IDENTIFICATION OF GENE REGULATORY ELEMENTS ASSOCIATED WITH
THE *MEIS* FAMILY OF HOMEOBOX GENES

A Thesis
by
KYLE CHRISTOPHER NELSON

Submitted to the Graduate School
Appalachian State University
In partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

August 2011
Department of Biology

IDENTIFICATION OF GENE REGULATORY ELEMENTS ASSOCIATED WITH
THE *MEIS* FAMILY OF HOMEOBOX GENES

A Thesis
by
KYLE CHRISTOPHER NELSON
August 2011

APPROVED BY:

_____
Dr. Ted Zerucha
Chairperson, Thesis committee

_____
Dr. Susan L. Edwards
Member, Thesis Committee

_____
Dr. Annkatrin Rose
Member, Thesis Committee

_____
Dr. Steven W. Seagle
Chairperson, Department of Biology

_____
Dr. Edelma D. Huntley
Dean, Research and Graduate Studies

FORWARD

The references, tables, and figures within this manuscript were prepared in accordance to the author submission requirements of *Cell*, the peer-reviewed journal published by Cell Press, a division of Elsevier, housed in the editorial offices in Cambridge Massachusetts.

ABSTRACT

IDENTIFICATION OF GENE REGULATORY ELEMENTS ASSOCIATED WITH
THE *MEIS* FAMILY OF HOMEOBOX GENES

Kyle Christopher Nelson, B.A., Appalachian State University

M.S., Appalachian State University

Chairperson: Ted Zerucha

Homologs of the *Meis* homeobox-containing gene family (originally named for myeloid ecotropic leukemia virus integration site because a disruption of the first member of this gene family discovered was found to lead to Leukemia) have been identified in all animals studied. The products of the *Meis* genes appear to function as cofactors, directly interacting with other transcription factors as well as DNA to facilitate transcriptional regulation. Most notably, they appear to act as co-factors of the evolutionarily well-conserved Hox proteins and have also been described as acting with members of other homeobox genes. The vertebrate *Meis* homeobox-containing gene family consists of at least three members, and while little to nothing is known about their regulation, they are expressed in conserved patterns throughout the embryonic development of those vertebrates that have been examined. Using comparative genomics/phylogenetic footprinting to search for regulatory elements associated with the *Meis* family of homeobox-containing genes, 4 highly conserved elements located downstream of the *Meis2* gene have been identified that are very well-conserved in sequence and relative

position amongst the genomes of all vertebrates examined, including human, mouse, chicken, zebrafish and the pufferfish *Takifugu rubripes*. All elements, named m2de1 (<u>M</u>eis <u>2</u> <u>D</u>ownstream <u>E</u>lement), m2de2, m2de3, and m2de4, contain several putative transcription factor binding sites. Conservation in sequence and position indicate the m2de1, m2de2, m2de3, and m2de4 elements play some important conserved role in animals with the most likely possibility being that they function as *cis*-regulatory elements.

DEDICATION

This thesis is dedicated to my future stepdaughter, Reesa Faith Devers, who has taught me what it means to truly love, what sacrifice really means, and what is important in life. My eternal thanks to her for accepting me into her life with open arms.

ACKNOWLEDGEMENTS

It is a pleasure to thank all of those who made this thesis possible. I would first like to thank my advisor Dr. Ted Zerucha for all of the invaluable advice and guidance he gave me as both an undergraduate and graduate student. I would also like to thank my committee members Dr. Susan Edwards and Dr. Annkatrin Rose for their assistance though out my graduate work. In addition, I would like to thank all of my fellow lab mates Brandon Carpenter, Caroline Cochrane, Brantley Graham, and Cody Barrett for making my experience at Appalachian State University a most enjoyable one.

I would like to show my gratitude to Jim Sobieraj, Dr. Ece Karatan, Dr. Mary Connell, Dr. Maryam Ahmed, and Dr. Jeff Butts for the countless amount of time they provided me with an ear and assistance with troubleshooting. I am also deeply grateful to AJ and Jackie Nelson, my parents John and Fran Nelson, as well as my grandparents and extended family, for without their never ending support this thesis would not have been possible. My utmost gratitude is also extended to Dr. Coleman McClenegan, Dr. Ken Shull, Dr. Richard Henson, and Dana Danger because their professional influences in my life have directed me toward my current path. I would also like to express my gratitude to Gary Sauls, Brother Gary Cregan, and Bethel Gerald, for without their influences in my life I would not be the person I am today nor in the position that I am.

Finally, but most importantly, I would like to thank the loves of my life, Mary Jane Carmichael and Reesa Faith Devers for their unwavering support. Without them my life would have no purpose.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The field of Evolutionary Developmental Biology (Evo-Devo) is a result of the merger of ideas from Evolutionary Genetics and Developmental Genetics (Carroll, 2005a). In the beginning of the 20th century Evolutionary Biology and Genetics were distinct fields of study and began to merge as biologists began to recognize that each field was related to the other. In the 1940s the Modern Synthesis used paleontological, systematic, and genetic information to explain how natural selection could drive the gradual evolution of current complex forms from simpler forms observable in the fossil record. Largely due to the seminal works of Julian Huxley in 1942 (*Evolution: The Modern Synthesis*) and Theodosius Dobzhansky (Towards a Modern Synthesis) in the journal *Evolution* in 1949, as well the work of Ernst Mayr and George Gaylord Simpson, the ides of Evolutionary Biology and Genetics were merged creating modern Evolutionary Biology and the field of Evolutionary Genetics (Carroll, 2009; Dobzhansky, 1949; Huxley, 1943). Because of the advent of Evolutionary Genetics, evolutionary theory began to push forward and address more difficult issues. The field, however, largely ignored Developmental Biology for many decades.

During the 1970s and 1980s many advances in technology pushed the field of Developmental Biology forward as well. The majority of studies in Developmental Biology focused on morphological development: the patterning of the embryo, identifying embryonic morphological structures, organizational regions, and following cell migrations. However, rapid advancement of the field was hampered by the

misconceived notion that increased complexity of the organism must be correlated to an increased number of genes, compounded by an inability to easily study the actual nucleotide sequence of DNA (Carroll et al., 2008). This idea was challenged in 1975 when Mary-Claire King and Allan Wilson published a paper that demonstrated almost no difference between proteins of Humans and Chimpanzees (King and Wilson, 1975). Because of their findings King and Wilson hypothesized that the evolutionary difference between humans and chimps was a result of changes in gene regulation and not due to gene mutation or increased gene number.

The idea that increased complexity was related to increased gene numbers was permanently dispelled in the 1980s and early 1990s. During this time period it was shown that bilaterian embryonic development was controlled by a specific set of genes, often called "tool-kit genes" (Carroll, 2006; Carroll et al., 2001). One of the most surprising observations during this time period was that all bilaterians shared the same group of tool-kit genes, so the question then became what makes one animal different from another (Carroll, 2005b; Carroll, 2006)? To address this question, in the 1980s the technologies and principles of Developmental Biology and Genetics were combined forming the field of Developmental Genetics and it was discovered that the major difference between bilaterians was not the genes that pattern their development. Instead it was found that the difference between bilaterians is when and where these tool-kit genes were used during embryonic development (Carroll, 2005b; Carroll et al., 2008).

The discovery that differences in bilaterian form was the result of altering the regulation of the same tool-kit genes had several implications. First, it proved that the hypothesis put forth by King and Wilson in 1975 was correct, and it was for the most part

differences in the regulatory mechanisms controlling the genes, not the genes themselves, that explain the morphological differences between different bilaterian organisms. Second, these discoveries demonstrated that the evolution of form occurs by tinkering with the genes that control development and not the creation of new genes. Third, the focus of Evolutionary Biology was shifted away from strictly looking at the genes themselves and toward investigating how developmentally regulated genes are controlled. Thus, in the mid 1990s the field of Evo-Devo was formed by merging the fields of Evolutionary Genetics and Developmental Genetics (Carroll, 2005a).

Since its inception Evo-Devo has facilitated an exponential growth in our understanding of the processes of both evolution by natural selection and embryonic development. With the availability of fully annotated and sequenced genomes from multiple organisms it has become possible to not only employ comparative developmental studies, but also to compare their genomic sequences (Allende et al., 2006; Carroll, 2005a; Carroll et al., 2008). By studying the regulatory mechanisms that control embryonic development in comparative studies, the field of Evo-Devo can in turn more efficiently study and elucidate the mechanisms that drive the evolution of form.

LITERATURE REVIEW

Embryonic development in bilaterians is a complex, organized, and highly regulated process. During this process body axes are established, and the body plan is determined through a complex series of molecular events. Some of these events are controlled, at least in part, by homeobox genes. Homeobox genes are "tool-kit genes" that derive their name from homeotic mutations. Discovery of homeobox genes in the fruit fly *Drosophila melanogaster* was facilitated by their resulting homeotic mutations. First described in 1894, homeotic mutations result in the transformation of one region of an organism into the identity of another: for example, ectopic expression of the *Antennapedia* (*Antp*) gene in *Drosophila* (fruit fly) larva redesignated antennae into second legs (Gehring, 1987; McGinnis et al., 1984a; Schneuwly et al., 1987). Through the study of several homeotic genes in *Drosophila*, including *Antp*, a conserved 180 base pair sequence of DNA shared by these homeotic genes was discovered in 1983 and was named the Homeobox (McGinnis et al., 1984a; McGinnis et al., 1984b; Scott and Weiner, 1984). The Homeobox sequence has been shown to be conserved in all animals studied to date. It was determined that the 180 base pair Homeobox sequence codes for a conserved 60 amino acid domain termed the Homeodomain (HD) that consists of a helix-loop-helix DNA-binding motif (Gehring, 1987; Goulding and Gruss, 1989; McGinnis et al., 1984a; Scott and Weiner, 1984).

The determination of the anterior/posterior (A/P) body axis is controlled in large part by a subfamily of Homeobox genes called *Hox* genes that have been identified in all bilaterian animals examined to date (Deschamps and van Nes, 2005; Gehring, 1987; Goulding and Gruss, 1989; Kessel, 1992; Kessel and Gruss, 1991; Lemons and McGinnis, 2006; Ogishima and Tanaka, 2007; Prince et al., 1998; Santini et al., 2003; Schneuwly et al., 1987). *Hox*-like genes have also been isolated from several species of Cnidarians indicating that these genes have been evolutionarily conserved for hundreds of millions of years (Finnerty and Martindale, 1999). *Hox* genes encode proteins that act as transcription factors and are expressed in restricted segments of the developing embryo functioning to activate regionally specific genes in a regimented manner determining A/P body axis pattern (Deschamps and van Nes, 2005; Gehring, 1987; Goulding and Gruss, 1989; Kessel, 1992; Kessel and Gruss, 1991; Lemons and McGinnis, 2006; Ogishima and Tanaka, 2007; Prince et al., 1998; Santini et al., 2003; Schneuwly et al., 1987). *Hox* genes are also responsible for the Proximal/Distal (P/D) patterning of the limb (Ahn and Ho, 2008). Due to their vital roles in body axis determination it has been hypothesized that alteration of *Hox* gene expression patterns could serve as a powerful evolutionary mechanism, which is supported by studies examining the function of *Hox* gene function in the development of many organisms (Ahn and Ho, 2008; Duboule, 1993; Gendron-Maguire et al., 1993; Salsi et al., 2008).

In extant bilaterians *Hox* genes are grouped into clusters within the genome and are often expressed colinearly according to their chromosomal arrangement (Amores et al., 1998; Duboule, 2007; Prince et al., 1998). Each cluster consists of an anterior, central, and posterior region where the 3′ genes are expressed anteriorly and earlier while

the more 5′ genes are expressed later and in more posterior regions (Amores et al., 1998; Lemons and McGinnis, 2006; Ogishima and Tanaka, 2007; Prince et al., 1998; Santini et al., 2003). Most vertebrates have four *Hox* clusters (A, B, C, D) on different chromosomes consisting of 39 individual genes (Duboule, 2007; Goulding and Gruss, 1989; Lemons and McGinnis, 2006). Teleosts have additional *Hox* clusters due to a proposed genome duplication event after their divergence from the tetrapod lineage (Amores et al., 1998; Prince et al., 1998). For example, the zebrafish (*Danio rerio*) has 7 *Hox* clusters consisting of 49 genes. Each *Hox* cluster contains a set of genes that are categorized into 13 paralog groups (ex: *hox*B1, *hox*B2) depending on their relationship to the *Hox* genes of *Drosophila* (Amores et al., 1998; Gehring, 1987; Goulding and Gruss, 1989; Lemons and McGinnis, 2006; Prince et al., 1998; Prohaska and Stadler, 2004).

The Hox transcription factors function by binding specific DNA sequences located within *cis*-regulatory elements associated with their respective target genes (Goulding and Gruss, 1989; Otting et al., 1990; Waskiewicz et al., 2001). It has been demonstrated that this binding can be inefficient and that Hox proteins require the help of other proteins called cofactors to assist in DNA binding (Moens and Selleri, 2006; Prince et al., 1998). Cofactors bind to Hox transcription factors, as well as DNA in a sequence specific manner increasing the protein complex's DNA binding specificity (Choe and Sagerstrom, 2005; Moens and Selleri, 2006; Waskiewicz et al., 2001). One such family of Hox cofactors is the Meis family (Burglin, 1997; Choe and Sagerstrom, 2005; Moens and Selleri, 2006; Williams et al., 2005).

The *Meis* family consists of a collection of HD containing genes that belong to the TALE (Three Amino acid Loop Extension) class of homeobox genes. The TALE

proteins are characterized by an additional three amino acids in the loop separating helix 1 and helix 2 of the HD (Burglin, 1997; Choe and Sagerstrom, 2005; Moens and Selleri, 2006; Williams et al., 2005). The basis of the *Meis* name is <u>M</u>yeloid <u>E</u>cotropic leukemia virus <u>I</u>ntegration <u>S</u>ite because the founding member of the *Meis* family was identified when it was upregulated in mouse leukemia cells after the Myeloid Ecotropic Leukemia Virus was found to induce leukemia upon integration into the gene's promoter region (Moskow et al., 1995). Subsequent identification of several paralogs in many species led to the establishment of the *Meis* family (Steelman et al., 1997). The *Meis* family consists of a solitary member in *Drosophila*, *Homothorax* (*Hth*) (Kurant et al., 1998; Pai et al., 1998; Rieckhof et al., 1997), 4 vertebrate *Meis* genes (*Meis1*, *Meis2*, *Meis3*, and *Meis4*) (Moskow et al., 1995; Nakamura et al., 1996; Waskiewicz et al., 2001), and 2 divergent vertebrate genes *Prep1* and *Prep2* (Berthelsen et al., 1998a; Fognani et al., 2002). Members of the *Meis* family have been identified in every vertebrate examined, including the zebrafish *Danio rerio* (Choe et al., 2002; Sagerstrom et al., 2001; Waskiewicz et al., 2001; Zerucha and Prince, 2001), the frog *Xenopus* (Salzberg et al., 1999; Steelman et al., 1997), the mouse *Mus musculus* (Cecconi et al., 1997; Moskow et al., 1995; Nakamura et al., 1996; Oulad-Abdelghani et al., 1997), the chicken *Gallus gallus* (Coy and Borycki, 2010), as well as humans (Smith et al., 1997; Steelman et al., 1997; Yang et al., 2000). A *C. elegans* homolog, *psa*-3, has been identified as well (Arata et al., 2006), and there is a significant amount of identity between *Meis* genes and *Knox* genes in plants (Burglin, 1997). The identification of significant *Meis* gene identities in protostomes, deuterostomes, as well as plants indicates that *Meis* genes are of ancient origin predating the divergence of the plant and animal kingdoms (Burglin, 1998). In spite of this ancient

ancestry, the general biochemical structure and protein-protein interactions of each member of the *Meis* family are highly conserved.

Each Meis cofactor can be divided into 2 structural regions that are highly conserved; an N-terminal bipartite domain, and the HD (Berthelsen et al., 1998a; Burglin, 1997; Chang et al., 1997; Jacobs et al., 1999; Knoepfler et al., 1997; Pai et al., 1998). A third region, the C-terminal region, is conserved among homologous genes but is variable among paralogs, and is the region that is subject to splice variance known to be exhibited in several *Meis* genes and which will be discussed in more detail later (Ahn and Ho, 2008; Burglin, 1997; Oulad-Abdelghani et al., 1997; Pai et al., 1998; Williams et al., 2005; Yang et al., 2000). Each structural domain has been demonstrated to provide a unique function to Meis transcription co-factors. In the N-terminal bipartite domain, amino acids 30-60 contain 2 conserved regions known as the Meis1 (M1) and Meis2 (M2) domains with the region between each domain being variable (Berthelsen et al., 1998b; Burglin, 1997; Chang et al., 1997; Knoepfler et al., 1999; Pai et al., 1998). The M1 and M2 domains are predicted to give rise to a coiled-coil motif, and are essential to facilitate Meis protein dimerization with Pbx proteins (Berthelsen et al., 1998b; Chang et al., 1997; Jacobs et al., 1999; Knoepfler et al., 1999). Both domains have been shown to be essential for Meis functionality outside of Pbx interaction, and the M1 domain may also contain an auxiliary binding site for other proteins (Choe et al., 2002).

The HD of the *Meis* family is highly conserved and structurally located in the center of the peptide. The HD provides Meis cofactors the ability to bind DNA, and each Meis cofactor has been shown to preferentially bind the sequence TGACAG / ACTGTC (Berthelsen et al., 1998a; Chang et al., 1997; Waskiewicz et al., 2001). Early studies

indicated that DNA binding by Meis cofactors was not important, but more recent studies have demonstrated that Meis DNA binding ability is indeed important to embryonic development in the vertebrate lineage (Waskiewicz et al., 2001). It has been suggested that Meis DNA binding ability may be dependent on or modulated by Meis cofactor interaction with Pbx cofactors (Berthelsen et al., 1998a; Chang et al., 1997). Further, amino acid residue 50 of the HD has been shown to vary among different TALE family members and may be necessary to provide Meis cofactors with a unique DNA binding sequence specificity (Burglin, 1997).

The C-terminal region of Meis proteins is far more variable than the N-terminal and HD regions and has been shown to be the site of splice variation (Burglin, 1997). There have not been splice variants described for *Prep1*, *Prep2*, *Hth*, or *Meis3*, but *Meis1* has 2 known variants (a and b) (Steelman et al., 1997), and *Meis2* has at least 4 splice variants (a, b, c, d) in mice (Oulad-Abdelghani et al., 1997) and 5 in humans with *Meis2*e having a truncated HD (Yang et al., 2000). The C-terminus of Meis cofactors can physically interact with Hox transcription factors, and splice variation appears to provide selective interaction with different Hox Paralog Groups (PG) (Huang et al., 2005; Shen et al., 1997; Williams et al., 2005). This would indicate that Meis cofactors can differentially interact with specific Hox PGs depending on C-terminal variation (Williams et al., 2005).

In addition to physical interactions with various Hox proteins many other functions have been ascribed to the C-terminal region of Meis proteins. It is specifically the C-terminal region of Meis proteins that is involved in the progression of Acute Myelogenous Leukemia (Mamo et al., 2006). This region has also been directly

demonstrated to be necessary to inhibit Histone Deacetylase (HDAc) activity, resulting in the acetylation of Histone 4 (H4) priming some promoters for transcriptional activity (Choe et al., 2009; Huang et al., 2005). To further support the importance of the C-terminus for functionality Huang et al. demonstrated that when the C-terminal region of Meis1 is lost, the mutant protein will still form higher order complexes, but the functionality of these complexes is inhibited (Huang et al., 2005).

As mentioned earlier the N-terminal region of every member of the *Meis* family has been shown to be necessary for physical interaction between Meis and Pbx cofactors. In fact, every member of the *Meis* family has been shown to directly interact with members of the *Pbx* family, both *in vitro* and *in vivo* (Abu-Shaar and Mann, 1998; Ferretti et al., 2006; Knoepfler and Kamps, 1997). Due to the high level of conservation in both the *Pbx* and *Meis* families, Meis-Pbx interactions appear to be promiscuous (Azcoitia et al., 2005). Meis and Pbx dimers cooperatively bind to consensus half sites *in vitro* and *in vivo* with the physical interaction providing DNA binding stability to the complex. These functional half sites can be located next to each other, but have also been shown to be flexible regarding distance between half sites, with distances of 7bp and 12bp known to exist when binding *in vivo* (Abu-Shaar and Mann, 1998; Jacobs et al., 1999; Knoepfler et al., 1999; Ryoo and Mann, 1999; Wang et al., 2001).

To date, Meis-Pbx dimers have been shown to bind numerous regulatory sequences *in vivo*, including *cis*-regulatory sequences of: *Hox*, *Myogenin*, *Sox*3, and the Malic Enzyme gene (Berkes et al., 2004; Choe et al., 2009; Knoepfler et al., 1999; Mojsin and Stevanovic, 2010; Wang et al., 2001). These dimers have been demonstrated to be transcriptionally active, and be able to physically interact with other transcriptional

complexes like: bHLH transcription factor MyoD, Retinoic Acid (RA) receptors RARα and RXR, as well as displacing HDACs associated with Pbx cofactors when bound to promoters *in vivo* (Choe et al., 2009; Knoepfler et al., 1999; Mojsin and Stevanovic, 2010; Wang et al., 2001).

Meis-Pbx dimerization, in addition to transcriptional modulation, has the function of influencing the nuclear localization of each cofactor. This was first identified by Rieckhof et al. in 1997 when they observed that the presence of Hth is necessary for the nuclear localization of the *Drosophila Pbx* homolog Extradenticle (Exd) (Rieckhof et al., 1997), which was further supported by the observation that the Hth HM domain (*Drosophila* M1 plus M2 domain) is sufficient to drive Exd nuclear transport (Ryoo et al., 1999). The reliance on Meis/Hth presence for Pbx/Exd nuclear localization has been shown to be consistent in mice as well (Abu-Shaar et al., 1999; Berthelsen et al., 1999; Capdevila et al., 1999; Kurant et al., 1998; Mercader et al., 1999). However, the opposite requirement has been identified in zebrafish, *Xenopus*, and spiders. In these lineages it appears that Pbx is constituently nuclear, and Meis cofactors require the presence of Pbx for nuclear localization (Deflorian et al., 2004; Maeda et al., 2002; Prpic et al., 2003; Vlachakis et al., 2001). Consistent with the species specific requirement for one member of the Meis-Pbx dimer to be present to drive the nuclear localization of the other, it has been observed that both *Pbx* and *Meis* expression domains frequently overlap (Toresson et al., 2000).

In spite of this correlation it has been observed that in some cases Meis is not sufficient to drive Pbx nuclear localization, and that Exd is nuclear in the absence of Hth in imaginal disc cells (Rieckhof et al., 1997; Swift et al., 1998). It was shown by Saleh et

al. that in species where Pbx peptides require the presence of Meis that the Pbx peptide contains 2 Nuclear Localization Signals (NLSs) located in its N-terminal region. They proposed a model where the NLSs, which are masked due to native protein folding, are exposed following a conformational change reliant on the binding of Meis, initiating nuclear localization of both cofactors (Saleh et al., 2000). Pbx cofactors also have a phosphorylation site in the N-terminal region. It has been proposed that this site is necessary for Pbx nuclear localization in the absence of Meis cofactors via phosphorylation, resulting in the same conformational change that Meis would initiate exposing the NLSs, driving Pbx nuclear localization. Subsequently Pbx remains cytoplasmic when the site is dephosphorylated, thus providing additional control over the nuclear localization of Pbx through the regulation of PKA activity (Kilstrup-Nielsen et al., 2003). The ability for both cofactors to be nuclear independent of each other in some situations provides an additional post-transcriptional regulatory mechanism that may facilitate differential gene regulation by modulating transcription factor availability in the nucleus.

Meis cofactors have been shown to form higher order complexes with Hox proteins in addition to Pbx. Meis-Hox dimer interactions were first observed in 1997 by Shen et al. when they demonstrated that Meis1 will cooperatively bind DNA on consensus half sites (TGACAG-TTAT / Meis-Hox binding sites respectively) with HoxA11, HoxD12, and HoxD13 in Electrophoretic-Mobility Shift Assays (EMSA) (Shen et al., 1997). Unlike what has been observed for Meis-Pbx consensus binding sites, the spacing of Meis-Hox consensus binding sites have been shown to be obligatory neighbors lacking flexibility (Ryoo et al., 1999). In the following years numerous Meis cofactors,

although not Prep1 or Prep2, have been demonstrated to directly interact with a variety of Hox gene products (Choe and Sagerstrom, 2005; Fujino et al., 2001; Pai et al., 1998; Ryoo et al., 1999; Shanmugam et al., 1999; Thorsteinsdottir et al., 2001). Of particular interest, Meis cofactors appear to dimerize with Adb-B like Hox PGs 9-13. In many cases Meis cofactors have been shown to form trimers with Hox and Pbx cofactors, where the Meis peptide interacts with a Pbx-Hox dimer. The interaction between Pbx and Hox is facilitated by a conserved tryptophan motif in the Hox transcription factor. However, Hox PGs 9-13 lack this tryptophan motif making them unlikely to physically interact with Pbx. This indicates that Meis specific interaction with Adb-B like Hox transcription factors may be of significance (Shen et al., 1997). However, as demonstrated by Williams et al., Meis cofactors, in addition to Adb-B like Hox transcription factors, physically interact with other Hox transcription factors. Using a yeast expression system the group was able to demonstrate that Meis1, Meis2, and Meis3 cofactors combined are capable of dimerizing with Hox PGs 2, 4, 5, 8, 9, 10, 11, 12, and 13. Furthermore, they demonstrated that different Meis splice variants have different PG binding preferences (Williams et al., 2005). This observation indicates that not only do individual Meis cofactors form functional dimers with Hox transcription factors from multiple PGs, each individual Meis gene preferentially dimerizes with different PGs depending on which variant is present. This provides a developing organism the ability to have multiple Meis cofactors present while simultaneously differentially regulating its genome by controlling which functional partners are present to interact.

In addition to dimerization with Pbx cofactors or Hox transcription factors, Meis cofactors are also known to form trimeric complexes with Hox and Pbx. Originally

proposed by Chang et al., every member of the Meis family has been demonstrated to

have the ability to form trimeric complexes with Hox and Pbx cofactors (Figure 1)

(Berthelsen et al., 1998a; Chang et al., 1997; Jacobs et al., 1999; Kurant et al., 1998;
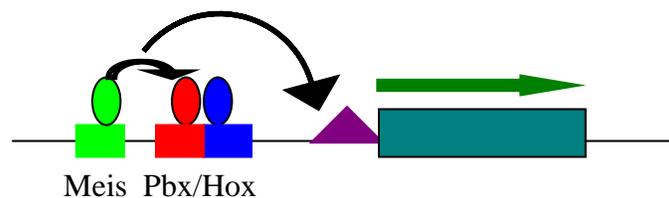
Meis Pbx/Hox

Figure 1: Meis-Pbx-Hox Trimeric Interactions
Ovals represent transcription factors, smaller rectangles
represent transcription factor binding sites, triangle
is a promoter, and the long rectangle is the target gene.
Meis, Pbx, and Hox transcription factors are known to
form higher order complexes regulating the expression
of their target gene.
Adapted from Berthelsen et al., 1998a and Jacobs et
al.,1999

Shanmugam et al., 1999; Shen et al., 1999; Vlachakis et al., 2001; Vlachakis et al., 2000).

The first several experiments performed showed that DNA binding by Meis was not a

requirement for higher order complex formation. It was shown through multiple

experiments that a Pbx-Hox dimer will bind to consensus Pbx-Hox half sites followed by

a Meis cofactor binding to the Pbx cofactor independent of Meis-DNA interaction. It

was even shown that the formation of higher order complexes was not dependent on the

integrity of the Meis HD. (Berthelsen et al., 1998a; Shanmugam et al., 1999; Shen et al.,

1999).

Following these initial experiments there was much disagreement as to the

functional input of the Meis cofactors in these trimeric complexes. However, these

original studies involved EMSAs on Meis-Pbx-Hox consensus sites located adjacent to

each other. Using cellular extracts or *in vitro* translated products, different groups were able to show that the mutation of the Meis binding site did not negatively impact the formation of trimeric complexes. In 1999, Jacobs et al. were the first to notice that there were *Meis* consensus sites present in the *Hox*B1 ARE and rhombomere (r) 4 regulatory elements *in vivo*. Unlike previous binding sites studied *in vitro* these *Meis* consensus sites were not directly adjacent to the *Pbx-Hox* core, but rather, were located 1 helical turn and 2 helical turns away from any present *Pbx-Hox* consensus sites respectively (Jacobs et al., 1999). The group subsequently demonstrated that Meis co-factors enter into trimeric complexes with Pbx and Hox at the ARE and r4 sites as a DNA binding partner (Jacobs et al., 1999). Almost every investigation that followed has found Meis cofactors to be DNA binding members of any higher order complexes that they form. In many cases, Meis cofactor DNA binding has been shown to modulate sequence binding specificity, to be required for complex formation, and contribute to the transcriptional activity of the complex (Choe et al., 2009; Fujino et al., 2001; Jacobs et al., 1999; Kurant et al., 1998; Sarno et al., 2005; Schnabel et al., 2000; Vlachakis et al., 2001; Vlachakis et al., 2000).

It was initially believed that Meis cofactors only form higher order complexes with Hox and Pbx, but it is now known that Meis cofactors form higher order complexes with several non-canonical binding partners. The first non-canonical binding partner identified was PDX1, when it was demonstrated by Swift et al. in 1998 that MEIS2 forms a trimeric complex with PBX1b and PDX1 in human Acinar cells. In this case MEIS2 was shown to be a DNA non-binding member of the complex, but it would be interesting to revisit the binding site and see if a detached *Meis* binding site exists given that the

work was done around the same time as when the flexibility in Meis binding site location was identified (Liu et al., 2001; Swift et al., 1998). Meis and Pbx cofactors have since been shown to function as DNA binding members of higher order complexes with MyoD-E12 dimers, TLX1, RARα, TR-RXR dimers, and hypothetically with FOXP2 in the primate forebrain (Knoepfler et al., 1999; Milech et al., 2010; Mojsin and Stevanovic, 2010; Takahashi et al., 2008; Wang et al., 2001). Meis co-factors have also been shown to bind *cis*-regulatory elements that control *Pax*6 transcription independent of Pbx in the developing eye or with Pbx in the Acinar cells (Zhang et al., 2002; Zhang et al., 2006). The identification of novel interactions and DNA binding partners in a plethora of different tissues indicates that there may be more novel partners to be identified.

An array of binding partners, in combination with the coupling of nuclear localization mechanisms and preferential interactions dependent on splice variation, lays the groundwork for an extremely complex and intricate interplay of regulatory mechanisms. Because the basic regulatory functions of Meis cofactors are mostly conserved among each member of the family, the complexity would make it possible for a developing organism to differentially regulate developmental cascades according to positional and temporal requirements. The built in potential for redundancy may make it possible for expressions patterns to be slightly altered, and tweaked with little negative side effects. This ability would allow for the establishment of the necessary variation for natural selection to act on without being detrimental to the organism.

In addition to the many homologous functions and structures conserved within the *Meis* family, each member of the *Meis* family has also evolved specific functions and roles in development. Phylogenetic analysis of the *Meis* family shows that the *Prep* and

Meis genes were the first members to diverge from each other; hence the *Prep* genes are the most ancient. According to all available phylogenetic information this divergence occurred prior to that of the protostome and deuterostome lineages, since the next *Meis* family to diverge was *Hth*, the *Drosophila* homolog. This is interesting because, to date, there has not been a *Prep* homolog identified in *Drosophila* or any other protostome, but homologs have been identified in all vertebrates examined. Next, *Meis3* diverged in vertebrates, followed by the divergence of *Meis1* and *Meis2* from a common ancestor gene (Biemar et al., 2001; Burglin, 1997; Coy and Borycki, 2010; Moens and Selleri, 2006; Nakamura et al., 1996; Nam and Nei, 2005; Williams et al., 2005; Zerucha and Prince, 2001). Thus *Meis1* and *Meis2* are the most recent genes to have evolved (Figure 2).

Prep genes are ubiquitously expressed during early zebrafish embryonic development. Deposited maternally, *prep1* exhibits its highest expression level at the germ-ring stage, with ubiquitous expression continuing through 24 hours postfertilization (hpf). At 72 hpf *prep1* expression becomes restricted to the brain and optic vesicles in the zebrafish central nervous system (CNS) (Choe et al., 2002; Deflorian et al., 2004; Waskiewicz et al., 2001). Functionally, *Prep* genes have been implicated in numerous developmental processes in multiple organisms. Overexpression studies in the zebrafish have demonstrated that *Prep1* is involved in the development of the posterior nervous system, with mutants exhibiting gradations of cyclopia like phenotypes, a reduced forebrain, and caudalization of anterior neural structures (Deflorian et al., 2004). This indicates that *Prep1* is functional in the specification of the presumably more ancient posterior structures of the CNS, because it appears that when there is an excess of *Prep1*

it can override the anterior patterning pathways, respecifying them to a posterior-like form.

zPrep2　cPrep2　xtPrep2　mPrep2　　zPrep1.1　zPrep1.2　cPrep1　mPrep1

xtMeis3

xlMeis3

zMeis3

mMeis3

zMeis1

cMeis1

xtMeis1

mMeis1

Hth

zMeis4.1　zMeis2.1

mMeis2　cMeis2　xtMeis2　zMeis2.2
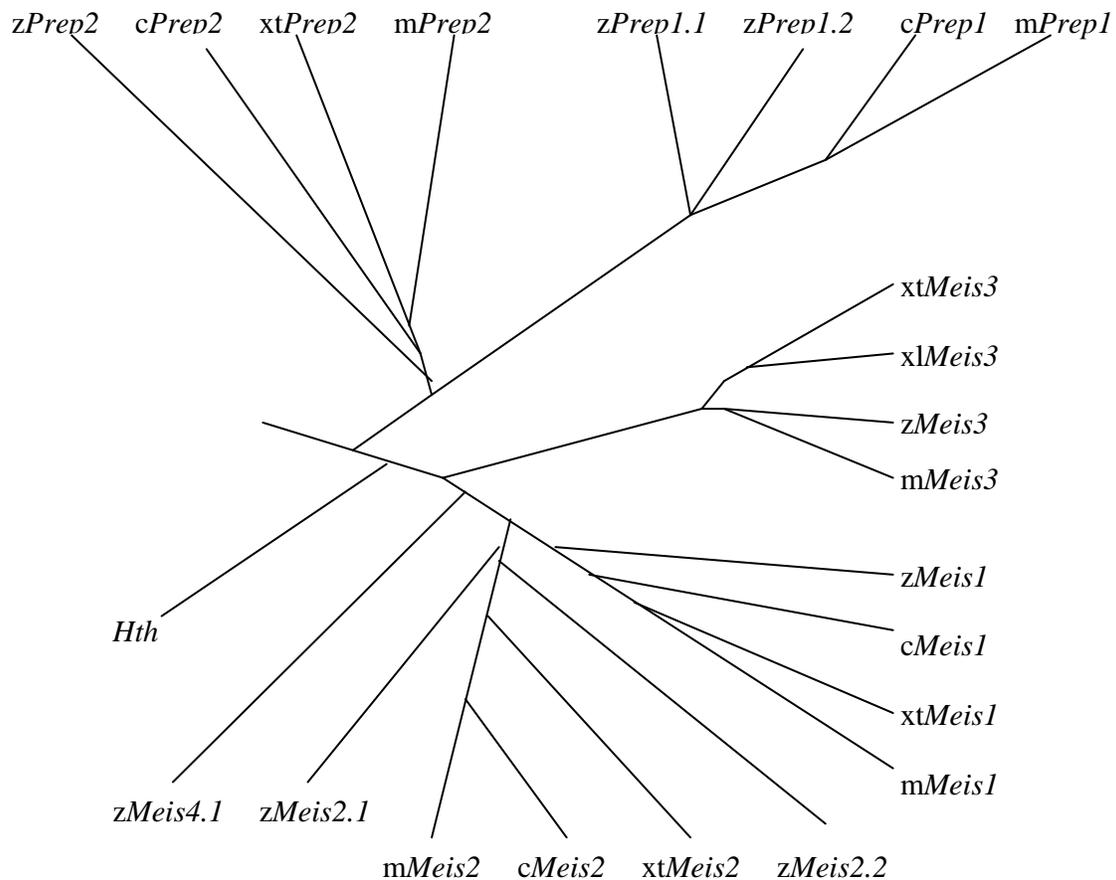
Figure 2: Phylogenetic Tree of Meis and Prep Families (Based on Coy and Borycki, 2010)
z stands for zebrafish (*Danio rerio*)
c stands for chicken (*Gallus gallus*)
xt stands for frog (*Xenopus tropicalis*)
xl stands for frog (*Xenopus laevis*)
m stands for mouse (*Mus musculus*)

Knockout studies in mice and knockdown studies in zebrafish have demonstrated even more drastic phenotypes. Double null *Prep* (⁻/⁻) mutant mice exhibit severe morphological phenotypes as well as alterations in genetic expression patterns. As the embryos of these mutants develop they exhibit hematopoietic failure and reduced eyes. They also have morphological deficiencies in the hematopoietic liver as well as other

organs. All of these morphological abnormalities are accompanied by almost no *Meis2* expression at all, no *Pax6* expression in the developing retina or lens, and significantly decreased *Meis1* and *Pbx* expression (Ferretti et al., 2006). In zebrafish *prep1* morpholino (MO) studies, the morphants exhibited severe neural deformities, including: loss of r1/2 and r6/7 boundaries, defective hindbrain patterning, reduced eyes, microcephaly, and a lack of neural crest derived cephalic cartilage and jaws. These morphants also exhibited a loss of rhombomeric specific gene expression, heightened neuroectoderm and CNS apoptosis, as well as reduced or impaired motor ability. In addition to the neural defects, *prep1* morphants also had severe mesodermal phenotypes exhibiting atrophic pectoral fins, an absence of swim bladders, and pericardial edema. Morpholino treatment was lethal by 6 to 7 days past fertilization (dpf) (Deflorian et al., 2004).

Based on morphological studies in both mice and zebrafish, it has been hypothesized that *Prep1* may function as a master regulatory gene, functioning to regulate members of the *Meis* and *Pbx* families (Ferretti et al., 2006). Because many of the *Prep1* phenotypes observed are also observed in *Meis* and *Pbx* mutants, combined with the ubiquitous early expression pattern of *Prep*1, it is feasible that the observed *Prep*1 effects could be caused by the downstream effects of the decrease in *Meis1* and *Meis2* expression observed in mice (Ferretti et al., 2006). Further support for this idea comes from an earlier study that found that *Prep*1 is negatively regulated while *Meis1* is positively regulated by RA in P19 cells (Ferretti et al., 2000).

In addition to *Meis* regulatory functions, it appears that *Prep* genes may also be necessary for proper development of endodermal structures. In the mouse pancreas,

Prep1 and Prep2 cofactors both synergistically bind to the *Pax*6 promoter with Pbx on a consensus *Meis-Pbx* site. *Pax*6, in addition to being a neural transcription factor, is also a pancreatic endodermal determinant, and this Prep-Pbx interaction is favored over any other Meis co-factors that might be present (Delporte et al., 2008; Zhang et al., 2006). This observation, when combined with mesodermal defects identified in other studies, would indicate that *Prep* functions in endodermally derived structures, potentially independently from other *Meis* genes.

Prep cofactors also have regulatory roles that have been ascribed to them, with some overlap with other Meis cofactors. While Prep1 has been shown to enhance Pbx/Hox related reporter activity (Berthelsen et al., 1998a) and drive ectopic *Hox* expression when overexpressed (Choe et al., 2002), it has also been demonstrated that Prep1 cofactors do not physically interact with PG 3 or PG 9 Hox transcription factors (Thorsteinsdottir et al., 2001). In addition, the C-terminal domain of Prep1 has no functional role in H4 acetylation like Meis cofactors do (Huang et al., 2005).

Combined, the lack of direct Hox interaction and an inability to influence H4 acetylation indicates that *Prep* gene function is more than likely less specialized than the *Meis* genes that they are related to. The wide array of *Meis* gene like phenotypes coupled with downregulation of *Meis1* and *Meis2* indicate a role for *Prep* genes as regulators of *Meis* genes. Especially when one considers the non-*Meis* like phenotypes of lacking swim bladders and pericardial edema, it is a logical conclusion that the functions of *Prep* genes represent a conserved, ancient mesodermal role.

The most perplexing aspect of *Prep* function is the phylogenetic history. Prep1 is not sufficient to substitute for Hth in *Drosophila*, indicating a vertebrate specific function

(Vlachakis et al., 2001). This observation is consistent with the phylogenetic analysis placing *Hth* more closely related to *Meis1*, *Meis2*, and *Meis3* than the *Prep* genes (Figure 2) (Coy and Borycki, 2010; Moens and Selleri, 2006; Nam and Nei, 2005). It has also been predicted that the Most Recent Common Ancestor (MRCA) between protostomes and deuterostomes had a single *Prep* gene, a single *Meis* gene, and subsequently the *Prep* gene was lost in the entire protostome lineage (Nam and Nei, 2005). But this prediction begs a major question: If *Prep* function is reminiscent of an ancient function, is essential to the patterning of mesodermal structures, as well as initiating subsequent *Meis* expression; how could the *Prep* gene be categorically lost in the protostome lineage without it being fatal? Furthermore, what transcription factor or factors took over the role in driving *Hth* expression? From a functional standpoint, it may make more sense to evaluate the possibility that the *Prep* genes could have diverged in the *Meis* gene lineage after the protostome-deuterostome lineage split, retaining only the basal *Meis* functions, while the *Meis* genes simultaneously maintained ancient higher order functions and evolved new functions.   This scenario would be in accordance with the Sub-functionalization scenario of the Duplication – Degeneration – Complementation (DDC) (Force et al., 1999) theory for gene evolution following duplication events (Figure 3). This would also explain the lack of function in the Prep C-terminus, but the conservation of N-terminal and HD function, while not being able to substitute for *Hth* functionally.
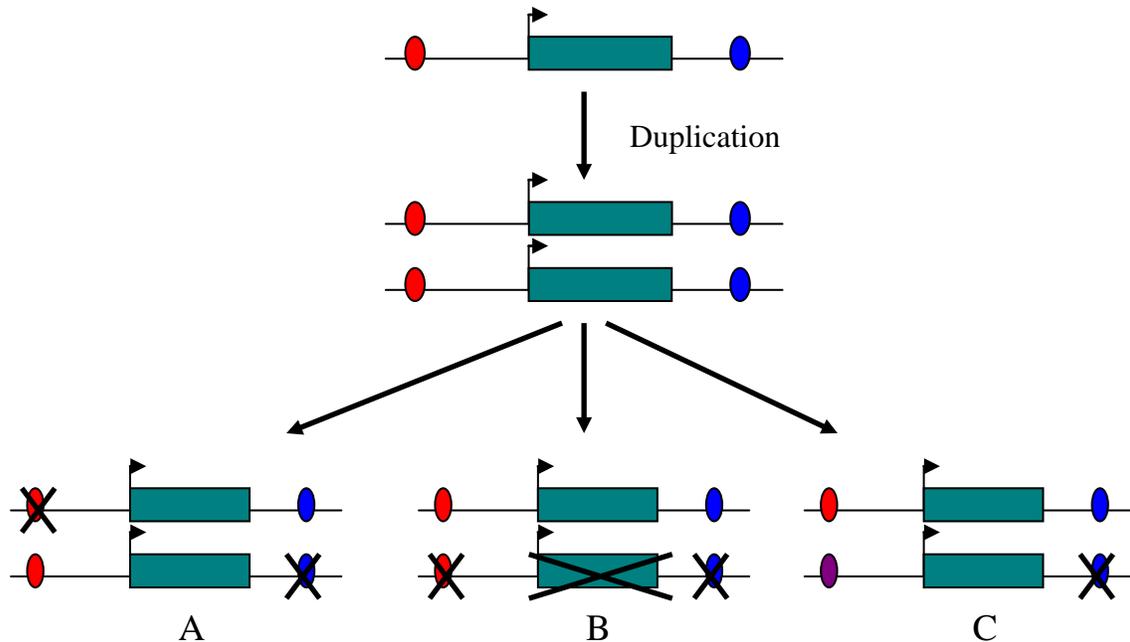
Figure 3: Duplication-Degeneration-Complementation Model for Gene Evolution

Occasionally mutations occur where genes, chromosomes, or entire genomes are duplicated. These mutations result initially in 2 identical paralogous genes and their *cis*-regulatory elements. Ovals are cis-regulatory elements, purple oval is a novel cis-regulatory element resultant of mutation.

(A) Sub-functionalization: Following duplication, individual regulatory elements of one paralog can be mutated or deleted (degenerated) and there will be no adverse affects to the organism because the other paralogous element can maintain ancestral function. As a result both genes will be evolutionarily maintained through natural selection to prevent a loss of gene function, but each gene will exhibit a subset of the original expression pattern.

(B) Non-functionalization: Following duplication, all regulatory elements of one paralog, or the gene its self, can be mutated or deleted (degenerated) and there will be no adverse affects to the organism because the other paralogous gene will be sufficient to preserve ancestral function. As a result, one paralog and its regulatory mechanisms will be evolutionarily maintained through natural selection to prevent a loss of gene function, while the other paralog will be lost over time resulting in a pseudogene.

(C) Neo-functionalization: Following duplication, one paralog can come under the control of a novel *cis*-regulatory element generated by random mutation in one of the ancestral elements resulting in a novel function for that particular paralog. If this novel function proves advantageous, natural selection will maintain it, while the original *cis*-regulatory elements can maintain ancestral function with no adverse affects to the organism. As a result, one paralog and its ancestral regulatory mechanisms will be evolutionarily maintained, while the other paralog will evolve some novel function due to its new regulatory mechanism.

Based on phylogenetics the next member of the *Meis* family to diverge was the *Drosophila* homolog *Hth* (Coy and Borycki, 2010; Moens and Selleri, 2006; Zerucha and Prince, 2001). *Hth* has been implicated in several basal *Drosophila* developmental functions, most of which are conserved among vertebrate *Meis* genes. It was first thought that *Meis* genes may function to modulate *Hox* genes when it was noticed that *Hth* loss-of-function phenocopies *Hox* phenotypes. In this case, loss of *Hth* resulted in an altered number of nerve roots in the *Drosophila* CNS (Kurant et al., 1998; Rieckhof et al., 1997). Functional studies have demonstrated that *Hth* functions possibly to suppress eye development (Heine et al., 2008; Pai et al., 1998), forms higher order complexes as a DNA binding cofactor with transcriptional activity, is essential for *Engrailed* function, and regulates *Distal-less* activity (Dibner et al., 2001; Inbal et al., 2001; Kobayashi et al., 2003; Ryoo et al., 1999). *Hth* has been shown to affect gene expression in *Drosophila* in a similar manner to that of *XMeis3* in *Xenopus*, as well as to be necessary for proper neuronal differentiation and positioning (Inbal et al., 2001). It has even been shown that *Meis1* can functionally substitute for *Hth* (Rieckhof et al., 1997).

In addition to neural and regulatory roles, *Hth* is also expressed in the developing limb of *Drosophila*, where its expression is restricted to the proximal region of the limb primordia, dividing the limb into proximal *Hth+* and distal *Hth-* regions. In the distal region, *Hth* is actively suppressed by *Wingless* and *Decapentaplegic*. This regional expression pattern is conserved across both the protostome and deuterostome lineages. The conservation indicates an ancient origin for the role of *Hth* in limb development (Abu-Shaar and Mann, 1998; Mercader et al., 1999).

All together, the high degree of functional and sequence conservation between protostomes and deuterostomes seems to indicate that the functions of *Hth* may closely resemble *Meis* function in the MRCA of the 2 lineages. Unlike the deuterostome lineage, however, there have not been large scale gene duplication events in the protostome lineage resulting in the existence of only a single *Meis* homologue in *Drosophila*.

The next step in the evolution of the *Meis* family was the divergence of *Meis3* in deuterostomes (Coy and Borycki, 2010; Moens and Selleri, 2006; Zerucha and Prince, 2001). Expression studies of *Meis3* in *Xenopus*, zebrafish, and mouse have all demonstrated a similar expression pattern during development. CNS expression is limited to the neural tube/spinal cord, and the posterior region of the hindbrain, with an anterior boundary at the r3/r4 boundary. Expression is initiated early in development and rapidly fades, becoming restricted to the spinal cord and small subpopulations of expression in the posterior hindbrain (Sagerstrom et al., 2001; Salzberg et al., 1999; Waskiewicz et al., 2001). Outside of the CNS *Meis3* is expressed early in the proximal region of the developing limb/pectoral fin, and somites, becoming endodermally restricted later in development (diIorio et al., 2007; Sagerstrom et al., 2001; Waskiewicz et al., 2001; Williams et al., 2005). In mice, *Meis3* is also expressed in the developing male and female reproductive tracts (Williams et al., 2005).

In accordance with CNS expression patterns, morphological studies have demonstrated an essential role for *Meis3* in the posterior CNS. Mutational studies in *Xenopus* and zebrafish have shown that *Meis3* functions to caudalize the CNS to an r4 hindbrain-like state. When expression is impaired, posterior truncations and expansion of anterior CNS structures are observed (Dibner et al., 2001). A partial loss of *Meis3*

function, as indicated by a loss of distinct boundary formation and Mauthner neurons misplacement in r4, results in a loss of rhombomeric identity, respecifying the hindbrain to r4. This demonstrates an anterior shift in tissue specificity, while a total loss of function results in a loss of segmentation altogether (Choe and Sagerstrom, 2004, 2005; Choe et al., 2002).

Overexpression studies agree with loss of function studies, showing that when *Meis3* is overexpressed a decrease in anterior markers and structures, coupled with an increase in posterior marker expression, and an anterior expansion of posterior structures occurs (Salzberg et al., 1999; Vlachakis et al., 2001). Morphological studies have prompted investigators to hypothesize that r4 represents a native state of the hindbrain, and *Meis3* functions to inhibit this fate in more posterior structures (Choe and Sagerstrom, 2004, 2005). In further support of this idea, genetic analysis shows that *Meis3* may function to specify the posterior hindbrain. It has been shown that *Meis3* expression is essential for the proper expression of *hoxb1a* and *hoxb2* in the zebrafish (Choe et al., 2002). Choe and Sagerstrom went on to show that PG1 *hox* genes function to specify an r4 type state while, in r5 and r6, another gene, *vhnf1*, functions to suppress this signal specifying a more posterior fate. They also showed that *vhnf*1 has multiple consensus *Meis* binding sites its promoter, and *Meis3* functions with *ho*x*b1a* to induce *vhnf*1 expression in r5 and r6, thus inhibiting the r4 specification program from proceeding (Choe and Sagerstrom, 2004, 2005).

In addition to developmental roles in the CNS, *Meis3* also has endodermal roles in zebrafish; namely in pancreatic development and the repression of the insulin pathway (diIorio et al., 2007). Studies have shown that *Meis3* is essential for *Shh* and *Foxa2*

expression in the anterior endoderm, indicating that *Meis3* is an upstream regulator of the pancreatic repression pathway. Additionally, *Meis3* is suspected of having a role in the diversification of the branchial arches in zebrafish (diIorio et al., 2007). This latter possibility has yet to be explored in detail, but given functional overlap between members of the *Meis* family, it is plausible.

Another important function that has been identified for *Meis3* involves H4 acetylation. Choe et al. have recently demonstrated that *Meis3* has the ability to promote H4 acetylation at *Hox* promoters *in vivo* (Choe et al., 2009). Believed to function by displacing HDAc, *Meis3* has been proposed to prime genes for activity by promoting H4 acetylation, and easing chromosomal compaction. The research group also noticed that neither Pbx nor Hox cofactors were bound to promoters in the absence of *Meis3* or an acetylated H4, causing them to suggest that Meis cofactor DNA binding may be a prerequisite for subsequent transcription factor binding (Choe et al., 2009). The biochemical source of this function has not been identified, but if the source of the function is in either the N-terminal region or the HD central region, it would indicate that all *Meis* genes may have the same activity, sharing an ancient evolutionarily conserved role in chromosomal modification.

The most recent lineages to diverge in the *Meis* family were *Meis1* and *Meis2* (Coy and Borycki, 2010; Moens and Selleri, 2006; Zerucha and Prince, 2001). *Meis1* and *Meis2* both have significant impacts on development and demonstrate restricted expression patterns similar to *Meis3*. *Meis2* expression has been evaluated in zebrafish, mouse, chicken, monkey, and several other mammals. While there is only one *Meis2* paralog in most vertebrates, due to a recent genomic duplication event in the teleost

lineage zebrafish have two paralogs: *meis2.1* and *meis2.2* (Waskiewicz et al., 2001; Zerucha and Prince, 2001). Among all animals examined there are several commonalities that occur in the expression patterns of *Meis2* as it is expressed in the proximal limb bud of every vertebrate examined to date, with the exception of *meis2.1* in zebrafish (Capdevila et al., 1999; Coy and Borycki, 2010; Mercader et al., 1999; Mercader et al., 2005; Waskiewicz et al., 2001; Zerucha and Prince, 2001). Restricted *Meis2* expression has also been described in the forebrain, midbrain, and hindbrain of every vertebrate examined to date, especially in the developing lens and retina. *Meis2* expression in the neural tube/spinal cord and branchial arches (exception of *meis2.1*) of vertebrates has also been described (Capellini et al., 2008; Conte et al., 2010; Coy and Borycki, 2010; Heine et al., 2008; Oulad-Abdelghani et al., 1997; Takahashi et al., 2008; Toresson et al., 1999; Vennemann et al., 2008; Waskiewicz et al., 2001; Zerucha and Prince, 2001). In addition to expression in the CNS, *Meis2* has been described in the somites (specifically in the sclerotome) (Capellini et al., 2008; Cecconi et al., 1997; Waskiewicz et al., 2001), and in both the male and female reproductive tracts (Crijns et al., 2007; Oulad-Abdelghani et al., 1997; Williams et al., 2005). The only germ layer where *Meis2* embryonic expression has not been described in is the endoderm (diIorio et al., 2007).

Functional studies have implicated *Meis2* in the development of many structures. Some of the less well studied are the pancreas where *Meis2* may serve as a switch between endocrine and exocrine determination in cultured Acinar cells (Liu et al., 2001; Swift et al., 1998), primate forebrain structures associated with the acquisition of language (Takahashi et al., 2008), and in muscle histogenesis (Cecconi et al., 1997). The

structures where *Meis2* has a function that have been studied in the most detail are the hindbrain, limb, and eye. Expression studies have shown that *Meis2* plays a role in the patterning of the hindbrain, where Meis2 cofactors have been shown to synergize with Hox/Pbx dimers, as well as Iroquois transcription factors on element C driving *Krox*20 expression in r3 (Stedman et al., 2009; Wassef et al., 2008). Ectopic expression of *Meis2* has been shown to cause midbrain structures to transform caudally (Vennemann et al., 2008), and to expand anterior hindbrain structures (r1 through r3) anteriorly (Wassef et al., 2008).

Meis2 has also been shown to be essential for normal development of the adult limb. Just as with *Hth*, *Meis2* expression is restricted to the proximal limb primordia where it has been shown to be responsive to RA, and function to repress the distal limb program (Capdevila et al., 1999; Kumar et al., 2007; Mercader et al., 2005). In the developing limb, *Meis2* is antagonized by BMP and Gremlin, two known distal factors. As a result it has been proposed that Pbx/Meis cofactors may function as proximal factors in the proximal limb bud, while distal progression of their expression domains is restricted by the *Shh*/*FGF* loop via BMP and Gremlin (Capdevila et al., 1999). *Meis2* is such a strong proximal activator that when *Meis2* is ectopically expressed distally, all distal structures are transformed to a proximal fate (Capdevila et al., 1999).

In the vertebrate eye, *Meis2* has been shown to be important for several different functions. Recently Conte et al. showed that dosage of *Meis2* is controlled by the microRNA miR-204 in the lens and optic cup of chickens. They demonstrated a role for *Meis2* in lens differentiation, dorsal-ventral patterning of the optic cup, and a necessity of *Meis2* expression for closure of the optic fissure (Conte et al., 2010). In the developing

eye Meis2 cofactors have been shown to be upstream regulators of *Pax*6 in both the retina and lens (Hisa et al., 2004; Zhang et al., 2002). In addition to *Pax*6 regulation, *Meis2* has been implicated in the maintenance of progenitor stem cell populations in the developing retina, noting a progressing wave of neural differentiation corresponding to a regressing wave of *Meis2* expression. It has also been shown that the duration of *Meis2* expression is species specific, and correlates to eye and retinal size (Heine et al., 2008; Hisa et al., 2004).

While *Meis2* has been implicated in the maintenance of progenitor stem cell populations in the retina, *Meis2* has also been shown to be a marker for specific subtypes of postmitotic neurons, including GABAergic Amacrine (AM) cells and striatal neurons (Bumsted-O'Brien et al., 2007; Toresson et al., 1999). This indicates that *Meis2* has a functional role in maintaining progenitor populations, as well as in determining neural transmitter subtype. These are seemingly opposite roles, but it is plausible since *Meis1* (as described later) has been implicated in cell cycle control (Bessa et al., 2008), and that *Meis2* has been implicated in determining exocrine vs. endocrine function in the pancreas. As a result it is quite conceivable that *Meis2* could differentially function as both a differentiation determination factor and a progenitor maintenance factor depending on the cellular context.

The final member of the *Meis* family is *Meis1*, which like *Meis2* has been studied in a wide array of vertebrates. *Meis1* has an expression pattern that is conserved across vertebrates and closely mimics that of *Meis2* expression. Just as with *Meis2*, *Meis1* is heavily expressed in the CNS with restricted expression in the forebrain, midbrain, and hindbrain (Capellini et al., 2008; Coy and Borycki, 2010; Maeda et al., 2002; Maeda et

al., 2001; Smith et al., 1997; Waskiewicz et al., 2001). *Meis1* is also expressed in the developing eye, both male and female reproductive tracts in mouse, in the branchial arches, in somites, as well as in the proximal limb bud (Bessa et al., 2008; Capellini et al., 2008; Coy and Borycki, 2010; Crijns et al., 2007; Maeda et al., 2002; Maeda et al., 2001; Mercader et al., 1999; Waskiewicz et al., 2001; Williams et al., 2005).

Mouse knockout studies have indicated that a major role for *Meis1* in development may be in hematopoiesis and vascular patterning. Mouse double knockouts exhibit morphological deformities in the brain, heart, lung, and kidney. Furthermore this mutation is lethal due to little to no hematopoiesis (Hisa et al., 2004). In addition to a lack of hematopoiesis, *Meis1* (-/-) embryos exhibit severe defects in their vasculature, as well as extreme hemorrhaging in the brain and trunk regions (Azcoitia et al., 2005; Hisa et al., 2004).

Outside of a role in hematopoiesis, *Meis1* has been implicated in the patterning of the CNS. Alterations in *Meis1* expression has been shown to result in the misplacement of neurons in the dorsal horn of the spinal cord, as well as an inhibition of trigeminal ganglion precursors resulting in decreased neural crest derived neurons in the facial ganglion of chickens (Rottkamp et al., 2008; Yang et al., 2008). Overexpression studies have indicated that *Meis1* has regulatory roles in r3 and r5, in RA mediated induction of neural transcription factor *Sox*3, as well as regulating *Hox* gene expression and function in the CNS (Azcoitia et al., 2005; Huang et al., 2005; Mojsin and Stevanovic, 2010; Schnabel et al., 2000; Stedman et al., 2009; Waskiewicz et al., 2001). Additionally, overexpression of *Meis1* has been shown to induce neural and neural crest markers like:

N-cam, N-tubulin, *Krox*20, *Zic*3, *Fgf*8, and *Cyp*26b1 (Maeda et al., 2002; Maeda et al., 2001; Stedman et al., 2009).

Meis1 has also been implicated in the development and patterning of the vertebrate limb and retina. Just like *Meis2*, *Meis1* expression is restricted to the proximal limb bud in mouse and chicken, and when *Meis1* is ectopically expressed distally a reduction in distal structures has been observed (Mercader et al., 1999; Mercader et al., 2005). In the developing retina, *Meis1*, like *Meis2*, is suspected to function upstream of *Pax6*, and help maintain retinal progenitor populations (Heine et al., 2008).

There are 2 additional functions that have been attributed to *Meis1* that have not been described for any other member of the *Meis* family. One of these roles is in controlling cell cycle regulation, specifically the Gap1 to DNA Synthesis phase transition, through regulation of *Cyclin*D1 and *c-myc* in eye development (Bessa et al., 2008). While other members of the *Meis* family have been implicated in the control of eye development, none have been reported to be involved in cell cycle regulation. Because of the number of shared functional roles, and the amount of similarity in the structure of each member of the *Meis* family, it would not be surprising if cell cycle regulation is also a shared function. This ability would explain why so many members of the *Meis* family have been demonstrated to be involved in the maintenance of progenitor populations, because progenitor cells are by definition proliferating.

The second novel function that has been attributed to *Meis1* is the ability to recruit the transcription factor MyoD to promoter regions of differentiating myoblasts and myotubes in culture (Berkes et al., 2004). In this context, Meis1 binds to a binding site, prior to the induction of *MyoD* expression, in a region of high chromosomal compaction

(Berkes et al., 2004). This finding, combined with the high degree of structural and functional conservation in the *Meis* family, has prompted some groups to hypothesize that *Meis* genes may function as master regulatory factors. The idea is that *Meis* genes can interact with inactive, compacted chromosomal regions, and remodel the chromosomal structure. This means that *Meis* genes can function as "homing beacons" for additional transcription factors marking promoters for activation, modulating differentiation (Berkes et al., 2004; Sagerstrom, 2004).

In support of this idea the Zhang group has shown that *Pax*6 has evolved a novel function as the initial neuroectoderm determinant in primates, prior to *Sox1* (Pankratz et al., 2007; Zhang et al., 2010). They have also noticed that just after *Pax*6 induction *Meis2* and *Meis1* are upregulated as well (Pankratz et al., 2007; Zhang et al., 2010). The expression of *Pax*6, *Meis2*, and *Meis1* is then followed by increased expression of anterior neural markers. These findings prompted the hypothesis that the novel *Pax*6 role in primate development may be tied to the significant increase in primate brain size (Pankratz et al., 2007; Zhang et al., 2010). If indeed *Meis* genes are expressed earlier in neural development in primates than other vertebrates, this could have significant evolutionary implications. Because *Meis* genes have been hypothesized to function as "homing beacons" for other transcription factors (Berkes et al., 2004; Sagerstrom, 2004), to maintain progenitor populations through cell cycle control (Bessa et al., 2008; Heine et al., 2008; Hisa et al., 2004), and to up-regulate anterior specific neural markers, an early expression of *Meis* genes could feasibly result in vastly increased anterior neural structure size in primates. This idea is further supported by the identified correlation of species specific duration of *Meis* expression in the developing retina, and species specific retinal

size (Heine et al., 2008; Hisa et al., 2004). All of these correlations point to the divergence of the *Meis* family in vertebrate evolution being necessary for the evolution of increasingly complex neural structures over evolutionary time.

While much is known about the function of the different *Hox* gene clusters, and their TALE family cofactors *Pbx* and *Meis*, little is known in regards to the mechanisms by which *Meis* genes are regulated during development. Every developmentally regulated gene is controlled by complex regulatory mechanisms (Davidson, 2006). The simplest form of regulation is through a gene's promoter. The promoter is a short nucleotide sequence where the basic machinery for transcription is assembled. The promoter is located upstream of, and in close proximity to the transcription start site, where it is bound by general transcription factors (Carroll et al., 2001).

Regulating the efficiency by which the general transcription factors interact with the promoter is imperative to ensure that developmentally regulated genes are only activated according to spatial and temporal necessity. This differential regulation is accomplished through short (hundreds of base pairs in length), non-coding DNA sequences called *cis*-regulatory elements (Carroll et al., 2008; Davidson, 2006). *Cis*-regulatory elements function by communicating with the promoter controlling transcriptional activity. This communication is facilitated by a physical interaction between transcription factors bound to the *cis*-regulatory element and the general transcription machinery located at the promoter (Carroll et al., 2008). This physical interaction is made possible by a looping of the DNA so that the *cis*-regulatory element is in the vicinity of the promoter (Tolhuis et al., 2002). Looping is necessary because *cis*-regulatory elements can be located thousands of nucleotides away from the promoter,

upstream or downstream of the gene that they control, on the opposite side of a neighboring gene, or even within an intron of either the gene the element is controlling or a neighboring gene (Allende et al., 2006; Carroll et al., 2008; Davidson, 2006; Dutton et al., 2008; Echelard et al., 1994; Kikuta et al., 2007).

Most developmentally regulated genes have multiple *cis*-regulatory elements associated with them (Carroll et al., 2008; Degenhardt et al., 2010; Kague et al., 2010; Kikuta et al., 2007; Tumpel et al., 2006) *Cis*-regulatory elements can act as enhancers, insulators, or repressors depending on the complement of transcription factors that bind to the element (Allende et al., 2006; Carroll, 2005; Carroll et al., 2008; Davidson, 2006; Echelard et al., 1994; Tumpel et al., 2006; Woolfe et al., 2005). They have even been shown to be able to control multiple genes (Duboule, 1993; Zerucha et al., 2000). However, the most important characteristic of *cis*-regulatory elements is that they function modularly. This means that each individual *cis*-regulatory element can function independently to drive expression of a developmentally regulated gene in a tissue, spatial, and temporal specific manner, while a second can drive expression of the same gene in another tissue at a different time. This modularity is the source of differential gene regulation during embryonic development (Carroll, 2005; Carroll et al., 2008; Davidson, 2006; Gompel et al., 2005; Tumpel et al., 2006).

In addition to differential regulation, c*is*-regulatory elements have played a major role during the course of evolution (Carroll, 2005; Carroll et al., 2008). As mentioned earlier, the majority of genes that control embryonic development are shared among all bilaterians. This means that differences in animal form are due to differences in the regulation of the genes that control their embryonic development, and not differences in

the genes that control embryonic development (Carroll, 2005; Carroll et al., 2008; King and Wilson, 1975). The beauty of *cis*-regulatory elements is that they can mutate with minimal effects on the developing organism. When a gene mutates the effects are usually pleiotropic and maladaptive. The protein produced by a mutant gene will usually not function properly, no matter where the mutant gene is expressed. But if a *cis*-regulatory element is mutated, the effect will alter the expression pattern of a gene, not the structure of the gene itself. Because of their ability to avoid the effects of a mutation to the coding region of the gene, it has been hypothesized that mutating *cis*-regulatory elements may be the primary mechanism for producing genetic variation in development (Carroll, 2005; Carroll et al., 2008; Gompel et al., 2005; Wittkopp et al., 2004).

Evolution by *cis*-regulatory modification can proceed in 2 possible ways. First, according to the idea of sub-functionalization in the DDC model of evolution (Figure 3), following duplication of a gene, negative mutation of one or several of a single paralog's associated *cis*-regulatory elements will place heavy selection on the maintenance of both genes in future generations in order to ensure that the appropriate expression pattern is maintained (Hurley et al., 2005; Kleinjan et al., 2008; Prince and Pickett, 2002; Taylor and Raes, 2004; Tumpel et al., 2006). The second way in which *cis*-regulatory modification functions in evolution is through neo-functionalization (Figure 3). This can occur after, or independent of, a duplication event. In either circumstance, a novel *cis*-regulatory element is randomly generated through mutation. The presence of a new *cis*-regulatory module can impart a novel expression pattern to the gene with which it is associated. If the new expression pattern is positively beneficial to the organism, it will

be selected for by natural selection, and maintained in future generations (Carroll, 2005; Cullen et al., 2004; Gompel et al., 2005; Rebeiz et al., 2009).

In further support of a role for *cis*-regulatory elements in evolution, it has been demonstrated that *cis*-regulatory elements mutate at much higher rates than the coding regions of the genes. This includes the observation that novel *cis*-regulatory elements appear at much greater frequencies than do gene duplication events (Carroll, 2005; Gompel et al., 2005; Rebeiz et al., 2009). So, while the DDC model of evolution represents a plausible and powerful mechanism for evolution, the opportunities for duplicate genes to influence the evolution of an organism are far less frequent than the evolution of novel *cis*-regulatory elements. This has caused some to speculate that it is the emergence of novel *cis*-regulatory elements that has given rise to the majority of variation observed in bilaterians (Gompel et al., 2005; Rebeiz et al., 2009). In spite of the importance of *cis*-regulatory elements to embryonic development, and evolution, identifying these regulatory elements has proven difficult.

One successful approach that is being utilized to identify *cis*-regulatory elements is phylogenetic footprinting (Allende et al., 2006; Santini et al., 2003). *Cis*-regulatory elements are often highly conserved across species (Allende et al., 2006; Kikuta et al., 2007; Santini et al., 2003). Phylogenetic footprinting makes use of the conserved nature of *cis*-regulatory elements; by comparing non-coding genomic sequences from multiple species associated with developmental genes, Highly Conserved Noncoding Elements (HCNEs) can be identified (Allende et al., 2006; Kikuta et al., 2007; Muller et al., 2002; Navratilova et al., 2009; Santini et al., 2003; Woolfe et al., 2005). By identifying HCNEs in multiple species separated by millions of years of evolution there is a high probability

that there is some sort of selective pressure maintaining the integrity of the HCNEs, and it has been shown that HCNEs often function as *cis*-regulatory elements (Allende et al., 2006; Kikuta et al., 2007; Muller et al., 2002; Navratilova et al., 2009; Santini et al., 2003; Woolfe et al., 2005; Zerucha et al., 2000).

After identification, HCNEs are commonly characterized through transgenic analysis of reporter gene activity. By linking a HCNE to a reporter gene driven by a minimal promoter, followed by introducing this expression cassette into an organism, the expression pattern of the reporter gene will recapitulate that of any gene that the HCNE regulates, if indeed the HCNE functions as a *cis*-regulatory element (Antonellis et al., 2008; Echelard et al., 1994; Gompel et al., 2005; Linney et al., 1999; Zerucha et al., 2000). A gene isolated from jellyfish, coding for Green Fluorescent Protein (GFP), is commonly used as a reporter gene in zebrafish transgenics because it can be easily visualized in the optically clear zebrafish embryos (Amsterdam et al., 1995; Linney and Udvadia, 2004).

Generating transgenic zebrafish originally involved injecting foreign DNA into a single celled embryo, and relied on spontaneous incorporation of the reporter construct into the genome (Stuart et al., 1988). This method proved to be highly inefficient with less than a 10% success rate reported, but in recent years transgenic efficiency has been vastly improved by using the Tol2 transposon to facilitate genome insertion (Kawakami, 2005; Kawakami et al., 2000; Kawakami et al., 2004; Stuart et al., 1988). Originally isolated from the medaka fish *Oryzias latipes*, Tol2 is a stable transposase that recognizes specific *cis* sequences, and uses them to facilitate genomic translocation (Kawakami et al., 1998; Kawakami and Shima, 1999). By flanking a reporter construct with Tol2 *cis*

sequences, and co-injecting the construct with mRNA that encodes the Tol2 transposase, up to 60% germ line transposition success rates can be achieved (Allende et al., 2006; Kawakami, 2005; Kawakami et al., 2004; Kotani and Kawakami, 2008; Urasaki et al., 2006). Since the advent of Tol2 mediated zebrafish transgenics, the system has been shown to function in *Xenopus*, chicken, and mice (Kawakami, 2007; Sumiyama et al., 2010). The system has been improved further by incorporating translocation sequences for Invitrogen's Gateway® Multisite and Invitrogen's Gateway® 2-way cloning systems (Fisher et al., 2006; Kwan et al., 2007). By using the Gateway® cloning system the molecular process of building complex transposable reporter constructs and developing transgenics has been vastly simplified.

As a result, by employing phylogenetic footprinting, it is possible to identify putative *cis*-regulatory elements in the form of HCNEs associated with *Meis2*. Once identified *in silico*, zebrafish transgenics can then be used to assess the functionality of HCNEs as *cis*-regulatory elements *in vivo*.

MATERIALS AND METHODS

Animal Care

Genetically controlled zebrafish strains AB, TU, and Mitfa, as well as non-genetically controlled wild-type strains were maintained at 27°C and on a 14-10 hr day-night light cycle in a Marine Biotech z-mod (Aquatic Habitats) closed system. All adult animals were fed Zeigler Adult Zebrafish Complete Diet dry food twice daily, and once daily live brine shrimp. Water conditions were monitored daily, with pH maintained above 7.2, and conductivity kept between 480 and 700 mS/m. Adult animals were housed in 1L aquaria housing no more than 6 individuals of approximately 50:50 male to female ratios. Females were not housed in solitude to prevent them from becoming egg-bound.

Fish were bred by placing 1 female with 2 males, or 2 females with 3 males, of the same strain, in breeding chambers the evening prior to desired breeding. Males and females were separated from each other by a physical divider to prevent premature breeding. The divider was pulled approximately 10 minutes after the system lights turned on, and the fish were allowed to spawn for 20 minutes. Upon successful spawning, the adults were relocated to a new aquarium, and the embryos were harvested by filtering the water with fine mesh to allow water flow through while catching the embryos without injury. The isolated embryos were subsequently washed with Reverse Osmosis (RO) water, and then transferred to a glass dish where they were kept in 1X Danieau Buffer (58 mM NaCl, 0.7 mM KCl, 0.4 mM $MgSO_4$, 0.6 mM $Ca(NO_3)_2$, 5 mM HEPES pH 7.6).

Embryos were reared in glass dishes for 5 days in a separate dry incubator at 27°C. At 5 dpf feeding with fine particulate dry food (Z-mod) commenced twice daily, and the fry were transferred to aquariums in the system where they were still housed in 1X Danieau Buffer until 20 dpf. During this time period 50% of the 1X Danieau Buffer in each aquarium was removed every 2 days with a turkey baster, and replaced with fresh 1X Danieau Buffer to reoxygenate and prevent nitrogen toxicity.

At 20 dpf fry were transferred from 1X Danieau Buffer to a gentle system water flow, gradually increasing the strength of flow over time as judged by size of the fry. As the fry developed, they were fed incrementally larger dry food (ZM-100, 200, 300, and 400 from Z-mod) twice daily according to size. As fry grew in size they were split into multiple aquaria, preventing overpopulation and feeding dominance, while promoting growth and development. Once individuals were large enough to be fed ZM-300 the developing fry began a once daily brine shrimp regiment. Upon reaching adulthood (2-3 months' time) fish were maintained as described previously.

HCNE Identification

HCNEs were identified by Allen Wellington and Ted Zerucha by scanning 1 million base pairs, upstream and downstream of the human *MEIS2* coding region, for known Hox transcription factor binding sites (ATTA) (Shen et al., 1997), making use of the publicly available human genome. Upon identification of known transcription factor consensus binding sites, NCBI's BLAST was used to determine if the surrounding sequences were conserved across species. If BLAST hits came back as highly conserved among multiple species, including human, mouse, and chicken (and zebrafish and

pufferfish when applicable), the sequences were then used for multiple sequence alignments to determine identity of each HCNE.

PCR Isolation of HCNEs from Genome

Each HCNE was isolated from genomic DNA by Polymerase Chain Reaction (PCR). Sequence and species specific primers (Table1) were designed for each HCNE and used for subsequent PCR reactions.

Table 1: PCR Primers and Sequences

Primer Name	Sequence
Dr-m2de1-3	GCTCATTATAAGGCCGTGCATG
Dr-m2de1-5b	TATACCATGGAGGTCGGGTTTAAAGGA
Mm-m2de1-3	ATGTCCTTGTCATCCCAG
Mm-m2de1-5	GCTTATCACTCTGTCACC
Mm-m2de2-3	CGACCCTAAGAAAAGCAG
Mm-m2de2-5	GAAGGATCAGCTGGTAC
Mm-m2de3-3	GAGATCTGTCTACTGTCC
Mm-m2de3-5	GGCTAAGAAGAAGGCATC
Mm-m2de4-3	TCTATCTCGATCATTCAG
Mm-m2de4-5	CTCCGTAATTTTATCCAC
M13 Forward	GTAAAACGACGGCCAG
M13 Reverse	CAGGAAACAGCTATGAC
5'-attB1-TOPO	GGGGACAAGTTTGTACAAAAAAGCAGGCT/GAGCTCGGATCCACTAGTAAC
3'-attB2-TOPO	GGGGACCACTTTGTACAAGAAAGCTGGGT/TCACTATAGGGCGAATTGGG
5'-attB2-TOPO	GGGGACCACTTTGTACAAGAAAGCTGGGT/GAGCTCGGATCCACTAGTAAC
3'-attB1-TOPO	GGGGACAAGTTTGTACAAAAAAGCAGGCT/TCACTATAGGGCGAATTGGG

One HCNE, (Mm m2de4), was isolated from the Mouse genome using *Taq* DNA polymerase (NEB M0273S), the primer pair mm-m2de4-3′ and mm-m2de4-5′, and standard *Taq* PCR solution concentrations and conditions [18.8 µl water, 0.5 µl dNTPs (10 mM), 2.5 µl Standard *Taq* 10X Buffer (NEB), 0.2 µl *Taq* Polymerase (NEB M0273S), 1.0 µl Mm-m2de4-5′ (50 ng/µl), 1.0 µl Mm-m2de4-3′ (50 ng/µl), and 1.0 µl

mouse genomic DNA (0.33 µg/µl); 1:30 min initial melt at 98.0°C, 35 cycles (30s 98.0°C melt, 30s 56.8°C annealing, and 2:00 min 72°C extension), 10:00 at 72°C completion and 4°C for ∞].

All other HCNEs (Dr m2de1, Mm m2de1, Mm m2de2, and Mm m2de3) were isolated from genomic DNA by gradient PCR using sequence and species specific primers [Dr m2de1 (dr-m2de1-3′ and dr-m2de1-5′), Mm m2de1 (mm-m2de1-3′ and mm-m2de1-5′), Mm m2de2 (mm-m2de2-3′ and mm-m2de2-5′), and Mm m2de3 (mm-m2de3-3′ and mm-m2de3-5′)] as well as proofreading Phusion® High-Fidelity DNA Polymerase (NEB M0530L). The preceding primer pairs (Table 1) were used for each PCR with standard Phusion® reaction conditions [16.3 µl water, 0.5 µl dNTPs (10 mM), 1.0 µl 3' primer (50 ng/µl), 1.0 µl 5' primer (50 ng/µl), 5.0 µl Phusion 5x HF Buffer (NEB), 1.0 µl mouse or zebrafish genomic DNA (0.33 µg/µl), and 0.2 µl Phusion® DNA Polymerase (NEB M0530L); 1:30 min initial melt at 95.0°C, 35 cycles (30s 98.0°C melt, 30s annealing (Table 2), and 1:00 min 72°C extension), 10:00 at 72°C completion and 4°C for ∞], and a logarithmic gradient of annealing temperatures (Table 2):

Table 2: Annealing Temperatures for Gradient PCRs

HCNE	Reaction Number	Temperature °C	Reaction Number	Temperature °C
Dr m2de1	1	58.9	7	64.4
	2	59.1	8	65.7
	3	59.7	9	66.9
	4	60.6	10	67.9
	5	61.8	11	68.6
	6	63.0	12	68.9
Mm m2de1	1	49.8	7	58.3
Mm m2de2	2	50.2	8	60.2
Mm m2de3	3	51.1	9	62.0
	4	52.5	10	63.5
	5	54.3	11	64.6
	6	56.2	12	65.1

After each PCR reaction a 5 µl sample of each product was run on a 1% agarose TBE gel with Ethidium-Bromide (0.3 µg/ml) at 100 mV for 1 hr in order to determine reaction success.

Cloning of HCNEs

Each HCNE was subsequently cloned into the pCR®2.1-TOPO® TA cloning vector after PCR isolation. TA cloning takes advantage of a 5′ Adenine (A) overhang that *Taq* DNA polymerase leaves by utilizing a linearized plasmid that ′has a 3′ Thymidine (T) overhang. By putting the compatible ends of the PCR product and the Plasmid together, the PCR product is ligated into the plasmid by a DNA ligase. Phusion® High-Fidelity DNA Polymerase does not however leave an A overhang, so 5′ A's must be added to each PCR product produced by Phusion® High-Fidelity DNA Polymerase before TA cloning. To add 5′ A's to the Dr m2de1, Mm m2de1, Mm m2de2 , and Mm m2de3 PCR products, 5 µl of each PCR reaction performed using Phusion® High-Fidelity DNA Polymerase was incubated at 72°C for 10 min with standard *Taq* DNA Polymerase and dATPs [38.5 µl water, 1.0 µl dATPs (10 mM), 5.0 µl Standard *Taq* 10X Buffer (NEB), and 0.5 µl *Taq* DNA Polymerase (NEB M0273S)].

Following the addition of A's, each HCNE PCR product was ligated into the pCR®2.1-TOPO® TA cloning vector according to standard TOPO protocol [4.0 µl PCR Product, 1.0 µl TOPO Salt Solution (Invitrogen), and 1.0 µl TOPO® Vector (Invitrogen KNM455001); incubated at Room Temperature (RT) for 10 min], followed by Heat Shock transformation: 50% volume of ligation is added to One Shot® TOP10 Chemically Competent *E. coli* Cells (Invitrogen C4040-10), incubated on ice for 20 min, incubated at 42°C for 45s, incubated on ice for 2 min. The cells were immediately transferred to 1 ml

SOC [20 g/L bacto-tryptone (BD 211705), 5 g/L bacto-yeast extract (BD 212750), 0.5 g/L NaCl, 20 mM glucose], recovered at 37°C in a shaking incubator at 200 RPM for 1 hr, plated on LB-Ampicillin (100 mg/µl) (Amp) plates with 50 µl X-Gal (20 mg/µl) for secondary selection, and incubated at 37°C overnight.

After overnight incubation, 10 white colonies were screened by PCR with standard *Taq* DNA polymerase for presence of correct insert. Each colony was sampled with a sterile toothpick, suspended in 25 µl RO water, vortexed, and 4.0 µl of the resulting solution was placed into each PCR reaction [15.8 µl water, 1.0 µl M13 Forward (50 ng/µl), 1.0 µl M13 Reverse (50 ng/µl), dNTPs (10 mM), 2.5 µl Standard *Taq* 10X Buffer (NEB), 0.2 µl *Taq* Polymerase (NEB M0273S)]. [10:00 minute initial 98.0°C cell lysis step, 35 cycles (30s 98.0°C melt, 30s 56.0°C annealing, and 2:00 min 72°C extension), 10:00 at 72°C completion and 4°C for ∞]. After each PCR reaction a 5 µl sample of each product was run on a 1% agarose TBE gel as previously described.

One to three positive colonies containing an insert of the predicted size were picked and grown overnight in 3.0 ml LB/Amp (100 mg/ml) liquid media at 37°C and 200 RPM. The following day minipreps were performed on each overnight culture using the PureYield^TM Plasmid Miniprep System (Promega A1223), eluting in RO water, to isolate the plasmid DNA [cultures were transferred to microcentrifuge tubes where cells were pelleted by centrifugation at maximum speed for 2 min; supernatant was discarded and cells resuspended in 200 µl of Cell Resuspension Solution by pipetting up and down; 200 µl of Cell Lysis Solution was added and inverted to mix thoroughly; 200 µl of Neutralization Solution was added and inverted to mix; lysate was centrifuged at maximum speed for 5 min; during centrifugation, 1.0 ml of Resin was added to each

minicolumn/syringe complex; supernatant was transferred to the syringe and plunged discarding flow through; 2 ml of Column Wash solution was added to each miniprep and plunged discarding flow through; the minicolumn was placed in a microcentrifuge tube and centrifuged at maximum speed for 2 min discarding flow through; the minicolumn was placed in a new microcentrifuge tube and 50 µl of RO water or TE (10 mM Tris·Cl, 1 mM EDTA pH 7.4) was added, and centrifuged at maximum speed for 20 seconds (s); DNA was quantified by spectrophotometry and stored at ⁻20°C]. Each plasmid was screened for the presence of an insert of the appropriate size by restriction digest for 2 hours at 37°C using the Restriction Endonuclease EcoRI [5.0 µl of plasmid DNA, 2.0 µl EcoRI 10X Buffer (NEB), 1.0 µl EcoRI (NEB R0101S), and 12.0 µl water]. Each digest screen was run on a 1% TBE gel (as previously described) in order to visualize insert size. After insert size confirmation, the sequence of each insert was confirmed by Sanger DNA sequencing (performed by Cornell University's Life Sciences Core Laboratories for using universal M13 Forward and M13 Reverse primers [Table 2]).

Generation of Transgenic Reporter Constructs

Using Invitrogen's Gateway® 2-Way cloning system modified to work with Tol2 (Figure 4) the HCNEs, Dr m2de1 and Mm m2de1, were cloned into the plasmids pDONR221 and pGW_cfosGFP (Fisher et al., 2006). pDONR221 and pGW_cfosGFP were transformed into One Shot® ccdB Survival™ 2 T1R Competent Cells (Invitrogen A10460) by heat shock (as previously described). pDONR221 was grown on LB-Chloramphenicol (30 mg/µl) + Kan (50 mg/µl), and pGW_cfosGFP was grown on LB-Chloramphenicol (30 mg/µl) + Amp (100 mg/µl) (Fisher et al., 2006; Kwan et al., 2007).

Each HCNE was amplified from the pCR®2.1-TOPO® plasmid by PCR with Phusion® DNA Polymerase [16.3 µl water, 0.5 µl dNTPs (10 mM), 1.0 µl 3' primer (50 ng/µl), 1.0 µl 5' primer (50 ng/µl), 5.0 µl Phusion 5x HF Buffer (NEB), 1.0 µl plasmid DNA (50 pg/µl), and 0.2 µl Phusion® DNA Polymerase (NEB M0530L); 1:30 minute initial melt at 95.0°C, 35 X (30s 95.0°C melt, 30s annealing (55.8°C or 62.9°C), and 1:00 min 72°C extension), 10:00 at 72°C completion and 4°C for∞], using both of the following 2 primer pairs individually (Table 1): 5′-attB1-TOPO and 3′-attB2-TOPO or 5′-attB2-TOPO and 3′-attB1-TOPO. PCR reactions using 5′-attB1-TOPO and 3′-attB2-TOPO were annealed at 55.8°C and PCR reactions using 5′-attB2-TOPO and 3′-attB1-TOPO were annealed at 62.9°C. Each primer is sequence specific for the pCR®2.1-TOPO® plasmid, with the 5′primer recognizing the 5′ polycloning region and the 3′ primer recognizing the 3′ polycloning region. Each primer has either an attB1 or an attB2 site (Figure 4 and Table 3) flanking the recognition sequence so that the resulting PCR product contains the HCNE flanked on both sides by an attB1 and an attB2 site (Figure 4 and Table 3). A 5 µl sample of each PCR was run on a 1% TBE gel as previously described to confirm that a product of appropriate size was amplified.
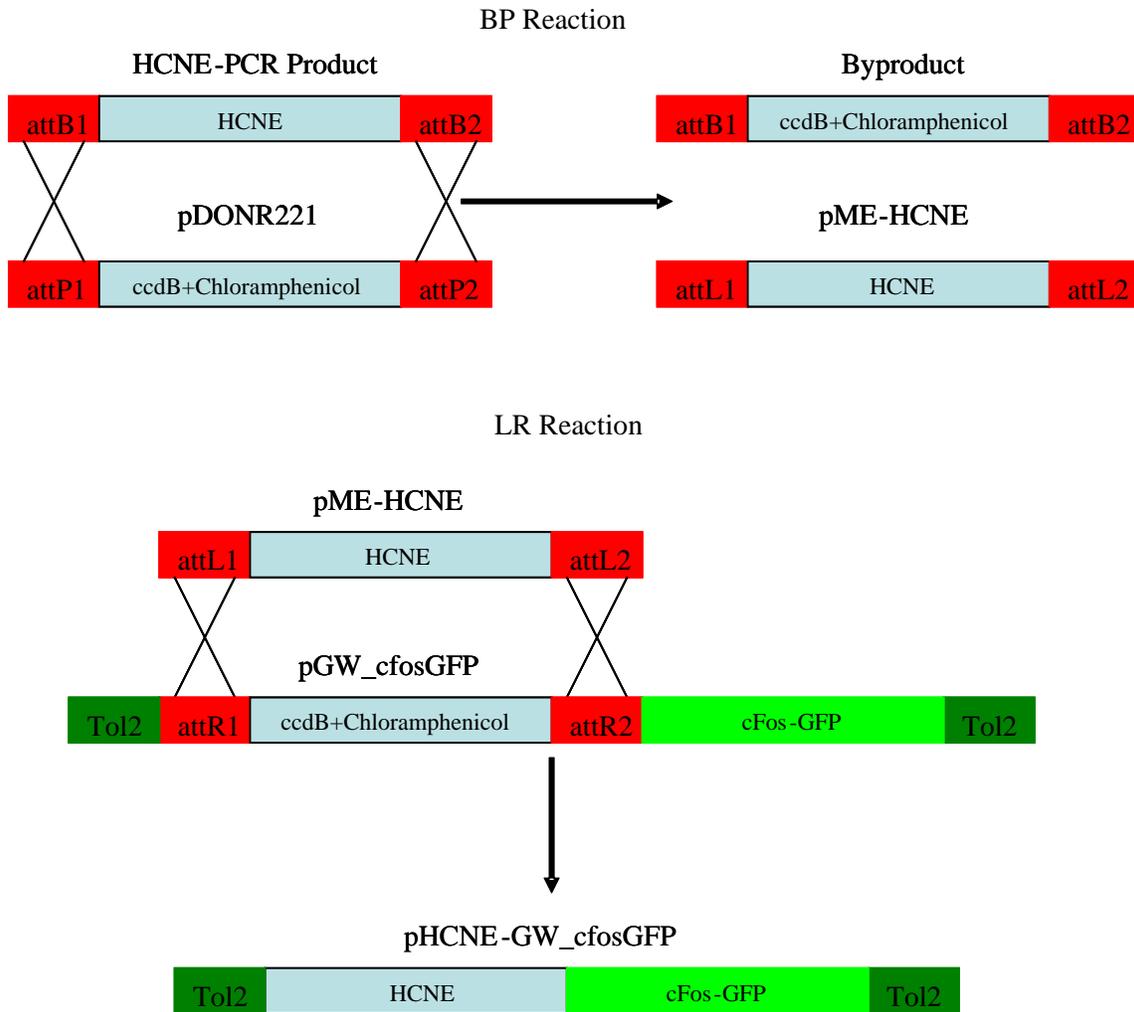
Figure 4: Schematic Diagram of Gateway® 2-way Cloning System
BP Reaction: The BP reaction recombines a PCR product flanked by attB1 and attB2 cloning sites (Table 3) with the donor vector pDONR221. The BP Clonase recognizes compatible sequences between attB1-attP1 and attB2-attP2, and translocates the sequence between attB1-attB2 into the donor vector, generating a middle entry vector.
LR Reaction: The LR reaction recombines a middle entry vector with a destination vector (pGW_cfosGFP) by recombining attL1-attR1 and attL2-attR2 cloning sites (Table 3). The LR Clonase recognized compatible sequences between attL1-attR1 and attL2-attR2, and translocates the sequence between attL1-attL2 into the destination vector, generating an expression construct.

Table 3: att Cloning Site Sequences and Sources

Att Site	Sequence	Source
attP1	AAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAC ATTGATGAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAGC TGAACGAGAAACGTAAAATGATATAAATATCAATATATTAAATTAG ATTTTGCATAAAAAACAGACTACATAATACTGTAAAACACAACATAT CCAGTCACTATGAATCAACTACTTAGATGGTATTAGTGACCTGTA	pDONR221
attP2	AATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAAT TGATAAGCAATGCTTTCTTATAATGCCAACTTTGTACAAGAAAGCTG AACGAGAAACGTAAAATGATATAAATATCAATATATTAAATTAGATT TTGCATAAAAAACAGACTACATAATACTGTAAAACACAACATATCC AGTCACTATGAATCAACTACTTAGATGGTATTAGTGACCTGTA	pDONR221
attL1	CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACA MATTGATGAGCAATGCTTTTTTATAATGCCAACTTTGTACAAAAAAG CAGGCT	pME-Dr/Mm-m2de1
attL2	ACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTGCTTA TCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCAT TATTTG	pME-Dr/Mm-m2de1
attR1	CAAGTTTGTACAAAAAAGTTGAACGAGAAACGTAAAATGATATAAA TATCAATATATTAAATTAGATTTTGCATAAAAAACAGACTACATAAT ACTGTAAAACACAACATATGCAGTCACTATGAATCAACTACTTAGAT GGTATTAGTGACCTGTA	pGW_cfos GFP
attR2	TACAGGTCACTAATACCATCTAAGTAGTTGRTTCATAGTGACTGCAT ATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTA ATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAACTTTC TTGTACAAAGTGG	pGW_cfos GFP
attB1	CAAGTTTGTACAAAAAAGCAGGCT	5'-attB1-TOPO and 3'-attB1-TOPO
attB2	ACCCAGCTTTCTTGTACAAAGTGG	5'-attB2-TOPO and 3'-attB2-TOPO

Each attB1/2 flanked HCNE PCR product was subsequently purified with the QIAquick PCR Purification Kit (QIAGEN 28104) according to the manufacturer instructions [5 volumes of Buffer PB was added to 1 Volume of reaction; 1:250 volume of pH indicator I was added to the reaction and inverted to mix; proceeded if solution was orange/yellow; the mixture was transferred to a column and spun at maximum speed for 1 min, discarding flow through; 750 µl of Buffer PE was added to the column and spun at

maximum speed for 1 min, discarding flow through; the column was spun again at maximum speed for 2 min; the column was transferred to a new microcentrifuge tube and 30 µl of TE (10 mM Tris·Cl, 1 mM EDTA pH 7.4) was added; the column was allowed to incubate at room temperature for 1 min, and spun at maximum speed for 1 min; elute was stored at ⁻20°C], and concentration was then quantified by spectrophotometry.

After each PCR product was purified they were inserted into the donor plasmid pDONR221, generating middle entry vectors (pME-HCNE) via a BP reaction (Figure 4), using the Gateway® Clonase II enzyme mix (Invitrogen 11789020) [25 fmol pDONR221, 25 fmol PCR Product, 2.0 µl BP Clonase[TM]II, to a final volume of 5.0 µl with TE (10 mM Tris·Cl, 1 mM EDTA pH 7.4)]. The volume for each DNA component was calculated by determining the nanogram (ng) equivalent for each DNA component that corresponds to 25 fmol using the following equation, then calculating the volume based on known concentration:

$$ng = 25 \text{ fmol} \cdot \text{size bp} \cdot (660 \text{ fg/fmol}) \cdot (1 \text{ ng}/10^6 \text{ fg})$$

The reaction was allowed to run at room temperature overnight, was then transformed into Chemically Competent DH5α *E. coli* by Heat-Shock transformation as previously described, and grown on LB-Kan (50 mg/µl) plates at 37°C overnight.

Following successful transformation, 2 colonies were used to inoculate 3 ml of liquid LB-Kan (50 mg/µl) cultures which were grown overnight in a shaking incubator at 37°C and 200 RPM. The following day PureYield[TM] Plasmid Minipreps (Promega A1223) were performed on each culture using TE (10 mM Tris·Cl, 1 mM EDTA pH 7.4) to elute the plasmid DNA (as previously described). Each plasmid prep was then tested by restriction digestion for 1 hour at 37°C with the Restriction Endonucleases EcoRI and

BglII [16.0 µl water, 1.0 µl plasmid DNA, 2.0 µl 10X EcoRI Buffer (NEB), 0.5 µl BglII (NEB R0144S), and 0.5 µl EcoRI (NEB R0101S)], followed by visualization on a 1% TBE gel (as previously described) to confirm the presence of an insert of the predicted size.

Upon confirmation of pME-(HNCE) generation, the plasmid was used to transfer the HCNE into the destination vector pGW_cfosGFP by an LR Reaction using the LR Clonase[TM] II Plus Enzyme Mix (Invitrogen 12538120), generating reporter constructs p(HCNE)-cfosGPF [10 fmol pDONR221, 10 fmol PCR Product, 2.0 µl LR Clonase[TM]II Plus, to a final volume of 5.0 µl with TE (10 mM Tris·Cl, 1 mM EDTA pH 7.4)]. The volume for each DNA component was calculated by determining the nanogram (ng) equivelant for each DNA component that corresponds to 25 fmol using the following equation, then calculating the volume based on known concentration:

$$ng = 10 \text{ fmol} \cdot \text{size bp} \cdot (660 \text{ fg/fmol}) \cdot (1 \text{ ng}/10^6 \text{ fg})$$

The reaction was allowed to run overnight at room temperature. After overnight incubation the LR reaction was transformed into One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen C4040-10) by Heat-Shock transformation, as previously described. Following successful transformation, 2 colonies were used to inoculate 3 ml of liquid LB-Amp (100 mg/µl) and incubated overnight in a shaking incubator at 37°C and 200 rpm. The following day, PureYield[TM] Plasmid Minipreps (Promega A1223) were performed as previously described, using RO water to elute the plasmid DNA. Resultant DNA was stored at ‾20°C. Each plasmid was then tested by restriction digestion [15.0 µl water, 2.0 µl plasmid DNA, 2.0 µl 10X EcoRI Buffer (NEB), and 1.0 µl EcoRI (NEB R0101S)] at 37°C for 1 hour with the Restriction Endonuclease EcoRI,

followed by visualization on a 1% TBE gel (as previously described) to confirm the presence of an insert of the correct size.

Upon successful LR reaction transformation the construct was prepared for future microinjection by purification using the PureYield™ Plasmid Maxiprep System (Promega A2392): Cells were grown in 500 ml liquid LB-Amp (100 mg/µl) overnight; cells were pelleted at 5000 Xg for 10 min at room temperature discarding the supernatant; cells were resuspended in 15 ml of Cell Resuspension Solution by gently pipetting up and down until no clumps were left; cells were lysed by adding 15 ml of Cell Lysis Solution and mixed by inversion; the reaction was allowed to run until solution is clear; 15 ml of Neutralization Solution was added and the solution was mixed by inverting; the lysate was cleared by spinning for 15 min at 14,000 Xg at room temperature; the lysate was filtered through an autoclaved coffee filter into 100 ml graduated cylinder, 0.5 volume of room temperature isopropyl alcohol (EMD PX1835-9) was added to the solution which was mixed by inversion; DNA was precipitated by spinning for 15 min at 14,000 Xg at room temperature, discarding the supernatant; the pellet was then resuspended with 2 ml of TE (10 mM Tris·Cl, 1 mM EDTA pH 7.4); 10 ml of DNA Purification Resin was added to the solution, swirling to mix; the DNA/Resin mix was transferred to a maxicolumn attached to a vacuum manifold, applying maximum vacuum; 25 ml of Column Wash Solution was added and cleared from the column by applying a vacuum, not allowing vacuum to continue once dry; 5 ml of 80% ethanol was added, and cleared from the maxicolumn by applying max vacuum; the vacuum was allowed to continue for 1 additional minute; the maxicolumn was transferred to a self provided 50 ml conical tube, centrifuged in swinging bucket rotor for 5 min at 1,300 Xg and room temperature;

the resin was dried by applying maximum vacuum for 5 min; 1.5 ml of preheated 65°C RO water was added to the maxicolumn and centrifuged for 5 min at 1,300 Xg at room temperature in a swinging bucket rotor; the elute was transferred to a syringe/filter apparatus and expelled into 1.5 ml microcentrifuge tube; the elute was spun at 14,000 Xg for 1 min; the supernatant was transferred to a new microcentrifuge tube and stored at ⁻20°C.

Production of Transposase mRNA

Transposase capped mRNA was transcribed using mMESSAGE mMACHINE® SP6 RNA Transcription Kit (Ambion® AM1340M). To prepare the template for transcription 20 µg of the plasmid pCS2FA-transposase (generously donated by Dr. Chi-Bin Chien) was digested with the restriction endonuclease NotI (NEB R0189S) over night at 37°C [68.19 µl water, 12.81 µl (20 µg) plasmid DNA, 10.0 µl 10X NEB Buffer 3 (NEB), 1.0 µl BSA, and 8.0 µl NotI (NEB R0189S)]. The enzyme was subsequently heat-killed by incubating the solution at 65°C for 25 min, and the linearized plasmid was then purified using the QIAquick PCR Purification Kit (QIAGEN 28104) as previously described.

The linearized pCS2FA-transposase template was then transcribed under complete RNase free conditions using mMESSAGE mMACHINE® SP6 RNA Transcription Kit (Ambion® AM1340M) [The 10X Reaction Buffer, 2X NTP/CAP, Nuclease Free Water, and Enzyme Mix were all removed and briefly spun to remove RNases; the Enzyme Mix and 2X NTP/CAP were placed on ice to thaw, while the 10X Reaction Buffer and Nuclease Free Water were left at room temperature; once thawed the 2X NTP/CAP and 10X Reaction Buffer were vortexed until any precipitated particles

were completely dissolved, then the solutions were returned to previous conditions; the reaction was assembled in a microcentrifuge tube at room temperature in the following order: 0.66 µl Nuclease Free Water, 5.34 µl pCS2FA-transposase DNA (2.0 µg), 10.0 µl 2X NTP/CAP, 2.0 µl 10X reaction Buffer, and 2.0 µl Enzyme Mix; the reaction was mixed by gently flicking, then spun briefly to aggregate reagents at the bottom of the microcentrifuge tube, and incubated in a 37°C water bath for 2 hrs; following incubation 1.0 µl of TURBO DNase was added, the solution mixed well, and the reaction was incubated in a water bath at 37°C for an additional 15 min; after incubation 30.0 µl of Nuclease Free Water and 30.0 µl of LiCl Precipitation Solution were added, and the solution was incubated at ⁻20°C for 1 hr; after incubation the reaction was centrifuged at 4°C for 15 min at maximum speed; the supernatant was removed and the pellet was washed with 1.0 ml of 75% Nuclease Free Ethanol, followed by centrifugation at 4°C for 10 min at maximum speed; the supernatant was then poured off and the tube inverted to gravity dry the pellet; the pellet was then resuspended in 30 µl Nuclease Free Water; mRNA concentration was quantified by spectrophotometry; mRNA was diluted to 175 ng/µl, aliquoted into 10 µl aliquots, and store at ⁻80°C].

Once aliquoted, 2.0 µl of mRNA was run on a 1%TBE gel (as previously described) to make sure that the product is of the appropriate size and not degraded.

Microinjections

For microinjections, zebrafish were bred as described earlier. Embryos were harvested 20 min after the divider was pulled and stored in 1X Danieau Buffer. Microinjections were performed on 1-4 cell embryos using a Nanoliter 2000 (Model

B203XVY) Microinjector mounted on a Marhauser MMJR Micromanipulator, both from World Precision Instruments.

Prior to injection, needles were created by pulling 3.5 nanoliter glass capillary tubes (World Precision Instruments 4878) with a David Kopf Instruments Vertical Pipette Puller (Model 700C) set at Heat 85 and Solenoid 0. Each needle was examined under a compound microscope with a micrometer to ensure that the needle point had a long taper with a width of 8 µm (Linney and Udvadia, 2004). Needle points were then broken with a neatly occluding pair of micro-tweezers, and then back filled with mineral oil. The needle was mounted in the Nanoliter 2000 Microinjector, and the mineral oil was expelled, followed by front loading the injection solution (125 ng Transposon Plasmid, 175 ng Transposon RNA, 2.0 µl of 0.5% Phenol Red (Sigma P0290), and RNase-free water to a final volume of 5.0 µl) (Fisher et al., 2006), leaving a non-compressible mineral oil buffer between the injection solution and the injection rod.

Approximately 50, 1-4 cell embryos were then lined up along a microscope slide sitting on the top of a bottom Petri dish and viewed under a dissecting microscope. The needle was then positioned next to the line of embryos with a slight incline (~20°), and the Petri dish was physically moved plunging the needle into the embryo's yolk just below the blastomeres. Four nanoliters (nl) of the red Tol2 injection solution was injected into the yolk immediately below the blastomeres. The process was then repeated, injecting between 100 and 200 embryos total per experiment.

After injections, embryos were housed at 27°C in 1X Danieau Buffer (58 mM NaCl, 0.7 mM KCl, 0.4 mM $MgSO_4$, 0.6 mM $Ca(NO_3)_2$, 5 mM HEPES pH 7.6). Three hours after injection each embryo was examined for successful injection under a

dissecting microscope, removing any dead embryos, or embryos that did not contain observable quantities of phenol red. The remaining embryos were allowed to develop until 24 hpf.

Imaging of Injected Embryos

Injected embryos were imaged using a Zeiss LSM 510 Confocal Microscope. Using the Argon laser and FITC filter to image GFP, between 8 and 15 embryos at 24 hpf were placed in a concave microscope slide and viewed using the 10X objective individually to screen for expression of GFP.

RESULTS

Identification of HCNEs

Bioinformatic analysis of the human *MEIS2* locus revealed 4 HCNEs that have been named <u>Meis2</u> <u>d</u>ownstream <u>e</u>lement (m2de) 1-4 (Wellington and Zerucha unpublished). The first HCNE, m2de1, is found approximately 220 Kb downstream of *Meis2* in mice and humans, 120 Kb downstream of *Meis2* in chicken, and 40 Kb downstream of *Meis2.2* in zebrafish. The other 3 HCNEs (m2de2, m2de3, and m2de4) are conserved in a downstream position in relation to *Meis2* in all tetrapods examined to date (Figure 5). The second HCNE, m2de2, is located approximately 210 Kb downstream of *Meis2* in mice and humans, and 130 Kb downstream of *Meis2* in chicken. The third HCNE, m2de3, is found approximately 250 Kb downstream of *MEIS2* in humans, 260 Kb downstream of *Meis2* in mice, and 150 Kb downstream of *Meis2* in chicken. The last HCNE, m2de4, is located approximately 440 Kb downstream of *MEIS2* in humans, 415 Kb downstream of *Meis2* in mice, and 215 Kb downstream of *Meis2* in chicken.

The first HCNE, m2de1, is 255 nucleotides in length and is highly conserved in sequence across all vertebrates examined, with 64% conservation between zebrafish and humans (Figure 6). Within m2de1 there are 2 conserved Hox binding sites (ATTA / TAAT) (Shen et al., 1997), 1 conserved Pbx binding site (ATCA / TGAT) (Chang et al., 1997), and 1 conserved Meis2 binding site (CTGTC / GACAG) (Shen et al., 1997). The other 3 HCNEs are highly conserved in the tetrapod lineage but have not been identified

in zebrafish to date. The second HCNE, m2de2, is approximately 260 nucleotides in length, with 65% conservation between human and chicken, a conserved transcription factor binding site for Meis proteins (CTGTCA / TGACAG), and a Hox transcription factor binding site (ATTA / TAAT) (Figure 7) (Shen et al., 1997). The third HCNE, m2de3, is approximately 975 nucleotides in length, 85% conserved between human and chicken, with 1 conserved Hox binding site (TAAT / ATTA) (Shen et al., 1997), and 1 conserved Pbx binding site (TGAT / ATCA) (Figure 8) (Jacobs et al., 1999). The final HCNE, m2de4, is approximately 520 nucleotides in length, 65% conserved between human and chicken, and contains 5 conserved Hox transcription factor binding sites (TAAT / ATTA) (Shen et al., 1997) (Figure9).
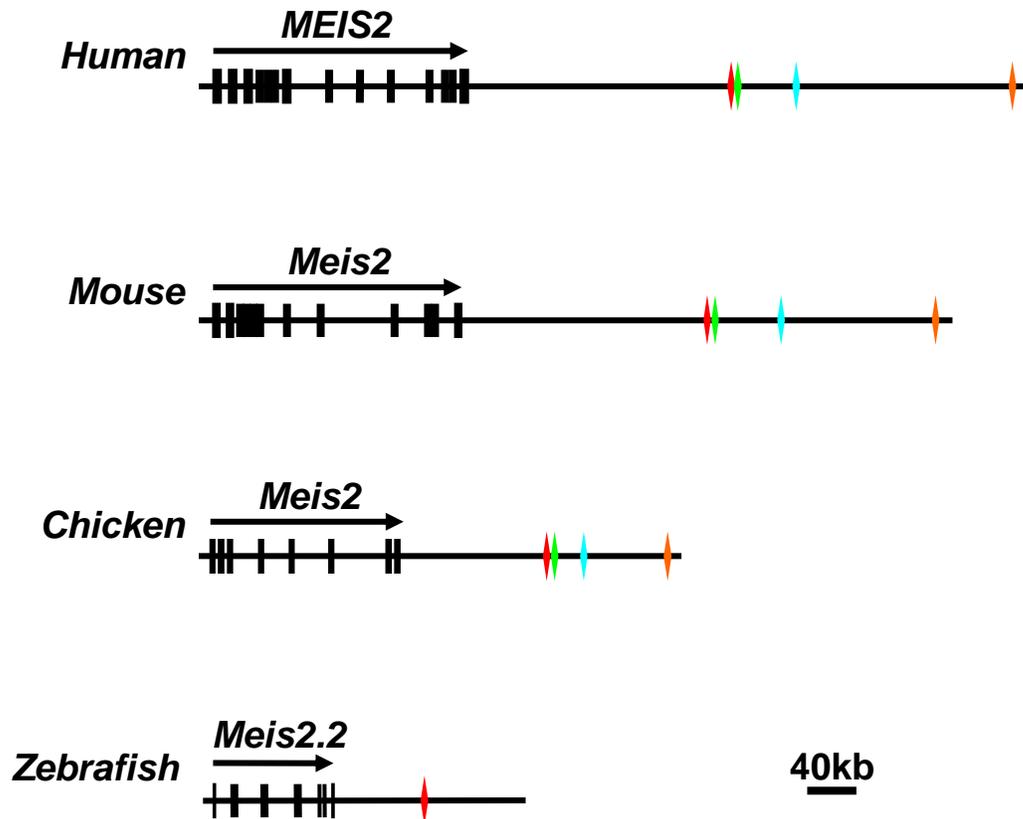
Figure 5: Position and Orientation of all 4 HCNEs relative to *Meis2*
Positional alignment of each HCNE relative to *Meis2* in Human, Mouse, Chicken, and the zebrafish homolog *Meis2.2*: Red is m2de1, Green is m2de2, Blue is m2de3, Orange is m2de4.

Figure 6: Multiple Sequence Alignment of HCNE m2de1

Multiple sequence alignment of m2de1: Gg is *Gallus Gallus* (Chicken), Mm is *Mus musculus* (Mouse), Hs is *Homo sapiens* (Human), Dr is *Danio rerio* (zebrafish), Tr is *Takifugu rubripes* (Pufferfish). Sequences in red are conserved across all vertebrates, light blue are conserved in zebrafish and pufferfish, green are conserved in land vertebrates, yellow are conserved in land vertebrates and pufferfish. Binding sites: Purple is a known Meis2 binding site, pink is a known Hox binding site, dark blue is a known Pbx binding site.

```
Ggm2de2  attacatacatttgac-taaataaagaaacat-catcta--ggtgtgatt-atat-aagtaag
Mmm2de2  ataaaaagtaagtaatgtg-acctagg-actgacag-tacagctgt-atcgaagggaagaaaa
Hsm2de2  ataaaaagaaagtaatgtg-atctagg-acagacag-taaagctgt-atctagg--aag-aaa

Ggm2de2  tcctgtgtaaatttgaaattgaatgtagtaactgt-aagtatctaagaggaagtttattactc
Mmm2de2  tcctatggaaggctgaaattgaacatggtaactgc-aattatctaaaaggaagtctattacct
Hsm2de2  tcctatggaaggttgaaattcaatgtggtaactgccaattatctaaaaggaagtctattactt

Ggm2de2  catgataaaactgtcaaaataatgactgggg-ctcctacaaatcgtgtcgaagatcaataaag
Mmm2de2  tctgataaaactgtcaaaataatgatttggtactcccacaaattgttcagagaatcaataaag
Hsm2de2  tatgataaaactgtcaaaataatgattgggtactcctacaaattgtacagagaatcaataaaa

Ggm2de2  ttagtcc--tat---gtcatagcttgagaatttagaaattgtattttgactt-taattca-ca
Mmm2de2  -----ccaatatttcgtcacagtttgagaatttaggaattccactttt--ctcctaagtcacaa
Hsm2de2  -----ccaatattacgtcacagtttgagaatttaggaattccacttttt-cttcttattcatga

Ggm2de2  gtgcagac
Mmm2de2  gagaatgc
Hsm2de2  gaagacgc
```

Figure 7: Multiple Sequence Alignment of HCNE m2de2
Multiple sequence alignment of m2de2: Gg is *Gallus Gallus* (Chicken), Mm is *Mus musculus* (Mouse), Hs is *Homo sapiens* (Human). Sequences in green are conserved in land vertebrates. Binding sites: Purple is a known Meis2 binding site, pink is a known Hox binding site.

```
Ggm2de3  gtgtaggtcatctccgagaagctgtgacagtgatccaaggttcctcagctgtgcagtgtacataacacaga
Mmm2de3  gtgcgggtcatcttcgtgatgttctgacagtgagctggggctcctcagctgtgccgtgtacataacacaga
Hsm2de3  gtgcaggtcagctccacgatgttctgacagtgatccagggctcctcagctgtgcagtgtacataacacaga

Ggm2de3  tgttgggagagtcatctgtcagctatgcctctcctatggagttagagttctaggggggaaaagaattcgaca
Mmm2de3  tgtt-ggggagtcatctgtcagccatgcctctcctatggagttaaagttctagagggaaaagaatttgaca
Hsm2de3  tgtt-ggggagtcatctgtcagccatgcctctcctatggagttaaagttctagagggaaaagaatttgaca

Ggm2de3  cagtgctgtcactgttacattttcacaatggaagagcatgcaaatattcaacctcctgcttagcagtagtg
Mmm2de3  ccctgctgtcactgttacattttcacaatagaagagcatgcaaatattgaacctcctgcttagccacagtg
Hsm2de3  cagtgctgtcactgttacattttcacaatagaagagcatgcaaatattgaacctcctgcttagcaatagtg

Ggm2de3  ccctgtctaatgtctgtgctggttagaataaaaaaatagtgggtgcatctagggctcatagatctgacaaa
Mmm2de3  tcctgtctaatgtctgtgctggttagaataaaaaaatca-tgggtgcatcttgggctctgagatccgacaaa
Hsm2de3  ccctgtctaatgtctgtgctggttagaataaaaaaatca-tgggtgcatcttgggctcagagatctgacaaa

Ggm2de3  ggtcatgct----atgctgggcatttgaattggaaattgtctggacttagattttataaagctcctactgt
Mmm2de3  ggtcatgctg---tt---gggcatttgaattggaaattgtctggacgtagattttataaagctcccactgt
Hsm2de3  ggtcatgctatgt-t---gggcatttgaattggaaattgtctggacttagattttataaagctcccactgt

Ggm2de3  tgtc---ggggaaaatttcagcagggttttgtccctagagaggcctcttaggcccttttctgtcctatgaa
Mmm2de3  tgtgtgtggggagaatttctacagggttttgtccctggagaggcctcttaggcccttg-ctgtcctatgaa
Hsm2de3  tgtgtgtggggagaatttcagcagggttttgtccctggagaggcctcttaggccctttttctgtcctatgaa

Ggm2de3  tgaatttagatgtgagggtttct-ctgccttaaaactgttaagttcattttgcatacatccccagagagaa
Mmm2de3  tgaatttagatgtgagggttactcctgccttaaaactgttaagttcattttgcatacatccctagagagaa
Hsm2de3  tgaatttagatgtgagggttactcctgccttaaaactgttaagttcattttgcatacatccctagagagaa

Ggm2de3  aaactggcagatgctttttgtcttggaagtgtttaaaagaaaactgcagaacaggaacaagggagagaaggg
Mmm2de3  aaactggcagatgcttttatcttggaagtgtttaaaagaaaactgtagagcaggaacaagagaggaaaggg
Hsm2de3  aaactggcagatgctttttgtcttggaagtgtttaaaagaaaactgcagagcaggaacaagggagaaaaggg

Ggm2de3  gccctgtgtggagtcccagaacattttggaaatggccaatatgcagttttcatcagtactaaggcggggca
Mmm2de3  gccctgtgtggagtcccagaacattttggaaatggccaatatgcagttttcatcagtatgaaggcggggtg
Hsm2de3  gccctgtgtggagtcccagaacattttggaaatggccaatatgcagttttcatcagtatgaaggcggggtg

Ggm2de3  caatatggtgcctgtagctcactgggggaatatgaatgttgattaagcatactcccaggctttgaaatcctg
Mmm2de3  caatatggtgcctgtagcccaccaggggaatatgaatgttgattaagcattctcccaggctttgaaattctg
Hsm2de3  caatatggtgcctgtagttcacgagggaatatgaatgttgattaagcatactcccaggctttgaaattctg

Ggm2de3  aaagcggtgtcagaataatggatgttgagcaaatgtcaagcatttgttaatttctctgttatttggagtttt
Mmm2de3  aaagcagtgtcagaataatggatgttgagcaaatgtcaagcatttgttaatttctctgttatttggagtttt
Hsm2de3  aaagcagtgtcagaataatggatgttgagcaaatgtcaagcatttgttaatttctctgttatttggagtttt

Ggm2de3  atctaccatgatcaatcaaataaaactagtgcttctctgttgtggcatgtgtcattgatatggtgattaggg
Mmm2de3  atctaccatgatcaatcaaataaaactagtgcttctctcttgtggcatgtgtcattgatatggtgattaggt
Hsm2de3  atctaccatgatcaatcaaataaaactagtgcttctctcttgtggcatgtgtcattgatatggtgattaggt

Ggm2de3  ggctagagaggccttctgccttttttttttttttctttttttttttttttttttctgttacaggtaacagatt
Mmm2de3  ggctagagagaccttctgctttttttttttc--------------------------caggtaacacatt
Hsm2de3  ggctagagagaccttctgcttttttttc------------------------------caggtaacagatt

Ggm2de3  acatatgaacag-ccataactttaaaaactgttggtggttggtttaatggga
Mmm2de3  ataaatgagcagaacagagcttctaa-attttgtttatggtttaaatttgt
Hsm2de3  ataaacgagtag-acagagcttttcagttatttgtatattaactgaatacta
```

Figure 8: Multiple Sequence Alignment of HCNE m2de3

Multiple sequence alignment of m2de3: Gg is *Gallus Gallus* (Chicken), Mm is *Mus musculus* (Mouse), Hs is *Homo sapiens* (Human). Sequences in teal are conserved in land vertebrates. Binding sites: Dark blue is a known Pbx binding site, pink is a known Hox binding site.

```
Ggm2de4  gac--tttatgatcctttcactacagtaactagcctttattctcacaatcagcatcaaccata
Mmm2de4  tgat-tttatgattctttcgctgaagtaactagtctttattctgacaatcagcaccaatgaca
Hsm2de4  ttgactttatgattctttcactgaagtaactagttttttattttttgcaatcagcaccaatgaca

Ggm2de4  actccaagaggaaaaaaatga----ttaaaatctagagaatagtgaggagaggtttgtttgat
Mmm2de4  attt-gggaggggg tagg-gaattattaaaatctggagaaagaagaagcgctatttgt-ttga
Hsm2de4  att--tgggagggggtggggaattattaaaatctggagaaagaagaggcactgattatttgat

Ggm2de4  gtggtacaatg-cattcttgcaca—-cacttcttctgtataataggtactgaatta-caggtg
Mmm2de4  tgtgaaatg-a-cacattcctgcacgcacttccctagtataatgggtgtcaaattcccaggaa
Hsm2de4  gtggaac---agcacattcctccacacacttctccagtatcataggtgctggattccaggaag

Ggm2de4  tttatagcctccaaagtagcacagtgaatagctgcctcaaagacataataagcaaccgatgcc
Mmm2de4  tcgctggcttccaaag---cacagggattagttgccacagagatggaataagccaccaatatc
Hsm2de4  tctctctggcttcca----taacacacttagttgccacagagacataataagcaaccaatgcc

Ggm2de4  taaataatcattacagatactccattagcagcatctaactgtgtggcttttaaccatacaaat
Mmm2de4  caaataattggtacacatcctccattaacaacatctaactgtgtggcttttaaccatacaaat
Hsm2de4  caaataatcattacagatactccattaacagcatctaactgtgtggcttttaaccatacaaat

Ggm2de4  tgttttttgacagcttgtaacctttccaaacagtttcctttgggcattataccacagtcagat
Mmm2de4  tgttttgtgacagcttgtaacctttttcaaatagtttttctctgggtagtacaccacactcaggt
Hsm2de4  tgtttttttgacagcttgtaacctttttcaaatactttcctttggacattacactacactcagat

Ggm2de4  ctaattttaacttttgctttgtctataatttattattttctttggaaacacctaatgtatttc
Mmm2de4  ctaattttaactttggctttgtctataattttatgagttttttgg-agatgtctaatatatttg
Hsm2de4  ctaattttaacttttgctttgtctataattttatgagtttctttggagctgcataatgtatttg

Ggm2de4  agtggaacggatataatccaaataa-aagataaaaaaaaatcaattgctgtttttaatctatga
Mmm2de4  aagagaaaggatacattcaaacca-aagctaaaaacaaaaccagttaccgtattaatccataa
Hsm2de4  aggagaaaggatacattcaaataaaaagctaaaaacaaaatcaattactgttttaatccatga

Ggm2de4  aattcttgctcactgg
Mmm2de4  acccc-tgtccaccag
Hsm2de4  aacctttgctcactgg
```

Figure 9: Multiple Sequence Alignment of HCNE m2de4

Multiple sequence alignment of m2de4: Gg is *Gallus Gallus* (Chicken), Mm is *Mus musculus* (Mouse), Hs is *Homo sapiens* (Human). Sequences in yellow are conserved in land vertebrates. Binding sites: Pink is a known Hox binding site.

Cloning of HCNEs

Dr m2de1:

The HCNE m2de1 was isolated from the zebrafish (*Danio rerio*) genome by PCR. Species and sequence specific PCR primers were designed flanking m2de1 in zebrafish (Table 2). The PCR was predicted to produce a product of 440 nucleotides. This product is larger than the 255 bp HCNE because the primers were designed to provide a wide berth by flanking either side of the HCNE to ensure isolation of the entire region. Analysis of the Gradient PCR (Figure 10A) demonstrated bands of the appropriate size decreasing in intensity in the first 5 lanes, corresponding to the annealing temperatures: 58.9°C, 59.1°C, 59.7°C, 60.6°C, and 61.8°C respectively. Due to the decreasing nature of the band intensity the PCR product from lane 1, produced when annealed at 58.9°C, was used for cloning Dr m2de1 into the pCR®2.1-TOPO® TA cloning vector.

Upon cloning and transformation, 9 colonies were screened for the presence of an insert of the appropriate size by PCR using M13 Forward and M13 Reverse primers (Table 2). A band of 641 nucleotides was expected because the primers recognized regions flanking the polycloning region (Figure 10D), and only colony 1 (Figure 10B) demonstrated a clean PCR product of appropriate size. This colony was then grown up, a plasmid prep performed on this culture and the plasmid DNA digested for determination of insert size. As is shown in (Figure 10D) dr m2de1-pCR2.1-TOPO has only 2 EcoRI cut sites, both flanking the insert site, so digestion with EcoRI should yield 2 bands: 457 nucleotides corresponding to the insert, and 3,913 nucleotides corresponding to the linearized plasmid. This is the pattern that is observed when the digests were visualized by agarose gel electrophoresis (Figure 10C), demonstrating that the appropriate clone is

present, which was confirmed by sequencing. The colony was grown up and used for subsequent generation of transgenic constructs.
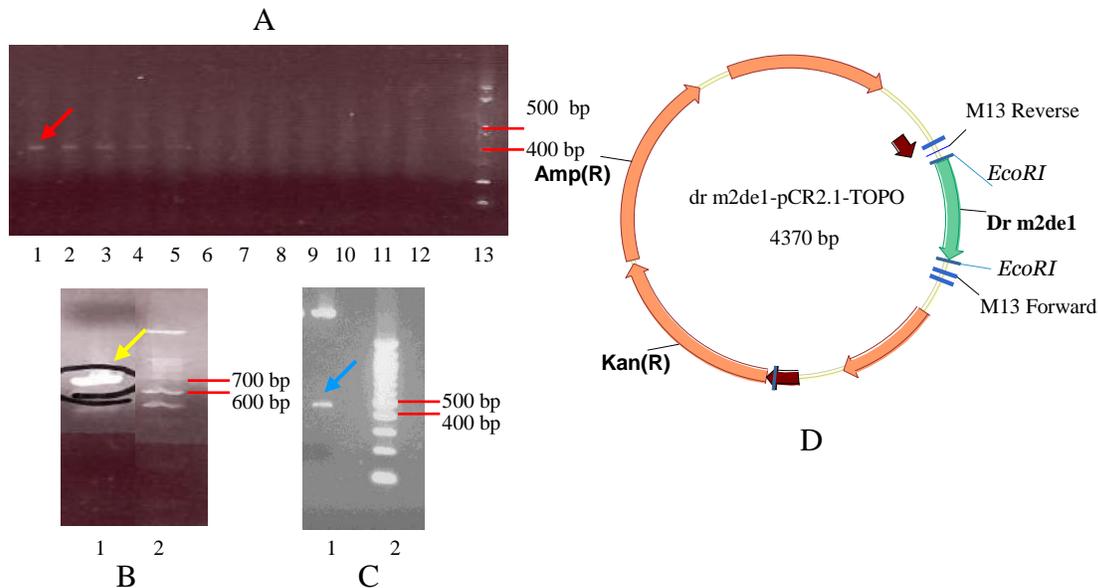
Figure 10: Cloning of Dr m2de1 into pCR®2.1-TOPO®
A) Gradient PCR: lanes 1-12 contain PCR products 1-12 respective to annealing temperatures listed in Table 2. Lane 13 contains a 100 bp DNA ladder (NEB N3231L). Red arrow indicates band of desired size.
B) Dr m2de1 colony PCR screen to identify positive ligation. Lane 1contains colony PCR screen of colony 8, lane 2 contains a 100 bp DNA ladder (NEB N3231L). Yellow arrow indicates band of desired size.
C) EcoRI (NEB R0101S) digest of Dr m2de1 colony 8. Lane 1 contains digest of colony 8, and lane 2 contains a 100 bp DNA ladder (NEB N3231L). Blue arrow indicates band of desired size.
D) Plasmid map of the clone dr m2de1-pCR2.1-TOPO.

Mm m2de1

After identification, m2de1 was isolated from the mouse (*Mus musculus*) genome. To do so species and sequence specific PCR primers were designed flanking m2de1 in zebrafish (Table 2). The PCR was predicted to produce a product of 1,054 nucleotides. This product is larger than the 255 bp HCNE because the primers were designed to provide a wide berth by flanking either side of the HCNE to ensure isolation of the entire region. Analysis of the Gradient PCR (Figure 11A) showed a consistent band of

appropriate size present in lanes 7-12 corresponding to annealing temperatures: 58.3°C,
60.2°C, 62.0°C, 63.5°C, 64.6°C, and 65.1°C respectively. The first 6 lanes do not
contain a band of appropriate size, but do contain numerous PCR products of smaller
length indicative of the primers binding to inappropriate sites at lower annealing
temperatures. As the annealing temperature increases, a band of the appropriate size
begins to appear while indiscriminant banding disappears. Because the PCR product in
lane 11 (annealing temperature 64.6°C) is robust and the cleanest, the PCR product from
lane 11 was used to clone Mm m2de1 into the pCR®2.1-TOPO® TA cloning vector.
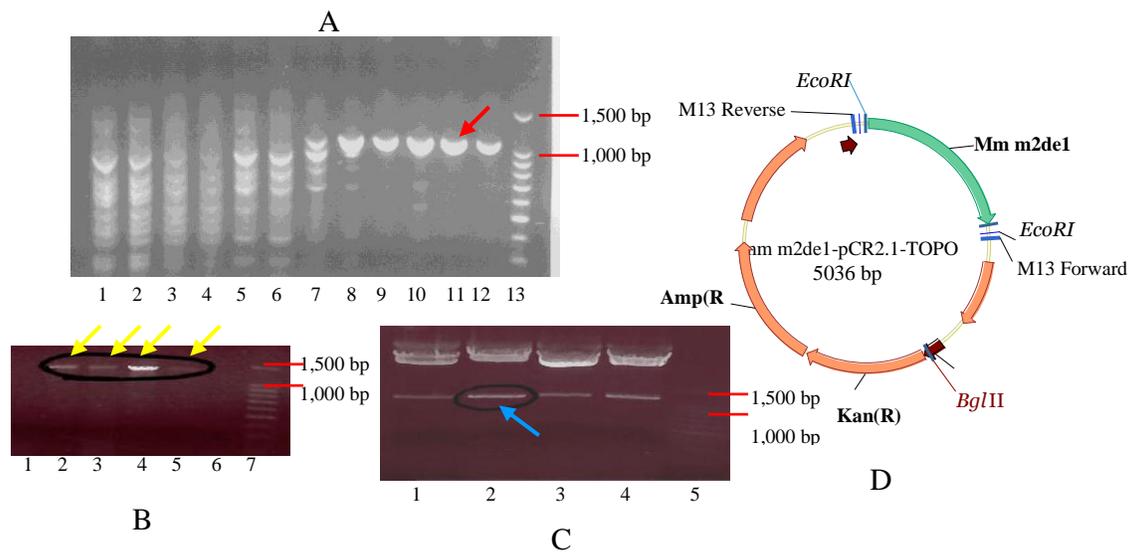
Figure 11: Cloning of Mm m2de1 into pCR®2.1-TOPO®
A) Gradient PCR: Lanes 1-12 contain PCR products 1-12 respective to annealing
temperatures Table 2. Lane 13 contains a 100 bp DNA ladder (NEB N3231L). Red
arrow indicates band of desired size.
B) Mm m2de1 colony PCR screen to identify positive ligation. Lanes 1-6 contain colony
PCR screens of colonies 1-6, lane 7 contains a 100 bp DNA ladder (NEB N3231L).
Yellow arrow indicates band of desired size.
C) EcoRI (NEB R0101S) digest of Mm m2de1 colonies 2, 3, 4, and 5. Lanes 1-4 contain
digests of colonies 2-5, and lane 5 contains a 100 bp DNA ladder (NEB N3231L). Blue
arrow indicates band of desired size.
D) Plasmid map of the clone mm m2de1-pCR2.1-TOPO.

Upon cloning and transformation 6 colonies were screened by PCR, using M13 Forward and M13 Reverse primers (Table 2), for the presence of an insert of the appropriate size. A band of 1,271 nucleotides was expected because the primers recognized regions flanking the polycloning region. Colonies 4, 5, 6, and 7 (Figure 11B) all demonstrated a clean PCR product of appropriate size. All 4 colonies were then grown up and digested for determination of insert size. As is shown in Figure 11D, mm m2de1-pCR2.1-TOPO has only 2 EcoRI cut sites, both flanking the insert site, so digestion with EcoRI (NEB R0101S) should yield 2 bands: 1,054 nucleotides corresponding to the insert, and 3,913 corresponding to the linearized plasmid. This is the pattern that is observed for all 4 colonies (Figure 11C), demonstrating that the appropriate clone is present, which was confirmed by sequencing. Colony 2 was grown up and used for subsequent generation of transgenic constructs.

Mm m2de2

The m2de2 element was isolated from the mouse (*Mus musculus*) genome. To do so species and sequence specific PCR primers were designed flanking m2de2 in mouse (Table 2). The PCR was predicted to produce a product of 1,308 nucleotides. This product is larger than the 260 bp HCNE because the primers were designed to provide a wide berth by flanking either side of the HCNE to ensure isolation of the entire region. Analysis of the Gradient PCR (Figure 12A) showed an increasingly intense band of appropriate size present in lanes 3-9 corresponding to annealing temperatures 50.2°C, 51.1°C, 52.5°C, 54.2°C, 56.2°C, 58.3°C, and 60.2°C respectively, with a weaker solitary band present in lanes 10-12 corresponding to annealing temperatures 62.0°C, 63.5°C, and 64.6°C respectively. Lane 2 does not contain a band of appropriate size, and lanes 2-8

contain numerous PCR products of inappropriate size indicative of the primers binding to inappropriate sites at lower annealing temperatures. As the annealing temperature increases, bands of inappropriate size begin to disappear, and a solitary band of appropriate size appears. Because the PCR product in lane 9 (annealing temperature 60.2°C) is robust and the cleanest, the PCR product from lane 9 was used to clone Mm m2de2 into the pCR®2.1-TOPO® TA cloning vector.
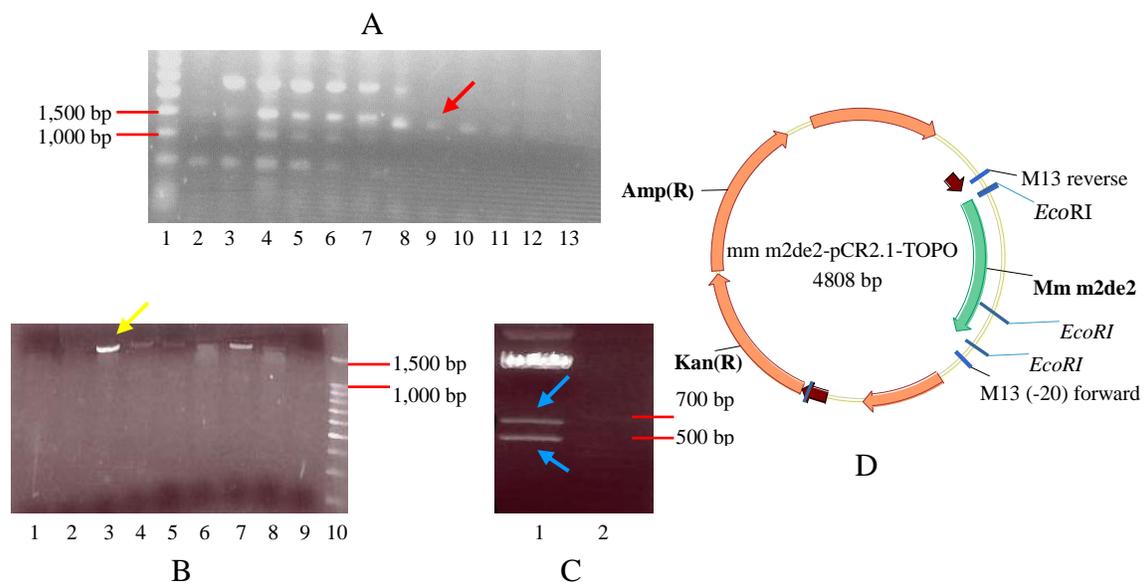
Figure 12: Cloning of Mm m2de2 into pCR®2.1-TOPO®
A) Gradient PCR: Lane 1 contains a 1 kb DNA ladder (NEB N0468S), and lanes 2-13 contain PCR products 1-12 respective to annealing temperatures listed in Table 2. Red arrow indicates band of desired size.
B) Mm m2de2 colony PCR screen to identify positive ligation. Lanes 1-9 contain colony PCR screens of colonies 1-9, lane 10 contains a 100 bp DNA ladder (NEB N3231L). Yellow arrow indicates band of desired size.
C) EcoRI (NEB R0101S) digest of Mm m2de2 colony 3. Lane 1 contains the digest of colony 3 and lane 2 contains a 100 bp DNA ladder (NEB N3231L). Blue arrow indicates band of desired size.
D) Plasmid map of the clone mm m2de2-pCR2.1-TOPO.

Upon cloning and transformation, 9 colonies were screened for the presence of an insert of the appropriate size by PCR using M13 Forward and M13 Reverse primers (Table 2). A band of 1,490 nucleotides was expected because the primers recognized

regions flanking the polycloning region. Colonies 3, 4, 5, and 7 (Figure 12B) all demonstrated a clean PCR product of appropriate size. Colony 3 was grown up and digested for determination of insert size. As is shown in Figure 12D, mm m2de2-pCR2.1-TOPO has 3 EcoRI cut sites, 2 flanking the insert site and 1 internal site, so digestion with EcoRI (NEB R0101S) was predicted to yield 3 bands: approximately 700 nucleotides and 500 nucleotides corresponding to a total insert of 1,308 nucleotides, and 3,913 corresponding to the linearized plasmid. This is the pattern that is observed for colony 3 (Figure 12C), demonstrating that the appropriate clone is present, which was confirmed by sequencing. Colony 3 was grown up to be used for generation of transgenic constructs in the future.

Mm m2de3

After identification, m2de3 was isolated from the mouse (*Mus Musculus*) genome. To do so species and sequence specific PCR primers were designed flanking m2de3 in mouse (Table 2). The PCR was predicted to produce a product of 1,250 nucleotides. This product is larger than the 975 bp HCNE because the primers were designed to provide a wide berth by flanking either side of the HCNE, ensuring isolation of the entire region. Analysis of the Gradient PCR (Figure 13A) showed a consistent band of appropriate size present in lanes 4-9 corresponding to annealing temperatures: 52.5°C, 54.3°C, 56.2°C, 58.3°C, 60.2°C, and 62.0°C respectively. The first 3 lanes do not contain a band of appropriate size, but do contain numerous PCR products of smaller length indicative of the primers binding to inappropriate sites at lower annealing temperatures. As the annealing temperature increases, the band of the appropriate size begins to appear, and indiscriminant banding gradually disappears. Because the PCR

product in lane 9 (annealing temperature 62.0°C) is robust and the cleanest, the PCR

product from lane 9 was used to clone Mm m2de3 into the pCR®2.1-TOPO® TA cloning

vector.
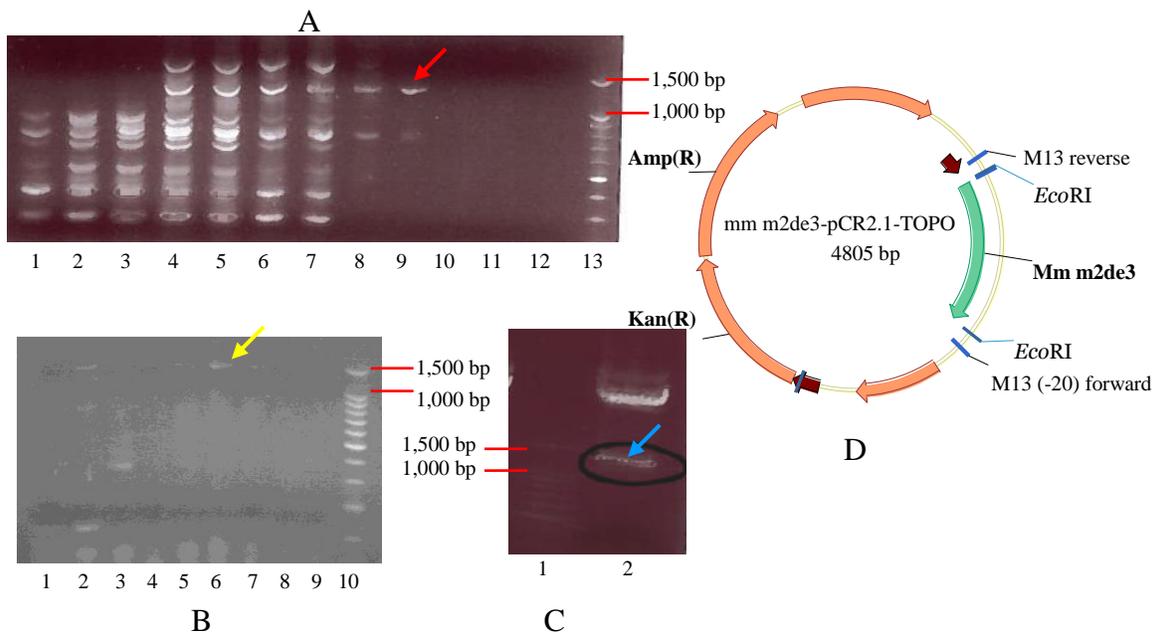
Figure 13: Cloning of Mm m2de3 into pCR®2.1-TOPO®
A) Gradient PCR: Lanes 1-12 contain PCR products 1-12 respective to annealing temperatures listed in Table 2. Lane 13 contains a 100 bp DNA ladder (NEB N3231L). Red arrow indicates band of desired size.
B) Mm m2de3 colony PCR screen to identify positive ligation. Lanes 1-9 contain colony PCR screens of colonies 1-9, lane 10 contains a 100 bp DNA ladder (NEB N3231L). Yellow arrow indicates band of desired size.
C) EcoRI (NEB R0101S) digest of Mm m2de3 colony 6. Lane 2 contains the digest of colony 6, and lane 1 contains a 100 bp DNA ladder (NEB N3231L). Blue arrow indicates band of desired size.
D) Plasmid map of the clone mm m2de3-pCR2.1-TOPO.

Upon cloning and transformation, 9 colonies were screened for the presence of an

insert of the appropriate size using PCR with M13 Forward and M13 Reverse primers

(Table 2). A band of 1,432 nucleotides was expected because the primers recognized

regions flanking the polycloning region. Colonies 2, 5, 6, and 7 (Figure 13B) all

demonstrated a clean PCR product of appropriate size. Colony 6 was then grown up and

digested for determination of insert size. As is shown in Figure 13D, mm m2de3-pCR2.1-TOPO has 2 EcoRI cut sites, both flanking the insert site, so digestion with EcoRI (NEB R0101S) should yield 2 bands: 1,250 nucleotides corresponding to the insert, and 3,913 corresponding to the linearized plasmid. This is the pattern that is observed for colony 6 (Figure 13C), demonstrating that the appropriate clone is present, which was confirmed by sequencing. Colony 6 was subsequently grown up to be used for future generation of transgenic constructs.

Mm m2de4

After identification, m2de4 was isolated from the mouse (*Mus Musculus*) genome. To do so species and sequence specific PCR primers were designed flanking m2de4 in mouse (Table 2). The PCR was predicted to produce a product of 588 nucleotides. This product is larger than the 520 bp HCNE because the primers were designed to provide a wide berth by flanking either side of the HCNE to ensure isolation of the entire region. . Analysis of the PCR (Figure 14A) showed a strong band of appropriate size present when annealed at 56.8°C, and the PCR product was used to clone Mm m2de4 into the pCR®2.1-TOPO® TA cloning vector.

Upon cloning and transformation, 9 colonies were screened for the presence of an insert of the appropriate size by PCR using M13 Forward and M13 Reverse primers (Table 2). A band of 641 nucleotides was expected because the primers recognized regions flanking the polycloning region. Colonies 4, 5, 6, and 8 (lanes 5, 6, 7, and 9) (Figure 14B) all demonstrated a clean PCR product of appropriate size. Colony 6 was then grown up and digested for determination of insert size. As is shown in Figure 14D, mm m2de4-pCR2.1-TOPO has only 2 EcoRI cut sites, both flanking the insert site, so

digestion with EcoRI (NEB R0101S) should yield 2 bands: one band of 588 nucleotides corresponding to the insert, and 3,913 corresponding to the linearized plasmid. This is the pattern that is observed for the 1 colony screened (Figure 14C), demonstrating that the appropriate clone is present, which was confirmed by sequencing. That colony was then grown up to be used for future generation of transgenic constructs.
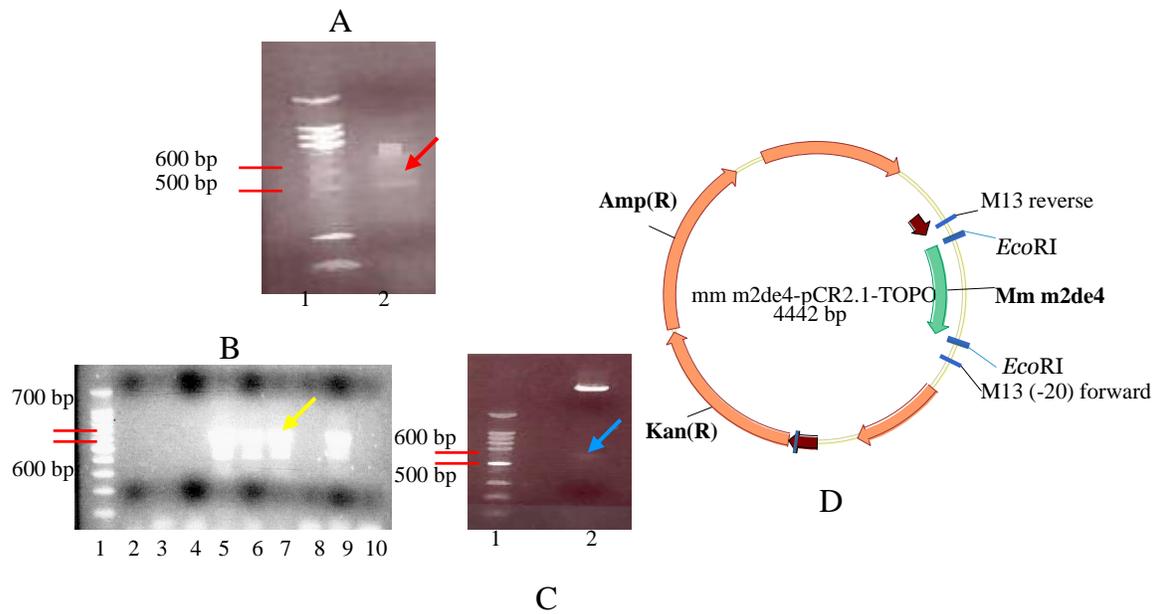
Figure 14: Cloning of Mm m2de4 into pCR®2.1-TOPO®
A) PCR: Lane 2 contains PCR product from annealing temperature 56.8°C. Lane 1 contains a 100 bp DNA ladder (NEB N3231L). Red arrow indicates band of desired size.
B) Mm m2de4 colony PCR screen to identify positive ligation. Lanes 2-10 contain colony PCR screens of colonies 1-9, lane 1 contains a 100 bp DNA ladder (NEB N3231L). Yellow arrow indicates band of desired size.
C) EcoRI (NEB R0101S) digest of Mm m2de4 colonies. Lane 2 contains the digest of colony 6, and lane 1 contains a 100 bp DNA ladder (NEB N3231L). Blue arrow indicates band of desired size.
D) Plasmid map of the clone mm m2de4-pCR2.1-TOPO.

Generation of Reporter Constructs

pDr m2de1-F-cfos-GFP

The Tol2 reporter construct containing Dr m2de1 in a forward orientation relative to the minimal promoter (cfos) and reporter gene GFP was generated using the Tol2-

Gateway 2-way system (Fisher et al., 2006). First, using the primer set 5'-attB1-TOPO and 3'-attB2-TOPO (Table 2), Dr m2de1 was isolated by PCR out of the plasmid dr m2de1-pCR2.1-TOPO (Figure 15A). The PCR was predicted to produce a product of 528 nucleotides, and when the PCR product was visualized by agarose gel electrophoresis, that is what was observed (Figure 15B). Once the PCR product was cleaned with the QIAquick PCR Purification Kit (QIAGEN 28104), it was used in a BP reaction to generate the middle entry vector pME-Dr m2de1-F (Figure 15C). The reaction was then transformed and grown on LB-Kan (50 mg/µl) plates. Because the middle entry vector has a ccdB gene that is removed upon successful translocation (Figure 4), the reaction was transformed into cells that lacked ccdB resistance. This meant that any colonies that grew would not be false positives that had failed to undergo successful translocation. Because of the selective measures built in it was fairly certain that any colony chosen would be positive for pME-Dr m2de1-F. So 2 colonies were grown up for miniprep, followed by analysis via restriction digestion.

Since both plasmids have Kan resistance, screening pME-Dr m2de1-F with EcoRI (NEB R0101S) and BglII (NEB R0144S) was necessary to rule out background transformation of residual dr m2de1-pCR2.1-TOPO from the PCR reaction. dr m2de1-pCR2.1-TOPO contains 2 EcoRI cut sites and 1 BglII site (Figure 15A), so if present, digestion will result in 3 bands of expected sizes: 2,931bp, 982bp, and 457bp. If the desired plasmid pME-Dr m2de1-F is present, there will only be 2 bands (2,656bp and 457bp) because the same EcoRI sites will be introduced into the plasmid, but pME-Dr m2de1-F does not contain a BglII restriction site (Figure 15C). Upon analysis of each digestion, each contained 2 bands in lanes 2 and 5, while the uncut lanes (1 and 4) only

had 1 band (Figure 15D). This indicated that both colonies contained the desired plasmid, pME-Dr m2de1-F.
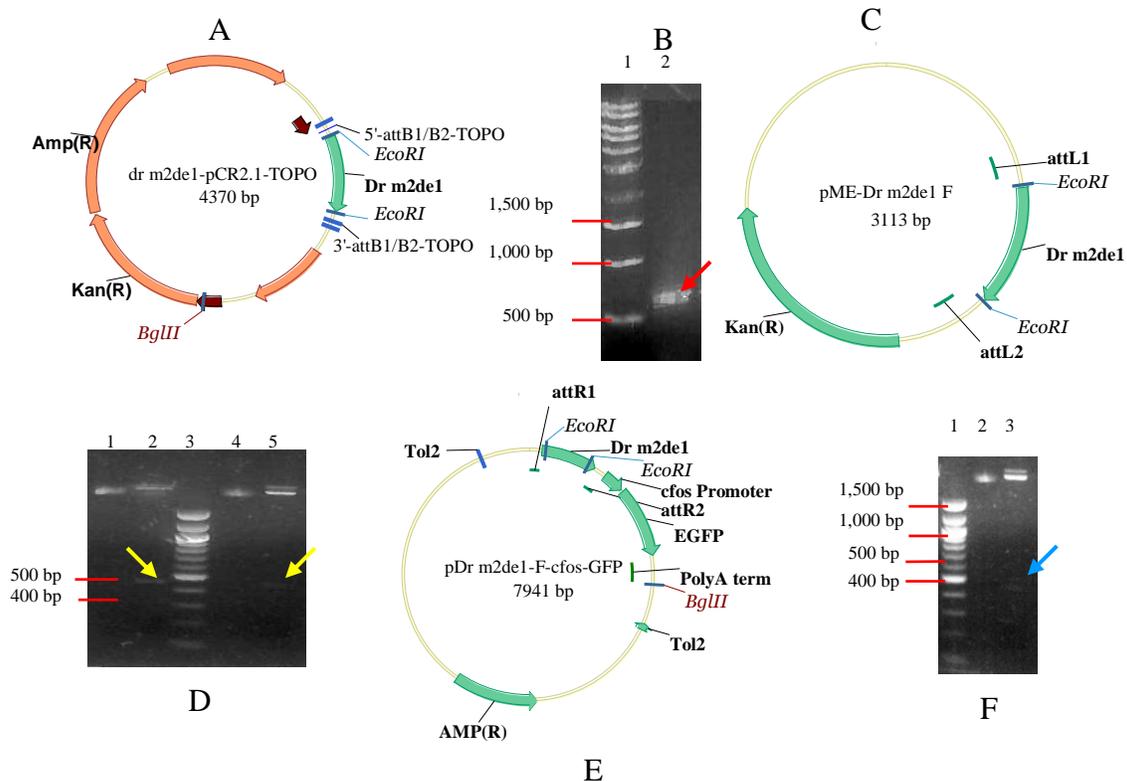
Figure 15: Generation of Forward Orientation Zebrafish m2de1 Reporter Construct
A) Plasmid Map of dr m2de1-pCR2.1-TOPO showing att primer binding sites, EcoRI sites, and BglII site.
B) PCR generating Dr m2de1-F for insertion into the middle entry vector pDONR221. Lane 1 contains 1 Kb ladder (NEB N0468S) and lane 2 contains 5 µl of PCR product. Red arrow indicates band of desired size.
C) Plasmid map of middle entry vector pME-Dr m2de1-F showing EcoRI sites, attL1, and attL2 sites.
D) Digest of pME-Dr m2de1-F with EcoRI and BglII testing for successful translocation. Lanes 1 and 4 contain uncut miniprep plasmid, lanes 2 and 5 contain miniprep plasmid cut with EcoRI and BglII, lane 3 contains 100 bp DNA ladder (NEB N3231L). Yellow arrows indicate bands of desired size.
E) Plasmid map of pDr m2de1-F-cfosGFP showing Tol2 *cis* sites, EcoRI sites, BglII site, cfos promoter, and GFP gene.
F) Digest of pDr m2de1-F-cfosGFP with EcoRI to test for successful translocation. Lane 1 contains 1 Kb ladder (NEB N0468S), lane 2 contains uncut miniprep plasmid, and 3 contains miniprep plasmid digested with EcoRI (NEB R0101S). Blue arrow indicates band of desired size.

After confirmation of pME-Dr m2de1-F, colony 1 was used for a subsequent LR reaction to clone Dr m2de1 into the reporter plasmid pGW_cfosGFP. The LR reaction was then transformed and grown on LB-Amp (100 mg/µl) plates. Because the destination vector pGW_cfosGFP has a ccdB gene that is removed upon successful translocation (Figure 4), the reaction was transformed into cells that lacked ccdB resistance. This meant that any colonies that grew would not be false positives that had failed to undergo successful translocation. In addition, pME-Dr m2de1-F has Kan resistance (Figure 15C) while pDr m2de1-F-cfosGFP has Amp resistance (Figure 15E), so any colony that grew should not be the middle entry vector. Because of the selective measures built in, it was fairly certain that any colony chosen would be positive for pDr m2de1-F-cfosGFP. So 1 colony was grown up for miniprep, followed by analysis via restriction digestion.

The resultant miniprep was digested with EcoRI (NEB R0101S), which is predicted to result in 2 bands: 8,200bp and 457bp. If somehow the middle entry vector is present, the same digestion will result in 2 bands of 2,656bp and 457bp. The difference in the size of the top band makes it easy to discern between the 2 plasmids. The restriction digestion resulted in 2 bands of approximately 8,000bp and 450bp, indicating that the transgenic reporter construct pDr m2de1-F-cfosGFP was present. The colony was maintained and used for future transgenic reactions.

pDr m2de1-R-cfos-GFP

The Tol2 reporter construct containing Dr m2de1 in a reverse orientation relative to the minimal promoter (cfos) and reporter gene GFP was generated using the Tol2-Gateway 2-way system (Fisher et al., 2006). First, using the primer set 5'-attB2-TOPO and 3'-attB1-TOPO (Table 2), Dr m2de1 was isolated by PCR from the plasmid dr

m2de1-pCR2.1-TOPO (Figure 16A). The PCR was predicted to produce a product of 528 nucleotides, which is what was observed (Figure 16B). Once the PCR product was cleaned, it was used in a BP reaction to generate the middle entry vector pME-Dr m2de1-R (Figure 16C). The reaction was then transformed and grown on LB-Kan (50 mg/μl) plates. Because the middle entry vector has a ccdB gene that is removed upon successful translocation (Figure 4), the reaction was transformed into cells that lacked ccdB resistance. This meant that any colonies that grew would not be false positives that had failed to undergo successful translocation. Because of the selective measures built in, it was fairly certain that any colony chosen would be positive for pME-Dr m2de1-R, so 2 colonies were grown up for miniprep, followed by analysis via restriction digestion.

Since both plasmids have Kan resistance, screening pME-Dr m2de1-R with EcoRI (NEB R0101S) and BglII (NEB R0144S) was necessary to rule out background transformation of residual dr m2de1-pCR2.1-TOPO from the PCR reaction. dr m2de1-pCR2.1-TOPO contains 2 EcoRI sites and 1 BglII site (Figure 16A), so if present, digestion will result in 3 bands of expected sizes: 2,931bp, 982bp, and 457bp. If the desired plasmid pME-Dr m2de1-R is present, there will only be 2 bands (2,656bp and 457bp) because the same EcoRI sites will be introduced into the plasmid, but pME-Dr m2de1-R does not contain a BglII restriction site (Figure 16C). Upon analysis of each digestion, each contained 2 bands in lanes 1 and 3 (Figure 16D). This indicated that both colonies had the desired plasmid, pME-Dr m2de1-R.
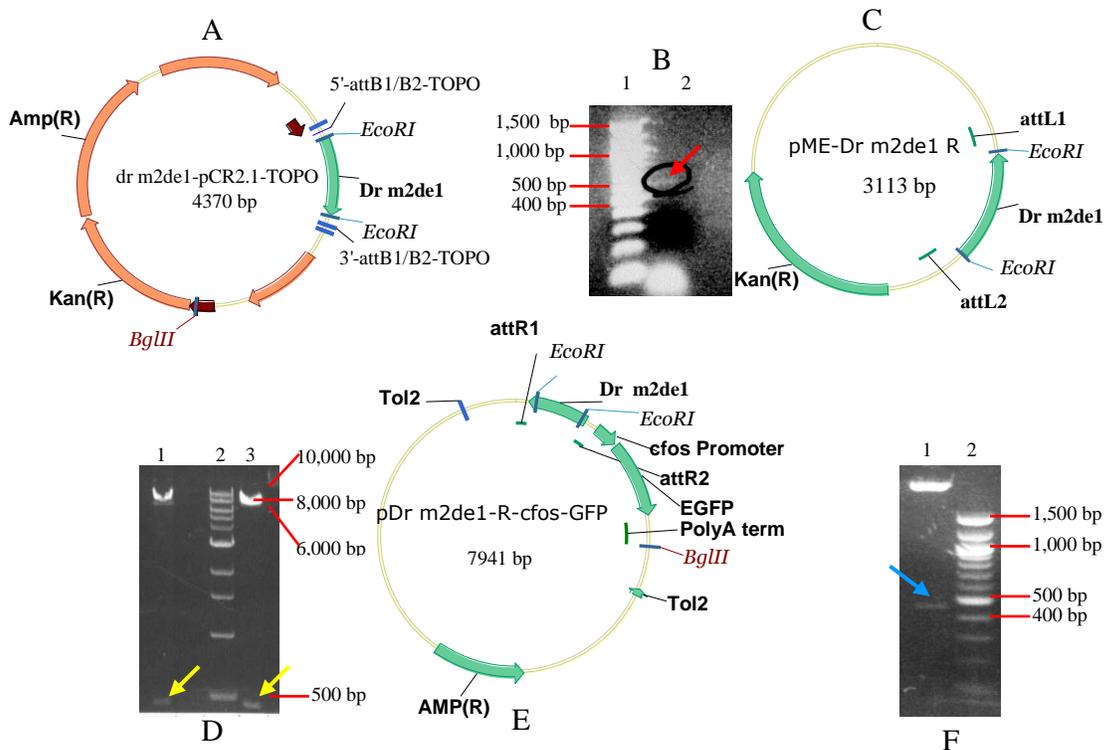
Figure 16: Generation of Reverse Orientation Zebrafish m2de1 Reporter Construct
A) Plasmid Map of dr m2de1-pCR2.1-TOPO showing att primer binding sites, EcoRI sites, and BglII site.
B) PCR generating Dr m2de1-R for insertion into the middle entry vector pDONR221. Lane 1 contains 100 bp DNA ladder (NEB N3231L) and lane 2 contains 5 µl of PCR product. Red arrow indicates band of desired size.
C) Plasmid map of middle entry vector pME-Dr m2de1-R showing EcoRI sites, attL1 and attL2 sites.
D) Digest of pME-Dr m2de1-R with EcoRI and BglII testing for successful translocation. Lanes 1 and 3 contain miniprep plasmid cut with EcoRI and BglII, lane 2 contains 1 Kb ladder (NEB N0468S). Yellow arrows indicate bands of desired size.
E) Plasmid map of pDr m2de1-R-cfosGFP showing Tol2 *cis* sites, EcoRI sites, BglII site, cfos promoter, and GFP gene.
F) Digest of pDr m2de1-R-cfosGFP with EcoRI to test for successful translocation. Lane 2 contains 100 bp DNA ladder (NEB N3231L), lane 1 contains miniprep plasmids from the colony digested with EcoRI (NEB R0101S). Blue arrow indicates band of desired size.

After confirmation of pME-Dr m2de1-R, colony 1 was used for a subsequent LR

reaction to clone Dr m2de1 into the reporter plasmid pGW_cfosGFP. The LR reaction

was then transformed and grown on LB-Amp (100 mg/µl) plates. Because the destination

vector pGW_cfosGFP has a ccdB gene that is removed upon successful translocation

(Figure 4), the reaction was transformed into cells that lacked ccdB resistance. This meant that any colonies that grew would not be false positives that had failed to undergo successful translocation. In addition, pME-Dr m2de1-R has Kan resistance (Figure 15C) while pDr m2de1-R-cfosGFP has Amp resistance (Figure 15E), so any colony that grew should not be the middle entry vector. Because of the selective measures built in, it was fairly certain that any colony chosen would be positive for pDr m2de1-R-cfosGFP, so 1 colony was grown up for miniprep, followed by analysis via restriction digestion.

The resultant miniprep was digested with EcoRI (NEB R0101S) which is predicted to result in 2 bands (8,200bp and 457bp). If somehow the middle entry vector is present the same digestion will result in 2 bands (2,656bp and 457bp). The difference in the size of the top band makes it easy to discern between the 2 plasmids. The restriction digestion resulted in 2 bands of approximately 8,000bp and 450bp, indicating that the transgenic reporter construct pDr m2de1-R-cfosGFP was present. The colony was maintained to be used for future transgenic reactions.

pMm m2de1-F-cfos-GFP

The Tol2 reporter construct containing Mm m2de1 in a forward orientation relative to the minimal promoter (cfos) and reporter gene GFP was created using the Tol2-Gateway 2-way system (Fisher et al., 2006). First, using the primer set 5'-attB1-TOPO and 3'-attB2-TOPO, Mm m2de1 was isolated by PCR out of the plasmid mm m2de1-pCR2.1-TOPO (Figure 17A). The PCR was predicted to produce a product of 1,193 nucleotides, and when the PCR product was viewed by agarose gel electrophoresis, that is what was observed (Figure 17B). Once the PCR product was cleaned it was used in a BP reaction to generate the middle entry vector pME-Mm m2de1-F (Figure 17C).

The reaction was then transformed and grown on LB-Kan (50 mg/µl) plates. Because the middle entry vector has a ccdB gene that is removed upon successful translocation (Figure 4), the reaction was transformed into cells that lacked ccdB resistance. This meant that any colonies that grew would not be false positives that had failed to undergo successful translocation. Because of the selective measures built in it was fairly certain that any colony chosen would be positive for pME-Mm m2de1-F. So 1 colony was grown up for miniprep, followed by analysis via restriction digestion.

Since both plasmids have Kan resistance, screening pME-Mm m2de1-F with EcoRI (NEB R0101S) and BglII (NEB R0144S) was necessary to rule out background transformation of residual Mm m2de1-pCR2.1-TOPO from the PCR reaction. mm m2de1-pCR2.1-TOPO contains 2 EcoRI cut sites and 1 BglII site (Figure 17A), so if present, digestion will result in 3 bands of expected sizes: 2,931bp, 982bp, and 1,054 bp. If the desired plasmid pME-Mm m2de1-F is present, there will only be 2 bands (2,656bp and 1,054bp) because the same EcoRI sites will be introduced into the plasmid, but pME-Mm m2de1-F does not contain a BglII restriction site (Figure 17C). Upon analysis of the digestion, the reaction resulted in 2 bands in lane 1 (Figure 17D). This indicated that both colonies had the desired plasmid pME-Mm m2de1-F.
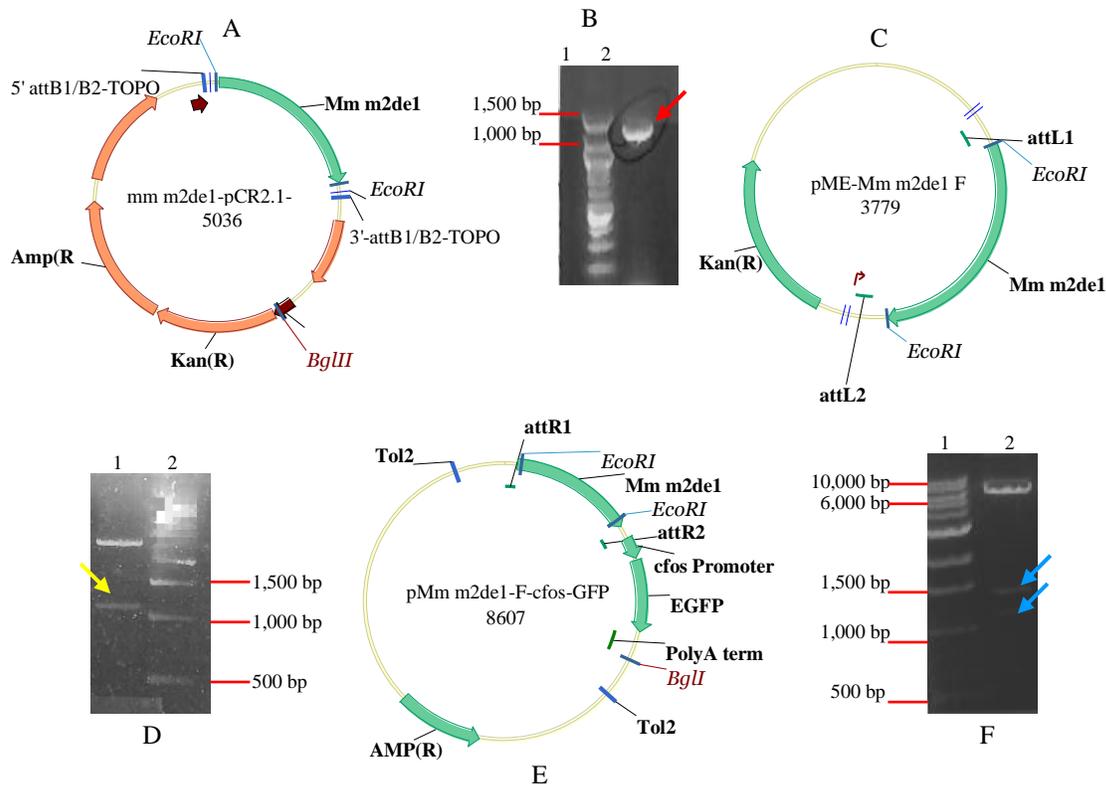
Figure 17: Generation of Forward Orientation Mouse m2de1 Reporter Construct
A) Plasmid Map of mm m2de1-pCR2.1-TOPO showing att primer binding sites, EcoRI sites, and BglII site.
B) PCR generating Mm m2de1-F for insertion into the middle entry vector pDONR221. Lane 1 contains 100 bp DNA ladder (NEB N3231L) and lane 2 contains 5 µl of PCR product. Red arrow indicates band of desired size.
C) Plasmid map of middle entry vector pME-Mm m2de1-F showing EcoRI sites, attL1 and attL2 sites.
D) Digest of pME-Mm m2de1-F with EcoRI and BglII testing for successful translocation. Lane 1 contains miniprep plasmid cut with EcoRI and BglII, lane 2 contains 100 bp DNA ladder (NEB N3231L). Yellow arrow indicates band of desired size.
E) Plasmid map of pMm m2de1-F-cfosGFP showing Tol2 *cis* sites, EcoRI sites, BglII site, cfos promoter, and GFP gene.
F) Digest of pMm m2de1-F-cfosGFP with EcoRI to test for successful translocation. Lane 1 contains 1 Kb ladder (NEB N0468S), lane 2 contains miniprep plasmid digested with EcoRI (NEB R0101S). Blue arrows indicate bands of desired size.

After confirmation of pME-Mm m2de1-F, the colony was used for a subsequent

LR reaction to clone Mm m2de1 into the reporter plasmid pGW_cfosGFP.   The LR

reaction was then transformed and grown on LB-Amp (100 mg/µl) plates. Because the destination vector pGW_cfosGFP has a ccdB gene that is removed upon successful translocation (Figure 4), the reaction was transformed into cells that lacked ccdB resistance. This meant that any colonies that grew would not be false positives that had failed to undergo successful translocation. In addition, pME- Mm m2de1-F has Kan resistance (Figure 17C) while pMm m2de1-F-cfosGFP has Amp resistance (Figure 17E), so any colony that grew should not be the middle entry vector. Because of the selective measures built in, it was fairly certain that any colony chosen would be positive for pMm m2de1-F-cfosGFP, so 1 colony was grown up for miniprep, followed by analysis via restriction digestion.

The resultant plasmid DNA was digested with EcoRI (NEB R0101S), which was predicted to result in 2 bands (8,200bp and 1,054bp). If somehow the middle entry vector is present the same digestion will result in 2 bands (2,656bp and 1,054bp). The difference in the size of the top band makes it easy to discern between the 2 plasmids. The restriction digestion resulted in 2 bands of approximately 8,000bp and 1,054bp, indicating that the transgenic reporter construct pMm m2de1-F-cfosGFP was present. The colony was maintained to be used for future transgenic reactions.

pMm m2de1-R-cfos-GFP

The Tol2 reporter construct containing Mm m2de1 in a reverse orientation relative to the minimal promoter (cfos) and reporter gene GFP was created using the Tol2-Gateway 2-way system (Fisher et al., 2006). First, using the primer set 5'-attB2-TOPO and 3'-attB1-TOPO (Table 2), Mm m2de1 was isolated by PCR out of the plasmid mm m2de1-pCR2.1-TOPO (Figure 18A). The PCR was predicted to produce a product

of 1,193 nucleotides, and when the PCR product was visualized by agarose gel electrophoresis, that is what was observed (Figure 18B). Once the PCR product was cleaned, it was used in a BP reaction to generate the middle entry vector pME-Mm m2de1-R (Figure 18C). The reaction was then transformed and grown on LB-Kan (50 mg/µl) plates. Because the middle entry vector has a ccdB gene that is removed upon successful translocation (Figure 4), the reaction was transformed into cells that lacked ccdB resistance. This meant that any colonies that grew would not be false positives that had failed to undergo successful translocation. Because of the selective measures built in it was fairly certain that any colony chosen would be positive for pME-Dr m2de1-R, so 1 colony was grown up for miniprep, followed by analysis via restriction digestion.

Since both plasmids have Kan resistance, screening pME-Mm m2de1-R with EcoRI (NEB R0101S) and BglII (NEB R0144S) was necessary to rule out background transformation of residual mm m2de1-pCR2.1-TOPO from the PCR reaction. mm m2de1-pCR2.1-TOPO contains 2 EcoRI sites and 1 BglII site (Figure 18A), so if present, digestion will result in 3 bands of expected sizes: 2,931bp, 982bp, and 1,054bp. If the desired plasmid pME-Mm m2de1-R is present, there will only be 2 bands (2,656 bp and 1,054 bp) because the same EcoRI sites will be introduced into the plasmid, but pME-Mm m2de1-R does not contain a BglII restriction site (Figure 18C). Upon analysis of the digestion, the reaction resulted in 2 bands in lane 2 (Figure 18D). This indicated that both colonies had the desired plasmid pME-Mm m2de1-R.
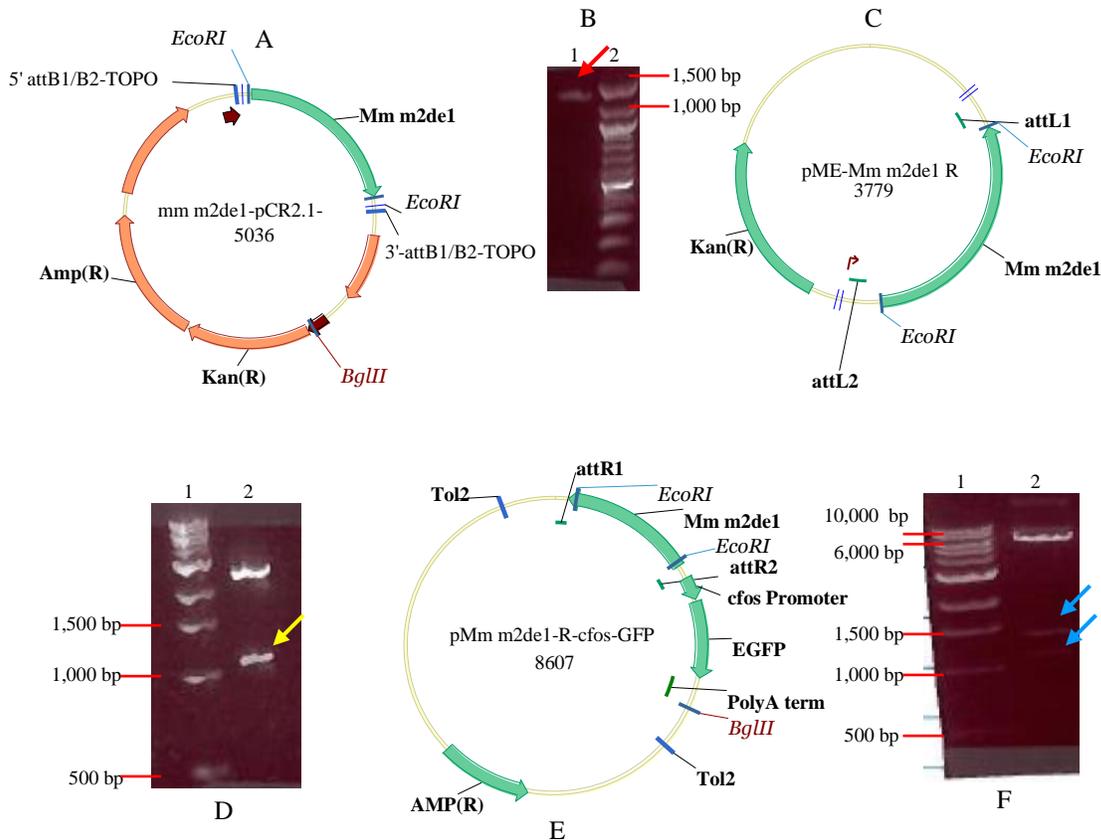
Figure 18: Generation of Reverse Orientation Mouse m2de1 Reporter Construct
A) Plasmid Map of mm m2de1-pCR2.1-TOPO showing att primer binding sites, EcoRI sites, and BglII site.
B) PCR generating Mm m2de1-R for insertion into the middle entry vector pDONR221. Lane 1 contains 5 µl of PCR product and lane 2 contains 100 bp DNA ladder (NEB N3231L). Red arrow indicates band of desired size.
C) Plasmid map of middle entry vector pME-Mm m2de1-R showing EcoRI sites, attL1 and attL2 sites.
D) Digest of pME-Mm m2de1-R with EcoRI and BglII testing for successful translocation. Lane 1 contains 1 Kb ladder (NEB N0468S) and lane 2 contains miniprep plasmid cut with EcoRI and BglII. Yellow arrow indicates band of desired size.
E) Plasmid map of pMm m2de1-R-cfosGFP showing Tol2 *cis* sites, EcoRI sites, BglII site, cfos promoter, and GFP gene.
F) Digest of pMm m2de1-R-cfosGFP with EcoRI to test for successful translocation. Lane 1 contains 1 Kb ladder (NEB N0468S); lane 2 contains miniprep plasmid digested with EcoRI (NEB R0101S). Blue arrows indicate bands of desired size.

After confirmation of pME-Mm m2de1-R, the colony was used for a subsequent

LR reaction to clone Mm m2de1 into the reporter plasmid pGW_cfosGFP. The reaction

was then transformed and grown on LB-Amp (100 mg/µl) plates. Because the

destination vector pGW_cfosGFP has a ccdB gene that is removed upon successful translocation (Figure 4), the reaction was transformed into cells that lacked ccdB resistance. This meant that any colonies that grew would not be false positives that had failed to undergo successful translocation. In addition, pME-Mm m2de1-R has Kan resistance (Figure 18C) while pMm m2de1-R-cfosGFP has Amp resistance (Figure 18E), so any colony that grew should not be the middle entry vector. Because of the selective measures built in, it was fairly certain that any colony chosen would be positive for pMm m2de1-R-cfosGFP, so 1 colony was grown up for miniprep, followed by analysis via restriction digestion.

The resultant minipreps were digested with EcoRI (NEB R0101S), which was predicted to result in 2 bands (8,200bp and 1,054bp). If somehow the middle entry vector is present, the same digestion will result in 2 bands (2,656bp and 1,054bp). The difference in the size of the top band makes it easy to discern between the 2 plasmids. The restriction digestion resulted in 2 bands of approximately 8,000bp and 1,054bp, indicating that the transgenic reporter construct pMm m2de1-R-cfosGFP was present. The colony was maintained to be used for future transgenic reactions.

Production of Transposase mRNA

To produce the Transposase mRNA for microinjections, the plasmid was first linearized with NotI (NEB R0189S), which would leave a linear construct. This was done so that when the SP6 promoter binds and begins transcribing the template the polymerase will run out of template just after the SV40 PolyA tail ends (Figure 19A). This reaction was performed in duplicate to ensure success. After digestion, each trial was run on a gel to confirm that digestion was complete, because any residual circular

plasmid will inhibit the transcription reaction. As is seen in Figure 19B, both digests

resulted in only 1 solitary band of approximately 6,000bp, the size of the plasmid.
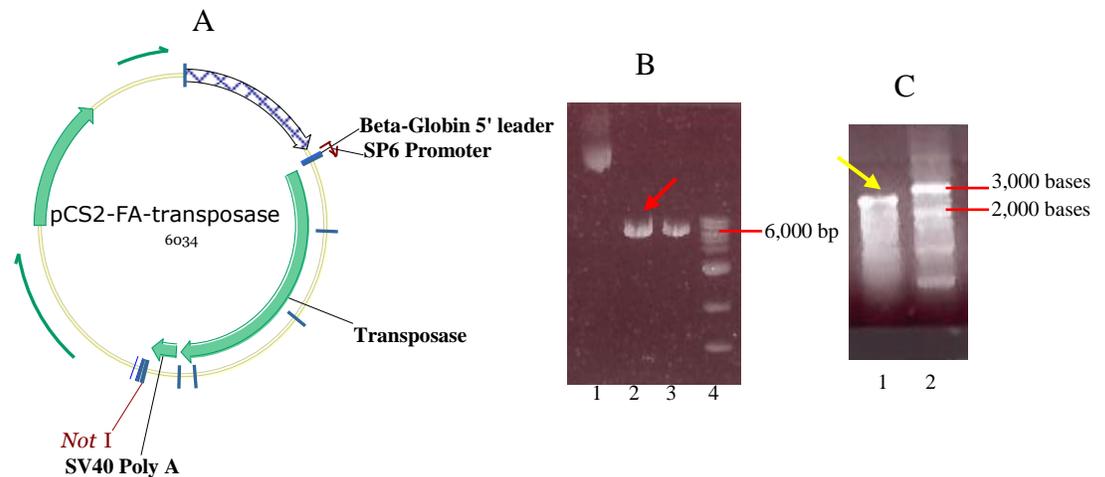
Figure 19: Production of Transposase mRNA for Injection
A) Plasmid map of pCS2-FA-Transposase showing SP6 Promoter, Transposase coding region, SV40 Poly A Tail, and NotI restriction site (NEB R0189S).
B) Gel showing the digestion of pCS2-FA-Transposase with NotI (NEB R0189S). Lane 1 contains undigested plasmid, lanes 2 and 3 contain individual linearization reactions, lane 4 contains 1 Kb ladder (NEB N0468S). Red arrow indicates band of desired size.
C) Gel showing the product of mMESSAGE mMACHINE® SP6 RNA Transcription Kit (Ambion® AM1340M) reaction. Lane 1 contains reaction product and lane 2 contains ssRNA ladder (NEB N0362). Yellow arrow indicates band of desired size.

The digest from lane 2 was chosen to be used as the transcription template. Upon

completion of the reaction, the mRNA product was purified, aliquoted, and a small

sample was run on a gel to determine integrity and size of the product (Figure 19C). A

band of 2,251 bases was expected, and upon visualization of the gel a single band

between 2,000 bases and 3,000 bases was observed. There is some smearing indicating

some shorter constructs or minor degradation. But the intensity of the band indicates that

the reaction yielded enough mRNA to proceed with microinjections. If the smear is due

to degradation, the product may have been degraded while running on the gel.

Microinjections

To determine if m2de1 was able to direct expression of a transgene consistent with the known expression pattern of *Meis2*, one to four cell embryos were injected with either a Tol2 positive control construct that has a known expression pattern, or pDR m2de1-F-cfos-GFP. A minimum of 668 embryos (Table 4) were injected with a positive control construct [125ng Transposon Plasmid, 175 ng Transposon RNA, 2.0 µl of 0.5% Phenol Red (Sigma P0290), and RNase-free water to a final volume of 5.0 µl] (Fisher et al., 2006), and screened for GFP expression. A total of 0 embryos exhibited reporter expression (Table 4). A minimum of 761 embryos (Table 4) were injected with pDR m2de1-F-cfos-GFP [125ng Transposon Plasmid, 175 ng Transposon RNA, 2.0 µl of 0.5% Phenol Red (Sigma P0290), and RNase-free water to a final volume of 5.0 µl] (Fisher et al., 2006), and screened for expression of GFP. A total of 0 embryos exhibited expression of the reporter gene (Table 4).

Table 4: Results of Microinjections

	Amount	Observed Expression Pattern
Positive Control Injections	668	0
pDr m2de1-F-cfos-GFP Injections	761	0

DISCUSSION

All 4 HCNEs (m2de1, m2de2, m2de3, and m2de4) exhibit all of the stereotypical signs of being *cis*-regulatory elements. All 4 are highly conserved across multiple species in sequence and orientation. All 4 HCNEs contain highly conserved transcription factor binding consensus sequences. It is also known that HCNEs oftentimes function as *cis*-regulatory elements when associated with developmentally regulated genes (Allende et al., 2006; Fisher et al., 2006; Kikuta et al., 2007; Woolfe et al., 2005; Zerucha et al., 2000). In addition, it is not unusual to find *cis*-regulatory elements shared between multiple genes as previously mentioned (Duboule, 1993; Zerucha et al., 2000), or for a *cis*-regulatory element to be controlling one gene, but located within an adjacent gene that is a bystander gene (Kikuta et al., 2007).

It has also been shown that linkage between developmentally important genes and bystander genes can be maintained over evolutionary time between zebrafish and humans. This linkage is often preserved by the presence of *cis*-regulatory elements that control the developmentally regulated gene that are located either within introns or outside of the bystander gene (Kikuta et al., 2007). These systems comprise what are called Gene Regulatory Blocks (GRB), and the presence of a GRB has been shown to be dependant on the presence of the *cis*-regulatory elements.

This is significant because a novel gene has been identified downstream of *Meis2*. This gene, temporarily named *M2LG* (*Meis*2 Linked Gene) is linked downstream of

Meis2 in vertebrates from zebrafish to humans in an inverted convergently transcribed orientation. In addition, *M2LG* appears to be expressed in a pattern that is similar to *Meis2* (Carpenter et al. unpublished). Of the 4 HCNEs that have been described, 2 (m2de1 and m2de2) are located within an intron of *M2LG*, and 2 (m2de3 and m2de4) lie outside and upstream of its coding region. The two most parsimonious explanations for the linkage between *Meis2* and *M2LG* are that the 2 genes are either sharing some or all of the regulatory elements, or that *M2LG* is a bystander gene and its location adjacent to *Meis2* is being maintained because if the 2 genes were to be segregated from each other by translocation the regulatory mechanisms controlling *Meis*2 would be lost causing it to lose function. Such a mutation, as has been previously described, would most likely be lethal. So because of the genomic organization of *Meis*2, *M2LG*, and all 4 HCNEs, the system appears to comprise a GRB, adding some support to the hypothesis that the HCNEs function as *cis*-regulatory elements.

While all of the bioinformatic information indicates that each HCNE functions as a *cis*-regulatory element, to definitively prove it requires a functional analysis of each HCNE in a model organism. To do so in zebrafish calls for transgenic analysis, but unfortunately all attempts to perform transgenics to this point have been unsuccessful. After optimization of the injection process through mock injections, a total of 1,429 embryos were injected with transgenic constructs. A total of 668 embryos were injected with positive control constructs, and 761 with the experimental expression construct pDr m2de1-F-cfosGFP. Not a single embryo injected with an experimental construct exhibited restricted expression patterns of the reporter gene. Just like the experimental injections, not a single embryo injected with a positive control construct exhibited

reporter expression. This result, while it does not prove that m2de1 functions as a *cis*-regulatory element, does also not rule out the possibility that m2de1 does function as a *cis*-regulatory element.

The injection process is a delicate one, and there are several steps where errors could have been made and should be evaluated. The Tol2 system has been shown to be extremely efficient, with the construct pGW_cfosGFP showing restricted, non-mosaic, reporter expression in 10-20% of injected embryos that survive the injection process (Fisher et al., 2006). It has also been shown that the Tol2 system can achieve successful germline transmission rates with transgenic F1 generations from up to 70% of injected embryos that exhibit reporter expression (Kawakami, 2007). The published high rate of success, when compared to the failure of 1,429 injected embryos to exhibit reporter expression, indicates a problem with the process that we are using.

The process involves using the reporter constructs that have been created to generate transgenic lines in order to quantify the reporter gene's expression pattern. To do so requires injecting each reporter construct (individually) along with Transposase RNA into a 1-cell embryo. The RNA is translated and the resulting transposase initiates insertion of the Tol2 reporter construct into the embryo's genome after injection. As subsequent cell divisions occur, each daughter cell would contain the insertion resulting in homogenous transgenics (Linney et al., 1999; Linney and Udvadia, 2004; Stuart et al., 1988; Xu, 1999).

In order to inject anything into a single cell extremely small needles are required. As the injection process began, it quickly became apparent that the appropriate working knowledge of how to inject into zebrafish embryos was absent. After a trip to the Prince

laboratory at the University of Chicago first hand knowledge of the injection process was gained.

Upon returning, the embryos were lined up on a Petri dish using a microscope slide to function as a brace for the embryos. The small gap between the dish and slide would serve to remove excess water through capillary action. The process of moving the Petri dish instead of the needle in order to orient the embryos into the appropriate position for injection was also initiated. Successful injections into the embryos soon followed these changes, at which point it was noticed that the vast majority of embryos that were injected died. There were many possibilities as to what the problem was: needles, mineral oil, or toxic solution. Changes to the puller settings were made to make sure that the needles had a long taper, a width of no more than 8 µm, all the while making sure to break the tips leaving a beveled edge (Linney and Udvadia, 2004). Following these changes, a marked improvement was noticed, but there was still a significant level of lethality. Subsequently, mock injections (518) were performed while altering types of water and varying the volume of 0.5% Phenol Red (Sigma P0290) in order to optimize the physical injection process.

In addition to tinkering with the injection process, post-injection care of the embryos was reevaluated in order to maximize survival rates. It was discovered that survival rates were drastically increased if, at approximately 3 hours following injection, embryos that did not survive the injection process, as well as embryos that did not show red blastomeres, were discarded.

After optimization of the injection process, both experimental (761 embryos) and positive control (668 embryos) constructs were injected, and the embryos screened for

reporter expression. There were no successful transformants however. During this time period, the concentrations of the expression construct, mRNA, and Phenol Red were varied to determine if there was an optimum concentration that varied from those that had been published. But there was no variation of concentration that resulted in positive transgenics.

Because component concentration is not the causal factor resulting in a lack of transgenesis, the problem must be either the quality of the components, or with some aspect of the procedure that has not yet been addressed. There are a couple of possibilities that have yet to be tested: mRNA integrity and the supercoiled nature of the expression construct. The integrity of the mRNA is of paramount importance, because integrity of the mRNA is necessary in order for the construct to be translated, making the transposase available to facilitate the translocation of the Tol2 construct into the embryo's genome. Even though the mRNA product showed a strong band, it is entirely possible that the reaction did not produce a product that will be effective. The mMESSAGE mMACHINE® SP6 RNA Transcription Kit (Ambion® AM1340M) has been known to occasionally produce products that lack a 5 cap , or for some other reason yield non-functional products (Linney, personal communication). To circumvent this, it may be optimal to perform multiple (5) reactions simultaneously, pooling all of the products into 1. This would help to maximize the probability of the presence of a functional mRNA product.

The second component issue that should be addressed is the supercoiled nature of the expression constructs. The ability for a foreign fragment of DNA to incorporate into the genome is contingent on the construct being extremely pure and of the highest

integrity (Chi-Bin Chien, personal communication). This is why maxi prep DNA is used, as opposed to miniprep DNA. However, it still remains possible that for some reason the maxi prep DNA does not have the necessary characteristics for the desired purposes. If DNA integrity is indeed the issue, there are several potential causes. Two examples would be that the maxiprep system used may be old or defective, or that the cells that the constructs are maintained in could be altering the construct in some way, such as by methylating the DNA. One way to circumvent these problems would be to use a cesium chloride DNA extraction protocol to ensure DNA integrity and purity.

Additional sources of trouble could be the setup itself. The microinjector that was used has never been published in zebrafish transgenic experiments. The microinjector uses mechanical movement of a rod to expel the injection solution. Most microinjectors utilize a pulse of gas to expel the injection solution. So utilizing the unpublished mechanical microinjector is a novel approach. Using gas backpressure to expel the solution, it is necessary to visually estimate the volume of solution injected into the embryo. By using the mechanical rod apparatus it is possible to precisely control the volume of solution injected. However, the system requires a mineral oil buffer region to be present between the rod and the injection solution, which no other system requires.

This mineral oil layer is necessary to provide a noncompressable buffer between the rod and the solution, ensuring even and constant back pressure. In addition, mineral oil will not cause rusting, minimizing the risk of oxidative damage to the rod itself. The problem with using mineral oil is that there is an additional variable added to the system that has not been addressed in the literature. The presence of mineral oil could be somehow altering the integrity of the expression construct, or mRNA. The mineral oil is

also not RNase free, so its presence could be introducing RNases leading to degradation of the mRNA, which would inhibit the Tol2 translocation reaction.

Given all of the potential areas for potential error that have been addressed, combined with the lack success of a positive control construct, it is highly likely that the issue with the procedure has not been addressed yet. In addition, the lack of reporter expression is not indicative of the HCNE not being a *cis*-regulatory element. Given the high degree of conservation in sequence and orientation, the presence of known transcription factor binding sites, and the existence of a putative GRB, the supporting bioinformatic evidence lends to a high probability that each HCNE functions as a *cis*-regulatory element. So, further analysis of each HCNE through transgenics should be carried out to definitively quantify the functionality of each HCNE as a *cis*-regulatory element. The procedure will need to be further optimized, and it may be desirable to look into alternative methods for quantifying functionality of putative *cis*-regulatory elements in the future.

GLOSSARY

List of Abbreviations

A/P	Anterior/Posterior
Amp	Ampicillin
Antp	*Antennapedia*
CNS	Central Nervous System
DDC	Duplication-Degeneration-Complementation
dpf	Days Postfertilization
Dr	*Danio rerio* (zebrafish)
EMSA	Electrophoretic-Mobility Shift Assays
Evo-Devo	Evolutionary Developmental Biology
GFP	Green Fluorescent Protein
Gg	*Gallus gallus* (chicken)
GRB	Gene Regulatory Block
H4	Histone 4
HCNE	Highly Conserved Noncoding Element
HD	Homeodomain
HDac	Histone Deacetylase
hpf	Hours Postfertilization
Hs	*Homo sapiens* (human)
Hth	*Homothorax*
Kan	Kanamycin
M1	Meis Domain 1
M2	Meis Domain 2
m2de	*Meis2* Downstream Element
M2LG	*Meis2 Linked Gene*
Mm	*Mus musculus* (mouse)
MRCA	Most Recent Common Ancestor
NLS	Nuclear Localization Signal
P/D	Proximal/Distal
PCR	Polymerase Chain Reaction
PG	Paralog Group
pGW	Plasmid Gateway
pME	Plasmid Middle Entry
RA	Retinoic Acid
r	Rhombomere
RO	Reverse Osmosis
TALE	Three Amino acid Loop Extension
Tr	*Takifugu rubripes* (pufferfish)

REFERENCES

Abu-Shaar, M., and Mann, R.S. (1998). Generation of multiple antagonistic domains along the proximodistal axis during Drosophila leg development. Development *125*, 3821-3830.

Abu-Shaar, M., Ryoo, H.D., and Mann, R.S. (1999). Control of the nuclear localization of Extradenticle by competing nuclear import and export signals. Genes Dev *13*, 935-945.

Ahn, D., and Ho, R.K. (2008). Tri-phasic expression of posterior Hox genes during development of pectoral fins in zebrafish: implications for the evolution of vertebrate paired appendages. Dev Biol *322*, 220-233.

Allende, M.L., Manzanares, M., Tena, J.J., Feijoo, C.G., and Gomez-Skarmeta, J.L. (2006). Cracking the genome's second code: enhancer detection by combined phylogenetic footprinting and transgenic fish and frog embryos. Methods *39*, 212-219.

Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L.*, et al.* (1998). Zebrafish hox clusters and vertebrate genome evolution. Science *282*, 1711-1714.

Amsterdam, A., Lin, S., and Hopkins, N. (1995). The Aequorea victoria green fluorescent protein can be used as a reporter in live zebrafish embryos. Dev Biol *171*, 123-129.

Antonellis, A., Huynh, J.L., Lee-Lin, S.Q., Vinton, R.M., Renaud, G., Loftus, S.K., Elliot, G., Wolfsberg, T.G., Green, E.D., McCallion, A.S.*, et al.* (2008). Identification of neural crest and glial enhancers at the mouse Sox10 locus through transgenesis in zebrafish. PLoS Genet *4*, e1000174.

Arata, Y., Kouike, H., Zhang, Y., Herman, M.A., Okano, H., and Sawa, H. (2006). Wnt signaling and a Hox protein cooperatively regulate psa-3/Meis to determine daughter cell fate after asymmetric cell division in C. elegans. Dev Cell *11*, 105-115.

Azcoitia, V., Aracil, M., Martinez, A.C., and Torres, M. (2005). The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. Dev Biol *280*, 307-320.

Berkes, C.A., Bergstrom, D.A., Penn, B.H., Seaver, K.J., Knoepfler, P.S., and Tapscott, S.J. (2004). Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. Mol Cell *14*, 465-477.

Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., and Zappavigna, V. (1999). The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. Genes Dev *13*, 946-953.

Berthelsen, J., Zappavigna, V., Ferretti, E., Mavilio, F., and Blasi, F. (1998a). The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity. EMBO J *17*, 1434-1445.
Berthelsen, J., Zappavigna, V., Mavilio, F., and Blasi, F. (1998b). Prep1, a novel functional partner of Pbx proteins. EMBO J *17*, 1423-1433.

Bessa, J., Tavares, M.J., Santos, J., Kikuta, H., Laplante, M., Becker, T.S., Gomez-Skarmeta, J.L., and Casares, F. (2008). meis1 regulates cyclin D1 and c-myc expression, and controls the proliferation of the multipotent cells in the early developing zebrafish eye. Development *135*, 799-803.

Biemar, F., Devos, N., Martial, J.A., Driever, W., and Peers, B. (2001). Cloning and expression of the TALE superclass homeobox Meis2 gene during zebrafish embryonic development. Mech Dev *109*, 427-431.

Bumsted-O'Brien, K.M., Hendrickson, A., Haverkamp, S., Ashery-Padan, R., and Schulte, D. (2007). Expression of the homeodomain transcription factor Meis2 in the embryonic and postnatal retina. J Comp Neurol *505*, 58-72.

Burglin, T.R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res *25*, 4173-4180.

Burglin, T.R. (1998). The PBC domain contains a MEINOX domain: coevolution of Hox and TALE homeobox genes? Dev Genes Evol *208*, 113-116.

Capdevila, J., Tsukui, T., Rodriquez Esteban, C., Zappavigna, V., and Izpisua Belmonte, J.C. (1999). Control of vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism of BMPs by Gremlin. Mol Cell *4*, 839-849.

Capellini, T.D., Zewdu, R., Di Giacomo, G., Asciutti, S., Kugler, J.E., Di Gregorio, A., and Selleri, L. (2008). Pbx1/Pbx2 govern axial skeletal development by controlling Polycomb and Hox in mesoderm and Pax1/Pax9 in sclerotome. Dev Biol *321*, 500-514.

Carroll, S.B. (2005a). Endless forms most beautiful : the new science of evo devo and the making of the animal kingdom, 1st edn (New York: Norton).

Carroll, S.B. (2005b). Evolution at two levels: on genes and form. PLoS Biol *3*, e245.

Carroll, S.B. (2006). The making of the fittest : DNA and the ultimate forensic record of evolution, 1st edn (New York, N.Y.: W.W. Norton & Co.).

Carroll, S.B. (2009). Remarkable creatures : epic adventures in the search for the origins of species (Boston, Houghton Mifflin Harcourt).

Carroll, S.B., Grenier, J.K., and Weatherbee, S.D. (2001). From DNA to diversity : molecular genetics and the evolution of animal design (Malden, Mass.: Blackwell Science).

Carroll, S.B., Prud'homme, B., and Gompel, N. (2008). Regulating evolution. Sci Am *298*, 60-67.

Cecconi, F., Proetzel, G., Alvarez-Bolado, G., Jay, D., and Gruss, P. (1997). Expression of Meis2, a Knotted-related murine homeobox gene, indicates a role in the differentiation of the forebrain and the somitic mesoderm. Dev Dyn *210*, 184-190.

Chang, C.P., Jacobs, Y., Nakamura, T., Jenkins, N.A., Copeland, N.G., and Cleary, M.L. (1997). Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. Mol Cell Biol *17*, 5679-5687.

Choe, S.K., Lu, P., Nakamura, M., Lee, J., and Sagerstrom, C.G. (2009). Meis cofactors control HDAC and CBP accessibility at Hox-regulated promoters during zebrafish embryogenesis. Dev Cell *17*, 561-567.

Choe, S.K., and Sagerstrom, C.G. (2004). Paralog group 1 hox genes regulate rhombomere 5/6 expression of vhnf1, a repressor of rostral hindbrain fates, in a meis-dependent manner. Dev Biol *271*, 350-361.

Choe, S.K., and Sagerstrom, C.G. (2005). Variable Meis-dependence among paralog group-1 Hox proteins. Biochem Biophys Res Commun *331*, 1384-1391.

Choe, S.K., Vlachakis, N., and Sagerstrom, C.G. (2002). Meis family proteins are required for hindbrain development in the zebrafish. Development *129*, 585-595.

Conte, I., Carrella, S., Avellino, R., Karali, M., Marco-Ferreres, R., Bovolenta, P., and Banfi, S. (2010). miR-204 is required for lens and retinal development via Meis2 targeting. Proc Natl Acad Sci U S A *107*, 15491-15496.

Coy, S.E., and Borycki, A.G. (2010). Expression analysis of TALE family transcription factors during avian development. Dev Dyn *239*, 1234-1245.

Crijns, A.P., de Graeff, P., Geerts, D., Ten Hoor, K.A., Hollema, H., van der Sluis, T., Hofstra, R.M., de Bock, G.H., de Jong, S., van der Zee, A.G.*, et al.* (2007). MEIS and PBX homeobox proteins in ovarian cancer. Eur J Cancer *43*, 2495-2505.

Cullen, M.E., Dellow, K.A., and Barton, P.J. (2004). Structure and regulation of human troponin genes. Mol Cell Biochem *263*, 81-90.

Davidson, E.H. (2006). The regulatory genome : gene regulatory networks in development and evolution (Burlington, Mass.: Academic Press, Inc.).

Deflorian, G., Tiso, N., Ferretti, E., Meyer, D., Blasi, F., Bortolussi, M., and Argenton, F. (2004). Prep1.1 has essential genetic functions in hindbrain development and cranial neural crest cell differentiation. Development *131*, 613-627.

Degenhardt, K.R., Milewski, R.C., Padmanabhan, A., Miller, M., Singh, M.K., Lang, D., Engleka, K.A., Wu, M., Li, J., Zhou, D*., et al.* (2010). Distinct enhancers at the Pax3 locus can function redundantly to regulate neural tube and neural crest expressions. Dev Biol *339*, 519-527.

Delporte, F.M., Pasque, V., Devos, N., Manfroid, I., Voz, M.L., Motte, P., Biemar, F., Martial, J.A., and Peers, B. (2008). Expression of zebrafish pax6b in pancreas is regulated by two enhancers containing highly conserved cis-elements bound by PDX1, PBX and PREP factors. BMC Dev Biol *8*, 53.

Deschamps, J., and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. Development *132*, 2931-2942.

Dibner, C., Elias, S., and Frank, D. (2001). XMeis3 protein activity is required for proper hindbrain patterning in Xenopus laevis embryos. Development *128*, 3415-3426.

diIorio, P., Alexa, K., Choe, S.K., Etheridge, L., and Sagerstrom, C.G. (2007). TALE-family homeodomain proteins regulate endodermal sonic hedgehog expression and pattern the anterior endoderm. Dev Biol *304*, 221-231.

Dobzhansky, T. (1949). Towards a modern synthesis. Evolution *3*, 376.

Duboule, D. (1993). The function of Hox genes in the morphogenesis of the vertebrate limb. Ann Genet *36*, 24-29.

Duboule, D. (2007). The rise and fall of Hox gene clusters. Development *134*, 2549-2560.

Dutton, J.R., Antonellis, A., Carney, T.J., Rodrigues, F.S., Pavan, W.J., Ward, A., and Kelsh, R.N. (2008). An evolutionarily conserved intronic region controls the spatiotemporal expression of the transcription factor Sox10. BMC Dev Biol *8*, 105.

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

Ferretti, E., Marshall, H., Popperl, H., Maconochie, M., Krumlauf, R., and Blasi, F. (2000). Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins. Development *127*, 155-166.

Ferretti, E., Villaescusa, J.C., Di Rosa, P., Fernandez-Diaz, L.C., Longobardi, E., Mazzieri, R., Miccio, A., Micali, N., Selleri, L., Ferrari, G.*, et al.* (2006). Hypomorphic mutation of the TALE gene Prep1 (pKnox1) causes a major reduction of Pbx and Meis proteins and a pleiotropic embryonic phenotype. Mol Cell Biol *26*, 5650-5662.

Finnerty, J.R., and Martindale, M.Q. (1999). Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. Evol Dev *1*, 16-23.

Fisher, S., Grice, E.A., Vinton, R.M., Bessling, S.L., Urasaki, A., Kawakami, K., and McCallion, A.S. (2006). Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat Protoc *1*, 1297-1305.

Fognani, C., Kilstrup-Nielsen, C., Berthelsen, J., Ferretti, E., Zappavigna, V., and Blasi, F. (2002). Characterization of PREP2, a paralog of PREP1, which defines a novel sub-family of the MEINOX TALE homeodomain transcription factors. Nucleic Acids Res *30*, 2043-2051.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics *151*, 1531-1545.

Fujino, T., Yamazaki, Y., Largaespada, D.A., Jenkins, N.A., Copeland, N.G., Hirokawa, K., and Nakamura, T. (2001). Inhibition of myeloid differentiation by Hoxa9, Hoxb8, and Meis homeobox genes. Exp Hematol *29*, 856-863.

Gehring, W.J. (1987). Homeo boxes in the study of development. Science *236*, 1245-1252.

Gendron-Maguire, M., Mallo, M., Zhang, M., and Gridley, T. (1993). Hoxa-2 mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. Cell *75*, 1317-1331.

Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., and Carroll, S.B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature *433*, 481-487.

Goulding, M.D., and Gruss, P. (1989). The homeobox in vertebrate development. Curr Opin Cell Biol *1*, 1088-1093.

Heine, P., Dohle, E., Bumsted-O'Brien, K., Engelkamp, D., and Schulte, D. (2008). Evidence for an evolutionary conserved role of homothorax/Meis1/2 during vertebrate retina development. Development *135*, 805-811.

Hisa, T., Spence, S.E., Rachel, R.A., Fujita, M., Nakamura, T., Ward, J.M., Devor-Henneman, D.E., Saiki, Y., Kutsuna, H., Tessarollo, L.*, et al.* (2004). Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. EMBO J *23*, 450-459.

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

Hurley, I., Hale, M.E., and Prince, V.E. (2005). Duplication events and the evolution of segmental identity. Evol Dev *7*, 556-567.

Huxley, J. (1943). Evolution : the modern synthesis (New York: London, Harper & brothers).

Inbal, A., Halachmi, N., Dibner, C., Frank, D., and Salzberg, A. (2001). Genetic evidence for the transcriptional-activating function of Homothorax during adult fly development. Development *128*, 3405-3413.

Jacobs, Y., Schnabel, C.A., and Cleary, M.L. (1999). Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity. Mol Cell Biol *19*, 5134-5142.

Kague, E., Bessling, S.L., Lee, J., Hu, G., Passos-Bueno, M.R., and Fisher, S. (2010). Functionally conserved cis-regulatory elements of COL18A1 identified through zebrafish transgenesis. Dev Biol *337*, 496-505.

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., 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 Shima, A. (1999). Identification of the Tol2 transposase of the medaka fish Oryzias latipes that catalyzes excision of a nonautonomous Tol2 element in zebrafish Danio rerio. Gene *240*, 239-244.

Kawakami, K., Shima, A., and Kawakami, N. (2000). Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. Proc Natl Acad Sci U S A *97*, 11403-11408.

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

Kessel, M. (1992). Respecification of vertebral identities by retinoic acid. Development *115*, 487-501.

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

Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A.Z., Engstrom, P.G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K.*, et al.* (2007). Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res *17*, 545-555.

Kilstrup-Nielsen, C., Alessio, M., and Zappavigna, V. (2003). PBX1 nuclear export is regulated independently of PBX-MEINOX interaction by PKA phosphorylation of the PBC-B domain. EMBO J *22*, 89-99.

King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107-116.

Kleinjan, D.A., Bancewicz, R.M., Gautier, P., Dahm, R., Schonthaler, H.B., Damante, G., Seawright, A., Hever, A.M., Yeyati, P.L., van Heyningen, V.*, et al.* (2008). Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence. PLoS Genet *4*, e29.

Knoepfler, P.S., Bergstrom, D.A., Uetsuki, T., Dac-Korytko, I., Sun, Y.H., Wright, W.E., Tapscott, S.J., and Kamps, M.P. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. Nucleic Acids Res *27*, 3752-3761.

Knoepfler, P.S., Calvo, K.R., Chen, H., Antonarakis, S.E., and Kamps, M.P. (1997). Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1. Proc Natl Acad Sci U S A *94*, 14553-14558.

Knoepfler, P.S., and Kamps, M.P. (1997). The Pbx family of proteins is strongly upregulated by a post-transcriptional mechanism during retinoic acid-induced differentiation of P19 embryonal carcinoma cells. Mech Dev *63*, 5-14.

Kobayashi, M., Fujioka, M., Tolkunova, E.N., Deka, D., Abu-Shaar, M., Mann, R.S., and Jaynes, J.B. (2003). Engrailed cooperates with extradenticle and homothorax to repress target genes in Drosophila. Development *130*, 741-751.

Kotani, T., and Kawakami, K. (2008). Misty somites, a maternal effect gene identified by transposon-mediated insertional mutagenesis in zebrafish that is essential for the somite boundary maintenance. Dev Biol *316*, 383-396.

Kumar, A., Gates, P.B., and Brockes, J.P. (2007). Positional identity of adult stem cells in salamander limb regeneration. C R Biol *330*, 485-490.

Kurant, E., Pai, C.Y., Sharf, R., Halachmi, N., Sun, Y.H., and Salzberg, A. (1998). Dorsotonals/homothorax, the Drosophila homologue of meis1, interacts with extradenticle in patterning of the embryonic PNS. Development *125*, 1037-1048.

Kwan, K.M., Fujimoto, E., Grabher, C., Mangum, B.D., Hardy, M.E., Campbell, D.S., Parant, J.M., Yost, H.J., Kanki, J.P., and Chien, C.B. (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev Dyn *236*, 3088-3099.

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

Linney, E., Hardison, N.L., Lonze, B.E., Lyons, S., and DiNapoli, L. (1999). Transgene expression in zebrafish: A comparison of retroviral-vector and DNA-injection approaches. Dev Biol *213*, 207-216.

Linney, E., and Udvadia, A.J. (2004). Construction and detection of fluorescent, germline transgenic zebrafish. Methods Mol Biol *254*, 271-288.

Liu, Y., MacDonald, R.J., and Swift, G.H. (2001). DNA binding and transcriptional activation by a PDX1.PBX1b.MEIS2b trimer and cooperation with a pancreas-specific basic helix-loop-helix complex. J Biol Chem *276*, 17985-17993.

Maeda, R., Ishimura, A., Mood, K., Park, E.K., Buchberg, A.M., and Daar, I.O. (2002). Xpbx1b and Xmeis1b play a collaborative role in hindbrain and neural crest gene expression in Xenopus embryos. Proc Natl Acad Sci U S A *99*, 5448-5453.

Maeda, R., Mood, K., Jones, T.L., Aruga, J., Buchberg, A.M., and Daar, I.O. (2001). Xmeis1, a protooncogene involved in specifying neural crest cell fate in Xenopus embryos. Oncogene *20*, 1329-1342.

Mamo, A., Krosl, J., Kroon, E., Bijl, J., Thompson, A., Mayotte, N., Girard, S., Bisaillon, R., Beslu, N., Featherstone, M.*, et al.* (2006). Molecular dissection of Meis1 reveals 2 domains required for leukemia induction and a key role for Hoxa gene activation. Blood *108*, 622-629.

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., and Gehring, W.J. (1984a). A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell *37*, 403-408.

McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A., and Gehring, W.J. (1984b). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature *308*, 428-433.

Mercader, N., Leonardo, E., Azpiazu, N., Serrano, A., Morata, G., Martinez, C., and Torres, M. (1999). Conserved regulation of proximodistal limb axis development by Meis1/Hth. Nature *402*, 425-429.

Mercader, N., Tanaka, E.M., and Torres, M. (2005). Proximodistal identity during vertebrate limb regeneration is regulated by Meis homeodomain proteins. Development *132*, 4131-4142.

Milech, N., Gottardo, N.G., Ford, J., D'Souza, D., Greene, W.K., Kees, U.R., and Watt, P.M. (2010). MEIS proteins as partners of the TLX1/HOX11 oncoprotein. Leuk Res *34*, 358-363.

Moens, C.B., and Selleri, L. (2006). Hox cofactors in vertebrate development. Dev Biol *291*, 193-206.

Mojsin, M., and Stevanovic, M. (2010). PBX1 and MEIS1 up-regulate SOX3 gene expression by direct interaction with a consensus binding site within the basal promoter region. Biochem J *425*, 107-116.

Moskow, J.J., Bullrich, F., Huebner, K., Daar, I.O., and Buchberg, A.M. (1995). Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice. Mol Cell Biol *15*, 5434-5443.

Muller, F., Blader, P., and Strahle, U. (2002). Search for enhancers: teleost models in comparative genomic and transgenic analysis of cis regulatory elements. Bioessays *24*, 564-572.

Nakamura, T., Jenkins, N.A., and Copeland, N.G. (1996). Identification of a new family of Pbx-related homeobox genes. Oncogene *13*, 2235-2242.

Nam, J., and Nei, M. (2005). Evolutionary change of the numbers of homeobox genes in bilateral animals. Mol Biol Evol *22*, 2386-2394.

Navratilova, P., Fredman, D., Hawkins, T.A., Turner, K., Lenhard, B., and Becker, T.S. (2009). Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. Dev Biol *327*, 526-540.

Ogishima, S., and Tanaka, H. (2007). Missing link in the evolution of Hox clusters. Gene *387*, 21-30.

Otting, G., Qian, Y.Q., Billeter, M., Muller, M., Affolter, M., Gehring, W.J., and Wuthrich, K. (1990). Protein--DNA contacts in the structure of a homeodomain--DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J *9*, 3085-3092.

Oulad-Abdelghani, M., Chazaud, C., Bouillet, P., Sapin, V., Chambon, P., and Dolle, P. (1997). Meis2, a novel mouse Pbx-related homeobox gene induced by retinoic acid during differentiation of P19 embryonal carcinoma cells. Dev Dyn *210*, 173-183.

Pai, C.Y., Kuo, T.S., Jaw, T.J., Kurant, E., Chen, C.T., Bessarab, D.A., Salzberg, A., and Sun, Y.H. (1998). The Homothorax homeoprotein activates the nuclear localization of another homeoprotein, extradenticle, and suppresses eye development in Drosophila. Genes Dev *12*, 435-446.

Pankratz, M.T., Li, X.J., Lavaute, T.M., Lyons, E.A., Chen, X., and Zhang, S.C. (2007). Directed neural differentiation of human embryonic stem cells via an obligated primitive anterior stage. Stem Cells *25*, 1511-1520.

Prince, V.E., Joly, L., Ekker, M., and Ho, R.K. (1998). Zebrafish hox genes: genomic organization and modified colinear expression patterns in the trunk. Development *125*, 407-420.

Prince, V.E., and Pickett, F.B. (2002). Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet *3*, 827-837.

Prohaska, S.J., and Stadler, P.F. (2004). The duplication of the Hox gene clusters in teleost fishes. Theory Biosci *123*, 89-110.

Prpic, N.M., Janssen, R., Wigand, B., Klingler, M., and Damen, W.G. (2003). Gene expression in spider appendages reveals reversal of exd/hth spatial specificity, altered leg gap gene dynamics, and suggests divergent distal morphogen signaling. Dev Biol *264*, 119-140.

Rebeiz, M., Pool, J.E., Kassner, V.A., Aquadro, C.F., and Carroll, S.B. (2009). Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population. Science *326*, 1663-1667.

Rieckhof, G.E., Casares, F., Ryoo, H.D., Abu-Shaar, M., and Mann, R.S. (1997). Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. Cell *91*, 171-183.

Rottkamp, C.A., Lobur, K.J., Wladyka, C.L., Lucky, A.K., and O'Gorman, S. (2008). Pbx3 is required for normal locomotion and dorsal horn development. Dev Biol *314*, 23-39.

Ryoo, H.D., and Mann, R.S. (1999). The control of trunk Hox specificity and activity by Extradenticle. Genes Dev *13*, 1704-1716.

Ryoo, H.D., Marty, T., Casares, F., Affolter, M., and Mann, R.S. (1999). Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. Development *126*, 5137-5148.

Sagerstrom, C.G. (2004). PbX marks the spot. Dev Cell *6*, 737-738.

Sagerstrom, C.G., Kao, B.A., Lane, M.E., and Sive, H. (2001). Isolation and characterization of posteriorly restricted genes in the zebrafish gastrula. Dev Dyn *220*, 402-408.

Saleh, M., Huang, H., Green, N.C., and Featherstone, M.S. (2000). A conformational change in PBX1A is necessary for its nuclear localization. Exp Cell Res *260*, 105-115.

Salsi, V., Vigano, M.A., Cocchiarella, F., Mantovani, R., and Zappavigna, V. (2008). Hoxd13 binds in vivo and regulates the expression of genes acting in key pathways for early limb and skeletal patterning. Dev Biol *317*, 497-507.

Salzberg, A., Elias, S., Nachaliel, N., Bonstein, L., Henig, C., and Frank, D. (1999). A Meis family protein caudalizes neural cell fates in Xenopus. Mech Dev *80*, 3-13.

Santini, S., Boore, J.L., and Meyer, A. (2003). Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. Genome Res *13*, 1111-1122.

Sarno, J.L., Kliman, H.J., and Taylor, H.S. (2005). HOXA10, Pbx2, and Meis1 protein expression in the human endometrium: formation of multimeric complexes on HOXA10 target genes. J Clin Endocrinol Metab *90*, 522-528.

Schnabel, C.A., Jacobs, Y., and Cleary, M.L. (2000). HoxA9-mediated immortalization of myeloid progenitors requires functional interactions with TALE cofactors Pbx and Meis. Oncogene *19*, 608-616.

Schneuwly, S., Klemenz, R., and Gehring, W.J. (1987). Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. Nature *325*, 816-818.

Scott, M.P., and Weiner, A.J. (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proc Natl Acad Sci U S A *81*, 4115-4119.

Shanmugam, K., Green, N.C., Rambaldi, I., Saragovi, H.U., and Featherstone, M.S. (1999). PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins. Mol Cell Biol *19*, 7577-7588.

Shen, W.F., Montgomery, J.C., Rozenfeld, S., Moskow, J.J., Lawrence, H.J., Buchberg, A.M., and Largman, C. (1997). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. Mol Cell Biol *17*, 6448-6458.

Shen, W.F., Rozenfeld, S., Kwong, A., Kom ves, L.G., Lawrence, H.J., and Largman, C. (1999). HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells. Mol Cell Biol *19*, 3051-3061.

Smith, J.E., Jr., Bollekens, J.A., Inghirami, G., and Takeshita, K. (1997). Cloning and mapping of the MEIS1 gene, the human homolog of a murine leukemogenic gene. Genomics *43*, 99-103.

Stedman, A., Lecaudey, V., Havis, E., Anselme, I., Wassef, M., Gilardi-Hebenstreit, P., and Schneider-Maunoury, S. (2009). A functional interaction between Irx and Meis patterns the anterior hindbrain and activates krox20 expression in rhombomere 3. Dev Biol *327*, 566-577.

Steelman, S., Moskow, J.J., Muzynski, K., North, C., Druck, T., Montgomery, J.C., Huebner, K., Daar, I.O., and Buchberg, A.M. (1997). Identification of a conserved family of Meis1-related homeobox genes. Genome Res *7*, 142-156.

Stuart, G.W., McMurray, J.V., and Westerfield, M. (1988). Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos. Development *103*, 403-412.

Sumiyama, K., Kawakami, K., and Yagita, K. (2010). A simple and highly efficient transgenesis method in mice with the Tol2 transposon system and cytoplasmic microinjection. Genomics *95*, 306-311.

Swift, G.H., Liu, Y., Rose, S.D., Bischof, L.J., Steelman, S., Buchberg, A.M., Wright, C.V., and MacDonald, R.J. (1998). An endocrine-exocrine switch in the activity of the pancreatic homeodomain protein PDX1 through formation of a trimeric complex with PBX1b and MRG1 (MEIS2). Mol Cell Biol *18*, 5109-5120.

Takahashi, K., Liu, F.C., Oishi, T., Mori, T., Higo, N., Hayashi, M., Hirokawa, K., and Takahashi, H. (2008). Expression of FOXP2 in the developing monkey forebrain: comparison with the expression of the genes FOXP1, PBX3, and MEIS2. J Comp Neurol *509*, 180-189.

Taylor, J.S., and Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet *38*, 615-643.

Thorsteinsdottir, U., Kroon, E., Jerome, L., Blasi, F., and Sauvageau, G. (2001). Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia. Mol Cell Biol *21*, 224-234.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell *10*, 1453-1465.

Toresson, H., Mata de Urquiza, A., Fagerstrom, C., Perlmann, T., and Campbell, K. (1999). Retinoids are produced by glia in the lateral ganglionic eminence and regulate striatal neuron differentiation. Development *126*, 1317-1326.

Toresson, H., Parmar, M., and Campbell, K. (2000). Expression of Meis and Pbx genes and their protein products in the developing telencephalon: implications for regional differentiation. Mech Dev *94*, 183-187.

Tumpel, S., Cambronero, F., Wiedemann, L.M., and Krumlauf, R. (2006). Evolution of cis elements in the differential expression of two Hoxa2 coparalogous genes in pufferfish (Takifugu rubripes). Proc Natl Acad Sci U S A *103*, 5419-5424.

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.

Vennemann, A., Agoston, Z., and Schulte, D. (2008). Differential and dose-dependent regulation of gene expression at the mid-hindbrain boundary by Ras-MAP kinase signaling. Brain Res *1206*, 33-43.

Vlachakis, N., Choe, S.K., and Sagerstrom, C.G. (2001). Meis3 synergizes with Pbx4 and Hoxb1b in promoting hindbrain fates in the zebrafish. Development *128*, 1299-1312.

Vlachakis, N., Ellstrom, D.R., and Sagerstrom, C.G. (2000). A novel pbx family member expressed during early zebrafish embryogenesis forms trimeric complexes with Meis3 and Hoxb1b. Dev Dyn *217*, 109-119.

Wang, Y., Yin, L., and Hillgartner, F.B. (2001). The homeodomain proteins PBX and MEIS1 are accessory factors that enhance thyroid hormone regulation of the malic enzyme gene in hepatocytes. J Biol Chem *276*, 23838-23848.

Waskiewicz, A.J., Rikhof, H.A., Hernandez, R.E., and Moens, C.B. (2001). Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning. Development *128*, 4139-4151.

Wassef, M.A., Chomette, D., Pouilhe, M., Stedman, A., Havis, E., Desmarquet-Trin Dinh, C., Schneider-Maunoury, S., Gilardi-Hebenstreit, P., Charnay, P., and Ghislain, J. (2008). Rostral hindbrain patterning involves the direct activation of a Krox20 transcriptional enhancer by Hox/Pbx and Meis factors. Development *135*, 3369-3378.

Williams, T.M., Williams, M.E., and Innis, J.W. (2005). Range of HOX/TALE superclass associations and protein domain requirements for HOXA13:MEIS interaction. Dev Biol *277*, 457-471.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004). Evolutionary changes in cis and trans gene regulation. Nature *430*, 85-88.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K.*, et al.* (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol *3*, e7.

Xu, Q. (1999). Microinjection into zebrafish embryos. Methods Mol Biol *127*, 125-132.

Yang, X., Zhou, Y., Barcarse, E.A., and O'Gorman, S. (2008). Altered neuronal lineages in the facial ganglia of Hoxa2 mutant mice. Dev Biol *314*, 171-188.

Yang, Y., Hwang, C.K., D'Souza, U.M., Lee, S.H., Junn, E., and Mouradian, M.M. (2000). Three-amino acid extension loop homeodomain proteins Meis2 and TGIF differentially regulate transcription. J Biol Chem *275*, 20734-20741.

Zerucha, T., and Prince, V.E. (2001). Cloning and developmental expression of a zebrafish meis2 homeobox gene. Mech Dev *102*, 247-250.

Zerucha, T., Stuhmer, T., Hatch, G., Park, B.K., Long, Q., Yu, G., Gambarotta, A., Schultz, J.R., Rubenstein, J.L., and Ekker, M. (2000). A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. J Neurosci *20*, 709-721.

Zhang, X., Friedman, A., Heaney, S., Purcell, P., and Maas, R.L. (2002). Meis homeoproteins directly regulate Pax6 during vertebrate lens morphogenesis. Genes Dev *16*, 2097-2107.

Zhang, X., Huang, C.T., Chen, J., Pankratz, M.T., Xi, J., Li, J., Yang, Y., Lavaute, T.M., Li, X.J., Ayala, M.*, et al.* (2010). Pax6 is a human neuroectoderm cell fate determinant. Cell Stem Cell *7*, 90-100.

Zhang, X., Rowan, S., Yue, Y., Heaney, S., Pan, Y., Brendolan, A., Selleri, L., and Maas, R.L. (2006). Pax6 is regulated by Meis and Pbx homeoproteins during pancreatic development. Dev Biol *300*, 748-757.

BIOGRAPHICAL SKETCH

Kyle C. Nelson was born in Durham, NC on September 12, 1981. He received his undergraduate degree in Biology from Appalachian State University in 2008, during which time he began working with his Thesis Advisor Dr. Ted Zerucha. As an undergraduate he participated in a summer internship at the pharmaceutical company GlaxoSmithKline and was awarded the Sigma Xi award for Undergraduate Research. After graduation Kyle returned to Appalachian State University to continue his undergraduate project and pursue his Masters in Cell and Molecular Biology. As a graduate student he served as the Biology Department representative to the Graduate Student Association Senate (GSAS) where he served as a Co-Treasurer. Kyle received his degree in 2011 and is continuing his education at Wake Forest University where he will pursue his Ph.D. in Molecular and Cellular Biosciences.