IDENTIFICATION AND DEVELOPMENTAL EXPRESSION OF THE ZEBRAFISH
ZGC:154061 GENE, A CONSERVED YET UNCHARACTERIZED MATERNALLY
EXPRESSED *MEIS2* LINKED GENE

A Thesis
by
BRANDON SCOTT CARPENTER

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

May 2010
Department of Biology

IDENTIFICATION AND DEVELOPMENTAL EXPRESSION OF THE ZEBRAFISH
ZGC:154061 GENE, A CONSERVED YET UNCHARACTERIZED MATERNALLY
EXPRESSED *MEIS2* LINKED GENE

A Thesis
by
BRANDON SCOTT CARPENTER
May 2010

APPROVED BY:

_____
Ted Zerucha
Chairperson, Thesis Committee

_____
Mary U. Connell
Member, Thesis Committee

_____
Susan L. Edwards
Member, Thesis Committee

_____
Steven W. Seagle
Chairperson, Department of Biology

_____
Edelma D. Huntley
Dean, Research and Graduate Studies

FOREWORD

The organization and formatting of this thesis strictly follows the instruction to the author for manuscript submission to *Gene Expression Patterns*, the official journal of The International Society of Developmental Biologists. The general organization of the text is similar to that of a *Nature* letter, with the whole text in a single main section headed "Results and Discussion."

ABSTRACT

IDENTIFICATION AND DEVELOPMENTAL EXPRESSION OF THE ZEBRAFISH
ZGC:154061 GENE, A CONSERVED YET UNCHARACTERIZED MATERNALLY
EXPRESSED*MEIS2* LINKED GENE
(May 2010)

Brandon Scott Carpenter, B.S., Appalachian State University

M.S., Appalachian State University

Chairperson: Ted Zerucha

We have identified a novel and previously undescribed gene, *zgc:154061*, located directly downstream of the zebrafish *meis2.2* gene. We have identified putative orthologs of this gene in all animals that we have been able to examine. The *zgc:154061* gene and its vertebrate orthologs are organized in a convergently transcribed manner with respect to the *Meis2* gene in all species we have examined (*meis2.2* in teleosts). It appears that the homologs of *Meis* and *zgc:154061* are also linked in amphioxus and sea urchins but that this linkage is not present in urochordates, nor in protostomes. During zebrafish development, transcripts of *zgc:154061* are observed in every cell of the embryo from the earliest stage through the shield stage indicating this gene is a maternal transcript since its expression precedes the activation of the zygotic genome at the midblastula transition. Expression of *zgc:154061* gradually decreases from its peak value at 0 hpf until 8 hpf and then is observed to be activated again at 12 hpf throughout the neural tube before becoming restricted to the retina and tectum opticum by 48 hpf.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

The vertebrate *Meis* homeobox-containing gene family consists of at least three members in vertebrates. The products of the *Meis* genes appear to function as cofactors, directly interacting with other homeodomain proteins as well as DNA to facilitate transcriptional regulation. During the process of searching for *cis*-regulatory elements associated with *Meis2* using phylogenetic footprinting, we identified a novel conserved gene sequence located directly downstream of *Meis2* in every vertebrate examined. While this novel gene appears to be present in every animal we have been able to examine, nothing is known concerning its function or expression. Here we report the first spatial and temporal expression pattern to date for any ortholog of this gene, in zebrafish.

1. RESULTS and DISCUSSION

1.1 Identification of zgc:154061

In zebrafish, the *zgc:154061* gene is found directly downstream of the *meis2.2* gene on chromosome 17 (Fig. 1A), the two genes separated by approximately 10.9-Kb. The *zgc:154061* gene codes for the production of a transcript 1914-nt in length (Fig. 1B) that is convergently transcribed with respect to *meis2.2* and is predicted to be encoded by 9 exons using the NCBI Model Maker tool for genomic sequence analysis. It does not appear that there is a second paralog of *zgc:154061*, associated with the second zebrafish *Meis2* homolog *meis2.1* or elsewhere in the zebrafish genome for that matter, based on BLAST analysis

1

against the available zebrafish genome. The gene encodes an open reading frame that would be predicted to produce a protein 300 amino acids in length (Fig. 1B). This protein has been previously reported as hypothetical protein LOC767755 as part of a large scale gene identification effort (Strausberg et al., 2002), and a cDNA clone of the gene made available through Open Biosystems, from whom we obtained it. Nothing else has been reported concerning this gene or its product in any species.

We have identified predicted orthologs of *zgc:154061* in all vertebrates with publicly available genome data that we have been able to examine. These vertebrate orthologs are present in the same convergently transcribed orientation directly adjacent to that species' *Meis2* ortholog, although the spacing separating the two genes varies between species. An amino acid alignment of representative vertebrate species is shown in Fig. 2 and summarized also in Table 1. As shown in Fig. 3, phylogenetic analysis revealed that the translated zebrafish *zgc:154061* protein is most closely related to the *Takifugu* ortholog amongst all species examined. The zebrafish predicted protein represents the longest protein from the orthologs examined. Furthermore there seems to be some variability in the amino terminus of the orthologous proteins with several different translation start points. The mouse and chicken proteins have similar start and stop sites as do the human and *Takifugu* proteins. The zebrafish protein includes an 18-aa amino terminal sequence not found in tetrapods. In addition the chicken protein appears to have an 11-aa deletion compared to human, mouse and zebrafish. There does not appear to be any similarity between *zgc:154061*, nor any of its orthologs, to any previously described gene to provide any clues to functional domains coded for by this gene or about the function of this gene in general.

Putative homologs of *zgc:154061* have also been identified in the urochordate *Ciona intestinalis*, the cephalochordate *Branchiostoma floridae* and the echinoderm *Strongylocentrotus purpuratus*. A comparison of representative homologs of the translated product of these genes is included in Table 1. In *C. intestinalis*, based on the available genome data, it appears that the *zgc:154061* ortholog and the *Meis* ortholog are on different chromosomes. However, in the amphioxus genome, the *zgc:154061* ortholog and the *Meis* ortholog are found adjacently on scaffold 120 and in the sea urchin the two genes are found adjacent to one another but it appears that they are transcribed from the same strand of DNA as opposed to the convergently transcribed orientation observed in vertebrates. Putative homologs of *zgc:15406* have also been identified in a number of invertebrate species, based on sequence similarities, including fruit fly (*Drosophila melanogaster*) as indicated in Table 1 and Fig. 3. It appears that all nonvertebrate putative orthologs of the product of *zgc:154061* share several short conserved domains in the C-terminal region as indicated in Fig 2 but none of the protostome orthologs appear to be adjacent to the invertebrate *Meis* orthologue *homothorax* in any of these species. Together this suggests that the genomic organization and linkage of *Meis2* and orthologs of *zgc:154061* is ancient and arose early in the deuterostome lineage, although this organization seems to have been lost in urochordates.

1.2 Spatial and temporal expression of zgc:154061

As a first step in characterizing *zgc:154061*, we examined the spatial and temporal expression of this gene using whole mount *in situ* hybridization (Fig. 4 and Fig. 5) and quantitative real-time PCR (Fig. 6) between the zygote and early larva periods. Strong

expression is observed at the one cell stage and expression persists ubiquitously in every cell beyond the onset of transcription of the zygotic genome, at the midblastula transition, to the shield stage (Fig. 4A-C). By 9 hpf, expression of *zgc:154061* is observed to be very low and almost undetectable by whole mount *in situ* hybridization (Fig. 4D). These observations are supported by quantitative real-time PCR where we see the quantity of transcripts of *zgc:154061* steadily decrease from their maximum level at the earliest stage of development to barely detectable levels at 8 hpf (Fig. 6). These observations would be consistent with that of a maternal transcript that is present in the zygote and that gradually degrades over time while not being replenished by new transcriptional activity. This is not unexpected since at least the first 3 h of zebrafish development are not accompanied by transcription of the embryo's genome. Transcripts of the zebrafish *ß-actin* gene, used as endogenous control for the quantitative real-time PCR experiments, have been reported to be present at nearly even levels throughout most of development, although with an inexplicable slight increase in expression observed prior to the onset of embryonic transcription (McCurley and Callard, 2008). The decrease in expression of *zgc:154061* relative to the steady expression *ß-actin* prior to activation of transcription of the embryo's genome suggests varying half lives of these two transcripts, or perhaps the existence of a mechanism by which *zgc:154061* is actively reduced. Following epiboly, low expression of *zgc:154061* is again observed, this later expression peaking at 12 hpf based on quantitative real-time PCR (Fig. 6). Whole mount *in situ* hybridization reveals that this expression is largely restricted to the developing neural tube. Expression is observed throughout the neural tube and the optic vesicle until 24 hpf (Fig. 4G and Fig. 5C,D) becoming more anteriorly restricted by 48 hpf of development where expression is observed throughout the forebrain, the tectum of the midbrain and very

faintly in the anterior hindbrain (Fig. 5G,H). The strongest expression at 48 hpf is observed in the retina and tectum opticum (Fig. 5H,J,K). Following 48 hpf, *zgc:154061* expression gradually decreases in these areas to below observable levels for the remainder of development.

A reasonable explanation for the conservation of the genomic linkage of *Meis2* and orthologs of *zgc:154061* would be that the two genes are sharing *cis*-regulatory elements. This logic has been used to help explain the clustered organization of *Hox* genes (Harding et al., 1985; Gould et al., 1997; Mann, 1997; Sharpe et al., 1998; Prince, 2002) as well as the clustered organization of the *Dlx* gene family, for example, which are found as convergently transcribed gene pairs in vertebrates (Ellies et al., 1997; Zerucha et al., 2000). If *Meis2* and *zgc:154061* orthologs are sharing *cis*-regulatory elements, one would predict that they should exhibit overlapping patterns of expression. Comparing the expression of *zgc:154061* described here to that of *meis2.2* shown here (Fig. 5) and reported previously (Waskiewicz et al., 2001; Thisse, 2005; Bessa et al., 2008; Santos et al., 2010), it appears that there is some overlap between these two genes in the developing eye as well as anterior neural tube and brain leading up to and at 24 hpf and particularly in the retina at 48 hpf. Specifically, *meis2.2* is expressed throughout the optic vesicle from 15 to 18 hpf (Bessa et al., 2008) but is completely lost from the eye by 24 hpf as shown in Fig 5A and as reported previously (Bessa et al., 2008). We see *zgc:154061* expressed throughout the optic vesicle during this same time frame (Fig. 4G and Fig. 5C,D), however its expression persists to 24 hpf whereas *meis2.2* expression has disappeared by this point. Expression of *meis2.2* is also observed in the forebrain at 24 hpf (Fig. 5A,B) and *zgc:154061* is expressed throughout the neural tube at this same time, including the forebrain (Fig. 5C,D). By 48 hpf, *meis2.2* expression reappears

5

in the retina and tectum of the midbrain and exhibits overlapping expression with *zgc:154061* in these regions (Fig.5E-H). The highly conserved nature of *meis2.2* and *zgc:154061* expression, particularly at 48 hpf in the retina suggests the possibility that these genes are sharing *cis*-regulatory elements, at least for one directing expression in the developing eye. Examples consistent with our observations of expression of *meis2.2* and *zgc:154061* earlier in development, such as at 24 hpf, also exist however, where tandem genes are thought to share *cis*-regulatory elements but do not share perfectly overlapping patterns of expression, likely because of interactions between the shared enhancers and the promoters as well as developmental stage-specific repressors associated with each gene (Irvine et al., 2007). It is also possible that the organization of these genes has been preserved because of the presence of *cis*-regulatory elements within introns of one of the genes but directing expression of the other gene, such as that which is seen for the limb-specific long range enhancer controlling sonic hedgehog expression and that is found in an intron of the *Lmbr1* locus (Lettice et al., 2002).

Recently, conserved syntenic regions containing long-range *cis*-regulatory elements distributed over long distances and encompassing conserved developmental regulatory genes as well as phylogenetically and functionally unrelated "bystander" genes have been termed genomic regulatory blocks, or GRBs (Kikuta et al., 2007). It has been proposed that following the whole genome duplication in teleosts, certain GRBs that contain conserved developmental genes have been maintained in only one of the duplicated syntenic regions, suggesting that evolutionary pressure acts to maintain only a single-copy of the GRB. For example *Pax6,* a highly conserved gene involved in vertebrate retinal and central nervous system development, has been duplicated in teleosts and as a result zebrafish contain two

copies, *pax6.1* and *pax6.2*. Using enhancer detection, Kikuta et al. (2007) located a region ~68 kb downstream of *pax6.2* within an intron of a neighboring gene, *elp4,* that was able to recapitulate the expression pattern of *pax6.2*. This suggests that the *cis*-regulatory sequence driving *pax6.2* expression exists within the *elp4* gene despite *elp4* expression being more wide spread and consequently not appearing to be regulated by the regulatory elements within its own introns. Interestingly, *elp4* was only maintained downstream of *pax6.2* similar to what we report here in terms of *zgc:154061* being maintained downstream of *meis2.2* but there being no paralog associated with *meis2.1*. Thus, the presence of any putative *cis*-regulatory element within an intron of *zgc:154061* may explain why this gene is maintained downstream of *meis2.2*. It is possible that the single-copy GRB containing *zgc:154061* and *meis2.2* is protecting this locus from chromosomal breakage while any paralogous regulatory region associated with *meis2.1* has been lost by neutral evolution; a phenomenon predicted by the duplication degeneration complementation model (Force et al., 1999).

We have described here a novel zebrafish gene *zgc:154061* that is highly conserved amongst vertebrates in terms of sequence identity and linkage to the *Meis2* (*meis2.2* in zebrafish) homeobox gene. It appears that the genomic organization of the *zgc:154061* and *Meis2* genes arose early in the deuterostome lineage as it is also observed in cephalochordates and echinoderms. The amino acid alignment of the vertebrate orthologs of zgc:154061 indicates a very well conserved C-terminal region, which may very well represent a functional domain. This gene is maternally expressed and also expressed in the developing neural tube and eye in a manner at least partially overlapping with its genomic neighbor *meis2.2*. The role of *zgc:154061* and its orthologs remain to be investigated and we

7

are currently exploring RNA over-expression and morpholino knockdown experiments in zebrafish to pursue this.

2. EXPERIMENTAL PROCEDURES

2.1 Zebrafish Husbandry

Zebrafish embryos were obtained by pair-wise breeding of adult AB line fish housed in an enclosed Z-Mod system (Aquatic Habitats) and maintained at 28 °C on a 16 h light / 8 h dark cycle. Embryos were staged according to (Kimmel et al., 1995). All experimental procedures involving zebrafish were approved by the Appalachian State University IACUC.

2.2 In situ hybridization

The full length 1914-bp cDNA clone of *zgc:154061* was isolated by the Zebrafish Gene Collection (ZGC), an NIH initiative, in conjunction with the NIH Mammalian Gene Collection (MGC) project (Strausberg et al., 2002) and made available through Open Biosystems (Clone Id: 8334609, Accession: BC124527) from whom we purchased it. So as to utilize a smaller probe size for our *in situ* hybridizations, a partial 393-bp fragment of the zebrafish *zgc:154061* gene (underlined in Fig. 1B) corresponding to positions 608-1000 of the full length cDNA clone was isolated by PCR using the primer sequences GGTCTGGAACATGAAGAC and CCTCATGCCATCAGAAAC (locations indicated in Fig. 1B). The PCR amplification was performed with Phusion High-Fidelity DNA Polymerase (New England BioLabs), the product subcloned into the pGEM-T Vector System (Promega, Madison WI) and confirmed by sequencing. An antisense DIG-labelled riboprobe

was synthesized directly from this subclone using T7 RNA Polymerase after linearizing the plasmid with *Not*I. A sense DIG-labelled riboprobe, used as a negative control, was synthesized from the same construct using SP6 RNA Polymerase after linearizing the plasmid with *Sac*II. Whole mount *in situ* hybridization was performed as described by Thisse and Thisse (2008) (Thisse and Thisse, 2008). Embryos were photographed using a Leica MZ6 dissecting microscope, Leica DFC320 digital camera and the Leica Application Suite Version 3.3.1. Images were compiled for Figures 4 and 5 using Adobe Photoshop 7.0. For sectioning, 48 hpf embryos were positioned in 1.5% agarose melted in 5% sucrose in PBS following *in situ* hybridization. Agarose blocks were incubated in 30% sucrose in PBS at 4°C overnight. Agar blocks were frozen in optical cutting temperature (O.C.T.) media and 20μm sections were cut using a Leica CM 1100 cryostat. Images of sections were obtained using an Olympus IX81 inverted microscope and processed with MicroSuite Biological Suite software. Images were compiled for Fig.5J,K using Adobe Photoshop 7.0.

2.3 RNA Extraction And Quantitative real-time *PCR*

Total RNA was extracted from 30-100 staged embryos by homogenizing with RNase, DNase-, pyrogen-free disposable pestles (Kontes) in TRIzol (Invitrogen) and following the protocol described by (Chomczynski and Mackey, 1995). RNA quality and quantity was determined by NanoDrop and denaturing gel electrophoresis. Reverse transcription of RNA samples into cDNA were performed using the Applied Biosystems High Capacity RNA-to-cDNA Master Mix and following the manufacturer's instructions.

Relative Quantitative real-time PCR was performed to analyze temporal expression of *zgc:154061* during zebrafish development using the Applied Biosystems 7300 real-time PCR System. For each stage examined, 200 ng of cDNA was used as template together with TaqMan Universal PCR Master Mix (Applied Biosystems) and gene specific primers and probe: *zgc:154061* forward 5'-GCAGACGCACTTCACACATCTC-3'; *zgc:154061* reverse 5'–TGCGCTTCATTCTCTTCTGGTA-3'; *zgc:154061* probe 5'-FAM-CGCGTCCACTCTGCTGAGCATCTTC-TAMRA-3' (see Fig. 1B for locations of primers and probe). The zebrafish *ß-actin* gene was used as an internal control with gene and specific primers and probe: ß -actin forward 5'–GCTGTTTTCCCCTCCATTGTTG-3'; ß-actin reverse 5'–TTTCTGTCCCATGCCAACCAT-3'; ß-actin probe 5'-FAM-CCCAGACATCAGGGAGTG-TAMRA-3'. Primers and probes were designed using Primer Express (Applied Biosystems) and purchased from Operon. The following amplification protocol was used: 50°C for 2 m; 95°C for 10 m; 95°C for 15 s, 60°C for 1 m repeated for 40 cycles. All reactions were performed in triplicate twice and two separate RNA extractions from each stage of development examined. Results were interpreted and are shown as level of relative expression calibrated to expression in an adult zebrafish using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Fig. 1 Genomic organization and sequence of *zgc:154061* (A) Position of *zgc:154061* with respect to *meis2.2* in zebrafish. Vertical lines/boxes represent exons. Arrows indicate the direction of transcription. The predicted exons of the *Zgc:154061* gene and its location with respect to *meis2.2* were identified using the publicly available genome sequence and the Model Maker tool in the NCBI Database. (B) Nucleotide and translated sequence of the *zgc:154061* gene. The sequence used as a probe for *in situ* hybridization is underlined and the primer sites used to amplify this sequence are double underlined. The primer sites used for quantitative real-time PCR are indicated by red double underlines and the site of the probe by a red single underline. Amino acid sequence is shown above each codon. Start and stop codons are indicated by boxes.

Table 1. Amino acid sequence identity percentages based on pairwise comparisons between putative, representative homologues of zgc:154061 that we have been able to identify based on searches of that organisms's corresponding publicly available genome data. Species examined are: human (*Homo sapien* - Hs); mouse (*Mus musculus* - Mm); chicken (*Gallus gallus* - Gg); African clawed frog (*Xenopus laevis* - Xl); zebrafish (*Danio rerio* - Dr); pufferfish (*Takifugu rubripes* - Tr); tunicate (*Ciona intestinalis* - Ci); lancelet (*Branchiostoma floridae* - Bf); sea urchin (*Strongylocentrotus purpuratus* – Sp); fruit fly (*Drosophila melanogaster* - Dm). Amino acid sequences were obtained from publicly available genome sequences through the NCBI database (accession numbers: Hs - **NP_115888.1**; Mm - **XP_001480310.1**; Gg - **NP_001026371.1**; Xl - **NP_001090210.1**; Dr - **NP_001070190.1**; Ci - **XP_002128870.1**; Bf - **XP_002221176.1**; Sp - **XP_780639.1**; Dm - **NP_648806.1**) except that of *Takifugu rubripes* which was obtained from the IMCB Fugu Genome Project database (gene:**SINFRUG00000137928** transcript:**SINFRUT00000146276**).

Fig. 2. Amino acid sequence alignments of products of *zgc:154061* orthologues for: human (*Homo sapiens* - hypothetical protein LOC84529); mouse (*Mus musculus* - hypothetical protein LOC399568); chicken (*Gallus gallus* - hypothetical protein LOC423293); African clawed frog (*Xenopus laevis* - hypothetical protein LOC779112); pufferfish (*Takifugu rubripes* – translation of SINFRUT00000146276) and; zebrafish (*Danio rerio* - hypothetical protein LOC767755). Yellow shaded regions represent complete identity amongst all six species, blue shading indicates identity between at least three of the species shown. Amino acid sequences are based on publicly available genome sequences though the NCBI database except that of *Takifugu rubripes* which was obtained from the IMCB Fugu Genome Project database. Amino acid alignments were generated using Vector NTI Advance Version 11.0 Align X (Invitrogen) with the following pairwise settings: K-tuple size-1, number of best diagonals-5, gap penalty-3 and multiple alignment settings: gap opening penalty-10, gap extension penalty-0.05, Gap separation penalty range-40, % identity for alignment delay-40. Underlined regions indicate domains that are also found in putative invertebrate orthologs of *zgc:154061*.

Fig. 3. Phylogenetic tree of zgc:154061 orthologous amino acid sequences. The phylogenetic tree was constructed using Phylogeny.fr (http://www.phylogeny.fr/) using MUSCLE 3.7 for multiple sequence alignment, GBlocks 0.91b for alignment refinement, PhyML 3.0 aLRT for phylogeny, and TreeDyn 198.3 for tree rendering (Dereeper et al., 2008). Numbers on branches represent the percentage of how many times clades grouped following 500 replications. Species examined are: human (*Homo sapien*); chimp (*Pan troglodytes*); mouse (*Mus musculus*); dog (*Canis familiaris*); cow (*Bos taurus*) chicken (*Gallus gallus*); African clawed frog (*Xenopus laevis*); zebrafish (*Danio rerio*); pufferfish (*Takifugu rubripes*); tunicate (*Ciona intestinalis*); lancelet (*Branchiostoma floridae*); sea urchin (*Strongylocentrotus purpuratus*); fruit fly (*Drosophila melanogaster*). Amino acid sequences were obtained from publicly available genome sequences through the NCBI database (accession numbers: *Homo sapien* - **NP_115888.1**; *Pan troglodytes* - **XP_510289.2**; *Mus musculus* - **XP_001480310.1**; *Canis familiaris* - **XP_849922.1**; *Bos taurus* - **NP_001015668.1**; *Gallus gallus* - **NP_001026371.1**; *Xenopus laevis* - **NP_001090210.1**; *Danio rerio* - **NP_001070190.1**; *Ciona intestinalis* - **XP_002128870.1**; *Branchiostoma floridae* - **XP_002221176.1**; *Strongylocentrotus purpuratus* - **XP_780639.1**; *Drosophila melanogaster* - **NP_648806.1**) except that of *Takifugu rubripes* which was obtained from the IMCB Fugu Genome Project database (gene:**SINFRUG00000137928** transcript:**SINFRUT00000146276**).

Fig. 4. Whole-mount *in situ* hybridization on analysis of early zebrafish *zgc:154061* expression. Embryos are shown as whole mounts with the anterior to the left. (A) 16-cell (1.5 hpf), (B) 4 hpf, (C) 6 hpf, (D) 9 hpf (90% epiboly), (E) 14 hpf, (F) 20 hpf, and (G) 20 hpf dorsal view with yolk removed. e, eye.

Fig. 5. Whole-mount *in situ* hybridization on analysis of late zebrafish *zgc:154061* and *meis2.2* expression. Embryos are shown as whole mounts the anterior to the left. (A) 24 hpf *meis2.2*, (B) 24 hpf *meis2.2* dorsal view with yolk removed, (C) 24 hpf *zgc:154061*, (D) 24 hpf *zgc:154061* dorsal view with yolk removed, (E) 48 hpf *meis2.2*, (F) 48 hpf *meis2.2* dorsal view with yolk removed, (G) 48 hpf *zgc:154061*, (H) 48 hpf *zgc:154061* dorsal view with yolk removed, (I) 48 hpf *zgc:154061* sense RNA probe negative control. (J) and (K) represent 48 hpf *zgc:154061* transverse sections through the brain and retina following *in situ* hybridization. e, eye; fb, forebrain; r, retina; teo, tectum opticum.

Fig. 6. Quantitative real-time PCR analysis of *zgc:154061* expression. Total zebrafish mRNA was isolated from 30-100, 0 hpf to 120 hpf embryos. Expression of *zgc:154061* relative to that of the *β-actin* endogenous control at each developmental stage indicated is shown, calibrated to expression in an adult zebrafish using the $2^{-\Delta\Delta Ct}$ method. Error bars reflect standard error of the mean for each sample. Each value on the graph is significantly different from the neighboring values as determined by t-test (p 0.05).

FIGURES

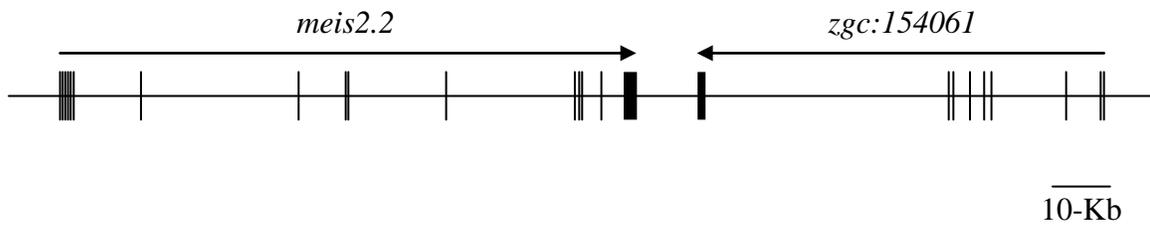Fig. 1A

Fig. 1B

```
   1    gtcgcgtgtgctgcagaactggagcttgtacagtgttattggtgctgtttgtgttgcaga
  60    gttgtgaaat
        M   D   A   V   S   G   G   S   A   T   G   T   G   E   Q
  71    ATG GAC GCA GTG TCA GGC GGC AGC GCC ACA GGA ACA GGT GAA CAG
        V   N   N   L   R   I   C   R   A   E   Y   R ··S   I   S
 116    GTG AAT AAT CTG AGG ATC TGC AGG GCA GAA TAC AGG AGC ATC AGC
        R   F   V   E   Q   L   R   P   T   R   Q   C   M   K · T
 161    AGG TTT GTG GAG CAG CTG CGT CCC ACA CGG CAG TGC ATG AAG ACC
        L   Q   T   H   F   T   H   L   P   A   S   T   L   L   S
 206    CTG CAG ACG CAC TTC ACA CAT CTC CCC GCG TCC ACT CTG CTG AGC
        I   F   S   Q   E   Y   Q   K   R   M   K   R   S   M   A
 251    ATC TTC TCC CAG GAG TAC CAG AAG AGA ATG AAG CGC AGT ATG GCC
        R   H   H   S   P   E   V   L   R   V   Y   Y   Q   R   Y
 296    AGA CAT CAC TCT CCT GAG GTG CTC AGA GTG TAT TAT CAG AGG TAT
        R   D   E   A   E   T   R   A   T   E   P   L   L   L   E
 341    CGA GAT GAA GCT GAG ACC AGA GCC ACA GAA CCA CTG CTG CTG GAG
        L   A   N   Q   V   D   L   S   P   A   L   L   A   R   L
 386    CTC GCT AAC CAG GTG GAT CTA TCT CCA GCT CTT TTG GCT CGT CTG
        M   L   E   C   F   L   E   E   R   N   A   S   V   P   S
 431    ATG CTG GAG TGT TTC CTA GAG GAG CGC AAC GCT TCA GTC CCT TCC
        R   Q   V   L   N   N   M   L   R   E   P   Y   L   I   P
 476    AGA CAA GTC CTC AAC AAC ATG CTG CGT GAG CCG TAT TTA ATT CCA
        D   L   V   L   A   K   H   I   E   Q   C   T   V   N   D
 521    GAT CTG GTG TTA GCC AAG CAC ATC GAG CAG TGC ACA GTA AAT GAC
        C   C   Y   G   P   L   V   D   C   I   K   H · A   I   G
 566    TGC TGT TAT GGA CCG CTG GTC GAC TGC ATC AAA CAT GCC ATC GGT
        L   E   H   E   D   T   L   R   D   K   L   R   E   R · N
 611    CTG GAA CAT GAA GAC ACT CTG AGA GAC AAA CTC AGA GAG AGG AAC
        L   S   F   L   D   E   N   Q   L   R   V   K   G   Y   D
 656    CTG TCG TTT TTA GAT GAG AAT CAG CTG CGG GTC AAA GGA TAC GAC
        K   T   P   D   I   I   L   E   V   P   I   A   V   D   G
 701    AAA ACC CCG GAC ATC ATC CTG GAG GTG CCG ATC GCT GTT GAT GGC
        H   I · V   H   W   I   E   S   K   A   S   F   G   D   D
 746    CAC ATC GTT CAC TGG ATC GAG AGT AAA GCT TCA TTT GGA GAT GAT
        H   S   H   N · T   Y   L   N   E   Q   F   W   S   Y   C
 791    CAC AGT CAC AAC ACA TAC CTG AAC GAG CAG TTC TGG AGC TAC TGC
        N   R   F   G   P   G   L   V   I   Y   W   F   G   F   I
 836    AAC AGG TTT GGT CCG GGT CTG GTC ATC TAC TGG TTC GGC TTC ATC
        S   E   L   D   C   Q   R · E   R   G   I   L   L   K   D
 881    TCA GAG CTG GAC TGC CAG CGG GAG CGA GGG ATC CTG CTG AAG GAC
        G   F   P   T   D   I   S   S   L   C   A   G   P   Q   R
 926    GGC TTC CCC ACG GAC ATC AGC AGC CTG TGT GCG GGA CCC CAG CGC
 971    TGAggacggggcgtttctgatggcatgaggatccggaagcaatttaagctttcctgcagc
1031    cagagactctgacaaacatgtgctgtgaactccatctgaaccatctgtgtgagtgtgtgt
1091    gtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgc
1151    tggttacactttatttgaagtctccttaactactatataggtacatttcaactactcatt
1211    gaaactgcacagtaaatgcaagtttttgtctggtgtttggtccagatatctgtgaactct
1271    taaataaagaagcatttttggatgtgcactgaatgtagtgttgttttcagaggtgatgtg
1331    ctaaagtgaagtgagttttttaattagagcaggctgaataatctgccagtgggggaagcag
1391    aataatcttatgtgtaaaggaaaaagccagtgatttccctcaccccactggcggattatt
1451    ctgcttgtttttgaggtaggactctctttattttggctcgttatttctgtaaacaacactg
1511    tattttatccctgtctataaaatgcttcctgatttaagaacgttcagatatttataccaa
1571    aaacaagacgaactctctaagaatacgtcagttttactgcgttgtactaatattttgaaa
1631    tcttttgcatgtaattacaattttttccgttaataattacaccctctcttacaccttaacc
1691    cactcttaacccttcccatatcactaaacctgtcaacaacccaaccccgatcccagatta
1751    atagcaccataactgttctgcaatgcattataaacaggaggagtacattgtgctgatgtt
1811    ttgatggtagttattgtgacttcaaataaagtgtatgattgtgttacatgctgatatgga
1871    ataaataaagagtattcttcgtaaaaaaaaaaaaaaaaaaaaaaa
```

15

Table 1

	Hs	Mm	Gg	Xl	Dr	Tr	Ci	Bf	Sp	Dm
Hs		94	82	76	70	77	42	52	51	38
Mm			82	72	61	77	34	44	50	27
Gg				68	59	74	32	42	58	24
Xl					64	70	38	44	52	31
Dr						76	34	41	50	26
Tr							42	49	50	38
Ci								36	41	28
Bf									54	33
Sp										36
Dm										

Fig. 2

```
Homo sapien        (1) ------------------------------------------------------------
Mus musculus       (1) ------------------MILTKAQYEEIAQCLVSVPPTRQSLRKLKQRFPSQSQATLLS
Gallus gallus      (1) ------------------MKLTKAQYDEIAQFLGHVQPTRQSLRKLKEKFPSQSQSTLLS
Xenopus laevis     (1) ------------------------------------------------------------
Takifugu repribes  (1) ------------------------------------------------------------
Danio rerio        (1) MDAVSGGSATGTGEQVNNLRICRAEYRSISRFVEQLRPTRQCMKTLQTHFTHLPASTLLS

Homo sapien        (1) ----------------------------------------------------------MARL
Mus musculus      (43) IFSQEYQKHIKRTHAKHHTPEAIESYYQRYLNGVGKNGAAPVLLELANEVDYAPSLMARI
Gallus gallus     (43) IFSQEYQKQIKRTHAKHHTAEAVETYYQRYLNGVMKNAAAPVLLELANEMDFAPSLMARI
Xenopus laevis     (1) ----------------------MYYARYLNEVARDPKVPILLELANEVDFSPALMART
Takifugu repribes  (1) ----------------------------------------------------------MARV
Danio rerio       (61) IFSQEYQKRMKRSMARHHSPEVLRVYYQRYRDEAETRATEPLLLELANQVDLSPALLARL

Homo sapien        (5) ILERFLQEHEETPPSKSIINSMLRDPSQIPDGVLANQVYQCIVNDCCYGPLVDCIKHAIG
Mus musculus     (103) ILERFLQGHEQTPPSKSVINSMLRDPSQIPDGVLANQVYQCIVNDCCYGPLVDCIKHAIG
Gallus gallus    (103) VLERFLQEQEQAIPSKTLINSMLRDPSQIPDGVLANQIYQCTVNDCCYGPLVDCIKHFIN
Xenopus laevis    (37) VLERFLQDHDGQPPTKPVLSSMLRDPSLIPDPVLANQVHLCIINDCFNGPLVDGIKHAIG
Takifugu repribes  (5) ILDRFLQDLFEGEMPSKTVLNSMLKEPSLIPDLILAQNIQQCTVNDCCYGPLVDCIKHAVG
Danio rerio      (121) MLECFLEERNASVPSRQVLNNMLREPYLIPDLVLAKHIEQCTVNDCCYGPLVDCIKHAIG

Homo sapien       (65) HEHEVLLRDLLLEKNLSFLDEDQLRAKGYDKTPDFILQVPVAVEGHIIHWIESKASFGDE
Mus musculus     (163) YEHEVLLRDLLLKKNLSFLDEDQLRAKGYDKTPDFILQVPVAVEGHIIHWIESKASFGDE
Gallus gallus    (163) -NR----------SCSLCVAEDQLRAKGYDKTPDFILEVPVAVEGHIIHWIESKASFGDE
Xenopus laevis    (97) HEHEVLLRQKLKEHNLAFLDEDQLRLKGYDKTPDVILEVPVAVDGHVIHWIESKASFGDE
Takifugu repribes (65) QEHEVLLCDKLKERNLSFLDETQLRAMGYDKTPDIILEVPVAVEGHIIHWIESKASFGDD
Danio rerio      (181) LEHEDTLRDKLRERNLSFLDENQLRVKGYDKTPDIILEVPIAVDGHIVHWIESKASFGDD

Homo sapien      (125) CSHHAYLHDQFWSYWNRFGPGLVIYWYGFIQELDCNRERGILLKACFPTNIVTLCHSIA-
Mus musculus     (223) CSHHAYLHGQFWSYWNRFGPGLVIYWYGFIQFLDCNRERGILLKASFPTDIVTLCHSTA-
Gallus gallus    (212) SSHQAYLQDQFWSYWNRFGPGLVIYWYGFIEFLDCHRERGILLKDCFPTDIVTLRHSMAQ
Xenopus laevis   (157) ASHKTYLHDQFWSYWNRFGPGLVIYWYGFIEDLDCNRERGILLKDGFPETLVMLGSCMAQ
Takifugu repribes(125) HSHRNYLNEQFWSYWNRFGPGLVIYWYGFIGELDCQRDRGILLKDCFPTDIVTLCHA---
Danio rerio      (241) HSHNTYLNEQFWSYCNRFGPGLVIYWFGFISELDCQRERGILLKDGFPTDISSLCAGPQR

Homo sapien      (184) --------------------
Mus musculus     (282) --------------------
Gallus gallus    (272) R-------------------
Xenopus laevis   (217) TDEHGQHTKNCLNETHQETES
Takifugu repribes(182) --------------------
Danio rerio      (301) --------------------
```
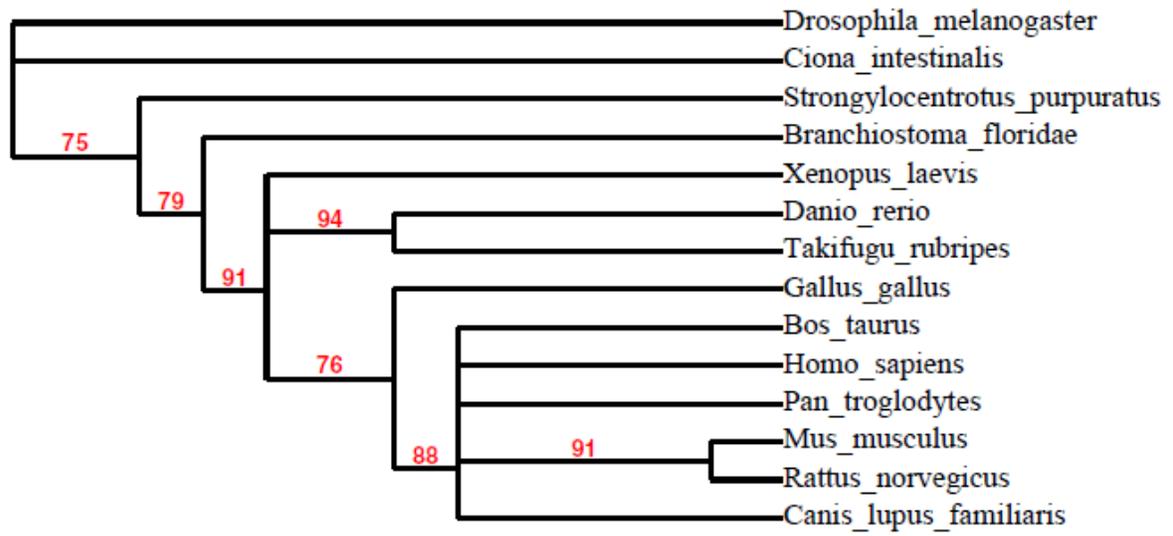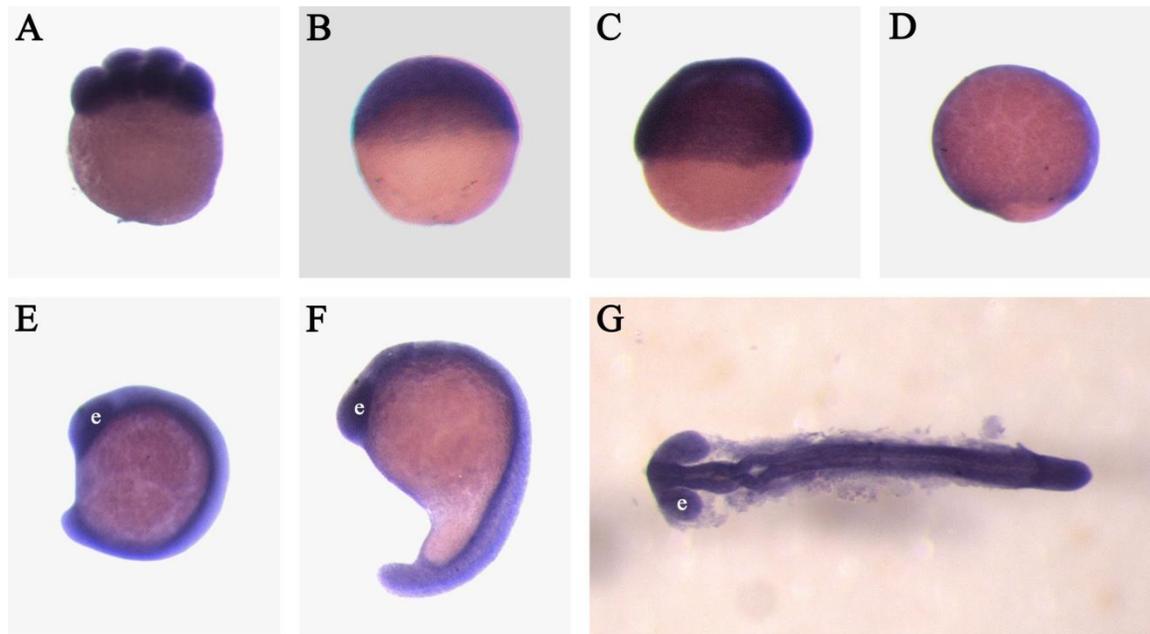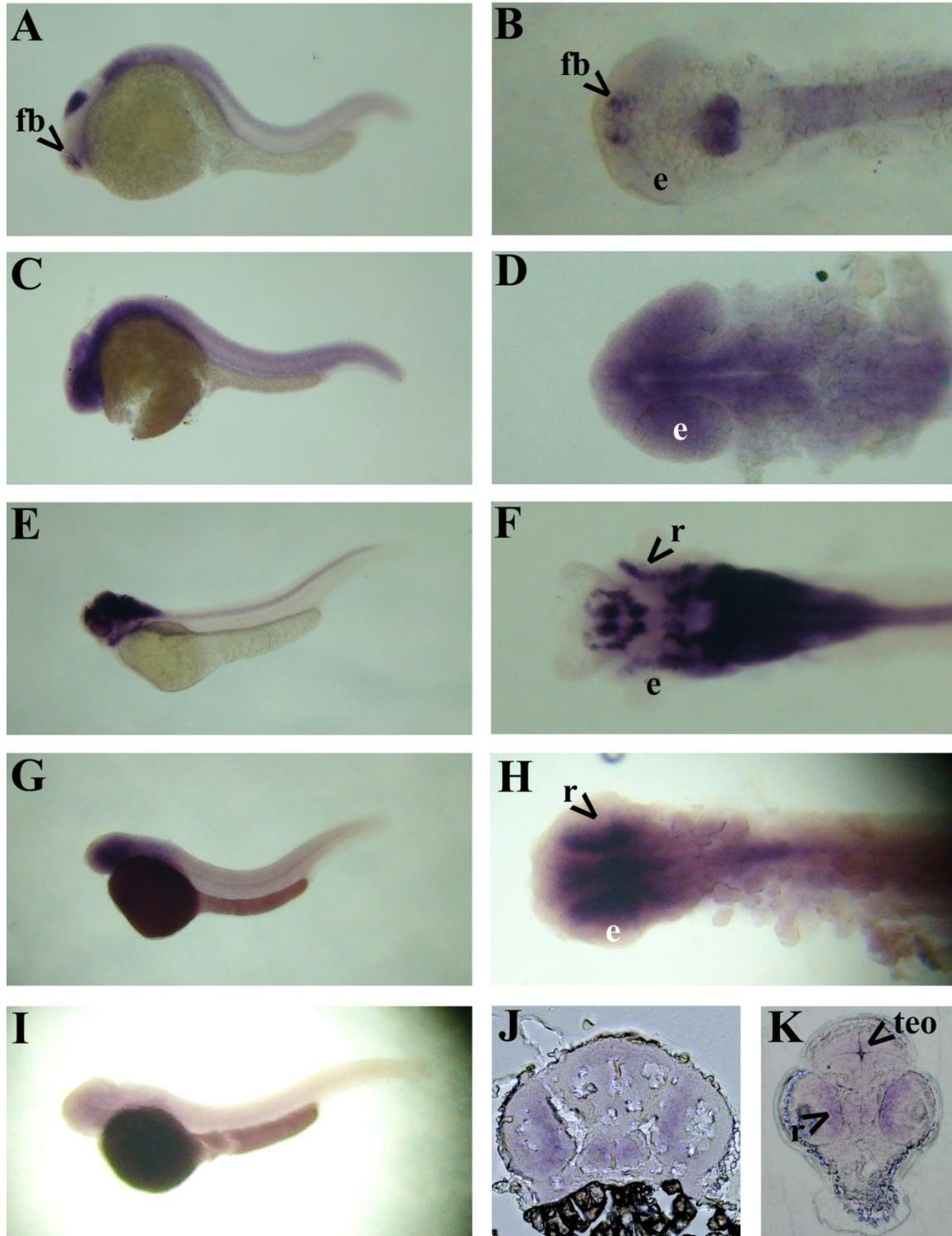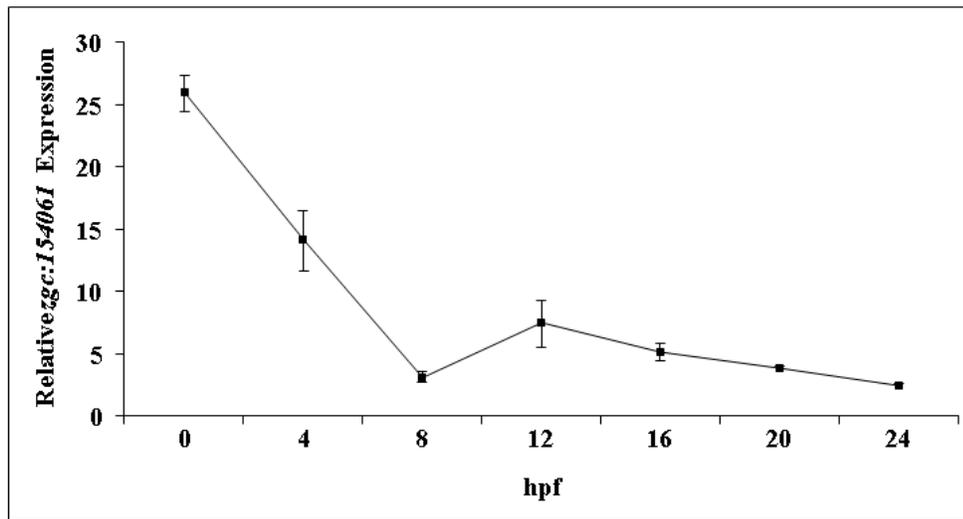
17

Fig. 3

Fig. 4

Fig. 5

Fig. 6

REFERENCES

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.

Chomczynski, P., and Mackey, K., 1995. Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. Biotechniques 19, 942-945.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M.*, et al.*, 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36, W465-469.

Ellies, D.L., Stock, D.W., Hatch, G., Giroux, G., Weiss, K.M., and Ekker, M., 1997. Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish dlx genes. Genomics 45, 580-590.

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

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

Harding, K., Wedeen, C., McGinnis, W., and Levine, M., 1985. Spatially regulated expression of homeotic genes in Drosophila. Science 229, 1236-1242.

Irvine, S.Q., Cangiano, M.C., Millette, B.J., and Gutter, E.S., 2007. Non-overlapping expression patterns of the clustered Dll-A/B genes in the ascidian Ciona intestinalis. J Exp Zoolog B Mol Dev Evol 308, 428-441.

Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A.Z., Engstrom, P.G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K.*, et al.*, 2007. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res 17, 545-555.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F., 1995. Stages of embryonic development of the zebrafish. Dev Dyn 203, 253-310.

Lettice, L.A., Horikoshi, T., Heaney, S.J., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N.*, et al.*, 2002. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci U S A 99, 7548-7553.

Livak, K.J., and Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402-408.

Mann, R.S., 1997. Why are Hox genes clustered? Bioessays 19, 661-664.

McCurley, A.T., and Callard, G.V., 2008. Characterization of housekeeping genes in zebrafish: male-female differences and effects of tissue type, developmental stage and chemical treatment. BMC Mol Biol 9, 102.

Prince, V., 2002. The Hox Paradox: More complex(es) than imagined. Dev Biol 249, 1-15.

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

Sharpe, J., Nonchev, 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.

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F*., et al.*, 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl Acad Sci U S A 99, 16899-16903.

Thisse, C., and Thisse, B., 2005. High Throughput Expression Analysis of ZF-Models Consortium Clones. In ZFIN Direct Data Submission (http://zfin.org).

Thisse, C., and Thisse, B., 2008. High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc 3, 59-69.

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.

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

BIOGRAPHICAL SKETCH

Brandon Scott Carpenter was born in Gastonia, North Carolina on June 18, 1984. He attended elementary schools in Cherryville, North Carolina and graduated as valedictorian of his senior class from Cherryville High School in 2002. The following autumn he entered Appalachian State University to study Biology, and in May of 2007 he graduated magna cum laude as the Outstanding Biology Senior with a B.S. in Biology and a minor in Chemistry. In the fall of 2007, he remained at Appalachian State University and began study towards a M.S. degree with a concentration in Cell and Molecular Biology. After receiving his M.S. in Biology from Appalachian State University in May of 2010, Brandon commenced work toward his Ph.D in Cell and Developmental Biology at the University of Michigan Ann Arbor.

Brandon is the son of Gary Carpenter of Belmont, North Carolina and Donna Thomas of Cherryville, North Carolina and is married to Heather Atwell Carpenter. Since entering Appalachian State University to study Biology in 2002, Brandon has been determined to research cancer as a result of his childhood battle with leukemia. In the future, he plans to attain his goal of becoming an independent research scientist and apply his knowledge and skills towards saving the lives of cancer patients.