

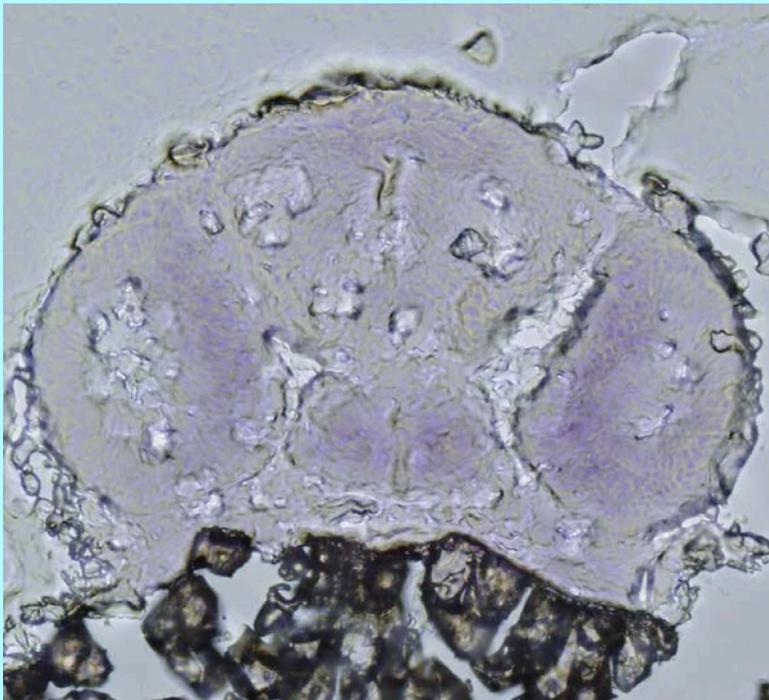
# Eastern Biologist

No. 5

2016

## Identification and Developmental Expression of the Zebrafish *zgc:154061* Gene

Brandon S. Carpenter, Brantley Graham, and  
Ted Zerucha



## ***The Eastern Biologist . . .***

- ◆ A peer-reviewed journal that publishes original articles focused on the many diverse disciplines of biological research, except for the natural history science disciplines (ISSN 2165-6657 [online]).
- ◆ Subject areas - The journal welcomes manuscripts based on original research and observations, as well as research summaries and general interest articles. Subject areas include, but are not limited to, biochemistry, biotechnology, cell biology, developmental biology, genetics and genomics, immunology, microbiology, molecular evolution, neurobiology, parasitology, physiology, toxicology, as well as scientific pedagogy.
- ◆ Offers article-by-article online publication for prompt distribution to a global audience
- ◆ Offers authors the option of publishing large files such as data tables, audio and video clips, and even PowerPoint presentations as online supplemental files.
- ◆ Special issues - The Eastern Biologist welcomes proposals for special issues that are based on conference proceedings or on a series of invitational articles. Special issue editors can rely on the publisher's years of experiences in efficiently handling most details relating to the publication of special issues.
- ◆ Indexing - As is the case with Eagle Hill's other journals, the Eastern Biologist is expected to be fully indexed in Elsevier, Thomson Reuters, Proquest, EBSCO, Google Scholar, and other databases.
- ◆ The journal staff is pleased to discuss ideas for manuscripts and to assist during all stages of manuscript preparation. The journal has a mandatory page charge to help defray a portion of the costs of publishing the manuscript. Instructions for Authors are available online on the journal's website (<http://www.eaglehill.us/ebio>) or by e-mail ([office@eaglehill.us](mailto:office@eaglehill.us)).
- ◆ Co-published with the *Northeastern Naturalist*, *Southeastern Naturalist*, *Caribbean Naturalist*, and *Urban Naturalist*. Each journal has a separate Board of Editors.
- ◆ Available online in full-text version on the journal's website (<http://www.eaglehill.us/ebio>). Arrangements for inclusion in other BioOne databases are being pursued.

**Cover Photograph:** Transverse section through the head of a 48-hour-old zebrafish embryo showing *zgc:154061* expression in the brain and retina by in situ hybridization

# *EASTERN BIOLOGIST*

---

---

## **Board of Editors**

Brian Burnes, Department of Biology, University of West Alabama, Livingston, AL

Judy Awong-Taylor, School of Science and Technology, Georgia Gwinnett College, Lawrenceville, GA

Chris R. Gissendanner, College of Pharmacy, University of Louisiana, Monroe, LA

Robert Gregerson, Florida Gulf Coast University, College of Arts & Sciences, Fort Myers, FL

Jill Lee, Eagle Hill Institute, Steuben, ME ... **Assistant Editor**

Tim H. Lindblom, Math and Science Division, Lyon College, Batesville, AR ... **Managing Editor**

Joerg-Henner Lotze, Eagle Hill Institute, Steuben, ME ... **Publisher**

Christi Magrath, Department of Biological and Environmental Sciences, Troy University, Troy, AL

Marilyn S. Mayer, Eagle Hill Institute, Steuben, ME ... **Production Editor**

Scott Merrell, Department of Microbiology and Immunology, Uniformed Services, University of the Health Sciences, Bethesda, MD

Ashley B. Morris, Department of Biology, Middle Tennessee State University, Murfreesboro, TN

J. Kenneth Shull, Department of Biology, Appalachian State University, Boone, NC (Emeritus)

Stephanie R. Songer, Department of Biology, North Georgia College and State University, Dahlonega, GA

Darla J. Wise, Biology Department, Concord University, Athens, WV

Dwayne A. Wise, Department of Biological Sciences, Mississippi State University, Mississippi State, MS

## Identification and Developmental Expression of the Zebrafish *zgc:154061* Gene

Brandon S. Carpenter<sup>1,2</sup>, Brantley Graham<sup>1,3</sup>, and Ted Zerucha<sup>1,\*</sup>

**Abstract** - We identified a previously undescribed gene in *Danio rerio* (Zebrafish), *zgc:154061*, located directly downstream of the *meis2a* gene. We also identified putative orthologs of this gene in all animal genomes that we could examine. The ortholog of this gene in *Homo sapiens* (L)(Human), *C15ORF41*, has recently been shown to play a role in erythropoiesis and has been suggested to be a member of the PD-(D/E)XK superfamily of nuclease proteins. The *zgc:154061* gene and its orthologs were organized in a convergently transcribed manner with respect to the *Meis2* gene (*meis2a* in Zebrafish) in the 6 vertebrate species we examined. During zebrafish development, transcripts of *zgc:154061* were observed in every cell of the embryo from the earliest stage through the shield stage indicating that this gene is a maternal transcript since it was present prior to the activation of the zygotic genome at the midblastula transition. Transcripts of *zgc:154061* mRNA gradually decreased from their peak value at 0 hpf until 8 hpf, and then gene expression was observed again at 12 hpf throughout the neural tube before becoming restricted to the brain, retina and developing olfactory region by 2 days of development.

### Introduction

The congenital dyserythropoietic anemias (CDAs) make up a group of rare blood disorders that affect erythropoiesis in humans (Wickramasinghe, S.N. and W.G. Wood 2005). These disorders are marked by ineffective erythropoiesis and often by erythroblasts that feature abnormal morphologies (Wickramasinghe, S.N. and W.G. Wood 2005). The CDAs have been classified into at least 7 types based upon characteristics that include erythroblast morphologies and serological characteristics (Heimpel, H. et al. 1968, Wickramasinghe, S.N. 1998, Wickramasinghe, S.N. 2000, Wickramasinghe, S.N. et al. 1998). Congenital dyserythropoietic anemia type I (CDA I) is inherited as an autosomal recessive disease that is characterized by macrocytosis and erythroblasts that are megaloblastic, exhibit internuclear chromatin bridges and also heterochromatin with a spongy appearance (Wickramasinghe, S.N. and W.G. Wood 2005). Mutations associated with CDA I have been mapped to the gene *Codanin-1* (*CDANI*) (Dgany, O. et al. 2002) as well as to *C15ORF41* (Babbs, C. et al. 2013). Interestingly, both of these genes are found on chromosome 15 in humans. The product of *CDANI* is hypothesized to associate with microtubules

---

<sup>1</sup>Cell and Molecular Biology Concentration, Department of Biology, Appalachian State University, Boone, NC 28608. <sup>2</sup>Present address: Department of Cell Biology, Emory University School of Medicine, Atlanta, GA 30322. <sup>3</sup>Current address – Spinal Cord and Brain Injury Research Center, University of Kentucky, Lexington, KY 40506. \*Corresponding author – zeruchat@appstate.edu.

Manuscript Editor: Robert Gregerson

and, through this association, to be involved in nuclear envelope integrity (Dgany, O. et al. 2002). The *C15ORF41* gene has been suggested to code for the production of a protein that contains two adjacent winged helix-turn-helix DNA-binding domains and to be a member of the PD-(D/E)XK superfamily of nuclease proteins (Babbs, C. et al. 2013). Members of this group of nuclease proteins are known to play roles in the repair of DNA damage, as Holliday junction resolvases and in RNA processing (Babbs, C. et al. 2013, Laganeckas, M. et al. 2011). The broad proposed functions of *CDANI* and *C15ORF41* suggest that both of these genes may have other biological relevance in addition to erythropoiesis.

We identified a novel, conserved gene in *Danio rerio* (Hamilton) (Zebrafish) that we propose to be an ortholog of *C15ORF41*, *zgc:154061*. We studied its location within the zebrafish genome and its orientation relative to the homeobox containing the *meis2a* gene. We also looked for homologs of this gene in every animal genome that we could examine and studied their position and orientation. Here we report the first spatial and temporal expression pattern for any ortholog of *C15ORF41* / *zgc:154061*, in zebrafish

## Methods

### Zebrafish care and maintenance

Zebrafish embryos were obtained by pair-wise breeding of adult AB\* fish housed in an enclosed Z-Mod system (Aquatic Habitats) and maintained according to standard conditions (Westerfield, M. 2000). Embryos were staged according to Kimmel et al. (Kimmel, C.B. et al. 1995). All experimental procedures involving Zebrafish were carried out according to the National Institutes of Health (NIH) Office of Laboratory Animal Welfare Guide for the Care and Use of Laboratory Animals and approved by the Appalachian State University Institutional Animal Care and Use Committee (Permit Number: 15-05).

### Sequence analysis

Amino acid sequences for *zgc:154061* orthologs from commonly used model systems and species representing major chordate lineages were obtained from publicly available genome sequences through the NCBI database (accession numbers: *Homo sapiens* (L.) (Human) - NP\_001277162; *Mus musculus* (L.) (House Mouse) - XP\_001480310.1; *Gallus gallus domesticus* (L.) (Domestic Chicken) - NP\_001026371.1; *Xenopus laevis* (Daudin) (African Clawed Frog) - NP\_001090210.1; *Danio rerio* (Hamilton) (Zebrafish) - NP\_001070190.1; *Ciona intestinalis* (L.) (Vase tunicate) - XP\_002128870.1; *Branchiostoma floridae* (Hubbs) (Florida lancelet) - XP\_002221176.1; *Strongylocentrotus purpuratus* (Stimpson) (Purple sea urchin) - XP\_780639.1; *Drosophila melanogaster* (Meigen) (Common fruit fly) - NP\_648806.1; except that of *Takifugu rubripes* (Temminck and Schlegel) (Japanese puffer fish) which was obtained from the IMCB Fugu Genome Project database (gene:SINFRUG00000137928 transcript:SINFRUT00000146276). Amino acid alignments were generated using Vector NTI Advance Version 11.0 Align

X (Invitrogen, ThermoFisher Scientific) 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. The phylogenetic tree of *zgc:154061* orthologous amino acid sequences 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, A. et al. 2008).

### ***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, R.L. et al. 2002) and made available through Open Biosystems (AL) (Clone Id: 8334609, Accession: BC124527) from whom we obtained it. To utilize a smaller probe size for our *in situ* hybridizations, we isolated by PCR, a partial 393-bp fragment of the zebrafish *zgc:154061* gene (Fig. 1B) corresponding to positions 608-1000 of the full length cDNA clone using the primer sequences GGTCTGGAACATGAAGAC and CCTCATGCCATCAGAAAC (Fig. 1B). We performed PCR amplification with Phusion High-Fidelity DNA Polymerase (New England BioLabs, MA), subcloned the product into the pGEM-T Vector System (Promega, Madison WI) and confirmed by sequencing. We synthesized an antisense DIG-labelled riboprobe directly from this subclone using T7 RNA Polymerase after linearizing the plasmid with NotI. For a negative control, we synthesized a sense DIG-labelled riboprobe from the same construct using SP6 RNA Polymerase after linearizing the plasmid with SacII. We performed whole mount *in situ* hybridization as described by Thisse and Thisse (Thisse, C. and B. Thisse 2008). Embryos were photographed using a Leica MZ6 dissecting microscope, Leica DFC320 digital camera and the Leica Application Suite Version 3.3.1. For sectioning, we positioned 48 hpf embryos 1.5% agarose melted in 5% sucrose in PBS following *in situ* hybridization. We incubated agarose blocks in 30% sucrose in PBS at 4°C overnight. Agar blocks were frozen in optical cutting temperature (OCT) media, and 20µm sections were cut using a Leica CM 1100 cryostat. We captured images of sections using an Olympus IX81 inverted microscope and processed them with MicroSuite Biological Suite software (Soft Imaging System GmbH).

### **RNA extraction and quantitative real-time PCR**

Total RNA was extracted from 30–100 staged embryos by homogenizing them with RNase-free, DNase-free, pyrogen-free disposable pestles (Kontes, NJ) in TRIzol (Invitrogen) following the protocol described by Chomczynski and Mackey (Chomczynski, P. and K. Mackey 1995). We determined RNA quality and quantity by NanoDrop (DE) and denaturing gel electrophoresis. Reverse transcription of RNA samples into cDNA were performed using the Applied Biosystems (Thermo-

Fisher Scientific) High Capacity RNA-to-cDNA Master Mix following the manufacturer's instructions.

We performed relative quantitative real-time PCR to analyze temporal expression of *zgc:154061* during zebrafish development using the Applied Biosystems 7300 real-time PCR System. For each stage examined, we used 200 ng of cDNA as a 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' (Fig. 1B). We used the zebrafish  $\beta$ -actin gene as an internal control with gene specific primers and probe:  $\beta$ -actin forward 5'-GCTGTTTTCCCCTCCATTGTTG-3';  $\beta$ -actin reverse 5'-TTTCTGTCCCATGCCAACCAT-3';  $\beta$ -actin probe 5'-FAM-CCCAGACATCAGGGAGTG-TAMRA-3'. Primers and probes were designed using Primer Express (Applied Biosystems) and purchased from Eurofins MWG Operon (LOCATION). We used the following amplification protocol: 50°C for 2 min; 95°C for 10 min; 95°C for 15 s, 60°C for 1 min repeated for 40 cycles. We performed all reactions in triplicate, repeated this and also examined two separate RNA extractions from each stage of development. We expressed results as level of relative expression calibrated to expression in an adult Zebrafish using the  $2^{-\Delta\Delta C_t}$  method (Livak, K.J. and T.D. Schmittgen 2001).

## Results

### Identification of *zgc:154061* and orthologs

In Zebrafish, the *zgc:154061* gene was found directly downstream of the *meis2a* gene on chromosome 17 (Fig. 1A) with 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 *meis2a* and is predicted to be encoded by 10 exons using the NCBI Model Maker (<http://www.ncbi.nlm.nih.gov>) tool for genomic sequence analysis. The sequence of the cDNA clone

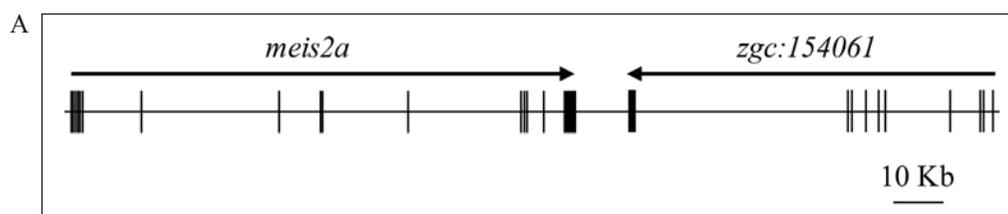


Figure 1. Genomic organization and sequence of *zgc:154061*. (A) Position of *zgc:154061* with respect to *meis2a* in zebrafish. Vertical lines/boxes represent exons. Arrows indicate the direction of transcription. The predicted exons of the *zgc:154061* gene and its location with respect to *meis2a* were identified using the publicly available genome sequence and the Model Maker tool in the NCBI Database. (Contd.)

B

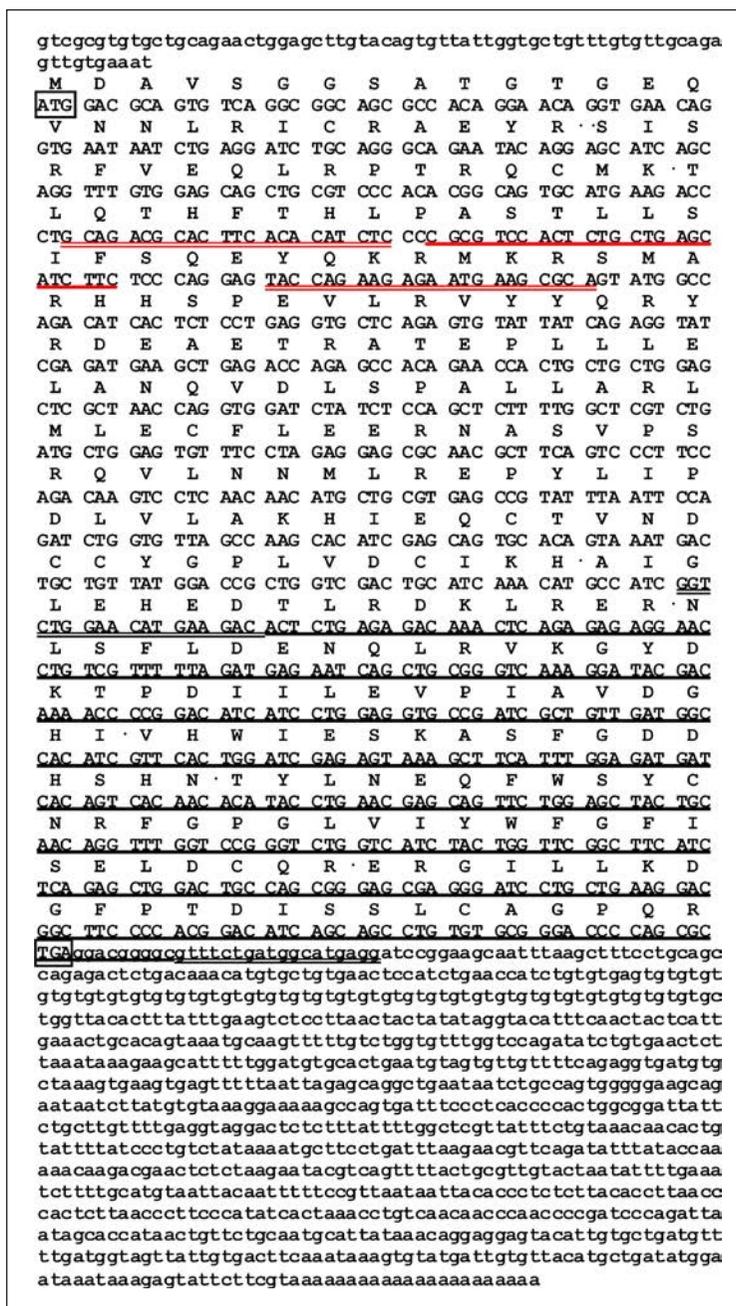


Figure 1 (contd.). (B) Nucleotide and translated sequence of the *zgc:154061* gene. The sequence used as a probe for *in situ* hybridization is underlined and the primer sites used to amplify this sequence are double underlined. The primer sites used for quantitative real-time PCR are indicated by red 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.

(Fig. 1B) confirmed the predicted splicing pattern. Based on BLAST analysis (<http://www.ncbi.nlm.nih.gov>) against the available zebrafish genome, there is not a second, predicted paralog of *zgc:154061* associated with the second zebrafish *Meis2* homolog, *meis2b*, or elsewhere in the zebrafish genome. The *zgc:154061* gene encoded 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, R.L. et al. 2002). We obtained a cDNA clone of the gene from Open Biosystems.

We identified predicted orthologs of *zgc:154061* in all vertebrates with publicly available genome data that we could obtain. These vertebrate orthologs were present in the same convergently transcribed orientation, directly adjacent to that species' *Meis2* ortholog, although the spatial separation of the two genes varied among species (data not shown). The zebrafish *zgc:154061* was most similar to the *Takifugu rubripes* ortholog, and the predicted zebrafish protein represented the longest protein from the representative orthologs examined (Fig. 2A, B). Furthermore, the *zgc:154061* orthologs displayed variability within their amino-termini, some of which contained different translation start points (Fig. 2A). The zebrafish protein included a 23-aa amino-terminal amino acids not found in tetrapods (Fig. 2A). In addition, the chicken protein appeared to have an 11-aa deletion compared to the other species examined (Fig. 2A). The most well conserved region within the amino-terminal portion of the predicted zebrafish protein lay between amino acids 23 and 125 (Fig. 2A). The carboxyl-terminal regions of the C15ORF41 protein were also very well-conserved across all species examined (Figure 2A).

We identified putative homologs of *zgc:154061* in the urochordate *Ciona intestinalis*, the cephalochordate *Branchiostoma floridae* and the echinoderm *Strongylocentrotus purpuratus* (Fig. 2B). In *C. intestinalis*, based on the available genome data, the *zgc:154061* ortholog and the *Meis* ortholog were on different chromosomes. However, in the *B. floridae* genome, the *zgc:154061* and *Meis* orthologs were found adjacently on scaffold 120. Furthermore, in *S. purpuratus*, the two genes were found adjacent to one another, but they appeared to be transcribed from the same strand of DNA as opposed to the convergently transcribed orientation observed in vertebrates. Based on sequence similarities, we also identified putative homologs of *zgc:154061* in a number of invertebrate species, including *Drosophila melanogaster* (Table 1 and Figure 2B). It appeared that all non-vertebrate, putative orthologs of the product of *zgc:154061* shared several short conserved domains in the carboxyl-terminal region that are all within the proposed nuclease domain of C15ORF41 (Babbs, C. et al. 2013) (Figure 2A). However, none of the protostome orthologs appeared to be adjacent to the invertebrate *Meis* ortholog in any of these species.

### **Spatial and temporal expression of *zgc:154061* and orthologs**

Our examination of the spatial and temporal expression of *zgc:154061* using whole mount *in situ* hybridization (Figs. 3 and 4) and quantitative real-time PCR (Fig. 5) between the zygote and early larva periods indicated that transcripts of

A

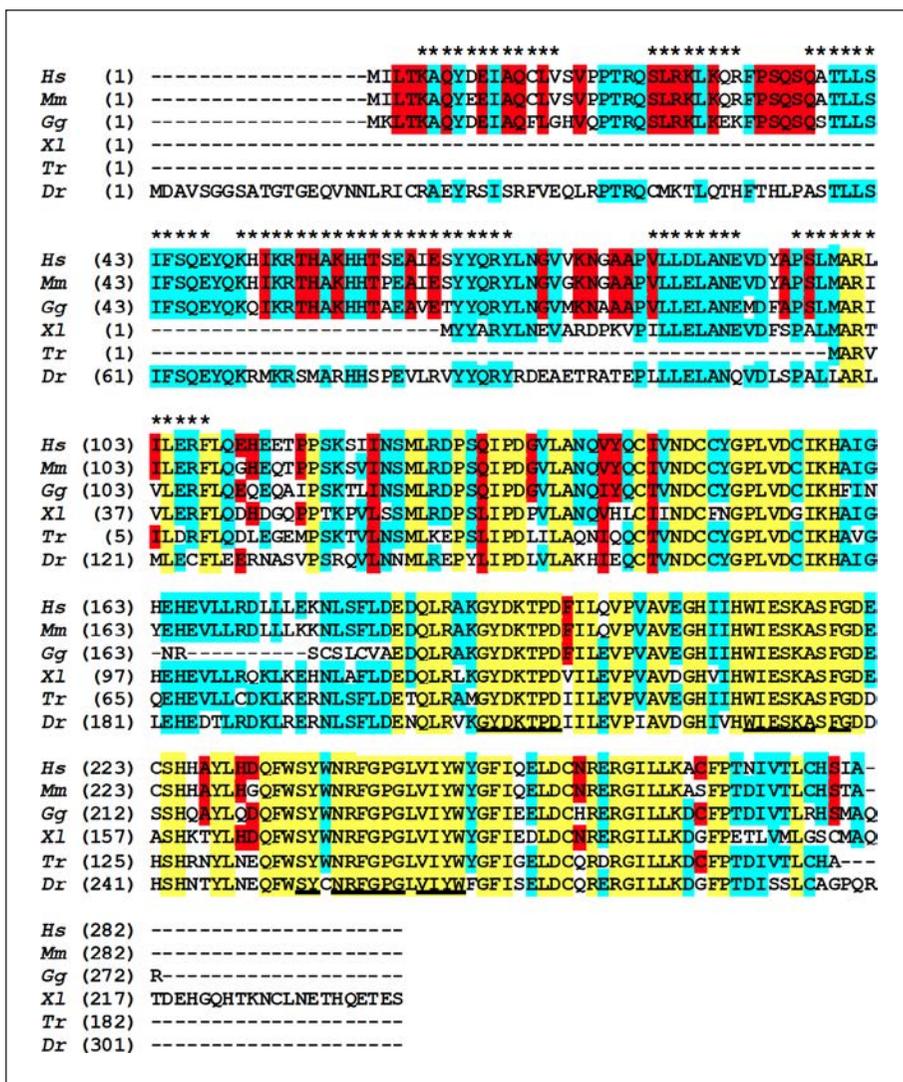


Figure 2. Amino acid sequence comparisons of *zgc*:154061 orthologs. (A) Amino acid sequence alignment of *zgc*:154061 in: Human (Hs, *Homo sapiens* – C15orf41), House Mouse (Mm, *Mus musculus* - hypothetical protein LOC399568), Domestic chicken (Gg, *Gallus gallus domesticus*- hypothetical protein LOC423293), African Clawed Frog (Xl, *Xenopus laevis* - hypothetical protein LOC779112), Japanese pufferfish (Tr, *Takifugu rubripes* – translation of SINFRUT00000146276 ), and Zebrafish (Dr, *Danio rerio* – *zgc*: 154061). Yellow shaded regions represent complete identity amongst all six species, blue shading indicates identity between four to five of the six species shown and red shading indicates identity between three of the species shown. Predicted locations of amino acids that make up the  $\alpha$ -helices that contribute to the two DNA-binding winged helix-turn-turn helix domains (Babbs, C. et al. 2013) are indicated with an “\*”. Underlined regions indicate domains that are also found in putative invertebrate orthologs of *zgc*:154061 and that also represent the proposed nuclease domain (Babbs, C. et al. 2013). (Contd.)

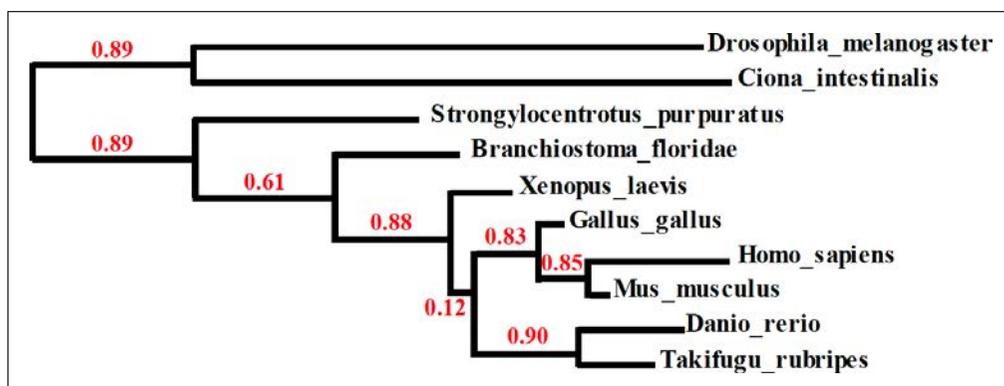


Figure 2 (contd.). (B) The phylogenetic tree of zgc:154061 orthologous amino acid sequences. Numbers on branches represent the percentage of how many times clades grouped following 500 replications. Representative species examined are: Human (*Homo sapien*), House Mouse (*Mus musculus*), Domestic chicken (*Gallus gallus domesticus*), African Clawed Frog (*Xenopus laevis*), Zebrafish (*Danio rerio*), Japanese pufferfish (*Takifugu rubripes*), Vase tunicate (*Ciona intestinalis*), Florida lancelet (*Branchiostoma floridae*), Purple sea urchin (*Strongylocentrotus purpuratus*), Common fruit fly (*Drosophila melanogaster*).

Table 1. Amino acid sequence identity percentages of zgc:154061 homologs. Amino acid sequence identity percentages based on pairwise comparisons between putative, representative homologs 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 sapiens* - Hs); house mouse (*Mus musculus* - Mm); chicken (*Gallus gallus domesticus*- Gg); African clawed frog (*Xenopus laevis* - XI); zebrafish (*Danio rerio* - Dr); Japanese pufferfish (*Takifugu rubripes* - Tr); vase tunicate (*Ciona intestinalis* - Ci); Florida lancelet (*Branchiostoma floridae* - Bf); purple sea urchin (*Strongylocentrotus purpuratus* - Sp); common fruit fly (*Drosophila melanogaster* - Dm).

	Hs	Mm	Gg	XI	Dr	Tr	Ci	Bf	Sp	Dm
Hs	100	94	82	76	70	77	42	52	51	38
Mm		100	82	72	61	77	34	44	50	27
Gg			100	68	59	74	32	42	58	24
XI				100	64	70	38	44	52	31
Dr					100	76	34	41	50	26
Tr						100	42	49	50	38
Ci							100	36	41	28
Bf								100	54	33
Sp									100	36
Dm										100

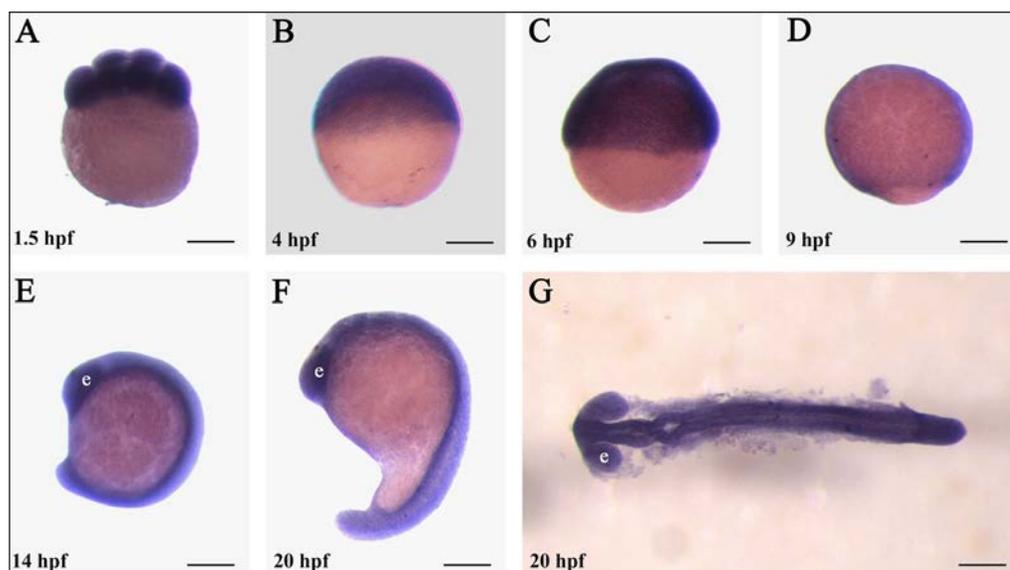


Figure 3. Early zebrafish *zgc:154061* expression. Whole-mount *in situ* hybridization analysis of early zebrafish *zgc:154061* expression. Embryos are shown as whole mounts with the anterior to the left. (A) 16-cell (1.5 hpf), (B) 4 hpf, (C) 6 hpf, (D) 9 hpf (90% epiboly), (E) 14 hpf, (F) 20 hpf, and (G) 20 hpf dorsal view with yolk removed. e, eye. Scale bars, 250  $\mu$ m

*zgc:154061* occurred at the one cell stage and in all progenitor cells. Transcripts were observed ubiquitously in every cell beyond the onset of transcription of the zygotic genome, at the midblastula transition to the shield stage which corresponds to 6 hpf and 50% epiboly during gastrulation (Fig. 3A-C). By 9 hpf, expression of *zgc:154061* was observed to be very low and almost undetectable by whole mount *in situ* hybridization (Fig. 3D). These observations were supported by quantitative real-time PCR where we saw 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. 4).

Following epiboly, low expression of *zgc:154061* was again observed during the segmentation period, and this later expression peaked at 12 hpf based on quantitative real-time PCR (Fig. 5). Whole mount *in situ* hybridization of *zgc:154061* revealed that its expression was largely restricted anteriorly to the developing neural tube at 24 h, with no obvious expression observed in the ICM during this time frame (Fig. 3E-G and Fig. 4C, D). We observed expression of *zgc:154061* throughout the neural tube and eye until 24 hpf (Fig. 3E-G and Fig. 4C, D), becoming more anteriorly restricted by 48 h of development where faint expression was observed throughout the forebrain, the retina, the olfactory region and the tectum of the mid-brain (Fig. 4G-K). After 48 hpf, expression gradually decreased in these areas to below observable levels for the remainder of development (data not shown).

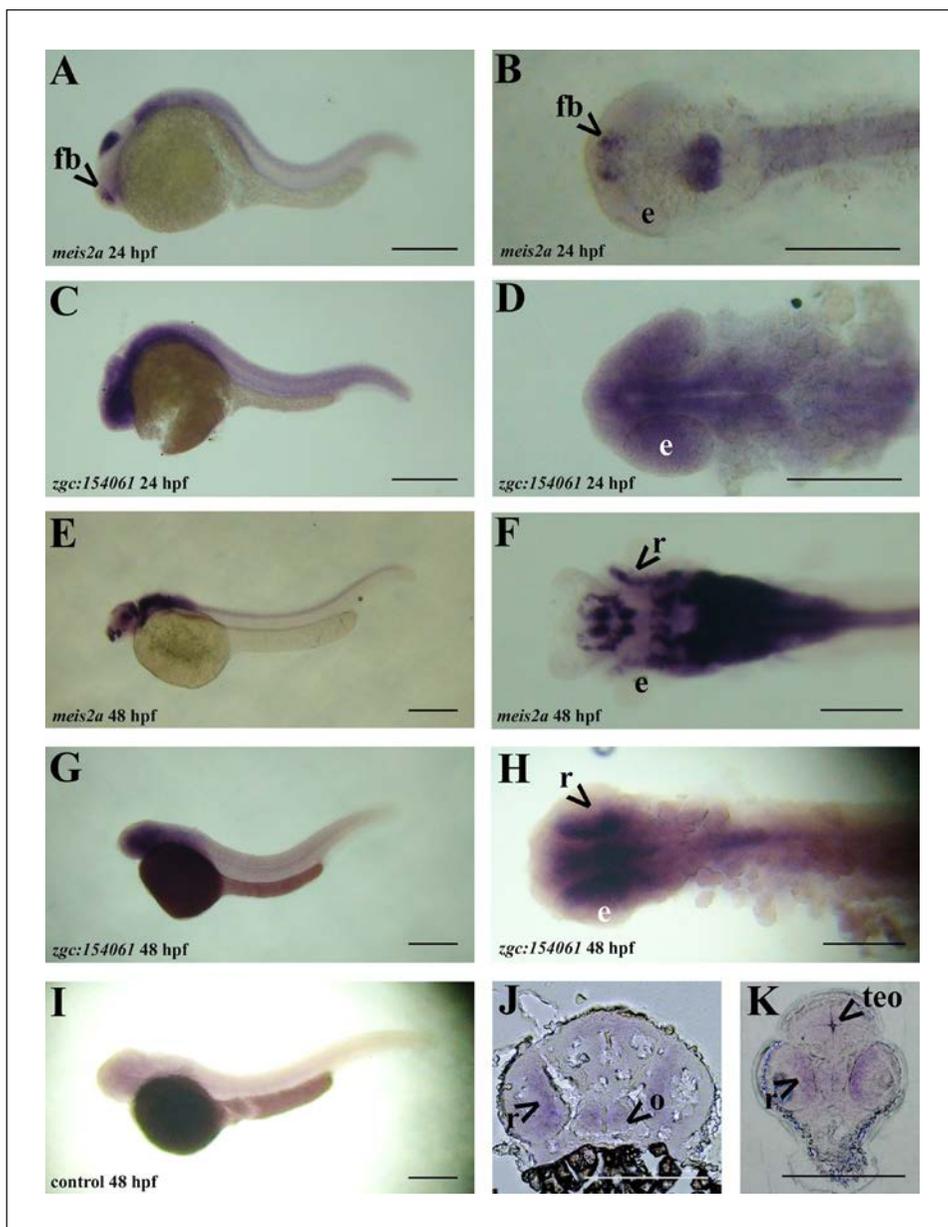


Figure 4. Late zebrafish *zgc:154061* and *meis2a* expression. Whole-mount *in situ* hybridization analysis of late zebrafish *zgc:154061* and *meis2a* expression. Embryos are shown as whole mounts the anterior to the left. (A) 24 hpf *meis2a*, (B) 24 hpf *meis2a* dorsal view with yolk removed, (C) 24 hpf *zgc:154061*, (D) 24 hpf *zgc:154061* dorsal view with yolk removed, (E) 48 hpf *meis2a*, (F) 48 hpf *meis2a* 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; o, olfactory region; r, retina; teo, tectum opticum. Scale bars, 250  $\mu$ m

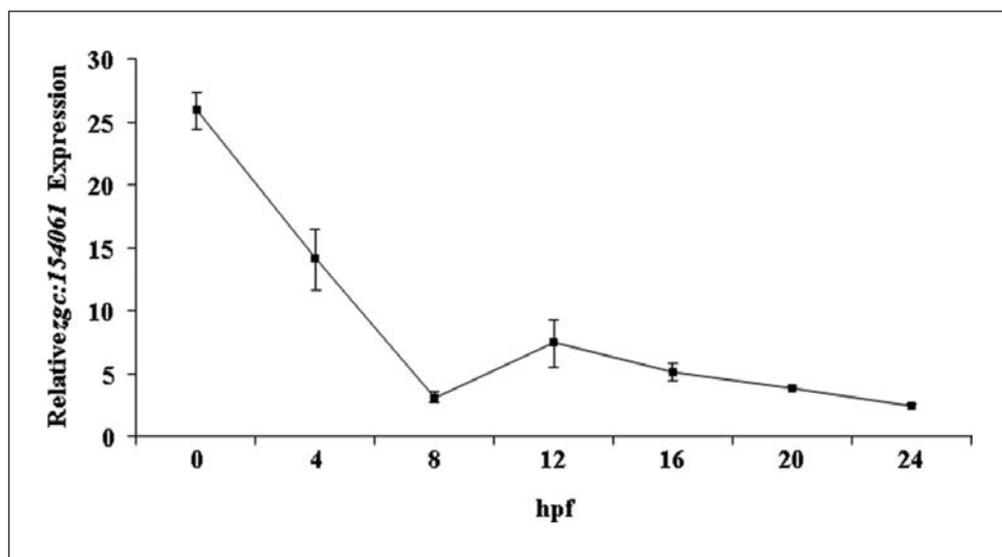


Figure 5. Quantitative real-time PCR analysis of *zgc:154061* expression. Total zebrafish mRNA was isolated from 30-100, 0 hpf to 120 hpf embryos. Expression of *zgc:154061* relative to that of the  $\beta$ -actin endogenous control at each developmental stage indicated is shown, calibrated to expression in an adult Zebrafish using the  $2^{-\Delta\Delta C_t}$  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 ( $p < 0.05$ ).

## Discussion

We identified a novel, conserved gene in Zebrafish, *zgc:154061*, that was located directly downstream of the homeobox containing *meis2a* gene with the two genes organized in an inverted, convergently transcribed orientation in the genome. We also identified predicted orthologs of *zgc:154061* in all animals with publicly available genome data that we could obtain. In vertebrates, *zgc:154061* orthologs were present in the same convergently transcribed orientation directly adjacent to that species' *Meis2* ortholog. In nonvertebrate chordates, this genomic organization was less well conserved. In the Florida lancelet and Purple sea urchin genomes, the *zgc:154061* and *Meis2* orthologs were found adjacent to one another in their genomes. However, in Purple sea urchin, the two genes appeared to be transcribed from the same strand of DNA as opposed to the convergently transcribed orientation observed in vertebrates. In Vase tunicate, the *zgc:154061* ortholog and the *Meis* ortholog were on different chromosomes, and this was also the case for invertebrates such as Common fruit fly. 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.

When comparing the orthologs of the predicted *zgc:154061* protein, the most well conserved region corresponded to that from amino acids 23 to 125 of the zebrafish protein (Fig. 2A). This region is proposed to contain the DNA-binding

domain of C15ORF41. The human C15ORF41 protein has been proposed to bind DNA via two adjacent winged helix-turn-helix domains formed from six  $\alpha$ -helices (Babbs, C. et al. 2013) that are indicated in Figure 2A. The regions of the amino-terminal portion of the predicted zebrafish protein that correspond to helices 3, 5 and 6 of the human protein were strikingly well conserved between mammals, chicken and zebrafish. This is somewhat expected as helices 4 and 6 have been proposed to directly interact with DNA (Babbs, C. et al. 2013). Thus the similarity of these helices would be important for the conservation across species of the specificity of the DNA recognition by this protein. The carboxyl-terminal region of the C15ORF41 protein has been proposed to contain a nuclease domain (Babbs, C. et al. 2013) and this region is also well conserved across species.

Transcripts of *zgc:154061* were observed throughout the zebrafish embryo at the earliest developmental stages. The developmental stages of zebrafish can be divided into seven broad periods: the zygote period (0-0.75 hpf), the cleavage period (0.75-2.25 hpf), the blastula period (2.25-5.25 hpf), the gastrula period (5.25-10 hpf), the segmentation period (10-24 hpf), the pharyngula period (24-48 hpf), and the hatching / early larval period (48-72 hpf) (Kimmel, C.B. et al. 1995). The observation of transcripts of *zgc:154061* during the zygote, cleavage and early blastula periods indicates that this gene is maternally expressed since the zygotic genome does not become active until the midblastula transition at approximately 2.75 hpf. Levels of these transcripts steadily decreased until 9 hpf. The steady decrease in quantities of transcripts of a maternally transcribed gene has not been described extensively in the zebrafish literature, and we have only been able to identify four other genes that exhibit even somewhat similar expression in the early zebrafish embryo: the transforming growth factor receptor *tgf $\beta$ 2* (Albertson, R.C. et al. 2005), the androgen receptor (*ar*) (Pikulkaew, S. et al. 2009), *bucky ball* (Bontems, F. et al. 2009), and *march8* (Kim, M.H. et al. 2014). This pattern of expression is consistent with that of a maternal transcript that is present in the zygote and then gradually disappears over time since it is not being replenished by new transcriptional activity. This expression dynamic is not unexpected, because at least the first 3 h of zebrafish development are not accompanied by transcription of the embryo's genome. However, it is interesting that this pattern of expression appears to only be observed infrequently. Transcripts of the zebrafish  $\beta$ -actin gene, used as an 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, A.T. and G.V. Callard 2008). The decrease in expression of *zgc:154061* relative to the steady expression of  $\beta$ -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.

We were surprised that the expression of *zgc:154061* observed later in development, following epiboly and peaking at 12 hpf was largely restricted to the develop-

ing neural tube, becoming more anteriorly restricted to the forebrain, the retina, the olfactory region and the tectum of the midbrain as development proceeded. The time period following the two-somite stage (~10.7 hpf) is when transcription factors required for hematopoiesis are first expressed as reviewed in Kulkeaw and Sugiyama (Kulkeaw, K. and D. Sugiyama 2012). In addition, between 12 and 24 hpf is when the anterior intermediate cell mass (ICM) develops anteriorly into myeloid cells and posteriorly into erythroid and some myeloid cells. We had expected to observe *zgc:154061* expression in the ICM based on the role that the human ortholog of this gene has been shown to play in erythropoiesis (Babbs, C. et al. 2013). The expression pattern of *zgc:154061* at 24 hpf, however, was remarkably similar to that of *slc22a5*, a gene with a demonstrated role in zebrafish erythropoiesis despite having no observable expression in the ICM as determined by the same method of *in situ* (Nilsson, R. et al. 2009). This lack of observed expression of *slc22a5* in the ICM has been suggested to be due to it being expressed at relatively low levels that are below the threshold detectable by *in situ* hybridization (Nilsson, R. et al. 2009). We cannot rule this out as an explanation for the lack of observable *zgc:154061* expression in presumptive erythropoietic cells of the ICM.

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 (Gould, A. et al. 1997, Harding, K. et al. 1985, Mann, R.S. 1997, Prince, V. 2002, Sharpe, J. et al. 1998) as well as the clustered organization of the *Dlx* gene family, for example, which are also found as convergently transcribed gene pairs in vertebrates (Ellies, D.L. et al. 1997, Zerucha, T. 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. We were particularly excited to observe expression of *zgc:154061* in the retina and olfactory placode, because *Meis2* has been shown to play a role in eye development (Zhang, X. et al. 2002) as well as olfactory neurogenesis (Agoston, Z. et al. 2014). Comparing the expression of *zgc:154061* reported here to that of *meis2a* (Bessa, J. et al. 2008, Thisse, C., And Thisse, B. 2005, Waskiewicz, A.J. et al. 2001) (Figure 4A-H), it appears that there is also some overlap between these two genes in the anterior neural tube and brain at 24 and 48 hpf.

A number of conserved non-coding elements (CNEs) have recently been reported as being associated with the *Meis2* gene (*meis2a* in Zebrafish). Three elements are of particular interest, because they have been demonstrated to direct expression to regions of the hindbrain as well as cranial ganglia in developing Zebrafish in a manner consistent with aspects of both *zgc:154061* and *meis2a* expression (Parker, H.J. et al. 2011). These elements are found within introns of the *zgc:154061* gene (Parker, H.J. et al. 2011), making them excellent candidates as *cis*-regulatory elements that are perhaps being shared by *Meis2* (*meis2a* in Zebrafish) and *zgc:154061* and its vertebrate orthologs. Examples also exist, where tandem genes are thought to share *cis*-regulatory elements, but do not share complete overlapping patterns of

expression, likely because of interactions between shared enhancers and gene and developmental stage-specific repressors (Irvine, S.Q. et al. 2007). This latter possibility could help explain why we do not observe a complete overlap of expression of *meis2a* and *zgc:154061*. It is also possible that the organization of these genes has been preserved, because of the presence of undiscovered *cis*-regulatory elements that are shared by both genes. Such yet to be identified elements may be in the intergenic region separating these genes or 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 that is found in an intron of the *Lmbr1* locus (Lettice, L.A. et al. 2002).

It is surprising that the *zgc:154061* expression we report here does not include areas of the embryo associated with erythropoiesis despite the demonstrated role of the human homolog of this gene in erythropoiesis. As previously mentioned, this may simply be due to it being expressed at relatively low levels in presumptive erythropoietic cells, levels below the threshold detectable by *in situ* hybridization. The roles of the PD-(D/E)XK superfamily of nuclease proteins in repairing recombination errors, degrading DNA and processing RNA suggest that *zgc:154061* and its homologs may be playing a much broader role than being restricted to erythropoiesis. The relatively high expression levels that we observed of this gene during early development prior to the initiation of hematopoiesis gives credence to this. It is possible that the mutations in *C15ORF41* that lead to CDA I have a relatively small impact on the protein which is manifested as the disease and that other mutations that have a major impact on the protein's function are lethal because of this broader role. It is also possible that the role this gene plays may have diverged at some point after the splitting of the teleost and tetrapod lineages. Preliminary mRNA overexpression studies have been inconclusive (data not shown), but it will be interesting to see what insights future functional inactivation studies will provide.

In summary, we have described a novel zebrafish gene *zgc:154061* that was maternally expressed and also expressed in the developing central nervous system. This gene was well conserved evolutionarily amongst vertebrates in terms of sequence identity and linkage to the *Meis2* (*meis2a* in zebrafish) homeobox gene. It appears that the genomic organization of the *zgc:154061* and *Meis2* genes arose early in the deuterostome lineage, because it is also observed in cephalochordates and echinoderms. It is possible that the conservation of this genomic organization is due to the sharing of regulatory regions as there is some overlap in expression of *zgc:154061* and *Meis2*, particularly in the developing eye, olfactory region and anterior neural tube. Previously, it has been proposed that the human homolog of *zgc:154061* is a PD-(D/E)XK nuclease (Babbs, C. et al. 2013). The amino acid alignment of the vertebrate orthologs of *zgc:154061* indicates conserved amino-terminal and carboxyl-terminal regions consistent with the previously proposed DNA-binding domains and a nuclease domain respectively (Babbs, C. et al. 2013). While the functional role of *zgc:154061* and its orthologs remains to be investigated more fully, our results provide a critical step towards characterizing this gene.

### Acknowledgements

We would like to thank: Jim Sobieraj for his expert technical assistance with the real-time PCR, Monique Eckerd and the Appalachian State University (ASU) College of Arts and Sciences Animal Care Facility for expert assistance with caring for our zebrafish colony, Guichuan Hou, director of the ASU College of Arts and Sciences Microscopy Facility, Sue Edwards for her assistance with our initial sectioning experiments, and Allen Wellington for his contributions at the onset of this project.

This work was supported by: a Sigma Xi Grants In Aid of Research Award to Carpenter, an American Cancer Society Scholarship to Carpenter, a Cratis D. Williams Graduate Student Research Grant to Carpenter, ASU Office of Student Research Awards to Carpenter and Graham, ASU Graduate Student Association Senate Awards to Carpenter and Graham, an ASU Graduate Research Associate Mentoring Award to Zerucha for mentoring Carpenter, an ASU University Research Council Award to Zerucha, an ASU Cratis D. Williams Graduate School/ Office of Research and Sponsored Programs Research Development Award to Zerucha, and support from ASU College of Arts and Sciences and Department of Biology.

### Literature Cited

- Agoston, Z., P. Heine, M.S. Brill, B.M. Grebbin, A.C. Hau, W. Kallenborn-Gerhardt, J. Schramm, M. Gotz , and D. Schulte. 2014. Meis2 is a pax6 co-factor in neurogenesis and dopaminergic periglomerular fate specification in the adult olfactory bulb. *Development* 141(1):28–38.
- Albertson, R.C., T.L. Payne-Ferreira, J. Postlethwait , and P.C. Yelick. 2005. Zebrafish *acvr2a* and *acvr2b* exhibit distinct roles in craniofacial development. *Developmental Dynamics* 233(4):1405–1418.
- Babbs, C., N.A. Roberts, L. Sanchez-Pulido, S.J. MCGowan, M.R. Ahmed, J.M. Brown, M.A. Sabry, W.G.S. Consortium, D.R. Bentley, G.A. Mcvean, P. Donnelly, O. Gileadi, C.P. Ponting, D.R. Higgs , and V.J. Buckle. 2013. Homozygous mutations in a predicted endonuclease are a novel cause of congenital dyserythropoietic anemia type i. *Haematologica* 98(9):1383–1387.
- Bessa, J., M.J. Tavares, J. Santos, H. Kikuta, M. Laplante, T.S. Becker, J.L. Gomez-Skarmeta , and F. Casares. 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(5):799–803.
- Bontems, F., A. Stein, F. Marlow, J. Lyautey, T. Gupta, M.C. Mullins , and R. Dosch. 2009. Bucky ball organizes germ plasm assembly in zebrafish. *Current Biology* 19(5):414–422.
- Chomczynski, P., and K. Mackey. 1995. Short technical reports. Modification of the tri reagent procedure for isolation of rna from polysaccharide- and proteoglycan-rich sources. *Biotechniques* 19(6):942–945.
- Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J.F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J.M. Claverie , and O. Gascuel. 2008. Phylogeny.Fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* 36 (Web Server issue):W465–469.
- Dgany, O., N. Avidan, J. Delaunay, T. Krasnov, L. Shalmon, H. Shalev, T. Eidelitz-Markus, J. Kapelushnik, D. Cattani, A. Pariente, M. Tulliez, A. Cretien, P.O. Schischmanoff, A.

- Iolascon, E. Fibach, A. Koren, J. Rossler, M. Le Merrer, I. Yaniv, R. Zaizov, E. Ben-Asher, T. Olender, D. Lancet, J.S. Beckmann, and H. Tamary. 2002. Congenital dyserythropoietic anemia type i is caused by mutations in codanin-1. *American Journal of Human Genetics* 71(6):1467–1474.
- Ellies, D.L., D.W. Stock, G. Hatch, G. Giroux, K.M. Weiss, and M. Ekker. 1997. Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish *dlx* genes. *Genomics* 45(3):580–590.
- Gould, A., A. Morrison, G. Sproat, R.A. White, and R. Krumlauf. 1997. Positive cross-regulation and enhancer sharing: Two mechanisms for specifying overlapping hox expression patterns. *Genes & Development* 11(7):900–913.
- Harding, K., C. Wedeen, W. McGinnis, and M. Levine. 1985. Spatially regulated expression of homeotic genes in drosophila. *Science* 229(4719):1236–1242.
- Heimpel, H., F. Wendt, D. Klemm, H. Schubothe, and L. Heilmeyer. 1968. [congenital dyserythropoietic anemia]. *Archiv fur klinische Medizin* 215(2):174–194.
- Irvine, S.Q., M.C. Cangiano, B.J. Millette, and E.S. Gutter. 2007. Non-overlapping expression patterns of the clustered *dll-a/b* genes in the ascidian *ciona intestinalis*. *Journal of Experimental Zoology Part B Molecular and Developmental Evolution* 308(4):428–441.
- Kim, M.H., M.L. Rebbert, H. Ro, M. Won, and I.B. Dawid. 2014. Cell adhesion in zebrafish embryos is modulated by *march 8*. *PLoS ONE* 9(4):e94873.
- Kimmel, C.B., W.W. Ballard, S.R. Kimmel, B. Ullmann, and T.F. Schilling. 1995. Stages of embryonic development of the zebrafish. *Developmental Dynamics* 203(3): 253–310.
- Kulkaew, K., and D. Sugiyama. 2012. Zebrafish erythropoiesis and the utility of fish as models of anemia. *Stem Cell Research & Therapy* 3(6):55.
- Laganeckas, M., M. Margelevicius, and C. Venclovas. 2011. Identification of new homologs of *pd-(d/e)xk* nucleases by support vector machines trained on data derived from profile-profile alignments. *Nucleic Acids Research* 39(4):1187–1196.
- Lettice, L.A., T. Horikoshi, S.J. Heaney, M.J. Van Baren, H.C. Van Der Linde, G.J. Breedveld, M. Joosse, N. Akarsu, B.A. Oostra, N. Endo, M. Shibata, M. Suzuki, E. Takahashi, T. Shinka, Y. Nakahori, D. Ayusawa, K. Nakabayashi, S.W. Scherer, P. Heutink, R.E. Hill, and S. Noji. 2002. Disruption of a long-range cis-acting regulator for *shh* causes preaxial polydactyly. *Proceedings of the National Academy of Sciences of the United States of America* 99(11):7548–7553.
- Livak, K.J., and T.D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative pcr and the  $2^{-\Delta\Delta C_t}$  method. *Methods* 25(4):402–408.
- Mann, R.S. 1997. Why are hox genes clustered? *Bioessays* 19(8):661–664.
- Mccurley, A.T., and G.V. Callard. 2008. Characterization of housekeeping genes in zebrafish: Male-female differences and effects of tissue type, developmental stage and chemical treatment. *BioMed Central Molecular Biology* 9:102.
- Nilsson, R., I.J. Schultz, E.L. Pierce, K.A. Soltis, A. Naranuntarat, D.M. Ward, J.M. Baughman, P.N. Paradkar, P.D. Kingsley, V.C. Culotta, J. Kaplan, J. Palis, B.H. Paw, and V.K. Mootha. 2009. Discovery of genes essential for heme biosynthesis through large-scale gene expression analysis. *Cell Metabolism* 10(2):119–130.
- Parker, H.J., P. Piccinelli, T. Sauka-Spengler, M. Bronner, and G. Elgar. 2011. Ancient *pbx-hox* signatures define hundreds of vertebrate developmental enhancers. *BioMed Central Genomics* 12:637.
- Pikulkaew, S., A. De Nadai, P. Belvedere, L. Colombo, and L. Dalla Valle. 2009. Expression analysis of steroid hormone receptor mRNAs during zebrafish embryogenesis. *General and Comparative Endocrinology* 165(2):215–220.

- Prince, V. 2002. The hox paradox: More complex(es) than imagined. *Developmental Biology* 249(1):1–15.
- Sharpe, J., S. Nonchev, A. Gould, J. Whiting, and R. Krumlauf. 1998. Selectivity, sharing and competitive interactions in the regulation of hoxb genes. *The EMBO Journal* 17(6):1788–1798.
- Strausberg, R.L., E.A. Feingold, L.H. Grouse, J.G. Derge, R.D. Klausner, F.S. Collins, L. Wagner, C.M. Shenmen, G.D. Schuler, S.F. Altschul, B. Zeeberg, K.H. Buetow, C.F. Schaefer, N.K. Bhat, R.F. Hopkins, H. Jordan, T. Moore, S.I. Max, J. Wang, F. Hsieh, L. Diatchenko, K. Marusina, A.A. Farmer, G.M. Rubin, L. Hong, M. Stapleton, M.B. Soares, M.F. Bonaldo, T.L. Casavant, T.E. Scheetz, M.J. Brownstein, T.B. Usdin, S. Toshiyuki, P. Carninci, C. Prange, S.S. Raha, N.A. Loquellano, G.J. Peters, R.D. Abramson, S.J. Mullahy, S.A. Bosak, P.J. Mcewan, K.J. Mckernan, J.A. Malek, P.H. Gunaratne, S. Richards, K.C. Worley, S. Hale, A.M. Garcia, L.J. Gay, S.W. Hulyk, D.K. Villalon, D.M. Muzny, E.J. Sodergren, X. Lu, R.A. Gibbs, J. Fahey, E. Helton, M. Kettelman, A. Madan, S. Rodrigues, A. Sanchez, M. Whiting, A. Madan, A.C. Young, Y. Shevchenko, G.G. Bouffard, R.W. Blakesley, J.W. Touchman, E.D. Green, M.C. Dickson, A.C. Rodriguez, J. Grimwood, J. Schmutz, R.M. Myers, Y.S. Butterfield, M.I. Krzywinski, U. Skalska, D.E. Smailus, A. Schnerch, J.E. Schein, S.J. Jones, and M.A. Marra. 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* 99(26):16899–16903.
- Thisse, C., and Thisse, B. 2005. High throughput expression analysis of zf-models consortium clones. *ZFIN Direct Data Submission* (<http://zfin.org>).
- Thisse, C., and B. Thisse. 2008. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nature Protocols* 3(1): 59–69.
- Waskiewicz, A.J., H.A. Rikhof, R.E. Hernandez, and C.B. Moens. 2001. Zebrafish meis functions to stabilize pbx proteins and regulate hindbrain patterning. *Development* 128(21):4139–4151.
- Westerfield, M. (2000). *The zebrafish book. A guide for the laboratory use of zebrafish (danio rerio)*, 4<sup>th</sup> Edition. Eugene, University of Oregon Press.
- Wickramasinghe, S.N. 1998. Congenital dyserythropoietic anaemias: Clinical features, haematological morphology and new biochemical data. *Blood Reviews* 12(3):178–200.
- Wickramasinghe, S.N. 2000. Congenital dyserythropoietic anemias. *Current Opinion in Hematology* 7(2):71–78.
- Wickramasinghe, S.N., R.L. Spearing, and G.R. Hill. 1998. Congenital dyserythropoiesis with intererythroblastic chromatin bridges and ultrastructurally-normal erythroblast heterochromatin: A new disorder. *British Journal of Haematology* 103(3):831–834.
- Wickramasinghe, S.N., and W.G. Wood. 2005. Advances in the understanding of the congenital dyserythropoietic anaemias. *British Journal of Haematology* 131(4):431–446.
- Zerucha, T., T. Stuhmer, G. Hatch, B.K. Park, Q. Long, G. Yu, A. Gambarotta, J.R. Schultz, J.L. Rubenstein, and M. Ekker. 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. *The Journal of Neuroscience* 20(2):709–721.
- Zhang, X., A. Friedman, S. Heaney, P. Purcell, and R.L. Maas. 2002. Meis homeoproteins directly regulate pax6 during vertebrate lens morphogenesis. *Genes & Development* 16(16):2097–2107.