

CHARACTERIZATION OF THE MEIS2A DOWNSTREAM REGULATORY  
ELEMENT *DR-M2DE1*

A Thesis  
By  
TYLER JAMES FERRARA

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

December 2015  
Department of Biology

CHARACTERIZATION OF THE MEIS2A DOWNSTREAM REGULATORY  
ELEMENT *DR-M2DE1*

A Thesis  
By  
TYLER JAMES FERRARA  
December 2015

APPROVED BY:

---

Dr. Ted Zerucha  
Chairperson, Thesis Committee

---

Dr. Ece Karatan  
Member, Thesis Committee

---

Dr. Darren F. Seals  
Member, Thesis Committee

---

Dr. Zack E. Murrell  
Chairperson, Department of Biology

---

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

Copyright by Tyler James Ferrara 2015  
All Rights

## Abstract

### CHARACTERIZATION OF THE MEIS2A DOWNSTREAM REGULATORY ELEMENT *DR-M2DE1*

Tyler James Ferrara, B.S., University of Rhode Island  
M.S., Appalachian State University

Chairperson: Dr. Ted Zerucha

The *Meis2* gene encodes a homeodomain containing protein that acts as a *Hox* cofactor to regulate development in vertebrate embryos. *Meis2* is also a member of the TALE superclass, or three amino acid loop extension, which is a subset of homeodomain proteins that is characterized by an extra three amino acids between two of the alpha helices of the homeodomain. We have identified four highly conserved noncoding elements associated with the vertebrate *Meis2* gene and named them *m2de1-4* (for *Meis2* downstream element). While *M2de2-4* have to date only been found in land vertebrates, *m2de1* is also found in teleosts like zebrafish. The *m2de1* sequence is approximately 450bp in length and its sequence and relative position to *Meis2* (*meis2a* in zebrafish) is highly conserved amongst all vertebrates that we have examined. Using the Tol2 system we have generated transgenic zebrafish in which the zebrafish element (*dr-m2de1*) has been able to direct reporter transgene expression the mid and hindbrain of developing embryos.

The *m2de1* sequence was recently described by another group (Parker et al., 2011), however the expression of the reporter transgene being driven by their element

was restricted to the area posterior and rhombomere 5 which was dissimilar from the expression that we have observed which was a more broad expression throughout the mid and hindbrain. Upon examination of the sequences, we determined that our *m2de1* sequence contains an additional 19bp on the 3' end and an additional 17bp on the 5' end that were not reported in the sequence that was published by the other group. This suggests the possibility that their sequence does not represent the full-length element. In this study I have found that that both the Parker paper element, called 3288, and *m2de1* both drive broad expression patterns in mid and hindbrain, similar to that which was seen originally with *m2de1*. I also observed that *m2de1* drove expression in the muscle fibers of developing zebrafish while 3288 and any other generated fragment of *m2de1* did not. This would appear to suggest that both 3' and 5' basepairs not found in 3288 are both necessary for the enhancer to drive complete expression.

## **Acknowledgements**

During my time spent at Appalachian State University, many people helped me along the way. By far though, the most influential and important person has been my advisor Dr. Ted Zerucha. I thank him for his motivation, advice and friendship that without, I surely would have not gotten as far as I have.

I would also be remiss if I did not thank all of my lab mates, past and present, for all the help they have given me. Without the help of lab members like Zach Williams, Alicia Ramsaran, Tucker Munday and Hannah Freundlich I would have been lost my first semester and members like Tray Nielson made the lab a fun place to be. Most notably I would like to thank lab members Cody Barrett and Kyle Nelson for the work they have done previously on this project, without which I would have no basis for my research.

To the other members of my committee, Dr. Ece Karatan and Dr. Darren Seals, I would just like to say how thankful I am for all the help they gave me with my thesis. Besides their academic duties, they also had graduate students of their own to help, so the fact that they used what little extra time they had to read draft after draft of my comps and thesis means a lot.

In general I would also just like to thank everyone else at Appalachian State who helped me in my time here like Monique Eckerd, for her help in the animal facility, and

those who work in the Admissions Office and Office of Student Research for admitting me and help funding my project.

Finally I would just like to thank all my friends and family for all of the support they have given me now and in the past.

## Table of Contents

|                                     |    |
|-------------------------------------|----|
| Abstract.....                       | iv |
| Acknowledgements.....               | vi |
| List of Figures & Tables.....       | ix |
| Introduction.....                   | 1  |
| Materials and Methods.....          | 27 |
| Results.....                        | 38 |
| Discussion.....                     | 44 |
| References.....                     | 51 |
| Supplementary Material Figures..... | 71 |
| Vita.....                           | 74 |



## List of Figures & Tables

|  |    |
|--|----|
| Table 1. PCR Primers for Generation of Enhancer Fragments.....   | 30 |
| Figure 1. Diagram of Hox Expression.....   | 13 |
| Figure 2. Diagram of Hox Gene Clusters in Various Organisms.....   | 14 |
| Figure 3. Diagram of the m2de1 fragments and attB sites in Topo 2.1.....   | 32 |
| Figure 4. Diagram of Gateway BP Reaction.....  | 33 |
| Figure 5. Diagram of Gateway LR Reaction.....  | 34 |
| Figure 6. Diagram of Enhancers in Injection Vector Flanked by Tol2.....  | 35 |
| Figure 7. Diagram of 3' Dr-m2de1 Transcription factor Binding Sites.....   | 39 |
| Figure 8. Image of Dr-m2de1 and Enhancer Fragments.....  | 40 |
| Figure 9. Dr-m2de1 Driven Expression of GFP at 48hpf and 54hpf.....  | 41 |
| Figure 10. 3288 Driven Expression of GFP at 48hpf and 54hpf.....   | 42 |
| Figure 11. 3288-3'/5' Driven Expression of GFP at 48hpf and 54hpf.....   | 43 |
| Supplementary material Figure 1. Complete Diagram of Putative Dr-m2de1<br>Transcription Factor Binding<br>Sites..... | 7  |

## **Introduction**

Embryonic development is a complicated process consisting of many cell divisions, differentiations and migrations. If any one of these processes happens at the wrong time or in the wrong location, then the embryo may become deformed or perish. These processes are coordinated by various genetic signals generated by transcription of either the maternal or the zygotic genome, and then translated by the embryo. These signals can be regulated at either of these two steps to ensure that the proper genes are expressed at the right time and location. To fully understand how these genes are regulated a firm grasp of how they are transcribed is required.

Each gene consists of three general regions, an upstream region, a coding region and a downstream region. The upstream region refers to the region of DNA that lies upstream of the coding sequence and is where the promoter can be found. The promoter is the region bound by RNA Polymerase II and the location where transcription is initiated. The promoter typically contains the protein binding regions known as the TATA box and CAAT box. The CAAT box is a DNA sequence that lies about 75-80bp upstream of the transcriptional start site. Proteins bind the CAAT box to open up the chromatin surrounding the promoter so that general transcription factors (GTF's) can access the DNA and ultimately recruit RNA polymerase II. The CAAT box is bound by the CAAT enhancer binding protein (c/EBP). The TATA box is another promoter element that is typically but not always found

upstream of every gene. The TATA box is characterized by a repetitious sequence of thymine and adenine base pairs, found usually 25-30 bp upstream of the transcription start site. Promoters that do not possess a TATA box do still contain a thymine and adenine rich region that serves a similar function. The TATA box and similar regions are bound by general transcription factors like TFIID, which then recruit RNA polymerase II.

Downstream of the promoter there is the coding region, which consists of a transcription start site (or cap sequence), the 5' untranslated region (UTR), a translation start site and then finally a series of protein coding exons and non protein coding introns. Both the introns and the exons are transcribed into mRNA. The transcription start site is where the RNA polymerase begins transcribing the gene into mRNA. It is also called the cap sequence because this beginning sequence will eventually be capped with modified nucleotides. The translation start site is found further downstream and is where the first codon that is translated can be found. The 5' UTR lies between these two sites, and as the name suggests is not translated and does not contribute to the sequence of amino acids in the protein but instead helps control translation initiation (Barrett et al., 2012).

Immediately following the stop codon begins the region downstream of the coding region, here is found the 3' UTR. Similar to the 5' UTR, the 3' UTR is untranslated and serves as a regulatory region for translation (Barrett et al., 2012). The 3' UTR is involved in transcript cleavage, mRNA stability, polyadenylation and translational regulation and does so by containing numerous binding sites for proteins and for microRNAs in the mRNA strand (Barrett et al., 2012). What lies downstream of the coding region is mostly intergenic DNA but it may also contain various types of regulatory regions that may influence the transcription of the gene located upstream of it or another gene entirely. These regulatory

regions also may be found upstream (Davidson, 2006; Dutton et al., 2008; Echelard et al., 1994; Kikuta et al., 2007; Valverde-Garduno et al., 2004).

Transcription happens in three general steps. The first step is commonly known as initiation. Initiation is the step where the preinitiation complex, a group of proteins including RNA polymerase II and general transcription factors, is formed and binds the promoter. In order for the preinitiation complex to bind to the promoter, the area of chromatin around the gene must first be remodeled. This is done by the acetylation of specific residues on the histones, which are special proteins upon which DNA is coiled around and while the DNA is bound in this fashion, other proteins cannot act upon it. This acetylation causes the histones to expose the region of DNA containing the gene's promoter (Garcia-Ramirez et al., 1995).

Next, the protein c/EBP binds the CAAT box, which helps to further expose the promoter and TATA box. The TATA boxes unique T-A-T-A-@-A-A-N sequence, @ representing either T-A or A-T, allows for expansion in the minor groove of the DNA strand and grants it flexibility and rigidity in certain locations (Juo et al., 1996; Thomas and Chiang, 2006). TFIID binds this region using its TATA binding protein (TBP) subunit, one of 14 total subunits that includes the TBP associated Factors or TAFs (Li et al., 1999; Thomas and Chiang, 2006). These other subunits allow TFIID to bind those promoters that lack a TATA box by binding to sites such as the initiator (Inr) sequence, downstream promoter element (DPE), downstream core element (DCE) and the motif ten element (MTE). This binding bends the DNA in the minor groove over an angle of 80° (Juo et al., 1996; Nikolov et al., 1996; Thomas and Chiang, 2006). It is thought that this binding and bending helps unwind and separate the DNA strands. In addition, the TATA box is rich in thymine/adenine base pairings, which contain only two hydrogen bonds, a much weaker interaction than that of

cytosine/guanine base pairings that possess three bonds, which facilitates the separation of the DNA strands.

Once TFIID has bound DNA it recruits other GTFs: TFIIB, TFIIE, TFIIA, TFIIF and TFIIH. Each general transcription factor has a specific role in creating the preinitiation complex, a complex that consists of RNA polymerase II and the various general and specific transcription factors that help initiate transcription. After TFIID, TFIIB binds, which acts to stabilize the growing preinitiation complex (Orphanides et al., 1996; Thomas and Chiang, 2006). This general transcription factor binds two different sequences that flank the TATA box known as the TFIIB-recognition elements, BRE<sup>u</sup> (upstream) and BRE<sup>d</sup> (downstream) (Deng and Roberts, 2005; Thomas and Chiang, 2006). BRE<sup>u</sup> is bound by the helix-turn-helix domain of TFIIB while BRE<sup>d</sup> is bound by TFIIB's recognition loop (Deng and Roberts, 2005; Thomas and Chiang, 2006). TFIIB also binds the TBP subunit of TFIID using its C-terminal domain (Nikolov et al., 1995; Thomas and Chiang, 2006). TFIIB is responsible for the unidirectional transcription of RNA polymerase II by creating an asymmetrical complex, and TFIIB also is thought to help TFIID binding by providing more points of contact to the DNA strand (Deng and Roberts, 2005; Thomas and Chiang, 2006). At roughly the same time that TFIIB joins the preinitiation complex so does TFIIA and they act together to help stabilize TFIID (Thomas and Chiang, 2006).

TFIIF and RNA polymerase II attach next to the complex. TFIIF is comprised of a heterodimer made of the proteins RAP30 and RAP74 (Flores et al., 1990; Thomas and Chiang, 1990). This transcription factor interacts with the polymerase, DNA, and TFIIB and serves multiple roles. First, TFIIF assists in the recruitment of RNA polymerase II to the transcription factors currently bound to the promoter, and secondly TFIIF aids in stabilizing

the whole complex to facilitate stronger binding to DNA (Flores et al., 1990; Thomas and Chiang 2006). TFIIF is also necessary for the recruitment of the final two transcription factors, TFIIE and TFIIH (Orphanides et al., 1996; Thomas and Chiang, 2006).

The next transcription factors that bind to the preinitiation complex are TFIIE and TFIIH. TFIIE consists of two subunits, alpha and beta, that form a heterodimer and binds to TFIIF, TFIIIB, RNA polymerase II and DNA and serves to recruit TFIIH to the promoter (Flores et al., 1989; Ohkuma et al., 1991; Peterson et al., 1991; Sumimoto et al., 1991; Thomas and Chiang, 2006; Watanabe et al., 2003). TFIIH possesses certain enzymatic activities that are essential for transcription including a DNA-dependent ATPase, an ATP-dependent helicase, and a carboxyl terminal domain (CTD) kinase (Conaway and Conaway, 1989; Drapkin et al., 1994; Feaver et al., 1991; Lu et al., 1992; Roy et al., 1994; Thomas and Chiang, 2006). This ATPase and helicase activity is required for transcriptional initiation and for RNA polymerase II promoter clearance. Clearance is achieved by phosphorylating serine 5 of the carboxyl terminal domain; this marks the transition from transcriptional initiation to transcriptional elongation (Choe et al., 2014; Lu et al., 1992). Once the polymerase has cleared the promoter TFIIH is no longer needed for transcription and leaves the complex (Goodrich and Tjian, 1994).

The majority of genes require more than just the general transcription factors to ensure that they are transcribed efficiently in the proper time and place. This optimization of transcriptional efficiency is mediated by *cis*-regulatory elements. A *cis*-regulatory element can be any nucleotide sequence that is found on the same molecule as the gene that it regulates. Currently there are three known types of *cis*-regulatory elements: enhancers, silencers, and insulators (Bell et al., 2001; Woolfe et al., 2005). All of these elements

function by containing binding sites for specific transcription factors and then exerting some influence over the rate of transcription of their target gene by interacting with other chromatin bound proteins or the preinitiation complex (Popham et al., 1989; Szutorisz et al., 2005).

All of these regulatory elements are non-protein coding, meaning they are not transcribed and eventually translated into a protein. These elements can be found either upstream or downstream of the target gene's promoter or even in the introns of other nearby genes (Davidson, 2006; Dutton et al., 2008; Echelard et al., 1994; Kikuta et al., 2007; Valverde-Garduno et al., 2004). Despite the fact that these regions are non-coding they are still highly conserved between evolutionarily diverse species, most likely due to their importance in directing gene expression (Engstrom et al., 2007; Navratilova and Becker, 2009)

Enhancer elements are bound by specific transcription factors, or activator proteins, that increase the efficiency of the transcription of the genes with which they are associated at specific times and in specific regions of the organism (Fiering et al., 2000; Popham et al., 1989; Walters et al., 1995). The proteins that bind the enhancer are thought to cause the DNA to fold and bend, and the enhancer and associated proteins loop around to the promoter (Blackwood and Kadonaga, 1998; Gibcus and Dekker, 2013; Tolhuis et al., 2002). Then, with the aid of a protein complex known as mediator, the enhancer and its proteins bind the preinitiation complex at the promoter (Blackwood and Kadonaga, 1998; Gibcus and Dekker, 2013; Tolhuis et al., 2002). Another proposed model for enhancer-promoter interaction is the DNA scanning model where proteins bind the enhancer and then creep along the DNA until they find a compatible promoter (Blackwood and Kadonaga, 1998). While this model works very well with proposed insulator models, there are flaws. This model does not explain how

an enhancer could activate transcription from a tailed hairpin extending outwards from the DNA strand, or how an enhancer on one chromosome could interact with a promoter on separate, paired chromosome as in transvection (Blackwood and Kadonag, 1998). Therefore a combinatorial method of enhancer function has been proposed, known as facilitated tracking. This model suggests that the enhancer binding protein complex does scan the DNA but does so in small jumps, utilizing the strengths of both the looping and scanning models (Blackwood and Kadonag, 1998). The additional binding of these other proteins increases the binding affinity of the polymerase to the DNA strand and thus increases the efficiency of transcription (Blackwood and Kadonaga, 1998). Enhancers can also help to remodel the chromatin and expose genes for transcription (Blackwood and Kadonaga, 1998). Some enhancers control multiple genes, but regardless, enhancers ensure that specific genes are expressed in specific spatial and temporal patterns in an organism (Blackwood and Kadonaga, 1998).

Silencers on the other hand serve to decrease and suppress gene expression in certain tissues (Alberts, 2008; Brand et al., 1985; Clark and Docherty, 1993). While enhancers are bound by activators and increase the rate of gene transcription, silencers bind aptly named suppressor proteins that stop or decrease the rate of transcription of target genes by either interacting with other transcriptional proteins or by blocking enhancer regions (Clark and Docherty, 1993).

Enhancers and silencers can also affect the chromatin structure of DNA. Eukaryotic DNA is packaged into chromatin, which includes it being coiled around proteins known as histones. Which need to be methylated and/or acetylated to expose genes for transcription (Grunstein, 1997). The activators and repressors that bind enhancers and silencers can



mediate how tightly or loosely the chromatin is wrapped around these histones by recruiting certain enzymes to either methylate or acetylate them (Struhl, 1999).

The first two *cis*-regulatory elements discussed, enhancers and silencers, dictate when and where certain protein coding genes are expressed by moderating chromatin structure and polymerase binding. The third type of *cis*-regulatory element, known as insulators, serves to moderate by ensuring the first two types do not act upon the incorrect genes. Without the presence of insulators, enhancers and silencers could affect other non-target genes nearby and lead to incorrect spatiotemporal gene expression. For example, insulators can help prevent transgenic animals from expressing a transgene phenotype that is dictated by positional effects, they do this by preventing nearby enhancers and silencers from acting on the transgene (Bell et al., 2001). How insulators accomplish this function is not completely understood, however there are a few models that propose a mechanism. The first proposed mechanism, known as the “Promoter Decoy” model, suggests that insulators act as mock promoters that bind various proteins and “trick” the enhancers into thinking that the insulator is a promoter (Gerasimova and Corces, 2001; Geyer, 1997). While this seems to hold true in yeast, whose insulators do contain regions that can bind certain promoter binding proteins, there is no definitive supportive evidence for this model in other eukaryotes (Gerasimova and Corces, 2001; Geyer, 1997). The second model proposes that an insulator that lies between an enhancer and a promoter can bind specific boundary complexes flanking the entire region (Gerasimova and Corces, 2001). By doing so the insulator theoretically forms two adjacent loops of DNA, one containing the enhancer and the other the promoter, which then prevents the two regions from interacting. This is currently the more supported mechanism for insulator function. These two models only pertain to enhancers, not silencers. The proposed

model for insulator mediated silencer repression is that they form boundaries that prevent areas of repressive chromatin from overlapping with that of an actively transcribing region (Gerasimova and Corces, 2001).

Like insulators, DNA methylation can also regulate enhancer and silencer activity. By adding methyl groups to cytosine residues within the *cis*-regulatory region, activators and repressors can no longer bind, rendering the enhancer or silencer inactive (Razin and Riggs, 1980; Rombauts et al., 2003; Weber et al., 2007). This methylation of the DNA is a heritable mark that can be passed down from parent to offspring via the enzyme methyltransferase, which methylates the newly synthesized strand of DNA in the same places as the template. The cytosine residues that are methylated are often found clustered together in what are known as CpG islands (Elgen and Reisine, 2013).

The regulation of gene expression helps dictate an organism's body plan. The body can be organized into three main axes, anterior-posterior (AP), dorsal-ventral (DV) and left-right (LR). The AP axis is the axis that runs the length of the organism and is typically the axis around which bilateral symmetry revolves. The anterior end of the organism is the most forward portion, where the head is located and the posterior is the most rear and generally terminates in a tail or anus. The DV axis on the other hand is the axis that dictates "back and belly", with the dorsal side corresponding to the back or top side of an organism and the ventral side referring to the bottom. The identities of these axes are often regulated by proteins that contain a homeodomain, which consists of three alpha helices, generally totaling 60 amino acids in size (Gehring, 1987; Gehring, 1993; McGinnis et al., 1984). This domain and its family members are typically found in proteins that are involved in embryonic development (Gehring, 1987; Gehring, 1993; McGinnis et al., 1984). Homeodomains

typically bind to DNA at a short palindromic repeat of nucleotides that often include TAAT with its helix-turn-helix DNA binding motif (Otting et al., 1998). Genes that encode homeodomain containing proteins are known as homeobox genes, and were first discovered in *Drosophila melanogaster*, the fruit fly. They were discovered when scientists noted that mutations in certain genes caused homeotic mutations in the fly where one body segment takes the identity of another, hence the name homeobox gene and homeodomain for the protein region it encodes (McGinnis et al., 1994; Morata and Lawrence, 1977). These particular mutated genes were later named the *Hox* genes, a set of important homeobox genes responsible for patterning in developing embryos that are now often included in the category of “tool-kit genes”. Since their discovery, the *Hox* genes have been found in both vertebrate and invertebrate organisms (Balavoine et al., 2002; Lemons and McGinnis, 2006).

Numerous other homeobox genes other than *Hox* have also been subsequently discovered and grouped into super classes based on sequence identity. One of these superclasses is known as the three amino acid loop extension (TALE) superclass. Often, TALE homeodomain proteins are found acting as cofactors for various *Hox* proteins during development (Choe et al., 2014). The proteins in this superclass differ from a typical homeodomain by the presence of an extra three amino acids within the loop between alpha helix 1 and 2 of their homeodomains. The amino acids that form this loop extension are typically proline-tyrosine-proline and make up residues 24-26 of the homeodomain (Burglin, 1997). While this is the most distinct difference between the TALE and standard homeodomains, there are also a few other differences in what is normally a very highly conserved amino acid sequence. One of these differences is in the composition of residues 16 and 20 of the TALE homeodomain. In the TALE class there can be a leucine, methionine,

phenylalanine, cysteine or serine at residue 16, which is typically a leucine residue in a typical homeodomain. Position 20 in the TALE homeodomain can be either a phenylalanine, tryptophan, leucine or methionine, but is typically either a serine or threonine normally (Burglin, 1997). Variation can also be found in the amino acid sequence of helix 3, the alpha helix that lies in the major groove of DNA (Burglin, 1997). Normally a polar residue at position 50 is required for proper DNA binding, but at residue 50 of the TALE homeodomain there is instead a small non-polar residue (Burglin, 1997). This change in the binding helix suggests that the nature of TALE/DNA binding is different from that of typical homeodomains like those found in Hox proteins (Burglin, 1997).

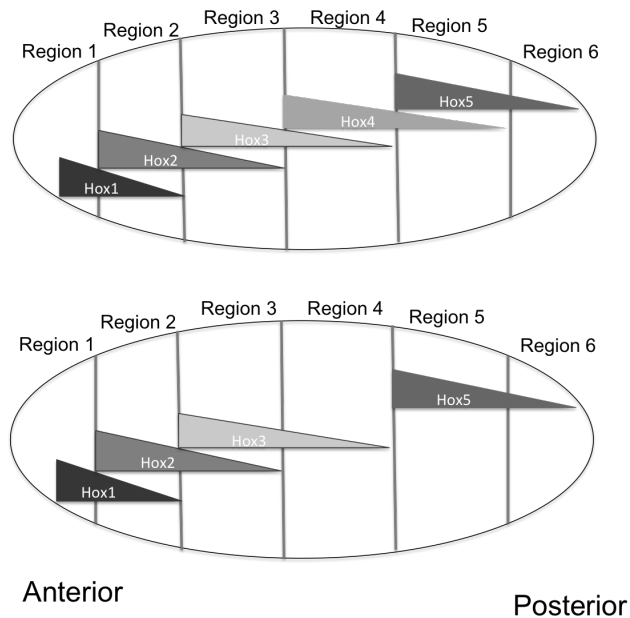
TALE homeodomain proteins can be divided into multiple subclasses, in animalia for instance, there are five classes: MEIS, TGIF, PBC, IRO and MKX (Burglin, 1997; Mukherjee and Burglin, 2007). The TALE proteins M-ATYP and CUP have been identified in fungi, and KNOX and BELL have been identified in plants. Interestingly, plants lack Hox proteins (Burglin, 1997; Lemons and McGinnis, 2006).

The most well-known homeodomain proteins are the Hox proteins of the ANTP class. Hox plays an important role in anterior-posterior (A/P) axis formation as evident by the presence of *Hox* genes in the genome of every animal species that have been examined (Holland and Garcia-Fernandez, 1996; Lemons and McGinnis, 2006). Typically *Hox* genes are found as linear clusters in genomes and exhibit spatial and temporal colinearity (Duboule, 1994; Lemons and McGinnis, 2006). This means that location of the *Hox* gene within the cluster determines where and when it is expressed during an organism's development. The *Hox* genes located closer to the 3' end of the cluster are expressed earlier and more anteriorly in the embryo than those that are found closer to the 5' end (Duboule, 1994; Lemons and

McGinni, 2006). The expression of these genes follows a pattern of a hard anterior border of expression followed by a decreasing gradient posteriorly, overlapping with the genes expressed more posteriorly to them (Fig. 1). This means that while multiple *Hox* genes are expressed in each region, only one of these genes however defines the region (Duboule, 1994). This defining gene is whichever *Hox* gene present that is found most closest to the 5' end of the cluster in that region (Fig. 1) (Duboule, 1994). This also means that if due to a mutation, a more posterior gene is expressed more anteriorly it will define that region instead of the normal gene (Duboule, 1994). This posterior expression gradient is also the cause of another type of homeotic mutation known as anteriorization. Anteriorization is when the *Hox* gene that normally defines a region is knocked out so the gene anterior to it in the cluster then dictates the segments fate instead (Gonzalez-Reyes and Morata, 1990). This occurs due to the expression patterns of *Hox* genes and their posterior expression gradients. The gene immediately anterior is now the most posterior gene in the region and now defines the region according to posterior prevalence (Fig. 1) (Gonzalez-Reyes and Morata, 1990). Without the more posterior *Hox* gene present, the gradient allows for the more anterior gene to function.

The number of the clusters and how many *Hox* genes they possess varies between different species, but the orthologs are typically located in similar locations within the clusters relative to their neighboring *Hox* genes (Amores et al., 1998; Lemons and McGinnis, 2006; Prince et al., 1998). In most tetrapods there are four, well conserved, *Hox* clusters (Fig. 2) (Greer et al., 2000; Lemons and McGinnis 2006). These four *Hox* clusters have been given the names *HoxA*, *HoxB*, *HoxC* and *HoxD* and each gene within each cluster has been given a paralog number consisting of the letter of what cluster it is found in along with what number *Hox* gene it is (Greer et al., 2000; Lemons and McGinnis, 2006). Further, each one

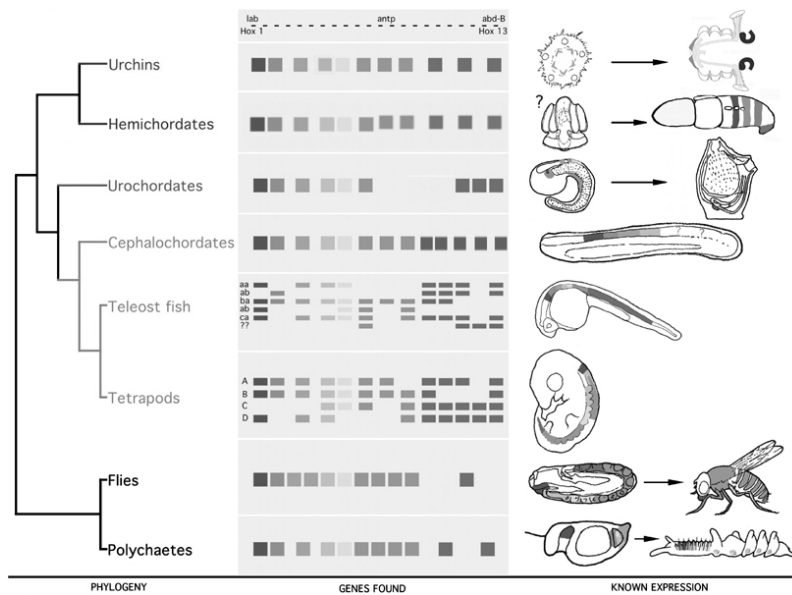
of these clusters contains a subset of the 13 different *Hox* genes, but no one cluster contains them all (Fig. 2) (Greer et al., 2000; Lemons and McGinnis 2006).



**Fig. 1. Diagram of Hox Expression.** This figure generally illustrates *Hox* expression in a developing embryo. In the top figure *Hox* genes are expressed highly in the region they define and then decrease in a gradient into the posterior segments. Each *Hox* gene defines the region associated with its number; for example, *Hox1* is defining region 1, region 6 is also defined by region 5 because it is the only *Hox* gene present. The bottom image displays an embryo that is missing the gene that defines segment 4, thus leaving *Hox3* to define that region.

Invertebrate organisms like *Drosophila melanogaster*, however, have only one *Hox* cluster consisting of eight different *Hox* genes (Fig. 2) (Amores et al., 1998; Holland and Garcia-Fernandez, 1996; Wagner et al., 2003). It is hypothesized that this discrepancy between vertebrates and invertebrates is due to the occurrence of multiple gene and genome

duplication events during metazoan evolution (Amores et al., 1998). During this time certain individual *Hox* genes underwent gene duplication events to increase the total number of *Hox* genes within a cluster from 8 to 13. As vertebrates evolved some of these genes were lost, most likely from non-functionalization, explaining why no one cluster has all 13 genes. These duplication events are thought to have played a role in the increasing complexity of the newly evolving animal body types, probably because an increased number of genes allowed for new patterns in development (Hokamp et al., 2003; Holland and Garcia-Fernandez, 1996; Venkatesh, 2003).



**Fig. 2. Diagram of Hox Gene Clusters in Various Organisms.** A diagram of the different *Hox* clusters in different taxa of organisms. Each colored block represents a different *Hox* gene and genes of the same color are paralogous and homologous to each other. Each row of shaded blocks represents a separate cluster. The spatial expression on the different genes is shown on the embryos on the right of the figure (Swalla, 2006).

Not all vertebrates have four *Hox* clusters, however ray-finned fishes or teleosts possess seven *Hox* clusters totaling 49 *Hox* genes (Amores et al., 1998; Brunet et al., 2006; Prince et al., 1998; Prohaska and Stalder, 2004). This is hypothesized to have occurred after teleosts, the lineage which includes zebrafish, underwent an additional whole genome duplication event when they diverged from the lineage which would give rise to tetrapods (Amores et al., 1998; Brunet et al., 2006; Prohaska and Stalder, 2004). Sometime after this, one of the now eight *Hox* clusters was lost, presumably because of nonfunctionalization resulting in the seven clusters now seen in teleosts (Amores et al., 1998; Brunet et al., 2006; Prohaska and Stalder, 2004).

As mentioned previously, the specific order that *Hox* genes are expressed follows a pattern where each gene has a distinct border of anterior expression with a gradual dissipation towards the posterior, which is at least part of the reason why we see anteriorization mutants (Fig. 1). This expression pattern is regulated by a number of different chemical signals such as retinoic acid (RA) or caudal (*Cdx*) protein, which acts as both an activator and repressor protein depending on which *Hox* gene is being regulated (Davidson and Zon, 2006; Mazzone et al., 2013; Shimizu et al., 2005; Shimizu et al., 2006). *Cdx* is another ANTP class homeobox gene and part of the *parahox* cluster consisting of *Xlox*, *Cdx*, and *Gsx* (Brooke et al., 1998). It is hypothesized that this cluster arose from an ancient *Hox* cluster that then evolved into a group of related homeobox genes (Brooke et al., 1998). In metazoans *Cdx* is expressed posteriorly in the embryo and is responsible for posterior identity (Gamer and Wright, 1993; Marom et al., 1997; Mlodzik et al., 1985; Skromne et al., 2007). *Cdx* paralogs have been discovered in vertebrate organisms: *CDX1*, *CDX2*, and *CDX4* in mammals, and *Cdx1a* and *Cdx4* in zebrafish (Davidson and Zon, 2006; Wingert et



al., 2007). In mice, *Cdx* knockouts result in posterior truncations with increasing severity depending on which and how many *Cdx* genes are knocked out. If all three *Cdx* genes are knocked out, the murine embryo dies at roughly 3.5 days post coitus (dpc), partially due to the function of *Cdx2* in placenta development (Lengerke and Daley, 2012). The *Cdx* genes typically exhibit overlapping expression. For instance, *Cdx1/4* have been implicated in both hindbrain and spinal cord development and *Cdx1/2* are both seen in varying levels during intestinal development and in the adult intestine to help regulate homeostasis (Grainger et al., 2013; Kromne et al., 2007; Shimizu et al., 2006). *Cdx4* has been shown to be important for proper development of the pancreas and the pronephros, an early developmental stage of the adult kidney (Kinkel et al., 2008; Wingert et al., 2007). Zebrafish that are deficient in *cdx4* have posteriorly shifted pancreas, liver and small intestines, and also have an increase in the number of beta cells that store and release insulin (Kinkel et al., 2008). Hematopoiesis is also *Cdx4* dependent. Zebrafish embryos that are deficient in *cdx4* exhibit a severe deficit in blood cells, and when both *cdx1a* and *cdx4* are knocked down the organism completely fails to produce blood cells (Davidson and Zon, 2006; Lengerke and Daley, 2012). *Cdx4* expression has been observed in mice starting at 7.5 dpc and is found in the primitive streak, neurectoderm, presomitic/lateral plate mesoderm, and hindgut endoderm. More staggered expression is observed in the anterior boundary of the paraxial mesoderm, the boundary changes dependent on which germ layer *Cdx* is found (Gamer and Wright, 1993). In zebrafish *cdx* genes have been shown to be responsible for mediating the boundary between hindbrain and spinal column and can be expressed as early as 12.5 hours post fertilization (hpf) in the posterior of the embryo (Kinkel et al., 2008; Shimizu et al., 2006; Skromne et al., 2007).

Another important protein for early *Hox* expression is nucleoporin 98 (Nup98) (Pascual-Garcia et al., 2014). Most nucleoporins form channels in the nuclear envelope but some have been shown to directly bind DNA (Garcia et al., 2014). Nup98, with the help of MBD-R2 and Trx, has been shown to drive expression of the *Hox* gene *ultrabithorax* in *Drosophila* (Garcia et al., 2014).

Hox proteins act as transcription factors, both as activators and repressors during anterior-posterior axis development (Dorn et al., 1994; Gehring et al., 1994; Nishimoto et al., 2014). As stated earlier, the homeodomain serves as a DNA binding domain for various *cis*-regulatory elements (Dorn et al., 1994; Gehring, 1993; Gehring et al., 1994). The third alpha helix of the homeodomain binds DNA in the major groove via hydrogen bonds and van der Waals contacts in a fashion unique to the homeodomain (Brennan and Matthews, 1989). Homeodomain-mediated DNA binding, however, is neither specific nor efficient (Ekker et al., 1994; Mann and Affolter, 1998; Moens and Selleri, 2006; Waskiewicz et al., 2001). One way Hox binding specificity is achieved is through multiple low affinity sites (Crocker et al., 2015). Studies in *D. melanogaster* found clusters of low affinity Hox DNA-binding sites known as homotypic clusters that interact with the *Hox* gene *ultrabithorax (ubx)* (Crocker et al., 2015). Studies with transgenic and mutant fly embryos along with Selex-seq testing showed that low affinity sites lent themselves to greater specificity while high affinity sites tended to be less specific, suggesting that these low affinity homotypic clusters are aiding Hox in binding specificity (Crocker et al., 2015). Hox proteins must also rely on various cofactors to ensure optimal DNA binding and specificity (Ekker et al., 1994; Mann and Affolter, 1998; Moens and Selleri, 2006; Waskiewicz et al., 2001).

In support of this necessity for co-factors, analysis of the amino acid residues of various Hox proteins showed that the greatest variability is not in their DNA binding domain but in the domains responsible for protein-protein interactions (Mann and Affolter, 1998; Sharkey et al., 1997). These cofactors increase the number of DNA binding sites in the protein complex and thus add specificity to the Hox:cofactor complex when binding DNA (Mann and Affolter, 1998; Moens and Selleri, 2006). Hox interacts with many different cofactors to give rise to these protein complexes, many of which belong to the TALE superclass of homeodomain proteins. Studies have shown that these TALE cofactors are responsible for much of the gene activation attributed to Hox complexes. The TALE cofactors first act to recruit the protein CBP (CREB-binding protein), which acetylates the histones surrounding the target gene, exposes the promoter, and recruits RNA polymerase II and P-TEFb, a protein which phosphorylates serine-2 on RNA polymerase II to activate it. Hox's role is to initiate phosphorylation of the enzyme in order for transcription to begin (Choe et al., 2014). This Hox:TALE complex in vertebrates consists of Hox and the TALE cofactors Pbx and Meis where Hox binds Pbx via the Hox N-terminal pentapeptide motif and Pbx binds Meis through its Meinox domain (Choe et al., 2002; Knoepfler and Kamps, 1995; Shanmugam et al., 1999).

The *Meis* genes were first discovered in BXH-2 mice, a model system for acute myeloid leukemia caused by viral genome integration into a previously uncharacterized region of DNA (Moskow et al., 1995). Upon further characterization of this region it was discovered that the insertion was within a previously undescribed gene that was named *Meis* (for Myeloid Ecotropic viral Integration Site) (Moskow et al., 1995). After this discovery scientists searched to see if more *Meis* genes were present in the genome. To accomplish

this, the entire mouse and *Xenopus* genomes were probed with with the murine *Meis1* homeobox using Southern blotting (Steelman et al., 1997). In this study, they discovered two more *Meis* genes, which at the time they named the *Mrg* genes (Meis related gene) (Steelman et al., 1997). Since then these genes and their homologs have been found in different vertebrate animals such as mice, humans, chickens, zebrafish and frogs (Bomgardner et al., 2003; Burglin, 1997; Chong et al., 2009; Coy and Borycki, 2010; Geerts et al., 2005; Lawrence et al., 1999; Mercader et al., 2000; Nakamura et al., 1996; Sagerstrom et al., 2001; Sanchez-Guardado et al., 2011; Smith et al., 1997; Steelman et al., 1997; Waskiewicz et al., 2001; Zerucha and Prince, 2001). A much later study discovered the presence of a fourth *Meis* gene that is only present in teleosts such as zebrafish (Waskiewicz et al., 2001). A homolog of the *Meis* genes is found in invertebrate organisms and is known as *homothorax* (*hth*) and serves a similar function (Choe et al., 2002, Irimia et al., 2011, Ryoo et al., 1999). A function shared between *Meis* and all of its family members, like *Prep1*, is to serve as a Hox cofactor and help to transcribe other proteins (Choe et al., 2002).

The various *Meis* genes are expressed throughout the developing embryo and are important for proper formation of rhombomeres 3 and 4, which are hindbrain precursors (Biemar et al., 2001; Cecconi et al., 1997; Choe et al., 2002; Santos et al., 2010; Waskiewicz et al., 2001; Zerucha and Prince, 2001). In zebrafish, the *meis* genes have been shown to be important in the segmentation of the hindbrain by increasing the levels of *Pbx* protein (Waskiewicz et al., 2001).

*Meis1*, the first *Meis* gene discovered, has been studied in the embryos of many different vertebrate species and is expressed in many different areas of the embryo. In mice, *Meis1* expression has been seen in the developing eye in the region of the retina (Bessa et al.,

2008; Hisa et al., 2004). This expression pattern is similar to that of *Pax6* in the eye (Zhang et al., 2002). This overlapping expression may be because *Pax6* expression is *Meis* dependent (Zhang et al., 2002). It is most likely that *Meis1* and some unknown cofactors bind one of the *Pax6* enhancers that controls eye formation. This would explain the eye mutations seen in *Meis1* mutants (Zhang et al., 2002). While *Meis* typically functions in a complex with *Pbx*, assays did not show its presence along side *Meis* when it is found at the *Pax* enhancer, suggesting it maybe be using a different cofactor (Zhang et al., 2002). Mice deficient in *Meis1* develop partially duplicated retinas and smaller lenses than those with proper gene function (Hisa et al., 2004). This is thought to happen because *Meis1* is a regulator of *cyclin D1* and *c-myc*, which are major growth 1 (G1) cell cycle regulators important for cell proliferation in the retina. By controlling the levels of these genes, *Meis1* can affect retinal formation (Bessa et al., 2008). *Meis1* mutants also develop malformed capillaries and suffer from massive hemorrhaging, eventually dying by embryonic day 14.5 (Hisa et al., 2004; Nakamura, 2005). This hemorrhaging could be magnified by the fact that *Meis1* deficient mouse embryos also lack megakaryocytes, a specialized cell type that is important for blood clotting (Hisa et al., 2004). Similar studies in zebrafish have shown that *meis1b* knockdowns, severely affected the function of *hoxd4a*, which resulted in poor vascularization and definitive hematopoiesis (Amali et al., 2013). This effect on hematopoiesis is not surprising considering that *Meis1* has previously been shown to have a role in the genesis of myeloid leukemias (Moskow et al., 1995). *Meis1* has been implicated in 15% of tumors in BXH-2 mice when they have been affected by viral integration, and this has led to acute myeloid leukemias from a decrease in *Meis1* expression (Moskow et al., 1995; Nakamura, 2005). However, the mutated *Meis1* protein alone is not enough to cause

leukemia in mice, it must be apart of the Hox:TALE complex in order to cause disease (Shen et al., 1999).

Besides expression in the developing eye and circulatory system, *Meis1* has also been seen in low levels in the ventricular zone of the telencephalon of mouse embryos starting at embryonic stage (E) 10.5 (Toresson et al., 2000). In the telencephalon, the structure that will become the cerebrum, the highest amounts of *Meis1* mRNA and protein have been found in the subventricular zone and mantel region of the ventral telencephalon (Toresson et al., 2000). *Meis1* is also partially responsible for proper limb development. Its expression is seen proximal to where the limb bud develops and is driven by the cell signaling molecule retinoic acid (Mercader et al., 2000). When *Meis1* is ectopically expressed near the distal end of the of the limb bud, *Meis1* was shown to proximilize limb tissues (Mercader et al., 2000).

When *Meis1* is overexpressed in tissues it has been shown to have a caspase-dependent apoptotic affect. This effect on apoptosis has been observed in both hematopoietic and non-hematopoeitic cells in human patients and mice (Nakamura, 2005). This apoptotic activity resembles that of PBX1 with the exception that HoxA9 is capable of suppressing Meis1 driven apoptosis (Nakamura, 2005).

As mentioned previously, after ray-finned teleosts diverged the lineage that would give rise to tetrapods their genomes underwent a duplication event (Amores et al., 1998). Over time however, the second *meis1* gene, *meis1a*, was lost (Irimia et al., 2011). Loss of function mutations in zebrafish *meis1b* result in a disorganization of the hindbrain compartments. In addition, a loss of vasculature has also been observed in zebrafish *meis1b*

knockdowns, similar to what is seen in murine *Meis1* mutants (Amali et al., 2013; Waskiewicz et al., 2001).

*Meis2* displays a similar pattern of expression to that of *Meis1* with some differences (Coy and Borycki, 2010). Early in mouse development *Meis2* protein can be found in the lateral somitic compartment and its derivatives, such as in the lateral muscle (Cecconi et al., 1997). As development progresses *Meis2* protein can be found in the overlying cells of the paraxial mesoderm in the dorso-ectodermal area at E8. By E10 it can be found in the developing forebrain, midbrain, hindbrain and spinal column (Cecconi et al., 1997). This *Meis2* midbrain activity has been shown, in part, to be regulated by the polycomb group of proteins, which bind to a 3' repressor sequence known as the RING1B-binding site (RBS) near the *Meis2* gene (Kondo et al., 2013). When *Meis2* is expressed, a specific midbrain enhancer loops around to the RBS to remove the repressor protein, which activates *Meis2* expression (Kondo et al., 2013). Similar to the expression of *Meis1* in the developing brain, in E10.5 mouse embryos *Meis2* can be found in high levels in the cells beneath the ventricular zone, lateral to the medial ganglionic eminence of the telencephalon (Toresson et al., 2000). The highest concentration of protein though is seen in both the subventricular zone and the mantle regions of the telencephalon (Toresson et al., 2000).

*Meis2* expression overlaps with that of *Pax6* in the eye, and actually drives *Pax6* expression since *Meis2* binds to one of *Pax6*'s enhancer elements (Zhang et al., 2002). If the amount of *Meis2* is increased or decreased then *Pax6* protein levels are affected similarly (Zhang et al., 2002). More recently *Meis2* and *Pax6* have been shown to play a role in olfactory bulb development in mice, along with *Dlx2* (Agoston et al., 2014).

*Meis2* has also been shown to be important for proper limb development in vertebrates (Capdevila et al., 1999). Retinoic acid activates *Meis2* transcription and helps ensure that it is expressed properly in the proximal end of the developing limb. Improper expression has been shown to cause severe limb defects in chickens resulting in the loss of digits and improper zeugopod formation (Capdevila et al., 1999; Mercader et al., 2000). Distally, *Meis2* functions to ensure proper digit formation by mediating the apoptosis of the interdigital tissue in mice and bats (Dai et al., 2014). In mice, expression is only seen in E15 embryos in the autopods of both fore and hindlimbs while in bats *Meis2* is strongly expressed in the developing wing membranes, suggesting that *Meis2* inhibits apoptosis (Dai et al., 2014).

A study published in 2013 examined the effects of *MEIS2* haploinsufficiency in humans and found that 7 out of 9 individuals with only one functioning copy of *MEIS2* possessed mild to severe cleft palates, suggesting that *MEIS2* is important for palate closure (Johansson et al., 2013). Also, all of the individuals examined possessed some level of learning disability with the worst falling into the mild intellectual disability range (Johansson et al., 2013). While the deficient individuals were not dysmorphic, the deficiency seemed to have other effects on the craniofacial structure, typically resulting in overlapping facial features like a broad forehead and finely arched eyebrows (Johansson et al., 2013).

As observed for *Meis1*, *Meis2* was also subject to the teleost duplication event, resulting in two *meis2* genes, *meis2a* and *meis2b* (Amores et al., 1998). Expression of the two zebrafish *meis2* genes can be first seen during the onset of gastrulation (Biemar et al., 2001; Zerucha and Prince, 2001). Later, expression can be seen in the central nervous system with the largest amount of expression in the hindbrain (Biemar et al., 2001). At the



same time *meis2a* and *meis2b* can also be seen in the isthmus, along the spinal cord and the lateral mesoderm (Biemar et al., 2001). This expression pattern is similar to that seen in mice but is subdivided between *meis2a* and *meis2b* with some overlap between them, (Santos et al., 2010; Zerucha and Prince, 2001). Specifically, *meis2b* expression is visible in the areas where the hindbrain will eventually arise from the embryo at around 60% epiboly (Zerucha and Prince, 2001). This expression spreads when the embryo is nearing 100% epiboly and by the 6 somite stage, at around 12 hours post fertilization, expression can be seen in the forebrain, midbrain and in the developing hindbrain in rhombomeres 2 and 3 (Zerucha and Prince, 2001). Expression of *meis2b* can also be seen in the anterior boundaries of the somites after segmentation has happened and by 14 hours expression has started to weaken in rhombomere 2 and has increased in rhombomere 4 (Zerucha and Prince, 2001). As development of the zebrafish embryo progresses *meis2b* expression remains localized to the developing brain in slightly different patterns but also appears in the retina at 36 hours post fertilization (Zerucha and Prince, 2001). The expression of *meis2a* is similar to *meis2b*, but *meis2a* is also found in the areas of the limb buds and branchial arches (Coy and Borycki, 2010; Mercader et al., 1999; Mercader et al., 2000; Waskiewicz et al., 2001).

*Meis3* is expressed in multiple tissues in the developing embryo. In *Xenopus* it is first seen in a single strip of cells in the early neural plate, but as development of the nervous system progresses the signal becomes localized to rhombomeres 2-4 and to the anterior of the spinal cord (Salzberg et al., 1999). When *Meis3* is expressed ectopically the posterior sections of the brain expand and disrupt the formation of anterior ones such as the fore and midbrain (Salzberg et al., 1999). Ectopic *Meis3* expression demonstrated no effect on pan-neural markers like *Nrp* but did increase expression of posterior markers signifying that

*Meis3* is important for posterior neural formation (Salzberg et al., 1999). When *Meis3* protein levels are knocked down a loss of neural crest lineages was seen in the hindbrain of *Xenopus* specimens. Further study demonstrated that this activity is regulated by *Pax3* and *Zic1/4*, and that *Meis3* moderates expression of *HoxD1* (Gutkovich et al., 2010). In zebrafish *meis3* can change mid and forebrain cell fates to that of the hindbrain, with the help of *hoxb1b* and *pbx4* (Vlachakis et al., 2001). *Meis3* also plays an important role in insulin and non-insulin producing beta-cell survival in the pancreas. *Meis1/2* proteins were also detected in pancreatic islets but the expression of *Meis3* was threefold higher than either of the other two *Meis* genes (Liu et al., 2010). *Meis3* helps suppress apoptotic activity in these cells by regulating the signaling pathways that effect caspase-3 cleavage. Beta-cells that are deficient in *Meis3* typically die (Liu et al., 2010).

To date, little is known about *meis4*. Currently it has only been identified in zebrafish and it has two splice variants: *meis4.1a* and *meis4.1b* (Waskiewicz et al., 2001). All *Meis* genes have proven to be important for vertebrate development and that their protein products function as Hox cofactors. Typically these genes are involved with central nervous system development but have also been implicated in limb development, pancreas function, palate formation, eye development, vascularization and hematopoiesis, and in somitic cell lines. When individuals become deficient in MEIS they can suffer any number of defects with one of the most severe being acute myeloid leukemia proving *Meis* genes are essential for healthy development and may act as a tumor suppressor.

While much is known about how *Meis* regulates gene expression, relatively little is known about how *Meis* is regulated. Analysis of the genomic region surrounding the gene *Meis2* revealed the presence of four highly conserved non-coding elements in tetrapods,

which were named the *Meis2* downstream elements (m2de) 1-4 and are hypothesized to be *Meis2* enhancers (Wellington and Zerucha, unpublished). It has been found though that teleosts only possess one of these enhancers, m2de1. Research by the Zerucha lab has shown that at 48hpf m2de1 drives expression of a GFP reporter gene in the mid and hindbrain of zebrafish embryos (Barrett, unpublished thesis). Research by the Gelgar lab however reported an enhancer called 3288 that is contained within m2de1 that drove a more isolated area of expression posterior to rhombomere 5. The difference between the size of 3288 and m2de1 is less than 20bp on both the 3' and 5' ends. The purpose of this study was to further characterize m2de1 driven expression in zebrafish, and also compare it to that of 3288 and to 3288 with either the m2de1 3' or 5' end in addition. Here we suggest that all of the fragments of m2de1 drive the mid and hindbrain expression previously seen by m2de1. Full length m2de1 also drives expression in the trunk of the developing embryo.

## **Materials and Methods**

### *Zebrafish Husbandry*

Zebrafish were housed in a Marine Biotech Z-mod closed system (Aquatic Habitats, Apopka, FL). The system maintained a constant temperature of 27°C and a day night cycle consisting of a 14 hour light period and 10 hour dark period. Water pH and conductivity were monitored daily and maintained between 7.0 and 7.4 and 450 and 600 milliSiemens per meter (mS/m). Water hardness was also monitored once a month and kept between 120-200ppm. The fish that were kept in the system were of two different strains, the genetically defined line of AB\* and wild type fish obtained from a local pet shop (Pet Supplies Plus). Six adult fish were kept in 1L tanks typically consisting of three male fish and 3 female fish.

In order to acquire embryos for injections from pure lines, male and female fish were separated and placed into special breeder tanks with dividers (Aquatic Habitats). The divider allowed control of when the fish spawned and increased the breeding vigor of the fish. It was found that the most embryos were produced when dividers were removed immediately after the system lights turned on at 9AM. The released eggs fell through the mesh bottom of the breeding tank, into the space between it and the 1L tank, and were then fertilized by the male fish. Fertilized eggs were removed from the tank after approximately 15 minutes to ensure that they were still at the single cell

stage when injected. The eggs were first isolated by filtering them through a fine mesh, then any fecal matter present was removed by rinsing the eggs with RO water. The cleaned eggs were then placed into a small beaker to be sorted and then into a petri dish containing 0.3x Danieau buffer (58 mM NaCl, 0.7 mM KCl, 0.4 mM MgSO<sub>4</sub>, 0.6 mM Ca(NO<sub>3</sub>)<sub>2</sub>, 5 mM HEPES pH 7.6 for 1x solution). If the embryos were to be raised to adulthood and not imaged they were placed in a 1x Danieau/methylene blue solution (0.66µL 1% methylene blue per liter of 1x Danieau buffer) in order to prevent fungal growth. Methylene blue has been shown to create auto fluorescence in embryos so those embryos that are to be imaged are raised in pure Danieau buffer. Embryos were incubated at 28°C and cleaned daily to remove dead embryos. After 5 days the fish began feeding on dry food (ZM-50; Zeigler) twice a day and were transferred to a bowl in the incubator. After 2 weeks the fry were transferred into the system and placed into a 1L tank. This tank was filled with 40% 1x Danieau buffer and 60% system water and is placed under a very slow drip of system water to ensure circulation of oxygen, faster drips can often result in fry death. Fry were still fed twice a day with the size of the dry food increasing according to fish size (ZM-100, ZM-200, ZM-300, ZM-400; Zeigler). Once the fry were consuming ZM-200 they also begin a diet of 2-day-old live brine shrimp once a day in addition to their normal dry food diet. After 3-4 months the zebrafish reached sexual maturity and were fed adult dry food once a day and a single concentrated drop of live brine shrimp per pair of fish.

#### *General Molecular Cloning Techniques*

In order to clone plasmids they must first be transformed into bacterial cells. This was done by first adding 2µl (10pg-100ng) of plasmid to 50µl of chemically competent DH5- $\alpha$  *E. coli*. These cells were then incubated on ice for 30 minutes and then heat shocked at

42°C for 30 seconds. Cells were then placed immediately on ice for a few minutes and then 250µl of SOC medium (20g bacto-tryptone, 5g bacto-yeast extract, 0.5g NaCl, 20mM glucose) was added and the culture incubated in the shaker for 1.5 hours at 37°C and 200rpm. After the incubation the cells were plated in a 1<sup>-10</sup> and 1<sup>-100</sup> dilution on LB agar plates with the appropriate antibiotic for selection. Plates were incubated overnight at 37°C and then wrapped in parafilm and stored at 4°C.

To screen large numbers of colonies, plasmid DNA was isolated for analysis by a boil mini-prep. 1.5ml of liquid bacterial culture from an overnight incubation was centrifuged at maximum speed for 30 seconds to pellet the cells, the supernatant was decanted and the pellet then resuspended with 350µl of STET w/ lysozyme (20g sucrose, 12.5ml triton, 12.5 ml 1M Tris pH 8, 35ml 0.5M EDTA, 0.5mg/ml lysozyme). Resuspended cells were then placed in a boiling water bath for 30 seconds and then centrifuged again for 15 minutes at maximum speed. The pellet at the bottom was removed with a sterile toothpick and the DNA was then precipitated with 350µl of isopropanol. Samples were centrifuged at 4°C for 20 minutes, the isopropanol was discarded and the pellet washed with 0.5ml 70% ethanol. The resulting plasmid DNA was then spun down for 30 seconds at maximum speed, the ethanol discarded and the pellet dried in a speed vac to remove all of the alcohol. To remove any RNA, the DNA pellet was resuspended in 20µl of water and 0.4µl RNase-A (10mg/ml). This mixture was incubated at 65°C for 15 minutes and then at 37°C for another 15 minutes.

#### *Generation of Injection Cassettes*

In order to generate the three different expression cassettes of *dr-m2de1* fragments (3288, 3288-3', 3288-5'), *dr-m2de1* in the TOPO 2.1 vector (Invitrogen) was PCR amplified with the specific primers found in Table 1 using Taq Phusion (NEB Labs). PCR product was

then prepared for ligation into the TOPO 2.1 vector by first adenylating the fragments by incubating them at 72°C for 10 minutes with Standard Taq Polymerase (NEB Labs) and dATP. The fragments were then ligated into TOPO 2.1 by mixing 4.0µl of the adenylated fragments with 1.0µl of TOPO 2.1 and 1.0µl of a salt solution from the Invitrogen TOPO TA Cloning Kit (pCR 2.1-TOPO Vector) in a 1.5ml centrifuge tube and incubated at room temperature for 30 minutes.

**Table 1. PCR Primers for Generation of Enhancer Fragments**

| <b>Forward Primer</b> |              |                                 |
|-----------------------|--------------|---------------------------------|
| <b>Fragment</b>       | <b>Names</b> | <b>Sequence</b>                 |
| 3288                  | Dr3288-3     | GCATGTTGAGGGTTCGATT             |
| 3288-3'               | dr-m2de1-3   | GCTCATTATAAGGCCGTGCATG          |
| 3288-5'               | Dr3288-3     | GCATGTTGAGGGTTCGATT             |
| <b>Reverse Primer</b> |              |                                 |
| 3288                  | Dr3288-5     | AGGAGGTCGGGTTTAAAG              |
| 3288-3'               | Dr3288-5     | AGGAGGTCGGGTTTAAAG              |
| 3288-5                | dr-m2de1-5b  | TATACCATGGAGGTCGGGTTTAA<br>AGGA |

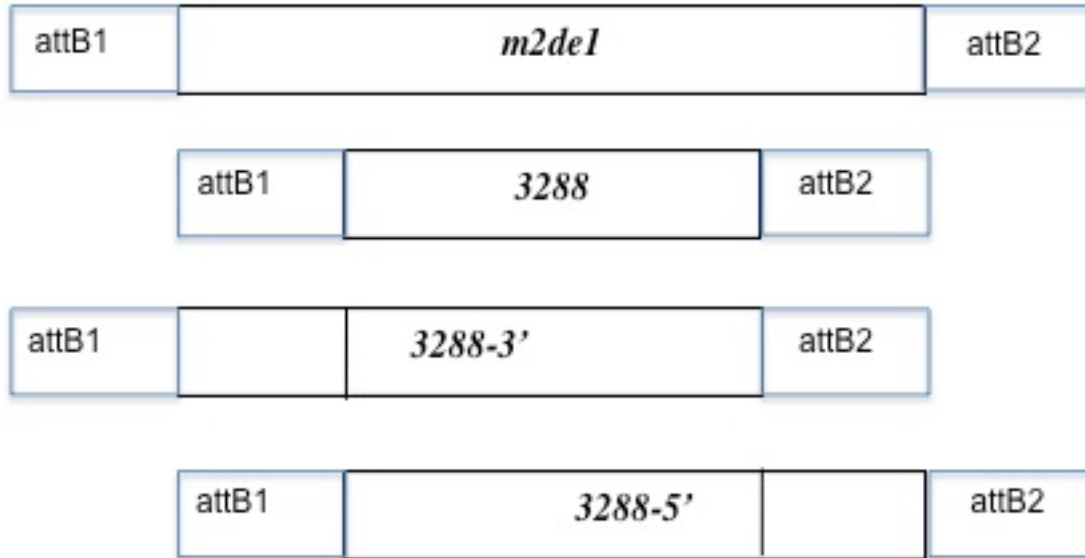
Ligated vectors were transformed into *DH5-α* chemically competent *E. coli* via heat shock and then plated on LB-ampicillin (100mg/ml) selective media previously spread with X-Gal (40mg/ml dimethylformamide). Plates were incubated at 37°C overnight and then examined; white colonies signified a successful ligation and were picked onto a fresh LB-ampicillin plate and incubated again 37°C overnight. Picked colonies were then further screened via boil mini-prep. The collected DNA was then digested with EcoR1 to screen for inserts of the correct size. One colony with the correctly ligated plasmid was then grown up in a 500ml culture of LB broth with ampicillin (100mg/ml) and the plasmid DNA isolated using the Qigen MaxiPrep kit following the manufacturer's protocol. This TOPO 2.1 vector was then used in the Gateway reaction, a process in which a target gene can be transferred into an injection vector via a two step process. The first step is for the gene of interest to be inserted into the pDONR221 expression vector. This reaction is known as the BP reaction and uses the attB sites and attP sites found in the TOPO 2.1 clones and pDONR221 plasmids to transfer the target gene into pDONR221 (Fig. 3,4). This reaction requires 25 f-mol of both pDONR221 and Topo 2.1 clones, and was calculated via the equations:

$$\text{ng Topo 2.1} = (25 \text{ fmoles}) \cdot (\text{HCNE size in base pairs}) \cdot (660 \text{ fg/fmoles}) \cdot (1 \text{ ng}/10^6 \text{ fg})$$

$$\text{ng pDONR221} = (25 \text{ fmoles}) \cdot (645 \text{ bp}) \cdot (660 \text{ fg/fmoles}) \cdot (1 \text{ ng}/10^6 \text{ fg})$$

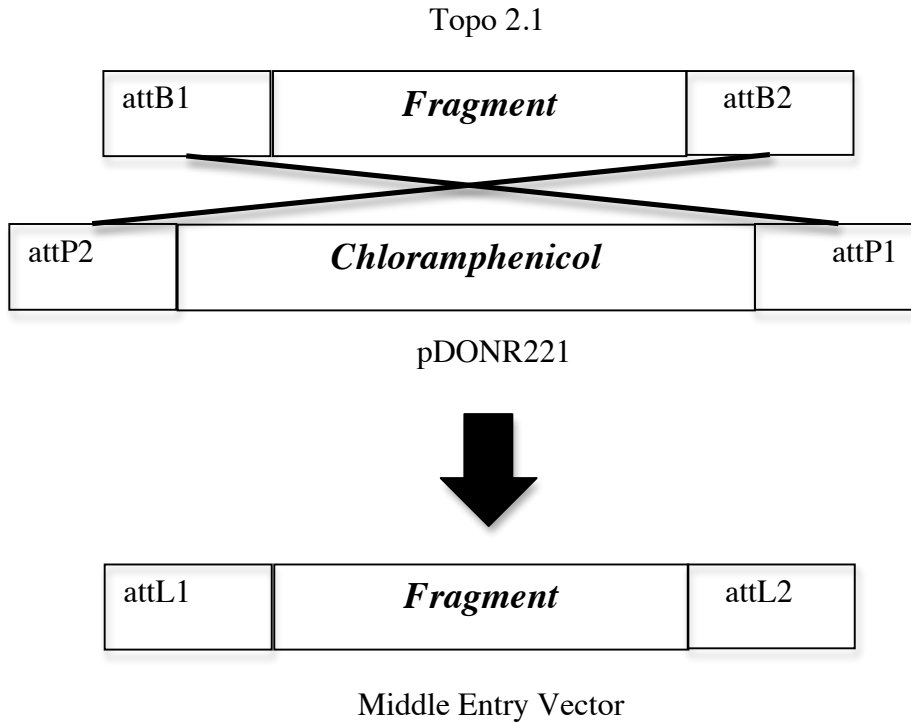
The calculated volume of both plasmids were then mixed in a 1.5ml centrifuge tube and incubated for 16 hours at room temperature. After this incubation, 1µl of proteinase K was added to each tube and incubated at 37°C in a water bath for 10 minutes. The plasmids were





**Fig. 3. Diagram of the m2de1 fragments and attB sites in Topo 2.1.** Diagram of the *m2de1* fragments flanked by the attB sites found in the Topo 2.1 Vector. The topmost image is the 3288 *dr-m2de1* segment and below it are the 3288-3' and 3288-5' *dr-m2de1* fragments.

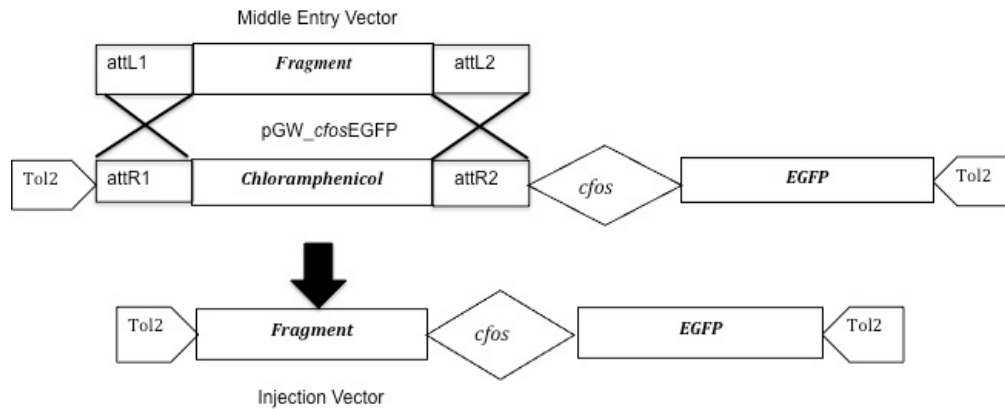
then separately transformed via heat shock into *DH5-α* and plated onto LB-kanamycin (50mg/ml) and incubated at 37°C overnight. Colonies were then picked onto a grid on two separate plates, another LB-kanamycin plate and a LB-kanamycin/chloramphenicol (50ug/ml kanamycin, 25ug/ml chloramphenicol) plate. If the BP reaction was successful then the chloramphenicol resistance gene in pDONR221 would be removed and there would be no growth on the chloramphenicol plate (Fig. 4). The colonies that did not grow on LB/kanamycin/chloramphenicol but did grow on LB/kanamycin were then isolated via the Qiagen Plasmid Plus Maxi Kit.



**Fig. 4. Diagram of Gateway BP Reaction.** The figure shows TOPO 2.1 on the top and pDONR221 on the bottom. The lines demonstrate how during the BP reaction that the attB sites and the attP sites will switch the enhancer fragment into the pDONR221 vector, replacing the chloramphenicol resistance gene. This creates the middle entry vector, which contains the enhancer fragment flanked by attL sites and containing a kanamycin resistance gene for selection.

Once the middle entry vector was created the enhancer fragment was then ready to undergo the LR reaction to insert it into the final expression vector, pGW\_*cfos*EGFP. The amount of each vector that was needed was calculated using the same equation used for the BP reaction with the exception that 10 fmoles of the middle entry vector and pGW\_*cfos*EGFP was required instead of 25 fmoles. The appropriate volume of each plasmid was then mixed in a 1.5mL centrifuge tube with 1X TE buffer and LR Clonase II and

incubated overnight at room temperature. The LR Clonase II catalyzes the insertion of the target gene into the final vector (Fig. 5).

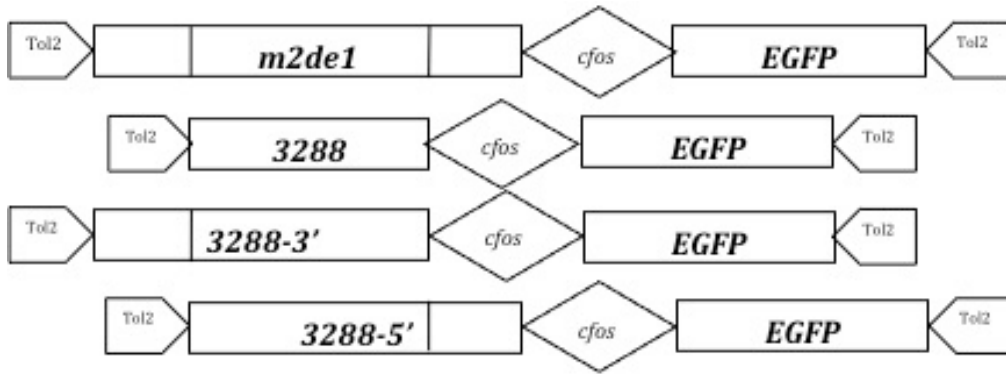


**Fig. 5. Diagram of Gateway LR Reaction.** The figure shows the middle entry vector and pGW\_cfosEGFP exchanging the enhancer fragment via the attL and attR sites. This creates the expression vector containing the enhancer fragments, the *cfos* minimal promoter, the EGFP reporter gene, and the flanking Tol2 sites.

This generated the final expression vector, inserting the enhancer fragments upstream of a *cfos* minimum promoter and the EGFP reporter gene (Fig. 6). The whole construct is flanked by a pair of Tol2 sites that allow the integration of the cassette into the embryonic genome during injections using Tol2 transposase mRNA.

#### *Injection of Zebrafish Embryos*

Embryos that were to be injected were collected from specialized breeder tanks roughly 15 minutes after they were laid. Embryos were rinsed with RO water and sorted in a beaker to remove any unfertilized embryos as described previously. Embryos were then lined up single file along the edge of a glass microscope slide fixed to the lid of a petri dish using a glass pipet. Then 125ng of injection vector along with 125ng of Tol2



**Fig. 6. Diagram of Enhancers in Injection Vector Flanked by Tol2.** A diagram of the three different injection cassettes as compared to the whole *dr-m2de1* enhancer. The arrows on either side of the constructs are the Tol2 sites, on the right of each construct is the EGFP reporter gene, then the *cfos* minimal promoter in the middle, and the enhancer fragments on the left. The vertical lines in the elements show the boundaries of the 3288 element in comparison to the other enhancers.

transposase mRNA were then mixed in a sterile 1.5 mL centrifuge tube and then brought up to 5 $\mu$ L with phenol red (0.5% in DPBS). This solution was kept on ice until used. The needles used for the injections were pulled in advance from glass capillary tubes using a David Kopf Instruments Vertical Pipet Puller (Model 700C). Heat was set at 54 and the solenoid was set to 10 and the needle pulled by gravity to a fine point. Needles were then collected and beveled using #5 watch maker forceps and filled with mineral oil to force out any air in the needle and to ensure that the tip had been beveled. This needle was then placed into a Nanoliter 2000 Microinjector (World Precision Instruments Model B203XVY) attached to a Marehauser MMJR Micromanipulator (World Precision Instruments). Once attached to the microinjector the needle was then filled with the mRNA/plasmid/phenol red solution. The embryos on the slide were then positioned under a stereoscopic microscope

and the needle was manually inserted into the yolk of the 1-cell stage embryo. To ensure that the solution did not get forced out of the yolk the needle was pulled out slightly after initially penetrating the yolk and then was injected with 4nl of the prepared solution. After being injected the embryos were placed into a glass petri dish with 0.3x Danieau buffer and kept in a 28°C incubator until they were to be prepared for imaging. The embryos were cleaned twice daily by removing any dead embryos and other debris and by adding fresh 0.3% Danieau buffer.

### *Imaging the Embryos*

Embryos younger than 72hpf were first dechorionated under the microscope using a pair of #5 watchmaker forceps. Once the embryos had been dechorionated, the chorions were removed from the petri dish containing the embryos to avoid their autofluorescence. The embryos were then collected and placed into a 1.5 ml centrifuge tube along with 1ml of 0.3% Danieau buffer. To this tube was then added several drops of 8% tricaine/1% Danieau buffer (10ml 1% Danieau buffer, 4mg tricaine) to anesthetize the embryos. Once the embryos were anesthetized, they were separated, 3 per 1.5ml centrifuge tube, and transferred with approximately 0.1ml of the liquid from the previous tube. To this tube an equal volume of melted Danieau agarose was added (80mg agarose, 10ml 0.3% Danieau buffer) that was then quickly mixed and placed in a deep well glass microscope slide. Before the agarose cooled fully the embryos were arranged with a needle to ensure that they were not grouped too closely together and that the proper angles could be imaged. Additional 8% tricaine/ 1% Danieau buffer was then added to the slide well to adhere the cover slip and any extra liquid was removed with a Kim wipe. If the same embryos were to be imaged multiple times in one day, agarose was only applied prior to the final round of imaging to reduce stress on the

embryos. To image the previous time points the embryos were simply suspended in 8% tricaine/ 1% Danieau buffer in a shallow well microscope slide and then left to recover in a petri dish of fresh 0.3x Danieau buffer at 37°.

Fluorescent images of the embryos were obtained using a Zeiss LSM 520 Confocal Microscope. Embryos were visualized using the 10x objective, the argon laser, mercury lamp, and the FITC filter to locate the fluorescence from EGFP. Once the EGFP expression had been located, a Z-Stack of the image was taken to compound the expression from the various focal layers of the embryos to create a stack of roughly 20 layers of variable thickness depending on the region of the embryo being imaged. Pixel quality was set to 1024x1024 and laser-scanning speed was slowed down to 6 to help increase the clarity of the images.

## Results

### *Bioinformatic Analysis of Dr-m2de1 Transcription Factor Binding Sites*

To first determine whether or not the extra basepairs in the Dr-m2de1 sequence were necessary to drive optimal expression of the Dr-m2de1 target gene, the element was examined for putative transcription factor binding sites. The element was examined using the free online program TFsearch, and a threshold of 85% certainty was used to identify transcription factor binding sites. Analysis revealed 15 putative transcription factors with 64 total possible binding sites (supplementary material Fig. S1). The five transcription factors with the highest rated binding sites were Cdx4, Sry, Gata-1, Sox-5 and AML-1a and the five transcription factors with the most binding sites were Cdx4 (29), Oct-1 (9), Gata-1 (7), Sry (6) and the S8, C/EBP and HFH-2 with 3 binding sites each. The majority of these binding sites lied within the shared region of Dr-m2de1 and 3288. No binding sites were observed in the 5' region of Dr-m2de1 that is not included in 3288, however three binding sites are observed in the 3' region of Dr-m2de1 that is not included in 3288 (Fig. 7; supplementary material Fig. S1). The three binding sites present are two Cdx4 binding sites and a single Oct-1 site. This would suggest the possibility that the 3' end of Dr-m2de1 could differentiate in the expression that would be driven by constructs with or without the binding sites.

|        |                            |            |            |            |            |                        |             |
|--------|----------------------------|------------|------------|------------|------------|------------------------|-------------|
| 451    | GTGTTTTTTT                 | TTAACAACAA | TCGACCCTCA | ACATGCACGG | CCTTATAATG | entry                  | score       |
| <----- |                            |            |            |            |            | <a href="#">M00148</a> | SRY 96.4    |
| -----  |                            |            |            |            |            | <a href="#">M00042</a> | Sox-5 94.1  |
|        |                            |            |            | -----      |            | <a href="#">M00137</a> | Oct-1 92.7  |
|        |                            |            |            | ----->     |            | <a href="#">M00101</a> | CdxA 90.7   |
|        |                            | ----->     |            |            |            | <a href="#">M00042</a> | Sox-5 88.9  |
| -----> |                            |            |            |            |            | <a href="#">M00130</a> | HFH-2 87.8  |
| -----> |                            |            |            | <-----     |            | <a href="#">M00101</a> | CdxA 86.4   |
|        |                            |            |            |            |            | <a href="#">M00130</a> | HFH-2 86.3  |
|        |                            |            | <-----     |            |            | <a href="#">M00075</a> | GATA-1 86.1 |
| 501    | <b><u>AGC</u></b> TTTTCTAA | TTTCTTTTAC | TTTATTAATT | CAGTCAACAA | AATTATACAA | entry                  | score       |
|        |                            |            | <-----     |            |            | <a href="#">M00101</a> | CdxA 100.0  |
|        |                            |            |            | ----->     |            | <a href="#">M00101</a> | CdxA 98.6   |
| ----   |                            |            |            |            |            | <a href="#">M00137</a> | Oct-1 92.7  |

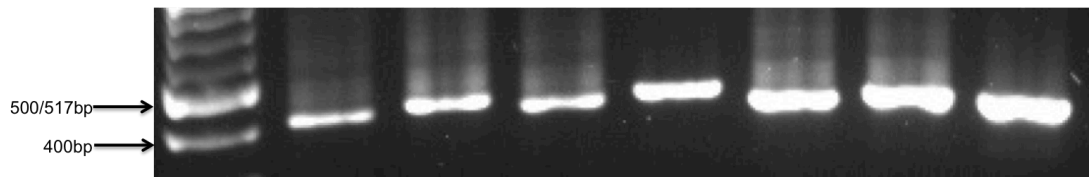
**Fig. 7. Diagram of 3' Dr-m2de1 Transcription factor Binding Sites.** This is a portion of the TFsearch results from *Dr-m2de1*. The entire data set is presented in supplementary Fig. S1. The bolded and underlined basepairs AGC represent the 3' end of the Dr-m2de1 sequence with all of the basepairs preceding it being part of the element. The arrows located below the sequence indicate where and in which direction the putative transcription factors bind to the DNA. To the right are the identification numbers and protein names. The score describes the percent certainty of DNA binding.

#### *Generation of Dr-m2de1 Fragments*

Fragments of Dr-m2de1 were generated from the plasmid vector Dr-m2de1-pCR2.1-TOPO by PCR amplification with specific primers (Table 1; Fig. 8). Three fragments were generated, the first of which was the original 3288 enhancer (Parker et al., 2011). The second fragment called 3288-3' consists of 432bp of Dr-m2de1, which includes the 3288 fragment and the additional 3' end basepairs of Dr-m2de1. The final fragment consisted of the 3288 enhancer with the addition of the extra 5' end basepairs of Dr-m2de1 and is called 3288-5'. These fragments were then Gateway cloned into the injection vector, pGW-*cfos*-EGFP, which would enable them to be inserted into the genome of injected zebrafish



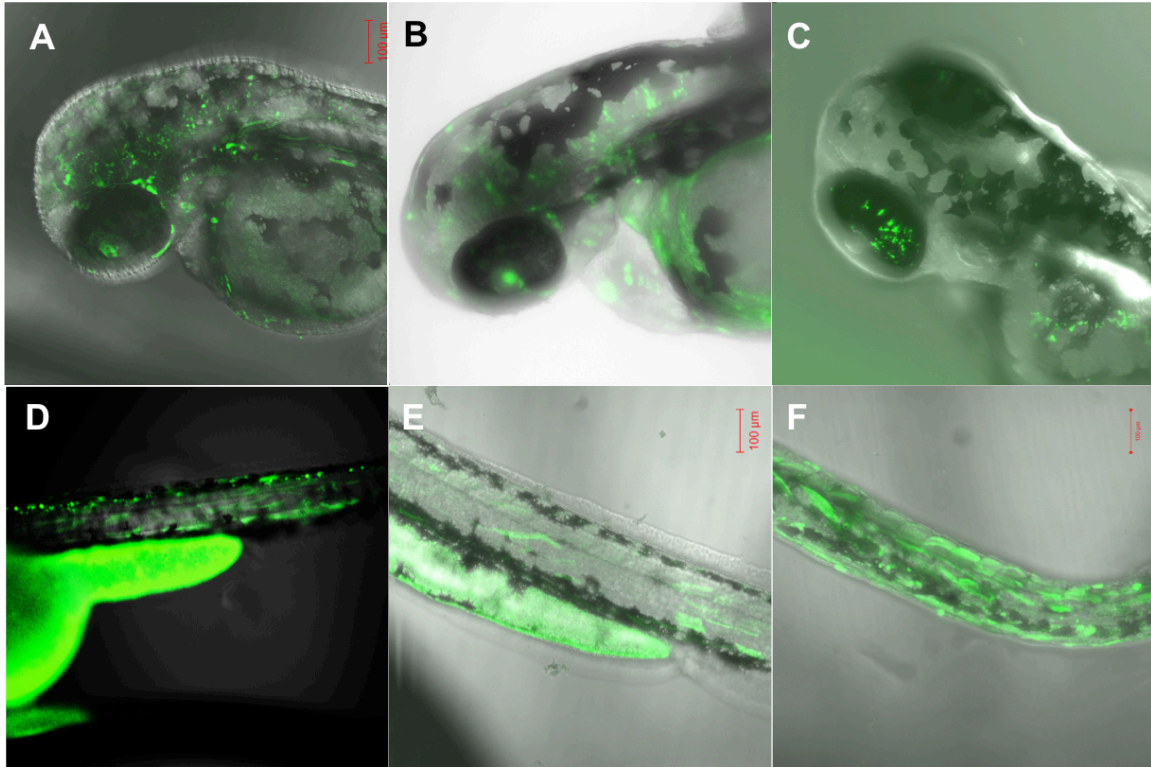
embryos with the aid of co-injected transposase mRNA. The resulting vectors from the process were p3288-*cfos*-pGW, p3288-3'-*cfos*-pGW, and p3288-5'-*cfos*-pGW (Fig. 6). These injection vectors would also allow for the observation of where the enhancer fragments drove expression of the EGFP reporter gene in the embryo, allowing for the comparison of the fragment expression patterns.



**Fig. 8. Image of Dr-m2de1 and Enhancer Fragments.** An image of the Dr-m2de1 enhancer and fragments PCR amplified out of TOPO 2.1 and *cfos*-pGW. Products in the first 3 lanes were amplified from *cfos*-pGW and are, in order, 3288, 3288-3', and 388-5'. The last four lanes were amplified from Topo 2.1 and are Dr-m2de1, 3288, 3288-3', and 388-5', in order.

#### *Expression Driven by Dr-m2de1*

Embryos were injected with pDr-m2de1-*cfos*-pGW injection vector and were imaged at both 48hpf and 54hpf using a Zeiss LSM 520 Confocal Microscope. Sixty-five Embryos were imaged, and at 48 hours, 21 exhibited the mid and hindbrain expression that had been seen previously (Fig. 9) (Barrett, unpublished thesis). The embryos that were imaged also showed expression in muscle fibers along the trunk, as well as the head (Fig. 9) (Barret, 2013). When the same embryos were imaged at 54hpf the same expression was seen in all of the same regions (Fig. 9).

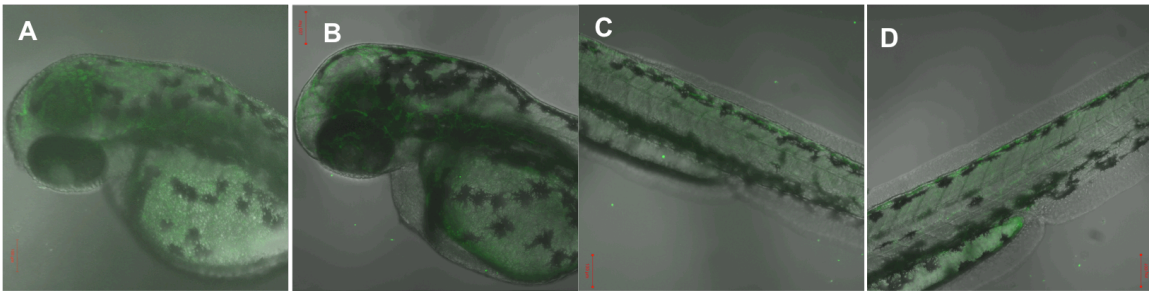


**Fig. 9. Dr-m2de1 Driven Expression of GFP at 48hpf and 54hpf.** Images of EGFP expression in muscle fibers driven by Dr-m2de1. Expression at 48hpf can be seen in the head (A) and the trunk (D) from the side of the embryo. Expression at 54hpf can be seen in the head (B) and the trunk (E) from the side and in the trunk dorsally (F). Image C shows a control embryo that was injected with EGFP-*cfos*-pGW without an enhancer to show that expression seen in the other images is driven by the enhancer and not the promoter.

#### *Expression Driven by 3288*

Embryos were injected with the p3288-*cfos*-pGW cassette and observed using the same microscope as before at 48hpf. Ninety-seven embryos were imaged and expression in the mid and hindbrain resembling that driven by the whole Dr-m2de1 enhancer was seen in 34 embryos (Fig. 10). Similar expression was seen at 54hpf in the same embryos but was brighter and more defined than what was seen at the previous time point (Fig. 10). No EGFP

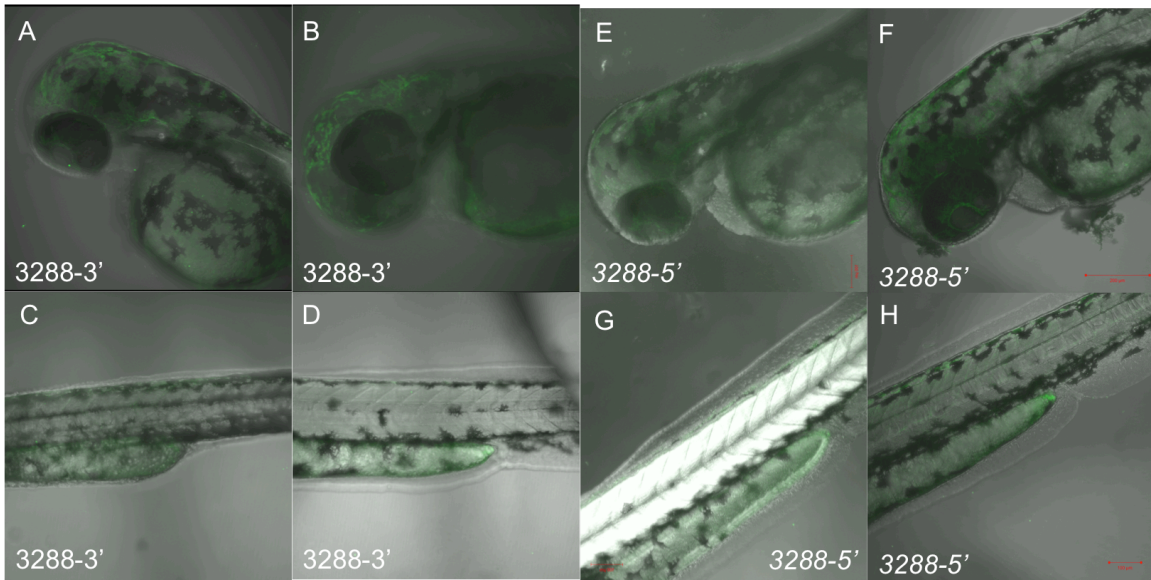
expression was seen however in the muscle fibers at either time point, unlike the results seen with Dr-m2de1. At 48hpf and 54hpf some expression was seen in the eye, but this is still thought to be in the brain and only apparent due to z-stacking of the images (Fig. 10).



**Fig. 10. 3288 Driven Expression of GFP at 48hpf and 54hpf.** Images of EGFP driven by the 3288 enhancer. Figure A displays the side of the head of a 48hpf embryo with visible expression in the mid and hindbrain. C is also from a 48hpf embryo and shows no visible expression in muscle fibers. Figures B and D are both from a 54 hpf embryo and show expression from the side.

#### *Expression Driven by 3288-3' and 3288-5'*

Sixty-eight embryos were imaged following injection with the p3288-5'-*cfos*-pGW injection cassette and 112 embryos were imaged with the p3288-3'-*cfos*-pGW cassette. Out of these embryos only 14 of the 3288-5' transgenic embryos and 24 of the 3288-3' transgenic embryos displayed EGFP expression. The expression that was seen was similar to the embryos that had been injected with p3288-*cfos*-pGW (Fig. 11). At 48hpf weak EGFP expression was seen in the mid and hindbrain of the embryos, with stronger expression in the same regions at 54hpf (Fig. 11). Again no muscle fiber expression was seen at either of the time points but some expression can still be seen in the eye.



**Fig. 11. 3288-3'/5' Driven Expression of GFP at 48hpf and 54hpf.** Images show EGFP expression driven by the 3288-3' and 3288-5' enhancers at both 48hpf (A, C, E, G) and 54hpf (B, D, F, H) in embryos. The top row of images display views of the embryos from the side while the bottom row shows views of the trunk from the same angle.

## Discussion

Currently, much is known about how Meis proteins and their cofactors regulate the expression of other genes but not much is known about the regulation of the *Meis* genes. Previously, our lab discovered four putative enhancers for *Meis2*, *m2de1-4*. All four of these enhancers are present in tetrapod organisms, like humans, chickens, and mice, and have been found in the same location downstream of *Meis2*. Zebrafish, however, only possess one of these enhancers, *m2de1*. Our lab has previously shown that this enhancer directs expression of a GFP reporter gene to the mid and hindbrain of developing zebrafish at 48hpf. Another paper, published in 2011, described a similar enhancer that drove expression to a much smaller region of the hindbrain (Parker et al., 2011). This enhancer, 3288, was slightly smaller than *m2de1*, it was missing roughly 19 basepairs on the 3' and 17 basepairs on the 5' end. In this study, I examined three fragments of *Dr-m2de1* to determine whether or not the basepairs on either end contained any important regions that would influence expression.

In zebrafish the *m2de1* enhancer is 451bp long and the difference between this enhancer and 3288 is relatively small, only 19bp on the 3' end and 17bp on the 5' end. To determine how necessary these ends were, the entire *Dr-m2de1* sequence was examined for putative transcription factor binding sites using the free online program TFsearch. The search revealed numerous binding sites within the shared region

between the two enhancers, and while none were found within the 17 bp on the 5' end, three binding sites were found in the 19bp on the 3' end. The three sites were two caudal-a (Cdx<sub>a</sub>) sites and a single Oct-1 binding site.

The two caudal binding sites were found at the same location in the sequence but in opposite directions, suggesting that the Cdx<sub>a</sub> protein could bind to either strand. The two sites, however, received different scores for DNA binding, the first with a 90.7% chance of binding and the second with an 86.4% chance, suggesting that the former is more likely. Zebrafish have two different caudal genes, *cdx1a* and *cdx4*, which are important for posterior identity and *Hox* regulation (Davidson and Zon, 2006; Gamer and Wright, 1993; Marom et al., 1997; Mlodzik et al., 1985; Skromne et al., 2007; Wingert et al., 2007). Cdx<sub>a</sub> is actually an older name for the chicken Cdx1 protein and while it and *cdx1a* only share 49% identity, they are 86% identical in their homeodomains, thus suggesting that *cdx1a* could be binding to these sites instead of Cdx<sub>a</sub>.

The Oct-1 site was reported to have a 92.7% chance for binding. In zebrafish Oct-1 is also known as *pou2f1a* or *pou1f1*. The *Oct* genes and other *Pou* genes are involved in gene regulation and are known to be expressed in the developing central nervous system. *Oct-1*, is ubiquitously expressed in the tissues of mice (Ming et al., 2001; Schöler et al., 1989). *Pou2f1a* has been found in great abundance in single cell stage embryos suggesting that it has an early function in embryogenesis (Ming et al., 2001). Oct1 has also shown to be very important for recruiting TBP to enhancer bound promoters, if the promoter is minimal or if it lacks proximal control elements (Bertolino and Singh, 2002). This data suggests that any of six Oct1 binding sites found in the Dr-m2de1 enhancer could be important for its regulation of expression of a target gene.

After the bioinformatics analysis of the enhancer, embryos were injected with one of the three generated injection cassettes, p3288-*cfos*-pGW, p3288-3'-*cfos*-pGW, and p3288-5'-*cfos*-pGW or with the pDr-m2de1-*cfos*-pGW injection cassette designed by Cody Barrett (Barrett, unpublished thesis). The embryos that were imaged in Parker et al. (2011) were imaged at 54hpf while those studied previously in the Zerucha lab were at 48hpf, so injected embryos were imaged at both time points for comparison.

Embryos injected with pDr-m2de1-*cfos*-pGW exhibited similar expression to that seen previously in the Zerucha lab at 48hpf. Expression was seen broadly within the regions of the mid and hindbrain. In addition to the neural expression, GFP could also be seen in muscle fibers in the trunk of the embryo. This had not been observed previously. Pbx/Meis has been previously shown, in both zebrafish and mice, to interact with the transcription factor MyoD to help dictate skeletal muscle cell fates (Berkes et al., 2004; Fong et al., 2015). The expression of *myoD* in zebrafish is strongest early in development from ~7hpf to 24hpf. It then drops off, but is still found in tissues throughout the tail and trunk (Weinberg et al., 1996). This dual expression in both the brain and muscle fibers helps support the idea that m2de1 is a *meis2a* enhancer. The same pattern of expression was seen when the embryos were imaged again at 54hpf.

Embryos that had been injected with the p3288-*cfos*-pGW injection cassette displayed different expression from what was seen in both Parker et al. (2011) and from the embryos injected with pDr-m2de1-*cfos*-pGW. Embryos with the 3288 enhancer fragment showed the same broad expression in the mid and hindbrain as Dr-m2de1 at both 48hpf and 54hpf as opposed to the more restricted expression seen previously in the Parker paper. 3288 however did not drive the same expression in muscle fibers that Dr-m2de1 did. This evidence

supports the hypothesis that the 3' and 5' end of Dr-m2de1 do play some important function in regulating target gene expression.

To further support this hypothesis, when embryos that had been injected with the cassette p3288-5'-*cfos*-pGW, which lacks the extra 3' binding sites, expression in muscle fibers was not seen. Expression was still seen in the brain at 48hpf and 54hpf with the intensity of GFP fluorescence increasing as development progressed. This implies that the transcription factor binding sites found in the 3' end of the enhancer are important for proper muscle fiber development. If this were true than the embryos that had been injected with the p3288-3'-*cfos*-pGW cassette should still show fluorescence in muscle fibers. When the embryos were imaged, however, GFP expression was identical to that driven by the 3288-5' enhancer with no muscle fiber expression.

The original analysis of Dr-m2de1 did not reveal any putative transcription factor binding sites contained in the 5' end of the enhancer so its loss should not have any effect on gene expression. What could explain this discrepancy is that the region could contain one or more binding sites that the algorithm in TFsearch ignored because they fell below a certain threshold. When the program analyzes a sequence for potential binding sites it ignores any sites that do not have a likelihood of binding below 85%, so it is plausible that there are binding sites in the 5' region that were just not recognized by the program. These yet to be identified binding sites could help explain why both the 3288-3' and 3288-5' enhancer fragments do not drive any expression in the trunk.

Another factor that could have impacted where expression was seen is the *cfos* minimal promoter that was used in all of the injection cassettes. It has been suggested recently that *cfos* tends to direct expression to the neural regions of an embryo instead of



globally. This is mostly likely not an issue, however for two reasons. First, one of the main objectives of this study was to compare the expression between Dr-m2de1 and the 3288 from Parker et al. (2011). In that paper they also used the *cfos* minimal promoter, so all of the data should be consistent and the promoter should not be responsible for any discrepancies between the two enhancers. The second reason why the promoter should not prove to be biased is that expression driven by Dr-m2de1 was seen regularly in muscle fibers, a non-neuronal tissue, as well as the brain. Regardless though we intend to replace the *cfos* promoter with one from carp to remove any possible bias that could be present in future studies.

The proximity between the promoter and the enhancer fragments 3288-5' and 3288 could also possibly cause false expression. The transcription factor Oct-1, which has a binding site in the 3' end of Dr-m2de1, is known to recruit TBP to promoters in the scenario where there is a distal regulatory element and a minimal promoter (Bertolino and Singh, 2002). This recruitment of TBP helps drive expression of minimal promoters that lack proximal control elements so that the gene can be expressed (Bertolino and Singh, 2002). Without oct-1, gene expression weakens (Bertolino and Singh, 2002). Both the 3288-5' and 3288 fragments lack this oct-1 site, which may cause decreased expression of a target gene in the genome. This change in expression however would most likely not be able to be detected due to the proximity of the enhancer to the promoter. The design of the injection cassettes turns the distal enhancer into a proximal one, thus hypothetically the promoter would no longer need oct-1 to help drive expression. A possible solution would be to knock-out m2de1 in zebrafish and then insert a mutated enhancer that lacks the oct-1 site. Its affect on transcription could then be observed via *in situ* hybridization for *meis2a* transcripts or

qRNAseq for a more quantitative method. if the amount of *meis2* mRNA is less than with the wild type enhancer, then the site is necessary.

Another future direction would be to first confirm that m2de1 is indeed a *Meis2* enhancer. Currently the location of m2de1 and the expression that it drives strongly suggests that its target gene is *Meis2*, but this has not yet confirmed. The best way to accomplish this would be to knock-out the entire enhancer using the CRISPR/Cas9 system and then see if there is any effect on *meis2* transcription. While this method has become quite common for protein coding genes it has only recently been used for the deletion of enhancers. Typically CRISPRs function by using a short sequence of RNA known as a guide RNA (sgRNA) that is complimentary to a specific sequence of DNA. This sgRNA guides a Cas9 endonuclease that then generates a double stranded break in the DNA between the sgRNA target sequence and a specific sequence known as a proto-spacer adjacent motif (PAM). This double stranded break then ligates back together via non-homologous end joining, causing a mutation that then non-functionalizes the protein. A single mutation is not enough however to non-functionalize an enhancer; instead two separate sgRNAs must be used that flank the enhancer. These will cut out the enhancer and hopefully the areas around it will ligate together, generating a mutant. A previous study that removed a super-enhancer region found that in 50% of embryos injected with the CRISPR/Cas9 system had a mutation and that 12% had the mutation in both alleles (Li et al., 2014). This experiment would provide strong evidence that *meis2* is the target gene of m2de1.

Another study that would need to be done would be to confirm the presence of *meis2* in the muscle fibers of zebrafish. This could be done by simply creating an RNA probe for *meis2*, which could then be visualized via *in situ* hybridization to confirm the presence of

*meis2* RNA. If the probe does not reveal the presence of any transcripts in the muscle fibers it could either be that the transcripts are only present in low levels or that the enhancer may actually be driving the expression of a different gene. A possible candidate could be one of the teashirt-related zinc finger (*tshz*) genes. The *tshz* gene has expression that resembles that of *meis2* and can be found in both the area of the brain and muscle fibers (Santos et al., 2010; Parker et al. (2011). To confirm that the enhancer is actually driving expression in muscle fibers the embryos could be co-injected with a separate construct that is known to drive expression only in muscle fibers, if the expression of the two constructs overlaps then we can be certain that *Dr-m2de1* is driving expression in muscle fibers.

In conclusion, when we imaged embryos with the 3288 element, we observed mid and hindbrain expression dissimilar to what was seen previously in Parker et al. (2011), and instead much more similar to what was seen previously with *Dr-m2de1*. The 3288 enhancer and other enhancer fragments also did not drive any expression in the muscle fibers of the developing zebrafish. This difference is most likely due to the lack of certain binding sites in both the 3' and 5' end. While this study does show that 3288 is a fragment of *m2de1*, further testing needs to be done to confirm whether *m2de1* definitively drives *meis2* expression.

## References

- Agoston, Z., Heine, P., Brill, M., S., Britta, M., G., Hau, A., Kallenborn-Gerhardt, W., Schramm, J., Gotz, M. and Schulte, D.** (2014). Meis2 is a Pax6 co-factor in neurogenesis and dopaminergic periglomerular fate specification in the adult olfactory bulb. *Development* **141**, 28-38.
- Alberts, B.** (2008). Molecular biology of the cell. Garland Science, New York.
- Amali, A., A., Sie, L., Winkler, C. and Featherstone, M.** (2013). Zebrafish *hoxd4a* acts upstream of *meis1.1* to regulate vasculogenesis, angiogenesis and hemotopoiesis. *PLoS ONE* **8(3)**, e58857.
- Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M. and Postlethwait, J. H.** (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science* **282** 1711-1714.
- Akiyama, Yutaka.** *TFsearch*. Computational Biology Research Center, AIST, Japan, 2001. <http://diyhpl.us/~bryan/irc/protocol-online/protocol-cache/TFSEARCH.html>. Web. 8 June, 2015
- Balavoine, G., de Rosa, R. and Adoutte, A.** (2002). Hox clusters and bilaterian phylogeny. *Mol. Phylogenet. Evol.* **24**, 366-373.

- Barrett, L.W., Fletcher, S. and Wilton, D.S.** (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* **69**, 3613-3634.
- Barrett, C.E.** (2013). Conserved non-coding element derived regulation of the Meis2.2 homeobox gene during embryonic development. [MS Thesis]. Appalachian State University. 130 pp.
- Bell, A. C., West, A. G. and Felsenfeld, G.** (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**, 447-450.
- Berkes, C. A., Bergstrom, D. A., Penn, B. H., Seaver, K. J., Knoepfler, P. S. and Tapscott, S. J.** (2004). Pbx marks genes for activation by MyoD indicating a pole for a homeodomain protein in establishing myogenic potential. *Molecular Cell* **14**, 465-477.
- Bertolino, E. and Singh, H.** (2002). POU/TBP cooperativity: a mechanism for enhancer action from a distance. *Molecular Cell* **10**, 397-407.
- Bessa, J., Tavares, M. J., Santos, J., Kikuta, H., Laplante, M., Becker, T. S., Gomez-Skarmeta, J. L. and Casares, F.** (2008). meis1 regulates cyclin D1 and c-myc expression, and controls the proliferation of the multipotent cells in the early developing zebrafish eye. *Development* **135**, 799-803.
- Biemar, F., Devos, N., Martial, J. A., Driever, W. and Peers, B.** (2001). Cloning and expression of the TALE superclass homeobox Meis2 gene during zebrafish embryonic development. *Mechanisms of Development* **109**, 427-431.

- Blackwood, E.M. and Kadonaga, J.T.** (1998). Going the distance: a current view of enhancer action. *Science* **281**, 60-63
- Bomgardner, D., Hinton, B. T. and Turner, T. T.** (2003). 5' hox genes and meis 1, a hox-DNA binding cofactor, are expressed in the adult mouse epididymis. *Biol. Reprod* **68**, 644- 650.
- Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R. and Nasmyth, K.** (1985). Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* **41**, 41-48.
- Brennan, R.G. and Matthews, B.W.** (1989). The helix-turn-helix DNA binding motif. *The Journal of Biological Chemistry* **264**, 1903-1906.
- Brooke, N.M., Garcia-Fernandez, J. and Holland, P.W.H.** (1998). The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Letters to Nature* **392(6679)**, 920-922.
- Brunet, F. G., Roest Crollius, H., Paris, M., Aury, J. M., Gibert, P., Jaillon, O., Laudet, V. and Robinson-Rechavi, M.** (2006). Gene loss and evolutionary rates following whole- genome duplication in teleost fishes. *Mol. Biol. Evol.* **23**, 1808-1816.
- Burglin, T. R.** (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Res.* **25**, 4173-4180.

**Capdevila, J., Tsukui, T., Rodriguez Esteban, C., Zappavigna, V. and Izpisua**

**Belmonte, J. C.** (1999). Control of vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism of BMPs by Gremlin. *Mol. Cell* **4**, 839-849.

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

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

**Choe, S., K., Ladam, F. and Sagerstrom, C. G.** (2014). TALE factors poise promoters for activation by Hox proteins. *Developmental Cell* **28**, 203-211.

**Chong, S. W., Korzh, V. and Jiang, Y. J.** (2009). Myogenesis and molecules insights from zebrafish *Danio rerio*. *J. Fish Biol.* **74**, 1693-1755.

**Clark, A. R. and Docherty, K.** (1993). Negative regulation of transcription in eukaryotes. *Biochem. J.* **296 (Pt 3)**, 521-541.

**Conaway, R.C. and Conaway, J.W.** (1989). An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. *Proc Natl Acad Sci USA.* **86**, 7356–7360.

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

- Crocker, J., Abe, N., Rinaldi, L., McGregor, A.P., Frankel, N., Wang, S., Alsaadi, A., Valenti, P., Plaza, S., Payre, F., Mann, R.S. and Stern, D.L.** (2015). Low affinity binding site clusters confer Hox specificity and regulatory robustness. *Cell* **160**, 191-203.
- Dai, M., Wang, Y., Fang, L., Irwin, D., M., Zue, T., Zhang, J., Zhang, S. and Wang, Z.** (2014). Differential Expression of *Meis2*, *Mab21/2* and *Tbx3* during limb development associated with diversification of limb morphology in mammals. *PLoS ONE* **9(8)**, 3106100.
- Davidson, A.J. and Zon, L.I.** (2006). The *caudal*-related homeobox genes *cdx1a* and *cdx4* act redundantly regulate *hox* gene expression and the formation of putative hematopoietic stem cells during zebrafish embryogenesis. *Developmental Biology* **292**, 506-518.
- Davidson, E. H.** (2006). The regulatory genome: gene regulatory networks in development and evolution. Academic Press, San Diego, CA.
- Deng, W. and Roberts, S.G.E.** (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev.* **19**, 2418–2423.
- Dorn, A., Affolter, M., Gehring, W. J. and Leupin, W.** (1994). Homeodomain proteins in development and therapy. *Pharmacol. Ther.* **61**, 155-184.
- Drapkin, R., Reardon, J.T., Ansari, A., Huang, J.C., Zawel, L., Ahn, K., Sancar, A. and Reinberg, D.** (1994). Dual role of TFIID in DNA excision repair and in transcription by RNA polymerase II. *Nature* **368**, 769– 772.



- Duboule, D.** (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development Supplement*, 135-142.
- Dutton, J. R., Antonellis, A., Carney, T. J., Rodrigues, F. S., Pavan, W. J., Ward, A. and Kelsh, R. N.** (2008). An evolutionarily conserved intronic region controls the spatiotemporal expression of the transcription factor Sox10. *BMC Dev. Biol.* **8**, 105.
- Echelard, Y., Vassileva, G. and McMahon, A. P.** (1994). Cis-acting regulatory sequences governing Wnt-1 expression in the developing mouse CNS. *Development* **120**, 2213-2224.
- Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E. and Beachy, P. A.** (1994). The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. *EMBO J.* **13**, 3551-3560.
- Elgen, R.M., and Reisine, T.** (2011). Screening for compounds that modulate epigenetic regulation of the transcriptome. *Journal of Biomolecular Screening* **16**, 1137-1152.
- Engstrom, P. G., Ho Sui, S. J., Drivenes, O., Becker, T. S. and Lenhard, B.** (2007). Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res.* **17**, 1898-1908.
- Feaver, W.J., Gileadi, O., Li, Y. and Kornberg, R.D.** (1991). CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* **67**:1223–1230.

- Fiering, S., Whitelaw, E. and Martin, D. I.** (2000). To be or not to be active: the stochastic nature of enhancer action. *Bioessays* **22**, 381-387.
- Flores, O., Maldonado, E. and Reinberg, D.** (1989). Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J Biol Chem.* **264**, 8913–8921.
- Flores, O., Ha, I. and Reinberg, D.** (1990). Factors involved in specific transcription by mammalian RNA polymerase II. Purification and sub- unit composition of transcription factor IIF. *J Biol Chem.* **265**, 5629– 5634.
- Fong, A. P., Yao, Z., Maves, L. and Tapscott, S. J.** (2015) Conversion of MyoD to a neurogenic factor: binding site specificity determines lineage. *Cell Reports* **10**, 1937-1946.
- Gamer, L.W. and Wright, C.V.E.** (1993). Murine *Cdx-4* bears striking similarities to the *Drosophila caudal* gene in its homeodomain sequence and early expression pattern. *Mechanisms of Development* **43**, 71-81
- Garcia-Ramirez, M., Rocchini, C. and Ausio, J.** (1995). Modulation of chromatin folding by histone acetylation. *J. of Biol. Chem.* **270**, 17923-17928
- Gehring, W. J.** (1987). Homeoboxes in the study of development. *Science* **236**, 1245-1252.
- Gehring, W. J.** (1993). Exploring the homeobox. *Gene* **135**, 215-221.
- Gehring, W. J., Affolter, M. and Burglin, T.** (1994). Homeodomain proteins. *Annu. Rev. Biochem.* **63**, 487-526.

- Geerts, D., Revet, I., Jorritsma, G., Schilderink, N. and Versteeg, R.** (2005). MEIS homeobox genes in neuroblastoma. *Cancer Lett.* **228**, 43-50.
- Geyer, P. K.** (1997). The role of insulator elements in defining domains of gene expression. *Curr. Opin. Genet. Dev.* **7**, 242-248.
- Gerasimova, T. I. and Corces, V. G.** (2001). Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annu. Rev. Genet.* **35**, 193-208.
- Gibcus, J. H. and Dekker, J.** (2013). The Hierarchy of the 3D genome. *Mol. Cell* **49**, 773-82.
- Gonzalez-Reyes, A. and Morata, G.** (1990) The developmental effect of overexpressing a Ubx product in drosophila embryos is dependent on its interactions with other homeotic products. *Cell* **61(3)**, 515-522.
- Goodrich, J.A. and Tjian, R.** (1994). Transcription factors IIE and IIIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* **77**, 145–156.
- Grainger, S., Hryniuk, A. and Lohnes, D.** (2013) Cdx1 and Cdx2 exhibit transcriptional specificity in the intestine. *Plos One* **8(1)**, 1-12
- Greer, J. M., Puetz, J., Thomas, K. R. and Capecchi, M. R.** (2000). Maintenance of functional equivalence during paralogous Hox gene evolution. *Nature* **403**, 661-665.
- Grunstein, M.** (1997). Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349-352.

- Gutkovich, Y., E., Ofir, R., Elkouby, Y., M., Dibner, C., Gefen, A., Elias, S. and Frank, D.** (2010). Xenopus Meis3 Protein lies at a nexus downstream to Zic1 and Pax3 proteins, regulating multiple cell-fates during early nervous system development. *Dev. Biol.* **338**, 50-62.
- Hisa, T., Spence, S. E., Rachel, R. A., Fujita, M., Nakamura, T., Ward, J. M., Devor-Henneman, D. E., Saiki, Y., Kutsuna, H., Tessarollo, L., Jenkins, N. A. and Copeland, N. G.** (2004). Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *EMBO J.* **23**, 450-459.
- Hokamp, K., McLysaght, A. and Wolfe, K.H.** (2003). The 2R hypothesis and the human genome sequence. *Journal of Structural and Functional Genomics* **3**, 95-110.
- Holland, P. W. and Garcia-Fernandez, J.** (1996). Hox genes and chordate evolution. *Dev. Biol.* **173**, 382-395.
- Irimia, M., Maeso, I., Burguera, D., Hidalgo-Sanchez, M., Puellas, L., Roy, S. W., Garcia-Fernandez, J. and Ferran, J. L.** (2011). Contrasting 5' and 3' evolutionary histories and frequent evolutionary convergence in Meis/hth gene structures. *Genome Biol. Evol.* **3**, 551-564.
- Johansson, S., Berland, S., Gradek, G., A., Bongers, E., de Leeuw, N., Pfundt, R., Fannemel, M., Rodningen, O., Brendehaug, A., Haukanes, B., I., Hovland, R., Helland, G. and Houge, G.** (2013). Haploinsufficiency of *MEIS2* is associated with orofacial clefting and learning disability. *Am. J. of Med. Genet. A.* **164(7)**. 1622-1626.

**Juo, Z.S., Chiu, T.K., Leiberman, P.M., Baikalov, I., Berk, A.J. and Dickerson R.E.**

(1996). How proteins recognize the TATA Box. *J. Mol. Biol.* **261**, 239-254

**Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A. Z., Engstrom, P. G.,**

**Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K., Ghislain, J., Pezeron,**

**G., Mourrain, P., Ellingsen, S., Oates, A. C., Thisse, C., Thisse, B., Foucher, I.,**

**Adolf, B., Geling, A., Lenhard, B. and Becker, T. S. (2007).** Genomic regulatory

blocks encompass multiple neighboring genes and maintain conserved synteny in

vertebrates. *Genome Res.* **17**, 545-555.

**Kinkel, M.D., Eames, S.C., Alonzo, M. R. and Prince, V.E. (2008).** Cdx4 is required in the

endoderm to localize the pancreas and limit  $\beta$ -cell number. *Development* **135**, 919-

929.

**Knoepfler, P.S. and Kamps, M.P. (1995).** The pentapeptide motif of Hox proteins is

required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and

enhances and DNA binding by Pbx1. *Molecular and Cellular Biology* **15(10)**, 5811-

5819.

**Kondo, T., Isono, K., Kondo, K., Endo, T., A., Itohara, S., Vidal, M. and Koseki, H.**

(2013). Polycomb potentiates meis2 activation in midbrain by mediating interaction

of the promoters with a tissue-specific enhancer. *Developmental Cell.* **28**, 94-101

**Lawrence, H. J., Rozenfeld, S., Cruz, C., Matsukuma, K., Kwong, A., Komuves, L.,**

**Buchberg, A. M. and Largman, C. (1999).** Frequent co-expression of the HOXA9

and MEIS1 homeobox genes in human myeloid leukemias. *Leukemia* **13**, 1993-1999.

- Lemons, D. and McGinnis, W.** (2006). Genomic evolution of Hox gene clusters. *Science* **313**, 1918-1922.
- Lengerke, C. and Daley, G.Q.** (2012). Caudal genes in blood development and leukemia. *Ann. N Y Acad. Sci.* **1266**, 47-54.
- Li, X., Virbasius, A., Zhu, X. and Green, M.R.** (1999). Enhancement of TBP binding by activators and general transcription factors. *Nature* **399**, 605-609
- Li, Y., Rivera, C. M., Ishii, H., Jin, F., Selvaraj, S., Lee, A. Y., Dixon, J. R. and Ren, B.** (2014). CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLOS One* **9(12)**: e114485. Doi: 10.1371/journal.pone.0114485.
- Liu, J., Wang, Y., Birnbaum, M., J. and Stoffers, D. A.** (2010). Three-amino-acid-loop-extension homeodomain factor Meis3 regulates cell survival via PDK1. *Proc. Natl. Acad. Sci. USA.* **107(47)**. 20494-20499.
- Lu, H., Zawel, L., Fisher, L., Egly, J.M. and Reinberg, D.** (1992). Human general transcription factor IIIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* **358**, 641-645.
- Mann, R. S. and Affolter, M.** (1998). Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* **8**, 423-429.
- Marom, K., Shapira, E. and Fainsod, A.** (1997). The chicken *caudal* genes establish an anterior-posterior gradient by partially overlapping temporal and spatial patterns of expression. *Mechanisms of Development* **64**, 41-52.

- Mazzoni, E.O., Mahony, S., Peljto, M., Patel, T., Thorton, S.R., McCuine, S., Reeder, C., Boyer, L.A., Young, R.A., Gifford, D.K. and Wichterle, H.** (2013). Saltatory remodeling of *Hox* chromatin in response to rostro-caudal patterning signals. *Nat Neurosci.* **16(9)**, 1191-1198.
- McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. and Gehring, W. J.** (1984). A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* **308**, 428-433.
- Mercader, N., Leonardo, E., Azpiazu, N., Serrano, A., Morata, G., Martinez, C. and Torres, M.** (1999). Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature* **402**, 425-429.
- Mercader, N., Leonardo, E., Piedra, M. E., Martinez, A. C., Ros, M. A. and Torres, M.** (2000). Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development* **127**, 3961-3970.
- Ming, L., Ying, C., Zhixing, Z., Shuo, L. and Anming, M.** (2001). Characterization and expression pattern of *pouIII*, a novel class II POU gene in zebrafish. Chinese *Science Bulletin* **46(18)**, 1523-1527
- Mlodzik, M., Fjose, A. and Gehring, W.J.** (1985). Isolation of *caudal*, a *Drosophila* homeo box-containing gene with maternal expression, whose transcripts form a concentration gradient at the pre-blastoderm stage. *Embo. J.* **4(11)**, 2961-2969.
- Moens, C. B. and Selleri, L.** (2006). Hox cofactors in vertebrate development. *Dev. Biol.* **291**, 193-206.

- Morata, G. and Lawrence, P. A.** (1977). The development of wingless, a homeotic mutation of *Drosophila*. *Dev. Biol.* **56**, 227-240.
- Moskow, J. J., Bullrich, F., Huebner, K., Daar, I. O. and Buchberg, A. M.** (1995). Meis1, a PBX1- related homeobox gene involved in myeloid leukemia in BXH-2 mice. *Mol. Cell Biol.* **15**, 5434-5443.
- Mukherjee, K. and Burglin, T. R.** (2007). Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution. *J. Mol. Evol.* **65**, 137-153.
- Nakamura, T., Jenkins, N. A. and Copeland, N. G.** (1996). Identification of a new family of Pbx- related homeobox genes. *Oncogene* **13**, 2235-2242.
- Nakamura, T.** (2005). Meis and Hox: a mighty pair defeats apoptosis. *Blood* **105**, 909-910.
- Navratilova, P. and Becker, T. S.** (2009). Genomic regulatory blocks in vertebrates and implications in human disease. *Brief Funct. Genomic Proteomic* **8**, 333-342.
- Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G. and Burley, S.K.** (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119–128.
- Nikolov, D.B., Chen, H., Halay, E.D., Hoffman, A., Roeder, R.G. and Burley, S.K.** (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc Natl Acad Sci USA.* **93**, 4862– 4867.



- Nishimoto, S., Minguillon, C., Wood, S. and Logan, M.P.O.** (2014). Expression of *Tbx5* and reveals Hox Protein Specificity. *Plos Genetics* **10(3)**, 1-13.
- Ohkuma, Y., Sumimoto, H., Hoffmann, A., Shimasaki, S., Horikoshi, M. and Roeder, R.G.** (1991). Structural motifs and potential  $\sigma$  homologies in the large subunit of human general transcription factor TFIIE. *Nature* **354**, 398–401.
- Orphanides, G., Lagrange, T. and Reinberg, D.** (1996). The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657-2683.
- Otting, G., Qian, Y. Q., Muller, M., Affolter, M., Gehring, W. and Wuthrich, K.** (1988). Secondary structure determination for the Antennapedia homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. *EMBO J.* **7**, 4305-4309.
- Parker, H. J., Piccinelli, P., Sauka-Spengler, T., Bronner, M. and Elgar, G.** (2011). Ancient Pbx-Hox signatures define hundreds of vertebrate developmental enhancers. *BMC Genomics* **12(637)**, 1471-2164.
- Pascual-Garcia, P., Jeong, J. and Capelson, M.** (2014). Nucleoporin Nup98 associates with Trx/Mll and NSL histone-modifying complexes and regulates Hox gene expression. *Cell Reports* **9**, 433-442.
- Prince, V. E., Joly, L., Ekker, M. and Ho, R. K.** (1998). Zebrafish hox genes: genomic organization and modified colinear expression patterns in the trunk. *Development* **125**, 407-420.

- Peterson, M.G., Inostroza, J., Maxon, M.E., Flores, O., Admon, A., Reinberg, D. and Tjian, R.** (1991). Structure and functional properties of human general transcription factor IIE. *Nature* **354**, 369–373.
- Popham, D. L., Szeto, D., Keener, J. and Kustu, S.** (1989). Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* **243**, 629-635.
- Prohaska, S. J. and Stadler, P. F.** (2004). The duplication of the Hox gene clusters in teleost fishes. *Theory Biosci.* **123**, 89-110.
- Razin, A. and Riggs, A. D.** (1980). DNA methylation and gene function. *Science* **210**, 604-610.
- Rombauts, S., Florquin, K., Lescot, M., Marchal, K., Rouze, P. and van de Peer, Y.** (2003). Computational approaches to identify promoters and cis-regulatory elements in plant genomes. *Plant Physiol.* **132**, 1162-1176.
- Roy, R., Schaeffer, L., Humbert, S., Vermeulen, W., Weeda, G. and Egly, J.M.** (1994). The DNA-dependent ATPase activity associated with the class II basic transcription factor BTF2/TFIIH. *J Biol Chem.* **269**, 9826– 9832.
- Ryoo, H. D., Marty, T., Casares, F., Affolter, M. and Mann, R. S.** (1999). Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. *Development* **126**, 5137-5148.

- Sagerstrom, C. G., Kao, B. A., Lane, M. E. and Sive, H.** (2001). Isolation and characterization of posteriorly restricted genes in the zebrafish gastrula. *Dev. Dyn.* **220**, 402-408.
- Salzberg, A., Elias, S., Nachaliel, N., Bonstein, L., Henig, C. and Frank, D.** (1999). A Meis family protein caudalizes neural cell fates in *Xenopus*. *Mechanisms of Development* **80**, 3-13.
- Sanchez-Guardado, L. O., Ferran, J. L., Rodriguez-Gallardo, L., Puellas, L. and Hidalgo- Sanchez, M.** (2011). Meis gene expression patterns in the developing chicken inner ear. *J. Comp. Neurol.* **519**, 125-147.
- Santos, J. S., Fonseca, N. A., Vieira, C. P., Vieira, J. and Casares, F.** (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**, 1010-1018.
- Schöler, R. H., Hatzopoulos, K. A., Balling, R., Suzuki, N. and Gruss, P.** (1989). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *Embo J.* **8(9)**, 2543-50.
- Shanmugam, K., Green, N. C., Rambaldi, I., Saragovi, H. U. and Featherstone, M. S.** (1999). PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins. *Mol. Cell Biol.* **19**, 7577-7588.
- Sharkey, M., Graba, Y. and Scott, M. P.** (1997). Hox genes in evolution: protein surfaces and paralog groups. *Trends Genet.* **13**, 145-151.

- Shen, W. F., Rozenfeld, S., Kwong, A., Kom ves, L. G., Lawrence, H. J. and Largman, C.** (1999). HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells. *Mol. Cell Biol.* **19**, 3051-3061.
- Shimizu, T., Bae, Y., Muraoka, O. and Hibi, M.** (2005). Interaction of Wnt and *caudal*-related genes in zebrafish posterior body formation. *Developmental Biology* **279**, 125-141.
- Shimizu, T., Bae, Y. and Hibi, M.** (2006). Cdx-Hox code controls competence for responding to Fgfs and retinoic acid in zebrafish neural tissue. *Development* **133**, 4709-4719.
- Skromne, I., Thorsen, D., Hale, M., Prince, V.E. and Ho, R.K.** (2007). Repression of the hindbrain developmental program by Cdx factors is required for the specification of the vertebrate spinal cord. *Development* **134(11)**, 2147-2158.
- Smith, J. E., Jr., Bollekens, J. A., Inghirami, G. and Takeshita, K.** (1997). Cloning and mapping of the MEIS1 gene, the human homolog of a murine leukemogenic gene. *Genomics* **43**, 99-103.
- Steelman, S., Moskow, J. J., Muzynski, K., North, C., Druck, T., Montgomery, J. C., Huebner, K., Daar, I. O. and Buchberg, A. M.** (1997). Identification of a conserved family of Meis1-related homeobox genes. *Genome Res.* **7**, 142-156.
- Struhl, K.** (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**, 1-4.

- Sumimoto, H., Ohkuma, Y., Sinn, E., Kato, H., Shimasaki, S., Horikoshi, M. and Roeder, R.G.** (1991). Conserved sequence motifs in the small subunit of human general transcription factor TFIIE. *Nature* **354**, 401– 404.
- Swalla, B. J.** (1996). Building divergent body plans with similar genetic pathways. *Heredity* **97**, 235-243.
- Szutorisz, H., Dillon, N. and Tora, L.** (2005). The role of enhancers as centers for general transcription factor recruitment. *Trends Biochem. Sci.* **30**, 593-599.
- Thomas, M.C. and Chiang, C.** (2006). The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology* **41**, 105-178.
- Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W.** (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell.* **10**, 1453-1465.
- Toresson, H., Parmar, M. and Campbell, K.** (2000). Expression of *Meis* and *Pbx* genes and their protein products in the developing telencephalon: implications for regional differentiation. *Mechanisms of Development* **94**, 183-187.
- Valverde-Garduno, V., Guyot, B., Anguita, E., Hamlett, I., Porcher, C. and Vyas, P.** (2004). Differences in the chromatin structure and cis-element organization of the human and mouse GATA1 loci: implications for cis-element identification. *Blood* **104**, 3106- 3116.

- Venkatesh, B.** (2003). Evolution and diversity of fish genomes. *Curr. Opin. Genet. Dev.* **13**, 588-592.
- Vlachakis, N., Choe, S. K. and Sagerstrom, C. G.** (2001). Meis3 synergizes with Pbx4 and Hoxb1b in promoting hindbrain fates in the zebrafish. *Development* **128**, 1299-1312.
- Wagner, G. P., Amemiya, C. and Ruddle, F.** (2003). Hox cluster duplications and the opportunity for evolutionary novelties. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14603-14606.
- Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M. and Martin, D. I.** (1995). Enhancers increase the probability but not the level of gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7125-7129.
- Waskiewicz, A. J., Rikhof, H. A., Hernandez, R. E. and Moens, C. B.** (2001). Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning. *Development* **128**, 4139-4151.
- Watanabe, T., Hayashi, K., Tanaka, A., Furumoto, T., Hanaoka, F. and Ohkuma, Y.** (2003). The carboxy terminus of the small subunit of TFIIE regulates the transition from transcription initiation to elongation by RNA polymerase II. *Mol Cell Biol.* **23**, 2914–2926.
- Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Paabo, S., Rebhan, M. and Schubeler, D.** (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457-466.

- Weinberg, E. S., Allende, M. L., Kelly, C. S., Abdelhamid, A., Murakami, T., Andermann, P., Doerre, O. G., Grunwald, D. J. and Riggleman, B. (1996)**  
Developmental regulation of zebrafish *MyoD* in wild-type, *no tail* and *spadetail* embryos. *Development* **122**, 271-280.
- Wingert, R.A., Selleck, R., Yu, J., Song, H., Chen, Z., Song, A., Zhou, Y., Thisse, B., Thisse, C., McMahaon, A.P. and Davidson, A.J. (2007).** The *cdx* genes and retinoic acid control the positioning and segmentation of the zebrafish pronephros. *Plos Genetics* **3(10)**, 1922-1938.
- Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y. J., Cooke, J. E. and Elgar, G. (2005).** Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7.
- Zerucha, T. and Prince, V. E. (2001).** Cloning and developmental expression of a zebrafish *meis2* homeobox gene. *Mechasnisms of Development* **102**, 247-250.
- Zhang, X., Friedman, A., Heaney, S., Purcell, P. and Maas, R. L. (2002).** Meis homeoproteins directly regulate Pax6 during vertebrate lens morphogenesis. *Genes Dev.* **16**, 2097- 2107.

## Supplementary Material Figures

**Fig. S1. Complete Diagram of Putative Dr-m2de1 Transcription Factor Binding Sites.**

Database: TRANSFAC MATRIX TABLE, Rel.3.3 06-01-1998  
 Query: drm2de1 (561 bases)  
 Taxonomy: Vertebrate  
 Threshold: 85.0 point

TFMATRIX entries with High-scoring:

|     |  |                        |              |
|-----|--|------------------------|--------------|
| 1   | AACATATTTA AACTCACTTT AATACACACA CACTAACACT AACTCTCTCA         | entry                  | score        |
|     | ----->   | <a href="#">M00101</a> | CdxA 92.9    |
|     | <-----   | <a href="#">M00240</a> | Nkx-2. 90.7  |
|     | ----->   | <a href="#">M00100</a> | CdxA 88.5    |
|     | ----->   | <a href="#">M00148</a> | SRY 87.3     |
|     | <-----   | <a href="#">M00101</a> | CdxA 85.0    |
|     |  |                        |              |
| 51  | CACAAACTCA CACACACACA <b>CAGG</b> AGGTCG GGTTTAAAGG AGTAAATCTG | entry                  | score        |
|     | ----->   | <a href="#">M00148</a> | SRY 87.3     |
|     | ----->   | <a href="#">M00032</a> | c-Ets- 87.3  |
|     |  |                        |              |
| 101 | TAGCTGCGTG CAGGGCTCTG TGCAGCGGCA GAGATTTGCG GATCTGTCTT         | entry                  | score        |
|     | <-----   | <a href="#">M00250</a> | Gfi-1 85.1   |
|     |  |                        |              |
| 151 | CTAGCATCTA ACAGCCTCAT CCATCACGGC CGCAAACACT CGGTTCCCTGC        | entry                  | score        |
|     | <-----   | <a href="#">M00075</a> | GATA-1 93.5  |
|     | ----->   | <a href="#">M00227</a> | v-Myb 86.1   |
|     |  |                        |              |
| 201 | ACTGTCTGTA AATGTTTTTA GATATTAGCC AATTTATATG CTCTCAGATT         | entry                  | score        |
|     | ----->   | <a href="#">M00101</a> | CdxA 100.0   |
|     | ----->   | <a href="#">M00100</a> | CdxA 96.2    |
|     | <-----   | <a href="#">M00101</a> | CdxA 87.9    |
|     | <-----   | <a href="#">M00100</a> | CdxA 87.2    |
|     | ----->   | <a href="#">M00162</a> | Oct-1 85.7   |
|     |  |                        |              |
| 251 | CATCATGGAA AATCAGCTTT AGCAGCGGCG GCGCATTATC AGCCGGCGTC         | entry                  | score        |
|     | ----->   | <a href="#">M00145</a> | Brn-2 93.1   |
|     | <-----   | <a href="#">M00128</a> | GATA-1 89.4  |
|     | <-----   | <a href="#">M00075</a> | GATA-1 89.0  |
|     | <-----   | <a href="#">M00076</a> | GATA-2 87.7  |
|     | <-   | <a href="#">M00075</a> | GATA-1 86.5  |
|     | <-----   | <a href="#">M00101</a> | CdxA 85.7    |
|     | <-----   | <a href="#">M00127</a> | GATA-1 85.5  |
|     | <-----   | <a href="#">M00077</a> | GATA-3 85.3  |
|     |  |                        |              |
| 301 | GCATCGTGAA GTTTGATATA CGACGTTGTT CACGCGGCGC TGTCTTATGA         | entry                  | score        |
|     | <-----   | <a href="#">M00137</a> | Oct-1 90.5   |
|     | -----  | <a href="#">M00075</a> | GATA-1 86.5  |
|     | <-----   | <a href="#">M00101</a> | CdxA 86.4    |
|     | --   | <a href="#">M00101</a> | CdxA 86.4    |
|     |  |                        |              |
| 351 | TTATAACCAC ACCGAGAAGG GCTTTAATTA ACAAACACAC TCCAGAGCTC         | entry                  | score        |
|     | <-----   | <a href="#">M00271</a> | AML-1a 100.0 |



|  |   |
|--|---|
| <pre> -----               -----&gt;               &lt;-----           &lt;-----           &lt;-----       -----&gt;           -----&gt;       --       ----&gt;           -----&gt;           &lt;-----         </pre> | <pre> M00148 SRY          92.7 M00137 Oct-1       90.5 M00099 S8          90.4 M00159 C/EBP      88.5 M00248 Oct-1      88.4 M00099 S8          88.0 M00101 CdxA       87.9 M00101 CdxA       86.4 M00101 CdxA       86.4 M00148 SRY        86.4 M00241 Nkx-2.    85.3         </pre> |
|--|---|

|   |   |
|---|---|
| <pre> 401 ATCAGTGTCA GTGCCGTAA CCGCTGACAT TCATTATTAT TAGTAGTATT               &lt;-----               &lt;-----                   &lt;----                   &lt;-----               &lt;-----                   &lt;-----               &lt;-----                   &lt;-----               &lt;-----                   &lt;-----               &lt;-----           </pre> | <pre> entry      score M00101 CdxA   98.6 M00101 CdxA   97.9 M00042 Sox-5  94.1 M00101 CdxA   92.9 M00137 Oct-1  90.2 M00130 HFH-2  87.8 M00137 Oct-1  87.3 M00101 CdxA   87.1 M00130 HFH-2  86.3 M00137 Oct-1  85.5         </pre> |
|---|---|

|   |   |
|---|---|
| <pre> 451 GTTGTTTTTT TTAACAACAA TCGACCCTCA ACATGCACGG CCTTATAATG &lt;----- -----               -----&gt; -----&gt;               -----&gt; -----&gt;               &lt;-----           </pre> | <pre> entry      score M00148 SRY          96.4 M00042 Sox-5       94.1 M00137 Oct-1      92.7 M00101 CdxA       90.7 M00042 Sox-5      88.9 M00130 HFH-2      87.8 M00101 CdxA       86.4 M00130 HFH-2      86.3 M00075 GATA-1     86.1         </pre> |
|---|---|

|   |   |
|---|---|
| <pre> 501 <b>AGCTTTCTAA</b> TTTCTTTTAC TTTATTAATT CAGTCAACAA AATTATACAA               &lt;----- ----&gt;               -----&gt;               -----&gt;               &lt;-----           &lt;-----               -----&gt;               &lt;-----           </pre> | <pre> entry      score M00101 CdxA   100.0 M00101 CdxA   98.6 M00137 Oct-1  92.7 M00269 XFD-3  91.8 M00099 S8     90.7 M00148 SRY    90.0 M00101 CdxA   90.0 M00096 Pbx-1  89.2 M00101 CdxA   88.6 M00101 CdxA   88.6 M00131 HNF-3b  87.3 M00100 CdxA   87.2 M00159 C/EBP  86.9 M00100 CdxA   85.9         </pre> |
|---|---|

|   |   |
|---|---|
| <pre> 551 ATAATTTAAC T -----&gt; ----- ----- -- ---</pre> | <pre> entry      score M00101 CdxA   88.6 M00101 CdxA   88.6 M00131 HNF-3b  87.3 M00100 CdxA   87.2 M00159 C/EBP  86.9         </pre> |
|---|---|

Total 64 high-scoring sites found.  
 Max score: 100.0 point, Min score: 85.0 point

Supplementary Material Fig. S1. Above are the TFsearch results from the *m2de1* sequence from zebrafish. Only binding sites with a score above 85 are featured. On the left is the sequence broken down into segments 50bp in length and each is annotated with an arrow depicting where each binding site is located with the head of the arrow indicating its direction. On the right is the name of the TF along with its score. Cdx4 is shown to be the highest scoring putative TF followed by SRY, GATA-1, SOX-5 and AML-1a. The basepairs bolded and underlined represent the 5' and 3' ends of *Dr-m2de1* in that orientation.

## **Vita**

Tyler James Ferrara was born in Warwick, Rhode Island to James and Sharon Ferrara. He attended Warwick Veterans Memorial High School and after graduating in 2007, he went on to receive a Bachelors of Science in Microbiology at the University of Rhode Island, located in Kingston, RI. Then in 2013 he began studying under Dr. Ted Zerucha at Appalachian State University in order to obtain his Master of Science degree in Cell and Molecular Biology.