ABSTRACT

LEF/TCF Family Transcription Factors Are Important in the Upregulation and Restricted Expression of Zic Genes in the Dorsal Neural Tube

A central question in biology is how different cell-types arise from an initially one-celled embryo to form an organism. Many cellular differences arise through differential gene expression in space and time via the action of transcription factors (TFs). Zic genes are TFs that contribute to the patterning of the developing vertebrate nervous system, but little is known about how their transcription is regulated. Consensus scoring across several vertebrate species identified possible loci for TF binding sites within a previously identified stretch of zic regulatory DNA, cis-regulatory module D5. The three most promising sites were mutated and injected in constructs upstream of eGFP reporters into zebrafish embryos at the one cell stage. Expression was observed in F1 transgenics by *in vivo* fluorescence at ≈24 hrs. The mutation of the LEF TF binding site suggests that Zic expression is upregulated and restricted to the dorsal neural tube by action of LEF TFs.

Kyle C. Webert – Biology/Zoology
Author Name/Major

Author Signature

5/14/09
Date

Yevgenya Grinblat - Dept. of Zoology/Anatomy
Mentor Name/Department

Mentor Signature

COVER SHEET

TITLE: LEF/TCF Family Transcription Factors Are Important in the Upregulation and Restricted Expression of Zic Genes in the Dorsal Neural Tube

AUTHOR'S NAME: ____Kyle C. Webert_____

MAJOR: ____Biology/Zoology_____

DEPARTMENT: _____Zoology_____

MENTOR: ____ Yevgenya Grinblat, Ph.D_____

DEPARTMENT:_____Zoology_____

MENTOR(2): ____N/A_____

DEPARTMENT(2):____N/A_____

YEAR: ____2009_____

(The following statement must be included if you want your paper included in the library's electronic repository.)

The author hereby grants to University of Wisconsin-Madison the permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Introduction and Background:

One of the chief aims in developmental biology is to understand how an animal's body plan is elaborated from what begins as a single cell. From this original cell arise hundreds of cell types that are properly organized into functional body parts. This process of growth and differentiation often requires cells that are initially identical to produce patterns of different cell types.

Generally speaking, cells initiate patterning by the spread of signaling molecules that are transduced into various levels of transcription factor (TF) activity. Active TFs enter the cell's nucleus where they bind to specific DNA sequences thereby regulating the expression of genes. Many of the genes regulated by TFs are themselves TFs that in turn regulate the expression of more genes. Since TFs are the product of gene expression, some TFs are able to regulate their own gene or the gene of the TF that regulated the transcription of their gene (Gilbert, 1997). These interactions form a complicated regulatory network which transforms the graded distribution of extracellular signaling molecules into clearly defined patterns of intracellular TF gene expression, eventually leading to new cell types (Stathopoulos and Levine, 2005).

To understand this transcriptional network, we need to understand both the parts that interact and what the consequences of those interactions are. We know that these properties are necessarily encoded in DNA as TF binding sites. In most cases, the presence of a TF binding site in close proximity to a gene implies that the gene is regulated by the specific TF that binds at that site. Also, the physical configuration of TF

binding sites around a gene can affect how TFs bind and therefore drive the regulation of transcription in a particular way (Istrail and Davidson, 2005). This type of interaction along with those mentioned above result in the net regulatory effect of the network being more than simply a sum of each separate effect (Istrail and Davidson, 2005).

These properties can be linked to how expression patterns form. If binding sites for several different TFs are present near a gene, it allows highly specific regulation. The cooperative regulatory effects of a particular binding sites configuration can explain how initial TF activity gradients become bound expression areas. The code of TF binding sites in close proximity to different TF genes specifies how patterns of new cell types emerge during development (Istrail and Davidson, 2005). Therefore, understanding the TF binding site code is key to gaining a mechanistic understanding of how body-plans are elaborated through development and into how they have evolved over time (Carroll et al, 2005).

Like all organs, the vertebrate nervous system develops from a mass of identical cells that are gradually patterned by transcriptional regulatory networks. Zic genes are a component of these networks, but little is know about how their transcription is regulated (Aruga, 2004). Using the zebrafish neural tube as a model of developing vertebrate nervous systems, a ~1 kilo base pairs (kbp) block of DNA,

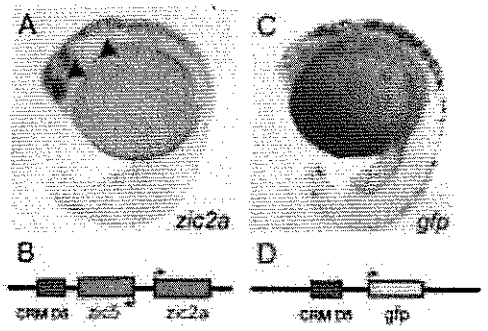

Figure 1: *In situ* hybridizations from Nyholm et al., 2007. (**A**) Endogenous expression of zic2a in a 24 hpf zebrafish embryo. Genomic position of ~1 Kbp CRM D5 is shown in (**B**). (**C**) Transient expression that results from injection of the construct shown in (**D**). Similar expression in A and C suggest that the CRM D5 contains regulatory activity.

located within a few kbps of all zic genes (A. Taylor and Y. Grinblat, unpublished data; Nyholm et al., 2007; Fig. 1) was demonstrated to cause a reporter, enhanced green fluorescent protein (eGFP), to be expressed in a pattern that is very similar to the zic genes endogenous expression pattern. These data strongly suggest that this block contains binding sites for the TFs that regulate zic gene expression. To further guide the search for these sites, we have also found that several sequence stretches within this block are conserved from sharks to humans and contain molecular 'marks' that suggest an involvement in transcriptional regulation (DNaseI hypersensitivity and H3K4 trimethylation; ENSEMBL Human Genome Browser).

Methods:

Sequences of the cis-regulatory module region downstream of zic5, known as D5, from dog (*Canis familiaris*), cow (*Bos taurus*), human (*Homo sapiens*), opossum (*Monodelphis domestica*), mouse (*Mus musculus*), fugu (*Takifugu rubripes*), tetradon (*Tetradon nigrovirdis*), medaka (*Oryzias latipes*), and zebrafish (*Danio rerio*) were downloaded from ENSEMBL genome database and aligned with CLUSTALW. Regions of high conservation across all species were used to create a consensus sequence (Figure 2B). This consensus sequence was then scored with positional weight matrices (PSWM) in Transfac and MatInspector databases. These algorithms detect specific TF binding sites. Potential TF binding sites had PSWM scores >0.9 (highest possible = 1). Also, sites occurring two or more times in a negative control sequence constructed with properties of non-coding zerbafish DNA of equal length were excluded due to likelihood of random occurrence.

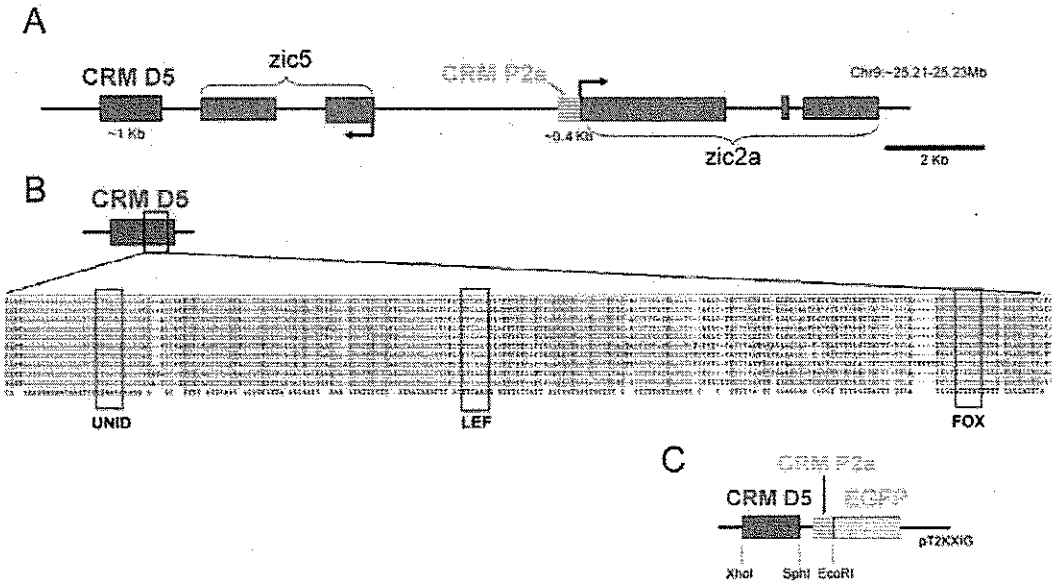

Figure 2: (A) Configuration of cis-regulatory module D5 (CRM D5) and zic5 along with zic 2a and its promoter, CRM P2a in zebrafish genome; (B) Close-up of CRM D5 vertebrate consensus sequence with regions selected for mutagenesis outlined in red; (C) Mutations were incorporated into this construct and injected into zebrafish embryos at the one-cell stage. Illustrations used with permission of A. Taylor.

Sites with high PSWM scores and low likelihood of random occurrence in non-coding DNA were located on the vertebrate alignment. Three potential TF binding sites of interest from the bioinformatics analysis that were located within regions of high conservation were chosen for further investigation. These sites included Forkhead box family (FOX), Tcf/Lef family (LEF), and an unidentified section of highly conserved sequence (UNID) (Figure 2A). High consensus among species was used as a surrogate for relative importance of the code under the logic that regulatory sequences would be conserved through time as they provide necessary functions in development.

We hypothesize that mutation of these sites will prevent bonding of the TF whose site is mutated and this should change the pattern of expression of the reporter eGFP. In

order to test this hypothesis, the sites were mutated. Recall that an intact D5 upstream of eGFP produces an expression pattern similar to endogenous expression of zics (Figure 1). These sections were mutated by the method from Zaret, Liu, and DiPersio, 1990 (Figure 3).

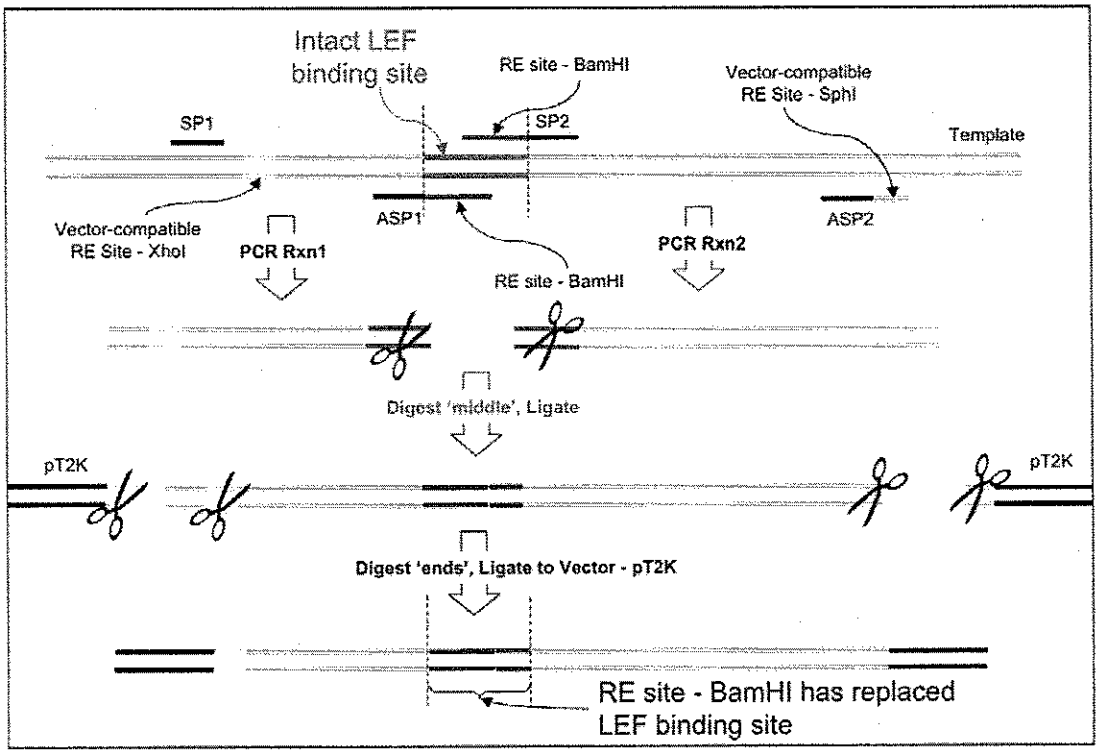

Figure 3: Mutagenesis method for Tcf/Lef binding site adapted from Zaret, Liu, and DiPersio, 1990.

The mutated sequences of FOX, LEF, and UNID; called mutFox, mutLef, and mutUNID respectively, were incorporated into a vector upstream of eGFP reporter and ampicillin resistance, creating three plasmid constructs with one mutation each. LEF was replaced with restriction enzyme (RE) site BamHI (Figure 3), FOX was replaced with RE site Xma I, and the UNID site was replaced with RE site BglII. These constructs were then cloned into electro-competent bacterial cells (Invitrogen OneShot® Top10 Electrocomp Cells) via electroporation. The bacteria cells were then incubated briefly in

Super Optimal broth with Catabolite repression (SOC) broth for higher transformation efficiency before plating on medium selecting for ampicllin resistance and allowed to incubate for approximately sixteen hours at 37°C. Single colonies selected from these plates were then grown overnight and plasmid DNA was minipreped with a QIAGEN kit. Sequencing of the cloned sequence was compared to expected sequence to confirm the product was what we expected (Figure 4).

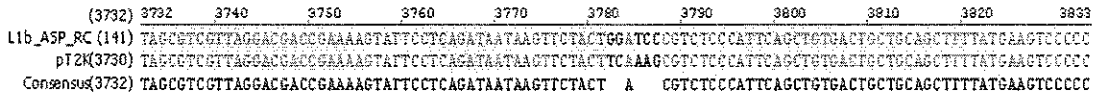

Figure 4: Sequencing results. The top sequence, L1b_ASP_RC, is the product of the anti-sense primer used to mutate the LEF binding site. The middle line, pT2K, is the sequence of the intact CRM D5 in the pT2K vector. The mutation in the top sequence is the six bp BamHI RE site beginning at bp 3783.

The constructs were then injected into wild-type zebrafish embryos at the one cell stage. All embryos were injected with approximately 100 pico-grams of DNA and 100 pico-grams of Tol2 RNA. Care was taken to ensure differences in expression were not resulting from injection of different amounts of DNA. Embryos showing mosaic expression of eGFP (Figure 5) were reared until adulthood. These adults were then mated to produce a F1 generation of transgene fish with stable expression of the injected constructs.

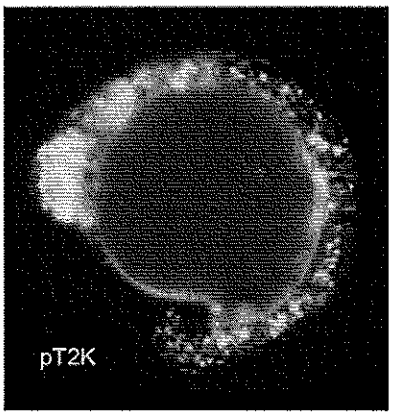

Figure 5: Embryo showing universal mosaic expression of eGFP by fluorescence (≈24 hrs).

Results and Discussion:

Comparison of the plasmid DNA from bacterial clones to the expected cloning outcome sequence shows that the method was successful in introducing our planned mutations into the zic cis-regulatory module D5 (Figure 4). This mutagenesis method is powerful in that it creates precisely the mutation desired while designing the initial primers for PCR. This method does have the limitation of requiring a palindromic restriction enzyme site within the new sequence. Our mutation simply replaced six base pairs of the TF binding DNA sequence with a restriction enzyme site, but mutations including more flanking sequence could be constructed.

Tcf/Lef:

In earlier findings (Nyholm et al., 2007) Tcf/Lef factors, as effectors of canonical Wnt signaling, were shown to activate zic2a and zic5 transcription during zebrafish development. Tcf/Lefs were also shown to be required for zic gene expression (Nyholm et al., 2007) by elimination of zic2a and zic5 expression through ΔTcf, a dominant inhibitor of Tcf/Lef factors (Lewis et al., 2004). These papers both recognize the importance of Tcf/Lefs in upregulation of zic genes. We hypothesized that embryos with mutated Tcf/Lef binding domains would show decreased or no expression of zics.

Embryos with mutLEF TF binding sites show universal low-level expression of eGFP (Figure 6D). When contrasted to expression driven by cis-regulatory modules with intact LEF TF binding sites, these data suggest that LEF TFs are involved in not only the upregulation of zic gene expression, but also the restriction of this expression to the dorsal neural tube.

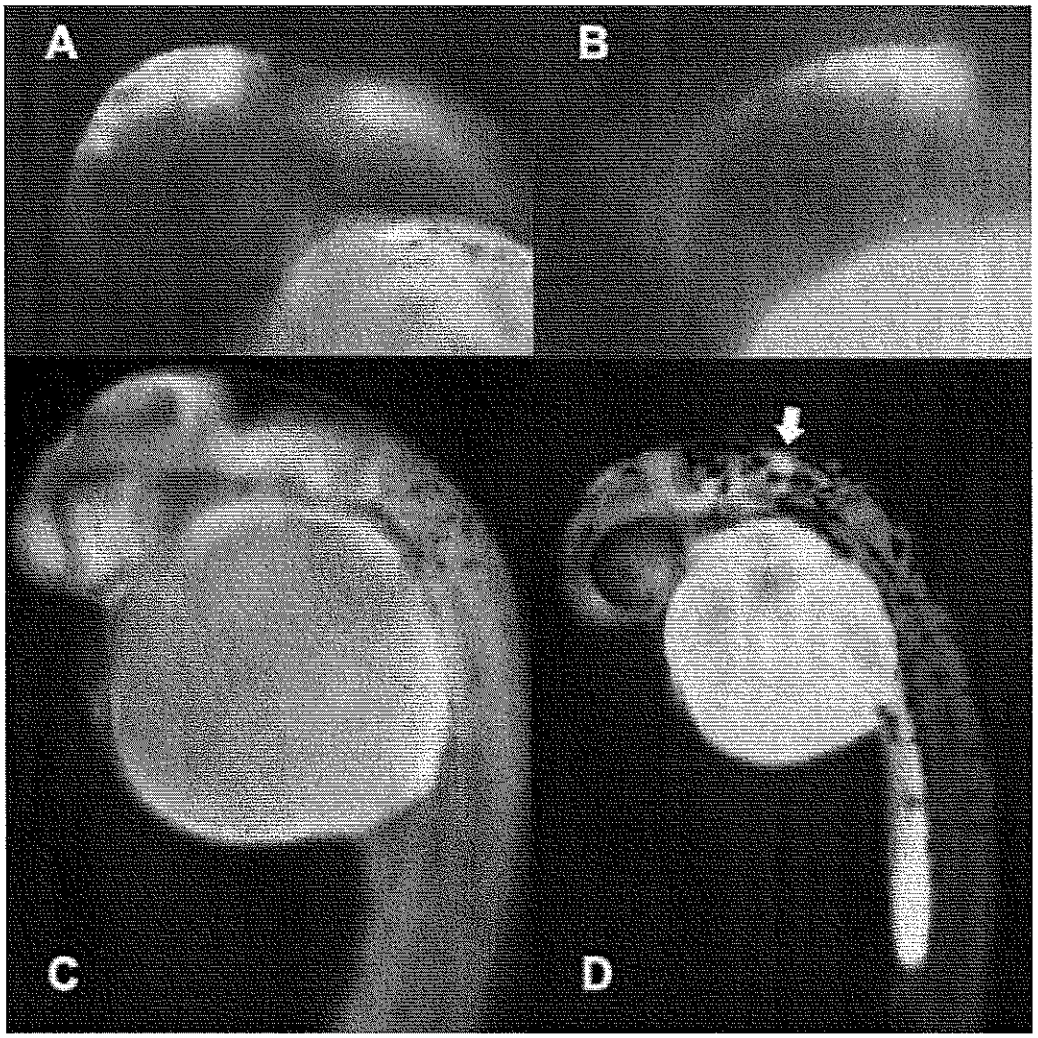

Figure 6: (A) eGFP expression in embryo with intact D5 CRM. Expression approximates endogenous zic expression (Figure 1) with two regions of expression in the mid- and hindbrain and along the dorsal neural tube (not pictured above). (B) mutFOX embryos express eGFP approximately in the same manner as controls in (A). (C) mutUNID embryos express a high level of eGFP universally with expression in the telencephalon and eye lens. (D) mutLEF embryos express a low level of eGFP universally. The arrow indicates an area of strong expression in ≈rhombomere five.

FOX:

Forkhead box factors are known to be expressed in neural crest cells and involved in the formation of the neural tube. *Fox,* especially FOXD3, and *zic* genes are both very early markers in neural crest cell induction (Aybar and Mayor, 2002; Odenthal and

Nusslein-Volhard, 1998). Zic genes are expressed on the boundary of the neural plate and ectoderm and play a role in the development and patterning of neural crest cells where FoxD3 genes are expressed (Grinblat and Sive 2001, Kos *et al.* 2001, Dottori *et al.* 2001, Meulemans and Bronner-Fraser 2004). The zebrafish gene *foxb1.1* is also widely expressed in the nervous system with the exception of the telencephalon (Grinblat *et al.* 1998, Mazet and Shimeld 2002). We hypothesized that embryos with mutated FOX TF binding sites would express eGFP differently than embryos with intact FOX binding sites.

The mutFOX embryos appear to express eGFP in approximately the same manner as embryos with the intact D5 cis-regulatory module (Figure 6B). These data suggests that forkhead factors do not directly play a role in the regulation of zic gene expression in CRM D5.

UNID:

We hypothesized that due to the highly conserved nature of the UNID section and its proximity to zic genes, the section may be an important locus for regulatory activity. We expected that embryos with altered UNID sites may show different expression of zic genes that embryos with intact UNID sites.

Embryos with mutated UNID sites express zics universally, including expression in the telencephalon and eye lens (Figure 6C). Because of this expression pattern, we hypothesize that the UNID segment may act as a general repressor of zic gene expression.

Implications for other cis-regulatory modules:

Two other cis-regulatory modules (CRMs) of zic genes have been identified. These CRMs, D4 and D6, are downstream of zic4 and zic6 respectively. Like D5, these CRMs contain several possible TF binding sites and regions of conservation across many vertebrate species. Alignments confirm that similar TF arrangements exist in all three CRMs. Our findings here may be extrapolated to the other CRMs. We hypothesize that Tcf/Lef TFs in the additional CRMs will be an important component in the regulation of the zic genes upstream. The CRM D4 also has a highly conserved region resembling the UNID site, which would be interesting to test for repressor function.

Acknowledgments:

I would sincerely like to thank Aaron Taylor and Yevgenya Grinblat whose expertise and guidance has been an invaluable resource in the preparation and realization of this study. I would also like to thank the other Grinblat lab members; Molly Wagner, Jessica Pierson, and Nick Sanek for their assistance and kindness in the lab.

References:

Aruga, J. (2004). The role of Zic genes in neural development. *Mol. Cell. Neurosci.* **26,** 205-221.

Aybar, M. J., Mayor, R. (2002). Early induction of neural crest cells: lessons learned from frog, fish and chick. *Cur. Op. in Gen. & Dev.* **12,** 452-458.

Carroll, S., Grenier, J., and Weatherbee, S. (2005). From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design, 2nd Edition: Blackwell Publishing.

Dottori, M., Gross, M. K., Labosky, P., Goulding, M. (2001). The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. *Development* **125,** 3111-3121.

Gilbert, S. (1997). *Developmental Biology*, 6th ed. Sunderland, MA: Sinauer Asc., Inc.

Grinblat, Y., Gamse, J., Patel, M., and Sive, H. (1998). Determination of the zebrafish forebrain: Induction and patterning. *Development* **125,** 4403–4416.

Grinblat, Y., Sive, H. (2001). *zic* gene expression marks anteroposterior pattern in the presumptive neurectoderm of the zebrafish gastrula. *Dev. Dyn.* **222,** 688-693.

Istrail, S., Davidson, E. H. (2005). Logic functions of the genomic cis-regulatory code. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 4954-4959.

Kos, R., Reedy, M. V., Johnson, R. L., Erickson, C. (2001). The winged-helix transcription factor FoxD3 is important for establishing the neural crest lineage and repressing melanogenesis in avian embryos. *Development* **128,** 1467-1479.

Lewis, J. L., Bonner, J., Modrell, M., Ragland, J. W., Moon, R. T., Dorsky, R. I., and Raible, D. W. (2004). Reiterated Wnt signaling during zebrafish neural crest development. *Development* **131,** 1299-1308.

Mazet, F., Shimeld, S. M. (2002). The Evolution of Chordate Neural Segmentation. *Dev. Bio.* **251,** 258-270.

Meulemans, D., Bronner-Fraser, M. (2004). Gene-regulatory Interactions in Neural Crest Evolution and Development. *Dev. Cell.* **7,** 291-299.

Nyholm, M. K., Wu, S., Dorsky, R. I., and Grinblat, Y. (2007). The zebrafish *zic2a-zic5* gene pair acts downstream of canonical Wnt signaling to control cell proliferation in the developing tectum. *Development* **134,** 735-746.

Odenthal, J., Nusslein-Volhard, C. (1998). Fork head genes in zebrafish. *Dev. Genes Evol.* **208,** 245-258.

Stathopoulos, A., Levine, M. (2005). Genomic regulatory networks and animal development. *Dev. Cell.* **9,** 449-462.

Zaret, K. S., Liu, J., and DiPersio, C. M. (1990). Site-directed mutagenesis reveals a liver transcription factor essential for the albumin transcriptional enhancer. *Proc. Natl. Acad. Sci. USA* **87,** 5469-5473.