Stony Brook University

Use of Native DNA Binding Domains Tethered to the Fok1 Endonuclease to Target Nodal Related Genes via the Activin Response Element

A Thesis Presented

by

Jessica L. Minder

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Master of Science

in

Biochemistry and Cell Biology

Stony Brook University

December 2012

Stony Brook University

The Graduate School

Jessica L. Minder

We, the thesis committee for the above candidate for the

Master of Science degree, hereby recommend

acceptance of this thesis.

Dr. Howard I. Sirotkin
Associate Professor
Department of Neurobiology and Behavior

Dr. Benjamin L. Martin
Assistant Professor
Department of Biochemistry and Cell Biology

This thesis is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Thesis

Use of Native DNA Binding Domains Tethered to the Fok1 Endonuclease to Target Nodal Related Genes via the Activin Response Element

by

Jessica L. Minder

Master of Science

in

Biochemistry and Cell Biology

Stony Brook University

2012

Use of zinc finger nucleases is an established method for gene targeting in zebrafish. Such strategies are efficient in targeting a single gene of interest. Our study attempts to target many Nodal related genes by targeting the Activin Response Element (ARE) through the use of native DNA binding domains tethered to the Fok1 endonuclease. In this light, genes are targeted in a Nodal-dependent manner, and can retain transcript levels sustained by other signaling pathways. This technique holds the potential to be a high throughput means of generating mutations in the regulatory regions of various Nodal responsive genes with the use of one set of constructs.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ºC	Degrees Celsius
µl	Microliter
Aa	Amin Acid
ActRIB	Activin Receptor Type IB
ActRIIa	Activin Receptor Type IIa
ActRIIb	Activin Receptor Type IIb
ALK-4	Activin Receptor-Like Kinase 4
Amp	Ampicillin
ARE	Activin Response Element
ARF	Activin Response Factor
BLAST	Basic Local Alignment Search Tool
BMP	Bone Morphogenic Protein
Bp	Base Pairs
Cyc	Cyclops
DBD	Fast-1 DNA Binding Domain
DIC	Differential Interference Contrast
DIC	Dye Injected Control
DNA	Deoxyribonucleic Acid
EGF	Epidermal Growth Factor
EMSA	Electrophoretic Mobility Shift Assay
Ephnb2	Ephrin Receptor 2
Fast-1/FoxHI	Forkhead Activin Signal Transducer I
Flh	Floating Head
Foxa3	Forkhead Box A3
FRET	Fluorescence Resonance Energy Transfer
Fynb	FYN Oncogene Related to SRC, FGR, YES b
Gsc	Goosecoid
Gta2	Gata3
Hpf	Hours Past/Post Fertilization
HRMA	High Resolution Melt Analysis
Lft1	Lefty 1
Lft2	Lefty 2
Lhx1a	LIM Homeobox 1a
MH1	Mad Homology 1
MH2	Mad Homology 2
Mg	Milligrams
Ml	Microliter
mRNA	Messenger RNA
Ndr1	Nodal-Related 1 (Squint)
NES	Nuclear Export Signal
Ng	Nanograms

List of Abbreviations Continued

NHEJ	Non Homologous End Joining
NLS	Nuclear Localization Signal
Ntl	No Tail
Oep	One-Eyed Pinhead
Otx2	Orthodenticle Homolog 2
PCR	Polymerase Chain Reaction
PK	Proteinase K
Pitx2	Paired-Like Homeodomain Transcription Factor 2
Pg	Picograms
Pmol	Pico Molar
Rarab	Retinoic Acid Receptor, Alpha B
RNA	Ribonucleic Acid
RO	Reverse Osmosis
S	Seconds
SBD	Fast-1 Smad2 Binding Domain
SDM	Site Directed Mutagenesis
Sqt	Squint
TARAM-A	Type I Activin Receptor
TGFβ	Transforming Growth Factor Beta
Tgif1	TGFB-Induced Factor Homeobox 1
TLB	Tupfel Long Fin Cross to Brian's Wild Type
Xnr1-6	Xenopus nodal related 1-6
Znf143	Zinc Finger Protein 143a

Acknowledgements

Thank you to all of the members of Dr. Sirotkin's Lab. Cara Moravec and Carolyn Milano, thank you for helping me get acquainted with the lab. To Erika Wunderlich who did a wonderful job in taking care of the fish, and Andrew Taibi who performed all the injections and helped with assessing injected embryos and photography. Special thanks to Dr. Sirotkin for taking me as a rotation student and for all that I have learned in the time as a member of your lab. Thank you to Dr. Sirotkin and Dr. Martin for your aid in the development of this work.

Introduction

Nodal signaling plays an essential role in many aspects of early vertebrate embryogenesis. The pathway involves differentiation of cells into mesoderm and endoderm, as well as left-right patterning (Schier, 2003). In having such an important role in development, it would be valuable to develop a technique to generate mutations at different downstream points in this pathway by disrupting Nodal-dependent regulation of target genes. The goal of this study is to generate various genetic mutants in the Nodal pathway using one set of constructs and to show proof of principle that in zebrafish, one can target regulatory elements by attaching the Fok1 endonuclease to native DNA binding domains. These native DNA binding domains are responsible for recruitment of Fok1 to the DNA. Targeting the ARE would result in lack of gene expression in a Nodal-dependent manner, in which the same genes would still be functional in other signaling pathways. Target specific mutant studies would give further insights into the roles of specific genes, and may possibly lead to the discovery of a novel gene in the Nodal pathway.

Nodal Signaling

Nodals are classified as part of the transforming growth factor beta (TGFβ) superfamily. Nodal ligands are restricted to deuterostomes and are not found in *D. melanogaster or C. elegans.* Nodal signaling is most involved in processes specific to vertebrate development, perhaps evolving for that purpose and explaining why the pathway is not as highly conserved as Wnt or Notch. The Nodal family consists of Nodal in *M. musculus*, Cyclops, Squint and Southpaw in *D. rerio*, and Xnr1-6 in *X. laevis,* as well as others not mentioned here (Schier, 2003).

In zebrafish, mutations in the *squint* (*sqt*) and *cyclops* (*cyc*) genes result in lack of mesendodermal tissue, and failure to establish cardiac and gene expression asymmetries (Sirotkin, Gates, Kelly, Schier, Talbot, 2000). The observed phenotypes in zebrafish embryos include cyclopia, bent ventral body axis, tail truncations, lack of notochord, and somite deformations. Overexpressed *sqt* and *cyc* done by microinjection of mRNA coding for these genes induces dorsal mesoderm and in some embryos, local formation of a secondary body axis. Injection of mRNA encoding *sqt* into the extraembryonic yolk syncytial layer acts as a morphogen by inducing differentiation into dorsal mesoderm in neighboring embryonic blastomeres (Schier, Talbot, 2001). This evidences a characteristic of Nodals, which is the ability to act locally or distally. Nodals also work via concentration gradients. Unlike Squint, Cyclops acts only within short ranges (Schier, 2003).

Nodal proteins are synthesized in a proprotein form and are cleaved into active form upon recruitment of proprotein convertases. Nodal signaling is propagated by type I and II Activin receptors that are serine/threonine kinases. The type I receptor is activated by phosphorylation and subsequently complexes with, and activates the type II receptor. This results in the phosphorylation of downstream Smads that act as transcription factors and regulate gene expression (Schier, 2003). The Activin type 1

receptor TARAM-A was identified in zebrafish and was linked to Nodal through its similarity to the Activin receptor-like kinase 4 (ALK-4), also known as Activin receptor type IB (ActRIB) Nodal-related receptor in *Xenopus* (Peyriéras, Strähle, Rosa, 1998). TARAM-A is part of the TGFβ superfamily and is known to be involved in signaling related to the Activin and Bone Morphogenic Protein (BMP) pathways (Chen, Wang, Lin, Chang, Chung, Ying, 2006). However, there is a convincing line of evidence suggesting it as the receptor for signaling via Cyclops and Squint (Schier, 2003). TARAM-A is present in the early embryo as a maternal transcript with zygotic transcripts being present in the mesoderm and migrating axial mesoderm (Renucci, Lemarchandel, Rosa, 1996). Injection of RNA encoding a constitutively active form of TARAM-A into a marginal blastomere of a 16 celled zebrafish embryo resulted in derivative cells being fated to mesoderm or endoderm (Peyriéras, *et al.,* 1998). Thus, TARAM-A is linked to tissues related to Nodal signaling, supporting evidence for cross-talk between the Nodal and Activin pathways. Specificity for the Nodal pathway arises from restricted expression patterns and interactions with factors such as one-eyed pinhead (oep), which does not aid in the transduction of Activin signaling (Schier, 2003).

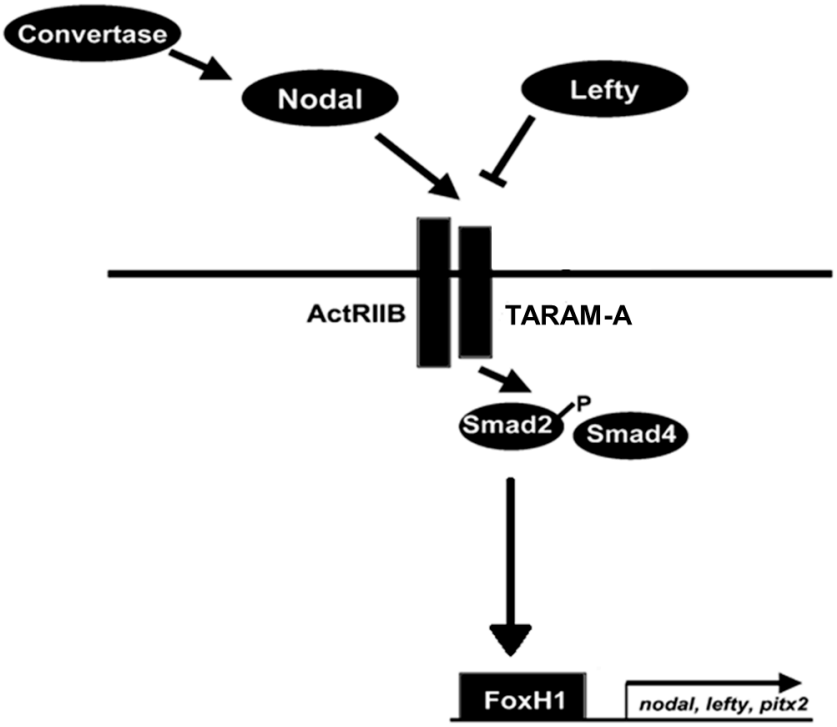
Activated receptors propagate signaling via Smad transcription factors. Receptor Smads are divided between receptor pathways in which Smad1, 5, 8 and 9 respond to BMP receptor kinases. Smads 2 and 3 act downstream to TGFβ and Activin receptors. Smads have two highly conserved domains that are conserved all the way from nematodes to humans. The amino terminal domain is the Mad homology 1 (MH1) domain and the carboxy terminal domain is the Mad homology 2 (MH2) domain. The regions falling in-between these domains are less conserved. Both domains are functionally important because the MH1 domain acts as an inhibitory domain in which the Smad2 protein folds upon itself in such a way that the MH1 domain blocks the MH2 domain from interacting with other molecules. Inhibition release occurs after an activated receptor kinase (e.g. TARAM-A) phosphorylates a receptor Smad at its carboxy end, releasing the MH2 domain. This allows the Smads to dimerize with co-Smad4. The remaining Smads are Smad6 and 7, which act as inhibitors of Smad signaling by blocking further transduction through binding with upstream receptors. After phosphorylation by a receptor kinase and heterodimerization with Smad4, receptor Smads are translocated to the nucleus to regulate gene transcription either independently or through recruitment of DNA binding factors. Smad2 and 3 do not have a high intrinsic DNA binding affinity. Therefore, they require co-recruitment with factors such as Smad4 and forkhead activin signal transducer 1 (Fast-1) (Müller, Blader, Rastegar, Fischer, Knöchel, Strähle, 1999). Co-Smad4 contains a nuclear localization signal (NLS) and nuclear export signal (NES). Smad4 enters the nucleus via importin-α and the Ran-GTPase pathway. Smad3 contains an NLS in its MH1 domain but Smad2 does not and enters the nucleus through a direct interaction between its MH2 domain and nucleoporins (Chen, *et al.*, 2006).

Specificity between receptor Smads that are related to TGFβ, Acitivin, and Nodal signaling versus those related to BMP signaling, is dependent upon structural similarities in the L45 loop of type I receptors. Type I receptors related to Smad2 and 3 have very closely related L45 loop structures that differ from the L45 loops of BMP type I receptors related to Smad1, 5 and 8. Smad2 and 3 also have a consensus L3 loop sequence that differs from Smad1, 5, and 8 (Chen, *et al.,* 2006). Studies of the ActRII

receptor also demonstrate cross-talk between the Activin and Nodal pathways. Northern blot analysis of poly $(A)^+$ RNA from ovarian and adult zebrafish tissues revealed that the highest level of *ActRIIa* transcripts were in the oocyte with levels decreasing up until 90% epiboly and rising again by the fourteen somite stage. *ActRIIb* transcripts were highest during sphere and decreased by 90% epiboly, also rising again by the fourteen somite stage. Expression is ubiquitous. Studies of overexpression of the type II receptors were done by synthetic mRNA injection into the blastomere of one or two cell stage embryos aiming for ubiquitous expression. These experiments resulted in embryos with phenotypes similar to when Cyclops and Squint are overexpressed, having an abnormal ventral body axis and development being heavily dorsalized. The dorsal mesoderm marker *goosecoid* (*gsc*) had expression expanded into ventral regions. Markers *gata3* (*gta3*) and *orthodenticle homolog 2* (*otx2*) for non-neural and neural anterior ectoderm respectively, were reduced in expression. The use of a dominant negative ActRIIb blocks both Activin and BMP signaling when injected in combination with Activin and BMP affinity beads. Loss of signaling was determined by a decrease in *no tail* (*ntl*) transcripts related to Activin signaling and a decrease in *gta3* expression. This suggests that the type II receptors do not have specificity between the Activin or BMP pathways and that such specificity arises from the interaction between a type II receptor and its ligand (Nagaso, Suzuki, Tada, Ueno, 1999).

Cofactors are often in involved in Nodal signaling. In zebrafish, a member of the epidermal growth factor (EGF) family, One-eyed pinhead (oep) has been identified as a co-factor of Nodal signaling. *Oep* mutants have phenotypes identical to *cyc* and *sqt* mutants, including lack of anterior axial mesoderm and endoderm, dorsal mesoderm, and defects in tail formation. This has been evidenced in mutants by observance of a lack of the prechordal plate (anterior axial mesoderm), and lack of *goosecoid* (*gsc*) expression in addition to tail abnormalities. *Oep* is expressed maternally and zygotically, similarly to TARAM-A (Schier, Talbot, 2001). Overexpression of the constitutively active form of TARAM-A is able to rescue *oep* mutants, providing further evidence for *oep* as a co-factor, as well as for TARAM-A as a Nodal related signal transducer (Peyriéras, *et al.,* 1998).

Two extracellular inhibitors of Nodal signaling are Lefty and Cerberus. Lefty1 and lefty2 (also known as Antivins) are closely related to TGFβ ligands. Overexpression of these genes generates phenotypes that resemble Cyclops and Squint homozygous mutants, and loss of lefty signaling seems to enhance mesoderm formation and left-right patterning. The mechanism by which lefty antagonizes Nodal signaling is yet to be elucidated. Lefty gene expression is dependent on Nodal signaling. Therefore, lefty functions in a negative feedback mechanism. *Xenopus* cysteine knot protein Cerberus binds to Xnr1-6 and blocks Nodal signaling. Injection of the truncated form Cer-S in zebrafish generates phenotypes similar to *cyc* and *sqt* mutants, such as lack of differentiation into mesoderm. An orthologue is yet to be found in zebrafish (Schier, 2003).



(Figure I: Key Components of Nodal Signaling)(Adapted from: Schier, 2003)

Fast-1 and Smad2/4 as Components of the Activin Response Element

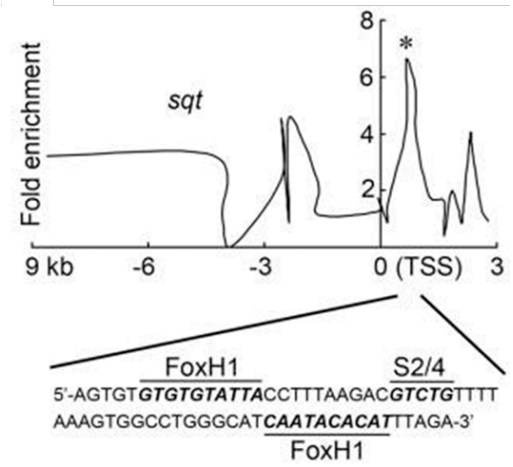
In order to be a downstream effector, Nodal target genes must have binding sites for various transcription factors, markedly Fast-1 and Smad2 (Fast-1 is also known as FoxH1). Due to cross-talk with the Activin pathway via use of TARAM-A and ActRII receptors, many Nodal-related genes contain an Activin Response Element (ARE) within their promoter regions. The main components of an ARE are binding sites for Smad2 and Fast-1. Upon binding of each molecule an Activin-Response Factor (ARF) has been created (Chen, Weisberg, Fridmacher, Watanabe, G.N. Whitman, M. Whitman, 1997). Fast-1 is an important transcription factor for cyc and sqt in zebrafish**.** Use of an antisense *fast-1* RNA probe in whole-mount *in situ* hybridization revealed that *fast-1* is expressed maternally and in similar regions to *oep* in zygotes. Zygotic and maternal mutants of Fast-1 also demonstrate phenotypes such as cyclopia, lack of floorplate, loss of posterior chordal plate, notochord deformities, and a bent ventral body axis. Once again, these phenotypes resemble loss of Nodal signaling, suggesting Fast-1 plays an important role in propagating cyc and sqt signals (Sirotkin, *et al.,* 2000). Fast-1 contains a Smad2 binding domain by which it binds Smad2 and helps Smad2 to regulate transcription at the gene level through its own DNA binding domain (Silvestri, Narimatsu, Both, Liu, Tan, Izzi, McCaffery, Wrana, Atisano, 2008). The role of Smad4 in the ARE was determined through a series of immunoprecipitation experiments. Smad2 and Smad4 were shown to complex together through an electrophoretic mobility shift assay (EMSA) supershift where the anti-tag antibodies resulted in a shift higher

than each of the molecules with their respective tag individually. Smad2 coprecipitated with Fast-1 with or without Activin stimulation, but Smad4 only coprecipitated under Activin activation. This suggests that Smad4 may not bind the ARE directly (Chen, *et al*., 1997).

Targeting the ARE is beneficial in obtaining a read-out that is related to Nodal signaling only. Genes targeted at the ARE will only lose function as those functions relate to Nodal, and will retain signaling capabilities with all other pathways. For example, in the case of *retinoic acid receptor b (rarab),* the receptor is responsive to both the Retinoic Acid and Nodal pathways. Upon binding a retinoid, rarab complexes with a retinoid-X receptor and forms an active dimer that binds a retinoic acid response element and regulates gene transcription. The Retinoic Acid pathway involves patterning of the central nervous system and paired appendages (Hale, Tallafuss, Yan, Dudley, Eisen, Postlethwait, 2006). Through this gene targeting strategy, if a mutation is sustained at the ARE, only the involvement of *rarab* in Nodal signaling will be affected while retaining its relationship to the Retinoic Acid pathway. The strategy presented targets genes that are under the regulation of an ARE, therefore genes within the Nodal pathway can still retain their responsiveness to pathways independent of Nodal. This facet of the construct design helps to eliminate Nodal-independent effects.

Gene Targeting Approach

Given the presented evidence, inducing mutations at the ARE where there are sites for both Smad2 and Fast-1 would be an effective way to generate a wide array of Nodal specific mutations. In fact, sites have been identified in the zebrafish genome where Fast-1 and Smad2 DNA binding sites exist in close proximity to one another. In one study, 556 Smad2/4 DNA binding regions were identified. These regions corresponded to 679 genes. Of these 556 Smad2/4 DNA binding regions, 30.17% (about 167 DNA sites) also contain a Fast-1 DNA binding site within 100 bp (Liu, Lin, Cai, Zhang, Han, Jia, Meng, Wang, 2011).



(Figure II: Smad2/4 DNA Binding Sites near a Fast-1(FoxH1) Binding Site as Evidence of an ARE)(Source: Liu, *et al.,* 2011).

Gene targeting in zebrafish is limited in the variety of techniques available to achieve full knock-out results. Morpholinos tend to only achieve knock-downs. Targeted knock-outs can be achieved through the use of tethered endonucleases, and thus is the inspiration for the technique used here (Händel, Cathomen, 2011).

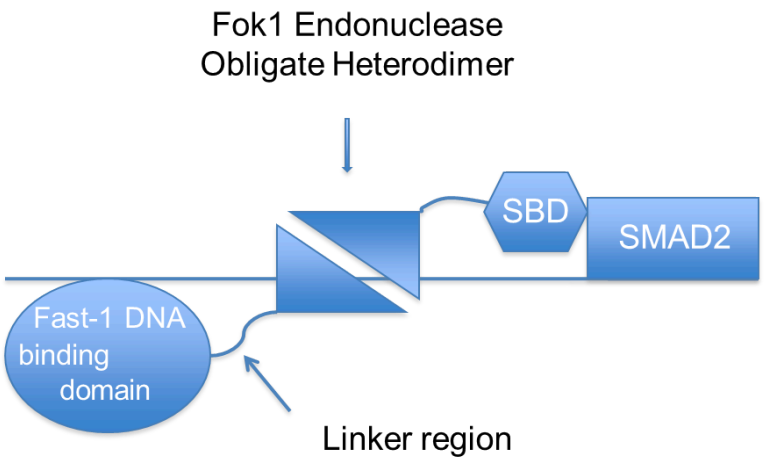
The use of zinc finger nucleases utilizes a construct consisting of zinc finger DNA binding elements connected by a linker region to a nonspecific nuclease. Zinc fingers are engineered $Cysteine_2$ – $Histadine_2$ residues that recognize about 3 bp of DNA and can be assembled modularly. The zinc finger domains are engineered to bind DNA on either side of the target region, with one set binding the sense, and the other binding the anti-sense strands. Attached to these DNA binding domains is a linker region made up of random nucleotides that are attached to a subunit of the nuclease. When the zinc fingers bind DNA, the two tethered nuclease subunits will bind and form a dimer. The nuclease is able to function only as a dimer. Therefore, the nuclease requires the binding and coming together of the two "halves" of the nuclease in order to create double strand DNA breaks. The DNA break will be generated in a spacer region that exists between the two regions where the zinc fingers bind on DNA. Once a break has been generated, the idea is that small insertions or deletions will be generated through non-homologous end joining (NHEJ), which is a faulty DNA repair system. A full zinc finger nuclease pair can recognize target sites of 18 to 36 bp. The exact number of zinc fingers to use, and the length of the linker region are questionable as to what is optimal, but also varies greatly with the genetic target (Händel, Cathomen, 2011).

A downfall to the use of zinc fingers is that complications can arise in achieving the desired specificity due to the highly modular nature of the zinc fingers. As a result, such constructs have the possibility of being highly toxic, and producing non-specific effects (Händel, Cathomen, 2011).

In our gene targeting strategy, the Fok1 endonuclease was used. In part as an attempt to bypass toxicity and non-specific effects, sequence encoding Fok1 was tethered to sequences encoding either full or partial endogenous ARE interacting proteins. The proteins of interest are Fast-1 and Smad2.

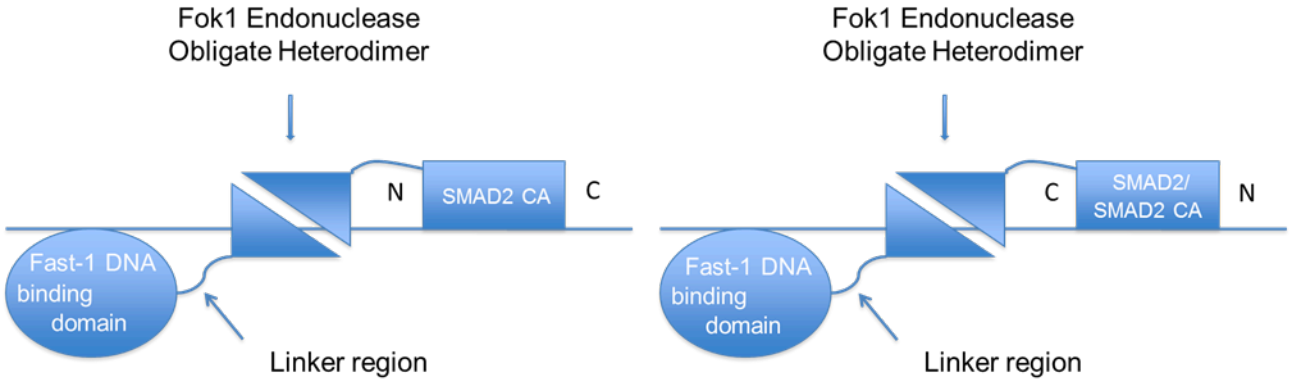
Fast-1 is proven to be an important transcription factor for Smad signaling via all the receptor Smads for the TGFβ pathway (Chen, *et al.,* 1997). This makes it a suitable candidate for use in construct design. Constructs were made tethering sequence encoding the DNA binding domain (DBD) of Fast-1 to Fok1 in a PCS2 vector containing one obligate heterodimeric subunit of Fok1 termed DD (Szcepek, Brondani, Büchel, Serrano, Segal, Cathomen, 2007). The N-terminus of Fok1 was linked the DBD by a nine or sixteen amino acid linker. Two linker lengths were used in order to test optimal spacing in the complex formed at the DNA binding site. mRNA from this construct was co-injected with a construct encoding the Smad2 binding domain (SBD) of Fast-1, having the other obligate heterodimeric subunit of Fok1 (termed RR) attached to the C-terminus of the SBD via a nine or sixteen amino acid linker as well (Meng, Noyes, Zhu, Lawson, Wolfe, 2008). In order for both the DD and RR subunits of the heterodimeric Fok1 complex to both be recruited to ARE, the DBD construct will need to bind DNA where Fast-1 would bind DNA and the SBD construct will need to bind Smad2 where Fast-1 would bind Smad2. Therefore, Smad2 needs to bind at the ARE via endogenous signaling to recruit Fok1 RR. Smad2 will bind DNA most likely through association with Smad4. Some endogenous Smad2 signaling may be blocked due to the lack of

association with Fast-1 in that the injected Fast-1 DBD will compete with endogenous Fast-1 for DNA binding. In addition the SBD will compete with Fast-1 for Smad2 binding. Constructs were injected with either like, or varying linkers for both DD and RR vectors.



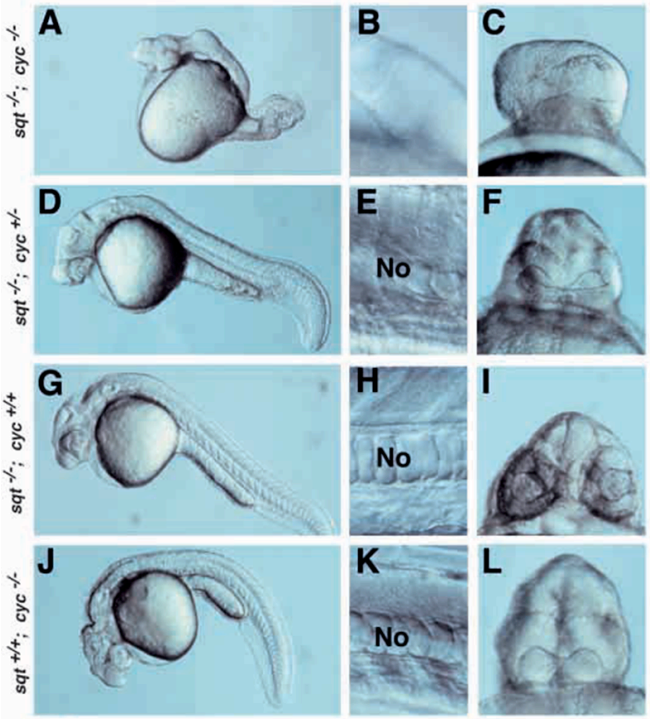
(Figure III: Strategy for Recruitment of Fok1 DD and Fok1 RR to the ARE)

Sequences encoding the full, and a constitutively active (CA) form of Smad2 were also inserted into the pCS2 Fok1 RR (inadvertently DD to be discussed later) vector to be injected with the Fast-1 DBD in pCS2 Fok1 DD. Another set of constructs were made inserting Fok1 RR (accidentally DD) into the pCS2MT vector containing sequence encoding a constitutively active form of Smad2. The intent here is to place Fok1 RR at the N-terminus of Smad2 due to evidence suggesting that attaching anything to the C-terminus of Smad2 may alter its function (Müller, *et al.,* 1999). This set of constructs was also co-injected with the Fast-1 DNA binding domain in a vector with the matching Fok1 subunit. All constructs were made with both a nine and sixteen amino acid linker and were injected in varying combinations. DNA from these constructs were linearized and synthesized into messenger RNA (mRNA), which was then injected into zebrafish embryos at the one cell stage.



(Figure IV: Placement of Fok1 N and C Terminal to Smad2 or Smad2 CA)

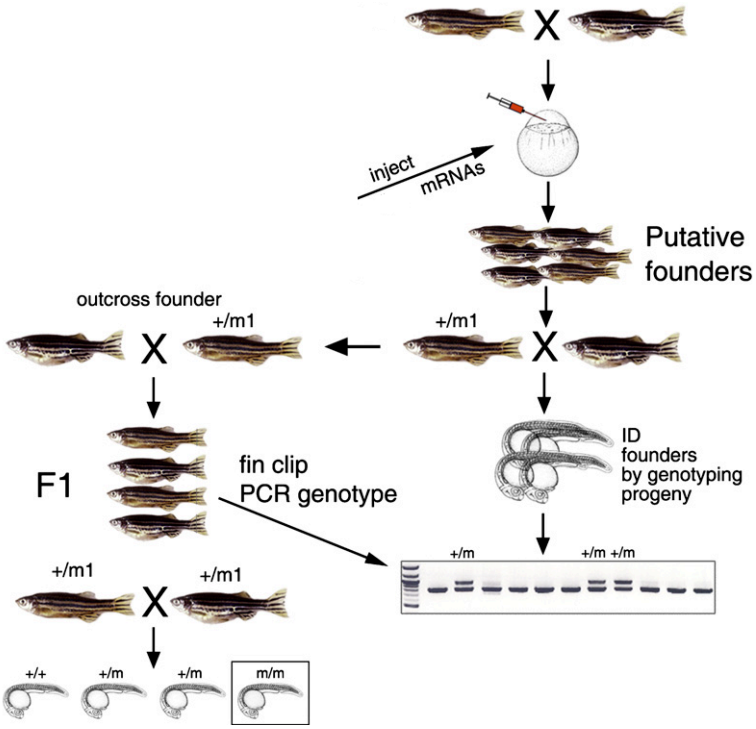
Due to the fact that an ARE is targeted, mutations will randomly be generated within the promoters of Nodal responsive genes. Not only are the selected genes random (within ARE responsive genes), but the mutations themselves are random as well. Given the faulty repair system of NHEJ, bases may be inserted or deleted resulting in a non-functional response element. With an altered genomic sequence the ARE would be unrecognizable to binding factors required for gene transcription. Mutations will be generated in genes that are part of the Nodal pathway. Some of the phenotypes that one would expect to see in zebrafish embryos that have been injected with the constructs would be those that resemble *cyc* and *sqt* mutants. As seen in Figure V, *cyc* and *sqt* mutants have deformed mesoderm and endodermal tissues demonstrated by deformed somites, or lack thereof. Increased blood formation in the tail is indicative of an abundance of mesodermal tissue. Missing or deformed notochords, as well as cyclopia are also mesoderm deformities. Additionally, a bent body axis and poor head and/or tail formation are seen in Nodal mutants (Dougan, Warga, Kane, Schier, Talbot, 2003). Injected embryos demonstrating many of these phenotypes could suggest disruptions in Nodal signaling. Therefore, the hope is that the amount of mRNA injected can be dialed down in such a manner so as to elicit minimal effects, generating zebrafish that are viable into sexually mature adults. Ideally fish that appear nearly wild type would have a desired mutation and would grow to viable adults. Any mutations generated in injected embryos should be heterozygous and should have a very low frequency of generating phenotypes.



(Figure V: Analysis of Genetic Interaction Between *cyc* and *sqt*)(Source: Dougan, *et al.,* 2003)

Assays

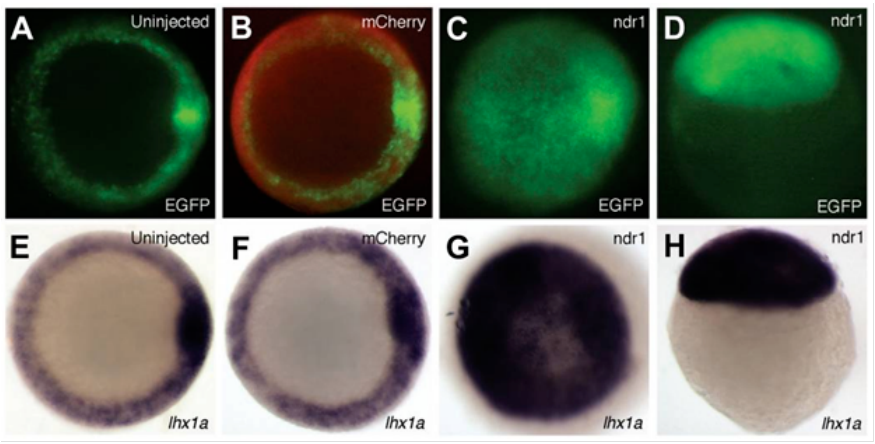
A select set of embryos from injected fish can be used for genotyping and fin clips can be performed on fish that mature to adulthood to confirm the presence of a mutation. It is highly likely that such adults will contain an array of various mutations if mutations are generated. Such fish will be mated to a wild type zebrafish, and one mutation will go germline and be passed on to the F1 generation of offspring. In this light, there is potential for one injected fish to generate offspring having varying mutations by passing on one mutation per offspring. The offspring will have mutations that can be characterized and will serve as a valuable resource in studying the downstream effects of the targeted gene upon loss of function. In order to generate homozygous knockouts, sibling heterozygous zebrafish can be mated, resulting in 25% homozygous in the F2 generation (Lawson, Wolfe, 2011).



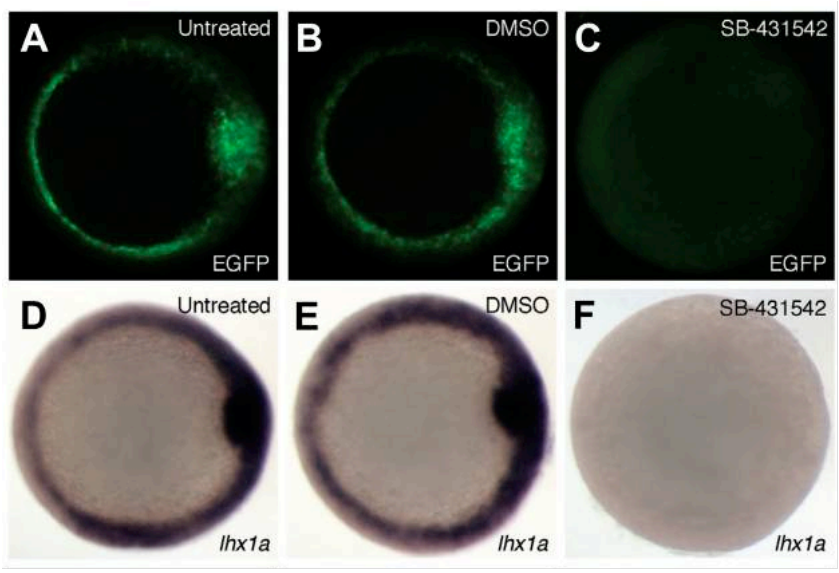
(Figure VI: Mating Schemes of Injected Zebrafish to Generate Germline Mutations)
(Adopted from: Lawson, Wolfe, 2011)

In addition to genotyping analysis, an *LIM homeobox 1a enhanced green fluorescent protein (lhx1a-EGFP)* transgenic zebrafish line was used to study the effectiveness of the knock-out constructs. This line expresses enhanced green fluorescent protein under the control of *lhx1a* regulatory regions and has the potential to be a more sensitive assay due to the visualization of individual cells. The *lhx1a* promoter contains an ARE and is a downstream nodal target as exemplified by Figures VII and VIII. Figure VII shows increased GFP expression by fluorescence microscopy and localization of *lhx1a* by in situ hybridization in response to injection with 25 pg *ndr1*

RNA. In contrast, Figure VIII demonstrates loss of EGFP expression by fluorescence microscopy as controlled by *lhx1a* due to inhibition of Nodal signaling via the small molecule inhibitor SB-431542. Evidence that the lack of GFP expression is due to *lhx1a* inhibition is seen in the in situ hybridization where no *lhx1a* expression is evident (Swanhart, Takahashi, Jackson, Gibson, Watkins, Dawid, Hukriede, 2010).



(Figure VII: *Lhx1a* Expression in Response to Overexpression of *Ndr1*)(Source: Swanhart, *et al*., 2010)



(Figure VIII: *Lhx1a* Responsiveness to Nodal Inhibition)(Source: Swanhart, *et al*., 2010)

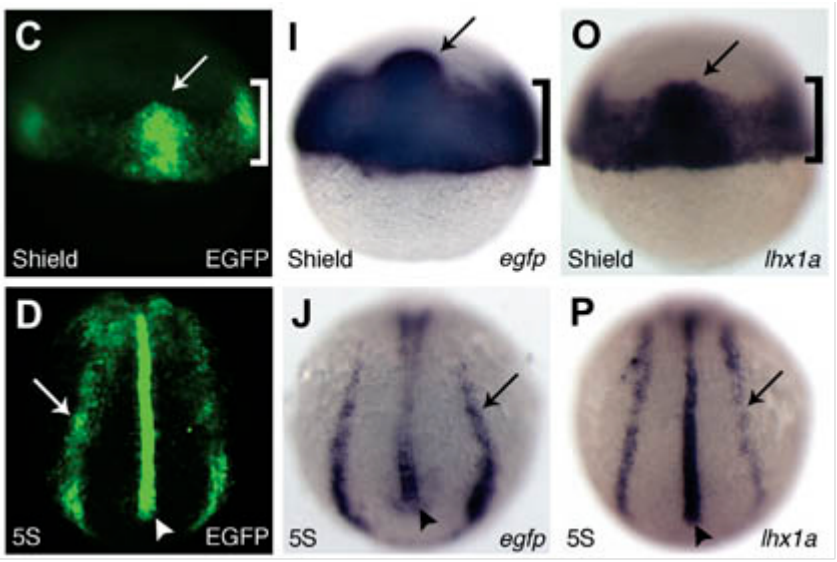
Materials and Methods

Husbandry

Tupfel long fin cross to Brian's wilt type (TLB) strains of zebrafish were used for mating to generate wild type embryos or to generate hemizygous fish when mated to fish with the *lhx1a-EGFP* transgene**.** Founder fish will also be outcrossed to TLB wild types.

Lhx1a-EGFP Assay

Lhx1a-EGFP transgenic fish were mated to TLB wild types to generate hemizygous fish that were injected with mRNA at the one cell stage. The purpose of this assay is to look via fluorescence microscopy for gaps in expression of EGFP and patterns that differ from dye injected fish. Observations were made in-between the 60-80% epiboly and shield stages due to an easily distinguishable wild type GFP expression pattern (Swanhart, *et al.,* 2010). Figure IX below shows the wild type fluorescence pattern of the lhx1a-EGFP transgene at shield and five somite stages. The goal is that the constructs will generate mutations in the ARE within the promoter of the *lhx1a* gene. The injected embryo would be mosaic for mutations since a mutation will not be generated in every cell, which is why one would expect changes in fluorescence instead of complete loss of fluorescence.

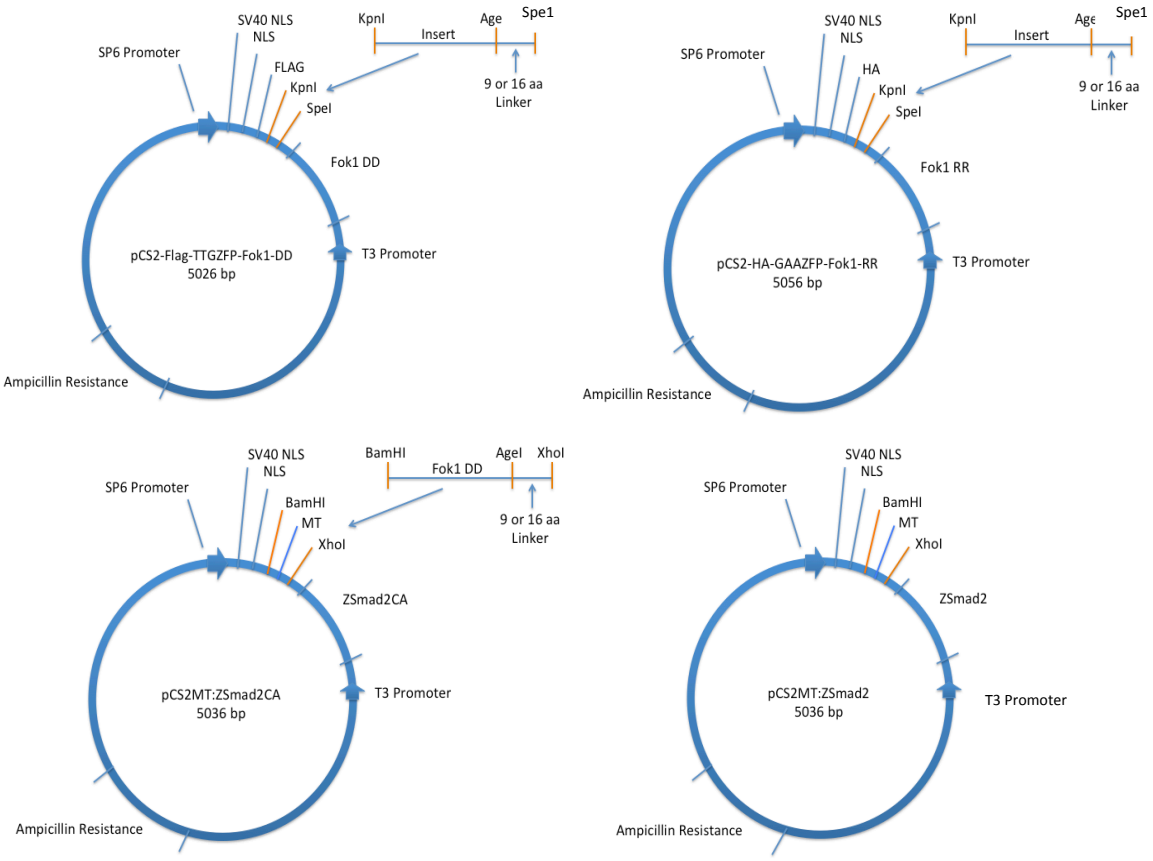


(Figure IX: *Lhx1a-EGFP* Expression Patterns at the Shield and Five Somite Stages)(Source: Swanhart, *et al.,* 2010)

Cloning

All cloning was done with a variant of the pCS2 plasmid containing ampicillin (amp) resistance. We obtained pCS2 plasmids with the Fok1 RR and Fok1 DD genes already cloned in from Scott Wolfe's lab via Addgene (Meng, *et al.,* 2008). We also obtained pCS2 Myc tagged plasmids with the Smad2 and Smad2 constitutively active (CA) genes cloned in from Uwe Strähle's lab (Müller, *et al.,* 1999). The majority of cloning was done by digesting PCR inserts in and out of the plasmids. PCR products were digested with Dpn1 to remove parental plasmid. Digested vectors and PCR products were gel purified and extracted using the MinElute Gel Extraction kit from Qiagen. Digests were followed by subsequent ligations. All restriction enzymes were from New England Biolabs and the T4 ligase used was from Invitrogen. The rAPid Alkaline Phosphatase kit distributed by Roche Applied Science was used to dephosphorylate the 5' ends of the vectors to prevent self-ligation of vector without insert. Insert DNAs and linkers were cloned into two restriction sites with a third Age1 site generated between the insert and the linker in order to be able to switch out linkers to obtain the desired constructs shown in Figure X. Competent DH5α cells from Invitrogen were used in transforming constructs in *E. coli* on LB/amp plates. Select colonies were grown in LB/amp nutrient broth and isolation of plasmid DNA was done with the use of the GeneJET Plasmid Miniprep kit from Thermo Scientific and the HiSpeed Plasmid Midi kit from Qiagen. Table I lists all of the clones generated with their respective insert cloning sites. All clones were cut with diagnostic enzymes to screen for integration of insert DNA. Positive clones were additionally sequenced and linearized with Not1 to make mRNA.

Some difficulty was faced in properly being able to ligate in the DNA binding domain of Fast-1 (DBD) PCR product. The pGEM-T Easy Vector System TA cloning kit from Promega was used to simply integrate the DBD into the pGEM-T Easy Vector in order to transform it and grow it up to digest and generate the proper ends for ligation into the desired vector. A HindIII diagnostic digest resulting in bands 510 bp, 750 bp and about 3 kb in size indicate proper DBD integration. The pGEM-T Easy Vector System protocol was followed as recommended and the supplied Platinum Taq was used for PCR amplification of the DBD to generate TA overhangs. Cycling conditions are listed in Table II.



(Figure X: Vector Maps)

Construct + Linker	Cloning Sites	Diagnostic Enzyme	Correct Fragment Sizes
DBD in pCS2:Fok1 DD 9aa	Kpn1/Spe1	HindIII	1.2 kb + ~ 5 kb vector
DBD in pCS2:Fok1 DD 16aa	Kpn1/Spe1	HindIII	1.2 kb + ~ 5 kb vector
SBD in pCS2:Fok1 DD 9aa	Kpn1/Spe1	HindIII	1.0 kb + ~ 5 kb vector
SBD in pCS2:Fok1 DD 16aa	Kpn1/Spe1	HindIII	1.0 kb + ~ 5 kb vector
Smad2 in pCS2:Fok1 DD 9aa	Kpn1/Spe1	HindIII	1.8 kb + ~ 5 kb vector
Smad2 in pCS2:Fok1 DD 16aa	Kpn1/Spe1	HindIII	1.8 kb + ~ 5 kb vector
Smad2 CA in pCS2:Fok1 DD 9aa	Kpn1/Spe1	HindIII	1.1 kb + ~ 5 kb vector
Smad2 CA in pCS2:Fok1 DD 16aa	Kpn1/Spe1	HindIII	1.1 kb + ~ 5 kb vector
Fok1 DD in pCS2MT:Smad2 CA (cloning removed MT)	Xho1/BamHI	HindIII	300 bp + ~ 5 kb vector
DBD in pCS2:Fok1 RR 9aa (Made to RR by SDM)	Kpn1/Spe1	HindIII	1.2 kb + ~ 5 kb vector
DBD in pCS2:Fok1 RR 16aa (Made to RR by SDM)	Kpn1/Spe1	HindIII	1.2 kb + ~ 5 kb vector
Fok1 RR Vector with no DNA Binding Domain	N/A	HindIII	715 bp + ~ 5 kb vector

SBD = Smad Binding Domain of Fast1
DBD = DNA Binding Domain of Fast1
CA = Constitutively Active
aa = Amino Acid Linker
SDM = Site Directed Mutagenesis

(Table I: Constructs from which mRNA was made)

Polymerase Chain Reaction

Insert DNAs for cloning purposes were amplified by polymerase chain reaction (PCR) using the high fidelity PrimeStar Taq polymerase obtained from Takara Bio. Table II lists insert DNA templates, amplification primers, and cycling conditions. Primers were designed to incorporate restriction sites and linkers into the PCR product. Linkers were incorporated into primers and their sequence translations are shown in Figure XI.

PCR Product	Template	Amplification Primers	Cycling Conditions	PCR Product Size
DBD	pCS2:ZFast-1	Zfast DBD F: 5′- CCG GTA CCA CAA AGC ACT GGG GAG GTC CAG GCT TG -3′ Zfast DBD R (16 aa Linker): 5′- TGA CTA GTT GCC CAG AGG CGC CAC TGC TAC CCG CCC TGG CGG CAG CAC CAG GAC CGG TTG AAC TGG CAG GCC TGA GGC TAT GCA GG -3′	98°C 30 s 98°C 10 s 62°C 10 s 72°C 1 min (29x) 72°C 5 min 4°C	792 bp
DBD amplified for TA cloning	pCS2:ZFast-1	Zfast DBD F: 5′- CCG GTA CCA CAA AGC ACT GGG GAG GTC CAG GCT TG -3′ Zfast DBD R (16 aa Linker): 5′- TGA CTA GTT GCC CAG AGG CGC CAC TGC TAC CCG CCC TGG CGG CAG CAC CAG GAC CGG TTG AAC TGG CAG GCC TGA GGC TAT GCA GG -3′	95°C 5 min 94°C 30 s 62°C 30 s 72°C 1 min (30x) 72°C 5 min 4°C	792 bp
SBD	pCS2:ZFast-1	Zfast SBD F: 5′- CCG GTA CCG CTG GGG AGG GGT TAC GGG AGA GAG -3′ Zfast SBD R (9aa Linker): 5′- TGA CTA GTT GCT TAG TAA CGC CTC TGT TTG GAC CGG TAA GAG AAT ATT TGC AAA GGG AAG GTC CGT TC -3′	98°C 30 s 98°C 10 s 62°C 10 s 72°C 1 min (29x) 72°C 5 min 4°C	623 bp
Smad2	pCS2MT:ZSmad2	Smad2 Kpn1 F: 5′- CAT GGT ACC ATG TCC TCC ATC TTG CCT TTC -3′ Smad2 Age1 R: 5′- GCA CCG GTG GAC ATA CTG GAG CAG CGT ACG -3′	98°C 30 s 98°C 10 s 55°C 10 s 72°C 1 min (29x) 72°C 5 min 4°C	1406 bp
Smad2 CA	pCS2MT:ZSmad2 CA	Smad2 CA Kpn1 F: 5′- CAT GGT ACC ATG GAC ACA GGT TCT CCT GC -3′ Smad2 Age1 R: 5′- GCA CCG GTG GAC ATA CTG GAG CAG CGT ACG -3′	98°C 30 s 98°C 10 s 55°C 10 s 72°C 1 min (29x) 72°C 5 min 4°C	681 bp
Fok1	pCS2:Fok1	Fok1 F: 5′- CGA GGA TCC GCT ATG GAG CAA CTA GTC AAA AGT GAA CTG -3′ Fok1 9aa R: 5′- GGC ACC TCG AGC TTA GTA ACG CCT CTG TTT GGA CCG GTA AAG TTT ATC TCG CCG TTA TTA AAT TTC CG -3′	98°C 30 s 98°C 10 s 56°C 30 s 72°C 1 min (27x) 72°C 5 min 4°C	626 bp
16 aa Linker (For Fok1 N terminal to Smad2 constructs)	Smad DNA Binding Domain in Fok1 DD 16 aa Linker	16 aa from SBD F: 5′- GAA CGG ACC TTC CCT TTG C -3′ 16aa with Xho1 R: 5′- CGG ATC TCG AGC CCA GAG GCG CCA CTG CTA CC -3′	98°C 30 s 98°C 10 s 60°C 30 s 72°C 30 s (26x) 72°C 5 min 4°C	100 bp

(Table II: Summary of Non-Genotyping PCR Products and Amplification Primers)

9 aa Linker

ACC GGT CCA AAC AGA GGC GTT ACT AAG

T G P N R G V T K

16 aa Linker

ACC GGT CCT GGT GCT GCC GCC AGG GCG GGT AGC AGT GGC GCC TCT GGG

T G P G A A A R A G S S G A S G

(Figure XI: Linker Translations)

Sequencing

To confirm proper integration of inserts and that no mutations were generated through the PCR amplification process, sequencing was performed using primers SP6, and T3 that are homologous to their respective promoter sites within the pCS2 vector. The Fok1 Reverse primer was also used. Template DNA ranged from 250-500 ng/ul and primers were added at a 3.2 pmol dilution. All sequencing was performed by Stony Brook University Office of Scientific Affairs DNA Sequencing Center and was analyzed using Sequencer 5.1 (GeneCodes).

Primer	**Sequence**
SP6	5′- TAC GAT TTA GGT GAC ACT ATA G -3′
Fok1 Reverse	5′- CGA AGT TCA GAT TTC TTC TCC TCC AG -3′
T3	5′- AAT TAA CCC TCA CTA AAG GG -3′

(Table III: Sequencing Primers)

RNA Synthesis

Messenger RNA was made from linearized template DNA with the Ambien mMESSAGE mMACHINE High Yield Capped mRNA Transcription kit, using the SP6 promoter and following the suggested protocol.

Site Directed Mutagenesis

Site directed Mutagenesis (SDM) was performed on some clones to convert Fok1 DD to Fok1 RR. Primers were designed to overlap the desired targeting region and to contain the base pairs encoding this change. The desired change will be inserted into the PCR product by incorporation of the change into the primer, from which the template plasmid is extended. LA taq, standing for long and accurate from Takara was used due to its proofreading ability and accuracy. Sequencing was performed across all coding regions to ensure no point mutations were introduced. The PCR product was transformed directly into DH5α competent cells from Invitrogen and DNA was isolated using the GeneJET Plasmid Miniprep kit from Thermo Scientific.

Desired Product	**Template**	**Amplification Primers**	**Cycling Conditions**	**PCR Product Size**
DBD in Fok1 RR with 9 and 16 aa linkers	DBD in Fok1 DD with 9 and 16 aa linkers	Fok1 RR SDM Forward: 5′- CTG CCA ATT GGC CAA GCA CGT GAA ATG CAA CGA TAT GTC GAA GAA AAT CAA AC -3′ Fok1 RR SDM Reverse: 5′- GTT TGA TTT TCT TCG ACA TAT CGT TGC ATT TCA CGT GCT TGG CCA ATT GGC AG -3′	94°C 1 min 95°C 30 s 55°C 1 min 68°C 5 min 72°C 10 min 4°C 18x	~ 5 kb

(Table IV: Summary of Site Directed Mutagenesis Conditions)

Injections

Injections were performed at the one cell stage after removal of the chorion using a microinjector with various concentrations of mRNA that ranges from 2 pg to 112.5 pg. Phenol red dye at a concentration of 2.0 mg/ml was also injected into embryos to control for handling and the injection process.

Injected embryos were kept in embryo media in an incubator kept between 26.5°C and 28.5°C. If genotyping a selection of injected embryos yielded the possibility of founder fish, the remaining fish were entered into the system when they reached five days old.

Analyses of the injections were performed at various time points to observe toxicity of the mRNA and/or success of the injection via light microscopy. Phenotypes and survival rates were noted. All injections and some injection analyses were performed by Andrew Taibi.

Photography

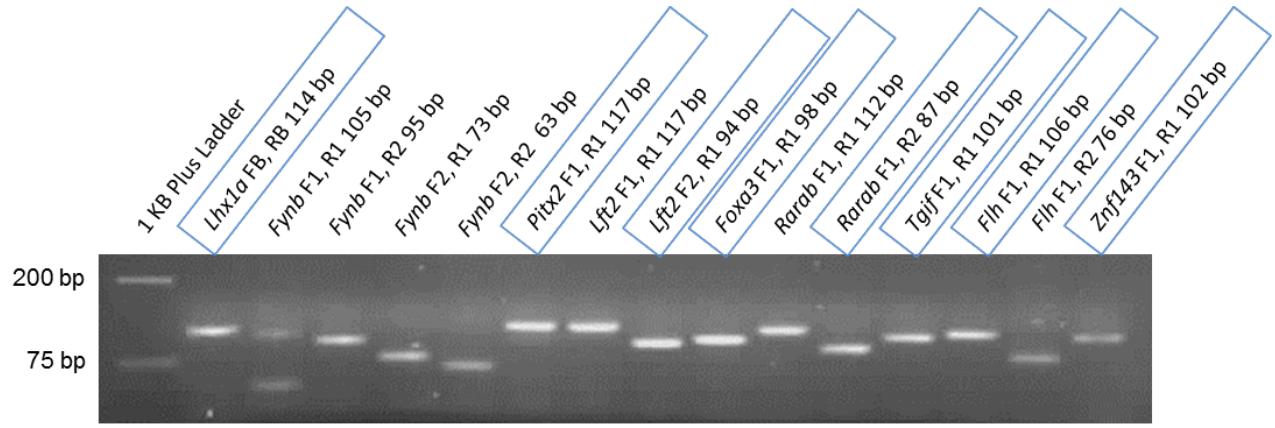
Interesting phenotypes were photographed using the Zeiss Axio Vision 3.1 System. Fish past 24 hours of growth were anesthetized in 0.4% tricaine, and photographed in 3.0% methyl cellulose to create a viscous solution limiting movement.

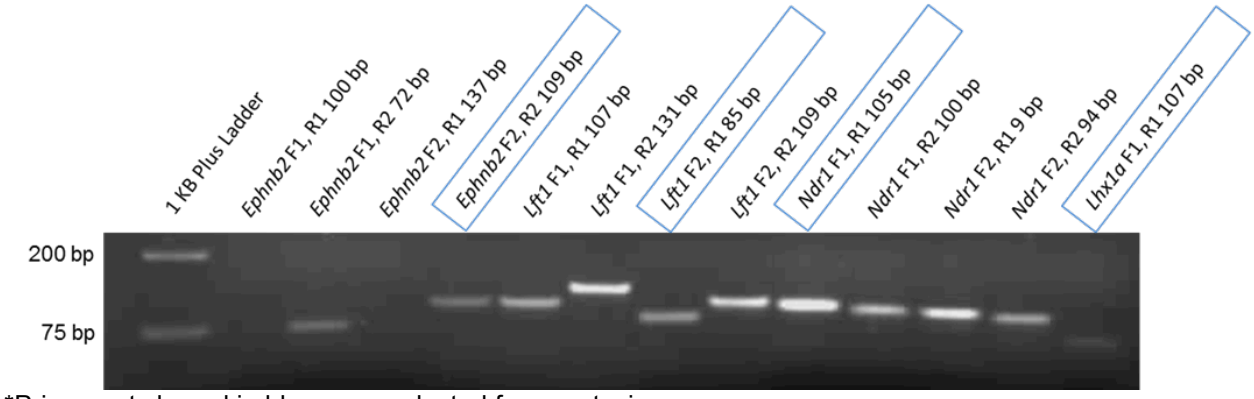
Mutation Analysis

A study was performed by Dr. Wang's lab at Tsinghua University in Beijing, China identifying sites in the zebrafish genome (Zv7) where both Fast1 and Smad2 bind within close proximity to one another (Liu, *et al.,* 2011). From this list we selected twelve annotated genes known to be downstream targets of Nodal to use in a preliminary screen for mutations. Table V below lists the genes selected. Primers were generated flanking the ARE with the aim to include at least one Smad2 and Fast-1 DNA binding site, with a PCR product about 100 bp or smaller. Primers were designed either by hand, with the help of the IDT Oligo Analyzer, or with the use of the Primer3 software created by Rozen and Skaletsky against Zv7. Primers were run through a BLAST search against the zebrafish genome (Zv9) for quality control. Figure XII shows the intensity of the PCR products of all primer sets on wild type DNA. The PCR product needs to be small in order to be able to detect two to three bp insertions or deletions by gel electrophoresis using 2.5 – 3.0% agarose gels. Any mutant bands would be excised, gel extracted, and sequenced. Embryos were prepped for genotyping no later than 24 hours after development based on phenotypes. Some embryos were prepped based on interesting *lhx1a-EGFP* expression, phenotypes, or at random. Embryos were prepped by heating at 98°C in 100 μl lysis buffer for 10 minutes, followed by the addition of 10 μl (10 mg/ml) proteinase K (PK) left at 55°C overnight. PK was inactivated by an additional 10 minute incubation at 98°C. PCR was performed on either individual embryos or embryo pools, diluting DNA either 1:5 or 1:10 in sterile RO water. EconoTaq from Lucigen was used. Genotyping PCR cycling conditions consisted of a 96°C hot start for 2 minutes, 96°C melting for 30 seconds (s), 55°C annealing for 30 s, 72°C extension for 30 s with a final extension of 72°C for 5 minutes after 44 cycles of melting through extension steps.

Gene with ARE and Amplification Primers	Distances (bp) of Two Adjacent Smad2 DNA Binding Sites Relative to a Fast1 Binding Site at the ARE	(Full Name) Role
Efnb2b F1: 5′- TTT GAC ACC CCT TGT TTT GA -3′ Efnb2b R1: 5′- GGG ATA ATT GAA TGT CGA GGT G -3′ Efnb2b F2: 5′- CCT AAT TCT TTC AAT CTT CAA CCT -3′ Efnb2b R2: 5′- TGT GAT GCG TTT GCA CTT -3′	10, 2nd distance NA	(Ephrin Receptor 2) Involved in Cell Adhesions
Flh F1: 5′- GAT GAT AAA GTG TGA GAC AAG AGT -3′ Flh R1: 5′- GAA AGA GCA CAA TCA CTA CTG C -3′ Flh R2: 5′- CAG AGA CAG TCA GAG AGC GCC -3′	0, 149	(Floating Head) Axial Mesoderm Morphogenesis
Foxa3 F1: 5′- GAT GCG CGA CTG TAG TAT G -3′ Foxa3 R1: 5′- CAT CTC CAT TCA TTA GCC ATT AGG -3′	12, 53	(Forkhead Box A3) Axial Mesoderm Morphogenesis
Fynb F1: 5′- GCA TAC AAC AAC AAA CAC CAC AG -3′ Fynb R1: 5′- CTA TAC ATG CCC TCA TTC TCC AC -3′ Fynb F2: 5′- CAT TAG GCC AAA TCC CGT CAC -3′ Fynb R2: 5′- CCT CAT TCT CCA CTG ATA TAA ACG -3′	6, 153	(FYN Oncogene Related to SRC, FGR, YES b) Gastrulation
Lft1 F1: 5′- TTC TTC CTG TGA GCT CCT TCA -3′ Lft1 R1: 5′- CAT TAG AAG AGG GGC TAG TGG -3′ Lft1 F2: 5′- TCT GTC TGA CAT CTT AGA GCA AAT -3′ Lft1 R2: 5′- CCA TGC CTT GTT TGG TCA C -3′	3, 416	(Lefty 1) Dorsal/Ventral Patterning
Lft2 F1: 5′- CCT TTG GAT CCT GTG TGG CG -3′ Lft2 R1: 5′- CTA AAT CCC AAA AGA AAG CAC -3′ Lft2 F2: 5′- CGC ACT CTT TTG GCC ATT TC -3′	25, 74	(Lefty 2)Heart Development
Lhx1a F1: 5′- AAC GAC TCT GTG ATG GAG ACC -3′ Lhx1a R1: 5′- TTC TCA CCA CTT AAT TAT GAC CAA -3′ Lhx1a FB: 5′- GTG ATG GAG ACC CTC AGA C -3′ Lhx1a RB: 5′- CAC CAT TGC AGT AAA TTC TCA CC -3′	29, 44	(LIM Homeobox 1a) Neural and Kidney Development
Ndr1 F1: 5′- ACG TTC GGG ACA AGC AAT TA -3′ Ndr1 R1: 5′- GTG TGT GTG TAT TAC CTT TAA GAC G -3′ Ndr1 F2: 5′- GGG ACA AGC AAT TAA TAA AGT AGA A -3′ Ndr1 R2: 5′- TGT GTA TTA CCT TTA AGA CGT CTG TTT -3′	11, 2nd site NA	(Nodal-Related 1) Dorsal Identity
Pitx2 F1: 5′- GTG TGT CTC CGC CTG CTT AG -3′ Pitx2 R1: 5′- GTC ACT TTT TAT CAG TCA CCC GG -3′	0, 14	(Paired-Like Homeodomain Transcription Factor 2) Digestive Tract and Left/Right Symmetry
Rarab F1: 5′- GTG CCA CAT GCT ATT AAC ACT GAC -3′ Rarab R1: 5′- GGC TCG CTC TGT CAT ATA C -3′ Rarab R2: 5′- ATG GCT GGA GTT GAA AAG CTG -3′	7, 25	(Retinoic Acid Receptor, Alpha B) Brain and Pectoral Fin Development
Tgif1 F1: 5′- CAC ATT GCT CAT ACA TCA GTC -3′ Tgif1 R1: 5′- CAT AAG AGA CCA GAG CTA AAA CG -3′	44, 206	(TGFB-Induced Factor Homeobox 1) Anterior/Posterior Patterning
Znf143 F1: 5′- TTA CCC AGT GTG CCT TAT TT -3′ Znf143 R1: 5′- ACG AGT TTA TTT CGT TCT GC -3′	45, 58	(Zinc Finger Protein 143a) Gene transcription

(Table V: Select Genes for Genotyping by PCR)(Roles Source: ZFIN.org)





*Primer sets boxed in blue were selected for genotyping.

(Figure XII: Primer Set Test on Wild Type DNA)

Results

Messenger RNA for all of the constructs listed in Table I were made and injected singly or in various combinations and concentrations in order to determine toxicity of individual constructs and effectiveness of combinations. All injections were performed by Andrew Taibi. Table VI below summarizes all injections performed.

Date	Construct 1	Construct 2	Concentration
6/14/12*	DBD-DD 16aa	SBD-RR 16aa (really DD)	10 pg both
			80 pg both
6/19/12*	DBD-DD 16aa	SBD-RR 16aa (really DD)	40 pg both
			80 pg both
6/21/12*	DBD-DD 9aa	None	40 pg
	SBD-RR 9aa (really DD)	None	40 pg
		DBD-DD 9aa	40 and 80 pg both
7/05/12*	DBD-DD 16aa	None	40 pg
	SBD-RR 16aa (really DD)	None	40 pg
		DBD-DD 16aa	40 and 80 pg both
7/10/12	SBD-RR 9aa (really DD)	None	40 pg
	SBD-RR 16aa (really DD)	None	40 pg
7/11/12	DBD-DD 9aa	None	20 and 40 pg
	DBD-DD 16aa	None	20 and 40 pg
7/17/12	DBD-DD 16aa	SBD-RR 16aa (really DD)	60 pg DBD and 10 pg SBD
8/15/12	Fok1 RR vector w/o insert (Vector mix-up, identity unknown)	None	100 pg
	DBD-DD 16aa	Unknown vector	40 pg DBD and 100 pg RR
8/16/12	Fok1 RR vector w/o insert (Vector mix-up, identity unknown)	None	20 pg and 40 pg
	DBD-DD 16aa	Unknown vector	40 pg both and 40 pg DBD with 20 pg RR
9/18/12	Smad2 CA-RR 9aa (really DD)	None	10, 40, and 80 pg
	Smad2 CA-RR 16aa (really DD)		10, 40, and 80 pg
9/19/12	DBD-DD 16aa	None	37.8 pg
	Smad2 CA-RR 16aa (really DD)	None	76 pg and 112.5 pg
		DBD-DD 16aa	76 pg Smad2 CA and 37.8 pg DBD
9/26/12	DBD-DD 9aa	None	30 pg
	Smad2 CA-RR 9aa (really DD)	None	80 pg and 112.5 pg
		DBD-DD 9aa	80 pg Smad2 CA and 30 pg DBD
10/23/12	Fok1 RR vector w/o insert	None	30 and 80 pg
	DBD-DD 9aa	Fok1 RR vector w/o insert	30 pg both
10/24/12	Fok1 RR vector w/o insert	None	2 and 10 pg
	DBD-DD 9aa	None	30 pg
		Fok1 RR vector w/o insert	10 pg Fok1 RR and 30 pg DBD
11/06/12	DBD-RR 9aa	None	10 and 40 pg
	DBD-RR 16aa	None	10 and 40 pg
	Fok1 DD into pCS2 Smad2 CA 9aa	None	40 pg
		DBD-RR 9aa	40 pg both
		DBD-RR 16aa	40 pg both
11/07112	DBD-RR 9aa	None	10 pg
	DBD-RR 16aa	None	10 pg
	Fok1 DD into pCS2 Smad2 CA 9aa	None	20 pg
		DBD-RR 9aa	20 pg Fok1 and 10 pg DBD
		DBD-RR 16aa	20 pg Fok1 and 10 pg DBD
11/08/12	Smad2 CA-DD 9aa	None	40 pg
		DBD-RR 9aa	40 pg Smad2 CA and 10 pg DBD
	Smad2 CA-DD 16aa	None	40 pg
		DBD-RR 9aa	40 pg Smad2 CA and 10 pg DBD

*Constructs injected into *Lhx1a-EGP* X TLB WT embryos.
aa= amino acid linker

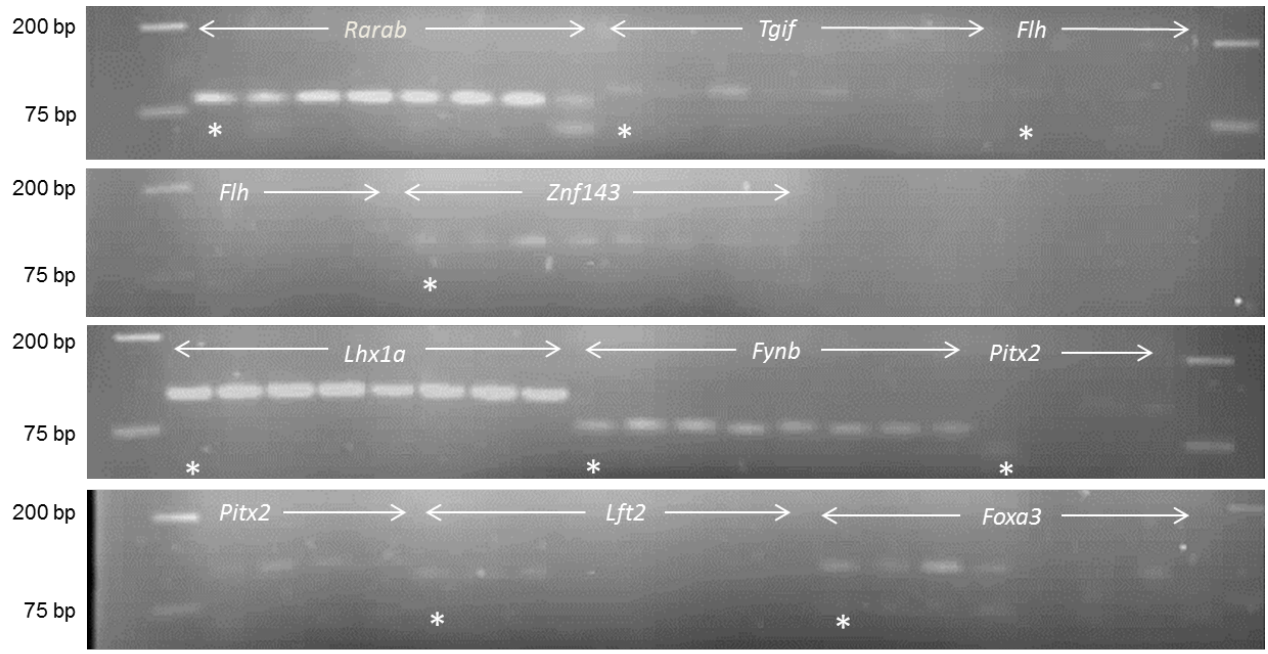
(Table VI: Summary of Injections)

Fast-1 DNA Binding Domain (DBD) + Fast-1 Smad2 Binding Domain (SBD) Injections

The first several injections were of the first construct pair made into mRNA using fish hemizygous for the *lhx1a-EGFP* transgene. This involved injection of the DBD in Fok1 DD 16aa linker and the SBD in Fok1 RR 16 aa linker (inadvertently DD). The SBD + DBD pair was injected at concentrations ranging from 10 to 80 pg in order to test for the optimal concentration. Observations of the conditions of the embryos injected are noted below and photographs of EGFP expression are presented in Figures XIV and XVI. Photographs were taken by Andrew Taibi. Select embryos were prepped and used for the PCR mutation analysis to detect mutations generated in the ARE of selected genes (Figures XIII and XV). Table VII summarizes the survival rates of injected embryos at 24 hours post fertilization (hpf).

Observations from 6-14-12 injection of DBD + SBD, both in Fok1 DD 16aa linker vector:

- EGFP expression was slight in all, including dye injected controls at 60-80% epiboly.
- Embryos injected with 10 pg of both constructs had wild type EGFP expression at 60-80% epiboly.
- Embryos injected with 80 pg of both constructs had some visual differences in EGFP expression at 60-80% epiboly.
- At 24 hpf some embryos had a bent body axis and enlarged yolk. A subset of embryos was prepped to PCR.
- Embryos were sorted for EGFP fluorescence and were put into the system as follows:
    - SB 3002: 31 GFP + of DOB 6-14-12 injected with 10 pg each of DBD and SBD both in Fok1 DD 16aa linker.
        ♂ SB 2858 *Lhx1a-EGFP* PT 303 Het
        ♀ TLB WT
    - SB 3003: 2 GFP + of DOB 6-14-12 injected with 80 pg each of DBD and SBD both inFok1 DD 16aa linker.
        ♂ SB 2858 *Lhx1a-EGFP* PT 303 Het
        ♀ TLB WT
    - SB 3004: 4 GFP – of DOB 6-14-12 injected with 10 pg each of DBD and SBD both in Fok1 DD 16aa linker.
        ♂ SB 2858 *Lhx1a-EGFP* PT 303 Het
        ♀ TLB WT



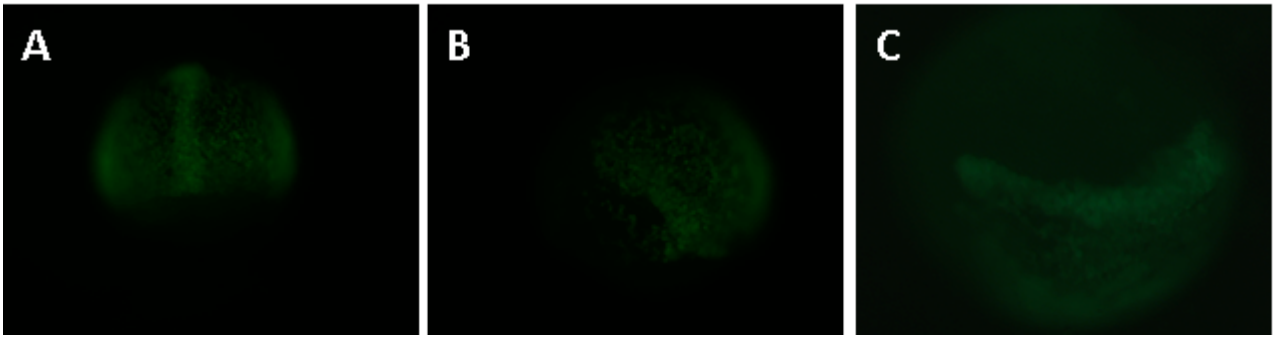
*Indicates Dye Injected Control

(Figure XIII: Sampling of PCR Results of Embryos Prepped from 6-14-12 Injection of DBD + SBD Both in Fok1 DD 16aa Linker Vector)

A sampling of PCR results are depicted for each primer set used at this point in time. PCR was done with a 1:10 dilution of DNA. Embryos were prepped at 60-80% epiboly and 24 hpf and were collected only from injections where both DBD-DD and SBD-DD (SBD was intended to be in RR but was actually DD) were injected. Selected embryos were chosen based on suggestive abnormalities in *Lhx1a-EGFP* expression at 60-80% epiboly, as well as from embryos whose expression pattern appeared wild type. Embryos prepped at 24 hpf were also chosen based on appearance being close to wild type and due to morphological abnormalities. In addition, DNA was harvested from embryos only injected with dye. No bands indicating small insertions or deletions were observed across all samples in the PCR reaction.

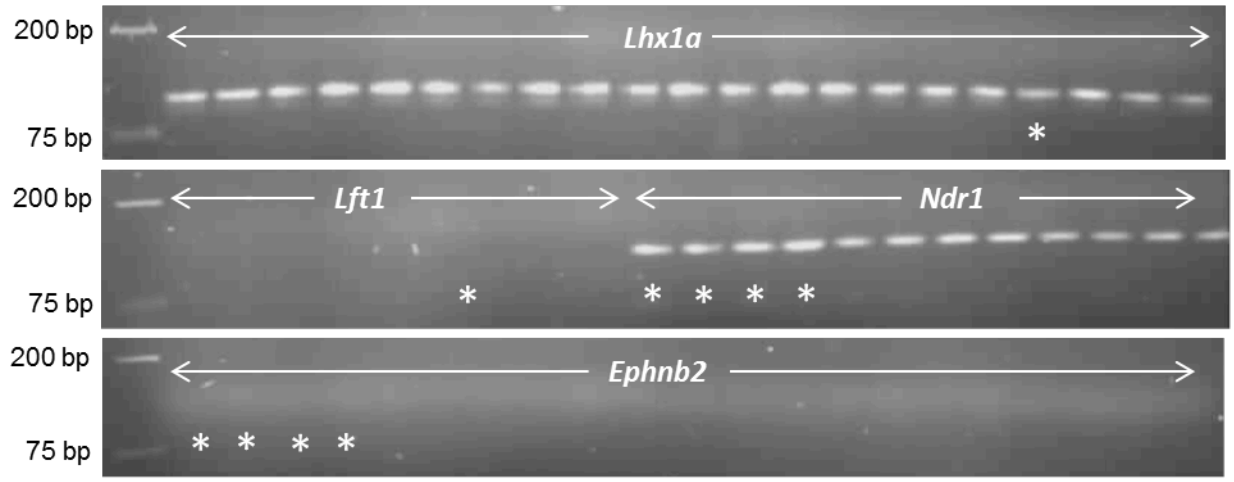
Observations from 6-19-12 injection of DBD + SBD both in Fok1 DD 16aa linker vector:

- Many embryos injected with both 40 and 80 pg were mangled with morphological defects but no changes in GFP fluorescence at 24 hpf.



(Figure XIV: Lhx1a Assay of 6-19-12 Injection of DBD + SBD Both in Fok1 DD 16aa Linker Vector)

A) Dye injected control embryo A under GFP fluorescence at 60-80% epiboly. B) Lhx1a-EGFP fluorescence of embryo B injected with 40 pg each of SBD + DBD both in Fok1 DD with the 16 amino acid linker at 60-80% epiboly (SBD was intended to be in RR, but was actually DD). C) Lhx1a-EGFP fluorescence of embryo B injected with 80 pg each of SBD + DBD both in Fok1 DD with the 16 amino acid linker at 60-80% epiboly (SBD was intended to be in RR, but was actually DD). Photographs by Andrew Taibi.



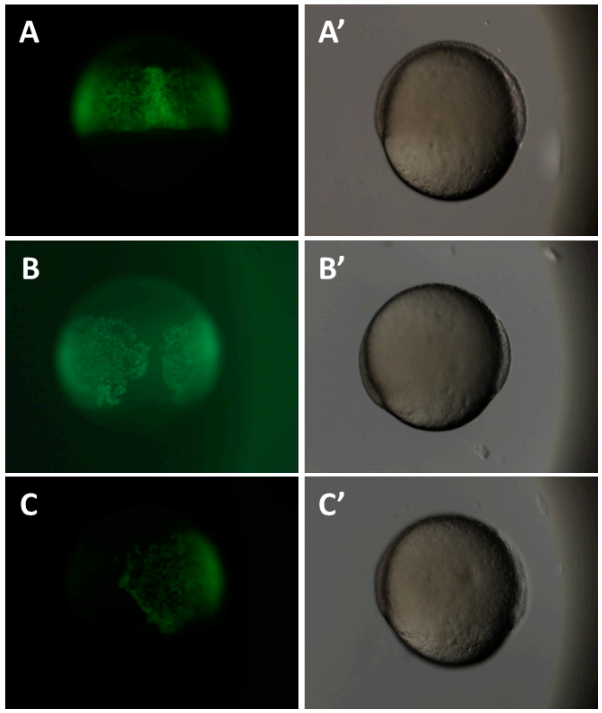
* Indicates Dye Injected Control

(Figure XV: Sampling of PCR Results of Embryos Prepped from June 14, 19, and 21st Injections)

A sampling of PCR results are depicted for each primer set used at this point in time. PCR was initially done with a 1:5 dilution of DNA (data not shown) and was repeated with a 1:10 dilution of DNA shown here due to a lack of intensity or presence of wild type bands. Embryos were prepped at 60-80% epiboly and 24 hpf and were collected only from injections where both DBD-DD and SBD-DD (SBD was intended to be in RR but was actually DD) were injected. Selected embryos were chosen based on suggestive abnormalities in *Lhx1a-EGFP* expression at 60-80% epiboly, as well as from embryos whose expression pattern appeared wild type. Embryos prepped at 24 hpf were also chosen based on appearance being close to wild type, as well as morphological abnormalities. In addition, DNA was harvested from embryos only injected with dye. No bands indicating small insertions or deletions were observed across all samples in the PCR reaction.

Observations from 7-05-12 injection of DBD + SBD both in Fok1 DD vector with the 16aa linker:

- In embryos injected with either 40 or 80 pg, the observed phenotypes were poor head development, large yolk, small overall size and bent body axis.
- Fish were not separated by GFP presence and were entered into the system as follows:
  - SB 3027: 10 of DOB 7-05-12 injected with 40 pg each of DBD and SBD both in Fok1 DD 16aa linker.
    - ♂ SB 2888 *Lhx1a-EGFP* PT 303 Het
    - ♀ SB 2738 TLB WT
  - SB 3028: 3 of DOB 7-05-12 injected with 80 pg each of DBD and SBD both in Fok1 DD 16aa linker.
    - ♂ SB 2888 *Lhx1a-EGFP* PT 303 Het
    - ♀ SB 2738 TLB WT



(Figure XVI: Lhx1a Assay of 7-05-12 Injection of DBD + SBD Both in Fok1 DD Vector with 16aa Linker)

A) Dye injected control embryo A under GFP fluorescence at 60-80% epiboly. A') Image of dye injected control embryo A from a differential interference contrast (DIC) microscope. B) Lhx1a-EGFP fluorescence of embryo B injected with 40 pg each of SBD + DBD both in Fok1 DD with the 16 amino acid linker at 60-80% epiboly (SBD was intended to be in RR, but was actually DD). B') DIC image of embryo B. C) Lhx1a-EGFP fluorescence embryo C injected with 40 pg each of SBD + DBD both in Fok1 DD with the 16 amino acid linker at 60-80% epiboly (SBD was intended to be in RR, but was actually DD). C') DIC image of embryo C. Photographs by Andrew Taibi.

Construct	Total	Alive	Dead	Survival (%)
DIC GFP-	3	3	0	100
DIC GFP+	13	13	0	100
DBD-DD 16aa 40 pg only GFP -	4	2	2	50
DBD-DD 16 aa 40 pg only GFP+	10	9	1	90
SBD-RR 16aa (accidentally DD) 40 pg only GFP-	7	3	4	43
SBD-RR 16aa 40 pg (accidentally DD) only GFP+	14	14	0	100
40 pg both DBD-DD 16 aa and SBD-DD 16aa GFP-	21	12	9	57
40 pg both DBD-DD 16aa and SBD-DD 16aa GFP+	37	31	6	86
80 pg both DBD-DD and SBD-DD GFP-	8	4	4	50
80 pg both DBD-DD 16aa and SBD-DD 16aa GFP+	16	12	4	75
R2 DIC GFP-	2	2	0	100
R2 DIC GFP+	13	13	0	100
R2 40 pg both DBD-DD 16aa and SBD-DD 16aa GFP-	14	14	0	100
R2 40 pg both DBD-DD 16aa and SBD-DD 16aa GFP+	40	34	6	85

DIC = Dye Injected Control
R2 = Round 2 Injections

This table sorts results of embryos by presence of *lhx1a-EGFP* transgene.

(Table VII: Survival Rates of 7-05-12 Injection of DBD + SBD Both in Fok1 DD Vector with 16aa Linker at 24 Hours Post Fertilization)

Constructs were made using both a 16aa and a 9aa linker in order to test for the optimal spacing of the tethered endonuclease at the ARE site. Observations are listed below for the DBD + SBD injection where both inserts are placed in the Fok1 DD vector

with the 9aa linker. Table VIII reports the survival of embryos injected with these constructs at 24 hpf. Select embryos from this injection were prepped and used for PCR analysis (Figure XV).

Observations from 6-21-12 injection of DBD + SBD both in Fok1 DD 9aa linker vector:

- Some embryos injected with both 40 and 80 pg of both constructs had poor head formation, enlarged yolks, and bent body axis at 24 hpf. No changes in EGFP fluorescence were seen at 60-80% epiboly.

Construct	Total	Alive	Dead	Survival (%)
DIC	20	19	1	95
SBD-RR (accidentally DD) 9aa only 40 pg	33	17	16	51
DBD-DD 9aa only 40 pg	24	20	4	83
40 pg both SBD-DD 9aa and DBD-DD 9aa	53	36	17	68
80 pg both SBD-DD 9aa and DBD-DD 9aa	64	45	19	70

DIC = Dye Injected Control

(Table VIII: Survival Rates of 6-21-12 Injection of DBD + SBD Both in Fok1 DD Vector with 9aa Linker at 24 Hours Post Fertilization)

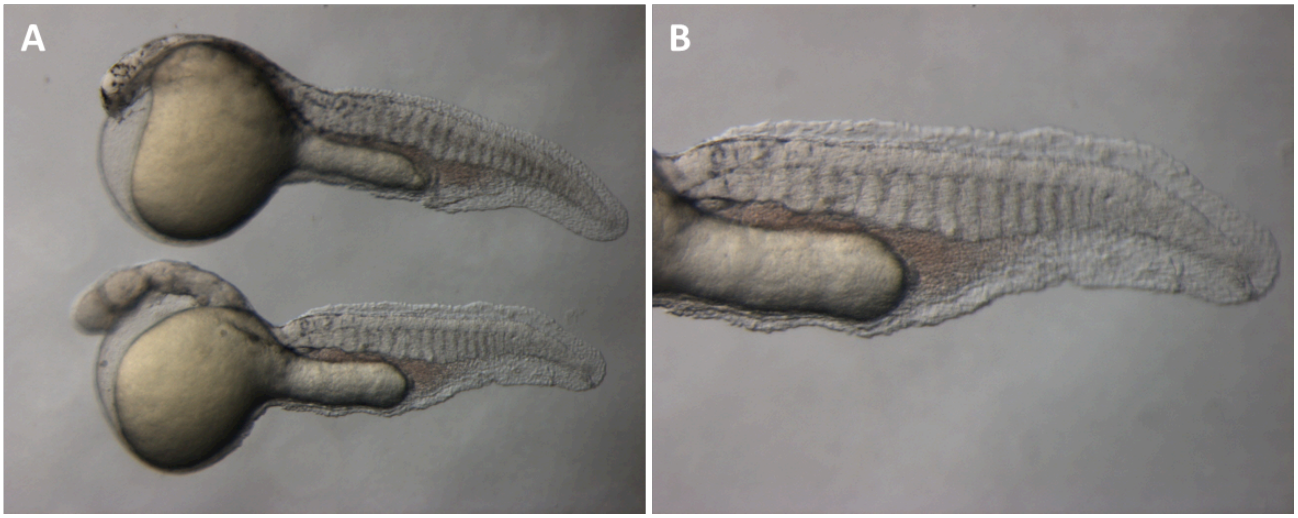
Single Construct Injections of Fast-1 DNA Binding Domain (DBD) and Fast-1 Smad2 Binding Domain (SBD) Containing Constructs

At this point in time the observation was made that when single constructs were injected by themselves phenotypes were fairly pronounced and survival rates at 24 hpf were low. The concern is that the results of DBD + SBD injections may not be a result of effective gene targeting, but due to off-target or toxic effects of each construct by itself. Injections were done using wild type embryos and both the DBD and SBD constructs by themselves in order to control for the DBD + SBD injections and to obtain a better idea of what is occurring with each construct individually using a larger sample size. Tables IX and X summarize the phenotypes observed and survival rates of injected embryos at 24 hpf. Figures XVII and XVIII are of fish representing the observed phenotypes.

Construct Injected	Total	Dead/ Survival (%)	Appear Wild Type/ (%)	Observed Phenotypes/ (% Occurrence from Total)
DIC	16	0/ (100)	16/ (100)	Appear Wild Type
SBD-RR 9aa 40 pg (accidentally DD)	42	12/ (71)	4/ (10)	16 pinhead (38) 10 head deformities (24) 4 deformed somites (10)
SBD-RR 16aa 40 pg (accidentally DD)	41	7/ (83)	11/ (27)	15 pinhead (37) 5 head deformities (12) 3 no head formation (7) 2 no tail formation (5) 8 deformed somites (20)

DIC= Dye Injected Control
Phenotypes are not mutually exclusive.

(Table IX: Survival Rates of 7-10-12 SBD Only Injection at 24 Hours Post Fertilization)



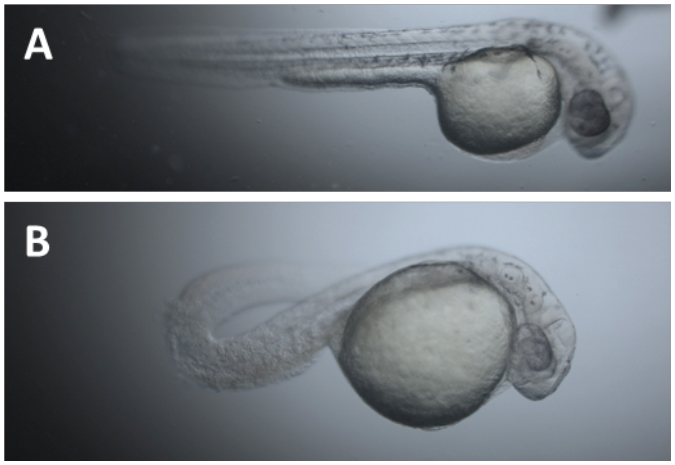
(Figure XVII: Phenotype of Representative Zebrafish from 7-10-12 SBD Only Injection)

A) DIC image of representative zebrafish injected with SBD in Fok1 DD (SBD was intended to be in RR, but was actually DD) at 24 hpf. These embryos demonstrate observed phenotypes such as pinhead, missing head (not shown), trunk deletions, and posterior deformities. B) Close up of tail of bottom fish seen in figure A. This is representative of a ventralizing phenotype marked by somite deformities, loss of notochord and increased blood levels. Photographs by Andrew Taibi.

Construct Injected	Total	Dead/ Survival (%)	Appear Wild Type/ (%)	Observed Phenotypes/ (% Occurrence from Total)
DIC	9	2/ (77)	7/ (77)	
DBD-DD 9aa 20 pg	43	4/ (91)	34/ (79)	1 no head (2) 4 poor tail development (9)
DBD-DD 16aa 20 pg	39	6/ (85)	31/ (79)	2 slightly bent body axis (5) 2 poor tail development (5)
DBD-DD 9aa 40 pg	39	6/ (85)	27/ (69)	6 slightly bent body axis (15)
DBD-DD 16aa 40 pg	34	7/ (79)	24/ (71)	3 severely bent body axis (8) 1 poor tail development (3)

DIC= Dye Injected Control
Phenotypes are not mutually exclusive.

(Table X: Survival Rates of 7-11-12 DBD Only Injection at 24 Hours Post Fertilization)



(Figure XVIII: Phenotype of Representative Fish from 7-11-12 DBD Only Injection)

A) DIC image of dye injected control embryo at 24 hpf. Embryo appears to be unaffected from the process of injection. B) DIC image of DBD injected embryo at 24 hpf. Although most embryos appeared wild type, this one represents the bent body axis observed. Other observed phenotypes were poor head and/or tail formation (data not shown).

Conclusions of Fast-1 DNA Binding Domain (DBD) + Fast-1 Smad2 Binding Domain (SBD) Injections

Data is not shown for the remaining DBD + SBD injection (7-17-12) due to redundancy of results. All of the constructs made were injected by themselves to

control for mRNA toxicity. From these injections it became apparent that the constructs containing the SBD produced some sort of an effect as compared to constructs containing the DBD. Referring to Table VIII, the SBD in Fok1 DD 9aa construct had a 51% survival rate vs. an 83% survival rate of the DBD containing construct with the same linker. Decreased rates of survival in embryos injected with the SBD construct alone (Table IX) is seen in comparison to survival rates of embryos injected with the DBD construct alone (Table X). Not only were the survival rates lower in embryos injected with the SBD, but a higher frequency of aberrant phenotypes was seen in these embryos. Figure XVII is an example of the typical phenotype seen. This consists of poor head and tail formation, somite deformities, loss of notochord and increased levels of blood.

Similar to the SBD in Fok1 DD construct, the DBD in Fok1 DD construct had some effect, although minimal in comparison to the SBD construct. The survival rates of embryos injected with the DBD construct was much higher and the frequency of observed phenotypes was lower, most of the fish looked wild type. Figure XVIII is representative of some of the phenotypes that were observed, such as a bent body axis and poor head or tail formation. The phenotypes observed in this construct were much less severe than seen in the SBD construct.

Aberrant EGFP expression patterns were seen in *lhx1a-EGFP* transgenic fish as compared to wild type. Intensity of EGFP was low overall, making it difficult to conclude changes in fluorescence being a result of the DBD and/or SBD vs. artifacts of injection.

Injections that contained both the SBD + DBD did not demonstrate survival rates significantly lower, or phenotypes that were more pronounced than when each of the constructs were injected singly. Due a vector mix-up in the lab, these injections were of a Fok1 DD variant with another Fok1 DD variant to be discussed later. No mutations were detected in the AREs of the *ephnb2, flh, foxa3, fynb, lft1, lft2, lhx1a, ndr1 (sqt), pitx2, rarab, tgif,* and *znf143* genes. This does not exclude the presence of mutations at other ARE sites, but the given data suggests that Fok1 DD in combination with Fok1 DD is unable to generate double strand DNA breaks as expected (Szcepek, *et al.,* 2007).

Use of Fok1 RR Vector with no DNA Binding Domain

Fok1 RR vector without a DNA binding domain was injected with the DBD in Fok1 DD 9 aa construct in order to see if Fok1 DD and Fok1 RR have enough affinity for one another to bind on their own without being separately recruited to the same DNA locus. Due to the off-target effects of the SBD, this injection was done to see if recruitment to the ARE by way of the DBD only provides sufficient specificity. Tables XI and XII summarize the survival rates and observed phenotypes of two injections with Fok1 RR + DBD-DD 9aa.

Construct Injected	Total	Dead/ Survival (%)	Appear Wild Type/ (%)	Observed Phenotypes/ (% Occurrence from Total)
Fok1 RR vector w/o insert 30 pg	44	3/ (93)	29/ (66)	5 slightly bent body axis (11) 3 severely bent body axis (7) 4 poor head development (9)
Fok1 RR vector w/o insert 80 pg	31	13/ (58)	0/ (0)	8 poor head/tail development, short (26) 10 no head or tail extension (32)
Fok1 RR vector w/o insert 30 pg and DBD-DD 9aa 30 pg	116	101/ (13)	3/ (3)	2 poor head/tail development and extension (2) 3 severely bent body axis (3) 7 no head or tail extension (6)

Phenotypes are not mutually exclusive and dye injected controls were deemed acceptable.

(Table XI: Survival Rates and Phenotypes of 10-23-12 Injection of Fok1 RR + DBD-DD 9aa at 24 Hours Post Fertilization)

Construct Injected	Total	Dead/ Survival (%)	Appear Wild Type/ (%)	Observed Phenotypes/ (% Occurrence from Total)
Fok1 RR vector w/o insert 2 pg	39	6/ (85)	24/ (62)	4 developmentally slower with large yolks (10) 4 partially necrotic (10) 1 small (3)
Fok1 RR vector w/o insert 10 pg	46	0/ (100)	44/ (96)	None
DBD-DD 9aa 30 pg	42	10/ (76)	16/ (38)	16 slightly bent body axis (38) 10 partially necrotic (24) 6 small head (14)
Fok1 RR vector w/o insert 10 pg and DBD-DD 9aa 30 pg	54	29/ (46)	0/ (0)	9 severely bent body axis (16) 9 partially necrotic (16) 10 no head or tail extension (19) 4 long (7)

Phenotypes are not mutually exclusive and dye injected controls were deemed acceptable.

(Table XII: Survival Rates and Phenotypes of 10-24-12 Injection of Fok1 RR + DBD-DD 9aa at 24 Hours Post Fertilization)

Conclusions of Fok1 RR Vector with no DNA Interacting Domain + Fast-1 DNA Binding Domain (DBD) in Fok1 DD with the 9aa Linker Injections

Referring to Tables XI and XII, the Fok1 RR vector without a DNA binding domain produced a strong effect at an 80 pg concentration, and nearly no effect at a 10 pg concentration. When injected in combination the two constructs produced a strong effect in which survival was low and embryos had a non-wild type phenotype. Phenotypes observed, such as a bent body axis, poor head and tail formation, or no head and tail formation are similar to Nodal knockout phenotypes (Dougan, *et al.,* 2003). Injections with Fok1 RR vector with no DNA binding domain were performed earlier than 10-23-12, on 8-15-12 and 8-16-12. These August injections were done with a vector later to be discovered was not Fok1 RR due to a delayed sequencing result. Plasmid was obtained from the wrong well on a 96-well stock plate.

Smad2 Constructs and Injections

To achieve greater specificity to the ARE, the two subunits of Fok1 should be recruited to the ARE by both Smad2 and Fast-1. Since use of the SBD appeared to block Smad signaling, a constitutively active (CA) form of Smad2 was used. The MH1 domain of Smad2 was removed in order to release the self-inhibition of the MH1 domain upon the MH2 domain. In this way, Smad2 CA could be recruited to the DNA without any dependence on signaling events. Smad2 CA was PCR amplified from a pCS2:MTZSmad2 CA plasmid and inserted into the SBD-Fok1 RR (inadvertently DD) construct in the place of the SBD (Müller, *et al.,* 1999). This construct placed Fok1 DD C-Terminal to Smad2 CA. The observations below, and Tables XII and XIV reflect the rates of survival and phenotypes seen at 24 hpf. Sequence encoding the entire Smad2 open reading frame was also PCR amplified from a pCS2MT:ZSmad2 plasmid and inserted into the same Fok1 RR (inadvertently DD) vector in the place of the SBD (Müller, *et al.,* 1999). Upon sequencing the Smad2 insert in a properly integrated clone it was discovered that the Smad2 DNA sequence contained a point mutation shown in Figure XIX.

Observations from 9-18-12 Injection of Smad2 CA in Fok1 RR (Accidentally DD) Vector with 9aa and 16aa linkers:

- Smad2 CA plasmid containing Fok1 DD was injected in concentrations of 10, 40 and 80 pg for each linker length. The majority of fish appeared wild type across all injections, with the appearance of some phenotypes such as a bent body axis, yolk deformities, poor head/tail development, and no head or tail extension.
- Dye injected controls sustained a fairly low rate of survival considering they were controls in that 8 out of 28 died by 24 hpf (71% survival rate). Therefore, phenotypes observed from this injection may be artifacts of the injection process rather than the Smad2 CA construct.

Construct Injected	Total	Dead/ Survival (%)	Appear Wild Type/ (%)	Observed Phenotypes/ (% Occurrence from Total)
DBD-DD 16aa 37.8 pg	41	0/ (100)	22/ (54)	6 slightly bent body axis (15) 10 severely bent body axis (24) 3 poor head development (7)
Smad2 CA-RR 16aa (accidentally DD) 76 pg	27	6/ (77)	17/ (63)	2 slightly bent body axis (7) 2 severely bent body axis (7)
Smad2 CA-RR 16aa (accidentally DD) 112.5 pg	30	10/ (66)	6/ (20)	9 slightly bent body axis (30) 5 severely bent body axis (17)
Smad2- CA DD16aa 76 pg with DBD-DD 16aa 37.8 pg	66	20/ (70)	31/ (47)	15 poor tail development (23) 6 poor head/tail development (9)

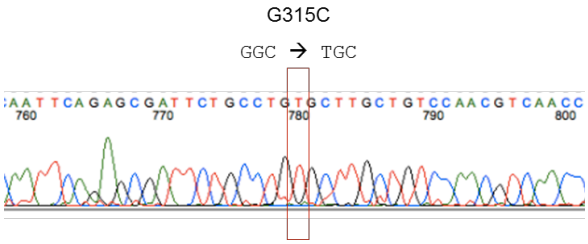
Phenotypes listed are not mutually exclusive and dye injected controls were deemed acceptable.

(Table XIII: Survival Rates and Phenotypes of 9-19-12 Injection of Smad2 CA + DBD Both in Fok1 DD Vector with 16aa Linker 24 Hours Post Fertilization)

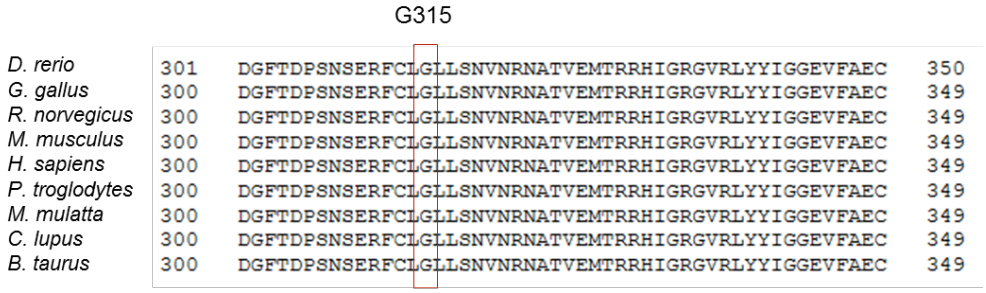
Construct Injected	Total	Dead/ Survival (%)	Appear Wild Type/ (%)	Observed Phenotypes/ (% Occurrence from Total)
DBD-DD 9aa 30 pg	41	3/ (93)	22/ (54)	6 slightly bent body axis (15) 10 severely bent body axis (24) 7 partially necrotic (17)
Smad2 CA-RR 9aa (accidentally DD) 80 pg	27	5/ (81)	17/ (63)	2 slightly bent body axis (7) 2 severely bent body axis (7) 1 small body (4)
Smad2 CA-RR 9aa (accidentally DD) 112.5 pg	30	13/ (57)	6/ (20)	9 slightly bent body axis (30) 2 severely bent body axis (7)
Smad2 CA-DD 9aa 80 pg with DBD-DD 9aa 30 pg	97	24/ (75)	33/ (34)	33 small body (34) 6 poor head formation (6) 7 severely bent body axis (7)

Phenotypes are not mutually exclusive and dye injected controls were deemed acceptable.

(Table XIV: Survival Rates and Phenotypes of 9-26-12 Injection of Smad2 CA + DBD Both in Fok1 DD Vector with 9aa Linker 24 Hours Post Fertilization)



(Figure XIX: Smad2 Point Mutation in Sequence of Smad2 Inserted into Fok1 DD)



(Figure XX: Smad2 Multiple Alignment Showing Conserved Glycine Residue)(Source: Homologene Result 21197)

Conclusions on Fok1 DD C-Terminal to Smad2 Constitutive Active (CA) and Smad2 Full Length Constructs

According to Tables XIII and XIV, the survival rates of the Smad2 CA in Fok1 DD with the 16 and 9 aa linkers respectively, were fairly acceptable when injected at about 80 pg, but decreased to about 50% survival when the injected concentration was increased to 112.5 pg. Phenotypes appearing wild type also decreased substantially from about 60% to 20% upon increasing the concentration injected. Phenotypes differing from wild type were less severe overall as opposed to other injections.

Upon sequence conformation of the Smad2 full length coding sequence placed in Fok1 DD, it was discovered that this clone sustained a point mutation encoding a cysteine residue instead of a glycine residue (Figure XIX). From a multiple sequence alignment this residue appears to be highly conserved and thus, injections with this construct were temporarily put on hold (Figure XX). This amino acid change was reported in the original zebrafish sequence deposited in Genbank by Uwe Strähle's lab who also reported that the pCS2MT:ZSmad2 construct was not biologically active (Müller, *et al.,* 1999).

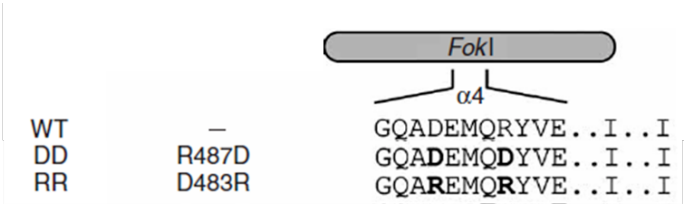
Fok1 DD N-Terminal to Smad2 Constitutive Active (CA)

Another construct in which Fok1 was placed N-terminal to Smad2 CA was generated with the belief that placing Fok1 DD at the C-terminus of the Smad2 CA protein may alter its ability to bind DNA (Müller, *et al.,* 1999). Fok1 DD was inserted into a Smad2 CA containing pCS2 plasmid attached at the N-terminus of Smad2 CA via a 9 aa linker (Müller, *et al.,* 1999). An attempt was made to place Fok1 DD in the

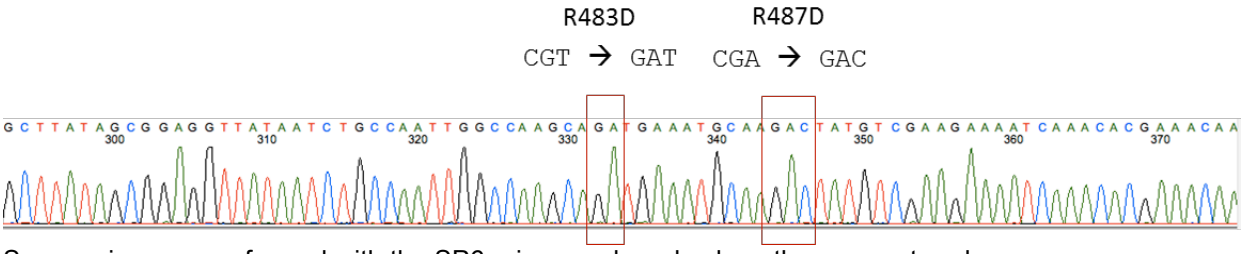
pCS2:ZSmad2 vector to obtain Fok1 DD N-terminal to the full length Smad2, but this vector yielded a high background of colonies containing self-ligated vector, even after many troubleshooting attempts.

** Fok1 RR and Fok1 DD Vector Mix-up*

All injections up until 10-23-12 are injections that consisted of Fok1 DD injected with Fok1 DD. This was discovered upon sequencing the construct in which 'Fok1 RR' was placed into pCS2:ZSmad2 CA in which 'RR' was really DD, as seen in Figures XXI and XXII. Fok1 was PCR amplified from the SBD in pCS2:Fok1 RR 9aa linker construct. Therefore, the Fok1 in this construct was also DD. The Smad2 CA and full PCR products were cloned into the SBD in pCS2:Fok1 RR 9 and 16 aa linker constructs, replacing the SBD. Since the original construct was actually Fok1 DD, the Smad2 CA and full constructs are also in DD and not RR. Fok1 works as an obligate heterodimer between Fok1 RR and Fok1 DD (Szczepek, *et al.,* 2007). Therefore, most likely Fok1 was not able to function as an endonuclease. In total there are more Smad related Fok1 DD clones vs. DBD in Fok1 DD clones, therefore the DBD in Fok1 DD clones were changed to RR via site directed mutagenesis to be injected along with the Smad DD clones as shown in figure XXIII.

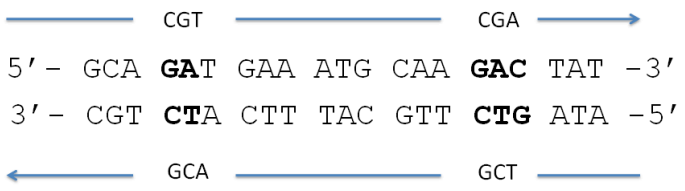


(Figure XXI: Amino Acid Sequence of Fok1 DD versus Fok1 RR)(Adopted from: Szczepek, *et al.,* 2007)



Sequencing was performed with the SP6 primer and reads along the sense strand.

(Figure XXII: Sequencing Result of Fok1 into pCS2:Smad2 CA 9aa Linker Demonstrating Fok1 is the DD Variant)



*Arrows represent primer sequence homologous to DNA sequence.

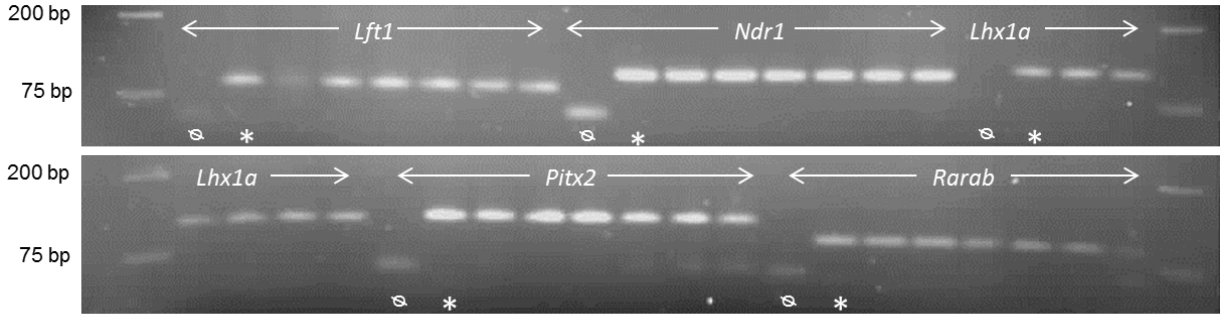
(Figure XXIII: Site Directed Mutagenesis of Fok1 DD into Fok1 RR)

After SDM and the generation of the DBD-RR clone, mRNA was made and injections were done using the constructs in which Fok1 DD was placed N-terminal to Smad2 CA, in combination with DBD-RR. Tables XV and XVI summarize the survival rates and phenotypes seen in these embryos. Select embryos were used for PCR analysis and results are shown in Figures XXIV and XXV.

Construct Injected	Total	Dead/ Survival (%)	Appear Wild Type/ (%)	Observed Phenotypes/ (% Occurrence from Total)
Dye Injected Controls	10	1/ (90)	9/ (90)	All look wild type
DBD-RR 9aa 10 pg	36	0/ (100)	31/ (86)	3 slightly bent body axis (8) 1 severely bent body axis (3) 1 no head or tail extension (3)
DBD-RR 9aa 40 pg	42	2/ (95)	35/ (83)	5 severely bent body axis (12)
DBD-RR 16aa 10 pg	51	1/ (98)	45/ (88)	2 partially necrotic (4) 3 poor tail formation (6)
DBD-RR 16aa 40 pg	72	21/ (71)	2/ (3)	16 severely bent body axis (22) 32 no head or tail extension (44) 1 partially necrotic (1)
Fok1 DD into pCS2: Smad2 CA 9aa 40 pg	102	46/ (55)	2/ (2)	15 severely bent body axis (15) 39 no head or tail extension (38)
Fok1 DD into pCS2: Smad2 CA 9aa 40 pg with DBD-RR 16aa 40 pg	146	36/ (75)	41/ (28)	14 short (10) 3 slightly bent body axis (2) 19 severely bent body axis (13) 33 no head or tail extension (23)

Phenotypes are not mutually exclusive.

(Table XV: Survival Rates and Phenotypes of 11-06-12 Injection of Fok1 DD N-Terminal to Smad2 CA + DBD-RR at 24 Hours Post Fertilization)



*Indicates Dye Injected Control
⌀ Indicates Negative Control

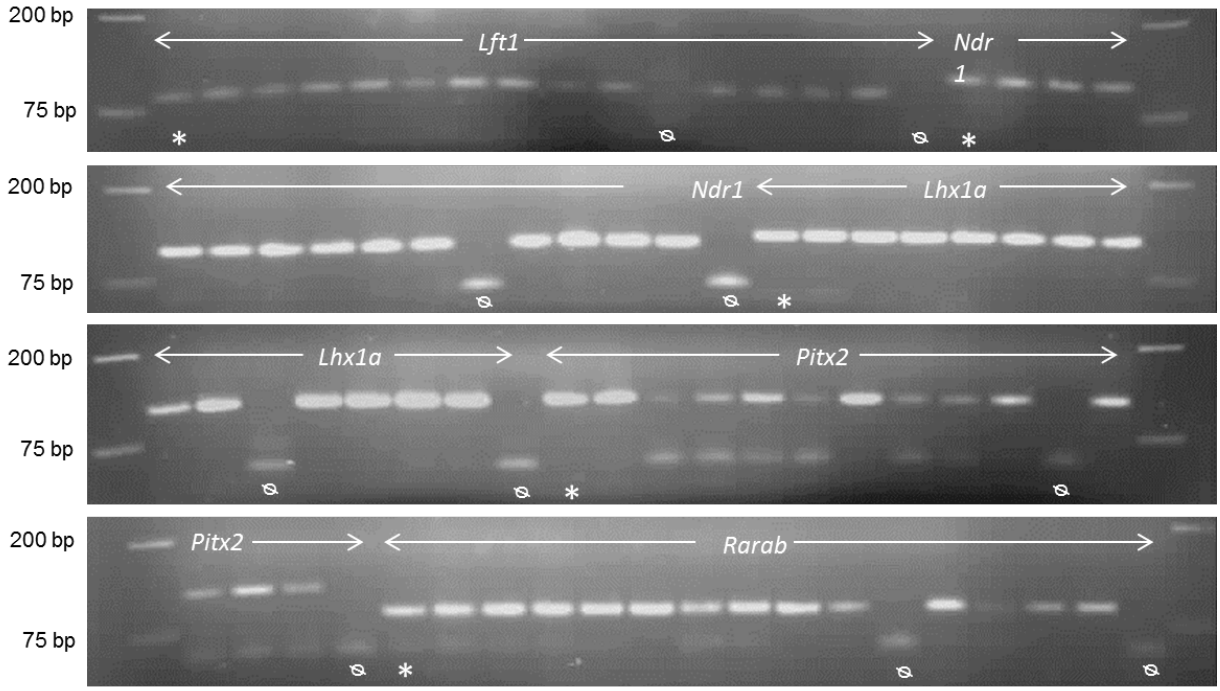
(Figure XXIV: PCR Results of Embryos Prepped from 11-06-12 Injection of Fok1 DD N-Terminal to Smad2 CA + DBD-RR)

PCR results are depicted for each primer set used at this point in time. PCR was done with a 1:10 dilution of DNA. Embryos from which DNA was prepped and used for PCR were selected only from injections in which both Fok1 DD into pCS2:ZSmad2 CA and DBD-RR (both 9 and 16 aa linkers) were injected. Embryos were also prepped from zebrafish only injected with dye as a control. The selection of embryos included ones demonstrating the phenotypes described in Table XV, as well as fish that appeared wild type. All embryos were collected at 24 hpf. No bands indicating small insertions or deletions were observed across all samples in the PCR reaction.

Construct Injected	Total	Dead/ Survival (%)	Appear Wild Type/ (%)	Observed Phenotypes/ (% Occurrence from Total)
DBD-RR 9aa 10 pg	26	0/ (100)	24/ (92)	1 slightly bent body axis (4)
DBD-RR 16aa 10 pg	33	1/ (97)	24/ (73)	4 short (12) 6 severely bent body axis (18) 2 no head or tail extension (6)
Fok1 DD into pCS2:Smad2 CA 9aa 20 pg	34	3/(91)	1/ (3)	19 expanded somites (56) 11 no head or tail extension (32)
Fok1 DD into pCS2:Smad2 CA 20 pg with DBD-RR 9aa 10 pg	38	18/ (53)	1/ (3)	9 severely bent body axis (24) 10 expanded somites (26) 2 short (5)
Fok1 DD into pCS2:Smad2 CA 20 pg with DBD-RR 16aa 10 pg	29	3/ (90)	1/ (3)	21 severely bent body axis (72) 1 expanded somites (3) 3 short (10)

Phenotypes are not mutually exclusive and dye injected controls were deemed acceptable.

(Table XVI: Survival Rates and Phenotypes of 11-07-12 Injection of Fok1 DD N-Terminal to Smad2 CA + DBD-RR at 24 Hours Post Fertilization)



*Indicates Dye Injected Control
⦸ Indicates Negative Control

(Figure XXV: PCR Results of Embryos Prepped from 11-07-12 Injection of Fok1 DD N-Terminal to Smad2 CA + DBD-RR)

PCR results are depicted for each primer set used at this point in time. PCR was done with a 1:10 dilution of DNA. Embryos from which DNA was prepped and used for PCR were selected only from injections in which both Fok1 DD into pCS2:ZSmad2 CA and DBD-RR (both 9 and 16 aa linkers) were injected. Embryos were also prepped from zebrafish only injected with dye as a control. The selection of embryos included ones demonstrating the phenotypes described in Table XVI, as well as fish that appeared wild type. All embryos were collected at 24 hpf. No bands indicating small insertions or deletions were observed across all samples in the PCR reaction.

Conclusions of Fok1 DD N-Terminal to Smad2 Constitutively Active (CA)

The injections with Fok1 DD N-terminal to Smad2 CA 9aa + DBD-RR 9 and 16 aa linkers were the first injections where a Fok1 DD + RR were injected. These were the first injections where the Fok1 endonuclease could theoretically function properly as an obligate heterodimer. Sequencing data of the 'Fok1 RR' (really DD) insert in the pCS2MT:ZSmad2 CA vector lead to the discovery of the vector mix-up. After SDM of the DBD-DD construct, the Fok1 DD N-terminal to Smad2 CA construct now had a partner to be injected with. To test the toxicity of the DBD as DBD-RR, the 9 and 16 aa variants of this construct were injected by themselves. Injections of the DBD in Fok1 RR with both the 9 and 16 aa linkers were successful in that there was a high percentage of survival and phenotypes that appeared wild type (Tables XV, XVI). The toxicity of the Fok1 DD N-terminal to Smad2 CA construct was also injected by itself. At a concentration of 40 pg survival was 55% at 24hpf and only 2% appeared wild type

(Table XV). When injected at 20 pg the survival rate increased to 91% but embryos still had phenotypes of a bent body axis, head and tail deformities, and somite deformities (Table XVI).

When the Smad2 CA with N-terminal Fok1 via a 9aa linker was injected with the DBD in Fok1 RR with 9 and 16 aa linkers at a concentration of 40 pg each, the survival rate was 75% and the appearance of wild type embryos reached 28% at 24hpf. Phenotypes of a bent body axis and head and tail deformities are suggestive of disruptions in Nodal signaling (Table XV). When injected at a lower concentration of 20 pg Fok1 DD in pCS2:Smad2 CA and 10 pg DBD in Fok1 RR 9 and 16aa linkers survival raged from 53% to 90% with respect to DBD linker, but with more obvious mesodermal effects observed through the increased incidence of somite deformities (Table XVI). This provides stronger evidence for the disruption of Nodal signaling. In the case of the Fok1 DD in pCS2:ZSmad2 CA 9aa linker construct injected with the DBD in Fok1 RR with the 9 and 16aa linkers, survival rates and wild type appearance rates are similar as compared to individual injections, but phenotypes seem to be more pronounced when injected in combination (Tables XV, XVI). The increased severity of phenotypes when constructs containing both Fok1 DD and Fok1 RR were injected suggest that Fok1 may be generating double strand DNA breaks, which may or may not be present at the ARE. The PCR results provide evidence that no mutations are being generated at the ARE in that no mutations were detected in the *lft1, lhx1a, ndr1, pitx2 and rarab* genes (Figures XXIV, XXV). This does not exclude the possibility that mutations still exist elsewhere in the genome, or that mutations are occurring at a frequency too low to detect with this method. It is also possible that phenotypes appear more severe from the nonspecific effects of each of the constructs individually, acting either in an additive or synergistic manner when injected together. If Fast-1 and Smad2 signaling are both independently affected from the individual constructs then that may have an indirect effect on proper recruitment of the necessary factors at the ARE to initiate Nodal-related gene transcription.

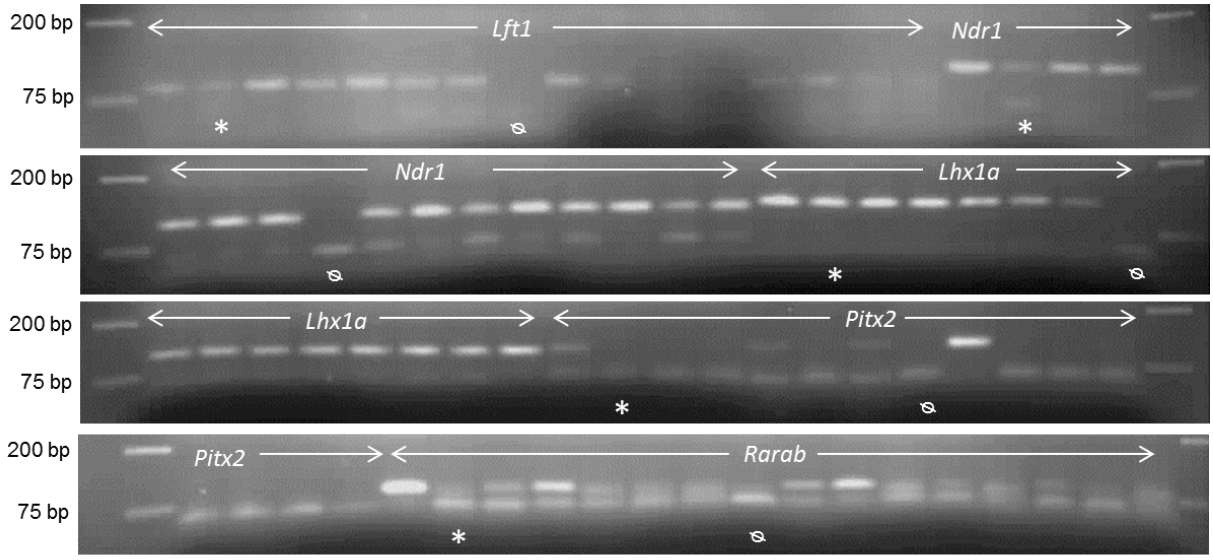
Fok1 DD C-Terminal to Smad2 Constitutively Active (CA) Re-Injected with Fast-1 DNA Binding Domain (DBD) in Fok1 RR Constructs

After the development of the DBD-RR construct by SDM, the Fok1 DD C-terminal to Smad2 CA both 9 and 16 aa linker constructs were re-injected with the DBD-RR 9aa construct to see if mutations could be generated. The Fok1 DD C-terminal to Smad2 CA constructs when injected alone did not demonstrate high toxicity or off-target effects as the SBD did. Therefore, this construct would be promising in reducing toxicity, while having both components of the Fok1 obligate heterodimer when injected with DBD-RR.

Construct Injected	Total	Dead/ Survival (%)	Appear Wild Type/ (%)	Observed Phenotypes/ (% Occurrence from Total)
Smad2 CA-DD 9aa 40 pg	29	6/ (79)	0/ (0)	19 slightly bent body axis (65) 3 poor head/tail formation (10) 1 poor tail formation (3)
Smad2 CA-DD 16aa 40 pg	21	7/ (66)	0/ (0)	12 slightly bent body axis (57) 1 severely bent body axis (5) 1 no tail formation (5)
Smad2 CA-DD 9aa 40 pg with DBD-RR 9aa 10 pg	29	3/ (90)	4/ (14)	21 slightly bent body axis (72) 1 poor head/tail formation (3)
Smad2 CA-DD 16aa 40 pg with DBD-RR 16aa 10 pg	29	4/ (86)	4/ (14)	13 slightly delayed development, about 26 somites (45) 1 delayed development, about 21 somites (3) 4 partially necrotic (14) 3 poor tail formation (10)

Phenotypes are not mutually exclusive and dye injected controls were deemed acceptable.

(Table XVII: Survival Rates and Phenotypes for 11-08-12 Injection of Fok1 DD C-Terminal to Smad2 CA + DBD-RR at 24 Hours Post Fertilization)



*Indicates Dye Injected Control
ø Indicates Negative Control

(Figure XXVI: PCR Results of Embryos Prepped from 11-08-12 Injection of Fok1 DD C-Terminal to Smad2 CA + DBD-RR)

PCR results are depicted for each primer set used at this point in time. PCR was done with a 1:10 dilution of DNA. Embryos from which DNA was prepped and used for PCR were selected only from injections in which both Smad2 CA in Fok1-DD and DBD-RR (both 9 and 16 aa linkers) were injected. Embryos were also prepped from zebrafish injected with dye as a control. The selection of embryos included ones demonstrating the phenotypes described in Table XVII, as well as fish that appeared wild type. All embryos were collected at 24 hpf. No bands indicating small insertions or deletions were observed across all samples in the PCR reaction.

Conclusions of Fok1 DD C-Terminal to Smad2 Constitutively Active (CA) Re-Injected with Fast-1 DNA Binding Domain (DBD) in Fok1 RR

When the constructs with Fok1 DD C-terminal to Smad2 CA with both the 9 and 16aa linkers were injected with the same linker version of DBD in Fok1 RR, the results showed some improvement. Survival rates increased slightly and the rate of wild type phenotypes was about the same but with a less severe phenotype. Phenotypes observed in these injections seemed to be more related to delayed and slowed growth vs. aberrant morphologies. At 24 hpf these embryos should be between the 26-somite stage of the segmentation period, and the prim-5 stage of the pharyngula period (Kimmel, Ballard, Kimmel, Ullmann, Schilling, 1995). The majority appeared to be at the 26-somite stage, which may be an artifact of the injection itself and not what was injected. Many injected embryos seemed to develop at a slower rate. A minority of embryos had head and tail deformities.

No mutations were observed in the *lft1, lhx1a, ndr1, pitx2,* and *rarab* genes tested by PCR. This does not exclude the presence of mutations elsewhere in the genome. Mutations may also exist in the AREs of these genes but may occur at a frequency much lower than the sensitivity of this method of detection.

Discussion

A goal of this study was to generate target specific mutations for Nodal-dependent genes by targeting the ARE. In this way, targeted genes could still signal in pathways independent of Nodal. In light of the fact that a response element is targeted, there is potential for generating many genetic mutations across different genes within the Nodal pathway itself. Nodal signaling plays an essential role in early developmental processes such as mesoderm or endoderm cell fate determination, neural development and left-right patterning (Schier, 2003). Concentrations and optimal linker lengths need to be assessed across all injections to determine what is optimal to avoid non-specific effects and generate founders that survive to be sexually mature adults. The results up until this point seem to suggest that the optimal constructs and/or concentrations have not yet been achieved.

Off-Target Effects of Fast-1 Smad2 DNA Binding Domain

Due to a vector mix-up in the lab, many of the early injections consisted of a Fok1 DD variant, with another Fok1 DD variant. Fok1 operates functionally as a dimer. In order to help minimize nonspecific effects and optimize the use of Fok1 for gene targeting, Fok1 was engineered into an obligate heterodimer in which the DD and RR variants were used. From injections in which Fok1 DD and Fok1 DD were injected, one would anticipate only seeing effects dealing with artifacts of injection or nonspecific mRNA toxicity. Therefore, it was unexpected to see some embryos that had a Nodal knock-out phenotype. It is even more unexpected to see these results from injections in which only a single construct was injected, indicating some of the constructs mimicked dominant negatives.

Injections of the DBD and SBD independently demonstrated that the SBD by itself had a more pronounced effect. Survival rates were low (Tables VII and VIII) and phenotypes such as poor head and tail formation, somite deformities, loss of notochord, and increased presence of blood were all observed (Figure XVII). All of these phenotypes are similar to those seen in *cyc* and *sqt* mutants, suggesting a disruption in Nodal signaling (Dougan, *et al*., 2003). Perhaps the SBD binds Smad2 and Smad2 recruits the Fok1 endonuclease to the DNA site via the SBD, allowing for the dimerization of Fok1 DD that has retained some self-affinity for other Fok1 DDs. This would allow for the possibility of double strand DNA breaks to still be generated under conditions in which one construct was injected. Due to the fact that Fok1 DD and RR were engineered specifically to avoid this event, it seems more likely that the SBD in Fok1 DD construct is acting as a competitive inhibitor of Smad signaling (Szczepek, *et al.,* 2007). Evidence against the self-dimerization of Fok1 DD can be seen in Table VIII. The survival rates of embryos injected with both SBD in Fok1 DD and DBD in Fok1 DD were higher than that of the SBD in Fok1 DD alone. If Fok1 DD retained some functionality through dimerization with itself, increasing the concentration of Fok1 DD in the embryo would likely be more toxic. The fact that injections with both constructs resulted in increased survival is most likely random and a consequence of small sample size. Further evidence can be seen in the PCR results shown in Figures XIII and XV. No mutations were detected in the *ephnb2, flh, foxa3, fynb, lft1, lft2, lhx1a, ndr1 (sqt),*

pitx2, rarab, tgif, and *znf143* genes. The SBD binds Smad2 where endogenous Fast-1 would bind; therefore the engineered SBD is competing with Fast-1 for Smad2 binding. Smad2 does not have a high affinity for DNA binding. Therefore, it needs other factors, such as Fast-1 for recruitment to the DNA (Müller, *et al.,* 1999). Binding of the SBD to Smad2 would prevent Smad2 from interactions with Fast-1, and based on the level of dependence of Smad2 recruitment to the DNA via Fast-1, binding of the SBD may prevent Smad2 recruitment to the DNA. Binding of the SBD may also prevent Smad2 interactions with other molecules. Given that Nodal signals through the Smad signaling pathway, blocks in Smad signaling would also block Nodal signaling, producing the phenotypes seen in Figure XVII.

Although the DBD construct elicited some effects, these are likely to be able to be controlled through proper dosing. This construct may have less of an off-target effect because it competes with Fast-1 for the DNA binding site, rather than altering endogenous Fast-1 function (Table X and Figure XVIII). In the 7-05-12 injection, both the DBD and SBD in Fok1 DD 16 aa linker appear to have the same off-target effects demonstrated by fairly similar survival rates at 24 hpf, although sampling size was low (Table VII).

Lhx1a-EGFP Mutation Detection Assay

Images from the *lhx1a-EGFP* assay demonstrate aberrant EGFP expression from wild type when injected with the SBD and DBD both in Fok1 DD with the 16aa linker (Figures XIV and XVI). This may suggest that the *lhx1a* gene is highly susceptible to the off target effects of both the DBD and SBD as described above. *Lhx1a* is involved in gastrulation and kidney development (Swanhart, *et al.,* 2010). Therefore, it makes sense that disruptions in the proper function of this gene would affect survival. Improper recruitment of Fast-1 and Smad2 to the ARE through signaling interferences introduced with the SBD and DBD constructs may result in the lack of *lhx1a* transcription without the formation of the proper ARF complex.

Difficulties with the *lhx1a* assay arose due to the fact that the intensity of EGFP expression was not as high as anticipated across all injections and in dye injected controls. Due to this, it was hard to qualitatively determine gaps in EGFP fluorescence accurately enough to make conclusions from. One cannot also rule out the possibility that aberrant EGFP expression may be an artifact of the injection itself due to typical slowed growth of injected embryos. Gaps may be more apparent if the assay was performed at later stages, such as at the five somite stage in which there is a more defined fluorescence pattern in tissues destined for notochord formation (Figure IX). The embryos injected were hemizygous for the *lhx1a-EGFP* transgene, if it was present. Therefore, a more robust assay may be produced if hemizygous fish were mated to generate homozygous transgenic fish. Images in the paper from the Hukriede lab that generated the transgenic line appeared very robust and were from embryos homozygous for the transgene (Swanhart, *et al.,* 2010). It may be more exhaustive to generate homozygous fish in light of the fact that additional mating would take longer and fish would need to be genotyped to screen for homozygotes, but this effort may prove invaluable as a preliminary screen considering the efforts required in screening for specific mutations. It may be beneficial to see how these results differed in an

injection in which a Fok1 RR and a Fok1 DD variant were co-injected. Currently the *lhx1a-EGFP* assay was only performed on injections in which two Fok1 DD constructs were injected.

Loss of Specificity with use of Fok1 RR Vector with no DNA Binding Domain

In order to determine if Fok1 RR had enough affinity for binding to Fok1 DD on its own, Fok1 RR in the absence of a DNA binding domain was injected with the DBD in Fok1 DD. In these injections survival was low with most embryos demonstrating non-wild type phenotypes at 24 hpf (Tables XI and XII). Due to the strong effect of these injections, it is likely that the use of the Fok1 RR vector without a DNA binding domain results in many non-specific effects. The fact that the Fok1 RR vector without a DNA binding domain, when used with the DBD in Fok1 DD produced an effect supports the idea that Fok1 RR and Fok1 DD have a high enough binding affinity to bind on their own. The downside to this is that the use of the DBD as the only determinant of specificity is most likely not enough. One could try dialing down the concentration injected of the DBD in Fok1 DD to inject with Fok1 RR to lower than 30 pg since previous injections show that by itself, the DBD did not generate significant effects at this concentration (Table X).

Smad2 Constitutively Active (CA) and Full Length Constructs

Injections with Fok1 DD placed N-terminal to Smad2 CA by itself resulted in low survival when injected at 40 pg and phenotypes such as a bent body axis, head and tail deformities, and somite deformities were observed across all injections of 20 and 40 pg (Tables XV and XVI). Such phenotypes are in concordance with a disruption in Nodal signaling. In this case, overexpression of a constitutively active form of Smad2 appears to have a toxic effect due to increased Smad signaling. Use of a constitutively active version of Smad2 has been previously shown to produce a dorsalizing phenotype in which the tail is curled up or missing (Müller, *et al.,* 1999). Injections with concentrations lower than 20 pg would be worth trying to minimize Smad2 overexpression.

In the case of the construct where Fok1 DD was placed C-terminal to Smad2 CA, survival rates at 24 hpf were acceptable but generated zero fish with a wild type phenotype when injected alone (Table XVII). Although the incidence of a wild type phenotype was low, the observed phenotypes were not as pronounced as in the constructs where Fok1 DD was placed N-terminal to Smad2 CA. This may suggest that placing Fok1 DD C-terminal to Smad2 CA has an effect on the proper folding and/or function of Smad2 CA. Fok1 DD and Fok1 RR may have had a decreased efficiency of inducing mutations in this case due to the loss of ability for Smad2 CA to bind at the ARE based on placement of Fok1. The PCR results support this hypothesis in that no mutations were detected in the *lft1, lhx1a, ndr1, pitx2,* and *rarab* genes for this injection (Figure XXVI). The evidence also suggests that the toxicity effects of the Smad2 CA construct are lessened when Smad2 CA is altered so as to lose some functionality, which may also support the idea that all of the effects observed are more likely a result of altered signaling vs. the induction of mutations at the desired genes.

In addition, further experiments need to be done injecting the construct in which sequence encoding a full length Smad2 was inserted into Fok1 DD (Fok1 C-terminal). In this light Smad2 would be overexpressed, but may not have increased activity since it is still under regulatory control of the MH1 domain. A full length Smad2 also retains the propensity for signaling in all other naturally occurring signaling events both inhibiting and activating Smad2, which may be a good or bad thing. The C-terminal attachment of Fok1 DD may interfere with the native folding of Smad2 and thus its function. Loss of function would not necessarily be a drawback so long as Smad2 can bind DNA at the ARE. In fact it would be ideal if DNA binding was all that the full length Smad2 Fok1 DD fusion would be capable of doing. Upon sequencing this construct it was discovered that there was a point mutation encoding a cysteine residue in place of the wild type glycine residue (Figure XIX). Figure XX shows that this glycine residue is highly conserved among several species of chordates. If this residue is important for Smad2 function, than loss of function due to the amino acid change may be helpful in this application so long as the Smad2 Fok1 DD fusion can still bind DNA. Previous studies injecting mRNA of the pCS2MT:ZSmad2 plasmid showed that the full length Smad2 (containing the mutation) was not biologically active, even at concentrations as high as 1µg/ml (Müller, 1999). If this loss of activity is due to loss of DNA binding, and thus loss of function as a transcription factor, than that may prove detrimental to the application set forth here. In the attempt to generate this construct some clones resulted in an incorporation of Smad2 in which a part of the sequence encoding the MH1 domain was randomly removed. It may also be worth it to make mRNA from this clone in order to see if it retains DNA binding capabilities.

Future Directions and Considerations

Enhanced Specificity

Smad2 is recruited to DNA not only through interactions with Fast-1, but with the aid of the co-smad, Smad4. Therefore, Smad2 does not solely rely on Fast-1 for recruitment to the DNA. Given that binding of the SBD seems to interfere with Smad2 recruitment to DNA, it may be beneficial to generate constructs designed to bind sites on Smad2 that are not involved in binding interactions and are of minimal functional importance. A possible desirable site would be the linker region between the MH1 and MH2 domains, which is less highly conserved than the MH1 and MH2 domains (Müller, *et al.,* 1999).

Spacer Length Variation

A factor to consider for the effectiveness of the recruitment of Fok1 DD and Fok1 RR to the ARE is the spacing of the Fast-1 DNA binding site in relation to the Smad2 DNA binding site. A question to ask would be, are the linker lengths appropriate for

Fok1 DD and Fok1 RR to be able to dimerize? If the distance between two amino acids is about 5 ½ Å, than the 9aa linker would have a length of about 49 ½ Å, and the 16aa linker would be about 88 Å long. This would mean a 9aa construct injected with a 9aa construct would bring Fok1 DD and Fok1 RR to meet at the center of a region spanning about 99 Å, and two 16aa constructs injected together would span over about 176 Å. In looking at the genomic sequence of many Nodal target genes, there are often multiple Fast-1 and Smad2 DNA binding sites (Smad2 more frequent). The region with the highest frequency of these sites has been inferred to be the ARE. Table V lists the first and second closest Smad2 DNA binding sites with respect to a Fast-1 binding site of select genes. The smallest distance between a Fast-1 and Smad2 DNA binding site is zero nucleotides and the largest among the first closest Smad2 sites is 45 nucleotides (about 153 Å). When referring to a second closest Smad2 binding site, distances can reach up to 416 nucleotides (about 1788 Å). For example, the distance between a Smad2 and Fast-1 DNA binding site in the *lhx1a* ARE is 29 nucleotides, which is about 99 Å, and would be ideal for two 9aa constructs injected in combination. In the case of *fynb* the distance is 6 nucleotides, spanning about 20 Å, in which two 9aa linker constructs injected together may overlap in such a way where Fok1 RR and Fok1 DD do not dimerize at all, or may dimerize in an area not related to the ARE. These estimations are limited in they are assuming linear sequence and disregard secondary structures. Given this wide variability, distances between Smad2 and Fast-1 DNA binding sites vary greatly in which two linker lengths may not provide enough variability to target all ARE containing genes. Conversely, the linker length may actually be used to generate more specificity in targeting a certain subset of genes that have similar distances between their Smad2 and Fast-1 DNA binding sites. An additional thing to consider may be the insertion of an increased frequency of flexible residues, such as glycines into the linker sequences in order to allow for flexibility of the linkers (Figure XI). With increased flexibility, larger linkers may still be effective across smaller distances between Fast-1 and Smad2 DNA binding sites on the genomic sequence.

Alternative Approaches to Detecting Mutations

A possible assay to quantify the effectiveness of the construct's ability to generate mutations in the promoters of Nodal target genes would be to create an ARE:GFP fusion transgenic line. In this way fluorescence would be detected in regions in which an endogenous ARE containing gene is expressed. In order to generate this construct an exact ARE needs to be more clearly defined due to the large variability in number of Fast-1 and Smad2 DNA binding sites. A clear transcription start sequence would also have to be integrated into the ARE. The assay would be helpful because it would allow for proof of principle in detecting mutations made at the ARE based on gaps in fluorescence. The way in which this differs from the *lhx1a:EGFP* assay is that more genes could be assayed with the use of an ARE as the reporter, and not just *lhx1a's* promoter region. Using only the ARE fused to a GFP will result in fluorescence in regions of all genes that contain an ARE, but will not be influenced by other regulatory elements in a gene's promoter. Although this assay would be beneficial in determining whether or not an ARE is effectively targeted, the assay is not exactly indicative of the mutations being generated in endogenous genes. Mutations may

either be generated in the ARE of the transgene, or an ARE of an endogenous gene. Both the transgene and endogenous genes are independently activated by the same transcription factors of Fast-1 and Smad2, among others. Since mutations are being generated in genomic sequence and do not necessarily affect the expression of the transcription factors involved, both the transgene and endogenous genes can be activated in an independent manner from the same factors. Injected embryos are mosaic and the mutations generated are random within genes containing an ARE. Therefore, within the same cell the transgene could be activated while the endogenous gene sustains a mutation. This assay would not detect that mutation due to continued detection of fluorescence from retained transgene expression. The reverse may be true in which the endogenous gene was targeted but the transgene was not. This may also be a downside to the *lhx1a-EGFP* assay as well.

For this application, the use of fluorescence resonance energy transfer (FRET) may be more beneficial. Using this assay, fluorescence would only be emitted upon Fok1 DD binding Fok1 RR. The assay would tell if the two subunits are able to effectively dimerize, in which the desired specificity and subsequent generation of a double strand DNA break are assumed. Sequence encoding a donor fluor would need to be integrated into the Fok1 DD construct, and sequence encoding an acceptor fluor would be integrated into the Fok1 RR construct, or vice versa. The idea is that excitation of the donor flour will result in excitation of the acceptor flour only if the two fluors are within 100 Å of one another. This is about on the same order of the distances observed between Fast-1 and Smad2 DNA binding sites within the ARE. Microscopic observations should be done at the wavelength emitted by the acceptor fluor, which would be indicative of FRET between the donor and acceptor fluors (Jares-Erijman, Jovin, 2003). Detection of fluorescence in this assay would be a promising sign that Fok1 DD and Fok1 RR were effectively recruited to an ARE site. A pull down assay would also demonstrate Fok1 RR and DD binding but FRET would be more beneficial for this application in that temporary associations could be observed, as well as spatial information as to where fluorescence is emitted. Detection of fluorescence with respect to space may demonstrate a preliminary idea of what genes may have been targeted. A possible downside to this assay would be determining the proper placement of the fluorescent protein in the construct to avoid disruption in DNA binding and to retain protein function.

Although better construct design and further tweaking of injected concentrations would improve the frequency by which mutations were generated, another problem to consider is not the generation of mutations, but the detection of them. Genotyping in this experiment thus far has been very labor intensive, involving manual identification of Fast-1 and Smad2 DNA binding sites, primer design flanking these regions, performance of individual PCRs and the running of each sample individually on agarose gels. One of the most beneficial tools would be the development of software that performs genome searches for Fast-1 and Smad2 DNA binding sites within a certain distance of one another, allowing for wobble in consensus sequences. Up to this point exact consensus sequences have been manually searched for in genes known to have Smad2/4 and Fast-1 DNA binding sites. Identifying more genes with the potential of containing an ARE would allow for the design of more primer sets to be used to amplify a larger number of Nodal-dependent genes. A more high-throughput method needs to

be used in order to analyze more genes at once. In this study a small subset of genes were used to get a preliminary idea before undertaking the exhaustive efforts of genotyping many genes by the current methods. About 167 genes have been identified thus far that contain both Fast-1 and Smad2 binding sites, a method to analyze all of these genes at once would be highly beneficial (Liu, *et al.,* 2011). Plates can be generated with up to 384 wells containing up to 384 different primer sets from Integrated DNA Technologies (IDT). In this way up to 384 target sequences can be PCR amplified at once (given the availability of a thermocycler with this capacity, 96 well plates can also be used). These PCR products could be assessed by running the samples on a gel (still exhaustive) or by high resolution melt analysis (HRMA).

HRMA can be used to detect mutations and can be applied for use with 96 well, and up to 384 well plates. A fluorescent dye with a high binding affinity for heteroduplex regions is added to the PCR mix and used for quantitative PCR. The PCR step occurs as usual, followed by HRMA. The idea is that the temperature is raised so as to cause the two DNA strands of the PCR product to separate. Each wild type PCR product should melt at a certain temperature. If small insertions or deletions are present in the PCR product (as a result of a Fok1 generated mutation) a heteroduplex region is formed and the melting temperature of this PCR product will differ from that of the wild type product. Melting curves are generated based on fluorescence emission in which fluorescence is higher in double stranded DNA and decreases once the DNA has effectively melted. This demonstrates the importance of the use of a dye that binds heteroduplex regions with high affinity. These dyes help to accurately demonstrate different melting temperatures in mutant DNAs during this process (Erali, Writtwer, 2010). This could be performed on the entire array plate at once, in which all samples would be analyzed for mutations at once and may prove to be a good strategy to screen many genes at once.

Sequencing analysis of all ARE containing genes could be performed with the use of high throughput sequencing technologies. Often concentrations of PCR products obtained in a PCR reaction are not high enough for a high throughput sequencing reaction to work. Multiple PCR reactions need to be carried out, followed by pooling. Fluidigm has developed a high throughput means to pool PCR products for high throughput sequencing in their Access Array IFC system. This system is optimized for use with most high throughput sequencing systems, such as the Illumina HiSeq2000. The Fluidigm array is available from 48 to 96 wells, and has a 99% call rate accuracy. The reaction involves generating primers to amplify the desired region (ARE) with a universal forward tag attached to the forward primer, and a universal reverse tag attached to the reverse primer. The array is composed of wells on one side containing primers, and wells on the other side containing sample. PCR is then carried out in 2,304 reaction chambers contained within a 2 cm^2 area on the middle of the plate. This means that each target region is amplified in 48 PCR reactions simultaneously, in which about 1.1 MB of DNA are obtained per amplicon. The plate is compatible with the BioMark HD System, FC1 Cycler, EP1 Reader, and IFC Controller HX systems. The samples are automatically pooled back into the 48 wells and can be run in another traditional PCR reaction with primers incorporating the Illumina next-generation forward and reverse adaptors. A unique barcode sequence could also be incorporated into either the forward, or reverse primer in order to label certain embryo DNAs if grouping is

desired based on observed phenotypes (in this case all primer sets would be incorporated in one well in order to tag a well to a specific embryo, but this may limit screening power). After the second PCR step all PCR products are pooled and ready to be run on a single lane of the Illumina HiSeq 2000 through a paired-end run of 2x100 bases. Detection of mutations is performed through barcode or specific primer sequence comparisons with respect to wild type as dictated by the Illumina protocol (Halbritter, Diaz, Chaki, Porath, Tarrier, Fu, Innis, Allen, Lyons, Stefanidis, Omran, Soliman, Otto, 2012). This method has the propensity to amplify the AREs of many Nodal-related genes at once, with ease and high sensitivity in order to detect any mutations sustained from exposure to Fok1.

Conclusions

The potential for gene targeting with the use of the Fok1 endonuclease tethered to endogenous DNA binding domains is promising in being used to target a vast number of genes. The use of native DNA binding domains offers an ideal situation in terms of obtaining the desired specificity. Although the number of genes targeted through one pair of constructs may be less due to optimal linker length with respect to the separation of Fast-1 and Smad2 DNA binding sites in the ARE, there is still the possibility of targeting a large number of genes at once. The use of the DNA binding domain of Fast-1 as a determinant of specificity seems reasonable in that constructs containing this element produced off-target effects that should be manageable through proper dosing. Since Smad2 plays a much more dominant role in general cell signaling, interference with Smad2 signaling has a greater potential for off-target effects. The idea of generating a construct which binds Smad2 at an unessential site holds great hope for success. This construct would not require the injection of additional Smad2, and would hopefully not alter its signaling so long as the protein translated from injected mRNA is small enough not to block other binding sites on Smad2. Moreover, a great deal of information is being lost through the current genotyping strategy in which many target genes are simply not being assessed. With a better detection method, one may discover the generation of additional mutations. Upon the generation of mutations, this method would demonstrate proof of principle and holds the potential for learning more about the roles of Nodal in early development and the discovery of new Nodal-regulated genes.

Bibliography

Chen YG, Wang Q, Lin SL, Chang CD, Chung J, Ying SY. 2006. Activin Signaling and Its Role in Regulation of Cell Proliferation, Apoptosis, and Carcinogenesis. *Exp. Biol. Med.* 231:534-44.

Chen X, Weisberg E, Fridmacher V, Watanabe M, Naco G, Whitman M. 1997. Smad4 and FAST-1 in the assembly of activin-responsive factor. *Nature* 389:85-89.

Dougan ST, Dougan ST, Warga RM, Kane DA, Schier AF, Talbot WS. The role of the zebrafish nodal-related genes squint and cyclops in patterning of mesendoderm. *Development* 130(9):1837-51.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792-7.

Erali M, Wittwer CT. 2010. High resolution melting analysis for gene scanning. *Methods.* 50(4):250-61. Epub 2010 Jan 18.

Halbritter J, Diaz K, Chaki M, Porath JD, Tarrier B, Fu C, Innis JL, Allen SJ, Lyons RH, Stefanidis CJ, Omran H, Soliman NA, Otto EA. 2012. High-throughput mutation analysis in patients with a nephronophthisis-associated ciliopathy applying multiplexed barcoded array-based PCR amplification and next-generation sequencing. *J. Med. Genet.* 49:756-767.

Hale LA, Tallafuss A, Yan YL, Dudley L, Eisen JS, Postlethwait JH. 2006. Characterization of the retinoic acid receptor genes *raraa*, *rarab* and *rarg* during zebrafish development. *Gene Expression Patterns* 6:546–555.

Händel EM, Cathomen T. 2011. Zinc-Finger Nuclease Based Genome Surgery: It's All About Specificity. *Current Gene Therapy* 11:28-37.

Lawson ND, Wolfe SA. 2011. Forward and Reverse Genetic Approaches for the Analysis of Vertebrate Development in the Zebrafish. *Dev. Cell* 21:48-64.

Jares-Erijman EA, Jovin TM. 2003. FRET Imaging. *Nat Biotechnol.* 21(11):1387-95.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. *Dev. Dyn.* 203(3):253-310.

Liu Z, Lin X, Cai Z, Zhang Z, Han C, Jia S, Meng A, Wang Q. 2011. Global Identification of SMAD2 Target Genes Reveals a Role for Multiple Co-regulatory Factors in Zebrafish Early Gastrulas. *J. Biol. Chem.* 286(32):28520-32.

Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nature Biotechnology* 26(6):695-701.

Müller F, Blader P, Rastegar S, Fischer N, Knöchel W, Strähle U. 1999. Characterization of zebrafish *smad1, smad2* and *smad5*: the amino-terminus of Smad1 and Smad5 is required for specific function in the embryo. *Mechanisms of Development* 88:73-88.

Nagaso H, Suzuki A, Tada M, Ueno N. 1999. Dual specificity of activin type II receptor ActRIIb in dorso-ventral patterning during zebrafish embryogenesis. *Develop. Growth Differ.* 41:119-133.

Peyriéras N, Strähle U, Rosa F. 1998. Conversion of zebrafish blastomeres to an endodermal fate by TGF-beta-related signaling. *Curr. Biol.* 13:783-6.

Renucci A, Lemarchandel V, Rosa F. 1996. An activated form of type 1 serine/threonine kinase receptor TARAM-A reveals a specific signaling pathway involved in fish head organizer formation. *Development* 12:3735-43.

Rozen S, Skaletsky HJ. 2000 Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386.

Schier AF. 2003. Nodal Signaling in Vertebrate Development. *Annu. Rev. Cell Dev. Biol.* 19:589-621.

Schier AF, Talbot WS. 2001. Nodal signaling and the zebrafish organizer. *Int. J. Dev. Biol.* 45:289-97.

Silvestri C, Narimatsu M, Both IV, Liu Y, Tan NBJ, Izzi L, McCaffery P, Wrana JL, Attisano L. 2008. Genome-Wide Identification of Smad/Foxh1 Targets Reveals a Role for Foxh1 in Retinoic Acid Regulation and Forebrain Development. *Dev. Cell* 14:411-412.

Sirotkin HI, Gates MA, Kelly PD, Schier AF, Talbot WS. 2000. Fast1 is required for the development of dorsal axial structures in zebrafish. *Curr. Biol.* 10:1051-54.

Swanhart LM, Takahashi N, Jackson RL, Gibson GA, Watkins SC, Dawid IB, Hukriede NA. 2010. Characterization of an lhx1a transgenic reporter in zebrafish. *Int. J. Dev. Biol.* 54(4):731-6.

Szczepek M, Brondani V, Büchel J, Serrano L, Segal DJ, Cathomen T. 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature Biotechnology* 25(7):786-793.