THE ROLE OF THE HIGHLY CONSERVED NON-CODING ELEMENT m2de1 ON THE REGULATION OF THE EXPRESSION OF THE *Meis2* AND *zgc:154061* GENES

A Thesis
by
Young Koun Jeon

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

December 2019
Department of Biology

THE ROLE OF THE HIGHLY CONSERVED NON-CODING ELEMENT m2de1 ON THE REGULATION OF THE EXPRESSION OF THE *Meis2* AND *zgc:154061* GENES

A Thesis
by
YOUNG KOUN JEON
December 2019

APPROVED BY

_____
Dr. Ted Zerucha
Chairperson, Thesis Committee

_____
Dr. Cortney Bouldin
Member, Thesis Committee

_____
Dr. Andrew Bellemer
Member, Thesis Committee

_____
Dr. Zack Murrell
Chairperson, Department of Biology

_____
Dr. Mike McKenzie
Dean, Cratis D. Williams School of Graduate Studies

Abstract

THE ROLE OF THE HIGHLY CONSERVED NON-CODING ELEMENT m2de1 ON THE REGULATION OF THE EXPRESSION OF THE *Meis2* AND *zgc:154061* GENES

Young Koun Jeon
B.S., Lees-McRae College
M.S., Appalachian State University
Chairperson: Dr. Ted Zerucha

The *Meis* genes are members of the three amino acid loop extension (TALE) class, of the homeobox super gene family. Three homologs of the *Meis* genes are found in vertebrate species, including human, chicken, mice and zebrafish. Transcription factors encoded by *Meis* genes act as cofactors with other homeodomain proteins to regulate the expression of target genes during early embryonic development. While the *Meis* genes have been fairly well-characterized for their molecular functions, and spatial and temporal expression pattern, the mechanism of their regulation has not been well studied. The Zerucha lab, has identified four highly conserved noncoding elements (m2de1-4) associated with the *Meis2* gene that are hypothesized to represent enhancers controlling *Meis2* expression. To date, only m2de1 has been found in teleosts, including zebrafish. In addition, m2de1 has been found to be in an intron of the *zgc:154061* gene. This genomic organization is well conserved among land vertebrates and teleosts. One possible reason for the linkage of the *Meis2* and *zgc:154061* genes is that they are sharing regulatory elements. I report here that *meis2a* and *zgc*:*154061* are expressed in a similar pattern as a transgene directed by m2de1. Most of these overlapping regions are in the eye and brain. Excision of a region of m2de1 using CRISPR/Cas9 results in decreases in expression of both *meis2a* and *zgc:154061* in these overlapping regions. The region of m2de1 that was

excised was determined to contain a putative binding site for Pax6 which has been shown to regulate *Meis2* expression and knocking down *pax6b* expression in zebrafish lead to a similar decrease in expression of *meis2a* and *zgc:154061*. Together these results suggest that m2de1 is indeed a shared regulatory element controlling *meis2a* and *zgc:154061* expression.

Acknowledgement

First and foremost, I would like to thank my advisor, Dr. Ted Zerucha, for all of his time and effort during my study as a graduate student. Without his guidance, I would not have been able to complete the program in two and a half years. I would also express my feeling of appreciation to my committee members, Dr. Cort Bouldin and Dr. Andrew Bellemer, for their assistant and knowledge. I am also grateful for Dr. Ece Karatan and Dr. Ray William who guided me numerous time through graduate school affairs. Lastly, I am thankful to Dr. Guichuan Hou in the microscopy facility and Monique Eckert in the animal facility for the time and effort spent on me over the past two and a half years. I am in much debts to my Biology graduate student peers for their support and assistance when I was desperate. I am also indebted to the Biology Department for training provided and Appalachian State University Office of Student Research for financial support. Finally, I thank my parents for their love and support during my graduate program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

This project describes the role of the putative regulatory element m2de1 (*Meis2* Downstream Element 1) in regulating the expression of *Meis2* and homologs of the zebrafish *zgc: 154061* gene. Both *Meis2* and *zgc: 154061* homologs are organized as convergently transcribed neighboring genes, a genomic arrangement that is conserved across vertebrates. A possible reason for the conservation of this genomic organization is that these genes are sharing one or more *cis*-regulatory regions. Previously, the Zerucha lab has identified four highly conserved non-coding elements (HCNEs) associated with this genomic locus and that are candidates for regulating *Meis2* and *zgc: 154061* expression (Tennant, 2018). To date only one of these elements, m2de1, has been found in all vertebrates. We propose that this is the most likely candidate to be maintaining the linked genomic organization between *Meis2* and *Zgc: 154061*. If m2de1 is indeed regulating *Meis2* and homologs of *zgc:154061*, I predict that compromising function of m2de1 will result in a change of the *Meis2* and *zgc: 154061* expression pattern.

Regulation of Gene Expression

Gene expression is one of the most fundamental and crucial processes in all life forms. This process involves the synthesis of functional ribonucleic acids (RNAs) and proteins for controlling cellular functions. Genes within certain tissues need to be activated during development, a process which controls the differentiation of cells. The regulation of gene expression is controlled by various regulatory molecules, including transcription factors and cell-signaling molecules. Knowing where and when a specific

regulatory molecule is expressed is crucial to understand its function (Hinman and Cary, 2017).

General Eukaryotic Transcription

Several important steps are required in the expression of a specific gene. These include transcription, RNA splicing, translation, and post-translational modification. While every process is crucial in the expression of the gene, the focus of this work is on transcription regulation. Transcription is the process to produce an RNA molecule based on the genetic information encoded in the double stranded DNA. RNA synthesized from a specific template DNA contains the same genetic information and consists of the complementary sequence from its specific template DNA. The initiation of transcription requires recruitment of the RNA polymerase (Pol) enzyme and a set of transcription factors (TFs) assembled at the promoter region of a gene (Heyn et al., 2015).

TFs are key regulators that play a crucial role in regulating transcription rates. While, in general, TFs bind regions of DNA, to regulate the specific expression of the gene they are associated with, the functions of TFs are varied with respect to the genes they associate with, stage of development when associated genes are expressed and how combinations of TFs form a complex. Based on the different combinations of TFs bound to a regulatory element, transcriptional rates will be upregulated or downregulated to enhance or suppress the transcription process (Spitz and Furlong, 2012). Examples of TFs that act in this way are coded for by members of the homeobox gene family, which plays crucial roles in morphological development of the embryo and cell proliferation/differentiation. The products of homeobox genes, homeodomain proteins,

often function by forming different combination of dimeric or trimeric complexes that bind to regulatory elements of specific target genes to upregulate or downregulate the gene expression process (Nunes et al., 2003).

Cis-Regulatory Elements

Cis-regulatory elements are non-coding DNA sequences that are involved in gene expression and transcriptional regulation. In the human genome, non-coding DNA sequence comprises approximately 98% of the entire genome, which suggests the presence of many regulatory elements with varied functions (Alexander et al., 2010). The types of regulatory elements include promoters, enhancers, silencers and insulators, which play crucial roles that either promote or suppress the Pol-template DNA interaction. During the transcription process, *cis*-regulatory elements change the transcription level by interacting with the promoter region of a target gene. Even though *cis*-regulatory elements and the promoter region can be located far away from each other, the rate of transcription can still be affected by *cis*-regulatory elements (Li et al., 2015).

Promoter sequences tend to be very well conserved from gene to gene to serve as binding sites for GTFs. For a *cis*-regulatory element, typically, many different binding sites are present for the binding of different TFs, which subsequently play the variation of functions. Different TFs are synthesized in cells from different genomic locations and tissues, and carry out their different functions at specific development stages. The interaction between TFs and *cis*-regulatory elements regulate expression of genes during the different development stages in various location of tissues and organs (Gan et al., 2014).

Cis-Regulatory Elements: Enhancer

An enhancer is a type of *cis*-regulatory element that is involved in a direct interaction with the promoter to increase transcription efficiency. Enhancers are diverse in the positions relative to the promoter (Kulaeva et al., 2012). During transcription, enhancers assist efficient PIC formation by placing both enhancer and promoter close to each other. Subsequently, the enhancer promotes assembly of components for the PIC and its binding to the promoter, which efficiently initiates the transcription initiation phase (Pennacchio et al., 2013).

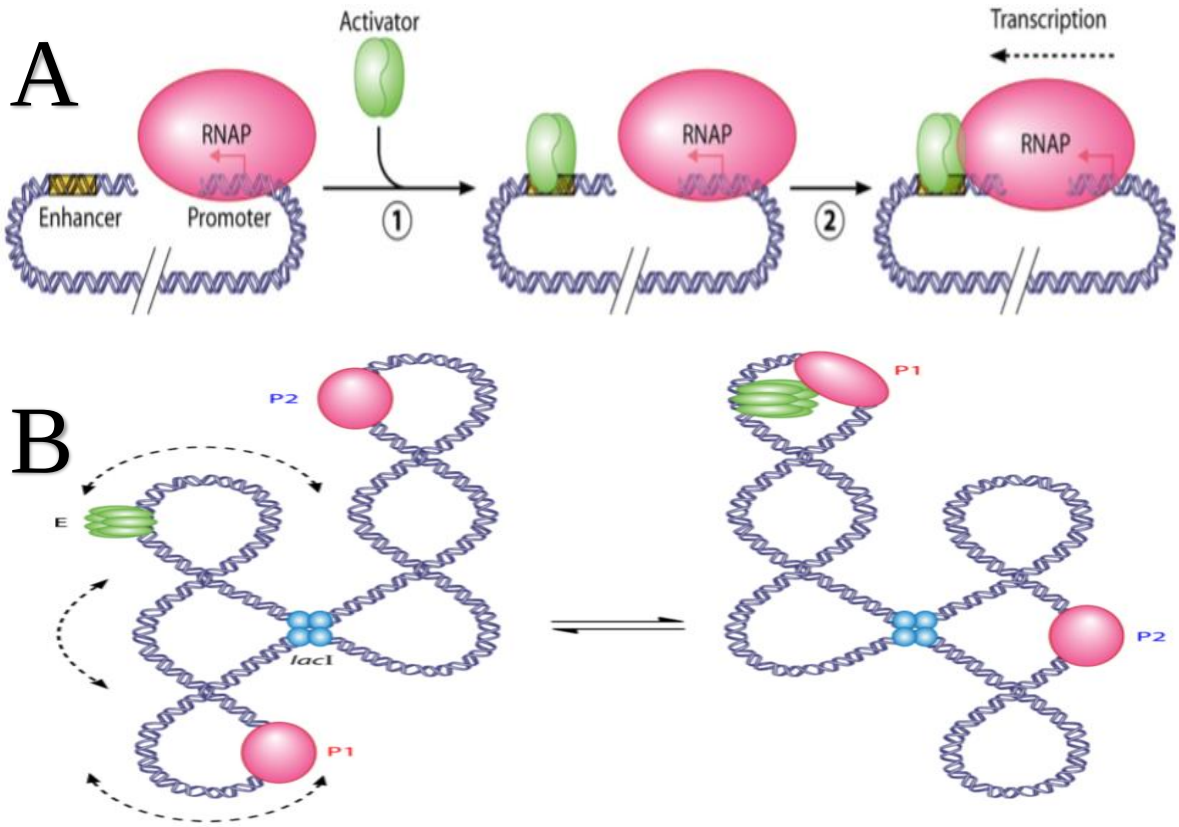

Fig. 1 Schematic image of enhancer mechanism. (A) DNA loop mechanism for enhancer action. The activator brings enhancer and promoter together to enhance the recruitment of GTFs. (B) Slithering or tracking mechanism of enhancer action. The enhancer moves toward the target promoter1 (P1) via DNA supercoiling and branches formation. Insulator lac1 is avoided through supercoiled DNA (Kulaeva et al., 2012).

Two different models have been proposed to describe how enhancers can influence the rate of transcription. Both models explain how an enhancer and promoter can interact despite the large distances between them. Although an enhancer can be located many kilobases away from a target promoter region, the efficiency of enhancer-promoter interaction will not be affected (Matthews, 1992). The first model suggests that the DNA forms a loop to draw an enhancer and the promoter close together thus allowing the interaction between them (Fig. 1-A). The efficiency of the enhancer-promoter interaction increases by compacting non-coding DNA between the enhancer and promoter, which brings the enhancer and promoter closer to one another, and corresponds with a greatly increased recruitment of PIC components. The second model, the slithering or tracking model suggests the fast movement of an enhancer towards the promoter by traveling along the chromatin (Fig. 1-B). However, this model requires DNA supercoiling and subsequent formation of DNA branches, which increase the juxtaposition proximity between distally located enhancer and promoter and helps to avoid the protein bridge built by an insulator. (Kolovos et al., 2012; Kulaeva et al., 2012).

Cis-Regulatory Elements: Insulator

An insulator is a type of *cis*-regulatory elements that prevent the function of enhancers. Insulators in different organisms, such as yeast, vertebrates and *Drosophila*, show sequence similarities, which suggest that this mechanism is highly conserved across species. During insulator evolution, some properties have been conserved that are similar to common properties of promoters, such as the ability of these regions of the chromosome to be modified to become accessible to TFs and mediating long distance

interaction (Raab and Kamakaka, 2010). This may explain the ability of insulator to change the conformation of a nucleosome to allow the accessibility to chromatin, and disrupt the long distance interaction of promoter and enhancer (West et al., 2002).

The function of insulators involves the prevention of an enhancer and promoter interaction, likely by blocking the formation of loops (West et al., 2002). Mutating or knocking out an insulator has been found to change the expression pattern of a gene and to lead to abnormalities during developmental processes. The most common example of a protein that functions as an insulator is the CCCTC-binding factor (CTCF). CTCF can prevent the activation of the maternal *Igf2* allele by mediating insulator-insulator and insulator-enhancer interactions to prevent the interaction between the enhancer and *Igf2* promoter region. Mutations in CTCF are associated with invasive breast cancers, prostate cancers, and Wilms' tumors, affecting the reactivation of *Igf2* allele (Kolovos et al., 2012).

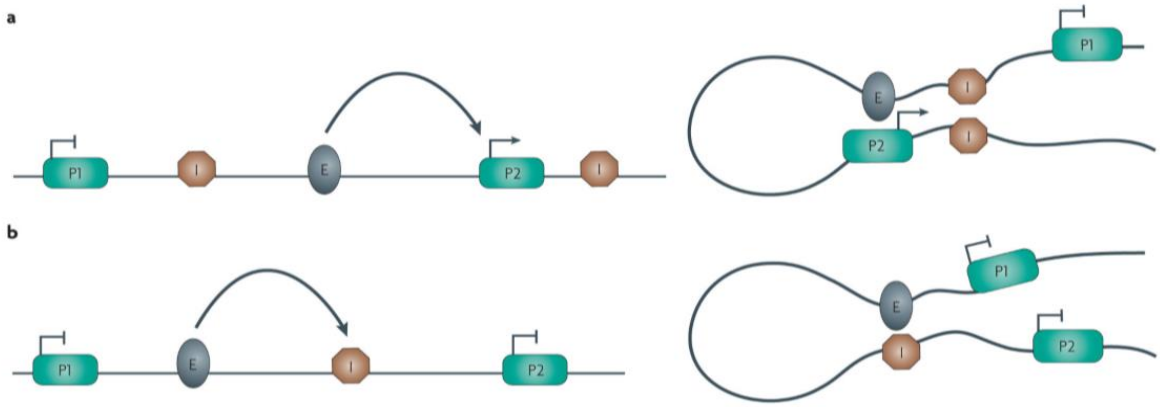

Fig. 2 Schematic image of interaction among enhancers, promoters and insulators. (A) A pair of insulators (I) interact each other to prevent the interaction between enhancer (E) and promoter 1 (P1). However, two insulators cannot prevent the loop formation between E and promoter 2 (P2). (B) An insulator directly binds to E and prevent its ability to interact with a promoter (Raab and Kamakaka, 2010).

Two models have been proposed to describe the mechanism of insulator function. First, the insulator can block the interaction between a distal enhancer and the target promoter (Fig. 2). However, this requires the insulator to be located between the enhancer and promoter. In this model, the insulator acts as a blocker and directly binds to the promoter, instead of the enhancer, or multiple insulators interact and act as a partition to prevent the interaction between enhancer and promoter. While the blocking of an enhancer will prevent the expression of the target gene, the promoter region can initiate gene expression together with another unblocked enhancer through the slithering model. Second, an insulator can function as a barrier to prevent the interaction with nearby condensed chromatin. This function requires the insulator to be located between the promoter and silencer, which is another type of *cis*-regulatory element. A barrier insulator will block the formation of a chromatin loop, which is likely required to increase the efficiency of enhancer and promoter interaction. In addition, some insulators can function in both ways as blocker and barrier (Raab and Kamakaka, 2010; West et al., 2002).

Cis-Regulatory Elements: Silencers

Silencers are a type of *cis*-regulatory elements that suppress gene expression during transcription and is also known as a negative regulatory elements (NREs). Silencers are located in the 5' region upstream of a promoter (Ogbourne and Antalis, 1998). Repressors are a type of DNA binding protein, which interact with silencers at specific binding sites and influence transcription by repressing TSS from the binding of GTFs required for the formation of PIC. Motifs that are putative silencers are rich with TF binding sites for repressors to bind (Jayaveluet al., 2018). However, the lack of

annotated chromatin signatures in silencers makes it difficult to identify them. Consequently, the distinction between silencer and other unknown non-coding DNA sequences is difficult. In addition to the difficulty in finding differences between putative silencers and non-coding DNA sequences, many putative silencers that interact with promoters have low or no expression. One possible method to identify putative silencer elements is to search for repressor TF binding sites and motifs (Jayaveluet al., 2018).

Silencers are capable of suppressing gene expression in various ways. For a position dependent silencer, NREs can suppress the transcription process by flanking the crucial regulatory proteins and promoter region. The competition between activator proteins for the same binding sites with repressor proteins can affect gene expression. For example, the NRE associated with the *c-fos* promoter can bind to nuclear factor Yin-Yang 1, which prevents the interaction between the *cis*-regulatory elements and *c-fos* promoter (Ogbourne and Antalis, 1998). Also, a silencer located in an intron of the CD4 gene can suppress gene expression by inhibiting CD4 expression in CD8+ T cells. The CD4 specific silencer represses the promoter, which inhibits the expression of CD4 while not affecting CD8 (Jayavelu et al., 2018). A position dependent NRE present in the first intron of the CD4 gene possesses a protein binding site, which is related to the repressor binding (Donda et al., 1996). For position independent silencers, silencers interact with chromatin to repress the conformation change by preventing the conversion of CC to OC, which is associated with between the components of the general transcriptional mechanism and promoter (Ogbourne and Antalis, 1998).

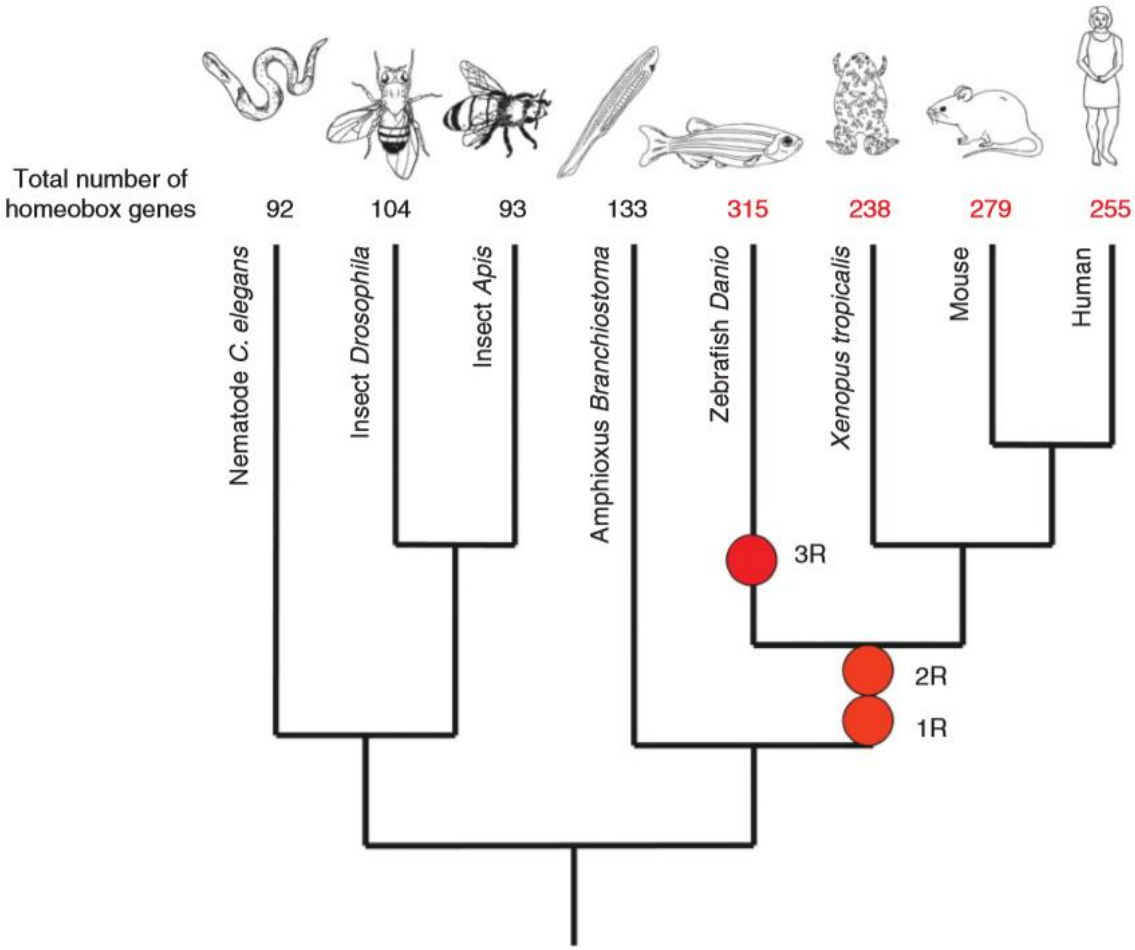

Fig. 3 Evolutionary lineage of homeobox genes. Total number of different homeobox genes are present in the genomes of particular species. The numbers may change with each release of a newly revised genome data. 1R, 2R and 3R are indicating whole genome duplication events (Holland, 2013).

Homeobox Gene Family

The TFs that bind to enhancers and silencers, unlike GTFs that bind the promoter, are specific to specific genes and cell types. An example of these types of specific TFs are homeodomain proteins, which are encoded by homeobox genes. The homeobox gene family is the superfamily of genes that encodes transcription factors that play roles as key regulatory proteins during embryonic development. The homeodomain transcription factors regulate morphogenesis and cellular proliferation/differentiation

during embryonic development in many organisms. Homeobox genes can be found in almost all eukaryotes, and are classified into 11 gene classes. The 11 different classes are composed of over 100 gene families. The largest class in vertebrates is the Antennapedia (ANTP) class, which includes the *Hox* genes and many other genes involved in development. The members of the homeobox gene family have gone through many evolutionary changes, which include gene and genome duplication events, to produce divergent sequences and/or carry out different functions. Some animal species possess more divergent homeobox genes than other species because of additional whole genome duplication events (Fig. 3) (Holland, 2013).

The homeobox gene family can be found in various species including animals, plants and fungi. The very first homeobox gene presumably evolved from early eukaryotic species and extensive gene duplications have produced the broad diversity of the homeobox gene family. However, it is unclear how many functional homeobox genes are present in the genomes of organisms. For example, studies on the human genome have identified multiple sequences that suggest over a hundred different Homeobox sequences (Holland and Takahashi, 2005). Because of the insufficient data on automated annotation, it is difficult to distinguish the functional homeobox genes and pseudogenes. However, the study successfully annotated functional homeobox genes, supported variants and pseudogenes in human and mouse genome by manual annotation (Wilming et al., 2015)

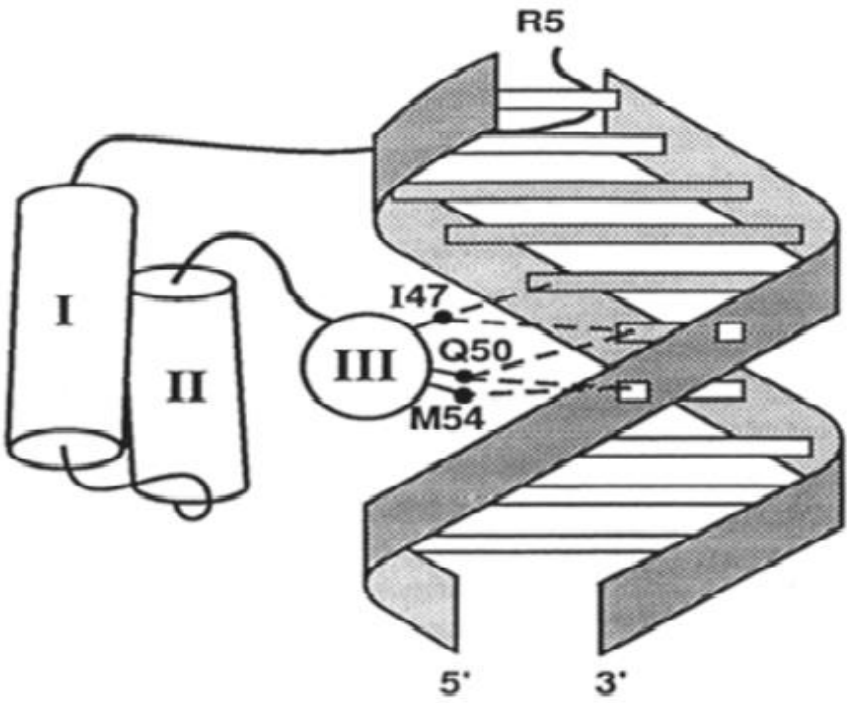

Fig. 4 Schematic diagram of the Antp homeodomain-DNA complex. Two alpha helices helix (I)-turn-helix (II) with homeodomain motif (III) are displacing. The contacts between amino acid residues of homeodomain and specific bases of DNA are shown (Gehring et al., 1990).

The first homeobox gene discovered was in the fruit fly *Drosophila melanogaster* in mutations of genes that resulted in homeotic transformations. These mutations transform the identity of one part of the *Drosophila* body into another part, such as the change of identity of a thoracic segment that does not normally have wings into one that does, which results in a fly with four wings instead of two. The sequences of all homeobox genes contain an approximately 180-nucleotide sequence, which encodes a specific protein motif known as the homeodomain. The homeodomain is an approximately 60-amino acid domain that acts as a DNA-binding domain, which is important for the overall function of the entire protein as a transcription factor. The structure of the homeodomain comprises a helix-turn-helix motif that interacts with the

major groove of DNA (Fig. 4). In addition, the peptide motif of the homeodomain is similar to the bacterial helix-turn-helix proteins, which suggests the homeobox sequence may have evolved from ancestral helix-turn-helix genes (Holland, 2013).

The most well known function of homeobox genes in embryonic development is the patterning of the antero-posterior (A-P) axis. However, homeobox genes are also involved in crucial biological processes of eukaryotic cells such as the regulation of cell fate, cell growth and cell differentiation. While the homeobox genes have both tumor-enhancing and tumor-repressing properties, the mutation of homeobox genes will often result in abnormal cell proliferation with the consequence being cancer (Cillo, 1994). For example, the study showed the role of homeobox genes in the development of Hepatocellular carcinoma (HCC), a type of liver cancer (Liu et al., 2019). In addition, homeobox gene mutations are involved in human limb defects such as the hand-foot-genital syndrome (Mortlock and Innis, 1997).

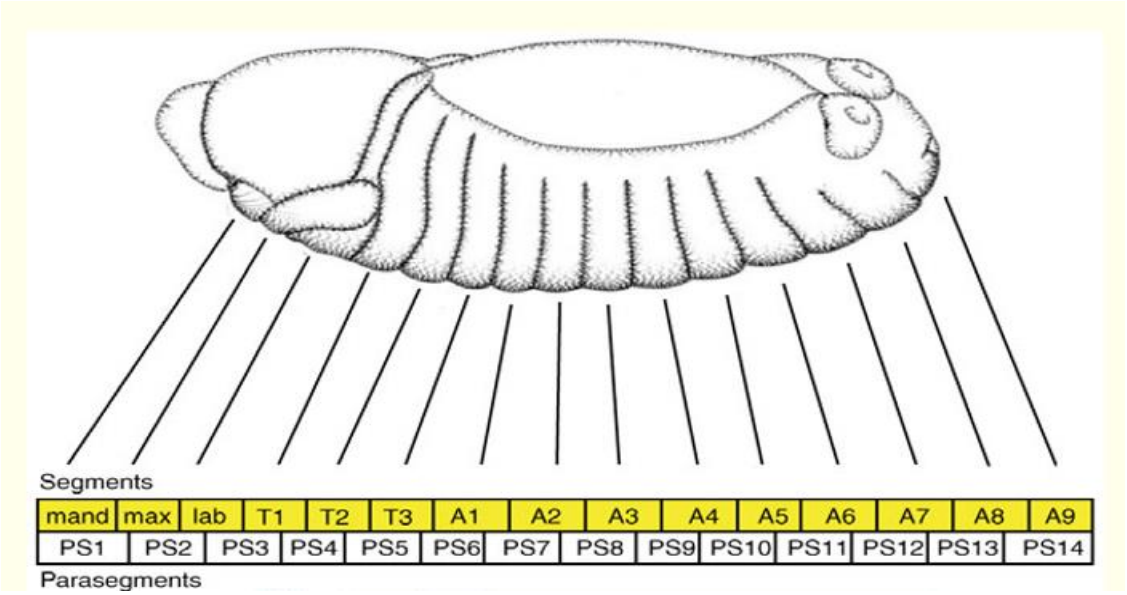

Fig. 5 Schematic image of segments and parasegments in the *Drosophila* larvae. Parasegments and segments are out of phase by one compartment each (Maeda and Karch, 2006).

In *Drosophila*, the ANTP class of homeobox gene family encodes the homeodomain containing proteins found in the Antennapedia and Bithorax complexes, which are referred to as the Homeotic complex (HOM-C). The ANTP class genes include *Labial* (*Lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), and *Abdominal-A* (*Abd-A*). Mutation of these genes can result in homeotic transformations in *Drosophila*. For example, *Lab* is expressed in the most anterior region of the developing fly embryo and mutating it can result in a missing head or defects in the head. As another example, *Abd-A* is expressed the posterior abdominal region and its mutation can result in the homeotic transformation of parasegments 7 through 12 to parasegment 6 through 11 (Fig. 5). The homeotic genes are expressed in a pattern that is often described as having spatial and temporal collinearity. Genes that are located more towards the 3' end of the cluster are expressed earlier and more anteriorly in the developing embryo while genes located closer to the 5' end of the cluster are expressed later and in more posterior regions. The presence of the ANTP class of homeobox genes is also found in vertebrate genomes with similar genomic organization and expression patterns to that observed in *Drosophila* (McGinnis and Krumlauf, 1992).

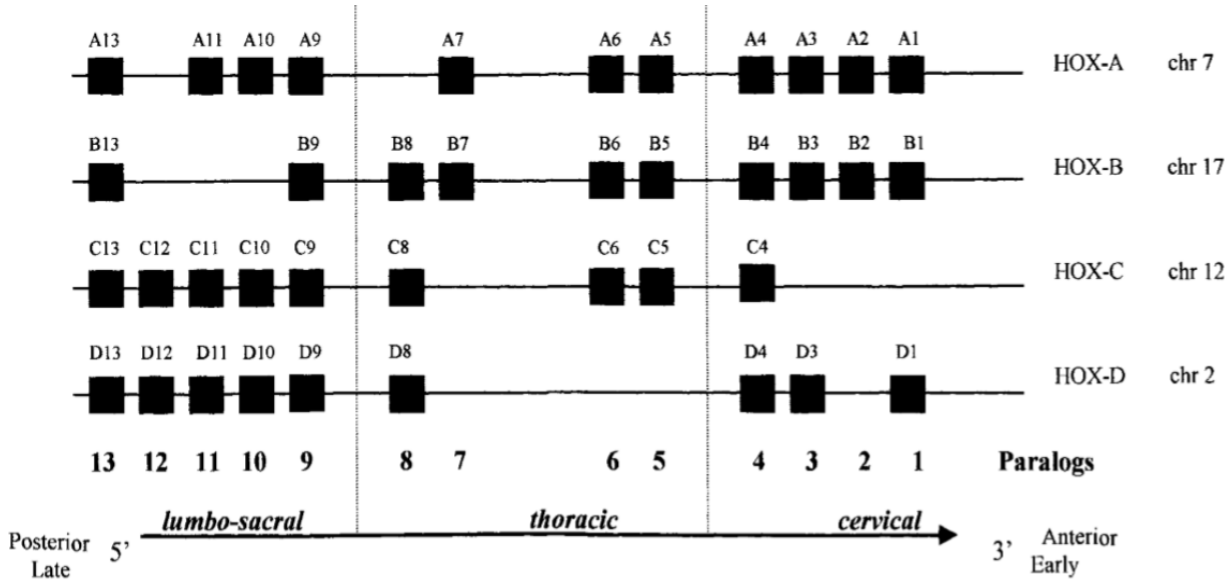

Fig. 6 Schematic representation of the four Hox clusters or loci in human that are aligned according to the position of chromosome with the spatial expression pattern of each gene (Alfredo, 2015).

Hox Genes

In vertebrate, the ANTP class *Hox* genes are organized into four clusters that are composed of multiple *Hox* genes and expressed along the body axis in a corresponding manner to their position along the chromosome, as is seen in *Drosophila*. The *Hox* genes play many crucial roles including patterning of the AP axis of the body, the limb bud axis, hematopoiesis, organogenesis, angiogenesis and more. In human and mice, there are at least 39 genes organized into four clusters (HOXA-D) (Fig. 6). Each cluster is located on a different chromosome and possesses a different number of *Hox* genes. Because the *Hox* genes participate in patterning the AP body axis, mutation of *Hox* genes can result in the malformation of the body plan (Moens and Selleri, 2006; Nunes et al., 2003).

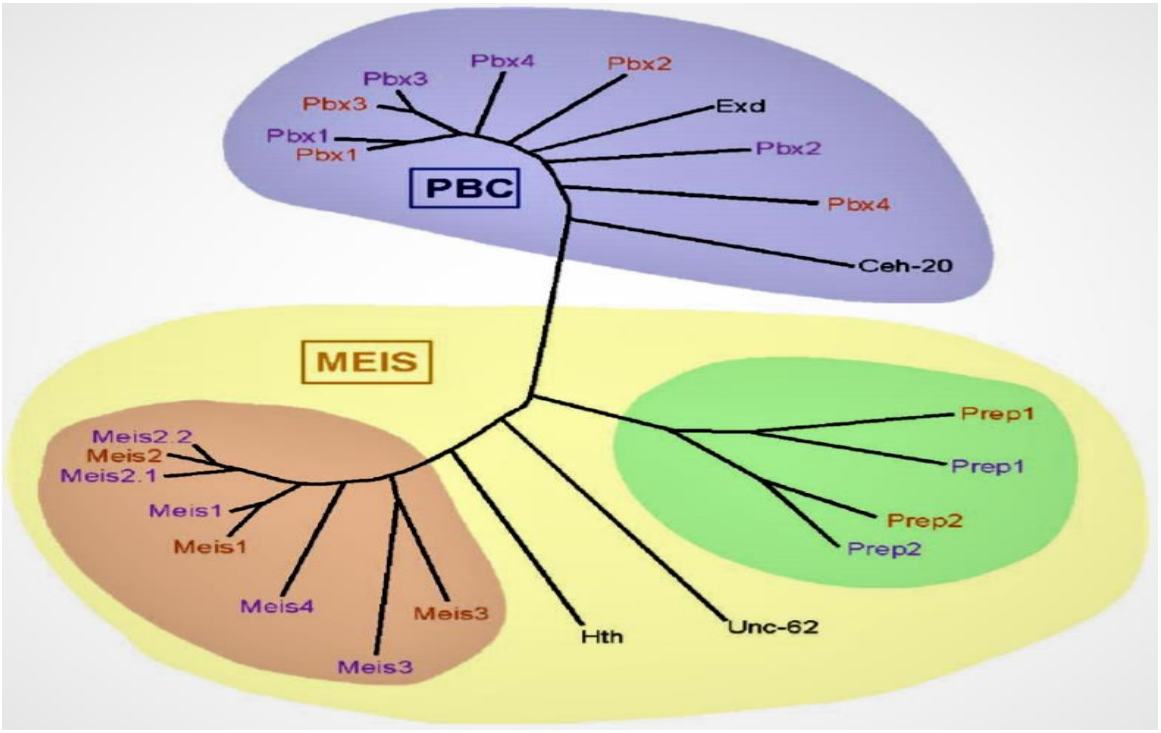

Fig. 7 Phylogeny and Hox cofactors structure. There are two groups including PBC and MEIS family. Orange letters are indicating mouse proteins and purple letter is indicating zebrafish orthologues (Moens and Selleri, 2006).

The *Hox* genes encode homeodomain-containing proteins that act as transcription factors to carry out specific functions yet they bind to remarkably similar DNA sequences. Additional specificity of DNA binding is achieved by these proteins working together with other DNA-binding proteins referred to as Hox cofactors. Without the interactions of cofactors, the specificity of Hox proteins binding DNA decreases. These cofactors include the PBC and MEIS classes in the TALE (Three Amino Acid Loop Extension) class of the homeobox gene family (Fig. 7). While the PBC class includes *Drosophila* Extradenticle (Exd) and vertebrate Pbx proteins, the MEIS class includes *Drosophila* Homothorax (Hth), and vertebrate Meis and Prep proteins (Moens and Selleri, 2006).

In *Drosophila*, Exd was first identified as a Hox cofactor based on the mutant phenotype including homeotic transformation of specific body segments. The studies found that the Exd-Pbx protein complex interacts with some Hox proteins to bind DNA with high specificity (Chan et al., 1994). Alternatively, depending on the target, Hox-Exd or Hox-Pbx dimers can either function as transcriptional activators or repressors. In vertebrates, similar complexes are observed between Meis and Prep proteins with Hox proteins. Meis and Prep proteins can also function to enhance the nuclear localization and the stability of Pbx proteins (Abu-Shaar et al., 1999). While some Hox cofactors are Hox-dependent, others are less dependent on Hox proteins, which suggests the broad activities of transcription factor complexes with varied combinations of Hox cofactors. For example, some Hox proteins interact with Pbx to increase DNA-binding affinity while other Hox proteins show a minimal effect of DNA-binding affinity (Moens and Selleri, 2006).

Hox proteins are remarkably well conserved among diverse species. Through studying changes in the expression patterns of *Hox* genes across species, these changes have come to be associated with the evolution of novel body plans of bilaterally symmetric animals. The function of *Hox* genes in the patterning of the AP axis for *Drosophila* is also conserved in vertebrates, which indicates evolutionary conservation (Dobuoule, 1989). During evolutionary history, the sponges diverged first from the metazoans, followed by cnidarians (jellyfish and corals), and both groups are more basal to the Bilaterians (Fig. 3). Sponges do not appear to have definitive *Hox* genes. On the other hand, definitive Hox-like genes are identified in cnidarians, however there is no

significant evidence of their involvement in AP patterning. Urbilateria, the hypothetical common ancestor to bilateral animals, is thought to have possessed an ancestral cluster of at least six *Hox* genes, which evolved from a single ancestral *Hox* gene that underwent a number of tandem gene duplications to form a cluster of multiple genes. Arthropods and vertebrates inherited their Hox clusters from Urbilateria. In invertebrates, typically, a single Hox cluster is present, however multiple clusters are present in vertebrates. For example, mammals possess four Hox clusters and the teleosts have up to seven Hox clusters (Amores et al., 1998).

Pbx Genes

The *Pbx* genes are a member of the TALE class homeobox genes. They are homologues of the *Exd* gene in *Drosophila*, and four homologs are present in the human genome (PBX1-4). The *Pbx* genes encode homeodomain-containing proteins that act as transcription factors like the *Hox* genes, but *Pbx* genes are not organized into clusters. Also, *Pbx* genes encode highly conserved regions outside of the homeobox that code for protein motifs involved in the recruitment of other transcription factors including Meis and Prep proteins. During recruitment, Pbx proteins form strong interactions with Hox paralog groups 1-11 proteins for the initiation of transcription. (Longobardi et al., 2014; Morgan et al., 2017).

The *Pbx* genes are also known to be involved in oncogenic functions. The most common example is functioning as a chimeric fusion partner in various leukemia and lymphoma. For example, in humans, the *Pbx1* gene and *E2A* gene are involved in encoding transcription factors determining the cell fate in pre-B cell and, before the

differentiation, pre-B cell acute lymphoblastic leukemia can fuse to produce a chimeric oncogene (Aspland et al., 2001). Also, the upregulation of *Pbx3* will stabilize and induce the transcription of the *Meis1* gene, which increases the chance of leukemia in mice (Garcia-Cuellar et al., 2015). The *Pbx* genes are also found to be overexpressed in various solid tumors. In colorectal cancer (CRC), *Pbx3* expression is correlated with invasiveness in the lymph node. In addition, the overexpression of *Pbx3* in cells with a low metastatic potential promotes high migration and invasion potential. Consequently, the overexpression of *Pbx3* will enhance tumor proliferation and increase colony formation with enhanced invasive properties. Although *Pbx3* is reported as the most oncogenic gene of the *Pbx* genes, *Pbx1* is involved with shorter post chemotherapy survival and more resistance to platinum based drugs in ovarian cancer, affecting survival rate negatively (Morgan et al., 2014).

Meis Genes

The *Meis* genes are members of the TALE class of homeobox genes. The *Meis* genes were first identified in mouse cells where the leukemogenic virus integrated into the *Meis1* gene. By using the sequence identity of *Meis1* gene as a search tool, *Meis2* and *Meis3* were subsequently identified as well (Moskow et al., 1995). The *Meis* genes share the sequence similarities with other gene encoding TALE class proteins within the N-terminal and C-terminal domains of the protein. The members of the TALE class also exhibit extensive alternative mRNA splicing. The splicing events are well conserved among the TALE class genes and result in the synthesis of variant proteins with functional differences. For Meis1 variant proteins, there are two homologous variants

known as Meis1b/Meis1d. For Meis2, there are four full-length different splice variants known as Meis2a-d. The differences in variant proteins are found in the C-terminal region, which is important for recruiting additional transcription regulator in the complex. The low similarity among variations of C-terminal domains results in the recruitment of varied molecules, such as TFs, to the Meis proteins for the formation of different transcriptional complexes, which subsequently carry out varied functions on the promoter region of the DNA (Geerts et al., 2005). For example, while four full-length Meis2 variants are distributed throughout different mouse tissues during embryogenesis, the Meis2e variant lacks two thirds of the homeodomain and possesses an incomplete C-terminal domain, which prevents its contribution to transcriptional regulation and negatively regulates Meis2 functions (Yang et al., 2000). The consequence of compromised *Meis2* expression includes decreased stem cell proliferation/differentiation, such as neuronal stem cells and mesenchymal stem cells (Agoston et al., 2012).

The *Meis* genes act as potent regulators in cell proliferation and cell fate determination during development. In *Xenopus*, *Meis* gene expression in the neural tube enhances the cell proliferation of neural crest cells and disrupts cell differentiation to result in cancerous cell masses (Maeda et al., 2001). In chicken, the *Meis* genes are the first expressed throughout the developing limb bud but later become restricted to the proximal region by BMP signaling (Geerts et al., 2005). The expression of Hth, the Meis homolog in *Drosophila*, is restricted to the proximal limb region during limb development and repressed by the signals Wnt and TGFβ (Gonzalez-Crespo et al., 1998). The ectopic expression of both Meis and Hth can disrupt normal limb development and

cell differentiation. In human leukemia, *MEIS*1 can be found in bone marrow cells of AML patients where *HOXA9* is co-expressed. The *HOXA9* expression has found to cause the transformation of normal hematopoietic cells to AML cells when co-expressed with *MEIS*1 (Geerts et al., 2005).

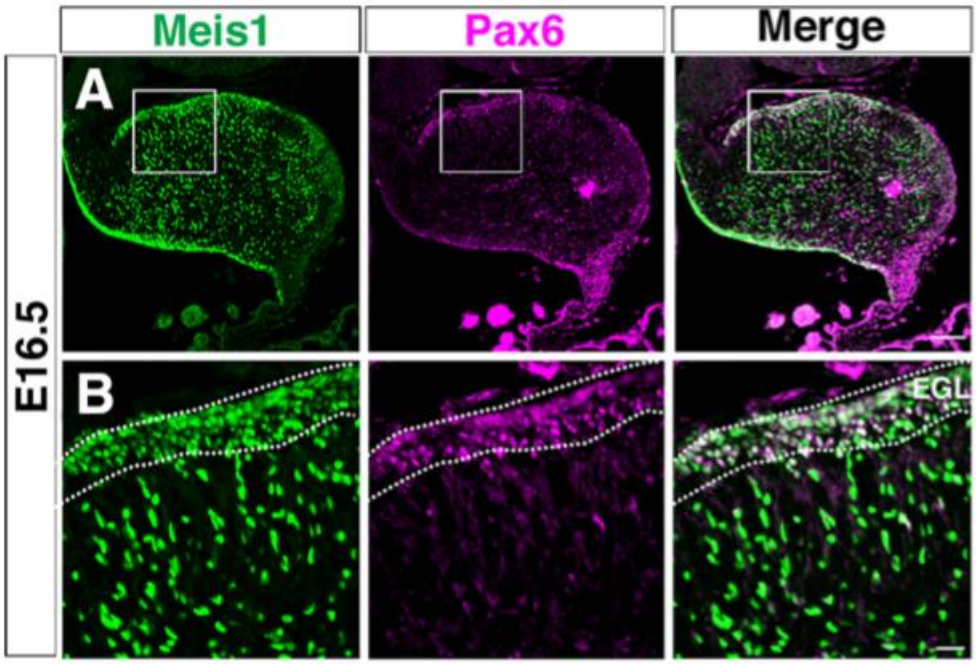

Fig. 8 Distribution of Meis1 and Pax6 expression identified during the cerebellum development. Both A and B images are indicating cerebellar sagittal section E16.5. EGL indicates external granule layer (Owa et al., 2018).

In the mouse, Meis1 plays crucial roles during the development of granule cells (GCs) in the cerebellum. Meis1 expression is found in GC lineage cells and astrocytes in the cerebellum (Fig. 8). Disruption in this expression results in a smaller cerebellum size with compromised morphogenesis of GCs. Meis1 function in cerebellum development is achieved via the formation of a dimeric complex with Pax6, which promotes Pax6 transcription in GCs and, subsequently, promotes the differentiation of GCs. In *Meis1* knockout mice, not only is the size of the cerebellum smaller, but also the expression of

Smad and BMP are decreased. BMP signaling is important to induce the GC precursors (GCPs) to mature and enhances the induction of transcription factor, Atoh1. Atoh1 is the key transcription factor that determines cell fate of GCs and proliferation of GCPs. Disrupted regulation of Atoh1 will result in the occurrence of medulloblastoma, which is a malignant brain tumor in the cerebellum (Owa et al., 2018).

During tectum development of chicken, *Meis2* expression is found in the mesencephalic vesicle and confined to the tectum anlagen, which are the very first stage of brain development. Also, the location of expression is found close to the diencephalon and mid-hindbrain boundary (MHB). Sufficient *Meis2* expression is required for tectum development, however decrease in *Meis2* will lead to abnormal development of tectum structures. During development of the MHB, the MHB organizer secretes signaling molecules, including Fgf8, to develop the mid- and hindbrain structures by regulating expression of the transcription factors Pax2/3/5/7/8 and En1/2, and the signaling molecules Wnt1/3a/10b. The unique role of *Meis2* during tectum development is to auto-regulate its own expression for stabilizing the tectum fate and, alternatively, induce a di- to mesencephalic fate change without the presence of signaling molecules secreted by MHB organizer. However, the mechanism to regulate *Meis2* expression in the mesencephalic alar plate, which is crucial to prevent the auto-activation of *Meis2* and subsequent overexpression is not yet known, (Agoston et al., 2012).

In zebrafish, meis3 transcription factors cooperate with pbx4 and hoxb1b in regulating hindbrain fates. The interaction between hoxb1b and pbx4 induces downregulated hoxb1a expression to compromise the development of rhombomere2 in

the hindbrain (Vlachakis et al., 2001). For other examples, meis3, hoxb1b and pbx4 interact to induce hoxb1a and hoxb2 expression, which results in the transformation of forebrain and midbrain fates to that of hindbrain (Vlachakis et al., 2001). These consequences are dependent on the recruitment of a meis3-hoxb1b complex to the interacting domains of pbx4. meis3, hoxb1b and pbx4 are co-expressed during early development of the caudal hindbrain in zebrafish embryos and form trimeric complexes to function as regulatory complexes. The trimeric complexes will regulate the differentiation of hindbrain fates, particularly in rhombomere4, during zebrafish embryogenesis (Vlachakis et al., 2001).

Putative regulatory elements: m2de1-m2de4

Over evolutionary time, genes can move to different locations in the genome. This suggests that genes that remain adjacent to one another in diverse species have this genomic organization being preserved by some sort of selective pressure. One mechanism that has been suggested to provide this selective pressure is the sharing of *cis*-regulatory elements (Irimia et al., 2012).

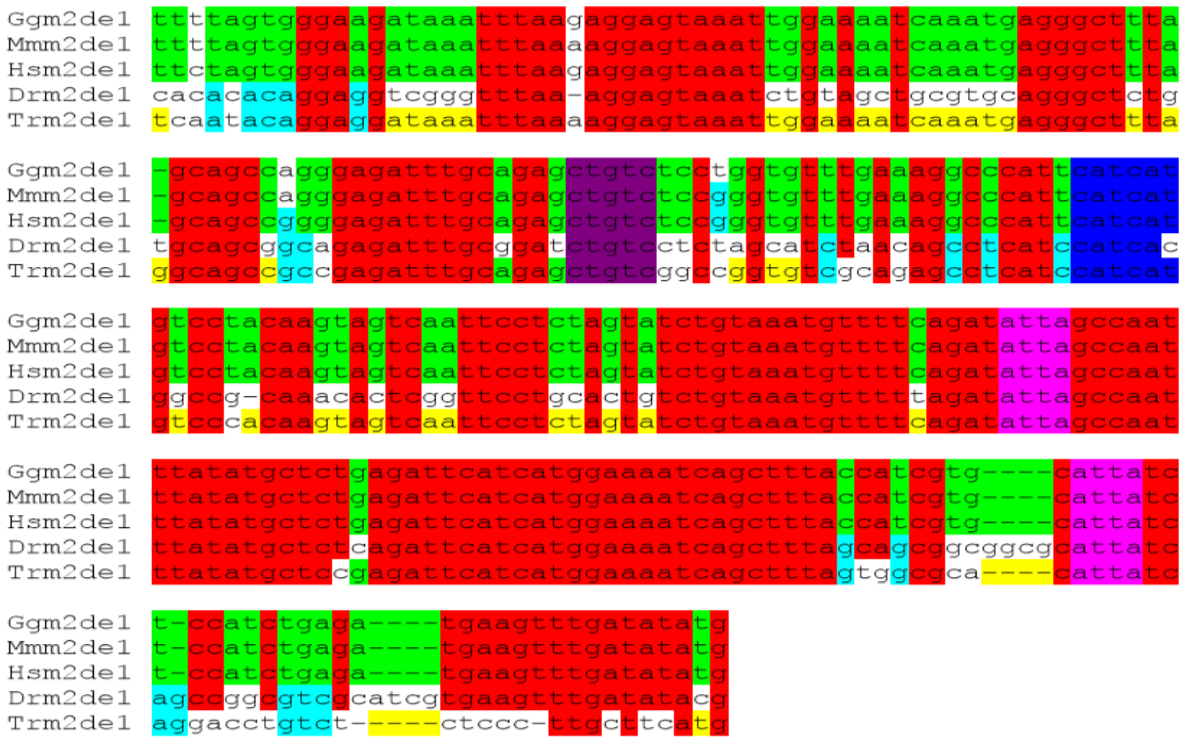

Fig. 9 Multiple sequence alignment of the m2de1 in five vertebrates. Gg represents chicken, Mm represents mouse, Hs represents human, Dr represents zebrafish, and Tr represents pufferfish. Red colored sequence represents conserved sequences among all vertebrates. Light blue colored sequences represent conservation between zebrafish and pufferfish. Green represents conservation among three land vertebrates. Yellow represents conservation between land vertebrates and pufferfish. Dark purple colored sequences represent Meis2 binding sites, pink sequences represent Hox binding site, and blue sequences represent Pbx binding site (Nelson, 2011).

The Zerucha laboratory has identified four highly conserved non-coding elements (HCNEs), which are located downstream of the *Meis2* gene, and hypothesized to function as *cis*-regulatory elements for regulating *Meis2* expression. The elements are named *Meis2* Downstream Elements 1-4 (m2de1-4), and possess multiple binding sites for transcription factors predicted to control *Meis* gene expression. To date, m2de1 has been identified in every vertebrate examined, however m2de2-4 have only been identified in land vertebrates. The binding sites in m2de1 include consensus sequences for Meis2, Hox and Pbx, suggesting this element may act in a cross-regulatory capacity for other genes and auto-regulatory capacity for *Meis2* gene (Fig. 9). In addition, both *meis2a* and m2de1

directed expression share similar spatial and temporal expression pattern, in midbrain and hindbrain of zebrafish embryos (Tennant, 2018).

Zgc: 154061

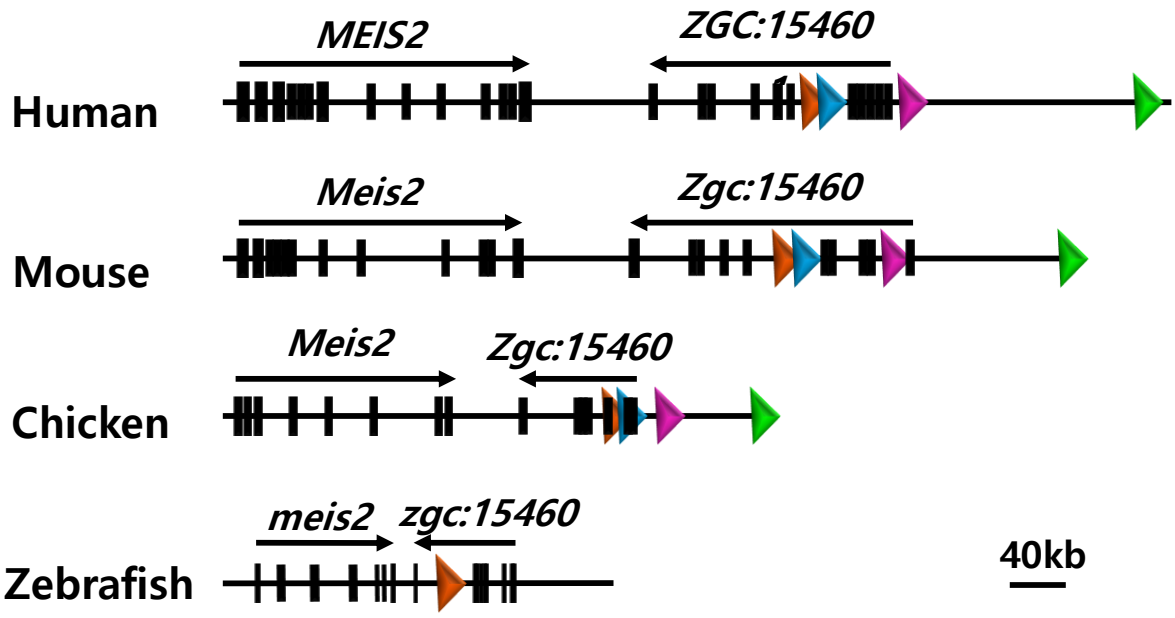

Fig. 10 Position and orientation *Meis2*, *Zgc:154061* and *M2de1-4* in human, mouse, chicken, and zebraish. Exons of each gene are indicated, labelled and arrows represent the direction of transcription. Orange represents m2de1, blue represents m2de2, purple represents m2de3, green represents m2de4 (Nelson, 2011).

The *zgc:154061* gene is a relatively newly discovered gene and a member of the PD-(D/E)XK superfamily of nuclease proteins (Babbs et al., 2013). The PD-(D/E)XK superfamily is known to synthesize nuclease proteins that function in repairing damaged DNA particularly in cells that function in erythropoiesis. In human, the ortholog of *zgc:154061* gene is *C15ORF41*, which is found on chromosome 15. A mutation in *C15ORF41* has been associated with congenital dyserthropoietic anemias (CDAs), a group of rare blood disorders that disturbs erythropoiesis in humans (Wickramasinghe and Wood, 2005). Among all vertebrate species with publicly available genome data, the

homologs of *zgc:154061* are consistently located adjacent to the *Meis2* gene sequence in an inverted convergently transcribed orientation downstream of *Meis2* (Fig. 10) (Carpenter et al., 2016).

In the zebrafish genome, there are two copies of the *Meis2* gene, which are known as *meis2a* and *meis2b* as a result of a proposed genome duplication in teleosts. The *zgc:154061* gene appears to only have one copy in the zebrafish genome and is found adjacent to *meis2a*. The expression of *zgc:154061* is observed during early developmental stages of erythropoiesis, retina formation, and olfactory placode formation in the brain in zebrafish (Babbs et al., 2013; Carpenter et al., 2016). The expression pattern of *zgc:154061* shows overlaps with the expression of *meis2a,* in the eye and brain, which may indicate the possibility of these two genes sharing regulatory elements between the two genes (Carpenter et al., 2016). Interestingly, in vertebrates, m2de1-2 are found in the introns of *zgc: 154061* homologs and m2de3-4 are found adjacent to *zgc: 154061* homologs. This consistent organization suggests the possibility of the sharing of regulatory elements, in particular m2de1, between *Meis2* and *zgc: 154061* homologs (Fig. 10).

To test the ability of the m2de1-4 elements to direct gene expression during development, previous members of the Zerucha lab made use of the Tol2 transposon system, an effective gene transfer tool to generate transgenic by using an insertional technique (Gaiano et al., 1996). The Tol2 transposable element was originally found in the genome of the medaka fish, *Oryzias latipes,* which is a small freshwater teleost native to Japan (Koga et al., 1996). The sequence of Tol2 is similar to the transposons that

belong to the hAT (hobo Activator Tam3) family. While many DNA transposons are present in the vertebrate genome, most are thought to be naturally inactive or non-autonomously active. The non-autonomous DNA transposons can be reinserted into the genome but cannot synthesize transposase proteins. Subsequently, without the presence of transposase proteins, the non-autonomous DNA transposons cannot be reinserted into the genome. On the other hand, the Tol2 element is the only exception found to be autonomously active. However, in zebrafish, Tol2 elements are not autonomously active, which active transposon elements are co-injected with the functional transposase protein to identify, excise, and reinsert the DNA element to generate transgenes (Kawakami et al., 1998; Koga et al., 1996; Ni et al., 2016).

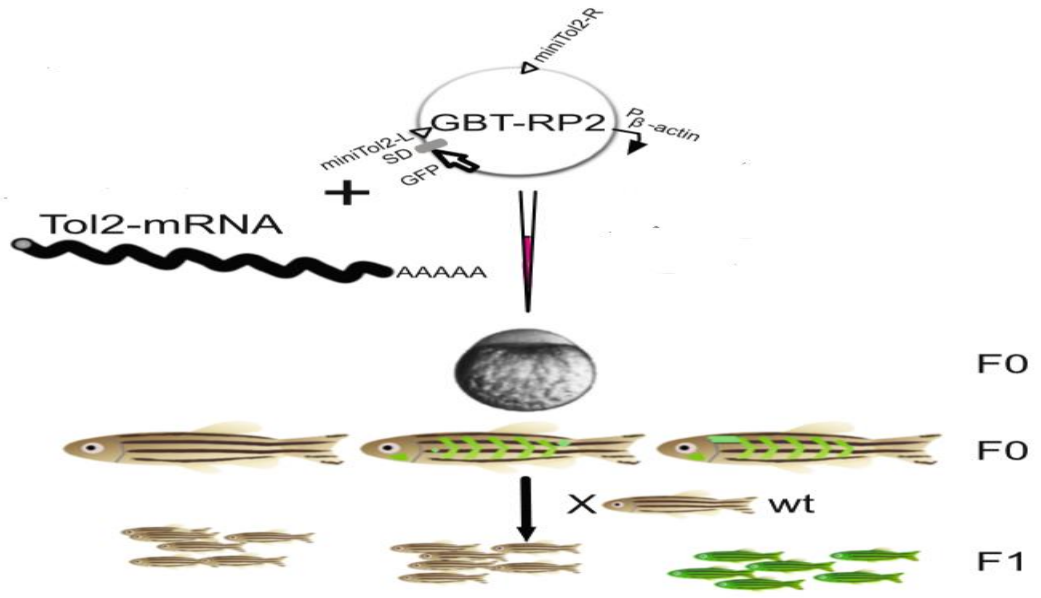

Fig. 11 Diagram of microinjection of zebrafish fertilized eggs at the single cell stage to generate a transgenic line. The constructs contain Tol2 transposon element and dominant marker inserted GBT-RP2 plasmid, and Tol2 mRNA. The crossing of F_0 generation fish and wild type zebrafish produces a one-third chance for transgenic zebrafish (Ni et al., 2016).

Previously in the Zerucha lab, constructs were generated where a single m2de sequence was placed adjacent to a minimal promoter and the eGFP gene. Microinjections were performed into single cell stage zebrafish embryos with these constructs mixed with a transposase-encoding mRNA, which when translated inserted the m2de-eGFP expressed cassette into the genome of fertilized zebrafish embryos. Each of the elements were determined to be able to direct expression during development. In particular, m2de1 was shown to direct expression to the developing brain and muscle fibers in zebrafish embryo. The expression of eGFP directed by m2de1 has some overlaps with both *meis2a* and *zgc:154061*.

A previous student in the Zerucha lab performed some preliminary experiments to investigate mutating m2de1 to further characterize the role of m2de1 in possibly controlling *meis2a* and/or *zgc:154061* expression (Tennant, 2018). This was done using the Clustered Regularly Interspaced Short Palindromic Repeat/CRISP associated protein 9 (CRISPR/Cas9) system. The CRISPR/Cas9 system was originally derived from bacteria and archaea, and in these organismsfunctions as an innate immune system to inactivate foreign nucleic acids (Deveau et al., 2010). The CRISPR system utilizes the Cas9 protein to target any gene of interest by a customized short guide RNA (sgRNA), which is a complementary sequence to the target DNA (Mali et al., 2013). As designed sgRNA facilitates the base-pairing with target sequence in dsDNA (double strand DNA), co-injected Cas9 protein facilitates the cleavage of the target sequence along with its dsDNA (Jinek et al., 2012).

After the cleavage of the target dsDNA, two cellular mechanisms to repair damaged DNA are activated. First repairing method is non-homologous end joining (NHEJ), leading to the insertion or deletion of a small number of nucleotides at the break site. Second repairing method is homology-directed repair (HDR), leading specific base-pair changes by introducing a donor template to the break site. Precise cellular DNA repairing is crucial for gene manipulation, which prevents unwanted byproducts at target sequence with insertions and deletions of nucleotide sequences (Symington and Gautier, 2011).

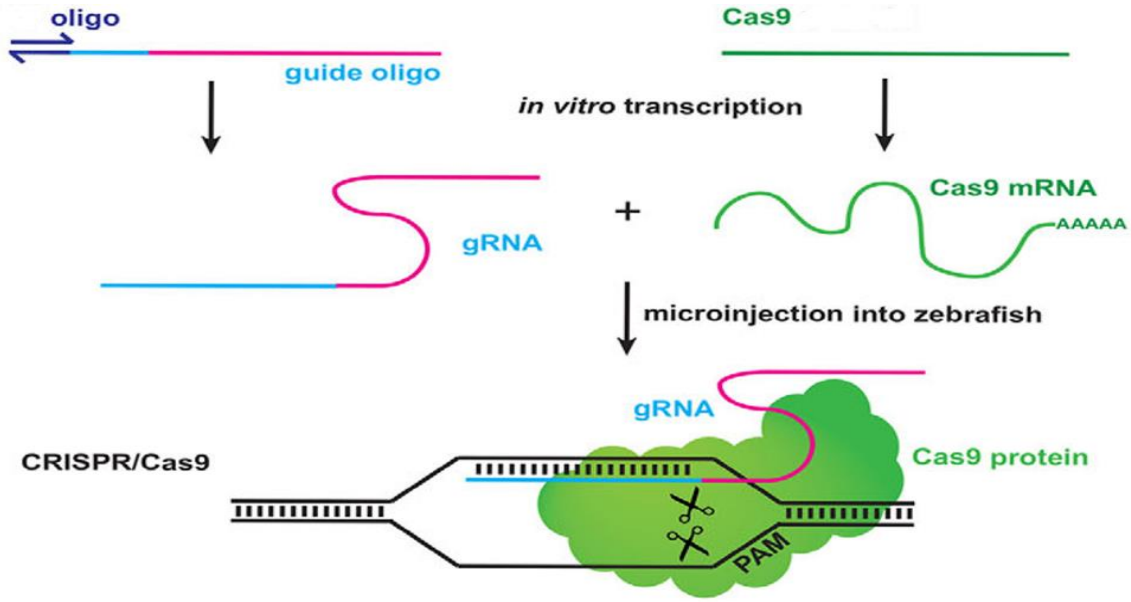

Fig. 12 Overview of the CRISPR/Cas9 system. Before the microinjection to be performed, both guide oligonucleotides and Cas9 are transcribed in vitro as gRNA and Cas9 mRNA. After the microinjection, the gRNA will mutate a target gene by forming the base pair with a target gene and recruiting Cas9 protein to break a double-strand (Hruscha et al., 2013).

The advantages of the CRISPR/Cas9 are high site specificity, flexible design, and ease of operation. CRISPR/Cas9 can be applied to a variety of applications other than gene knockout. These include functional genome editing, such as specific gene knockout,

transcriptional perturbations, the labelling of specific gene in living cells, and epigenetic modulation with collaboration with specific functional effector domains (Hsu et al., 2013). In this work, I describe the excision of a region of m2de1 from the zebrafish genome and the consequence of this mutation on the expression of *meis2a* and *zgc:154061*. The role of m2de1 in the regulation of *meis2a* and *zgc:154061* expression may be observed within compromised function of m2de1.

Methods and Materials

Zebrafish Husbandry

Zebrafish husbandry in this project was performed according to *The Zebrafish Book: A guide for the laboratory use of zebrafish (Danio rerio)* (Westerfield, 2000) in compliance with the Appalachian State University Institutional Animal Care and Use Committee (IACUC). Zebrafish were raised and maintained in 1L tanks within a Marine Biotech Z-mod closed system (Aquatic Habitats, Apopka, FL) with each tank containing 7 adult zebrafish at most. Water quality of the system was maintained with the following parameters: water temperature maintained between 26-28° C, water pH level between 6.8 and 7.2, and water conductivity level between 450 and 600 microSiemens. The monitoring of these parameters was performed daily to prevent dramatic changes in the water quality. Also, the water hardness was measured monthly and maintained between 120-220ppm. Lastly, the maintenance of system for a systematic light cycle of 14 hours of light and 10 hours of dark was maintained. For the feeding of zebrafish, adult fish were fed dry food every morning at 9 am, and fed 2 days old live brine shrimp at 3 pm.

Zebrafish Breeding

To obtain fertilized zebrafish embryos for the project, male and female fish were placed in a special breeding tank (Aquatic Habitats), which kept 2 male and 3 female separated by a divider. The breeding tanks with the divider in place were kept overnight in the system and the divider was removed at 9 am when the system light turned on. After the divider was removed, the female fish began to lay eggs and, simultaneously, the male fish fertilized these eggs. The embryos fell through the mesh bottom of the breeding tank and were collected between 15-30 minutes post fertilization. The embryos were collected by a fine mesh net and rinsed with RO water to clean embryos from debris such as fecal matter. In addition, any unfertilized embryos were removed via a Pasteur pipet.

The cleaned embryos were placed into a petri dish containing 1.0x E3 medium solution (0.29g NaCl, 0.01g KCl, 0.048g $CaCl_2$ $2H_2O$, 0.082g $MgSO_4$ $7H_2O$ per 1L of RO water). For those embryos to be raised to adulthood, they were placed in a 28℃ incubator until 5 days post fertilization (dpf). After 5 dpf, larval fish were transferred to a bowl with 200mL of 1.0x E3 medium at 28℃. Larval fish were fed Golden Pearl Reef & Larval fish 50-100 (Active Spheres) 3 times a day (9 am, 12 pm, 3 pm), and cleaned daily by replacing 1.0x E3 medium. If necessary, the bowl was switched to prevent the buildup of food waste. After 2 weeks post fertilization, the larval fish were transferred into a standard 1L tank, which was filled with 50% 1.0x E3 medium solution and 50 % system water, on a slow dripping system water. Based on the size of larval fish, the food size and number of dry food feedings changed. Golden Pearl Reef 50-100 (Active Spheres) food was fed three times a day from 5 dpf to 2 weeks post fertilization and Golden Pearl Reef

100-200 (Active Spheres) food was fed three times a day from 2 weeks to 3 weeks post fertilization. Next, dry food 200 (Zeigler) was given at 9 am between 3 to 6 weeks post fertilization, and, at 3 pm, a drop of concentrated 2 days old live brine shrimp per fish were fed. Until 3 months fertilization, dry food 300 and 400 (Zeigler) were provided at 9 am and brine shrimp provided at 3 pm. After approximately 3 months, zebrafish were grown to the adulthood and fed standard adult food (Zeigler Adult Zebrafish Complete Diet) at 9 am and brine shrimp at 3 pm.

Fixing Embryos

To perform in situ hybridization (ISH), zebrafish embryos were fixed at specific hours post fertilization (hpf). Embryos were dechorionated and transferred to Eppendorf tubes after euthanizing embryos via 0.4% Tricaine (in Danieau buffer: 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). Excess E3 medium was removed and replaced with 750µL of 4% paraformaldehyde (PFA). Then, Eppendorf tubes were placed on a shaker table (Labrat Gyrotwister) at a speed between 45-50 RPM for 10 minutes at room temperature. After 10 minutes, the 4% PFA was removed and replaced with fresh 750µL of 4% PFA. Eppendorf tubes stayed on a shaker for 4 hours at room temperature or overnight at 4℃. Next, embryos in Eppendorf tubes were transferred to a petri dish for a brief wash with 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 were then transferred into a new petri dish for depigmentation with 3% H_2O_2 solution (1 mL 30% H_2O_2, 0.05g KOH, raised up to 10mL with DI H_2O) until the pigment disappeared from the head/eyes of embryos (approximately 30-45 min). After

depigmentation, embryos were transferred into a clean petri dish with 1x PBS to wash off the bubbles. After the wash, embryos were transferred into fresh Eppendorf tubes, remaining PBS was then removed, and replaced with 800µL of 50% MeOH/ 50% PBS. Eppendorf tubes were placed on a shaker for 5 minutes and replaced with 500µL of 100% MeOH, which was repeated 3 times. Lastly, Eppendorf tubes filled with fixed embryos were stored at -20℃.

In situ Hybridization

The generation of anti-sense and sense *meis2a* and *zgc:154061* RNA probes was accomplished according to the previous work from the Zerucha lab (Tennant, 2018). Typically, In situ hybridization (ISH) would take three days to complete. Day 1 started with preheating both 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) and 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) with probe at the concentration of 1:100 at 65℃ for approximately 1 hour. The fixed embryos were washed with 50% MeOH/ 50% PBS resting for 5 minutes at room temperature. This was followed by two PBT (0.1% Tween-20 in 1x PBS) wash for 5 minutes on the shaker. After 2 wash steps, 500µL of the preheated prehybridization buffer was added to the embryos and incubated in the water bath at 65℃ for 2 hours. After incubation, the buffer was removed and 200µL of the hybridization buffer with probe at the concentration of 1:100 was added into Eppendorf tubes. Eppendorf tubes then stayed in the water bath at 65℃ for overnight.

Day 2 started with preheating of prehybridization buffer, 50% prehybridization buffer/ 50% 2x SSC, 2x SSC (17.53g of NaCl, 8.82g of sodium citrate in 1L of DI H_2O, pH 7.0) and 0.2x SSC at 65℃ water bath for approximately 1 hour. After the preheating, the hybridization buffer with probe in Eppendorf tubes was removed, saved, and stored at -20℃. The four wash steps started with 500µL of prehybridization buffer at 65℃ for 60 seconds, second incubation with 500µL of 50% prehybridization buffer/ 50% 2x SSC at 65℃ for 45 minutes, third incubation with 500µL of 2x SSC at 65℃ for 15 minutes, and final incubation was with 500uL of 0.2x SSC at 65℃ for 1 hour, respectively. After 4 steps, the contents of the tubes were replaced with 1mL of PBT (0.1% Tween-20 in 1x PBS) and placed on a shaker at room temperature for 5 minutes, which repeated twice so done 3 times all together. After the last PBT wash, 500µL of blocking solution (2% goat serum, 2 mg/mL BSA in PBT) was added the tube and placed on a shaker at room temperature for 1 hour. After 1 hour, the solution was removed and replaced with 200µL of antibody solution (1µL of Anti-Digoxigenin-AP-Fab fragments (Roche) per 1mL of blocking solution). Day 2 ended with leaving Eppendorf tubes on a shaker table at 4℃ for overnight.

The last day, day 3, started with removing antibody solution and quickly rinsing the embryos with 500µL of PBT. After removing the PBT solution, embryos were washed 6 times wash with 500µL of PBT followed by agitation on a shaker at room temperature for 15 minutes each. Then, embryos were washed 3 times with 500µL of alkaline phosphatase buffer (500 µL of 1M Tris 9.5, 250 µL of 1M $MgCl_2$, 100 µL of 5M NaCl, 25µL 20% Tween-20, 4.125mL of DI H_20) on a shaker at room temperature for 5

minutes. After the last wash, 500µL of alkaline phosphatase buffer with color substrate (3.5 µL BCIP and 4.5 µL NBT) were added to the embryos in the Eppendorf tubes. Eppendorf tubes were kept in a closed drawer without light and sealed with aluminum foil. Periodically, the color development of embryos was checked. After observing adequate color development, embryos were washed 3 times with 500µL of PBT with shaking at room temperature for 5 minutes to stop the color development. After the last wash, 500µL of 100% MeOH was added and the embryos placed on a shaker at room temperature for 5 minutes. Finally, solution on the embryos was replaced with fresh 500µL of 100% MeOH and the embryos were stored at -20℃.

Screening in situ hybridization

The screening of embryos that had gone through ISH with anti-sense probes to *meis2a* or *zgc:154061* was performed to observe the displayed expression pattern. Stained embryos were washed with 50% MeOH/ 50% PBS on a shaker at room temperature for 5 minutes and washes were repeated six times with PBT at room temperature for 5 minutes were performed. Embryos were then suspended in 25% glycerol/ 75% PBS and 50% glycerol/ 50% PBS on a shaker table at room temperature for 10 minutes each, respectively. Lastly, the suspension with 80% glycerol/ 20% PBS on shaker at room temperature for 3 hours was performed. To screen, the embryos were placed on a 1.0 mm thick microslide (VWR 48312-004) and covered with a 1-ounce 22x22 mm microscope cover glass (VWR 16004-094). An Olympus IX81 inverted microscope was used for screening embryos and images were taken through Olympus cellSens Software.

Bacterial Transformation: Transformation of Plasma DNA

In this project, the generation of transgenic zebrafish line with GFP expression directed by m2de1 was performed. To obtain sufficient plasmid DNA vector carrying dr.m2de1-pGW-cfos-EGFP, this plasmid was transformed into DH5α *E. Coli*. 50ng/2µL of plasmid and 50µL of chemically competent DH5α *E. coli*, were mixed solution in an Eppendorf tube and placed on ice for 30 minutes, followed by a heat shock in a 42℃ water bath for 45 seconds. After the heat shock, the solution was rested for 5 minutes at room temperature and then 1mL of SOC medium (20g bacto-tryptone, 5g bacto-yeast extract, 0.5g NaCl, 20mM glucose) was added. The mixed solution was incubated at 37℃ for 1 hour and placed on LB agar plates+AMP antibiotic. The plates were incubated at 37℃ for overnight and then stored at 4℃.

Minipreps: Preparation of Plasmid DNA

To generate the plasmid DNA construct for Tol2 microinjection, two glass tubes were prepared containing 2mL of liquid LB, 2µL of 100mg/ml in RO water AMP antibiotic and a single colony of cultivated bacterial cells from the bacterial transformation. The glass tubes were placed in a shaker at 37℃ for overnight at 200 RPM. On the following day, the sample was transferred to fresh Eppendorf tubes until the tubes were almost full. Then, Eppendorf tubes were centrifuged at 13.2 RPM for 5 minutes and the liquid was decanted off. Eppendorf tubes were refilled with the remaining solution from the overnight culture and centrifuged again for 5 minutes, followed by decanting off the liquid. Next, the Wizard® *Plus* Midipreps DNA Purification System (Promega A7510) was utilized by adding 250µL of cell resuspension

solution, 250µL of cell lysis solution, 10µL of alkaline protease solution and 350µL of neutralization solution respectively, and mixed well by inverting Eppendorf tubes multiple times gently. After mixing, the sample was placed on ice for 5 minutes. After 5 minutes, Eppendorf tubes were centrifuged at 13.2 RPM for 5 minutes. Only liquid from Eppendorf tubes were transferred into fresh Eppendorf tubes for phenol/chloroform extraction.

An equal amount of Phenol/Chloroform/Isoamyl alcohol (25:24:1) was added to the Eppendorf tubes. The Eppendorf tubes were then mixed by gentle flicking of the tube before centrifuging for 10 minutes at 13.2 RPM. After centrifugation, the upper layer, which contained the DNA, was transferred into a fresh Eppendorf tube and 100µL of Chloroform/Isoamyl alcohol was added. Again, the sample went through centrifugation for 10 minutes. The upper layer of samples was again transferred into a fresh Eppendorf tube, and 2.5 volumes of 100% ethanol with total volume of 0.2M sodium chloride were added. The sample was stored at -20℃ for 2 hours. After 2 hours, the sample was centrifuged at 4℃ for 30 minutes. Then, the solutions were removed completely by decanting off and remaining liquid was removed via micropipette. Next, 1mL of 70% ethanol was added to the Eppendorf tube, mixed with gentle flicking, and centrifuged at room temperature for 5 minutes. After the centrifugation, the ethanol was removed as described previously. The Eppendorf tube was left for 15-20 minutes to allow the evaporation of the remaining ethanol in the chemical hood. After drying, the samples were resuspended with 10µL of nuclease free water.

Transcribing Transposase mRNA

Transposase mRNA was transcribed using the mMessage mMachine Sp6 kit (Ambion, Life Technologies, Grand Island, NY) according to the manufacturer's protocol. After the transcription, mRNA was cleaned and concentrated using the RNA Clean & Concentrator™-5 kit (Zymo Research Corporation) according to the manufacturer's protocol. The resuspension of mRNA was performed with nuclease free water and used for the subsequent microinjection with the dr.m2de1-pGW-cfos-EGFP construct.

Microinjection of Tol2 Expression Constructs

Breeding steps to acquire fertilized zebrafish eggs were followed as previously described. Fertilized zebrafish embryos at the single cell stage were collected and placed on a petri dish with E3 medium. Embryos were washed with RO water and cleaned through a Pasteur pipet. Needles were prepared by pulling a 3.5nl capillary previously heated at 260℃ for preventing RNase activity on needles on a David Kopf Instruments Vertical Pipette Puller model 700C. The Pipette Puller was set at 54 volts, and the solenoid was set at 10 amps. Each top and bottom part of machine pulled a capillary tube into two needles with sharp ends. The tip of the needles was precisely removed with watchmaker forceps. The needles were filled with mineral oil and then placed on the Nanoliter 2000 Microinjector (World Precision Instruments Model B203XVY), which was supported by a Marhauser MMJR Micromanipulator (World Precision Instruments). By adding pressure to the needle, approximately 20% of the mineral oil was removed to prevent the clogging of needle tip.

After preparation of needle, the construct was prepared by mixing 180ng of transposase mRNA, 135ng of plasmid DNA and nuclease free water to a total volume to 3μL. The collected single-cell stage zebrafish embryos were placed adjacent to a 1.0-millimeter-thick VWR micro slide (VWR international 48300-025) in a line, which was stabilized via taping a cover of a plastic petri dish. The injection was aimed directly into the single cell of the zebrafish embryos to deliver approximately 1nl per embryo. After microinjection, the injected embryos were placed into a petri dish with fresh E3 medium and allowed to develop as described previously.

Screening Tol2 Transposase

The screening of embryos that were injected with a construct composed of dr.m2de1-pGW-cfos-EGFP plasmid DNA and transposase mRNA was performed to observe the expression pattern of m2de1 by locating fluorescent expression. After raising those injected embryos until adulthood, outcrossing between male Tol2 transgenic fish and female wild type fish was performed. As embryos were laid and collected, those embryos were raised until they reached to 48 hpf. Then, embryos were anesthetized with 0.4% Tricaine to prevent their movement while screening. The preparation to screen embryos was followed as previously described in screening ISH. The Zeiss LSM 880 Confocal Microscope, using the Argon laser with wavelength of 488, was used to collect images of embryos with fluorescent expression. The quality of images was constant at 1024x1024, the scanning speed was ranged between 5 to 9 for better resolution, and the green levels were regulated to reduce autofluorescence.

Table 1. Designed two gRNAs

Oligo Name	**5'-Sequence-3'**
Dr-m2de1-ia	AATTAATACGACTCACTATA**GTTTGCGGCCGTGATGGATG** GTTTTAGAGCTAGAAATAGC
Dr-m2de1-ib	AATTAATACGACTCACTATA**CTGCACAGAGCCCTGCACGC** GTTTTAGAGCTAGAAATAGC
Excised sequence between two gRNAs	AGGCTGTTAGATGCTAGAGGACAGATCCGCAAAT CTCTGCCG

Designed gRNAs = consisting of 3 different sequences, including T7 promoter sequence, **target sequence**, and overlapping sequence with scaffolding oligo. First 3 nucleotide sequences provide the stabilization at the 5' end of the oligo. The scaffolding oligo anneals with the short guide oligo to form a complete template for gRNA transcription. The sequence between two gRNAs will be excised when two gRNAs are co-injected with Cas9 protein.

Transcribing guide RNAs for CRISPR/Cas9 Microinjection

The generation of two gRNAs targeting a portion of the m2de1 sequence was followed according to previous work from the Zerucha lab (Tennant, 2018). A specific oligonucleotide for gRNA synthesis containing the T7 promotor sequence (5'-TAATACGACTCACTATA-3'), the short guide oligo (targeting sequence), and a complementary sequence that anneals to a guide constant (scaffolding) oligo was transcribed (Table. 1). The samples for running polymerase chain reaction (PCR) were prepared for a total volume of 20µL, consisting of 4µL of 5x Phusion buffer, 0.4µL of dNTPs, 3µL of 10µM short guide oligo, 3 µL of 10µM scaffolding oligo, 0.2µL of Phusion and 9.4µL of nuclease free water. The samples were placed in the thermocycler to perform PCR with the following settings: initial denaturation of 98℃ for 30 seconds, 45 cycles of normal denaturation of 98℃ for 10 seconds, annealing at 60℃ for 10

seconds and extension at 72℃ for 15 seconds, and a final extension at 72℃ for 5 minutes. Once the PCR reaction was completed, the sample was kept at 4℃ until its removal from the thermocycler. Then, the sample volume was increased to 100µL with nuclease free water before purification with phenol-chloroform extraction and precipitation with ethanol. After spinning down the DNA as described previously the sample was rehydrated with nuclease free water at room temperature, and the DNA was used as a template to transcribe gRNAs.

Transcription of the template short guide oligo was performed using MAXIscript T7 Transcription Kit™ (Ambion, Life Technologies, Grand Island, NY) essentially following the manufacturer's protocol. However, two changes were made, including the addition of 1µL of RNase inhibitor (Promega N2515) to the sample and the incubation at 37℃ for overnight instead of 1 hour as the manufacturer described. The addition of RNase inhibitor prevented the degradation of the transcribed gRNAs and the overnight incubation ensured transcription of the majority of the template DNA into gRNA. On the following day, 1µL of turbo DNase was added to the sample and incubated at 37℃ for 15 minutes. After 15 minutes, the RNA Clean & Concentrator™-5 kit (Zymo Research Corporation) was utilized according to the manufacturer's protocol. After the RNA clean step, gel electrophoresis was run. To confirm RNA size, a 2% agarose gel was prepared in 1x TBE. The gel preparation steps were repeated as previously described for 1% gel electrophoresis. The RNA ladder as a size marker was prepared consisting of 2µL of ssRNA ladder, 3µL of loading dye and 5µL of nuclease free water. For the second lane, the sample was prepared consisting of 5µL of PCR sample, 3µL of purple loading dye

and 2μL of nuclease free water. After 30 minutes ran at 115 volts, the gel was observed for appropriate nucleotide sequence length. After the observation of the gel, the sample was stored at -20℃ for subsequent CRISPR/Cas9 microinjection.

Microinjection of the CRISPR/Cas9 system

Breeding and collecting steps of single-cell stage zebrafish embryos, and needle preparation steps were followed as previously described during Tol2 microinjection. Before the removal of the divider in the breeding tanks, the construct solutions were thawed. After the solutions thawed, 40ng/μl of each gRNA, 250ng/μl of EnGen® Cas9 NLS (NEB M0646T), and nuclease free water were combined to a total volume of 3μL. Then, the construct was incubated at 37℃ water bath for 10 minutes. After the incubation, 1nl of construct was injected into single cell embryos.

Screening CRISPR

After the microinjection of CRISPR/Cas9, the injected embryos were raised to adulthood. Then, potential transgenic males and females were crossed to produce transgenic embryos. The screening of injected embryos was not performed for generating the stable transgenic zebrafish line via CRISPR/Cas9 system. As subsequent generation of the embryos reached 48 hpf, the fixation of embryos was performed and followed by ISH as described previously via anti-sense and sense probes of *meis2a* and *zgc:154061*. After the 3 days process of ISH, the screening of embryos was followed according previously described.

DNA Isolation by Zebrafish Fin Clipping

DNA isolation of CRISPR/Cas9 injected zebrafish embryos was performed to determine partial knockout of the m2de1 sequence. A small amount of the caudal fin was clipped from transgenic zebrafish and utilized for target DNA extraction. Before the fin clipping, zebrafish were placed in ice water for anesthetization, preventing the movement of zebrafish. Sterile tweezers and scissors were used to grab and clip small amount of the caudal fin. The clipped fin was placed into an Eppendorf tube containing 50µl of activated genomic extraction buffer (50µL of 1M tris pH 8.0, 100µL of 0.5M EDTA pH 8.0, 1mL of 1M NaCl, 250µL of 10% SDS, 1mg of Proteinase K and raise the stock until 5mL with sterile RO). The fin clipped zebrafish were quickly placed in the recovery tanks individually and checked for their activity in the Marine BioTech ZMod. Zebrafish were isolated in the recovery tanks for 2 weeks until the clipped fin regeneratd.

The clipped tissue in the genomic extraction buffer was incubated in a shaking incubator at 56℃ and 200 RPM for at least 3 hours or until complete dissolving of the tissue was confirmed. After the incubation, 100µL of 100% ethanol was added and then stored at -20℃ for overnight. On the following day, the Eppendorf tube was centrifuged for 10 minutes at 13,200 RPM and supernatant was decanted off. Next, 200µL of 70% ethanol was added for wash and briefly vortexed. The Eppendorf tube was again centrifuged for 5 minutes at 13,200 RPM. The supernatant was decanted off again and the Eppendorf tube was placed in the chemical hood for drying remaining ethanol for 15-20 minutes. After drying, 20µL of TE+RNase buffer (20µL of 1M Tris pH 8.0, 4µL of 0.5M EDTA pH 8.0, 20µL of 100µg/mL RNase and raise the stock until 2mL with sterile RO)

was added for resuspension. Then, the Eppendorf tube was incubated at 37℃ for 1 hour. After the incubation, the sample was cleaned via phenol chloroform extraction.

Table 2. m2de1 Primers

Primer Name	**5'-Sequence-3'**
Dr-m2de1-3	GCTCATTATAAGGCCGTGCATG
Dr-m2de1-5b	TATACCATGGAGGTCGGGTTTAAAGGAG

After the phenol chloroform extraction, 15μL of nuclease free water was added for the resuspension. The resulting DNA was amplified by performing PCR with the thermocyler. The sample was prepared for the total of 25μL containing 50ng of isolated DNA, 2.5μL of 10x Thermopol buffer, 0.5μL of dNTPs, 1μL of 10μM Dr-m2de1-3 primer (Table. 2), 1μL of 10μM Dr-m2de1-5b primer (Table. 2), 0.125μL of Taq Polymerase (NEB Labs), and raise the volume to 25μL with nuclease free water. The settings of thermocycler for the polymerase chain reaction (PCR) was performed: initial denaturation of 95℃ for 5 minutes, and followed by 30 cycles of normal denaturation of 95℃ for 30 seconds, annealing at 55℃ for 30 seconds and extension at 68℃ for 1 minute and 45 seconds. After 30 cycles were completed, the sample proceeded to a final extension at 68℃ for 10 minutes. Once the PCR reaction was completed, 1% gel electrophoresis was run to ensure the absence of degradation. Gel was prepared with 0.5g of agarose and 50mL of 1x TBE, microwaved for 1.5 minutes, cooled down at room temperature for 5 minutes, and then added 3.5μL of EtBr. As liquid gel was poured on gel

box for the solidification, the comb was set on for the wells. After the gel was solidified, 1x TBE was filled in gel box. For the first well, the ladder was prepared for consisting 1.5µL of 100bp ladder, 3.5µL of purple loading dye and 5µL of nuclease free water. For the second well, the sample was prepared for consisting 1.5µL of PCR sample, 3.5µL of purple loading dye and 5µL of nuclease free water. After 30 minutes with 115 volts currents, the gel was observed for appropriate nucleotide sequence length.

Microinjection using the Morpholino reagent

MO-Pax6b v1 (Table. 3) and PCO-RandomControl were purchased from Gene Tools, LLC. The MO-Pax6b v1 and PCO-RandomControl were brought to 2mM with distilled water. 2mM of MO concentration is approximately equivalent to 16ng/nL. The MO stock solutions were aliquoted and sealed tightly to avoid evaporation, and stored at room temperature. In the rare case of incomplete dissolution of the morpholino, solutions were autoclaved t. Before the microinjection, the aliquoted stock solution was diluted to 2.5ng/nl and at 65℃ for 10 minutes to denature.

Breeding and collecting steps of single cell stage zebrafish embryos, and needle preparation steps were followed as previously described. Morpholinos were injected into single cell of zebrafish embryos at a concentration of 0.6mM and inject approximately 2.5ng/nl was injected per embryo.

Screening Morphant

After the microinjection of mopholino reagents MO-Pax6b v1 and PCO-RandomControl, morphants were raised until they reached 48 hpf. Then, ISH was performed using anti-sense probes of *meis2a* and *zgc:154061* as described previously.

Results

Expression pattern of m2de1 in zebrafish

To observe the pattern of expression directed by m2de1, adult male and female pDr-m2de1-F-cfos-eGFP transgenic zebrafish were crossed. This transgenic line was generated by a previous graduate student in the Zerucha lab and features eGFP directed by the m2de1 element. Embryos were screened at 48 hpf when m2de1 is known to actively be directing eGFP. Embryos exhibited eGFP expression in the eye, forebrain, midbrain and hindbrain (Fig. 13). This pattern of expression is identical to that observed from previous work from the Zerucha lab (Ferrara, 2015; Tennant, 2018).

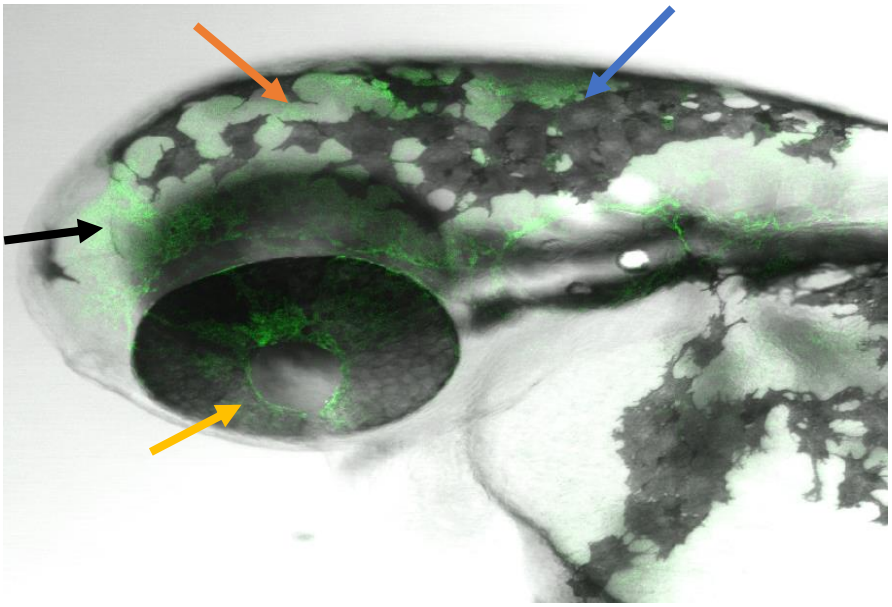

Fig. 13 The expression pattern of m2de1 directed with eGFP in zebrafish embryo at 48 hpf. Expression is observed in the forebrain (black arrow), midbrain (orange arrow), hindbrain (blue arrow) and eye (yellow arrow). Embryo is positioned to view anterior side on left and dorsal on top.

Expression pattern of meis2a and zgc:154061

With the knowledge of the expression pattern directed by m2de1, the expression patterns of *meis2a* and *zgc:154061* at the same stage of development were examined to determine overlapping regions of expression. In order to determine any overlapping expression pattern of two genes with directed by m2de1, in situ hybridization (ISH) was performed. Anti-sense probes of *meis2a* and *zgc:154061* were used on fixed wild-type zebrafish embryos at 48 hpf. The expression pattern of *meis2a* gene was found mainly in the anterior portion of an embryo, including the forebrain, midbrain and hindbrain. Similarly, *zgc:154061*was also found in the forebrain, midbrain, hindbrain and eye as well, however, signal expression was lower than *meis2a*. The expression patterns are consistent with the previously published studies (Carpenter et al., 2016). As the negative

control, ISH was performed using a sense probe of *meis2a* and *zgc:154061*. The result showed no expression, indicating the effectiveness of my ISH.

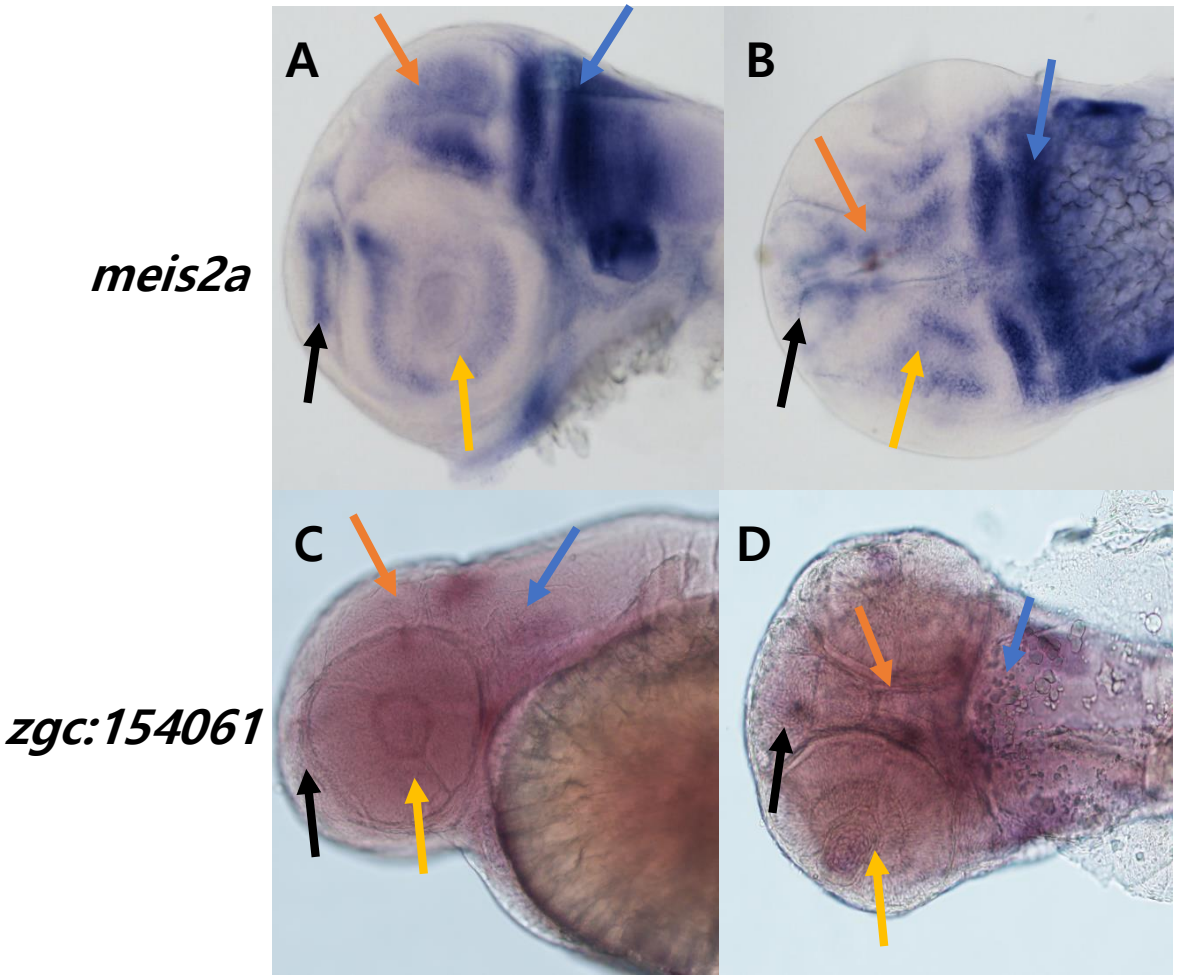

Fig. 14 Whole mount in in situ hybridization was performed on zebrafish embryos at 48 hpf to observe *meis2a* and *zgc:154601* expression pattern. Expression is observed in the forebrain (black arrow), midbrain (orange arrow), hindbrain (blue arrow) and eye (yellow arrow). (A, B) *meis2a* expression in embryos positioned in the (A) lateral orientation and the (B) dorsal orientation. (C, D) *zgc:154061* expression in embryos positioned in the (C) lateral orientation and the (D) dorsal orientation.

m2de1 mutation by CRISPR/Cas9

Because of the overlapping expression pattern of eGFP directed by m2de1 and that of *meis2a* and *zgc:154061*, the CRISPR/Cas9 system was employed to excise a partial sequence of m2de1. Initially, to determine the efficacy of this technique in my

hands, I used CRISPR/Cas9 to target the *Spadetail* gene to use as a positive control. Spadetail is a well characterized mutant that has also been demonstrated to be generated by CRISPR/Cas9. 379 single cell zebrafish embryos were injected with Spadetail gRNA as described in Burger, et al., (2016) and Tennant, (2018). Of these embryos injected, 105 (39%) exhibited the Spadetail phenotype at 24 hpf as shown in Table 3 and Fig 15. The frequency of Spadetail phenotype in my study is 39%, which is lower than the Burger, et al. 2016. One possible reason for this discrepancy is that the concentration of Spadetail gRNA I used is 10ng/ul lower than the concentration used by the Burger, et al. (2016).

Table 3. Concentration of Spadetail gRNA and Cas9 with total number of injected embryos and result of microinjection.

Lists	**Numbers**
Concentration of Spadetail gRNA (DK960)	**40ng/ul**
Concentration of Cas9	**250ng/ul**
Total number of injected embryos	**379**
Number of alive at ~48hpf	**269 (71%)**
Number of death at ~48hpf	**110 (29%)**
Number of embryos displaying WT phenotype	**164 (61%)**
Number of embryos displaying Spadetail phenotype	**105 (39%)**

Fig. 15 Phenotype of wild type (WT) embryos and Spadetail (ST) phenotype of CRISPR injected embryos within 24 hpf. The comparison between wild type and Spadetail injected embryos suggest the changes in compromised development at the posterior end of tail.

Because of the success of the Spadetail experiments I proceeded to attempted removal of the m2de1 element. Two gRNAs that recognize different regions of m2de1 and are approximately 82bp apart were injected into single-cell zebrafish embryos. Previous graduate student attempted to target whole m2de1 sequence, however, the mortality rate was approximately 100%, suggesting the importance of m2de1. Subsequently, the partial sequence of m2de1 was targeted, consisting putative binding sites for homeodomain containing TFs. The injected embryos were raised to adulthood. 830 embryos were injected and of these 598 (72%) survived to 48 hpf. Of these 104 survived to adulthood. To determine if any of these fish raised from injected embryos contained the m2de1 excision, they were screened by fin clip. The caudal fin of the CRISPR injected zebrafish was clipped partially and the genomic DNA isolated and used as template for PCR with m2de1 specific primers (Fig. 16). The expected amplicon size of wild type m2de1 using these primers is ~450bp and for m2de1 lacking the region excised by CRISPR/Cas9 ~360bp. Fig 16 clearly shows that the fish represented by lane

B is heterozygous for the mutated version of m2de1 as it exhibits a full length m2de1 sequence as well as the truncated one.

Table 4. Concentration of each gRNA and Cas9 with total number of injected embryos and result of microinjection.

Lists	**Numbers**
Concentration of gRNA m2de1 ia gRNA	**40ng/ul**
Concentration of gRNA m2de1 ib gRNA	**40ng/ul**
Concentration of Cas9	**250ng/ul**
Total number of injected embryos	**830**
Number dead at ~48 hpf	**232 (28%)**
Number alive at ~48hpf	**598 (72%)**

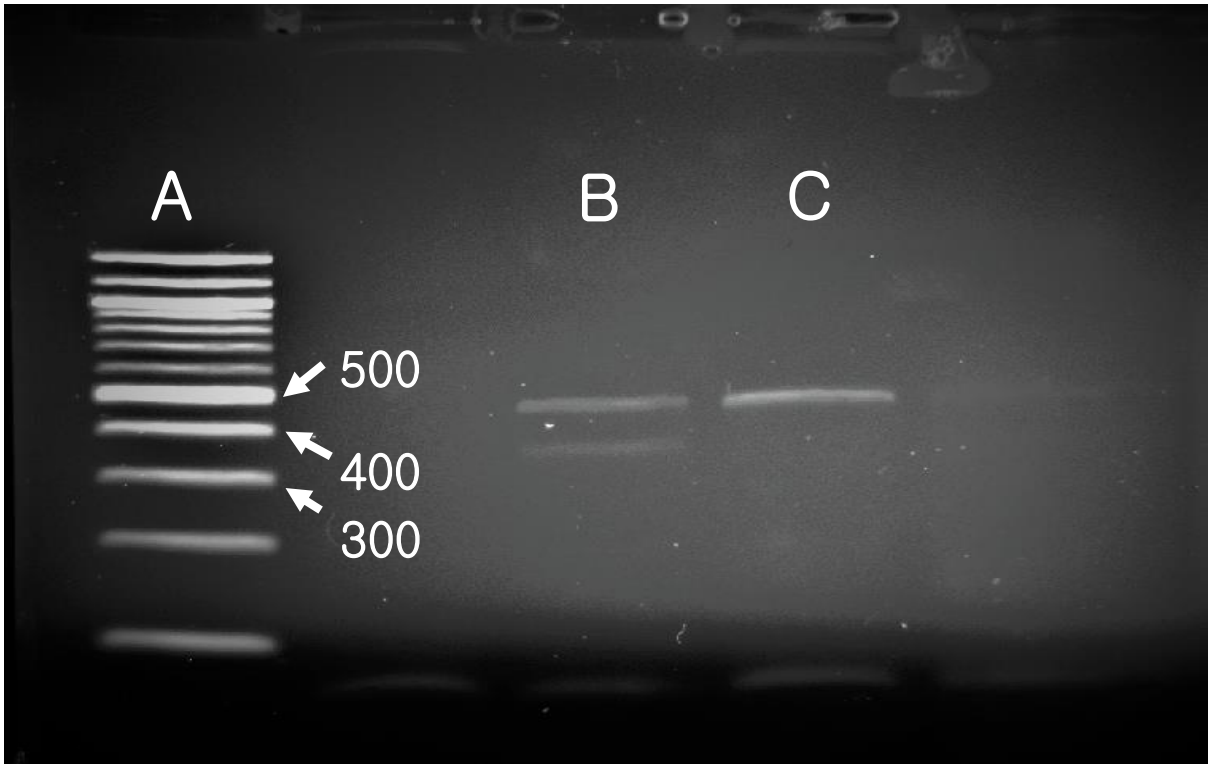

Fig. 16 1% agarose gel of isolated DNA demonstrating the partial deletion of the m2de1 sequence. Lane A of the gel shows a 100bp DNA ladder. Lane B shows a PCR reaction with DNA isolated from a potential m2de1 mutant zebrafish. Lane C shows a positive control PCR reaction with DNA isolated from wild type zebrafish. While both lanes B and C show ~450bp band indicating the whole m2de1 sequence, lane B also shows an additional band at ~360bp. In lane A, the 300bp, 400bp and 500bp bands are labeled.

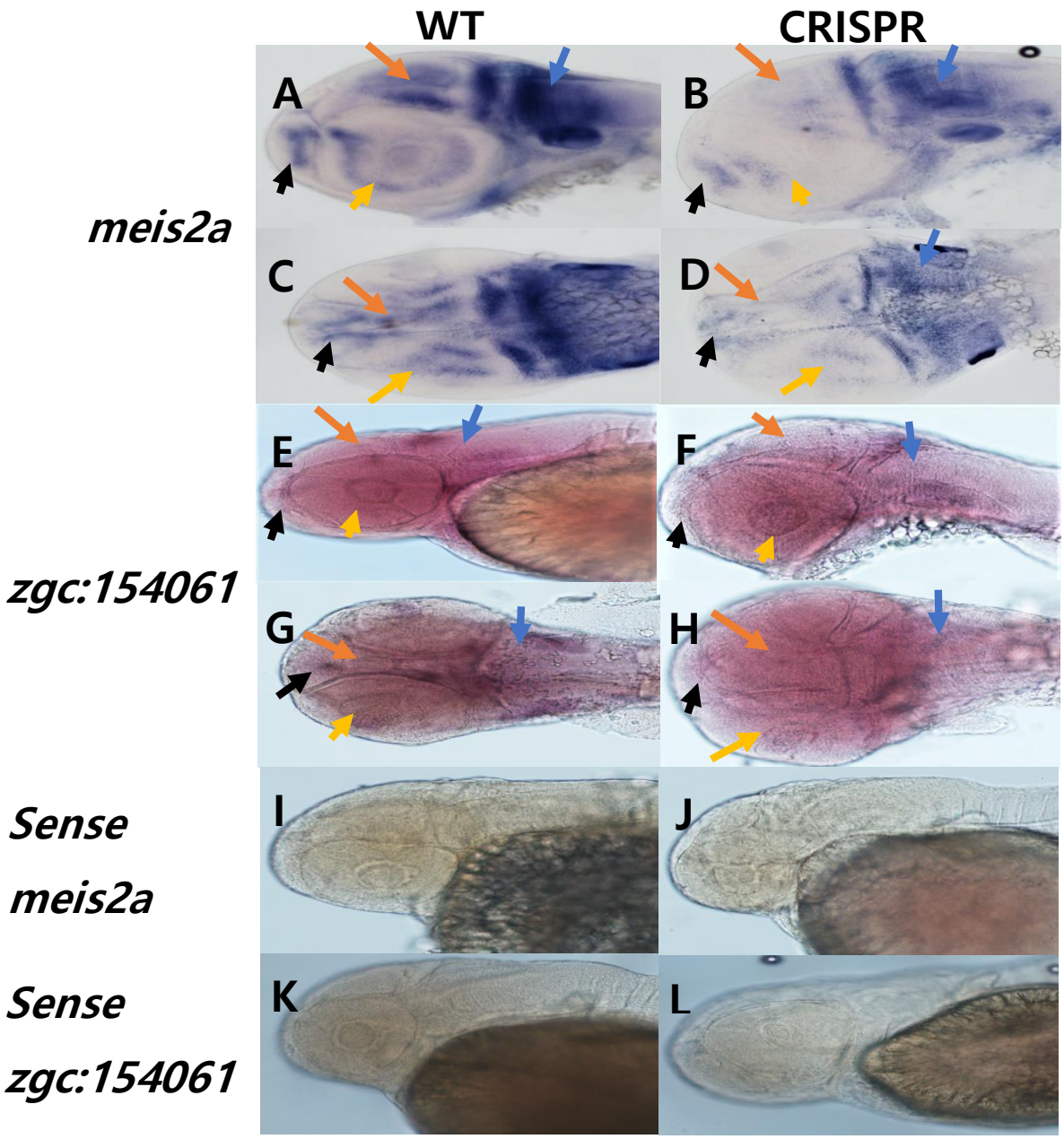

Fig. 17A-L Whole mount in situ hybridization on zebrafish embryos at 48 hpf to observe *meis2a* and *zgc:154601* expression in the wild type and transgenic CRISPR injected zebrafish. (A-D) embryos show *meis2a* expression. (A, C) wild type embryos and (B, D) CRISPR m2de1 embryos. (E-H) embryos show *zgc:154601* expression. (E, G) wild type embryos. (F, H) CRISPR m2de1 embryos. (I, J, K, L) images are the negative control that ISH performed with sense probe of *meis2a* and *zgc:154061*. (I) is wild type sense *meis2a*, (J) is wild type sense *zgc: 154061*, (K) is a CRISPR m2de1 embryos using sense *meis2a*, and (L) is CRISPR embryos using sense *zgc:154061*. (A, B, E, G, I, J, K, L) images show lateral orientation of embryos and (C, D, F, H) dorsal orientation. Expression patterns are indicated in forebrain (black arrow), midbrain (orange arrow), hindbrain (blue arrow) and eye (yellow arrow).

After confirming the successful excision of the partial m2de1 sequence in zebrafish, ISH was performed to observe the changes in the expression pattern of *meis2a* and *zgc:154061* in mutant embryos. CRISPR injected male and female zebrafish heterozygous for the m2de1 mutation were crossed and, embryos fixed at 48 hpf and examined for *meis2a* and *zgc:154061* expression. The mutant embryos were screened by DNA isolation followed by PCR reaction to confirm the consistent excision of the partial m2de1 sequence with the parent generation. Changes in *meis2a* and *zgc:154061* expression were observed in approximately 44% of analyzed embryos compared to the expression pattern observed from the wild type (Fig. 17A-H). For *meis2a*, the expression level decreased significantly in the forebrain, midbrain, hindbrain and eye (Fig, 17A-D). For *zgc:154061*, the expression level decreased as well, which is similar to*meis2a* pattern (Fig. 17E-H).

Searching for target Morpholino

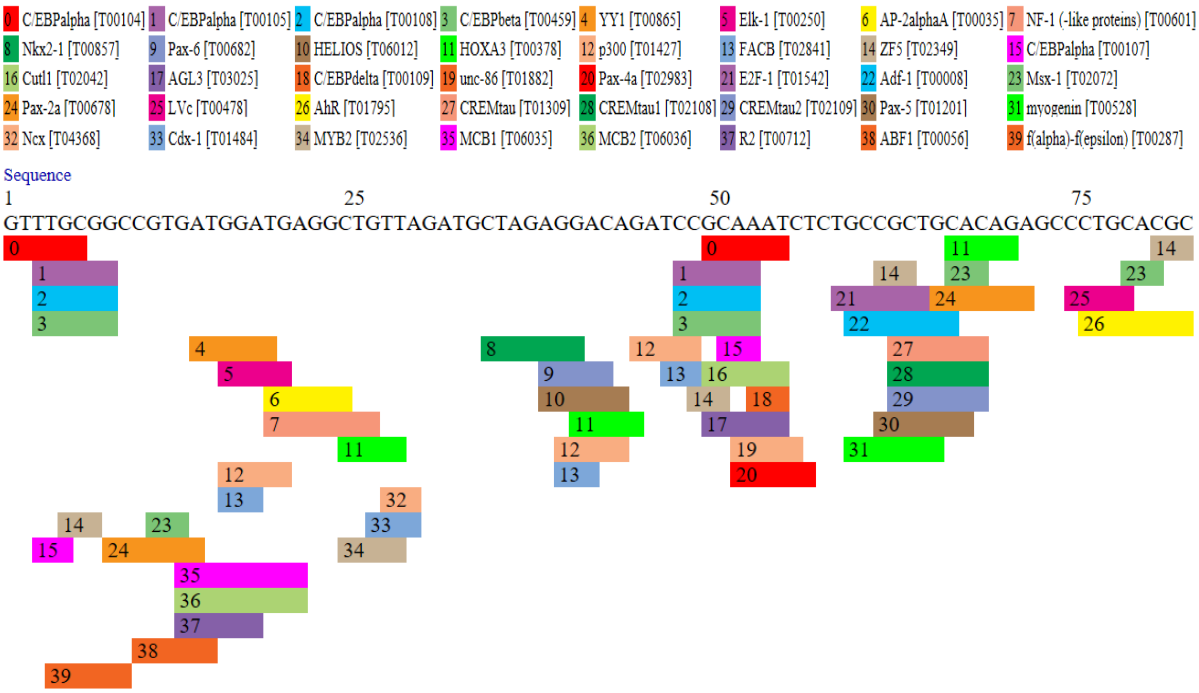

Fig. 18 The schematic result of PROMO search for finding putative transcription factor binding sites in the partial m2de1 sequence, which were knockout by CRISPR/Cas9 system. Each number represents different transcription factors, and the color marking represents binding site sequences, which indicating the types of transcription factor can bind. Some transcription factors have multiple binding sites, whereas other transcription factors have single binding site.

To further study the importance of m2de1 in directing the expression of *meis2a* and or *zgc:154061*, I examined the region of m2de1 excised by CRISPR/Cas9 to determine if the binding sites of any transcription factors known to regulate *Meis2* were present. The 82 nucleotide sequences excised in m2de1, was searched using the transcription factor binding site searching tool PROMO (http://alggen.lsi.upc.es/cgibin/promo_v3 /promo/promoinit.cgi?dirDB=TF_8.3) with parameter of less than 5% dissimilarity (Fig. 18). Matrix dissimilarity rate states that how similar between the TFBSs and putative TFs. After examining the results of putative transcription factor binding sites, the search for finding out similar temporal and spatial

expression pattern with *meis2a, zgc:154061* and m2de1 in zebrafish via ZFIN The Zebrafish Information Network (https://zfin.org/) was performed. This search identified 39 different transcription factors, whose binding sites were present in this region. One of these transcription factors is Pax6b, which has been hypothesized to regulate *Meis2* expression and is also known to be expressed in the developing eye (Coutinho et al., 2011), sharing similar expression pattern with *Meis2* and *zgc:154061*. To further examine if Pax6b might be regulating the expression directed by m2de1 I obtained a morpholino to Pax6b that is well-characterized including an easily identifiable small eye phenotype. that was previously described (Coutinho et al., 2011). Both *Pax6b* and Control MO injected embryos were fixed at 48 hpf and ISH was performed to observe the changes in the expression patterns for *meis2a* and *zgc:154061*. Approximately 54% of the Pax6b-MO injected zebrafish embryos exhibited smaller than wild-type eyes. In addition, in these injected embryos I observed decreases in *meis2a* and *zgc:154061* expression in the eye and hindbrain (Fig. 19).

Morpholino Microinjection

Table 5. MO-pax6b v1 Morpholino Reagent

Morpholino Reagent Name	**5'-Sequence-3'**
Pax6b v1-MO	**GCCTGAGCCCTTCCGAGCAAATCAG**

Table 6. Concentration of MO-Pax6b v1 with total number of injected embryos and result of microinjection.

Lists	**Numbers**
Concentration of MO-Pax6b v1	**2.5ng/nl**
Number of total injected embryos	**664**
Number dead at ~48 hpf	**185 (28%)**
Number alive at ~48hpf	**479 (72%)**
Number of embryos displaying WT phenotype	**220 (46%)**
Number of embryos displaying altered phenotype	**259 (54%)**

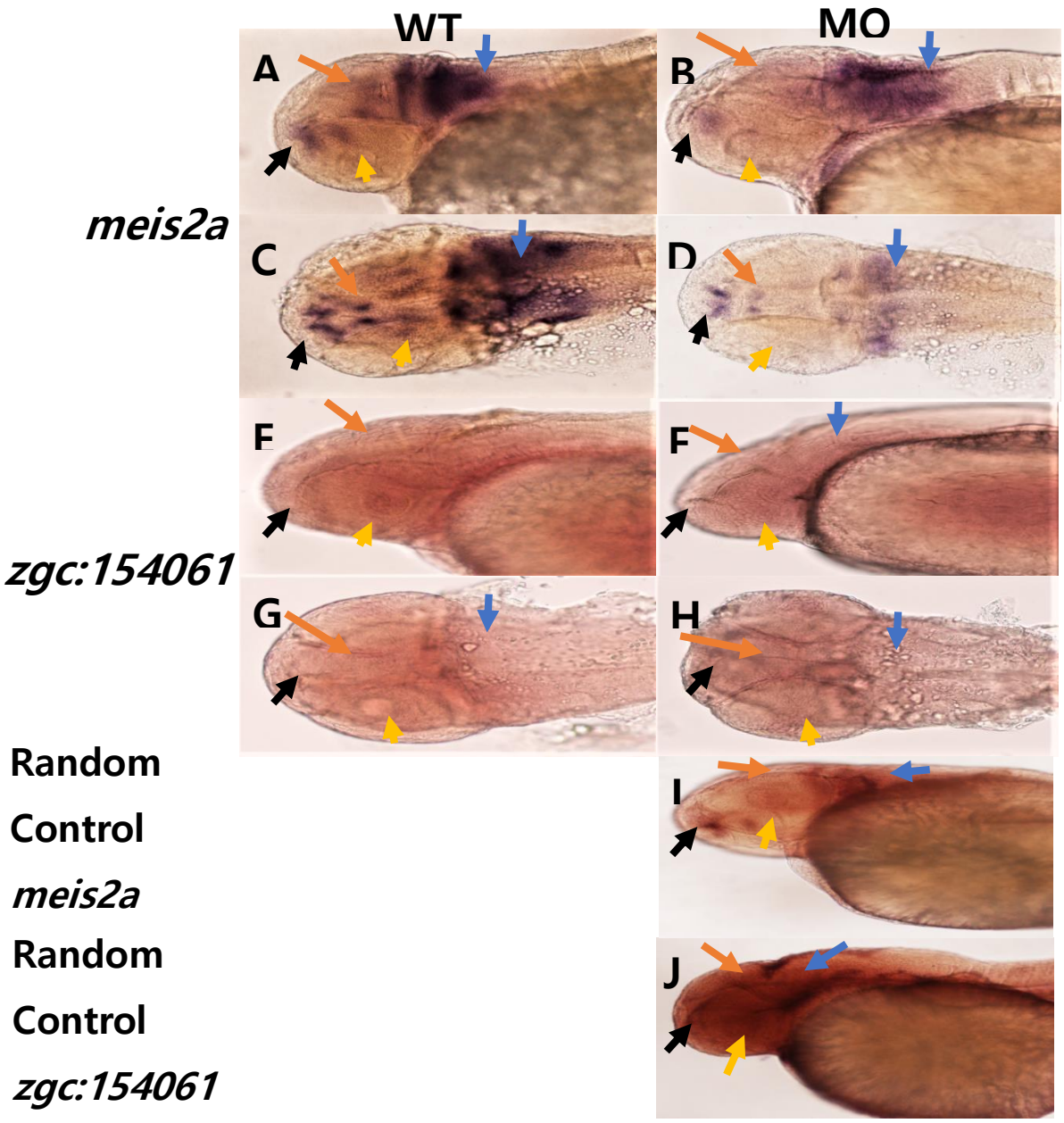

Fig. 19 Whole mount in in situ hybridization on zebrafish embryos at 48 hpf to observe *meis2a* and *zgc:154601* expressions on wild type and morpholino (MO) injected zebrafish embryos. (A-D) embryos show *meis2a* expression. (A, C) wild type embryos show *meis2a* expression. (B, D) MO injected embryos are showing *meis2a* expression. (E-H) embryos show *zgc:154601* expression. (E, G) wild type embryos show *zgc:154061* expression. (F, H) MO injected embryos are showing *zgc:154061* expression. (I, J) images are the negative control that RandomControl MO injected embryos. (I) is *meis2a* expression, (J) is *zgc: 154061* expression. (A, B, E, F, I, J) images show lateral orientation of embryos and (C, D, G, H) dorsal orientation. The expression patterns are found in forebrain (black arrow), midbrain (orange arrow), hindbrain (blue arrow) and eye (yellow arrow).

Discussion

The expression pattern of the *Meis* genes during embryonic development have been well characterized, however the mechanisms controlling *Meis* expression are less well-known. Here I describe the further characterization of the noncoding element, m2de1, that I propose is involved in regulating the expression of *meis2a* as well as its adjacent gene *zgc: 154061*. In this study, using the CRISPR/Cas9 system I demonstrate that excising a portion of the m2de1 sequence in zebrafish impacts the expression of *meis2a* and *zgc:154061*. In addition, knocking down the expression of *Pax6b* in zebrafish embryos using a morpholino also leads to an apparent decrease in expression of both *meis2a* and *zgc:154061*. This provides further evidence that m2de1 does indeed control the expression of *meis2a* and *zgc:154061* and also reveals a mechanism by which *Pax6* may control *Meis2*, a relationship that has been previously suggested (Agoston et al., 2012; Manuel et al., 2015).

The Zerucha lab has previously shown the expression pattern directed by m2de1 using eGFP as a reporter. Previous work indicated that m2de1 was able to direct expression in developing zebrafish embryos found in the forebrain, midbrain and hindbrain of zebrafish embryos at 48 and 54 hpf (Barrett, 2013; Ferrara, 2015; Tennant, 2018). In addition, a previous student in the Zerucha lab was able to generate a putative stable transgenic zebrafish line in which m2de1 directs eGFP expression (Tennant, 2018). The establishment of a stable transgenic zebrafish is important for additional studies to ensure that we are able to observe the complete pattern of expression directed by m2de1 and don't have to consider the possibility of mosaic expression. I began my project by

crossing fish from this putative line to ensure we had fish homozygous for the m2de1-eGFP transgene. As the result of producing homozygous fish to produce embryos that all exhibit a consistent expression pattern. We now have a clear understanding of the complete expression pattern being directed by m2de1, particularly in the forebrain, midbrain, hindbrain and eye (Fig. 13).

The expression patterns of *mesi2a* and *zgc:154061* have been previously studied and observed in the forebrain, midbrain, hindbrain and eye in zebrafish at 48 hpf (Carpenter et al., 2016; Tennant, 2018). In this study, whole mount *in situ* hybridization (ISH) was performed to confirm these previous studies and importantly to allow a direct comparison of the overlaps in expression of these two genes as well as that directed by m2de1. At 48 hpf, *meis2a* expression was observed in the forebrain, midbrain, hindbrain and eye, suggesting the data is consistent with previous work (Fig. 14-A, B). For *zgc:154061*, overlapping patterns of expression with *meis2a* were observed at 48 hpf (Fig. 14-C, D). However, the expression was harder to detect than with *meis2a*. Overlapping expression regions in the forebrain, midbrain, hindbrain and eye of *meis2a*, *zgc:154061* as well as that directed by m2de1 suggest the sharing of m2de1 as a *cis*-regulatory element between these two genes.

To further examine the role of m2de1 in controlling the expression of *meis2a* and *zgc:154061*, a region of m2de1 was excised from the zebrafish genome using CRISPR/Cas9. Using this approach, I was able to generate several fish that appeared to be heterozygous for this excision. These fish were raised to adulthood and screened by fin clip/PCR. A male and female fish were crossed and these embryos were screened for the

expression of *meis2a* and *zgc:154061*. Embryos were screened by DNA isolation/PCR before ISH, confirming the excision of partial sequence of m2de1. I would expect the inheritance pattern to be nonmendelian that ¼ of these embryos would be homozygous for the m2de1 excision, ½ would be heterozygous and ¼ would display a wild type phenotype. Because of this I would anticipate observing 3 expression patterns of varying intensities of *meis2a* and *zgc:154061* expression. In approximately 44% of the embryos I observed decreases in *meis2a* expression (Fig. 17A-D) in the forebrain, midbrain, hindbrain and eye. For *zgc:154061*, the decrease in the expression level was observed similarly to the observed for *meis2a* in approximately the same proportion of embryos (Fig. 17E-H) These ISH results suggest that the partial knockout of m2de1 via CRISPR/Cas9 system decreased the expression pattern of both *meis2a* and *zgc:154061*. This being said, it will be important to continue crossing the existing heterozygous m2de1 mutants in an attempt to generate a homozygous mutant, as long as this mutation does not turn out to be homozygous lethal.

Upon examination of the region of m2de1 that was excised using CRISPR/Cas9 and that seems to lead to a decrease in *meis2a* and *zgc:154061* expression, multiple putative binding sites for transcription factors were identified (Fig. 18). Among the 39 putative transcription factors binding sites one was of particular interest. Pax6b has previously been proposed to require Meis2 transcription factor as cofactor (Agoston et al., 2014) and the presence of a putative binding site for pax6b in the region of m2de1 that was excised resulting in decreases of *meis2a* and *zgc:154061* expression was intriguing. This was particularly interesting because *pax6b* has been shown to be expressed in eye

and hindbrain. Furthermore, a morpholino (MO) to successfully knockdown *pax6b* expression resulting in an easily identifiable morphological phenotype has also been described (Coutinho, P., et al. 2011).

I was able to duplicate the small eye phenotype in zebrafish embryos injected with the pax6b MO with a similar degree of success as previously described 46% (Table. 6). Importantly, *meis2a* expression in MO injected embryos showed a decrease of expression levels in eye and hindbrain at 48 hpf in comparison with wild type embryos (Fig. 19A-D). The MO injected embryos also showed a slight decrease of expression level in eye and hindbrain of *zgc:154061* at 48 hpf (Fig. 19E-H). Because of the inherent low levels of expression of *zgc:154061* at these time points, this data is not as convincing, however the CRISPR/Cas9 results do suggest the importance of this region of m2de1 to be important for controlling this gene as well.

To conclude, my results presented in this study supports the proposal that m2de1 is a *cis*-regulatory element shared by *meis2a* and *zgc:154061*. The expression pattern of m2de1, *meis2a* and *zgc:154061* shows an overlapping pattern spatially and temporally. The result of the CRISPR/Cas9 system shows the partial knockout of m2de1 decreases the expression of *meis2a* and *zgc:154061*. In addition, the result of MO microinjection shows the knockdown of Pax6b affected the expression pattern of *meis2a* and *zgc:154061* in the regions where all of these genes are expressed and where m2de1 is able to direct expression. For the future studies, the examination of subsequent generations for the m2de1 CRISPR transgenic line will be required to observe more consistent results. It will also be interesting to study the effect of some of the other putative transcription binding

sites in m2de1 on the expression of *meis2a* and *zgc:154061*. Final follow up experiment I would like to mention is utilizing electrophoretic mobility shift assay (EMSA) to identify Pax6b TF and m2de1 interaction. My data supports that Pax6b has putative binding site in m2de1. Performing EMSA will identify the binding of Pax6b TF to m2de1 and play its role during expression of *meis2a* and possibly *zgc:154061*.

References

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

Agoston, Z., Heine, P., Brill, M.S., Grebbin, B.M., Hau, A.C., Kallenborn-Gerhardt, W., Schramm, J., Gotz, M., 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.

Agoston, Z., Li, N., Haslinger, A., Wizenmann, A., Schulte, D., 2012. Genetic and physical interaction of Meis2, Pax3 and Pax7 during dorsal midbrain development. BMC Dev Biol 12, 10.

Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M., Gerstein, M.B., 2010. Annotating non-coding regions of the genome. Nat Rev Genet 11, 559-571.

Alfredo, P., 2015. The HOXD Locus in the Pathogenesis of the Celian Disease. American Journal of Food Science and Health 1, 82-85.

Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., Westerfield, M., Ekker, M., Postlethwait, J.H., 1998. Zebrafish hox clusters and vertebrate genome evolution. Science 282, 1711-1714.

Aspland, S.E., Bendall, H.H., Murre, C., 2001. The role of E2A-PBX1 in leukemogenesis. Oncogene 20, 5708-5717.

Babbs, C., Roberts, N.A., Sanchez-Pulido, L., McGowan, S.J., Ahmed, 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., Buckle, V.J., 2013. Homozygous mutations in a predicted endonuclease are a novel cause of congenital dyserythropoietic anemia type I. Haematologica 98, 1383-1387.

Barrett, C.E., 2013. Conserved Non-Coding Element derived regulation of the MEIS2.2 Homeobox Gene during Embryonic Development, Biology Department. Appalachian State University., p. 130.

Burger, A., Lindsay, H., Felker, A., Hess, C., 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. Development 11, 10.1242/dev.134809.

Carpenter, B.S., Graham, B., Zerucha, T., 2016. Identification and Developmental Expression of the Zebrafish zgc:154061 Gene. Eastern Biologist 5, 1-7.

Chan, S.K., Jaffe, L., Capovilla, M., Botas, J., Mann, R.S., 1994. The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. Cell 78, 603-615.

Cillo, C., 1994. HOX genes in human cancers. Invasion Metastasis 14, 38-49.

Coutinho, P., Pavlou, S., Bhatia, S., Chalmers, K.J., Kleinjan, D.A., van Heyningen, V., 2011. Discovery and assessment of conserved Pax6 target genes and enhancers. Genome Res 21, 1349-1359.

Deveau, H., Garneau, J.E., Moineau, S., 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol 64, 475-493.

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

Donda, A., Schulz, M., Burki, K., De Libero, G., Uematsu, Y., 1996. Identification and characterization of a human CD4 silencer. Eur J Immunol 26, 493-500.

Ferrara, T.J., 2015. Characterization of the Meis2a downstream regulatory element DR-M2DE1, Biology Department. Appalachian State University., p. 83.

Gaiano, N., Amsterdam, A., Kawakami, K., Allende, M., Becker, T., Hopkins, N., 1996. Insertional mutagenesis and rapid cloning of essential genes in zebrafish. Nature 383, 829-832.

Gan, Y., Guan, J., Zhou, S., Zhang, W., 2014. Identifying Cis-Regulatory Elements and Modules Using Conditional Random Fields. IEEE/ACM Trans Comput Biol Bioinform 11, 73-82.

Garcia-Cuellar, M.P., Steger, J., Fuller, E., Hetzner, K., Slany, R.K., 2015. Pbx3 and Meis1 cooperate through multiple mechanisms to support Hox-induced murine leukemia. Haematologica 100, 905-913.

Geerts, D., Revet, I., Jorritsma, G., Schilderink, N., Versteeg, R., 2005. MEIS homeobox genes in neuroblastoma. Cancer Lett 228, 43-50.

Gehring, W.J., Muller, M., Affolter, M., Percival-Smith, A., Billeter, M., Qian, Y.Q., Otting, G., Wuthrich, K., 1990. The structure of the homeodomain and its functional implications. Trends Genet 6, 323-329.

Gonzalez-Crespo, S., Abu-Shaar, M., Torres, M., Martinez, A.C., Mann, R.S., Morata, G., 1998. Antagonism between extradenticle function and Hedgehog signalling in the developing limb. Nature 394, 196-200.

Heyn, P., Kalinka, A.T., Tomancak, P., Neugebauer, K.M., 2015. Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences. Bioessays 37, 148-154.

Hinman, V., Cary, G., 2017. The evolution of gene regulation. Elife 6.

Holland, P.W., 2013. Evolution of homeobox genes. Wiley Interdiscip Rev Dev Biol 2, 31-45.

Holland, P.W., Takahashi, T., 2005. The evolution of homeobox genes: Implications for the study of brain development. Brain Res Bull 66, 484-490.

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

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., Zhang, F., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31, 827-832.

Irimia, M., Tena, J.J., Alexis, M.S., Fernandez-Minan, A., Maeso, I., Bogdanovic, O., de la Calle-Mustienes, E., Roy, S.W., Gomez-Skarmeta, J.L., Fraser, H.B., 2012. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res 22, 2356-2367.

Jayavelu, N.D., Jajodia, A., Mishra, A., and Hawkins, R.D., 2018. An atlas of silencer elements for the human and mouse genomes. bioRxiv 10.1101, 252304.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821.

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

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

Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R., Papantonis, A., 2012. Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin 5, 1.

Kulaeva, O.I., Nizovtseva, E.V., Polikanov, Y.S., Ulianov, S.V., Studitsky, V.M., 2012. Distant activation of transcription: mechanisms of enhancer action. Mol Cell Biol 32, 4892-4897.

Li, Y., Chen, C.Y., Kaye, A.M., Wasserman, W.W., 2015. The identification of cis-regulatory elements: A review from a machine learning perspective. Biosystems 138, 6-17.

Liu, K.Y., Wang, L.T., Hsu, S.H., Wang, S.N., 2019. Homeobox Genes and Hepatocellular Carcinoma. Cancers (Basel) 11.

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, 59-75.

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

Maeda, R.K., Karch, F., 2006. The ABC of the BX-C: the bithorax complex explained. Development 133, 1413-1422.

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

Manuel, M.N., Mi, D., Mason, J.O., Price, D.J., 2015. Regulation of cerebral cortical neurogenesis by the Pax6 transcription factor. Front Cell Neurosci 9, 70.

Matthews, K.S., 1992. DNA looping. Microbiol Rev 56, 123-136.

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

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

Morgan, R., Boxall, A., Harrington, K.J., Simpson, G.R., Michael, A., Pandha, H.S., 2014. Targeting HOX transcription factors in prostate cancer. BMC Urol 14, 17.

Morgan, R., El-Tanani, M., Hunter, K.D., Harrington, K.J., Pandha, H.S., 2017. Targeting HOX/PBX dimers in cancer. Oncotarget 8, 32322-32331.

Mortlock, D.P., Innis, J.W., 1997. Mutation of HOXA13 in hand-foot-genital syndrome. Nat Genet 15, 179-180.

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

Nelson, K.N., 2011. Identification of gene regulatory elements associated with the *MEIS* family of Homeobox Genes, Biology Department. Appalachian State University., p. 119.

Ni, J., Wangensteen, K.J., Nelsen, D., Balciunas, D., Skuster, K.J., Urban, M.D., Ekker, S.C., 2016. Active recombinant Tol2 transposase for gene transfer and gene discovery applications. Mob DNA 7, 6.

Nunes, F.D., de Almeida, F.C., Tucci, R., de Sousa, S.C., 2003. Homeobox genes: a molecular link between development and cancer. Pesqui Odontol Bras 17, 94-98.

Ogbourne, S., Antalis, T.M., 1998. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem J 331 (Pt 1), 1-14.

Owa, T., Taya, S., Miyashita, S., Yamashita, M., Adachi, T., Yamada, K., Yokoyama, M., Aida, S., Nishioka, T., Inoue, Y.U., Goitsuka, R., Nakamura, T., Inoue, T., Kaibuchi, K., Hoshino, M., 2018. Meis1 Coordinates Cerebellar Granule Cell Development by Regulating Pax6 Transcription, BMP Signaling and Atoh1 Degradation. J Neurosci 38, 1277-1294.

Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., Bejerano, G., 2013. Enhancers: five essential questions. Nat Rev Genet 14, 288-295.

Raab, J.R., Kamakaka, R.T., 2010. Insulators and promoters: closer than we think. Nat Rev Genet 11, 439-446.

Spitz, F., Furlong, E.E., 2012. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13, 613-626.

Symington, L.S., Gautier, J., 2011. Double-strand break end resection and repair pathway choice. Annu Rev Genet 45, 247-271.

Tennant, M.D., 2018. Transcription Regulation of the MEIS2 Locus, Biology Department. Appalachian State University., p. 119.

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

West, A.G., Gaszner, M., Felsenfeld, G., 2002. Insulators: many functions, many mechanisms. Genes Dev 16, 271-288.

Westerfield, M., 2000. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio), Univ. of Oregon Press, Eugene.

Wickramasinghe, S.N., Wood, W.G., 2005. Advances in the understanding of the congenital dyserythropoietic anaemias. Br J Haematol 131, 431-446.

Wilming, L.G., Boychenko, V., Harrow, J.L., 2015. Comprehensive comparative homeobox gene annotation in human and mouse. Database (Oxford) 2015.

Yang, Y., Kwang, C., D'Souza, U., Lee, S., Junn, E., and Mouradian, M., 2000. Three-amino acid extension loop homeodomain proteins MEIS2 and TGIF differentially regulate transcription. . J. Biol. Chem. 275, 20734-20741.

Vita

Young Koun Jeon was born in Daejeon, South Korea to Gwang Woon and Jung Bok Jeon. He graduated from Patterson School in North Carolina in May 2009. The following autumn, he entered Lees-McRae College to study biology and graduated in May 2013 with the Bachelor of Science degree award. In June 2015, he entered in the army in South Korea to complete the mandatory military service. After his 21 months of service in the army, he was accepted in a graduate program of General Biology at Appalachian State University in the summer of 2017, and started the study toward a Master of Science degree.