TRANSCRIPTIONAL REGULATION OF THE *MEIS2 L*OCUS

A Thesis
by
Megan Daley Tennant

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

May 2018
Department of Biology

TRANSCRIPTIONAL REGULATION OF THE *MEIS2 L*OCUS

A Thesis
by
MEGAN DALEY TENNANT
May 2018

APPROVED BY

_____
Dr. Ted Zerucha
Chairperson, Thesis Committee

_____
Dr. Cortney Bouldin
Co-chairperson, Thesis Committee

_____
Dr. Mary Kinkel
Member, Thesis Committee

_____
Dr. Zack Murrell
Chairperson, Department of Biology

_____
Max C. Poole, Ph.D.
Dean, Cratis D. Williams School of Graduate Studies

Abstract

TRANSCRIPTIONAL REGULATION OF THE *MEIS2 L*OCUS

Megan Daley Tennant
B.S., Bluefield College
M.S., Appalachian State University

Chairperson: Dr. Ted Zerucha

The *Meis* genes are a member of the three-amino acid loop extension (TALE) superclass of the homeobox super-family of genes. Homologs of the *Meis* genes have been identified in all animals examined and have been found to be expressed in similar patterns during the embryonic development of those animals. The *Meis* genes code for the production of proteins that act as transcription factors. These transcription factors are able to directly regulate the expression of target genes but are most well-known for functioning as cofactors, directly interacting with other transcription factors and DNA to facilitate transcriptional regulation. Most notably, they appear to act as co-factors of the evolutionarily well-conserved Hox proteins and have also been described to act with other transcription factors on DNA. Although the *Meis* genes are fairly well-characterized in terms of their molecular function and expression pattern during development, little is known regarding how their expression is regulated. Previously, the Zerucha lab has 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 only been found in land vertebrates to date, m2de1 is also found in teleosts, including zebrafish, where it is downstream of *meis2a*.

Interestingly, these elements are found within the introns of an adjacent gene, *zgc*:*154061* in zebrafish, whose orthologs are always found in an inverted convergently transcribed orientation directly downstream of *Meis2* in vertebrates. The evolutionarily conserved genomic organization of these two genes hints to the sharing of *cis*-regulatory elements.

When observing the expression patterns of *meis2a* and *zgc*:*154061* as well as the expression of a reporter transgene directed by m2de1 in zebrafish, there were overlaps in expression, specifically around the eye and brain. To determine whether or not m2de1 played a role in directing the expression of either of these genes, a portion of m2de1 was excised from the zebrafish genome via the CRISPR/Cas9 system. Embryos lacking this portion of m2de1 displayed faint *meis2a* expression in the forebrain and optic tectum but lacked expression in the cerebellum and rhombencephalon, as compared to wildtype embryos. When observing *zgc:154061* expression in mutant fish compared to wildtype, there was no difference; however, this could be due to the already low levels of expression of *zgc:154061* and the inability to distinguish any changes in expression to this already low and difficult to visualize expression pattern in wildtype embryos. The change in *meis2a* expression in mutant embryos suggests that m2de1 is an important player in directing the expression of *meis2a.* While further testing is needed to identify the transcription factor binding sites that are being excised and how this could effect *meis2a* expression, this project served to further characterize m2de1, a highly conserved non-coding element associated with the *Meis2* gene.

Acknowledgements

First and foremost, I would like to thank my co-advisors, Dr. Ted Zerucha and Dr. Cort Bouldin, for all of their time and assistance during my time here as a graduate student. I cannot express how thankful I am for the guidance that they both offered. I would also like to extend this thanks to my third committee member, Dr. Mary Kinkel for offering her assistance and expertise along the way. I am also grateful to Dr. Ece Karatan, for her assistance with numerous protocols and graduate school affairs. To Dr. Guichuan Hou in the microscopy facility and Monique Eckerd in the animal facility, thank you both for your time and effort in assisting me over the past two and a half years.

I owe much appreciation to my Biology graduate student peers for providing support, assistance and stress relief when needed. I would also like to extend my gratitude to the Appalachian State University Office of Student Research for the financial support I have received during my years of graduate school. Finally, to my parents, thank you for the love and support that you have shown me.

Table of Contents

List of Tables

List of Figures

Foreword

The references, tables and figures within this thesis were prepared in accordance to the author submission requirements of *Development,* a peer-reviewed journal by The Company of Biologists.

Introduction

This project serves to gain a better understanding of transcriptional regulation of the *Meis2* locus. In order to have a better understanding of embryogenesis, it is important to understand transcriptional control, specifically how homeobox genes are regulated. While both the molecular function and expression patterns of the *Meis* genes are fairly well characterized, little is known regarding how the expression of the *Meis* genes is regulated throughout development. The Zerucha laboratory has identified four highly conserved noncoding elements associated with the vertebrate *Meis2* gene. These elements were named m2de1-4 (for *Meis2* downstream element 1-4). While M2de2-4 have only been found in land vertebrates to date, m2de1 is also found in teleosts, including zebrafish where it is downstream of *meis2a*. Interestingly, these elements are found within the introns of an adjacent gene, *zgc*:*154061* in zebrafish, whose orthologs are always found in an inverted convergently transcribed orientation directly downstream of *Meis2* in vertebrates. Previous studies have shown overlaps in the expression patterns of *meis2a* and *zgc*:*154061* as well as the expression of a reporter transgene directed by m2de1 in zebrafish (Carpenter et al., 2016). We propose that the genomic organization of these two genes has been evolutionarily conserved due to the sharing of cis-regulatory elements. This project serves to further characterize one of the highly conserved elements, m2de1 within this context. Using the CRISPR/Cas9 system, zebrafish m2de1 was targeted in order to excise a portion of the element in order to determine the effects on the expression of both *meis2a* and *zgc:154061*.

Regulation of transcription is a complex process that controls when, where and how often a gene is transcribed. The regulation of gene transcription plays an important role in embryogenesis, specifically through directing cell division, differentiation and patterning

(Wolpert, 2007). During development, homeobox genes have been shown to play important roles in the context of the regulation of gene expression (Wolpert, 2007).

Homeobox genes make up a large super-family of genes that are highly conserved across species. Originally discovered in *Drosophila melanogaster* (the common fruit fly), it was noted that mutations to these genes caused homeosis, where one body segment of the fly takes on the characteristics of another (McGinnis et al, 1984; Morata and Lawrence, 1977). Since their discovery in *Drosophila*, they have been identified in all multicellular eukaryotes and over 200 homeobox genes have been identified ("Homeoboxes," 2012). These genes encode transcription factors that regulate downstream target genes (Chariot et al., 1999) and are crucial for embryogenesis as they often act at or near the top of genetic hierarchies and thus direct expression of suites of downstream developmental genes (Burglin et al. 1997; Chang et al., 1996; Chariot et al., 1999; Mukherjee and Burglin, 2007). Homeobox genes contain a highly conserved, approximately 180 base pair sequence, known as the homeobox. The homeobox of each gene encodes for a specific homeodomain, which comprises approximately 60 amino acids (Gehring, 1987; Gehring, 1993; McGinnis et al., 1984). Generally, the homeodomain acts as a DNA-binding motif that binds to *cis* regulatory elements. When homeodomain transcription factors bind to *cis* regulatory elements, they are able to control spatial and temporal regulation of the target gene (Blackwood and Kadonaga, 1998).

The homeodomain structure consists of a 3 α-helix bundle. Helices one and two form a helix-loop-helix structure. The third helix lies perpendicular to the other two helices, forming a helix-turn-helix with the second α-helix. The helix-turn-helix structure is what allows the homeodomain protein to bind to DNA (Fig. 1) (Banarjee-Basu et al., 2001;

Burglin, 1997; Longobardi et al., 2014; Mukherjee and Burglin, 2007; Shang et al., 1994). The third α-helix lies in the major groove of the DNA as the numerous amino acids interact with both strands of the DNA helix through intermolecular and hydrophobic interactions (Dror et al, 2014; Gehring et al., 1994; Lappin et al., 2006). In addition to the interactions occurring between the protein and the DNA in the major groove, there are also interactions occurring within the minor groove of the DNA. The N-terminal tail of the homeodomain protein interacts with the minor groove of the DNA (Dror et al., 2014; Gehring et al., 1994; Lappin et al., 2006; Shang et al., 1994) while the C-terminal end forms salt bridges to further the interaction between the protein and the DNA (Gehring et al., 1994)

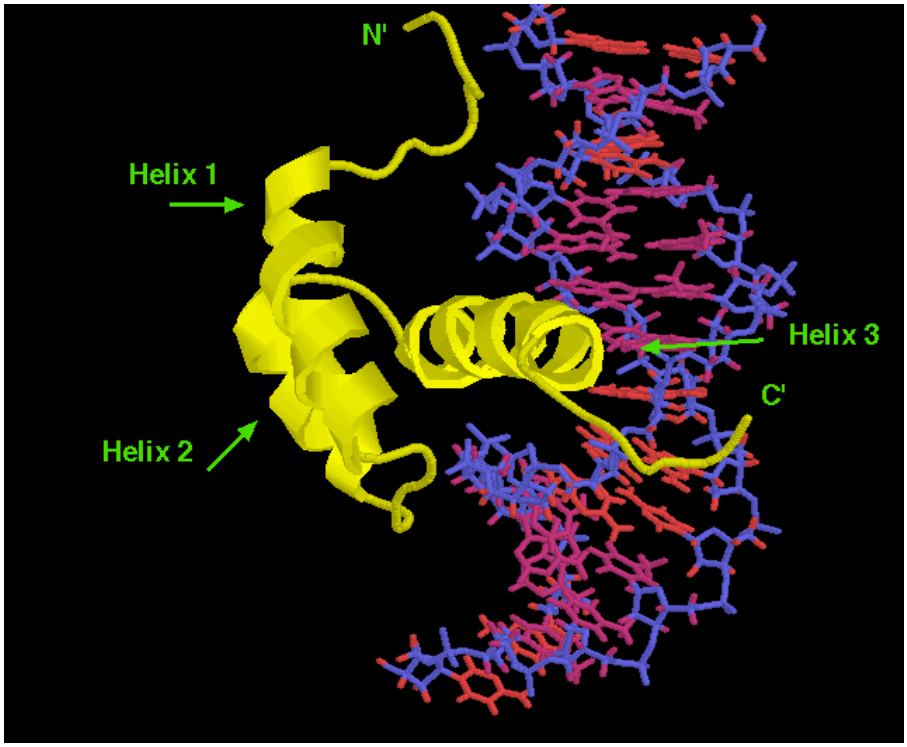

Figure 1: Schematic of an Antennapedia homeodomain protein interacting with double stranded DNA. Alpha helix one and two lie parallel to one another while alpha helix two and three are perpendicular to one another. The third alpha helix lies in the groove of the double stranded DNA. The N-terminal domain interacts with other transcription factors in order to increase stability of the complex, while the C-terminal domain also interacts with transcription factors, like Hox. (Based on The Homeobox Page, maintained by Burglin, T. R., http://homeobox.biosci.ki.se/)

Some homeodomain proteins, such as Hox proteins, play significant roles in the anterior and posterior patterning of bilateral animals (Amores et al., 1998; Moens and Selleri, 2006; Waskiewicz et al., 2001). The *Hox* genes are well conserved in sequence and function across species. While different animal species have varying numbers of *Hox* genes, they have been identified in all species tested thus far (Lemons and McGinnis, 2006). *Hox* genes are normally found in clusters, with teleosts having seven clusters, tetrapods having four *Hox* clusters and *Drosophilia* having one cluster (Amores et al., 1998; Krumlauf, 1994; Lappin et al., 2006; Lemons and McGinnis, 2006). The order and location of the *Hox* genes within a cluster on a chromosome correlate to the anterior posterior expression and timing of expression of these genes. For example, *Hox* genes that are positioned nearer the 3' end tend to be expressed earlier in development and more towards the anterior (head) region of the developing embryo as compared to genes on the 5' end. Likewise, the *Hox* genes that are expressed later in development and more toward the posterior (tail) area of the embryo, are located closer to the 5' end (Amores et al, 1998; McGinnis and Krumlauf, 1992). This specific organization that is observed for *Hox* genes assists in both spatial and temporal regulation of gene specific expression and is referred to as collinearity (Amores et al., 1998; McGinnis and Krumlauf, 1992; Prince et al., 1998). Interestingly, the disruption of *Hox* genes that alter their spatial and temporal expression can result in homeotic transformations (Amores et al., 1998; Lemons and McGinnis, 2006; Noordermeer et al., 2011). For example, in *Drosphila*, a mutation that expands the anterior expression of the *Antp Hox* gene results in a fly developing a leg where an antenna should develop (McGinnis et al., 1984).

As previously discussed, homeodomain proteins, including members of the Hox family, function by binding DNA to direct the transcription of target genes. Often,

homeodomain proteins, including Hox, bind to TAAT nucleotide sequences within the major groove of the DNA sequence (Burglin, 1997). This implies these proteins bind DNA with relatively low specificity since many of them recognize this same core sequence. In addition to binding with low specificity, studies in *D. melanogaster* have shown that homeodomain proteins typically bind DNA with relatively low affinity (Crocker et al., 2015). This low specificity and low affinity of homeodomain binding is difficult to reconcile with the important and specific roles they play during development. In order to increase binding affinity and increase binding specificity, homeodomain proteins often interact with other proteins to form complexes. These protein complexes can contain multiple homeodomain proteins, acting as cofactors, and are able to increase binding specificity by increasing the number of binding sites that they occupy. In addition, multiple proteins binding DNA as a complex will also increase the affinity of the binding. This increase in affinity and specificity helps to better regulate the complex process of gene transcription (Mann and Affolter, 1998; Moens and Selleri, 2006).

There are many homeodomain proteins that act as cofactors. One group of cofactors belongs to the Three Amino Acid Loop Extension (TALE) class of homeodomain proteins. The TALE class of homeodomain proteins contains three additional amino acids between alpha helices one and two of the homeodomain. In animals, including vertebrates, gene families in the TALE class include *Meis*, *Pbc*, *Tgif* and *Iro* (Burglin, 1997; Longobardi et al., 2014; Mukherjee and Burglin, 2007). TALE genes have also been identified in fungi and plants (Burglin, 1997). Such cofactors have the ability to bind to one another as well as other homeodomain proteins while also binding to DNA at *cis* regulatory elements. This allows for

the formation of protein complexes, containing multiple transcriptional proteins and DNA (Mann and Affolter, 1998).

TALE proteins, such as Meis and Pbx, often interact with Hox proteins in binding DNA and controlling gene transcription (Choe et al., 2014). Not only do these proteins bind to DNA, they also have the ability to bind to one another, through structures within their N-terminal and C-terminal domains. The C-terminal region interacts with the N-terminal region of the Hox protein (Chang et al.,1997). As mentioned previously, Hox commonly interacts with a TAAT binding sequence, resulting in low binding specificity. However, in mice, when Hox interacts with the Pbx cofactor, the two proteins bind DNA at a TGATTGAT sequence. Similar changes are seen in other models (Waskiewicz et al., 2001). This increases the binding specificity of Hox, ultimately helping to regulate gene transcription with greater specificity.

In addition to working in conjunction with Hox proteins, sequences of the N-terminal domain of Pbx proteins allow them to bind to the Meis cofactor (Chang et al., 1997; Choe et al., 2002; Jacobs et al., 1999). This specific dimer forms on TGATTGACAG sequences, with Pbx bound to the TGAT sequence and Meis bound to the remaining, unbound nucleotides. Interestingly, Meis and Pbx will form dimers even when the specific binding sequence for Meis is not present; this indicates that Meis and Pbx have a high affinity for each other (Chang et al., 1997). Some of the Hox proteins can also use their N-terminal domain to interact with the C-terminal domain of Meis proteins, though the Hox-Meis dimer is not common as only a few of the different Hox proteins have the ability to interact directly with Meis (Shen et al., 1997).

In addition to these dimers, Meis, Pbx and Hox can also form trimers (Chang et al., 1997; Waskiewicz et al., 2001). The Pbx-Meis-Hox trimers are composed of a Meis protein binding to its specific DNA sequence while also interacting with a Pbx-Hox dimer (Chang et al., 1997; Jacobs et al., 1999). The trimer is formed when Hox binds Pbx via the Hox N-terminal domain and Pbx binds Meis through their N-terminal domains (Fig. 2) (Choe at al., 2002; Knoepfler and Kamps, 1995; Shanmugam et al., 1999). Both trimers and dimers have been found to bind in the absence of DNA, indicating that these proteins have a high affinity for each other that is DNA independent (Chang et al., 1997; Shen et al., 1999). Furthermore, the incorporation of the Meis cofactor to form trimers rather than dimers, has been shown to increase the stability and function of transcriptional control (Jacobs et al., 1999).

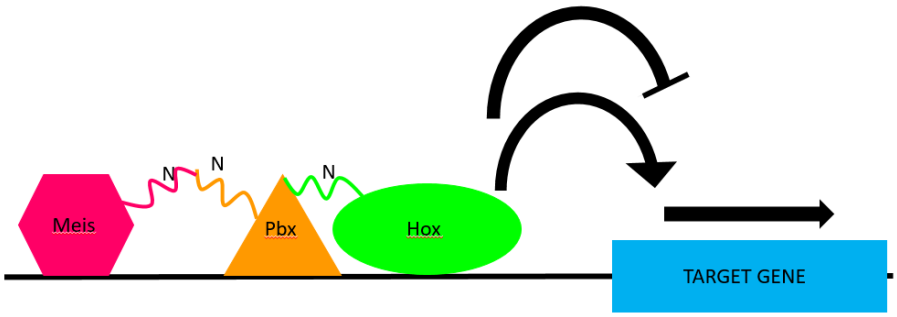

Figure 2: A schematic diagram of Meis, Pbx and Hox proteins working together as transcription factors. Each protein binds to a specific sequence on the black strand of DNA while also working together to form a trimer. The trimer is formed when the N-terminal domain of the Meis protein attaches to the N-terminal domain of the Pbx protein. At the same time, the N-terminal domain of the Hox protein is also interacting with Pbx. The resulting conformation acts as a trimer to either activate or repress target genes. The Meis protein is shown in pink, the Pbx protein is shown in orange and the Hox protein in shown in green. The target gene is shown in blue and the direction of transcription is indicated by the arrow.

While homeodomain proteins can work together as dimers and trimers to increase binding specificity, some homeobox genes like PBC, MEIS and KNOX are known to use

alternative splicing events outside of the conserved regions to generate variations within the C-termini (Burglin, 1997; Longobardi et al., 2014; Tamaoki et al., 1995). These variations within the C-terminal tails are believed to slightly alter the binding specificity allowing the protein to bind to slightly different DNA sequences (Tamaoki et al., 2015). These alternative splicing events can lead to different homeodomain products being generated using the same promotor.

The interactions of transcription factors acting upon *cis* regulatory elements play a significant role in organismal development. More specifically, homeodomain proteins acting as transcription factors play a crucial role in spatial and temporal gene regulation during embryogenesis. These interactions allow for complex anatomical systems to develop in a precise, tightly controlled manner. While the role of individual homeodomain proteins, like the Hox family, in development is relatively well understood, there are still many unanswered questions concerning how they act as complexes. In order to gain a better insight into the complex molecular interactions occurring during development, it is important to gain a better understanding of the cofactors, such as the *Meis* genes, that facilitate the optimal function of these proteins.

The *Meis* genes were first discovered in a study of mouse myeloid ecotropic leukemia virus. It was found that the mouse leukemia virus integrated into a specific site that was later identified as a gene. Because of this, the gene was named the myeloid ecotropic leukemia virus integration site (*Meis*) (Moskow et al., 1995). Paralogs of the gene were subsequently identified and this led to the establishment of the *Meis* gene family. *Drosophila* have a single member of the *Meis* family, *homothorax (hth)* (Kurant et al., 1998; Pai et al., 1998). Tetrapods have at least three *Meis* genes, *Meis1*, *Meis2,* and *Meis3* (Moskow et al., 1995;

Nakamura et al., 1996; Waskiewicz et al., 2001). Tetrapods also have two related *Meis* genes, *Prep1* and *Prep2* (Berthelsen et al., 1998). However, a genome duplication event occurred following the divergence of the teleost lineage from the lineage that would give rise to tetrapods. This resulted in ohnologs in teleosts of the *Meis1* and *Meis2* genes: *meis1a, meis1b, meis2a* and *meis2b* (Amores et al., 1998). To date, *Meis* gene(s) have been identified in all vertebrates examined (Table 1), while *Knox* genes in plants show very similar sequence identity to the *Meis* genes. The similarities between the *Meis* and *Prep* gene families suggest that they diverged from a common ancestor (Burglin, 1997).

Table 1: Species and common name of vertebrates with *Meis* genes, including resources for each species listed.

Species	Common name	Resource(s)
Danio rerio	Zebrafish	(Choe et al., 2002; Sagerstrom et al., 2001; Waskiewicz et al., 2001; Zerucha and Prince, 2001)
Xenopus	Frog	(Salzberg et al., 1999; Steelman et al., 1997)
Gallus gallus	Chicken	(Coy and Borycki, 2010)
Mus musculus	Mouse	(Cecconi et al., 1997; Moskow et al., 1995; Nakamura et al., 1996)
Homo sapiens	Human	(Smith et al., 1997; Steelman et al., 1997)

Like many homeodomain proteins, the homeodomains across the *Meis* family are highly conserved, with a preferred binding sequence of TAAT, causing Meis to bind sequences with poor affinity and low specificity (Moens and Prince, 2001; Waskiewicz et al., 2001). However, when Meis forms a transcription factor complex by interacting with cofactors such as Pbx and Hox, binding specificity is increased (Moens and Prince, 2002; Waskiewicz et al., 2001). As mentioned on pages 6 and 8, Meis proteins act as cofactors, binding both transcription factors and regions of DNA known as *cis*-regulatory sequences. In order for Meis proteins to carry out these functions, the homeodomain and N-terminal domain must carry out specific roles (Burglin, 1997; Chang et al., 1997; Jacobs et al., 1999; Shanmugam et al., 1999). The homeodomain is the structure that the Meis protein uses to bind DNA (Chang et al., 1997). Alone, the interaction between the Meis homeodomain and DNA has been determined to be rather weak (Burglin, 1997). The N-terminal domain functions to increase the strength of the interactions. This is the region of the protein that enables Meis to interact with other transcriptional proteins, or cofactors, ultimately increasing stability and specificity when binding DNA as a complex (Chang et al, 1997; Choe et al., 2002; Jacobs et al., 1999; Shanmugam et al., 1999). The C-terminal region of Meis proteins has also been shown to interact with numerous Hox proteins (Fig. 1) (Shen et al., 1997; Williams et al., 2005). The C-terminal domain has also been shown to accelerate the progression of Acute Myelogenous Leukemia when *Hoxa9* and *Meis* transactivate genes that lead to cancer cell proliferation (Mamo et al., 2006).

One way in which Meis proteins function during development is by forming a transcription factor complex with Pbx and Hox, as previously mentioned. The transcription factor complex then binds *cis*-regulatory elements (Moens and Prince, 2002; Waskiewicz et

al., 2001). When this complex is formed, it has a different preferred binding sequence as compared to the non-specific binding sequences of the individual proteins. Together Meis and Pbx bind a TGATTGAC sequence, with Pbx binding to the TGAT sequence and Meis binding the TGAC sequence (Knoepfler et al., 1999). As the binding specificity and stability increase when Meis and Pbx act together as cofactors, the same is true for when the they act as cofactors with the homeodomain protein Hox (Choe et al., 2002; Knoepfler et al., 1999; Moens and Prince, 2002; Waskiewiez et al., 2001). In addition to working as cofactors with Hox and Pbx, Meis also works with a number of other genes during development.

Since the identification of the *Meis* genes in the myeloid ecotopic leukemia virus, it has been shown that an overexpression of Meis1 causes blood cancer, or leukemia, in both mice and humans (Hisa et al., 2004, Moskow et al., 1995). Leukemia cells with high levels of Meis1 have slower apoptotic pathways (Grubach et al., 2008). It has been proposed that viral integrations in *Meis1* result in a dysregulation of the Meis1 protein, which leads to the inability to differentiate. This can also result in an increase in proliferation of developing myeloid cells. This mutation affecting differentiation leads to the formation of the myeloid leukemia (Moskow et al., 1995).

It has also been suggested that the inability of the Meis/ Pbx complex to regulate Hox can lead to cancer development and increased malignancy. This is due to Hox's role in differentiation and apoptosis (Grubach et al., 2008). Studies have also shown that forced co-expression of *Hoxa9* and *Meis1* in murine marrow leads to the rapid development of leukemia. Using this information, human leukemic cell lines were tested for co-expression of *HOXA9* and *MEIS1* and results showed that co-activation of the two genes is exclusive to acute myelogenous leukemia and is common in such cases (Lawerence et al., 1999). In

addition, Hoxa9, Hoxa7 and Meis1 proteins are found to be co-localization in leukemia cell lines in both mouse and human. This suggests that while *Meis1* is a common site for retroviral insertions, both *Meis* and *Hox* genes could operate together to cause myeloid leukemia (Lawerence et al., 1999; Nakamura et al., 1996).

As mentioned on page 8, there are three members of the *Meis* gene family. Meis1 works with proteins such as HoxA9, HoxB3, Pbx1 and Prep1 (Berkes et al., 2004; Cvejic et al., 2011; Hisa et al., 2004; Melvin et al., 2013; Thorsteinsdottir et al., 2001). *Meis1* has been studied in many different vertebrate embryos. *Meis1* is expressed during the early and later stages of embryogenesis. During the early stages of development, *Meis1* is expressed in the posterior regions of the embryo, specifically in the primitive streak, somites and mesoderm (Coy and Boryicki, 2010; Maeda et al., 2002). *Meis1* also plays an important role in hematopoiesis, as well as heart, muscle and eye development in numerous species, including zebrafish, mice and human (Berkes et al., 2004; Cvejic et al., 2011; Hisa et al., 2004; Melvin et al., 2013; Thorsteinsdottir et al., 2001). The expression of *Meis1* has also been observed in the developing limbs and branchial arches of chick (Coy and Borycki, 2010), xenopus (Maeda et al., 2002) and zebrafish (Waskiewicz et al., 2001).

In mice, *Meis1* expression has been seen in the developing eye and brain (Bessa et al., 2008; Hisa et al., 2004; Toresson et al., 2000). Specifically, *Meis1* has been observed around the retina (Bessa et al., 2008; Hisa et al., 2004) in a similar pattern to that of *Pax6* (Zhang et al., 2002). Meis1 has been shown to bind an enhancer for *Pax6* that controls eye development, suggesting that *Pax6* expression is *Meis1* dependent (Zhang et al., 2002). Mice lacking proper *Meis1* function develop smaller lenses and partially duplicated retinas as compared to wild type mice (Hisa et al., 2004). *Meis1* expression has also been observed in

low levels in the telencephalon of embryonic day 10.5 mouse embryos, around the ventricular zone. The telencephalon ultimately forms the cerebrum of the brain (Toresson et al., 2000). In addition to the expression seen in the eyes and brain area, *Meis1* also plays a role in limb development where expression has been observed proximal to the site of limb bud development, being driven by the signaling molecule retinoic acid (Mercader et al., 2000). In previous studies, *Meis1* mutant mice have developed malformed capillaries and suffered from hemorrhaging, ultimately dying by embryonic day 14.5 (Hisa et al., 2004; Nakamura, 2005).

Due to a genome duplication, teleosts originally had two *Meis1* genes: *meis1a* and *meis1b* (Amores et al., 1998). Over time, it appeared as if the *meis1a* gene was lost in teleosts, likely due to a loss of coding exons (Irimia et al., 2011). However, recent annotations show that not all teleosts have lost the *meis1a* gene. For example, zebrafish, a member of the teleost family, originally had a gene identified as *meis4,* and recent annotations show that *meis4* in zebrafish is indeed a second *meis1* gene, *meis1a*. This indicates that zebrafish still have two *meis1* genes, yet the *meis1a* gene has yet to be characterized (Waskiewicz et al., 2001). The second *meis1* gene, *meis1b,* is expressed throughout the developing face and hindbrain of zebrafish and also plays a role in hematopoiesis (Cvejic et al., 2011; Melvin et al., 2013). Similar to the previously mentioned mice studies, studies in zebrafish have shown that when *meis1b* is knocked down, embryos suffered hematopoiesis defects and poor vascularization (Amali et al., 2013).

Much like *Meis1, Meis2* shows similarities in expression patterns across model organisms, such as mouse, zebrafish and chicken. *Meis2* is expressed in the developing brain, eye, branchial arches, spinal cord, heart and limbs (Bumsted-O'Brien et al., 2007; Cecconi et

al., 1997; Coy and Borycki, 2010; Mercader et al., 2005; Oulad-Abdelghani et al., 1997; Waskiewicz et al., 2010). During mouse embryonic development, *Meis2* is expressed in the somitic compartment and related areas, such as lateral muscle tissue. As development continues, the protein can be found in the mesoderm around embryonic day 8 (E8). By E10, it can be observed in the developing forebrain, hindbrain and spinal cord (Cecconi et al., 1997). The hindbrain activity of *Meis2* is partially regulated by polycomb proteins. These proteins bind to a repressor element near the *Meis2* gene. When the gene is expressed, an enhancer loop, specific to the midbrain, loops around the repressor, ultimately removing the repressor protein and activating the expression of *Meis2* (Kondo et al., 2013). Similar to the expression of *Meis1, Meis2* is active in the developing brain at high levels. In E10.5 mice embryos, Meis2 is found in high levels under the ventricular zone. The highest concentration of the Meis2 protein is found in the subventricular zone and the mantel regions of the telencephalon (Toresson et al., 2000).

The expression of *Meis2* overlaps significantly with *Pax6* in the eye. *Meis2* binds to an enhancer element of *Pax6* and drives the expression of the gene. This results in decreased levels of *Pax6* when *Meis2* is decreased and higher levels of *Pax6* when *Meis2* is increased (Zhang et al., 2002). Together, the two proteins have been shown to play a role in the development of the olfactory bulb in mice (Agoston et al., 2014).

Meis2 has also been shown to play a role in the development of limbs in vertebrates (Capdevilla et al., 1999). Retinoic acid activates the transcription of *Meis2* and this helps to ensure proper expression in the proximal end of developing limbs. When *Meis2* is not expressed properly, it results in loss of digits and limb defects in chickens (Capdevila et al., 1999; Mercader et al., 2000). *Meis2* is expressed in low levels in the developing forelimbs

and hindlimbs of E15 mouse embryos. However, in bats, *Meis2* is highly expressed in the interdigital tissues between the digits. This suggests that *Meis2* plays a role in the sculpting of the wing. Further, it is suggested that *Meis2* plays a role in the retention of interdigital tissue between the digits in bats, ultimately forming the webbing of the bat wing. The low levels of *Meis2* in developing mice limbs, as compared to, the high levels seen in the developing bat wing suggests that mice may not have sufficient *Meis2* expression to form interdigital tissue. Conversely, the higher levels of *Meis2* in the developing bat wing suggests that the high expression is enough to retain the interdigital tissues between the digits of a bat (Dai et al., 2014).

Meis2 has also been shown to play a role in craniofacial development. A study from 2013 showed that 7 of 9 people lacking a functional copy of *MEIS2* possessed some level of cleft palate. This suggests that *MEIS2* is important for closure of the palate in humans. Of these individuals lacking a functional copy of *MEIS2*, all individuals possessed some form of a learning disability, with the worst being a mild intellectual disability. Lastly, this study showed that *MEIS2* deficient individuals also possessed other craniofacial deformities besides cleft palate. While these were only mild defects, typically resulting in overlapping facial features such as a wide forehead or shortened philtrum, it is an interesting finding (Johansson et al., 2013).

In zebrafish, the genome duplication event that occurred in teleosts, following the divergence from the tetrapod lineage, resulted in two functional *Meis2* genes, *meis2a* and *meis2b* (Biemar et al., 2001; Waskiewicz et al., 2001; Zerucha and Prince, 2001). In zebrafish, *meis2b* is expressed during the first part of gastrulation (Biemar et al., 2001; Zerucha and Prince, 2001). Later during development, expression is observed in the central

nervous system, the developing eye, the forebrain, midbrain and hindbrain, spinal cord and the lateral mesoderm (Bessa et al., 2008; Biemar et al., 2001; Moens and Prince, 2002). Specifically, *meis2b* is expressed throughout the hindbrain. The expression within the hindbrain spreads throughout development, and by 12 hours post fertilization (hpf), expression is observed in the forebrain, midbrain and hindbrain. Expression of *meis2b* is also observed in the anterior boundaries of the somites after segmentation has occurred. However, by 14 hpf, expression appears to be down-regulated. Expression of *meis2b* also appears in the retina of developing zebrafish around 36 hpf (Zerucha and Prince, 2001). The expression of *meis2a* is similar to the expression of *meis2b;* however, *meis2a* is also seen in the developing branchial arches and limb buds (Coy and Borycki, 2010; Mercader et al., 1999; Waskiewicz et al., 2001) The overlapping expression pattern of *meis2a* and *meis2b* in zebrafish is similar to the expression pattern of *Meis2* observed in mice (Santos et al., 2010; Zerucha and Prince, 2001).

The third member of the *Meis* family, *Meis3* also shows similar expression patterns between species. Much like *Meis1* and *Meis2, Meis3* was originally duplicated in teleosts. However, it appears as if the second *Meis3* gene (*meis3.2*) in zebrafish has been subsequently lost since the genome duplication event (Waskiewicz et al., 201). The remaining *meis3* gene is first seen observed in the developing hindbrain of zebrafish and *Xenopus* (Choe et al., 2002; Sagerstrom et al., 2001; Salzberg et al, 1999; Waskiewicz et al, 2001). Later during development, zebrafish *meis3* is observed in the somites, pectoral fin and neural tube (Sagerstrom et al., 2001; Waskiewicz et al., 2001). *Meis3* is also active in the developing pancreas of mice and zebrafish, as it plays a role in insulin beta-cell survival (Liu et al., 2010). *Meis3* has also been observed to play an important role in posterior neural

development in *Xenopus* (Salzberg et al., 1999). In zebrafish, *meis3* can change the fate of midbrain and forebrain cells into hindbrain cells (Vlachakis et al., 2001). In loss of function studies of *Meis3,* posterior neural tissues in the hindbrain were disturbed (Choe et al., 2002; Dibner et al., 2001). This demonstrates the importance of Meis3 in the developing hindbrain.

It is clear that *Meis* genes play an important role during embryonic development across species. Though there are differences within each individual *Meis* gene's expression patterns, it is apparent that they share developmental roles. Structurally, all Meis proteins have homeodomains; functionally, all Meis proteins bind DNA and regulate gene transcription during embryonic development and the coding regions of the genes are conserved across species (Irimia et al., 2011). While the basic molecular function and expression patterns of *Meis* are known, it is unknown what is actually controlling these genes during transcription. In order to have a better understanding of how the *Meis* genes are regulated, *cis*-regulatory elements that direct their expression must be identified (Fisher et al., 2006; Hughes et al., 2005). To do this, phylogenetic footprinting can be used to locate highly conserved non-coding elements (HCNEs) near the *Meis* genes. HCNEs have been shown to often act as *cis*-regulatory elements (Fisher et al., 2006; Hughes et al. 2005, Woolfe et al., 2005). Once potential *cis*- regulatory elements have been identified, they can be tested in order to confirm whether or not they are able to regulate the transcription of *Meis* genes.

Previously, four HCNEs that possibly are *cis*-regulatory elements associated with the *Meis2* gene (*meis2a* in zebrafish) were identified by Zerucha and Wellington (unpublished data). Currently, the Zerucha lab is focusing on characterizing these elements to determine if they are functioning as *cis*-regulatory elements and regulating the function of *Meis2a.* One of the elements has also been described by Parker et al. (2011). Work by both Parker et al.

(2011) and the Zerucha lab (unpublished) have shown that the 3288 element (Parker et al., 2011) and m2de1 (Zerucha lab, unpublished data) overlap in nucleotide sequence and genomic location. In addition, they direct expression of reporter transgenes in a pattern broadly similar to a subset of known *Meis2* expression in the midbrain and hindbrain in zebrafish, while the 3288 element also shows similar expression patterns in lampreys (Parker et al., 2011; Parker et al., 2014). The m2de1 element also directs expression to muscle fibers throughout the trunk of the developing zebrafish embryo later in development (Zerucha lab, unpublished data). After further examination, the 3288 element is 17 basepairs shorter than m2de1 on the 5' end and 19 basepair shorter on the 3' end, implying that 3288 may only be a fragment of the entire m2de1 element. This is likely the cause of the discrepancy between expression patterns between 3288 and m2de1 (Zerucha lab, unpublished data). Both m2de1 and 3288 have binding sites for the Pbx-Hox dimer, an important transcription factor complex for the Meis proteins (Parker et al., 2011).

The identification and characterization of these elements reveals that potential *cis*-regulatory elements can be identified through phylogenetic footprinting, which makes use of the conserved nature of *cis*-regulatory elements and compares non-coding genomic sequences across multiple species associated in order to identify Highly Conserved Noncoding Elements (HCNEs) (Kikuta et al., 2007, Muller et al., 2002; Santini et al., 2003; Woolfe et al., 2005). This offers a starting point in de-coding regulatory elements that play a role in the development of vertebrates (Parker et al., 2011). The data from both zebrafish and lamprey show that ancient conserved enhancer sequences are suggestive of conserved developmental mechanisms across vertebrates (Parker et al., 2014). By better understanding the regulatory mechanisms associated with the *Meis* family, further insight will be gained

into the developmental process. Furthermore, understanding the regulatory mechanisms associated with *Meis* genes will also give insight into cell proliferation and the effect on cancers, such as leukemia.

While observing the expression directed by m2de1 can give great insight into the characterization of the element, other experiments are also needed to further characterize the element. The Zerucha lab has previously shown that the expression directed by m2de1 is similar to the expression pattern of *meis2a* in developing zebrafish embryos. This suggests that m2de1 is acting as an enhancer and directing the expression of *meis2a.* In order to better understand the relationship between *meis2a* and m2de1 in zebrafish, m2de1 can be knocked out by using genome engineering, such as the CRISPR/Cas9 system. The expression patterns of *meis2a* can then be observed both before and after the knockout of the element to determine whether or not m2de1 plays a role in the regulatory mechanisms associated with *meis2a.*

Genome engineering is a broad term referring to the techniques and mechanisms established to generate targeted modifications within the genetic information, or the genome, of living organisms. In the past, genetic engineers have been able to insert and delete genes in a variety of organisms; however, these insertions and deletions were not precise, as their location could be difficult to specify (Pennisi, 2013). In recent years, genome editing tools have improved. Zinc Finger Nucleases (ZFNs) and truncated Transcription Activator-Like Effector Nucleases (TALENs) are genome editing tools that induce double stranded breaks at desired loci. These breaks can be repaired by error prone non-homologous end joining, ultimately generating small insertions or deletions at the cut sites (Wood et al., 2011). With ZFNs and TALENs, both technologies require custom designed proteins for each DNA

target; however, designing custom proteins is both expensive, time consuming and difficult to engineer. Although both technologies have begun to enable targeted genome modifications, there is still a need for genome editing technology that is easy to engineer, affordable and easily scalable (Cong et al., 2013).

While TALENs and ZFNs continue to be used and improved in the world of genetic engineering, a new system was identified in which Cas9 (CRISPR associated protein 9) nucleases are directed by short RNAs to induce precise cleavage at desired endogenous loci (Cong et al., 2013). The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system was first identified in bacteria in 1987 when Nakata and colleagues observed a repetitive sequence at the end of a gene in *Escherichia coli* (Ishino et al., 1987). While the full implications of these repeat sequences were not clear at first, in 2005, bioinformatics groups were able to match the spacing between these repetitive sequences to the sequences of phages (Mojica et al., 2000; Pourcel et al., 2005).

Bacteria, as well as archaea, have adaptive immune defenses against plasmids and viruses that use short RNA fragments to direct the degradation of foreign nucleic acids (Jinek et al., 2012; Mali et al., 2013). It was found that when CRISPR acts against invading phages, bacteria transcribe spacer DNA and palindromic DNA into a single RNA molecule. The cell then cuts this single molecule into short CRISPR RNAs (crRNAs) which act with trans-activating RNAs (tracrRNAs) and Cas9 proteins to attack foreign DNA that matches the sequence of the crRNA (Deltcheva et al., 2011).

To better understand the mechanisms associated with this newly identified CRISPR system, a study was performed in *Streptococcus pyogenes* type II. Here, the CRISPR system showed that a CRISPR RNA (crRNA) fused to a trans-activating crRNA (tracrRNA) to form

a two RNA structure. The structure proved sufficient to direct the CRISPR associated Cas9 protein to a specific DNA sequence, cleaving the target site and ultimately generating a double stranded break (Jinek et al., 2012). The mechanisms observed in *S. pyogenes* suggested that the CRISPR system might function in eukaryotic organisms. By using RNA sequences to cleave target genomic sites, a system like CRISPR could improve the current field of genetic engineering (Mali et al., 2013).

Effective genome editing requires that sequences are targeted with both precision and efficiency (Cong et al., 2013). In order to observe the efficiency of the CRISPR/ Cas9 system, a stable cell line containing a GFP coding sequence disrupted by a stop codon was established. Due to the stop codon, the protein fragment did not fluoresce. Next, crRNA and tracrRNA molecules were fused together to generate a single guide RNA (gRNA). Two separate gRNAs were designed to target the fragment containing the disrupted GFP sequence. If the gRNAs were able to delete the stop codon within the GFP sequence, the protein would be able to fluoresce (Mali et al., 2013). The efficiency of this system was compared to a TALEN system targeting the same region (Mali et al., 2013; Sanjana et al., 2012). The successful correction rates of the CRISPR system ranged from 3% using the first gRNA (T1) to 8% using the second gRNA (T2). The correction rate using TALEN was similar. Compared to the TALEN system, the RNA-mediated editing process was twice as quick with results appearing within 20 hours as compared to 40 hours with TALEN. Results suggested that RNA-guided genome editing was promising, as using the CRISPR/Cas9 system was much more efficient and economical than both ZFNs and TALENs (Mali et al., 2013).

Recently, two studies (Jinek et al., 2014; Nishimasu et al., 2014) have shed light on the structural conformation of Cas9 and the structural mechanisms of RNA-guided DNA cleavage by Cas9. Through these studies, it was observed that the Cas9 family of proteins is characterized by six nuclease domains. The HNH domain is a single nuclease domain, while the RuvC domain contains three subdomains. Both the HNH domain and RuvC domain initiate single-stranded DNA cuts. The RuvC I subdomain is near the N-terminal region of Cas9 while the RuvC II and RuvC III subdomains flank the HNH domain near the middle of the protein. All Cas9 proteins also have two REC domains. The REC I domain is responsible for binding gRNA while the protospacer-adjacent motif (PAM) interacting domain is responsible for initiating binding of target DNA. The arginine-rich bridge helix then contacts between 8 to 10 nucleotides on the 3' end of the target DNA and initiates cleavage. The function of the REC II domain is unknown to-date (Fig. 3) (Jinek et al., 2014; Nishimasu et al., 2014).

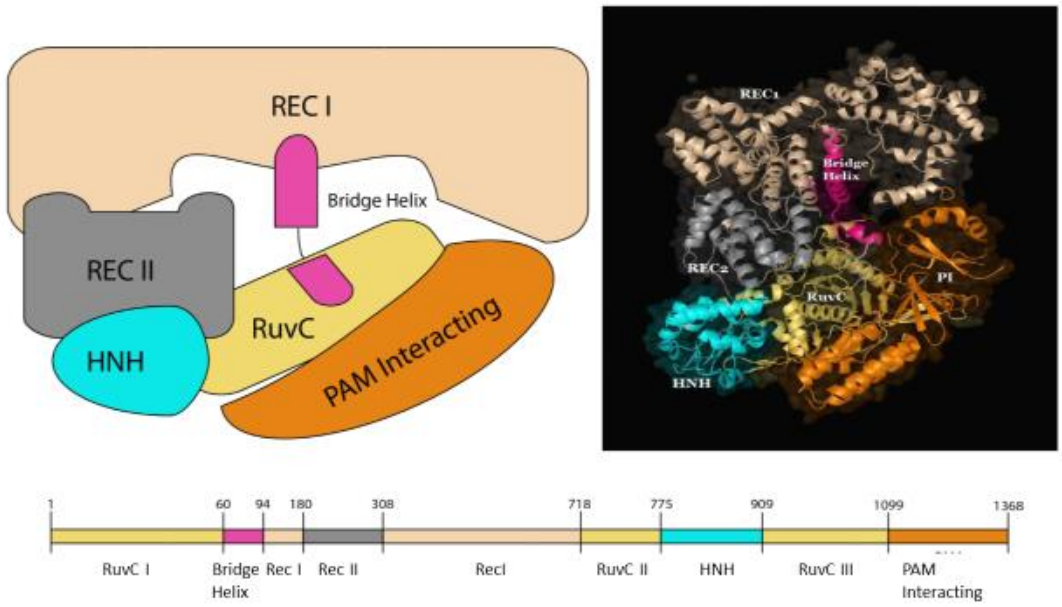

Figure 3: Schematic diagram, crystal structure and map form of the Cas9 protein. The Cas9 protein contains 6 domains: RuvC, Bridge Helix, Rec I and Rec II, HNH and PAM interacting. (Modified from https://sites.tufts.edu/crispr/crispr-mechanism/ website maintained by Tufts University and Jinek et al., 2014)

Currently, *Streptococcus pyogenes* Cas9 (SpyCas9) is the Cas9 enzyme that is used in most CRISPR genome engineering applications. Through a series of biochemical assays and single particle electron microscopy, the crystal structure of SpyCas9 was observed more closely in order to better understand how the enzyme functions within the CRISPR/Cas9 system. Through these experiments, it was observed that in the absence of nucleic acids, the enzyme maintains an auto-inhibited conformation which prevents binding. In other words, when the enzyme is not bound to target DNA or gRNA, SpyCas9 exhibits a conformation in which the HNH domain active site is blocked by the RuvC domain, while also being positioned away from the REC lobe. However, when SpyCas9 comes in contact with gRNA, the enzyme reorients itself into a position where the catalytic centers are on opposite ends. This RNA- induced repositioning is likely to enable the enzyme to bind to target DNA.

(Jinek et al., 2014). In addition, binding of the gRNA to the DNA can be sterically inhibited by the orientation of the C-terminal domain, causing Cas9 to be unable to bind and cleave the target DNA (Nishimasu et al., 2014). Together, these studies suggest that the gRNA serves as a scaffold for Cas9, allowing it to fold and organize its various domains (Jinek et al., 2014; Nishimasu et al., 2014).

Using the CRISPR system, Cas9 can be directed to the target site in two ways. The first way is by a pair of RNA molecules containing a crRNA and a tracrRNA. Alone, neither of these molecules can direct DNA cleavage via Cas9. Both the crRNA and tracrRNA must be present for Cas9 to cleave target DNA (Jinek et al., 2012). The second way to direct the Cas9 is with a chimeric short guide RNA (sgRNA). Both the crRNA and the sgRNA contain a 20-nucleotide sequence that directly match the target DNA sequence. The RNA molecule is able to recognize the DNA via Watson-Crick basepairing. (Choe et al., 2013; Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013).

Alone, crRNA is unable to direct Cas9 to cleave target DNA. Using an electrophoretic mobility shift assay (EMSA), it has been observed that there is very little specific DNA binding with Cas9 alone, or with both Cas9 and crRNA together; however, when using Cas9 and crRNA in combination with tracrRNA, Cas9 is able to cleave the target sequence at a substantially higher rate. This indicates that tracrRNA is required for target DNA recognition and cleavage by Cas9. This is because the tracrRNA is able to properly orient the crRNA for interaction with the complimentary target DNA (Jinek et al., 2012).

The location of site-specific cleavage is dictated by both base-pair complementarity between crRNA and DNA, as well as the PAM of Cas9. While the required PAM recognition sequence varies depending on the species of Cas9, a typical PAM sequence for SpyCas9 is 5’

N-G-G 3’ where the sequence begins with any nucleotide immediately followed by two guanines. The PAM recognition sequence is located immediately downstream of the target DNA and directs the target search mechanisms of the Cas9 protein (Jinek et al., 2012). The Cas9-gRNA complex associates with PAM sequences throughout the genome; however, if the gRNA does not match the DNA sequence, the Cas9 continues to search for the correct target sequence. Ultimately, the Cas9 will interact with a PAM sequence that directly follows the correct target DNA sequence (Fig. 4) (Jinek et al., 2014; Nishimasu et al., 2014; Sternberg et al., 2014). Once this occurs, the Cas9 nuclease domains are activated and the Cas9 will be able to flank the 3’ end of the target DNA (Nishimasu et al., 2014; Shah et al., 2013).

Once the Cas9 protein has been guided to the target DNA via the RNA molecule(s), the Cas9 nuclease cleaves DNA through both the RuvC domain and HNH domain. Each domain then cuts one strand of DNA, the HNH domain cleaves the complimentary strand while the RuvC domain cleaves the non-complementary strand, ultimately generating double stranded breaks (DSBs) with blunt ends (Fig. 4) (Gasiunas et al., 2012; Jinek et al., 2012; Sapranauskas et al., 2011). These DSBs then stimulate a repair mechanism. Normally, DSBs initiate repair through non-homologous end joining (NHEJ) or homology-directed repair (HDR) which are both active in most cell types and organisms (Cong et al., 2013; Maruyama et al., 2015). HDR occurs only during the S and G2 phases of the cell cycle, meaning it occurs less frequently than NHEJ which occurs throughout all phases of the cell cycle (Maruyama et al., 2015).

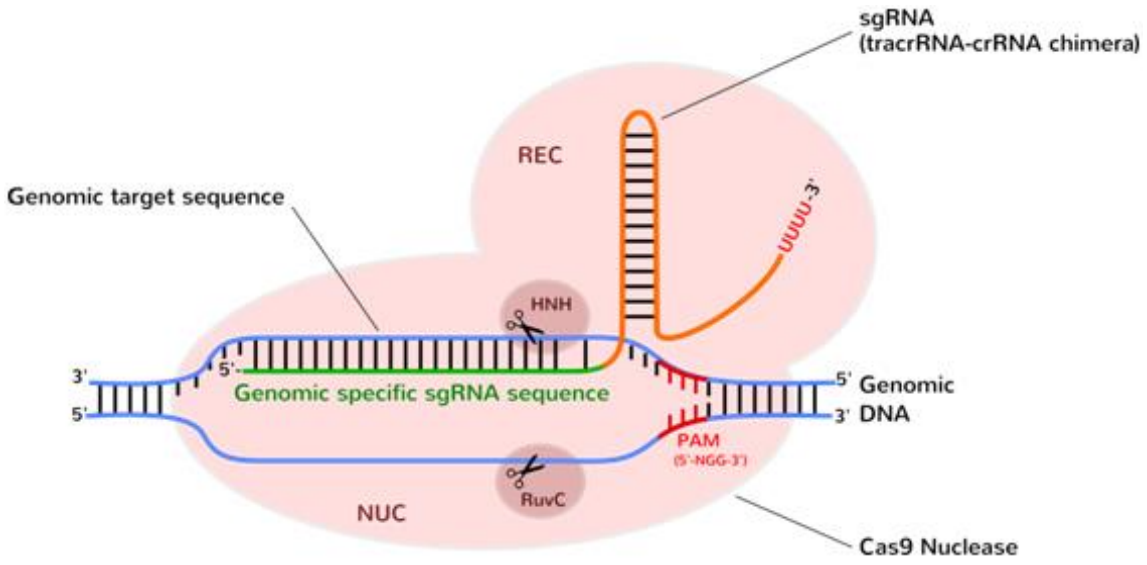

Figure 4: Schematic diagram depicting the interaction between the Cas9 nuclease (pink), the gRNA sequence (green) and the double stranded DNA (blue). Prior to binding, the Cas9/ gRNA complex targets the PAM sequence (red). Once the complex interacts with the correct PAM sequence and the gRNA is complimentary to the DNA, the gRNA and DNA form a heteroduplex. Next, the RuvC domain and the HNH domain of the Cas9 enzyme each cut a single DNA strand, ultimately generating a double stranded break. (Image acquired from www.ozbiosciences.com)

Non-homologous end joining is not the most precise method of repair for efficient genome editing. Often, NHEJ repair mechanisms leads to imperfect or frameshift repairs. This type of repair is thought to introduce insertions and deletions of random lengths which can lead to mosaicism in CRISPR-modified organisms (Maruyama et al., 2015). While HDR and NHEJ occur concurrently, HDR is increased in cells that are incapable of repair through NHEJ. Since NHEJ is not the most precise method to repair DSB, HDR was stimulated at the expense of NHEJ by targeting DNA ligase IV, a key enzyme in the repair mechanism used by the NHEJ pathway. Using the CRISPR/ Cas9 system, DSB were introduced in mammalian cells. At the same time, Scr7, a previously proven inhibitor of DNA ligase IV, was used to inhibit the NHEJ pathway. By inhibiting the NHEJ repair mechanisms, the efficiency of HRD mediated repair in mammalian cells increased 19-fold. Not only can this

be used to increase the precision and efficiency of repair after DSB from the CRISPR/ Cas9 system, this application could also be applied to other genome editing tools such as ZFNs and TALENs. Also, because NHEJ and HDR use similar molecular mechanisms throughout eukaryotes, this application could be applied to other mammalian cells, as well as non-mammalian cells with sufficiently conserved mechanisms (Maruyama et al., 2015). Ultimately, this finding could lead to a huge increase in repairing DSBs after genome editing, leading to higher efficiency and better results.

Another way to improve the repair mechanisms associated with breaks generated by Cas9 is to mutate the Cas9 endonuclease. Cas9 can be mutated by inactivating either the RuvC domain or the HNH domain. After being mutated, Cas9 acts as a DNA nickase that creates a single stranded break (SSB). Single stranded breaks are repaired through the high-fidelity base excision repair (BER) pathway, generating a more precise and specific repair mechanism than through NHEJ and HDR. Because SSB are repaired more effectively than DSB, another way to possibly improve on-target DSB specificity would be by using a double-nicking approach, similar to the approach used in ZFNs and TALENs (Mali et al., 2013b; Ran et al., 2013). By using a pair of gRNAs and a mutated Cas9 endonuclease with only one active domain (HNH+/Ruvc- or HNH-/RuvC+), nicks can mimic DSBs and mediate efficient insertions and deletions. Also, by using a double-nicking approach, off-target nicks are precisely repaired. Theoretically, this would improve specificity by 1,500X as compared to the wildtype Cas9 endonuclease (Ran et al., 2013).

The CRISPR/Cas9 system has been previously shown to work effectively for both knock-in and knock-out applications. These alterations can range from a single base-pair to multiple sites throughout the genome (Shalem et al., 2014). Compared to previous genome

engineering technology, CRISPR has proved to be efficient at generating deletions/ insertions and genetically altered karyotypes (Guo et al., 2015). Multiple studies have looked at altering human stem cells via the CRISPR/Cas9 system (González et al., 2014; Mali et al., 2013; Zhu and Huangfu, 2013). Human pluripotent stem cells (hPSC) have the ability to generate all adult cell types, including uncommon and isolated cell populations. This offers a unique way to study diseases that may not be available otherwise.

In addition to being effective *in vitro*, the CRISPR/ Cas9 system has also been effective *in vivo*. A breakthrough in the genome engineering field came when the CRISPR/Cas9 system was shown to be able to induce mutations at rates similar to, or better than, both TALENs and ZFNs in zebrafish embryos. In 2013, Hwang et al. generated several gRNAs that targeted the *fh* gene of zebrafish. The majority of these targets in the zebrafish *fh* gene were successfully mutated and showed robust mutations using the CRISPR/Cas9 system. Two of these successful targets had previously been unable to be mutated through either TALENs or ZFNs. For the other mutations, success rates were comparable to TALENs and ZFNs. These results show that using gRNA to direct site-specific DNA cleavage via Cas9 is a rather fast and highly efficient method for altering genes *in vivo,* opening the door for using the CRISPR/Cas9 system in other organisms (Hwang et al., 2013).

In addition to being comparable to other genome engineering techniques *in vivo* (Hwang et al., 2013), CRISPR/Cas9 also appears to be both highly efficient and precise in generating heritable mutations in zebrafish. In a different study conducted in 2013, Hruscha et al. designed four different gRNAs to target four different genomic loci. Injected fish were screened via PCR and RFLP and it was observed that mutagenesis reached up to 86.0% while off target effects were relatively low ranging from 1.1-2.5%. Because injected fish display

mosaicism, the transmission from injected fish (PO) to offspring (F1) was also observed as three of seven F1 clutches (43%) carried the mutation. This rate of germline transmission efficiency is comparable to the 45% that has been previously observed with ZFNs. Overall, this study demonstrates that CRISPR/Cas9 is efficient for inducing heritable mutagenesis with low off-target effects, with rates comparable to other genome editing tools. Compared to ZFNs and TALENs, the CRISPR/Cas9 system is cheaper and less time consuming to generate, making it a great tool for genome editing in zebrafish (Hruscha et al., 2013).

A common standard for targeted mutagenesis in zebrafish is *gol*. The gene encodes the slc24a5 cation exchanger which has been shown to effect pigmentation in both humans and zebrafish (Lamason et al., 2005). In 2013, Jao et al. used the CRISPR/Cas system to target the *gol* gene in zebrafish. Mutation in the *gol* gene display lighter melanocyte pigmentation that is detectable in zebrafish embryos as early as 48 hpf. Zebrafish carrying the homozygous *gol* mutation show hypopigmentation in skin melanophores and retinal pigment epithelium (RPE) (Lamason et al., 2005). At 48 hpf, mosaic hypopigmentation was observed in skin melanophores and RPE in 94% of the injected embryos, while mutagenesis rates ranged from 78-98%, indicating that most cells contained bi-allelic mutations. Germ-line transmission of the Cas9 induced mutation was also observed, as well as, recessive null-like phenotypes suggesting high rates of bi-allelic gene disruption. These results suggest that CRISPR/Cas9 is an efficient way to induce bi-allelic gene disruptions in zebrafish (Jao et al., 2013). This rate of germline transmission efficiency is comparable to the 45% that has been previously observed with ZFNs. This results in mosaicism of mutant and wildtype alleles. While this incomplete mutagenesis is necessary for generating germline mutants and permits higher rates of embryo survival, recent reports have shown increased mutagenesis efficiency

upon injection of *in vitro* assembled Cas9-sgRNA ribonucleoprotein complexes (RNPs) (Gagnon et al.,2014; Kotani et al., 2015; Sung et al., 2014).

In 2016, Burger et al. used *in vitro* assembled Cas9-sgRNA ribonucleoprotein (RNPs) complexes to recapitulate loss of function phenotypes in zebrafish embryos (2016). *tbx16* is a T-box gene that is regulated by FGF signaling. In homozygous mutant spadetail (spt) embryos, *tbx16* is defective resulting in embryos displaying a broad posterior tail. This mutant phenotype is attributed to the failure to differentiate mesoderm progenitor cells from different lineages (Kimmel et al., 1989). When Burger et al. used RNPs to target spt, more than 60% of all embryos injected displayed a strong phenotype while almost 20% displayed a mild or partial phenotype. Using MiSeq along with a customized software tool, CrispRVariants, for quantifying and visualizing mutagenesis, it was shown that RNPs allow for efficient bi-allelic mutagenesis for both *golden* and spadetail (Burger et al., 2016).

Zebrafish are not the only model organism being used in CRISPR/Cas9 experiments. In 2014, Long et al. was able to correct a germline mutation in mouse embryos to ultimately prevent muscular dystrophy using the CRISPR/ Cas9 system. Duchenne muscular dystrophy (DMD) is a genetic disease that mainly effects males, although female cases do occur. On average, 1 in 3,500 boys are affected. DMD is caused by mutations to the gene that encodes for dystrophin, an important protein in muscle fibers. To-date, there is no cure for muscular dystrophy and those effected normally have a shortened life span. A transgenic line of mice was generated that possessed the mutation for DMD. Using the CRISPR/Cas9 system, a recent study looked at the possibility of correcting the nonsense mutation in mouse embryos in vivo. By injecting Cas9, gRNA and an HDR template into mouse zygotes, Long and colleagues were able to correct the nonsense mutation in the germ line of mice possessing the

mutated gene. The genetically altered animals contained 2 to 100% correction of the mutation, with the variation being due to mosaicism. However, the degree of phenotypic muscle rescue greatly exceeded the degree of gene correction. Genetically altered mice that had just 17% gene correction (due to mosaicism) displayed between 47-60% corrected dystrophin fibers as compared to mice suffering from DMD. This striking result could possibly allow for mice with only a small percentage of gene correction to live a near normal life as compared to those mice with 0% corrected dystrophin fibers as seen in DMD. Also observed throughout this experiment was that mice expressed higher and higher levels of gene correction over time, leading to the hypothesis that corrected genes are rescuing mutated genes as time proceeds. Important to note, all genetically altered mice developed to the adult stage with no signs of tumors or other side effects associated with the injections. This study presents a hopeful approach to promote muscle repair in patients with DMD (Long et al., 2014).

Current applications of the CRISPR/Cas9 system include, but are not limited to, changing sequences in pluripotent embryonic stem cells (González et al., 2014; Mali et al., 2013a; Zhu and Huangfu, 2013) and correcting genetic defects in animals (Long et al., 2014). While these studies are laying the foundation for refined approaches that could one day treat human disease and genetic defects, there are many bioethical questions to address before the first human germline engineering project can begin. While engineering the human germline is not a new idea, both excitement and unease have long surrounded the topic. To begin with, both the safety and efficiency of the technology must be insured. Next is the ethical question of whether or not it is appropriate to alter a disease-causing mutation to a more typical sequence? It is also critical to think about the on-target events that produce unintended

consequences, as well as the off-target effects of these modifications. Aside from the ethical questions and concerns, using the CRISPR/Cas9 system on human subjects would need to be approved by the Food and Drug Administration (FDA) in the United States (Baltimore et al., 2015).

In conclusion, genetic engineers have been able to insert and delete genes for decades; however, they were unable to dictate the precise location of these modifications. Around a decade ago, ZFNs and TALENs allowed DNA to be broken at specific targets but having to generate custom designed proteins is time consuming and expensive. The CRISPR/Cas9 system uses the same protein and only requires custom designed gRNAs, making it a much simpler and less expensive system. The benefits of the CRISPR/Cas9 system include efficiency, precise targeting, flexibility to be used in numerous systems and the economical price as compared to genetic engineering technology of the past (Doudna and Charpentier, 2014). The CRISPR system has already assumed a position beside TALENs and ZFNs in the genome engineering toolbox and the ease of retargeting the CRISPR/Cas9 system to modify genomic sequences far exceeds that of ZFNs and TALENs, while offering similar rates of efficiency (Mali et al., 2013b). Aside from the bioethical questions concerning the system, it could potentially revolutionize genome editing as it could be used to treat diseases in humans such as HIV, correct birth defects such as trisomy 21 (down syndrome), and more. The CRISPR system has quickly become one of the most versatile and powerful tools in genome engineering and with much more to learn, the possibilities are endless. Here I make use of the system to excise a regulatory element from the zebrafish genome to study the role it plays in regulating genes with which I believe it to be associated.

This project increases the knowledge and understanding of the regulatory mechanisms associated with the *Meis2* gene. The activity of the m2de1 element was observed in order to determine when and where the element is active during the development of zebrafish embryos through use of a reporter transgene. This pattern was then compared to the spatial and temporal expression patterns of *meis2a* and *zgc:154061* in order to visualize overlaps between the expression directed by m2de1 and the expression of both *meis2a* and *zgc:154061*. Upon observation of overlapping expression patterns, a portion of m2de1 was excised from the zebrafish genome and the expression patterns of *meis2a* and *zgc:154061* will be compared in wildtype embryos and m2de1 mutant embryos. As differing expression patterns were observed between wildtype and mutant embryos, insight has been gained into the regulatory mechanisms associated with the zebrafish *meis2a* gene.

Materials and Methods

Zebrafish husbandry

Zebrafish were cared for as described in *The Zebrafish Book: A guide for the laboratory use of zebrafish (Danio rerio)* (Westerfield, 2000) and all animal use was approved in accordance with the Appalachian State University Institutional Animal Care and Use Committee (IACUC) and. Fish were housed in a Marine Biotech Z-mod closed system (Aquatic Habitats, Apopka, FL). Within the system, fish were kept in one-liter tanks with a maximum of 7 adults per tank. Temperature was maintained between 26-28° C, pH was maintained between 6.8 and 7.2, while conductivity was kept between 450 and 600 microSiemens. These environmental parameters were monitored daily. Water hardness was monitored monthly and maintained between 120-200ppm. Fish were on a systematic light cycle of 14 hours of light and 10 hours of dark. During light hours, adult fish were fed dry food once a day before noon and live brine shrimp daily in the afternoon.

Breeding

In order to obtain embryos for injections, male and female fish were separated by a divider in a special breeding tank (Aquatic Habitats). By dividing the fish, we were able to control when the fish spawned. Fish remained divided overnight and the divider was removed at approximately 9:00 in the morning when the system lights first turned on. Females began to release eggs that were fertilized by the male fish and then fell through the mesh bottom of the breeding tank. Fifteen minutes after the divider was pulled, embryos were collected. This allowed time to ensure that embryos were still at the single cell stage for injections. Collected embryos were filtered through a fine mesh net and then rinsed with RO water to clean them from debris such as fecal matter.

The clean eggs were then placed into a petri dish containing 0.3x Danieau buffer (58 mM NaCl, 0.7 mM KCl, 0.4 mM $MgSO_4$, 0.6 mM $Ca(NO_3)_2$, 5 mM HEPES pH 7.6 for 1x solution). If the embryos were to be raised to adulthood without injections or imaging, they were placed in a 1x Danieau/ methylene blue solution (66.66µL 1% methylene blue per liter of 1x Danieau buffer) in order to prevent fungal growth. The embryos were then placed in a 28°C incubator for the first five days post fertilization. After this time, larval fish were moved into standard 1L tanks. The tank was filled with 50% 1x Danieau buffer and 50% system water and placed on a very slow drip of system water. The larval fish were fed dry food (Zeigler) twice a day based on their size. ZM-50 food was fed approximately between 5 dpf until 11 dpf. ZM-100 was fed approximately between 2 and 3 weeks post fertilization. ZM-200 was fed between 3 and 6 weeks fertilization before moving to ZM-300 and ZM-400 until approximately 3 months post fertilization. After approximately 3 months, zebrafish were fed standard adult food (Zeigler Adult Zebrafish Complete Diet). Once larval fish began eating ZM-200, they were also fed 2-day-old brine shrimp. Once larval fish began their diet of brine shrimp, they only received dry food once a day in the morning and live brine shrimp in the afternoon. After fish reached sexual maturity between 2-4 months of age, they were fed adult dry food in the morning and one concentrated drop of live brine shrimp per fish in the afternoon.

Transformations

In order to obtain sufficient amounts of plasmid DNA, it was first transformed into bacterial cells. To do this, approximately 25-50 ng of plasmid (no more than 2µL) was added to 50µL of chemically competent DH5α *E. coli.* The competent cells were then incubated on ice for 30 minutes before being heat shocked at 42°C of 45 seconds. Cells were then placed

back on ice for 5 minutes before carefully being added to 1mL of SOC medium (20g bacto-tryptone, 5g bacto-yeast extract, 0.5g NaCl, 20mM glucose) followed by an incubation period of 1 hour at 37°C and 180rpm. After the incubation, the cells were plated at volumes of 25µL, 50µL and 100µL on LB agar plates that contained the appropriate antibiotic for selection. Plates were then incubated overnight at 37°C before being wrapped in parafilm wax the next morning and being stored at 4°C.

Transcribing Transposase mRNA

Transposase mRNA was transcribed using the mMessage mMachine Sp6 kit (Ambion, Life Technologies, Grand Island, NY). The protocol provide by the manufacturer was used. Once the DNA had been transcribed into mRNA, the mRNA was cleaned and concentrated using the RNA Clean & Concentrator™-5 kit (Zymo Research Corporation) by following the protocol provided by the manufacturer. The transposase mRNA was then resuspended in RNAse free DEPC water and coinjected with the expression construct, dr.m2de1-pGW-cfos-EGFP, into single cell zebrafish embryos.

Generating Transgenic Embryos through Microinjections with Tol2 mRNA

Fertilized zebrafish embryos were collected within 15-30 minutes of mating. Embryos were washed with RO water and cleaned of any debris by using a Pasteur pipet. The clean embryos were then placed in a glass petri dish. Excess water was removed from the dish and replaced with Danieau buffer solution (50x 2.9 M NaCl, 35 mM KCl, 20 mM $MgSO_4$, 30 mM $Ca(NO_3)_2$, 250 mM HEPES pH 7.600, adjusted to 1L in RO water).

While adult zebrafish were mating, the injection solutions were thawed. Once thawed, 180 ng of transposase mRNA, 135 ng plasmid DNA (expression vector plasmid carrying

dr.m2de1-pGW-cfos-EGFP) were combined and brought to a final volume of 3µL with DEPC water. Needles for injections were pulled from a 3.5nl capillary that had been baked at a temperature of 260°C in order to insure the inactivity of RNases. The capillary tube was secured in a David Kopf Instruments Vertical Pipette Puller model 700C, where the heater was set to 54 volts and the solenoid set at 10 amps. The weight on the bottom of the machine was used to pull and divide the needle. The tip of the needle was then beveled using size 5 forceps and filled with mineral oil. The needle was then placed on the Nanoliter 2000 Microinjector (World Precision Instruments Model B203XVY) which was attached to a Marhauser MMJR Micromanipulator (World Precision Instruments). Using the controls on the micromanipulator, approximately 20-25% of the mineral oil was pushed out of the beveled needle to ensure that the tip was not clogged. The mineral oil was then replaced with the injection construct.

Once the construct filled needle was ready, between 40 and 60 single cell embryos were placed against a 1.0-millimeter-thick VWR micro slide (VWR international 48300-025) which was taped to the outside of a plastic petri dish to insure stabilization. Single cell embryos were then injected with approximately 1 nl of construct. To ensure that the construct was injected directly into the single cell of the developing embryo, a translucent bolus was observed as the construct was dispensed into the single cell. Inside the embryo, the mRNA was translated into transposase protein. The protein then uses the Tol2 sites (Kawakami et al., 2004; Kwan et al., 2007) on the expression construct in order to insert the expression construct into the zebrafish embryo genome.

Screening Tol2 Injected Embryos to Identify Transgenics

In order to generate a stable transgenic line of zebrafish using the m2de1 element to direct the expression of eGFP, primary injected embryos were screened for eGFP expression to confirm the presence of the construct. Previous experiments in primary injected embryos had noted eGFP expression between 48 and 54 hpf, so these time points were used for imaging in order to identify eGFP positive embryos from those embryos that were eGFP negative. Because embryos expressing eGFP were to be raised to adulthood for the purpose of breeding, a non-invasive method of screening was used. Embryos younger than 72 hours were dechorionated under a microscope using #5 watchmaker forceps. The chorions were then removed from the petri dish to avoid autofluorescence. Embryos were anasthesized in a Pyrex 60x15 mm petri dishes (Corning Like Sciences, New York) that contained about 6 ml of 0.3x Danieau buffer and 6 drops of 20% Tricaine suspended in a 0.3x Danieau buffer. The Tricaine served as an anesthesia for the embryos, which allowed them to be imaged under the fluorescent lamp of the microscope. A Zeiss LSM 510 Confocal Microscope was used to view and sort the embryos. Specifically, the 10x objective lens, the FITC filter and bright field were used to visualize the fluorescence from eGFP. Embryos that expressed eGFP were transferred into a separate dish specifically for eGFP positive embryos.

Imaging of Transgenic Embryos

Once eGFP positive embryos were sorted from eGFP negative embryos, a Zeiss LSM 880 Confocal Microscope was used to image the anesthetized embryos. The Argon laser was used to collect images of the fluorescent embryos. Image collection ranged from single slice pictures to Z-stacks. The quality of the pictures remained constant at 1024x1024 and the scanning speed ranged between 5 to 9. Size rulers were added to images via Zeiss' built-in

software and then using Adobe Photoshop, the green levels throughout the entire picture were revised in order to reduce autofluorescence.

Designing guide RNAs

To knockout a portion of m2de1 in zebrafish using the CRISPR/Cas9 system, gRNAs were designed using CRISPRscan developed by the Giraldez Lab at Yale University. In zebrafish, the m2de1 element is located on chromosome 17: 52,911,113-52,911,563 according to GRCz10/danRer10. This region was examined for potential gRNAs (Fig. 11). When examining potential gRNAs, the sequence in between two gRNAs was entered into the Patch 1.0 database on gene-regulation.com, in order to determine whether or not there were transcription factor binding sites between the two gRNAs (Fig. 12).

Transcribing guide RNAs

To generate templates for gRNA transcription, a specific oligonucleotide containing the T7 promotor sequence (5'-TAATACGACTCACTATA-3'), the target sequence (short guide oligo), and a complementary sequence were annealed to a guide constant (scaffolding) oligo (Table 2). Samples were prepared for the thermocyler containing 5μL 5x Phusion buffer, 0.5 μL dNTPs, 3μL of 10μM short guide oligo, 3 μL of 10 μM scaffolding oligo and 0.25μL of Phusion. The sample was brought to a total volume of 25μL with nuclease free water. The sample was then placed in the thermocycler. The following settings were used to carry out the polymerase chain reaction (PCR): initial denaturation of 98°C for 30 seconds followed by 45 cycles of normal denaturation of 98°C for 10 seconds, annealing at 60°C for 10 seconds and extension at 72°C for 15 seconds. Once 45 cycles were completed, the sample underwent a final extension at 72°C for 5 minutes. The samples then remained at 4°C

until they were removed from the thermocycler. Once the samples were removed from the thermocycler, they were purified using phenol chloroform extraction. The reaction volume was raised to 100µL using nuclease free water. An equal amount of Phenol/Chloroform/Isoamyl alcohol (25:24:1) was added to the sample. The sample was then mixed by gently flicking the tube before centrifuging for 5 minutes at 16,000 rpm. After centrifugation, there were two distinct layers in the tube. The aqueous phase containing the DNA was less dense than the organic phase, resulting in the aqueous phase being the top layer. The top layer containing DNA was transferred into a new Eppendorf tube and 100µL Chloroform/Isoamyl alcohol was added. The sample was mixed and centrifuged as described previously. The upper layer was again transferred into a new Eppendorf tube and 2.5 volumes of 100% ethanol was added. Sodium chloride was then added to the mixture at a total volume of 0.2M. Once salt and ethanol were added, the DNA was allowed to precipitate for at least 2 hours at -20°C. After incubation, the sample was centrifuged for 30 minutes at 16,000 rpm at 4°C. The ethanol was then decanted off, with remaining ethanol being removed via micropipette. To wash the pellet, 1mL of 70% ethanol was added to the Eppendorf tube. The sample was gently mixed by flicking the tube before being centrifuged one last time at room temperature at 16,000 rpm for 5 minutes. Again, the ethanol was decanted before removing small amounts of ethanol via micropipette. The Eppendorf tube was then left for 10-20 minutes to allow the remaining ethanol to evaporate off of the DNA. Once the pellet was dry, the DNA was resuspended with 10µL of nuclease free water. After allowing the sample to rehydrate, the DNA was used as a template to transcribe gRNAs.

Table 2: A template short guide oligo for gRNA synthesis.

Oligo Name	Sequence
Short Guide Oligo	AATTAATACGACTCACTATAGGNNNNNNNNNNNNNNNNN GTTTTAGAGCTAGAAATAGC
Guide Constant Oligo	AAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGAC TAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC

***** Guide oligos are made of three main parts: a T7 promoter, shown in blue, a variable targeting sequence, underlined, and an overlap with the guide constant oligo. A short clamp, shown in red, is provided to stabilize the 5' end of the sequence. Using PCR, the guide constant oligo, shown in purple, and short guide oligo are annealed together to generate a full-length template for gRNA synthesis. The guide constant oligo is used as a Cas9 binding scaffold.

gRNAs were then transcribed using the MAXIscript T7 Transcription Kit (Ambion, Life Technologies, Grand Island, NY) according to the manufacturer's protocol with two exceptions. To ensure that the transcribed gRNA was not immediately degraded, 1 µL of RNAse inhibitor (Promega N2515) was added to the transcription reaction. Also, to ensure that the majority of the template DNA was transcribed into gRNA, the reaction was incubated overnight rather than the recommended one hour. After the overnight incubation period, turbo DNAse was added and the sample was incubated for 15 more minutes. Afterwards, the gRNA was cleaned and concentrated using the RNA Clean & Concentrator™-5 kit (Zymo Research Corporation) by following the manufacturer's protocol. The gRNA was then resuspended in DEPC water and coinjected with EnGen® Cas9 NLS (NEB Mo646T) into single cell zebrafish embryos.

Generating Transgenic Embryos through Microinjections using the CRISPR/Cas9 system

Fertilized zebrafish embryos were collected and cleaned as previously described. While adult zebrafish were mating, the injection solutions were thawed. Once thawed, 50 ng/µl of each gRNA and 250 ng/µl of EnGen® Cas9 NLS (NEB M0646T) were combined and brought to a final volume of 4µL with nuclease free water. The construct was then incubated at 37°C for 10 minutes to allow the Cas9 and the gRNAs to reconstitute (Sternberg et al., 2014). A needle for injections was pulled and prepared as described previously. Once the needle and construct were ready, single cell embryos were aligned and injected as previously described. Each embryo was injected with approximately 1 nl of construct.

DNA Isolation from Zebrafish Fin Clips

In order to isolate genetic material from adult fish for the purpose of identifying transgenesis, a small amount of tissue was clipped from the distal tip of the caudal fin. The tissue was then used to extract DNA which was used for further analysis using PCR. Fish were anesthetized with ice cold water. Once the fish was no longer responsive, but still had gill movement, sterile tweezers and scissors were used to hold and clip the caudal fin. The fin clip was then transferred into 50µl of activated genomic extraction buffer (10mM tris pH 8.0, 10mM EDTA, 200 mM NaCl, 0.5% SDS, 200 µg/ml Proteinase K in sterile RO water right before use). Fish were quickly returned to a recovery tank containing system water and a bubbler. Once they began swimming, fish were placed back in the Marine BioTech ZMod and maintained in isolation for 2 weeks or until at least 50% of the fin has regenerated.

The tissue in the genomic extraction buffer was then incubated in a shaking incubator at 56°C and 100 rpm for a minimum of three hours or until tissue was completely dissolved. Once the tissue was completely dissolved, 2 volumes, or 100µL of 100% ethanol was added to the microcentrifuge tube and stored at -20°C overnight. The next morning, the tissue was

centrifuged for 10 minutes at 13,000 rpm, afterwards the supernatant was decanted off, being careful not to disturb the visible pellet. To wash the pellet, 200µL of 70% ethanol was added to the tube and briefly vortexed. The tube was again placed in a microcentrifuge for 2 minutes at 13,000 rpm. The supernatant was removed once again and the pellet was left to dry. Once the pellet was dry, it was resuspended in 20µL of TE+RNAse buffer (10mM Tris, 1 mM EDTA pH 8.0, RNAse in sterile RO water at a concentration of 100µg/mL) and incubated at 37°C for 1 hour. The genomic DNA was then cleaned of contaminants using Phenol/Chloroform/ Isoamyl alcohol.

DNA extraction from zebrafish embryos

In order to isolate genetic material from zebrafish embryos for the purpose of identifying transgenesis, larvae were first euthanized with Tricaine (4% in Danieau buffer) between 48 and 72 hpf. Larvae were then washed three times with sterile RO water. Once washed, 5 larvae were placed in an Eppendorf tube containing 50µl of activated genomic extraction buffer (10mM tris pH 8.0, 10mM EDTA, 200 mM NaCl, 0.5% SDS, 200 µg/ml Proteinase K in sterile RO water right before use). The protocol from DNA Extraction from Zebrafish Fin Clips (pages 41-42) was then followed for the remainder of the DNA extraction.

Screening Primary Injected CRISPR/ Cas9 Fish to Identify Transgenics

In order to determine whether a portion of m2de1 was knocked out of the genome in m2de1 CRISPR injected embryos, DNA was isolated from 1/3 of a clutch (approximately 15 embryos) of injected embryos at 36 hours post fertilization. DNA was amplified using specific primers (Table 3) to amplify the entire m2de1 sequence. Samples were prepared for

the thermocyler containing 50ng of genomic DNA, 2.5µL 10x Thermopol buffer, 0.5 µL dNTPs, 1µL of 10µM Dr-m2de1-3 primer, 1 µL of 10µM Dr-m2de1-5b , and 0.125µL of Taq Polymerase (NEB Labs). The sample was brought to a total volume of 25µL with nuclease free water. The sample was then placed in the thermocycler. The following settings were used to carry out the polymerase chain reaction (PCR): initial denaturation of 95°C for 5 minutes followed by 30 cycles of normal denaturation of 95°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 68°C for 1 minute and 45 seconds. Once 30 cycles were completed, the sample underwent a final extension at 68°C for 10 minutes. The samples then remained at 4°C until they were removed from the thermocycler. Samples were then run on a 1% agarose gel at 90 volts for approximately 45 minutes to determine the fragment size.

Table 3: m2de1 Primers

Oligo Name	Sequence
Dr-m2de1-3	GCTCATTATAAGGCCGTGCATG
Dr-m2de1-5b	TATACCATGGAGGTCGGGTTTAAAGGAG

Generating Anti-sense and Sense RNA Probes

Previously, through an NIH initiative and the NIH Mammalian Gene Collection (MGC) project, the full length (1914 basepairs) cDNA clone of *zgc*:*154061* was isolated by the Zebrafish Gene Collection (ZGC) (Strausberg et al., 2002). The clone was purchased from Open Biosystems (Clone ID: 8334609, Accession: BC124527). To generate an anti-sense probe, the plasmid was linearized with SalI. Similarly, the cDNA for the *meis2a* gene was

also purchased from Open Biosystems (Clone ID: 6789004, Accession: BC056515) and subcloned into pCMV-sport6.1. The plasmid was then linearized with EcoRI to generate an anti-sense probe. To generate sense probes, the plasmid containing *zgc:154061* was linearized with Xho1 while the pCMV-sport6.1 containing *meis2a* was linearized with Not1.

After linearization, the samples were purified through Phenol/Chloroform precipitation. The purified product was then used as a template for generating the anti-sense RNA probe. The anti-sense DIG-labeled RNA probe was transcribed using the MAXIscript T7 Transcription Kit (Ambion, Life Technologies, Grand Island, NY) according to the protocol provided by the manufacturer, with three exceptions. Similarly, the sense DIG-labeled RNA probe was transcribed using the mMessage mMachine Sp6 kit (Ambion, Life Technologies, Grand Island, NY) according to the protocol provided by the manufacturer, with the same three exceptions. The first exception to the protocol was using DIG RNA Labeling Mix, 10x concentration (Roche) in the anti-sense probe transcription reactions rather that the uracil nucleotide from the MAXIscript T7 Transcription Kit. Likewise, DIG RNA Labeling Mix, 10x concentration (Roche) was added to the sense probe transcription reactions in addition to the 2x NTP/CAP provided in the mMessage mMachine Sp6 kit. Secondly, to ensure that the transcribed RNA was not immediately degraded, 1 μL of RNAse inhibitor (Promega N2515) was added to the transcription reaction. The last exception was to ensure that the majority of the template DNA was transcribed into RNA. For this purpose, all reactions were incubated overnight rather than one hour. After the overnight incubation period, turbo DNAse was added and the sample was incubated for 15 more minutes. Afterwards, the RNA was cleaned and concentrated using the RNA Clean & Concentrator™-5 kit (Zymo Research Corporation) by following the protocol provided by the manufacturer.

Afterwards, the RNA was resuspended in 15µL nuclease free water, 137µL of hybridization buffer (500 mL formamide, 250 mL 20x SSC, 10 mL 10% Tween-20, 10 mL 0.9 M Na Citrate stock, 10 mL of 50 mg/mL torula tRNA stock, 1 mL of 50 mg/mL heparin, 219 mL DI H_2O) was added to the sample. The resulting RNA was stored at -20°C.

Fixing Embryos

Prior to performing in situ hybridization, embryos were fixed at 12, 18, 24, 36, 48, 54, 60, 66 and 72 hours post fertilization (hpf). Embryos younger than 18 hpf were fixed in an Eppendorf tube while still in their chorions. However, embryos older than 18 hpf were first dechorionated using size 5 forceps. After dechorionating older embryos, between 30 and 50 embryos were placed in an Eppendorf tube. Excess Danieau was removed from the tube and replaced with approximately 750µL of 4% paraformaldehyde (PFA). Eppendorf tubes were then placed on a shaker (Labrat Gyrotwister) at a speed between 45-50 RPM for ten minutes. After ten minutes, PFA was removed and replaced with 750µL of fresh PFA. Embryos were then left on the shaker overnight at 4°C or for 4 hours at room temperature. Afterward, embryos were transferred into a petri dish containing 1x PBS (8g NaCl, 0.2g KCl, 1.44g Na_2HPO_4, 0.24g KH_2PO_4, brought to 1L with DI H_2O). Embryos that had not been previously dechorionated were dechorionated at this point. Embryos older than 24 hpf began to develop melanocytes and were depigmented at this time by transferring embryos to a 3% H_2O_2 solution 1 mL 30% H_2O_2, 0.05g KOH, brought to 10mL with DI H_2O). Embryos remained in the solution for approximately 30 minutes or until the pigment disappeared from the head/ eyes. After depigmenting, embryos were transferred into a clean petri dish containing fresh 1x PBS in order to remove the bubbles from the H_2O_2. Embryos were then transferred into a new Eppendorf tube and PBS was removed and replaced with

approximately 800µL 50% MeOH/ 50% PBS. The tubes containing embryos were then placed on the shaker table as previously described for 5 minutes. The solution was removed after 5 minutes and replaced with 100% MeOH and placed on the shaker table for 5 minutes. This was repeated 3 times before storing the fixed embryos at -20°C.

A second, slightly different method of fixing embryos was also used at times. For this method, rather than letting embryos develop melanocytes, embryos were raised in a 1-Phenyl-2-Thiourea (PTU) solution. PTU was dissolved in a dimethyl sulfoxide (DMSO) solution (7.813 grams DMSO brought to a final volume of 100 mL with RO water) (300µL PTU in 10mL DMSO). The PTU/DMSO solution was added to 1x Danieau buffer at a final concentration of 0.003%. This allowed embryos to continue to develop in the PTU solution and melanocytes were prevented from arising. Embryos were then dechorionated and fixed in 4% PFA as previously described. After fixation, the PFA solution was removed and embryos were placed in a PBS solution on a shaker table at room temperature for 5 minutes. After 5 minutes, the PBS was removed and replaced with fresh PBS. After a second 5 minute incubation on the shaker table, embryos were slowly dehydrated in a 50% MeOH/ 50% PBS solution before going into a 100% MeOH solution and stored at -20°C.

After rehydrating embryos, embryos were digested with Proteinase K. Embryos between 24 and 30 hpf were digested with Proteinase K at a concentration of 10ug/mL for approximately 16 minutes. Embryos over 30 hpf were digested with Proteinase K at a concentration of 10ug/mL (10ug Proteinase K in 1 mL PBT) for 30 minutes. During digestion, embryos were placed on a shaker table at room temperature with gentle agitation. After digestion, the Proteinase K solution was removed and replaced with 4% PFA. Embryos were placed on a shaker table in the PFA solution for 20 minutes. After the incubation, the

PFA solution was removed and replaced with PBT. Embryos were placed on a shaker table with gentle agitation for 5 minutes. The PBT wash was repeated 2 times. After two PBT washes, embryos are ready to begin the prehybridization step where they are incubated with 500µL of the preheated prehybridization buffer in the water bath at 65°C for 2-5 hours. This is described below.

In situ Hybridization

Once embryos were fixed, whole mount *in situ* hybridization was performed. This is normally completed over a three-day period. On day 1, both the prehybridization buffer (500 mL formamide, 250 mL 20x SSC, 10 mL 10% Tween-20, 10 mL 0.9 M Na Citrate stock, 230 mL DI H_2O) buffer and the hybridization buffer (500 mL formamide, 250 mL 20x SSC, 10 mL 10% Tween-20, 10 mL 0.9 M Na Citrate stock, 10 mL of 50 mg/mL torula tRNA stock, 1 mL of 50 mg/mL heparin, 219 mL DI H_2O) were preheated at 65°C for approximately 1 hour. In order to rehydrate previously fixed embryos stored in 100% MeOH, embryos went through 3 wash steps. The first wash step was with 50% MeOH/ 50% PBS. While in this solution, embryos were placed on a shaker table as previously described for 5 minutes. The solution was removed and replaced with PBT (0.1% Tween-20 in 1x PBS) and again placed on a shaker table for 5 minutes. This wash step was performed three times, ensuring that all PBT was removed. At this point, embryos that were bleached were ready to begin prehybridization.

After rehydration, embryos that were grown in PTU were digested with Proteinase K. Embryos between 24 and 30 hours hpf were digested with Proteinase K at a concentration of 10ug/mL for approximately 16 minutes. Embryos over 30 hpf were digested with Proteinase K at a concentration of 10ug/mL (10ug Proteinase K in 1 mL PBT) for 30 minutes. During

digestion, embryos were placed on a shaker table at room temperature with gentle agitation. After digestion, the Proteinase K solution was removed and replaced with 4% PFA. Embryos were placed on a shaker table in the PFA solution for 20 minutes. After the incubation, the PFA solution was removed and replaced with PBT. Embryos were placed on a shaker table with gentle agitation for 5 minutes. The PBT wash was repeated 2 times. After two PBT washes, embryos are ready to begin the prehybridization step

Once bleached embryos were rehydrated and embryos grown in PTU were rehydrated and digested, 500µL of the preheated prehybridization buffer was added to the embryos. Embryos were incubated in the water bath at 65°C for 2-5 hours. After incubation, the buffer was removed and 200µL of the hybridization buffer containing the probe at a concentration of 1:100 was added. The embryos were left in the buffer overnight at 65°C. The following morning, the prehybridization buffer, 2x SSC (17.53g of NaCl, 8.82g of sodium citrate in 1L DI H_2O, pH 7.0) and 0.2x SSC were preheated to 65°C. Once solutions were preheated, the hybridization buffer containing the probe was removed and stored at -20°C. Approximately 500µL of prehybridization buffer was added and incubated at 65°C for 60 seconds. The buffer was removed and replaced with 500µL of 50% prehybridization buffer/ 50% 2x SSC and this was then incubated at 65°C. The buffer was removed after 45 minutes and replaced with 500µL 2x SSC. This was then incubated for 15 minutes at 65°C. The buffer was removed after incubation and replaced with 500 uL 0.2x SSC and incubated for 1 hour at 65°C.

Following the previous wash steps at 65°C, the last buffer containing 0.2x SSC was removed and 1mL of PBT was added. The Eppendorf tubes were placed on a shaker table to room temperature and agitated for approximately 5 minutes. After the incubation, the PBT was

removed and replaced with 1mL of fresh PBT and incubated on the shaker table at room temperature. These PBT washes were repeated a total of three times. After the last wash with PBT, 500µL of blocking solution (2% goat serum, 2 mg/mL BSA in PBT) was added to embryos, which are then placed on a shaker table for 1 hour at room temperature. Following blocking, the buffer was removed and replaced with 500µL of antibody solution (1µL of Anti-Digoxigenin-AP-Fab fragments (Roche) per 1mL of blocking solution). Embryos were left overnight at 4°C on a shaker table with gentle agitation.

On the following day, the antibody solution was first removed from the Eppendorf tube. The embryos were then washed quickly with 500µL of PBT, which was then immediately removed. After the quick wash, embryos were washed six times with 500µL of PBT, with each wash lasting for 15 minutes on the shaker table. Next, the embryos were washed with 500µL alkaline phosphatase buffer (500 µL 1M Tris 9.5, 250 µL 1M $MgCl_2$, 100 µL 5M NaCl, 25µL 20% Tween-20, 4.125 mL H_20). Each wash with alkaline phosphatase buffer lasted 5 minutes on the shaker table. A total of 3 alkaline phosphatase buffer washes occurred. Following the last wash, 300µL of alkaline phosphatase buffer containing color substrate (3.5 µL BCIP & 4.5 µL NBT per 500 µL AP buffer) was added to the embryos. The color was allowed to develop in the embryos while keeping the embryos in a dark space. Embryos were checked for color development periodically. If color was developing slowly, embryos were left in color substrate solution overnight. Once color had developed, the reaction was quenched. To quench the color, the color substrate was removed and 500µL of PBT was added. Embryos were rinsed three times in PBT. Each rinse lasted for approximately 5 minutes and the embryos were on the shaker table with gentle agitation during each wash. After the last wash, 500µL of 100% methanol was added. Embryos were

incubated at room temperature on a shaker plate for 5 minutes before removing methanol and replacing with 500µL of fresh 100% methanol. Embryos were stored in methanol at -20°C.

Imaging in situ hybridization

To examine spatial and temporal expression of *meis2a* and *zgc:154061*, embryos were suspended in a 75% glycerol/ 25% PBS solution overnight. The following day, embryos were mounted on a 1.0 mm thick microslide (VWR 48312-004) before covering the sample with a 1-ounce 22x22 mm microscope cover glass (VWR 16004-094). Images were taken using an Olympus IX81 inverted microscope. Images were then processed with Olympus cellSens Software.

Results

Expression driven by Dr-m2de1

In order to observe when and where the m2de1 element was active in developing zebrafish embryos, single cell zebrafish embryos were injected with pDr m2de1-F-*cfos*-EGFP and transposase mRNA. Embryos were then imaged through 72 hpf. It was observed that between 48 hpf and 54 hpf, the m2de1 element was actively directing the expression of eGFP. At 48 hpf embryos displayed eGFP in the forebrain, midbrain and hindbrain, as well as muscle fibers in the head and the trunk (Fig. 5A, B). When embryos were imaged at 54 hpf, the expression was observed in these same areas (Fig. 5C, D). The expression in each of these areas has been observed previously in primary injected embryos (Barrett, 2013; Ferrara, 2015).

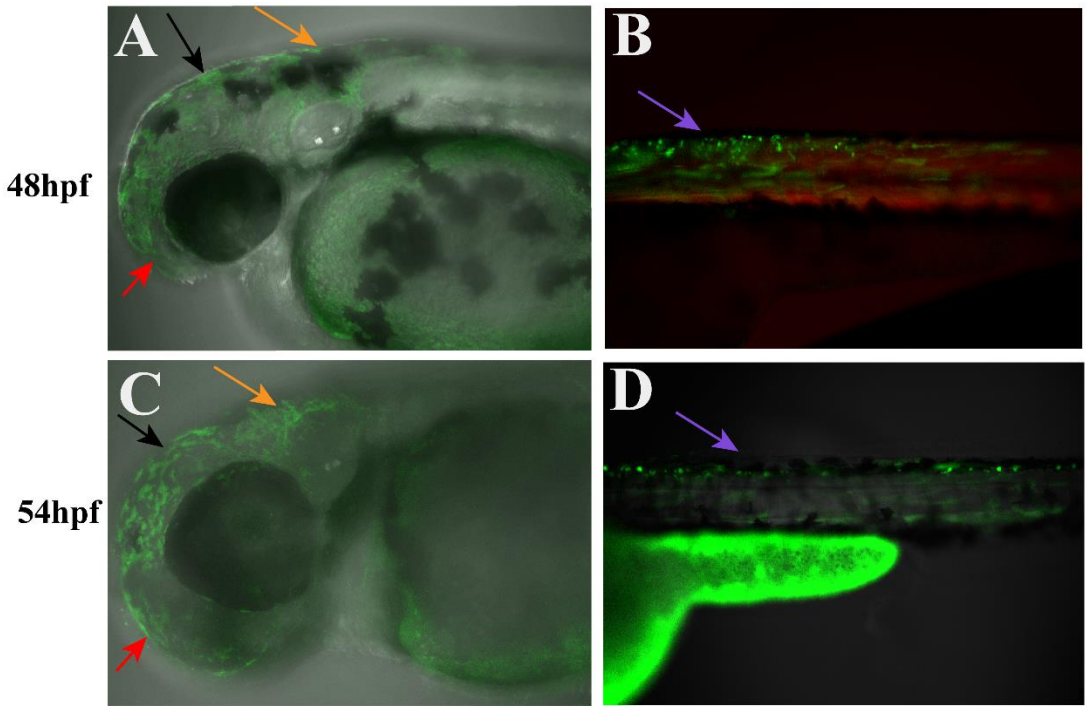

Figure 5: eGFP expression driven by Dr-m2de1 in the forward orientation (5'-3') at 48 hpf and 54 hpf. eGFP expression is seen in the developing **(A)** forebrain (red arrow), midbrain (black arrow) and hindbrain (orange arrow) and **(B)** muscle fibers of the trunk (purple arrow) of 48 hpf embryos. At 54 hpf, expression can be seen in the **(C)** developing forebrain (red arrow), midbrain (black arrow) and hindbrain (orange arrow), as well as, **(D)** the trunk (purple arrow). All embryos are positioned so that the anterior is to the left and the dorsal is to the top.

Spatial and Temporal Expression of meis2a and zgc:154061

After observing when and where the m2de1 element was active during development, the spatial and temporal expression of *meis2a* and *zgc:154061* was observed in order to determine whether the expression directed by the element overlapped with the expression of either gene. Previously, spatial and temporal expression of *meis2a* has been reported from the 2-somite stage until 48 hpf (Bessa et al., 2008; Carpenter et al., 2016; Thisse and Thisse, 2005; Waskiewicz et al., 2001). Similarly, spatial and temporal expression of *zgc:154061* has been reported from 1.5 hpf until 48 hpf (Carpenter et al., 2016). In order to observe the expression patterns of *meis2a* and *zgc:154061* that have been previously described (Bessa et

al., 2008; Carpenter et al., 2016; Thisse and Thisse, 2005; Waskiewicz et al., 2001), *in situ* hybridization was completed at 24 hpf. At this time, *in situ* hybridization shows overlaps in the expression of both *meis2a* and *zgc:154061*. Both genes are restricted anteriorly and show expression in the telencephalon and diencephalon (forebrain), optic tectum, cerebellum (midbrain) and rhombencephalon (hindbrain) (Fig. 6A-D).

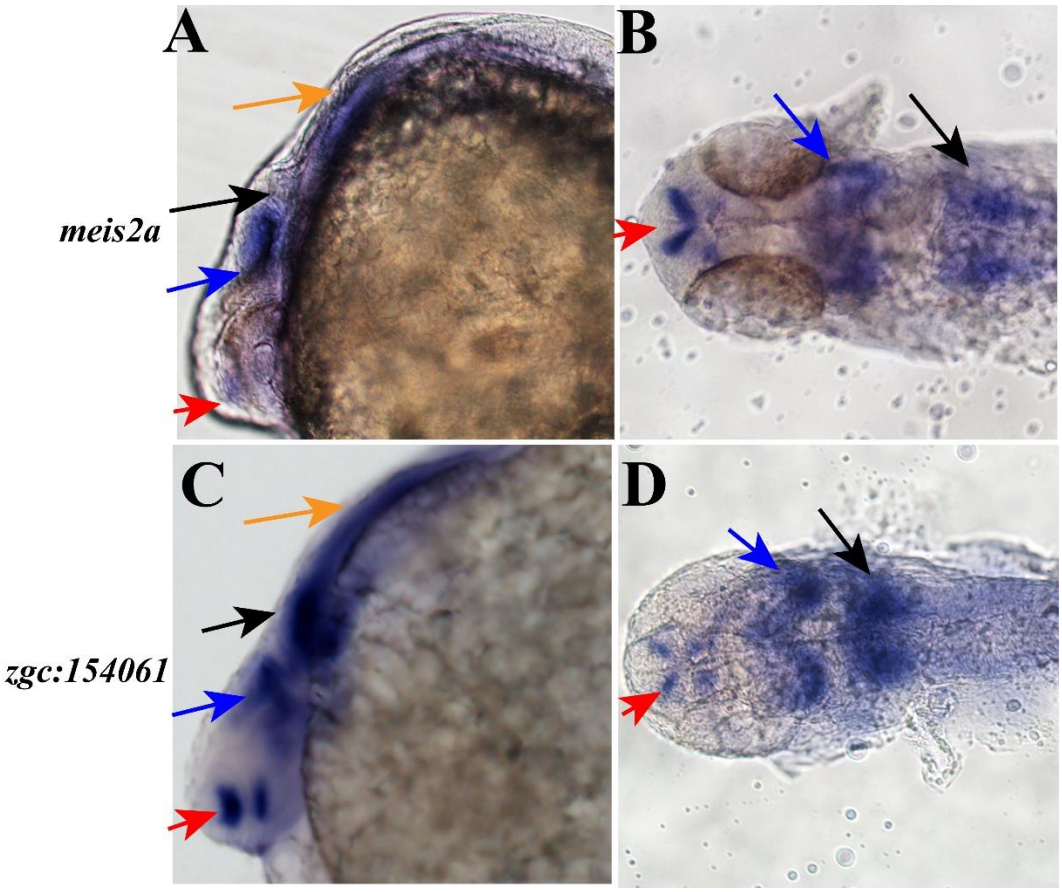

Figure 6: Whole mount *in situ* hybridization on 24 hpf zebrafish embryos to visualize *meis2a* and *zgc:154061* expression. *meis2a* and *zgc:154061* expression was observed to be localized in the forebrain (red arrow), optic tectum (blue arrow), cerebellum (black arrow) and rhombencephalon (orange arrow). **(A, B)** embryos displaying *meis2a* expression in the (A) lateral orientation and the (B) dorsal orientation. **(C, D)** embryos displaying *zgc:154061* expression in the (C) lateral orientation and the (D) dorsal orientation. All embryos are positioned so the anterior is to the left.

Between 48 and 54 hpf, both *meis2a* and *zgc:154061* exhibit overlapping expression with that directed by m2de1. At 48 hpf, *meis2a* (Fig. 7A, D) has an expression pattern similar to *meis2a* expression at 24 hpf. Expression was still seen in the telencephalon and diencephalon (forebrain), optic tectum, cerebellum and rhombencephalon. In contrast, at 48 hpf, *zgc:154061* exhibits less localized expression and seems to be expressed at lower levels. Expression was seen in the forebrain and retina (Fig. 7B, E). Similarly, at 48 hpf, m2de1 directed expression to areas around the forebrain, the eye and the midbrain (Fig. 7C, F).

At 54 hpf, the expression of both genes still overlapped with the expression directed by m2de1. At this time point, *meis2a* became more anteriorly restricted. Expression of *meis2a* was still observed in the forebrain, optic tectum and cerebellum, while expression in the retina was also present at this time (Fig. 7G, J); however, the expression observed in the rhombencephalon at earlier time points was no longer present. While *zgc:154061* still lacked definitive localized expression, low levels of expression were observed around the forebrain, the retina and the cerebellum (Fig. 7H, K). Again, the expression of both genes overlapped with the expression directed by m2de1 at 54 hpf. The expression directed by m2de1 at 54 hpf was very similar to the pattern observed at 48 hpf where expression of eGFP was observed around the forebrain, the eye and the midbrain (Fig 7I, L).

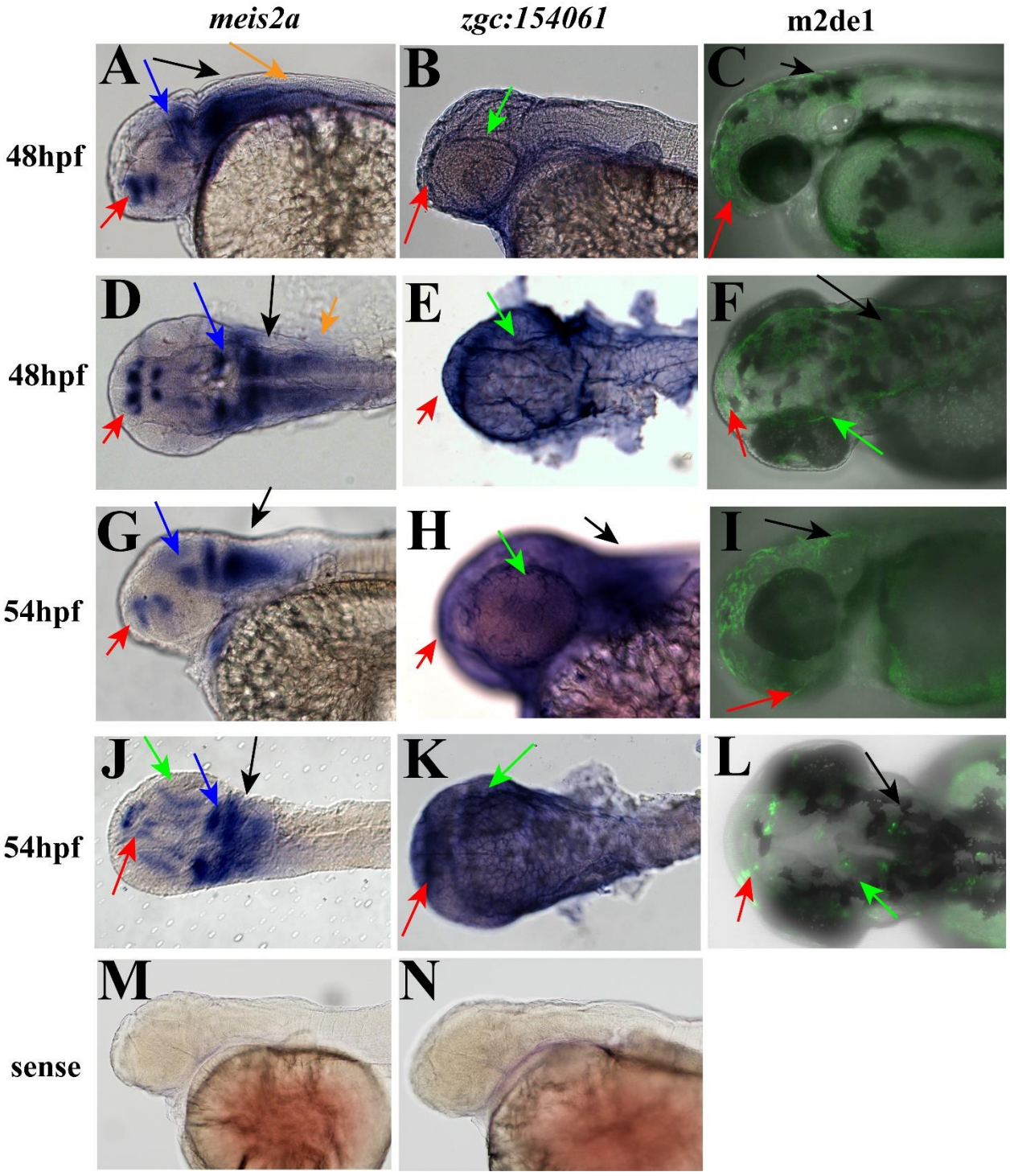

Figure 7: Whole mount *in situ* hybridization on 48 hpf and 54 hpf wildtype zebrafish embryos to visualize *meis2a* and *zgc*:*154061* expression and expression directed by m2de1 in live transgenic embryos of the same age. *meis2a* and *zgc:154061* expression was observed to overlap the expression directed by m2de1 in 48 and 54 hpf embryos. At 48 hpf, **(A, D)** embryos displayed *meis2a* expression in the forebrain (red arrow), optic tectum (blue arrow), cerebellum (black arrow) and rhombencephalon (orange arrow) **(B, E)** embryos

displayed *zgc:154061* expression in the forebrain (red arrow) and retina (green arrow), while **(C, F)** m2de1 directed expression to areas around the forebrain (red arrow), eye (green arrow) and midbrain (black arrow). At 54 hpf, **(G, J)** *meis2a* was expressed in the forebrain (red arrow), retina (green arrow), optic tectum (blue arrow) and cerebellum (black arrow) **(H, K)** *zgc:154061* was expressed in the forebrain (red arrow), retina (green arrow) and cerebellum (black arrow), while **(I, L)** m2de1 continued to direct expression to areas around the forebrain (red arrow), eye (green arrow) and midbrain (black arrow). Sense RNA probes served as negative controls at 48 hpf for **(M)** *meis2a* and **(N)** *zgc15406.* All embryos are positioned so the anterior is to the left. Images A-C, G-I and M-N are in the lateral orientation and images D-F and J-L are in the dorsal orientation.

After 54 hpf, the expression directed by m2de1 was no longer detected. However, *meis2a* and *zgc:154061* still displayed overlapping expression patterns, specifically within the forebrain. Both genes are restricted anteriorly between these time points. At 60 hpf, *meis2a* was no longer observed in the retina as it was at 54 hpf; however, expression was still observed in the forebrain, optic tectum and cerebellum (Fig. 8A, G), while embryos for *zgc:154061* at 60 hpf displayed expression around the retina and forebrain (Fig. 8D, J). At 66 hpf, *meis2a* expression was still observed in the forebrain, optic tectum and cerebellum (Fig. 8B, H). At the same time, z*gc:154061* was expressed around the forebrain and retina (Fig. 8E, K). At 72 hpf *meis2a* was observed in a similar pattern as observed at 66 hpf. At this time, *meis2a* was still observed in the forebrain, optic tectum and cerebellum (Fig. 8C, I) while *zgc:154061* expression was observed in the forebrain and retina (Fig. 8F, L).

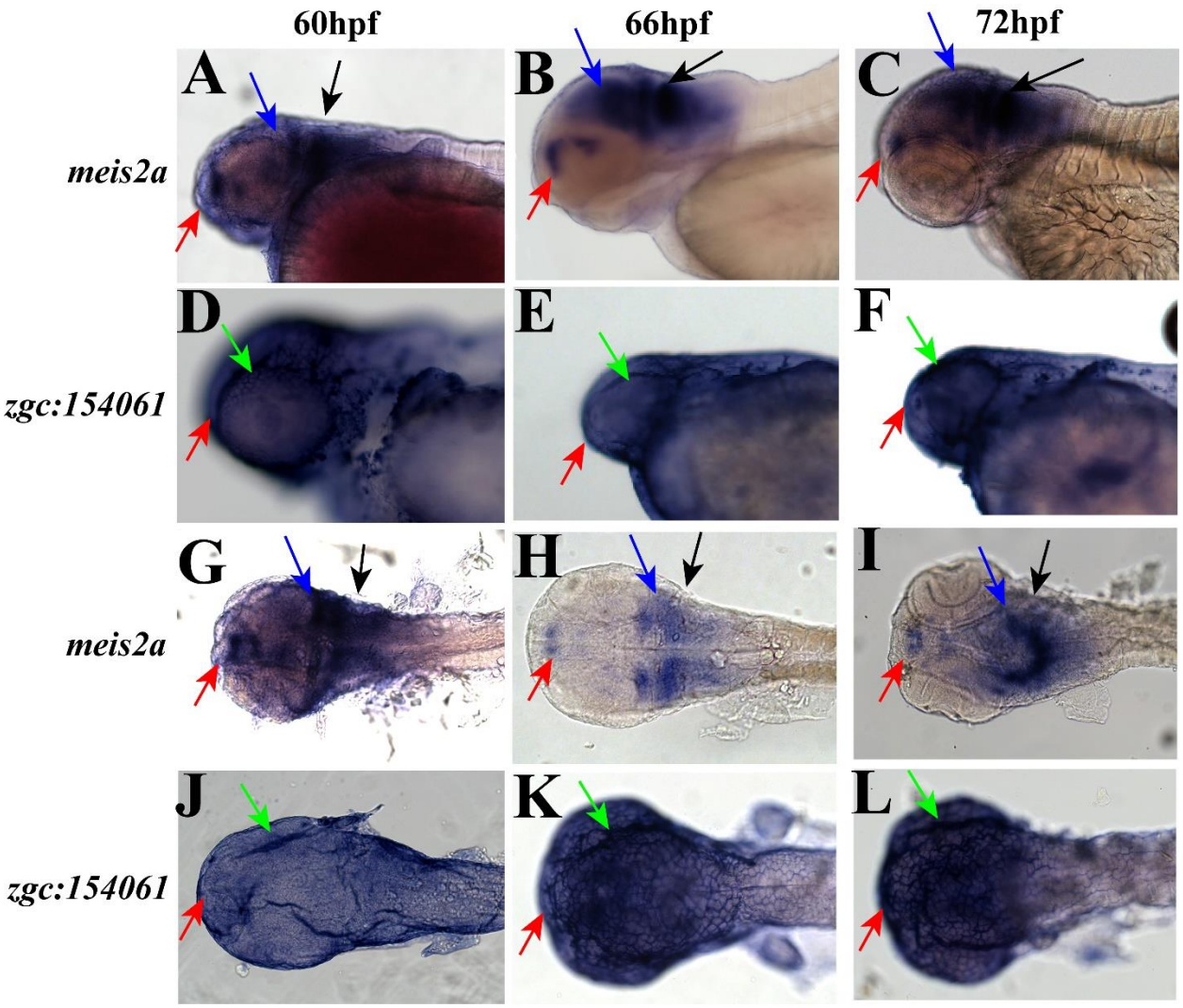

Figure 8: Whole mount *in situ* hybridization to visualize *meis2a* and *zgc:154061* expression. At 60 hpf, **(A, G)** *meis2a* expression can be seen the forebrain (red arrow), optic tectum (blue arrow) and cerebellum (black arrow) **(D, J)** while *zgc:154061* expression is seen in the forebrain (red arrow) and retina (green arrow). At 66 hpf **(B, H)** *meis2a* expression is observed in the forebrain (red arrow), optic tectum (blue arrow) and cerebellum (black arrow) and **(E, K)** *zgc:154061* expression is observed in the forebrain (red arrow) and retina (green arrow). At 72 hpf, **(C, I)** *meis2a* is observed in the forebrain (red arrow), optic tectum (blue arrow) and cerebellum (black arrow) while **(F, L)** *zgc:154061* is observed in the forebrain (red arrow) and retina (green arrow). All embryos are positioned so the anterior is to the left. Embryos A-F are positioned in the lateral orientation while G-L are in the dorsal position.

CRISPR Injections using a positive control

To further study the potential role of m2de1 in the expression of *meis2a* and/or *zgc:154061,* CRISPR/Cas9 was utilized to attempt to remove the element from the zebrafish genome. As a first step towards this, several approaches were used to optimize this approach, including positive control injections, Cas9 mRNA and Cas9 protein. While Burger et al. previously maximized zebrafish mutagenesis with solubilized CRISPR/Cas9 RNPs, generating such proteins can be both timely and costly. In our initial test to assess the efficacy of microinjections using CRISPR, we used a known genomic sequence to design a gRNA targeting the *spadetail* gene (Supplementary Table 1). Spadetail gRNA and Cas9 mRNA were co-injected into single cell zebrafish embryos at a range of different concentrations (Supplementary Table 2). After allowing embryos to reach approximately 24 hpf, the number of live embryos was counted and recorded. Living embryos were examined for the spadetail phenotype (Burger et al., 2014). Although different concentrations of gRNA and Cas9 mRNA were injected into the different clutches, no embryos displaying the spadetail phenotype when examined.

Due to the absence of the spadetail phenotype after our initial test, we co-injected the spadetail gRNA and Cas9 protein at a variety of concentrations (Table 4). At approximately 24 hpf, we recorded the number of living embryos and examined them for the spadetail phenotype (Figure 9F). All clutches contained embryos displaying the spadetail phenotype (Table 4). Embryos that were injected with 50 pg/nl of gRNA and 250 pg/nl of Cas9 protein (n= 195) showed the highest mutation rate (87.2%).

To further test the efficacy of Cas9 mRNA versus commercially obtained Cas9 protein in the CRISPR/Cas9 system, we used a second known genomic sequence to design a gRNA targeting the *golden* pigmentation gene (Supplementary Table 1). Golden gRNA and Cas9 protein were co-injected into single cell embryos at varying concentrations (Table 5). Between 36 and 48 hpf, embryos were examined for hypopigmentation in skin melanophores and retinal pigment epithelium (RPE) (Fig. 9B, D). All clutches contained embryos displaying hypopigmentation (Table 5). Again, embryos that were injected 50 pg/nl of gRNA and 250 pg/nl of Cas9 protein (n= 287) showed the highest mutation rate (82.0%). The similar results in terms of efficacy observed using gRNAs targeting the *tbx16* and *golden* genes suggested that the coinjection of 50pg/nl gRNA and 250pg/nl of commercially obtained Cas9 protein represented an optimal condition for subsequent experiments.

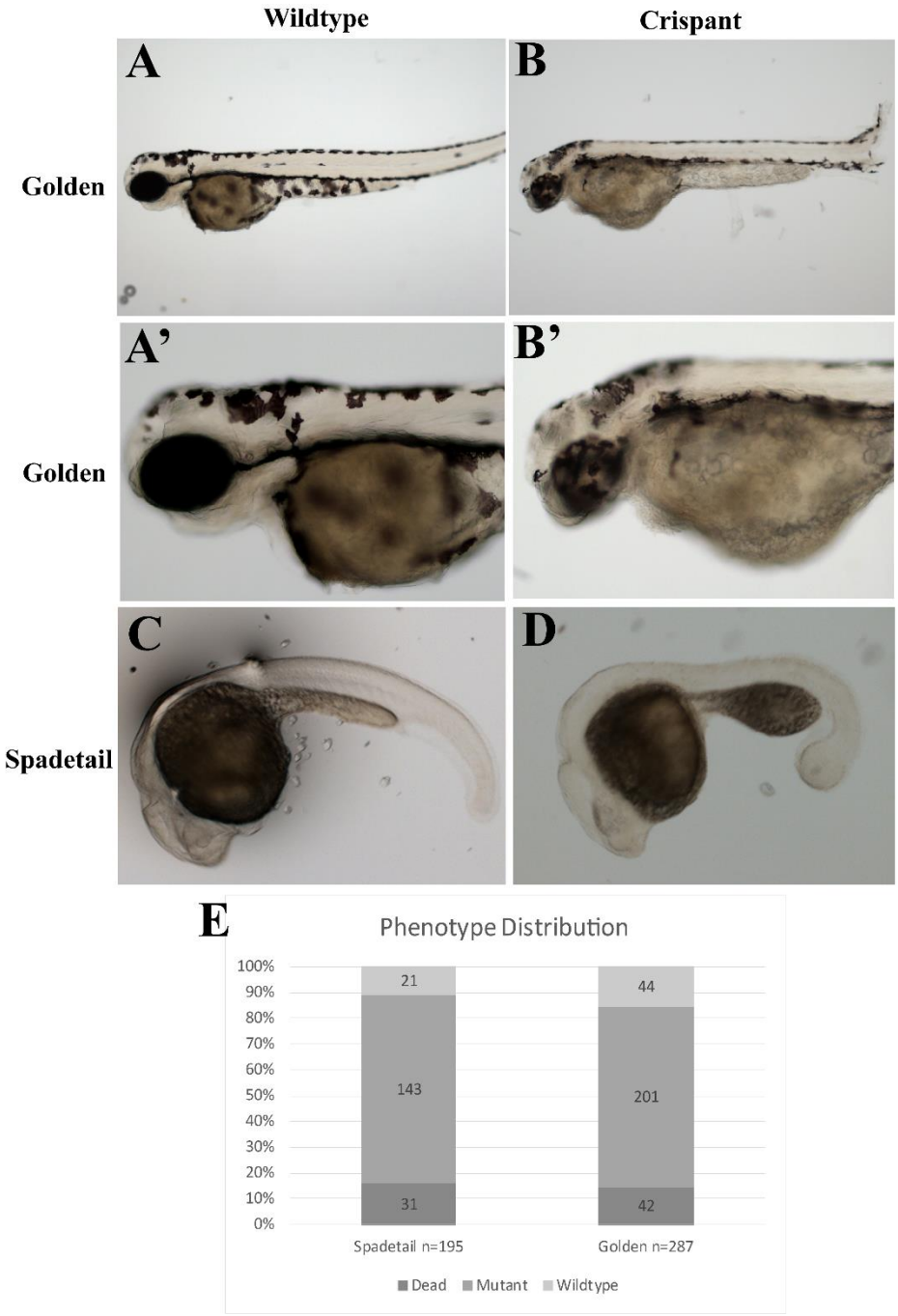

Figure 9: Mutant phenotypes in CRISPR injected embryos. (A-B') Mutagenesis of *golden* (*gol)* pigmentation gene encoding slc24a5 **(A, A')** wildtype uninjected control **(B, B')** targeting *golden* with Cas9 protein generates a phenotype displaying mosaicism within the pigmented cells **(C-D)** Mutagenesis of *tbx16* generating the spadetail phenotype **(C)** wildtype uninjected control **(D)** targeting *tbx16* with Cas9 protein generates a phenotype displaying a broad posterior tail **(E)** Observed mutagenesis rates of *golden* and spadetail when injecting 50 pg/nl of gRNA and 250 pg/nl of Cas9 protein into single cell embryos. All embryos are positioned in the lateralorientation and so that the anterior is to the left.

Table 4: Embryos Injected with Spadetail gRNA + Cas9 Protein

Concentration gRNA/ Cas9 (pg/nl)	100/250	50/250	25/250	12.5/250
Number injected	73	195	104	112
Number alive	36	164	86	84
Number w/ phenotype	8	143	63	39
% successfully targeted	22.2%	87.2%	73.3%	46.4%

Table 5: Embryos Injected with Golden gRNA + Cas9 Protein

Concentration gRNA/ Cas9 (pg/nl)	50/250	25/250
Number injected	287	154
Number alive	245	123
Number w/ phenotype	201	79
% successfully targeted	82.0%	74.2%

Dr-m2de1 gRNA Design

After observing high efficiency rates using CRISPR/Cas9 with both the spadetail and *golden* gRNAs, we moved to designing gRNAs to excise the m2de1 sequence from the zebrafish genome. Originally, we hoped to excise m2de1 in its entirety. Using CRISPRscan, a novel scoring algorithm developed by the Giraldez Lab at Yale University, we were able to identify potential gRNAs upstream and downstream of m2de1. Upstream of the element, there were no potential gRNAs with high efficacy scores and no off-target effects. However, we selected the upstream gRNA with the lowest off-target score and selected a downstream gRNA with a high efficacy score and no off-target effects. These two gRNAs together would lead to the deletion of approximately 1,229 basepair region of the genome that would include the entire

m2de1 sequence. After selecting the best possible gRNAs as indicated by CRISPRscan, embryos were injected with the CRISPR/Cas9 system and each of the gRNAs (Figure 10). Embryos that were injected in this manner consistently died at approximately 24hfp. This experiment was repeated four times (Table 6).

Table 6: The number of embryos injected with the original m2de1 gRNAs during each of the four trials of the original m2de1 CRISPR injections.

Trial #	1	2	3	4
Concentration upstream gRNA/ downstream gRNA/ Cas9 (pg/nl)	50/50/250	50/50/250	50/50/250	50/50/250
Number injected	78	69	112	101
Number alive at ~24 hpf	14	11	17	9
Number dead at ~ 24 hpf	64	58	95	92
% Dead at ~24 hpf	82.1%	84.1%	84.8%	91.1%

*The number of alive and dead embryos after approximately 24 hpf and finally the percentage of injected embryos that were dead by 24 hpf are also shown.

To determine whether one or both of the gRNAs was responsible for the lethal phenotype, embryos were injected with either 50 pg/nl of the upstream gRNA and 250 pg/nl of the Cas9 protein or 50pg/nl of the downstream gRNA and250 pg/nl of the Cas9 protein. After injections, embryos that had been injected with the downstream gRNA and Cas9 developed normally; however, embryos injected with the upstream gRNA and Cas9 died approximately 24 hpf. Due to the lethality generated by the upstream gRNA, gRNAs were redesigned.

gcatccactgcattaaacatatgctggaata**gttggcggttcattccgctgtgg**tgacccctgatcaataaagagactaagctgaggg
aaaatgaatgaatgaatgaaaaatctgtgtaatgctctgaacactaatttttaaattgtttttctagttaaattatttgtataattttgttgactgaa
ttaataaagtaaaagaaattag**aaagctcattataaggccgtgcatgttgagggtcgattgttgttaaaaaaaacaacaatacta**
ctaataataatgaatgtcagcggttacggccactgacactgatgagctctggagtgtgtttgttaattaaagcccttctcggtgt
ggttataatcataagacagcgccgcgtgaacaacgtcgtatatcaaacttcacgatgcgacgccggctgataatgcgccgcc
gctgctaaagctgattttccatgatgaatctgagagcatataaattggctaatatctaaaaacatttacagacagtgcaggaac
cgagtgtttgcggccgtgatggatgaggctgttagatgctagaggacagatccgcaaatctctgccgctgcacagaccctgca
cgcagctacagatttactcctttaaacccgacctcctgtgtgtgtgtgtgagtttgtgtgagagagttagtgttagtgtgtgtgtattaa
agtgagtttaaatatgtttatatatgtgtgcataaatgtgtgtatatgtatatacacacatatacacatacacacacatatataaacatatacac
atataaacatacacacacatatataaacacatatacacacnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnntatatatatatatatatatatat
atatatatatatatatatatatatatatatatatatatatgtatatgtgtgtgtgtgtattttgtgagtgttatatctggagctctctcagcaaggt
tagtggggcagaactgatcttactctgtctgacagcagtgttgcggctcggtccgttatgcacatgtgtgtgtgagagctccacaactcc
acgctagtaattagtaaacttgtcaagatgctgcttcgctcaccgcctcccgcagctgctaatgaagagtgtgtgtgtcaggatttacagc
aga**ccgacgcgctgcacagaactccg**cacaacacactttctgtctcg

Figure 10: Genomic DNA sequence located on chromosome 17 of zebrafish displaying the entire m2de1 sequence as well as upstream and downstream sequences. The bold and underlined sequences represent the DNA sequences that were originally selected to be targeted by the gRNA/ Cas9 complex to excise m2de1, which is indicated by the bold sequence. The first target DNA sequence is the complement of the actual sequence recognized by the gRNA.

When attempting to redesign gRNAs to excise m2de1, there were no gRNAs with high efficacy scores and low off-target effects upstream of the element, as described on page 61. Rather than selecting another upstream gRNA with potential off-target effects, we decided to aim to knockout a portion of m2de1, rather than the entire element. To determine whether or not the Dr-m2de1 sequence contained any potential gRNA sequences within the element, the element was screened for potential gRNAs using CRISPRscan. Within this region, only two potential gRNAs (Fig. 11) with no off-target effects were identified. These two gRNAs are separated by 42 basepairs (Fig. 11).

GCTCATTATAAGGCCGTGCATGTTGAGGGTCGATTGTTGTTAAAAAAAACAACAATACT
ACTAATAATAATGAATGTCAGCGGTTACGGCCACTGACACTGATGAGCTCTGGAGTGTG
TTTGTTAATTAAAGCCCTTCTCGGTGTGGTTATAATCATAAGACAGCGCCGCGTGAACAA
CGTCGTATATCAAACTTCACGATGCGACGCCGGCTGATAATGCGCCGCCGCTGCTAAAG
CTGATTTTCCATGATGAATCTGAGAGCATATAAATTGGCTAATATCTAAAAACATTTACA
GACAGTGCAGGAACCGAGT**GTTTGCGGCCGTGATGGATG**AGGCTGTTAGATGCTAGA
GGACAGATCCGCAAATCTCTGCCG**CTGCACAGAGCCCTGCACGC**AGCTACAGATTTAC
TCCTTTAAACCCGACCTCCTGTGTGTGTGTGTGAGTTTG

Figure 11: Genomic DNA sequence located on chromosome 17 of zebrafish displaying the entire m2de1 sequence and gRNAs targets within the element. The entire sequence represents the m2de1 element while the bold sequences represent the target DNA sequence for gRNAs used to direct the Cas9 enzyme to excise a portion of m2de1. The protospacer adjacent motifs (PAMs) are shown in red. The second target DNA sequence is the complement, as the gRNA recognizes the 3' strand of DNA. The region to be excised is underlined.

Bioinformatical Analysis of Transcription Factor Binding Sites

Since I was unable to excise the entire m2de1 sequence from the zebrafish genome, I wanted to ensure that I was excising an important part of the element as shown in figure 11. By taking the sequence in between the two gRNAs (Fig. 11) and entering it into the Patch 1.0 database on gene-regulation.com, we were able to determine that there were indeed transcription factor binding sites between the two gRNAs. Predicted transcription factor binding sites include HOXA1/B1, Pbx-1a/1b, PKNOX1, Meis-1a/1b and GATA1 (Fig. 12).

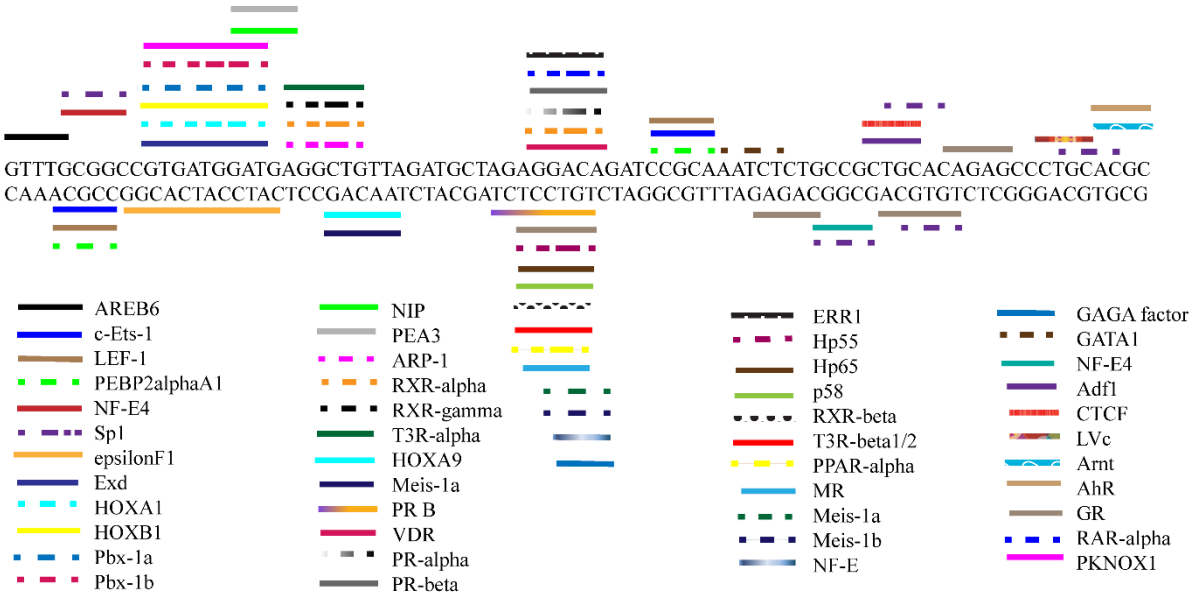

Figure 12: **Schematic of double stranded DNA containing the 82 basepairs that will be excised via m2de1 CRISPR injections.** Patch1.0 was used to predict transcription factor binding sites. The transcription factor is shown either below or above the sequence depending on which strand of the double stranded DNA the factor is predicted to bind to. Transcription factors are color coordinated and are listed below the schematic.

Dr-m2de1 knockout using CRISPR

After confirming that there were transcription factor binding sites located between the 2 target DNA sequences shown in Figure 11/12, the gRNA sequences were transcribed and m2de1 CRISPR injections were performed on single cell zebrafish embryos. After injections, in order to confirm that a portion of m2de1 was excised from the genome after CRISPR injections, genomic DNA was isolated from m2de1 CRISPR injected embryos. These embryos were randomly divided into 2 tubes for DNA extraction. This DNA was then used as a template for PCR in order to amplify m2de1. Samples from the PCR were run on an agarose gel (Fig. 13). Lane A contains a 100 basepair ladder, wildtype DNA was used as a control in Lane B, while Lanes C and D contained different samples of template DNA from injected embryos. Lanes B-D all displayed a band at approximately 450 basepairs, the size of the fully intact m2de1 element. However, Lane C also contained a lower molecular weight

band at approximately 350 basepairs (Fig. 13). This second band in Lane C indicates that some of the injected embryos were indeed transgenic and that the 82 basepairs between the gRNAs in m2de1 were likely removed.

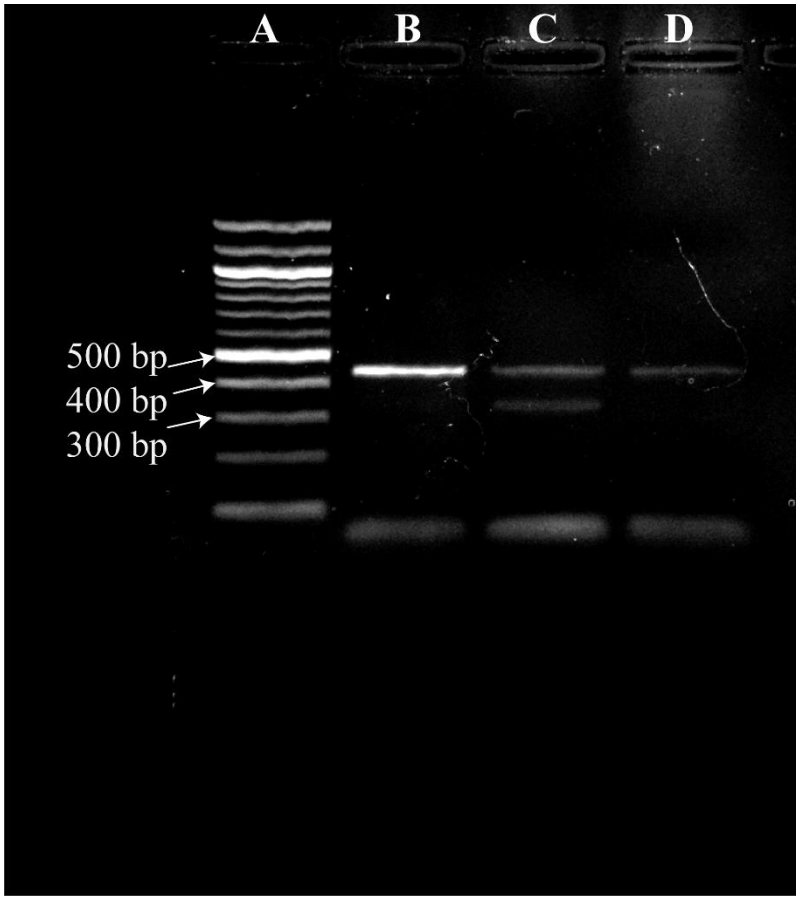

Figure 13: Agarose gel showing m2de1 partial knockout. Lane A contains a 100bp ladder. Lane B contains a positive control reaction using wildtype DNA. Lanes C and D contain the experimental reactions containing DNA extracted from m2de1 CRISPR injected embryos. Lanes C and D contain the same ~450 bp band as seen in wildtype. However, Lane C also contains a slightly lower molecular weight band at ~360 bp. This suggests that the knockout was at least partially successful in some of the injected embryos.

After confirming that m2de1 CRISPR injections were at least somewhat successful by observing a PCR product consistent in length to that of the element missing the region excised in Figure 13, Lane C, whole mount *in situ* hybridization was performed in order to identify any potential differences in expression patterns in wildtype embryos verses

knockouts. m2de1 CRISPR injected embryos were fixed at 48 hpf in order to examine expression of *meis2a* and *zgc:154061*. Embryos were fixed at 48 hpf due to the presence of the m2de1 element directing expression of eGFP beginning at 48 hpf. In wildtype embryos, *meis2a* expression is observed in the forebrain, optic tectum, cerebellum and rhombencephalon (Figure 14A). In embryos that had been injected with the m2de1 CRISPR system, *meis2a* expression was observed faintly in the forebrain and optic tectum; however, expression in the cerebellum and rhombencephalon was not observed (Figure 14B). There was no observable difference in expression patterns of *zgc*:*154061* in CRISPR injected embryos when compared to wildtype (data not shown). The number of injected embryos displaying an altered expression pattern is shown in Table 7.

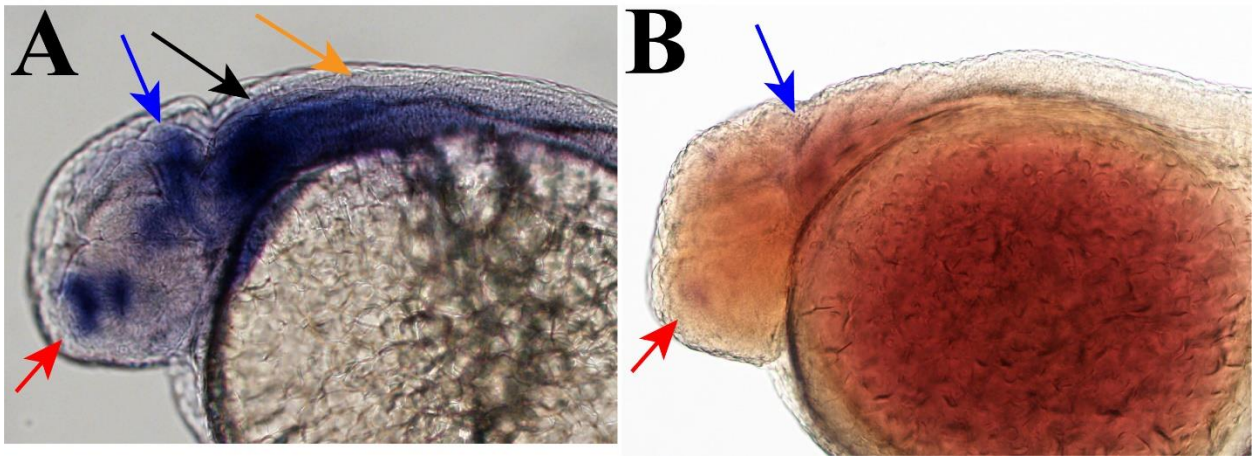

Figure 14: *W*hole mount *in situ* hybridization for *meis2a* on wildtype and partial knockout embryos. At 48 hpf, **(A)** wildtype embryos displayed *meis2a* expression in the forebrain (red arrow), optic tectum (blue arrow), cerebellum (black arrow) and rhombencephalon (orange arrow) (panel A was taken from Figure 7) while **(B)** m2de1 CRISPR injected embryos displayed mild *meis2a* expression in the forebrain (red arrow) and optic tectum (blue arrow) but no expression is observed in the cerebellum or rhombencephalon. While these *in situs* were not developed at the same time, the same probe was used for both embryos. All embryos are positioned in the lateral orientation with the anterior to the left.

Table 7: The concentration of gRNAs and Cas9 protein injected into each embryo for m2de1 CRISPR injections.

Embryos injected with m2de1 CRISPR system	Trial 1
Concentration	
upstream gRNA/ downstream gRNA/ Cas9 (pg/nl)	50/50/250
Number injected	89
Number dead at ~48 hpf	51
% Dead at ~48 hpf	42.7%
Number alive/ fixed at ~48 hpf	51
Number of fixed embryos displaying WT *meis2a* expression pattern	42
Number of fixed embryos displaying altered *meis2a* expression pattern	9

* The number of injected embryos is also shown, as well as a breakdown of dead/ alive embryos and the number of embryos showing altered expression patterns as compared to wildtype.

After observing a different expression pattern in embryos that had previously been injected with the m2de1 CRISPR system, DNA was extracted from these embryos in order to confirm deletion. Individual embryos that displayed a different expression pattern as compared to wildtype embryos were placed into individual tubes for DNA extraction. This DNA was then used as a template for PCR in order to amplify m2de1. Samples from the PCR were run on an agarose gel (Fig. 15). Lane A contains a 100 basepair ladder, while wildtype DNA was used as a control in Lane B. Lanes C contains DNA from embryos that were injected with the CRISPR/Cas9 system but show no difference in *meis2a* expression pattern when compared to wildtype (data not shown). Lastly, Lane D contains the DNA extracted from the mutant embryo in Figure 14B. While Lanes B-D all contain a band at approximately 450 basepairs, indicative of the unmutated m2de1 element, Lane D also contains a lower

molecular weight band around 350 basepairs indicating that m2de1 was partially excised on one allele through CRISPR injections.

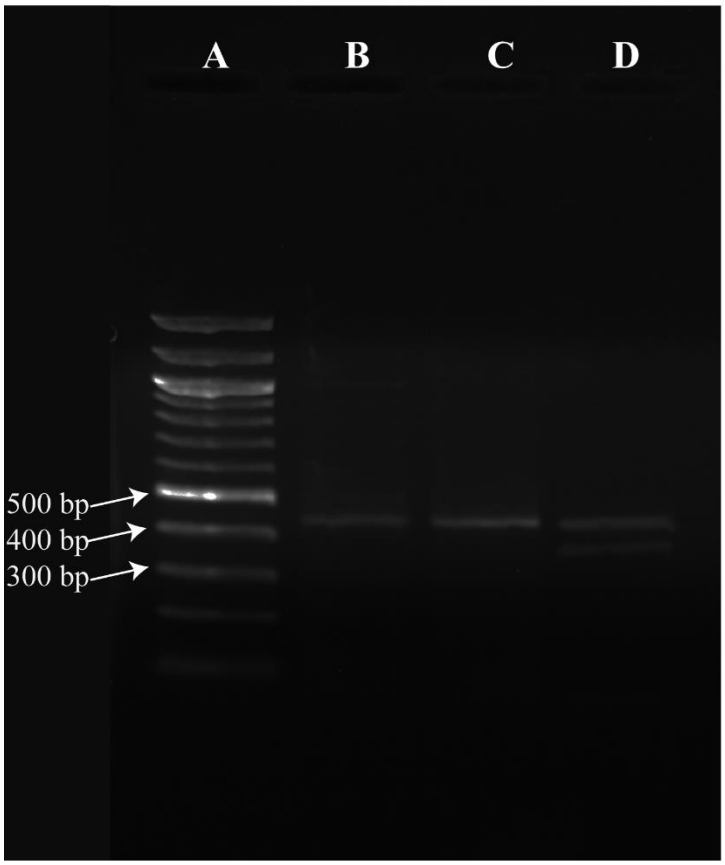

Figure 15: Agarose gel showing m2de1 partial knockout. Lane A contains a 100bp ladder. Lane B contains a positive control reaction using wildtype DNA. Lanes C contains extracted DNA from CRISPR injected embryos showing the same *meis2a* expression as wildtype embryos. Lane D contains the DNA extracted from an m2de1 CRISPR injected embryo that displayed an altered expression pattern of *meis2a*. Lanes C and D contain the same ~450 bp band as seen in wildtype. Lane D also contains a slightly lower molecular weight band at ~360 bp.

Discussion

By examining developmental genes for their specific function and expression patterns, insight can be gained into the regulation of patterning, morphogenesis, differentiation and growth. While the expression patterns of the *Meis* genes are fairly well characterized, the mechanisms that direct the *Meis* genes expression at the correct time and to the correct place during development are largely unknown. Previously, the Zerucha lab has identified four highly conserved noncoding elements associated with the vertebrate *Meis2* gene and named them m2de1-4 (for *Meis2* downstream element). The four putative enhancers are present in humans, chicken, mice and other tetrapods and in each organism, they are found downstream of *Meis2*. However, in zebrafish, only one of these presumed enhancers has been identified. Interestingly, these elements are found within the introns of an adjacent gene, *zgc*:*154061* in zebrafish, whose orthologs are always found in an inverted convergently transcribed orientation directly downstream of *Meis2* in vertebrates. I have observed overlaps in the expression patterns of *meis2a* and *zgc*:*154061* as well as the expression of a reporter transgene directed by m2de1 in zebrafish. I propose that the genomic organization of these two genes has been evolutionarily conserved due to the sharing of cis-regulatory elements. In this study, I have used the CRISPR/Cas9 system to excise a portion of the m2de1 element in zebrafish to determine whether or not *meis2a* or *zgc:154061* expression is affected.

Generation of a stable transgenic line expressing eGFP

Our lab has previously shown that m2de1 directs eGFP expression to the forebrain, midbrain and hindbrain of zebrafish embryos at 48 and 54 hpf (Barrett, C., 2013; Ferrara, T., 2015). The element has shown to direct expression to the same areas whether in forward or reverse orientation (Barrett, C., 2013). This is expected because the proteins interacting with

m2de1 should be able to recognize the DNA sequence regardless of forward or reverse orientation. While the expression directed by m2de1 has been observed in primary injected embryos multiple times, a stable transgenic line has yet to be generated. A stable line would eliminate mosaicism, a common issue in trying to interpret data form primary transgenic organisms. A homozygous line would allow us to visualize the full pattern of expression driven by the m2de1 element, especially expression that has been observed only faintly previously, such around the eye (Ferrara, 2015). Further, a stable transgenic line would allow us to more specifically determine where the element directs localized expression. To date, we are able to observe relative expression patterns, specifically around the head. However, the zebrafish head contains many overlapping sections and a stable transgenic line would allow us to differentiate between the overlapping segments. Currently, pDr m2de1-F-*cfos*-EGFP injected embryos are being raised in hopes to breed the primary injected fish to generate a stable line.

Whole mount in situ hybridization

Expression of *meis2a* has previously been shown to be expressed in the forebrain, hindbrain and neural plate in zebrafish at the 2-somite stage. At the 5-somite stage, *meis2a* expression in the hindbrain expands anteriorly and by the 10-somite stage, *meis2a* is expressed in the eyes, the forebrain, the developing tectum and within the neural rod (Bessa et al., 2008; Thisse and Thisse, 2005; Waskiewicz et al., 2001). At later stages during development, *meis2a* expression has been observed in the forebrain, retina, eye, midbrain, hindbrain, spinal cord and neural tube at 24 hpf. At 48 hpf, the *meis2a* expression is no longer observed in the neural tube but is still observed in the forebrain, eye, retina, midbrain, hindbrain and spinal cord, as well as, the optic tectum and tegmentum (Carpenter et al., 2016;

Thisse and Thisse, 2005). In this study, whole mount *in situ* hybridization was performed in order to visualize *meis2a* expression at later timepoints during development. At 24 and 48 hpf, *meis2a* expression was observed in the forebrain, which correlates with the expression observed by Carpenter et al. and Thisse and Thisse. Additionally, *meis2a* was observed in optic tectum, cerebellum and rhombencephalon at both 24 and 48 hpf. However, *meis2a* was not observed in the spinal cord or neural tube as observed previously (Carpenter et al., 2016; Thisse and Thisse, 2005). The *meis2a* expression in the spinal cord and neural tube observed previously by Carpenter et al. and Thisse and Thisse was faint. In this experiment, we developed embryos for *in situ* hybridization for a shorter period of time in order to eliminate any background color. Due to this shortened period of color development, we most likely quenched our reactions before the expression in the spinal cord and neural tube became visible.

At 54 hpf, *meis2a* was observed in the forebrain, retina, optic tectum and cerebellum. At 60 hpf, the retina expression disappeared and expression was still observed in the forebrain, optic tectum and cerebellum. At 66 hpf, the optic tectum expression was no longer observed; however, at both 66 and 72 hpf, *meis2a* was observed in the forebrain and cerebellum. Our results coincide with previously published data as expression of *meis2a* becomes more anteriorly restricted as development progresses (Bessa et al., 2008; Carpenter et al., 2016; Thisse and Thisse, 2005; Waskiewicz et al., 2001).

Previously, the Zerucha lab identified a conserved gene, *zgc:154061* in zebrafish, that is located downstream of *meis2a.* Whole mount *in situ* hybridization was performed in order to visualize spatial and temporal expression of *zgc:154061.* Analysis revealed transcripts of the gene were present as early as the zygotic period (0-0.75 hpf). Transcripts were also observed

during the cleavage (0.75-2.25 hpf) and blastula (2.25-5.25) periods. The zygotic genome does not become active until the midblastula transitition (around 2.75 hpf), therefore, the observation of transcripts of *zgc:154061* during the zygote, cleavage and blastula periods indicates that the gene is expressed maternally. Around 12 hpf, the expression of *zgc:154061* was observed to be anteriorly restricted, with expression appearing in the developing neural tube around the retina, forebrain, olfactory region and the tectum of the midbrain. At 24 hpf, expression was observed to be restricted to the neural tube and eye, while at 48 hpf expression was observed in the retina, forebrain, olfactory region and the tectum of the midbrain. At 48 hpf, expression of *zgc:154061* gradually begins to decrease, so much that localized expression becomes unobservable (Carpenter et al., 2016). In this study, whole mount *in situ* hybridization was performed in order to visualize *zgc:154061* expression at later timepoints during development. At 24 hpf, *zgc:154061* expression was observed in the developing neural tube around the forebrain, optic tectum, cerebellum and rhombencephalon which correlates with the expression observed by Carpenter et al. in 2016; however, the faint expression observed by Carpenter et al. in the developing eye at 24 hpf was not observed in this study. In this experiment, in order to eliminate any background color, we developed *in situs* for a relatively short period of time and we most likely quenched our reactions before the expression in the eye became visible. At 48 hpf, expression does seem to decrease as suggested by Carpenter et al. However, in this study, whole mount *in situ* hybridization was overdeveloped in order to attempt to identify areas of expression. Between 48 and 72 hpf, localized expression was not identified; however, relative expression was identified around the retina and forebrain at 48, 54, 60, 66 and 72 hours. Additionally, *zgc:154061* was observed in the cerebellum at 54 hpf.

The expression of *zgc:154061* is anteriorly restricted between 48 and 72 hpf, which is similar to *meis2a* expression, which is also restricted anteriorly during these timepoints. Both *meis2a* and *zgc:154061* are observed in the developing neural tube, specifically around the forebrain between 24 hpf until 72 hpf. Further, *Meis2* is known to play a role in the development of the eye (Zhang et al., 2002) and we observed *zgc:154061* expression around the eye between 48 and 72 hpf. If *meis2a* and *zgc:154061* are sharing cis-regulatory elements as we predict, the overlapping expression patterns that we observe would be expected. While each gene normally has its own set of *cis*-regulatory elements, elements commonly direct the expression of multiple genes (Gould et el., 1997; Zerucha et al., 2000).

CRISPR injections using a positive control

Before beginning CRISPR injections to excise a portion of m2de1, positive control CRISPR injections were completed. In accordance to Talbot and Amacher (2014) 40ng/µl of spadetail gRNA and 80ng/µl of Cas9 mRNA were injected into single cell zebrafish embryos. At 24 hpf, the spadetail phenotype was not observed. In 2013, Hwang et al. generated several gRNAs that targeted the *fh* gene of zebrafish. By co-injecting the gRNAs with Cas9 mRNA, mutation rates ranged from 10-52.7%. The highest average mutation rate occurred with an injection concentration containing 12.5ng/uL gRNA and 300ng/uL Cas9 mRNA. Using this information, we completed a second round of injections with 12.5ng/uL spadetail gRNA and 300ng/uL Cas9 mRNA into single cell zebrafish embryos. Still, no phenotype was observed. The concentration of gRNA and Cas9 mRNA was altered two additional times (Supplementary Table 2) in order to observe the spadetail phenotype. Concentrations came from unpublished data and personal communications. However, after each round of subsequent injections, the phenotype was never observed (Supplementary

Table 2). Rather than continuing to troubleshoot CRISPR injections, EnGen® Cas9 NLS (NEB Mo646T) was commercially obtained for CRISPR injections.

In accordance with the zebrafish CRISPR injection protocol by Jeffrey Essner (2016), 50ng/ul gRNA and 250 ng/µl Cas9 protein was used for CRISPR injections. At approximately 24 hpf, the spadetail phenotype was observed. The concentration of gRNA was both increased and decreased in subsequent rounds of injections to determine the most efficient concentration. These concentrations were also tested with *golden* injections. For each positive control, 50ng/µl gRNA and 250ng/µl Cas9 protein showed to be most efficient.

In 2016, in order to compare the efficiency of *in vitro* assembled Cas9-sgRNA ribonucleoprotein complexes, Burger et al. assembled and injected Cas9-sg RNA RNPs into single cells zebrafish embryos. Burger et al. targeted *gol* and compared their results to those results obtained by Jao et al. 2013. Jao et al. injected single cell embryos with 50pg of *gol* gRNA and 150pg of *nls-zCas9-nls* RNA. At 48 hpf, mosaic hypopigmentation was observed in skin melanophores and RPE in 94% of the injected embryos, while mutagenesis rates ranged from 78-98%, indicating that most cells contained bi-allelic mutations (Jao et al., 2013). In comparison, when *gol* was targeted with Cas9-sgRNA RNPs, nearly 80% of successfully injected embryo displayed strong mutant phenotypes. The remaining 20% of embryos displayed different degrees of the mutant phenotype (Burger et al., 2016). Furthermore, the injected embryos maintained the mutant phenotype as compared to morpholino knockdowns, which only maintain the gene knockdown for a few days (Bill et al., 2009; Burger et al., 2016). When the *gol* mutants were grown to adulthood, the mutant phenotype was preserved. The fish were then in-crossed resulting in the F1 offspring displaying complete mutant phenotypes, showing that when the *gol* loci was targeted with

Cas9-sgRNA RNPs, the germline of injected embryos was mutated (Burger et al., 2016). As mentioned previously, in addition to targeting *golden*, Burger et al. also used RNPs to recapitulate loss of function phenotypes, like spadetail. Using RNPs to target spt, more than 60% of all embryos injected displayed a strong phenotype while almost 20% displayed a mild or partial phenotype. Analysis via MiSeq and CrispRVariants revealed that RNPs allow for efficient bi-allelic mutagenesis for both *golden* and spadetail (Burger et al., 2016).

Burger et al. reported the most efficient concentration of Cas9 protein to be 831ng/µl, while the optimal gRNA concentration varied depending on the DNA sequence and the molar mass. Burger et al. determined gRNA concentration using a free online calculator, CrispantCal, available at http://lmweber.github.io/CrispantCal/. Contrary to these results, our results showed the highest efficacy of all gRNAs at a concentration of 50ng/ul gRNA. While we did not test different concentrations of Cas9 protein, we showed that 250ng/ul Cas9 was efficient. Burger et al. displayed 80% strong mutant phenotypes for *gol* injections, while the remaining 20% showed mild phenotypes. In our test, we saw 82% of embryos displaying the *gol* phenotype. Using spadetail, Burger et al. showed about 80% of injected embryos with the mutant phenotype while our results displayed 87.2% mutant phenotype. Although our Cas9 protein differed from the protein utilized by Burger et al., our commercially obtained protein shows to be effective at a comparable rate, indicating this to be more efficient and cost effective than generating our own *in vitro* translated Cas9.

In our original injections with Cas9 mRNA, we consistently observed no embryos displaying mutant phenotypes; therefore, we hypothesize that the integrity of the plasmid was compromised and are not taking our mutagenesis rates using Cas9 mRNA into consideration when comparing Cas9 mRNA and Cas9 protein. Using Cas9 mRNA, Jao et al. reported *gol*

injected embryos showed mutagenesis rates between 78-98%. Burger et al. reported 100% of *gol* injected embryos showed a mutant phenotype, 80% with a strong phenotype and 20% with a weak phenotype, when using *in vitro* assembled Cas9. Finally, our results show 82.0% mutant phenotype using Cas9 protein. Combined, these results suggest that protein is not necessarily more efficient than Cas9 mRNA; however, *in vitro* assembled Cas9 provides an effective reagent that can be replicated for direct and measurable knockout and loss-of-function studies.

Identifying CRISPR Target

Originally, in order to excise m2de1 in its entirety, we examined the zebrafish genome upstream and downstream of m2de1 for potential gRNAs. We were aiming to keep the knockout under 1500 basepairs in hopes of preventing NHEJ from occurring between cut sites. Using CRISPRscan to identify potential gRNAs, there were no upstream gRNAs within our region with low off-target scores. High off-target scores indicated that our chance of knocking out m2de1 in its entirety could be compromised due to the possibility of Cas9 cutting at these off-target locations rather than upstream of m2de1. Although the upstream gRNA scores predicted off-target effects, we chose the gRNA in the area with the lowest off-target score. Downstream of the element, we chose a gRNA with no off-target effects. Upon injections with these gRNAs and Cas9, we observed that the embryos consistently died by 24 hpf. Upon further experimentation, we observed that the upstream gRNA and Cas9 generated the same lethal phenotype, while the downstream gRNA and Cas9 did not generate a lethal phenotype. We hypothesized that the high off-target score associated with the upstream gRNA could be the reason we were observing the lethal phenotype. If the upstream gRNA was directing Cas9 to a location rather than upstream of m2de1, there is a chance we were

knocking out some other unknown but necessary sequence within the genome that caused lethality.

Due to the observed lethal phenotype generated by the upstream gRNA, gRNAs were redesigned. We had established that there were no upstream gRNAs within our region with no off-target effects. Because of this, rather than excising m2de1 in its entirety, we aimed to excise a portion of m2de1. In order to excise a portion of m2de1, we looked within the element for potential gRNAs. Two gRNAs with zero off-target effects within m2de1 were identified. To determine how necessary the sequence between these gRNAs was, the sequence was examined for putative transcription factor binding sites. The search revealed numerous deuterostome putative binding sites between the potential gRNAs. While Figure 12 shows putative transcription factor binding sites for various deuterostomes rather than only zebrafish, these sites are relevant because possible homologous transcription factors could bind to the zebrafish sequence. Particularly of interest, the sequence contains predicted binding sites for homeodomain proteins HOX, Pbx and Meis. As mentioned previously, Meis and Pbx commonly interact with Hox proteins in binding DNA and controlling gene transcription (Choe et al., 2014). Not only do these proteins bind to DNA, they also have the ability to interact with one another, binding through structures within their N-terminal and C-terminal domains (Chang et al.,1997). It is interesting that putative binding sites for each of these three proteins were identified within the 82 basepair sequence between gRNAs. This suggests that this portion of m2de1 may contain binding sites for a combination of important regulatory proteins. The presence of a Meis recognition sequence also suggest the possibility of auto- or cross-regulatory interactions controlling *Meis2* expression.

Further, the sequence contains a putative GATA1 binding site. Previously, *GATA1* has been shown to encode a critical hematopoietic transcription factor (Valverde-Garduno et al., 2004). This is of particular interest considering both *Meis* and *C15ORF41* (the human *zgc:154061* ortholog) have been linked to blood. *Meis* genes were first discovered in a study of mouse myeloid ecotropic leukemia virus and an overexpression of *Meis1* causes blood cancer in humans and mice (Hisa et al., 2004, Moskow et al., 1995). Additionally, *C15ORF41* has been shown to play a role in erythropoiesis (Babbs et al., 2013). After the bioinformatical analysis of the sequence between the gRNAs identified in Figure 11, it was decided that excising this portion of m2de1 could indeed affect the element's potential ability to act as a *cis*-regulatory element for *meis2a* and/or *zgc:154061*.

m2de1 knockout, in situ hybridization and DNA extraction

The second band in Figure 13, Lane C indicated that some of the injected embryos were indeed transgenic and that a portion of the sequence between the gRNAs in m2de1 was excised. Because DNA was extracted from multiple embryos, it is impossible to determine whether the portion of m2de1 was excised from both alleles from any of the embryos, as the second band could indicate heterozygotic embryos. However, observing that the m2de1 CRISPR injections were somewhat successful, injected embryos were fixed at approximately 48 hpf. Embryos were fixed at 48 hpf because this is the earliest stage where m2de1 has been observed to direct the expression of eGFP. Whole mount *in situ* hybridization for both *meis2a* and *zgc:154061* was completed. Upon examination, no difference in expression pattern was observed between wildtype embryos and CRISPR injected embryos for *zgc:154061* (data not shown). However, this could be due to the already low levels of expression of *zgc:154061* and the inability to distinguish any changes in expression to this

already low and difficult to visualize expression pattern in wildtype embryos. When observing *meis2a* expression for CRISPR injected embryos, 42 embryos did not display an altered expression pattern (data not shown). When DNA was extracted from these embryos and PCR was utilized to amplify m2de1, it was observed that these embryos only contained the wildtype m2de1 element. While 42 embryos did not show a difference in *meis2a* expression, 9 m2de1 CRISPR injected embryos showed a variation in *meis2a* expression verses the wildtype expression pattern (Figure 14) (Table 7). After these embryos were imaged, DNA was extracted and used to amplify the m2de1 element. After PCR amplification and gel electrophoresis, it was confirmed that the embryo pictured in Figure 14B had a wildtype version of m2de1, as well as a version of m2de1 missing the described region.

As mentioned previously, predicted transcription factor binding sites within the described area include HOXA1/B1 and Pbx-1a/1b. *Hoxa1*/*b1* have been shown to play a role in hindbrain and craniofacial development. In mice, *Hoxa1* is expressed in the forebrain and midbrain during embryonic development (McClintock et al., 2002; Shih et al., 2001). Further, *Hoxa1* null mutants display reduced or absent rhombomeres-4 and 5, leading to defects in the hindbrain (Carpenter et al., 1993; Chisaka et al., 1992; Dolle et al., 1993; Mark et al., 1993). In *Hoxb1* null mutants, the identity of r4 is altered and mutant mice fail to develop the somatic motor component of the VIIth (facial) nerve (Goddard et al., 1996; Studer et al., 1996). Similarly, *Pbx1* is expressed in high levels in the developing rat brain (Roberts et al., 1995). *PBX1* is also expressed abundantly in the human fetal brain, while *PBX2* is expressed at lower levels in the human fetal brain (Monica et al., 1991). If the excised portion of m2de1 contains binding sites for Hox and Pbx proteins as predicted, this

could explain the lack of expression in the cerebellum and rhombencephalon of the mutant embryo.

Further, there are also predicted transcription factor binding sites for Meis-1a/1b within the described region. Murine Meis1 and Meis2 and human MEIS2 have an identical amino acid sequence within their homeodomains, resulting in human Meis2 binding to the same consensus DNA sequence as murine Meis1. This sequence has been identified as 5'-TGACAG-3' (Yang et al., 2000). Because these Meis proteins bind to identical sequences, it is possible that meis2a, as well as meis1a/1b, could contain a binding site within the described portion of m2de1. Given this information, it is possible that Meis proteins auto- and/or cross-regulate *Meis* genes. If m2de1 is directing *meis2a* expression, knocking out a Meis binding site within the element could alter when, where and how much *meis2a* is expressed which would also help to explain the altered expression pattern observed in Figure14B. Together, the role of *Hox* and *Pbx* in the developing brain, as well as the possibility of auto- and/or cross-regulation between Meis proteins, could help to explain why we see an altered expression pattern of *meis2a* in mutant embryos. If m2de1 is directing the expression of *meis2a,* removing putative binding sites for transcription factors that play a role in brain and craniofacial development could affect when and where *meis2a* is expressed.

In the future, sequencing can be completed in order to confirm exactly which portion of m2de1 has been excised out of different fish and this could help to explain the mosaicism observed and help to understand which putative binding sites might be crucial. It would also be of interest to study the interactions between the DNA in the knockout region and specific proteins using chromatin immunoprecipitation (ChIP) in order to understand exactly which

transcription factors are playing a role, whereas now we are basing our information off of putative transcription factor binding sites.

It was interesting that individual knockout embryos contained a wildtype and mutant version of m2de1. In both of our positive control CRISPR injections, we observed biallelic knockouts as both *gol* and spadetail phenotypes are only observed in null mutants (Burger et al., 2016; Jao et al., 2013). However, a possible explanation is that m2de1 null embryos display a lethal phenotype. We observed that 42.7% of m2de1 CRISPR injected embryos died before reaching 48 hpf (Table 7). It is possible that the high death rate was due to biallelic disruption of m2de1 in some of these embryos. When DNA was extracted from these m2de1 CRISPR injected embryos and used as a template to amplify the m2de1 element, the agarose gel displayed bands indicative of both wildtype and mutant m2de1 (data not shown). Because the template DNA came from a combination of embryos, it is impossible to tell whether any of these dead embryos were m2de1 null mutants; however, the mutated version of m2de1 was present in a portion of the embryos and it will be interesting in the future to determine the differences, if any, between heterozygotic and null mutants.

While heterozygotic and homozygotic mutants could display different phenotypes, there are other variables that could affect m2de1 CRISPR knockouts as well. The protospacer-adjacent motif (PAM) interacting domain is responsible for initiating binding of target DNA. The arginine-rich bridge helix contacts between 8 to 10 nucleotides on the 3' end of the target DNA and initiates cleavage (Figure 3) (Jinek et al., 2014; Nishimasu et al., 2014). Further, NHEJ repair mechanisms leads to imperfect or frameshift repairs. This type of repair is thought to introduce insertions and deletions of random lengths, normally between 1-18 basepairs, which can lead to mosaicism in CRISPR-modified organisms

(Maruyama et al., 2015). Together, the variation in where Cas9 cleaves the DNA and the length of insertions and deletions due to NHEJ, CRISPR injected embryos can contain a wide variety of mosaicism. This could mean that while some embryos have 70-100 basepairs of m2de1 excised from their genome, others may only contain a 30-40 basepair excision. This could greatly affect whether or not the injections are lethal while also affecting the expression pattern of *meis2a.* For example, the putative transcription factor binding sites in the outer most 10-15 nucleotides may be excised in some embryos and not excised in other embryos. If these putative binding sites are crucial in directing *meis2a* expression, embryos with a larger portion of the sequence excised may die or display weak expression of *meis2a* while embryos with smaller knockout regions display a wildtype expression pattern. As mentioned previously, it will be important to sequence this genomic region where the excision occurs in order to confirm its absence. This will help to better understand the mosaicism that is observed in knockouts.

In conclusion, the data presented in this study shows expression of *meis2a* and *zgc:154061* at older timepoints throughout development, as previous studies had only examined expression patterns up to 48 hpf (Bessa et al., 2008; Carpenter et al., 2016; Thisse and Thisse, 2005; Waskiewicz et al., 2001). The expression patterns of both genes show overlap between 48 and 72 hpf. Further, at 48 and 54 hpf, the expression of both genes overlaps with the expression directed by m2de1. When a portion of the m2de1 element that contains numerous putative binding sites is excised from the zebrafish genome, a difference in *meis2a* expression at 48 hpf is observed when compared to wildtype embryos. While a direct relationship between *meis2a* and m2de1 cannot yet be confirmed, we hope to raise a stable transgenic line of m2de1 knockouts in order to better understand the relationship

between m2de1and *meis2a* and *zgc:154061*. The work completed for this project assisted in further characterizing m2de1, a novel element hypothesized to be involved in the regulation of *meis2a* and *zgc:154061* and their orthologs.

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.

Akoulitchev, S., Makela, T. P., Weinberg, R. A., and Reinberg, D. (1995). Requirement for TFIIH kinase activity in transcription by RNA polymerase II. *Nature* **377**, 557-560.

Amali, A., A., Sie, L., Winkler, C. and Featherstone, M. (2013). Zebrafish *hoxd4a* acts upstream of *meis1.1* to regulate vasculogenesis, angiogenesis and hemotopoesis. *PLoS ONE* **8(3)**, e58857.

Amores, A., Forcce, A., Yan, Y. L., Joly, L., Ameniya, 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.

Babbs, C., Roberts, N.A., Sanchez-Pulido, L., McGowan, M.R., Brown, J.M., Sabry, M.A., Consortium, W.G.S., Bentley, D.R., Mcvean, G.A., Donnelly, P., Gileadi, O., Ponting, C.P., Higgs, D.R. and Buckle, V.J. (2013). Homozygous mutations in a predicted endonuclease are a novel cause of congenital dyserythropoietic anemia type i. *Haematologica* **98(9)**, 1383-1387.

Baltimore, D., Berg, P., Botchan, M., Carroll, D., Charo, R. A., Church, G., Corn, J. E., Daley, G. Q., Dounda, J. A., Fenner, M., Greely, H. T., Jinek, M., Martin, G. S., Penhoet, E., Puck, J., Sternberg, S. H., Weissman, J. S. and Yamamoto, K. R. (2015). A prudent path forward for genomic engineering and germline gene modification. *Science, **348(6230),*** 36-38.

Banerjee-Basu, S., Sink, D., and Baxevanis, A. (2001). The Homeodomain Resource: Sequences, Structures, DNA binding sites and genomic information. *Nucleic Acids Res.* **25**, 4173-4180.

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 pages.

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. *Mol. Cell* **14**, 465- 477.

Berthelsen, J., Zappavigna, V., Mavilio, F. and Blasi, F. (1998). Prep1, a novel functional partner of Pbx proteins. *EMBO J.* **17**, 1423-1433.

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

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

Bill, B.R., Petzold, A.M., Clark, K.J., Schimmenti, L.A. and Ekker, S.C. (2009). A primer for morpholino use in zebrafish. *Zebrafish* **6,** 69-77.

Blackwood, E.M. and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. *Science* **281**, 60-63

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.

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

Burger, A., Lindsay, H., Felker, A., Christopher, H., Anders, C., Chiavacci, E., Zaugg, J., Weber, L.M., Catena, R., Jinek, M., Robinson, M.D., Mosimann, C. (2016). Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Dev.* **143**, 2025-2037.

Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A. and Felsenfeld, G. (2002). The insulation of genes form external enhancers and silencing chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **99** (Suppl. 4), 16433, 16437.

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

Carpenter, B.S., Graham, B. and Zerucha, T. (2016). Identification and Developmental Expression of the Zebrafish *zgc:154061* Gene. *Eastern Biologist* **5**, 1-17.

Carpenter, E. M., Goddard, J. M., Chisaka, O., Manley, N. R. and Capecchi, M. R. (1993). Loss of Hox-A1 (Hox-1.6) function results in the reorganization of the murine hindbrain. *Development* **118**, 1063-1075.

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

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

Chariot, A., Gielen, J., Merville, M.P., and Bours, V. (1999). The homeodomain-containing proteins. *Biochem. Pharm.* **58**, 1851-18

Chisaka, O., Musci, T. S. and Capecchi, M. R. (1992). Developmental defects of the ear, cranial nerves and hindbrain resulting from targeted disruption of the mouse homeobox gene Hox-1.6. *Nature **355***,516 -520.

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.

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

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. and Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, **339** (6121), 819-823.

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., Alsawadi, 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.

Cvejic, A., Serbanovic-Canic, J., Stemple, D. and Ouwehand, W. (2011). The role of *meis1* in primitive and definitive hematopoiesis during zebrafish development. *Hematologica* **96**, 190-198.

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, E. H. (2006). The regulatory genome: gene regulatory networks in development and evolution. Academic Press, San Diego, CA.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J. and Charpentier, E. (2011) CRISPR RNA maturation by trans- encoded small RNA and host factor RNase III. *Nature*, **471**(7340), 602-607.

Deschler, B., and Lübbert, M. (2007). Acute myeloid leukemia: Epidemiology and etiology. *Cancer*, **107**, 2099-2107.

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

Dolle, P., Lufkin, T., Krumlauf, R., Mark, M., Duboule, D. and Chambon, P. (1993). Local alterations of Krox-20 and Hox gene expression in the hindbrain suggest lack of rhombomeres 4 and 5 in homozygote null Hoxa-1 (Hox-1.6) mutant embryos. *Proc. Natl. Acad. Sci. USA* **90**, 7666-7670.

Dounda, J.A. and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR/Cas9. *Science* **346**, doi:1258096.

Drapkin, R., Reardon, J.T., Ansari, A., Huang, J.C., Zawel, L., Ahn, K., Sancar, A. and Reinberg, D. (1994). Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature* **368**, 769– 772.

Dror, I., Zhou, T., Mande-Gutfreund, Y., and Rohs, R. (2013). Covariation between homeodomain transcription factors and the shape of their DNA binding site. *Nucleic Acids Res*. **72**,1-12.

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.

Essner J. (2016) Zebrafish embryo microinjection: Ribonucleoprotein delivery using the Alt-R CRISPR-Cas9 System. [Online] Coralville, Integrated DNA Technologies. [Accessed 21 July, 2017.]

Faderl, S., Talpaz, M., Estrov, Z., O'Brien, S., Kurzrock, R., and Kantarjian, H.M. (1999). The Biology of Chronic Myeloid Leukemia. *N Engl J Med.* **341**, 164-172.

Ferrara, T.J. (2015). Characterization of the Meis2a Downstream Regulatory Element Dr-m2de1. [MS Thesis]. Appalachian State University. 74 pages.

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.

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

Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C. and Schier, A.F. (2014) Efficient mutagenesis

by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single -guide RNAs. *PLoS One* **9**, e98186.

Garneau, J. E., Dupuis, M. E., Vollion , M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A. H. and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacterioplage and plasmid DNA. *Nature*, **468**, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediated specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA***, 109(39),** E2579-E2586.

Gehring, W. J. (1987). Homeoboxes in the study of development. *Science* **236**, 1254-1252.

Gehring, W. J. (1993). Exploring the homeobox. *Gene* **135**, 215-221.

Gehring, W., Quian, Y., Billeter, M., Furukubo-Tokunaga, K., Schier, A., Resendez-Perez, D., Affolter, M., Otting, G., and Wutrich, K. (1994). Homeodomain-DNA Recognition. *Cell*, **78**, 221-223.

Geyer, P. K. (1997). The Role of insulator elements in defining domains of gene expression. *Curr. Opin. Genet. Dev.* **7**, 242-248.

Goddard, J. M., Rossel, M., Manley, N. R. and Capecchi, M. R. (1996). Mice with targeted disruption of Hoxb-1 fail to form the motor nucleus of the VIIth nerve. *Development* **122**, 3217-3228.

González, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q.V. and Huangfu, D. (2014). An iCRISPR Platform for Rapid, Multiplexable, and Inducible Genome Editing in Human Pluripotent Stem Cells. *Cell Stem Cell* **15**, 215-226.

Gould, A., Morrison, A., Sproat, G., White, R.A. and Krumlauf, R. (1997). Postive cross- regulation and enhancer sharing; two mechanisms for specifying overlapping Hox expression patterns. *Genes Dev*. **11**, 900-913.

Griffin, K.J.P., Amacher, S.L., Kimmel, C.B. and Kimelman, D. (1998). Molecular identification of *spadetail:* regulation of zebrafish trunk and tail mesoderm formation by T-box genes. *Development* **125**, 3379-3388.

Grubach, L., Juhl-Christensen, C., Rethmeier, A., Olesen, L. H., Aggerholm, A., Hokland, P. and Østergaard, M. (2008). Gene expression profiling of Polycomb, Hox and Meisgenes in patients with acute myeloid leukemia. *European Journal of Haematolog,* **81**, 112-122.

Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M. Q., Ren, B., Krainer, A. R., Maniatis, T. and Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, **162**, 900–910.

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.

Homeoboxes. (2012). http://ghr.nlm.nih.gov/geneFamily/homeobox

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C. and Schmid, B. (2013). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development*, **140**, 4982-4987.

Hughes, J.R., Cheng, J.F., Ventress, N., Prabhakar, S., Clark, K., Anguita, E., DeGobbi, M., deJong, P., Rubin, E. and Higgs, D.R. (2005). Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9830-9835.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J. R. and Joung, J. K. (2013). Efficient *In Vivo* Genome Editing Using RNA- guided Nucleases. *Nat. Biotechnol.* **31**(3), 227-229.

Irimia, M., Maeso, I., Burguera, D., Hidalgo-Sanchez, M., Puelles, 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.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987). Nucleotide Sequence of the iap Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia coli, and Identification of the Gene Product. *J. of Bac.* **169** (12) 5429-5433.

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

Jao, L.E., Wente, S.R., and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Dev. Bio.* **140**, 4982-4987.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpertier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science,* **337** (6096), 816-821.

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E. and Doudna, J.A. (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science,* **343**(6176), doi:1247997.

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.

Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., Mishina, M. (2004). A transposon- mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev. Cell.* **7**(1), 133-144.

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.

Kimmel, C.B., Kane, D.A., Walker, C., Warga, R.M. and Rothman, M.B. (1989). A mutation that changes cell movement and cell fate in the zebrafish embryo. *Nature* **337**, 358-362.

Knoepfler, P. S., Bergstrom, D. A., Uetsuki, T., Dac-Korytho, I., Sun, Y. H., Wright, W. E., Tapscott, S. J. and Kamps, M. P. (1999). A conserved motif N-terminal to the DNA- binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-meis1/prep1. *Nuc. Acids Res.* **27***(*18*)*, 3752-3761.

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 DNA binding by Pbx1. Mol and Cell Bio. **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

Kotani, H., Taimatsu, K., Ohga, R., Ota, S., Kawahara, A. (2015). Efficient Multiple Genome Modification INdiced by the crRNAs, tracrRNA and Cas9 Protein Complex in zebrafish. *PLoS ONE* **10**(5), e0128319.

Krumlauf, R. (1994). Hox genes in vertebrate development. *Cell*, **78**, 191-201.

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

Kwan, K., Fujimoto, M. E., Grabher, C., Magnum, B. D., Hardy, M. E., Campbell, D. S., Chien, C. B. (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev. Dyn.* **23**(11*),* 3088-3099.

Lamason, R.L., Mohideen, M.A.P.K., Mest, J.R., Wong, A.C., Norton, H.L., Aros, M.C., Jurynec, M.J., Mao, X., Humphreville, V.R., Humbert, J.E., et al. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782-1786.

Lappin, T., Grier, D., Thompson, A., and Halliday, H. (2006). HOX GENES: Seductive science, mysterious mechanisms. *Ulster Med. J.* **75**, 23-31.

Lawerence, H. J., Rozenfeld, S., Cruz, C., Matsukuma, K., Kwong, A., Komuves, L., Buchberg, A. M. and Largman, C. (1999). Frequent co-expression of 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.

Li, X. Y., Virbasius, A., Zhu, X. and Green, M. R. (1999). Enhancement of TBP binding by activators and general transcription factors. *Nature* **399**, 605-609.

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. U.S.A.* **107**(47), 20494-20499.

Long, C., McAnally, J. R., Shelton, J. M., Mireault, A. A., Bassel-Duby, R. and Olson, E. N. (2014). Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science* **345**(6201), 1184-1188.

Longobardi, E., Penkov, D., Mateos, D., De Florian, G., Torres, M., Blasi, F. (2014). Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. *Dev. Dyn.* **243**(1), 59-75.

Maeda, R., Ishimura, A., Mood,K., Park, E. K., Buchberg, A. M. and Daar, I. O. (2002). Xpbx1b and Mmeis1b play a collaborative role in hindbrain and neural crest gene expression in Xenopus embryos. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5448-5453.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Mosuri, S., Yang, L. and Church, G. M. (2013). Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotech.*, **31**, 833-838.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., Dicarlo, J. E., Norville, J. E. and Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, **339** (6121), 823-826.

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

Mann, R. S., and Affolter, M. (1998). Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* **8,** 423-429.

Mark, M., Lufkin, T., Vonesch, J. L., Ruberte, E., Olivo, J. C., Dolle, P., Gorry, P., Lumsden, A. and Chambon, P. (1993). Two rhombomeres are altered in Hoxa-1 mutant mice. *Development*, **119**, 319-338.

Maruyama, T., Dougan, S. K., Truttman, M., Bilate, A. M., Ingram, J. R. and Ploegh, H. L. (2015). Inhibition of non-homologous end joining increases the efficientcy of CRISPR/Cas9-mediated precise [TM: inserted] genome editing. *Nat. Biotechnol.*, **33**(5), 538-542.

Mathis, D. J., and Chambon, P. (1981). The SV4- early region TATA box is required for accurate in vitro initiation of transcription. *Nature* **290**, 310-315.

McClintock, J. M., Kheirbek, M. A. and Prince, V. E. (2002). Knockdown of duplicated zebrafish hoxb1 genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. *Development* **129**, 2339-2354.

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes are axial patterning. *Cell* **68**, 283- 302.

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

Melvin, V., Feng, W., Hernandez-Lagunas, L., Artinger, K. B. and Williams, T. (2013). A Morpholino Based Screen to Identify Novel Genes Involved in Craniofacial Morphogenesis. *Dev. Dyn.* **242**, 817-831.

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.

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

Moens, C. and Prince, V. (2002). Constructing the hindbrain: insights from the zebrafish. *Dev. Dyn*, **224**(1), 1-17. Dio:10.1002/dvdy.10086

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

Mojica, F. J., Diez-Villasenor, C., Soria, E. and Juez, G. (2000). Biological significance of a family of regularly interspaces repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.,* **36**, 244-246.

Monica, K., Galili, N., Nourse, J., Saltman, D. and Cleary, M. L. (1991) PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1. *Mol Cell Biol* **11**, 6149–6157

Morata, G. and Lawerence, P. A. (1977). The development of wingless, a homeotic mutation of Drosophila. *Dev. Bio.* **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 Bio.* **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.

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

Nakamura, T. (2005). Meis and Hox: a mighty pair defeats apoptosis. *Blood* **105**, 909-910.

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., Largaespada, D. A., Shaughnessy, J. D., Jenkins, N. A. and Copeland, N. G. (1996). Cooperative activation of *Hoxa* and *Pbx1*-related genes in murine myeloid leukemias. *Nature Genetics* **12**, 149-153.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F. and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, **156**, 935-949.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., DeLaat, W. and Duboule, D. (2001). The dynamic architecture architecture of *Hox* gene clusters. *Science*. **334***(*6053*)*, 222-225.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev*. **10**, 2657-2683.

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

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

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.

Parker, H. J., Sauka-Spengler, T., Bronner, M. and Elgar, G. (2014). A reporter assay in lamprey embyos reveals both functional conservation and elaboration of vertebrate enhancers. *PLoS ONE,* **9**(1), e85492.

Pennisi, E. (2013). A bacterial immune system yields a potentially revolutionary genome-editing technique. *Science*, **341** (6148), 883-836.

Pourcel, C., Salvignol, G. and Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA and provide additional tools for evolutionary studies. *Microbiol.,* **151**, 653-663.

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.

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F. (2013). Double Nicking RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380- 1389.

Roberts, V.J., van Dijk, M.A. and Murre, C. (1995). Localization of Pbx1 transcripts in developing rat embryos. *Mech. Dev.* **51**, 193–198.

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.

Sanjana, N. E., Cong, L., Zhou, Y., Cunniff, M. M., Feng, G. and Zhang, F. (2012) A transcription activator-like effector toolbox for genome engineering. *Nat. Prot.*, **7,** 171- 192.

Santos, J. S., Fonseca, N. A., Vieira, C. P., Vieira, J. and Casares, F. (2010). Phylogeny of the Teashirt-related zinc finger (tshz) gene gamily and analysis of the developmental expression of *tshz2* and *tshz3b* in the zebrafish. *Developmental Dynamics* ***239***, 1010- 1018.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrougou, R., Horvath, P. and Siksnys, V. (2011). The *Streptococcus thermophiles* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.,* **39**(21), 9275-9282.

Shah, S. A., Erdmann, S., Mojica, F. J. and Garrett, R. A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.,* **10**, 891-899.

Shalem O., Sanjana N. E., Hartenian E., Shi X., Scott D. A., Mikkelsen T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G. and Zhang, F. (2014) Genome-scale CRISPR- Cas9 knockout screening in human cells. *Science,* **343**(6166), 84-87.

Shang, Z., Isaac, V. E., Li, H., Patel, L., Catron, K. M., Curran, T., Montelione, G. T., and Abate, C. (1994). Design of a "minimal" homeodomain: The N-terminal arm modulates DNA binding affinity and stabilizes homeodomain structure. *PNAS* **91**, 8373-8377.

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

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

Shen, W. F., Rozenfeld, S., Lawrence, H. J. and Largman, C. (1997). The Abd-B-like Hox homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target. *J. Bio. Chem.* **272**, 8198-8206.

Shih, L. J., Tsay, H. J., Lin, S. C. and Hwang, S. P. (2001). Expression of zebrafish Hoxa1a in neuronal cells of the midbrain and anterior hindbrain. *Mech. Dev.* **101**, 279 -281.

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.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507(7490)**, doi:10.1038.

Studer, M., Lumsden, A., Ariza-McNaughton, L., Bradley, A. and Krumlauf, R. (1996). Altered segmental identity and abnormal migration of motor neurons in mice lacking Hoxb-1. *Nature* **384**, 630-634.

Sung, Y.H, Kim, J.M., Kim, H.T., Lee, J., Jeon, J., Jin, Y., Choi, J.H., Ban, Y.H., Ha, S.J., Kim, C.H., Lee, H.W. and Kim, J.S. (2014). Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. *Genome Res.* **24**, 125-131.

Talbot, J.C. and Amacher, S.L. (2014). A streamlined CRISPR pipeline to reliably generate zebrafish frameshifting alleles. *Zebrafish* **11(6)**, 583-585.

Tamaoki, M., Tsugawa, H., Minami, E., Kayano, T., Yamamoto, N., Kano-Murakami, Y., and Matsuoka, M. (19995). Alternative RNA products from a rice homeobox gene. *Plant Journal.* **7(6)**, 924-938.

Thisse, C. and Thisse B. (2005). High throughput expression of zf-models consortium clones. ZFIN Direct Data Submission. (www.zfin.org)

Thomas, M.C. and Chiang, C. (2006). The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology* **41**, 105-178.

Thorsteindottir, U., Kroon, E., Jerome, L., Blasi, F. and Sauvageau, G. (2001). Defining roles for HOX and MEIS1 Genes in Induction of Acute Myeloid Leukemia. *Molecular and Cellular Biology* **21(1)**, 224-234.

Toresson, H., Parmar, M. and Campbell, K. (2000). Expression of *Meis* and *Pbx* genes and their protein products in the developing telencephalon: implications for regional differentiation. *Mech. of Dev.* **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. *Bloof* **104**, 3106-3116.

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.

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**, 7125-7179.

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**, 4139–51.

Westerfield, M. (2000). *The Zebrafish Book: A guide for the laboratory use of zebrafish (Danio rerio).* Oregon: Institute of Neuroscience, University of Oregon.

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

Wolpert, L. (2007). Principles of development. Oxford University Press, Oxford, U.K.

Wood, A. J., Lo, T., Zeitler, B., Pickle, C. S., Ralston, E. J., Lee, A. H., Amora, R., Miller, J. C., Leung, E., Meng, X., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D. and Meyer, B. J. (2011). Targeted genome editing across species using ZFNs and TALENs. *Science*, **333** (6040), 307.

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

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

Zawel, L., Kumar, K. P. and Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev.* ***9***, 1479-1490.

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

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

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

Zhu Z. and Huangfu, D. (2013). Human pluripotent stem cells: an emerging model in developmental biology. *Development* **140**, 705–717.

Supplementary Material

Supplementary Table 1: Target sequence for Spadetail gRNA and Golden gRNA

gRNA	Target Sequence
Spadetail	5'-GGGTGCAGGTACGTCCTGTA-3'
Golden	5'-GGTCTCTCGCAGGATGTTGC-3'

Supplementary Table 2: Embryos injected with Spadetail gRNA + Cas9 mRNA

Concentration gRNA/ Cas9 (pg/nl)	40/80	25/100	75/250	12.5/300
Number injected	112	75	467	523
Number alive	78	59	309	298
Number w/ phenotype	0	0	0	0
% successfully targeted	0%	0%	0%	0%

Vita

Megan Daley Tennant was born in Crossnore, North Carolina. She graduated from Avery County Highschool in North Carolina in May 2012. The following autumn, she entered Bluefield College to study Biology and in May 2015 she was awarded the Bachelor of Science degree. In the spring of 2016, she began her study toward a Master of Science in Cell and Molecular Biology at Appalachian State University. She is scheduled to graduate in May 2018. She resides in Crossnore, NC.