

IDENTIFICATION AND CHARACTERIZATION OF A HIGHLY CONSERVED
NONCODING ELEMENT ASSOCIATED WITH THE *MEIS2* GENE; M2DE2

A Thesis
By
Hannah Hemingway Freundlich

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

May 2016
Department of Biology

IDENTIFICATION AND CHARACTERIZATION OF A HIGHLY CONSERVED
NONCODING ELEMENT ASSOCIATED WITH THE *MEIS2* GENE; M2DE2

A Thesis
by
HANNAH HEMINGWAY FREUNDLICH
May 2016

APPROVED BY:

Dr. Ted Zerucha
Chairperson, Thesis Committee

Dr. Ece Karatan
Member, Thesis Committee

Dr. Annkatrin Rose
Member, Thesis Committee

Dr. Zack E. Murrell
Interim Chairperson, Department of Biology

Dr. Max Poole
Dean, Cratis D. Williams School of Graduate Studies

Copyright by Hannah Hemingway Freundlich 2016
All Rights Reserved

Abstract

IDENTIFICATION AND CHARACTERIZATION OF A HIGHLY CONSERVED NONCODING ELEMENT ASSOCIATED WITH THE *MEIS2* GENE; M2DE2

Hannah Hemingway Freundlich

B.S., North Carolina State University, M.S., Appalachian State University

Chairperson: Dr. Ted Zerucha

The *Meis* genes are highly conserved across species and play important roles in embryogenesis. There are four known members of the *Meis* gene family in vertebrates, *Meis1-Meis4*. Because of the genome duplication event that occurred in the teleost lineage following the divergence from the lineage that would give rise to land vertebrates, zebrafish have two copies of the *Meis2* gene, *meis2a* and *meis2b*, in contrast to the single *Meis2* gene in tetrapods.

We have identified four highly conserved noncoding elements (CNEs) in tetrapods that we hypothesize direct *Meis2* expression. We have named these m2de1-4 (for Meis2 downstream element). To date only one of these has been identified in zebrafish.

The purpose of this study was to characterize m2de2 using zebrafish as a model organism. Using the Tol2 system, expression constructs containing mouse m2de2 that drove expression of eGFP through the *cfos* minimal promoter were microinjected into zebrafish embryos at the single cell stage. Confocal microscopy was used to determine eGFP

expression at different time points during development. Expression was observed in specific neurons in the brain of the developing zebrafish embryos in a pattern consistent with that observed for the murine *Meis2* gene. eGFP was also observed in developing muscle fibers in the trunk of developing zebrafish embryos.

Acknowledgements

I am sincerely grateful to Dr. Ted Zerucha for his patient guidance and understanding during this research and the development of the thesis. I consider myself lucky to have been one of his graduate students. I would like to thank Dr. Ece Karatan from the bottom of my heart for her help and guidance with my thesis. My grateful thanks are extended to Dr. Annkatrin Rose for her willingness to work with me as a member of my committee. I would also like to thankfully acknowledge past and current Zerucha lab graduate students, specifically Kyle Nelson, Cody Barrett, Zach Williams, Alicia Ramsaran, Tucker Munday, Tyler Ferrara, and Tray Neilson, for all their help and advice with this research. I have received help from several undergraduate students and would like to thank each of them as well; this includes Patrick Bearden, Rachel Lemmond, Harlie Walkup, Riley Parr, and Mackenzie Trapp. I would also like to thank Lucy Edy, the Zerucha lab high school intern, for her help in caring for zebrafish regardless of transgenic state.

Several members of the Appalachian faculty have assisted in this project with advice, or assistance and I would like to thank them for their time. Monique Eckerd runs a wonderful animal facility where the fish are housed. Dr. Hou has trained me in the use of the LSM confocal microscope that was used for each of the GFP pictures. Dr. Estep helped me troubleshoot and perfect genome extractions and PCR screen transgenic fish. Dr. Kinkel assisted with breeding and raising zebrafish even before she began her time at Appalachian State University. Thank you each for your help with these issues.

Additionally, advice and assistance given by graduate students not in the same lab as me has been very helpful and greatly appreciated. These helpful souls include Laura Ellis, Libby Villa, Richard Sobe, Julie Ragsdale, Audrey Brown, Blake Sanders, Cameron Houser, Will Brennen, and Randy Cockerell. Thank you to each of you. The entire Appalachian family has been wonderful and I would like to thank the school as a whole. The Department of Biology has been very supportive of my education and research. The Appalachian State University Cratis D. Williams Graduate School, and the Appalachian State University Office of Student Research have supported my research with funding and presentation opportunities. Without their support I would not have been able to present this thesis. Also, I would like to thank Dr. Cama Duke for help with writing.

Finally, I would like to thank my family for their endless support. My parents, Hilary Hemingway and Jeff Freundlich, never gave up on me and pushed me forward. I would like to express my appreciation to Dr. Mary Held for her moral and technological support. I would like to offer my thanks to Anne Feuer and Bill Feuer for their support with my thesis. Thank you to Dr. August Freundlich for his advice.

Dedication

For Hilary Hemingway and Jeff Freundlich, who have dedicated so much to me in their lives.

Table of Contents

Abstract.....	iv
Acknowledgements.....	vi
Dedication.....	viii
Foreword.....	x
Introduction.....	1
Materials and Methods.....	31
Results and Discussion.....	48
Conclusions and Future Directions.....	69
References.....	72
Vita.....	83

Foreword

This thesis manuscript, methods, Figures, tables and references were structured after given instructions detailed for manuscript submission to the peer-reviewed science journal *Developmental Biology*, the official journal for the Society for Developmental Biology, published by Elsevier.

Introduction

Cells are the basic unit of life, yet to survive they need to be able to produce a myriad of proteins. To control which proteins are produced and which genes are being expressed at any given time, cells must regulate a number of steps, arguably the most important being the regulation of DNA transcription (Butler and Kadonaga, 2001; Cianfrocco et al., 2013). Gene regulation in eukaryotic organisms is a complex process.

The most frequently occurring aspects of gene regulation include transcription initiation, *cis* regulatory elements, and transcription factors (Butler and Kadonaga, 2001). Transcription of a gene can either be repressed or activated through the use of *cis* regulatory elements and transcription factors, all of which leads to determination of cell fate (Andersson et al., 2015; Bell et al., 2001; Buratiwski et al., 1989; Butler and Kadonaga, 2001; Forget et al., 2010; Langelier et al., 2001; Venters and Pugh, 2009). In eukaryotes, the basic transcription process begins with a series of ubiquitous transcription factors that must first bind to DNA, seen in Fig. 1 (Akoulitchev et.al., 1995; Buratiwski et al., 1989; Butler and Kadonaga, 2001; Forget et al., 2010; Langelier et al., 2001; Pardee et al., 1998; Venters and Pugh, 2009). These general transcription factors (GTFs) assist in recruiting RNA polymerase II (RNAPII) to the promoter region, separating the DNA strands, and then releasing RNAPII to transcribe the DNA, seen in Fig. 1 (Akoulitchev et.al., 1995; Buratiwski et al., 1989; Forget et al., 2010; Langelier et al., 2001; Orphanides, et al., 1996; Pardee et al., 1998; Venters and Pugh, 2009).

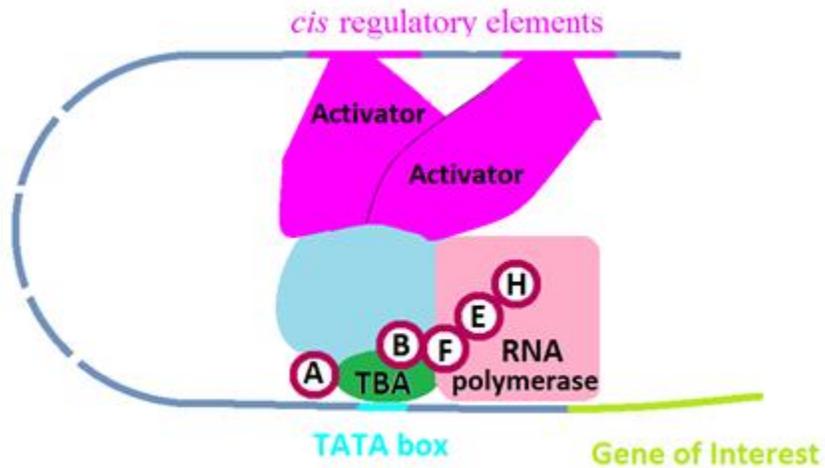


Fig. 1. Simplified drawing depicting common elements to the start of eukaryotic transcription. The blue grey line represents DNA, while purple fragments indicate *cis* regulatory elements. The single lettered circles represent general transcription factors attached to the green TATA-binding protein and pink RNA polymerase. The breaks in the blue DNA strand represent a pictorial cut in the same chromosome but several hundred base pair have been hidden between them. The Fig. also shows activators bound to *cis* regulatory elements, upstream of the core promoter, that assist with transcription. (Based on Fig. 3 Tefferi et al., 2002)

Transcription factor IID (TFIID) is the first GTF to bind to the DNA (Buratiwsky et al., 1989; Pardee et al., 1998; Thomas and Chiange, 2006). TFIID is a core promoter-binding factor and it assists with identifying promoters; it can interact with the TATA box, and the TATA-containing promoters, or the TATA-less promoters, allowing for site-specific transcription (Buratiwsky et al., 1989; Cianfrocco et al., 2013; Pardee et al., 1998). TFIID is composed of multiple proteins including the TATA binding protein (TBP) which binds to an upstream region of DNA known as the TATA box (Buratiwsky et al., 1989; Butler and Kadonaga, 2001; Cianfrocco et al., 2013; Thomas and Chiange, 2006). While the TATA box is not associated with all genes, it is common and frequently located around 30 base pairs

(bp) upstream from the start of transcription (Butler and Kadonaga, 2001; Cianfrocco et al., 2013; Venters and Pugh, 2009). Additionally, the thymine and adenine found in the TATA box makes it easier for the TBP to assist with the DNA unwinding process (Butler and Kadonaga, 2001). A transcription initiation complex is then formed with transcription factor II B, E, and H, and completed with the final addition of RNAPII (Akoulitchev et.al., 1995; Cianfrocco et al., 2013; Pardee et al., 1998).

TFIIB is next to act in transcription initiation by attaching to the TBP, increasing stability of the complex and it assists in recruiting RNAPII to form a TBP-TFIIB-DNA structure (Pardee et al., 1998; Schaeffer et al., 1993; Thomas and Chiange, 2006; Venters and Pugh, 2009). TFIIB, in turn, gains its structural integrity from the zinc ribbon motif (Pardee et al., 1998). Polymerase II and Transcription Factor IIF (TFIIF) interact with this zinc ribbon during the recruitment process that follows (Langelier et al., 2001; Thomas and Chiange, 2006). The N-terminal region of the Zinc ribbon motif in TFIIB contains the charged cluster domain known as the B-finger which is highly conserved; in humans it spans amino acids 44 to 75 (Thomas and Chiange, 2006). TFIIF works closely with RNA polymerase II and, in humans, is composed of two subunits, Rap74 and Rap30, which form a winged helix domain that results in suppressing non-specific binding between RNA polymerase II and DNA (Chen et al., 2010).

Transcription factor IIE (TFIIE) and Transcription factor IIH (TFIIH) work together to complete the list of general transcription factors that attach before transcription can begin;

these have been presented above in the order they attach. TFIIE interacts with RNAPII (Buratiwski et al., 1989; Stewart, and Stargell, 2001; Thomas and Chiange, 2006). TFIIE uses an ATP-independent mechanism to separate the DNA around the promoter near the transcription initiation site (Langelier et al., 2001). TFIIE is composed of α and β subunits, which form a heterotetramer (Thomas and Chiange, 2006). The α subunit has many relevant motif features that allow it to perform its function, including a zinc-finger between amino acids 113 and 174 and helix-turn-helix (Thomas and Chiange, 2006). The zinc-finger found in TFIIE is unusual because it has an antiparallel β -sheet followed by a middle α -helix followed by three β -strands (Thomas and Chiange, 2006). Polymerase II interacts with the β subunit and the N-terminal end of the α subunit (Thomas and Chiange, 2006). The C-terminal end of the α subunit attaches to transcription factor IIH (Thomas and Chiange, 2006). TFIIH works with TFIIE, but also acts uniquely as a helicase (Langelier et al., 2001; Schaeffer et al., 1993; Venters and Pugh, 2009). Using ATP, TFIIH dislodges histones and unwinds the DNA similar to helicase (Langelier et al., 2001; Schaeffer et al., 1993; Venters and Pugh, 2009). TFIIH also has a role during DNA repair (Schaeffer et al., 1993). It is after these general transcription factors bind that transcription can begin.

While general transcription factors are needed to begin transcription, they are not the only factors that control gene transcription (Andersson et al., 2015; Bell et al., 2001; Butler and Kadonaga, 2001; Forget et al., 2010; Venters and Pugh, 2009). Not all DNA is transcribed in every cell and many genes are only active during specific time points in an

organism's life (Bhatia et al., 2014). *Cis* regulatory elements are one way eukaryotic organisms control gene transcription (Bhatia et al., 2014; Butler and Kadonaga, 2001; Duboule, 1998; Wittkopp and Kalay, 2012). These elements are found along the same chromosome of DNA as the gene whose expression they control (Andersson et al., 2015; Bhatia et al., 2014; Blackwood and Kadonaga, 1998; Duboule, 1998; Venters and Pugh, 2009; Wittkopp and Kalay, 2012). Because DNA is flexible, the *cis* regulatory elements can be located upstream or downstream of the gene they work with (Bhatia et al., 2014; Duboule, 1998; Venters and Pugh, 2009; Wittkopp and Kalay, 2012). *Cis* regulatory elements contain binding sites for specific transcription factors and other regulatory molecules (Bhatia et al., 2014; Duboule, 1998; Wittkopp and Kalay, 2012). Through mediators, they are able to interact with the general transcriptional machinery at the promoter through DNA looping. Because they work to regulate when and where a gene is transcribed into mRNA and then translated into protein, *cis* regulatory elements can have a huge potential impact on the development of an organism and can play a role in the phenotypic differences between species that show very similar sequences of the genes they are associated with (Bhatia et al., 2014; Wittkopp and Kalay, 2012). Even though they do not code for proteins themselves, *cis* regulatory elements are often well conserved between species likely due to their developmental importance (Bhatia et al., 2014). Changes to the nucleotide sequences of *cis* regulatory elements particularly within transcription factor binding sites can change the spatial and temporal expression of the gene they control (Bhatia et al., 2014; Wittkopp and Kalay, 2012). Furthermore, a long range *cis* regulatory element can be used to tightly

regulate and control the expression of specific genes, especially during development; this is the case with *Pax6* (Bhatia et al., 2014). There are over 500 highly conserved sequences between chimps and other mammal genomes that are absent in the human genome; these sequences seem to represent *cis* regulatory elements and have been hypothesized to cause the noticeable differences between humans and chimps (Wittkopp and Kalay, 2012). Conversely, some *cis* regulatory elements, like *HACNS1*, are more active in humans than they are in other animals, like chimps, seen in Fig. 2; while both chimp and human *HACNS1* activity occurs in the ventral portion of the neck, the developing ear, and in the eye, human activation also occurs in the anterior limb buds of both the forelimb and hind limb, seen in Fig. 2 (Wittkopp and Kalay, 2012).

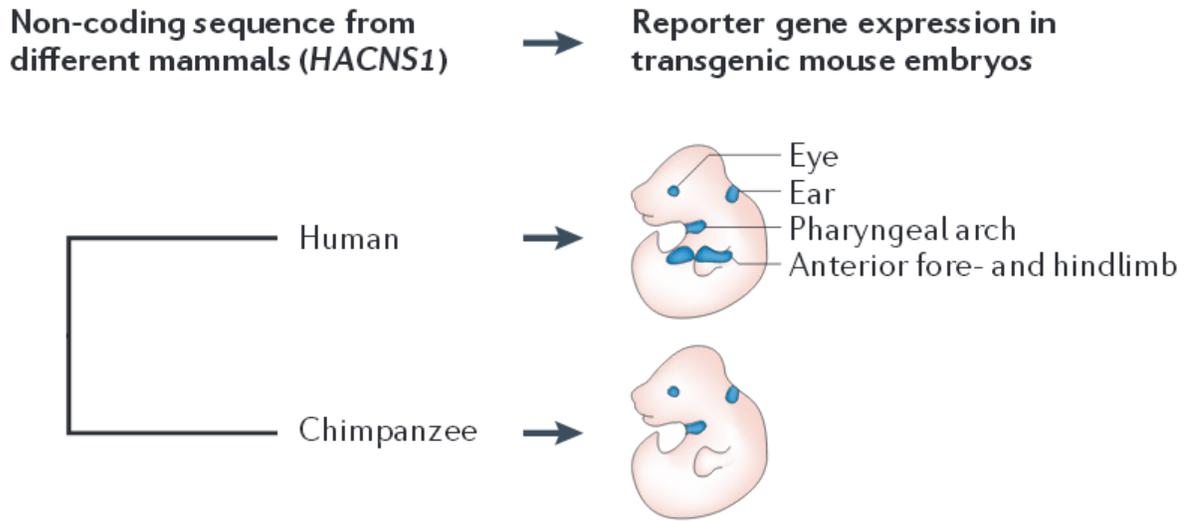


Fig. 2. *HACNS1* enhancer activity in human and chimpanzee. The *HACNS1* orthologs found in human and chimpanzee are active in many of the same locations but also express in some difference locations when expressed in transgenic mice at embryonic day 11.5 and represented in the drawing. The blue portion is where the expression was noted (Wittkopp and Kalay, 2012, Modified Fig. 3b).

Some examples of transcription factors that interact with *cis* regulatory elements include, Meis, Hox, Pbx, and MyoD. When transcription factors bind to *cis* regulatory elements, they are able to regulate the spatial and temporal expression of the gene with which they are associated. Transcription factors often contain well characterized DNA-binding motifs that are found in multiple proteins. They can include motifs like zinc fingers, and helix-turn-helix motifs and the homeodomain. The zinc finger motif has been mentioned previously and is found in more transcription factors than the general transcription factors. Zinc finger proteins attach to the major groove of DNA following specific sequences

(Pavletich and Pabo, 1991). Proteins with zinc finger motifs can bind to either RNA or DNA but usually not both (Shi and Berg, 1995).

Eukaryotic cells express numerous different genes that serve many functions; the homeobox genes, which encode homeodomain proteins, are of particular importance as they have been shown to play many crucial roles during development by regulating the expression of other genes. One example are the homeobox containing *Hox* genes (Allen et al., 2000; Choe et al., 2002; Jacobs et al., 1999; Krumlauf, 1994; Lemons and McGinnis, 2006; McGinnis and Krumlauf, 1992). Depending on the genome, the number of *Hox* genes each eukaryote possesses differ and the number of these genes can be between 4 and 51 which are arranged in clusters (Choe et al., 2002; Jacobs et al., 1999; Jave-Suárez and Schweizer, 2006; Krumlauf, 1994; Lemons and McGinnis, 2006; McGinnis and Krumlauf, 1992).

By examining the overall function of homeobox genes, it is clear that they are important for the development of eukaryotes and direct expression of other genes during development, and are highly conserved across multiple species (Bürglin, 1997; Chang et al., 1996; Chariot et al., 1999; Jacobs et al., 1999; McGinnis and Krumlauf, 1992; Mukherjee and Bürglin, 2007). Most homeobox genes code for proteins that function as transcription factors (Jacobs et al., 1999; McGinnis and Krumlauf, 1992; Schnabel et al., 2000). Their further subclassification depends on their sequence and to a certain extent how and where they are expressed. Homeobox genes were first studied in *Drosophila* with the identification of the homeotic or *Hox* genes (Chariot et al., 1999; Krumlauf, 1994; Lappin et al., 2006;

McGinnis and Krumlauf, 1992; Shang et al., 1994). Homeobox genes encode for homeodomain proteins. Some homeodomain proteins specify identity along the anterior-posterior axis of vertebrates during development, and play a role in the function of the hematopoietic system (Chang et al., 1996; Jacobs et al., 1999; Krumlauf, 1994; McGinnis and Krumlauf, 1992; Schnabel et al., 2000).

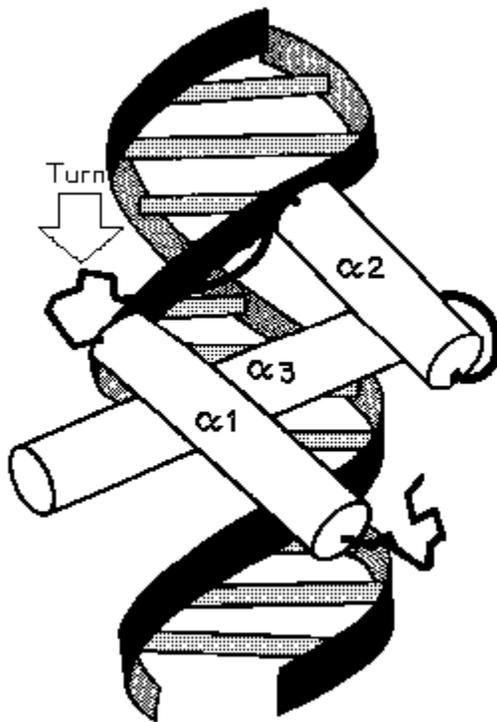


Fig. 3. Configuration of homeodomain protein attachment to DNA. Alpha helix one and two are parallel to each other and connected by a turn. The third alpha helix, the recognition helix, is in contact with both strands of DNA. (Based on Gehring et al., 1994)

Within the structure of the homeodomain proteins lies a DNA-binding structural motif known as the homeodomain. The homeodomain structure has been well conserved in eukaryotes and is involved in the control of transcription for many developmental genes

(Banerjee-Basu et al., 2001; McGinnis and Krumlauf, 1992; Shang et al., 1994). A major component of this protein domain is its helix-turn-helix motif which allows it to bind to DNA, seen in Fig. 3. The homeodomain is composed of approximately 60 amino acids that forms a 3- α helix bundle, seen in Fig. 3 (Banerjee-Basu et al., 2001; Bürglin, 1997; Dror et al., 2014; Fognani et al., 2002; Jave-Suárez and Schweizer, 2006; Lappin et al., 2006; Longobardi et al., 2014; McGinnis and Krumlauf, 1992; Mukherjee and Bürglin, 2007; Shang et al., 1994). The first two helices (I, II) lie parallel to each other and form a helix loop helix, while the third sits across from them forming a helix-turn-helix with helix two, as seen in Fig. 3 (Banerjee-Basu et al., 2001; Lappin et al., 2006). Several amino acids throughout the protein interact directly with the DNA backbone through either intermolecular electrostatic interactions or hydrophobic contacts (Gehring et al., 1994). The third α helix, interacts with both the α and β strand of DNA; this interaction stabilizes the complex on the DNA (Gehring et al., 1994). In the *Drosophila* Antp homeodomain, Gln-44 and Met-54 bind to the phosphate group on both strands of DNA, while Lys-46, 57, and Arg-53 bind to the α strand (Gehring et al., 1994). In contrast, Arg-43, 52, Lys-55 and Ile-47 bind to the β strand of DNA (Gehring et al., 1994). On the first turn of the Antp homeodomain, Lys-46 and Arg-43 connect to DNA, while the C-terminal turn of the Antp recognition helix forms four salt bridges (Gehring et al., 1994). The recognition helix provides a distinctive structure that inserts into DNA and acts as a functional group (Banerjee-Basu et al., 2001; Dror et al., 2014; Lappin et al., 2006). When interacting with DNA, the third helix lies in the major groove, often at a TAAT or ATTA nucleotide sequence on the β strand of the DNA (Dror et

al., 2014; Gehring et al., 1994; Jave-Suárez and Schweizer, 2006; Lappin et al., 2006). In the major groove there are two mechanisms used by the protein for DNA recognition. The first is the sequence dependent shape of the helix, while the second mechanism relies on the formation of hydrogen bonds within the major groove (Rohs et al., 2009). The formation of hydrogen bonds between amino acid side chains involves hydrogen bond donors and acceptors of individual base pairs and is frequently involved with nucleotide sequence-specific interactions (Dror et al., 2014; Rohs et al., 2009). The Antp homeodomain, which has a DNA recognition site of GAAAGCCATTAGAG, contains a ATTA core; this previously mentioned sequence is frequently present in DNA recognition sites bound by homeodomain proteins although there is greater variability in the nucleotide sequences flanking this sequence (Gehring et al., 1994).

The N-terminal tail is crucial to the shape read out, which occurs when this part of the protein attaches to the minor groove of the DNA sequence (Dror et al., 2014; Gehring et al., 1994; Lappin et al., 2006; Shang et al., 1994). Residues 1-6 of the Antp homeodomain, in the N-terminal tail, attach to base pairs 11-13 of the recognition sequence in the minor groove (Gehring et al., 1994). Without the N-terminal region of the protein attaching to the DNA, the protein complex's binding affinity is greatly hindered (Gehring et al., 1994). Arg-3 connects to the phosphate group of G12 through a salt bridge, while Arg-5 connects to G12, A13, T11 and G12 by hydrophobic interactions with the sugar moieties (Gehring et al.,

1994). On the β strand of the DNA, Gln-6 and Tyr-8 contact the backbone of A10 (Gehring et al., 1994).

As previously mentioned, homeodomains frequently bind to TAAT DNA sites, but within the *Hoxb1* enhancer, Hox works in conjunction with the homeobox protein Pbx and when they work together the binding site changes to TGATTGAT within the mouse model, and a similar sequence is seen in *Drosophila* with the Pbx homolog Exd (Ferretti et al., 2006; Krumlauf, 1994; McGinnis and Krumlauf, 1992; Waskiewicz et al., 2001). While there are many conserved amino acids within different homeodomains, it is the amino acids found within the three alpha helices that bind to the TGATTGAT sites (Ferretti et al., 2006; Krumlauf, 1994; McGinnis and Krumlauf, 1992; Waskiewicz et al., 2001). Non-conserved residues are found in the N-terminal arm that allow for high affinity DNA-binding activity in the minor groove of DNA and also interactions with other proteins, as in the case with homeodomain proteins Hox, Meis, and Pbx (Shang et al., 1994; Steelman et al., 1997).

Specific nucleotide sequences of binding sites forms unique three-dimensional shapes that transcription factor proteins recognize and preferentially bind (Dror et al., 2014). Arginine bound in the minor groove assists in the protein-DNA recognition process by narrowing the minor groove; this enhances the negative electrostatic potential of the DNA (Rohs et al., 2009). Variations in these sequences of even a single nucleotide can alter the overall shape and change the binding affinity (Dror et al., 2014).

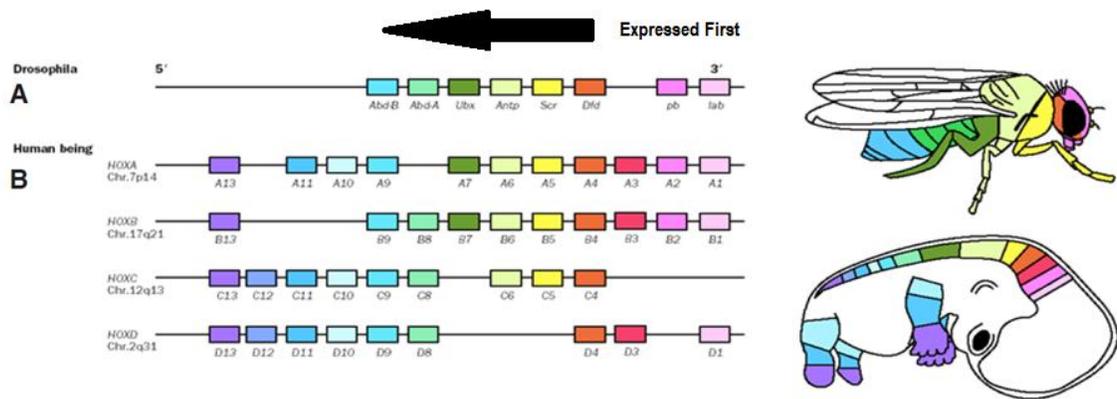


Fig. 4. Hox colinearity expression patterns. The *Hox* genes and orthologs follow a consistent pattern across various organisms, where 3' genes are expressed more anteriorly, and before the 5' genes, which are expressed more posteriorly. (modified from Fig. 1. Durston et al., 2011)

Hox genes encode proteins that bind DNA, guiding cell fate along the anterior-posterior axis (AP). These genes are organized into clusters that are first active during early gastrulation and continue as they pattern this major body axis (Amores et al., 1998; Krumlauf, 1994; Lappin et al., 2006; Lemons and McGinnis, 2006; McGinnis and Krumlauf, 1992; Waskiewicz et al., 2001). *Hox* genes are conserved across species, (Fig. 4), and are master embryonic development gene regulators. However, *Hox* activity continues throughout the organism's life (Krumlauf, 1994; Lappin et al., 2006; McGinnis and Krumlauf, 1992). *Hox* loss-of-function experiments show homeotic transformations along the anterior posterior axis, such as when a fly has legs develop where their antenna should be (Krumlauf, 1994; McGinnis and Krumlauf, 1992; Waskiewicz et al., 2001). In teleosts there are seven *Hox* clusters, while tetrapods have four *Hox* clusters (Amores et al., 1998; Krumlauf, 1994; Lappin et al., 2006). This presence of additional *Hox* genes in the teleosts is thought to be

due to a whole genome duplication event in the teleosts lineage after the divergence of the tetrapod lineage (Amores et al., 1998; Lappin et al., 2006; Lemons and McGinnis, 2006). The order and location of the *Hox* genes within a cluster on a given chromosome correlate to the anterior-posterior location and timing of expression of these genes; where the 3' genes are expressed more anteriorly and earlier during development than 5' genes, this is known as spatial and temporal “colinearity” (Krumlauf, 1994; Lappin et al., 2006; McGinnis and Krumlauf, 1992). As can be seen in Fig. 5, *hoxb* expression in zebrafish at the 20 somite stage is ordered and colinear, where *hoxb4* is expressed more anteriorly than *hoxb10*, which is confined posteriorly to somite 7 (Prince et al., 1998).

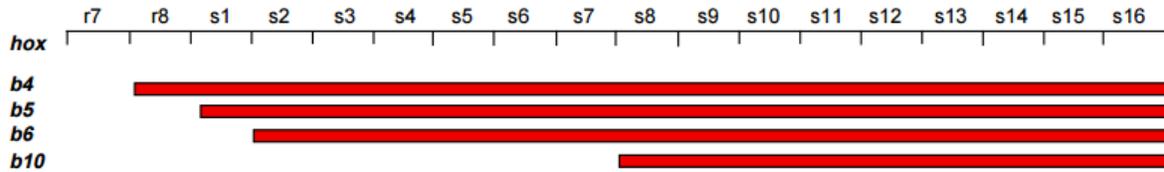


Fig. 5. *hoxb* cluster colinearity expression limits in the developing zebrafish embryo. The anterior boundary for zebrafish *hoxb* at the 20 somite stage is depicted as a bar graph above, where r stands for the rhombomere, and s represents the somite number where the expression patterns end. As can be seen, *hoxb4* ends between r7 and r8 while each subsequent *hoxb* member expression ends more posteriorly. (Modified Fig. 5 from Prince et al., 1998)

Additionally, mutations in *Hox* genes can lead to congenital lung defects, limb deformations, such as hand-foot-genital syndrome, and cancers (Krumlauf, 1994; Lappin et al., 2006; Schnabel et al., 2000). For example, almost two dozen *Hox* genes are active in the chick embryo to develop one limb, and many limb formation abnormalities in multiple

animals can be linked back to *Hox* mutations (Krumlauf, 1994; Lappin et al., 2006; McGinnis and Krumlauf, 1992).

Aside from the increase in number of zebrafish *hox* genes, they are generally similar in patterning to murine and avian counterparts with differences in timing and spatiality, underlining functional changes during the evolutionary split of teleosts and tetrapods (Krumlauf, 1994; Prince et al., 1998). This similar *Hox* patterning within animals is to be expected given the similarity seen in these genes that can be traced across the animalia kingdom.

While expression patterns can be similar across species, it is the subtle differences in the location proteins are expressed that cause the separation, and distinction that form species. While duplication and mutations within the duplicated genes can give rise to these species differences, alternative splicing can also cause differences. Alternative splicing within homeobox genes allows different homeodomain products to be produced using the same promoter; however, this includes transcription products generated that do not encode a homeodomain (Bürglin, 1997; Jave-Suárez and Schweizer, 2006; Longobardi et al., 2014; Magnani and Hake, 2008). The lack of a homeodomain in splice variants is interesting as it would be likely that these proteins are playing some role that is not dependent on binding DNA (Bürglin, 1997; Jave-Suárez and Schweizer, 2006; Longobardi et al., 2014; Magnani and Hake, 2008). PBC, MEIS and KNOX (plant) are known to use alternative splicing events outside the conserved regions to generate variations within the C-termini (Bürglin, 1997;

Jave-Suárez and Schweizer, 2006; Longobardi et al., 2014; Magnani and Hake, 2008; Tamaoki et al., 1995). The variations within the C-terminal region are believed to alter the binding specificity allowing the protein to bind to slightly different DNA sequences (Tamaoki et al., 1995). An example of alternative splicing takes place with *Pax6* in both zebrafish and mice (Puschel et al., 1992.). Two splice variants have been identified for *Pax6* and both protein products are found within both zebrafish and mice (Puschel et al., 1992). Both alternate splicing *Pax6* products function within the mouse developing eyes and brain (Epstein et al., 1994). Additionally, the zebrafish protein product has a similar expression pattern to the mice protein product, despite the evolutionary distance between the organisms (Epstein et al., 1994; Puschel et al., 1992).

Within the homeodomain family lies a super class of proteins that contain three additional amino acids between the first and second helix, generating an extended loop between helix I and helix II of their homeodomains (Bürglin, 1997; Derelle et al., 2007; Fognani et al., 2002; Jackson et al., 2011; Liu et al., 2001; Longobardi et al., 2014; Magnani and Hake, 2008; Mercader et al., 2005). These three amino acids are typically proline-tyrosine-proline and are consistently found at positions 24-26 of the homeodomain. The Three Amino acid Loop Extension (TALE) super class of homeobox genes include the *MEIS*, *PBX*, and *PREP* genes (Bellaoui et al., 2001; Bürglin, 1997; Choe et al., 2002; Ferretti et al., 2006; Fognani et al., 2002; Jacobs et al., 1999; Jave-Suárez and Schweizer, 2006; Liu et al., 2001; Longobardi et al., 2014; Magnani and Hake, 2008; Mercader et al., 2005; Selleri et al.,

2004; Stankunas et al., 2008; Steelman et al., 1997). While non-TALE homeodomains contain leucine and phenylalanine or tyrosine at positions 16 and 20 respectively, the TALE homeodomains show a greater variety of amino acids at those positions (Bürglin, 1997). Non-TALE homeodomains contain a polar amino acid at position 50 because it assists with DNA binding specificity at this very critical position, yet TALE homeodomains have smaller non-polar amino acids at this location in many cases (Bürglin, 1997; Chang et al., 1996). This unique residue suggests TALE proteins bind to DNA differently than non-TALE homeodomains (Bürglin, 1997).

TALE proteins, such as Meis and Pbx, assist to control *Hox* gene activation (Choe et al., 2002). These TALE homeodomain proteins have been shown to interact with non-TALE homeodomain proteins, like Hox, to form dimers or trimers; it is thought that this increases the stability of Hox or other homeodomain proteins on DNA (Allen et al., 2000; Bürglin, 1997; Choe et al., 2002; Jackson et al., 2011; Jacobs et al., 1999; Lappin et al., 2006; Liu et al., 2001; Longobardi et al., 2014; Magnani and Hake, 2008; Mukherjee and Bürglin, 2007; Selleri et al., 2004; Schnabel et al., 2000; Stankunas et al., 2008; Waskiewicz et al., 2001). This protein-protein interaction is typically through a pentamer amino acid sequence frequently found upstream of the homeodomain that binds to TALE cofactors (Lappin et al., 2006). For example, Hoxb1 has been shown to use both a Meis-binding element and Pbx in rhombomere 4 of the mouse developing hindbrain (Waskiewicz et al., 2001).

TALE proteins play additional roles in gene regulation. Meis and Pbx recruit the transcriptional coactivator CREB-binding protein (CBP) to actively transcribe *Hox* (Choe et al., 2002). TALE factors also recruit histone modifying enzymes independent of Hox and prior to gene activation; these factors can increase acetylation of already highly acetylated histones at *Hox* loci beyond basal levels (Choe et al., 2002). Additionally, TALE factors can activate silent chromatin at *Hox* loci (Choe et al., 2002; Fognani et al., 2002). While they can start transcription without Hox, these factors work significantly better when working with Hox as a complex (Choe et al., 2002; Fognani et al., 2002).

The common ancestor of all eukaryotes is thought to have had at least two different types of homeodomain proteins that were unique to eukaryotes (Bürglin, 1997; Derelle et al., 2007; Iyer et al., 2008). The Helix-Turn-Helix (HTH) proteins that bind to DNA, such as homeodomain proteins, are unique to eukaryotes but distantly related to helix-turn-helix motifs found in prokaryotes in sequence suggesting their evolution occurred after the split between eukaryotes and prokaryotes (Iyer et al., 2008). It is thought that homeodomain proteins evolved from the common ancestor of plants, fungi, and animals; during this time the genes that encode them underwent a duplication event prior to the divergence from the first eukaryotes and Protists (Derelle et al., 2007). Within single cell eukaryotes, previously known as the kingdom Protista, TALE homeodomains are found sporadically suggesting either an earlier origin for TALE with frequent losses or their evolution within already radiated Protistans (Derelle et al., 2007; Iyer et al., 2008). The *Trichomonas* (Protista)

genome contains one non-TALE homeobox gene while it has eight TALE homeobox genes, which are thought to have arisen from five different duplication events (Derelle et al., 2007).

A comparison of plant, animal, and fungi TALE homeodomain proteins showed enough similarity to form an archetypal group dubbed MEINOX, as seen in Fig. 6 (Bellaoui et al., 2001; Bhatt et al., 2004; Bürglin, 1997; Fognani et al., 2002). This grouping of proteins includes MEIS, KNOX (plant), CUP (fungi), BEL (plant), and TGIF. This grouping is based on both their DNA-binding characteristics and the similarity at position nine of helix three (Bellaoui et al., 2001; Bhatt et al., 2004; Bürglin, 1997; Longobardi et al., 2014). This MEINOX protein presence in the kingdoms of Fungi, Plantae, and Animalia suggests its evolution prior to the separation from single celled eukaryotes, previously known as the kingdom Protista; the presence of TALE homeodomain proteins found in some Protists supports this hypothesis (Bürglin, 1997; Derelle et al., 2007).

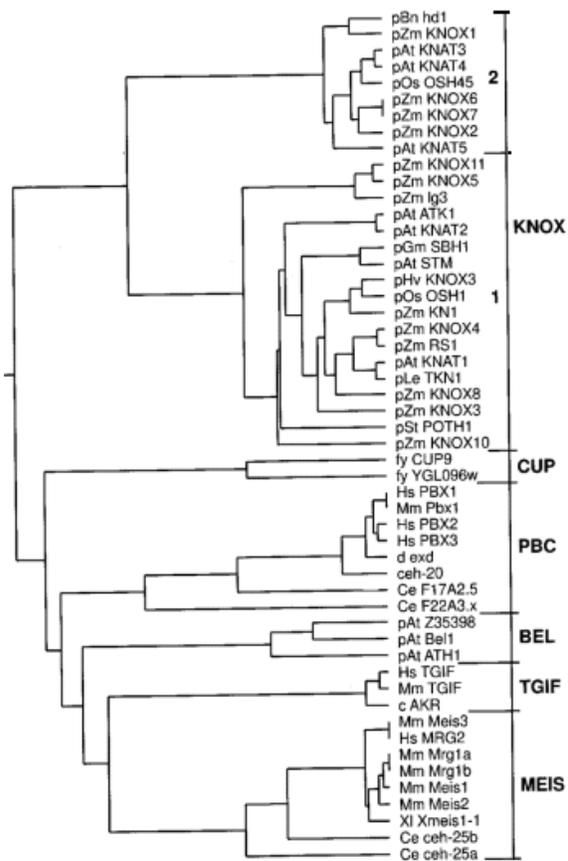


Fig. 6. Phylogenetic tree for MEINOX shows multiple species of TALE proteins; MEIS and KNOX are at either end of the comparison suggesting they are the least related. (Fig. edited from Bürglin, 1997)

The ability to trace the evolution of homeodomain proteins across kingdoms indicates the importance of these transcription factors and their roles in gene regulation being crucial for the development of all eukaryotic organisms. Further study of the homeodomain may provide answers to not only evolutionary questions but also further our understanding of embryogenesis through understanding of gene activation and regulation.

My thesis work is centered on a sequence of DNA found downstream of the homeobox gene *Meis2*. The *Meis* genes code for Meis homeodomain proteins. As individual proteins, the homeodomain proteins, like Meis, bind generally to small DNA sites, such as TAAT, and do so with poor specificity and affinity (Moens and Prince, 2002; Waskiewicz et al., 2001). When these transcription factors combine to form a complex on DNA with cofactors, like Pbx and Hox, specificity and stability are added (Moens and Prince, 2002; Waskiewicz et al., 2001). The *Meis* genes are vertebrate homologs of the *Drosophila homothorax (hth)* gene (Waskiewicz et al., 2001). The vertebrate *Meis* gene was discovered through the study of the mouse myeloid ecotropic leukemia virus (Moskow et al., 1995). It was noted that the mouse leukemia virus integrated into a specific site that was later identified as a gene, because of this, the gene received the name myeloid ecotropic leukemia virus integration site (*Meis*) (Moskow et al., 1995). There are four known *Meis* gene paralogs in mice, *Meis1*, *Meis2*, *Meis3* and *Prep1* (Waskiewicz et al., 2001). While there are multiple *Meis* genes within all known vertebrates, the zebrafish, and other members of the teleost infraclass, underwent a genome duplication event following the divergence between the lobe-finned and ray-finned fish (Amores et al., 1998). This divergence also occurred before the teleost radiation from the ray-finned fish (Amores et al., 1998). This led to ohnologs of *meis1* and *meis2* that are currently called *meis1a*, *meis1b*, *meis2a* and *meis2b*. These four, with *meis3*, are the *meis* genes found in zebrafish. The Meis proteins work in conjunction with Hox and Pbx protein products and are a member of the Three Amino Acid Loop Extension or TALE class of homeobox genes (Bumsted-O'Brien et al., 2007; Choe et al.,

2002; Hisa et al., 2004; Melvin et al., 2013; Moens and Prince, 2002; Moens and Selleri, 2006; Thorsteinsdottir et al., 2001; Vlachakis et al., 2001). Meis proteins are known to function in a number of ways during vertebrate development. One way includes forming a complex with Pbx and Hox that acts as a protein cofactor to stabilize DNA transcription by binding *cis* regulatory elements associated with the target gene (Cvejic et al., 2011; Moens and Prince, 2002; Waskiewicz et al., 2001). This complex shows differential affinity for TGATTGAC where Meis shows binding affinity with TGAC 3', and Pbx shows binding affinity with 5' TGAT (Knoepfler et al., 1999). Pbx was also first identified in relation to a leukemia study and is a TALE homeodomain protein, like its working partner Meis (Bumsted-O'Brien et al., 2007; Hisa et al., 2004; LeBrun, 2003; Melvin et al., 2013).

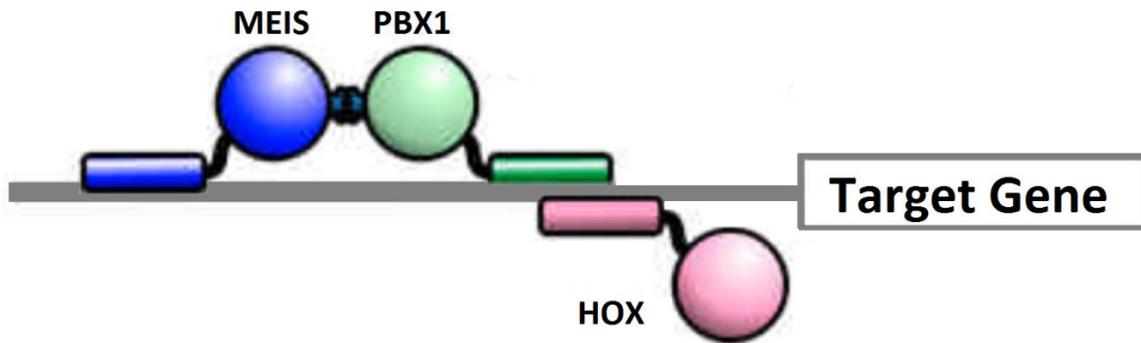


Fig. 7. A pictorial representation of Meis and Pbx working together with Hox as transcription factors that bind to specific DNA sequences. The blue Meis protein is attached to both the grey strand of DNA upstream of the gene of interest and the teal Pbx protein. The teal Pbx protein is also attached to the grey DNA upstream of the gene of interest, and the pink Hox protein completes the protein trimer through its attachments to the DNA and the Meis/Pbx complex, though not clearly depicted here due to oversimplification (Modified from Fig. 2A; LeBrun, 2003).

The Pbx/Meis complex is known to work with the homeodomain protein Hox, as seen in Fig. 7 (Aamar and Frank, 2004; Choe et al., 2002; Cvejic et al., 2011; Hisa et al., 2004; Knoepfler et al., 1999; Moens and Prince, 2002; Vlachakis et al., 2001; Waskiewicz et al., 2001). As mentioned above, Hox plays a key role in development along the head to tail axis, and disruptions of Hox can cause homeotic transformations (Amores et al., 1998; Lemons and McGinnis, 2006; Noordermeer et al., 2011). Not all Hox proteins are consistently expressed throughout the entire developing embryo; instead a *Hox* gene may be expressed only in one region, while a closely linked but separate *Hox* gene will be expressed a little later in development, in a slightly different place (Amores et al., 1998; Lemons and McGinnis, 2006; Noordermeer et al., 2011). Hox proteins function by binding to *cis* regulatory elements associated with target genes and regulating their transcription. As individual proteins, Hox proteins bind generally to small DNA sites, such as TAAT, but do so with poor specificity and affinity (Amores et al., 1998; Moens and Prince, 2002; Waskiewicz et al., 2001). When Hox transcription factors combine with other transcription factors, like Meis, to form a complex, there is an increase in specificity and stability (Moens and Prince, 2002; Waskiewicz et al., 2001). The *Hox* genes are also known for their relationship between the number of digits that develop on a limb and the quantity of *Hox* genes in the genome (Sheth et al., 2012). This observation has been supported by noting the number of digits in ray-finned fish fins and the respective number of *Hox* genes found in these teleosts versus the number of digits on a land vertebrate and the *Hox* genes they have (Amores et al., 1998; Sheth et al., 2012). *Hox* genes are arranged into clusters that have been

shown to share enhancer sequences despite sometimes being expressed at different times. For example 3' genes in a cluster typically are expressed before 5' genes within the cluster despite these genes sometimes sharing *cis* regulatory elements (Duboule, 1998; Noordermeer et al., 2011). It is believed this clustering of these genes is due to their shared enhancer sequences (Duboule, 1998; Noordermeer et al., 2011).

Another important regulatory gene in embryonic development is *MyoD*; the MyoD protein helps to orchestrate muscle-cell specific genes during differentiation (Knoepfler et al., 1999). If Pbx is bound to Hox, it will not bind to MyoD; however if Pbx is bound to Meis, it may also bind to MyoD (Knoepfler et al., 1999). This makes Meis an important protein for the Pbx complex to work with MyoD to bind to DNA, and this complex has been found to be important for *in vivo* function (Knoepfler et al., 1999). It is these differences in binding preferences that allow for the regulation of different genes and thus the differentiation of different types of cells within the multicellular system.

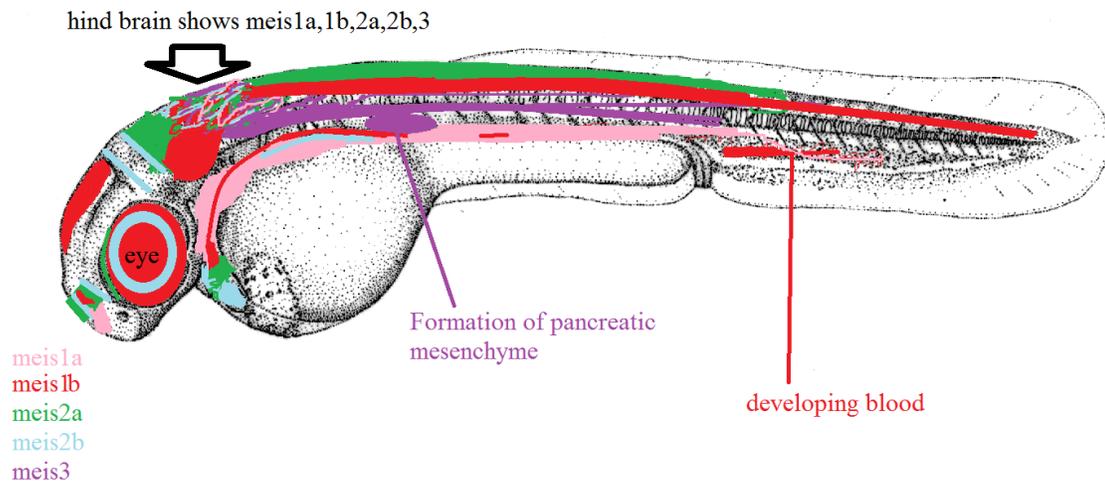


Fig. 8. Expression patterns of *meis* superimposed on a 48 hour post fertilization zebrafish embryo. The pink represents *meis1a* expression seen in the nose, lateral line, gut and hindbrain. The red represents *meis1b* found in the eye, olfactory bulb, midbrain, hindbrain, gut and extending distally from the hindbrain to the end of the tail. While *meis2a*, in green, is in the olfactory bulb, midbrain near the eye, hindbrain, heart, and expressed weakly in the trunk. The blue representing *meis2b* is seen in the nose, eyes, midbrain, hindbrain, heart, and in the developing gut. The purple represents *meis3*, found in the hindbrain and the developing pancreatic mesenchyme. Data expressed was based on in situ hybridization information found on zfin.org for each *meis* gene. (Original drawing modified from Kimmel et al., 1995; Fig. made from expression patterns noted in DiIorio et al., 2007; Melvin et al., 2013; Noël et al., 2008; Santos et al., 2010; Thisse and Thisse, 2004, 2005; Wilfinger et al., 2013; Zerucha and Prince, 2001)

While each member of *Meis* works with several of the same genes, there are some variations and these cause a unique expression pattern for each *Meis* gene (Fig. 8). In the mouse, *Meis1* works with *HoxA9*, *HoxB3*, *Pbx1* and *Prep1* and it is important for proper blood, heart, muscle and eye development (Berkes et al., 2004; Cvejic et al., 2011; Hisa et al., 2004; Mahmoud et al., 2013; Melvin et al., 2013; Thorsteinsdottir et al., 2001). In

addition to its role in heart development, *Meis1* has also shown a role in mouse heart regeneration; this trait is particularly successful in helping the heart regenerate within the first few days of life (Mahmoud et al., 2013). In zebrafish, *meis1* is expressed in facial cartilage, the hindbrain, and hematopoiesis sites along the developing embryo, seen in Fig. 8 (Cvejic et al., 2011; Melvin et al., 2013). If *meis1* is knocked down in zebrafish embryos, the lumen of blood vessels does not develop properly (Cvejic et al., 2011). Similarly, in *Meis1* deficient mice embryos, the smaller capillaries do not develop properly, leading to significant hemorrhaging causing the embryos to not be able to survive to birth (Hisa et al., 2004). In contrast, overexpression of both *Meis1* and *HoxA9* causes Leukemia in mice; this has been shown to have the same effect in humans (Hisa et al., 2004). The similarity in function and expression patterns seen within species could be due to the highly conserved amino acid sequence of *Meis1*, which is responsible for protein form and function (Cvejic et al., 2011). A lack of *Meis1* causes abnormal eye development in zebrafish, mice, and chickens (Erickson et al., 2010; Hisa et al., 2004). Specifically in mice, the absence of *Meis1* causes the retina to partially duplicate, and produces a lens smaller than wild type embryos (Hisa et al., 2004). In zebrafish, there is a loss of temporal identity within the developing retina (Erickson et al., 2010). Similarly, the *Meis* homolog found in the fruit fly, *hth*, is also involved in eye development (Erickson et al., 2010).

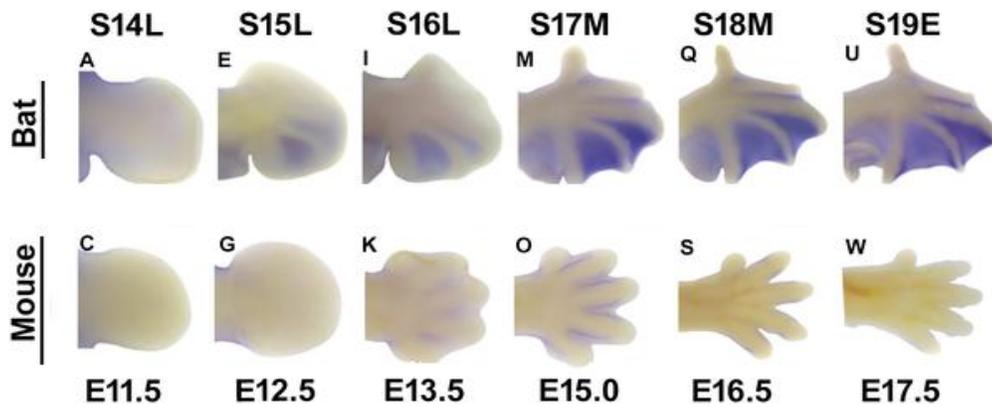


Fig. 9. *Meis2* expression in the forelimbs of embryonic bats and mice. The top row shows bat forelimbs at different stages of development, where the left side is earlier in development and the right side is later in development. As the bat wing develops, *Meis2* intensifies in the interdigit region of the limb. The bottom row highlights the differences seen in the expression of *Meis2* in the mouse forelimb. Again the left side is earlier in development than the right side and expression is observable in the interdigit region of the limb. Unlike the bat, the mouse *Meis2* does not continue to intensify as development progresses; *Meis2* expression is reduced after embryonic day 15, when the digits become more defined. All images are the dorsal view with anterior pointing up (modified from Fig. 3, Dai et al., 2014).

Zebrafish *meis2a* and *meis2b* expression is also observed in the eye. *Meis2* expression across multiple species seems to be active in many parts of the developing embryo and adult animals. It is active in the hindbrain and during the formation of hindbrain identity of the zebrafish (Moens and Prince, 2002). *Meis2* expression is tightly regulated in the eye, forebrain, midbrain, hindbrain, in the developing spinal cord, somites, limbs and hearts of mice (Bumsted-O'Brien et al., 2007; Cecconi et al., 1997; Machon et al., 2015; Oulad-Abdelghani et al., 1997). *Meis2* expression in the retina is not limited to the developing mouse embryo, but has been found in both developing and adult mice and humans (Bumsted-

O'Brien et al., 2007). In the chicken, *Meis2* is crucial to the correct development of the limbs and retina differentiation (Capdevila et al., 1999; Heine et al., 2009).

Many of the animal models mentioned above show cross species similarities. These similarities suggest *Meis2* expression is likely to be found in similar places in animals not used as model organisms, like humans. Specifically, *Meis2* may play a similar role in development and adult maintenance. The similarities of *Meis2* expression are further highlighted when comparing its expression in the developing limb of mice and bats, seen in Fig. 9 (Dai et al., 2014). While bats and mice are both mammals with fore and hind limbs, use of the limbs differs; bats fly and mice walk. Despite this difference, *Meis2* is expressed in mouse and bat limbs in comparably the same place, the proximal limb bud and interdigital tissue (Fig. 9); the only difference is in the levels of intensity of *Meis2* expression (Dai et al., 2014). In addition to the role in limb formation, *meis2b*, found in zebrafish, is expressed in the heart (Paige et al., 2012). This is interesting because it was *Meis1*, not *meis2b*, which is known to work with muscle specific genes.

Meis3 continues the comparative trend in that it is found to be expressed in many of the same places between species. As described for the previously discussed *Meis* genes, the *Meis3* protein product is also found in the hindbrain (Aamar and Frank, 2004; Choe et al., 2002; Vlachakis et al., 2001). In both Zebrafish and *Xenopus*, *Meis3* is needed for the proper formation of the embryo's hindbrain (Aamar and Frank, 2004; Choe et al., 2002). In knockdown *Xenopus meis3* experiments, the hindbrain was lost and the forebrain expanded

into what should have been the hindbrain (Aamar and Frank, 2004). In addition to the developing hindbrain, *meis3* is also active in the developing pancreas of the zebrafish and mouse (Liu et al., 2010; Manfroid et al., 2007). There is a relative upregulation of *meis3* found in the developing pancreas of mice, specifically the β -cell found within that organ, which seems to depend on *meis3* for their survival (Liu et al., 2010). In addition, *meis3* zebrafish knockdown experiments show a partial loss of the developing exocrine tissue (Manfroid et al., 2007). Interestingly, Meis3 works in conjunction with another anterior posterior regulator protein, Wnt, in both the hindbrain and formation of the pancreas (Aamar and Frank, 2004; Elkouby et al., 2010).

Despite the differences between each *Meis* gene's expression profile and functionality, it is easy to see they have some shared developmental roles, seen in Fig. 9. On a structural level, all Meis proteins have homeodomains and interact with other proteins. On a functional level, they all attach to DNA and regulate genes during embryogenesis assisting in tissue differentiation and formation of specific cells found within the vertebrate body. Across species, the coding region is highly conserved (Irimia et al., 2011). The number of introns the Meis homologous sequences have seem to be about the same, ten or eleven, across metazoans, from flies to humans (Irimia et al., 2011). This all suggests *Meis* came from one common ancestor and is very important to animal survival.

While the expression patterns of the *Meis* genes have been fairly well-characterized, nothing is currently known about how the expression of these genes is directed during

development. A number of putative regulatory elements that are associated with the *Meis2* gene have been identified in the Zerucha laboratory. The purpose of this study was to examine the function of one of these elements, the mouse m2de2 element using zebrafish as a model organism. To test the ability of m2de2 to direct expression during development, I made an eGFP expression cassette. This expression construct allowed m2de2 to direct expression of EGFP through a minimal promoter in transgenic zebrafish. Zebrafish are well suited to the study of vertebrate development because their eggs are transparent allowing for the observer to study the developing embryo (Kawakami, 2007).

Materials and Methods

Zebrafish Care

All Zebrafish were housed and handled in accordance with the Institutional Animal Care and Use Committee (ICAUUC) and maintained based on *The Zebrafish Book: a guide for the laboratory use of zebrafish (Danio rerio)* (Westerfield, 2000). Adults were housed in 1L adult tanks with a 14hrs light/ 10hrs dark cycle. The pH was kept between 7.2 & 8 and conductivity between 400 & 600 milisiemens with daily monitoring. Temperature was kept at 27° C. Adults were fed Zeigler adult zebrafish complete diet (Zeigler, Pennsylvania) and live brine shrimp once a day. To raise fish, after the embryos hatch they were placed in a small bowl with Danieau buffer solution (50x adjusted to 1L in RO water: 2.9 M NaCl, 35 mM KCl, 20 mM MgSO₄, 30 mM Ca(NO₃)₂, 250 mM HEPES pH 7.600), kept in an incubator (27° C) and were not fed for the first five days. At 5 days post fertilization (dpf), the fry were placed into a standard adult tank filled with Danieau buffer solution. These young were fed dry ZM (ZM Fish Food, Winchester, UK) food specific to size and stage of development twice a day. The type of food given was dependent on the size of the fish. As the rate of growth varies between tanks and even between siblings, the fish were checked at least once a week to ensure they were receiving the correct size food for their size. The ZM food system by ZM systems (Hampshire, UK) provides several different particle sizes of

food for zebrafish; our lab uses ZM-100, 200, 300, & 400. As the numbers increase the particle size also increases. At 20 dpf, they were slowly transitioned, drop wise, from Danieau buffer solution to the system water and continue to be fed increasingly larger food. When they were large enough to eat ZM200, they were also fed live brine shrimp. After ZM400 the fish were transitioned to Zeigler Adult Zebrafish Complete Diet and continue to receive live brine shrimp.

Isolation of HCNE MM.m2de2 from TOPO

The mouse m2de2 (MM.m2de2) element (Fig. 10) was originally isolated from mice genomic DNA using PCR and subcloned into the PCR®2.1 TOPO® Vector (Nelson, 2011). The GFP expression cassettes were constructed using the Gateway Tol2kit (Kwan et al., 2007), as described by Fisher et al. (2006). The MM.m2de2 element was PCR amplified from the TOPO vector using the primers attB1 5' and attB2 3' (Fig. 10 & Table 1). The PCR conditions used were as follows: 35 cycles, each cycle consisting of a melt at 95.0° C for 30 seconds, annealing at 58° C for 30 seconds, extension at 72° C for 30 seconds, post-cycle completion at 72° C for 10 minutes, and cycle completion held at 4° C until retrieval. QIAquick PCR purification kit (Qiagen, Valencia, CA) was then used to purify the amplified DNA. All maps (Fig. 14, 17, 19) were synthesized with Vector NTI (Thermo Fisher Scientific, Waltham, MA).

Table 1. The attB Primers

Oligo Name	Sequence
5'-attB1-TOPO	GGGGACAAGTTTGTACAAAAAAGCAGGCCCT /GAGCTCGGATCCACTAGTAAC
5'-attB2-TOPO	GGGACCACTTTGTACAAGAAAGCTGGGT/GAGCTCGGATC CACTAGTAAC
3'-attB2-TOPO	GGGGACCACTTTGTACAAGAAAGCTGGGT /TCACTATAGGGCGAATTGGG
3'-attB1-TOPO	GGGGACAAGTTTGTACAAAAAAGCAGGCT/TCACTATAGG GCGAATTGGG

```

001 GAAGGATCAGCTTGGTACCAAATTA TG TCACTGCATTTCTGCTTTT CAGTTG TACACTTT
061 CAATATCAGCTCTTTTTTATACATTTATTACCCTAAACCCCAAACTCTGTGACAGT
121 CTTAGAATGAGAAAAACAATGACCAGGGAGCTCATCTCTTGCCCTCTCATTGTGTTTGTGC
181 ACCTTCACCCCAAGGGCTGAGGCCACAGGCTTTGTTTGGTAGTTACTTATCACTAAAGGCA
241 AATCAAATAGT CAGCTG TAGTATCTGA CACTTG CAGAGCAAGTATAACATGCCAAGCATC
301 TCCTAAGTGCTTGGTTTATTGTGGCTAGTTCATTGTCCAGG TAAT GGTATAAGGAAAGGA
361 GAAGTTTTATATCCTTGGAGGTACTGTGACCAGGACCTGAACACTGAAGAGATCTGCCAA
421 AGTCCCAGTAAGAGCTGATGACACAGCGTCTGAGGCACCCTCTAACAGCTATGCCCAAA
481 GAGCCTCTGTATT CAGGGCAGTATTGGGTTACACAGTTTGATGTTGTGAGAGTTGGTGG
541 TTAAGTAGTTG TAAT GAATCAGAAAGCTTGTAAGTGGATACTCTAGTTAGC TAAT GTCAT
601 TGCCTAGAAGCAAACCTTTCAGATTTTCTAGTATACCCAAAAAGGAAGTCCCTCTTCTTTT
661 TTTAACCCTCCGATAGAGCTGTG TAAT TCAGTTACCTATGTCACCAGCACTATATGTC
721 TTTATAAAAAGGAGTCTTAACATATACGTGCTTAACCTTTACTTCT TAAT GATGAGTGCT
781 TAAATTGTTATTGTGCTGCCA TAATTA GGTTTAAGGCCTAAAATGCTGTGTGTATAGC
841 TATAAAAAGTAAG TAAT GTGACCTAGGACTGACAGTA CAGCTG ATCGAAGGGAAGAAAA
901 TCCTATGGAAGGCTGAAATTGAACATGGTAACTGCAATTATCTAAAAGGAAGTCTATTAC
961 CTTCTGATAAACTGTCAAAA TAAT GATTTGGTACTCCACAAAATTGTT CAGAGAATCAA
1021 TAAAGCCAATATTTTCGTCACAGTTTGAGAATTTAG GAATTC CACTTTCCTAAGTCACA
1081 AGAGAATGC TAAT ATAGTTTAAGAATGAGATAAT TTTACTGTTTTGTTCTGTTTCATCT
1141 TGTAACCTCT ATTA GAGTAAATATATTTGCTTGGGAAACCTGATGAACACCATTCTTGGA
1201 AAGCATAAATAGTATTTATATT CAGTAGTTCTGCACAGGATTTATACAAATATACTACTT
1261 TGCTTTGCAGGAATACAAAATGCACGATTTTGAGTAGAGTATAGATAAT TCCTACTGCAG
1321 TCACTGCTTTTCTTAGGGTCG

```

Fig. 10. Nucleotide sequence of MM.m2de2. This method Fig. depicts the second downstream element associated with *Meis* isolated from mouse. This specific sequence was used in this research in association with the minimal promoter *cfos* and *eGFP*. The red letters represent the PCR primer sequences used to amplify this element while the sequences highlighted in purple symbolize TAAT binding sites, an internal EcoRI site is highlighted in yellow while the E-box sites are highlighted in blue.

Construction of the Middle Entry Vector

The gateway system consists of two reactions, BP and LR (Alberti et al., 2007; Fisher et al., 2006). MM.m2de2 was amplified with attB sites flanking it (Fig. 11). The donor vector pDONR221 contains attP sites that recombine with the elements' attB sites during the BP

recombination reaction (Fisher et al., 2006). The name BP refers to the attB site on the element attaching to the attP site on pDONR 221(Fisher et al., 2006).

The BP reaction was set up with 75 ng pDonR221, 25 fmol clean PCR product, 1 µl BP clonase, and raised to a final volume of 5 µl with TE.

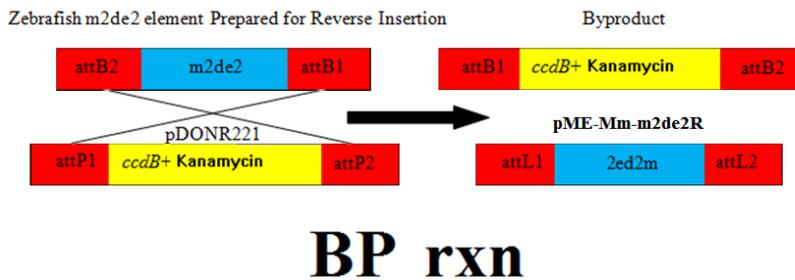


Fig. 11. Pictorial representation of the Gateway® BP reaction. During this reaction, the mouse m2de2 element (the blue box) which is flanked by the initial attB2 and attB1 cloning sites (indicated by red boxes) recombined with pDONR221 forming attL sites. The attB1 recognizes the attP1 (also red box) site while the attB2 recognizes the attP2 through the use of BP Clonase™ II. Because of this, m2de2 located in between the attB2 and attB1 sites is then translocated into the pDONR221 vector. This generates the middle entry vector with the element m2de2 inserted in reverse orientation (indicated by 2ed2m).

An eppendorf tube filled with the BP reaction was vortexed 2 sec, spun in mini centrifuge for 2 sec and then left on bench top overnight or approximately 16 hrs. The following day the BP reaction was stopped with the addition of 0.5 µl proteinase K, which degrades the clonase II enzyme, and the reaction was incubated at 37° C for 10 min. Approximately half of the BP reaction (2.5 µl) was transformed into competent *Escherichia*

coli by heat shock. The BP reaction DNA (2.5 μ l) was added to 50 μ l of competent cells (DH5 α or Top10). This was slowly and carefully pipetted up and down to gently mix, then placed on ice for 30 min; after which, it was heat shocked at 42° C for 30 sec, then incubated on ice for 2 min. To this, 1 ml SOC media was added the suspension was incubated at 37° C for 90 min with shaking. The bacteria were plated onto LB/Kanamycin plates (kanamycin 50 mg/ml) at a low concentration (50 μ l of transformation) and a high concentration (remainder of transformation concentrated by a quick spin in mini centrifuge).

Construction of the Transgenic Reporter Construct

The second part of the gateway cloning process completed was the LR reaction. This starts with the middle entry vector which was bordered by attL sites and was transformed into the destination vector pGW-cfos-eGFP; the destination vector has attR re-combination sites upstream of the eGFP gene (Fisher et al., 2006). The LR reaction was a recombination event, in which the m2de2 element was transferred from the middle entry vector to the pGW-cfos vector through recombination of the middle entry vector's attL sites and destination vector's attR sites (Alberti et al., 2007; Fisher et al., 2006; Kwan et al., 2007). As seen in Fig. 12, after the LR reaction, the element was in a cassette with a minimal *cfos* promoter and eGFP (a green florescent protein sequence more stable than GFP) and flanked by Tol2 sites (Fisher et al., 2006; Kawakami, 2007).

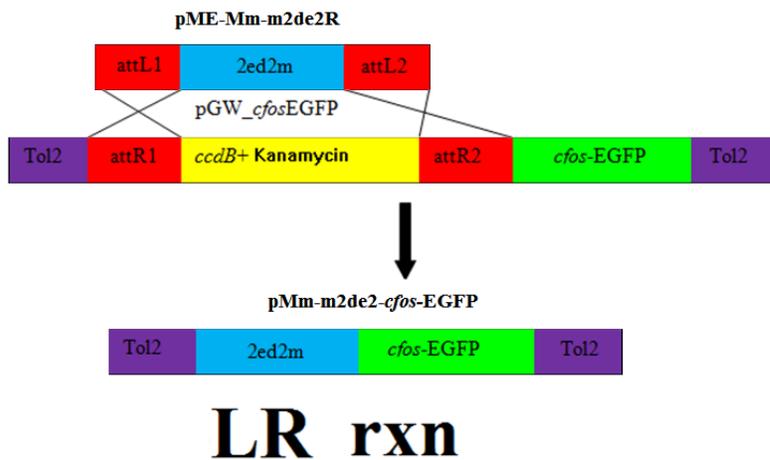


Fig. 12. Pictorial representation of the Gateway LR reaction. The LR reaction uses the middle entry vector generated by the BP reaction to recombine with the destination vector pGW-cfos-eGFP. During this event, the att sites recombine to insert the m2de2 sequence in place of the kanamycin resistance gene. The final product shows Tol2 sites flanking the m2de2 and cfos-eGFP sequences.

The amount of BP product DNA used for the LR reaction was calculated using the following formula:

$$(10\text{fmol})(\text{size of BP insert})(660\text{fg/fmol})(1\mu\text{g}/10^6\text{fg}) = \underline{\mathbf{B}}\text{ng}$$

$$(\text{BP dilute concentration})(\underline{\mathbf{Y}} \mu\text{l}) = \underline{\mathbf{B}}\text{ng}$$

$\underline{\mathbf{Y}}\mu\text{l}$ BP product used

$(5\mu\text{l} - \mathbf{Y} - \mathbf{X} - 1)$ TE pH8 buffer

An eppendorf tube was filled with 58.39 ng of pGW-cfos-eGFP, 1 μl of LR Clonase II, 10.01 ng of BP product were added using the LR calculation above, and the reaction brought to 5 μl with TE pH8 buffer. The solution was then vortexed, briefly spun, and left on bench top

overnight (approximately 16 hours). The following day the LR reaction was stopped with the addition of 0.5 μ l proteinase K, and then incubated 10 min at 37° C. Then 2 μ l of the LR reaction was added to X μ l Top10 cells (X from previous page equation). These cells were then incubated on ice for 30 min, followed by a heat shock at 42° C for 30 sec, kept on ice to add 250 μ l of room temp SOC medium, then incubated at 37° C for 90 min with shaking. Cells were then plated on LB/ampicillin plates overnight. Two plates were made, a low concentration plate (50 μ l from transformation onto LB/ampicillin plate), and a high concentration (remainder of transformation concentrated by a quick spin in mini centrifuge).

Confirmation Sequencing

Constructs generated from the BP and LR reactions were sequenced to confirm the identity of the insert (Cornell University Biotechnology Resource Center). This was done using T7 and M17 primers to confirm the middle entry construct resulting from the BP reaction; 5' and 3' gateway specific primers (Table 2) to confirm the destination construct resulting from the LR reaction.

Table 2. Oligos used to confirm middle entry and destination construct sequence

Oligo Name	Sequence
M17	GTAAAACGACGGCCAG
T7	TAATACGACTCACTATAGGG
Sense: 5'- Gateway seq	GCAATCCTGCAGTGCTGAAA
Antisense: 3'- Gateway seq	GGACTTCCTACGTCCTGGA

Transcribing Transposase mRNA

Transposase mRNA was transcribed with the use of mMessage mMachine kit (Ambion, life technologies, Grand Island, NY) using the recommended protocol and then precipitated with Lithium chloride. This transposase mRNA was resuspended in DEPC water and coinjected with the expression construct into zebrafish embryos where it was translated into a transposase protein that catalyzes insertion of the expression cassette into the host genome via the Tol2 sites (Kawakami et al., 2004; Kwan et al., 2007). A pictorial representation of this mechanism can be seen in Fig. 13.

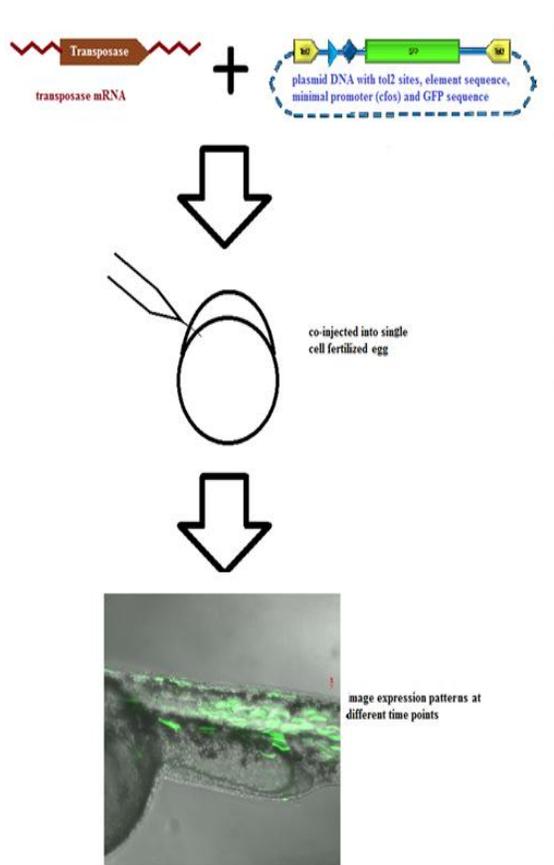


Fig. 13. Pictorial representation of the process of microinjection into zebrafish embryos. Transposase mRNA was combined with the MM.m2de2-pGW-*cfos*-eGFP featuring the minimal promoter *cfos* and the reporter gene eGFP. This solution was then microinjected into one cell zebrafish embryos. The embryos were cared for and photographed under a LSM confocal microscope at different time points to observe expression.

Generating Transgenic Embryos through Use of Microinjections

Freshly laid zebrafish eggs were collected following mating and washed with RO water. They were then added to a beaker with 170 ml system water and 100 µl Bleach/RO stock (5.25% Bleach). Eggs were swirled in this solution for exactly two minutes and then transferred into another beaker with 170 ml system water only. Eggs were swirled in this solution for exactly two minutes and then transferred into a third beaker containing 170 ml system water only. Eggs were swirled in this solution for exactly two minutes and then transferred into a petri dish. The excess water was removed and replaced with Danieau buffer solution (50x adjusted to 1L in RO water: 2.9 M NaCl, 35 mM KCl, 20 mM MgSO₄, 30 mM Ca(NO₃)₂, 250 mM HEPES pH 7.600).

While this was occurring, the solution used in the injections was thawed and combined. Specifically 175 ng transposase mRNA, 2 µl of Phenol red (0.5% in Dulbecco's Phosphate Buffered Saline) and 125 ng plasmid DNA was brought to a final volume of 5 µl with DEPC water. Additionally, the needle used for the injections was pulled from a 3.5 nl capillary tube that was baked prior to pulling at 260°C in order to inactivate any RNases. To pull the capillary tube into a needle a David Kopf Instruments Vertical Pipette Puller (Model 700C) was used with the heat set kept at 54 and the solenoid fixed at 10. Once gravity divided the needle, the tip was beveled using watchmaker forceps (size 5), it was filled with mineral oil and placed on the Nanoliter 2000 Microinjector (World Precision Instruments Model B203XVY) attached to a Marhauser MMJR Micromanipulator (World Precision

Instruments). Using the Microinjector controls, between one fourth and one fifth of the mineral oil was pushed out of the beveled end of the needle and replaced with the injection solution.

Approximately 50 one cell embryos were then placed against a 1.0 millimeter thick VWR micro slide (VWR International 48300-025) which had been taped to the outside bottom of a plastic Petri dish. Single celled zebrafish embryos were injected with 4 nl of transposase mRNA, the destination vector plasmid carrying MM.m2de2-pGW-*cfos*-eGFP and phenol red into the yolk sack preferably directly beneath the developing animal pole.

Immobilization of Transgenic Zebrafish Embryos for Imaging

To image, embryos were embedded in agarose after being anesthetized in a dilute tricaine mixture. Specifically, hatched larvae or dechorionated embryos were anesthetized in 1 mL of 0.8% tricaine/danieau buffer. After 10 minutes, the embryos were gently tilted to confirm they were anesthetized and embryos were individually mixed with 0.1 ml of 0.8% Danieau buffer/agarose at 30° C and suspended in a deep-welled glass microscope slide for confocal imagery. Embryos were adjusted within the slide to optimal orientation and the slides set aside to allow the agarose to solidify. Once the agarose was solid, 4% tricaine/Danieau buffer was added to raise the volume to fill the well, then a cover slip was added.

Screening Transgenic Zebrafish Embryos Through Microcopy

To identify transgenic embryos for the purpose of forming a transgenic line, a non-invasive, survivable method was needed. Because eGFP was noted between 54 and 60 hpf in the injected embryos, this time point can be used to separate transgenic embryos from their non-transgenic siblings. Embryos were imaged in Pyrex 60 x 15 mm petri dishes (Corning Life Sciences, New York) containing about 6 ml of 0.3x Danieau buffer and 4 drops of 20% Tricaine in 0.3x Danieau buffer solution. This amount of Tricaine anesthetized the embryos, without killing them, allowing them to be briefly observed under the fluorescent lamp of the microscope and transferred into another dish for eGFP positive embryos. A Zeiss LSM 510 Confocal Microscope was used to view and sort injected embryos. The dishes were examined under 10x objective with FITC under a mercury lamp to confirm possible eGFP expression within the embryos.

Confocal Imaging of Zebrafish Embryos

A Zeiss LSM 510 Confocal Microscope was used to view and image immobilized embryos. Slides were individually examined under 10x objective with FITC and bright field to position the sample and confirm possible eGFP expression within the sample; the Argon laser was then used for image collection with either a single slice picture or a Z-stack. Image pixel quality was kept at 1024 X 1024, while scan speed ranged from 5 to 9. The images

were modified with Zeiss software's built-in projection tool to add a size overlay and Adobe Photoshop to adjust the green, and white levels in the whole picture, both eGFP and autofluorescence would be observed as green. This was done because the images were captured in a dark room and needed to be adjusted to be more defined in a brighter room. No one part of the image was altered over another part, but Photoshop was used to adjust brightness levels and compile a single image from many smaller frames. All whole embryo photos are composites connected through use of Photoshop.

DNA Isolation from Zebrafish Fry

Genomic DNA from zebrafish larvae was isolated to determine if the transgenic element could be detected by PCR screening following injections. Zebrafish larvae were euthanized with Tricaine (4% in Danieau buffer solution) at 5 dpf or directly following imaging and frozen. Larvae were washed three times with sterile RO water and then placed in groups of 5 into 50 μ l of activated genomic extraction buffer (10 mM Tris pH 8.2, 10 mM EDTA, 200 mM NaCl, 0.5% SDS, 200 μ g/ml Proteinase K added just before use in sterile RO water) and incubated a minimum of 3 h at 56° C with shaking. This was followed by the addition of 100 μ l of 100% ethanol (-20° C) and placed at -20° C overnight.

The following day the samples were centrifuged for 10 min at 13000 rpm. The pellet was washed with 200 μ l of 70% ethanol, vortexed, then spun again under the same settings for 2 min. The supernatant was removed and the pellet dried. The DNA was then re-suspended in

20 μ l TE+RNase buffer (10 mM Tris, 1 mM EDTA pH 8.0, 100 μ g/ml RNase added just before use in sterile RO water) and incubated at 37° C for 1 h. Samples were then treated with phenol. Specifically; 450 μ l of phenol was added to an eppendorf tube containing the genomic DNA and the solution carefully inverted 3x to mix, then the tube was spun for 15 min at 1500 x g; The aqueous phase was transferred to a clean eppendorf tube and DNA precipitated with 100% ethanol at -80° C for 30 min. The genomic DNA was pelleted by centrifugation for 15 min at 15000 x g. The supernatant was discarded and the pellet was washed with 500 μ l of 70% ethanol by inverting it 3x, and the tube was again spun for 5 min at 15000 x g. The supernatant was discarded and the pellet dried before it was resuspended in 17 μ l of TE buffer. Multiple genomic preparations from the same set of injected embryos were combined, quantified, and stored at -20° C. The presence of MM.m2de2 element was determined in this genomic DNA using PCR and the primers attB1 and attB2 (Fig. 10 & Table 1). The PCR conditions used were as follows: 35 cycles, each cycle consisting of a melt at 95.0° C for 30 seconds, annealing at 60° C for 30 seconds, extension at 72° C for 1 minute, post-cycle completion at 72° C for 10 minutes, and cycle completion held at 4° C until retrieval. This was followed with gel electrophoresis to determine the size of the fragments amplified.

DNA Isolation from Fin Clip

This procedure was performed in order to isolate genetic material from individual adult fish for the purpose of genotyping and confirming the presence of the MM.m2de2-pGW-*cfos*-eGFP. A small amount of tissue was clipped from the end of the caudal fin in order to extract DNA which was used for further analysis such as PCR. If done correctly, the caudal fin regenerates within two weeks and it has been previously shown that female fish regenerate quicker than male fish (Nachtrab et al., 2011). Fish were anesthetized with Tricaine (150 mg/ml) prior to clipping. Once the fish was unresponsive but gill flaps were still moving, sterile tweezers and scissors were used to hold and clip the caudal fin. The tail portion removed was not greater than halfway between the tip of the fin and the point where the scales end. The entire process was designed to take less than 1 min, fish were not anesthetized for more than 5 min and the procedure did not result in bleeding. The clip was transferred to 50 μ l of activated genomic extraction buffer (10 mM Tris pH 8.2, 10 mM EDTA, 200 mM NaCl, 0.5% SDS, 200 μ g/ml Proteinase K added just before use in sterile RO water). The fish were quickly transferred into a recovery tank containing system water and a bubbler.

The gills and fish in general were observed for the following hour to ensure they continued to recover from the anesthesia and procedure. Although fish usually recovered fully within 2 min, some did not eat that day. Following the hour recovery period, the fish was transferred into another individual tank and placed back into the system. Fish were

maintained in isolation and allowed to recover for 2 weeks (or until fin was at least 50% regenerated).

The fin clip in the genomic extract buffer was incubated a minimum of 3 h at 56° C, with shaking at 100 rpm to dissolve the tissues. This was followed by the addition of 100 µl of 100% Ethanol (-20° C) and the solution was placed at -20° C overnight. The following day the fin clip was centrifuged for 10 min at 13000 rpm. The supernatant was removed, 200 µl of 70% Ethanol was added and the tube was briefly vortexed. Then the sample was spun again under the same settings for 2 min, and the supernatant was removed and the pellet dried. The DNA was then re-suspended in 20 µl TE+RNase buffer (10 mM Tris, 1 mM EDTA pH 8.0, 100 µg/ml RNase in sterile RO water) and incubated at 37° C for 1 h.

The DNA was then treated with phenol. Specifically, 450 µl of phenol was added to the eppendorf tube under the hood. The solution was carefully inverted 3x to mix, then centrifuged for 15 min at 1500 x g. The supernatant was transferred to a new tube and DNA precipitated with 100% ethanol at -80° C for 30 min. The genomic DNA was centrifuged for 15 min at 15000 x g. The supernatant was then poured off and the pellet was washed with 500 µl of 70% ethanol by inverting it 3x and the DNA centrifuged for 5 min at 15000 x g. The supernatant was removed and the pellet dried before it was resuspended in 17 µl of TE buffer and quantified. The MM.m2de2 element was amplified from this genomic DNA using PCR and the primers attB1 and attB2 (Fig. 10 & Table 1) in the same manner as previously described in *DNA isolation from 5dpf zebrafish fry*.

Results and Discussion

Isolation of HCNE MM.m2de2 from TOPO

Mouse *Meis2* downstream element 2 (MM.m2de2) was previously isolated (Nelson, 2011) and cloned into pCR2.1-TOPO (Fig. 14). It was then transformed into One Shot® TOP10 Chemically Competent *E. coli* cells and stored as a glycerol stock at -80 °C. DNA isolated from liquid cultures of this TOPO MM.m2de2 stock was digested with EcoRI to confirm size. The MM.m2de2 element is 1350 base pairs (bp) in length. An additional 200 base pairs were added from the pCR2.1-TOPO vector making it 1530 bp or 1.53 kb (kilobase pairs) when amplified with attB1 5' and attB2 3' primers (Table 1). These primers bind to sequences of DNA upstream and downstream of MM.m2de2 in pCR2.1-TOPO. This plasmid DNA was used to proceed with Gateway cloning to generate expression constructs. There is an EcoRI site within the m2de2 sequence, putative homeodomain protein binding sites (TAAT/ATTA) and E-Box binding sites (CANNTG), shown in Fig. 10. Because of the internal EcoRI site, following EcoRI digestion MM.m2de2.pCR2.1-TOPO is visualized in three pieces with gel electrophoresis (Fig. 15). In Fig. 15, the EcoRI digest of Mm-mM2de2-TOPO displays fragment 1068 bp and 295 bp in length, these sizes confirm the plasmid is the correct size.

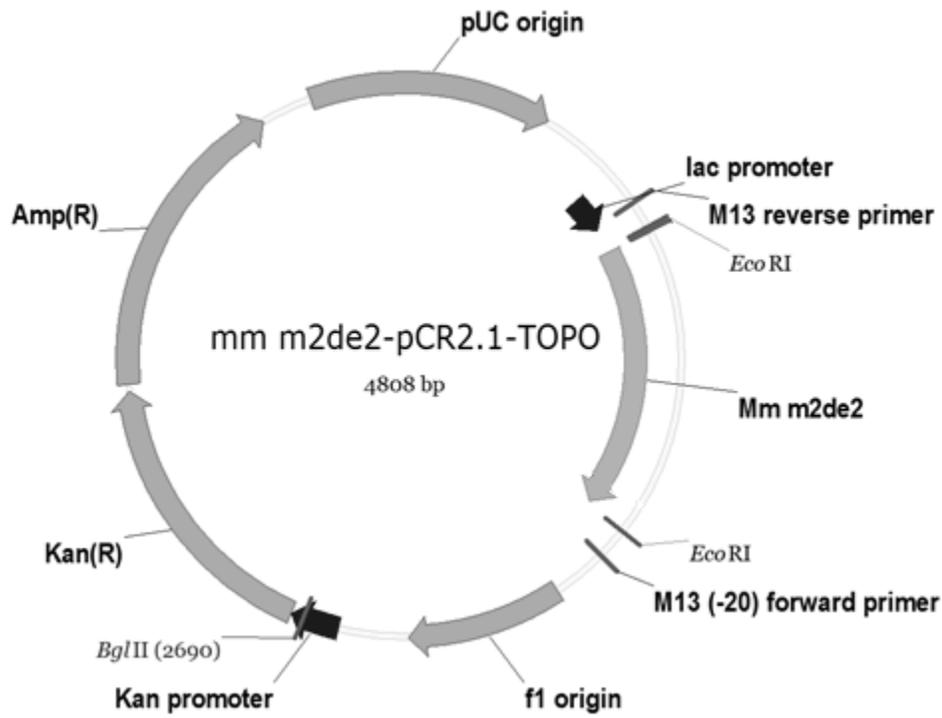


Fig. 14. Site map of MM.m2de2 within the pCR2.1-TOPO vector. The TOPO vector contains the genes for ampicillin and kanamycin resistance. The mouse *Meis2* downstream element 2 sequence is between EcoRI sites.

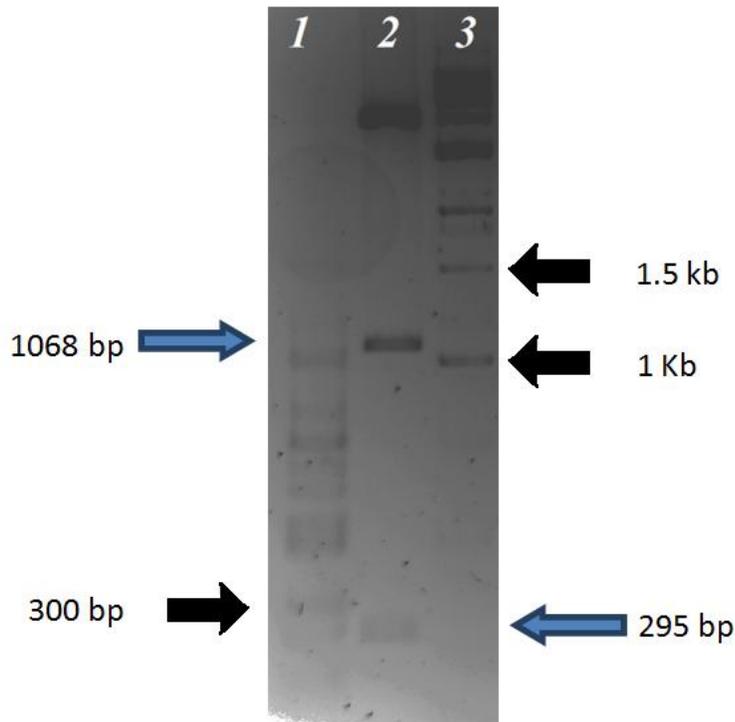


Fig. 15. MM.m2de2.pCR2.1-TOPO. Lane 1, 100 bp ladder (BioLabs N3231S) with the 300 bp band indicated by a wide arrow. Lane 2, EcoRI digest of Mm-mM2de2-TOPO with DNA sizes of 1068 bp and 295 bp indicated with thinner, longer arrows. The top band represents the backbone of TOPO. Lane 3, 1 kb ladder (BioLabs N3232S) with 1.5 kb, and 1 kb bands indicated by wide arrows. This image has been altered with photoshop to adjust exposure levels.

Gateway Cloning

As described in the methods section, under the BP reaction, m2de2 was amplified with attB sites flanking it (Fig. 11). The donor vector pDONR221 contains attP sites that

recombine with the elements' attB sites during the BP recombination reaction (Fisher et al., 2006). An eppendorf tube containing pDonR221, m2de2 PCR product, and BP clonase produced Mm-mM2de2-pDonR221. This was transformed into competent cells (DH5 α , or Top10 ThermoFisher Scientific, Waltham, MA) and grown on LB/Kanamycin plates. Plasmid DNA from these colonies was isolated, digested with EcoRI, and size was confirmed with gel electrophoreses (Fig. 16). The 1068 bp, and 295 bp bands observed on the gel are due to additional EcoRI site within m2de2. The top band represents pDonR221 backbone. The map for Mm-mM2de2-pDonR221 is seen in Fig. 17 and shows the EcoRI sites flanking and within the m2de2 sequence; additionally the attL sites important for the LR reaction are labeled.

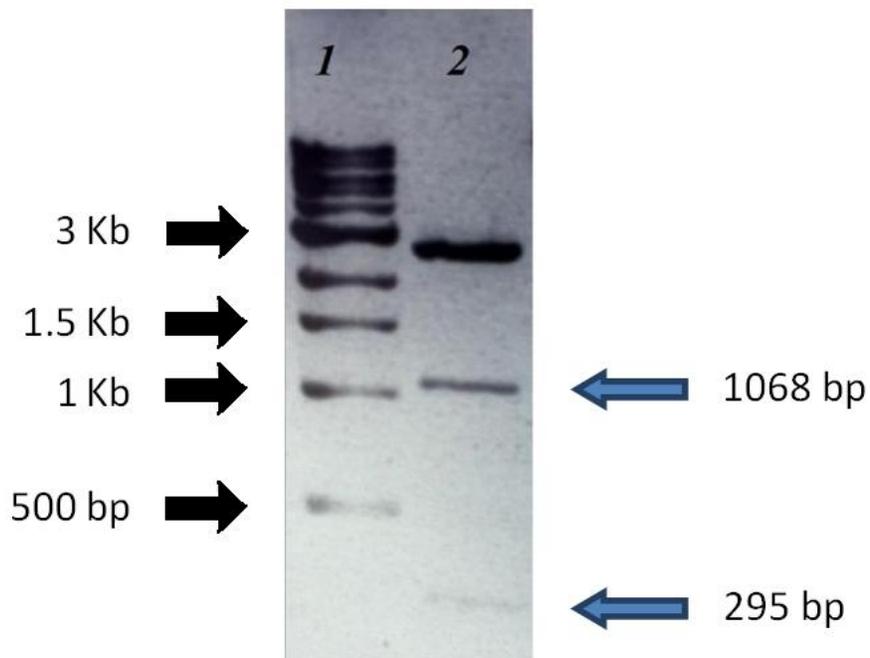


Fig. 16. Mm-mM2de2-pDonR221 test digest to confirm identity of clone.MM.m2de2 product cut with EcoRI. Lane 1, 1 kb ladder (BioLabs N3232S) with 3 kb, 1.5 kb, 1 kb and 500 bp bands indicated by arrows. Lane 2, EcoRI digest of Mm-mM2de2-pDonR221 with DNA sizes of 1068 bp and 295 bp indicated with thinner arrow. The top band of just under the 3 kb band on the ladder represents the backbone of pDonR221, 2647 bp in length.

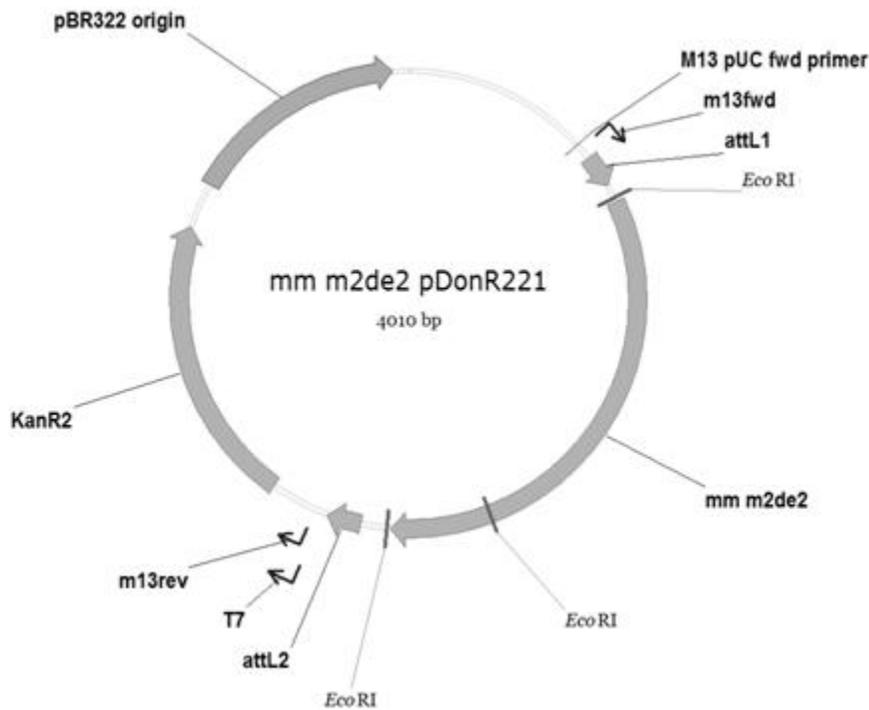


Fig. 17. Mm-mM2de2-pDonR221. The EcoRI sites are indicated. The pDonR221 vector contains the gene for kanamycin resistance. The mouse *Meis2* downstream element 2 sequence is between EcoRI sites and has an EcoRI site within it.

The second half of the gateway cloning process is the LR reaction. As described in the methods section, this starts with the middle entry vector (Mm-mM2de2-pDonR221), bordered by attL sites, transformed into the destination vector pGW-cfos-eGFP; the destination vector has attR recombination sites upstream of the eGFP gene (Fisher et al., 2006). The LR reaction was a recombination event, in which the m2de2 element was transferred from the middle entry vector to the pGW-cfos vector through recombination of

the middle entry vector's attL sites and destination vector's attR sites (Alberti et al., 2007; Fisher et al., 2006; Kwan et al., 2007). As seen in Fig. 12, after the LR reaction, the element was in a cassette with a minimal *cfos* promoter, the eGFP gene (a green fluorescent protein DNA sequence more stable than GFP) and flanked by Tol2 sites (Fisher et al., 2006; Kawakami, 2007). The MM.m2de2-pGW-*cfos-eGFP* was grown in Top10 cells on a LB/ampicillin plate. Plasmid DNA isolated from these cells was digested with EcoRI and confirmed by gel electrophoresis (Fig. 18). Because of the previously mentioned EcoRI site within m2de2, there are three bands on the LR product gel, they are expected to be 1363 bp, 295 bp and 1068 bp in length and seen in Fig. 18. The 1363 bp band is an incomplete digest and was not seen before because previous digests did not contain an incomplete digest. The map for the final product can be seen in Fig. 19 and includes the EcoRI sites flanking m2de2, as well as within it and the attL sites important for the LR reaction.

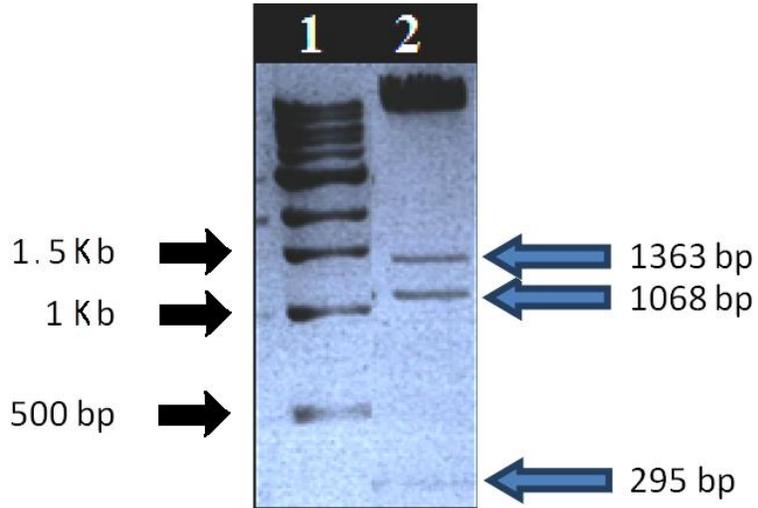


Fig. 18. MM.m2de2-pGW-*cfos-eGFP*. Test digest to confirm identity of clone. Lane 1, 1 kb ladder (New England BioLabs N3232S) with 1.5 kb, 1 kb and 500 bp bands indicated by arrows. Lane 2, EcoRI digest of MM.m2de2-pGW-*cfos-eGFP* with DNA at sizes of 1363, 1068 and 295 bp. The top band represents the backbone of pGW-*cfos-eGFP* without m2de2. This image has been altered with photoshop to adjust exposure levels.

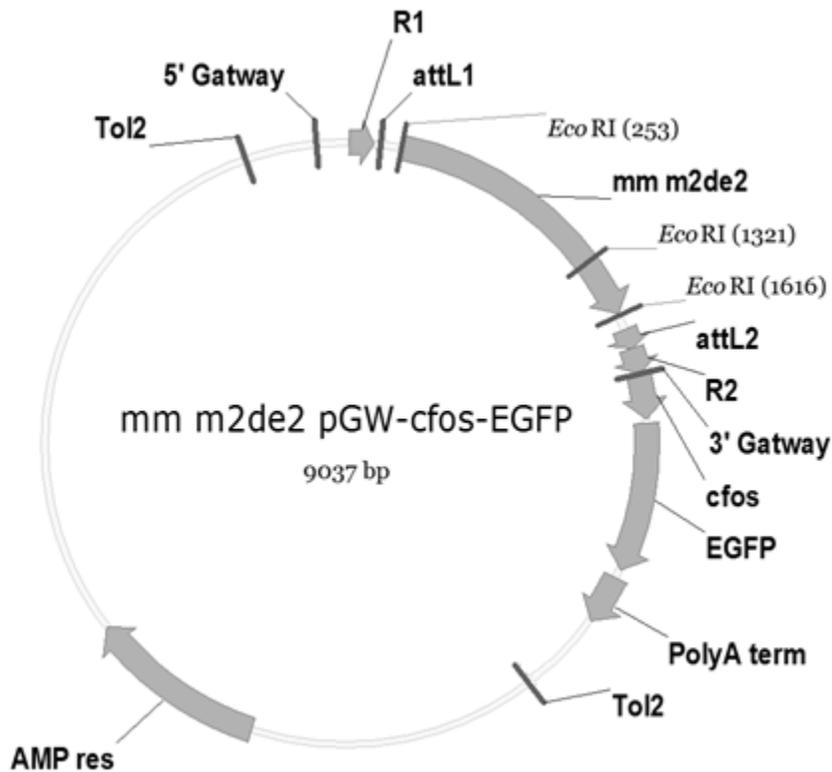


Fig. 19. MM.m2de2-pGW-*cfos*-eGFP. MM.m2de2 is upstream of the *cfos* minimal promoter and eGFP gene and this cassette is flanked by Tol2 sites. The TOPO vector contains the gene for ampicillin resistance. The mouse *Meis2* downstream element 2 sequence is between EcoRI sites and has an EcoRI site within it. The promoter for a *cfos* gene and the *eGFP* sequences are also labeled.

Both the Mm-mM2de2-pDonR221, and the final MM.m2de2-pGW-*cfos*-eGFP plasmid, LR product, were sequenced to confirm the identity of the insert (Cornell University Biotechnology Resource Center) using T7 and M17 primers and 5' and 3' gateway specific

primers (Table 2) respectively. MM.m2de2-pGW-*cfos*-eGFP was used in embryonic injections.

Overview of Transgenic Embryos MM.m2de2 Expression Patterns

Zebrafish embryos were injected at the one cell stage with transposase mRNA and MM.m2de2-pGW-*cfos*-eGFP. Injections were into the yolk and the mRNA and plasmid DNA were subsequently taken up into the cells of the developing embryo. Upon translation of the transposase mRNA, the transposase protein recognizes the Tol2 sites of the expression construct and inserts the expression cassette into the host genome. The whole clutch of approximately 200 embryos was injected, of these between 80 to 150 would survive the injection process. Of these, 10 to 60 embryos would then express eGFP. Clutch size, survivability, and incorporation rates varied widely; a few times clutch size was under 20 embryos and once all embryos that survived to imaging expressed eGFP. At approximately 48 hpf, injected fish containing MM.m2de2- *cfos*-eGFP are observed to first express eGFP in the midbrain and hindbrain (Fig. 20). At earlier developmental time points to 48 hpf no expression has been observed. Expression observed in the head is similar to expression observed to be directed by m2de3 and m2de4 (Alicia Ramsaran, Laiton Steele, Tucker Munday, Zerucha unpublished data). In mice, *Meis2* has been shown to be expressed in the hindbrain (Cecconi et al., 1997; Cunningham et al., 2011). The expression directed by m2de2 to cells of the developing zebrafish brain in a pattern consistent with *Meis2* expression in the

mouse provides preliminary evidence in support of this element being able to control *Meis2* expression.

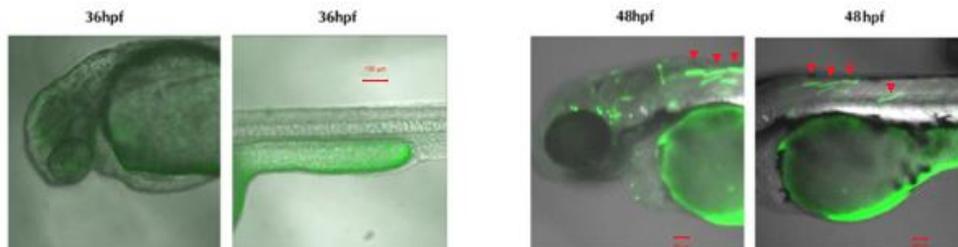


Fig. 20. Expression in head and trunk at different time points. No expression is observed at 36 hpf. Head expression (arrows) of eGFP directed by MM.m2de2 through the *cfos* promoter at 48 hpf. These images have been modified from their original form to increase contrast between background. Anterior is to the left and dorsal is to the top and size bars represent 100 μm in the 36 hpf picture and 50 μm in the 48 hpf.

As development proceeds, eGFP expression is observed to no longer be present in the brain and becomes more pronounced as long striations of expression within the somites. This expression is consistent with genes associated with muscle development (Akitake et al., 2011; Bessarab et al., 2008; Xu et al., 2003). This expression frequently continues caudally to the tail of the embryo in primary transgenic embryos and persists until 5 dpf (Fig. 21). The eGFP expression observed in primary transgenic embryos was likely mosaic, with no embryo displaying the complete expression pattern driven by m2de2. However, while any expression between individuals exhibited some variation, the general pattern of expression within somites was consistent. This is not unexpected in primary transgenic animals as there is no guarantee that every cell of the embryo contains MM.m2de2-pGW-*cfos*-eGFP in its genome. The injection solution is inserted into the yolk and not the single cell; incorporation may not

occur until after the first cell division and may not occur in both cells. A representative expression pattern can be seen in Fig. 21. This pattern was not noted in any control fish which included non-injected wild type, mock injected embryos, and embryos injected with *cfos*-eGFP without the MM.m2de2 sequence found in MM.m2de2-pGW-*cfos*-eGFP. *Meis1* is known to play a role in muscle development (Berkes et al., 2004). While muscle in the trunk of zebrafish develops before 48 hpf, the time point eGFP was observed, *MyoD*, a crucial muscle development gene, often persists for many hours after the initial formation and development of skeletal muscle (Berkes et al., 2004). Other muscle cofactors, such as *pax9*, are also active at this time point (Berkes et al., 2004; Chatterjee et al., 2011).

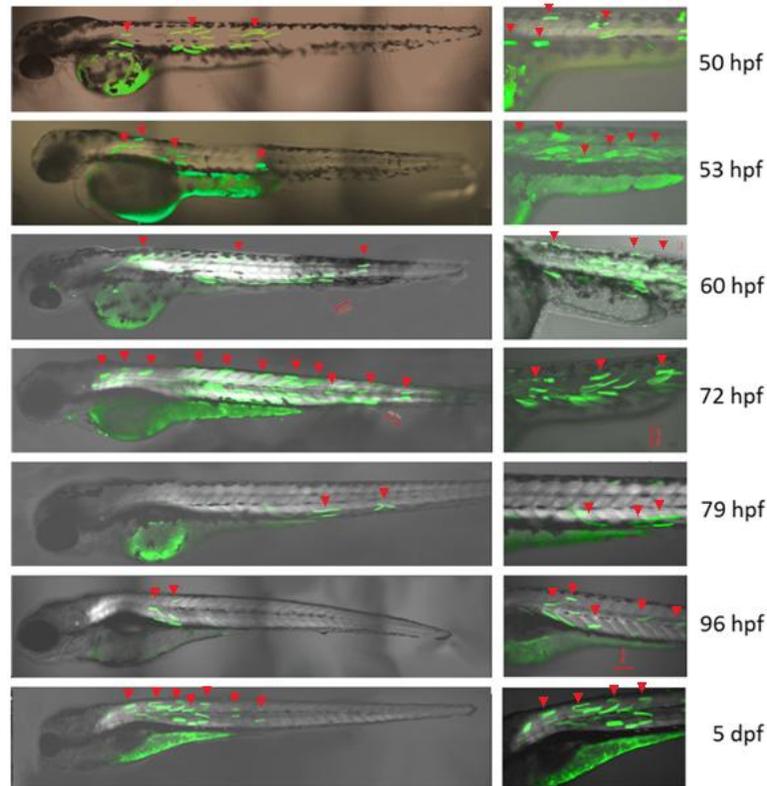


Fig. 21. MM.m2de2-pGW-*cfos*-eGFP injected embryos imaged at 50, 53, 60, 72, 79, 96 hpf and 5 dpf show eGFP in muscle fibers. This was common to this time point in general. (Left) A composite image of the whole embryo at 50, 53, 60, 72, 79, 96 hpf and 5 dpf shows eGFP starting just posterior to the brain and continuing posteriorly. (Right) The posterior of the 50, 53, 60, 72, 79, 96 hpf and 5 dpf MM.m2de2-pGW-*cfos*-eGFP injected embryo shows eGFP expression in muscle fibers. These images have been modified to increase contrast between the embryo and its background.

Expression Observed within Positive Control Injected Embryos

To ensure that expression seen with MM.m2de2-pGW-*cfos*-eGFP was unique, and caused by the m2de2 sequence, three types of negative control embryos were examined. The first type were wild type or AB* embryos that were not injected but imaged at the same time points eGFP was observed with MM.m2de2-pGW-*cfos*-eGFP to ensure that what I observed was not a result of autofluorescence. The second type of control embryos were mock injected

with only water and phenol red to determine if the injection procedure altered the autofluorescence observed. The final control group was injected with the *pgw-cfos-eGFP* plasmid but lacking the *MM.m2de2F* sequence along with Tol2 mRNA to determine if the *cfos* minimal promoter could drive the eGFP expression pattern observed (Fig. 22). All three control groups were imaged and no eGFP was observed. In all cases, no eGFP was observed in somites or the hindbrain expression pattern consistent with that directed by *MM.m2de2F*.

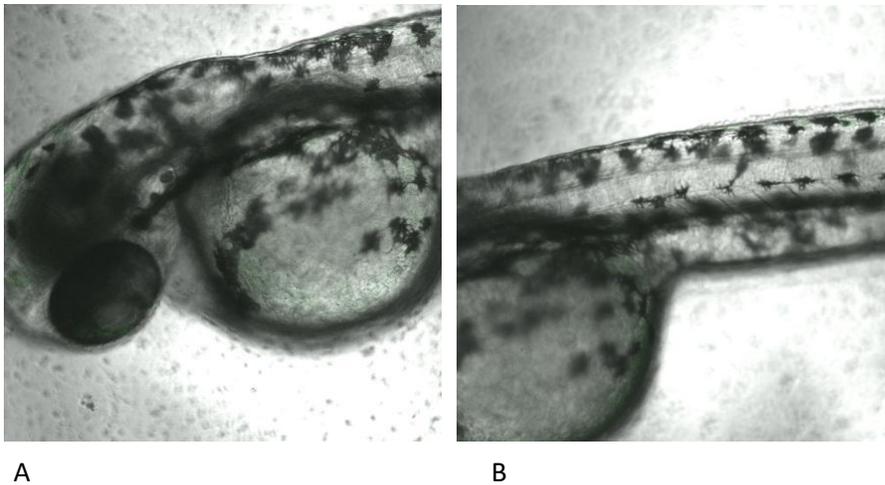


Fig. 22. Zebrafish embryos injected with transposase mRNA and *pgw-cfos-eGFP* and imaged at 52 hpf. No eGFP expression is observed in the head or trunk.

Transgenic Confirmation from MM.m2de2 Injected Embryos

To confirm *MM.m2de2* was integrating into the genome of the injected fish, genomic DNA was isolated initially from an entire clutch of *MM.m2de2-pGW-cfos-eGFP* injected 5 dpf fry and screened by PCR for the presence of *m2de2*. I was not able to detect the presence

of m2de2 in lane 3 (Fig. 23). After this I employed a different approach and only began to screen larvae from injected embryos that exhibited eGFP expression at 60 hpf. These eGFP positive 60 hpf embryos had their genomic DNA isolated, and a PCR screen with MM.m2de2 specific primers using this genomic DNA as template was performed. Using this approach, I observed a band consistent with MM.m2de2 (Fig. 24).

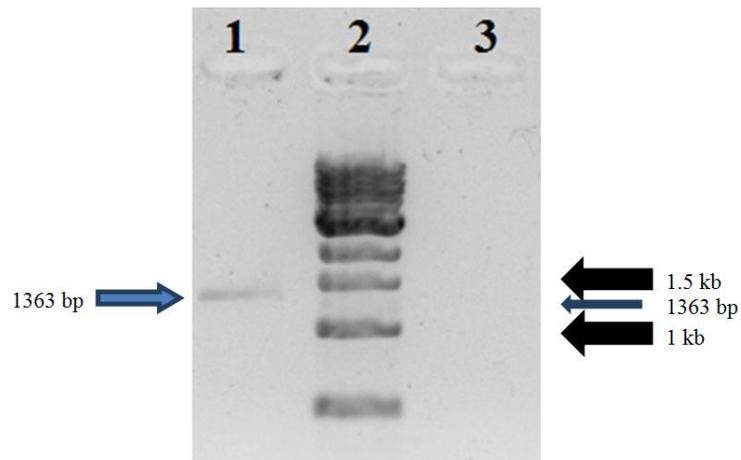


Fig. 23. MM.m2de2-pGW-*cfos*-eGFP injected embryo DNA PCR results when not selecting DNA from GFP expressing embryos. Lane 1, m2de2 PCR product of MM.m2de2-pGW-*cfos*-eGFP, thinner arrow indicates the predicted size of 1363 bp. Lane 2, 1.5 kb band on the 1 kb ladder (BioLabs N3232S). Lane 3, m2de2 PCR product of genomic DNA isolated from 5 dpf injected embryos, thinner arrow points to where the predicted size of band is should be indicated but is not present.

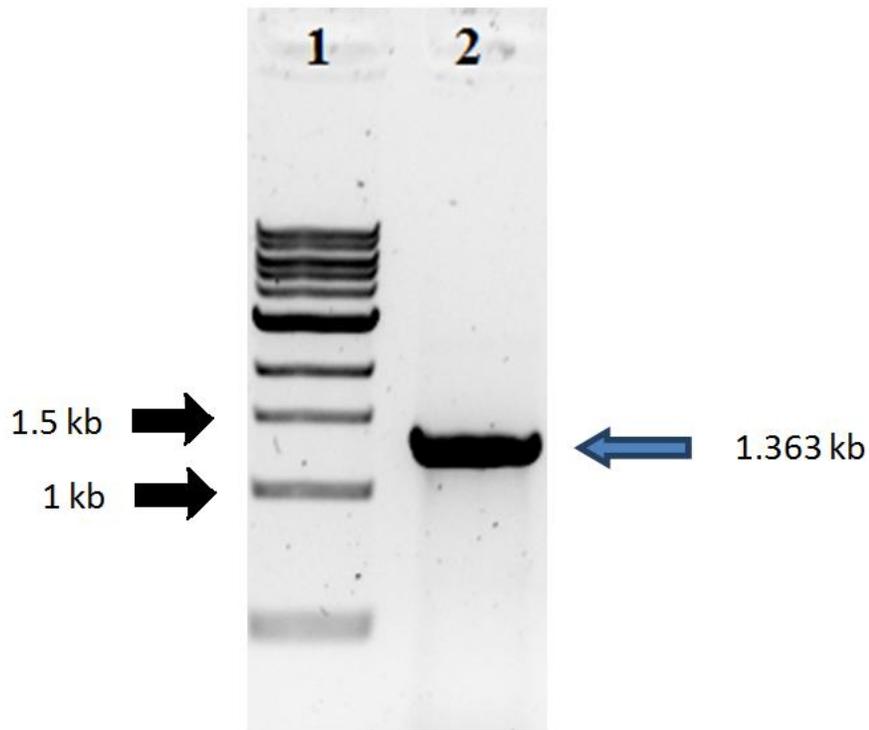


Fig. 24. MM.m2de2-pGW-*cfos*-eGFP injected embryo DNA PCR results when selecting DNA from GFP expression embryos. Lane 1, 1.5 kb band on the 1 kb ladder (BioLabs N3232S). Lane 2, PCR product of genomic DNA isolated from 60 hpf injected and screened embryos, arrow points to predicted size, 1363 bp, band indicated.

Transgenic confirmation from adult fin clip

To generate stable transgenic lines of zebrafish containing MM.m2de2-pGW-*cfos*-eGFP, injected embryos were allowed to reach sexual maturity and crossed. The offspring of the injected parents did not exhibit eGFP consistent with injected embryos of the same age. Two female and two male adults that were raised from embryos injected with MM.m2de2-

pGW-*cfos*-eGFP were screened for the presence of MM.m2de2 using PCR with MM.m2de2 specific primers and genomic DNA obtained from fin clips. Unfortunately the PCR from the fin clip did not show a band, suggesting these fish were not transgenic (Fig. 25).

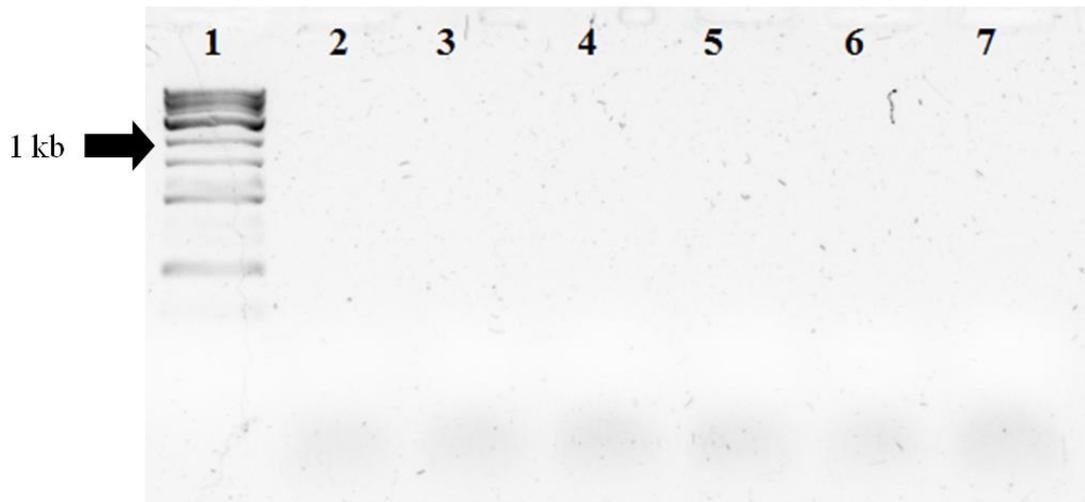


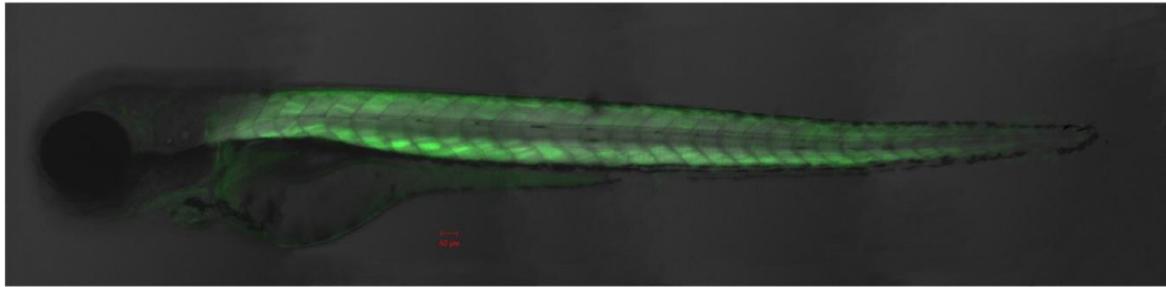
Fig. 25. PCR results for MM.m2de2-pGW-*cfos*-eGFP injected adult fish. Lane 1, 1 kb ladder (BioLabs N3232S) with 1.5 kb band indicated by arrow. Lanes 2-6, should show bands where arrow is if the adult fish were transgenic.

This result was similar to what happened when embryos were screened for m2de2 without first isolating eGFP positive fish. Because screening at 60 hpf proved helpful to that process, it was applied to the generation of a m2de2/eGFP transgenic line.

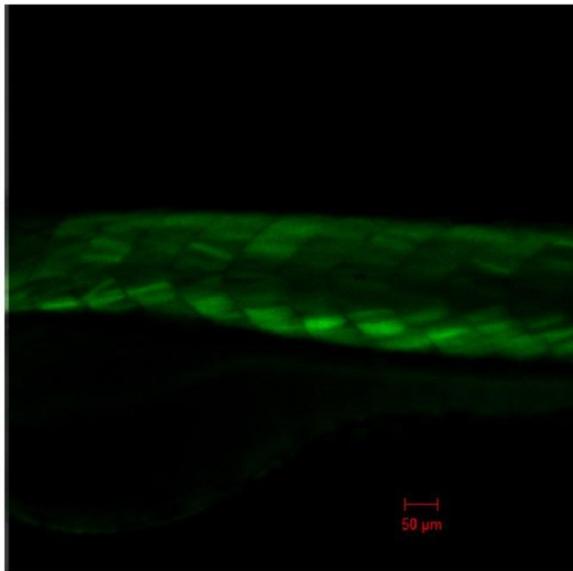
Generation of a Transgenic Line from Injected Individuals

The establishment of a transgenic line, from fish that were injected with the MM.m2de2-pGW-*cfos*-eGFP was performed in AB*embryos that were injected with the construct and then screened for eGFP expression at 60 hpf. Only those embryos that

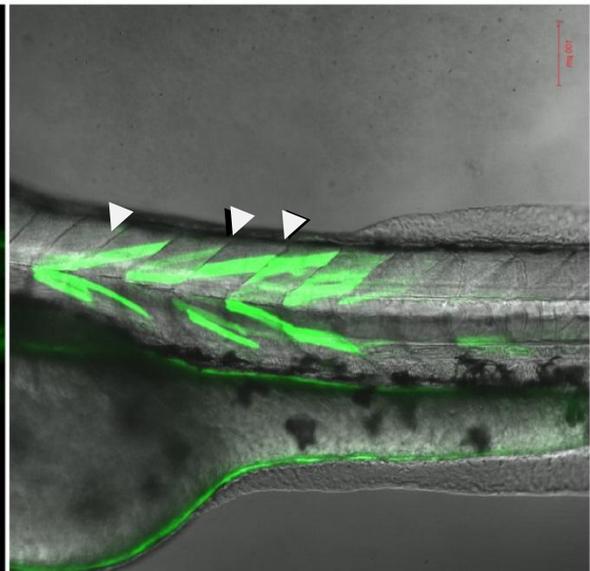
expressed eGFP in a pattern consistent with that of m2de2 were allowed to reach sexual maturity. Upon reaching sexual maturity the fish were crossed to each other to determine if they were able to produce offspring that exhibited eGFP expression consistent with what had been observed previously. Of the 12 screened fish that reached sexual maturity, two fish have produced eGFP expressing offspring. The eGFP expression pattern observed is similar to MM.m2de2-pGW-*cfos*-eGFP injected fish (Fig. 26).



A



B



C

Fig. 26. Progeny of m2de2 injected fish produce eGFP in muscle fibers. (A) 80 hpf eGFP expressing in skeletal muscle fibers within the somites of whole embryo posterior to head; photo collage was assembled from multiple body segment images. (B) 80 hpf eGFP trunk body segment displaying only fluorescence found in some skeletal muscle fibers. (C) 60 hpf eGFP embryo shows fluorescence in some skeletal muscles fibers. Images have been resized and rotated so that anterior is to the left and dorsal is up.

Transgenic MM.m2de2/eGFP Muscle Expression Pattern Seen with Non-Meis Genes

The striations observed with the MM.m2de2-pGW-*cfos*-eGFP injected fish have been documented in other transgenic zebrafish embryos (Akitake et al., 2011; Bessarab et al., 2008; Xu et al., 2003). Most of the constructs that show similar expression patterns are from conserved non-coding regions upstream of developmental genes that play roles in muscle development, such as *pax9* (Chatterjee et al., 2011). This suggests that m2de2 may be directing Meis2 expression to cells destined to become muscle within the somites.

Interestingly, m2de2 contains binding sites for not only other homeodomain proteins but also E-box sequences (Fig. 10). E-boxes are DNA sequences that have been linked to muscle development within the developing zebrafish embryo (Berkes and Tapscott, 2005; Du et al., 2003). They are sequences frequently found in regulatory regions of muscle specific genes (Berkes and Tapscott, 2005). One protein that has been shown to bind to the E-box sequence is MyoD which has been shown to play an important role in skeletal muscle development (Berkes and Tapscott, 2005) and that also can require Meis as a cofactor. This suggests the possibility that MyoD, a protein that requires Meis proteins as a partner, is able to regulate the expression of its partner during somitogenesis. It will be interesting to examine the m2de2 sequence more closely for the presence of other binding sequences for additional proteins that have been shown to play roles in the development of skeletal muscle as part of future studies and to also begin to make a closer examination of what proteins are able to bind m2de2 and regulate gene expression through it. As development proceeds, eGFP

expression is observed to no longer be present in the brain and becomes more pronounced as long striations of expression within the somites (Fig. 21).

Conclusions and Future Directions

In the course of this project, mouse *Meis2* downstream element 2 was amplified out of the MM.m2de2.pCR2.1-TOPO. Using Gateway cloning, MM.m2de2 was transformed into the vector pDonR221, and then was transformed into the destination vector pGW-*cfos*-eGFP. Following sequence confirmation, MM.m2de2-pGW-*cfos*-eGFP DNA was co-injected into single cell zebrafish embryos with transposase mRNA, Phenol red (0.5% in DPBS) and was brought to a final volume with DEPC water. Confocal imaging of the resulting transgenic zebrafish embryos revealed eGFP activity starting at the 48 hpf stage and continuing past 5 dpf. At approximately 48 hpf injected fish containing the MM.m2de2- *cfos*-eGFP were observed to first express eGFP in the midbrain and hindbrain (Fig. 20). As development proceeded, eGFP expression was no longer observed in the brain and became more pronounced as long striations of expression within the somites and continues past 5 dpf (Fig. 21). This pattern was not noted in any control fish which included non-injected wild type, mock injected embryos, and embryos injected with *cfos*-eGFP without the MM.m2de2 sequence found in the MM.m2de2- *cfos*-eGFP.

To confirm MM.m2de2 was integrating into the genome of the injected fish, genomic DNA was isolated from an entire clutch of MM.m2de2-pGW-*cfos*-eGFP injected 5 dpf fry and screened by PCR for the presence of m2de2 but was not detected (Fig. 23). Different

injected embryos were first screened for eGFP expression at 60 hpf. Genomic DNA isolated from these eGFP positive 60 hpf PCR screen for MM.m2de2 showed a band consistent with MM.m2de2 (Fig. 24). Only fish positive for eGFP were also positive for m2de2; this suggests that eGFP expression correlates to the presence of MM.m2de2. Another clutch of injected embryos were screened at 60 hpf; embryos that expressed eGFP in a pattern consistent with MM.m2de2 were crossed upon reaching sexual maturity. This was done to determine if they were able to produce offspring that exhibited eGFP expression consistent with what had been observed with MM.m2de2-pGW-*cfos*-eGFP injected fish. The resulting embryos were screened using confocal microscopy and eGFP was expressed in the offspring similar to the MM.m2de2-pGW-*cfos*-eGFP injected parents (Fig. 26).

Currently, the eGFP expressing offspring are being screened and raised in hopes to breed MM.m2de2-pGW-*cfos*-eGFP to homozygosity and have a stable line. Additionally, *meis2.2* and *meis2.1* expression patterns are being examined in zebrafish embryos older than 48 hpf to determine if expression patterns more closely resemble the somite expression found with MM.m2de2-pGW-*cfos*-eGFP injected embryos. A construct with the reverse orientation of m2de2 within MM.m2de2-pGW-*cfos*-eGFP is ready to be injected in a similar manner to how it was reported here. This is done to determine if the orientation will alter expression patterns; it should not change expression patterns because the proteins that interact with the element should still recognise the DNA sequence despite its orientation.

The E-boxes within the m2de2 sequence and the observed somite expression suggest MM.m2de2 may be interacting with muscle specific genes or regulatory genes. To examine this we will first look at MyoD. The expression pattern of myoD will be compared to MM.m2de2-pGW-*cfos*-eGFP embryos to examine if the two expression patterns overlap. Further examination of the relationship between *MyoD* and MM.m2de2 will also occur to determine how closely they work together and if one is able to regulate the other. We will determine if the MyoD protein is able to bind to the m2de2 Ebox sites.

In conclusion, the data presented in this study shows expression of the previously uncharacterized putative *Meis2* linked element m2de2. The pattern of expression in the trunk suggests this element may be working as a *cis* regulatory element with developmental genes and possibly with muscle specific developmental genes and/or proteins. This hypothesis is further supported by the E-box sites within the m2de2 sequence. However, which muscle specific developmental genes and/or proteins working with m2de2 remains to be determined. Additionally, while m2de2 is consistently found near *Meis2* in tetrapods, a direct relationship between *Meis2* and m2de2 can not be confirmed at this time. As a whole, the work completed for this project assisted in characterizing a novel element possibly involved in the regulation of *Meis2* and other unknown developmental genes and or proteins.

References

- Aamar E., D. Frank. 2004. *Xenopus* Meis3 protein forms a hindbrain-inducing center by activating FGF/MAP kinase and PCP pathways. *Development*. 131, 153-163
- Akitake C., M. Macurak, M. Halpern, M. Goll. 2011. Transgenerational analysis of transcriptional silencing in zebrafish. *Developmental Biology*. 352(2),191-201
- Akoulitchev S., T.P. Mäkelä, R.A. Weingberg, D. Reinberg. 1995. Requirement for TFIID kinase activity in transcription by RNA polymerase II. *Nature*. 377, 557-560
- Alberti, S., A. G. Gitler, & S. Lindquist. 2007. A suite of gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast*. 24(10), 913–919.
- Allen T.D., Y. Zhlp, T.S. Hawley, R.G. Hawley. 2000. TALE homeoproteins as HOX11-interacting partners in T-cell leukemia. *Leukemia & Lymphoma*. 39(3-4), 241-256
- Amali A., C. Lin, W. Wang, H. Gong, C. Lee, Y. Ko, J. Lu, G. Her, T. Chen, J. Wu. 2004. Up-regulation of muscle-specific transcription factors during embryonic somitogenesis of zebrafish (*Danio rerio*) by knock-down of myostatin-1. *Developmental Dynamics*. 229 (4), 847-856
- Amores A., A. Force, Y. Yan, L. Joly, C. Amemiya, A. Fritz, R.K. Ho, J. Langeland, V. Prince, Y. Wang, M. Westerfield, M. Ekker, J.H. Postlethwait. 1998. Zebrafish *hox* clusters and vertebrate genome evolution. *Science*. 282(5394),1711-1714
- Andersson R., A. Sandelin, C.G. Danko. 2015. A unified architecture of transcriptional regulatory elements. *Trends in Genetics*. 31(8),426–433
- Banerjee-Basu S., D. Sink, A. Baxevanis. 2001. The homeodomain resource: sequence, structures, DNA binding sites and genomic information. *Nucleic Acids Research*. 29(1), 291-293
- Bell A.C., A.G. West, G. Felsenfeld. 2001. Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science*. 291(447),447-450

- Bellaoui M., M. S. Pidkowich, A. Samach, K. Kushalappa, S.E. Kohalmi, Z. Modrusan, W.L. Crosby, G.W. Haughn. 2001. The Arabidopsis BELL1 and KNOX TALE homeodomain proteins interact through a domain conserved between plants and animals. *Plant Cell*. 13(11),2455-2470
- Berkes C., D. Bergstrom, B. Penn, K. Seaver, P. Knoepfler, S. Tapscott. 2004. Pbx marks genes for activation by MyoD indicating a role for homeodomain protein in establishing myogenic potential. *Molecular Cell*. 14,465-477
- Berkes C., S. Tapscott. 2005. MyoD and the transcriptional control of myogenesis. *Seminars in Cell & Developmental Biology*. 16(4-5),585-595
- Bessarab D., S. Chong, B. Srinivas, V. Korzh. 2008. Six1a is required for the onset of fast muscle differentiation in zebrafish. *Developmental Biology*. 323(2),216-228
- Bhatia S., J. Monahan, V. Ravi, P. Gautier, E. Murdoch, S. Brenner, V. van Heyningen, B.Venkatesh, D. A. Kleinjan. 2014. A survey of ancient conserved non-coding elements in the *PAX6* locus reveals a landscape of interdigitated *cis regulatory* archipelagos. *Developmental Biology*. 387(2),214-228
- Bhatt A.M., J.P. Etchells, C. Canales, A. Lagodienko, H. Dickinson. 2004. VAAMANA—a BEL1-like homeodomain protein, interacts with KNOX proteins BP and STM and regulates inflorescence stem growth in *Arabidopsis*. *Gene*. 328(17), 103–111
- Blackwood E.M., J.T. Kadonaga. 1998. Going the distance: a current view of enhancer action. *Science*. 281,60-63
- Bumsted-O'Brien K., A. Hendrickson, S. Haverkamp, R. Ashery-Padan, D. Schulte. 2007. Expression of the homeodomain transcription factor Meis2 in the embryonic and postnatal retina. *The Journal of Comparative Neurology*. 505, 58-72
- Buratiwski S., S. Hahn, L. Guarebte, P.A. Sharp. 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*. 56,549-561
- Bürglin T. 1997. Analysis of TALE superclass homeobox genes (MEIS, PBC KNOX, IROQUOIS, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acid Research*. 25(21),4173-4180

- Butler J.E.F., J.T. Kadonaga. 2001. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes & development*. 15,2515-2519
- Capdevila J., T.Tsukui, C. Esteban, V. Zappavigna, J. Belmonte.1999. Control of vertebrate limb outgrowth by the proximal factor *Meis2* and distal antagonism of BMPs by gremlin. *Molecular Cell*. 4(5),839–849
- Cecconi, F., G. Proetzel, G. Alvarez-Bolado, D. Jay, P. Gruss. 1997. Expression of *Meis2*, a *Knotted*-Related murine homeobox gene, indicates a role in the differentiation of the forebrain and the somitic mesoderm. *Developmental Dynamics*. 210, 184-190
- Chang C., L. Broechieri, W. Shen, C. Largman, M. Cleary. 1996. Pbx modulation of Hox homeodomain amino-terminal arms established different DNA-binding specificities across the Hox locus. *American society Microbiology*. 16(4),1734-1745
- Chariot, A., J. Gielen, M.-P. Merville, V. Bours. 1999. The homeodomain containing proteins: an update on their interacting partners. *Biochemical Pharmacology*. 58(12), 1851–1857
- Chatterjee S., G. Bourque, T. Lufkin. 2011. Conserved and non conserved enhancers direct tissue specific transcription in ancient germ layer specific developmental control genes. *BMC Developmental Biology*. 11, 63
- Choe S., N. Vlachakis, C. Sagerström. 2002. Meis family proteins are required for hindbrain development in the zebrafish. *Development*. 129, 585-595
- Chen Z., A. Jawhari, L. Fischer, C. Buchen, S. Tahir, T. Kamenski, M. Rasmussen, L. Lariviere, J. Bukowski-Wills, M. Nilges, P. Cramer, J. Rappsilber. 2010. Architecture of the RNA polymerase II -TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO Journal*. 29(4),717-726
- Choe S.K., F. Ladam, C.G. Sagerström. 2014. TALE factors poise promoters for activation by Hox proteins. *Developmental Cell*. 28(2),203-11
- Cianfrocco M.A., G.A. Kassavetis, P. Grob, J. Fang, T. Juven-Gershon, J.T. Kadonaga, E. Nogales. 2013. Human TFIID Binds to Core Promoter DNA in a Reorganized Structural State. *Cell*. 152(1-2),120-131
- Cunningham T., C. Chtazi, L. Sandell, P. Trainor, G. Duester, 2011. *Rdh10* mutants deficient in limb field retinoic acid signaling exhibit normal limb patterning but display interdigital webbing. *Developmental Dynamics*. 240 (5),1142-1150

- Cvejic A., J.Serbanovic-Canic, D.. Stemple, W.. Ouwehand. 2011. The role of *meis1* in primitive and definitive hematopoiesis during zebrafish development. *Haematologica*. 96, 190-198
- Dai M, Y. Wang, L. Fang, D.M. Irwin, T. Zhu, J. Zhang, S. Zhang, Z. Wang. 2014. Differential expression of *Meis2*, *Mab2112* and *Tbx3* during limb development associated with diversification of limb morphology in mammals. *PLoS One*. 9(8),e106100
- Derelle, R., P. Lopez, H. Guyader, M. Manuel. 2007. Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes. *Evolution & Development*. 9, 212–219.
- DiIorio, P., K. Alexa, S. Choe, L. Etheridge, C. Sagerström. 2007. TALE-Family homeodomain proteins regulate endodermal sonic hedgehog expression and pattern the anterior endoderm. *Developmental Biology*. 304(1),221-231
- Dror I., T. Zhou, Y. Mande-Gutfreund. R. Rohs. 2014. Covariation between homeodomain transcription factors and the shape of their DNA binding site. *Nucleic Acids Research*. 42(1),430-441
- Du S., J. Gao, V. Anyangwe. 2003. Muscle-specific expression of *myogenin* in zebrafish embryos is controlled by multiple regulatory elements in the promoter. *Comparative Biochemistry part B: Biochemistry & Molecular Biology*. 134(1),123-134
- Duboule D. 1998. Vertebrate *Hox* gene regulation: clustering and/or colinearity? *Current Opinion in Genetics & Development*. 8(5),514-518
- Durston A., H. Jansen, P. Rieden, M. Hooiveld. 2011. *Hox* collinearity-a new perspective. *The International Journal of Developmental Biology*. 55, 899-908
- Elkouby Y, S. Elias, E. Casey, S. Blythe. N. Tsabar, P. Klein, H. Root, K. Liu, D. Frank. 2010. Mesodermal Wnt signaling organizes the neural plate via *Meis3*. *Development*. 137(9), 1531-41
- Erickson T., C. French, A. Waskiewicz. 2010. *Meis1* specifies positional information in the retina and tectum to organize the zebrafish visual system. *Neural Development*. 5, 55
- Epstein J., T. Glaser, J. Cai, L. Jepeal, D. Walton, R. Maas. 1994. Two independent and interactive DNA binding subdomains of the *Pax6* paired domain are regulated by alternative splicing. *Genes & Development*. 8,2022-2034

- Ferretti J., C. Villaescusa, P. Di Rosa, L.C. Fernandez-Diaz, E. Longobardi, R. Mazzieri, A. Miccio, N. Micali, L. Selleri, G. Ferrari, F. Blasi. 2006. Hypomorphic mutation of the TALE Gene *Prep1* (*pKnox1*) Causes a major reduction of Pbx and Meis proteins and a pleiotropic embryonic phenotype. *Molecular & Cellular Biology*. 26(15),5650-5662
- Fisher S., E.A. Grice, R.M. Vinton, S.L. Bessling, A. Urasaki, K. Kawakami, A.S. McCallion. 2006. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nature Protocols*. 1(3), 1297–1305.
- Fognani C., C. Kilstrup-Nielsen, J. Berthelsen, E. Ferretti, V. Zappavigna, F. Blasi. 2002. Characterization of PREP2, a paralog of PREP1, which defines a novel sub-family of the MEINOX TALE homeodomain transcription factors. *Nucleic Acids Research*. 30(9),2043-2051.
- Forget D., A. Lacombe, P. Cloutier, R. Al-Khoury, A. Bouchard, M. Lavallée-Adam, D. Faubert, C. Jeronimo, M. Blanchette and B. Coulombe. 2010. The protein interaction network of the human transcription machinery reveals a role for the conserved GTPase RPAP4/GPN1 and microtubule assembly in nuclear import and biogenesis of RNA polymerase II. *Molecular & Cellular Proteomics*. 9, 2827-2839
- Gehring W., Y. Quian, M. Billeter, K. Furukubo-Tokunaga, A. Schier, D. Resendez-Perez, M. Affolter, G. Otting, K. Wüthrich. 1994. Homeodomain-DNA recognition. *Cell*. 78, 221-223
- Heine P., E. Dohle, D. Schulte. 2009. Sonic hedgehog signaling in the chick retina accelerates Meis2 downregulation simultaneously with retinal ganglion cell genesis. *NeuroReport*. 20(3),279-284
- Hisa T., S. Spence, R. Rachel, M. Fujita, T. Nakamura, J. Ward, D. Devor-Henneman, Y. Saiki, H. Kutsuna, L. Tessarollo, N. Jenkins, N. Copeland. 2004. Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *EMBO Journal*. 23(2),450-459
- Irimia M., I. Maeso, D. Burguera, M. Hidalgo-Sánchez, L. Puelles, S. Roy, J. Garcia-Fernández, J. Ferran. 2011. Contrasting 5' and 3' evolutionary histories and frequent evolutionary convergence in *Meis/hth* gene structures. *Genome Biology & Evolution*. 3, 551-564
- Iyer L.M., V. Anantharaman, M. Wolf, L. Aravind. 2008. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitology*. 38(1),1–31

- Jacobs Y., C.A. Schnabel, M. L. Cleary. 1999. Trimeric association of Hox and TALE homeodomain proteins mediates *Hoxb2* hindbrain enhancer activity. *Molecular & Cellular Biology*. 19(7),5134-5142
- Jackson B., S.J. Brown, A. A. Avilion, R. F. L. O'Shaughnessy. K. Sully, O.Akinduro, M. Murphy, M.L. Cleary, C. Byrne. 2011. TALE homeodomain proteins regulate site-specific terminal differentiation, *LCE* genes and epidermal barrier. *Journal of Cell Science*. 124, 1681-1690.
- Jave-Suárez L.F., J. Schweizer. 2006. The HOXC13-controlled expression of early hair keratin genes in the human hair follicle does not involve TALE proteins MEIS and PREP as cofactors. *Archives Dermatological Research*. 297(8),372-376.
- Kawakami, K. 2007. Tol2 : a versatile gene transfer vector in vertebrates. *Genome Biology*. 8(Suppl 1), 1–10.
- Kawakami, K., H. Takeda, N. Kawakami, M. Kobayashi, N. Matsuda, M. Mishina. 2004. A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Developmental Cell*. 7(1), 133–44
- Kimmel C.B., W.W. Ballard, S.R. Kimmel, B. Ullmann, T.F. Schilling. 1995. Stages of embryonic development of the zebrafish. *Developmental Dynamics*. 203, 253-310
- Krumlauf R. 1994. Hox genes in vertebrate development. *Cell*. 78, 191-201
- Knoepfler P.S., D.A. Bergstrom, T. Uetsuki, I. Dac-Korytko, Y.H. Sun, W.E. Wright, S.J. Tapscott, M.P. Kamps. 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 Research*. 27(18),3752-3761
- Kwan, K. M., E. Fujimoto, C. Grabher, B.D. Mangum, M.E. Hardy, D.S. Campbell, C.B. Chien. 2007. The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Developmental Dynamics : An Official publication of the American Association of Anatomists*. 236(11),3088–99
- Langelier M., D. Forget, A. Rojas, Y. Porlier, Z.F. Burton, B. Coulombe. 2001. Structural and functional interactions of Transcription Factor (TF) IIA with TFIIE and TFIIIF in transcription initiation by RNA polymerase II. *Journal of Biological Chemistry*. 276, 38652-38657

- Lappin T., D. Grier, A. Thompson, H. Halliday. 2006. HOX genes: seductive science, mysterious mechanisms. *Ulster Medical Journal*. 75(1),23-31
- LeBrun D. 2003. E2A Basic Helix-loop-helix Transcription factors in human Leukemia. *Frontiers in Bioscience*. 8, 206-222
- Lemons D., W. McGinnis. 2006. Genomic evolution of Hox gene clusters. *Science*. 313, 1918-1922
- Liu Y, R.J. MacDonald, G.H. Swift. 2001. DNA binding and transcriptional activation by a PDX1.PBX1b.MEIS2b trimer and cooperation with a pancreas-specific basic helix-loop-helix complex. *Journal of Biological Chemistry*. 276, 17985–17993
- Liu J., Y. Wang, M. Birnbaum, D. Stoffers. 2010. Three-amino-acid-loop-extension homeodomain factor Meis3 regulates cell survival via PDK1. *PNAS*. 107(47),20494-20499
- Longobardi E., D. Penkov, D. Mateos, G. De Florian, M. Torres, F. Blasi. 2014. Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. *Developmental Dynamics*. 243(1),59-75
- Machon O., J. Masek, O. Machonova , S. Krauss , Z. Kozmik. 2015. Meis2 is essential for cranial and cardiac neural crest development. *BMC Developmental Biology*. 15(1),40
- Magnani E., S. Hake. 2008. KNOX Lost the OX: The *Arabidopsis KNATM* gene defines a novel Class of KNOX transcriptional regulators missing the homeodomain. *Plant Cell*. 20(4),875-887.
- Mahmoud A., F. Kocabas, S. Muralidhar, W. Kimura, A. Koura, S. Thet, E. Porrello, H. Sadek. 2013. Meis1 regulates postnatal cardiomyocyte cell cycle arrest. *Nature*. 497(7448),249-253
- Manfroid I., F. Delporte, A. Baudhuin, P. Motte, C. Neumann, M. Voz, J. Martial, B. Peers, 2007. Reciprocal endoderm-mesoderm interactions mediated by *fgf24* and *fgf10* govern pancreas development. *Development*. 134, 4011-4021
- McGinnis W., R. Krumlauf. 1992. Homeobox genes and axial patterning. *Cell*. 68, 283-302
- Melvin V., W. Feng, L. Henandez-Lagunas, K. Artinger, T. Williams. 2013. A morpholonia-based screen to identify novel genes involved in craniofacial morphogenesis. *Developmental Dynamics*. 242(7),817-831

- Mercader N., E.M. Tanaka, M. Torres. 2005. Proximodistal identity during vertebrate limb regeneration is regulated by Meis homeodomain proteins. *Development*. 132, 4131-4142
- Moens, C. B., L. Selleri. 2006. Hox cofactors in vertebrate development. *Developmental Biology*. 291(2), 193–2066
- Moens, C., V. Prince. 2002. Constructing the hindbrain: insights from the zebrafish. *Developmental Dynamics*. 224(1),1–17
- Moskow, J., F. Bullrich, K.Huebner, I. Daar, A. Buchberg. 1995. Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice. *Molecular & Cellular Biology*. 15(10),5434-5443
- Mukherjee K., T. Bärzlin. 2007. Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution. *Journal of Molecular Evolution*. 65, 137–153
- Nachtrab G., M. Czerwinski, K. Poss. 2011. Sexually dimorphic fin regeneration in zebrafish controlled by androgen/GSK3 signaling. *Current Biology*. 21, 1912-1917
- Nelson, K.C. 2011. Identification of gene regulatory elements associated with the meis family of homeobox genes. Master's Thesis, Department of Biology. Appalachian State University, Boone, North Carolina
- Noël, E., A. Casal-Sueiro, E. Busch-Nentwich, H. Verkade, P. Dong, D. Stemple, E.Ober. 2008. Organ-specific requirements for Hdac1 in liver and pancreas formation. *Developmental Biology*. 322(2),237-250
- Noordermeer D., M. Leleu, E. Splinter, J. Rougemont, W. DeLaat, D. Duboule. 2011. The dynamic architecture of *Hox* gene clusters. *Science*. 334(6053),222-225
- Orphanides G., T. Lagrange, D. Reinberg. 1996. The general transcription factors of RNA polymerase II. *Genes & Development*. 10(21),2657-2683
- Oulad-Abdelghani M., C. Chazaud, P. Bouillet, V. Sapin, P. Chambon, P. Dollé. 1997. Meis2, a novel mouse Pbx-related homeobox gene induced by retinoic acid during differentiation of P19 embryonal carcinoma cells. *Developmental Dynamics*. 210(2),173-183

- Paige, S. L., S. Thomas, C. Stoick-Cooper, H. Wang, L. Maves, R. Sandstrom, L. Pabon, H. Reinecke, G. Pratt, G. Keller, R. Moon, J. Stamatoyannopoulos, C. Murry. 2012. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell*. 151(1),221–232.
- Pardee T.S., C.S. Bangur, A.S. Ponticelli. 1998. The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. *Journal of Biological Chemistry*. 273, 17859-17864
- Pavletich N.P., C.O. Pabo. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*. 252 (5007),809-817
- Prince V., C. Moens, C. Kimmel, R. Ho. 1998. Zebrafish hox genes: expression in the hindbrain region of wild-type and mutants of the segmentation gene, valentine. *Development*. 125, 393-406
- Puschel A.W., P. Gruss, M. Westerfield. 1992. Sequence and expression pattern of Pax-6 are highly conserved between zebrafish and mice. *Development*. 114, 643-651
- Rohs R., Sean West, A. Sosinsky, P. Liu, R. Mann, B. Honig. 2009. The role of DNA shape in protein-DNA recognition. *Nature*. 461(7268),1248-1253
- Santos, J., N. Fonseca, C. Vieira, J. Vieira, F. Casares. 2010. Phylogeny of the teashirt-related zinc finger (tshz) gene family and analysis of the developmental expression of tshz2 and tshz3b in the zebrafish. *Developmental Dynamics*. 239(3),1010-1018
- Schaeffer L., R. Roy, S. Humbert, V. Moncollin, W. Vermeulen, J.H.J. Hoeijmakers, P. Chambon, J. Egly. 1993. DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science*. 260(5104),58-63
- Schnabel C., Y. Jacobs, M. L. Cleary. 2000. HoxA9-mediated immortalization of myeloid progenitors requires functional interactions with TALE cofactors Pbx and Meis. *Oncogene*. 19(5),608-616
- Selleri L., J. DiMartino, J. van Deursen, A. Brendolan, M. Sanyal, E. Boon, T. Capellini, K. S. Smith, J. Rhee, H Pöpper, G. Grosveld, M. L. Cleary. 2004. The TALE Homeodomain protein Pbx2 is not essential for development and long-term survival. *Molecular & Cellular Biology*. 24(12),5324-5331

- Shang Z., V. E. Isaac, H. Li, L. Patel, K. M. Catron, T. Curran, G. T. Montelione, C. Abate. 1994. Design of a "minimal" homeodomain: the N-terminal arm modulates DNA binding affinity and stabilizes homeodomain structure. *PNAS*. 91(18),8373-8377
- Sheth R., L. Marcon, M.F. Bastida, M. Junco, L. Quintana, R. Dahn, M. Kmita, J. Sharpe, M.A. Ros. 2012. *Hox* genes regulate digit patterning by controlling the wavelength of a turing-type mechanism. *Science*. 338(6113),1476-1480
- Shi Y., J. Berg. 1995. Specific DNA-RNA hybrid binding by zinc finger proteins. *Science*. 268 (5208),282
- Stankunas K., C. Shang, K. Y. Twu, S.Kao, N. A. Jenkins, N. G. Copeland, M. Sanyal, L. Selleri, M.L. Cleary, C.Chang. 2008. Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circulation Research*. 103,702-709
- Steelman S., J.J. Moskow, K. Muzynski, C. North, T. Druck, J.C. Montgomery, K Huebner, I.O. Daar, A.M. Buchberg. 1997. Identification of a conserved family of Meis1-related homeobox genes. *Genome Research*. 7, 142-156
- Stewart J.J, L.A. Stargell. 2001. The stability of the zon-TBP-DNA complex is dependent on the sequence of the TATAAA element. *Journal of Biological chemistry*. 276, 30078-30084
- Tamaoki M., H. Tsugawa, E. Minami, T. Kayano, N. Yamamoto, Y. Kano-Murakami, M. Matsuoka. 1995. Alternative RNA products from a rice homeobox gene. *Plant Journal*. 7(6),924-938
- Tefferi A., E. Wieben, G.W. Dewald, D.A.H. Whiteman, M.E. Bernard, T.C. Spelsberg. 2002. Primer on medical genomics part II: Background principles and methods in molecular genetics. *Mayo Clinic Proceedings*. 77(8),785-808
- Thisse B., C. Thisse. 2004. Fast Release Clones: A High Throughput Expression Analysis. ZFIN Direct Data Submission, <http://zfin.org>. not published.
- Thisse C., B. Thisse. 2005. High throughput expression analysis of ZF-models consortium clones. ZFIN Direct Data Submission, <http://zfin.org>. not published.
- Thomas M.C., C.M. Chiang. 2006. The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology*. 41,105-178

- Thorsteinsdottir U., E. Kroon, L. Jerome, F. Blasi, G. Sauvageau. 2001. Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia. *Molecular & Cellular Biology*. 21(1),224-234
- Venters B.J., B.F. Pugh. 2009. How eukaryotic genes are transcribed. *Critical Reviews in Biochemistry and Molecular Biology*. 44(2-3),117-141
- Vlachakis N., S. Choe, C. Sagerström. 2001. Meis3 synergizes with Pbx4 and Hoxb1b in promoting hindbrain fates in the zebrafish. *Development*. 128, 1299-1312
- Waskiewicz A., H. Rikhof, R. Hernandez, C. Moens. 2001. Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning. *Development*. 128, 4139-4151
- Westerfield, Monte. 2000. *The Zebrafish Book: a guide for the laboratory use of zebrafish (Danio rerio)*. 4. Eugene: University of Oregon press. print.
- Wilfinger, A., V. Arkhipova, D. Meyer. 2013. Cell type and tissue specific function of islet genes in zebrafish pancreas development. *Developmental Biology*. 378(1),25-37
- Wittkopp P., G. Kalay. 2012. *cis regulatory* elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews. Genetics*. 13, 59-69
- Xu C., G. Wu, Y. Zohar, S. Du. 2003. Analysis of *myostatin* gene structure, expression and function in zebrafish. *Journal of Experimental Biology*. 406, 4067-4079
- Zerucha, T.,V. Prince. 2001. Cloning and developmental expression of a zebrafish meis2 homeobox gene. *Mechanisms of Development*. 102(1-2),247-250

Vita

Hannah Hemingway Freundlich was born in Burbank, CA to Hilary Hemingway and Jeff Freundlich. Following her time at Mariner high school, Hannah began her Bachelors of Science at North Carolina State University in 2007. She received a Bachelor of Science in both Animal Science and Poultry Science in 2012. Hannah started her master's research with Dr. Ted Zerucha in the spring of 2013. Following this publication, she intends to work for Appalachian State University in the office of research protections.