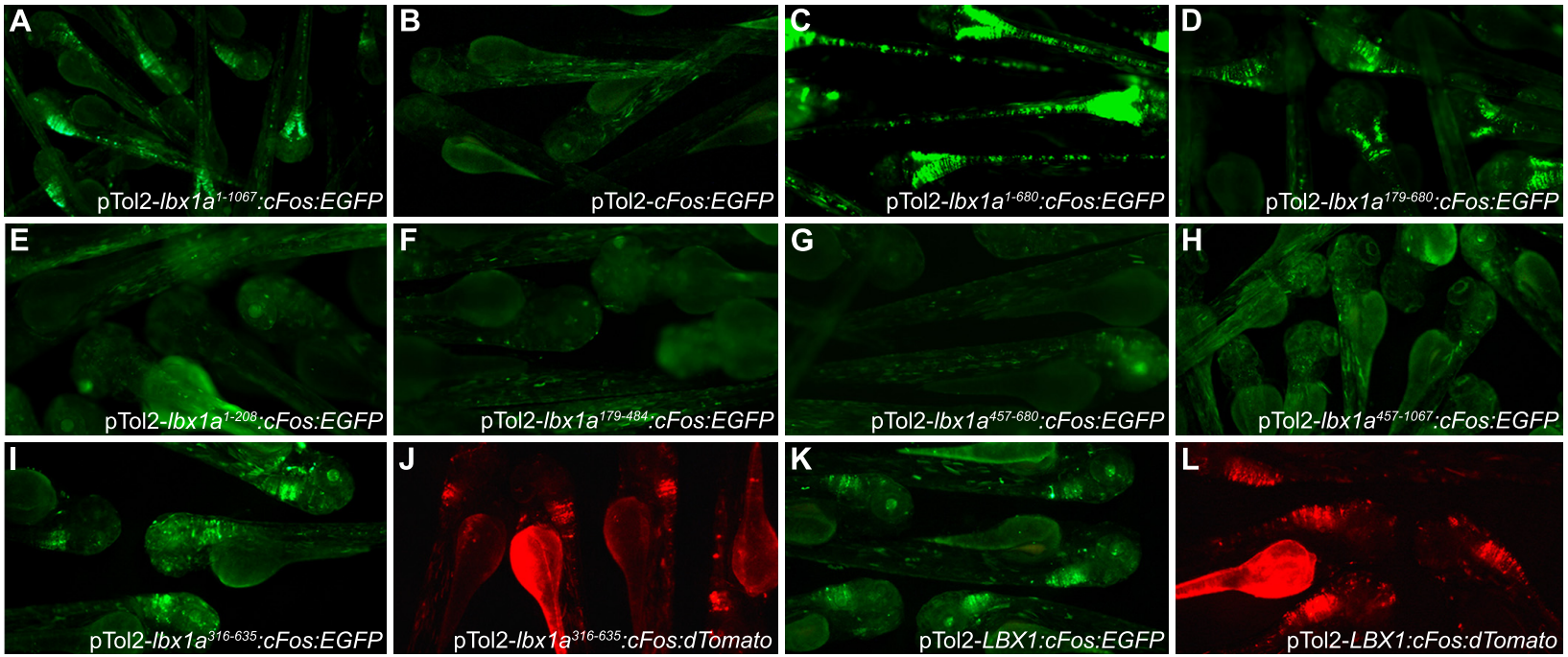

Figure 4-15. **Transient transgene expression driven by various *lbx1a* enhancer fragments in 72 hpf zebrafish embryos**. One-cell zebrafish embryos were injected with the indicated transgenic construct and *Tol2*-mRNA and monitored for fluorescent reporter expression. Superscript numbers denote the base pairs encompassed in each construct of the 1067 bp zebrafish *lbx1a* neural enhancer. As a reference, transient expression from the full length enhancer (A) is provided, as well as an empty vector control (B) which shows typical background expression in trunk muscle. Consistent results were obtained for several fragments, some yielding robust expression (C, D) or very little to no expression (E-H). A small fragment containing base pairs 315-635 (I) was shown to drive consistent hindbrain expression, however, also had significant non-specific EGFP expression. This fragment was also sub-cloned into the pTol-GW:*cFos*:*dTomato* transgenic vector (J) and transient expression confirmed consistent hindbrain expression. The corresponding region at the human *LBX1* locus was also sub-cloned and assayed for transient expression (K, L) and showed consistent hindbrain expression. Each panel shows a group of live 72 hpf embryos that had been injected at the 1-cell stage.

encompassing bp 1-208, 179-484, and 457-680 of the 1067 bp enhancer yielded very little to no consistent *EGFP* expression (Figure 4-15, E, F and G). Fragment 457-1067 (Figure 4-15, H) resulted in consistent transgene expression above that of the empty vector control but it was not restricted to any specific embryonic tissues. Fragment 1-680 (Figure 4-15, C) produced the most robust and consistent pattern in the hindbrain and spinal cord identical to the full length enhancer. Base pairs 179-680 also yielded consistent *EGFP* expression in a pattern similar to the full length construct (Figure 4-15, D), however, EGFP was restricted to only the dorsal hindbrain and very few $EGFP^{+}$ cells were detected in the spinal cord.

The smallest fragment that was able to generate a discernable and consistent *EGFP* expression pattern with a resemblance to the full length enhancer included bp 315-635 (Figure 4-15, I), although the injected DNA concentration was increased from 25 nM to 100 nM to obtain this result. This fragment produced a similar pattern of robust hindbrain expression when tested in the pTol2-GW:*cFos*:*dTomato* vector (Figure 4-15, J). Interestingly, the 489 bp enhancer region cloned from human genomic DNA, which is most similar to fragment 315-635 of the zebrafish *lbx1a* enhancer, also showed transient transgene expression in the zebrafish hindbrain for both the *EGFP* and *dTomato* reporters (Figure 4-15 K and L). Taken together, these results suggest spinal cord expression is driven by sequences in the 5' end of the *lbx1a* enhancer, hindbrain expression is driven by the central region and the 3' end is not essential for transgene expression.

Transient transgene expression is only an indicator of potential enhancer activity and it is likely that very discrete expression patterns will go unnoticed in a transient expression assay as not all of the cells of the embryo have successfully integrated the transgenic construct. Therefore, stable transgenic strains were created for several of the fragments previously mentioned, including ones that did not show any consistent pattern, and examined over the first 72 hours of development. Fragments 1-208 and 179-484, which did not display consistent transient expression patterns, did not display *EGFP* expression in stable transgenic embryos (Figure 4-16, A and B). When combined, an enhancer fragment containing base pairs 1-484 did display weak expression at the very dorsal edge of the hindbrain and spinal cord in stable transgenic embryos (Figure 4-16, C). Although fragment 179-680 had relatively low or undetectable transient expression in the spinal cord, stable Tg(*lbx1a*$^{179\text{-}680}$:*EGFP*) embryos did display spinal cord expression as well as robust hindbrain expression (Figure 4-16, D). Bases 1-680 of the *lbx1a* neural enhancer drove *EGFP* expression in both the hindbrain and spinal cord with a pattern and intensity similar to the full length enhancer (Figure 4-16, F). From transient expression results, it was expected that stable Tg(*lbx1a*$^{457\text{-}1067}$:*EGFP*) transgenic embryos would not show any *EGFP* expression, however, the pattern observed was one with specific differences between rhombomeres (Figure 4-16, F). Mainly, r2 and r4 had high levels of EGFP, while r3 and r5-8 had very low levels of EGFP, and interestingly, the caudal end of the hindbrain did exhibit elevated EGFP.

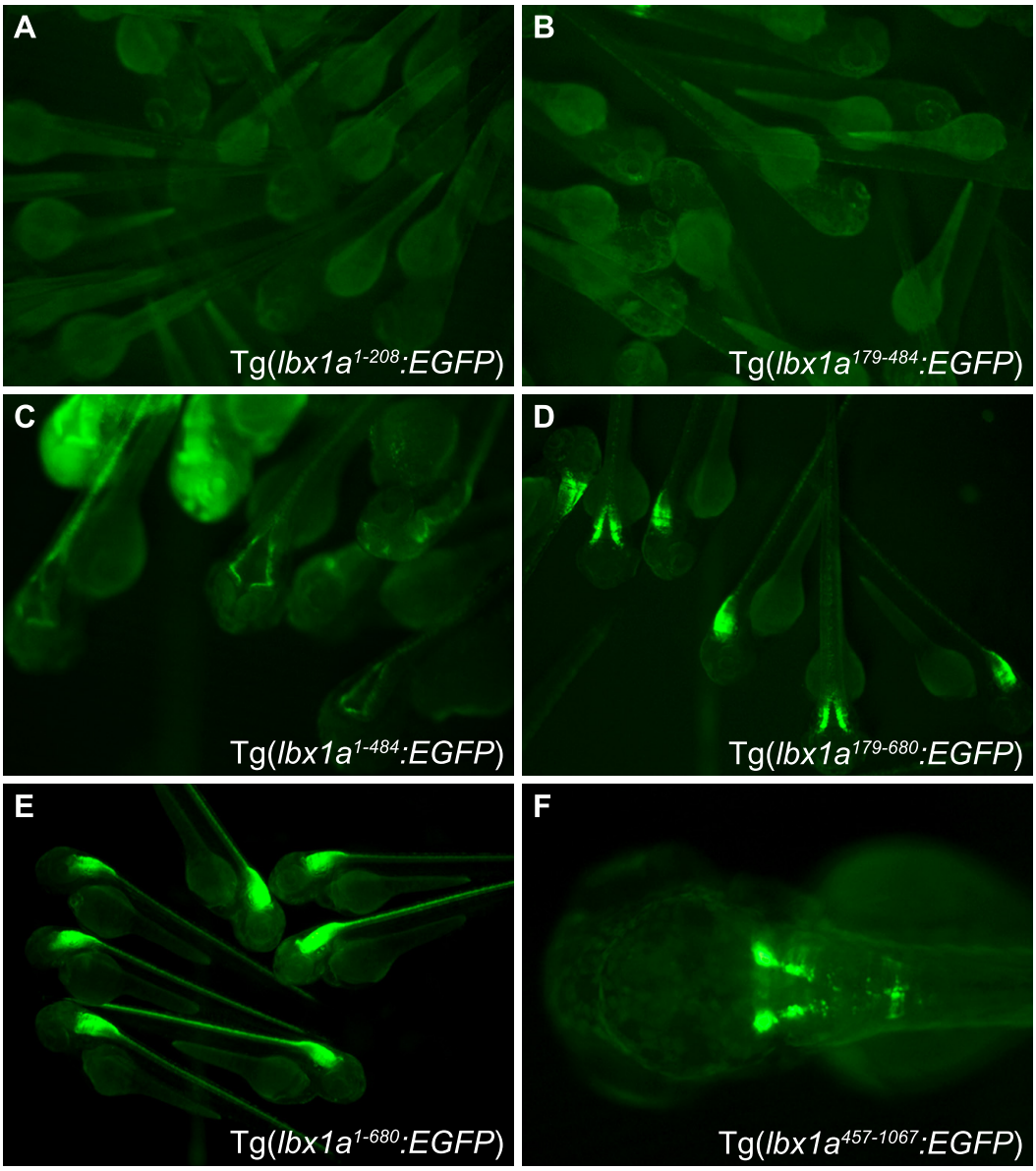

Figure 4-16. **Stable transgenic *EGFP* expression from *lbx1a* enhancer fragments**. Injected embryos showing consistent enhancer fragment transient expression (see Figure 4-12) were raised to adulthood and bred to recover stable transgenic progeny. Stable transgenic embryos offer a more precise representation of enhancer fragment activity as all non-specific transient expression due to multiple random insertions are eliminated. Fragments 1-208 (A) and 179-484 (B) yielded no visibly consistent pattern, although combining the two in fragment 1-484 (C) did show very faint dorsal hindbrain and spinal cord expression. Base pairs 179-680 (D) drove robust hindbrain and faint spinal cord expression while 1-680 (E) mimicked the pattern of the full length *lbx1a* neural enhancer. An interesting pattern within rhombomeres was generated by fragment 457-1067 (F). All panels are images of live 72 hpf stable transgenic zebrafish embryos with enhanced contrast to show the outline of each embryo.

Transient expression of *lbx1a* enhancer constructs containing small deletions of various highly conserved sequence blocks also provided insight into the key regions within the 1067 bp enhancer. Deletion constructs (see Figure 4-14) were inserted into both the EGFP and dTomato variants of the pTol-GW:*cFos* vector and injected into the converse stable transgenic line of the full length *lbx1a* neural enhancer. This provided a quick comparison of transient expression between the injected construct and the pattern generated by the entire 1067 bp enhancer. The variability of transient expression prompted at least three independent rounds of injections to be performed for each construct. Embryos were raised through the first week of development and examined for two main criteria: 1) whether the pattern generated by transient construct expression was consistent and 2) how similar the pattern was to stable transgene expression. The intensity of reporter protein was also assessed though this characteristic can yield some variability between embryos injected with the same construct.

Regions chosen for deletion were highly conserved between the zebrafish *lbx1a* locus and other fish *lbx1b* loci. Deletions of bp 244-270, 372-412, 470-619, 522-570 and 692-727 did not significantly reduce or disrupt hindbrain expression (Figure 4-17, A, C, E, G and I) while Δ372-619 restricted hindbrain expression to the dorsal edge (Figure 4-17, K). In the spinal cord, deletion of base pairs 244-270, 372-412 and 470-619 did not affect the dorsoventral plane of transgene expression or the consistency of expression throughout the spinal cord (Figure 4-17, B, D and F). Constructs bearing deletions of bp 692-727 did not disrupt

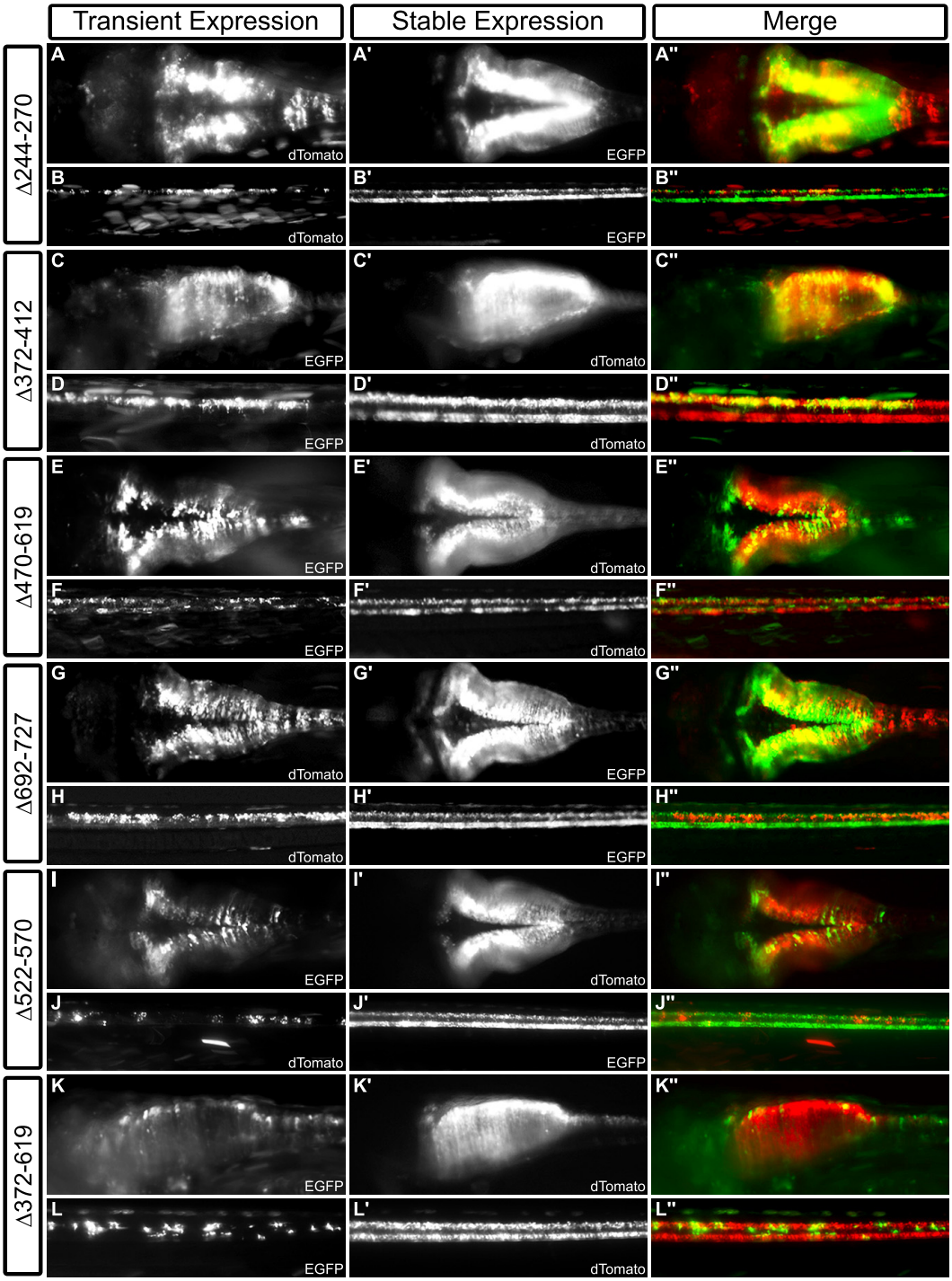

Figure 4-17. **Transient transgene expression from deletion constructs of the *Ibx1a* neural enhancer**. Tg(*Ibx1a*:*EGFP*) or Tg(*Ibx1a*:*dTomato*) stable transgenic embryos were injected with pTol2-*Ibx1a*:*cFos*:*dTomato* or pTol2-*Ibx1a*:*cFos*:*EGFP*, respectively, containing a variety of deletions (left) from the 1067 bp enhancer element. Embryos were examined at 4 dpf by live fluorescence microscopy. Shown is injected construct reporter expression (A-L), stable transgene expression (A'-L') and merged images (A''-L'').

consistency of expression in the spinal cord, although interestingly, expression was expanded throughout the dorsoventral plane (Figure 4-17, H). Δ522-570 and Δ372-619 constructs displayed patchy expression in the spinal cord, however, the areas of the spinal cord that did have transient expression were expanded throughout the dorsoventral axis (Figure 4-17, J and L).

The transient expression data from all of the various fragments and deletion constructs of the *lbx1a* neural enhancer (Figure 4-18) points to several conclusions. Constructs that lack the 5' half show a significant reduction of spinal cord expression and the central regions appears to be responsible for hindbrain expression. The 3' half drives variability in axial expression such as changes in the dorsoventral domain in the spinal cord and rostrocaudal differences in the hindbrain. Transient expression in the floor plate was also somewhat variable with several constructs having very little expression and others resembling stable transgene expression from the full length *lbx1a* enhancer.

The analyses of transgenic zebrafish bearing the *lbx1a* and *lbx1b* enhancers confirm that the Tol2 system is a quick and effective method of identifying highly conserved genomic sequences with enhancer activity. It is however, intriguing that such large blocks of non-coding genomic sequence are conserved between vertebrates when it is known that transcription factor binding sites are relatively short, ranging from 4 to 12 base pairs. Combining transcription factor binding site prediction software algorithms could potentially make the Tol2 system an efficient method for identifying critical sequences within enhancers

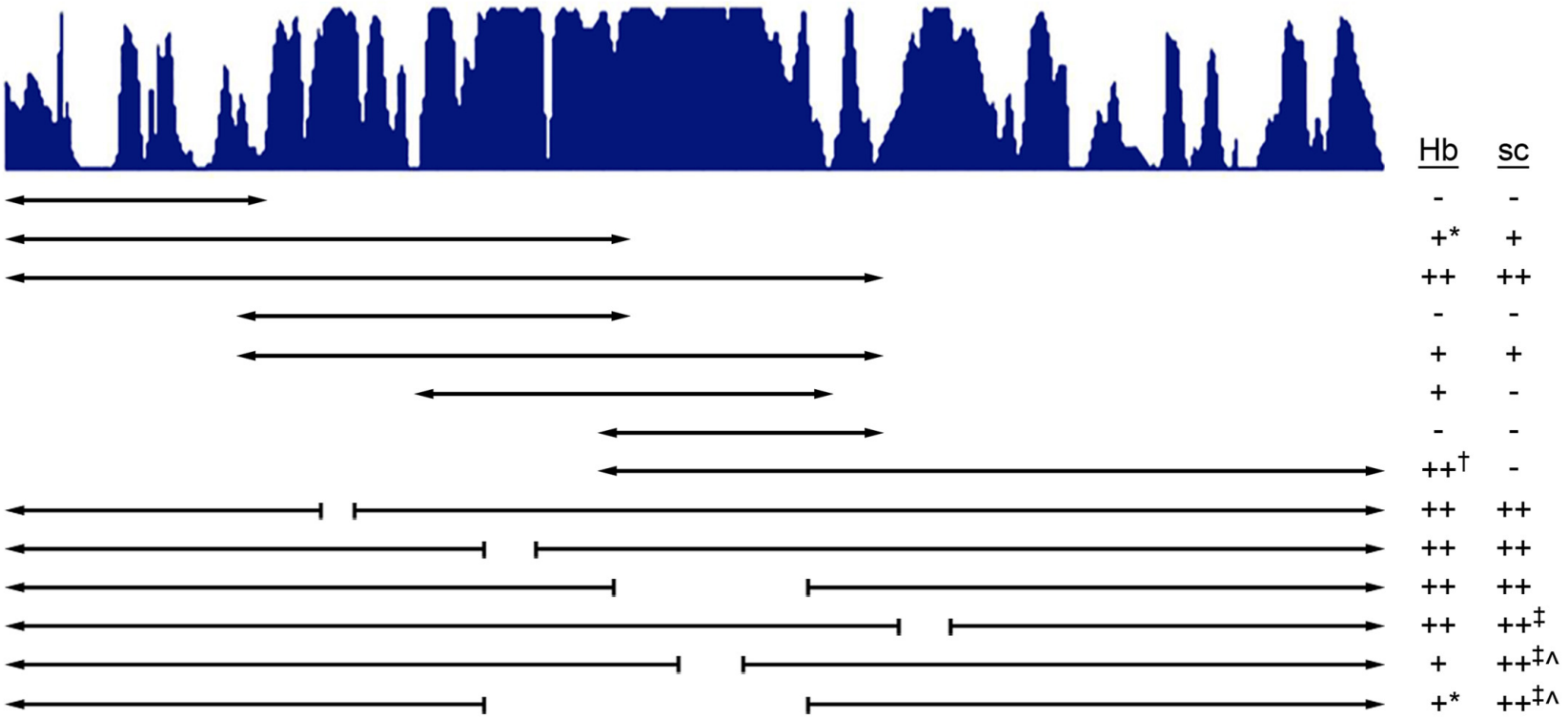

Figure 4-18. **Summary of transgene expression results obtained from various fragments and deletion constructs of the zebrafish *lbx1a* neural enhancer**. Both transient and stable expression of enhancer fragments and transient expression data for deletion constructs is summarized on the right for the hindbrain (Hb) and spinal cord (sc). - indicates very little to no transgene expression; + indicates reduced expression compared to the full length enhancer and ++ denotes no discernable difference to the full length enhancer. Expression patterns were assigned to several classes; * indicates expression only at the dorsal edge; † indicates variable expression was observed between rhombomeres; ‡ designates expression throughout the dorsoventral axis of the spinal cord and ^ denotes unevenly distributed expression within the spinal cord.

through deletion analysis. The vast number of different short deletion constructs that can be created from a single enhancer element, such as the 1 kbp *lbx1a* neural enhancer makes this a tedious task. It is clear that generating smaller fragments of an enhancer can narrow down the position of critical regions, although, limitations exist, such as the need to examine the progeny of numerous stable transgenic adults to identify the true expression pattern generated by each fragment.

3) References

Anderson JD, Lowary PT and Widom J, **2001**. Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol*. 307:977-985.

Babu MM, Luscombe NM, Aravind L, Gerstein M and Teichmann SA, **2004**. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol*. 14:283-291.

Barbazuk WB, Korf I, Kadavi C, Heyen J, Tate S, Wun E, Bedell JA, McPherson JD and Johnson SL, **2000**. The syntenic relationship of the zebrafish and human genomes. *Genome Res*. 10:1351-1358.

Bartova E, Krejci J, Harnicarova A, Galiova G and Kozubek S, **2008**. Histone modifications and nuclear architecture: a review. *J Histochem Cytochem*. 56:711-721.

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS and Crawford GE, **2008**. High-resolution mapping and characterization of open chromatin across the genome. *Cell*. 132:311-322.

Brohmann H, Jagla K and Birchmeier C, **2000**. The role of Lbx1 in migration of muscle precursor cells. *Development*. 127:437-445.

Cheng L, Samad OA, Xu Y, Mizuguchi R, Luo P, Shirasawa S, Goulding M and Ma Q, **2005**. Lbx1 and Tlx3 are opposing switches in determining GABAergic versus glutamatergic transmitter phenotypes. *Nat Neurosci*. 8:1510-1515.

Choi JK and Howe LJ, **2009**. Histone acetylation: truth of consequences? *Biochem Cell Biol*. 87:139-150.

Chudakov DM, Lukyanov S and Lukyanov KA, **2005**. Fluorescent proteins as a toolkit for in vivo imaging. *Trends Biotechnol*. 23:605-613.

Crawford GE, Holt IE, Mullikin JC, Tai D, Blakesley R, Bouffard G, Young A, Masiello C, Green ED, Wolfsberg TG and Collins FS, **2004**. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc Natl Acad Sci U S A*. 101:992-997.

De Graeve F, Jagla T, Daponte JP, Rickert C, Dastugue B, Urban J and Jagla K, **2004**. The ladybird homeobox genes are essential for the specification of a subpopulation of neural cells. *Dev Biol*. 270:122-134.

Fisher S, Grice EA, Vinton RM, Bessling SL and McCallion AS, **2006a**. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science*. 312:276-279.

Fisher S, Grice EA, Vinton RM, Bessling SL, Urasaki A, Kawakami K and McCallion AS, **2006b**. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat Protoc*. 1:1297-1305.

Follows GA, Dhami P, Gottgens B, Bruce AW, Campbell PJ, Dillon SC, Smith AM, Koch C, Donaldson IJ, Scott MA, Dunham I, Janes ME, Vetrie D and Green AR, **2006**.

Identifying gene regulatory elements by genomic microarray mapping of DNaseI hypersensitive sites. *Genome Res*. 16:1310-1319.

Frazer KA, Pachter L, Poliakov A, Rubin EM and Dubchak I, **2004**. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 32:W273-279.

Froschauer A, Braasch I and Volff JN, **2006**. Fish Genomes, Comparative Genomics and Vertebrate Evolution. *Current Genomics*. 7:43-57.

Gilligan P, Brenner S and Venkatesh B, **2002**. Fugu and human sequence comparison identifies novel human genes and conserved non-coding sequences. *Gene*. 294:35-44.

Gross MK, Dottori M and Goulding M, **2002**. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron*. 34:535-549.

Gross MK, Moran-Rivard L, Velasquez T, Nakatsu MN, Jagla K and Goulding M, **2000**. Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. *Development*. 127:413-424.

Guth SI, Bosl MR, Sock E and Wegner M, **2009**. Evolutionary conserved sequence elements with embryonic enhancer activity in the vicinity of the mammalian Sox8 gene. *Int J Biochem Cell Biol*.

Hamlet MR, Yergeau DA, Kuliyev E, Takeda M, Taira M, Kawakami K and Mead PE, **2006**. Tol2 transposon-mediated transgenesis in Xenopus tropicalis. *Genesis*. 44:438-445.

International, Chicken, Genome, Sequencing and Consortium, **2004**. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 432:695-716.

Jagla T, Bellard F, Lutz Y, Dretzen G, Bellard M and Jagla K, **1998**. ladybird determines cell fate decisions during diversification of Drosophila somatic muscles. *Development*. 125:3699-3708.

Karolchik D, Hinrichs AS and Kent WJ, **2007**. The UCSC Genome Browser. *Curr Protoc Bioinformatics*. Chapter 1:Unit 1 4.

Kawakami K, **2004**. Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element. *Methods Cell Biol*. 77:201-222.

Kawakami K, **2005**. Transposon tools and methods in zebrafish. *Dev Dyn*. 234:244-254.

Kawakami K, **2007**. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol*. 8 Suppl 1:S7.

Kawakami K, Imanaka K, Itoh M and Taira M, **2004a**. Excision of the Tol2 transposable element of the medaka fish Oryzias latipes in Xenopus laevis and Xenopus tropicalis. *Gene*. 338:93-98.

Kawakami K, Koga A, Hori H and Shima A, **1998**. Excision of the tol2 transposable element of the medaka fish, Oryzias latipes, in zebrafish, Danio rerio. *Gene*. 225:17-22.

Kawakami K and Noda T, **2004**. Transposition of the Tol2 element, an Ac-like element from the Japanese medaka fish Oryzias latipes, in mouse embryonic stem cells. *Genetics*. 166:895-899.

Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N and Mishina M, **2004b**. A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev Cell*. 7:133-144.

Kent WJ, **2002**. BLAT--the BLAST-like alignment tool. *Genome Res*. 12:656-664.

Khandekar M, Suzuki N, Lewton J, Yamamoto M and Engel JD, **2004**. Multiple, distant Gata2 enhancers specify temporally and tissue-specific patterning in the developing urogenital system. *Mol Cell Biol*. 24:10263-10276.

Kleinjan DA, Seawright A, Mella S, Carr CB, Tyas DA, Simpson TI, Mason JO, Price DJ and van Heyningen V, **2006**. Long-range downstream enhancers are essential for Pax6 expression. *Dev Biol*. 299:563-581.

Koga A, Suzuki M, Inagaki H, Bessho Y and Hori H, **1996**. Transposable element in fish. *Nature*. 383:30.

Latchman DS, **1997**. Transcription factors: an overview. *Int J Biochem Cell Biol*. 29:1305-1312.

Li B, Carey M and Workman JL, **2007**. The role of chromatin during transcription. *Cell*. 128:707-719.

Lin CH, Stoeck J, Ravanpay AC, Guillemot F, Tapscott SJ and Olson JM, **2004**. Regulation of neuroD2 expression in mouse brain. *Dev Biol*. 265:234-245.

Lobe CG, **1992**. Transcription factors and mammalian development. *Curr Top Dev Biol*. 27:351-383.

Marcus RC and Easter SS, Jr., **1995**. Expression of glial fibrillary acidic protein and its relation to tract formation in embryonic zebrafish (Danio rerio). *J Comp Neurol*. 359:365-381.

Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S and Wingender E, **2003**. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*. 31:374-378.

McCarthy AA, **2005**. Broad institute: bringing genomics to real-world medicine. *Chem Biol*. 12:717-718.

Mennerich D, Schafer K and Braun T, **1998**. Pax-3 is necessary but not sufficient for lbx1 expression in myogenic precursor cells of the limb. *Mech Dev*. 73:147-158.

Messeguer X, Escudero R, Farre D, Nunez O, Martinez J and Alba MM, **2002**. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics*. 18:333-334.

Metcalfe WK, Myers PZ, Trevarrow B, Bass MB and Kimmel CB, **1990**. Primary neurons that express the L2/HNK-1 carbohydrate during early development in the zebrafish. *Development*. 110:491-504.

Meyer A and Van de Peer Y, **2005**. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays*. 27:937-945.

Miller KA, Barrow J, Collinson JM, Davidson S, Lear M, Hill RE and Mackenzie A, **2007a**. A highly conserved Wnt-dependent TCF4 binding site within the proximal enhancer of the anti-myogenic Msx1 gene supports expression within Pax3-expressing limb bud muscle precursor cells. *Dev Biol*. 311:665-678.

Miller W, Rosenbloom K, Hardison RC, Hou M, Taylor J, Raney B, Burhans R, King DC, Baertsch R, Blankenberg D, Kosakovsky Pond SL, Nekrutenko A, Giardine B, Harris RS, Tyekucheva S, Diekhans M, Pringle TH, Murphy WJ, Lesk A, Weinstock GM, Lindblad-Toh K, Gibbs RA, Lander ES, Siepel A, Haussler D and Kent WJ, **2007b**. 28-way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Res*. 17:1797-1808.

Mitchell PJ and Tjian R, **1989**. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*. 245:371-378.

Muller T, Brohmann H, Pierani A, Heppenstall PA, Lewin GR, Jessell TM and Birchmeier C, **2002**. The homeodomain factor lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron*. 34:551-562.

Narlikar GJ, Fan HY and Kingston RE, **2002**. Cooperation between complexes that regulate chromatin structure and transcription. *Cell*. 108:475-487.

Ohno S, **1972**. So much "junk" DNA in our genome. *Brookhaven Symp Biol*. 23:366-370.

Parker SC, Hansen L, Abaan HO, Tullius TD and Margulies EH, **2009**. Local DNA topography correlates with functional noncoding regions of the human genome. *Science*. 324:389-392.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A and Rubin EM, **2006**. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*. 444:499-502.

Pevsner J, **2009**. Analysis of genomic DNA with the UCSC genome browser. *Methods Mol Biol*. 537:277-301.

Powles N, Marshall H, Economou A, Chiang C, Murakami A, Dickson C, Krumlauf R and Maconochie M, **2004**. Regulatory analysis of the mouse Fgf3 gene: control of embryonic expression patterns and dependence upon sonic hedgehog (Shh) signalling. *Dev Dyn*. 230:44-56.

Ptashne M and Gann A, **1997**. Transcriptional activation by recruitment. *Nature*. 386:569-577.

Ratnere I and Dubchak I, **2009**. Obtaining comparative genomic data with the VISTA family of computational tools. *Curr Protoc Bioinformatics*. Chapter 10:Unit 10 16.

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP and Young RA, **2000**. Genome-wide location and function of DNA binding proteins. *Science*. 290:2306-2309.

Rossi CC, Kaji T and Artinger KB, **2009**. Transcriptional control of Rohon-Beard sensory neuron development at the neural plate border. *Dev Dyn*. 238:931-943.

Rubtsov MA, Polikanov YS, Bondarenko VA, Wang YH and Studitsky VM, **2006**. Chromatin structure can strongly facilitate enhancer action over a distance. *Proc Natl Acad Sci U S A*. 103:17690-17695.

Sato Y, Kasai T, Nakagawa S, Tanabe K, Watanabe T, Kawakami K and Takahashi Y, **2007**. Stable integration and conditional expression of electroporated transgenes in chicken embryos. *Dev Biol*. 305:616-624.

Schafer K and Braun T, **1999**. Early specification of limb muscle precursor cells by the homeobox gene Lbx1h. *Nat Genet*. 23:213-216.

Shabalina SA and Spiridonov NA, **2004**. The mammalian transcriptome and the function of non-coding DNA sequences. *Genome Biol*. 5:105.

Shin JT, Priest JR, Ovcharenko I, Ronco A, Moore RK, Burns CG and MacRae CA, **2005**. Human-zebrafish non-coding conserved elements act in vivo to regulate transcription. *Nucleic Acids Res*. 33:5437-5445.

Uchikawa M, Takemoto T, Kamachi Y and Kondoh H, **2004**. Efficient identification of regulatory sequences in the chicken genome by a powerful combination of embryo electroporation and genome comparison. *Mech Dev*. 121:1145-1158.

Urasaki A, Mito T, Noji S, Ueda R and Kawakami K, **2008**. Transposition of the vertebrate Tol2 transposable element in Drosophila melanogaster. *Gene*. 425:64-68.

Urasaki A, Morvan G and Kawakami K, **2006**. Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics*. 174:639-649.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM and Pennacchio LA, **2009**. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 457:854-858.

Won KJ, Agarwal S, Shen L, Shoemaker R, Ren B and Wang W, **2009**. An integrated approach to identifying cis-regulatory modules in the human genome. *PLoS ONE*. 4:e5501.

Woolfe A, Goode DK, Cooke J, Callaway H, Smith S, Snell P, McEwen GK and Elgar G, **2007**. CONDOR: a database resource of developmentally associated conserved non-coding elements. *BMC Dev Biol*. 7:100.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE and Elgar G, **2005**. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol*. 3:e7.

Zweig AS, Karolchik D, Kuhn RM, Haussler D and Kent WJ, **2008**. UCSC genome browser tutorial. *Genomics*. 92:75-84.

Chapter 5: Conclusions and Future Directions

1) Conclusions

Mammalian species possess two *Ladybird* genes that are similar but exhibit significant differences in their predicted proteins. Survey of teleost fish genomes for *ladybird*-like genes has yielded both expected and unexpected results. Synteny analysis confirms the presence of two *Lbx1* paralogs, which is consistent with the theory of a fish specific genome duplication (Meyer and Van de Peer, 2005; Froschauer *et al.*, 2006), however, a third *Lbx1*-like coding region exists in the genomes of teleost fish. Currently, this third *ladybird* gene is designated *lbx2* in fish, yet the predicted amino acid sequences resemble vertebrate Lbx1. The genomic location and phylogenetic analysis of gene clusters (Wotton *et al.*, 2008) suggests it is an *Lbx2* ortholog, although the defining amino acid sequences of Lbx2 protein family members are only shared between mammalian species. Perhaps the divergence of Lbx2 sequences, particularly the homeodomain, arose concomitantly with mammalian evolution and ancient Ladybird sequences were more similar than perceived prior to the vertebrate lineage genome duplication. All of the identified teleost fish *ladybird* genes encode proteins with high similarity to Lbx1 proteins, although unique regions exist along the protein that allow orthologous name designations to be assigned between fish species.

The expression pattern of zebrafish *lbx1a* and *lbx1b* very closely resemble *Lbx1* expression in other vertebrates examined such as frog, chicken and mouse. These expression patterns are consistent with what appears to be an ancient role

for Ladybird in muscular and neuronal development, which was initially determined in *Drosophila* (Jagla *et al.*, 1993; Jagla *et al.*, 1998). Zebrafish *lbx1* paralogs do exhibit a remarkably similar expression pattern when observed at later stages of embryonic development. However, it is their early expression patterns which point to a rationale that both genes have been maintained. Zebrafish *lbx1a* is expressed very early throughout the hindbrain with specific variation between rhombomeres while *lbx1b* is restricted in this tissue to a subset of neurons that occupy a fine stripe approximately midway between the medial and lateral edges. In pectoral fin muscle, zebrafish *lbx1b* is expressed much earlier than *lbx1a*, at a point where precursors appear to be migrating into the developing pectoral fin bud. Zebrafish *lbx2* is also expressed in the pectoral fin and dorsal neural tube but also in the ventral somites, posterior fin precursors, pharyngeal arches and in clusters of presumptive muscle around the eye. Functional analysis in zebrafish suggests Ladybird proteins positively regulate the *Wnt* pathway as *Lbx*-mRNA injected embryos exhibit a phenotype very similar to mutant fish where the *Wnt* pathway is ectopically activated in anterior tissues.

While the expression patterns and functions of *Ladybird* have been well documented in a variety of species, there is very little evidence as to the nature of transcriptional regulation at *Ladybird* loci. The rapidly advancing field of comparative genomics has identified a common trend of extremely high sequence conservation exhibited amongst regulatory elements of developmentally controlled genes. Using this methodology, several highly conserved genomic

regions in the vicinity of *lbx1* coding sequences in zebrafish were found to recapitulate the neuronal expression pattern observed by endogenous transcripts. This non-coding genomic region is conserved at all vertebrate *Lbx1* loci examined and the corresponding region at the human *LBX1* locus is able to function in zebrafish, suggesting the mode of transcriptional regulation at this enhancer region is also conserved. Breakdown of the cloned zebrafish *lbx1* enhancer region into fragments or with small deletions indicates the 5' end is important for spinal cord expression and the central region drives hindbrain expression. Analysis of the 3' end provides clues to the anteroposterior and dorsoventral variation observed for endogenous *lbx1a* expression patterns.

Genomic analysis of fish *ladybird* loci provides insight into the evolution of teleost fish species. The conserved regions surrounding both *lbx1a* and *lbx1b*, when taken together, are very similar in location and spacing to mammalian *Lbx1* loci. On the other hand, comparisons between fish species only shows that zebrafish, which belongs to a different superorder than Medaka, Stickleback, *Tetraodon*, and Fugu, has retained a different set of conserved non-coding regions at each *lbx1* locus. This observation, along with more recent evidence examining other loci such as the *Hox* clusters (Hoegg *et al.*, 2007), points to a theory that differential loss of duplicated regulatory elements is just as important in speciation as sub- or neo-functionalization of paralogous gene products. There is much that can be learned about regulatory elements through the use of zebrafish

transgenic technology now that efficient transgenesis is readily accomplished via the Tol2 transposable element.

The Tg(*lbx1a*:*dTomato*) transgenic zebrafish strain provides a useful tool for assessing the development of dorsal spinal cord neurons as some differences are apparent between zebrafish and amniotes like chicken and mouse. Firstly, the dorsoventral plane in which *lbx1* is initially expressed in the zebrafish spinal cord occurs just beneath the dorsal edge. In mice however, early expression of *Lbx1* occurs in the middle of the spinal cord and those neurons subsequently migrate to occupy the dorsal horn of the neural tube. A plausible explanation for this difference is the fact that anamniote species such as zebrafish and frog develop a primary nervous system to facilitate simple motor movements after hatching. The primary nervous system is specified prior to neurulation (Clarke, 2009) and thus specific positions along the mediolateral axis of the neural plate determine neuronal position along the dorsoventral axis of the neural tube, rather than cues from the roof and floor plate. Similar signaling molecules such as BMPs, Shh and Wnts pattern neurons along the mediolateral axis of the zebrafish spinal cord (Lewis and Eisen, 2003). The dorsoventral position of secondary spinal cord neurons, which are more analogous to the mammalian nervous system, is dependent on the relative position of primary neurons as secondary neuron progenitors are set aside from primary neuron differentiation.

The patterning of zebrafish spinal cord interneurons is poorly understood when compared to mice and further comparative studies are needed to determine

the transcription factor code of these neurons. Transgenic zebrafish strains with labeled interneuron populations, such as the Tg(*lbx1a*:*dTomato*) line, will facilitate this line of research as these neurons can be isolated from whole embryos or dissected spinal cords through fluorescence activated cell sorting (FACS). High throughput microarray analysis can then be used to identify genes

2) Future Directions

a) Further Analysis of Zebrafish *Lbx1* Regulatory Elements

Work remains to identify the transcription factors that regulate neural expression of zebrafish *lbx1* genes. Analysis of transient expression from the deletion constructs has yielded some interesting results towards the regions required for hindbrain and spinal cord expression. The stable transgenic strains created for the various enhancer fragments did follow the results collected from transient expression analysis but unexpected results were also obtained. Because such variable results can be observed from simply varying the amount of transgenic vector delivered to a 1-cell zebrafish embryo (Figure 5-1), transient expression results can only be used as an indication of the possible tissues that are capable of driving the transgene. Stable transgenic strains are a far more reliable output of the regulatory capacity of a putative enhancer element since it is known that every cell in the embryo harbors the transgenic construct. The creation of transgenic strains carrying single insertions for each of the *lbx1a* enhancer deletion constructs is an important experiment that remains to be carried out.

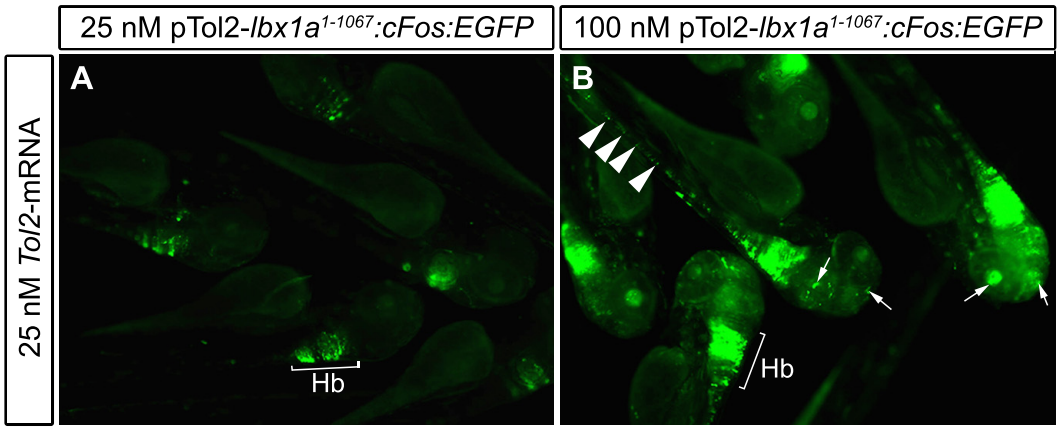

Figure 5-1. **Transgenic construct concentration greatly affects transient transgene expression**. Tol2 plasmid harboring the full length *lbx1a* neural enhancer was injected into one-cell zebrafish embryos in varying amounts to assess the concentration range appropriate for transient transgene expression analysis. Each panel shows a group of live 4 day old zebrafish embryos expressing the *EGFP* transgene photographed with the same exposure time. The volume injected was kept constant between 2-3 nL as this was found to be optimal in both delivery of the construct and overall embryo survival. The Tol2 mRNA concentration was unchanged at 25 nM. The low range of plasmid concentration that yielded a discernable expression pattern was approximately 25 nM (A), however, consistent expression was only observed in the hindbrain (Hb). Higher concentrations near 100 nM (B) gave more intense expression in the hindbrain as well as the spinal cord (arrowheads) expression was also apparent. However, an increase in non-specific expression (arrows), abnormal embryonic development and embryonic lethality was observed at concentrations of transgenic construct exceeding 100 nM.

Determining the expression pattern generated by stable transgenic progeny containing the various deletion constructs is critical in resolving the location of specific regulatory elements. Since the *lbx1a* neural enhancer deletion constructs have only been viewed as a transient expression assay from injected plasmid at the 1-cell stage, their mosaic results are not as clear. A major drawback of maintaining stable transgenic strains is the requirement for numerous adults to be kept in a fish facility that need to be screened several times for germline transgenesis. Stable transgene expression generated by each deletion construct will identify short stretches of sequence which are necessary to drive dorsal hindbrain and spinal cord patterns as well as the differences between rhombomeres. Transcription factor binding site prediction software can then be applied to these short regions and, in combination with site directed mutagenesis, will identify the precise regulatory input of zebrafish *ladybird* genes. Alternatively, the full length *lbx1a* neural enhancer can be used as a substrate to pull out transcription factors which bind there from zebrafish protein extracts which can then be identified through mass spectrometry. DNase footprinting (Galas and Schmitz, 1978) may also be helpful to elucidate the precise sequences which are necessary for transcriptional regulation of the *ladybird* locus.

The genomic region responsible for muscle expression of vertebrate *Ladybird* genes has not yet been identified, although several conserved regions surrounding the *Lbx1* locus have not yet been tested in a transient zebrafish transgenesis assay. However, the region driving muscle expression may not

demonstrate significant conservation to be recognized through BLAT analysis of the genomic regions surrounding the vertebrate *Lbx1* locus. Since the Tol2 vector is capable of supporting up to 5 kbp in sequence comfortably, a straightforward way to identify this enhancer would be to test large chunks of genomic sequence in a transient assay. With the small size of the pectoral fins it may be difficult to positively identify a muscle-specific construct simply by observing transient expression, although as shown previously, manipulating the injected plasmid amount may alleviate this limitation. Identifying this enhancer element will allow comparisons to be made to mammalian limb muscle development.

b) Determining the Function of Zebrafish *Ladybird* Gene Products

Morpholino-mediated knockdown of zebrafish *lbx1a*, *lbx1b* and *lbx2* has been attempted with a lack of conclusive results, as is the case with many duplicated genes in zebrafish. Injecting morpholinos into the stable transgenic Tg(*lbx1a*:*dTomato; isl1:GFP*) embryos may provide a more clear result. It is expected that knockdown of *lbx1b* would produce defects in pectoral fin development since it is the earliest *ladybird* gene expressed there. However, since *lbx1a* and *lbx2* are expressed in the pectoral fin shortly after *lbx1b*, they may compensate and mask distinct defects of *lbx1b* morphants. The combination of all 3 morpholinos was also attempted although to maintain low levels of embryonic lethality the total morpholino amount injected remained constant with single or double morpholino injections, effectively lowering the amount of each individual

morpholino. With recently developed zinc finger nucleases as an emerging tool to create mutant zebrafish (Meng *et al.*, 2008) it may be possible to assess morpholino knockdown of one *ladybird* gene in a strain carrying a mutation in another. Additionally, large-scale zebrafish retroviral mutagenesis carried out by Znomics yielded several strains with insertions located in and around the zebrafish *lbx2* coding region.

High-throughput analysis identifying Lbx1-regulated genes has been done in mice, revealing a complex network controlling the dorsoventral patterning of the neural tube (Kioussi *et al.*, 2008). It would be interesting to determine if any of these predicted targets are up-regulated in zebrafish embryos injected with *lbx* mRNA. Since the phenotype of *lbx* mRNA injected zebrafish embryos very closely resembles zebrafish *Wnt* pathway mutants, it is worthy to seek out the regulatory roles of Lbx on *Wnt* signaling by assessing the expression patterns of Wnt target genes. The sequence conservation observed amongst teleost Ladybird proteins suggest they play similar functions, although they have very distinct spatiotemporal expression patterns during development. One longstanding question with regards to zebrafish Ladybird proteins is the potential dual activator-repressor function indicated by the domain architecture of an engrailed homology domain near the amino terminus and a highly acidic carboxy terminus. Injecting zebrafish embryos with mRNA encoding truncated ladybird proteins as well as translational fusions to replace the acidic carboxy-terminus with a VP16 activation domain would confirm the predicted activity of the carboxy-terminus.

c) Control of Dorsoventral Patterning in the Zebrafish Spinal Cord

In mice, it is thought that since *Lbx1*$^{+}$ spinal cord neuron development is not perturbed by the loss of roof or floor plate signaling, a default neuronal sub-type exists which escapes the Sonic hedgehog-Bmp gradient that patterns the dorsoventral axis of the spinal cord (Muller *et al.*, 2002). In zebrafish, the dorsal position of *Lbx1*$^{+}$ neurons suggests they could potentially be patterned by roof plate signaling, although analysis of zebrafish Bmp mutants indicates RB sensory neurons and neural crest are the only cell types lacking from the loss of roof plate signaling (Nguyen *et al.*, 2000). To avoid the severe dorsalized phenotype associated with loss of early Bmp signaling and observe *lbx1*$^{+}$ neurons, treatment of zebrafish embryos with dorsomorphin, a chemical inhibitor of Bmp signaling (Yu *et al.*, 2008), was attempted at the tailbud stage of Tg(*lbx1a*:*dTomato*) embryos. Issues including the photostability of dorsomorphin and extremely low rate of penetrance prevented any definitive results from being obtained. To circumvent these issues and determine if *lbx1a*$^{+}$ neurons rely on Bmp signaling, it is suggested that a higher dose (>100 µM) be administered 1-2 hours prior to tailbud stage.

Conversely, the effects of zebrafish floor plate signaling appear to extend to the dorsal-most population of RB neurons such that an expansion of RB neurons occurs in the absence of notochord-derived signals (Nguyen *et al.*, 2000). It would be interesting to observe the effects of a chemical inhibitor of sonic

hedgehog signaling, cyclopamine, on the distribution and abundance of *lbx1a*$^{+}$ spinal cord neurons. If the domain of dorsally-situated RB neurons is expanded by the loss of hedgehog signaling in zebrafish and the distribution of *lbx1*$^{+}$ neurons is not, then it would suggest this neuronal class develops independently of floor plate signals. A recent study examining several zebrafish spinal cord neuron markers suggests the dorsal interneurons are not perturbed by cyclopamine treatment although the analysis of *ladybird* expression was not examined (Guner and Karlstrom, 2007).

Transplantation of cells from Tg(*lbx1a*:*dTomato*) or Tg(*lbx1a*:*EGFP*) donor embryos into WT host embryos may provide information on the neuronal types that express *ladybird*. It is difficult to track the axonal projections of single neurons in an entire neuronal class, such as the labeled neuronal populations in the Tg(*lbx1a*:*dTomato*) or Tg(*lbx1a*:*EGFP*) strains. Mosaics from transplantation will yield single labeled neurons sparsely distributed throughout the neural tube and in combination with confocal microscopy will determine exactly where axons project. These results can then be compared to what is known in the mouse to elucidate whether *lbx1+* neurons in the zebrafish neural tube innervate the same types of tissues that *Lbx1+* neurons do in the mouse. Some caveats to this procedure may be a limitation in the amount of fluorescent reporter protein produced in a single neuron and the use of other forms of fluorescent proteins such as membrane localized GFP (Kalejta *et al.*, 1997) will visualize axons better.

d) Effectiveness of Tol2 Technology

One major caveat to making transgenic strains for analysis of an identified regulatory element is the fidelity of the pattern generated by that fragment. Since the Tol2-GW:*cFos* constructs are susceptible to neighbouring promiscuous enhancers, meticulous sorting of embryos must be done before setting aside a potential founder adult. First generation stable transgenic embryos are heterozygous for usually one or multiple insertions and often have variable expression patterns. If inserted transgenic constructs are in non-ideal chromosomal locations, visual inspection by live fluorescence microscopy will inadvertently overlook low expression levels from the activity of weak neighbouring enhancers. Such was the case for the Tg(*lbx1a*:*EGFP*) strain, where heterozygous embryos did not produce noticeable expression in addition to the hindbrain, spinal cord and floor plate. When made homozygous, embryos exhibited EGFP in the ventral diencephalon and mediolateral hindbrain (Figure 5-2). This is likely a result of neighbouring enhancer activity as in comparison to the Tg(*lbx1a*:*dTomato*) strain, which does not exhibit dTomato in those regions. Additionally, the floor plate exhibits elevated transgene expression levels when compared to the Tg(*lbx1a*:*dTomato*) strain, which could be explained by synergism from the enhancer driving ectopic medial hindbrain expression as this region lies in the same plane as the floor plate.

A more precise method to determine the “true” expression pattern generated by any given enhancer element cloned into the Tol2-GW:*cFos* vectors

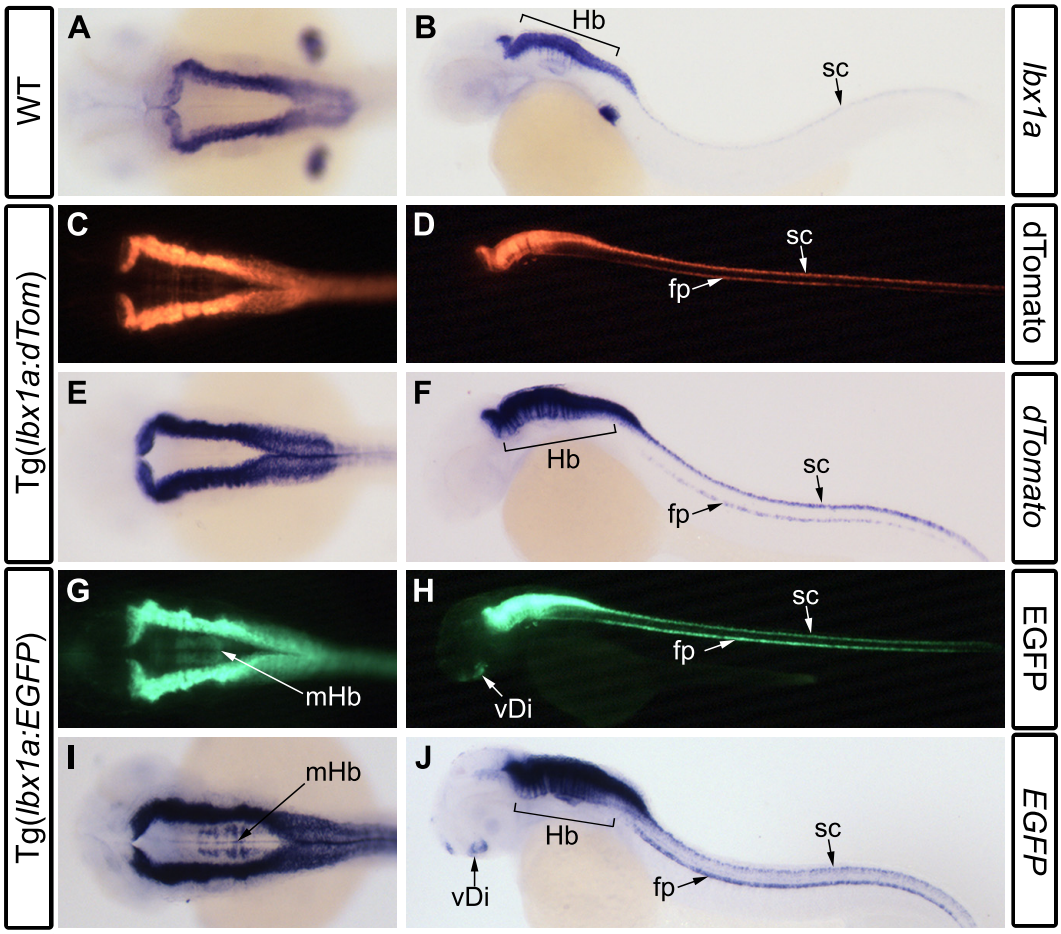

Figure 5-2. **Transgene expression discrepancies between stable Tg(*lbx1a*:*EGFP*) and Tg(*lbx1a*:*dTomato*) zebrafish lines**. The founder Tg(*lbx1a*:*EGFP*) adult yields progeny that have nearly identical expression in the hindbrain (Hb), spinal cord (sc) and floor plate (fp) to Tg(*lbx1a*:*dTomato*) embryos. However, transgene expression in a discrete region of the ventral diencephalon (vDi) and medioventral hindbrain (mHb) is also observed. Embryos were examined by mRNA *in situ* hybridization (A, B, E, F, I, J) and live fluorescence microscopy (C, D, G, H) to confirm *dTomato* (C-F) and *EGFP* (G-J) transgene expression patterns in relation to endogenous *lbx1a* (A, B). Panels show dorsal views of the hindbrain (A, C, E, G, I) or lateral views of the whole embryo (B, D, F, H, J).

is to examine numerous different first generation stable transgenic progeny. If the cloned genomic region within the construct is capable of driving a specific spatiotemporal expression pattern, it should always be visible in conjunction with any patterns generated from promiscuous enhancers in proximity to the inserted construct. By closely examining all of the first generation transgenic progeny for transgene expression, one can get a good idea of the common pattern that is likely created by the cloned enhancer region. This method was applied to the various *lbx1a* enhancer fragments generated in order to determine the real expression pattern. Stable Tg(*lbx1a*$^{457\text{-}1067}$:*EGFP*) embryos recovered are a clear indication of how the construct-driven expression pattern can be distinguished from ectopic transgene expression (Figure 5-3).

Ectopic transgene expression driven by neighbouring enhancers cannot always be uncovered through live fluorescent microscopy since the detection threshold for fluorescent proteins is not as sensitive as mRNA *in situ* hybridization. Several of the *lbx1a* enhancer fragments analyzed displayed a wide assortment of transgene expression patterns, all with a common underlying pattern in the dorsal neural tube. In particular, Tg(*lbx1a*$^{1\text{-}484}$:*EGFP*) stable heterozygous transgenic embryos showed consistent dorsal hindbrain and spinal cord expression with additional expression such as in the somites, optic tectum and forebrain (Figure 5-4). Many Tg(*lbx1a*$^{1\text{-}680}$:*EGFP*) stable transgenic embryos did not exhibit significant levels of ectopic EGFP tissue localization, however, mRNA *in situ* hybridization revealed that several isolated embryos had weak

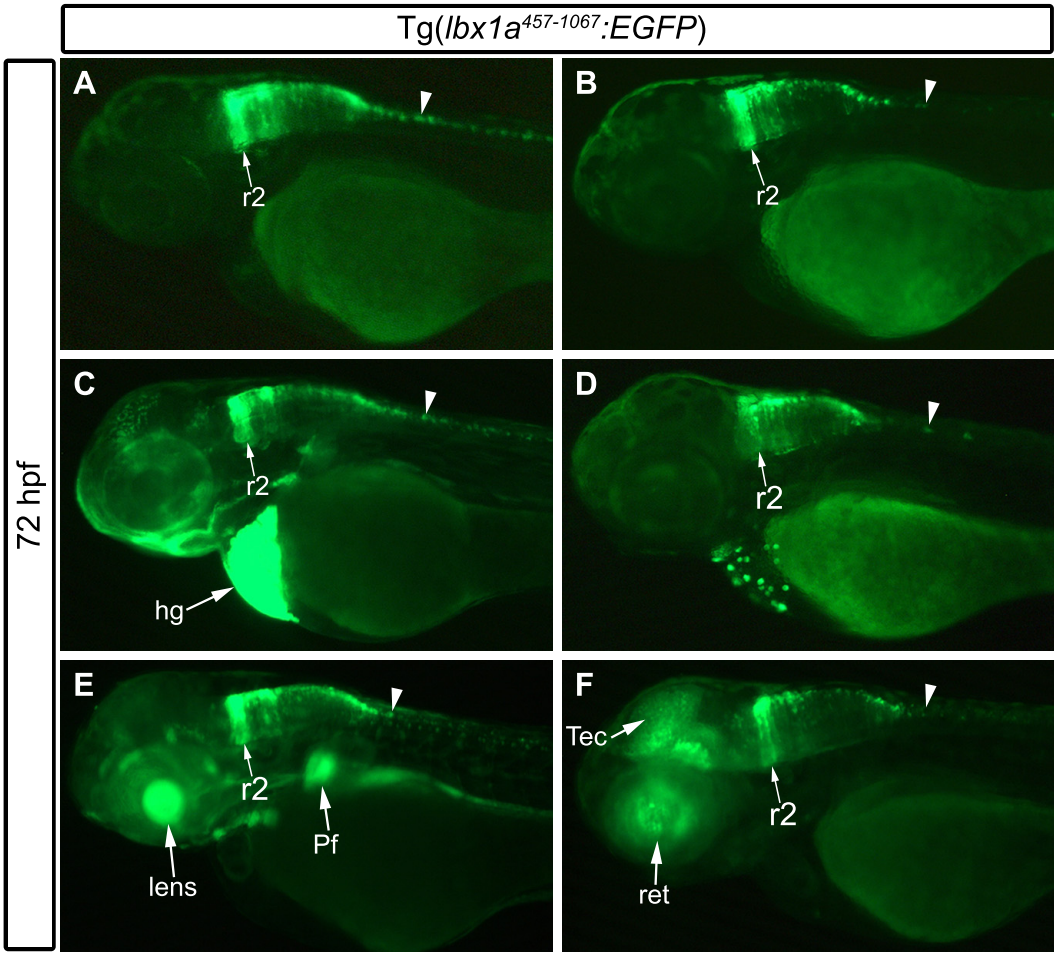

Figure 5-3. **Susceptibility of the Tol2 transgenic construct to nearby enhancers**. The phenomenon of random integration exhibited by the Tol2 construct is displayed above from several independent germline insertions recovered from embryos injected with the pTol2-*lbx1a*$^{457\text{-}1067}$*:cFos:EGFP*. The baseline pattern of strong r2 and dorsal spinal cord (arrowheads) expression generated from the cloned enhancer within the construct is visible in all transgenic embryos isolated, thus facilitating the selection of a true stable reporter zebrafish line without extraneous reporter expression from neighboring promiscuous enhancers. All six embryos shown were isolated from separate adults that were injected at the one-cell stage with pTol2-*lbx1a*$^{457\text{-}1067}$*:cFos:EGFP*. Embryos in panels A and B appear similar and are putative true transgenic reporters as they did not exhibit any additional expression during early development. The embryos in panels C-F also have the pattern shown in A and B, but show additional expression in tissues like the hatching gland (hg), lens, pectoral fin (Pf), optic tectum (Tec) and retina (ret).

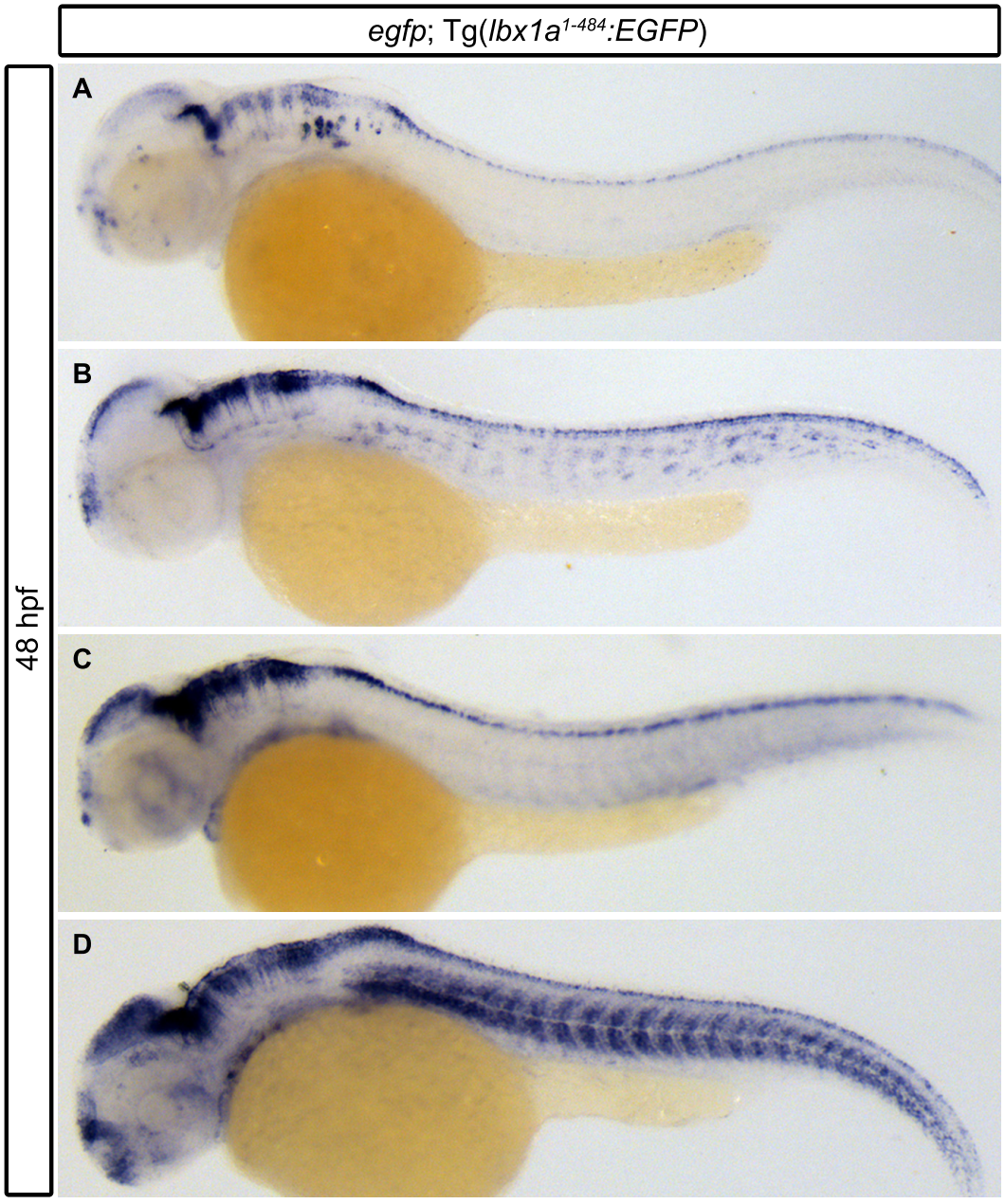

Figure 5-4. **Variability of transgene expression in Tg(*lbx1a*$^{1-484}$*:EGFP*) embryos**. Transient *EGFP* expression driven by the first 484 bp of the *lbx1a* neural enhancer did not yield consistent patterns, therefore, embryos were raised to adulthood and examined for stable transgenic progeny. GFP-positive F1 embryos were fixed at 48 hpf and examined for expression of the *EGFP* transgene by mRNA *in situ* hybridization to identify a common pattern between different stable transgenic embryos. Note that all embryos show a common pattern of dorsal hindbrain and spinal cord expression. All panels have anterior to the left and dorsal up.

EGFP expression in other tissues such as the optic tectum (Figure 5-5). The amount of variability observed from screening Tg(*lbx1a*$^{179\text{-}680}$:*EGFP*) embryos (Figure 5-6) indicates the great efficiency of the Tol2 vector's use in enhancer-trapping as previously reported (Choo *et al.*, 2005). Unfortunately, the ectopic expression driven by neighbouring enhancers is a highly undesirable characteristic when attempting to dissect critical regions of an enhancer.

The Tol2-GW:*cFos* vectors are a straightforward tool that can be used in combination with simple bioinformatic analysis of genomic DNA sequence to create stable transgenic zebrafish embryos. While this is a great method to examine putative enhancer activity, the thorough dissection of an enhancer into smaller components requires much more resources in terms of adult fish care and embryo screening to precisely identify critical regions. As more and more enhancers are identified using this method they can be incorporated into multisite gateway technology (Kwan *et al.*, 2007). Multisite gateway technology offers the modular assembly of independent clones to generate an enhancer-promoter-transgene-tag cassette with much higher ease and efficacy over typical Tol2-GW:*cFos* vectors. A major question that remains is the function of the vertebrate dorsal interneurons which are specified by *Ladybird*. The majority of phenotypic defects in Lbx1-knockout mice occur in muscular development and no apparent neuronal abnormalities exist (Schäfer and Braun, 1999; Gross *et al.*, 2000) although the relatively quick death of newborn pups prevents the assessment of neuronal circuit activity.

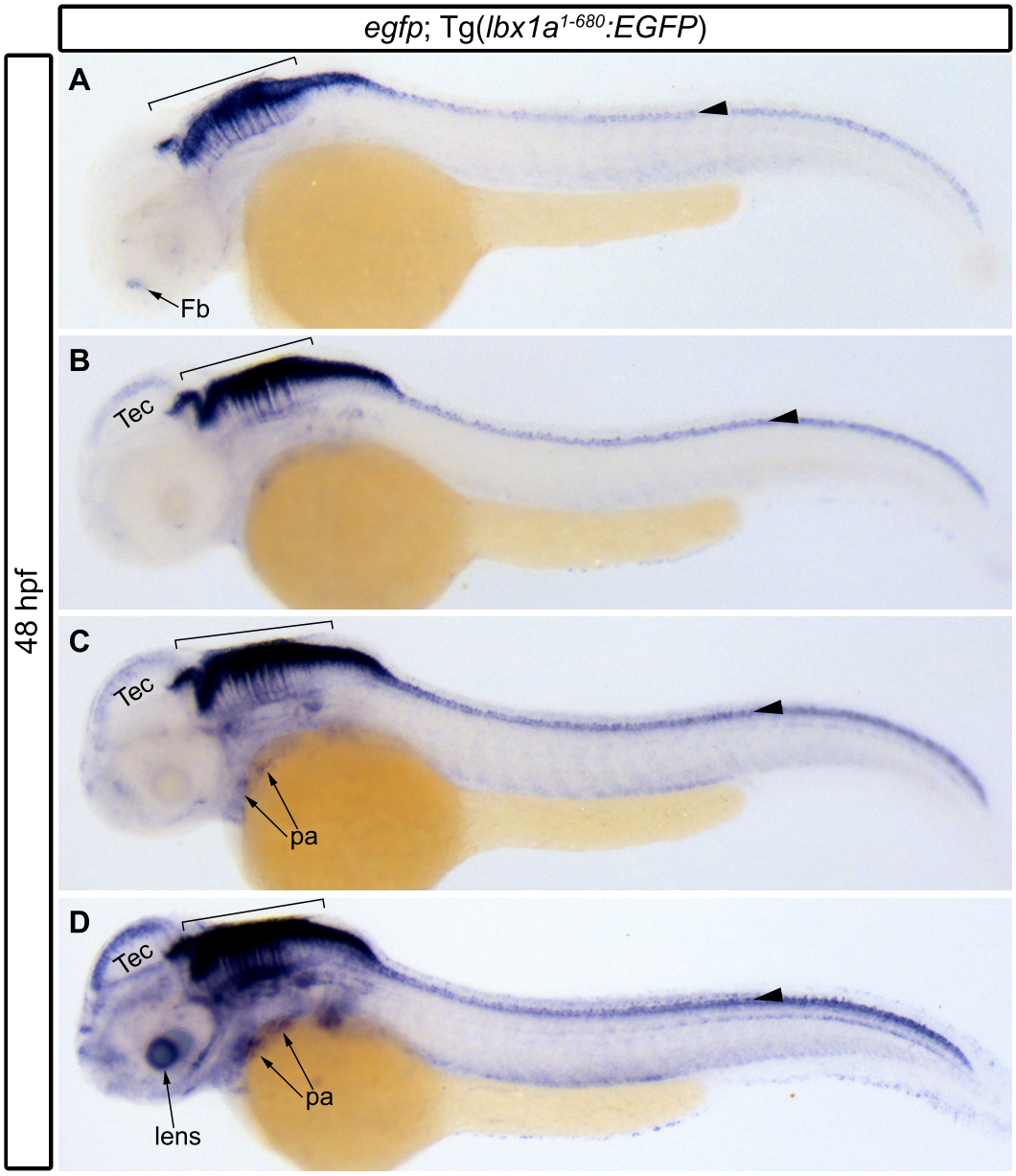

Figure 5-5. **Variability of transgene expression in Tg(*lbx1a*$^{1-680}$*:EGFP*) embryos**. Shown are recovered stable F1 transgenic progeny from different adults that had been injected at the one-cell stage with pTol2-*lbx1a*$^{1-680}$:*cFos*:*EGFP*. The first 680 bp of the *lbx1a* neural enhancer are able to drive *EGFP* expression in the dorsal hindbrain (bracket) and spinal cord (arrowheads) in numerous independent lines isolated. However, the construct is still susceptible to neighbouring enhancers as shown by forebrain (Fb; A), optic tectum (Tec; B-D), pharyngeal arch (pa; C, D) and lens (D) expression. Expression of the *EGFP* transgene was assessed by mRNA *in situ* hybridization to identify a common pattern between different stable transgenic embryos. All panels have anterior to the left and dorsal up.

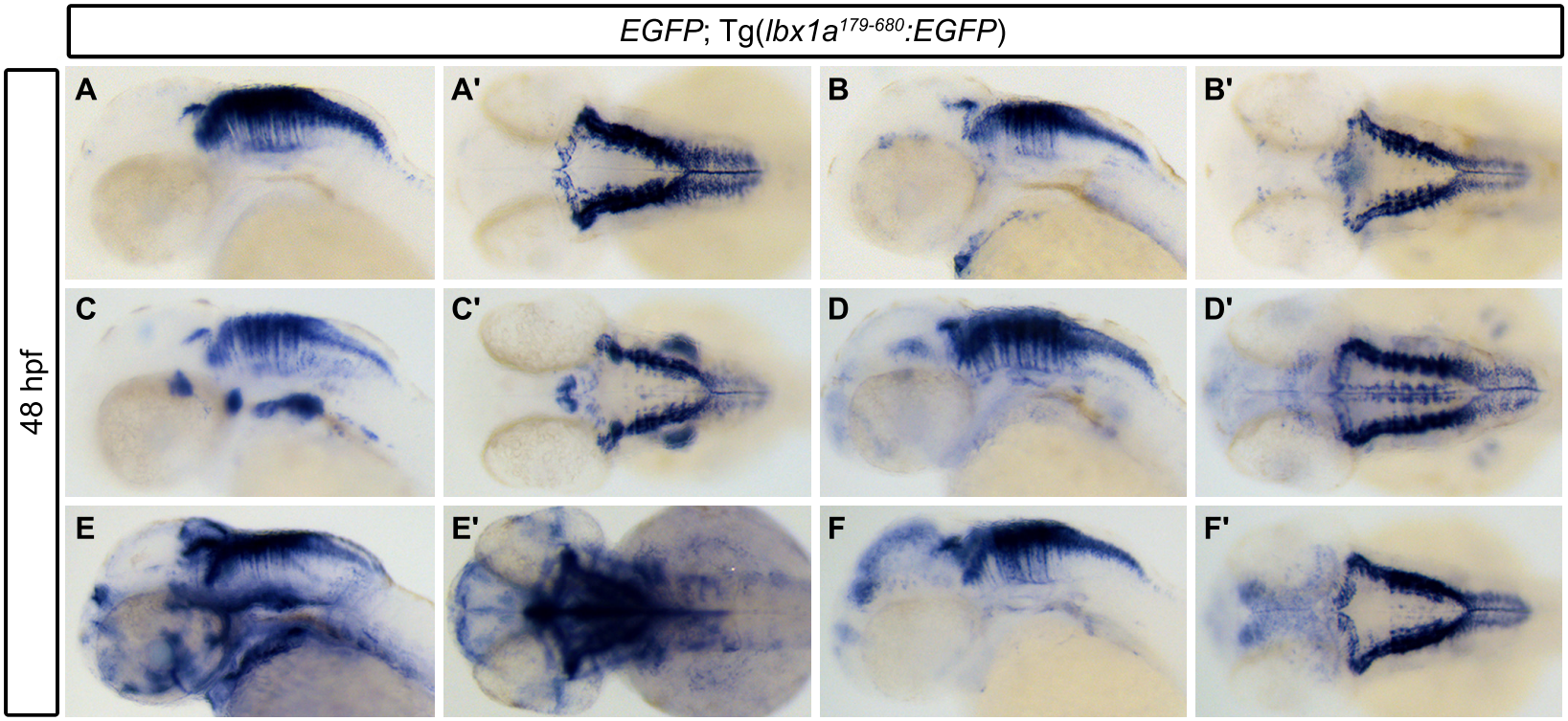

Figure 5-6. **Variability of transgene expression in Tg(*lbx1a*$^{179-680}$*:EGFP*) embryos**. Base pairs 179-680 of the *lbx1a* neural enhancer were cloned into pTol2-GW:*cFos*:*EGFP* and assayed for transient expression. Embryos with strong transient expression were selected, raised to adulthood, and individualy crossed to wild type adults to recover single insertion transgenic F1 progeny. Those progeny were fixed at 48 hpf for mRNA *in situ* hybridization analysis of the *EGFP* transgene. The array of expression patterns shown indicates the susceptibility of the Tol2 transgenic construct to nearby enhancers. Embryos were photographed in lateral (A-F) and corresponding dorsal (A'-F') views.

Conditional ablation of cells within a zebrafish embryo, such as the interneurons expressing *ladybird*, would provide much insight into the functional information these neurons relay. Empty Tol2-GW:*cFos* vector injections show that the *cFos* promoter exhibits ectopic expression in trunk muscle and heart tissue, making it an undesirable vector for introducing cyto-toxicity genes. The bacterial nitroreductase and pro-drug metronidazole system that has been used for tissue regeneration studies in zebrafish (Curado *et al.*, 2007) would be a great system whereby the *lbx1a* neural enhancer would drive cell-specific death which could be controlled in a temporal manner. Examining the development of dorsal interneurons through the use of the Tg(*lbx1a*:*dTomato*) strain and its corresponding enhancer element will greatly facilitate comparative studies to mammalian models as very little is known about interneuron development in teleost fish.

3) References

Choo BG, Kondrichin I, Parinov S, Emelyanov A, Go W, Toh WC and Korzh V, **2006**. Zebrafish transgenic Enhancer TRAP line database (ZETRAP). *BMC Dev Biol*. 6:5.

Clarke J, **2009**. Role of polarized cell divisions in zebrafish neural tube formation. *Curr Opin Neurobiol*. 19:134-138.

Curado S, Anderson RM, Jungblut B, Mumm J, Schroeter E and Stainier DY, **2007**. Conditional targeted cell ablation in zebrafish: a new tool for regeneration studies. *Dev Dyn*. 236:1025-1035.

Froschauer A, Braasch I and Volff JN, **2006**. Fish Genomes, Comparative Genomics and Vertebrate Evolution. *Current Genomics*. 7:43-57.

Galas DJ and Schmitz A, **1978**. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res*. 5:3157-3170.

Guner B and Karlstrom RO, **2007**. Cloning of zebrafish nkx6.2 and a comprehensive analysis of the conserved transcriptional response to Hedgehog/Gli signaling in the zebrafish neural tube. *Gene Expr Patterns*. 7:596-605.

Hoegg S, Boore JL, Kuehl JV and Meyer A, **2007**. Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish Astatotilapia burtoni. *BMC Genomics*. 8:317.

Jagla K, Georgel P, Bellard F, Dretzen G and Bellard M, **1993**. A novel homeobox nkch4 gene from the Drosophila 93E region. *Gene*. 127:165-171.

Jagla T, Bellard F, Lutz Y, Dretzen G, Bellard M and Jagla K, **1998**. ladybird determines cell fate decisions during diversification of Drosophila somatic muscles. *Development*. 125:3699-3708.

Meng X, Noyes MB, Zhu LJ, Lawson ND and Wolfe SA, **2008**. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol*. 26:695-701.

Meyer A and Van de Peer Y, **2005**. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays*. 27:937-945.

Muller T, Brohmann H, Pierani A, Heppenstall PA, Lewin GR, Jessell TM and Birchmeier C, **2002**. The homeodomain factor lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron*. 34:551-562.

Nguyen VH, Trout J, Connors SA, Andermann P, Weinberg E and Mullins MC, **2000**. Dorsal and intermediate neuronal cell types of the spinal cord are established by a BMP signaling pathway. *Development*. 127:1209-1220.

Wotton KR, Weierud FK, Dietrich S and Lewis KE, **2008**. Comparative genomics of Lbx loci reveals conservation of identical Lbx ohnologs in bony vertebrates. *BMC Evol Biol*. 8:171.

Yu PB, Hong CC, Sachidanandan C, Babitt JL, Deng DY, Hoyng SA, Lin HY, Bloch KD and Peterson RT, **2008**. Dorsomorphin inhibits BMP signals required for embryogenesis and iron metabolism. *Nat Chem Biol*. 4:33-41.

Appendix: Creation of Tg(*insm1a*:*EGFP*) Transgenic Zebrafish

1) Introduction

Gateway technology is an excellent means to efficiently transfer cloned sequences from a donor plasmid into a suite of recipient vectors, each having their own unique features and properties. One such recipient vector, pTol2-GW, contains Tol2 recombination sites flanking a construct containing a gateway (GW) cassette upstream of a minimal *cFos* promoter linked to a fluorescent reporter transgene (Fisher *et al.*, 2006; Kawakami, 2007). Enhancer analysis using such vectors is accomplished by recombining sequences from donor plasmids, via the *attR* and *attL* sites in the gateway cassette, upstream of the 255 bp *cFos* promoter. While the *cFos* promoter can be driven by putative enhancer sequences, it is unclear how putative promoter sequences cloned into the gateway cassette will behave in tandem with the *cFos* promoter. Multiple promoter elements may facilitate promoter competition that could interfere with tissue-specific expression of the transgene. RNA polymerase complex formation could potentially be initiated at 2 different positions and generate undesirable transcripts that may enter the non-sense mediated decay pathway (Chang *et al.*, 2007). The aberrant RNAs produced may prevent efficient translation of the reporter coding sequences such that transgene expression is difficult to detect.

The position of the gateway cassette in the pTol2-GW:*cFos:EGFP* vector allows the entire gateway cassette and minimal *cFos* promoter to be excised and replaced with any putative promoter, using XhoI and BamHI. The promoter to be inserted is PCR amplified with forward primer containing a XhoI restriction site

and reverse primer containing a BamHI restriction site designed into the sequence. The subsequent fragment can act as the sole promoter in the transgenic construct to drive *EGFP* expression. The gateway cassette of the pTol2-GW:*cFos:dTomato* vector is in the opposite orientation, which places an EcoRI site near the *dTomato* coding sequence so that any putative promoter sequences amplified with a forward primer having a XhoI site and a reverse primer having an EcoRI site, can replace the minimal *cFos* promoter and gateway cassette. To examine the ability of endogenous promoter sequences in driving transgene expression from the pTol2-GW constructs, the well-characterized promoter region of *insulinoma-associated 1* (*insm1*) was cloned from zebrafish (Breslin *et al.*, 2003; Pedersen *et al.*, 2006).

The mammalian *Insm1* promoter contains several conserved regulatory E-boxes upstream of the transcription start site that have the ability to drive reporter expression (Breslin *et al.*, 2003). *Insm1* is expressed in neuronal progenitors, the pancreas and the eye during early vertebrate development (Breslin *et al.*, 2003; Lukowski *et al.*, 2006) and a single E-box is thought to drive the majority of tissue-specific expression by recruiting a heterodimeric transcription factor complex consisting of the basic helix-loop-helix proteins, NeuroD and E47. This E-box and neighbouring genomic sequences, as well as several other short distal sequences in the *Insm1* proximal promoter, are highly conserved in vertebrates. Using the zebrafish *insm1a* locus, 695 bp of proximal promoter sequences were amplified by PCR with forward (5'-GTC TCG AGG AGA CCG AGA TGA GTC

CGT TTG-3') and reverse (5'-GTG GAT CCA GCT GAA AGG CAC TTC AGT CGG-3') primers. The indicated BamHI and XhoI restriction sites in each primer facilitated ligation of the insert into pTol2-GW:*cFos:EGFP* to replace the gateway cassette and minimal *cFos* promoter. Zebrafish embryos were injected with pTol2-GW:*insm1a:EGFP* and *Tol2* transposase mRNA at the 1-cell stage and examined at later stages of development for transgene expression. *EGFP* expression in stable transgenic progeny was compared to endogenous *insm1a* expression in zebrafish through mRNA *in situ* hybridization and immunofluorescence.

2) Results and Discussion:

a) Genomic Sequence Analysis of Vertebrate *Insm1* Loci

Insm1 is a zinc finger transcription factor that is highly conserved among vertebrates (Lukowski *et al.*, 2006) and has homologs in invertebrates such as *C. elegans* and *Drosophila*. A single exon encodes the entire protein, which ranges in size from as large as 521 amino acids in mouse to 383 amino acids for zebrafish Insm1a. The protein family shares conserved C_2H_2 zinc fingers which are thought to regulate neurogenic genes throughout pro-neural domains within the developing embryo. The high degree of sequence conservation and similarity in expression patterns suggests conserved regulatory elements exist near the *Insm1* locus. Analysis of genomic sequences surrounding vertebrate *Insm1* loci revealed several conserved regions upstream of transcript sequences while no conservation

was observed downstream of the 3' UTR (Figure A-1, A). A well-characterized E-box sequence in the proximal promoter seems to drive the majority of *Insm1* expression as the pro-neural protein, NeuroD and its co-activator, E47, have been shown to bind there (Breslin *et al.*, 2003). Out of all the conserved sequences amongst the various vertebrate *insm1* homologs and paralogs, this E-box sequence, and its position relative to the transcription start site, is conserved between all *insm1* loci, including zebrafish *insm1a* (Figure A-1, B).

Available vertebrate genomic sequence at the *Insm1* locus was analyzed for conservation using a variety of alignment parameters to identify highly conserved proximal promoter elements and, potentially, distal enhancers. The genome duplication in teleost fish species allows for a detailed view into the maintenance or loss of certain regulatory regions between paralogous genes. Predicted open reading frames for *insm1a* and *insm1b* were identified in Stickleback, *Takifugu rubripes* and *Tetraodon nigroviridis*, however, *insm1*-like genes have yet to be found in available Medaka genomic sequences. For *insm1*, it appears as though zebrafish paralogs have lost or retained a very different set of conserved intergenic regions when compared to other fish species. Compared to mammalian *Insm1*, the conservation seen at fish *insm1a* and *insm1b* loci is very high, but homologs between zebrafish and other teleost fish, share very little similarity in putative promoter, distal enhancer sequences, or untranslated regions (UTR). The zebrafish *insm1a* proximal promoter showed the highest level of sequence conservation (Figure A-1, C) and was cloned into pTol2-

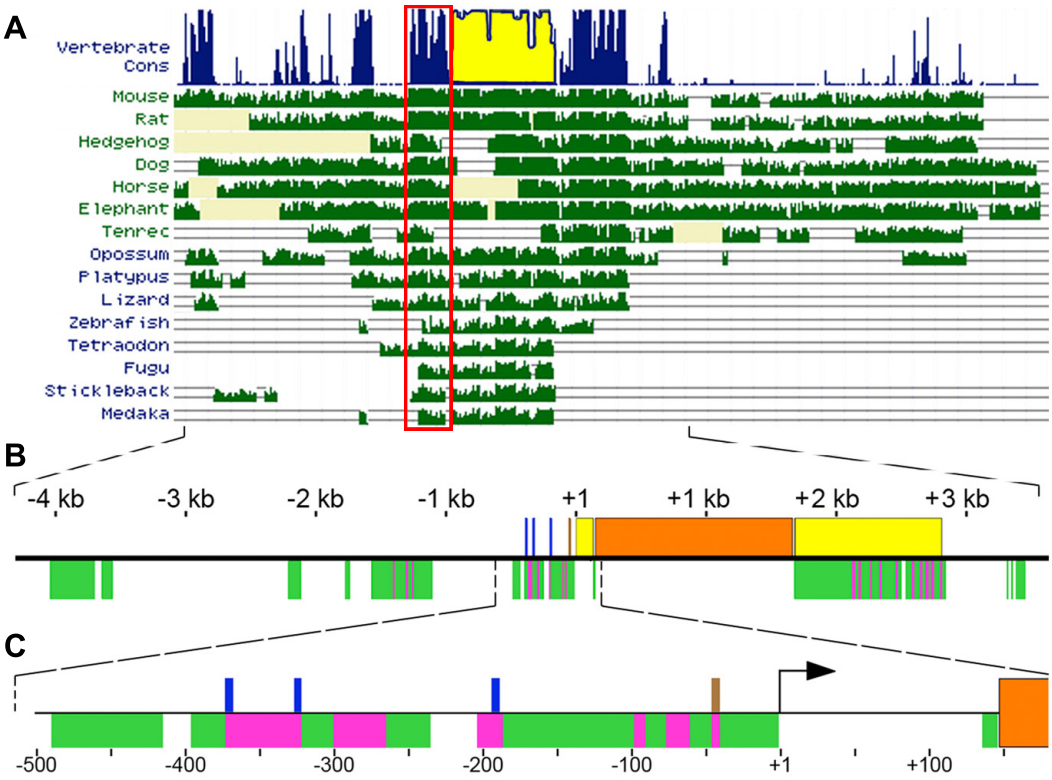

Figure A-1. **Analysis of genomic sequence surrounding the Human *INSM1* locus**. Genomic analysis of the human *INSM1* locus using the BLAT algorithm (A) and nearby intergenic sequence demonstrates several highly conserved regions as a histogram (navy) with a single exon (yellow) and pairwise species comparison to human (green). Proximal promoter sequences are boxed in red and were analyzed further for putative regulatory elements. More detailed analysis of sequence conservation (B) reveals the regions shared between humans and mammals (green) or between humans and zebrafish (magenta). Characterized E-boxes (blue), a conserved TATA box (brown), UTR (yellow) and coding sequences (orange) are shown above. A magnified view of the proximal *INSM1* promoter (C) shows zebrafish *insm1a* contains all three E-boxes as well as the putative TATA box found in mammalian species. The start of transcription is indicated by +1 (arrow) and distances relative to +1 are indicated in kbp (B) or bp (C).

GW:*cFos*:*EGFP* via replacement of the entire minimal *cFos* promoter and gateway cassette, leaving the *insm1a* promoter to drive *EGFP*.

b) Identifying Neuronal Regulatory Elements of the *Insm1* Promoter

Transient expression of *EGFP* in zebrafish embryos driven by the proximal zebrafish *insm1a* promoter was assessed by co-injection of purified transgenic construct plasmid DNA and *Tol2* mRNA into 1-cell zebrafish embryos (Figure A-2, A and B). Remarkable expression starting very early in embryonic development was observed throughout the central nervous system in pro-neural domains and followed a pattern similar to endogenous *insm1a*. The onset of transgene expression was slightly delayed in relation to endogenous transcripts, however, the dynamic expression pattern was mimicked by transgene expression. The presence of EGFP persisted in neurons well beyond that of gene expression, labeling the central nervous system for up to 2 weeks into embryonic development. The stability of EGFP allows for examination of these neurons in late stages of development. Stable transgenic progeny were raised to adulthood to act as founders for the Tg(*insm1a*:*EGFP*) zebrafish strain.

Immunofluorescence with an anti-GFP antibody reveals a pan-neural pattern of transgene expression (Figure A-2), namely in the olfactory bulb, hindbrain and throughout the spinal cord. Some tissues, such as the optic tectum, exhibit much more persistent visualization of EGFP, likely due to the stable nature of the encoded protein. Thought of as a pro-neural gene (Lukowski *et al.*,

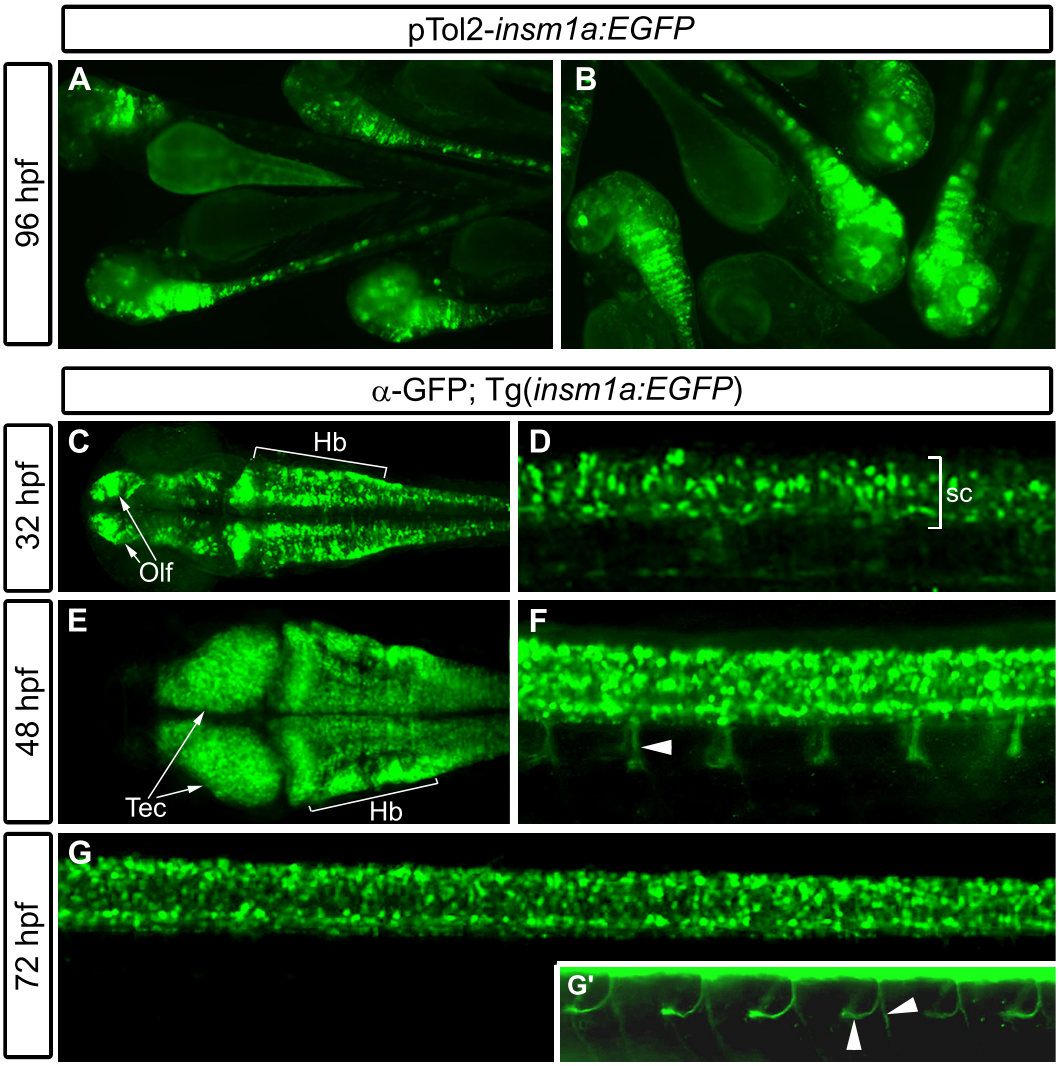

Figure A-2. **Transgenic zebrafish expressing *EGFP* driven by the *insm1a* promoter**. One-cell wild type AB zebrafish embryos were injected with Tol2-mRNA and pTol2-*insm1a:EGFP* and observed for consistent transient reporter expression (A, B). The transgenic construct contains 695 bp upstream of the zebrafish *insm1a* start codon linked to an EGFP open reading frame, which was accomplished through replacement of the gateway cassette and minimal *cFos* promoter in pTol2-GW:*cFos:EGFP*. Transgenic embryos were raised to adulthood and stable progeny were examined for the presence of GFP at 32 hpf (C, D), 48 hpf (E, F) and 72 hpf (G, G') by whole mount immunofluorescence. As expected, the distribution of GFP follows a pro-neural pattern, most notably in the olfactory bulb (Olf), optic tectum (Tec), hindbrain (Hb) and spinal cord (sc). GFP is also detected in the axons (arrowheads) of secondary motor neurons in the spinal cord, however, GFP is not detected in the pancreas nor the retina, two tissues that express high levels of *insm1a*. Panels C-G' have anterior to the left with dorsal (C, E) or lateral (D, F, G) views. Inset G' is an higher exposure image of panel G.

2006), it is not surprising the entire dorsoventral axis of the spinal cord is labeled with GFP from the *insm1a* proximal promoter (Figure A-2, F and G). The promoter is able to drive *EGFP* expression so well that axonal projections of spinal cord motor neurons are clearly visible in live embryos and, upon fixation, can still be visualized with an anti-GFP antibody (Figure A-2, G'). Interestingly, no transgene expression was observed in the pancreas or the eye, two tissues that robustly express endogenous *insm1a*. Several conserved regions upstream of the cloned promoter (Figure A-1, A) could function as distal enhancers that drive gene expression in the eye and pancreas as it has been shown in mice that 1.7 kbp upstream of the *Insm1* locus is enough to drive near complete tissue specific expression of *lacZ* (Breslin *et al.*, 2003).

In situ hybridization for *EGFP* mRNA in these stable Tg(*insm1a*:*EGFP*) transgenic embryos gave insight into the ability of the cloned 695 bp *insm1a* proximal promoter to respond dynamically during development much like endogenous *insm1a* expression. While visualization of fluorescent protein suggests the promoter acts on the transgene in much the same way as the native *insm1a* promoter, it was interesting to observe the latency of *EGFP* expression. As with all fluorescent reporters, it takes time for the protein to mature and become concentrated enough to be visualized under the microscope. *In situ* hybridization is far more sensitive and can detect reporter gene expression prior to the detection of fluorescent protein in a live embryo. In Tg(*insm1a*:*EGFP*) zebrafish embryos, *EGFP* mRNA is detected in a pan-neural pattern with an

apparent delay when comparing embryos of the same stage (Figure A-3). This is not unexpected as the transgene likely resides in a chromosomal environment different to that of endogenous *insm1a* and it gene transcription is not as efficient. Additionally, distal enhancer elements that could potentially enhance the rate of transcription are not present in the vicinity of the transgene

These results show that the replacement of the gateway cassette and minimal *cFos* promoter within the Tol2-GW series of vectors is a viable option for examining promoter sequences. The pattern of transgene expression under the control of the zebrafish *insm1a* proximal promoter appears to be more similar to that of *insm1b* as it is not expressed in the eye and is very weakly expressed in the pancreas. Further analysis of conserved sequences at the zebrafish *insm1a* locus may uncover an eye-specific enhancer, which may have been lost at the *insm1b* locus.

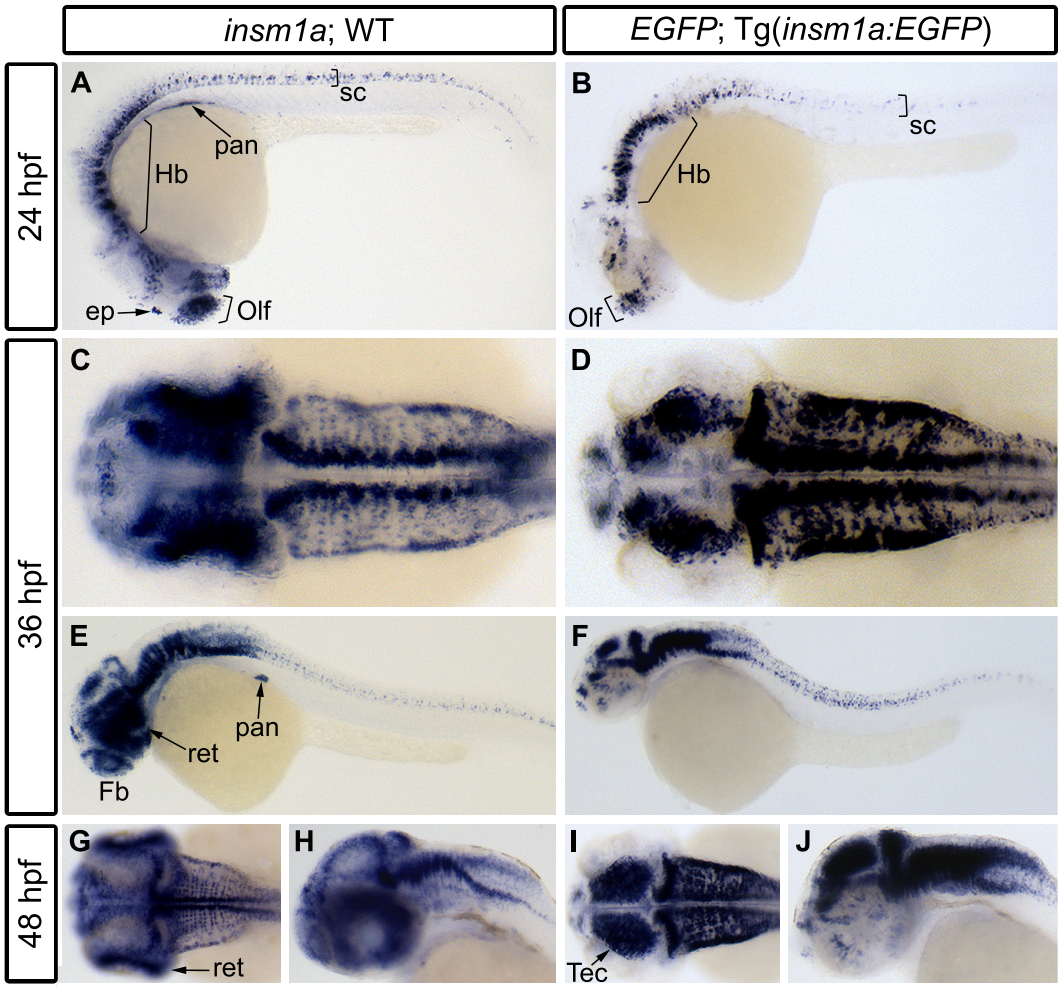

Figure A-3. **Pan-neural *EGFP* expression in Tg(*insm1a:EGFP*) transgenic embryos**. *In situ* hybridization for *EGFP* mRNA more accurately reflects the activity of the proximal 695 bp zebrafish *insm1a* promoter. Transgene expression is slightly delayed at 24 hpf (A, B) when compared to endogenous *insm1a*, however, *EGFP* is expressed in a pro-neural pattern within key regions of the embryo. The olfactory bulb (Olf), ventral hindbrain (Hb) and spinal cord (sc) are regions of early *insm1a* expression. At 36 hpf (C-F) transgene expression resembles the pattern of endogenous *insm1a* except for parts of the forebrain (Fb), retina (ret) and pancreas (pan). Slight differences persist at 48 hpf (G-J), namely in the optic tectum (Tec) as it is an area of transient *insm1a* expression, yet *EGFP* expression appears more stable and persistent in this tissue. The epiphysis (ep), pancreas and retina are tissues with high levels of endogenous *insm1a*, however, the regulatory elements that drive expression in those tissues appear to lie outside of the cloned 695 bp proximal promoter. All panels have anterior to the left with lateral (A, B, E, F, H, J) or dorsal (C, D, G, I) views.

3) References

Breslin MB, Zhu M and Lan MS, **2003**. NeuroD1/E47 regulates the E-box element of a novel zinc finger transcription factor, IA-1, in developing nervous system. *J Biol Chem*. 278:38991-38997.

Chang YF, Imam JS and Wilkinson MF, **2007**. The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem*. 76:51-74.

Kawakami K, **2007**. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol*. 8 Suppl 1:S7.

Lukowski CM, Ritzel RG and Waskiewicz AJ, **2006**. Expression of two insm1-like genes in the developing zebrafish nervous system. *Gene Expr Patterns*. 6:711-718.

Pedersen N, Pedersen MW, Lan MS, Breslin MB and Poulsen HS, **2006**. The insulinoma-associated 1: a novel promoter for targeted cancer gene therapy for small-cell lung cancer. *Cancer Gene Ther*. 13:375-384.