MQP-BIO-DSA-5506
MQP-BIO-DSA-9137

THE LOCATION AND TEMPORAL EXPRESSION OF *FOXA3* IN THE DEVELOPING ZEBRAFISH PANCREAS

A Major Qualifying Project Report

Submitted to the Faculty of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

in

Biology and Biotechnology

by

Amy Castonguay

Holly Thompson

April 27, 2006

APPROVED:

Charles Sagerstrom, Ph.D.
Biochemistry and Mol. Pharmacology
Umass Medical Center
Major Advisor

David Adams, Ph.D.
Biology and Biotechnology
WPI Project Advisor

// ABSTRACT

Numerous genes are expressed and regulated during the development of the pancreas in zebrafish embryos, and the mechanism is currently being researched. Although the cellular locations and times of *pdx-1* and *insulin* expression are known in the developing pancreas, less is known about other genes such as *foxa3*. This project analyzed *foxa3* temporal and spatial expression using a GFP reporter under the control of the *foxa3* promoter.

TABLE OF CONTENTS

Signature Page ………………………………………………………………………... 1

Abstract ……………………………………………………………………………….. 2

Table of Contents ……………………………………………………….……….. 3

Acknowledgements ………………………………………………………………… 4

Background …………………………………………………………………………… 5

Project Purpose ………………………………………………………………… 14

Methods ……………………………………………………………………………… 15

Results ……………………………………………………………………………….. 21

Discussion …………………………………………………………………………… 34

References …………………………………………………………………………… 37

Appendix A ……………………………………………………………………...... 39

Appendix B ………………………………………………………………………….. 41

ACKNOWLEDGEMENTS

First and foremost we would like to thank Charles Sagerstrom, PhD at the University of Massachusetts for allowing us to perform our research in this lab and use the outstanding facilities provided to us. We are extremely grateful to Professor Sagerstrom for providing us with this excellent learning opportunity. Next we would like to thank Kristen Alexa for the large amounts of time she spent developing our project, teaching us every research technique we learned, and assisting us with questions and every day tasks around the lab. We would like also like to thank Letitiah Etheridge and Nicolas Hirsh for their help with techniques in the lab. Finally, we would like to thank David Adams PhD at Worcester Polytechnic Institute for assisting with the initiation of our project, his helpful feedback and knowledge during the project, and the report writing.

1. BACKGROUND

1.1. Zebrafish as Models

Danio rerio, more widely known as the zebrafish, is one of several vertebrate models used for mutagenesis research. Zebrafish are an attractive biological model, at times even more so than previously used models, such as *Drosophila* (fruit flies) or *C. elegans* (nematode worms), because they are more closely related to humans, yet they retain the ease of experimental manipulation that made the fly and worm models useful. Zebrafish are also more attractive models than rodents due to their rapid development time. In some cases screening thousands of mutations was performed during embryo development (Driever et al., 1996). In addition, during development, mutations in pancreatic organogenesis are quite common, allowing for accessible studies on this organ (Pack et al., 1996).

1.2. Zebrafish Embryo Development

During Zebrafish embryonic gastrulation, three germ layers are established. First is the ectoderm on the outer sides of the embryo (shown as blue in Figure 1). The ectoderm gives rise to the epidermis, the brain, and the nervous system. Inside the ectoderm is the mesoderm (shown as orange in Figure 1) with the notochord in the middle running from the head to the tail of the developing fish. The mesoderm develops into the backbone of the adult fish. The mesoderm gives rise to the bones, cartilage, dermis, the circulatory system, and muscle tissue. Just under the mesoderm nearest to the yolk is the endoderm (shown as red in Figure 1), giving rise to the gut. The embryonic

gut consists of the lining of the digestive tract, respiratory tract, and the glands. This is divided into three gut segments. The first section, closest to the head of the embryo, is the anterior gut where the esophagus develops. The midgut, middle section, gives rise to three organs: the liver, the stomach, and the pancreas. At the tail end is the hindgut, where the intestines form.

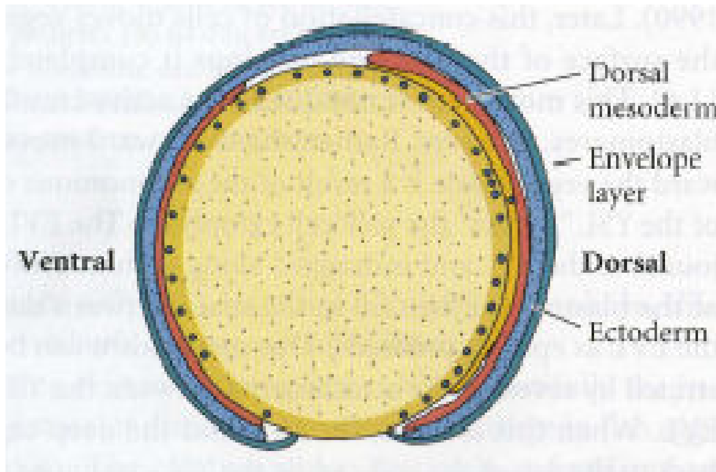
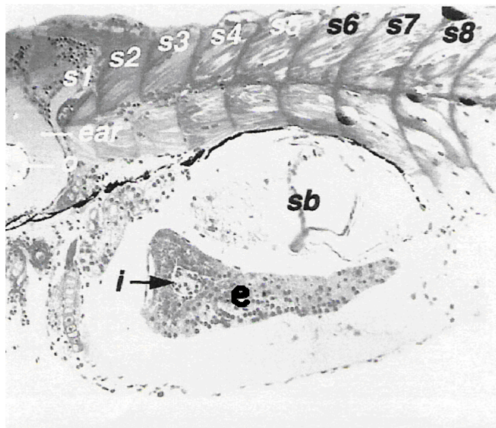


Figure 1. The Primary Sections of a Developing Zebrafish Embryo. Figure taken from Gilbert, 2003.

1.3. Pancreatic Development

The pancreas is an organ which controls glucose homeostasis and produces digestive enzymes in vertebrates. This organ usually resides at the right side of the body (see Fig. 2, right panel) (Biemar et al., 2001). The pancreas is located near somite 4. At 40 hours post-fertilization (hpf), the pancreas first begins as two buds, a ventral anterior bud, and a dorsal posterior bud (Figure 3). By 52 hpf (Fig. 4), these two buds have fused together to later develop into the adult pancreas (Field et al., 2003).



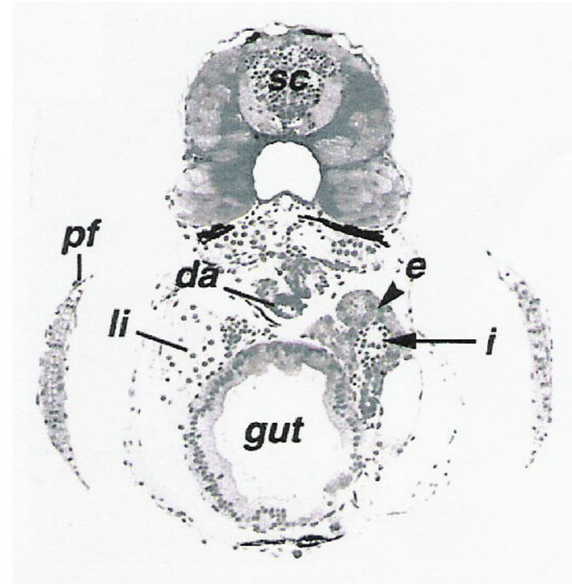


Figure 2. The Zebrafish Pancreas. The left panel shows a sagittal view of a 6-day old larvae. Endocrine pancreatic tissue is denoted with an (i) and the exocrine pancreatic tissue (e). The right panel shows a cross section of the same larvae, with the pancreatic tissue developing on the right side of the larvae. Figure taken from Biemar et al., 2001.

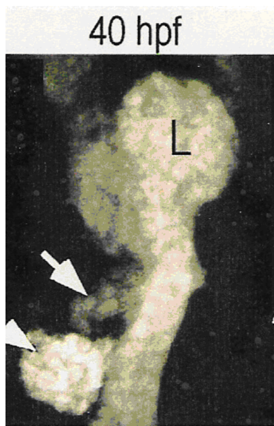


Figure 3. The Zebrafish Pancreas 40 Hrs Post-Fertilization (hpf). The pancreas first develops as two buds: a ventral anterior bud (arrow in the center of the diagram) and a dorsal posterior bud (arrowhead, lower left). Figure taken from Field et al., 2003.

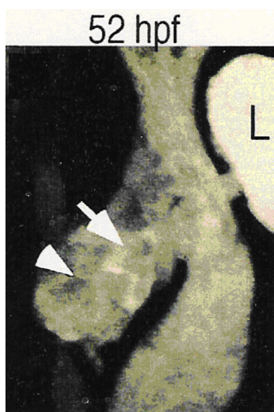


Figure 4. The Zebrafish Pancreas 53 Hrs Post-Fertilization. By 52 hpf, the two buds have fused to form one bud to develop into the adult pancreas. Again, the arrow represents the old anterior bud, and the arrowhead, the posterior bud. Figure taken from Field et al., 2003.

The adult pancreas is comprised of four types of cells: alpha, beta, gamma, and Polypeptide (PP). These four cell types express the following hormones in zebrafish: *glucagon, insulin, somatostatin*, and *pancreatic-polypeptide*, respectively. Of these main hormones, *insulin* has been determined to be the first hormone expression in the primordium, at the 15 somite (15s) development stage of the pancreas (Huang et al., 2001).

1.4. Genes Regulating Pancreatic Development

The first gene expressed in the endocrine and exocrine pancreas, as well as the duodenum, is *pdx-1*. It encodes a transcription factor protein that is necessary for the expression of insulin (Milewski et al., 1998). To help define the pancreatic primordium, *pdx-1* and insulin expression was analyzed and it was determined that the pancreas develops from a left and right primordium on the dorsal side (Fig. 5) (Argenton et al., 1999).

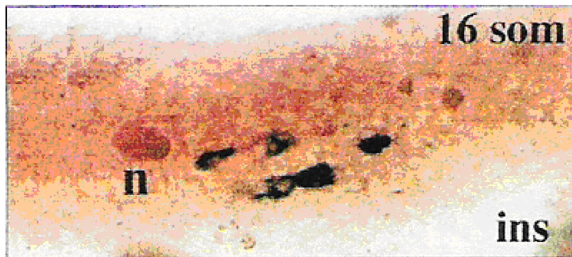


Figure 5. Insulin Immunostain of the Developing Pancreas. At the 16 somite stage, this *pdx-1* domain contains insulin producing cells, the endocrine primordium. Figure taken from Argenton et al., 1999.

Pdx-1 encoded transcription factor is essential in pancreatic development. A knock out of *pdx-1* eliminates development of the pancreas (Jonsson et al., 1994). Expression of *pdx-1* is detected as early as the 10 somite (10s, also 14 hpf) stage of embryo development (Fig. 6) in two bilateral rows of cells next to the midline (Biemar et

al., 2001), and continues into the 16s stage below the notochord (Fig. 7) (Argenton et al., 1999). After 24 hours, the two rows of *pdx-1*-expressing cells merge to form a single region at the midline (Fig. 8).

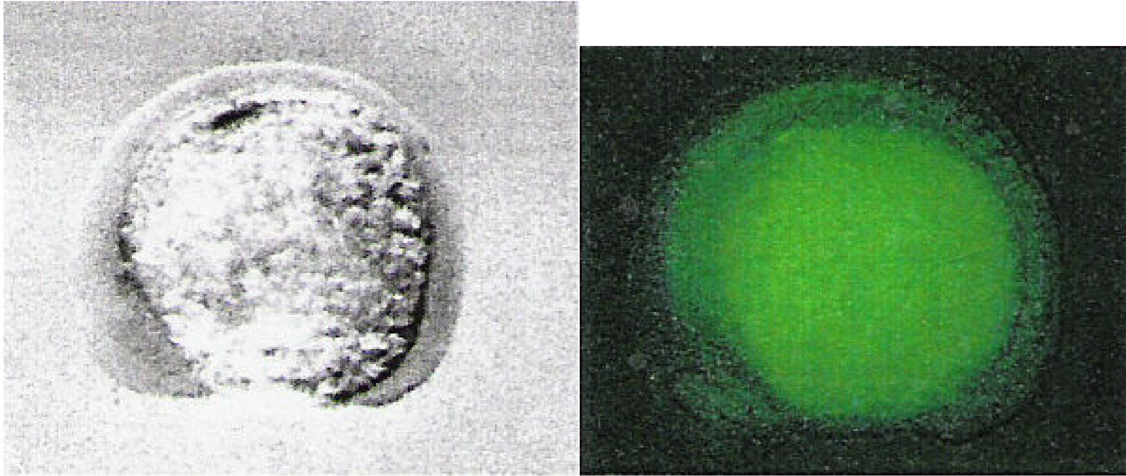


Figure 6. Cells Expressing Pdx-1 at the 10 Somite Stage. On the left panel, pdx-1 expression is seen as a dark shaded region at the top of the embryo. On the right panel, expression can be seen at the same location as the figure on the left, in brighter green fluorescence. Figure taken from Biemar et al., 2001.

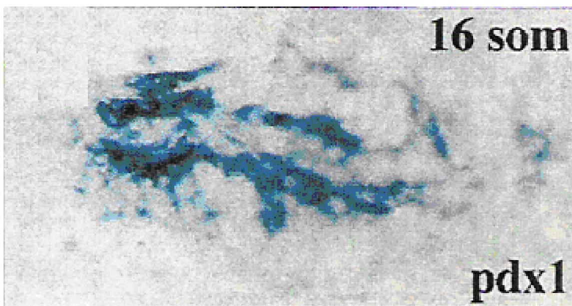


Figure 7. Cells Expressing Pdx-1 at the 16 Somite Stage. At the 16 somite stage, cells containing *pdx-1* form two rows of cells, seen in the dark blue stain. Figure taken from Argenton et al., 1999.

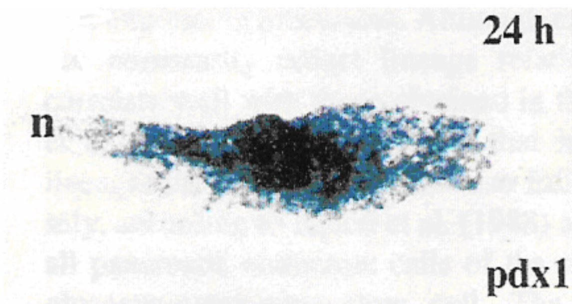


Figure 8. Cells Expressing Pdx-1 at the 24 Hour Stage. By 24 hours, the *pdx-1* expressing cells have merged from two rows into one mass. of Expression is located near the dorsal side of the embryo. Figure taken from Argenton et al., 1999.

Some genes actively expressed in the mesoderm are also important in the formation and development of the endoderm and pancreas. Signaling factors in the mesoderm such as RA, Sonic Hedgehog (*Shh)*, and TGF-β are important to begin the activation of *pdx-1*. *Shh* expression in the mesoderm is necessary to initiate the pancreatic cell differentiation during the early stages of development (Fig. 9) and in gastrointestinal development (Roy et al, 2001; Hebrok, 2003). Expression of *Shh* is also important in the endoderm to restrict *pdx-1* activation and prevent the pancreas from continual growth (Ramalho-Santos et al, 2000; Hebrok et al., 1998).

Figure 9. Immuno-staining of Shh and Pdx-1 24 Hrs Post-Fertilization. After 24 hfp, *Shh* (blue stain) and *pdx-1* (red stain) show no overlapping expression. Figure taken from Hebrok, 2003.

Pax4 and *pax6* are other transcription factors that aid the development of the pancreas (Huang, 2000). Other genes that appear to be involved in pancreatic development include: *islet-1, foxa2, foxa3, nkx2.2, nrd, pax 6.1, pax6.2*, and *prox-1*.

1.5. Analysis of Pancreatic Gene/Promoter Expression by GFP

To analyze the temporal and spatial expression of a specific gene, a promoter construct can be created fused to a reporter gene (for example, green fluorescence protein, GFP) ligated downstream from the promoter of the gene of interest (for example, *foxa3*). This construct (Fig. 10) is then microinjected into the embryo, where it is

integrated into the genome and transcribed and translated by the organism's machinery, and GFP fluorescence is monitored throughout development. By using the promoter region of a gene that is native to the animal, the GFP construct will be activated in the same areas and at the same time as the native gene. The function of the gene in relation to the development can be deduced, as well as determining the role and position a gene may have in a developmental pathway.

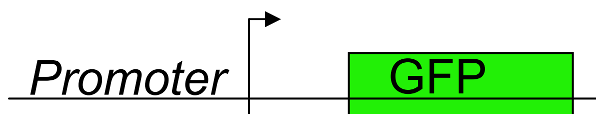


Figure 10. An Example of a Gene Promoter Construct Used to Activate GFP Expression Downstream. Made by Kristen Alexa, UMass Medical School Department of Biochemistry and Molecular Pharmacology.

Promoters can be cloned at different lengths, but some of the shorter promoters may not contain all the essential elements or enhancers for activating or repressing the expression of the gene, in the correct location, it regulates. Promoters analyzed in zebrafish endoderm development (not necessarily pancreatic) using the GFP approach include: *casanova (cas), her5, foxa1, and foxa2*. *Her5* is expressed in the midhindbrain boundary and endoderm (Fig. 11) (Tiso et al., 2002). *Cas* (also called *sox32*) is important for endoderm development; if this gene is mutated, the gut is not formed (Shivdasani, 2002). Expression of GFP can also be used to observe endoderm formation in embryonic cells. A *cas* mRNA injection in conjunction with GFP RNA drives endoderm formation in the injected cell. Fluorescence from the GFP RNA allows the endodermal development in the daughter cells of the injected cell to be seen as the embryo develops. This endoderm formation can be seen in Fig. 12 down the entire length of the developing fish below the notochord.

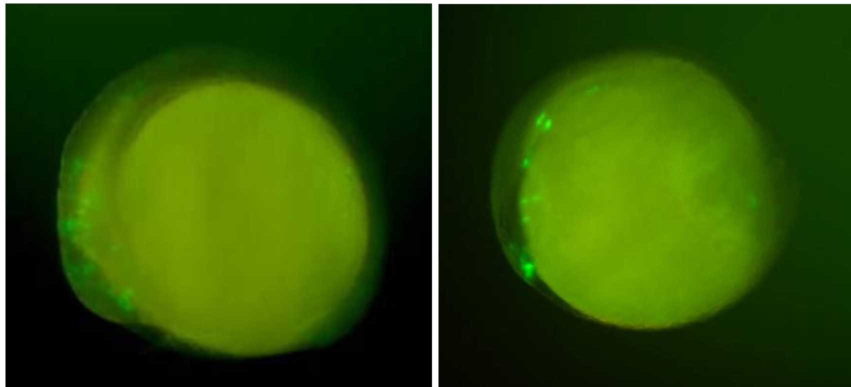


Figure 11. GFP Analysis of the *Her5* 2Kb Promoter in the Developing Embryo. The left panel denotes 10 hpf, and the right panel 12 hpf. Research and pictures taken by Kristen Alexa, UMass Medical School Department of Biochemistry and Molecular Pharmacology.

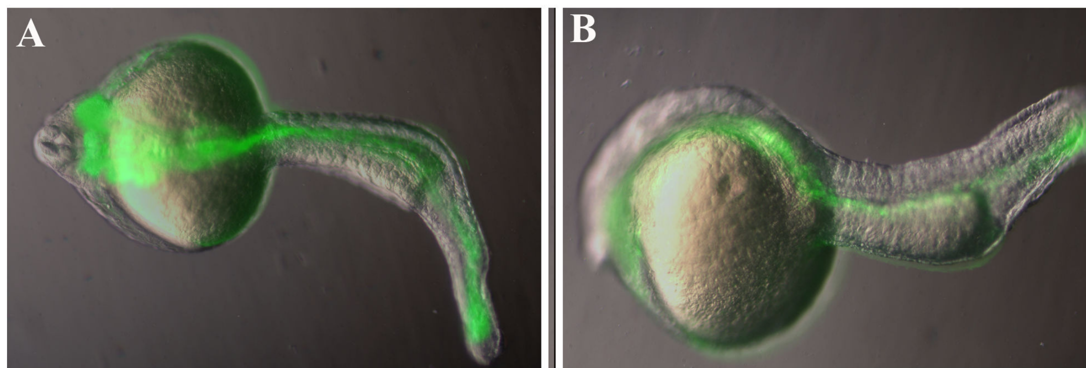


Figure 12. GFP Analysis of GFP Expression via Cas RNA Injections During Zebrafish Development 24 Hrs Post-Fertilization. Cas mRNA and GFP RNA were injected into one cell during the 32 cell stage to promote endoderm development in that cell. The left panel denotes the top of the embryo, while the right side denotes the side view. Research and pictures taken by Philip Dilorio, UMass Medical School.

Foxa genes are from the family Hepatocyte Nuclear Factor-3 (HNF3). Genes from the HNF3 family have primary roles of differentiating developing tissue in zebrafish embryos into functional adult organs, such as the pancreas and the liver (Kaestner, 2000). This family consists of three major genes: *foxa1, foxa2*, and *foxa3* in zebrafish. *Foxa1* and *2* have similar expression patterns; both are expressed in the notochord and floor plate, and are necessary to maintain the expression of *Shh*. Current research suggests that *foxa1* and *2* genes are upstream in the gene pathway of *Shh*, which regulates the expression of pancreatic genes at somite 4 in the endoderm.

Little is currently known about the function of *foxa3* (Odenthal, 1998). *Foxa3* is expressed in the endoderm (Matsui et al., 2005). The expression of *foxa3* in the developing zebrafish can be seen in Figure 13 (Matsui et al., 2005). From a dorsal view, the expression of the *foxa3* gene (indicated by the arrow) is localized to the endoderm of the embryo. Seiliez et al. recently discovered that *foxa3*, along with another transcription factor known as *Goosecoid*, are responsible for controlling the formation of the anterior region in a developing zebrafish embryo. Having similar expression patterns during the early stages of development, these two genes control anterior region formation by the repression of Wnt8a activity prior to development of the gut (Seiliez et al, 2006). Seiliez et al's studies showed that a lack of *foxa3* expression during development had little effect on the development of the anterior region, but lack of both *foxa3* and *Goosecoid r*esulted in truncation of the head. Therefore, it is concluded that *foxa3* contributes to the normal development of the early stages of the head. Interestingly, as previously mentioned by Matsui et al and Seiliez et al's research, *foxa3* is present in the endoderm and brain, respectively, during development. Knowing this, it may be assumed that *foxa3* could exhibit similar expression locations as *her5* in the endoderm and the midhindbrain, which also shows expression in these two locations. Finally, it can be concluded that *foxa3* may also play a role in pancreatic development.

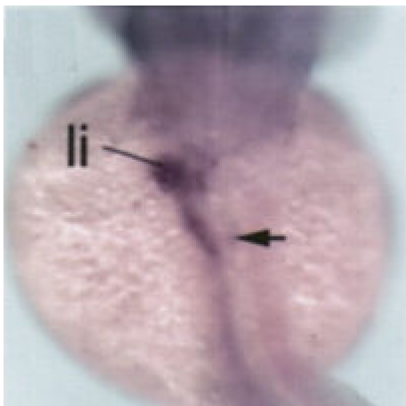


Figure 13. *Foxa3* Expression in the Zebrafish Endoderm. View is anterior to the top, dorsal side up at 48hpf. Figure taken from Matsui et al., 2005.

2. PROJECT PURPOSE

Several genes are known to be expressed in the developing pancreas, including the homeobox *pxd-1*, *insulin, glucagon, somatostatin*, *sonic hedgehog*, *casanova*, and *her5*. Recent data indicated *foxa3* may play a role in midhindbrain development and may be expressed in the endoderm. The Sagerstrom laboratory is currently analyzing the temporal and spatial expression of the *foxa3* protein in zebrafish embryos by engineering plasmid constructs containing the *foxa3* promoter upstream from a GFP reporter, and microinjecting the linearized constructs into embryos to establish transgenic lines. By comparing the GFP expression patterns with various promoters for pancreatic genes, hopefully a pathway can be deduced. The purpose of this MQP was to get a *foxa3* promoter long enough for GFP expression in the correct location in the endoderm. This promoter would then be used in further endoderm experiments.

3. METHODS

3.1. PCR

Various lengths of primers for *foxa3* promoters were prepared by PCR using 1.0 µl of zebrafish genomic DNA as template. Also added to the 0.5 mL PCR tube was 2.0 µl of 10x Hot Master Taq Buffer, 1.0 µl of dNTPs, 0.5 µl of Hot Master Taq, 1.0 µl of *foxa3* 1.5kb forward primer, 1.0 µl of *foxa3* reverse primer, and 13.5 µl of dH2O to bring the total to 20 µl. The PCR mixture was run on the following cycle: 94˚C for 2 minutes, 35 times: 94˚C for 45 seconds, 50˚C for 30 seconds, 68˚C for 5 minutes, and 68˚C for 5 minutes at the end of the entire cycle. The Thermocycler reduces its temperature to 4˚C at a hold at the conclusion of the cycle.

3.2. DNA Electrophoresis

PCR amplicons were analyzed on 1% agarose gels, made from 2.5 g of agarose, 250 ml 1X TAE Buffer, and 12.5 µl of 10 mg/ml EtBr. Gels were typically electrophoresed for 15-30 minutes at 96-106 V, then photographed under UV transillumination.

3.3. Extraction of DNA Bands From Agarose Gels

DNA samples were electrophoresed in 1% agarose gels (as described above) distributed across 6 wide lanes, approximately 5-6 µl in each well. The correct size band was then extracted by excising it out using a razor blade. After the sample was weighed, the Qiagen Gel Extraction Kit was used to purify the DNA. Buffer QG was added at three times the mass (100 mg = 100 µl) of the gel. The sample, in a conical tube, was

incubated at 50˚C and vortexed until the gel dissolved. Isopropenol at one times the gel mass was added. Next, the sample was applied to a Qiagen Spin Column and spun in the table top centrifuge at 12,000 RPM for one minute per 800 µl of sample. Between centrifugations the flow through was discarded. 750 µl of Buffer PE with ethanol was applied to the spin column, spun down, discarded, and spun down again to dry. Finally, the DNA in the spin column was eluted with Elution Buffer (EB), and the eluate containing DNA was collected into a clean Eppendorf Tube.

3.4. Digestion of Insert/Vector

Restriction enzyme digestions were performed using enzymes purchased from New England BioLabs Inc. with the manufacturer's recommended buffers. To prepare the insert for ligation into the GFP vector, 2.8 µl of 10 µg plasmid DNA was digested with *BstXI* restriction enzyme out of the pCRII – Topo vector in which it was cloned into after PCR. For the preparation of GFP vector for insert ligation, 3.1 µl of 10 µg plasmid DNA was digested overnight with 1 µl (10 units) of *BstXI* restriction enzyme at 55˚C, in a 25 µl total reaction. For the screening of GFP plasmids post-insert ligation to determine insert orientation, typically 40 ng of minipreped GFP plasmid was digested overnight with 1 µl (5 units) of *HpaI* enzyme at 37˚C, in a 25 µl total reaction.

3.5. SAP

Shrimp Alkaline Phosphatase (SAP) was used to prevent the GFP vector from self-ligation by removing a phosphate group from the 5' end of the DNA. Typically, 10 µg of GFP plasmid DNA previously linearized with *BstXI* enzyme was digested with 1.5

μl (1 unit) of SAP for 1 hr at 37˚C, followed by incubation at 65˚C for 15 minutes to inactivate the SAP.

3.6. Ligation of Insert/Vector

Promoter DNA fragments and GFP vector (previously cut with *BstXI* restriction enzyme) were ligated in 12 μl reactions containing 1.0 μl vector, 5.0 μl insert, and 6.0 μl of TakaRa's SolI Ligase. The reactions were incubated at 25˚C for 2-24 hrs. Negative control tubes contained dH_2O in place of insert or vector.

3.7. Transformation and Plating of Recombinant Bacterial Cells

Ligated recombinant DNAs were transformed into competent DH5α *E. coli* cells purchased from Invitrogen. DH5α cells were thawed on ice, and 50 μl of cells was mixed with 5 μl of ligation reaction. The samples were incubated on ice for 20 minutes, then heat shocked for 1 minute at 37˚C, and placed on ice for another 5 minutes. Each 55 μl sample was supplemented with 1.0 ml of LB medium (made from 25 LB medium capsules per 1L), in a round bottom tube, and was incubated at 37˚C for 1 hr on a vertical rotator. Cells were then pelleted by centrifugation in an Eppendorf microcentrifuge at 12,000 RMP for 1 min. The supernatant was removed, and the cells were resuspended in 110 μl of LB. Aliquots of 10 μl and 100 μl of cell suspension were plated on LB-amp plates, and the plates were incubated overnight at 37˚C.

3.8. Isolation of Plasmid DNA (Mini/Midi/Maxipreps)

Qiagen Mini/Midi/Maxiprep Kits were used to isolate plasmid DNA from *E. coli* bacterial cultures, using the manufacturer's recommended protocol with no exceptions. See Appendix A for the instructions from the Qiagen Manual for mini/midi/maxipreping. Miniprep kits were typically used with 3-5 ml of bacterial culture, while midi and maxi kits used 300 and 500 ml, respectively. While using the Qiagen kits, yields should be obtained between 0.04-0.2 μg of plasmid DNA/mL of LB. The following yields of plasmid DNA were typically obtained from the 3 kits, respectively: 0.14 μg, 2.8 μg, and 0.07 μg. Note: Only one maxiprep was performed, obtaining an unusually low yield.

3.9. Microinjection of Linearized GFP DNA into Embryos

Before setting up the microscope and needle, the male fish was added to the same tank as the female fish to allow fertilization of the eggs. A porous divider was inserted in the bottom of the tank to keep the males and females separate to prevent crossing over and laying before the lights turn on to stimulate the fish in order to inject before the 1 cell stage, and to prevent the adults from eating the young eggs once laid. Once the female fish laid the fertilized eggs, both adult fish while still kept in the crossing tank were removed and put into new water, and the eggs were collected into a net and placed in a petri dish with embryo media. Viable embryos were transferred from the petri dish to a new plate with an agarose mold containing grooves for the embryos to line up and orient the embryos. Embryos were lined up singly in the grooves. Next, the needle was prepped by adding 1μl of DNA sample containing 0.5 μg of plasmid DNA mixed with a red dye to the blunt end of the needle by gravity force. Once the pressure through the

needle was set between 1-10 psi, and the tip was snipped if needed, the needle was injected into the first embryo through the chorion and into the embryo. The embryos must be injected before the 1-cell stage, leaving only 15-20 minutes between fertilization and injection. On average, 200 embryos were injected for each construct. Fish were crossed staggered to prevent laying embryos, causing developing before injection.

Prior to the morning of injection, needles need to be made, injection plates need to be set, and the fish need to be set up into cross tanks. A machine was used to pull needles. The temperature and rate of the needle pulling can be set. The temperature was set between 6 and 7 while the rate was set between 3 and 4. The longest resulting needles were chosen for the injections. Agarose injection plates were made by pouring agarose media into a petri dish. When the agarose is cool enough, a plate with grooves was placed on top of the agarose and allowed to float on top until the agarose hardened. Also, two males were placed into the bottle of a normal tank. Three females were put into the cross tanks with netting at the bottom. The cross tank was placed into the normal tank. Four sets of 2 males and 3 females were set up the night before the injection.

3.10. Embryo Photography for GFP Fluorescence

Embryos were initially screened in their original petri dish using a fluorescence Leica microscope with a GFP filter. When positive embryos were found, a transfer pipet was used to transfer the embryos from the large petri dish to a new smaller plate to examine further. When photographing, the embryo was transferred again from the small petri dish to a plate with agarose in the bottom. A groove was made in the agarose with

forceps to prevent the embryo from moving. A switch on the microscope transferred vision from the eyepiece to the Leica camera.

4. **RESULTS**

4.1. PCR of *Foxa3* 1.5 kb and 2.0 kb Promoters

The purpose of this project was to attempt to define the minimal length of *foxa3* promoter that confers correct spatial temporal localization of the transcript in the endoderm. The experimental design entailed amplifying different *foxa3* promoter lengths by PCR, cloning them into a Topo vector, subcloning them upstream from a GFP gene in a reporter vector, microinjecting the reporter plasmids into zebrafish embryos, and visualizing GFP fluorescence at various stages of development.

Two different lengths of the *foxa3* promoter 1.5 and 2.0 kb long were chosen for analysis because the full-length 4.0 kb promoter would not amplify by PCR. Figure 14 shows a map of the 4.0 kb *foxa3* promoter. Note that the two fragments analyzed in this project overlap; the 2.0 kb sequence (both the two blue boxes) begins in the middle of the 4.0 kb fragment sequence, and the 1.5 kb sequence (the light blue box only) begins 500 bp into the 2.0 kb sequence.

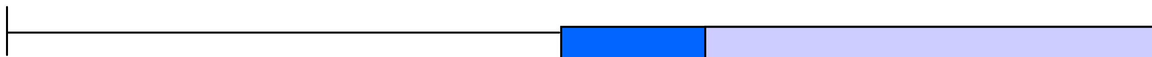


Figure 14. Map of the 4.0 kb Zebrafish *Foxa3* Promoter and the Fragments Analyzed in this Study. The *Foxa3* 4.0 kb promoter is represented in its entire length. The *Foxa3* 2.0 kb sequence is denoted with the dark and light blue boxes. The *Foxa3* 1.5 kb sequence is denoted with the light blue box only.

PCR was used to amplify the two different promoter lengths using zebrafish genomic DNA as template. PCR bands of the expected 1.5 and 2.0 kb sizes were obtained (Figure 15, upper and lower panels, respectively).

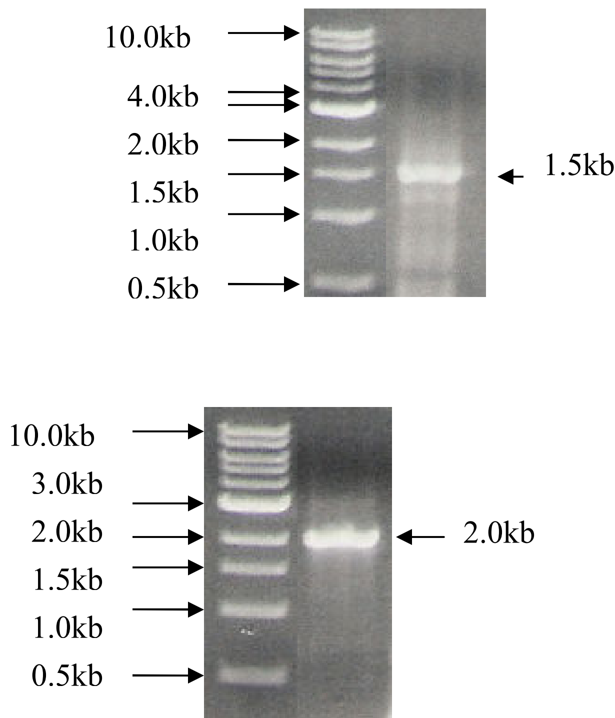


Figure 15. PCR of the *Foxa3* 1.5 kb Promoter (upper panel) and *Foxa3* 2.0 kb Promoter (lower panel). The first lane of each gel shows a 1.0 kb DNA ladder as marker.

4.2. Sequence Analysis of the *Foxa3* 1.5 kb & 2.0 kb Promoter Sequences

Taq polymerase is capable of inducing point mutations during PCR, so sequence analysis was performed directly on the uncloned PCR bands to prove they were indeed *foxa3* promoter with no serious mutations relative to the genbank entry. The sequence alignments are shown in Appendix B. The alignment for the 2.0 kb fragment is quite strong, with overall 99% homology, while the alignment for the 1.5 kb band is weak (40 % homology) perhaps indicating that this particular PCR product is not the actual *foxa3* promoter. Nevertheless, these differences were thought to be insignificant, so both PCR bands were cloned into the Topo vector (see below).

4.3. Cloning *Foxa3* Promoter PCR Bands into the Topo Vector

The *foxa3* promoters were cloned into the Topo vector by TA ligation. Potential positives were screened by restriction digestion. Figure 16 shows the screening of a positive clone containing the 1.5 kb insert.

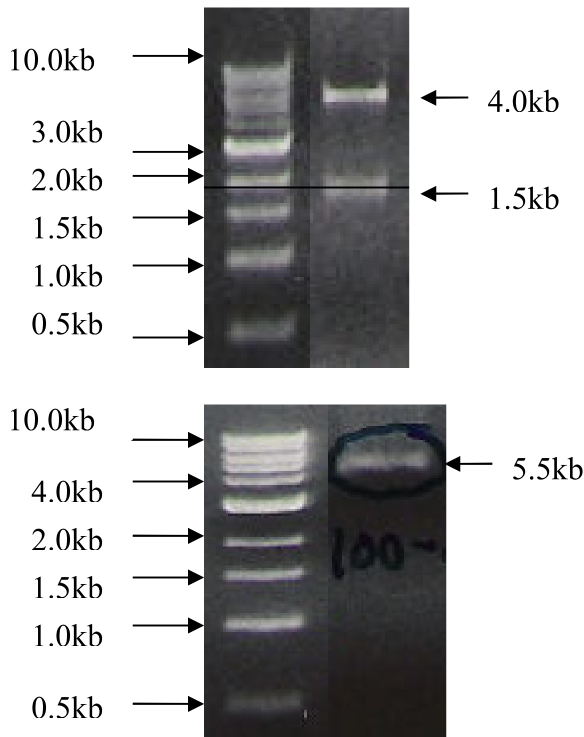


Figure 16. Screening of a *Foxa3* 1.5 kb Promoter Topo Clone Digested with *EcoRI* (upper panel) or *SmaI* (lower panel). The first lane of each gel shows a 1.0 kb DNA ladder as marker.

The plasmid in the upper panel was cut with *EcoRI* which flanks the insert site, releasing a 1.5 kb insert band and a 4.0 kb plasmid band. The sample in the lower panel was cut with *SmaI*, which cuts only once in the vector, producing a single 5.5 kb band (4.0 kb plasmid + 1.5 kb insert).

Figure 17 shows the screening of 16 plasmids for the *foxa3* 2.0 kb promoter fragment inserted in the Topo Vector.

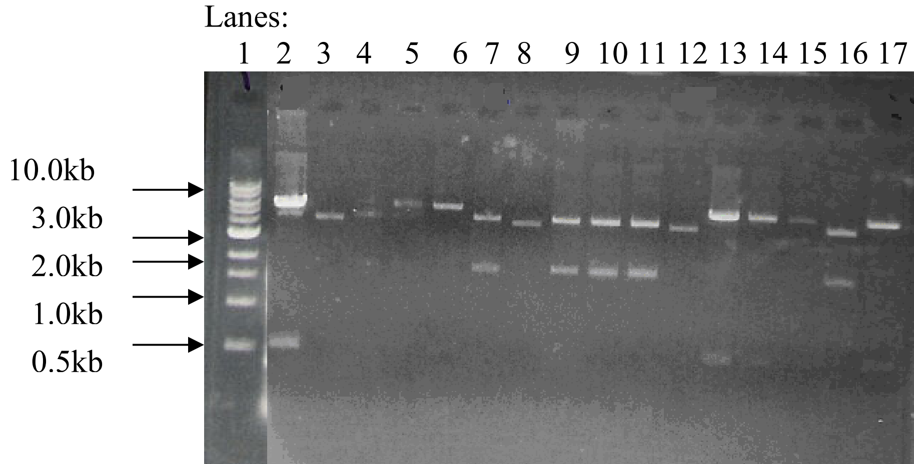


Figure 17. Restriction Enzyme Screening of Potential Positives for the *Foxa3* 2.0 kb Promoter Fragment Cloned in the Topo Plasmid. Digestion was performed with *ApaI*, which provides a directional screening. Lane 1 shows a 1.0 kb DNA ladder as marker.

In this case, the plasmids were cut with *ApaI* which cuts once in the vector, and once in the insert, to provide an orientation screen (although such orientation is not critical in the Topo vector). A forward promoter in the vector gives rise to band sizes of 500 bp and about 5.3 kb, while reverse promoters produce band sizes of 1.5 kb and 4.3 kb. Lanes 2, 13, 14, and 17 contained forward promoters, while lanes 7, 9, 10, 11, and 16 had reverse promoters. Since lane 13 had a strong positive band, this sample was chosen for further experiments.

4.4. Subcloning the Promoter Bands into a GFP Reporter Plasmid

The promoter-containing Topo plasmids and the GFP vector were both digested with *BstXI* since that site is upstream of the GFP gene, and *BstXI* flanks the promoter inserts (does not cut internal to the promoters). After the insert and vector were digested, a gel was run to determine if the digestion worked correctly (Figure 18). 1 µl of the digested sample was run versus 0.5 µl of the undigested sample. Figure 18 shows undigested and *BstXI* digested, respectively, GFP vector (lanes 3 and 4), 1.5 kb promoter

plasmid (lanes 5 and 6), and 2.0 kb promoter plasmid (lanes 7 and 8). In each case, bands of the correct sizes were obtained. Once it was determined that the GFP plasmid digestion was complete, the vector was digested with alkaline phosphatase to prevent self-ligation.

Lanes:
1 2 3 4 5 6 7 8

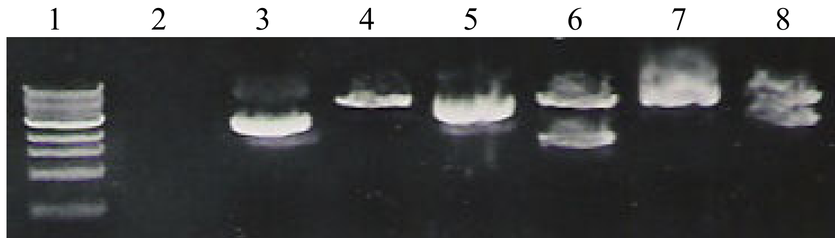


Figure 18. *BstXI* Digestion of the GFP Vector, *Foxa3* 1.5 kb Promoter, and *Foxa3* 2.0 kb Promoter. Lane 1 shows a 1.0 kb DNA ladder as marker. Lanes 3 and 4 contain undigested and digested GFP vector, respectively. Lanes 5 and 6 contain undigested and digested 1.5 kb promoter Topo plasmid, respectively. Lanes 7 and 8 contain undigested and digested 2.0 kb promoter Topo plasmid, respectively.

4.5. Concentration Gel for Promoter Inserts and GFP

During the ligation of insert into plasmids, the ratio of insert to plasmid can be critical to the cloning success, so it was necessary to approximate the mass of DNA present following the *BstXI* digestions. A sample dilution series was performed with different sample volumes: 5 µl , 2.5 µl , and 1 µl, loaded in that order. Figure 19a shows the *foxa3* 1.5 kb promoter series in lanes 2-4, and the GFP vector series in lanes 5-7. Figure 19b shows the 2.0 kb promoter series. When choosing a concentration for the ligation, the 5 µl volume was used for *foxa3* 1.5 kb promoter and the 2.0 kb promoter, while and 1 µl volume of the GFP vector was used. The reason for using a high volume for the inserts was to increase the chance of the insert ligating into the digested vector.

Lanes:
1 2 3 4 5 6 7

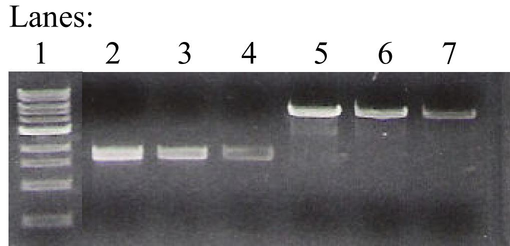


Figure 19a. ***Foxa3*** **1.5 kb Promoter and GFP Vector Sample Dilution Series Prior to Ligation.** Lane 1 shows a 1.0 kb ladder as marker.

Lanes:
1 2 3 4

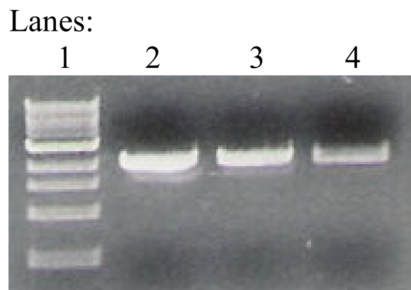


Figure 19b. ***Foxa3*** **2.0 kb Promoter Sample Dilution Series**. Lane 1 shows a 1.0 kb DNA ladder.

4.6. Symmetrical Cloning of *Foxa3* 1.5 kb and 2.0kb Promoters into the GFP Vector

Hypothetical plasmid maps were constructed for GFP vector containing the *foxa3* 1.5 kb promoter (Figure 20a) or 2.0 kb promoter (Figure 21a) to help design strategies for determining insert orientation. Although some promoters confer tissue specificity in either the forward or reverse orientations, we sought clones only in the forward promoter position to ensure GFP expression. The green mark at the twelve o'clock position denotes the first nucleotide position inside the vector in both figures, and the count of nucleotides increases clockwise; however, both plasmid vectors were transcribed counter clockwise. Ligation of insert into the *BstXI* site of the GFP vector was at the end of the insert not adjacent to the GFP sequence; this end begins the nucleotide count for the *foxa3* insert.

For the 1.5 kb promoter plasmid, restriction enzyme *HpaI* cuts once inside the GFP gene at site 855, and once in the promoter at site 1330. Ligation of the 1.5 kb promoter into the 4.0 kb plasmid generates a 5.5 kb plasmid. If the promoter was inserted in the forward direction, and the vector is cut with *HpaI*, bands 1.0 and 4.5 kb long are expected. If the promoter was inserted in the reverse orientation, bands of 2.2 and 3.3 kb are expected. The *HpaI* screening is shown in Figure 20b. Only 2 reverse orientation positives were obtained.

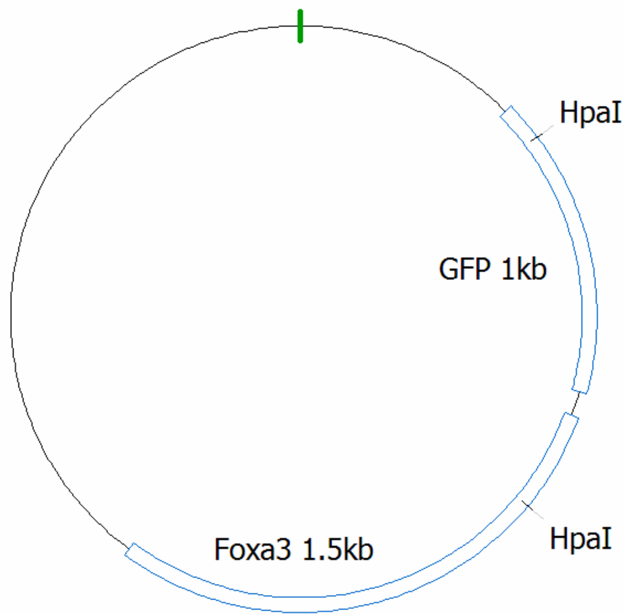


Figure 20a. Plasmid Map of the *Foxa3* 1.5 kb Promoter Ligated into the GFP Vector. Cut sites for *HpaI* are shown.

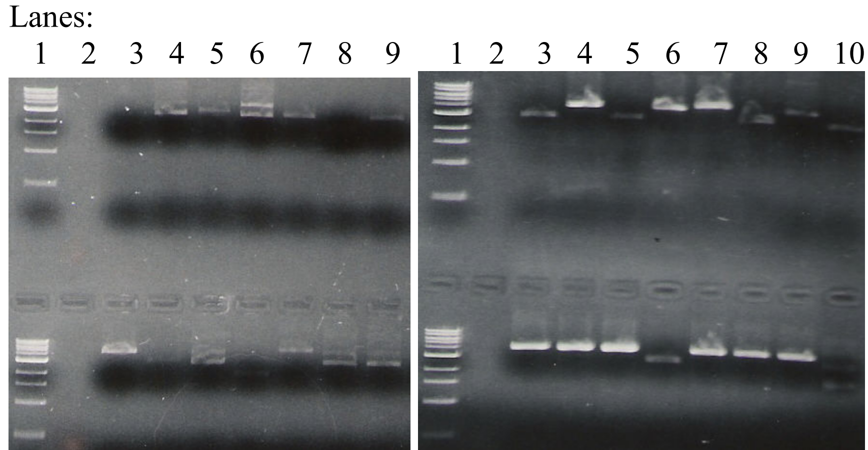


Figure 20b. *HpaI* Screening of the Cloning of the *Foxa3* 1.5 kb Promoter in the GFP Vector. The first lane of each panel shows a 1.0 kb DNA ladder.

For the vector containing the 2.0 kb insert, if the insert is in the forward position we expect bands of 1.0 and 5.0 kb, and in the reverse orientation we expect two 3.0 kb bands. The screen shown in figure 21b demonstrates that of 2 positives, both were in the negative orientation.

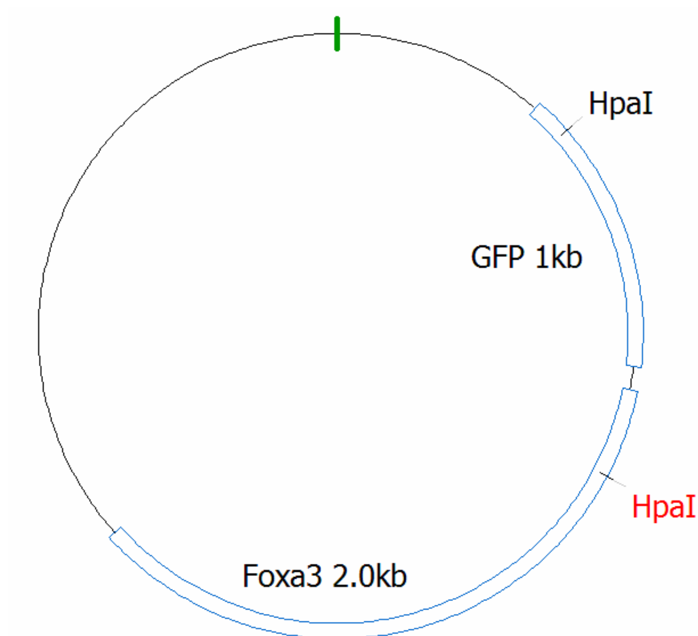


Figure 21a. Plasmid Map of the *Foxa3* 2.0 kb Promoter Ligated into the GFP Vector. Cut sites for *HpaI* are present.

Lanes:
1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9

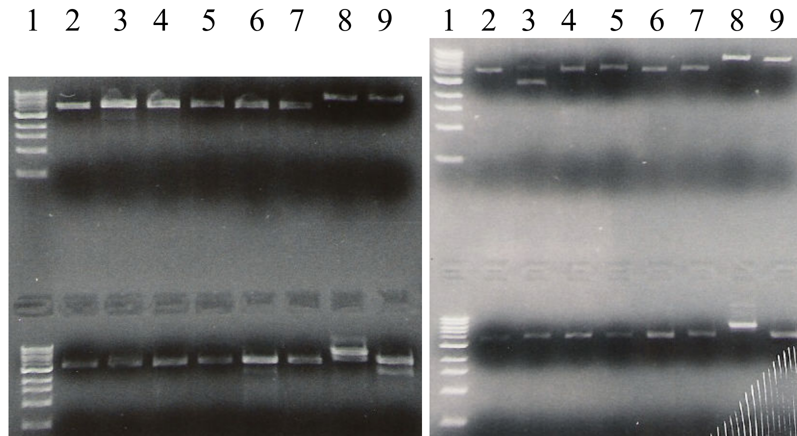


Figure 21b. *HpaI* Screening of the Cloning of the *Foxa3* 2.0 kb Promoter in GFP Vector. The first lane of each panel shows a 1.0 kb DNA ladder.

4.7. Directional Cloning of *Foxa3* 1.5 kb and 2.0 kb into the GFP Vector

As a second attempt to obtain a forward promoter, directional cloning was used. Both the insert and vector were cut with two restriction enzymes. These enzymes were determined by choosing one that cut at the 5' end of the *foxa3* promoter and another at the 3'end as it is oriented in the Topo vector. The orientation of *foxa3* 1.5 kb in the Topo vector was not known. Thus, the insert was digested for both orientations. For a forward promoter, *SacI* was used as the 5' enzyme and *NotI* was used for the 3' enzyme. In the case of the reverse promoter, *NotI* as used for the 5' end and *SpeI* was used for the 3'end. The *foxa3* 2.0 kb promoter was oriented backwards in the Topo vector, and thus the same enzymes were used to digest the insert as if the 1.5 kb promoter was backwards.

Figure 22 shows the confirmation of the digestion of the *foxa3* insert and the GFP vectors. Lane 1 in both rows is the 1kb marker. Lane 2 in the top row shows an undigested GFP vector. Lane 3 shows the *SacI-NotI* digested GFP vector, lane 4 and 5 show the *NotI-SpeI* digested GFP vectors. In the bottom row, lane 2 shows an undigested

foxa3 1.5 kb promoter, lane 3 and 4 show the digested samples of *SacI-NotI* and *NotI-SpeI*, respectively, lane 5 shows an undigested *foxa3* 2.0 kb promoter and lane 6 shows the digest *foxa3* 2.0 kb. In each case, bands of the correct sizes were obtained.

Lanes:
1 2 3 4 5

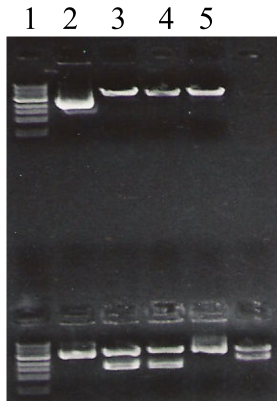


Figure 22. Digestion of the GFP Vector (Upper Row), *Foxa3* 1.5 kb Promoter (Bottom Row, Left), and *Foxa3* 2.0 kb Promoter (Bottom Row, Right). Lane 1 shows a 1.0 kb DNA ladder as marker.

Figure 23 shows a sample dilution series performed with different sample volumes: 5 µl , 2.5 µl , and 1 µl, loaded in that order. The *foxa3* 1.5 kb promoter series is shown in lanes 2-7, and the 2.0 kb promoter is shown in lanes 8-10. The GFP vector series is shown in the bottom row in lanes 2-10. When choosing a concentration for the ligation, the 5 µl volume was used for *foxa3* 1.5 kb and 2.0 kb promoter, and 1 µl volume of the GFP vector was used.

Lanes:
1 2 3 4 5 6 7 8 9 10

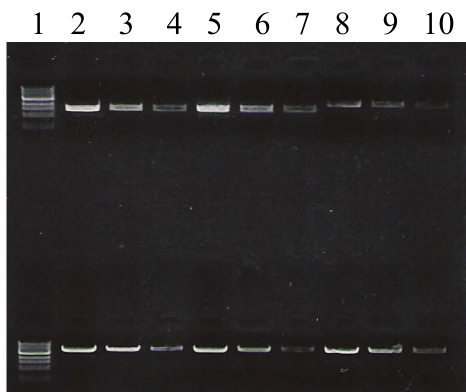


Figure 23. *Foxa3* 1.5 kb and 2.0 kb Promoter (Top Row) and GFP Vector (Bottom Row) Dilution Series Prior to Ligation. Lane 1 shows a 1.0 kb ladder as marker.

As with the symmetrical cloning, restriction enzyme *HpaI* creates bands 1.0 and 4.5 kb long for the 1.5 kb promoter plasmid and bands of 1.0 and 5.0 kb for the 2.0 kb promoter plasmid. The *HpaI* screenings for *foxa3* 1.5 kb and 2.0 kb promoters are shown in Figure 24 a and b. No forward promoters were obtained of *foxa3* 1.5 kb (Figure 24a), and 9 out of 12 samples are possible forward promoters of *foxa3* 2.0 kb. These forward promoters, seen in Figure 24b are in lanes 2, 3, 6, and 7 of the top row, and lanes 3-7 of the bottom row.

Lanes:

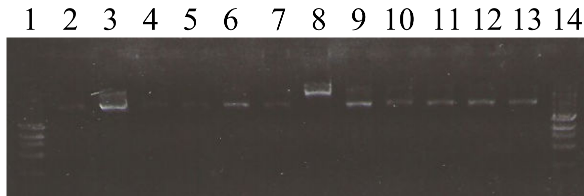


Figure 24a: Screening of the Directional Cloning of the *Foxa3* 1.5 kb Promoter in GFP Vector. The first and last lane of each panel shows a 1.0 kb DNA ladder.

Lanes:

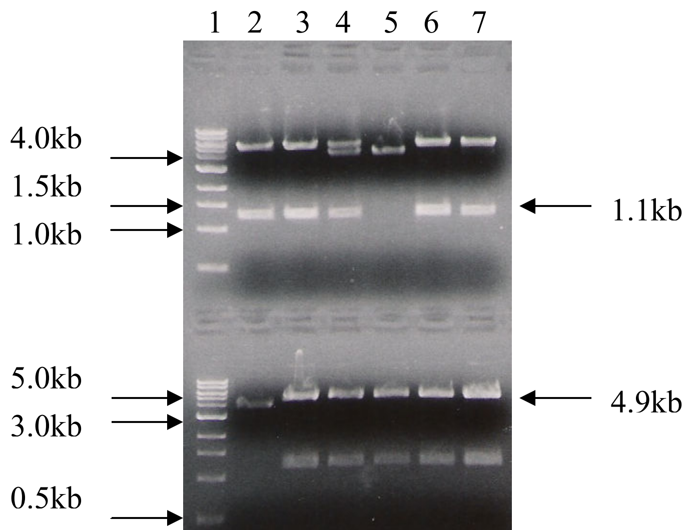


Figure 24b: Screening of the Directional Cloning of the *Foxa3* 2.0 kb Promoter in GFP Vector. The first lane of each panel shows a 1.0 kb DNA ladder.

4.8. Test of GFP Plasmid Microinjection and GFP Fluorescence

Previous experiments in our lab by Kristen Alexa with the promoter for gene *her5* indicated that a promoter length of 2.0 kb was sufficient for activating GFP expression in the expected anterior location (Figure 25). An embryo at 10 hours post fertilization (hpf) is shown in the left panel, and at 12 hpf in the right panel. The anterior of the embryo, oriented left in the figure, shows green fluorescence near the yolk at the 10 hpf stage, which was still present at 12 hpf. Expression occurred in the endoderm and the hindbrain regions, as expected for protein *her5*.

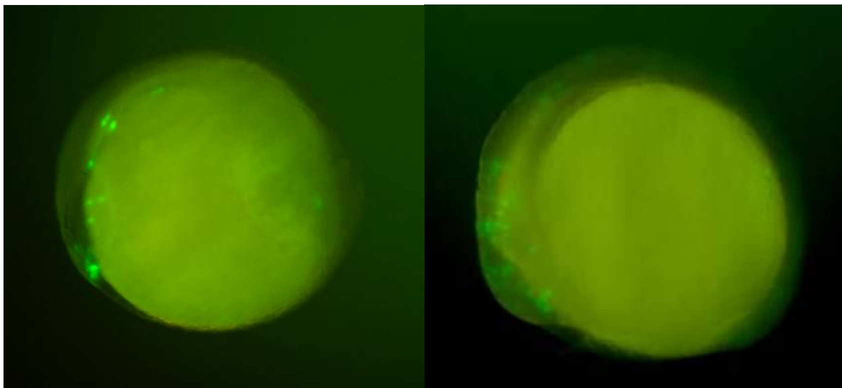


Figure 25. Sagerstrom Lab Data Showing GFP Fluorescence of the *Her5* 2.0 kb Promoter. The embryos are oriented anterior side to the left. The left panel denotes an embryo at 10 hpf, and the right panel 12 hpf. Research and pictures taken by Kristen Alexa, UMass Medical School Department of Biochemistry and Molecular Pharmacology.

We duplicated these experiments using the previously cloned *her5* 2.0 kb promoter plasmid (Figure 26) to test our microinjection technique, which can destroy a high percentage of embryos if not performed correctly. Two sets of zebrafish were crossed to produce embryos for the injection. In the first set, 265 embryos were injected with the *her5* 2.0 kb promoter, 169 lived to be scanned at 10 hpf (64% survival rate), while 96 were dead. Thirty-one, 18%, of the living embryos showed green fluorescence in the embryos, most in the anterior portion of the embryo as expected from previous experiments in our lab. Figure 26 shows a picture of an embryo with the protective

chorion on, with GFP in the endoderm and hindbrain, as previously observed in our lab. In some cases, the embryos fluoresced in undesirable locations, such as the ectoderm, but this was most likely due to an overinjection of plasmid caused by the injection pressure being set too high. The second set of embryos showed no GFP positive embryos out of the 246 embryos injected. Of the 246, 157 or 64% of the embryos survived. Overall, the *her5* injection experiment demonstrated our ability to microinject into newly fertilized zebrafish embryos, with embryo survival at least to the 10 hpf stage.

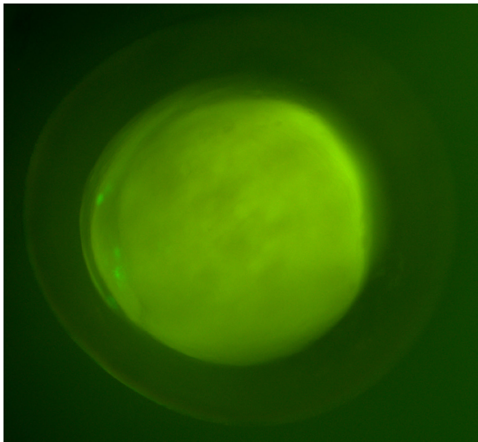


Figure 26. Test of Microinjection Technique Using Previously Cloned Her5 2.0 kb Promoter Plasmid. Embryo shown at approximately 10 hpf. Embryo anterior side is to the left, showing GFP fluorescence as expected.

5. DISCUSSION

5.1. Main Conclusions

The PCRed samples of *foxa3* 1.5 kb and 2.0 kb promoters were cloned into the Topo vector and midipreped. PCR can make transcriptional mistakes, so to ensure that the samples obtained experimentally were *foxa3,* the midipreped samples were sequenced from two different primer locations, T7-1 and R-1. From the comparison (shown in Appendix A) of the 1.5 kb and 2.0 kb sequences of *foxa3* from Ensembl (a website contain zebrafish gene sequence information) versus the sequences generated from our PCR samples it was determined that the cloned bands are *foxa3*.

In order to 'ligate' the GFP sequence to the *foxa3* promoter, the vector which already contained the GFP sequence, was digested with a single restriction enzyme to allow the promoter sequence to be inserted into the vector upstream of the GFP sequence. However, this allows for the promoter sequence to be inserted in both forward and reverse orientations with respect to the GFP sequence. It was necessary that the promoter sequence be inserted in the forward orientation with respect to the GFP to ensure transcription of the GFP once the promoter sequence was activated. Currently, only reverse oriented promoters have ligated into the vector. A forward promoter would be injected into the zebrafish embryos to verify Matsui et al.'s claim in their 2005 article that *foxa3* is an endoderm marker. GFP fluorescence would also show if there was expression in the midhindbrain as described in the experiments by Seiliez et al. in 2006 with *foxa3* and Goosecoid.

To avoid random orientation determination, directional cloning was also attempted. After the *foxa3* insert and GFP vector were digested with the same two

enzymes, one cutting at the 5' end of the insert and the other cutting at the 3' end of the insert, the two were ligated together as with the symmetrical cloning. Although this technique results in lower colony counts, all colonies contain forward inserts in the GFP vector.

Figure 12 in the Background Section and Figure 21 in the Results Section both show that a promoter length of 2,000 base pairs from the *her5* gene sequence was long enough to contain the proper elements to activate the GFP sequence downstream. The figures, with the anterior to the left, show green fluorescence near the yolk at the 10hpf stage, and by the 12hpf stage expression was still visible.

Figure 22 in the Results Section demonstrated our ability to microinject DNA into newly fertilized zebrafish embryos. The same techniques were used when injecting these embryos as was previously done in our lab with *her5* promoters. Fluorescence was seen in similar locations in the endoderm and midhindbrain as with the previous figures.

5.2. Future Experiments

As previously mentioned in the Background Section, not all promoters are long enough to contain the proper elements to activate the GFP sequence downstream from the promoter. The promoter lengths of 1.5 kb and 2.0 kb, even if forward orientations are obtained in the vector, may still not be long enough for *foxa3* to initiate GFP. Obtaining a longer promoter using PCR, such as the 4.0 kb sequence, would provide a greater chance that the promoter would activate GFP and green fluorescence would be seen in the embryo. Also attempt another round of directional cloning for *foxa3* 1.5 kb promoter if *foxa3* 2.0 kb promoter is long enough to activate GFP.

A transgenic fish is created after the *foxa3* promoter with GFP sequence is injected into the fish embryo. These first generation transgenic fish can be crossed to create a second generation of zebrafish that also fluoresce where *foxa3* is expressed. This technique would prove that the GFP expression regulated by the *foxa3* promoter is hereditary and can be passed through multiple generations. It also would allow for other transgenics to be created and injected into zebrafish and passed on to further the exploration of endoderm and pancreatic development genes and how they are regulated during gastrulation.

Another valuable experiment would be to implement a *foxa3* knock-down gene to observe if a zebrafish embryo develops normally. During the development of the embryo, endoderm and anterior region development would be monitored to find abnormalities in the formation of the endoderm, but more specifically the pancreas, and the head of the embryo. These data can help interpret how *foxa3* is involved during development and gastrulation, as well as where *foxa3* is vital for normal development.

REFERENCES

Argenton F, Zecchin E, Bortoliussi M (1999) Early Appearance of Pancreatic Hormone-Expressing Cells in the Zebrafish Embryo. *Mechanisms of Development* (**87**): 217-221.

Biemar F, Argenton F, Schmidtke R, et al. (2001) Pancreas Development in Zebrafish: Early Dispersed Appearance of Endocrine Hormone Expressing Cells and their Convergence to Form the Definitive Islet. *Developmental Biology* (**230**): 189-203.

Driever W, Solnica-Krezel L, Schier A, Neuhauss, S, et al. (1996) A Genetic Screen for Mutations Affecting Embryogenesis in Zebrafish. *Development* (**123**): 37-46.

Field H, Si Dong PD, Beis D, Stainier D (2003) Formation of the Digestive System in Zebrafish. *Developmental Biology* (**261**): 197-208.

Gilbert S (2003) *Developmental Biology.* 7th ed. Sunderland: Sinauer Associates Inc.

Hebrok M, Kim S, Melton D (1998) Notochord Repression of Endodermal Sonic Hedgehog Permits Pancreas Development. *Genes Development* (**12**): 1705-1713.

Hebrok M (2003) Hedgehog Signaling in Pancreas Development. *Mechanisms of Development* (**120**): 45-57.

Huang H, Tsai M (2000) Transcription Factors Involved in Pancreatic Islet Development. *J. Biomedical Scientific* (**7**): 27-34.

Huang H, Vogel S, Liu N, et al. (2001) Analysis of Pancreatic Development in Living Transgenic Zebrafish Embryos. *Molecular and Cellular Endocrinology* (**117**): 117-124.

Jonsson J, Carlsson L, Edlund T, Edlund H (1994) Insulin-Promoter-Factor 1 is Required for Pancreas Development in Mice. *Nature* (**371**): 606-609.

Kaestner K (2000) The Hepatocyte Nuclear Factor 3 (HNF3 or FOXA) Family in Metabolism. *Trends Endocrinology Metabolism* (**7**): 281-285.

Matsui T, Raya Á, Kawakami Y, Callol-Massot C, et al (2005) Noncanonical Wnt Signaling Regulates Midline Convergence of Organ Primordia During Zebrafish Development. *Genes and Development* (**19**): 164-175.

Milewski W, Duguay S, Chan S, Steiner D (1998) Conservation of PDX1 Structure, Function, and Expression in Zebrafish. *Endocrinology* (**139**): 1440-1449.

Pack M, Solnica-Kezel L, Malicki J, Neuhauss, S, et al. (1996) Mutations Affecting Development of Zebrafish Digestive Organs. *Development* (**123**): 321-328.

Ramalho-Santos M, Melton D, McMahon A (2000) Hedgehog Signals Regulate Multiple Aspects of Gastrointestinal Development. *Development* (**127**): 2763-2772.

Roy S, Qiao T, Wolff C, Ingham, P (2001) Hedgehog Signaling Pathway is Essential for Pancreas Specification in the Zebrafish Embryo. *Current Biology* (**11**): 1358-1363.

Seiliez I, Thisse B, Thisse C (2006) *Foxa3* and Goosecoid Promote Anterior Neural Fate Through Inhibition of Wnt8a Activity Before the Onset of Gastrulation. *Developmental Biology* (**290**): 152-163.

Shivdasani R (2002) Molecular Regulation of Vertebrate Early Endoderm Development. *Developmental Biology* (**249**): 191-203.

Tiso N, Filippi A, Pauls S, Bortolussi M, Argenton F (2002) BMP Signalling Regulates Anteroposterior Endoderm Patterning in Zebrafish. *Mech Development* (**118**): 29-37.

APPENDIX A:

Bench Protocol: QIAprep Spin Miniprep Kit Using a Microcentrifuge

This protocol is designed for the purification of up to 20 µg high-copy plasmid DNA from 1–5 ml overnight *E. coli* culture in LB medium. New users and users wanting to purify low-copy plasmids and cosmids, large plasmids (>10 kb), and DNA prepared using other methods should refer to the detailed protocols provided in the *QIAprep Miniprep Handbook*, 2nd ed.

Things to do before starting

- Add RNase A solution to Buffer P1.
- Optional: Add LyseBlue reagent to Buffer P1.
- Add ethanol (96–100%) to Buffer PE.
- Check Buffers P2 and N3 for salt precipitation and redissolve at 37°C if necessary.

Procedure

1. **Resuspend pelleted bacterial cells in 250 µl Buffer P1 and transfer to a microcentrifuge tube.**
2. **Add 250 µl Buffer P2 and mix thoroughly by inverting the tube 4–6 times.**
   If using LyseBlue reagent, solution turns blue.
3. **Add 350 µl Buffer N3 and mix immediately and thoroughly by inverting the tube 4–6 times.**
   If using LyseBlue reagent, solution turns colorless.
4. **Centrifuge for 10 min at 13,000 rpm (~17,900 x *g*) in a table-top microcentrifuge.**
5. **Apply the supernatant (from step 4) to the QIAprep spin column by decanting or pipetting.**
6. **Centrifuge for 30–60 s. Discard the flow-through.**
7. **Recommended: Wash the QIAprep spin column by adding 0.5 ml Buffer PB and centrifuging for 30–60 s. Discard the flow-through.**
   This step is only required when using *endA*+ or other bacteria strains with high nuclease activity or carbohydrate content (see *QIAprep Miniprep Handbook* for more details)
8. **Wash QIAprep spin column by adding 0.75 ml Buffer PE and centrifuging for 30–60 s.**
9. **Discard the flow-through, and centrifuge for an additional 1 min to remove residual wash buffer.**
10. **To elute DNA, place the QIAprep column in a clean 1.5 ml microcentrifuge tube. Add 50 µl Buffer EB or water to the center of each QIAprep spin column, let stand for 1 min, and centrifuge for 1 min.**

Bench Protocol: QIAfilter Midi and Maxi Kits

New users are strongly advised to familiarize themselves with the detailed protocol provided in the *QIAfilter Plasmid Purification Handbook* before using this bench protocol.

Things to do before starting

- Add RNase A solution to Buffer P1.
- Optional: Add LyseBlue reagent to Buffer P1.
- Check Buffer P2 for SDS precipitation.
- Pre-chill Buffer P3 to 4°C.
- Grow bacterial culture as described in QIAfilter Plasmid Purification Handbook.

Procedure

A) Bacterial culture, harvest, and lysis

1. **Pellet ▲ 25 ml or ● 100 ml (high copy) or ▲ 50–100 ml or ● 250 ml (low copy) overnight LB culture at 6000 x g for 15 min at 4°C.**
2. **Homogeneously resuspend the bacterial pellet in ▲ 4 ml or ● 10 ml Buffer P1.**
3. **Add ▲ 4 ml or ● 10 ml Buffer P2, mix thoroughly by vigorously inverting 4–6 times, and incubate at room temperature for 5 min.**

   During the incubation prepare the QIAfilter Cartridge:

   Screw the cap onto the outlet nozzle of the QIAfilter Midi or QIAfilter Maxi Cartridge. Place the QIAfilter Cartridge in a convenient tube or the QIArack.
4. **Add ▲ 4 ml or ● 10 ml chilled Buffer P3, mix thoroughly by vigorously inverting 4–6 times.**

B) Bacterial lysate clearing

5. **Pour the lysate into the barrel of the QIAfilter Cartridge. Incubate at room temperature (15–25°C) for 10 min. Do not insert the plunger!**
6. **Equilibrate a ▲ QIAGEN-tip 100 or ● QIAGEN-tip 500 by applying ▲ 4 ml or ● 10 ml Buffer QBT, and allow the column to empty by gravity flow.**
7. **Remove the cap from the QIAfilter Cartridge outlet nozzle. Gently insert the plunger into the ▲ QIAfilter Midi or ● QIAfilter Maxi Cartridge and filter the cell lysate into the previously equilibrated QIAGEN-tip.**

C) Bind, wash, and elute on QIAGEN-tip

8. **Allow the cleared lysate to enter the resin by gravity flow.**
9. **Wash the QIAGEN-tip with ▲ 2 x 10 ml or ● 2 x 30 ml Buffer QC.**
10. **Elute DNA with ▲ 5 ml or ● 15 ml Buffer QF.**

D) Precipitate, wash, and redissolve plasmid DNA

11. **Precipitate DNA by adding ▲ 3.5 ml or ● 10.5 ml (0.7 volumes) room-temperature isopropanol to the eluted DNA and mix. Centrifuge at ≥15,000 x *g* for 30 min at 4°C. Carefully decant supernatant.**
12. **Wash DNA pellet with ▲ 2 ml or ● 5 ml room-temperature 70% ethanol and centrifuge at ≥15,000 x *g* for 10 min. Carefully decant supernatant.**
13. **Air-dry pellet for 5–10 min and redissolve DNA in a suitable volume of buffer.**

APPENDIX B

Foxa3 1.5kb Promoter Alignment with T7-1.

```
EMBOSS_001         1 TTGTAGGCAAATAGAAAAACTTAAATAAACCGTTTCAAATGCGCGCCCAA
50
                                           ........|..|....||||.|
EMBOSS_001         1                       NNNNNGNNCNNTNNGCATGCTC------
22

EMBOSS_001        51 TATTACCGCCAATTTGTATAATTTTCACTTGCATAGTGGATTTTGATTTG
100
                     .....||||||                   |.....|||||    ||.||
EMBOSS_001        23 NNNGGCCGCCA-------------------GTGNNNTGGAT----ATCTG
49

EMBOSS_001       101 TATACTTAGTTATTTAAGAAACCGATGTTCTTACAATTTAAAACAATCTC
150
                     .|.|.||.|           .||    ||||  ||.||   ..|...||.
EMBOSS_001        50 CAGAATTCG-----------CCC----TTCT--CATTT---GCCTNNCTG
79

EMBOSS_001       151 TTGAGTATTAATCACCATCCT-TCACTTACTGTGACCGTCGTGAAAGTGA
199
                     |.|   |....||||..|.|. .|..||.|||.|...||||     |.|.
EMBOSS_001        80 TAG---AAAGGTCACTGTTCAGGCCTTTGCTGGGTTAGTCG-----GCGT
121

EMBOSS_001       200 TAGCCGAGGGGATTTAACAACTTTTTAAACTGGCACCTCGGATTACTGAT
249
                     |. |.|.|.|||.|||.||..||          |.|||.|..||
EMBOSS_001       122 TT-CTGTGTGGAGTTAGCAGGTT----------CTCCTTGTGTT------
154

EMBOSS_001       250 GTAGACCCCCACCGTGATGAGTTTAAGCCCTGATGAGCTTATACCACCCT
299
                              |||    ||..||||...||   ..|.|||
EMBOSS_001       155 ---------CAC----ATAGGTTTCCTCC---GGGTGCT-----------
177

EMBOSS_001       300 GGCCTGAACACATCACTGATGACAGAAGAAGCCATTAAATGATTGACTGA
349
                       ||.|....|.|||| .|||.||    |||.|||
EMBOSS_001       178 --CCGGTTTTCCTCAC-AATGCCA----AAGACAT---------------
205

EMBOSS_001       350 TGGTGGTTTACGACCATCAAAAGCAGGACTGTCCGCGTGCCTTAGGCTAC
399
                       |.|||                    ||.|     |||..||.|
EMBOSS_001       206 --GAGGT--------------------ACAG-----GTGAATTGG-----
223

EMBOSS_001       400 CAGAAAAAAAAAATAACTAAAATGTAAACTATCTATCTATCTATCTATCT
449
```

```
                                                |||..|||..|
EMBOSS_001       224 -----------------------GTAGGCTAAAT----------------
234

EMBOSS_001       450 ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTGTCTAT
499
                                                                ||||.|
EMBOSS_001       235 -------------------------------------------TGTCCA-
240

EMBOSS_001       500 CTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT
549
                         |.||...|.|||.|||.|||..||.|.|.||..|||.|..|.|
EMBOSS_001       241 ----CAGTGATTGTGTGTGTGTGTGGGTGTATATGGATGTTTCCCAG---
283

EMBOSS_001       550 GTCTGTCTAATGATTTCACTCGCTGCGAAAACAGCAAGCTTTTTGTTAT
599
                     |..||      |.||.||   ||||..|...||.|.| |||.||
EMBOSS_001       284 GGATG------GGTTGCA---GCTGGAAGGGCATCTA-CTTCTT------
317

EMBOSS_001       600 TTAAAAAAGAATTCTAATAGACGTCTAATGGACGTCTAAACATAGACGGC
649
                                      .|.||.|.|..||||         ..||..|||
EMBOSS_001       318 -----------------AAAACATATGCTGGA---------TAAGTTGGC
341

EMBOSS_001       650 ATGGTTAATGCTAAACTTGGGCTATCAGTTAAAATCTAATCGACA-TCTA
698
                     .  |||.||.|  ..||..|||.|                ||.|| ..||
EMBOSS_001       342 T--GTTCATTC--CGCTGTGGCGA----------------CGGCAGATTA
371

EMBOSS_001       699 AGAATAGCCCAAATGTAGTCGTCAAATACAAAAGAATGAATGACTACAAG
748
                     |.|||.|..|.    |||.||  |||.|.|||..|.||.|..||.||.|.
EMBOSS_001       372 ATAATGGGACT----TAGCCG--AAAAAAAAATAACTGTAAAACGACTAC
415

EMBOSS_001       749 TGTAAAGTCTGTTTAAAATGTATTTGATGACGAATCTAGTTTTATTGTTA
798
                     |.|.||...|||..|...||.|.|||.        |....|.|||||...
EMBOSS_001       416 TTTGAAAAATGTGCACTCTGAACTTGG--------CATAATATATTGAGC
457

EMBOSS_001       799 GATTG-TTTTAGCCAAGCCAAATGCTTATGTCTATTAGACGTCTATTTAA
847
                     .|.|| .|||.|....||...||.||| ||..|..|.      |.|||..
EMBOSS_001       458 AAATGCCTTTGGATTGGCTTTATACTT-TGCATGATT------TTTTTTT
500

EMBOSS_001       848 CATAAAATTGCCTGGTGAGGTTGTTGGAACC--CCTGTTTGGGTCAACTA
895
                     ..|..||||.|||       ||.|.||||||  ||        .||||..
EMBOSS_001       501 TTTTTAATTACCT-------TTTTGGGAACCAACC--------GCAACCT
535
```

```
EMBOSS_001       896 TGGACAACTGTTG--AGTTAAAAAGTTTAATTTATATTTAATTA-----A
938
                     |..|| |||...|  ||.|||||...||||...|.|..|||.||     |
EMBOSS_001       536 TAAAC-ACTAAAGCAAGCTAAAAGTATTAAAAAAAAAATAAATAAAGAGA
584

EMBOSS_001       939 CAACATT--AACCAAGTCCACATCAGACTCAACATTGGGTTAAACAACCC
986
                     .|.|.||  ||..||....|.||.|....|||||||..           .
EMBOSS_001       585 AATCTTTTAAATGAAAAAAAAATAAATACCAACATTAN-----------T
623

EMBOSS_001       987 AGCATTTTTGCCTAGTCTGAGTTGCAATAGATGTCCCAATATCTTCTCCT
1036
                     ||.||||||..||.||.  ||||        ||.||.||||
EMBOSS_001       624 AGTATTTTTATCTGGTT--AGTT--------TGGCCTAATA---------
654

EMBOSS_001      1037 TTTTACTCCGTTCATTTTCATTGTAGGCCTGGAGTTGTCTGGGATGACAA
1086
                                    ||..|||.||.||.|            ||||...|
EMBOSS_001       655 ----------------TTAAATTTTAAGCAT------------GATGTATA
677

EMBOSS_001      1087 TTGGCATGAACGTGTCGGCCGGCCAATATACAGAACAGCAGGGGGCAAGA
1136
                     ||...||.||    .|..||....||...||..||           ||.|
EMBOSS_001       678 TTTCAATAAA----ACTACCNAGAAAATAACCNAA-----------AATA
712

EMBOSS_001      1137 GAGACCTTTAGGGACGCCTTCCTGAGTCCCGGGGCCCGTAGGGCGGTCCC
1186
                     .|...|||       |.|||.|||.            ||||||||
EMBOSS_001       713 AAACACTT-------GTCTTACTGT------------GTAGGGCG-----
738

EMBOSS_001      1187 TTACCTGC-GCGTGTTTAAAGCCTGAGGTGAGGCTTTCCTACATTACTCA
1235
                        ||||| |..|.||||||.||                 |.|||
EMBOSS_001       739 ---CCTGCTGNTTTTTTAAAACC-----------------AAATT-----
763

EMBOSS_001      1236 CCGCACAGAGCTGACACGGGCTCCACAGCTTCACTTTGGACTTCTCAGGA
1285
                                          |.||||..||.|||.|||       |..|
EMBOSS_001       764 ---------------------TTCACAAATTGACTGTGG-------AAAA
785

EMBOSS_001      1286 TTATACTTGAAAAT-GTTTACGGGTCAC---CAAAGTGAACTATAAGTTA
1331
                     |||||.||...||| |.|||....||||   |...|||.||    ||.|
EMBOSS_001       786 TTATATTTTTTAATGGGTTAATAATCACATGCCCTGTGCAC----AGCT-
830

EMBOSS_001      1332 ACTAATCA-TGAAAAAACAACTGATTGATCTACGGGACAGAAGATAACAG
1380
```

```
                     .|....|| |.|..|||.....|||||......|.|.      |..||.|
EMBOSS_001       831 GCNNGGCACTAATTAAAAGTAGGATTGCNAATTGNGT------ACNACTG
874

EMBOSS_001      1381 AAGAGCCTGACTTCCAAAACTTTATAGTTCATATTACGTGGAATTACAAA
1430
                     |.| ||.||        ..|||||...||...|||||        ||...
EMBOSS_001       875 ANG-GCATG--------TGCTTTAATTTTTCAATTAC--------ACTGN
907

EMBOSS_001      1431 AGTTTCTCTTCTTGTGGGACGTTTT-GTTTGTTTATTTATTTTAAAAGCG
1479
                     .|    |||||.|.|   ||.|||| |||||.||.      ..|||..|.
EMBOSS_001       908 GG----TCTTCATTT---ACATTTTAGTTTGNTTG------GGAAACNCN
944

EMBOSS_001      1480 AAATCCTTATAACAGCTCGAGGGCATTTTAGTACACGTTAGTCGTCAGTC
1529
                     ...|...| |.|.|||.|....|||         .|.||.|.|..|   |
EMBOSS_001       945 GGCTAAAT-TGATAGCNCNNNNGCA---------GCATTNGGCAAC---C
981

EMBOSS_001      1530 ACTATCC-CGTTTCTTACGAGTTACAACTTTATTTGGGGATGTTGAGCTC
1578
                     |..|.|. ..|||...|..
EMBOSS_001       982 AANAACTGNNTTTNNNATN
1000

EMBOSS_001      1579 CGTGAAGATGGAATCTCATGAAATTCCGGAGTGGAATCCTTTCTACAGTG
1628

EMBOSS_001      1001
1000

EMBOSS_001      1629 AGGCAAATGAG   1639

EMBOSS_001      1001              1000
```

Foxa3 1.5kb Promoter Alignment with R-1.

```
EMBOSS_001         1 TTGTAGGCAAATAGAAAAACTTAAATAAACCGTTTCAAATGCGCGCCCAA
50

EMBOSS_001         1
0

EMBOSS_001        51 TATTACCGCCAATTTGTATAATTTTCACTTGCATAGTGGATTTTGATTTG
100

EMBOSS_001         1
0

EMBOSS_001       101 TATACTTAGTTATTTAAGAAACCGATGTTCTTACAATTTAAAACAATCTC
150

EMBOSS_001         1
0

EMBOSS_001       151 TTGAGTATTAATCACCATCCTTCACTTACTGTGACCGTCGTGAAAGTGAT
200

EMBOSS_001         1
0

EMBOSS_001       201 AGCCGAGGGGATTTAACAACTTTTTAAACTGGCACC---TCGGATTACTG
247
                                        .......|.||.|..||   |||. ....|.
EMBOSS_001         1                    NNANNNNAGCTNGGTCCNGNTCGN-NNCNTA
30

EMBOSS_001       248 ATGTAGACCCCCACCGTGATGAGTTTAAGCCCTGATGAGCTTATACCACC
297
                     .|...|.||.|||..|||.|| |..|..|||||            .|.|.
EMBOSS_001        31 GTAACGGCCGCCAGTGTGCTG-GAATTCGCCCT------------TCTCA
67

EMBOSS_001       298 CTGGCCTGAACACATCACTGATGACAGAAGAAGCCATTAAATGATTGACT
347
                     .|.|||        ||||||..||.||     ||||.|.||.|      |
EMBOSS_001        68 TTTGCC--------TCACTGTAGAAAG-----GCCAATCAAAG------T
98

EMBOSS_001       348 GATGGTGGTTTACGACCATCAAAAGCAGGACTGTCCGCGTGCCTTAGGCT
397
                     |||...|.|.|.|...|| |||||.||..||..|.|.|...|.|.|.|.|
EMBOSS_001        99 GATATAGATATGCTTGCA-CAAAAACACCACAATTCACCATCATCATGTT
147

EMBOSS_001       398 ACCAGA----AAAAAAAAATAACTAAAATGTAAACTATCTATCTATCTAT
443
                       ||||    ..||...||||||.|||||.||....|.|   ||...|..
EMBOSS_001       148 --CAGATGAGCTAACCCAATAACAAAAATATATTTGACC---CTGAATGA
192
```

```
EMBOSS_001       444 CTAT--CTATC-TATCTATCTAT--CTATCTATCTATCTATCTATCTATC
488
                     |.|.  |..|| |.||.||.|.|  |||..||...||...|.....|.|.
EMBOSS_001       193 CAAAACCAGTCGTGTCAATTTTTGGCTAATTAAGAATAATTAAGCATTTA
242

EMBOSS_001       489 TATCTG-TCTATCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG---
534
                     .||..| |.||..||..||..||| ||.|....|.||||..|.||||
EMBOSS_001       243 AATAAGCTTTACATGGATGGATGT-TGGCAAAATATCTGAATATCTGGAA
291

EMBOSS_001       535 TCTGTCTGTC-TGTCTGTCTGTCTAATGATTTCACTCGCTGCGAAAACAG
583
                     |||.....|. |.....||||.||..||  |.||...||       ||.|
EMBOSS_001       292 TCTAAAAATAGTAAAATTCTGACTCTTG--TACATAAGC-------ACTG
332

EMBOSS_001       584 CAAGCTTTTTGTTATTTAAAAAAGAATTCTAATAGACGTCTAATGGACG
633
                     ||         ||.|||          .||||....||     .|||.|
EMBOSS_001       333 CA---------GTGATT----------ATCTATGGTAC-----TTGGGC-
357

EMBOSS_001       634 TCTAAACATAGACGGCATGGTTAATGCTAAACTTGGGCTATCAGTTAAAA
683
                              .|||..|.| ||||..||...||.|   .|.|...|.||.|
EMBOSS_001       358 ---------TGACCTCGT-GTTACCGCATTACAT---ATTTATATGAACA
394

EMBOSS_001       684 TCTAATCGACATCTAAGAATAGCCCAAATGTAGTCGTCAAATACAAAAGA
733
                     .|||.|   ..|||..|.||      ..||||...| |||.|||.|...|
EMBOSS_001       395 CCTATT---TTTCTTGGGAT------TTTGTAAAAG-CAATTACCATTTA
434

EMBOSS_001       734 ATGAATGACTACAAGTGTAAAGTCTGTTTAAAATGTATTTGATGACGAAT
783
                     .|.|||.|||.|||         |||||.||||| .|.||.|||.|  .|
EMBOSS_001       435 TTAAATTACTTCAA---------CTGTTGAAAAT-CACTTAATGTC--TT
472

EMBOSS_001       784 CTAGTTTTATTGTTAGATTGTTTTAGCCAAGCCAAATGCTTATGTCTATT
833
                     | .||||.||||          .|.||...|.....|.|||.  |.||||
EMBOSS_001       473 C-GGTTTAATTG----------ATTGCATTGAATCTTTCTTT--TATATT
509

EMBOSS_001       834 AGACGTCTATT-TAACATAAAATTGCCTGGTGAGGTTGTTGGAACCCCTG
882
                     ..|..|||||| ||.||      ||||     ||.....||  ||.|.|
EMBOSS_001       510 GTATCTCTATTATAGCA------TGCC-----AGAAAAATG--ACACAT-
545

EMBOSS_001       883 TTTGGGTCAACTATGGACAACTGTTGAGTTAAAAAGTTTAATTTATATTT
932
```

```
                            ||    ||..||  |||        ||||
EMBOSS_001        546 ------TC----ATTAAC--CTG--------AAAA---------------
560

EMBOSS_001        933 AATTAACAA--CA-TTAACCAAGTCCACATCAGACTCAACATTGGGTTAA
979
                      |.||..|||  || ||.||||.|  ||||||.||
EMBOSS_001        561 ACTTGGCAAGGCATTTCACCATG--CACATCTGA----------------
592

EMBOSS_001        980 ACAACCCAGCATTTTTGCCTAGTCTGAGTTGCAATAGATGTCCCAATATC
1029
                        ||..|||.            |||..||||           |||||.
EMBOSS_001        593 --AAGACAGT------------TCTTGGTTG-----------CCAATG--
615

EMBOSS_001       1030 TTCTCCTTTTTACTCCGTTCATTTTCATTGTAGGCCTGGAGTTGTCTGGG
1079
                        ||.||.|||     ||.|  |.|||.|.|| |||| |.|||..|
EMBOSS_001        616 --CTGCTGTTT-----GTGC--TATCAATTTA-GCCT-GTGTTTCC----
650

EMBOSS_001       1080 ATGACAATTGGCATGAACGTGTCGGCCGGCCAATATACAGAACAGCAGGG
1129
                      |....||.|...|||.|..||..|.|          ||||.
EMBOSS_001        651 AAACAAACTAAAATGTAAATGAAGAC----------ACAGT---------
681

EMBOSS_001       1130 GGCAAGAGAGACCTTTAGGGACGCCTTCCTGAGTCCCGGGGCCCGTAGGG
1179
                       |.||..||.|..||.|.|  |.|.|              ||||
EMBOSS_001        682 -GTAATTGAAAAATTAAAG--CACAT--------------GCCC------
708

EMBOSS_001       1180 CGGTCCCTTACCTGCGCGTGTTTAAAGCCTGAGGTGAGGCTTTCCTACAT
1229
                         ||..||.  |.|.|...||                ||..||||||.|
EMBOSS_001        709 ---TCAGTTG--TACACAATTT----------------GCAATCCTACTT
737

EMBOSS_001       1230 TACTCACCGCACAGAGCTGACACGGGCTCCACAGCTTCACTTTGGACTTC
1279
                      |                |.|...|.||.|..|||||.          |.|
EMBOSS_001        738 T----------------TAATTAGTGCCCTGCAGCTG----------TGC
761

EMBOSS_001       1280 TCAGGATTATACTTGAAAATGTTTACGGGTCACCAAAGTGAACTATAAGT
1329
                      .||||     .|.||..|.|.||.||      |||.....||.|||||.|
EMBOSS_001        762 ACAGG-----GCATGTGATTATTAAC------CCATTAAAAAATATAATT
800

EMBOSS_001       1330 TAACT-AATCA---TGAAAAAACAACTGATTGATCTACGGGACAGAAGAT
1375
                      |..|. |.|||   ||..||||      .|||.|.|.        ||.|.
EMBOSS_001        801 TTCCACAGTCAATTTGTGAAAA------TTTGGTTTT--------AAAAA
836
```

```
EMBOSS_001      1376 AACAGAAGAGCCTGACTTCCAAAACTTTATAGTTCATATTACGTGGAATT
1425
                     |.|||      |.|.|..||                   |||...|.|..
EMBOSS_001       837 AGCAG------CAGGCGCCC-------------------TACACAGTAAG
861

EMBOSS_001      1426 ACAAAAG-TTTCTCTTCTTGTGGGACGTTTTGTTTGTTTATTTATTTTAA
1474
                     ||||..| |||.|.||||     |...||||.|.|||...||||||..||
EMBOSS_001       862 ACAAGTGTTTTATTTTCT-----GTTATTTTCTCTGTAGTTTTATTGAAA
906

EMBOSS_001      1475 -AAGCGAAATCCTTATAACAGCTCGAGGGCATTTTAGTACACGTTAGTCG
1523
                      |..|...||.||||.||    ||.|    ||.||||.|||
EMBOSS_001       907 TATACATCATGCTTAAAA----TCTA----ATATTAGGACA---------
939

EMBOSS_001      1524 TCAGTCACTATCCCGTTTCTTACGAGTTACAACTTTATTTGGGGATGTTG
1573
                          .||||.||.|.|     ..|..|||.|||         .|||||
EMBOSS_001       940 -----AACTAACCAGAT-----AAAAATACTACT---------AATGTT-
969

EMBOSS_001      1574 AGCTCCGTGAAGATGGAATCTCATGAAATTCCGGAGTGGAATCCTTTCTA
1623
                          |||                   |||         .||..|||.|.
EMBOSS_001       970 -----CGT-------------------ATT---------TATTTTTTTTT
986

EMBOSS_001      1624 CAGTGAGGCAAATGAG   1639
                     ||.|.|....|.|.
EMBOSS_001       987 CATTTAAAAGATTN     1000
```

Foa3 2kb Promoter Alignment with T7-1.

```
EMBOSS_001         1
0

EMBOSS_001         1 NNNNNGGCCNNTNNNNATGCTCNGNGGCCGCCAGNNNNNNNNNNTATNNN
50

EMBOSS_001         1               TTCTTATTGTGAAATTTCTCATTGCAAATAGGGTGT
36
                                   |||||||||.|||..||||||||||||||.||||||
EMBOSS_001        51 NNNNATTCGCCCTTTTCTTATTGGGAANNTTCTCATTGCAAATTGGGTGT
100

EMBOSS_001        37 GATATTGCGTTAGACTGACTGATAATCAACAATCGAATAGCCTAAATTCT
86
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       101 GATATTGCGTTAGACTGACTGATAATCAACAATCGAATAGCCTAAATTCT
150

EMBOSS_001        87 GATCATTTTAACTATGATCTCTTTTAGATGATTGACTATGTGGGAACGTT
136
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       151 GATCATTTTAACTATGATCTCTTTTAGATGATTGACTATGTGGGAACGTT
200

EMBOSS_001       137 GCATTTTGAGGGTCAGGTCTCTCCTAATCCTGTCGACCCCCCGCCAATTG
186
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       201 GCATTTTGAGGGTCAGGTCTCTCCTAATCCTGTCGACCCCCCGCCAATTG
250

EMBOSS_001       187 ATCAATTACATCTCCATTCATTAGCCATTAGGAAATGCATTGACATTGTG
236
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       251 ATCAATTACATCTCCATTCATTAGCCATTAGGAAATGCATTGACATTGTG
300

EMBOSS_001       237 ATGTTCCTGTCTGCGCCTAACAACAGATGTGTATTCACATACTACAGTCG
286
                     |||||||||||||||||||||||||||||||.|.|..|||||||||||||
EMBOSS_001       301 ATGTTCCTGTCTGCGCCTAACAACAGATGTGGANTNNCATACTACAGTCG
350

EMBOSS_001       287 CGCATCTTCACGGGTCATTTCAAGCCAGAAAGTTTTAGCAAAAATGTTCA
336
                     ||||||||||||||||||||||||||||||||||||||||||||||||.|
EMBOSS_001       351 CGCATCTTCACGGGTCATTTCAAGCCAGAAAGTTTTAGCAAAAATGTTTA
400

EMBOSS_001       337 GTTCACACACAGAAAATGGAGCACAATTCGAATCTGGTGTTGAGCAGAAA
386
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       401 GTTCACACACAGAAAATGGAGCACAATTCGAATCTGGTGTTGAGCAGAAA
450
```

```
EMBOSS_001         387 CAAATTTGCACCTCAAACACGCAAAATAAAAGCAGCTGCTTAAAGATTAG
436
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         451 CAAATTTGCACCTCAAACACGCAAAATAAAAGCAGCTGCTTAAAGATTAG
500

EMBOSS_001         437 CTGCCTAAAAAATATTAGAATCCATCAGAACGACTGTAATGTAGAAGTAA
486
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         501 CTGCCTAAAAAATATTAGAATCCATCAGAACGACTGTAATGTAGAAGTAA
550

EMBOSS_001         487 AATAAGGAAAACATTTGTAGGCAAATAGAAAAACTTAAATAAACCGTTTC
536
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         551 AATAAGGAAAACATTTGTAGGCAAATAGAAAAACTTAAATAAACCGTTTC
600

EMBOSS_001         537 AAATGCGCGCCCAATATTACCGCCAATTTGTATAATTTTCACTTGCATAG
586
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         601 AAATGCGCGCCCAATATTACCGCCAATTTGTATAATTTTCACTTGCATAG
650

EMBOSS_001         587 TGGATTTTGATTTGTATACTTAGTTATTTAAGAAACCGATGTTCTTACAA
636
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         651 TGGATTTTGATTTGTATACTTAGTTATTTAAGAAACCGATGTTCTTACAA
700

EMBOSS_001         637 TTTAAAACAATCTCTTGAGTATTAATCACCATCCTTCACTTACTGTGACC
686
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         701 TTTAAAACAATCTCTTGAGTATTAATCACCATCCTTCACTTACTGTGACC
750

EMBOSS_001         687 GTCGTGAAAGTGATAGCCGAGGGGATTTAACAACTTTTTAAACTGGCACC
736
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         751 GTCGTGAAAGTGATAGCCGAGGGGATTTAACAACTTTTTAAACTGGCACC
800

EMBOSS_001         737 TCGGATTACTGATGTAGACCCCCACCGTGATGAGTTTAAGCCCTGATGAG
786
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         801 TCGGATTACTGATGTAGACCCCCACCGTGATGAGTTTAAGCCCTGATGAG
850

EMBOSS_001         787 CTTATACCACCCTGGCCTGAACACATCACTGATGACAGAAGAAGCCATTA
836
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         851 CTTATACCACCCTGGCCTGAACACATCACTGATGACAGAAGAAGCCATTA
900

EMBOSS_001         837 AATGATTGACTGATGGTGGTTTACGACCATCAAAAGCAGGACTGTCCGCG
886
```

```
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       901 AATGATTGACTGATGGTGGTTTACGACCATCAAAAGCAGGACTGTCCGCG
950

EMBOSS_001       887 TGCCTTAGGCTACCAG-AAAAAAAAAATAACTAAAATGTAAACTATCTAT
935
                     |||||||.|||||||| |||||||||||||||.||||||||||||||||.
EMBOSS_001       951 TGCCTTANGCTACCAGAAAAAAAAAAATAACTNAAATGTAAACTATCTAN
1000

EMBOSS_001       936 CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT
985

EMBOSS_001      1001
1000

EMBOSS_001       986 ATCTATCTGTCTATCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT
1035

EMBOSS_001      1001
1000

EMBOSS_001      1036 CTGTCTGTCTGTCTGTCTGTCTAATGATTTCACTCGCTGCGAAAACAGCA
1085

EMBOSS_001      1001
1000

EMBOSS_001      1086 AGCTTTTTTGTTATTTAAAAAAGAATTCTAATAGACGTCTAATGGACGTC
1135

EMBOSS_001      1001
1000

EMBOSS_001      1136 TAAACATAGACGGCATGGTTAATGCTAAACTTGGGCTATCAGTTAAAATC
1185

EMBOSS_001      1001
1000

EMBOSS_001      1186 TAATCGACATCTAAGAATAGCCCAAATGTAGTCGTCAAATACAAAAGAAT
1235

EMBOSS_001      1001
1000

EMBOSS_001      1236 GAATGACTACAAGTGTAAAGTCTGTTTAAAATGTATTTGATGACGAATCT
1285

EMBOSS_001      1001
1000

EMBOSS_001      1286 AGTTTTATTGTTAGATTGTTTTAGCCAAGCCAAATGCTTATGTCTATTAG
1335

EMBOSS_001      1001
1000
```

```
EMBOSS_001      1336 ACGTCTATTTAACATAAAATTGCCTGGTGAGGTTGTTGGAACCCCTGTTT
1385

EMBOSS_001      1001
1000

EMBOSS_001      1386 GGGTCAACTATGGACAACTGTTGAGTTAAAAAGTTTAATTTATATTTAAT
1435

EMBOSS_001      1001
1000

EMBOSS_001      1436 TAACAACATTAACCAAGTCCACATCAGACTCAACATTGGGTTAAACAACC
1485

EMBOSS_001      1001
1000

EMBOSS_001      1486 CAGCATTTTTGCCTAGTCTGAGTTGCAATAGATGTCCCAATATCTTCTCC
1535

EMBOSS_001      1001
1000

EMBOSS_001      1536 TTTTTACTCCGTTCATTTTCATTGTAGGCCTGGAGTTGTCTGGGATGACA
1585

EMBOSS_001      1001
1000

EMBOSS_001      1586 ATTGGCATGAACGTGTCGGCCGGCCAATATACAGAACAGCAGGGGGCAAG
1635

EMBOSS_001      1001
1000

EMBOSS_001      1636 AGAGACCTTTAGGGACGCCTTCCTGAGTCCCGGGGCCCGTAGGGCGGTCC
1685

EMBOSS_001      1001
1000

EMBOSS_001      1686 CTTACCTGCGCGTGTTTAAAGCCTGAGGTGAGGCTTTCCTACATTACTCA
1735

EMBOSS_001      1001
1000

EMBOSS_001      1736 CCGCACAGAGCTGACACGGGCTCCACAGCTTCACTTTGGACTTCTCAGGA
1785

EMBOSS_001      1001
1000

EMBOSS_001      1786 TTATACTTGAAAATGTTTACGGGTCACCAAAGTGAACTATAAGTTAACTA
1835
```

```
EMBOSS_001        1001
1000

EMBOSS_001        1836 ATCATGAAAAAACAACTGATTGATCTACGGGACAGAAGATAACAGAAGAG
1885

EMBOSS_001        1001
1000

EMBOSS_001        1886 CCTGACTTCCAAAACTTTATAGTTCATATTACGTGGAATTACAAAAGTTT
1935

EMBOSS_001        1001
1000

EMBOSS_001        1936 CTCTTCTTGTGGGACGTTTTGTTTGTTTATTTATTTTAAAAGCGAAATCC
1985

EMBOSS_001        1001
1000

EMBOSS_001        1986 TTATAACAGCTCGAGGGCATTTTAGTACACGTTAGTCGTCAGTCACTATC
2035

EMBOSS_001        1001
1000

EMBOSS_001        2036 CCGTTTCTTACGAGTTACAACTTTATTTGGGGATGTTGAGCTCCGTGAAG
2085

EMBOSS_001        1001
1000

EMBOSS_001        2086 ATGGAATCTCATGAAATTCCGGAGTGGAATCCTTTCTACAGTGAGGCAAA
2135

EMBOSS_001        1001
1000

EMBOSS_001        2136 TGAG   2139

EMBOSS_001        1001        1000
```

Foxa3 2kb Promoter Alignment with R-1.

```
EMBOSS_001           1 TTCTTATTGTGAAATTTCTCATTGCAAATAGGGTGTGATATTGCGTTAGA
50

EMBOSS_001           1
0

EMBOSS_001          51 CTGACTGATAATCAACAATCGAATAGCCTAAATTCTGATCATTTTAACTA
100

EMBOSS_001           1
0

EMBOSS_001         101 TGATCTCTTTTAGATGATTGACTATGTGGGAACGTTGCATTTTGAGGGTC
150
                                                   ..|.....||.|     ||..|
EMBOSS_001           1                             NNANNNNAGCTT-----GGTNC
17

EMBOSS_001         151 AGGTCTC----TCCTAATCCTGTCGACCCCCCGCCAAT-TGATCAATTAC
195
                       .|| |||    .||||.|       |.|..||||||.| ||.|..|.|.|
EMBOSS_001          18 CGG-CTCGGNNCCCTAGT-------AACGGCCGCCAGTGTGCTGGAATTC
59

EMBOSS_001         196 ATCTCCATTCATTAGCCATTAGGAAATGCATTGACATTGTGATG-TTCCT
244
                       . |.|...|||||.|||.          ||.||   |.|..|.| ||||.
EMBOSS_001          60 G-CCCTTCTCATTTGCCT----------CACTG---TAGAAAGGATTCCA
95

EMBOSS_001         245 GTCTGCGCCTAACAACAGATGTGTATTCACATACTACAGTCGCGCATCTT
294
                       .||.| |..|..|     |||.| ||||                ||||||
EMBOSS_001          96 CTCCG-GAATTTC-----ATGAG-ATTC----------------CATCTT
122

EMBOSS_001         295 CACGGGTCATTTCAAGCCAGAAAGTTTTAGCAAAAATGTTCAGTTCACAC
344
                       |||||..|....||..||            ||||    |..||||...||
EMBOSS_001         123 CACGGAGCTCAACATCCC------------CAAA----TAAAGTTGTAAC
156

EMBOSS_001         345 AC---AGAAAATGGAGCACAATTCGAATCTGGTGTTGAGCAGAAACAAAT
391
                       .|   |||||..|||              |.|||..|.|    .||.||.
EMBOSS_001         157 TCGTAAGAAACGGGA--------------TAGTGACGTG----TACTAAA
188

EMBOSS_001         392 TTGCACCTCAAACACGCAAAATAAAAGCAGCTGCT-TAAAGATTAGCTGC
440
                       .|||.|.|                   .||.||.| ||||||||. | ||
EMBOSS_001         189 ATGCNCTT-------------------GAGTTGTTATAAAGATTT-C-GC
217
```

```
EMBOSS_001       441 CTAAAAAATATTAGAATCCATCAGAACGACTGTAATGTAGAAGTAAAATA
490
                     .|..||||||....||| .|.||.||||.|
EMBOSS_001       218 TTNNAAAATAAATAAAT-AAACAAAACGTC--------------------
246

EMBOSS_001       491 AGGAAAACATTTGTAGGCAAATAGAAAAACTTAAATAAACCGTTTCAAAT
540
                           .||        |||..|||.||||||...|||     |||
EMBOSS_001       247 ------CCA--------CAAGAAGAGAAACTTTTGTAA-----TTC----
273

EMBOSS_001       541 GCGCGCCCAATATTACCGCCAATTTGTATAATTTTCACTTGCATAGTGGA
590
                        |.|..|||||.|.|       |.||.|.||||......||    ||.
EMBOSS_001       274 ---CACGTAATATGAAC-------TATAAAGTTTTGGAAGTCA----GGC
309

EMBOSS_001       591 TTTTGATTTGTATACTTAGTTATTTAAGAAACCGATGTTCTTACAATTTA
640
                     |.||          ||  |||||.|......|||..|.||...||.||
EMBOSS_001       310 TCTT----------CT--GTTATCTTCTGTCCCGTAGATCAATCAGTT--
345

EMBOSS_001       641 AAACAATCTCTTGAGTATTAATCACCATCCTTCACTTACTGTGACCGTCG
690
                       ....|.||.|||.||.|.|.|. .||..|||||||. .||||||  ||
EMBOSS_001       346 --GTTTTTTCATGATTAGTTAACT-TATAGTTCACTTT-GGTGACC--CG
389

EMBOSS_001       691 TGAAAGTGATAGCCGAGGGGATTTAACAACTTTTTAAACTGGCACCTCGG
740
                     |                         .|||.||||.||.|...|.|
EMBOSS_001       390 T-------------------------AAACATTTTCAAGTATAATC----
410

EMBOSS_001       741 ATTACTGATGTAGACCCCCACCGTGATGAGTTTAAGCCCTGATGAGCTTA
790
                         |||| |.||.   |||..||||.|          |||..||||
EMBOSS_001       411 ----CTGA-GAAGT---CCAAAGTGAAG----------CTGTGGAGC---
439

EMBOSS_001       791 TACCACCCTGGCCTGAACACATCACTGATGACAGAAGAAGCCATTAAATG
840
                          ||.||        .||.|.|||           .||
EMBOSS_001       440 -----CCGTG--------TCAGCTCTG-----------TGC---------
456

EMBOSS_001       841 ATTGACTGATGGTGGTTTACGACCATCAAAAGCAGGACTGTCCGCGTGCC
890
                               ||||..|           ||.|.||||..| ||.|  .||
EMBOSS_001       457 ----------GGTGAGT-----------AATGTAGGAAAG-CCTC--ACC
482

EMBOSS_001       891 TTAGGCTACCAGAAAAAAAAAATAACTAAAATGTAAACTATCTATCTATC
940
```

```
                          |.|||||.                         |||||
EMBOSS_001            483 TCAGGCTT-------------------------TAAAC------------
495

EMBOSS_001            941 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA
990
                                                                          .|..
EMBOSS_001            496 ----------------------------------------------ACGC
499

EMBOSS_001            991 TCTGTCTATCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTC
1040
                          .|.|       ||..|  .|.|.|.|...|.|.|  .|.|  .|.|  ||
EMBOSS_001            500 GCAG-------GTAAG--GGACCGCCCTACGGGC--CCCG--GGAC--TC
534

EMBOSS_001           1041 TGTCTGTCTGTCTGTCTAATGATTTCACTCGCTGCGAAAACAGCAAGCTT
1090
                          .|...|.| |||  .|||..|||.||.||.||..|     |.||
EMBOSS_001            535 AGGAAGGC-GTC--CCTATAGATCTCTCTTGCCCC-----CTGC------
570

EMBOSS_001           1091 TTTTGTTATTTAAAAAAGAATTCTAATAGACGTCTAATGGACGTCTAAAC
1140
                             ||||.|                      ||.||.|||.||.|
EMBOSS_001            571 ---TGTTCT----------------------GTATATTGGCCGGC-----
590

EMBOSS_001           1141 ATAGACGGCATGGTTAATGCTAAACTTGGGCTATCAGTTAAAATCTAATC
1190
                               ||.||. |||.||||              ||.||...|||..|
EMBOSS_001            591 -----CGACAC-GTTCATGC--------------CAATTGTCATCCCA--
618

EMBOSS_001           1191 GACATCTAAGAATAGCCCAAATGTAGTCGTCAAATACAAAAGAATGAATG
1240
                          ||||.||         ||      ||.|.|.||||..|||.|||.|   |
EMBOSS_001            619 GACAACT---------CC------AGGCCTAAAATGAAAATGAACG---G
650

EMBOSS_001           1241 ACTACAAGTGTAAAGTCTGTTTAAAATGTATTTGATGACGAATCTAGTTT
1290
                          |.||.||..||.|||            .||||.|  |||  |||||
EMBOSS_001            651 AGTAAAAAGGTGAAG------------ATATTGG--GAC--ATCTA----
680

EMBOSS_001           1291 TATTGTTAGATTGTTTTAGCCAAGCCAAATGCTTATGTCTATTAGACGTC
1340
                                       ||..|.|..||.| .|.||
EMBOSS_001            681 -------------TTGCAACTCAGAC-TANGC------------------
698

EMBOSS_001           1341 TATTTAACATAAAATTGCCTGGTGAGGTTGTTGGAACCCCTGTTTGGGTC
1390
                                    ||||.|||    || |||||||..|..|..|| |||.|||
EMBOSS_001            699 ----------AAAAATGC----TG-GGTTGTTTAACNCNATG-TTGAGTC
732
```

```
EMBOSS_001      1391 AACTATGGACAACTGTTGAGTTAAAAAGTTTAATTTATATTTAATTAACA
1440
                     ...|.|||||     ||| |||||.....||||||.|...|.||||||..
EMBOSS_001       733 TGATGTGGAC-----TTG-GTTAATGTNGTTAATTAAATATAAATTAAAC
776

EMBOSS_001      1441 ACATTAACCAAGTCCACATCAGACTCAACATTGGGTTAAACAACCCAGCA
1490
                     ...|||||    ||.|||...|.|                        ||
EMBOSS_001       777 TTTTTAAC----TCAACAGTTGTC------------------------CA
798

EMBOSS_001      1491 TTTTTG--CCTAGTCTGAGTTGCAATAGATGTCCCAATATCTTC-TCCTT
1537
                     |..|||  ||||..|...|.                 |..|..| .|||
EMBOSS_001       799 TANTTGACCCTANACNNGGN-----------------TTNCNNCNNCCT-
830

EMBOSS_001      1538 TTTACTCCGTTCATTTTCAT--TGTAGGCCTGGAGTTGTCTGGGATGACA
1585
                         |.|....||.|||..|  ..|.|.|.|....|.|.|..|  |..||
EMBOSS_001       831 ----CNCNNGGCANTTTGGTNNNNTNGNCGTTTNNTCGNCNTG--TAGCA
874

EMBOSS_001      1586 ATTGGCATGAACGTGTCGGCCGGCCAATAT-----ACAGAACAGCAGGGG
1630
                     ...|||.|||..............|||||.     |.....||.||...
EMBOSS_001       875 TNNGGCTTGACTANNNNNNNNTAACAATAANNNNGATNNGNCANCANTN-
923

EMBOSS_001      1631 GCAAGAGAGACCTTTAGGGACGCCTT-------------------CCTGA
1661
                                |.||........|||                   ...|.
EMBOSS_001       924 -----------CNTTTNANNNNACTTNNNANNNNNNNNNNNNNTNNTTNGN
962

EMBOSS_001      1662 GTCCCGGGGCCCGTAGGG----------CGGTCCCTTACCTGCGCGTGTT
1701
                     .|...|....|......|          ..|.|..|..
EMBOSS_001       963 NTNTTGNNNNCTNNNNNGNNNNNNNNNNNNGNCNNTNN
1000

EMBOSS_001      1702 TAAAGCCTGAGGTGAGGCTTTCCTACATTACTCACCGCACAGAGCTGACA
1751

EMBOSS_001      1001
1000

EMBOSS_001      1752 CGGGCTCCACAGCTTCACTTTGGACTTCTCAGGATTATACTTGAAAATGT
1801

EMBOSS_001      1001
1000

EMBOSS_001      1802 TTACGGGTCACCAAAGTGAACTATAAGTTAACTAATCATGAAAAAACAAC
1851
```

```
EMBOSS_001       1001
1000

EMBOSS_001       1852 TGATTGATCTACGGGACAGAAGATAACAGAAGAGCCTGACTTCCAAAACT
1901

EMBOSS_001       1001
1000

EMBOSS_001       1902 TTATAGTTCATATTACGTGGAATTACAAAAGTTTCTCTTCTTGTGGGACG
1951

EMBOSS_001       1001
1000

EMBOSS_001       1952 TTTTGTTTGTTTATTTATTTTAAAAGCGAAATCCTTATAACAGCTCGAGG
2001

EMBOSS_001       1001
1000

EMBOSS_001       2002 GCATTTTAGTACACGTTAGTCGTCAGTCACTATCCCGTTTCTTACGAGTT
2051

EMBOSS_001       1001
1000

EMBOSS_001       2052 ACAACTTTATTTGGGGATGTTGAGCTCCGTGAAGATGGAATCTCATGAAA
2101

EMBOSS_001       1001
1000

EMBOSS_001       2102 TTCCGGAGTGGAATCCTTTCTACAGTGAGGCAAATGAG   2139

EMBOSS_001       1001                                          1000
```