Developmental Biology 326 (2009) 456–470

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Genomes & Developmental Control

Functional analysis of the evolutionarily conserved *cis*-regulatory elements on the *sox17* gene in zebrafish

Tzu-Min Chan [a,b], Chung-Hao Chao [a,b], Horng-Dar Wang [b], Yen-Ju Yu [a], Chiou-Hwa Yuh [a,c,d,*]

[a] Division of Molecular and Genomic Medicine, National Health Research Institutes, 35 Keyan Road, Zhunan Town, Miaoli County 350, Taiwan, ROC
[b] College of Life Science and Institute of Biotechnology, National Tsing-Hua University, Hsin-Chu, Taiwan, ROC
[c] College of Life Science and Institute of Bioinformatics and Structural Biology, National Tsing-Hua University, Hsin-Chu, Taiwan, ROC
[d] Department of Biological Science and Technology, National Chiao Tung University, Hsin-Chu, Taiwan, ROC

ARTICLE INFO

Article history:
Received for publication 30 May 2008
Revised 31 October 2008
Accepted 11 November 2008
Available online 24 November 2008

Keywords:
Sox17
Sox32
Pou5f1
Transcription regulation
Evolutionarily conserved non-coding region
Zebrafish endoderm development
Gene Regulatory Networks (GRNs)

ABSTRACT

The Sox17 is an important transcription factor for endodermal cells (*Danio rerio*). According to the predictions of the GRNs, based on perturbation experiments and literature search, the *sox17* gene is engaged with two other regulatory genes, *sox32* and *pou5f1*. Nodal signaling operated on several endoderm-specific transcription factors to determine the endoderm specification. In addition, endoderm specification requires the Fgf and Bmp signaling pathways to be repressed in the cells which will become endoderm. It is predicted that Nodal activates *sox32* and works synergistically with Pou5f1 to activate *sox17*. Bmp represses the expression of *sox17* on the ventral side and Fgf represses it on the dorsal side. The regulatory inputs of *sox17* at the genomic sequence level are not known. Here, we have uncovered the relevant *sox17 cis*-regulatory elements, and examined the specific input predictions of the GRNs. We discovered three conserved modules, A, B, and C, with a synergistic effect among them. We revealed that the Pou5f1-binding element on the B module and the Sox32-binding element on the C module work synergistically. Furthermore, an evolutionarily non-conserved *R* module exhibits a repressive effect on both the ventral and dorsal side. We have directly demonstrated the structural and functional relationships of the genomic code at this key node of the endoderm GRNs in zebrafish development. This information provides new insight into the complexity of endoderm formation and serves as a valuable resource for the establishment of a complete endoderm gene regulatory network.

Introduction

Zebrafish (*Danio rerio*) has long been a model organism for developmental biology (Schoenebeck and Yelon, 2007; Zacchigna et al., 2008; Zorn and Wells, 2007) and the mechanisms for axis formation (Schier and Talbot, 2005), endoderm differentiation (Alexander and Stainier, 1999; Zorn and Wells, 2007) and muscle development (Ochi and Westerfield, 2007) have been studied. Development is controlled by the coordinated regulation between signaling pathways and transcription factors. Of all the signaling pathways, the Nodal pathway is the most important for determining endoderm formation in different species. The Nodal pathway is conserved among zebrafish (Le Good et al., 2005; Schier and Talbot, 2001), *Xenopus* (Zhou et al., 1993), chicken (*Gallus gallus*), and human in endoderm and mesoderm differentiation; and is associated with vertebrate gastrulation and axial patterning (Jones et al., 1995). By cell transplantation, the activation of *nodal* signaling is sufficient to commit cells to both endodermal fate and behavior in zebrafish (David and Rosa, 2001). Sustained Nodal signaling is required to ensure endoderm formation, but transient Nodal signaling is sufficient for mesoderm formation (Aoki et al., 2002). The interaction of Nodal and its receptor initiate phosphorylation and activation of Smad2/3 to bind to Smad4. The Smad2/3/4 complex in combination with other specific partners like FoxH1 and Bon, allows the transcriptional activation of some specific target genes (Kunwar et al., 2003). The detailed gene regulatory network (GRN) operated by Nodal has been intensively studied.

Nodal signaling operates several endoderm-specific transcription factors, including *bon* (*bon*/*clyde*), *gata5* (*gata5*/*faust*; *fau*), *og9x*, *pou5f1*, *sox32* (*sox32*/*casanova*; *cas*) and *sox17* (Schier and Talbot, 2001). Mutation of those genes caused an endodermal tissue defect in zebrafish (Poulain and Lepage, 2002). Embryos deficient in *sox32* lack endoderm and develop cardiac bifida (Kikuchi et al., 2001). *Bon*, *gata5*, and *sox32* were identified downstream of Nodal signaling but upstream of the early endodermal marker *sox17*. *Sox32*, expressed in the endoderm, was cloned in a subtractive screen for *Nodal*-responsive genes (Dickmeis et al., 2001; Sakaguchi et al., 2001). The knockdown of *sox32* using a morpholino antisense oligonucleotide results in a lack of *sox17* expression in endodermal precursor cells during gastrulation. The over-expression of *sox32* restores endoderm markers in the

* Corresponding author. Division of Molecular and Genomic Medicine, National Health Research Institutes, 35 Keyan Road, Zhunan Town, Miaoli County 350, Taiwan, ROC. Fax: 886 37 586459.
E-mail address: chyuh@nhri.org.tw (C.-H. Yuh).

0012-1606/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2008.11.010

absence of Nodal signaling (Aoki et al., 2002). This revealed that *sox32* is located downstream of Nodal signaling and upstream of *sox17*.

The activation of *sox32* depends upon *gata5*, *bon*, and *eomesodermin* (Reiter et al., 2001). *Gata5* is expressed in the endodermal progenitors from the late blastula stages. The *gata5* mutant expresses lower levels of *sox17* and *foxa2* than wild-type. Using a complementary mutant and over-expression analysis, it was shown that *gata5* and *bon* are required for *sox32* activation in endoderm formation (Reiter et al., 2001). Eomesodermin, a maternal T-box protein, acts with *gata5* and *bon* to regulate the expression of *sox32* via an element located 1476 bp upstream of the translational initiation site of *sox32* (Bjornson et al., 2005). The regulation of endoderm formation by *eomesodermin* does not require Nodal signaling or *og9x* (Bjornson et al., 2005), which is a novel paired-like homeobox protein that depends on a functional Nodal signaling pathway. It was demonstrated that *og9x*, *bon*, and *sox32* are all immediate early targets of Nodal signaling. Expressing a constitutively active Nodal receptor in the presence of translation inhibitors showed that *sox17* requires protein synthesis in order to be induced (Poulain and Lepage, 2002). These results highlight the complexity of the transcriptional network operating during endoderm formation.

In addition to Nodal signaling, Bmp and Fgf have important roles in determining the expression of those endoderm-specific transcription factors. The Bmp pathway suppresses the endoderm specification on the ventral side as demonstrated by over-expression of the *bmp2*, *bmp4*, and *bmp7* cocktails that reduced endoderm formation on the ventral side (Poulain et al., 2006). The Fgf pathway suppresses the endoderm specification on the dorsal side, where the activation of Fgf signaling decreases the number of *sox32*-expressing endodermal cells (Mizoguchi et al., 2006). Conversely, inhibition of Fgf signaling increases the number of endodermal cells without affecting the expression of Nodal. The inhibition of Fgf signaling in endoderm mutants suggests that this signaling negatively regulates *sox32* expression by a pathway parallel to *bon* and *gata5* in the molecular cascades leading to endoderm formation. The largest number of endodermal precursors was observed when a noggin mutant inhibited Bmp signaling and Fgf signaling by the morpholino antisense oligonucleotides (Poulain et al., 2006). These results demonstrate that the Nodal, Fgf and Bmp signaling pathways are antagonistic, which ensures that at the margin most cells become endoderm cells.

Although there have been many studies of zebrafish development, there are very few studies regarding gene regulation networks in zebrafish. Nevertheless, considerable numbers of zebrafish maternal and zygotic regulatory factors and signaling pathways have been discovered. We have recently constructed the zebrafish GRNs (http://www.zebrafishGRNs.org) (Chan et al., 2008) and integrated the evidence of interaction in the literature and the spatial/temporal expression patterns extracted from the Zebrafish Information Network (ZFIN) (Sprague et al., 2008). We also established the network architecture on the basis of the results of morpholino oligonucleotide perturbation coupled with expression profiles in real time RT-PCR and whole-mount *in situ* hybridization (Yuh et al., unpublished data). The work on zebrafish GRNs is still an ongoing project; and direct evidence of interaction is often missing. Nevertheless, the GRNs can be experimentally verifiable at the DNA level by *cis*-regulatory analysis and direct testing for the existence of functional modules. The authenticity of the GRNs comes from verification of the key nodes by *cis*-regulatory analysis (Davidson and Levin, 2005; Davidson et al., 2003). Such verification provides an opportunity to confirm and correct the prediction of the GRNs, and to produce a direct structural and functional explanation of development in terms of the DNA regulatory system.

The objective of this study is to identify the node of gene regulatory subcircuits of endoderm formation. A node of GRNs is defined as a gene and relevant *cis*-regulatory modules that receive inputs from elsewhere in a network, and provide outputs destined for targets elsewhere in a network (Davidson, 2005). Here, we focused on the node of the *sox17* gene engaged with two other regulatory genes, *sox32* and *pou5f1*. We also tested the hypothesis that the Bmp and Fgf signaling pathways repress the transcription level of *sox17* through a specific repressive module. Using comparative genomics method, lost of function, gain of function, a direct mutagenesis study, co-injection of reporter constructs and the morpholino antisense oligonucleotides against specific transcription factors, we tested the *cis* and *trans* alteration directly. Thus, we have identified the important nodes of *sox17* in the gene regulatory networks of endoderm formation. This discovery provides a foundation for establishing a complete endoderm gene regulatory network in zebrafish.

Materials and methods

Zebrafish strains

Adult zebrafish were maintained at 28.5 °C, and embryos were maintained and staged as described previously (Westerfield, 1995). Strain AB zebrafish were purchased from Zebrafish International Resource Center (ZIRC at the University of Oregon, Eugene, OR). Natural spawning from strain AB provided the embryos used in this study.

DNA constructs and site-directed mutagenesis

The genomic DNA for *sox17* was obtained either from zebrafish genomic DNA or *sox17* BAC clone CH211-24F23, which was purchased from the BACPAC Resource Center (BPRC) at the Children's Hospital Oakland Research Institute, Oakland, CA. The genomic sequence of *sox17* was extracted from the Ensembl Genome Browser Database (http://www.ensembl.org/Danio_rerio/index.html). The conserved *cis*-regulatory regions on the *sox17* genome between several species were identified in the UCSC Genome Browser Database (http://genome.ucsc.edu/cgi-bin/hgGateway). To obtain the genomic sequence for our constructs, the BAC DNA was amplified with two primers (Forward, F and Reverse, R) containing restriction enzyme sites for ligation into the pEGFP-N1 vector (BD Biosciences, San Jose, CA). The genomic DNA insert was amplified by PCR, digested with restriction enzyme, gel-purified and inserted into the EGFP-N1 vector. The primer sequences are given below with restriction enzyme cutting sites underlined.

Primer name	Sequence	Restriction enzyme
Promoter-*sox17*-F−1712	5′-GATTGGTACCTTTGAATGCTAACATGTGAGGCA	KpnI
Promoter-*sox17*-F−1213	5′-GATTGGTACCCATGAGTCAGACGGAAGTTGTAT	KpnI
Promoter-*sox17*-F−660	5′-GATTGGTACCAACCTATTAAACTCCTTCAAAGC	KpnI
Promoter-*sox17*-F−350	5′-GATTGGTACCGTGCTATGATAAACGTTACCAAC	KpnI
Promoter-*sox17*-R	5′-GATTGGATCCATCCACAGTGAAAGTTTCAGGC	BamHI
sox17 Region A–F	5′-GATTCTCGAGTGTCAAATAGGTCACTGATAGGT	XhoI
sox17 Region A–R	5′-GATTGAATTCTTATCTGCCTCCTAAGTCGTTCT	EcoRI
sox17 Region B–F	5′-GATTGAATTCACTCGAACTAGTGACCTTCTTG	EcoRI
sox17 Region B–R	5′-GATTGTCGACGTGAGCCATAACCATTTTAACAG	SalI
sox17 Region C–F	5′-GATTGTCGACTTGTATATGGGAGAGGTCTTCCA	SalI
sox17 Region C–R	5′-GATTGGTACCTGCGTGTGTAAGTGGTGAGTGT	KpnI
EGFP-N1 r (polyA)	5′-ATTGTCTCATGAGCGGATACATAT	

Site-directed mutagenesis was performed using the QuickChange Site-Directed Mutagenesis kit from Stratagene. This method utilizes *PfuTurbo* DNA polymerase, which replicates both plasmid strands with high fidelity without displacing the mutant oligonucleotide primers. A 50 ng sample of template DNA was used in each reaction. Following temperature cycling (95 °C for 30 s, then 18 cycles of 95 °C for 30 s, then 55 °C for 1 min, and 68 °C for 12 min), the product was treated with DpnI. The nicked vector DNA containing the desired mutations was then transformed into XL2-Blue ultracompetent cells. The oligonucleotides used to generate the mutations are given below.

The underlined sequences are the 5′ and 3′ parts of the primer sequence. The internal sequences were deleted after site-directed mutagenesis.

(A)del-L 5′ GATGAGTTTGGGTATGGATGGATAATGGTGTAAGAT 3′
(A)del-R 5′ ATCTTACACCATTATCCATCCATACCCAAACTCATC 3′
GATGAGTTTGGGTATGGAGGTTGCCAGGAGAATGGGACTTGCCTGAC-TGCATTGTGCCAATTGGTATGTTTGGTGGTGGATAATGGTGTAAGAT

(B)del-L 5′ ATCTGTACAGCCTTCATTAATCATAATACTTCATAATATA 3′
(B)del-R 5′TATATTATGAAGTATTATGATTAATGAAGGCTGTACAGAT 3′
ATCTGTACAGCCTTCATTAAAATAGAGAATTGTTCTTTATTTTATAATG-CTGCTTCATTTTAGAAAAAAAATCATAATACTTCATAATATA

(C1)del-L 5′ AGCCTGCATTTCCTGTAGACAAACACACCCTC 3′
(C1)del-R 5′ GAGGGTGTGTTTGTCTACAGGAAATGCAGGCT 3′
AGCCTGCATTTCCTGTCCCCCACAAACTGCTTCCTGTCCTTGGCGTTG-GAGCACATACTGTCCGCCTTCAGACAAACACACCCTC

(C2)del-L 5′ ATACACACATACTGTCAGCATTGGCTCCTGAG 3′
(C2)del-R 5′ CTCAGGAGCCAATGCTGACAGTATGTGTGTAT 3′
ATACACACATACTGTCGCCAGGCAAGCCTTTGTGGGGAATTACACAA-TAGCATTGGCTCCTGAG

PCR primers for fine deletion constructs are given below.

Primer	Sequence
B-*sox17-pou5f1*-F	5′ GCCTTCATTAAAATAGAGAATTGT 3′
B-*pou5f1*-F	5′ ATTTTATAATGCTGCTTCATTTTAG 3′
B-0-F	5′ AGAAAAAAAATCATAATACTTCATAA 3′
C-*sox2-otx-sox17*-F	5′ AAGCCTTTGTGGGGAATTACAC 3′
C-*otx-sox17*-F	5′ GGAATTACACAATAGCATTGGCT 3′
C-0-F	5′ AGCATTGGCTCCTGAGTGCCA 3′

Morpholino antisense oligonucleotide injections

The morpholino oligonucleotides (MO) were injected at different dosages according to the titration results where we injected different amounts of MO and choose the dosage which generated most alive embryos with phenotype (Yuh et al., unpublished data). Injections were done into the yolk at the one-cell or two-cell stage. The sequence and the amount of each MO used for injection are given below.

Gene	Morpholino sequence	Dose (ng)	Stock (μg/μl)	MO (μl) per 10 μl	Injected (nl)
bon	5′ CACGACTGCCATTGTGCTGCTGTCC 3′	4.6	4	5	2.3
pou5f1	5′ CGCTCTCTCCGTCATCTTTCCGCTA 3′	4.6	4	5	2.3
otx2	5′ TTGTTTGCGTCTTCAGCGGTGGAGG 3′	2.3	4	2.5	2.3
gata5	5′ ATCCAGTGAATAAGCTAGATTTCGA 3′	6.9	4	7.5	2.3
gata6	5′-AGCTGTTATCACCCAGGTCCATCCA-3′	4.6	4	5	2.3
sox17	5′ CGCATCGGGACTGCTCATCTCAAAC 3′	6.9	4	7.5	2.3
sox32	5′ CGGTCGAGATACATGCTGTTTTGCG 3′	4.6	4	5	2.3

Microinjection and microscopic photography

The embryos were injected using either PCR-amplified DNA from the construct with the specific forward and reverse primers (EGFP-N1 r (poly(A) that contains the SV40 poly(A) signal) or the linearized constructs formed by digestion with XhoI.

For microinjection, the MOs or DNAs were prepared in PBS with 0.05% (w/v) phenol red. Embryos were injected at the one-cell stage with different amounts of MOs or 2.3 nl of 25 ng/μl Sox17-green fluorescent protein (GFP) constructs by Nanoject II (Drummond Scientific Co., Broomall, PA). Embryos were collected at different stages for GFP visualization and photography under a Leica DMIRB inverted fluorescence microscope coupled with a CoolSNAP™ Cooled CCD camera (Roper Scientific, Trenton, NJ).

RNA extraction and real time RT-PCR

Fifty embryos of *sox17* morphants or control morphants were collected at 5, 8, 11, 16, and 24 h postfertilization (hpf). RNA was extracted with an RNeasy Mini kit (catalog #74106, QIAGEN Co. Valencia, CA). Complementary DNA was synthesized by MultiScribe™ Reverse Transcriptase (Applied Biosystems, Foster City, CA) and real time PCR was carried out with Sybr Green (Applied Biosystems) in an ABI PRISM 7900 sequence detection system (Applied Biosystems). After normalization with *18S* RNA or actin, the expression ratio between the experiment and the control group was calculated. The data are shown as mean ± standard error.

Double fluorescence in situ hybridization

One cell stage embryos were injected with A-Bp-GFP, B-Bp-GFP, C-Bp-GFP and ABCR-Bp-GFP, then embryos were collected at 6 hpf. *Foxa2* and *sox17* riboprobes were labeled with digoxigenin, while GFP was labeled with Dinitrophenyl. The reagents used in the double fluorescence *in situ* hybridization are as follows: Sheep anti-DIG-POD (Roche #1 207 733), anti-DNP-POD (PerkinElmer # NEL747B), TSA-Fluorescein system (PerkinElmer NEL753) and TSA-Plus Cyanine 3 system (PerkinElmer NEL753). The protocol was as previously described (Julich et al., 2005; Ma and Jiang, 2007). Images were taken with a fluorescence microscope (Nikon Digital Camera Dxm1200).

Over-expression sox32 and pou5f1

The full-length *sox32* and *pou5f1* cDNAs were amplified and separately cloned into a yT&A vector. The sequence of the amplify primers are given below.

Primer	Sequence
sox32-F	5′-CGCAAAACAGCATGTATCTCGA
sox32-R	5′-CAGTCATCATTCAAGTGTTCATC
pou5f1-F	5′-TCTTGTTTGAAATCTCAACAACC
pou5f1-R	5′-AGCAAGTTAGCTGGTGAGATGA

The mMESSAGE mMACHINE T7 Ultra kit (Ambion, catalog #1345) was used for *in vitro* transcription from the cloned cDNAs. One-cell stage embryos were co-injected with 2.3 nl of full-length *sox32* or *pou5f1* RNA (25 ng/μl) and *sox17*-green fluorescent protein (GFP) constructs.

Electrophoretic mobility shift assays

Nuclear extracts were prepared from 5 hpf embryos by Nuclear Extraction Kit (Panomics, catalog #AY2002). 5 μg of protein were incubated with biotin-labelled (biotin label on 5′ end of the forward primer) double stranded oligonucleotides at 4 °C for 30 min. Protein extracts were incubated first with mismatch double-stranded oligonucleotides for 10 min in a competition experiment. DNA/protein complexes were separated by 10% PAGE as previously described (Yuh et al., 2004), and then transferred onto nitrocellulose (NC) paper. A chemiluminescent Nucleic Acid Detection Module kit (PIERCE, catalog #89880) was used for staining and exposure. The sequences of the probes are given below.

Primer	Sequence
B-pou5f1 site-F	5′-ATAGAGAATTGTTCTTTATT
B-pou5f1 site-R	5′-AATAAAGAACAATTCTCTAT
B-Mismatch-F	5′-ATAGAGACGGTGGCTTTATT
B-Mismatch-R	5′-AATAAAGCCACCGTCTCTAT

Chromatin immunoprecipitation

The full length cDNA of *sox32* was cloned, and then a polyclonal antibody was generated by LTK BioLaboratories. The Sox32 cDNA was cloned in the pREST vector fused with 6-His-tag, the tag was used to purify the Sox32 protein and then enzymatic removal of the tag was done before generate the antibody. About 400 zebrafish embryos at 8 hpf and the EZ-Magna ChIP™ G kit (catalog #17-409) from Upstate Corp. were used for the immunoprecipitation experiment. The PCR primers used to check the Chromatin immunoprecipitation are given below.

Primer	Sequence
B site-F	5′-ACTCGAACTAGTGACCTTCTTG
B site-R	5′-GTGAGCCATAACCATTTTAACA
C site-F	5′-CCTCTCCCTCACAGTCAAATCT
C site-R	5′-TGTGGACGCATATACAACTTCC
A–B space-F	5′-TGGTGCCTCACATGTTAGCATT
A–B space-R	5′-TTGTTCTACTGACTCTGACCC
B–C space-F	5′-ACATTGCAAGAACAGCGTCCAA
B–C space-R	5′-GCTGTATCTACTGGCAGGATAT

Results

Identification of the functional modules regulating sox17 in zebrafish development

Cis-regulatory elements are modular in structure and are functionally important with many transcription factor-binding sites. A DNA sequence conserved among different genomes often indicates its importance; providing a very powerful method to identify the important regulatory regions of a gene (Chen and Blanchette, 2007; Werner et al., 2007). To search for the functional *cis*-regulatory modules of the *sox17* gene, we first compared the genomic sequence of *sox17* to orthologous genomic regions from nine different species and identified three conserved non-coding regions spanning the *sox17* locus (Fig. 1A). They are named module A, B and C, from the 5′ end towards the transcription start site (Fig. 1A). To be specific, module A is conserved between zebrafish and *Tetraodon*, module B is conserved between zebrafish and human, and module C is conserved between zebrafish and fugu. To check the functions of these predicted modules, GFP reporter constructs containing different regions of the *sox17* genomic DNA were generated (Fig. 1B). For each construct, three different batches of embryos were injected with linearized DNA, the GFP expression patterns were observed at different stages, and the percentage of GFP expression are shown in Table 1. A typical expression pattern for each construct was selected and is shown in Figs. 2A and B.

Firstly, four constructs (1712-GFP, 1213-GFP, 660-GFP and 350-GFP (i.e. Bp-GFP)) were made and are shown in Fig. 1B to identify the proximal regulatory element. Comparing the GFP expression from the four deletion mutants, we identified the basal promoter, Bp-GFP (Bp stands for basal promoter), spanning 350 bp upstream to 133 bp downstream of the transcriptional start site. About 50–60% of the injected embryos carrying the Bp region express GFP weakly in the embryos at 5–8 hpf (Table 1). We observed fewer embryos with GFP expression for the Bp construct at 11 hpf and no activity at 16 hpf. The 660-GFP construct containing more upstream sequence (to −660) displayed no GFP expression at the early stage (up to 8 hpf), and similar expression to Bp at 11 hpf and at 16 hpf. This indicates that the sequence between positions −660 and −350 contains a suppressive module for the GFP expression in the early stage compared to Bp (Table 1). This sequence was named as the R module, denoting its repressive function. The repressive module is not evolutionarily conserved, since it was not found from the cross-species comparison. The construct 1213-GFP contains the DNA region from −1213 to +133, which includes the entire R module and 96 bp of the C module. After injection of 1213-GFP into embryos, we did not observe GFP expression in the early stage (up to 8 hpf, Table 1). However, the additional sequence increased the GFP expression from 6% to 12% at 11 hpf, and to 36% at 16 hpf (Table 1, third row). The longer construct 1712-GFP containing the region −1712 to +133 did not express GFP in the early stage (up to 11 hpf, Table 1), and showed yet more embryos expressing GFP at a later stage, about 15% at 11 hpf and 32% at 16 hpf (Table 1, the fourth row). With the presence of the R module, the intensity of GFP in the embryos was much weaker, and the early expression of the basal promoter was repressed completely. The results indicated the repression function occurs from the late blastula to early gastrula stage. The constructs containing module C (1213-GFP or 1712-GFP) exhibit higher transcriptional activity in the presence of the repressor module with limited amplitude (Table 1), indicating that module C may provide some anti-repressive activity. Thus, we defined the region −350 to +133 as the basal promoter (Bp), and sequence between positions −660 and −350 as suppressive module for *sox17*. The other constructs were designed by adding different lengths of the regulatory elements to the basal promoter.

The synergistic effect among the A, B and C modules

To understand the function of each module, the A, B, or C module was linked to the basal promoter and GFP expression was observed at 5 hpf (Figs. 2A, A–I), 8 hpf (Figs. 3B, A–I), 11 hpf and 16 hpf. To directly address whether the transgenes recapitulate endogenous *sox17* expression and are restricted to the endoderm, we performed double staining with endoderm markers (e.g. *foxa2* and *gfp*, or *sox17* and *gfp*) and the results for various constructs are shown in Fig. 2C. Double fluorescent *in situ* hybridization confirmed that GFP was expressed in the endoderm and some ectoderm lineage cells from each of the three constructs (Figs. 2C, A–I). GFP expression from each of the three constructs was increased compared to that from the basal promoter only (Tables 2 and 3). Interestingly, the A module promoted a stronger regulatory function at later stages: embryos carrying A-Bp exhibited a lower percentage of GFP expression at 5 hpf and at 8 hpf compared to 11 hpf and 16 hpf. The B module is the strongest module among all three; about 60–70% of B-Bp-GFP-injected embryos exhibited strong GFP expression at all four time-points. The C module had a moderate regulatory function, 20–30% of the C-Bp-GFP-injected embryos expressed GFP at the four different time-points. Interestingly, when the B and C modules were present on the same construct (BC-Bp-GFP), the intensity of GFP expression was increased (Table 2), indicating there is a synergistic effect between the B and C modules. The synergistic effect between three conserved modules can be observed from the expression pattern of ABC-Bp-GFP (Tables 2 and 3). We observed a high level of GFP expression at 5 hpf and at 8 hpf, and GFP was expressed in the entire embryo. All embryos injected with the ABC-Bp-GFP died at 11–16 hpf. We do not know the exact reason, and suspect too much ectopic GFP expression interferes with the development of the embryo. GFP expression from the A, B and C module alone, and the BC module was localized to only half of the embryo (Tables 2 and 3). The entire module of A, B and C together activates *sox17* expression efficiently, but with ectopic expression in the ectoderm at 5 hpf and at 8 hpf. This result indicates that the transcription factors binding to the A, B and C modules must interact with each other to enable the synergistic effect.

The repressive effect of the R module together with ABC module

We found a synergistic effect of the enhancer activity with the ectopic GFP expression in the ectoderm from ABC-Bp-GFP injected embryos. This ectopic GFP expression may be due to the lack of

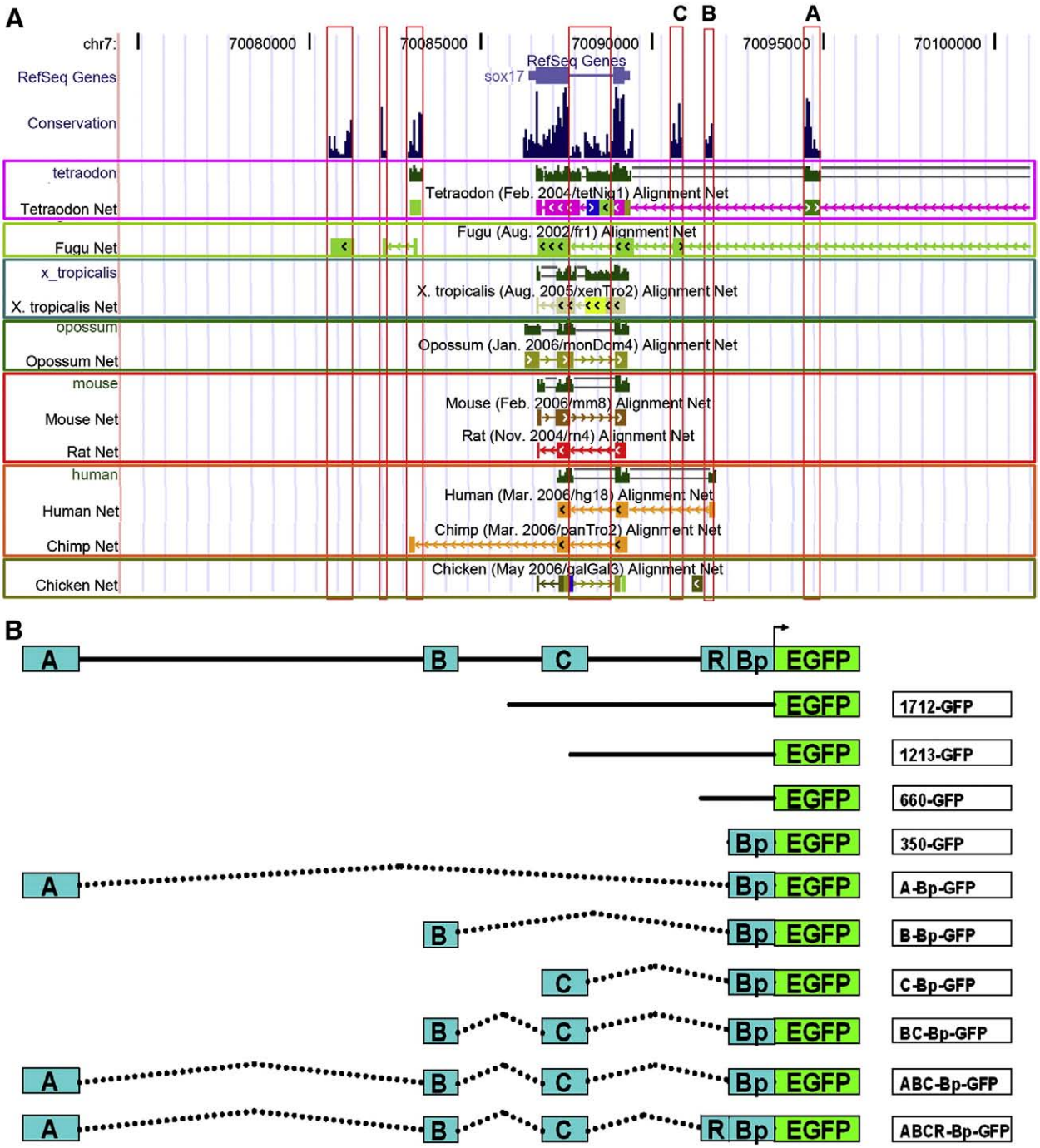


Fig. 1. Sox17 conserved genomic modules and GFP constructs in this study. (A) Conservation of the genomic sequence of the *sox17* locus among different species. The genomic sequence from zebrafish is compared with those from nine other species. The conserved regions are shown as a dark blue histogram in the first row below the *sox17* transcript. The transcription direction of *sox17* is from right to left. The alignment results show that the A module is conserved between *Tetraodon* sp. and zebrafish, the B module is conserved between human and zebrafish, and the C module is conserved between fugu (*Fugu rubripes*) and zebrafish. (B) The ten *sox17* GFP reporter constructs used to study the *sox17 cis*-elements. Four constructs are serial upstream deleted mutants; the number indicates the upstream nucleotide number of transcription start site on the *sox17* gene. Six constructs are the combination of different modules linked to Bp-GFP or Bp-GFP plus R. The light blue boxes labeled with A, B and C are the non-coding regions identified as evolutionarily conserved among seven species. The box labeled with R indicates the repression region, which is not evolutionarily conserved among the seven species, that was defined as a repression function after the expression analysis. There are five domains containing the A module ranging from −5425 to −4979, the B module −2376 to −2117, and the C module −1422 to −1117. The R module is between −660 and −350, and the basal promoter (Bp-GFP) spans −350 to +133.

repressor activity of the R module that suppresses an early ectopic expression from the basal promoter. To examine the role of the R module with the ABC module, we constructed the ABCR module linked to the basal promoter and GFP coding sequence (ABCR-Bp-GFP). The embryos injected with ABCR-Bp-GFP survived after 11 hpf up to 16 hpf, and the ectopic ectoderm expression was suppressed at 50% epiboly stage (Table 2). This result may help to explain the lethality of the ABC-Bp-GFP-injected embryos seen at 11 hpf and at 16 hpf. The GFP expression pattern of ABCR-Bp-GFP was very similar to *sox17* expression at 5 hpf (Figs. 2A, J–L) and 8 hpf (Figs. 2B, J–L).

Double fluorescence *in situ* hybridization analysis is a useful technique for localizing gene expression patterns based on well known marker genes (Hsiao et al., 2007; Zimonjic et al., 1997). Our A-

Table 1
Summary of the GFP expression percentage of the injected embryos

Construct	Average GFP expression rate at different stages (hpf)			
	5	8	11	16
350-GFP	52.6%	59.2%	6%	0%
660-GFP	0%	0%	6.4%	6.5%
1213-GFP	0%	5.8%	11.8%	36%
1712-GFP	0%	0%	15.4%	31.6%
A-Bp-GFP	8.8%	14.8%	51%	59%
B-Bp-GFP	60%	70%	67%	67%
C-Bp-GFP	19%	33%	33%	34%
BC-Bp-GFP	23.4%	54.9%	50%	53.8%
ABC-Bp-GFP	23%	82.3%	Dead	Dead (76>*N*>26)

The first column indicates the construct for each analysis. Four observation stages are 5 hpf, 8 hpf, 11 hpf, and 16 hpf. The portions of injected embryos with GFP expression are shown as percentages. The construct ABC-Bp-GFP-injected embryos were prematurely lethal after 11 hpf. The number within parentheses indicates the total number of embryos injected with each construct from two or three different batches of injection.

Bp-GFP, B-Bp-GFP and C-Bp-GFP constructs showed expression patterns in the half embryo, but we could not pinpoint their location in the embryos. Whether ABCR-Bp-GFP co-localizes with endogenous *sox17* expression needed to be determined. To resolve this question, we used two endoderm markers, *foxa2* and *sox17*, to trace the GFP expression. According to fluorescent double *in situ* hybridization results, we found that the expression from A-Bp-GFP, B-Bp-GFP and C-Bp-GFP co-localized with *foxa2* in the hypoblast of 6 hpf embryos, but was also widely ectopically expressed (Figs. 2C, D, H and L). These results indicated A-Bp-GFP, B-Bp-GFP and C-Bp-GFP expression patterns are on the dorsal side of the embryos and ectopic in the ectoderm. We further confirmed the GFP expression of ABCR-Bp-GFP by double *in situ* hybridization with both endogenous *sox17* (red, Figs. 2C, N) and *GFP* mRNA (green, Figs. 2C, O). We verified that GFP expression from the ABCR modules are co-localized with *sox17* mRNA (organ, Figs. 2C, P). This result implies that the function of the R module is to modulate the synergistic effect of the ABC module to an appropriate level by suppression of ectopic expression during zebrafish development.

The GFP expression patterns of different constructs from the 50% epiboly stage are summarized in Table 2. First, only the ABCR-Bp-GFP construct recapitulated endogenous *sox17* expression. ABC-Bp-GFP was expressed in the whole embryo, in both mesendoderm and ectoderm at the 50% epiboly stage. BC-Bp-GFP was expressed in half of the embryo at 50% epiboly. The individual modules alone (A-Bp-GFP and C-Bp-GFP), except for B-Bp-GFP, were expressed more weakly. All of the constructs without the R module were expressed in the ectoderm. The 660-GFP containing the R module expressed a very low level of GFP compared to the expression pattern obtained from 350-GFP (Bp-GFP). The data indicates that the R module functions as an early repressor to suppress the ectoderm expression. Data of the GFP expression patterns of the ten constructs at four different time-points are summarized in Table 3. Bp is expressed at an early stage, A-Bp-GFP is expressed at a late stage, B-Bp-GFP shows strong expression, and C-Bp-GFP shows a moderate activity at all stages. With all of the positive modules together (ABC-Bp-GFP), the synergistic effect may cause GFP over-expression ectopically, which was reduced by the presence of the R module (ABCR-Bp-GFP).

Quantification of sox17 cis-element regulatory regions by real time RT-PCR

After we revealed the spatial expression pattern of the transgenes, we used real time RT-PCR analyses for those *cis*-elements at different stages. First, we measured the GFP mRNA level generated from Bp-GFP-injected compared to control (uninjected) embryos. Embryos injected with Bp-GFP exhibited a basal level of GFP expression above 1×10^6 GFP RNA molecules per embryo (Fig. S1 A). Next, we analyzed GFP RNA molecules generated from three positive modules alone compared with Bp-GFP. At all stages of development tested, all of three modules (A, B and C) exhibited higher levels of GFP expression than Bp-GFP (Fig. S1 A). The A-Bp-GFP and B-Bp-GFP of GFP expression exhibited about 1×10^7 GFP RNA molecules per embryo, while C-Bp-GFP of GFP expression exhibited about 2×10^7 to 6×10^7 GFP RNA molecules per embryo. This quantified RNA result is consistent with the visualization of GFP under the microscope. Therefore, we confirmed that modules A, B and C of *sox17* function as enhancers. Furthermore, we found the combined BC module had a higher activity than the B or C module alone, except at 11 hpf for the C module. The ABC module showed the highest level of GFP activity, almost 100-fold greater than the basal promoter alone. The quantitative data further confirmed the synergistic effect observed by GFP visualization.

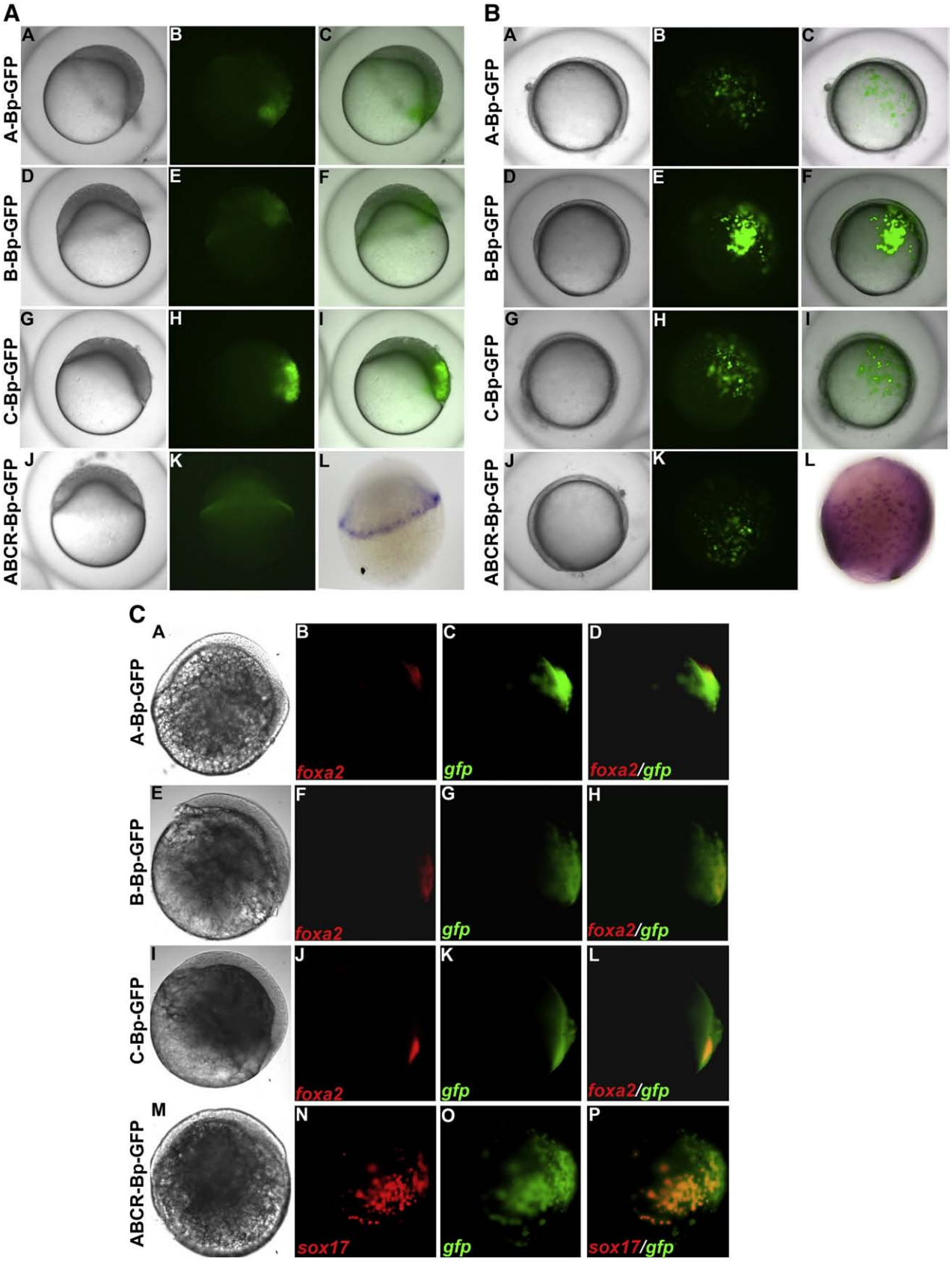
In the endoderm development, Pou5f1 has been known to interact physically with Sox32 to regulate the expression of *sox17* (Reim et al., 2004). An endoderm specification pathway has also been proposed in which Nodal activates *og9x*, *bon*, and *gata5* to activate the expression of *sox32* (Poulain and Lepage, 2002; Reiter et al., 2001). According to our morpholino oligonucleotide injection coupled with real time RT-PCR data, *otx2*, *sox32*, and *gata5* positively regulated *sox17* (Yuh et al., unpublished data). We found *gata6* positively regulated many endoderm-specific transcription factors, including *gata5*, *foxa2*, *foxa3*, and *sox32* (Yuh et al., unpublished data). Therefore, we examined whether those transcription factors work through interaction directly with the C module of *sox17* (Fig. S1 B). We co-injected C-Bp-GFP with the morpholino oligonucleotides against *gata5*, *gata6*, *bon*, *sox32*, or *pou5f1* into zebrafish embryos and then performed real time RT-PCR analysis. As shown in Fig. S1 B, morpholino oligonucleotides against *gata5*, *bon* and *sox32* reduced GFP expression of C-Bp-GFP. This result indicated that Sox32, Bon, and Gata5 might regulate *sox17* positively through interaction with the C module. The morpholino oligonucleotide against *gata6* had no effect on the GFP activity of C-Bp-GFP, indicating that Gata6 does not function in the C module of *sox17*. In fact, the morpholino oligonucleotide against *gata6* has no effect on endogenous *sox17* expression either (Yuh et al., unpublished data). Surprisingly, we found the morpholino against *pou5f1* caused a higher level of GFP expression, indicating that Pou5f1 might prevent the activated GFP expression through the C module of *sox17* by an unknown mechanism.

Mutational analysis of the three modules of the sox17 regulatory region

To further identify the direct transcription factor binding sites on the putative functional modules, we searched for target sites in the highly conserved elements inside the three modules. We identified a 58 bp highly homologous element that is super-conserved between zebrafish and *Tetraodon* sp. in the 459 bp A module region. There are Gsc and Sox17 overlapping binding sites within the 58 bp super-conserved element (Fig. S2). In the 251 bp B module, we discovered a 29 bp super-conserved element containing Sox17 and Pou5f1 binding sites. Two conserved elements were identified in the 309 bp C module region; one is the 40 bp super-conserved element (C1), and the other is the 22 bp highly conserved element (C2) containing Sox2, Otx2 and Sox17 binding sites. To verify the function of those super-conserved elements, we made specific element-deleted GFP constructs for each module and measured GFP expression. The sequences of those three modules and the deletion regions are illustrated in Fig. S2.

We compared the GFP expression of the deletion mutations against the wild-type modules. There is no significant change of GFP expression for the constructs with deletion of the 58 bp

element within the A module [(A) del] or the 40 bp element (C1) in the C module [(C1) del] at the early stage from 5 hpf to 11 hpf (Fig. 3A). In contrast, deletions of the 29 bp element in the B module [(B)del] and the 22 bp element (C2) in the C module [(C2) del] reproducibly abolished GFP expression reproducibly (Figs. 3A, B). This result indicates that the 29 bp element in the B module and the 22 bp element in the C module (C2) are positive regulatory elements.



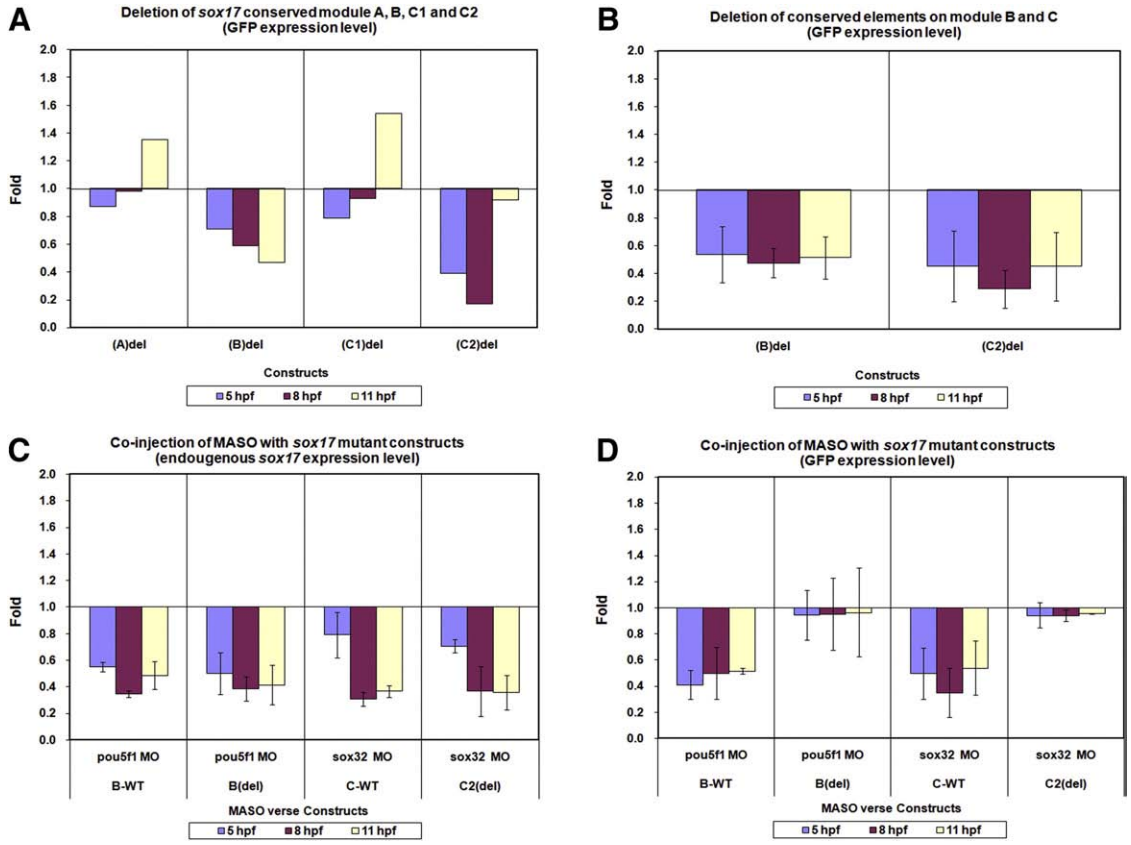


Fig. 3. Identification of the Pou5f1 target site on the B module, and the Sox32 target site on the C module. (A) Quantitative analysis of *sox17*transgenes, and evaluation of the fold change of the mutated constructs. Expression level of four deletion constructs: (A)del, (B)del, (C1)del, and (C2)del are compared to the wild-type *sox17*, A, B, and C at three stages: 5 hpf (blue); 8 hpf (red); and 11 hpf (yellow). The Y axis indicates the fold difference of mutant versus wild-type. (B) The GFP expression difference between (B)del versus wild-type B module, and (C2)del versus wild-type C module. Three different batches of injections were carried out and the standard deviation is indicated as a line extended from the mean. (C) The fold change of endogenous *sox17* expression after injection of the B-Bp-GFP construct with the *pou5f1* MO, and the C-Bp-GFP construct with the *sox32* MO. The fold change of the *sox17* was obtained from compared co-injection with MO versus injection of the construct alone. Three replicates were used to calculate the mean and the standard deviation. (D) The *gfp* expression fold change of the B-Bp-GFP construct co-injected with the *pou5f1* MO, and the C-Bp-GFP construct co-injected with the *sox32* MO. The *gfp* expression levels were normalized by the *gfp* level for injection of the construct alone. The mean result is for three different batches of injections and the standard deviation is indicated as a line extended from the mean.

Since the B element contains a Pou5f1-binding site and the C2 element contains a Sox2-binding site, we further investigated the possibility of Pou5f1 targeting the B element and Sox32 acting through the C2 element. To clarify the *trans*-effect of Pou5f1 and Sox32 on those constructs, we used the same strategy of co-injection of specific morpholino oligonucleotides with the mutated constructs. We injected the *pou5f1* MO with the wild-type B module or the B element-deleted B module, and we injected the *sox32* MO with the wild-type C module or the C2 element-deleted C module. The *pou5f1* MO reduced the endogenous *sox17* expression in both the wild-type and the (B)del module B, and the *sox32* MO reduced the endogenous *sox17* expression in both the wild-type and (C2)del module C (Fig. 3C). The control experiment demonstrated that the effects of the *pou5f1* MO and *sox32* MO are functional in both the wild-type and the mutant. When we determined the GFP mRNA expression using real time RT-PCR, wild-type B-Bp-GFP displayed reduced GFP expression when injected with the *pou5f1* MO compared to the control MO (Fig. 3D). On the contrary, GFP expression was not decreased upon injection of the *pou5f1* MO with the B element-deleted construct (B)del-Bp-GFP. This result demonstrated that the Pou5f1 transcription factor works through the 29 bp B element in the B module. Similarly, the wild-type C-Bp-GFP showed reduced GFP expression when injected with the *sox32* MO compared to the control MO (Fig. 3D). On the other hand, GFP expression was not decreased upon injection of *sox32* MO with the C2 element-deleted construct (C2)del-Bp-GFP, which indicated that Sox32 functions on the C2 element in the C module.

Loss of function assay proved that Sox32 and Pou5f1 directly bind to the genomic region of sox17

A functional node on a DNA element provides the binding site for a specific transcription factor to work. In theory, interference with the functionally important DNA–protein interaction by *cis*-

Fig. 2. (A) GFP expression patterns of different *cis*-element constructs during late blastula stage. The injected embryos were collected at 5 hpf for GFP visualization and photography. The embryos shown in lateral view with dorsal to the right. The A-Bp (A–C), B-Bp (D–F) and C-Bp (G–I) constructs show strong GFP expression in half of the embryo. The ABCR-Bp (J–K) construct shows strong GFP expression in the mesendoderm lineage. (A, D, G, J). The first column is a bright field image, the second column is green fluorescence and third column is a merged picture. (L) Endogenous *sox17* expression pattern at 5 hpf. (B) Embryos injected with *sox17* constructs have different GFP expression patterns during the gastrula stage. The injected embryos were collected at 8 hpf for GFP visualization and photography. The embryos were shown in lateral view with dorsal to the right. The A-Bp (A–C), B-Bp (D–F) and C-Bp (G–I) constructs show strong GFP expression in half of the embryo. The ABCR-Bp (J–K) construct shows strong GFP expression (A, D, G, J). The first column is a bright field image, the second column is green fluorescence and the third column is a merged picture. (L) Endogenous *sox17* expression pattern at 8 hpf. (C) Double fluorescent *in situ* hybridization for *egfp* and endogenous genes The expression pattern for A-Bp-GFP (A–D), B-Bp-GFP (E–H), C-Bp-GFP (I–L), and ABCR-Bp-GFP (M–P) are obtained from embryos at 6 hpf. The *egfp* mRNA are indicated as green (C, G, K and O), *foxa2* (B, F and J) and *sox17* (N) mRNA are shown as red. The merged image (D, H, L and P) indicated the co-localization of *egfp* with endogenous *foxa2* or *sox17* at endoderm cells.

Table 2
Summary of the GFP constructs at 50% epiboly

Construct	GFP expression domain at 5 hpf			
	Whole	Half	Mesendoderm	Ectoderm
350-GFP	−	+	+	−
660-GFP	−	−	−	−
1213-GFP	−	−	−	−
1712-GFP	−	+	+	−
A-Bp-GFP	−	+	++	+
B-Bp-GFP	−	+	++	++
C-Bp-GFP	−	+	++	+
BC-Bp-GFP	−	+	+++	+++
ABC-Bp-GFP	+	−	+++	+++
ABCR-Bp-GFP	+	−	+++	−

The injected embryos were observed for the GFP expression pattern at 50% epiboly stage. The first column indicates the construct for each analysis. Four expression patterns are whole, half, mesendoderm and ectoderm. Whole represents the GFP expressed in the entire embryo; half indicates the GFP expressed in half of the embryo; Mesendoderm denotes the GFP expressed in the margin; and Ectoderm indicates the GFP expressed in the animal pole. The sign indicates the intensity of the GFP signal, −: no expression, +: weak expression, ++: moderate expression, and +++: strong expression.

perturbation that mutates or deletes the target site, or by *trans*-perturbation that blocks the protein binding to this element, should produce the same effect. The equivalent effects of *cis*- and *trans*-perturbations have has been shown on all three of the Otx2 inputs (Yuh et al., 2004). Here, we demonstrate this result for the Pou5f1 and Sox32 inputs on the *sox17* gene.

The functional elements of B and C2 contain more than one transcription factor-binding site. To further dissect the interaction node for Pou5f1 and Sox32, we made more fine deletion constructs for the microinjection illustrated in Fig. 4A. We made three constructs to dissect the B elements; the first contains the Sox17 and Pou5f1 sites (B-sox17-pou5f1), the second contains only the Pou5f1 site (B-pou5f1), and the third removes the Sox17 and Pou5f1 sites (B-0). We also made three constructs to dissect the C2 element, the first contains all three sites Sox2, Otx, and Sox17 (C-sox2-otx-sox17), the second removes the Sox2 site (C-otx-sox17), and the third removes all three sites (C-0).

We first determined the effect of the removal of the different sites on the promoter's activity. As shown in Fig. 4B, if we used the B-sox17-pou5f1 as a control, the removal of Sox17 site (B-pou5f1) seemed to have no significant effect on the reporter GFP expression at 5 hpf or at 8 hpf, and had a slight reduction of GFP expression in the 11 hpf embryos. Further removal of the 23 bp containing the Pou5f1 site (B-0) decreased the GFP expression dramatically, suggesting that the Pou5f1 target site is an important site in the B module for *sox17* expression independent of the neighboring Sox17 site. When compared to C-sox2-otx-sox17, the deletion of the 12 bp containing the Sox2 binding site (C-otx-sox17) decreased the promoter activity dramatically, and further removal of the Otx and Sox17 binding sites (C-0) had no additional effect. These results argue that in the C module, the Sox2 binding site has a dominant role in activating *sox17* expression.

We measured the endogenous *sox17* expression after injection of MOs with various deletion constructs. The *pou5f1* MO and *sox32* MO consistently decrease the endogenous level of *sox17* mRNA (Fig. 4C). However, the deletion construct (B-0) with the removal of the Pou5f1 site cannot respond to the *pou5f1* MO effect, whereas a construct (B-pou5f1) that has 23 bp containing the Pou5f1 binding site responded to the *pou5f1* MO and decreased GFP expression upon injection with the *pou5f1* MO, like the wild-type (Fig. 4D). Similarly, the deletion construct (C-otx-sox17) with the removal of the Sox32 site cannot respond to the *sox32* MO effect, whereas a construct (C-sox2-otx-sox17) that has 12 bp containing the Sox2 binding site responded to the *sox32* MO and decreased GFP expression upon injection with this *sox32* MO, like the wild-type (Fig. 4D).

Gain of function assay for identification of the Pou5f1 and Sox32 functional node on the sox17 module B and C respectively

Previous studies have shown that over-expression of *sox32* and *pou5f1* can induce ubiquitous *sox17* expression, and further promote endodermal specification (Alexander and Stainier, 1999; Lunde et al., 2004; Reim et al., 2004). We used loss of function assay (promoter mutations and morpholino injection) that showed that Sox32 and Pou5f1 can target to the genomic region of *sox17* to regulate its expression. We also performed a gain of function study to further demonstrate Pou5f1 and Sox32 can positively and directly work through module B and C via their binding sites. The full length *pou5f1* and *sox32* cDNAs were separately cloned into the yT&A vector, then the mRNA for both genes were *in vitro* transcribed and microinjected into embryos with various constructs. We analyzed the embryos at 5, 8 and 11 hpf.

Endogenous *sox17* expression reproducibly increased upon injection of the *pou5f1* and *sox32* mRNA (Fig. 5A). Interestingly, we found the *pou5f1* RNA has stronger activation function at 5 hpf, and the *sox32* RNA has stronger activation effect at 8 hpf. Similarly, the GFP expression from B-Bp-GFP was enhanced by co-injection of *pou5f1* RNA, and the GFP level of C-Bp-GFP was increased in the presence of *sox32* RNA (Fig. 5B). These results demonstrated that the *cis*- and *trans*-perturbations produce essentially similar results, and that either loss of function or gain of function studies provides the same evidence for the hypothesis. We have found the direct interaction site for both Sox32 and Pou5f1 on the *sox17* promoter. Furthermore, we tested the direct binding of Sox32 on module C via chromatin immunoprecipitation and Pou5f1 on module B via gel-shift assays.

EMSA and ChIP assays show that Sox17 cis-elements are specifically bound by Pou5f1 and Sox32 in zebrafish embryos

Electrophoretic mobility shift analysis (EMSA) is a commonly used technique for studying the protein–DNA interaction *in vitro* (Muller et al., 2000; Yuh et al., 2004). We tested whether proteins from zebrafish nuclear extract could bind to regions of the *sox17* B module. Nuclear extracts were extracted from 5 hpf zebrafish embryos and then incubated with a biotin-labeled module B probe, covering the Pou5f1 binding site (Fig. S3). A specific competitor, the non-labeled double strand oligonucleotide for the same region, was used to test the specificity. As shown in Fig. S3, two complexes were formed in the presence of 5 hpf nuclear extracts in a dose

Table 3
Summary of the GFP expression at four different stages

Construct	GFP expression intensity at different stages (hpf)			
	5	8	11	16
350-GFP	+	+	−	−
660-GFP	−	−	+	+
1213-GFP	−	−	−	−
1712−GFP	+	+	+	−
A-Bp-GFP	++	+	+	+
B-Bp-GFP	++	++	++	++
C-Bp-GFP	++	+	+	+
BC-Bp-GFP	+++	+++	+++	+++
ABC-Bp-GFP	+++	+++	Dead	Dead
ABCR-Bp-GFP	++	++	++	++

The first column indicates the construct for each analysis. Four observation stages are 5 hpf, 8 hpf, 11 hpf, and 16 hpf. The plus sign indicates the level of GFP intensity observed under the microscope, the more plus signs, the stronger the GFP intensity. Dead means the injected embryos died prematurely.

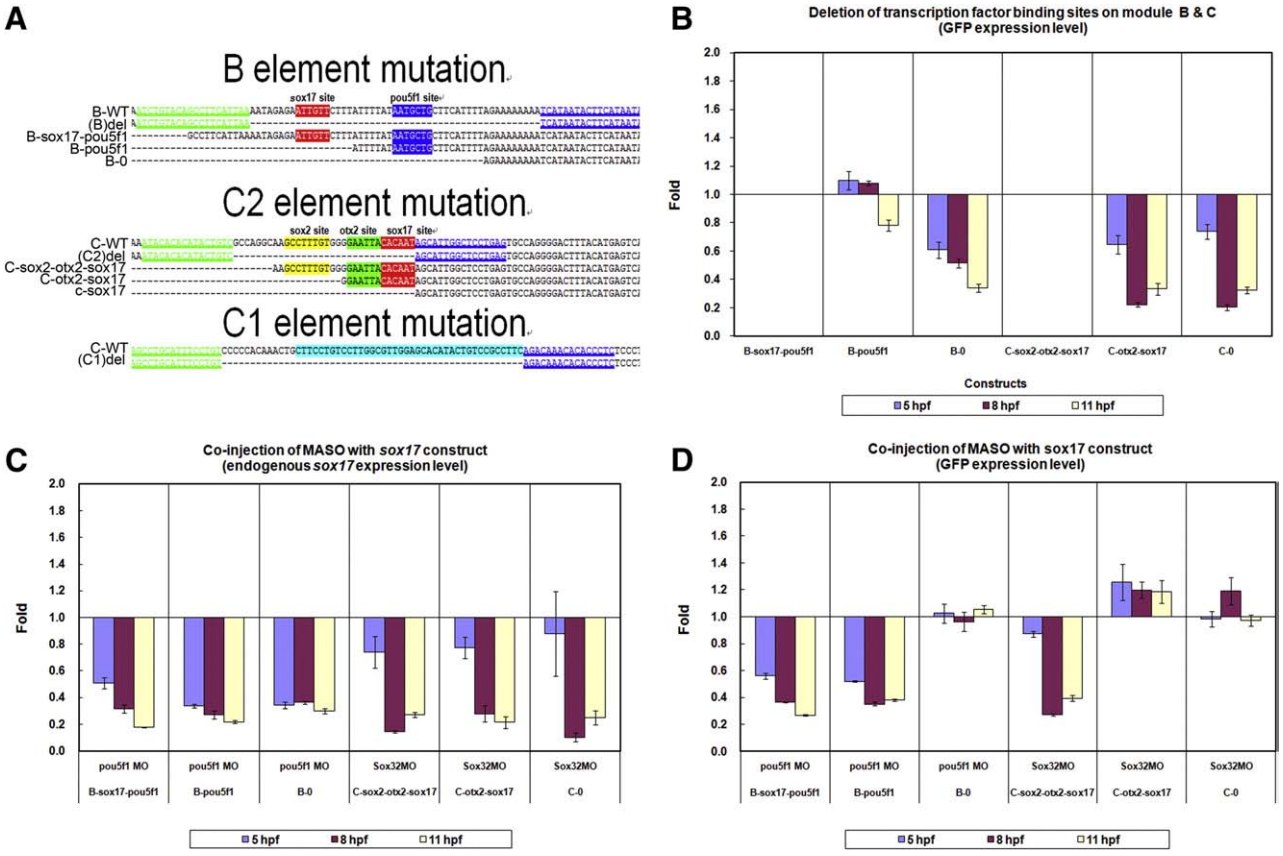


Fig. 4. Functional analysis of transcription factor-binding sites by using gene-specific MO perturbation with real time RT-PCR. (A) The nucleotide sequence of the serial deletion mutants covering the super-conserved region in B, C1, and C2 elements. The construct (B)del lost a 41 bp super-conserved fragment containing the Sox17 and Pou5f1 predicted binding sites. The B-sox17-pou5f1 construct contains the Sox17 and Pou5f1 predicted binding sites, but does not contain sequence 5′ upstream of B-Bp-GFP, and the B-pou5f1 construct contains the Pou5f1 predicted binding site, but without the Sox17 binding site. The B-0 construct has no predicted binding sites. The construct (C2)del lost a 32 bp super-conserved fragment containing the Sox2, Otx2 and Sox17 predicted binding sites. The C-sox2-otx2-sox17 construct contains the Sox2, Otx2, and Sox17 predicted binding sites, but has no sequence 5′ upstream of C-Bp-GFP. The C-otx2-sox17 construct contains the Otx2 and Sox17 predicted binding sites, but has no Sox2 binding site. The C-0 construct has no predicted binding site. The C1(del) construct lost a 53 bp super-conserved fragment. (B) Quantitative analysis of constructs with specific transcription factor binding sites deleted from the *sox17 cis*-element. The expression levels of the six deleted constructs, B-sox17-pou5f1, B-pou5f1, B-0, C-sox2-otx2-sox17, C-otx2-sox17, and C-0, were normalized with B-0 and C-0. The Y axis indicates the fold change after normalization. The different stages are indicated: blue, 5 hpf; red, 8 hpf; and yellow, 11 hpf. (C) The endogenous *sox17* expression fold change of the B-sox17-pou5f1, B-pou5f1, and B-0 constructs injected with the *pou5f1* MO, and the C-sox2-otx2-sox17, C-otx2-sox17, and C-0 constructs injected with the Sox32 MO. The fold changes of the constructs injected with a specific MO were normalized with respect to the injection of the construct alone. (D) The GFP expression fold change of the B-sox17-pou5f1, B-pou5f1, and B-0 constructs injected with the *pou5f1* MO; and the C-sox2-otx-sox17, C-otx-sox17, and C-0 constructs injected with the *sox32* MO. All experiments were done at least three times to calculate the mean and the standard deviation.

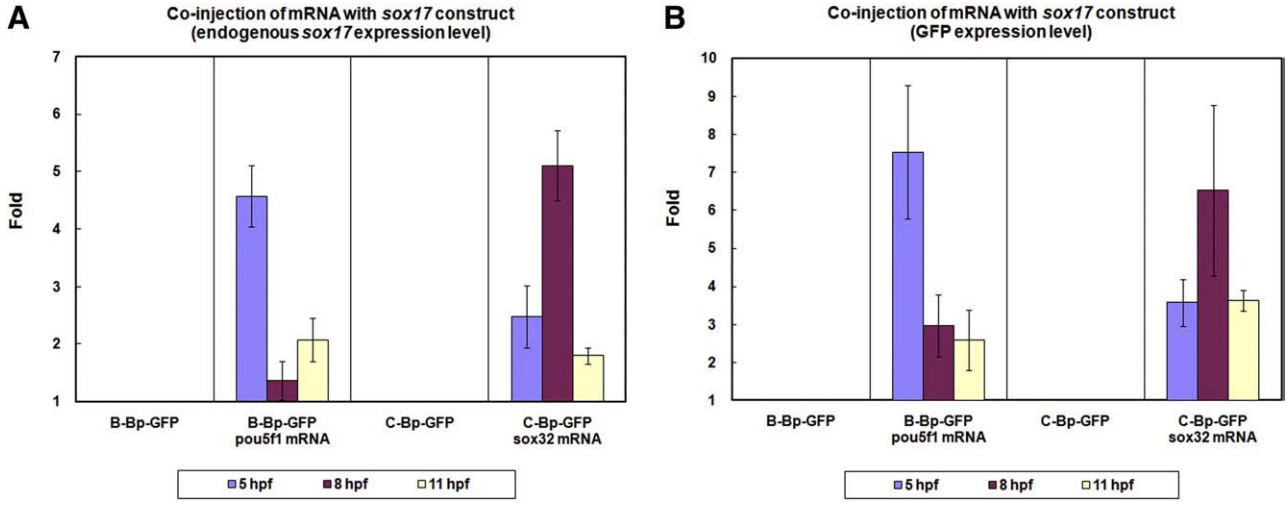


Fig. 5. Over-expression *pou5f1* or *sox32* mRNA increased the expression level of module B and C. *Sox17* B/Bp was co-injected with *pou5f1* mRNA, and *sox17* C/Bp was co-injected with *sox32* mRNA. At 5 hpf, 8 hpf or 11 hpf stage, twenty injected embryos were collected and Q-PCR was performed. Each column represents the average of the expression fold compared with module B alone, or module C alone, with the standard deviation shown as a line across the bar. (A) The fold change of the *sox17* following injection of the constructs with *pou5f1* mRNA or *sox32* mRNA were normalized with the *sox17* level by injection of the construct alone. (B) The fold change of the *gfp* following injection of the constructs with *pou5f1* mRNA or *sox32* mRNA were normalized with the *gfp* level by injection of the construct alone.

dependent manner. The upper complex specifically competed with the specific competitor in a dose dependent manner (Fig S3), but not with the non-specific competitor (data not shown). The lower complex competed with either the specific or non-specific competitor, suggesting it is a non-specific interaction. This test provides evidence for the direct binding of a protein present in the 5 hpf nuclear extract to the pou5f1 binding site in the module B of *sox17*. We noticed that the binding to the module B is very weak. This could be due to the lack of the other modules in this assay. From our functional study, module B and module C of *sox17* show a synergistic effect. The electrophoretic mobility shift assay cannot mimic the *in vivo* interaction between protein and DNA, and the cooperatively between modules. Therefore, we used chromatin immunoprecipitation (ChIP) to verify if the Sox32 protein binds to the module C, and if the binding of Sox32 on the target DNA can recruit the proteins that bind to module B.

The ChIP assay has become the mainstream technology in the study of protein–DNA interaction recently. The ChIP assay provides snapshots of protein–DNA interactions at a particular time point, and hence is useful for kinetic analysis of events occurring on chromosomal sequences *in vivo* (Aparicio et al., 2004). Coupled with microarray technology, ChIP on chip has become a powerful tool for studying genome-wide DNA–protein interactions (Moqtaderi and Struhl, 2004). Chromatin immunoprecipitation is a useful method to study *in vivo* relationships during zebrafish development (Hart et al., 2007; von Hofsten et al., 2008). It has been applied to zebrafish embryos development (Havis et al., 2006; Wardle et al., 2006). To further confirm that Sox32 specifically interacts with module C and has a synergistic effect with module B via protein–protein or protein–DNA interactions, we synthesized an antibody against the full-length Sox32 protein for use in a ChIP assay using the 8 hpf embryos. Four pairs of primers were designed to check the binding of Sox32 protein on the *cis*-regulatory region of sox17 (Fig. 6A). Two negative control primer pairs were designed, one located in the space region between the A and B modules, the other located in the space region between modules B and C. Two specific pairs of primers were designed for module B and C respectively. One pair is on module B for the Pou5f1 target site (Fig. 6B), and the other is on module C for the Sox32 target site (Fig. 6C). As predicted, Sox32 did not bind to the DNA non-specifically; there was no enrichment of Sox32 binding to the space region between the modules when compared to input DNA (Fig. 6D, yellow and light blue column). Sox32 binding to module C was enriched more than 20-fold compared to the input DNA (Fig. 6D, dark blue column), indicating that the Sox32 protein does indeed bind to the target site on module B. Furthermore, the Pou5f1 target site in module B also enriched Sox32 binding, albeit with a lower level (3.3-fold) compared to the input DNA (Fig. 6D, red column). This implied that the molecular basis for the synergistic effect between module C and B could be through the protein–protein interaction between the complex on the module C centered with Sox32 protein and the Pou5f1 complex on module B.

Our study has generated many new discoveries. Firstly, we found that the regulation of *sox17* requires three positive modules and one repressive module, and the synergistic effect between those modules is essential for the proper expression of the endoderm marker-*sox17*. Secondly, by using *cis*- and *trans*-perturbations and gain-of-function studies, we found the direct interaction site for Sox32 on module C and Pou5f1 on module B of the *sox17* promoter. Thirdly, direct binding of Sox32 on the module C was shown by a chromatin Immunoprecipitation (ChIP) assay, and the synergistic effect between module B and C was also demonstrated by the ChIP assay. These data demonstrate that the functional response of the *sox17* transcription unit to the Sox32 input and to the Pou5f1 input are indeed encoded in the DNA sequence of the *sox17 cis*-regulatory element.

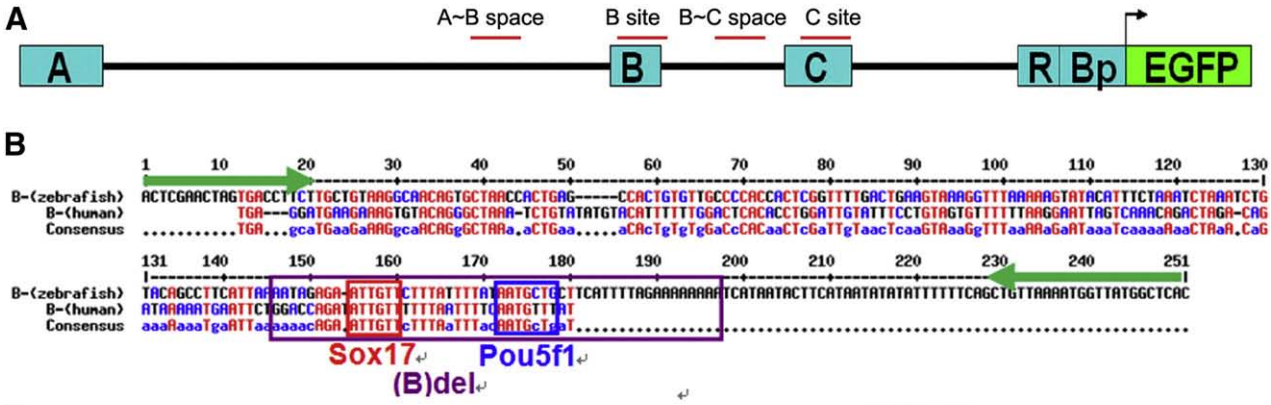
Discussion

A summary of expression data for sox17 cis-element constructs

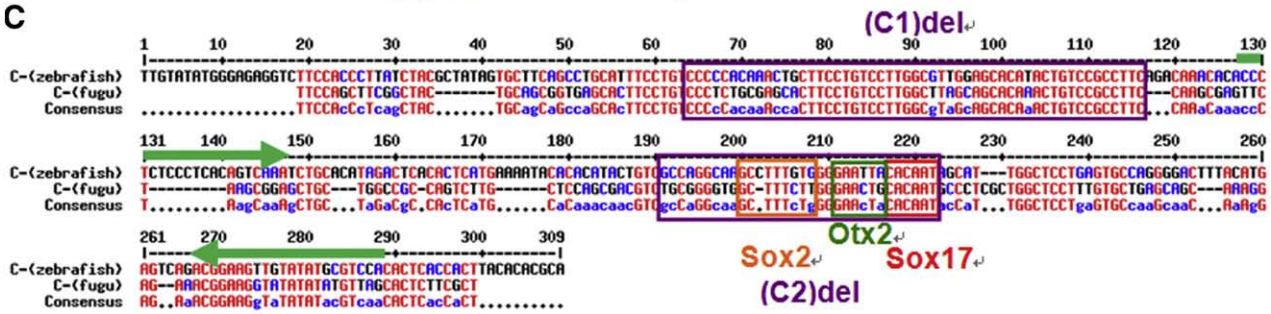
We have identified three enhancers and one repressor in the *sox17* regulatory region. From the deletion analysis, we found the shortest promoter region for *sox17* was contained in the 350-GFP construct, which showed non-specific expression everywhere in the embryo. The longer construct 660-GFP was not expressed anywhere in the embryo, indicating that the addition of the 310 bp upstream sequence to the construct restricted the expression of GFP. This implied that the 310 bp region functions as a repressor, which limits the non-specific ectopic expression of *sox17*. We called this module the R module, denoting its repression function. Longer constructs (−1213 and −1712) showed a slight increase of GFP expression. We analyzed the expression pattern of constructs containing the A, B, and C modules, and found adding module B to the basal promoter increased the expression in half of the embryos. Interestingly, the addition of both B and C modules increased the promoter activity further, indicating that there is a synergistic effect between the B and C modules. Furthermore, the expression of the BC-Bp-GFP was still restricted to half of the embryo, indicating that the signal transmitted from the BC module is restricted to half of the embryo. When all three conserved non-coding regions are included along with the basal promoter (ABC-Bp-GFP), expression was strongly enhanced by the entire embryo. There is a synergistic effect between the A module and the B and C modules, since the A, B, and C modules alone were expressed very weakly and in only in dorsal side of the embryo.

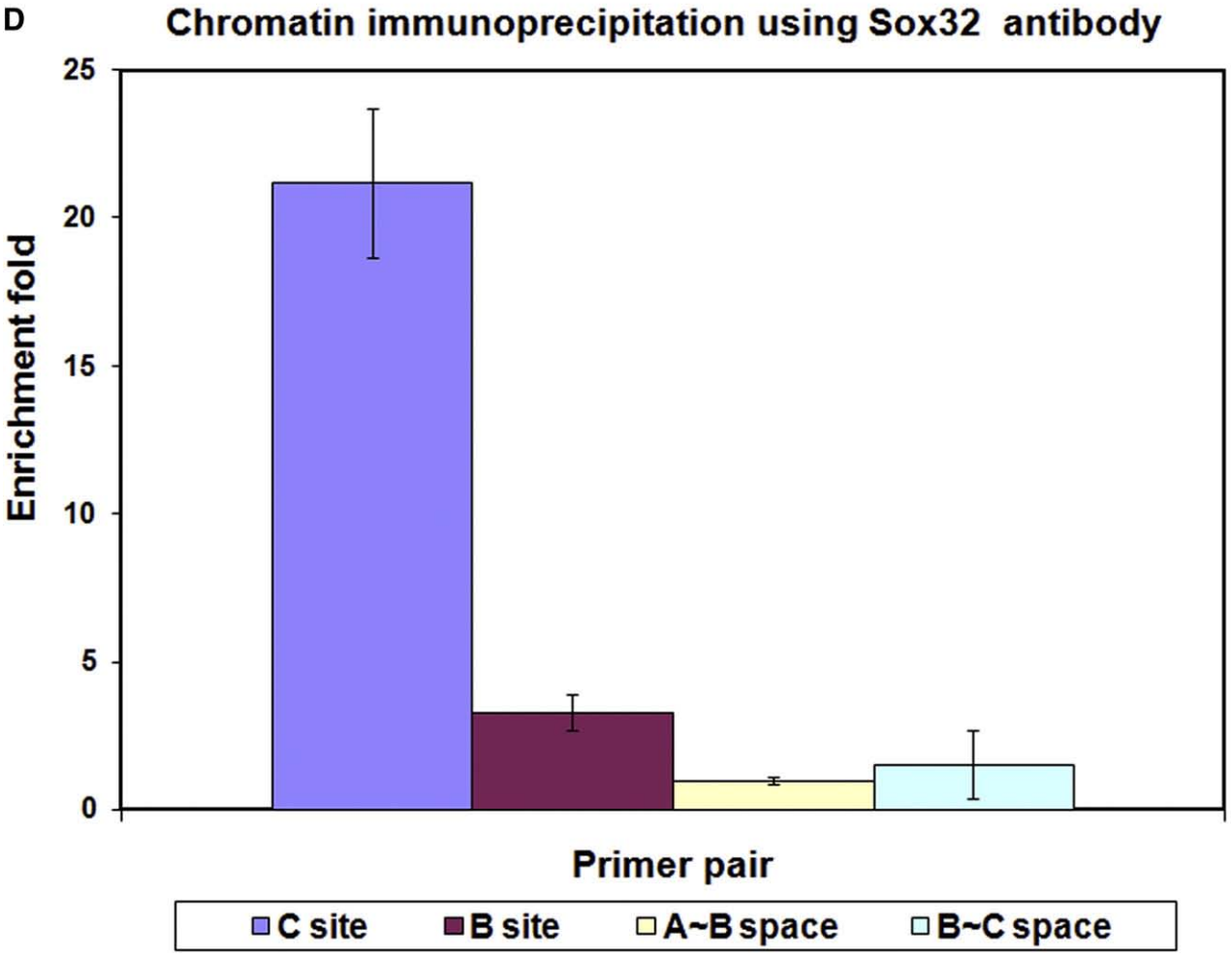
Another explanation for the half expression pattern for most of the weaker constructs might be the mosaic effect of the injected DNA. The exogenous DNA will not integrate into all of the cells in the first division during cleavage of the embryos. The expression pattern of the injected DNA-driven GFP will not be expressed in the whole region where endogenous gene is expressed. If the construct contains a stronger promoter, such as ABCR-Bp-GFP, the mosaic effect might be weaker. We used double florescence *in situ* hybridization in an attempt to verify the GFP expression pattern *versus* the dorsal marker foxa2, and found the A, B, C-Bp-GFP was expressed at dorsa; side as foxa2. Thus, we can rule out the possibility that the half embryo expression pattern of A, B, C, BC-Bp-GFP was due to mosaic expression. The pattern of the GFP expression in half of the embryo is different from the expression of endogenous *sox17* mRNA, which has a ring-like pattern in the endoderm lineage. Nevertheless, we found the ABCR-Bp-GFP construct expressed as a ring structure in the whole embryo, indicating that the A, B and C modules probably integrate the different inputs from various domains. It is possible that the A module is the downstream node for the Bmp or Fgf signal pathway,

Fig. 6. *In vivo* binding analysis of Sox32 on the sox17 *cis*-element verified by chromatin immunoprecipitation. (A) There are four pair primers, A~B space, B site, B~C space and C site, used for checking DNA from chromatin immunoprecipitation. A~B space and B~C space locate at a non-conserve region used as negative control test. B site and C site pair primers detect the fragment conserved within other species. (B) The sequence of the *sox17* B module aligns with human genomic sequence. Green arrows indicate the primer locations for identification of the enrichment fold, and the amplicon containing Pou5f1 binding sites. (C) The sequence of the *sox17* C module aligns with fugu genomic sequence. Green arrows indicated the primer locations for identification of the enrichment fold after chromatin immunoprecipitation, and the amplicon containing Sox32 binding sites. (D) Real time PCR shows that the *sox32* binding site has a 21.1-fold enrichment after ChIP; it comes from Sox32 antibody compared to input. Pou5f1 binding site also has 3.3-fold enrichment, but two negative controls have no significant enrichment.


A
A~B space
B site
B~C space
C site
A
B
C
R
Bp
EGFP

B
B-(zebrafish)
B-(human)
Consensus
Sox17
Pou5f1
(B)del


C
(C1)del
C-(zebrafish)
C-(fugu)
Consensus
Otx2
Sox2
Sox17
(C2)del


D
Chromatin immunoprecipitation using Sox32 antibody
Enrichment fold
25
20
15
10
5
0
Primer pair
C site
B site
A~B space
B~C space

while the B and C modules are the downstream node for the Nodal signaling. Furthermore, the R module represses the ectopic expression of GFP driven by the ABC modules, indicating that the R module should contain the binding sites for the Bmp or Fgf-activated transcription factors.

The cooperation between Sox32 and Pou5f1 activates sox17 expression

The study of the *sox17* promoter by Brand's group indicated that a 700 bp element from −2502 to −1809 is required for the synergistic effect between Sox32 and Pou5f1 (Reim et al., 2004). However, we found that −2502 to −1809 might contain a functional element for Pou5f1. Reim et al. reported that *pou5f1* mRNA enhanced the *sox17* promoter activity when injected with a DNA fragment from −2502 linked to a luciferase gene (−2503-Leu), but not with the DNA fragment from −1809 linked to the luciferase gene (−1809-Leu). Thus, the Pou5f1 functional element is located between −2503 and −1809 by Reim et al. They also reported that *sox32* mRNA enhanced the *sox17* promoter activity when injected with a DNA fragment from −1899 linked to luciferase (−1899-Leu), but not with a DNA fragment from −447 linked to luciferase (−447-Leu). Thus, the Sox32 functional element is located between −1899 and −447 by Reim et al. Our module B is between −2376 and −2132, a region that contains the Pou5f1 acting element, and our module C is between −1422 and −1117, a region that has the Sox32 acting element, both of which correlate with the results presented by Brand's group. We hypothesized that the Nodal signal induces *sox32* expression, and Sox32 binds to module C of *sox17*. Moreover, the maternal transcription factor Pou5f1 can bind to module B of the *sox17* gene and regulate the expression of *sox17*. Our data revealed also that the synergistic effect between modules B and C is important for the full activity of *sox17*.

Transcription factors occupying the basal promoter are necessary for interaction with the enhancersome. An earlier study of the *sox17* promoter showed that the 660 Bp-GFP upstream of the ATG site is not expressed in the zebrafish embryo (Reim et al., 2004). We used the DNA from −350 to +133 as a basal promoter, and observed a stronger non-specific GFP activity. The difference between our experiments and the previous study could be explained by the presence of a repressor that might reduce the GFP activity. Furthermore, we injected 46 pg of the PCR product that contains only the regulatory region linked to a GFP reporter. We found that the PCR product produced a 5-fold stronger signal than the linearized plasmid of the same construct (Yuh et al., unpublished data). Moreover, since the endogenous *sox17* gene is expressed from early embryogenesis to about 16 hpf, we observed the GFP activity as early as 5 hpf and every 3 h until 16 h. The results given above might explain the discrepancy of the effect of the 660-Bp-GFP in the *sox17* promoter between Rein's work and ours. We injected 46 pg of 660-GFP constructs into one-cell stage embryos. There was no GFP signal from early to late stages. The presence of a strong repressor region between −660 and −350 is obvious when compared with the basal promoter construct.

Other signals affect the endoderm formation

Signals other than Nodal have some important roles in determining the expression of the endoderm-specific transcription factors. The *Bmp* pathway decreases endoderm formation on the ventral side (Poulain et al., 2006), and Fgf inhibits endoderm formation on the dorsal side (Reim et al., 2004). Moreover, Fgf signaling negatively regulates *sox32* expression by a pathway parallel to *Bon* and Fau/Gata5 (Poulain et al., 2006). However, Fgf signaling increases the number of endodermal precursors and potentiates the ability of Nodal signaling to induce endoderm at the animal pole. These results demonstrated that Nodal, Fgf, and Bmp are antagonistic to each other, and to ensure at the margin most cells become an endoderm lineage cells.

The injection of *fgf8* mRNA or CA-Mek, a downstream effector of the Fgf signal, inhibits the expression of *sox32* and *foxa2* selectively (Reim et al., 2004). The treatment of the embryos with SU5402, an inhibitor of Fgf receptor 1, increased expression of the *sox32* and *foxa2* genes (Reim et al., 2004). Although the expression levels of *nodal* (*ndr1*, *ndr2*), Nodal antagonists (*lefty1* and *left2*), *bon* and *gata5* are unaffected by Fgf signaling, the expression of *og9x* was decreased markedly by SU5402 and increased after the injection of CA-Mek. These results demonstrated clearly that the Fgf pathway is parallel to Bon and Gata5 in the molecular cascades leading to endoderm specification. However, the Fgf signal may work on several other target genes, and not only through *sox32*.

Other transcription factors might indirectly regulate sox17 by regulating sox32

In our endoderm subcircuits, *sox17* is activated by Otx2 and Sox32 at 5 hpf, and by Gata5 and Sox32 at 8 hpf and 11 hpf. When we injected the *gata5* MO and *bon* MO with the C-Bp construct, we observed the same decrease of GFP expression level as that seen for the *sox32* MO (Fig. S1). However, we could not detect the Gata5 or Bon binding sites on the C module. It is possible that Gata5 and Bon activate *sox17* by activating *sox32*. It has been shown that both Gata5 and Bon can activate the expression of *sox32* (Aoki et al., 2002; Reiter et al., 2001), which suggested an alternative interpretation for the GRNs determined by perturbation coupled with a real time PCR experiment. Similarly, we could not detect any Otx2 binding site, which has a positive function in the ABCR modules of *sox17*. Interestingly, even though the *sox17* mRNA level is decreased at 5 hpf, none of the A-Bp, B-Bp, C-Bp, or ABCR-Bp-driven GFP levels was affected by *otx2* MO co-injection (Supplementary data). This indicates that either Otx2 acts through elements outside of those three positive modules, or the effect of Otx2 is also indirect. Further experiments are needed to clarify this node.

From our knockdown and real time RT-PCR data, Sox17, Gbx2, and Otx2 repress *sox17* at late stages. Interestingly, even though the *sox17* mRNA level is increased at 11 hpf by injection of *otx2* MO, none of the A-Bp, B-Bp, C-Bp, and ABCR-Bp-driven GFP levels was affected by *otx2* MO co-injection (Supplementary data). In addition, even though the *sox17* mRNA level is increased by injection of Sox17 MO, none of the A-Bp, B-Bp, C-Bp, or ABCR-Bp-driven GFP levels was affected by *sox17* MO co-injection (Supplementary data). This indicates that either Otx2 or Sox17 repressed the expression of *sox17* by acting through the elements outside of those four modules, or the effect of *otx2* and *sox17* is also indirect. Further experiments are needed to clarify this repression effect.

Evolutionary comparison of the role of sox17 in endoderm differentiation and the regulatory networks

Sox17 has been shown to be an evolutionary conserved endoderm regulator. The *sox17* gene plays an essential role in the gene regulatory networks (GRNs) for endoderm specification in human (Seguin et al., 2008), mouse (Qu et al., 2008) and *Xenopus* (Howard et al., 2007) development. In a human embryonic stem cell study, constitutive expression of *sox17* produced definitive endoderm progenitors (Seguin et al., 2008). In mouse stem cell, over-expression of *sox17* caused up-regulation of a set of endoderm-specific gene markers (Qu et al., 2008). It has been shown that the expression of *sox17*, but not other endoderm markers, is decreased in *dusp4* MO injected embryos (Brown et al., 2008). The foregut and pancreases are defective in the *dusp4* morphant at later stages; this specific loss of *sox17* establishes a new class of endoderm specification defect.

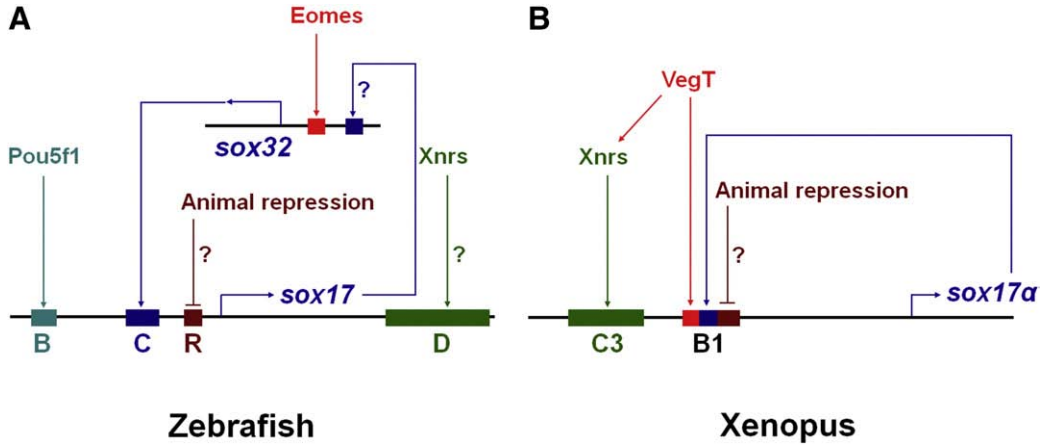


Fig. 7. Summary of the regulation of the zebrafish *sox17 cis*-element compared with *Xenopus sox17*α. Arrow lines indicate the input of activation from upstream to downstream target gene, flat line indicated the input of repression from upstream to downstream target gene. Small colored boxes indicate the transcription factor binding sites found on the regulatory region of the target genes. (A) Summary of the regulation of the zebrafish *sox17* promoter. *Emos* (*VegT* homologous gene in zebrafish) induced *sox32* directly, via the upstream target site. *Sox32* activates *sox17* through the binding to the target site on module C, *sox17* induces *sox32* but the target site is not known. Sox32 and sox17 form an autoregulation loop that mimics the *Xsox17* autoregulation loop. Pou5f1 binds to the module B, and the repressive module is required for the proper expression of *sox17*. This synergistical effect through the Sox binding site on the C module, Pou5f1 binding site on the B module and the R module ensure the correct expression of *sox17*. A Foxh1/Smad site has been found in module D, but the function of this target site is not yet known. (B) Summary of the regulation of *Xsox17*α promoter from the study of Howard et al. (2007). *VegT* induces *Xsox17* directly through the T-box half site on B1 module, *Xsox17* autoregulates itself through the Sox binding site on the B1 module. *VegT* and *Xsox17* work synergistically through the adjacent binding sites. Xnrs positively regulates the *Xsox17* perhaps by positive induction of C3 via the Foxh1/Smad sites. An unknown animal repressive function was found in the module B1, but the detailed mechanism is not known.

Dusp4 is a dual functional phosphatase and is a MAPK inhibitor. In zebrafish, *sox32* is specifically evolved and is not found in other species. The combined action of *sox32* and *sox17* are equivalent to the overall action of *sox17* in other species.

The regulation of *sox17* in *Xenopus* had been studied by transgenesis, and two important control elements (B1 and C3) had been identified, which reside about 9 kb upstream at the start of transcription of sox17 (Howard et al., 2007). By a gain of function study, Howard et al. showed that B1 responds positively to *VegT* and *sox17*. On the other hand, C3, which contains Smad and Foxh1 binding site, is induced by constituently active Smad2. Although *Xenopus* and zebrafish contains an evolutionary conserved *sox17* gene, the transcription regulation of *sox17* appears to be different (Fig. 7A). First, in zebrafish, the *sox32* is a very important transcription factor regulating the expression of *sox17*, but *sox32* is missing from *Xenopus*. Second, in zebrafish, the VegT homologous T-box gene, *eomesodermin*, positively regulates *sox32* via an element located 1476 bp upstream of the translational initiation site of *sox32* (Bjornson et al., 2005). Third, the autoregulation of *sox17* on itself is not seen in zebrafish, we have the knockdown expression of *sox17* and did not observe the down-regulation of *sox17* mRNA using real time RT-PCR. On the other hand, we observed that *sox17* positively regulates *sox32* (unpublished data). Lastly, we have compared the genomic sequence of *sox17* from zebrafish and *Xenopus*, and no similarity can be found (Fig. 1). The regulatory interactions Howard et al. found in *Xsox17*α, where *VegT* induces *TGF-*β and *Xsox17*, and autoregulation of *Xsox17* is most significant in endodermal differentiation in *Xenopus* (Fig. 7B). In zebrafish, this autoregulation of *sox17* evolved into positive regulation of *sox32* to *sox17*, and then *sox17* can positively regulate *sox32*; this autoregulation of two *sox* family genes is most significant in endodermal differentiation in zebrafish.

Acknowledgments

We thank Te-Hsuan Jang for maintaining the zebrafish stocks. Funding support from National Health Research Institute grants MG-094-PP-14, MG-095-PP-08, MG-096-PP-05, and MG-097-PP-07 to Dr Yuh Chiou-Hwa is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2008.11.010.

References

Alexander, J., Stainier, D.Y., 1999. A molecular pathway leading to endoderm formation in zebrafish. Curr. Biol. 9, 1147–1157.
Aoki, T.O., David, N.B., Minchiotti, G., Saint-Etienne, L., Dickmeis, T., Persico, G.M., Strahle, U., Mourrain, P., Rosa, F.M., 2002. Molecular integration of *casanova* in the Nodal signalling pathway controlling endoderm formation. Development 129, 275–286.
Aparicio, O., Geisberg, J.V., Struhl, K., 2004. Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. Curr. Protoc. Cell. Biol. 7 Chapter 17, Unit 17.
Bjornson, C.R., Griffin, K.J., Farr III, G.H., Terashima, A., Himeda, C., Kikuchi, Y., Kimelman, D., 2005. Eomesodermin is a localized maternal determinant required for endoderm induction in zebrafish. Dev. Cell. 9, 523–533.
Brown, J.L., Snir, M., Noushmehr, H., Kirby, M., Hong, S.K., Elkahloun, A.G., Feldman, B., 2008. Transcriptional profiling of endogenous germ layer precursor cells identifies *dusp4* as an essential gene in zebrafish endoderm specification. Proc. Natl. Acad. Sci. U. S. A. 105, 12337–12342.
Chan, T.M., Longabaug, W., Bolouri, H., Chen, H.L., Tseng, W.F., Chao, C.H., Jang, T.H., Lin, Y.I., Hung, S.C., Wang, H.D., Yuh, C.H., 2008. Developmental gene regulatory networks in the zebrafish embryo. Biochim. Biophys. Acta (Epub 2008 Oct 8).
Chen, H., Blanchette, M., 2007. Detecting non-coding selective pressure in coding regions. BMC Evol. Biol. 7 (Suppl 1), S9.
David, N.B., Rosa, F.M., 2001. Cell autonomous commitment to an endodermal fate and behaviour by activation of Nodal signalling. Development 128, 3937–3947.
Davidson, E., Levin, M., 2005. Gene regulatory networks. Proc. Natl. Acad. Sci. U. S. A. 102, 4935.
Davidson, E.H., McClay, D.R., Hood, L., 2003. Regulatory gene networks and the properties of the developmental process. Proc. Natl. Acad. Sci. U. S. A. 100, 1475–1480.
Dickmeis, T., Mourrain, P., Saint-Etienne, L., Fischer, N., Aanstad, P., Clark, M., Strahle, U., Rosa, F., 2001. A crucial component of the endoderm formation pathway, CASANOVA, is encoded by a novel sox-related gene. Genes Dev. 15, 1487–1492.
Hart, D.O., Raha, T., Lawson, N.D., Green, M.R., 2007. Initiation of zebrafish haematopoiesis by the TATA-box-binding protein-related factor Trf3. Nature 450, 1082–1085.
Havis, E., Anselme, I., Schneider-Maunoury, S., 2006. Whole embryo chromatin immunoprecipitation protocol for the in vivo study of zebrafish development. Biotechniques 40, 34, 36, 38 passim.
Howard, L., Rex, M., Clements, D., Woodland, H.R., 2007. Regulation of the *Xenopus* Xsox17alpha(1) promoter by co-operating VegT and Sox17 sites. Dev. Biol. 310, 402–415.
Hsiao, C.D., You, M.S., Guh, Y.J., Ma, M., Jiang, Y.J., Hwang, P.P., 2007. A positive regulatory loop between foxi3a and foxi3b is essential for specification and differentiation of zebrafish epidermal ionocytes. PLoS ONE 2, e302.

Jones, C.M., Kuehn, M.R., Hogan, B.L., Smith, J.C., Wright, C.V., 1995. Nodal-related signals induce axial mesoderm and dorsalize mesoderm during gastrulation. Development 121, 3651–3662.

Julich, D., Hwee Lim, C., Round, J., Nicolaije, C., Schroeder, J., Davies, A., Geisler, R., Lewis, J., Jiang, Y.J., Holley, S.A., 2005. *beamter/deltaC* and the role of Notch ligands in the zebrafish somite segmentation, hindbrain neurogenesis and hypochord differentiation. Dev. Biol. 286, 391–404.

Kikuchi, Y., Agathon, A., Alexander, J., Thisse, C., Waldron, S., Yelon, D., Thisse, B., Stainier, D.Y., 2001. *casanova* encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish. Genes Dev. 15, 1493–1505.

Kunwar, P.S., Zimmerman, S., Bennett, J.T., Chen, Y., Whitman, M., Schier, A.F., 2003. Mixer/Bon and FoxH1/Sur have overlapping and divergent roles in Nodal signaling and mesendoderm induction. Development 130, 5589–5599.

Le Good, J.A., Joubin, K., Giraldez, A.J., Ben-Haim, N., Beck, S., Chen, Y., Schier, A.F., Constam, D.B., 2005. Nodal stability determines signaling range. Curr. Biol. 15, 31–36.

Lunde, K., Belting, H.G., Driever, W., 2004. Zebrafish pou5f1/pou2, homolog of mammalian Oct4, functions in the endoderm specification cascade. Curr. Biol. 14, 48–55.

Ma, M., Jiang, Y.J., 2007. Jagged2a-notch signaling mediates cell fate choice in the zebrafish pronephric duct. PLoS Genet. 3, e18.

Mizoguchi, T., Izawa, T., Kuroiwa, A., Kikuchi, Y., 2006. Fgf signaling negatively regulates Nodal-dependent endoderm induction in zebrafish. Dev. Biol. 300, 612–622.

Moqtaderi, Z., Struhl, K., 2004. Defining in vivo targets of nuclear proteins by chromatin immunoprecipitation and microarray analysis. Curr. Protoc. Mol. Biol. 9 Chapter 21, Unit 21.

Muller, F., Albert, S., Blader, P., Fischer, N., Hallonet, M., Strahle, U., 2000. Direct action of the nodal-related signal cyclops in induction of sonic hedgehog in the ventral midline of the CNS. Development 127, 3889–3897.

Ochi, H., Westerfield, M., 2007. Signaling networks that regulate muscle development: lessons from zebrafish. Dev. Growth Differ. 49, 1–11.

Poulain, M., Lepage, T., 2002. Mezzo, a paired-like homeobox protein is an immediate target of Nodal signalling and regulates endoderm specification in zebrafish. Development 129, 4901–4914.

Poulain, M., Furthauer, M., Thisse, B., Thisse, C., Lepage, T., 2006. Zebrafish endoderm formation is regulated by combinatorial Nodal, FGF and BMP signalling. Development 133, 2189–2200.

Qu, X.B., Pan, J., Zhang, C., Huang, S.Y., 2008. Sox17 facilitates the differentiation of mouse embryonic stem cells into primitive and definitive endoderm in vitro. Dev. Growth Differ.

Reim, G., Mizoguchi, T., Stainier, D.Y., Kikuchi, Y., Brand, M., 2004. The POU domain protein spg (pou2/Oct4) is essential for endoderm formation in cooperation with the HMG domain protein casanova. Dev. Cell 6, 91–101.

Reiter, J.F., Kikuchi, Y., Stainier, D.Y., 2001. Multiple roles for Gata5 in zebrafish endoderm formation. Development 128, 125–135.

Sakaguchi, T., Kuroiwa, A., Takeda, H., 2001. A novel sox gene, 226D7, acts downstream of Nodal signaling to specify endoderm precursors in zebrafish. Mech. Dev. 107, 25–38.

Schier, A.F., Talbot, W.S., 2001. Nodal signaling and the zebrafish organizer. Int. J. Dev. Biol. 45, 289–297.

Schier, A.F., Talbot, W.S., 2005. Molecular genetics of axis formation in zebrafish. Annu. Rev. Genet. 39, 561–613.

Schoenebeck, J.J., Yelon, D., 2007. Illuminating cardiac development: advances in imaging add new dimensions to the utility of zebrafish genetics. Semin. Cell Dev. Biol. 18, 27–35.

Seguin, C.A., Draper, J.S., Nagy, A., Rossant, J., 2008. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. Cell Stem Cell 3, 182–195.

Sprague, J., Bayraktaroglu, L., Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Haendel, M., Howe, D.G., Knight, J., Mani, P., Moxon, S.A., Pich, C., Ramachandran, S., Schaper, K., Segerdell, E., Shao, X., Singer, A., Song, P., Sprunger, B., Van Slyke, C.E., Westerfield, M., 2008. The Zebrafish Information Network: the zebrafish model organism database provides expanded support for genotypes and phenotypes. Nucleic Acids Res. 36, D768–D772.

von Hofsten, J., Elworthy, S., Gilchrist, M.J., Smith, J.C., Wardle, F.C., Ingham, P.W., 2008. Prdm1- and Sox6-mediated transcriptional repression specifies muscle fibre type in the zebrafish embryo. EMBO Rep. 9, 683–689.

Wardle, F.C., Odom, D.T., Bell, G.W., Yuan, B., Danford, T.W., Wiellette, E.L., Herbolsheimer, E., Sive, H.L., Young, R.A., Smith, J.C., 2006. Zebrafish promoter microarrays identify actively transcribed embryonic genes. Genome Biol. 7, R71.

Werner, T., Hammer, A., Wahlbuhl, M., Bosl, M.R., Wegner, M., 2007. Multiple conserved regulatory elements with overlapping functions determine Sox10 expression in mouse embryogenesis. Nucleic Acids Res. 35, 6526–6538.

Westerfield, O., 1995. A prescription for hospital safety: treating workplace violence. Healthc. Facil. Manag. Ser. 1–8.

Yuh, C.H., Dorman, E.R., Howard, M.L., Davidson, E.H., 2004. An otx *cis*-regulatory module: a key node in the sea urchin endomesoderm gene regulatory network. Dev. Biol. 269, 536–551.

Zacchigna, S., Ruiz de Almodovar, C., Carmeliet, P., 2008. Similarities between angiogenesis and neural development: what small animal models can tell us. Curr. Top. Dev. Biol. 80, 1–55.

Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L., Kuehn, M.R., 1993. Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. Nature 361, 543–547.

Zimonjic, D.B., Kelley, M.J., Rubin, J.S., Aaronson, S.A., Popescu, N.C., 1997. Fluorescence in situ hybridization analysis of keratinocyte growth factor gene amplification and dispersion in evolution of great apes and humans. Proc. Natl. Acad. Sci. U. S. A. 94, 11461–11465.

Zorn, A.M., Wells, J.M., 2007. Molecular basis of vertebrate endoderm development. Int. Rev. Cytol. 259, 49–111.