FURTHER CHARACTERIZATION OF A HIGHLY CONSERVED NOVEL MEIS2 LINKED GENE AND PROTEIN PRODUCT

A Thesis
by
ZACHARY SCOTT WILLIAMS

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

August 2014
Department of Biology

FURTHER CHARACTERIZATION OF A HIGHLY CONSERVED NOVEL MEIS2 LINKED GENE AND PROTEIN PRODUCT

A Thesis
by
ZACHARY SCOTT WILLIAMS
August 2014

APPROVED BY:

Dr. Ted Zerucha
Chairperson, Thesis Committee

Dr. Ece Karatan
Member, Thesis Committee

Dr. Mary U. Connell
Member, Thesis Committee

Dr. Susan L. Edwards
Chairperson, Department of Biology

Dr. Edelma D. Huntley
Dean, Cratis Williams Graduate School

Abstract

FUTHER CHARACTERIZATION OF A HIGHLY CONSERVED NOVEL MEIS2 LINKED GENE AND PROTEIN PRODUCT

Zachary Scott Williams
B.S., Appalachian State University
M.S., Appalachian State University

Chairperson: Dr. Ted Zerucha

We have identified a novel gene linked to the *Meis2* gene (*meis2.2* in zebrafish) in all vertebrates with publicly available genome data. This gene is always located immediately downstream of *Meis2* (*meis2.2* in zebrafish) and is organized in an inverted convergently-transcribed manner. Transcripts of this gene are maternally expressed ubiquitously at high levels during early zebrafish development with the highest level of expression at fertilization and decreasing until 8 hour past fertilization (hpf). The zygotic genome produces transcripts of this gene again around 12 hpf and low levels of expression are then localized to the developing neural tube with further restriction to the retina of the developing zebrafish at 48 hpf. Using an antibody raised against a short peptide portion within the predicted zebrafish protein product we have shown that the gene is translated into protein within the developing embryo and that it is expressed at various stages throughout development. Western blots show that the protein is expressed as early as 2 hpf and is present in significant amounts until 24 hpf, at which point its expression is significantly decreased. Immunohistochemistry on 48 hpf zebrafish embryo cross-sections show that the protein is present and is highly localized to the optic area, including the retina, optic nerve and optic cup, as well as the olfactory bulb and epithelium.

Whole-mount *in situ* IHC hybridization on embryos 24 hpf and 48 hpf reveals localized expression the developing olfactory bulb and epithelium along with the optic cup and brain.

Dedication

To my younger sister, Sam, for always being by my side and being my best friend. To my parents Scott and Teresa who have given me unconditional support and love and have always pushed me to reach my goals. To my girlfriend, Emily, for putting up with me and my antics throughout my time in graduate school, and to all my friends for keeping me sane and active.

Acknowledgements

The research work that I have done in pursuit of a Master's at Appalachian State University would not have been possible without the aid I have received during my time here. I would first like to extend my gratitude towards the Appalachian State University Office of Student Research and the Appalachian State University Graduate Student Association Senate for all of the research funding support I have received during my years of graduate school. I would also like to thank the Society of Developmental Biology for making international conference travel possible.

I owe much appreciation to Dr. Ece Karatan, Dr. Howie Neufeld, and Dr. Mark Venable for their mentoring roles as an undergraduate and guidance toward pursuing a Master's degree. I especially thank Dr. Ece Karatan for her continued mentoring and assistance in graduate school affairs, in addition to her constant willingness to provide invaluable research project assistance and feedback. I am forever grateful to Dr. Mary Connell for entrusting me with teaching molecular biology to her undergraduate students. Her instruction and guidance were instrumental and I am very thankful for the opportunities presented. I would also like to thank Monique Eckerd for her time and effort in the animal facility assisting us in taking care of our zebrafish.

I would like to thank all of my Biology graduate student peers for providing camaraderie, research feedback, and conversation at the saloon. I owe a big thanks to my family for keeping me grounded and providing unconditional support and to my Uncle Drew and Aunt Renae for always providing a home away from home and encouragement.

Lastly, I owe the most gratitude to my primary advisor, Ted Zerucha. Ted has been an amazing role model, mentor, and boss while remaining a personal friend. He has helped me reach my goals and I cannot thank him enough. Without Ted's understanding, a side business and personal time in the woods would not have been possible. Under his guidance I have learned so much and am forever thankful to have worked in the Zerucha lab.

Foreword

This thesis manuscript and the figures, tables, and references herein were formatted according to the author instructions for manuscript submission to the peer-reviewed journal *Development*, published by The Company of Biologists.

Table of contents

Introduction

Embryonic development is an elaborate and complex process that depends on a network of precise and accurate gene expression patterns. Fertilization initially triggers a cascade of developmental events giving rise to a complex organism from a single cell (Gilbert, 2000). The organization and timing of these development events are stored as information within the cell's DNA. DNA is utilized by each cell as a blueprint plan especially during embryonic development; directing processes involved in cell proliferation, differentiation, functioning, and regulation (Gilbert, 2000; Holliday and Pugh, 1990; Wolpert, 2007). DNA contains codes for genes, which can be transcribed into RNA and translated into functional protein via gene expression. The temporal and spatial expression of each gene is heavily controlled, along with the availability and amount of transcripts, specifically limiting and regulating gene expression. These regulation processes are especially important during early embryonic development (Bjerke et al., 2011; Bumsted-O'Brien et al., 2007; Holliday and Pugh, 1975).

Facilitated by early cellular interaction, morphological and body plan characteristics begin developing in an organized manner. This is one of the most important developmental processes and results in early body plan patterning and formation of the body axis (Gehring, 1993; Gilbert, 2000; Holliday and Pugh, 1990; Wolpert, 2007). Body plan patterning includes the proper development of the anterior-posterior (AP) axis, dorsal-ventral (DV) axis, and the right-left axis, running head to tail, topside to bottom, and right to left respectively. Successful body plan formation is essential for proper early development and enables proper embryonic developmental progression through gastrulation, neurulation, and organogenesis. The

developmental events leading to the formation of a body plan and axis formation are tightly regulated and under strict control, often in association with homeobox genes (Duboule, 1998; Gilbert, 2000; Holliday and Pugh, 1990; Wolpert, 2007).

Homeobox genes contain a homeobox sequence that includes a highly conserved 180 base pair coding region. This region codes for a homeodomain, a conserved 60 amino acid domain encoding for a helix-turn-helix DNA binding motif. This binding motif enables the homeodomain protein to bind to, and subsequently regulate, a specific area of DNA, thus acting as transcription factors (Duboule, 1998; Gehring, 1993; McGinnis et al., 1984). The DNA binding motif allows the protein to preferentially activate or repress significant genes spatially and temporally by controlled binding within regulatory regions of developmental genes. The homeobox genes were initially discovered in the 1970's while examining body segmentation control in the fruit fly *Drosphila melanogaster*. It was determined that genes located in specifically organized clusters were responsible for axis formation and segmentation specification. If mutations were featured within these genes, segment identity was disrupted and malformation occurred during early development (Morata and Lawrence, 1977; Wolpert, 2007). Similar gene clusters were observed in almost every other multi-cellular organism, containing the same conserved homeobox region, thus giving rise to a superfamily of genes.

Within the homeobox superfamily of genes, the *Hox* genes play a significant role in the establishment and maintenance of the anterior-posterior axis during embryonic development (Alharbi et al., 2012; Amores et al., 1998; Krumlauf, 1994). Discovered in 1978, the *Hox* genes were first observed in *Drosophila* and presumed responsible for controlling body segmentation (Carroll 1995; Lewis, 1978). Since then, the *Hox* genes have been identified as conserved in every animal with a distinct AP axis and have additionally been observed in cnidarians, an organism lacking a distinct axis. *Hox* genes, however, have no homolog within plants, protozoa, or sponges (Amores et al., 1998; Balavoine et al., 2002; Krumlauf, 1994; Lemons and McGinnis,

2006; Maconcohie et al., 1996). Found in clusters, *Hox* genes exhibit colinearity, meaning they are physically organized in the same order they are transcribed and expressed temporally and spatially (Amores et al., 1998; Bomgardner et al., 2003; Carroll 1995; Duboule, 1998; Duboule, 2007; Krumlauf, 1994; McGinnis and Krulauf, 1992; Prince et al., 1998). This unique feature is found in the same manner in all organisms observed (Mark et al., 1997). Furthermore, disruption of *Hox* genes have been widely shown to be associated with varying malignancies, often being observed with tissue-specific function and expression within cancer cells (Alharbi et al., 2012; Argiropoulos and Humphries, 2007; Crijns et al., 2007; Mark et al., 1997; Milech et al., 2010). In fact, out of 6,817 genes examined in patients with acute myeloid leukemia, human *HOXA9* was shown to be the most correlated gene to treatment failure (Afonja et al., 2000; Fujino et al., 2001). It is proposed that within leukemic stem cells, *Hox* genes play imperative roles facilitating pathways associated with the cancer cells' 'self-renewal program' (Argiropoulos and Humphries, 2007; Krumlauf, 1994; Lawrence et al., 1996). Likewise, *Meis1* has been shown to also play a role interacting with *Hox* in the self-renewal of hematopoietic stem cells, presumably sharing a similar function in both normal and cancerous blood cell lineages (Hisa et al., 2004, Lawrence et al., 1996; Pillay et al., 2010).

Hox genes regulate axial formation and cell fate diversity by utilizing the homeodomain of the proteins they encode to bind to DNA preferentially and regulate developmental gene expression (Gehring, 1993; Krumlauf, 1994; McGinnis and Krulauf, 1992; Maconochie et al., 1996). By acting as a transcription factor, the protein is able to bind to *cis*-regulatory elements associated with these specific genes as an activator or repressor. Binding by the Hox protein by itself, however, has been shown to be relatively nonspecific and inefficient. With the regulation of major axis formation and development under the control of *Hox* genes, the Hox proteins additionally rely on interactions with other proteins (Caroll, 1995; Ekker et al., 1994; Moens and Selleri, 2006; Waskiewicz et al., 2001). After residue mapping studies, it was concluded

that not only do Hox proteins bind to DNA, but they bind even more specifically to other homeodomain proteins via proteins, termed co-factors, via protein-protein interactions (Carroll, 1995; Mann and Affolter, 1998; Rivas et al., 2013; Sharkey et al., 1997). Binding with additional co-factor proteins provides even further selectivity and specificity resulting in better gene regulation. This cooperative teamwork with cofactors results in the formation of transcription factor complexes, increasing the surface area of DNA bound by the proteins and subsequently, specificity (Hoey et al., 1988; Mann and Affolter, 1998).

One group of Hox cofactors of particular interest is encoded by the TALE family of genes. This family is characterized by a three amino-acid loop extension between the two helixes within the homeodomain (Biemar et al., 2001; Gehring, 1993; Yang et al., 2000). The TALE family is the largest set of *Hox* cofactors and includes the *Meis*, *Pbc*, *Iro*, and *Tgip* in animals and the *Knox* and *Bel* gene subfamilies within plants; all classified based on the presence of conserved motifs upstream of the homeodomain (Burglin, 1997). These motifs are responsible for the appropriate specific interactions between both Hox and the target DNA (Affolter et al., 1999; Burglin, 1997). It appears that the characteristic amino acid loop extension plays an important role in binding between proteins rather than direct DNA binding (Hyman-Walsh et al., 2010; Yang et al., 2000). Additionally, it has been observed that interactions occur between these protein classes even in the absence of DNA, suggesting novel purposes for the protein-protein interactions observed away from regulating DNA (Berthelsen et al., 1999; Waskiewicz et al., 2001). In fact, studies have shown that this interaction is highly efficient and stable (Affolter et al., 1999). Not only have these genes been shown to cooperatively bind together, but evidence has also suggested regulatory mechanisms between the genes themselves. In *Drosophila*, it has been observed that direct interaction between cofactors is required in order to translocate to the nucleus where DNA binding occurs (Affolter et al., 1999, Agoston and Schulte, 2009; Berthelsen et al., 1998).

Proteins encoded for by the *Meis* family of genes were initially discovered in a murine model when the Myeloid Ecotropic Leukemia Virus integrated itself into a then novel *Meis* gene's coding sequence and disrupted gene expression (Moskow et al., 1995). Commonly, the same murine genes that cause leukemia when they are impaired by virus insertions are often associated with human leukemia as well (Afonja et al., 2000). This family of genes quickly became identified using DNA-DNA hybridization studies. Using the *Meis1* homeobox sequence as a probe, *Meis2* and *Meis3* were identified; however, when using a probe made from the untranslated region of *Meis1* neither gene was detected. The lack of binding from the probe created from the *Meis1* untranslated regions suggests unique untranslated regions amongst genes sharing a highly conserved core homeodomain region (Nakamura et al., 1996; Steelman et al., 1997). These newly discovered *Meis* paralogs were identified in a wide variety of diverse organisms, ranging from three genes (*Meis1*, *Meis2*, and *Meis3*) in land vertebrates to four within the teleosts (addition of *meis4*), seemingly due to additional gene duplication events (Geerts et al., 2005; Nakamura et al., 1996; Waskiewicz et al., 2001). Orthologs have also been identified within invertebrate species as well as similarly acting genes identified in plants, both evidence suggesting divergence from a common ancestor (Becker et al., 2002; Burglin, 1998; Kurant et al., 1998). In *Drosophila* an ortholog for *Meis* has been identified and named *hth* (Affolter et al., 1999). Interestingly, in addition to a C-terminal domain, N-terminal domain, and a homeodomain, all *Meis* proteins have an hth domain. This domain was named after the *Meis* ortholog found within fruit flies. This domain increases binding specificity to Pbx and has implications in autoinhibition pathways (Hyman-Walsh et al., 2010).

Within the vertebrates, the *Meis* family of genes have been identified within humans (*Homo sapiens*), chickens (*Gallus gallus*), mice (*Mus musculus*), and zebrafish (*Danio rerio*), all with a high level of conservation and specificity (Biemar et al., 2001; Bomgardner et al., 2003; Cecconi et al., 1997; Geerts et al., 2005; Nakamura et al., 1996; Sanchez-Guardado et al., 2011;

Steelman et al., 1997; Waskiewicz et al., 2001; Zerucha and Prince, 2001). While the *Meis* genes are highly conserved across all species examined, alternative splice variants have also been identified, adding to the diversity and complexity of their products (Burglin, 1997; Geerts et al., 2005; Huang et al., 2005; Maeda et al., 2001; Sanchez-Guardado et al., 2011). Alternative splicing is proposed to give rise to a number of different proteins, from variants that lack DNA binding motifs or homeodomains to an alternative C-termini, all based off of the exons encoded (Burglin, 1997).

In order to gain insight into the function of the Meis proteins, the protein's structure was first examined. Acting as both a transcription factor and cofactor, Meis proteins utilize three domains in order to bind to DNA and other transcriptional proteins. The homeodomain serves to bind to the actual DNA, while a flexible N-terminal, or Meinox, domain plays a role in the binding to a Pbx protein, a specific transcriptional cofactor protein (Berthelsen et al., 1998; Burglin, 1997; Chang et al., 1997; Choe et al., 2002; Ekker et al., 1994; Jacobs et al., 1999; Shanmugam et al., 1999). A third domain, the C-terminal domain, serves to bind to variable Meis binding co-factors, including the Hox proteins, in order to form transcription factor complexes thus furthering binding specificity (Burglin, 1997; Huang et al., 2005; Moskow et al., 1995; Williams et al., 2005). The C-terminal domain is also one area of sequence difference between the families of proteins and includes additional activation domains, such as domains that respond to cell signaling and others that are responsible for the proteins transcription activity (Huang et al., 2005). As a complex, the transcription factors are specifically bound to act as regulators for a target gene situated downstream or upstream (Mann and Affolter, 1998). A preference toward trimeric formation (between Meis, Pbx, and Hox; See Fig. 1) has been shown to not only increase stability but also binding specificity (Chang et al., 1997; Shanmugam et al., 1999). Not only protein cofactors bind in complex to target DNA sequences, but studies have also shown dimer formation occurring between homeodomian proteins in the absence of

DNA (Shanmugam et al., 1999; Waskiewicz et al., 2001). Meis proteins have been shown to bind to Pbx proteins in the absence of DNA, *in vitro*, in order to regulate the export of Pbx proteins from the cytoplasm, suggesting novel stabilization and regulatory roles for these transcription factors outside of the known cofactor binding functions (Berthelsen et al., 1999; Mann and Abu-Shaar, 1996; Waskiewicz et al., 2001). The striking similarities between the *Meis* and *Pbx* genes, another *Hox* co-factor, also suggest further protein interactions and functions (Moskow et al., 1995). In addition to binding in the absence of DNA, dimers have been also shown to form complexes on DNA without the presence of one of the pair's target sequence, further confirming the strong affinity between TALE protein class members (Chang et al., 1997; Shanmugam et al., 1999). Observed functional similarities between class members suggest a shared conserved domain is responsible (Choe et al., 2002).

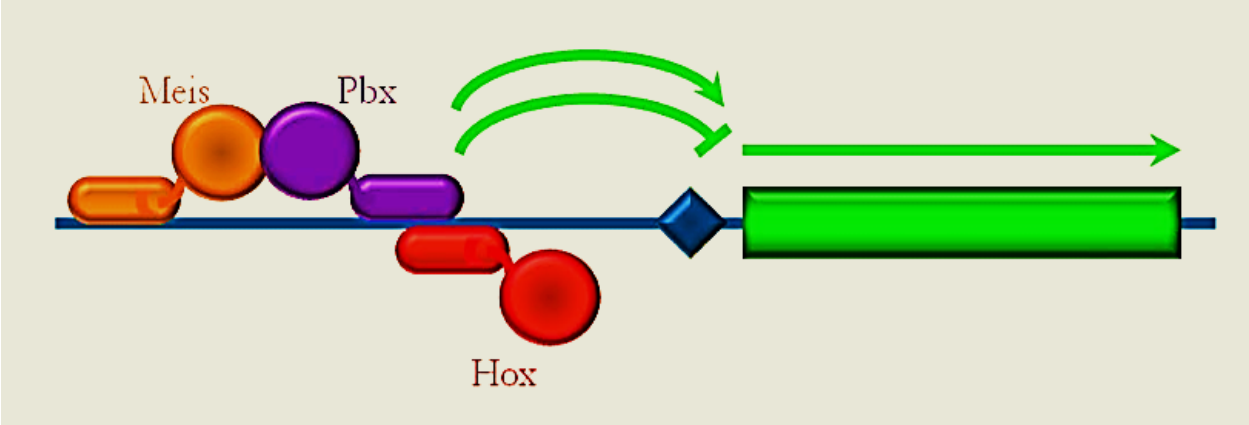

Fig. 1. A generalized depiction of a trimeric complex between Meis, Pbx, and Hox cooperatively regulating a target gene (green) downstream through a proximal promoter (blue). Formation of a trimeric complex increases binding stability and specificity, thus enabling more efficient gene regulation. The combination and utilization of multiple differing binding sites, as compared to a single binding site for a lone transcription factor, yields higher specificity and binding precision. (Image borrowed from Cochrane, 2012).

In total, four families of *Meis* genes have been identified: *Meis1*, *Meis2*, *Meis3*, and *Meis4*. Discovered first, the Meis1 family of proteins have since been shown to be highly expressed during early development. *Meis1* activation is nearly always accompanied by the expression of *Hox* genes, specifically *Hoxa7* and *Hoxa9* in murine models (Afonja et al., 2000; Nakamura et al., 1996). Expression has been demonstrated throughout the somites and mesoderm during embryonic development in a variety of organisms and expression patterns progress mainly anteriorly as development continues (Coy and Borycki, 2010; Jiang et al., 2009; Maeda et al., 2002; Waskiewicz et al., 2001). Late in development expression is readily seen anteriorly in the midbrain, hindbrain, and the developing optic area while observed posteriorly in the mesoderm and neural tube (Choe et al., 2002, Erickson et al., 2010). In addition to being patterned throughout the developing anterior and posterior axis in its associated role interacting with Hox proteins, Meis1 is also expressed in limb buds and the brachial arches (Coy and Borycki, 2010; Maeda et al., 2002; Waskiewicz et al., 2001). The observed wide expression pattern of Meis hints at the variety of roles and functions Meis is involved with throughout development.

In zebrafish, studies have shown that if the *meis1.1* gene is mutated in order to make its protein non-functional, the resulting hindbrain is not segmented or compartmentalized correctly, suggesting an important role in hindbrain segmentation and patterning through interactions with *Hox* (Waskiewicz et al., 2001). In addition to interacting with *Hox* during development, *Meis1* has been shown to interact with a variety of developmental genes in a variety of processes. During embryonic development, the Meis1 protein interacts with *Sox3* in order to regulate the neuron positioning in the spinal cord (Mojsin and Stevanovic, 2010). In *Xenopus*, MEIS protein expression is localized in neural crest cells and plays important roles facilitating neuroblastoma proliferation (Geerts et al., 2003; Jones et al., 2000; Maeda et al., 2001; Maeda et al., 2002; Salzberg et al., 1999).

During pancreatic development, the Meis1 protein interacts and regulates *Pax6* gene expression, driving cellular differentiation in the early organ (Carbe et al., 2012; Zhang et al., 2002; Zhang et al., 2006). Meis1 expression is observed in the developing retinal area and has been shown to play roles in the proliferation, positioning, and upkeep of retinal progenitor cells utilizing interactions with cell cycle proteins cyclin D1 and c-myc (Bessa et al., 2008; French et al., 2007; Heine et al., 2008; Royo et al., 2012). It appears that *Meis1* plays a specific role in axial positioning and cellular fate mapping of cell axons and studies have shown disruption in the developing retina and tectum in *meis1* zebrafish knockdowns (Erickson et al., 2010; French et al., 2007; Hisa et al., 2004). Furthermore, in chicken, *Meis1* plays a role in the developing proximal limbs through the control of retinoic acid, although only during early limb formation and limb regeneration. Meis1 localized in these areas enables cell differentiation along the proximal axis (Mercader et al., 2000; Mercader et al., 2005). Lastly, Meis1 proteins have been shown to be involved with hematopoietic processes after experiments with *Meis1* deficiencies showed deficiencies in blood vessel and arterial formation (Azcoitia et al., 2005; Hisa et al., 2004, Minehata et al., 2008). To date, the only *Meis* mutant produced has been a *Meis1* mutant. Mice with this mutation, a dysregulated *Meis1*, display deficiencies in proper eye development and angiogenesis (Agoston and Schulte, 2009). Meis1 is also expressed in adult bone marrow and both retinal and liver hematopoietic stem cells and has been found to be overexpressed in many types of leukemia and ovarian cancer, all evidence suggesting the importance *Meis1* plays in a variety of functions, while often tissue-specific (Afonja et al., 2000; Crijns et al., 2007; Imamura et al., 2002; Milech et al., 2010; Pineault et al., 2002; Royo et al., 2012). *Meis1* and *Meis2* have also been shown to play a role in the development of the inner ear, both expressed in the semicircular canals, cristae, and otic epithelium, functioning in differentiation and patterning (Sanchez-Guardado et al., 2011).

After subsequently being discovered, the *Meis1* gene was used as a probe to identify other homologs (Steelman et al., 1997). *Meis2* was the first homolog identified and was named so after being found to contain an identical homeobox sequence to *Meis1* (Oulad-Abdelghani et al., 1997). While *Meis2* shares the same homeodomain sequence and binds in complex with the same other homeodomain proteins (Hox and Pbx), *Meis2* has its own specific expression patterns and roles observed, being first expressed strongest within the hindbrain during gastrulation, and later during embryogenesis (Biemar et al., 2001; Chang et al., 1997; Zerucha and Prince, 2001). *Meis2* expression patterns are localized in a variety of areas including the developing forebrain, midbrain, hindbrain, somites, mesoderm, limb buds, root ganglia, branchial arches, neural tube, spinal cord, and optic cup (Biemar et al., 2001; Cecconi et al., 1997; Coy and Borycski, 2010; Heine et al., 2008; Mercader et al., 2005; Oulad-Abdelghani et al., 1997; Waskiewicz et al., 2001; Zerucha and Price, 2001).

Many of the functions and roles of *Meis2* are similar to that of *Meis1*. Vertebrate *Meis2* functions in developing and regenerating limb buds, by acting as a target for retinoic acid, to promote proximal out growth, development, and organization (Mercader et al., 2000; Mercader et al., 2005; Oulad-abdelghani et al., 1997). *Meis2* has also been shown to play a role directing cellular differentiation in the lateral ganglionic eminence (Bumbsted-O'Brien et al., 2007). While both *Meis1* and *Meis2* play a role in the initial pancreatic cellular proliferation, their mechanisms are dissimilar. As mentioned before, the Meis1 protein plays a role in *Pax6* gene regulation in order to drive differentiation within the developing organ, whereas the Meis2 protein (isoform: *Meis2b*) has been shown to form transcription factor complexes with Pbx1b and Pdx1 proteins in order to drive organogenesis (Carbe et al., 2012; Stoffers et al., 1997). Interestingly, *Meis2* has also been shown to interact with the *Pax6* gene in dual roles within the olfactory bulb neurogenesis. Recent research has suggested Meis2 expression is localized in the developing neuroblasts of the olfactory bulb along with some interneurons in the adult

olfactory bulb, suggesting cooperation with Pax6 to facilitate proper neuron proliferation, fate, and migration in this developing structure (Agoston et al., 2014). In mice, *Meis2* has been shown to play a role in the development of the cardiac septum along with cellular differentiation in the somatic mesoderm (Cecconi et al., 1997). *Meis2* has also been shown to directly interact with *Hox* to play a role in murine myeloid leukemogenesis and leukemia cell lines, possible through blocking cell differentiation pathways (Fujino et al., 2001).

More so than *Meis1*, *Meis2* has been shown to play an important role in the proper development of the eye, retina, and optic area in studies examining a variety of vertebrates including humans, mice, and chickens. Similar to *Meis1*, the Meis2 protein interacts with *Pax6* during the development of the eye functioning as a cell differentiation regulator (Bumsted-O'Brien et al., 2007; Carbe et al., 2012; Heine et al., 2008; Zhang et al., 2002; Zhang et al., 2006). Similar to the roles *Meis2* plays in the development of the optic area, *Meis2* has also been implicated in the developing neural system, including proper development of the neural tube, branchial arches, and somites (Cecconi et al., 1997). In monkeys, evidence suggests *Meis2* functions in the developing forebrain, specifically expression within the striatum (Takahashi et al., 2008). *Meis2* also interacts with *Otx2* in the chicks to facilitate proper tectal fate within the developing midbrain (Agoston and Schulte, 2009). Lastly, the Meis2 protein cooperatively mediates *Hox* enhancer activity as a complex in the hindbrain in murine models (Jacobs et al., 1999).

Out of the last two *Meis* genes, *Meis3* and *Meis4*, only the earlier has been examined expansively in vertebrates. In fact, only the identification of a *Meis4* gene, via presence of the homeobox in zebrafish, is currently available (Waskiewicz et al., 2001). *Meis3*, however, has been examined and found to be localized spatially and temporally similar to that of the first two *Meis* genes. *Meis3* is expressed early in development and has been shown to play an important role in hindbrain patterning in *Xenopus*. During this early patterning and development, *Meis3*

interacts with *Hoxb1b* and *Pbx4* in order to regulate target genes and facilitate cellular differentiation (Choe et al., 2002; Salzberg et al., 1999; Waskiewicz et al., 2001). Similarly, in zebrafish, *Meis3* is involved with hindbrain patterning and differentiation (Choe et al., 2002; Waskiewicz et al., 2001) as well as proper development of the neural plate (Choe et al., 2009). These studies provide further evidence of the interaction between these developmentally important genes. *Meis3* expression has also been shown localized later in development in the somites, neural tube, and budding fin in zebrafish (Sagerström et al., 2001; Waskiewicz et al., 2001).

While all three *Meis* genes have been shown to be expressed in pancreatic cells, *Meis3* has a higher level of expression in pancreatic islet and β-cells. In these cells, *Meis3* regulates cell survival gene *Pdk1*. *Pdk1* is involved in the regulation of apoptosis and when *Meis3* is mutated pancreatic cell death is higher (Choe et al., 2009; diLorio et al., 2007). *Meis3* has also been shown to play a role in proper pancreatic development by enabling proper patterning of the anterior endoderm through regulation of the sonic hedgehog (*shh*) gene in zebrafish (Choe et al., 2009; diLorio et al., 2007).

Excluding *Meis4*, of which little is known, all three studied *Meis* genes have been shown to play significant roles and functions in a variety of cellular processes. While some of the functions involved are diverse, developmental processes are an underlining theme associated with the genes. During development it appears the *Meis* genes play roles in processes involving cellular proliferation and differentiation, especially during organogenesis (Agoston and Schulte, 2009). From *Meis1* playing roles hematopoietic processes to *Meis2* acting as a target for retinoic acid in developing limb buds, this family serves important roles that are continuing to be examined and studied. While many of the regulator roles *Meis* plays have been studied, the actual regulation of *Meis* genes themselves has been relatively unstudied and is an area of particular interest.

Mammalian genomes contain thousands of different genes and each gene is specifically regulated in a variety of ways. It could easily be assumed that in such complex organisms the assortment of these genes would be well organized. This is, however, not always the case. Instead, within a genome, genes are constantly assorted and organized through evolution (Morey et al., 2009). This assortment through evolution can cause pairs of genes, or even clusters of genes, to become physically linked. While this gene linkage is often between related genes that share similar functions, sometimes it is between completely non-related genes (Cajiao et al., 2004; Spitz and Duboule, 2008). This genetic linkage, and associated bystander effect, is currently being further examined as a possible transcription regulation mechanism, conservation vector, or simply a byproduct of transcriptional regulation.

In general, there are two main ways genes can be linked. Genes can be linked physically based on their location near each other on a chromosome and through shared regulatory regions. In both cases, it has been observed that this linkage is often maintained through evolution, signifying its genetic importance. To understand how evolution has shaped linked genes the different linkage conditions and origins must be examined (Akalin et al., 2009; Meyer and Schartl, 1999).

Linked genes often share similar functional characteristics, even identical functions in many cases, which is commonly attributed to gene duplication events (Yanai et al., 2001). Gene duplication events are responsible for major gene diversification and subsequently evolving organismal complexity (Meyer and Schartl, 1999). Data suggest that the base vertebrate genome has had two duplication events, resulting in a potential for up to four paralogs within a gene family, while many species, like *Danio rerio* have undergone additional duplications (Hittinger and Carroll, 2007; Itoh and Ornitz, 2004; Meyer and Schartl, 1999). It appears that there was a single genome duplication event before the deviation between deuterostomes and protostomes and another during the early emergence of vertebrates (Itoh and Ornitz, 2004).

Some examples of vertebrate gene clusters that formed as a result of genomic duplication include the *Hox*, *hemoglobin*, and *keratin* gene families (Force et al., 1999; Lemons and McGinnis, 2006). As a result of nondisjunction during meiosis, whole genome duplication events result in two copies of every gene, each with its own copy of the respective regulatory element sites (Spitz and Duboule, 2008). Additionally, gene clusters can undergo more common large scale duplication events, which result in a duplication of a region rather than the entire genome. The Fibroblast growth factors (*Fgf*) gene family is an example of a family of genes that have not only undergone the two major genome duplication events, but additionally many more smaller scale duplications. As a result, there are 26 related genes within the family, identified in mice and humans alone (Itoh and Ornitz, 2004).

Initially, each duplicated gene shares completely the same expression patterns and functions. Unless the amount of gene expression is detrimental, having redundant genes is not harmful (Force et al., 1999). While it was originally assumed that the duplicate genes could become functionally independent, research has shown that the majority of these genes become lost, due to repetitive function, or become sub functionalized (Hittinger and Carroll, 2007). Recent findings using comparative genomics has revealed the co-elimination of many functionally related genes that were previously gained through gene duplication events through evolutionary lineage, often times through the accumulation of mutations (Aravind et al., 2000; Force et al., 1999). Over time duplicate genes that remain can share functions or subdivide multiple functions between the genes. For example, the *Gal* gene family of *S. cerevisiae* includes *Gal1* and *Gal3*, which both share a very similar sequence and diverged from a common ancestral gene after a duplication event. Over time, two specialized and sub functionalized genes emerged from evolved regulatory sites, suggesting that changes in regulatory regions play a role in the fate of duplicated genes (Hittinger and Carroll, 2007).

On rare occasions, one duplicate gene develops a novel function while the other gene preserves the original function (Force et al., 1999). Driven by evolution, this genetic makeup contributes to adaptivity and genetic fitness. The preservation of these gene linkages between different species is referred as conserved synteny (Akalin et al., 2009; Kikuta et al., 2007). Maintaining conservation synteny is accomplished when the gene cluster provides advantages to the organism's fitness, mutations in one of the linked genes is fatal, or through unrelated bystander gene effects (Zurovcova et al., 2006). While gene duplication events can explain some of the related expression within related gene clusters, notably, it does not tell the entire story. Several gene clusters, in fact, show the opposite. The *KRAB* gene family, for example, is clustered together, presumably due to a distant gene duplication event; however, they do not share co-expression given their close proximity (Spitz and Duboule, 2008).

One of the most common ways genetic variation is introduced within a genome is through chromosomal crossover. During meiosis, homologous chromosomes interchange matching regions. This recombination results a novel arrangements of alleles. While in theory genes adhere to independent assortment, they instead have a recombination frequency based off their genetic distance, or physical proximity on the chromosome. This results in physically linked genes being inherited together and conserved through evolution. Since they are physically close to each other, the likelihood of them separating is significantly lower.

Until recently, it was commonly thought that many of these highly conserved regions were just clusters of genes that had not yet undergone any chromosomal rearrangement (Kikuta et al., 2007). One of the exceptions to this hypothesis has always been the *Hox* gene cluster. Important during development, functioning in the patterning and morphogenesis of the anteroposterior body axis, the *Hox* genes are an important gene family that has been extensively conserved (Lemons and McGinnis, 2006; Mann, 1997; Moens and Selleri, 2006; Sharpe et al., 1998). *Hox* genes are found in clusters, arranged in a colinear style, ranging from

4-48 genes within four complexes in tetrapods (Sharpe et al., 1998). Colinearity is a property that describes a direct correlation between the physical positioning of genes within a cluster to the expression patterns of the genes in axial positioning (Gould et al., 1997; Lemons and McGinnis, 2006; Sharpe et al., 1998). Recent data have suggested an increased diversity amongst varying organisms in the number of *Hox* genes, organization, and expression patterns; however, overall general arrangement characteristics have been conserved due to gene linkage conditions (Lemmons and McGinnis, 2006; Sharpe et al., 1998). In vertebrates, more so than *Drosophila*, the *Hox* gene cluster has remained linearly conserved through evolution with a more complex cluster and fewer non-related bystander genes (Gould et al., 1997; Mann, 1997; Moens and Selleri, 2006). This maintenance suggests strict evolutionary constraints. Studies have shown evidence that overlapping patterns of expression between *Hox* homologs may cause these constraints. In vertebrate models, shared regulatory elements and cross regulation are proposed mechanisms for overlapping *Hox* gene expression patterns (Gould et al., 1997; Hittinger and Carroll, 2007; Lemons and McGinnis, 2006; Sharpe et al., 1998). By cross regulating each other (up and down regulating), and additionally sharing regulatory elements, adjacent *Hox* genes within the collinearly arranged gene cluster have been shown to exhibit overlapping expression and conserved synteny (Gould et al., 1997; Mann, 1997; Sharpe et al., 1998).

Each gene duplication event previously mentioned, not only results in a replicate gene, but also the associated regulatory region. These regulatory regions on the nearby DNA are non-coding but instead act as binding sites for regulatory proteins, encompassing a region around a target gene known as a genomic regulatory block (Akalin et al., 2009; Bjerke et al., 2011). These areas generally contain areas known as enhancers, repressors, promoters, and insulators. As their names imply, enhancers and promoters act together to encourage transcription (accomplished by the binding of transcription factors to the enhancer and then to the

promoter) while repressors and insulators act to limit transcription of a target gene or its neighbors (Holliday and Pugh, 1990). All of these transcriptional binding sites play a role in the regulation of a target gene and while it may seem like they should be close in proximity to their target gene, this is not always the case. Some enhancers can be found thousands of base pairs away from their target gene upstream and downstream, acting long range (Akalin et al., 2009; Holliday and Pugh, 1990; Kikuta et al., 2007). These are termed distal enhancers while the area around a target gene that includes all of these binding sites is called the locus control region (Akalin et al., 2009; Cajiao et al., 2004; Tena et al., 2011). Since genomic regulatory blocks can span a large distance, other related or non-related genes can be contained within the regulatory block and are referred to as 'bystander genes' (Akalin et al., 2009). Additionally, over time, gene duplicates can lose a duplicated regulatory region and become regulated by only one, subsequently sharing a regulatory element.

The simplest of mechanisms for co-expression of neighboring genes is the sharing of these regulatory elements. This bystander effect usually consists of a shared enhancer or promoter region between two genes, that ultimately govern the expression of both genes simultaneously (Hittinger and Carroll, 2007; Morey et al., 2009). The most notable example of direct enhancer sharing are the *Hox* genes, while this occurrence has been observed even within the yeast genome (Hittinger and Carroll, 2007; Kmita et al., 2000; Mann, 1997). The clustered *Hox* genes are regulated by 'gobal enhancer sequences,' which enable the sharing of enhancers that in turn control multiple *Hox* genes simultaneously. This kind of gene regulation is thought to have aided in genomic stabilization, organization, and synteny within the clusters (Kmita et al., 2000; Mann, 1997). Likewise, the *Dlx* gene family is arranged as bigene clusters within mammalian species examined (Zerucha et al., 2000). This gene family plays important roles during early morphogenesis of the head region. Stemming from an original homeobox gene, there are now at least six *Dlx* genes in vertebrates due to gene duplication events (Sumiyama et

al., 2002; Zerucha et al., 2000). Notably, *Dlx* genes are closely linked to *Hox* genes and are thought to have undergone duplication together. Additionally, just like observations within the *Hox* gene family, *Dlx* genes have been documented to share enhancer sequences as well, exhibiting bigene control (Sumiyama et al., 2002; Zerucha et al., 2000). Yet in another example of enhancer sharing and cross regulation, the vertebrate *Iroquois* (*Irx*) gene cluster has been confirmed to utilized long range enhancer sharing throughout the gene cluster (Tena et al., 2011). Iroquois-class homeodomain proteins play an important role in embryonic patterning and have been highly evolutionarily conserved, including the gene cluster organization. Recently data suggests that the physical structure of a mutually used transcription factor enables enhancer sharing and subsequently co-regulation (Tena et al., 2011).

This sharing can result in polycistronic mRNA, where a single piece of RNA contains multiple coded genes in succession. These genes can be translated simultaneously, or the polycistronic mRNA can be alternatively spliced after transcription, leaving two strands each coding for a single gene (Holliday and Pugh, 1990; Spitz et al., 2003). This mechanism can also be accomplished through the use of a bidirectional promoter, activating both upstream and downstream of two different genes (Ebisuya et al., 2008). Another type of shared element mechanism is the use of a universal long range enhancer. In this situation a single 'promiscuous' enhancer is bound by transcription factors and goes on to bind to and activate all the promoters for different genes within its reach (Holliday and Pugh, 1990). This type of control is dependent on the distances between neighboring genes within its control radius, along with being limited by boundary elements. *α-Fetoprotein* (*AFP*) and *albumin* are adjacent genes on human chromosome 4 that display enhancer sharing. Both proteins are similar in structure and both show evidence of coordinated high levels of expression in the fetal liver. Remarkably, these expression patterns are even seen in cell lines displaying an inactive *AFP* enhancer (Jin et al., 1995). Data has demonstrated that these two proteins share three

enhancer sites to stimulate both respective promoters ensuing high levels of transcription, sometimes competitively (Jin et al., 1995).

Genes do not, however, have to share a regulatory element in order to be co-expressed as a bystander. Sometimes, in fact, co-expression of genes occurs within genes with absolutely no related function. This bystander effect is observed between many different neighboring dissimilar genes. The mechanism by which transcription factor proteins bind and regulate transcription is varied and is thought to often be a mixture of different mechanisms used specifically at different times or conditions. The main principle in the different mechanisms is the binding of transcription factors to the different binding domains and then clumping together to form complexes between the proteins, while excess DNA is looped out of the complex (Morey et al., 2009). There are differences in where the complex forms, whether in a proximal position to the target gene, far upstream or downstream, or even directly on top of the transcription start site, however, ultimately the complex migrates to the target gene to initiate transcription (Holliday and Pugh, 1990). The highly varied location and distance of all these transcriptional binding sites along the DNA, and their subsequent binding proteins, adds a level of complexity to the transcription model. Instead of having a single gene proximally flanked by its transcription factor binding sites, there are instead often multiple genes, along with their associated binding domains, all located within the same locus control region, or genomic regulatory block (Akalin et al., 2009; Cajiao et al., 2004; Morey et al., 2009). The close proximity of all of these genes and factors, along with the extended distance from which transcriptional control can be exerted can cause a 'ripple' interference effect between the transcriptions of neighboring genes (Ebisuya et al., 2008). This effect is known as a bystander effect. The effect this co-expression has on gene expression regulation along with how this effect contributes to evolutionary changes are important questions currently being explored (Morey et al., 2009).

At a basic level, a bystander gene is a gene that is transcribed as a byproduct of a nearby target gene's transcription (Kikuta et al., 2007). While the expression level of a bystander gene is not always as strong as the expression of the target gene itself, it is often accomplished in the absence of its own transcriptional binding sites and factors (Cajiao et al., 2004). In order for a gene to be transcribed a 'transcriptional factory' must be set up. This 'factory' includes both the target gene along with transcription factors, polymerases, and nucleotides (Holliday and Pugh, 1990; Spitz and Duboule, 2008). These are all important to transcription and must be present at the promoter/target gene. DNA, however, is bundled around histones, which serve to protect the DNA and serve to regulate transcription as insulators. In order for a gene to be transcribed some level of chromatin and histone modifications must be made to 'expose' the target gene to these transcriptional elements (Cajiao, 2004; Holliday and Pugh, 1990; Jaenisch and Bird, 2003). One of the proposed mechanisms for co-expression is that the modification and removal of histones in order to expose a target gene region to transcriptional elements results in the exposure of neighboring genes and elements to transcriptional machinery as well (Holliday and Pugh, 1990; Kosak and Groudine, 2004). Thus, neighboring genes are indirectly transcribed, often times at lower levels, as a byproduct of the chromatin structure being modified, opened up, and exposed for free transcription (Spitz et al., 2003). Likewise, another mechanism is explained by the pulling of a target gene out of its chromatin toward a 'transcriptional factory' within the nuclear matrix where it is then transcribed. In this case, neighboring genes are also pulled along as a byproduct where they are consequently transcribed in low levels based on their presence and proximity to the transcriptional factory (Kosak and Groudine, 2004).

One example of co-expression of unrelated genes was described by Isabela Cajiao and colleagues in 2004. They examined the human *IgB* gene, which is situated directly between the pituitary specific *hGH* gene and its locus control region. While the *hGH* gene is pituitary specific, the *IgB* protein is B-cell specific and its only known functional role is in B-cell receptor

signaling. Interestingly, *IgB* is found to be transcribed at high levels in the pituitary and in the complete absence of B-cell specific transcription factors. Instead of being shielded by regulatory elements within the *hGH* locus control region, it is activated by it. The mRNA, however, is never translated into functional protein within the pituitary, suggesting that the transcript is unnecessary, but rather transcribed as a byproduct of the transcription of neighboring *hGH* gene (Cajiao et al., 2004). This is an example of the bystander effect in action and is most likely explained by the proximal location of the *IgB* gene to the target *hGH* gene along with activation and bind of similarly related transcription factor proteins (Cajiao 2004). Other studies have confirmed the presence of these bystander genes that unrelated and share minimal expression or regulatory elements, yet are kept in synenty with a target gene (Akalin et al., 2009).

So if these co-expressed genes are functionally irrelevant, and in some cases never even translated, what has kept these genes together throughout evolution? This is the question researchers are trying to analyze in order to better understand gene order and expression regulation. It may be that the random arrangement of genes within the genome is so complex and entangled that is too difficult to rearrange without detrimental errors, essentially the synteny is conserved by proximity to an essential gene and its regulatory block (Akalin et al., 2009). It may also be cost effective for the cells to leave it be or too costly to evolve complex silencers, repressors, and boundary elements (Spitz et al., 2003). Yet is it also possible that bystander genes play additional roles in the ever complex gene expression regulation system. There are a few different mechanisms that contribute to the presence of bystander genes. It is important to keep in mind that there is not one clear mechanism, but instead multiple mechanisms are used by different cells, for different conditions and often times multiple mechanisms are used in conjunction with one another (Williams and Hurst, 2000).

While the activity of these genes have proven to be difficult to study, developing zebrafish embryos offer a good model with transparent, externally developed and fertilized embryos that can be easily genetically manipulated (Hunter et al., 2011). Previously, in our laboratory, a novel gene, *zgc:154061*, has been identified. The *zgc:154061* gene is always localized downstream of the *Meis2* gene and has been found highly conserved among all vertebrates examined (Fig. 2). In addition to its proximity to the developmentally important *Meis2* gene, the *zgc:154061* gene is highly conserved among examined organisms, suggesting evolutionary importance. The *zgc:154061* gene is 1914 nucleotides in length, has a coded sequence of 900 amino acids in zebrafish, and is organized in an inverted convergently transcribed manner (Carpenter, 2010; Graham, 2009) (Fig. 3). The translated protein sequence was predicted and aligned between humans (*Homo sapiens*), chickens (*Gallus gallus*), mice (*Mus musculus*), and zebrafish (*Danio rerio*) (Fig. 4). In order to begin characterizing the novel gene, the transcription and translation profile was previously explored both spatially and temporally. Using quantitative real time PCR, the transcripts were found to be maternally expressed initially at high levels until they decrease at 8 hours past fertilization before the zygotic genome begins and maintains expression at 12 hours past fertilization (Carpenter, 2010; Graham, 2009). Using *in situ* hybridization, the *zgc:154061* mRNA expression was shown to overlap significantly with *meis2.2* expression in zebrafish, with a more diffused overall expression during early development (Carpenter, 2010). We have also hypothesized that the observed absence of the *zgc:154061* next to *meis2.1* may be a result of non-functionalization post-gene duplication and as a result, it is possible that the *zgc:154061* gene is acting as a bystander gene, as previously mentioned, physically linked with the *meis2.2* gene in zebrafish (Nelson, 2011). Early protein and transcript expression profile characterization has also provided evidence for localization in the developing optic and central nervous system, further supporting the possibility of physically linkage as a bystander with the *meis2.2* gene (Carpenter, 2010; Cochrane, 2012). While the

zgc:154061 gene has been found to be highly conserved, its function and expression have not been fully characterized yet. The research offered in this thesis project serves to further characterize this previously novel gene, additionally adding support for a postulated bystander effect.

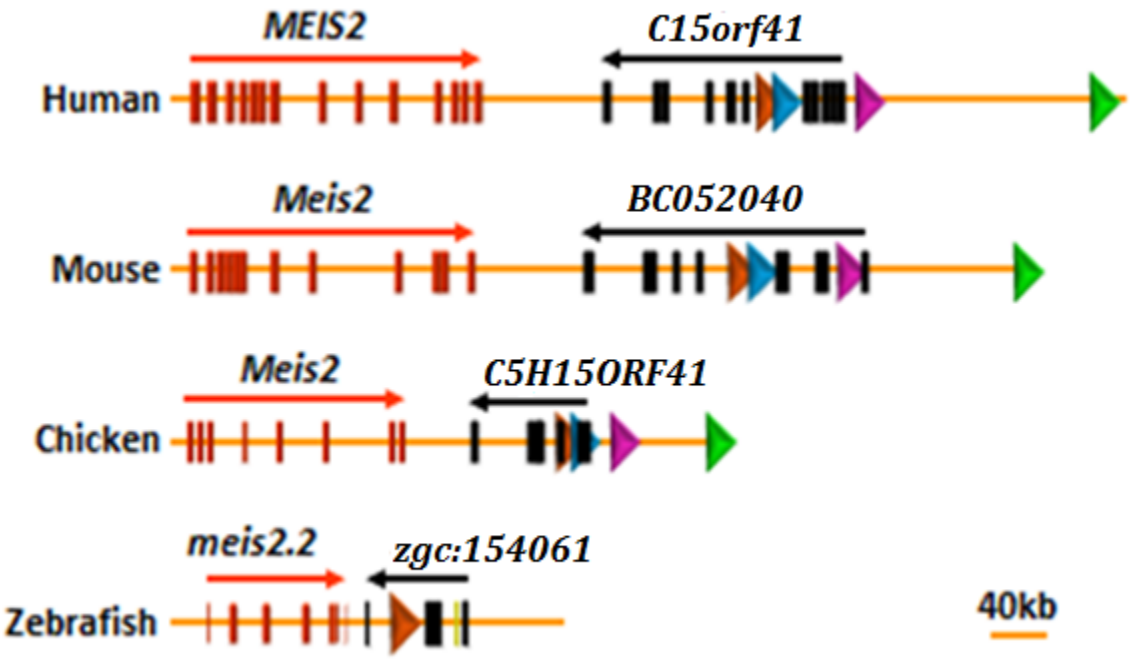

Fig. 2. Location of *zgc:154061* in relation to *Meis2* gene in human, mouse, chicken, and zebrafish genomes. The Meis2 gene is shown in red and upstream of the *zgc:154061* gene, shown in black, in all four species examined. Also shown are conserved non-coding elements (orange, blue, purple, and green arrowheads) possibly associated with the two neighboring genes. (Imaged borrowed from Carpenter, 2010)

gtcgcgtgtgctgcagaactggagcttgtacagtgttattggtgctgtttgtgttgcagagttgtgaaatatggacgcagtgtcaggcggcag
cgccacaggaacaggtgaacaggtgaataatctgaggatctgcagggcagaatacaggagcatcagcaggtttgtggagcagctgcgtc
ccacacggcagtgcatgaagaccctgcagacgcacttcacacatctccccgcgtccactctgctgagcatcttctcccaggagtaccagaa
gagaatgaagcgcagtatggccagacatcactctcctgaggtgctcagagtgtattatcagaggtatcgagatgaagctgagaccagagc
cacagaaccactgctgctggagctcgctaaccaggtggatctatctccagctcttttggctcgtctgatgctggagtgtttcctagaggagcg
caacgcttcagtcccttccagacaagtcctcaacaacatgctgcgtgagccgtatttaattccagatctggtgttagccaagcacatcgagca
gtgcacagtaaatgactgctgttatggaccgctggtcgactgcatcaaacatgccatcggtctggaacatgaagacactctgagagacaaa
ctcagagagaggaacctgtcgtttttagatgagaatcagctgcgggtcaaaggatacgacaaaaccccggacatcatcctggaggtgccg
atcgctgttgatggccacatcgttcactggatcgagagtaaagcttcatttggagatgatcacagtcacaacacatacctgaacgagcagtt
ctggagctactgcaacaggtttggtccgggtctggtcatctactggttcggcttcatctcagagctggactgccagcgggagcgagggatcc
tgctgaaggacggcttccccacggacatcagcagcctgtgtgcgggaccccagcgctgaggacggggcgtttctgatggcatgaggatcc
ggaagcaatttaagctttcctgcagccagagactctgacaaacatgtgctgtgaactccatctgaaccatctgtgtgagtgtgtgtgtgtgtg
tgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgctggttacactttatttgaagtctccttaactactatataggtac
atttcaactactcattgaaactgcacagtaaatgcaagtttttgtctggtgtttggtccagatatctgtgaactcttaaataaagaagcatttttt
ggatgtgcactgaatgtagtgttgttttcagaggtgatgtgctaaagtgaagtgagtttttaattagagcaggctgaataatctgccagtggg
ggaagcagaataatcttatgtgtaaaggaaaaagccagtgatttccctcaccccactggcggattattctgcttgttttgaggtaggactctc
tttattttggctcgttatttctgtaaacaacactgtattttatccctgtctataaaatgcttcctgatttaagaacgttcagatatttataccaaaa
acaagacgaactctctaagaatacgtcagttttactgcgttgtactaatattttgaaatcttttgcatgtaattacaatttttccgttaataatta
caccctctcttacaccttaacccactcttaacccttcccatatcactaaacctgtcaacaacccaaccccgatcccagattaatagcaccata
actgttctgcaatgcattataaacaggaggagtacattgtgctgatgttttgatggtagttattgtgacttcaaataaagtgtatgattgtgtta
catgctgatatggaataaataaagagtattcttcgtaaaaaaaaaaaaaaaaaaaaa

Fig. 3. Sequence of the *zgc:154061* gene. Sequence highlighted in green is the start codon, while the sequence highlighted in red corresponds to the stop codon. Sequence in orange represents binding site (on complementary stand) for the Dr-zgc154061-GST-5', while the sequence in green represents the binding site for the Dr-zgc154061-GST-3' primer. Sequence in highlighted in light blue represents binding site (on complementary stand) for the Dr-zgc154061-internal-5', while the sequence highlighted in yellow represents the binding site for the Dr-zgc154061-internal-3' primer (all primers used are detailed in Table 1). Total length of gene sequence is 1914 nucleotides.

Fig. 4. Multiple sequence alignment of the predicted protein sequence for zgc:154061 in zebrafish (Dr), mouse (Mm), human (Hs), and chicken (Gg). Zebrafish [Danio *rerio* (Dr), 300 aa (amino acids)], mouse [Mus *musculus* (Mm), 281 aa], human (Homo *sapiens* (Hs), 281 aa], and chicken (Gallus *gallus* (Gg), 272 aa] with identical conserved sequences in red blocks. The region encompassed within the black box was determined by Biosynthesis, Inc. (Lewisville, TX), to be the most probable region of the sequence to elicit a sufficient immune response in rabbit. The blue underlined region is the sequence, from zebrafish, that was synthesized by Biosynthesis, Inc., and used to inject into a New Zealand white rabbit for antibody production. (Image borrowed from Cochrane, 2012).

Materials and methods

Model Organism

For this research project the model organism used was *Danio rerio*, or zebrafish. These tropical freshwater fish belong to the minnow family and are an ideal organism for studying developmental biology due to their short maturation time, many offspring produced, rapid embryonic development, and fully sequenced genome. Zebrafish are also characterized by their large, robust and clear external embryos, which make them ideal for this study.

Zebrafish Husbandry

A Marine Biotech Z-mod closed system (Aquatic Habitats, Apopka, FL) was used to house zebrafish used for experimental purposes. This system was maintained at a constant water temperature of 27°C and programmed on an automatic 14 hour light/10 hour dark cycle. Daily water quality checks were performed to ensure pH was kept between 7.0-7.4 and conductivity was maintained between 500-600 mS/m. Genetically controlled zebrafish strains AB and AB* (Zebrafish International Resource Center, Eugene, OR) were maintained and used in the studies as well as non-genetically controlled wild-type strains (Carolina Biological, Burlington, NC). Up to six zebrafish were housed together in 1 L aquaria housing, with male to female ratios kept equal.

In order to harvest and utilize embryos for experiments, fish were routinely bred, always within the same genetic strain. Adult male and female zebrafish were divided overnight within 1 L aquariums by plastic dividers within specialized breeding chambers (Aquatic Habitats). Within an hour of the onset of the light cycle dividers were removed to allow breeding to occur. The fertilized eggs were separated from the fish through a plastic mesh

bottom, allowing efficient harvesting. Fertilized eggs were collected and rinsed repeatedly with reverse osmosis (RO) H_2O. If fertilized eggs were not immediately used, they were incubated in 1 X Danieau buffer [58 mM NaCl, 0.7 mM KCl, 0.4 mM $MgSO_4$, 0.6 mM $Ca(NO_3)_2$, 5 mM HEPES pH 7.6] at 27°C until the correct developmental stage was reached. Embryos raised to adulthood were relocated to a 1 L aquarium with approximately 500 mL 1X Danieau buffer. The developing zebrafish were fed twice a day with fine particulate dry food (Zeigler, Gardners, PA). Tanks were cleaned every other day using a turkey baster and replacing the 1 X Danieau with fresh buffer. After 20 days past fertilization, the larvae were introduced to the closed system on a slow drip of system water in a 1 L aquarium. Feedings twice a day continued as fish equilibrated to the system. Food particles used for feedings increased as the growth progressed (ZM-100, ZM-200, ZM-300, ZM-400; Zeigler) and embryos were separated into different tanks based on growth and size weekly. Larvae were also fed 2-day-old live brine shrimp (INVE Aquaculture, Salt Lake City, UT) daily, once feeding with ZM-200 dry food was begun. Approximately 3 months after fertilization, embryos reach adult size and are reproductively mature. At this point fish were divided into 1 L aquariums as previously described and began daily feedings with Zeigler Adult Zebrafish Complete Diet and 2-day-old live brine shrimp.

Gel Electrophoresis

An agar gel was prepared [1% w/v agar in 1 X TBE (89 mM Tris base, 89 mM Boric acid, 2 mM EDTA] and microwaved for 30 seconds before ethidium bromide (0.3 μg/ml) was added to the liquefied gel. The gel was allowed to cool to touch before being poured into a gel mold with a well comb. After the gel solidified, the comb was removed and the gel was submerged in ~300 mL 1 X TBE buffer within a Fisher Biotech mini horizontal electrophoresis unit (Fisher Scientific, Waltham, MA). Loading dye (6 X) was added to samples prior to well loading and electrophoresis was generally accomplished with 120 V for 75 minutes. Gels were visualized and imaged under UV light using an Alpha Innotech Fluorchem imager (San Leandro, CA).

Antibody Purification

A rabbit polyclonal antibody was generated against a small peptide portion of the predicted zgc:154061 protein sequence (Refer to Fig. 4; black outlined box containing the blue underlined predicted amino acid sequence for the peptide portion used for antibody generation). The small peptide was predicted to be optimal for eliciting an immune response by Biosnythesis, Inc. (Lewisville, TX) and purchased from the same company. The peptide was injected into a naïve New Zealand white rabbit (RSI Biotechnology, Mocksville, NC) in 2010 to elicit a secondary immune response (Cochrane, 2012). The pre-injection serum and post-injections serum was collected, aliquoted, stored at -80°C, and used for all subsequent experiments.

In order to purify the antibody from the serum collected, a NuncTM ProPur™ Mini Protein Purification kit for Protein A (Thermo Scientific, Rockwood, TN) was used according to the manufacturer's instructions. This kit first filtered the serum and then utilized a binding affinity column to capture IgG antibodies. Eluted IgG antibodies were recovered and stored at -20°C for future confirmation. In order to confirm antibody was purified, SDS-PAGE gel analysis was conducted. Filtered serum, diluted serum, wash steps, and eluted IgG antibody were run out on a gel following SDS-PAGE gel method listed. Antibody purification gels were Coomassie stained for visualization and correct band size confirmation.

Purified antibody samples were quantitatively analyzed using a Bradford assay (Pierce Biotechnology BCA Protein Assay Kit, Thermo Scientific, Rockwood, TN) using the provided Bovine Serum Albumin (BSA) as a standard according to the manufacturer's protocol. BSA concentrations used were as follows: 2000 µg/µL, 1500 µg/µL, 1000 µg/µL, 750 µg/µL, 500 µg/µL, 250 µg/µL, 125 µg/µL, 25 µg/µL, and 0 µg/µL diluted in RO H_2O. Samples were analyzed in triplicate using a plate reader (Soft Max Pro 5.2, Molecular Devices LLC, Sunnyvale, CA) for standard curve creation. Antibody eluate samples were also analyzed in triplicate both

undiluted and diluted in RO H_2O (1:10 and 1:25). Results were compared to the standard curve and average antibody concentration was obtained. Samples were stored in 5% glycerol at -20°C for further use.

Construction of a Recombinant GST-zgc:154061 Fusion Protein

A GST fusion construct was generated by subcloning the zebrafish *zgc:154061* gene in frame and downstream of a GST coding sequence using the pGEX-3X plasmid (Han and Colicelli, 1995). The expression vector used, pGEX-3X, contains a Glutathione S-transferase, or GST, coding region. When translated by the bacterium, this sequence will code for a GST protein that will be fused to the desired zgc:154061 encoded protein and can be further isolated.

A stock of *Escherichia coli* (DH5-α) bacteria was initially streaked for isolation on a fresh LB plate [2 mg/mL LB, 1.5 mg/mL Agar] and incubated at 37°C for 16 hours. Using an autoclaved toothpick, a single isolated colony was scraped and placed in a test tube containing 5 mL LB [2 mg/mL LB]. Culture was incubated for 16 hours at 37°C shaking and 1 mL of culture was added to a 50 mL conical tube containing 20 mL LB after the 16-hour incubation. The 20 mL culture was incubated at 37°C for approximately 50 minutes or until the absorbance reading (OD600) was ~0.6. Culture was then spun down using a bench top Legend XTR Sorvall centrifuge (Thermo Scientific, Rockwood, TN) at 4,700 rpm for 5 minutes at 4°C. The supernatant was discarded, pellet re-suspended in 11 mL ice cold 100 mM $CaCl_2$, and incubated on ice for 15 minutes. $CaCl_2$ suspended cells were then spun down at 4,700 rpm for 5 minutes at 4°C, supernatant removed, and pellet re-suspended in 1.5 mL ice cold 100 mM $CaCl_2$. Re-suspended cells were now chemically competent and were kept on ice between 30 minutes and 2 hours before 500 µL of 80% glycerol was added and competent cell solution was stored at -80°C in 100 µL aliquots.

Transformation of the expression vector was accomplished by initially adding 0.1 µg of the pGEX-3X plasmid (GE Healthcare) to a thawed competent cell aliquot. The cells and DNA

were incubated on ice for 15 minutes. The competent cells were subjected to heat shock by quickly removing the tubes from the ice and immediately putting them into a 42°C hot water bath for 65 seconds. After the heat shock, cells were immediately returned to the ice bath and incubated for 3 minutes before 1 mL of LB was added. Transformed cell cultures were incubated in the LB for 1 hour shaking at 37°C. Cells were plated (100 µL) on LB plates with ampicillin [2 mg/mL LB, 1.5 mg/mL Agar, 100 mg/mL ampicillin] and incubated at 37°C for 16 hours. After overnight incubation, plates were examined for isolated colonies. Using an autoclaved toothpick, a single isolated colony was scraped and placed, with the toothpick, in a test tube containing 5 mL LB with ampicillin [2 mg/mL LB, 100 mg/mL ampicillin]. Culture was incubated shaking for 16 hours at 37°C. Expression vector plasmid was purified using a Wizard Plus SV Miniprep DNA purification system and the manufacturer's supplied protocol. pGEX-3X plasmid miniprep samples were then measured using a NanoDrop® ND-1000 spectrophotometer to determine concentration (ng/µL) and purity (absorbance readings at 280/260 and 280/230 nm).

Next the plasmid vector was digested using enzymes to confirm correct isolation. Approximately 1 µg of the isolated plasmid (miniprep) was digested with *EcoR1* (10% v/v) and *Pst1* (10% v/v) in 1X NEB buffer 4 for 16 hours at 37°C. The digested plasmid was then confirmed to be the correct expression vector pGEX-3X using gel electrophoresis. After confirmation, ~ 10 µg of the plasmid was linearized with *EcoR1* [10% v/v *EcoR1* in 1X *EcoR1* enzyme buffer] at 37°C for 16 hours. Linearization of the plasmid was confirmed using gel electrophoresis the following day. In order to reduce self-annealing of the cut sites on the linearized pGEX-3X vector, the 5' phosphate groups were removed by treatment with alkaline phosphatase (CIP). CIP Phosphate (1% v/v) was added to the linearized plasmid sample and incubated at 37°C for 20 minutes before storage at -20°C.

After the expression vector was linearized, the insert was amplified using PCR. The insert was the *zgc:154061* gene, amplified from the pExpress clone (317.8 ng/µl; purchased from OpenBioSystems). Oligonucleotides were designed, specific to the 5' and 3' ends of the *zgc:154061* coding sequence and containing added *EcoRI* restriction sites at their 5' ends in frame with the *EcoRI* restriction site in pGEX-3X (primers Dr-zgc154061-GST-5' and Dr-zgc154061-GST-3'; Table 1). A PCR reaction [1.0 µl pExpress plasmid containing target sequence pExpress (317.8 ng/µl), 50pmol Dr- zgc154061-GST-5' primer, 50 pmol Dr-zgc154061-GST-3' primer, 20% v/v Phusion® 5 X HF Buffer, 10 mM dNTPs, and 1% v/v Phusion® High-Fidelity DNA Polymerase (NEB M0530L) in RO H_2O] was mixed, pulsed down in a thin walled PCR tube, and placed into a GeneAmp® PCR System 9700 (Applied Bioscience). The amplification conditions were: an initial DNA melting step at 98°C for 1:30 minutes; 40 amplification cycles (98°C for 30 seconds, 55°C for 30 seconds, and 72°C for 70 seconds); followed by 7 minutes at 72°C. The PCR reaction was maintained at 4°C until storage at -20°C.

Table 1. Oligonucleotide primers used for initially PCR amplification of the target gene, screening for gene presence, and sequencing confirmation of the constructed recombinant molecule.

Oligonucleotide Primers	
Name of Primer Used	*Primer Sequence*
Dr-zgc154061-GST-5'	gggaattcaatggacgcagtgtcaggc
Dr-zgc154061-GST-3'	atgaattccggatcctcatgcca
Dr-zgc154061-internal-5'	gtgcacagtaaatgactgc
Dr-zgc154061-internal-3'	gtcgactgcatcaaacatg

The amplified fragment was confirmed to be the correct size for the *zgc:154061* insert sequence using gel electrophoresis. Once PCR was confirmed, the product was purified using a Wizard® SV Gel and PCR Clean-Up System (Promega TB308) following manufacturer's instructions involving purification by centrifugation. Following PCR reaction clean up, ~ 5.0 µg of the PCR amplified insert was digested with *EcoR1* [10% v/v *EcoR1* in 1X *EcoR1* enzyme buffer] at 37°C for 16 hours in order to make ends compatible with the previously linearized pGEX-3X expression vector.

After both the PCR amplified insert and pGEX-3X expression vector were digested with *EcoR1* to create compatible ends, both samples were separately purified by phenol:choloform extraction. In separate microcentrifuge tubes, 100 mL of each sample was combined with 100 mL ice-cold phenol. Samples were vigorously vortexed for ~15 seconds and then centrifuged at 16,000x g for 5 minutes. The aqueous layer from each sample was then carefully transferred using a micropipette to a new microcentrifuge tube containing a 1:1 phenol:chloroform mixture [50 mL ice cold phenol and 50 mL chloroform:isoamyl alcohol 24:1], vortexed for 1 minute, and centrifuged at 16,000x g for 5 minutes. The aqueous layer was then carefully transferred to another tube containing 100 mL chloroform:isoamyl alcohol 24:1, vortexed, and spun at 16,000x g for 5 minutes. The wash step with chloroform:isoamyl alcohol 24:1 was repeated one time to ensure removal of all remaining phenol. After the second spin with chloroform:isoamyl alcohol 24:1, the aqueous layer was removed and the final volume was quantitatively measured using a micropipette. In a new microcentrifuge tube, the aqueous layer was combined with 7.5 M ammonium acetate (1/2 X the final aqueous layer volume) and 200 proof ethyl alcohol (2.5 X the final aqueous layer volume). The samples were vortexed for 30 seconds and stored at -80°C for 2 hours or -20°C for 16 hours. After cold incubation, samples were centrifuged at 16,000x g at 4°C for 10 minutes. The supernatant was removed following centrifugation and the visible pellet was washed twice with 70% ethyl alcohol. All remaining supernatant was carefully

removed and the tubes containing the pellet left were air dried on the bench top for 4-16 hours until pellet was completely dry. Each dried pellet was resuspended in 50 µL RO H_2O and each concentration quantified by NanoDrop® ND-1000 spectrophotometry.

In order to insert the purified and amplified *zgc:154061* sequence into the purified and linearized pGEX-3X (both with compatible 'sticky ends'), a ligation reaction was performed in a total volume of 10 µL. In a single centrifuge tube, 100 ng of the linearized pGEX-3X expression vector and 60 ng of the PCR amplified sequence insert were combined with T4 DNA ligase [10% v/v, New England BioLabs] in 1 X T4 DNA ligase buffer. The sample tube was thoroughly mixed using a micropipette and pulsed down with a microcentrifuge before incubating at 16°C for 16 hours or room temperature for 2 hours.

Chemically competent bacteria cells (*Escherichia coli* (DH5-α) prepared as previously described) were transformed with 5 µL (~80 ng) of ligated DNA. DNA was added to a thawed competent cell aliquot and incubated on ice for 15 minutes. The competent cells were subjected to heat shock by quickly removing the tubes from the ice and immediately putting them into a 42°C hot water bath for 65 seconds. After the heat shock, cells were immediately returned to the ice bath and incubated for 3 minutes before 1 mL of LB was added. Transformed cells were incubated in LB for 1 hour shaking at 37°C. Cells were plated (100 µL) on LB plates containing ampicillin [2 mg/mL LB, 1.5 mg/mL Agar, 100 mg/mL ampicillin] and incubated at 37°C for 16 hours. Following this incubation, plates were examined for isolated colonies. Using an autoclaved toothpick, a single isolated colony was scraped and placed, with the toothpick, in a test tube containing 5 mL LB containing ampicillin [2 mg/mL LB, 100 mg/mL ampicillin]. Culture was incubated for 16 hours at 37°C with shaking. Plasmid DNA was purified using a Wizard Plus SV Miniprep DNA purification system following the manufacturer's supplied protocol. Plasmid DNA was quantitatively measured using a NanoDrop® ND-1000 spectrophotometer.

In order to confirm the correct recombinant molecule was constructed and transformed into the bacteria, test digests were performed. Approximately 1 µg of the miniprep sample was digested in a 20 µL final volume with three different sets of enzymes in separate microcentrifuge tubes, *EcoR1/Pst1* [*EcoR1* (10% v/v) and *Pst1* (10% v/v) in 1X NEB buffer 4], *EcoR1/EcoRV* [*EcoR1* (10% v/v) and *EcoRV* (10% v/v) in 1X NEB buffer 3], and *Pst1* [*Pst1* (10% v/v) in 1X NEB buffer 4]. All digestion reactions were incubated for 16 hours at 37°C. The test digests of the plasmid were confirmed to exhibit the correct restriction patterns using gel electrophoresis (120 V for 70 minutes) then sequenced to further confirm (performed by Cornell University's Life Sciences Core Laboratories using internal primers [Dr-zgc154061-internal-5' and Dr-zgc154061-internal-3'; Table 1] for confirmation).

Induced zgc:154061 protein expression in bacteria

Plasmid DNA of *zgc:154061* in pGEX-3X was transformed into *Escherichia coli* strain BL21-DE3 bacteria for induced protein expression. Chemically competent BL21-DE3 bacteria were made using methods previously described. Plasmid DNA was transformed as previously described and plated (100 µL) on LB plates containing ampicillin [2 mg/mL LB, 1.5 mg/mL Agar, 100 mg/mL ampicillin] and incubated at 37°C for 16 hours. After overnight incubation, plates were examined for isolated colonies. Using an autoclaved toothpick, a single isolated colony was scraped and placed, with the toothpick, in a test tube containing 5 mL LB with ampicillin [2 mg/mL LB, 100 mg/mL ampicillin]. This culture was incubated for 16 hours at 37°C with shaking. This culture was then added to a flask containing 300 mL LB containing ampicillin [2 mg/mL LB, 100 mg/mL ampicillin]. This culture was incubated at 37°C for approximately 4-6 hours or until the absorbance OD600 reading was ~0.6. At the mid-log growth phase, (OD600 ~0.6) fresh IPTG (0.8 M) was aseptically added to the culture to induce protein expression. Incubation resumed at 37°C for 4 hours. An additional culture was

simultaneously incubated without the induction of IPTG. The cultures were then spun down using a Sorvall Instruments RC5C centrifuge at 3,500x g for 15 minutes at 4°C.

Protein Purification

Pelleted cell cultures, from the induced cultures were resuspended in 10 mL of a GST Binding/Lysis buffer [50 mM Tris-HCl (pH 7.4), 150 mM NaCl and 1% Triton X100] followed by the addition of 3 mg of lysozyme and 5 mM DTT to protect the protein from proteases. This mixture was incubated on ice for one hour. Using a Heat Systems Ultrasonics Processor (Farmingdale, NY) re-suspended cells were sonicated three times for 15 seconds, with 15 second breaks in an ice bath in between each sonication. The lysate was spun at 10,000x g for 30 minutes. The supernatant for both the induced and un-induced cultures was removed and stored at -20°C. Total protein extract from both cultures was examined with SDS-PAGE analysis and assessed for differences between the induced cultures and the un-induced using Coomassie staining and Western blot.

In order to isolate the target protein from the total protein extract, an affinity column was used to bind the GST portion of the fusion protein (GST-zgc:154061). Using a Pierce GST Spin Purification Kit, 5 mL aliquots of the induced target protein preparation was run through the GST affinity binding column following manufacturer's guidelines for Spin Purification of GST Tagged Proteins. Eluted fusion protein was confirmed with Coomassie staining based off sizing and Western blot analysis using the peptide antibody.

SDS-PAGE

First, a SDS-PAGE gel consisting of a 5% stacking gel [70% RO H_2O, 16.5% acrylamide: Bis 29:1, 125 mM Tris pH 6.8, 0.1% w/v SDS, 0.1% w/v ammonium persulfate (APS), 0.1% *N,N,N'N'* tetramethylethylenediamine (TEMED)] and a 12% resolving gel [32% RO H_2O, 40% 30% acrylamide:Bis 29:1, 390 mM Tris pH 8.8, 0.1% w/v SDS, 0.1% w/v APS, 0.04% TEMED] was prepared between two glass plates with a spacing of 1 mm. A 1mm comb was used and

removed after gel was solidified and rig was placed in a 1X Tris-Glycine Running Buffer [25 mM Tris Base, 192 mM Glycine, 3 mM SDS].

Protein samples were prepared by adding 10 µL of each sample to 5 µL 3 X Laemmli loading dye [240 mM Tris pH 6.8, 6% sodium dodecyl sulfate (SDS), 30% Glycerol, 0.16% β-mercaptoethanol, 0.006% w/v Bromophenol Blue]. Protein samples and dye were then heated to 70°C for 5 minutes for denaturing followed by placement in an ice bath. Solutions were then loaded on the SDS-PAGE gel and electrophoresis was performed at 240 volts for 90 minutes. After the electrophoresis was complete, the stacking gel was cut off using a razor blade. Gels were then either used for Western blotting or Coomassie staining.

Gels to be Coomassie stained were first placed in 100 mL RO H_2O after the stacking gel was removed, microwaved for 30 seconds, and washed in RO H_2O for 5 minutes at room temperature with shaking. After 5 minutes on the shaker, the water was replaced by fresh RO H_2O, microwaved for 30 seconds, and placed on the shaker for another 5 minutes. The water was then replaced with 100 mL Coomassie stain [70 mg Brilliant Coomassie Blue, 1 L RO H_2O, 3 mL hydrochloric acid (HCl)], microwaved for 10 seconds, and placed on the shaker for 16 hours at room temperature. The Coomassie stain was removed following the overnight incubation, replaced with RO H_2O, and the gel placed back on the shaker for destaining. Gels were imaged using an Alpha Innotech FluorChem imager.

Western Blot

In order to determine antibody specificity, Western blots were performed using protein samples isolated from bacterial extracts. Following electrophoresis, SDS-PAGE gels had stacking gels removed and were placed into a glass dish with 250 mL 1 X Tris- Glycine Transfer Buffer [48 mM Tris Base, 39 mM Glycine, 1.3 mM SDS, 20% (v/v) Methanol]. Polyvinylidene difluoride (PVDF, Thermo Scientific, Rockwood, TN) membrane was cut to the size of the gel (5 cm x 9 cm) and placed in 100% methanol for 30 seconds for priming. The PVDF membrane, gel,

and 4 pieces of 1 mm filter paper (cut to same size as membrane) were all incubated at room temperature, shaking in the transfer buffer for 20 minutes. Using a Trans-Blot® SD Semi-Dry Transfer Cell (Bio-Rad, Hercules, CA), protein samples embedded in the SDS-PAGE gel were transferred to the PVDF membrane. The protein bands were permanently fixed to the solid support membrane by placing the membrane on top of the gel and in between two pairs of filter papers and applying 15 volts for 20 minutes.

Following transfer electrophoresis, the PVDF membrane, containing the transferred protein bands, was placed in 50 mL blocking solution [5% Blotto in 1 X Phosphate Buffered Saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4)]. The membrane was incubated in blocking solution at 4°C for 16 hours or room temperature for 2 hours with gentle shaking. After blocking, the membrane was removed from solution, rinsed with RO H_2O, and placed in a primary antibody solution [1:100 purified anti-zgc:154061 antibody diluted in 5% blotto/1 X PBS blocking solution]. The membrane was incubated with gentle shaking for either 16 hours at 4°C or 2 hours at room temperature. Following incubation with primary antibody, membrane was removed from solution, rinsed with RO H_2O, and placed in 100 mL wash solution [0.1% Tween 20/1X PBS] with gentle shaking for 5 minutes. Washing was repeated three times, at room temperature, using fresh wash solution every 5 minutes to remove unbound antibodies. The membrane was then placed protein side up in a 50 mL conical tube. A secondary antibody solution [1:2500 Goat anti-rabbit alkaline phosphatase conjugated secondary antibody (Bio- Rad, Hercules, CA) in 10 mL 5% Blotto blocking solution] was added to the conical tube and set on rotation at 4°C for 16 hours, or room temperature for 2 hours. After incubation with the secondary antibody solution, the membrane was removed from the conical tube and rinsed using forceps and RO H_2O. The membrane was subsequently washed three times for 5 minutes in 100 mL wash solution [0.1% Tween 20/1X PBS] with gentle shaking.

Following the wash steps, and post-incubation with the secondary antibody, the membrane was placed protein side up on a piece of Saran™ plastic wrap. Any bubbles were removed and the membrane flattened before adding approximately 2 mL of Immun-Star™ AP Substrate (Bio-Rad, Hercules, CA) directly onto the membrane and incubating at room temperature for 10 minutes. The substrate solution was then poured off the membrane and the plastic wrap was folded over the membrane, and then cut to size. Membranes were exposed to x-ray film for 1-5 minutes and developed using a Konica Minolta SRX-101A developer (Konica Minolta Medical & Graphic, Inc., Shanghai, China).

Immunohistochemistry

Developing embryos at the 24, 36, and 48 hours past fertilization (hpf) stages were examined using immunohistochemistry (IHC) experiments. Both whole mount and 15 µm cross-sections of embryos were used to determine zgc:154061 protein localization within the embryo. Embryos were first manually dechorionated using No.5 Dumount forceps (Dumostar, Williston, VA) and then fixed in 4% paraformaldehyde (PFA) [1.3 M PFA, 0.002 N NaOH, 10% v/v 10 X PBS (1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4)]. Approximately 50 embryos were fixed at a time in 1.5 mL microcentrifuge tubes and incubated at room temperature with gentle shaking. PFA was replaced after 10 minutes with new 4% PFA and incubated at 4°C for 16 hours. Embryos were then washed three times with 1 X PBS and gentle shaking for 10 minutes.

Fixed embryos at the 24 hours past fertilization stage or older were depigmented for proper colorimetric visualization. Fixed embryos were placed in a petri dish containing approximately 5 mL of a H_2O_2 solution [3% H_2O_2, 1% KOH] for at least 30 minutes. Embryos were observed until depigmentation was complete and then individually removed from solution and placed into a 1 X PBS solution using a glass pipette. Depigmented and fixed embryos were dehydrated for storage if not immediately used for IHC experiments. To

dehydrate, embryos were washed for 10 minutes in a 25% methanol/1 X PBS solution shaking, followed by a 10 minute wash in a 50% methanol/1 X PBS solution. Embryos were then washed in a 75% methanol/1 X PBS solution for 10 minutes followed by a 10 minute shaking wash in 100% methanol. Embryos were then stored in a new 100% methanol solution at -20°C.

Embryos were rehydrated in groups of 10 embryos per well using 12 well plates. After being removed from the methanol using a glass pipette, embryos were placed in a 1 ml 75% methanol/1 X PBS solution with gentle shaking. After 5 minutes each, solutions in each well were replaced with 1 ml of 50% methanol/1 X PBS, 25% methanol/1 X PBS, and 100% 1 X PBS respectively. Embryos were then washed in new 1 X PBS for 30 minutes at room temperature. 1 X PBS was replaced after 30 minutes with freshly made, filter sterilized blocking solution [1% v/v dimethyl sulfoxide (DMSO), 5 mg/mL BSA, 0.8% v/v Triton, 10% v/v 10 X PBS (1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4)]. Embryos in blocking solution were either incubated for one hour at room temperature or 4 hours at 4°C. The blocking solution was removed and replaced with a primary antibody solution [1:100 purified anti-zgc:154061 antibody diluted in blocking solution] and incubated at 4°C for 16 hours. Primary antibody was substituted with pre-immune serum samples, under the same conditions, for negative controls. Following primary antibody incubation, embryos were washed for 2 hours in a wash solution PBS/DMSO/Triton [1% v/v dimethyl sulfoxide (DMSO), 0.8% v/v Triton, 10% v/v 10 X PBS (1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4)] with fresh solution changes every 30 minutes at room temperature. After the forth wash, the embryos were placed in a secondary antibody solution [1:1000 Goat anti-rabbit alkaline phosphatase conjugated secondary antibody (Bio- Rad, Hercules, CA) in 0.3% 1 X PBS/Triton X-100] and incubated at 4°C for 16 hours. Following incubation with the secondary antibody, embryos were washed three times with 1 mL 1 X PBS for 5 minutes at room temperature before being incubated with AP detection

buffer [10% v/v 1 M Tris-HCl (ph=9.50), 0.1 M NaCl, 0.1 M $MgCl_2$] for 15 minutes. Fresh AP detection buffer was replenished every 5 minutes.

Embryos were developed and visualized colorimetrically using BCIP/NBT substrate solution [0.66% v/v NBT (70% v/v N,N, Dimethylformamide, 50 mg/mL NBT) and 0.33% v/v BCIP (50 mg/mL BCIP in N,N, Dimethylformamide) in AP detection buffer). Embryos were first removed from the AP detection buffer and placed in 500 µL of the substrate solution in separate wells. Embryos were observed under a dissecting microscope until colorimetric visualization was detected. Embryos were then individually placed into a new petri dish containing distilled water to stop color development and the distilled water was replaced with fresh distilled water every 5 minutes until colorimetric development was completely stopped.

IHC embryos at 24, 36, and 48 hours past fertilization were imaged as whole mounts and embryos at 48 hours past fertilization were also imaged as cross sections. Whole mount embryos were removed from the distilled water and placed into an agarose solution [8 mg/mL agarose in 10 mL 1X Danieau buffer]. The agarose solution was prepared by bring the agarose and Danieau buffer to a boil in the microwave and then storing at 4°C. From storage, the hardened solution was liquefied and maintained in a boiling water bath. Using a Leica $mz7_5$ dissecting microscope, approximately 500 µL of the heated agar solution was place on a deep-dish microscope slide followed by a single embryo using a glass pipette. Prior to the agarose solution solidifying, the embryos were positioned correctly for imaging using needle probes. Whole mount slides were inverted and imaged using either a Zeiss Confocal Laser Scanning Microscope 510 (Leica Microsystems, Buffalo Grove, IL) or an Olympus IX81 (Shinjuku, Tokyo, Japan) inverted bright field microscope.

Embryos at the 48 hpf stage were also examined via cross sectional analysis. An agarose gel [1.5% agarose and 5% sucrose in 1 X PBS] was prepared and kept in boiling water bath to remain liquefied. Colorimetrically developed embryos were removed from the distilled

water in pairs and placed in mini petri dishes with approximately 1.5 mL of the agarose gel solution. Prior to gel solidification, the embryos were individually positioned with the anterior regions facing downwards with needle probes. Gels were allowed to cool and solidify and then placed in 4°C for 30 minutes. Gel blocks (5 mm x 5 mm) containing the positioned embryos were sliced and removed from the petri dish gel and placed in a large petri dish containing a sucrose solution (30% sucrose in 1 X PBS). The suspended gel blocks were left at 4°C for 16 hours.

Gel blocks containing embryos in the sucrose solution were then placed in aluminum foil wells [made by wrapping small pieces of aluminum around the base of a marker pen, taping around cup formed, and cutting height down to 1 cm]. The blocks were positioned so the anterior region facing down was closest to the base of the well and in the center of the well. Optimal Cutting Temperature (O.C.T.,Tissue-Tek®, Torrance, CA) compound was placed around the block within the aluminum well, and the well was placed on a block of dry ice to freeze. Completely frozen wells containing embryos were stored at -80°C. Embryos were subsequently cut into 15 µm cross-sections from the tip of the head region to the hindbrain. Sections were permanently bound to VistaVision™ HistoBond® Adhesive Slides (VWR, Radnor, PA) and sectioned using a Leica CM-1100 Bench Cryostat (Leica Microsystems, Buffalo Grove, IL). Slides were stored at -20°C and imaged using a Zeiss Confocal Laser Scanning Microscope 510 (Leica Microsystems, Buffalo Grove, IL) on bright field settings.

Immunoprecipitation

The zgc:154061 protein was isolated from zebrafish embryo protein extract using immunoprecipitation techniques. Approximately 200 embryos (1.5 mL), at the 2 hours past fertilization stage, were harvested as previously described. Embryos were combined in a microcentrifuge tube and all excess solution was removed before being flash frozen in liquid nitrogen and stored at -80°C. Frozen embryos were combined with 1 mL of a homogenization

buffer [250 mM sucrose, 30 mM Tris, 1 mM EDTA, pH 7.8, 5 µL proteinase inhibitor cocktail (PIC), and 10 µL phenylmethylsulfonyl fluoride (PMSF)] before being homogenized using the Tissue Tearor™ Homogenizer (Cole-Parmer, Vernon Hills, IL) for two 30 second applications on ice. Homogenized samples were microcentrifuged for 10 minutes at 4°C at 16,000x g and the resulting supernatant was stored in 200 µL at -20°C.

The concentration of the total protein in the supernatant samples was quantified using a Bradford assay (Pierce Biotechnology BCA Protein Assay Kit, Thermo Scientific, Rockwood, TN) as previously described. Target zgc:154061 protein was isolated from quantified protein extract samples using a Pierce Co-Immunoprecipitation Kit (Rockford, IL). Briefly, 50 mL of a resin slurry was incubated with 50 µL primary antibody solution [purified anti-zgc:154061 protein as previously described] for bead-antibody coupling. A 1.2 mL volume of the quantified protein extract (2 hpf) was incubated with the coupled antibody beads at 4°C for 16-24 hours. Using a resin column, unbound proteins are washed away with three washes followed by three elution steps of the bound protein using elution buffer. Elution samples of the target protein were stored at -20°C and analyzed via SDS-PAGE. As previously described, protein samples embedded in the SDS-PAGE gel were transferred to a PVDF membrane, blocked with Blotto [5% blotto/1 X PBS blocking solution], overnight at 4°C with shaking, and incubated with a primary antibody solution [1:100 purified anti-zgc:154061 antibody diluted in 5% blotto/1 X PBS blocking solution] for 2 hours at room temperature. After the washing solutions previously described, following secondary antibody incubation, the membrane was placed in AP detection buffer [10% v/v 1 M Tris-HCl (ph=9.50), 0.1 M NaCl, 0.1 M MgCl2] for 30 minutes. The membrane was then placed in a heat sealed plastic bag (3" x 4") and 10 mL of a BCIP/NBT substrate solution was added [0.66% v/v NBT (70% v/v N,N, Dimethylformamide, 50 mg/mL NBT) and 0.33% v/v BCIP (50 mg/mL BCIP in N,N, Dimethylformamide) in 1 X AP detection buffer). After all bubbles were removed, the bag was sealed on all four ends, placed on a

shaker, and incubated for 15 minutes to 16 hours based on colorimetric development. Reaction was stopped in 100 mL RO H2O. The membrane was dried on a paper towel for examination and imaging using a scanner.

Results

Construction and Expression of a Recombinant Molecule

In order to further study and characterize the novel *zgc:154061* gene, target protein expression was induced in bacteria using a constructed recombinant molecule. By producing the *zgc:154061* gene's protein product in bacteria, the size of the protein and specificity of the antibody current being used could be assessed. A target recombinant molecule was first constructed and transformed into bacteria for induced expression and subsequent target protein isolation (Fig. 5).

First, the expression vector, pGEX-3X, was transformed into bacteria from a small stock DNA sample, isolated, and confirmed using a restriction digest and gel electrophoresis (Fig. 6). The presence of two bands at ~4,000 and ~1,000 base pairs (bp) correspond to the expected restriction pattern for an *EcoR1/Pst1* double digestion (3,991 bp and 961 bp). In order to be used as the expression vector for the recombinant molecule, the pGEX-3X was linearized using *EcoR1* enzymes. Gel electrophoresis confirmed linearization of the plasmid occurred based on a 5,000 base pair length band, without the presence of multiple uncut conformation bands (Fig. 7). The linearized sample was then purified prior to the ligation being set up.

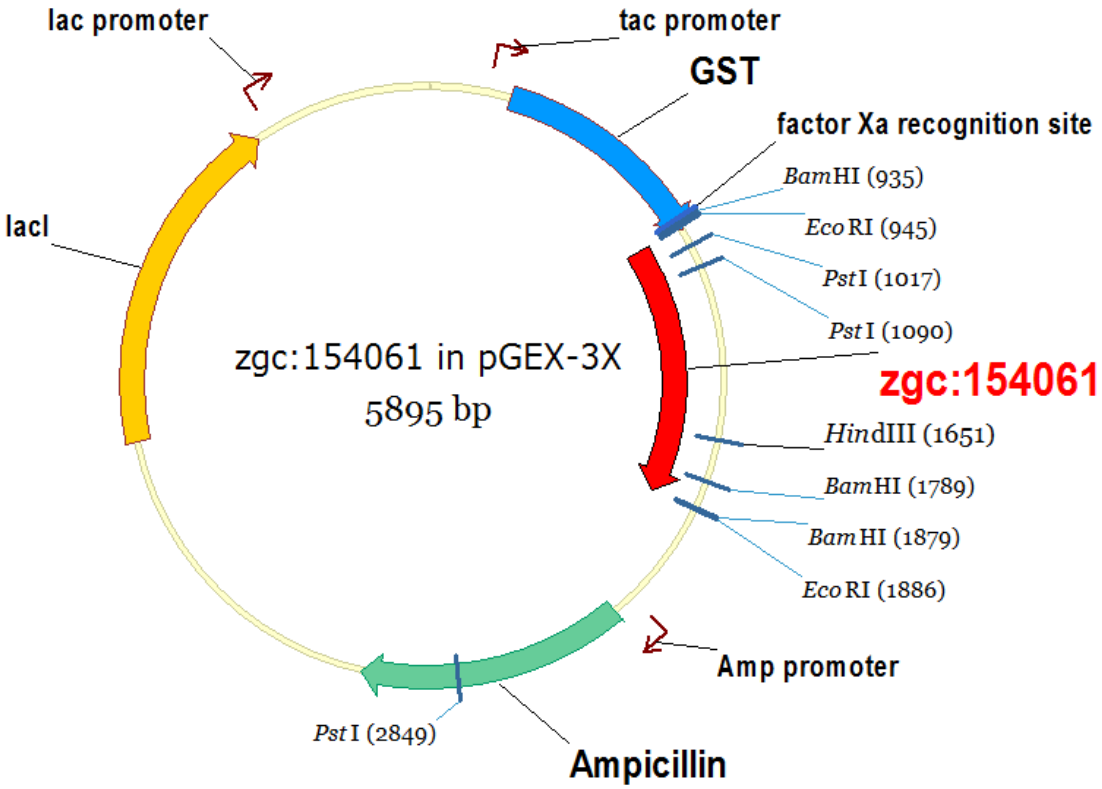

Fig. 5. Target GST-zgc:154061 Recombinant Molecule. A target recombinant molecule was constructed after insertion of a PCR amplified *zgc:154061* gene sequence into a linearized pGEX-3X expression vector. After transformation into bacteria, a GST- *zgc:154061* fusion protein expression was induced. The presence of an ampicillin resistance gene enabled ampicillin use as a selective agent. The numerous enzyme cut sites within the constructed plasmid allowed for proper plasmid identification via restriction analysis.

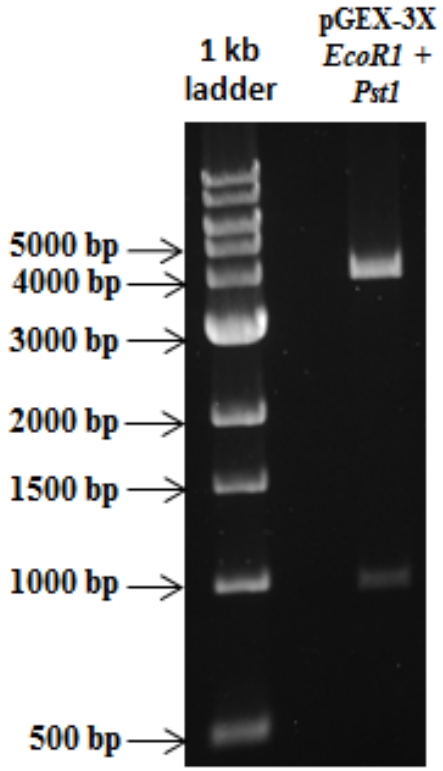

Fig. 6. Restriction analysis of purified pGEX-3X expression vector. The two bands observed in the right lane correspond to the expected sizes (3,991 bp and 961 bp) for the double digestion of pGEX-3X (with *EcoR1* and *Pst1*) confirming the correct molecule was purified.

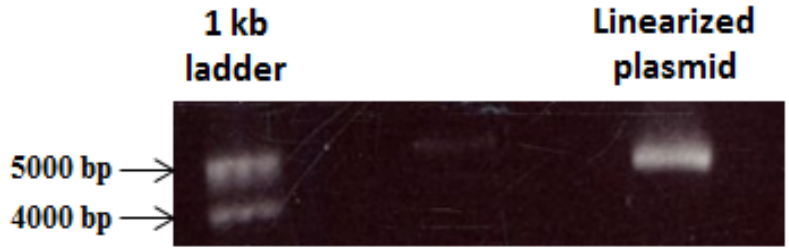

Fig. 7. Linearization of the pGEX-3X expression vector. pGEX-3X was digested with *EcoR1* overnight and analyzed via gel electrophoresis. The only band in the right lane correctly corresponds to the expected 4,952 base pair band of the linearized pGEX-3X vector.

Next, the insert containing the target *zgc:154061* gene sequence, was amplified from a previously constructed and isolated plasmid (*zgc:154061* in *pExpress*) containing the gene of interest. PCR amplification of the target gene area on the donor plasmid, using the primers Dr-zgc154061-GST-5' and Dr-zgc154061-GST-3' (Table 1), was performed and purified using a Wizard® SV Gel and PCR Clean-Up System (Promega TB308). Each end of the PCR product was made compatible with the linearized pGEX-3X vector via *EcoR1* enzyme digestion and confirmed using gel electrophoresis (Fig. 8).

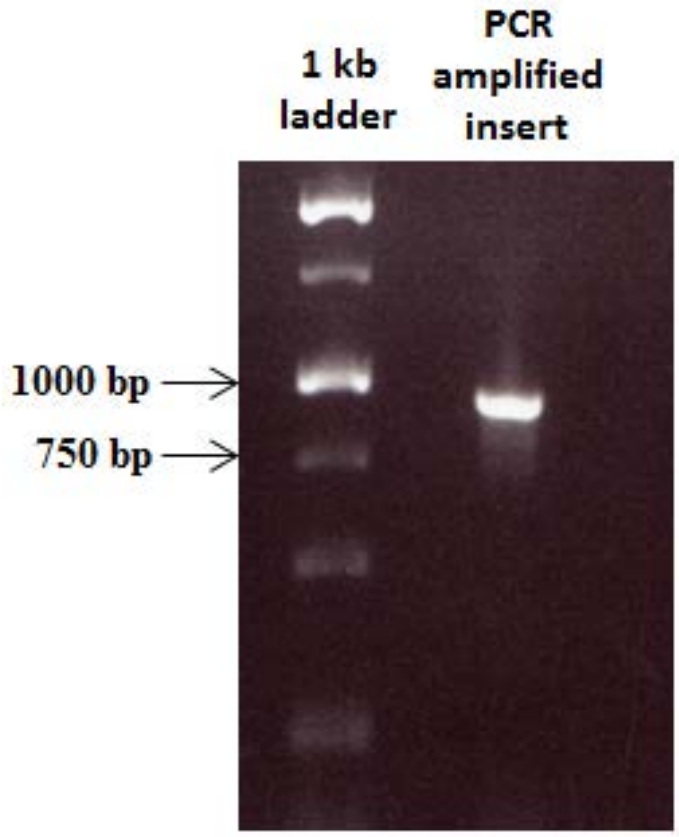

Fig. 8. PCR amplification of the *zgc:154061* gene sequence. The target gene sequence was PCR amplified, purified, and digested with *EcoR1*, yielding compatible ends for insertion into the vector. The presence of a single band (at ~900 base pairs) in the right lane corresponds to the desired 943 base pair amplified sequence.

A single band observed in the gel image (right lane) correctly corresponds to the desired target gene sequence (943 base pairs). Once the linearized vector and the PCR amplified insert were isolated, made end compatible, and purified, a ligation reaction was performed and transformed into bacteria.

Transformed bacteria were plated on selective media containing ampicillin and isolated colonies were screened for the correct recombinant molecule. Colony plasmid DNA was prepared from candidate colonies and separately digested with two different enzyme combinations (*EcoR1* and *EcoRV* or *Pst1*) to confirm that the insert was the correct size. DNA samples from two colonies matched the desired restriction pattern when visualized using gel electrophoresis (Fig. 9).

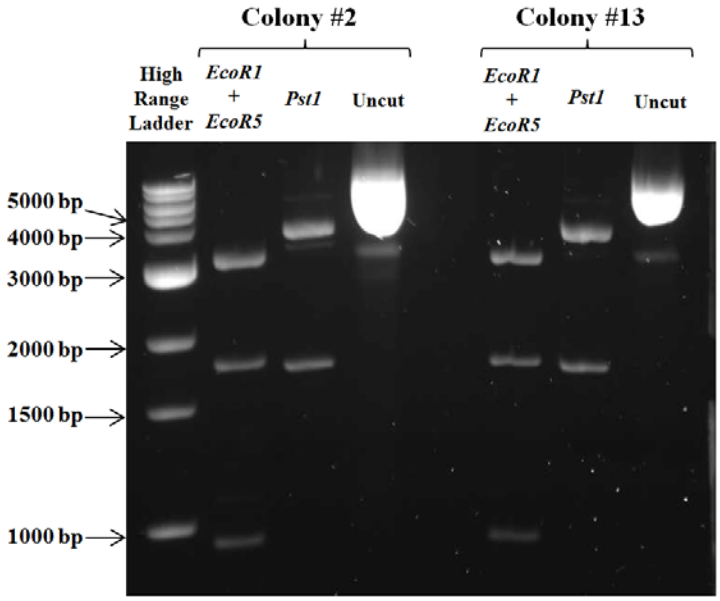

Fig. 9. Confirmation of the construction of a target recombinant molecule. Plasmid DNA samples from two transformed colonies were digested with *EcoR1/EcoR5* and *Pst1* enzymes for restriction analysis. Three bands observed in each double digestion lane (3,158 bp, 1,794 bp, and 943 bp) along with the presence of two bands in each *Pst1* digest lane (4,136 bp and 1,759 bp) are as predicted if the the target recombinant molecule was correctly constructed and isolated.

Plasmid DNA samples isolated from two different colonies (#2 and #13) exhibited restriction patterns that match the target recombinant molecule. The double digestion (*EcoR1/EcoR5*) of both samples yielded three distinct bands at ~3,200 bp, ~1,800 bp, and ~1,000 bp matching the expected 3,158 bp, 1,794 bp, and 943 bp size bands respectively for the target recombinant molecule. The two bands observed in each of the single digest (Pst1) lanes at ~4,100 bp and ~1,800 bp match the expected 4,136 bp and 1,759 bp size bands respectively. Both isolated samples were confirmed to be the correctly constructed target recombinant molecule via restriction analysis and sequencing. Sequencing results were compared to the known *zgc:154061* gene sequence and confirmed to be identical.

The confirmed recombinant molecule (isolated from colony #13) was then transformed into *Escherichia coli* strain BL21-DE3 bacteria for induced expression. Cultures at mid-log phase and 500 mL volume were chemically induced during the mid-log growth phase and the total protein from the bacteria was isolated via sonication and centrifugation. The target fusion protein produced from induction (GST - *zgc:154061*) was isolated from the total protein using an affinity-binding column. Samples were run on a SDS-PAGE gel and transferred to a membrane. Western blot analysis was performed on the protein samples affixed to the membrane with a primary antibody against the target *zgc:154061* protein. Antibody binding was detected using fluorescence and visualized using X-ray film (Fig. 10). The membranes containing the protein samples were stripped and Western blot analysis was repeated using an antibody specific for GST. Antibody binding was again detected using bioluminescence and visualized using X-ray film (Fig. 10).

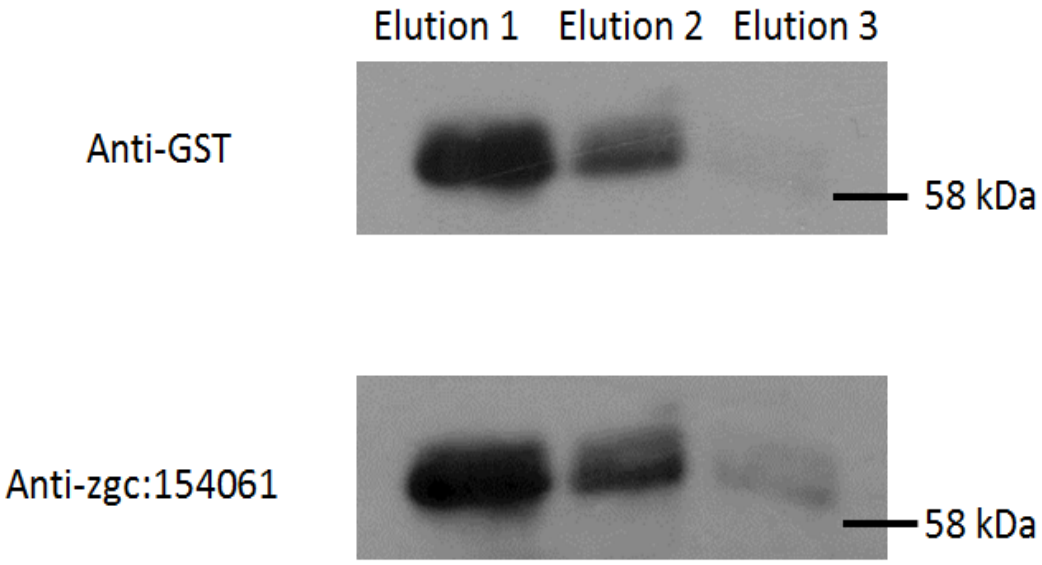

Fig. 10. Western blot analysis of isolated target fusion protein. Target fusion protein was isolated using an affinity-binding column, eluted, and separated via SDS-PAGE. Samples were transferred to a membrane and first incubated with a primary antibody recognizing the target *zgc:154061* protein. The bottom gel image displays binding by the antibody recognizing the target protein in the isolated fusion protein. The membrane was stripped and incubated a second time with a primary antibody recognizing the GST protein in the fusion protein. Bands at the exact same spot displayed binding by the antibody recognizing GST. The same band detected with the two different antibodies indicates the presence of the target fusion protein in all three elution lanes.

In all three protein elution lanes, the same sized bands were experimentally detected after binding and visualization of two different antibodies, one that recognized the target *zgc:154061* protein and another that recognized Glutathione S-transferase (GST). The desired induced protein was a fusion protein (GST - *zgc:154061*) with an estimated total protein size of 60.4 kDa, correctly matching all observed bands. The same exact band detected with the two

different antibodies at the correct estimated size indicates the presence of the target fusion protein in all three elution lanes.

Immunoprecipitation

Embryos were collected at the 2 hours past fertilization stage and the embryos were tissue homogenized. Isolated total protein from the homogenization was exposed to agarose beads cross-linked with anti-*zgc:154061* antibodies. Bound protein was eluted and analyzed via SDS-PAGE analysis using Coomassie staining (Fig. 11). Eluted protein sample bands were observed in all three elution lanes at ~58 kDa. No other bands were observed in each elution lane, however, bands were not strong enough for mass spectrometry analysis.

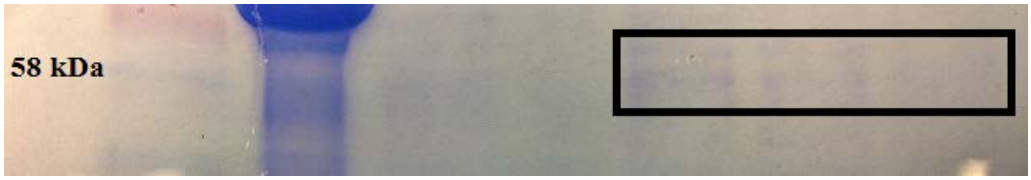

Fig. 11. Immunoprecipitation of the *zgc:154061* protein. Immunoprecipitation of the zgc:154061 protein from homogenized zebrafish embryos using agarose beads cross-linked with anti- *zgc:154061* antibodies. The black box highlights three lanes (eluted protein samples) shows faint bands corresponding to the eluted *zgc:154061* protein.

Immunohistochemistry

In order to further characterize *zgc:154061* gene expression, the spatial and temporal expression of the gene's protein product was explored using immunohistochemistry on both whole mount and cross sections of zebrafish embryos. Embryos at the 24, 36, and 48 hours past fertilization stage were harvested, depigmented, and fixed before being exposed to primary

antibodies against the target *zgc:154061* protein. The presence of the primary antibody, using a secondary antibody with an alkaline phosphatase enzyme tag, was detected colorimetrically at the three different stages (Figs. 12, 13, 14).

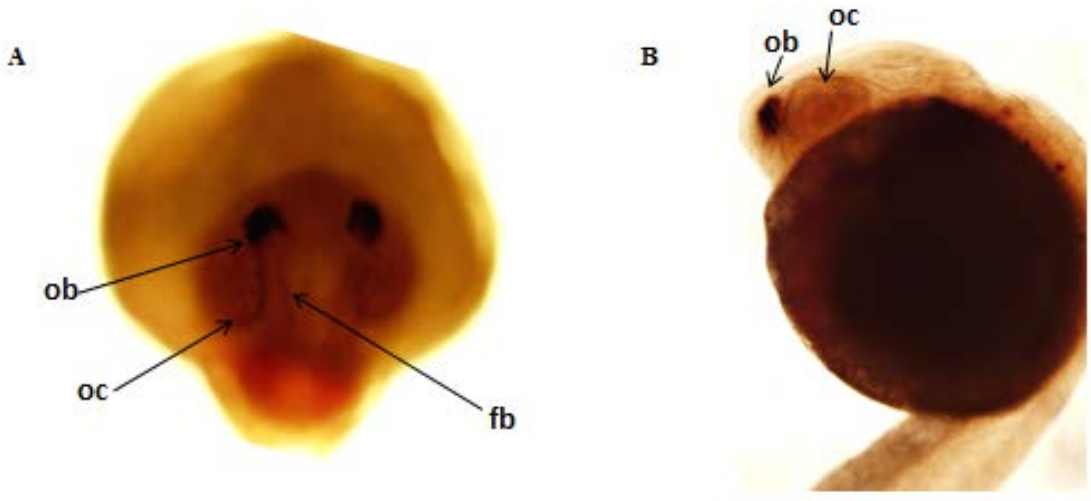

Fig. 12. Immunohistochemistry and colorimetric detection of the *zgc:154061* protein using whole mount embryos at 24 hours past fertilization. *zgc:154061* protein expression was observed to be localized in the developing optic region, brain, and olfactory bulb **(A-B).** Two orientations were utilized for viewing the embryos: a cranial (anterior head-on) orientation in which the anterior portion of the zebrafish head was visible (A) and a sagittal orientation in which the lateral areas of the zebrafish head could be analyzed (B). fb, forebrain; ob, olfactory bulb; oc, optic cup/retina.

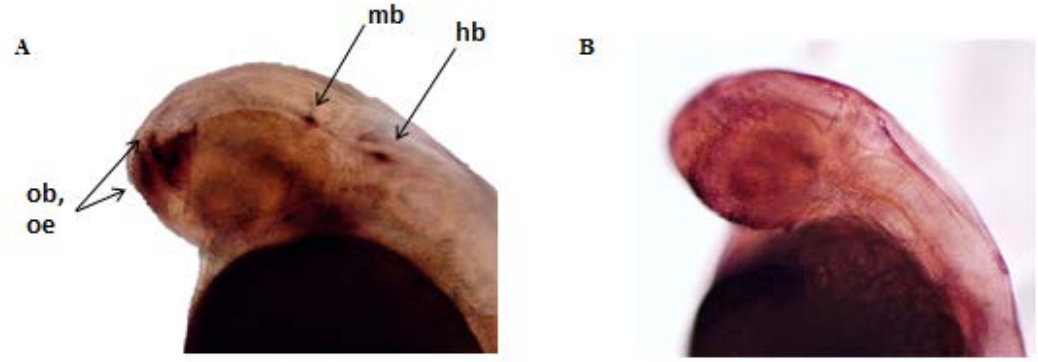

Fig. 13. Immunohistochemistry and colorimetric detection of the *zgc:154061* protein using whole mount embryos at 36 hours past fertilization. *zgc:154061* protein expression was observed to be localized in the developing brain and olfactory bulb and epithelium **(A).** Negative controls **(B)** were performed using pre-immune serum in lieu of a primary antibody. One orientation was utilized for viewing the embryos: a sagittal orientation in which the lateral areas of the zebrafish head could be analyzed. mb, midbrain; hb, hindbrain; ob, olfactory bulb; oe, olfactory epithelium.

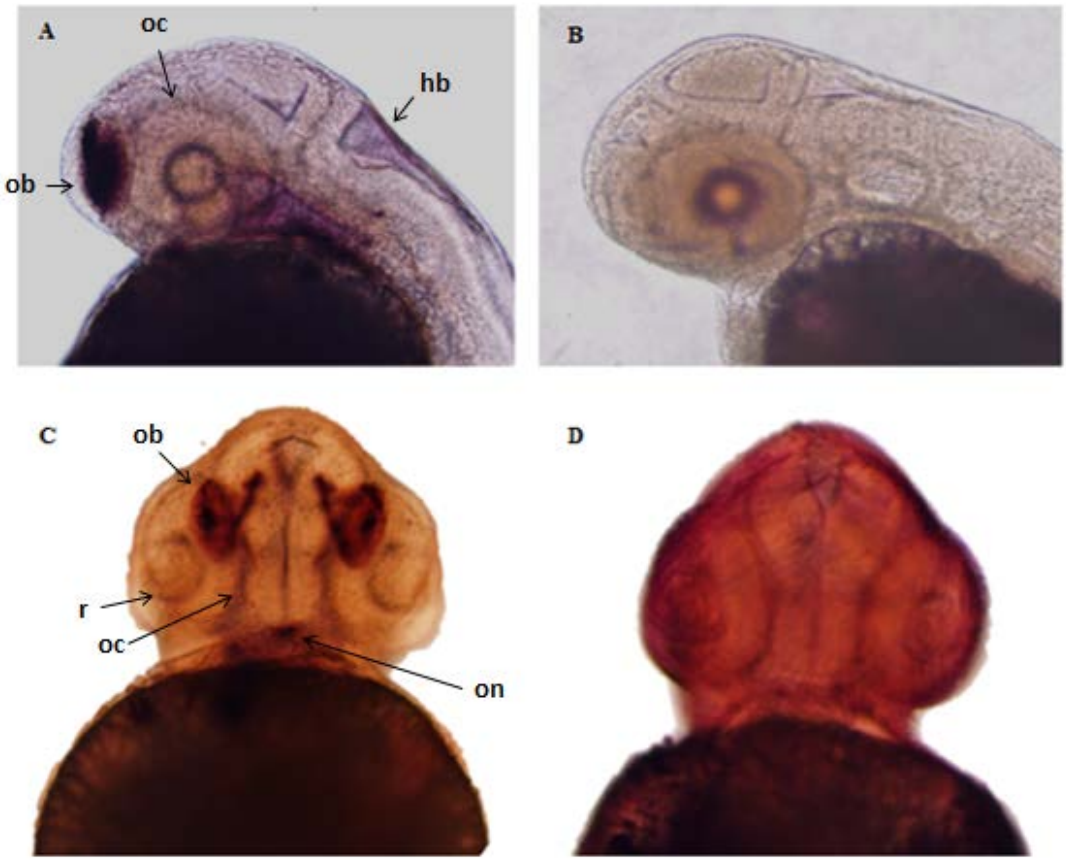

Fig. 14. Immunohistochemistry and colorimetric detection of the *zgc:154061* protein using whole mount embryos at 48 hours past fertilization. *zgc:154061* protein expression was observed to be localized in the optic region, developing brain, developing olfactory bulb **(A, C).** Negative controls **(B, D)** were performed using pre-immune serum in lieu of a primary antibody. Two orientations were utilized for viewing the embryos: a sagittal orientation in which the lateral areas of the zebrafish head could be analyzed (A-B) and a ventral (bottom-up) orientation in which the ventral portion of the zebrafish head was visible (B). hb, hindbrain; ob, olfactory bulb; oc, optic cup; on, optic nerve; r, retina.

At all three developmental time points examined (24, 36, and 48 hours past fertilization), *zgc:154061* protein expression was localization of the anterior region of the embryo. Specifically, expression observed throughout each time point was localized to the optic region, along the developing optic nerve, optic cup, and retina. Expression was also observed in the developing brain region, more often in the forebrain and midbrain. Pronounced expression in a distinct anterior structure was always observed. This structure has been identified as the developing olfactory bulb and epithelium.

In order to further explore target protein expression, especially expression localized in the unidentified region, embryos at 48 hours past fertilization were also examined using cross-sectional immunohistochemistry analysis. Embryos were subject to immunohistochemisty colorimetric analysis prior to being frozen in a Optimal Cutting Temperature compound and subsequently sectioned into cross-sections from the tip of the head region to the hindbrain using a bench cryostat. Cross-sections were bound to slides and imaged using microscopy (Fig. 15).

As the cross section analysis moves anteriorly from the hindbrain to the front of the zebrafish head, protein expression is observed diffused in the forebrain and developing optic area while pronounced expression is observed in the developing olfactory bulb and epithelium.

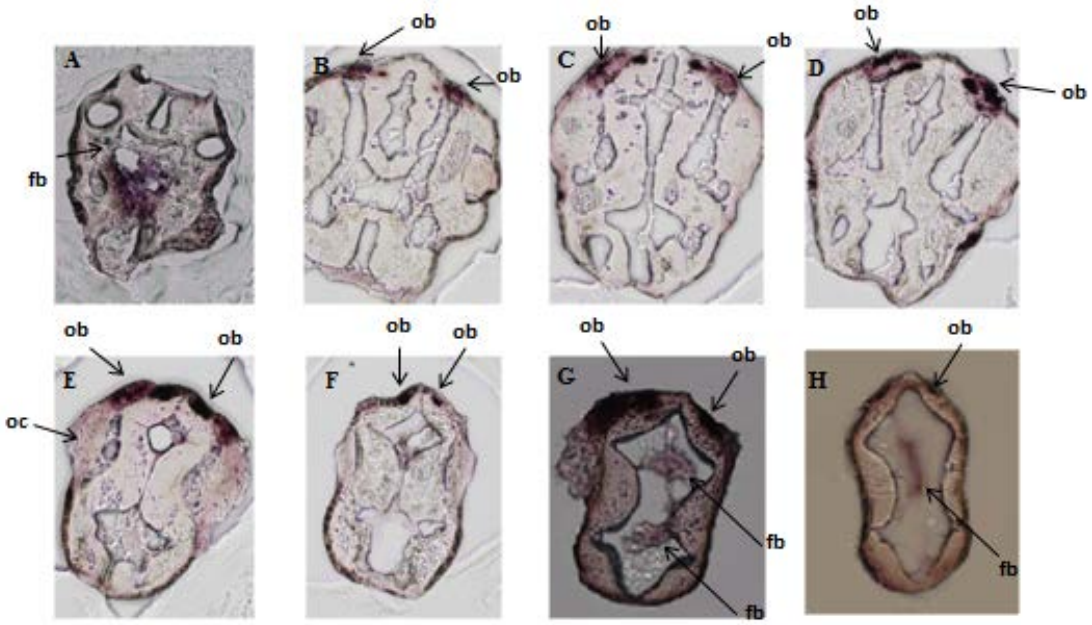

Fig. 15. Immunohistochemistry (IHC) and colorimetric detection of the *zgc:154061* protein using cross sections of whole mount embryos at 48 hours past fertilization. Protein expression was localized in the optic region, developing brain, and pronounced in an unidentified distinct anterior structure **(A-H).** fb, forebrain; ob, olfactory bulb/epithilium; oc, optic cup.

Discussion

While recent sequencing projects have identified over 20,000 protein-coding genes within the human genome, numerous genes have not yet been characterized. Many of these genes are highly conserved among species suggesting their evolutionary importance and enabling research using ideal model organisms. In this study, we further characterize the expression profile of the novel zebrafish *zgc:154061*gene. Previous work in the Zerucha lab provided the groundwork for the project allowing for further characterization. Gene *zgc:154061* has been found to be present in an inverted, convergently transcribed orientation downstream of the *Meis2* gene (*meis2.2* or *meis2a* in zebrafish), in all vertebrates examined using publically available genome data. Specific highly conserved regions of the gene sequence were also previously examined, however, none of the presumed protein domains have an identified function. The high level of conservation among divergent species observed coupled with the conserved positioning near the *Meis2* gene suggested possible importance (Carpenter, 2010; Graham, 2009).

When characterizing a gene, both the transcription and translation of the gene must be explored spatially and temporally. Prior experiments showed *zgc:154061* was transcribed early throughout development, with pronounced maternal expression at fertilization that gradually decreased until 8 hpf. Transcription of the gene was activated again at approximately 12 hpf and expression was observed throughout the neural tube before becoming restricted to increasingly anterior regions of the neural tube (Carpenter, 2010). In order to explore the protein expression of the gene, an antibody was generated against a small peptide portion of the predicted zgc:154061 protein. The protein expression matched the transcription profile of

the gene showing strong bands in Western blots at 0-12 hfp. Cross-sectional immunohistochemistry experiments also showed localization of the protein to the developing optic area at 48 hours past fertilization (Cochrane, 2012).

The next step in further characterizing the novel gene was expanding the protein expression profile. Developmentally important genes are sensitive to specific timing, concentration, and localization. Alternations in gene expression early in embryonic development are often detrimental or fatal, thus understanding the functional protein better can provide insight into the importance of a gene's role. Previous work with the protein's expression was done with an antibody that was generated against the small peptide portion of the predicted zgc:154061 protein and the antibody was used for determining temporal expression. The Western blot results, however, indicated that the protein was consistently 58 kDa in size as opposed to the predicted 34.4 kDa size. It could be hypothesized that this size discrepancy is due to posttranslational modification to the protein but could also be due to nonspecificity of the antibody. The specificity of the antibody was confirmed using a peptide competition experiment, however additional confirmation was required (Cochrane, 2012).

Confirming Antibody Specificity

In order to further confirm the specificity of the antibody, we created a recombinant molecule to express *zgc: 154061* in a bacteria system. By expressing *zgc:154061* in bacteria, any posttranslational modifications observed in eukaryotes should not occur. Linearization of the vector and amplification of the insert was properly confirmed. After extensive troubleshooting, the ligation reaction successfully yielded the desired recombinant molecule: the pGEX-3X vector containing the gene sequence in frame. This molecule was successfully transformed into bacteria and protein expression was induced. The target fusion protein was isolated and was confirmed using immunohistochemistry with the primary antibody previously mentioned. The size of the bands bound by both the anti-GST antibody and anti-zgc:154061 antibody were the

expected size of 60.4 kDa for the fusion protein produced. While this size is similar to the size of the unexpected experimental protein previously recognized by the anti-zgc:154061 antibody (58 kDa) in Western blot analysis of protein extracted from zebrafish embryos, it should be noted that this is due to the correct 34.4 kDa size zgc:154061 protein being linked to a 26 kDa sized Glutathione S-transferase (GST) protein.

Thus the specificity for the antibody to recognize the correct target protein was further confirmed. In order to further examine the size discrepancy, immunoprecipitation experiments using the primary antibody exposed to total protein from homogenized embryos. After eluting the bound protein and analysis using Coomassie stain and Western blot revealed the antibody recognized again a protein at 58 kDa. This result mirrored the prior results observed while characterizing the temporal protein expression. It appears that the protein *in vivo* is approximately 24 kDa larger in size then the predicted protein size confirmed by inducing protein expression in bacteria. We hypothesize that post-translational modifications *in vivo* cause the size differences experimentally observed.

Possible Post-translational modifications

Posttranslational modifications largely account for the human proteome containing over 1 million different proteins coded by only 20,000 genes within genome, along with the diversity, functionally and physically, observed within the proteome (Jensen et al., 2002; Jensen, 2004; Kamath et al., 2011; Wang et al., 2014). Post-translational modifications are a type of epigenetic mechanism, meaning they are heritable without directly affecting DNA sequences (Adamson, 1992; Jaenisch and Bird, 2003; Wood et al., 2009). These modifications have been documented to play roles in thousands of different key cellular control mechanisms, processes, and pathways, and in a plethora of mechanisms, both spatially and temporally (Jensen, 2000; Kamath et al., 2011; Kroger et al., 2001).

Enzymes recognize target proteins and use their cleaving, conjoining, and catalyzing abilities to add or remove functional groups, subunits, and side chains to or from proteins, subsequently changing the functional diversity of a protein by physical alternation (Kamath et al., 2011; Mann and Jensen, 2003). These chemical modifications influence the functionality of each protein by affecting localization, interaction, life span, and activity. Post-translational modifications can be broadly categorized in two categories, modifications that covalently add a chemical/functional group or modifications that cleave a protein side group (Kamath et al., 2011). As a result, post-translation modifications led to discrepancies in estimated protein mass and observed protein mass, supporting the possibility of a post translational modification causing the ~24 kDa increase in size from experimentally examined zebrafish protein samples compared to estimated target protein size (Jensen, 2004, Wilkins et al., 1999). In addition to diversifying protein possibilities, the mechanisms of post-translational modifications, as a whole, are also being researched for the role they may play in carcinogenesis. It was originally postulated that carcinogens might disrupt the structure of enzymes responsible for employing important modifications in repressors (Hancock, 1978).

In order to discern which post translation modification was responsible for the mass discrepancy experimentally observed, the average change in mass was examined for common post-translational modifications. Only two modifications result in a similar size change, formylation modifications (~28 kDa) and ethlyation modifications (~28 kDa) (Mann and Jensen, 2003).

Formylation is the addition of a formyl functional group to select protein side chains, which consists of a carbonyl bonded to a hydrogen, by formylphosphate. Recently a novel modification, formylation, was identified in human cell lines and mice using mass spectrometry techniques (Wood et al., 2009). In addition to adding an average of 28 kDa to a target protein's mass, formylation has been observed in all kingdoms, as a modification for all amino acids, and

at any position within a protein (Wilkins et al., 1999). As a mechanism to protect the protein from degradation by proteases, a formyl group is added to the N-terminus of the protein for protection (Schoenafinger et al., 2006). This unique modification is believed to arise post-oxidative damage and appears within the histone bundles wrapped around DNA. It is believed that oxidative damage triggers a reaction forming the secondary modification acting in interference of other modifications to that site and possibly preventing the utilization of the damaged DNA's binding sites (Wisniewski et al., 2007). The fact that this specific modification is only observed in histones, and possibly as a protection mechanism, adds support for this modification being an endogenous mechanism (Jiang et al., 2007). It is also thought that the formylation modification can compete in an epigenetic mechanism against normal chromatin function, possibly playing a role in diseases (Wisniewski et al., 2007).

Another less common modification has also been induced in vitro, ethylation (Xing et al., 2008). Ethylation, like formylation, is an adduct modification, adding an ethyl to the protein's side chains along with an additional ~28 kDa. This was confirmed with *in vitro* ethylation of aspartate and glutamate documented. It was notable, however, that identifying ethylation modification is difficult since it results in a similar size increase as that aforementioned formyl or two methyl additions (Xing et al., 2008). Ethylation modification events have been association with carcinogenesis by possibly altering a repressor for a target gene (Hancock, 1978).

Additionally, ethylation has been previously postulated to play a role in protein binding interference, regulating the DNA binding mechanisms of the human estrogen receptor (Obourn et al., 1993). Using mass spectrometry and labeling techniques, ethylation modifications were recently specifically identified within the human lens cortex. Ethylation modifications were observed on modified crystallin structures, in a relatively large number of proteins mainly in the cortical region (Asomugha et al., 2010).

It is possible that the zebrafish zgc:154061 protein is being modified with the addition of a formyl or ethyl group to the protein's side chains, although there is still a ~4 kDa size difference. Additionally, the size change could be due to a glycosylation event. Glycosylation, the addition of a carbohydrate glycan group to side chains, however, can vary immensely in size of the addition. The absence of proper protein glycosylation has also been linked with many developmental defects, specifically within the developing nervous system (Freeze et al., 2012). To further add to the complexity of this topic, many proteins undergo multiple modifications, at both the same positions within a lifespan, at varying positions simultaneously, or even in overlapping residues. They are also subjected to reversible modifications frequently, thus adding to the overall regulatory possibilities and mechanisms (Prabakaran et al., 2012). It is therefore possible that the protein undergoes a combination of post-translational modifications in order to reach the observed protein size. It will be beneficial in the future to analyze the zebrafish zgc:154061 protein via mass spectrometry to characterize the possible post-translational modifications.

Zebrafish zgc:154061 Protein Localization

Prior work in the Zerucha lab showed that the *zgc:154061* zebrafish gene is transcribed early during development, with ubiquitous expression early and more localized expression in the anterior and optic region in the later stages. Just because a gene is transcribed, however, does not necessarily mean it is always translated. This is especially the case with genes important during development. Developmental genes are often transcribed in quantity in order to have mRNA readily available, and on demand, for faster translation, thus enabling further gene control.

Using Western blot analysis, it was previously observed that the zebrafish zgc:154061 protein is translated throughout early development (2-12 hpf), but the localization needed to be further explored. Using whole mount embryos and cross sections, immunohistochemistry was

performed at 24, 36, and 48 hours past fertilization. Zgc:154061 protein expression was found to be localized to the anterior region of the developing embryos, specifically in the brain and optic region, partially overlapping with *meis2.2* expression in the developing zebrafish brain and optic region (Waskiewicz et al., 2001). This overlapping expression suggests the possibility of a shared regulatory element and should be examined in future experimentation. These results also matched the previous localization experiments, as well as the transcription localization specific to the brain and optic region observed late in development.

The protein expression observed in an anteriorly expressed distinct structure, however, was unexpected. Using the Atlas of Zebrafish Development, the anterior localization of the zgc:154061 protein is consistent with the location and approximate shape of the developing olfactory epithelium and bulb (Bryson-Richardson et al., 2011; Mueller and Wullimann, 2005). The olfactory epithelium is composed of neurons with receptors that recognize odors, relaying the information to the olfactory bulb. Together, the olfactory bulb and epithelium are located in the most frontal part of the brain, in the inferior region, and act to relay the perception of odors to the brain. Experimental results support the protein's localization in this distinct anterior structure at 24, 36, and 48 hours past fertilization, along with the expected developing brain, optic cup, retina, and optic nerve. Additionally, recent research has shown that Meis2 cooperates with Pax6 in the developing olfactory bulb, within the subventriucular zone, functioning in neurogenesis and cell fate specifications (Agoston et al., 2014). Meis2 was also previously used as a broad marker of olfactory bulb interneuron subtypes after it was found that Meis2 is expressed in the developing olfactory interneurons during development, specifically in *Dlx5/6*-derived periglomerular cells (Allen et al., 2007; Warclaw et al., 2009). Future experimentation could include examining additional time points and co-staining with structural markers.

The work done in this study serves to further the characterization of a previously novel gene, the zebrafish *zgc:154061* gene. Experimentation characterizing the expression profiles of the transcript and expressed protein indicate a role in the developing nervous system; overlapping *meis2.2* expression possibly as a bystander gene or through shared regulatory elements. The possibility of the zebrafish zgc:154061 protein undergoing post-translation modifications was further supported and more experimentation will need to be done to elicit the correct modification and function. Further research will also provide insight into expression at more temporal points along with determining the proteins function through overexpression and gene knockout or knockdown experiments.

Bibliography

Adamson, P., Marshall, C. J., Hall, A. and Tilbrook, P. A. (1992). Post-translational modifications of p21rho proteins. *J. Biol. Chem.* **28**, 20033-20038.

Affolter, M., Marty, T. and Vigano, M. A. (1999). Balancing import and export in development. *Genes Dev.* **13**, 913-915.

Afonja, O., Smith, J. E., Cheng, D. M., Goldenberg, A. S., Amorosi, E., Shimamoto, T., Nakamura, S., Ohyashiki, K., Ohyashiki, J., Toyama, K. and Takeshita, K. (2000). MEIS1 and HOXA7 genes in human acute myeloid leukemia. *Leukemia Res.* **24**, 849-855.

Agoston, Z. and Schulte, D. (2009). Meis2 competes with the Groucho co-repressor Tle4 for binding to Otx2 and specifies tectal fate without induction of a secondary midbrain-hindbrain boundary organizer. *Development* **136**, 3311-3322.

Agoston, Z., Heine, P., Brill, M. S., Grebbin, B. M., Hau, A. C., Kallenborn-Gerhardt, W., Schramm, J., Gotz, M. and Schulte, D. (2014). Meis2 is a Pax6 co-factors in neurogenesis and dopaminergic periglomerular fate specficiation in the adult olfactory bulb. *Development* **141**, 28-38.

Akalin, A., Fredman, D., Arner, E., Dong, X., Bryne, J. C., Suzuki, H., Daub, C. O., Hayaskizaki, Y. and Lenhard, B. (2009). Transcriptional features of genomic regulatory blocks. *Genome Biol.* **10**: 15.

Alharbi, R. A., Pettengell, R., Pandha, H. S. and Morgan, R. (2012). The role of HOX genes in normal hematopoiesis and acute leukemia. *Leukemia* **27**, 1000-1008.

Allen, Z. J., Waclaw, R. R., Colbert, M. C. and Campbell, K. (2007). Molecular identity of olfactory bulb interneurons: transcriptional codes of periglomerular neuron subtypes. *J. Mol. Histo.* **38**, 517-525.

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

Aravind, L., Watanabe, H., Lipman, D. and Koonin, E. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *PNAS* **97**, 11319-11324.

Argiropoulos, B. and Humphries, R. K. (2007). HOX genes in hematopoiesis and leukemogenisis. *Oncogene* **26**, 6766-6776.

Asomugha, C. O., Gupta, R. and Srivastava, O. P. (2010). Identification of crystalline modifications in the human lens cortex and nucleus using laser capture microdissection and CyDye labeling. *Mol. Vision* **16**, 476-494.

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

Balavoine, G., Rosa, R. and Adoutte, A. (2002). HOX clusters and bilaterians phylogeny. *Mole. Phylogen. Evo.* **24**, 366-373.

Becker, A., Bey, M., Bürglin, T. R., Saedler, H. and Theissen, G. (2002). Ancestry and diversity of BEL1-like homeobox genes revealed by gymnosperm (Gnetum gnemon) homologs. *Dev. Genes Evol.* **212**, 452-457.

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

Berthelsen, J., Zappavigna, V., Ferretti, E., Mavilio, F. and Blasi, F. (1998). The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperatively. *EMBO J.* **17**, 1434-1445.

Bessa, J., Tavares, M. J., Santos, J., Kikuta, H., Laplante, M., Becker, T. S., Gomez-Skarmeta, J. L. and Casares, F. (2008). Meis1 regulates cyclin D1 and c-myc expression, and controls the proliferation of the multipotent cells in the early developing zebrafish eye. *Development* **135**, 799-803.

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

Bjerke, G. A., Hyman-Walsh, C. and Wotton, D. (2011). Cooperative Transcriptional Activation by Klf4, Meis2, and Pbx1. Mole. *Cell Biol.* **31**, 3723-3733.

Bomgardner, D., Hinton, B. T. and Turner, T. T. (2003). Hox Genes and Meis 1, a Hox- DNA Binding Cofactor, Are Expressed in the Adult Mound Epididymis. *Biol. Reprod.* **68**, 644-650.

Bryson-Richardson, R., Berger, S. and Currie, P. (2011). ATLAS OF ZEBRAFISH DEVELOPMENT, **1st edition**. *Academic Press*, Waltham; Massachusetts.

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

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

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

Cajiao, I., Zhang, A., Yoo, E. J., Cooke, N. E. and Liebhaber, S. A. (2004). Bystander gene activation by a locus control region. *EMBO J.* **23**, 3854-3863.

Carbe, C., Hertzler-Schaefer, K. and Zhang, X. (2012). The functional role of the Meis/Prep-binding elements in Pax6 locus during pancreas and eye development. *Dev. Biol.* **363**, 320-329.

Carpenter, B. S. (2010). Identification and developmental expression of the zebrafish zgc:154061 gene, a conserved yet uncharacterized maternally expressed meis2 linked gene. [MS Thesis]. *Department of Biology, Appalachian State University*, Boone, North Carolina.

Carroll, S. B. (1995). Homeotic genes and the evolution of arthropods and chordates. *Nature*. **376**, 479-485.

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

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

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

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

Cochrane, A. C. (2012). Identification and embryonic expression of a highly conserved *meis*-linked gene. [MS Thesis]. *Department of Biology, Appalachian State University,* Boone, North Carolina.

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

Crijns, A. P. G., Graeff, P., Geerts, D., Hoor, K. A., Hollema, H., Sluis, T., Hofstra, R. M. W., Bock, G. H., Jong, S., Zee, A. J. G. and Vries, E. G. E. (2007). MEIS and PBX homeobox proteins in ovarian cancer. *Euro. J. Cancer.* **43**, 2495-2505.

DiLorio, P., Alexa, K., Choe, S., Etheridge, L. and Sagerström, C. (2007). TALE-Family homeodomain proteins regulate endodermal sonic hedgehog expression and pattern the anterior endoderm. *Dev. Bio.* **304**, 221-231.

Duboule, D. (1998). Vertebrate hox gene regulation: clustering and/or colinearity? *Curr. Opin. Genetics Dev.* **8**, 514-518.

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

Ebisuya, M., Yamamoto, T., Nakajima, M. and Nishida, E. (2008). Ripples from neighbouring transcription. *Nat. Cell. Biol.* **10**, 1106-113.

Ekker, S. C., Jackson, D. G., Kessler, D. P., Sun, B. I., Young, K. E. and Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. *EMBO J.* ***13***, 3551-3560.

Erickson, T., French, C. and Waskiewicz, A. (2010). Meis1 specifics positional information in the retina and tectum to organize the zebrafish visual system. *Neur. Dev.* **5**, 22.

French, C., Erickson, T., Callander, D., Berry, K., Koss, R., Hagey, D., Stout, J., Wuennenberg-Stapleton, K., Ngai, J., Moens, C. and Waskiewicz, A. (2007). Pbx homeodomain proteins pattern both the zebrafish retina and tectum. *BMC Dev. Biol.* **7**, 85.

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

Freeze, H. H., Eklund, E. A., Ng, B. G. and Patterson, M. C. (2012). Neurology of inherited glycosylation disorders. *Lancet. Neurol.* **11**, 453-466.

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

Geerts, D., Schilderink, N., Jorritsma, G. and Versteeg, R. (2003). The role of the MEIS homeobox genes in neuroblastoma. *Cancer Letters* **197**, 87-92.

Geerts, D., Revet, I., Jorritsma, G., Schilderink, N. and Versteeg, R. (2005). MEIS homeobox genes in neuroblastoma. *Cancer Letters* **228**, 43-50.

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

Gilbert, S. (2000). DEVELOPMENTAL BIOLOGY, **6th ed**. *Sinauer Associates*, Sunderland; Massachusetts.

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

Graham, B. S. (2009). Developmental expression pattern of the zgc:154061 gene in zebrafish. [MS Thesis]. *Department of Biology, Appalachian State University*, Boone, North Carolina.

Han, L. and Colicelli, J. (1995). A human protein selected for interference with Ras function interacts directly with Ras and competes with Raf1. *Mol. Cell. Biol.* **15**, 1318-1323.

Hancock, R. L. (1978). Theoretical mechanisms for synthesis of carcinogen-induced embryonic proteins II. Perturbations of post-translational nuclear protein modifications. *Med. Hypoth.* **4**, 290-301.

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

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

Hittinger, C. T. and Carroll, S. B. (2007). Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449**, 677-682.

Hoey, T. and Levine, M. (1988). Divergent homeo box proteins recognize similar DNA sequences in Drosophila. *Nature* **332**, 858-861.

Holliday, R. and Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science* **187**, 226-232.

Holliday, R. and Pugh, J. E. (1990). Mechanisms for the control of gene activity during development. *Bio. Rev.* **65**, 431–471.

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

Hunter, P. R., Nikolaou, N., Odermatt, B., Williams, P. R., Drescher, U. and Meyer, M. P. (2011). Localization of Cadm2a and Cadm3 proteins during development of the zebrafish nervous system. J Comp. *Neurol.* **519**, 2252-2270.

Hyman-Walsh, C., Bjerke G. and Wotton, D. (2010). An autoinhibitory effect of the homothorax domain of Meis2. *FEBS J.* **227**, 2584–2597.

Imamura, T., Morimoto, A., Ikushima, S., Kakazu, N., Hada, S., Tabata, Y., Yagi, T., Inaba, T., Hibi, S., Sugimoto, T. and Imashuku, S. (2002). A novel infant acute lymphoblastic leukemia cell line with MLL-AF5q31 fusion transcript. *Leuk.* **16**, 2302-2308.

Itoh, N. and Ornitz, D. M. (2004). Evolution of the Fgf and Fgfr gene families. *Trends Gen.* **20**, 7.

Jacobs, Y., Schnabel, C. and Cleary, M. (1999). Trimeric association of hox and tale homeodomain proteins mediates hoxb2 hindbrain enhancer activity. *Mol. Cell Biol.* **19**, 5134-5142.

Jaenisch, R. and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Gen. Supp.* **33**, 245-254.

Jensen, O.N. (2000). Modification-specific proteomics: systemic strategies for analyzing post-transnationally modified proteins. *Proteomics* **136**, 42.

Jensen, O.N. (2004). Modification-specific proteomics: characterizations of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **8**, 33-41.

Jensen, L. J., Gupta, R., Blom, N., Devos, D., Tamames, J., Kesmir, C., Nielsen, H., Staerfeldt, H. H., Rapacki, K., Workman, C., Andersen, C. A. F., Knudsen, S., Krogh, A., Valencia, A. and Brunak, S. (2002). Prediction of human protein function from post-translational modifications and localization features. *J. Mol. Biol.* **319**, 1257-1265.

Jiang, Y., Shi, H. and Liu, J. (2009). Two Hox cofactors, the Meis/Hth homolog UNC-62 and the Pbx/Exd homolog CEH-20, function together during C. elegans postembryonic mesodermal development. *Dev. Biol.* **334**, 535-546.

Jiang, T., Zhou, X., Taghizadeh, K., Dong, M. and Dedon, P. C. (2007). N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *PNAS* **104**, 60-65.

Jin, J., Wen, P. and Locker, J. (1995). Enhancer sharing in a plasmid model containing the α-Fetoprotein and albumin promoters. *DNA and Cell Biol.* **14**, 276-272.

Jones, T., Flomen, R., Senger, G., Nizetic, D. and Sheer, D. (2000). The homebox gene MEIS1 is amplified in IMR-32 and highly expressed in other neuroblastoma cell lines. *Euro. J. Cancer* **36**, 2368-2374.

Kamath, K. S., Vasavada, M. S. and Srivastava, S. (2011). Proteomic databases and tools to decipher post-translational modifications. *J. Proteomics* **75**, 127-144.

Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A. Z., Engstrom, P. G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K., Chislain, J., Pezeron, G., Mourrain, P., Ellingsen, S., Oates, A. C., Thisse, C., Thisse, B., Foucher, I., Adolf, B., Geling, A., Lenhard, B. and Becker, T. S. (2007). Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.* **17**, 545-555.

Kmita, M., Kondo, T. and Duboule, D. (2000). Targeted inversion of a polar silencer within the HoxD complex re-allocates domains of enhancer sharing. *Nat. Gen.* **26**, 451-454.

Kosak, S. T. and Groudine, M. (2004). Gene order and dynamic domains. *Science* **306**, 644-647.

Kroger, N., Deutzmann, R. and Sumper, M. (2001). The chemical structure of silaffin-1a from *cylindrotheca fusiformis*. *J. Bio. Chem.* **276**, 26066-26070.

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

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

Lawrence, J., Sauvageau, G., Humphries, R. and Largman, C. (1996). The Role of Hox Homeobox Genes in Normal and Leukemic Hematopoiesis. *Stem Cells* **14**, 281-291.

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

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-570.

Maconochie, M., Nonchev, S., Morrison, A. and Krumlauf, R. (1996). Paralogous hox genes: function and regulation. *Annu. Rev. Genet.* **30**, 529-556.

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

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

Mann, M. and Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nat. Biotech.* **21**, 255-261.

Mann, R. S. (1997). Why are Hox genes clustered? *BioEssays* **19**, 661-664.

Mann, R. S. and Abu-Shaar, M. (1996). Nuclear import of the homeodomain protein extradenticle in response to Wg and Dpp signalling. *Nature* **383**, 630-633.

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

Mark, M., Rijli, F. M. and Chambon, P. (1997). Homeobox genes in embryogenesis and pathogenesis. *Ped. Res.* **42**, 421-429.

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

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

Mercader, N., Leonardo, E., Piedra, M. E., Martinez, A. C., Ros, M. A. and Torres, M. (2000). Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development* **127**, 3961-3970.

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

Meyer, A. and Schartl, M. (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Cur. Opin. Cell Biol.* **11**, 699-704.

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

Minehata, K., Kawahara, A. and Suzuki, T. (2008). Meis1 regulates the development of endothelial cells in zebrafish. *Biochem. Biophys. Res. Commun.* **374**, 647-652.

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

Mojsin, M. and Stevanovic, M. (2010). Pbx1 and Meis1 up-regulate Sox3 gene expression by direct interaction with a consensus binding site within the basal promoter region. *Biochem. J.* **425**, 107-116.

Morata, G. and Lawrence, P. A. (1977). The development of wingless, a homeotic mutation of Drosophila. *Dev. Biol.* **56**, 227-240.

Morey, C., Kress, C. and Bickmore, W. A. (2009). Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression. *Genome Res.* **19**, 1184-1194.

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

Mueller, T. and Wullimann, M. F. (2005). ATLAS OF EARLY ZEBRAFISH BRAIN DEVELOPMENT: A TOOL FOR MOLECULAR NEUROGENETICS, **1st edition**. *Gulf Professional Publishing*, Houston; Texas.

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

Nelson, K. C. (2011). Identification of gene regulatory elements associated with the meis family of homeobox genes. [MS Thesis]. *Department of Biology, Appalachian State University*, Boone, North Carolina.

Obourn, J. D., Koszewki, N. J. and Notides, A. C. (1993). Hormone- and DNA-binding mechanisms of the recombinant human estrogen receptor. *Amer. Chem. Soc.* **32**, 6229-6236.

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

Pillay, L., Forrester, A., Erickson, T., Berman, J. and Waskiewicz, A. (2010). The Hox cofactors Meis1 and Pbx act upstream of gata1 to regulate primitive hematopoiesis. *Dev. Biol.* **340**, 306-317.

Pineault, N., Helgason, C. D., Lawrence, H. J. and Humphries, R. K. (2002). Differential expression of Hox, Meis1, and Pbx1 genes in primitive cells throughout murine hematopoietic ontogeny. *Exp. Hematol.* **30**, 49-57.

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

Prabakaran, S., Lippens, G., Steen, H. and Gunawardena, J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *WIREs Syst. Biol. Med.* **10**, 1185.

Rivas, M., Epinosa-Vazquez, J., Sambrani, N., Greig, S., Merabet, S., Graba, Y. and Hombria, J. (2013). Antagonism versus cooperatively with TALE cofactors at the base of the functional diversification of Hox protein function. *PLOS Gen.* **9**, 14.

Royo, J., Bessa, J., Hidalgo, C., Fernandez-Minan, A., Tena, J., Roncero, Y., Comez-Skarmeta, J. and Casares, F. (2012). Identification and analysis of conserved cis-regulatory regions of the MEIS1 gene. *PLOS One* **7**, 10.

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

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

Sanchez-Guardado, L. O., Ferran, J. L., Rodriguez-Gallardo, L., Puelles, L. and Hidalgo-Sanchez, M. (2011). Meis gene expression patterns in the developing chicken inner ear. *J. Comp. Neurol.* **519**, 125-147.

Schoenafinger, G., Schracke, N., Linne, U. and Marahiel, M. A. (2006). Formylation Domain: An essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. *JACS Comm.* **128**, 7406-7407.

Sharkey, M., Graba, Y. and Scott, M. P. (1997). Hox genes in evolution: protein surfaces and paralog groups. *Trends Genet.* **13**, 145-151.

Sharpe, J., Novchev, S., Gould, A., Whiting, J. and Krumlauf, R. (1998). Selectivity, sharing and competitive interactions in the regulation of Hoxb genes. *EMBO J.* **17**, 1788-1798.

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

Spitz, F. and Duboule, D. (2008). Global control regions and regulatory landscapes in vertebrate development and evolution. *Adv. Genet.* **61**,176-204.

Spitz, F., Gonzalez, F. and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the *HoxD* cluster. *Cell* **113**, 405-417.

Sumiyama, K., Irvine, S. Q, Stock, D. W., Weiss, K. M., Kawasaki, K., Shimizu, N., Shashikant, C. S., Miller, W. and Ruddle, F. H. (2002). Genomic structure and functional control of the Dlx3-7 bigene cluster. *PNAS* **99**, 780-785.

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

Stoffers, D. A., Thomas, M. K. and Habener, J. F. (1997). Homeodomain protein IDX-1: a master regulator of pancreas development and insulin gene expression. *Trends Endocrin. Metab.* **8**, 145-151.

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

Tena, J. J., Alonso, M. E., Calle-Mustienes, E., Splinter, E., Laat, W., Manzanares, M. and Gomex-Skarmeta, J. L. (2011). An evolutionarily conserved three-dimensional structure in the vertebrate Irx clusters facilitates enhancer sharing and coregulation. *Nat. Comm.* **2**, 310.

Waclaw, R. R., Wang, B., Pei, Z., Ehrman, L. A. and Campbell, K. (2009). Distinct temportal requirements for the homeobox gene Gsx2 in specifying striatal and olfactory bulb neuronal fates. *Neuron*. **63**, 451-465.

Wang, Y. C., Peterson, S. E. and Loring, J. F. (2014). Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* **24**, 143-160.

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

Wilkins, M. R., Gasteiger, E., Gooley, A. A., Herbert, B. R., Molloy, M. P., Binz, P. A., Ou, K., Sanchez, J. C., Bairoch, A., Williams, K. L. and Hochstrasser, D. F. (1999). High-throughput mass spectrometric discovery of protein post-translational modifications. *J. Mol. Biol.* **289**, 645-657.

Williams, E. and Hurst, L. (2000). The proteins of linked genes evolve at similar rates. *Nature* **407**, 900-903.

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

Wisniewski, J. R., Zougman, A. and Mann, M. (2007). N-Formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. *Nucl. Acids Res.* **36**, 570-577.

Wolpert, L. (2007). PRINCIPLES OF DEVELOPMENT, **3rd ed**. *Oxford University Press*, Oxford; New York.

Wood, C., Snijders, A., Williamson, J., Reynolds, C., Baldwin, J. and Dickman, M. (2009). Post-translational modifications of the linker histone variants and their association with cell mechanisms. *FEBS J.* **276**, 3685-3697.

Xing, G., Zhang, J., Chen, Y. and Zhao, Y. (2008). Identification of four novel types of *in vitro* protein modifications. *J. Proteome Res.* **7**, 4603-4608.

Yang, Y., Hwang, 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.

Yanai, I., Derti, A. and DeLisi, C. (2001). Genes linked by fusion events are generally of the same functional category: A systematic analysis of 30 microbial genomes. *PNAS* **98**, 7940-7945.

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

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

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

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

Zurovcoa, M., Tatarenkov, A. and Berec, L. (2006). Differences in the pattern of evolution in six physically linked genes of Drosophila melanogaster. *Gene* **381**, 24-33.

Vita

Zachary Scott Williams was born in Alameda, California. He has since lived in seven different places with his family, from Louisiana to Rhode Island, while his father worked for the US Coast Guard. Zach attended Great Bridge High School in Chesapeake, Virginia graduating in the top 10% of his class. As his parents relocated to Kodiak, Alaska, Zach moved to the mountains in Boone, North Carolina to attend college. At Appalachian State University, Zach graduated with double majors in Biology and Chemistry while serving as a Student Government Association Senator and his Fraternity's Chapter President. In the fall of 2011, Zach began pursuing a Master of Science concentrating in Cell and Molecular Biology completing his research thesis in the summer of 2014. In the fall of 2014, he commenced work towards earning his Doctor of Medicine at the Brody School of Medicine at East Carolina University in Greenville, North Carolina. In the future, Zach plans to obtain his M.D. and practice medicine wherever the hunting is best.