THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 32, pp. 28520–28532, August 12, 2011


# Global Identification of SMAD2 Target Genes Reveals a Role for Multiple Co-regulatory Factors in Zebrafish Early Gastrulas*[S]

Received for publication, March 2, 2011, and in revised form, June 10, 2011 Published, JBC Papers in Press, June 13, 2011, DOI 10.1074/jbc.M111.236307

**Zhaoting Liu[‡1], Xiwen Lin[§1], Zhaoping Cai[‡], Zhuqiang Zhang[§], Chunsheng Han[§], Shunji Jia[¶], Anming Meng[‡¶2], and Qiang Wang[‡3]**

*From the [‡]State Key Laboratory of Biomembrane and Membrane Biotechnology and [§]State Key Laboratory of Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, 100101 Beijing and the [¶]Protein Science Laboratory of Ministry of Education, School of Life Sciences, Tsinghua University, 10084 Beijing, China*

**Nodal and Smad2/3 signals play pivotal roles in mesendoderm induction and axis determination during late blastulation and early gastrulation in vertebrate embryos. However, Smad2/3 direct target genes during those critical developmental stages have not been systematically identified. Here, through ChIP-chip assay, we show that the promoter/enhancer regions of 679 genes are bound by Smad2 in the zebrafish early gastrulas. Expression analyses confirm that a significant proportion of Smad2 targets are indeed subjected to Nodal/Smad2 regulation at the onset of gastrulation. The co-existence of DNA-binding sites of other transcription factors in the Smad2-bound regions allows the identification of well known Smad2-binding partners, such as FoxH1 and Lef1/β-catenin, as well as many previously unknown Smad2 partners, including Oct1 and Gata6, during embryogenesis. We demonstrate that Oct1 physically associates with and enhances the transcription and mesendodermal induction activity of Smad2, whereas Gata6 exerts an inhibitory role in Smad2 signaling and mesendodermal induction. Thus, our study systemically uncovers a large number of Smad2 targets in early gastrulas and suggests cooperative roles of Smad2 and other transcription factors in controlling target gene transcription, which will be valuable for studying regulatory cascades during germ layer formation and patterning of vertebrate embryos.**

During early development of vertebrate embryos, signaling molecules play essential roles in axis determination and germ layer induction and patterning. Members of the transforming growth factor-β (TGF-β) superfamily are among the major players in those developmental processes. The *Nodal* gene was first identified as a novel TGF-β gene expressing in mouse node during gastrulation, and its interruption causes missing of the primitive streak and mesodermal derivatives with excess ectoderm (1–3). In zebrafish, two *nodal* genes, *squint* (*sqt*) and *cyclops* (*cyc*), are expressed in blastulas, and *cyc;sqt* double mutant embryos lack endodermal and mesodermal tissues, also suggesting an essential role of Nodal ligands in zebrafish mesendoderm induction (4, 5). In amphibians, Nodal signals form a gradient across the future dorsoventral axis with the highest concentration on the dorsal-most region, which leads to different cell fates along the axis (6). Moreover, Nodal signals are required for the establishment of the left-right axis in various vertebrate species (7, 8). Bone morphogenetic proteins, constituting a subfamily of the TGF-β superfamily, mainly act as ventral players because they induce epidermis from the ventral ectoderm and specify ventral mesodermal and endodermal fates (9–11).

Nodal/TGF-β signal transduction leads to the activation of the downstream effectors Smad2 and Smad3 through serine phosphorylation at their C terminus (12, 13). The activated Smad2 or Smad3 forms heteromeric complexes with Smad4 in the cytoplasm and translocates into the nucleus where they bind to the promoters/enhancers and regulate the expression of target genes (12). Like nodal null mutant mouse embryos, Smad2 knock-out embryos fail to undergo gastrulation and lack mesoderm (14, 15); however, Smad3 null embryos develop normally (16, 17). In zebrafish, all of three *smad2/3* genes, *smad2*, *smad3a,* and *smad3b*, are ubiquitously expressed during early embryogenesis (18); depletion of Smad2/3 activities by overexpression of dominant negative *smad2/3* mutants also leads to mesendodermal defects as well as additional defects in neural induction and patterning (19, 20). It is likely that Nodal signals exert their effects in mesendoderm induction and patterning mainly through Smad2/3.

Given that Nodal/Smad2 signaling is essential for early embryogenesis in vertebrates, efforts have been made to identify Nodal-regulated genes. It has been found in zebrafish that *sqt* overexpression caused alteration of transcription levels of 376 genes at the 30% epiboly stage, and 2045 genes showed altered transcription levels in the Nodal signal-deficient *MZoep* (maternal-zygotic oep) mutants at a similar stage (21). Another assay in zebrafish identified 72 Nodal-regulated genes that are expressed in blastodermal margins at the shield stage (22).

* This work was supported by National Basic Research Program of China Grants 2011CB943904 and 2011CB943800, Strategic Priority Research Program of the Chinese Academy of Sciences Grant XDA01010104, and the National Natural Science Foundation of China Grants 90919058, 30830068, and 30921004/C061003.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental "Experimental Procedures," Figs. S1 and S2, and Tables S1–S3.

⌘ *Author's Choice*—Final version full access.

[1] Both authors contributed equally to this work.

[2] To whom correspondence may be addressed. Tel.: 86-10-62772256; Fax: 86-10-62794401; E-mail: mengam@mail.tsinghua.edu.cn.

[3] To whom correspondence may be addressed. Tel.: 86-10-64807895; Fax: 86-10-64807895; E-mail: qiangwang@ioz.ac.cn.

However, few Nodal-regulated genes have proved to be direct targets of Smad2/3.

Chromatin immunoprecipitation (ChIP) combined with microarray-based analysis (ChIP-chip) has been successfully used to identify on a large scale direct targets of a known DNA-binding protein and to search for its novel cofactors during embryogenesis (23–27). In this study, we performed the ChIP-chip analysis using an anti-Smad2 antibody in zebrafish embryos to find Smad2 direct targets and its important cofactors during embryonic induction and patterning. We identified a total of 556 Smad2-bound regions (SBRs)[4] in the genome of early gastrulas, which were allocated to the promoters of 679 genes. The identified Smad2 target genes in early gastrulas are enriched in transcription factors, signaling molecules, and developmental regulators. The SBRs of the Smad2 target genes scatter in the promoter and enhancer regions. In many cases, the Smad2/4-binding motif (CAG(A/C)C) within the SBRs clusters with the DNA-binding sites of other transcription factors, including previously known and unknown factors. We demonstrate for the first time that Oct1 and Gata6 can physically associate with and regulate the activity of Smad2 during mesendoderm induction of zebrafish embryos. Our results collectively indicate that Nodal/Smad2 signaling exerts its developmental functions through initiating a cascade of regulatory events by controlling the expression of transcription factors, signaling molecules, and developmental regulators and reveal novel regulatory mechanisms of Smad2/4-induced transcription.

## EXPERIMENTAL PROCEDURES

*Fish Embryos*—The Tuebingen line of zebrafish was used. Embryos were collected from natural matings, raised in Holtfreter's solution at 28.5 °C, and staged by morphology as described previously (28).

*Antibodies and Chromatin Immunoprecipitation Assays*—For immunoblotting, we used affinity-purified anti-Smad2/3 (3102, Cell Signaling Technology), anti-Smad3 (9523, Cell Signaling Technology), anti-FLAG (F3165, Sigma) and anti-Myc (R950–25, Invitrogen) antibodies. Anti-Smad2/3 and anti-β-catenin (M24002, Abmart) antibodies were used in ChIP assays.

For each immunoprecipitation, ~1500 embryos at the 50% epiboly stage were enzymatically dechorionated and then fixed for 10 min in 1.85% formaldehyde at room temperature. ChIP-chip experiments were performed as described previously (29) with modifications. Additional details for the ChIP-chip experiments can be provided on request.

*Promoter Chip Assay*—The zebrafish promoter genome ChIP-on-chip microarray set (Agilent Technologies) was used, and promoter chip assays were performed by Shanghai Biochip Co., Ltd. This oligonucleotide-based promoter array set covers 9 kb upstream and 3 kb downstream of the transcription start sites (TSSs) of 11,512 zebrafish genes (based on Zv6). Purified DNA from immunoprecipitation was blunted and ligated to a universal linker and amplified by PCR. The enriched DNA was labeled with Cy5, and DNA eluted from control samples was labeled with Cy3. The labeled DNAs were combined and hybridized to the promoter array sets according to the manufacturer's instruction.

*Data Normalization, Analysis, and Identification of Bound Regions*—The two-channel raw data of ChIP-chip were normalized with intensity-dependent Loess method (30). The probes were re-mapped to the Zv7 genome, and only unique-mapped ones were retained, which finally represented promoter regions of 10,117 zebrafish genes. Those probes with signal intensity of less than 300 were excluded to decrease the false-positives.

The neighborhood model was used to identify the Smad2-bound regions (peaks). A region was considered to be a peak if more than two adjacent probes were significantly enriched ($p < 0.05$), and the distance between them was less than 1000 bp (29).

*Gene Ontology (GO) Analysis*—GO terms were extracted from Gene Ontology data base (31), and terms of the 5th layer were analyzed statistically. Terms that met some criterion were regarded as the enriched ones.

*Motif Analysis*—We expanded 500 bp upstream and downstream from the peak central position, respectively, and we scanned these sequences for the consensus Smad2/4-binding motif (CAG(A/C)C) (32, 33). The 1000-bp sequences upstream of TSS of zebrafish Refseq mRNA were chosen as the control set. To determine whether this motif is significantly enriched in foreground *versus* background, Z-score and *p* values were calculated as described previously (34, 35). Z-score was calculated as shown in Equation 1,

$$z = \frac{x - \mu - 0.5}{\sigma} \quad \text{(Eq. 1)}$$

$$\mu = B \times \frac{n}{N},\ \sigma = \sqrt{np(1-p)},\ p = B/N$$

where $x$ is the total length (bp) of observed sites in the expanded peaks; $B$ is the total length of the sites in control set; $n$ is the total length in expanded peaks; and $N$ is the total length of the control set.

The $p$ value was calculated according to the binominal distribution shown in Equation 2,

$$P(X \geq x) = 1 - P(X < x) = 1 - \sum_{i=0}^{x=1} \binom{n}{i} p^i (1-p)^{n-i} \quad \text{(Eq. 2)}$$

where $x$ is the observed number of sequences that contains this motif in the foreground set; $p$ is the proportion of sequences containing a motif in the background set; and $n$ is the peak number in the foreground set. We empirically chose $Z \geq 10$ and $p < 0.01$ as the criterion of significantly enriched motifs (34, 35).

Peaks containing the Smad2/4-binding motif (CAG(A/C)C) were selected, and each expanded 100 bp upstream and downstream of the Smad-binding motifs. The combined 200-bp sequences were considered as the foreground set. 200-bp sequences upstream of 5′ end of all zebrafish Refseq mRNAs were retrieved as our control set (background). Vertebrate

[4] The abbreviations used are: SBR, Smad2-bound region; TSS, transcription start site; GO, gene ontology.

position weight matrices from TRANSFAC (36) and JASPAR (37) databases were scanned on both sets (FoxH1-binding site was added manually as described previously (38)). To determine whether a motif is significantly enriched in foreground *versus* background, *Z*-score and *p* value were also calculated. We empirically chose $Z \geq 10$ and $p < 0.01$ as the criterion of significantly enriched motifs (34, 35). We also investigated the emergence frequency of significantly enriched motifs in the 200-bp peaks containing the Smad2/4-binding motif. The percentage was calculated by Equation 3,

$$\frac{n}{N} \times 100\% \qquad \text{(Eq. 3)}$$

where *n* is the number of peaks containing a motif, and *N* is the total number of peaks containing the Smad2/4 motif.

*Site-specific PCR Analysis*—To confirm the enrichment of selected Smad2 target sites, a subset of primers (see supplemental "Experimental Procedures" for details) corresponding to the predicted binding sites was designed to amplify a 200–300-bp region around the peak probes. Using chromatin immunoprecipitation with no antibody as a control, immuno-enriched DNA was amplified, and the products were visualized on a 1.5% agarose gel.

*Gene Expression Profiling*—Total RNA was extracted from 50% epiboly zebrafish embryos using TRIzol reagent (Invitrogen). The microarray was performed on the Agilent zebrafish (V2) gene expression microarray by Shanghai Biochip Co., Ltd. The microarray data were normalized across all arrays using quantile normalization (39). Duplicates were averaged, and up-regulated or down-regulated genes were determined by the signal log ratio.

*RNA Synthesis, Morpholinos, Microinjection, and Whole Mount in Situ Hybridization*—The *lefty1*, *sqt*, constitutively active *smad2* (*casmad2*), and dominant negative *smad2* (*dnsmad2*) mRNAs were synthesized *in vitro* from corresponding linearized plasmids using the mMessage mMachine kit (Ambion). Digoxigenin-UTP-labeled antisense RNA probes were transcribed *in vitro* using Megascript kit (Ambion) according to the manufacturer's instructions. Gene Tools synthesizing *smad2*, *smad3a*, and *smad3b* morpholino oligonucleotides were designed as reported previously (19). Microinjection and whole mount *in situ* hybridization were performed as before (19, 20).

*Generation of Efnb2b-GFP Constructs*—The pGL3-Basic vector was modified by replacing the firefly luciferase open reading frame with the enhanced GFP sequence from pEGFP-N1 vector. The promoter of *efnb2b* was amplified from zebrafish genomic DNA using the specific upper primer (5′-CGACGCTAGCAGAGTGAAAAAGAGCC-3′) and lower primer (5′-GTCAGTTTTAGTCGACAAGCGAACAC-3′) and was inserted into the NheI and XhoI sites of the modified pGL3-Basic vector to generate the efnb2b-GFP construct. The efnb2b(mut)-GFP construct was generated by mutating the tandem Smad2/4-binding motifs in the *efnb2b* promoter using a modified lower primer (5′-CACTTTCATCGTTTGGTTTGCGAAATC-3′, mutated bases are underlined) was used. All of the sequences were confirmed by DNA sequencing. The efnb2b-GFP or efnb2b(mut)-GFP plasmid DNA was injected into the cytoplasm of one-cell embryos.

*Electrophoretic Mobility Shift Assay*—The 45-bp oligonucleotide (5′-GTTTTGACCAATACACAATTGATTTCGCAGACCAGACGATGAAAG-3′), which was derived from the *efnb2b* promoter region, and its complementary oligonucleotides were synthesized and annealed to form a double-stranded oligonucleotide. Its mutant form with base substitutions (5′-GTTTTGACCAATACACAATTGATTTCGCAAACCAAACGATGAAAG-3′, the two substituted bases are underlined) was similarly prepared. The oligonucleotides were labeled with [$\gamma$-$^{32}$P]ATP using T4 polynucleotide kinase (Fermentas). The wild-type and *sqt* mRNA-injected embryos were harvested at the 50% epiboly stage, and embryonic extracts were prepared in TNE buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM EDTA, and 0.5% Nonidet P-40), 3 $\mu$l per embryo, and were spun to remove membrane debris and yolk lipids. 2 $\mu$l of embryonic extracts and 10 ng of radiolabeled probe were incubated at 4 °C for 30 min in a 10-$\mu$l reaction volume. The reaction buffer contains 5% glycerol, 20 mM Na-HEPES, pH 7.9, 100 mM KCl, 2 mM EDTA, 2 mM DTT, 4 mM Tris-HCl, pH 7.9, and 0.1 mg/ml poly(dI-dC) to reduce nonspecific binding. For supershift experiments, 1 $\mu$l of anti-Smad2/3 antibody was added to the binding mixture and incubated at 4 °C for another 30 min. The mixtures were resolved in 6% nondenaturing polyacrylamide gel containing 2.5% glycerol in 0.5× TBE buffer. Electrophoretic mobility shift assay using purified full-length GST fusion Smad4 protein was performed as before (40).

*GST Pulldown*—GST-Smad2 or GST plasmid along with FLAG-Oct1 or FLAG-Gata6 expression plasmids were transfected into HEK293T cells using the cationic polymer polyethyleneimine and harvested after 48 h. Whole cell lysates were prepared using TNE lysis buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM EDTA, and 0.5% Nonidet P-40) containing a protease inhibitor mixture. The lysates were incubated with glutathione-Sepharose beads at 4 °C for 4 h. The beads were washed four times with TNE buffer. The bound proteins were separated by SDS-PAGE and visualized by Western blotting.

*DNA Oligonucleotide Precipitation*—Biotinylated double-stranded oligonucleotides were synthesized by Sangon (Shanghai, China). The oligonucleotide sequences were as follows: *arl4l*-oligo-WT, 5′-GTGTTGTACAGACTACGCTTC-3′; *arl4l*-oligo-MT, 5′-GTGTTGTACAAACTACGCTTC-3′; *lefty1*-oligo-WT, 5′-TTGAATAGGTCTGTCTTTCGT-3′; *lefty1*-oligo-WT, 5′-TTGAATAGGTTTGTCTTTCGT-3′, where the Smad2/4-binding motifs are underlined and the mutated bases are in boldface type. HEK293 cells transfected with indicated expression plasmids were cultured for 48 h and then treated with TGF-$\beta$1 (5 ng/ml) for another 2 h before harvest. DNA oligonucleotide precipitation was carried out as described previously (41). DNA-bound proteins were then subjected to SDS-PAGE and visualized by Western blotting.

*Luciferase Reporter Assay*—HaCaT cells were cultured in 24-well plates and transfected with the indicated constructs and the internal control pRenilla-TK vector. The transfected cells were serum-starved for 8 h before addition of TGF-$\beta$1 (5 ng/ml). The luciferase activity was quantified about 16 h later

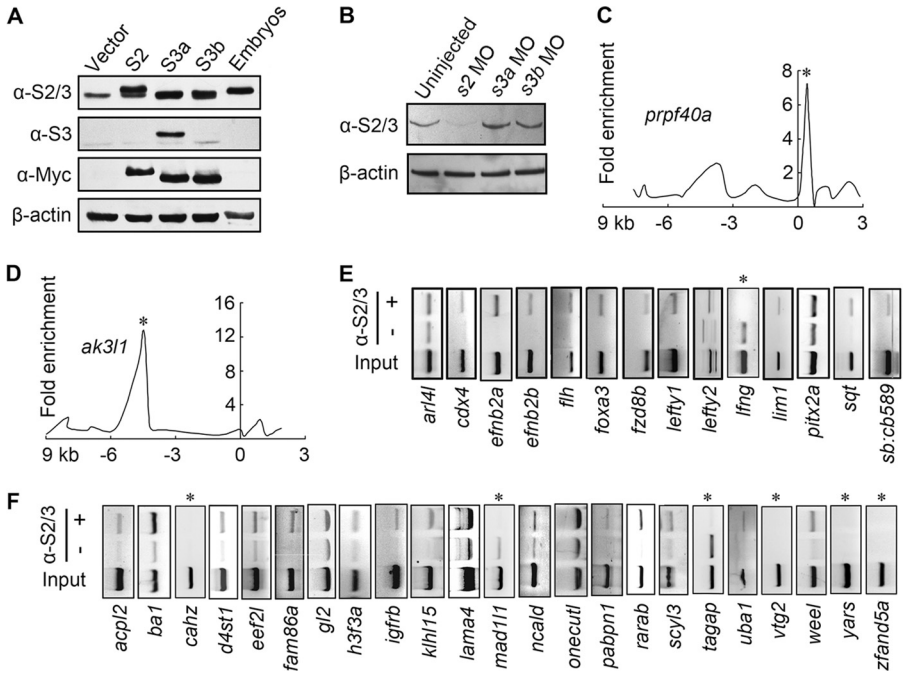


FIGURE 1. **Antibody specificity test and confirmation of Smad2-bound regions in zebrafish gastrulas.** *A* and *B*, affinity and specificity of the Smad2/3 antibody were examined by Western blotting. *A*, total proteins from HEK293T cells transfected with Myc-tagged zebrafish Smad2 (*S2*), Smad3a (*S3a*), and Smad3b (*S3b*) or from zebrafish 50% epiboly stage embryos were immunoblotted with the indicated antibodies. Note that endogenous Smad2 was detected by Smad2/3 antibody only. *B*, endogenous Smad2 in 50% epiboly stage embryos injected with *smad2*, *smad3a*, or *smad3b* (morpholino-oligonucleotides (*MOs*)) was immunoblotted with the Smad2/3 antibody. Injection doses are as follows: *smad2* morpholino-oligonucleotide, 10 ng; *smad3a* morpholino-oligonucleotide, 10 ng; *smad3b* morpholino-oligonucleotide, 10 ng. *C* and *D*, examples of signal distribution on the promoters of two genes. The peaks (Smad2-bound regions) were indicated by an *asterisk*. The transcription start sites were located at position 0. All the 556 identified SBRs to the promoters of 679 genes are shown in supplemental Table S1. *E* and *F*, SBRs of 14 previously known Nodal targets (*E*) or of 23 randomly chosen candidate targets (*F*) were validated by conventional ChIP assays with anti-Smad2/3 antibody or no antibody. 1/5× titration of input chromatin DNA was amplified by PCR as a positive control. Negatives were indicated by an *asterisk*.

using the Dual-Luciferase Assay (Promega). Each experiment was performed in triplicate, and the data represent the means ± S.D. of three independent experiments after normalization to *Renilla* activity.

## RESULTS

*Genome-wide Screening by ChIP-chip Uncovers 679 Smad2 Target Genes in Zebrafish Early Gastrulas*—To better understand the role of the Nodal signal in the early zebrafish gastrulas, we performed the ChIP-chip assay to identify direct target genes of the Nodal effectors Smad2 and Smad3. We first tested the affinity and specificity of the Smad2/3 and Smad3 antibodies, purchased from Cell Signaling Technology, by Western blot. As shown in Fig. 1*A*, the Smad2/3 antibody was able to detect Myc-tagged fish Smad2, Smad3a, and Smad3b, which were overexpressed in mammalian HEK293T cells, as well as endogenous Smad2 in 50% epiboly zebrafish embryos; in contrast, the Smad3 antibody detected Myc-Smad3a overexpressed in HEK293T cells but not the corresponding endogenous protein. Furthermore, the Smad2/3 antibody-recognized band was dramatically reduced in the embryo extract following knockdown of *smad2* with a specific morpholino (Fig. 1*B*), which was unaltered in *smad3a* or *smad3b* morphants. These results indicate that the Smad2/3 antibody is able to specifically recognize zebrafish endogenous Smad2. In a pilot ChIP-PCR experiment, this Smad2/3 antibody successfully enriched Nodal-responsive sequences of zebrafish *sqt* and *lim1/lhx1a* genes from the chromatin immunoprecipitates (data not shown), two previously characterized Nodal targets (42, 43), demonstrating its suitability for immunoprecipitating zebrafish chromatins.

For subsequent chromatin immunoprecipitation, zebrafish embryos at the 50% epiboly stage, at which gastrulation just started, were collected and used for ChIP with the Smad2/3 antibody. The immunoprecipitated DNA fragments were amplified and hybridized against the zebrafish promoter ChIP-on-chip microarray (Agilent Technology), which covers the promoter regions (−9 to 3 kb) representing 10,117 transcripts. As a control, we also carried out chromatin immunoprecipitation in the absence of specific antibodies. The SBRs were identified as peaks with the highest signal (for examples see Fig. 1, *C* and *D*). The genome-wide analysis, based on the seventh assembly of the zebrafish genome, Zv7, mapped a total of 556 identified SBRs to the promoters of 679 genes (supplemental Table S1). We observed high enrichments of the promoter sequences of 14 previously reported Nodal-regulated genes, including *sqt* and *lim1/lhx1a*. Then we tested whether the promoters of these 14 Nodal-regulated genes are actually bound by Smad2 in zebrafish early gastrulas using conventional ChIP/site-specific PCR analysis. As shown in Fig. 1*E*, the promoters of 13 of these 14 genes were confirmed to associate with Smad2 *in vivo*, suggesting an estimated positive rate of 92.8%. To further validate the identified Smad2 target genes, we randomly picked up 23 in alphabetical order from the identified 679 genes, which were not known previously to be Nodal targets, for site-specific PCR analysis after chromatin immunoprecipitation with the

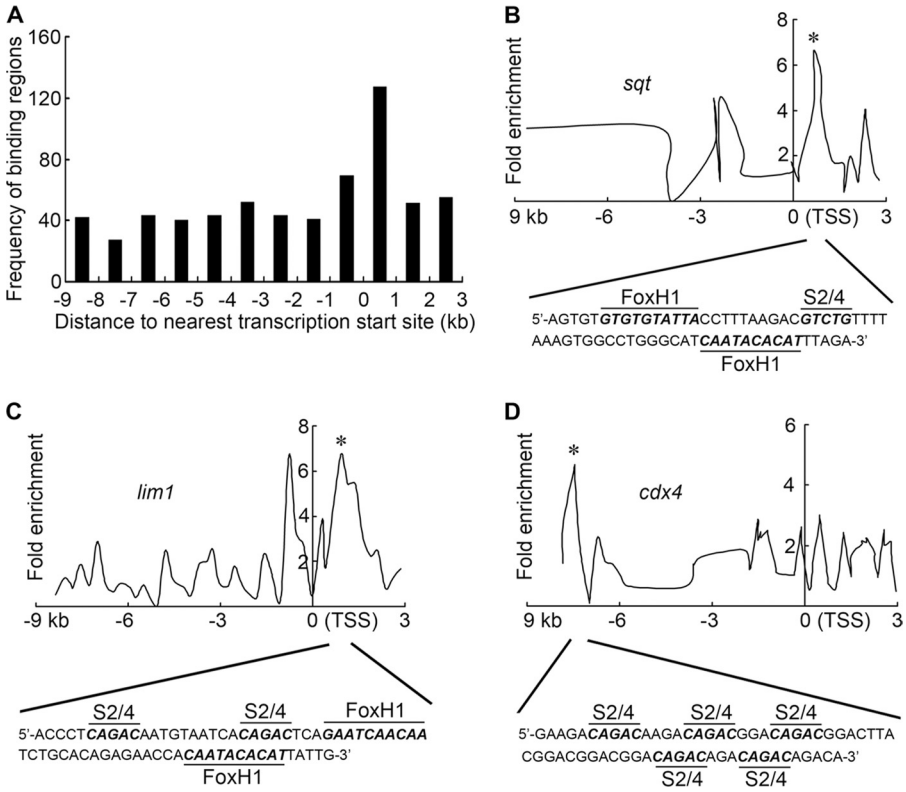


FIGURE 2. **General distributions of SBRs and identification of Smad2/4-binding motifs.** *A,* distribution of the identified SBRs along the interrogated promoter region. *B–D,* identification of Smad2/4-binding motifs within SBRs of three representative genes. The SBRs are indicated by an *asterisk*.

Smad2/3 antibody. Results indicated that 17 of them were bound by Smad2 (Fig. 1*F*), giving a positive ratio of 73.9%. Thus, the ChIP-chip approach was efficient and reliable for the genome-wide identification of Smad2 targets in our assay. We conclude that in zebrafish early gastrulas, the promoters of several hundreds of genes have already been occupied by Smad2, and as a result their transcription may be subjected to Smad2 regulation.

*SBRs Scatter in the Promoter/Enhancer Regions*—An interesting question is how the identified SBRs are distributed in the zebrafish genome. For revealing the distribution features, the 5′ end of each mapped transcript was defined as the TSS. We found that the SBRs were moderately enriched in close proximity to the TSS (1 kb around the TSS) and showed fairly even distribution in the other regions (Fig. 2*A*), which are typical of enhancers. It have been reported that Smad proteins usually form complexes with other transcription factors at nearby promoter or enhancer sites and function in an orientation-, position-, and distance-independent manner (44). Therefore, the enhancer-like distribution of the identified SBRs may reflect genuine locations of Smad2-bound sites in the zebrafish genome at the onset of gastrulation.

It is believed that Smad2 is not able to directly associate with DNA, but Smad2/4 complexes recognize and bind to the CAG(A/C)C motif via Smad4 (32). Then we looked into the existence of the consensus Smad2/4-binding motif CAG(A/C)C within the 556 SBRs, and we found that the Smad2/4-binding motif was significantly enriched in the SBRs compared with that in the background ($z$ score = 33.9 and $p < 10^{-9}$, which sufficiently exceeded the significance cutoffs). Most (537) of these SBRs, albeit scattering from distant promoter/enhancers to intron regions, contained at least one CAG(A/C)C motif within 500 bp from the peak signal position. For example, the first intron of *sqt* or *lim1/lhx1a* bears an SBR that showed the highest signal peak and contains one or two Smad2/4-binding motifs, respectively (Fig. 2, *B* and *C*). The examination of the SBR with the highest signal peak of a previously unknown Smad2 target gene, *cdx4*, also identified five Smad2/4-binding motifs in tandem (Fig. 2*D*). The results would suggest a lower possibility of off-targeting in our ChIP-chip assay.

*Smad2 Target Genes Are Mainly Involved in Regulation of Transcription, Signal Transduction, and Development and Responsive to Nodal Deficiency in Zebrafish Embryos*—We next asked in what kinds of functions the identified Smad2 target genes in early gastrulas may be involved. To address this question, we examined enrichment of GO in terms of the identified targets. Because the annotations of zebrafish genes are still underway, only 5876 of 10,117 genes represented in the promoter microarray and only 294 of the identified Smad2 target genes have been comprehensively annotated; thus, these 294 genes were used for the GO term analysis. We found that three GO terms were markedly enriched. The first one was the signaling component ontology, which was enriched to 12.9 from 4.5% (Fig. 3*A*). More importantly, most (29/38) of the signaling components that were identified as Smad2 targets fell into Nodal, bone morphogenetic protein, Wnt, FGF, and G-protein signaling pathways, which are known to be active at the onset of gastrulation. The second enriched group was developmental regulators, accounting for 28.6% of the identified Smad2 target genes compared with 9.5% of the anno-

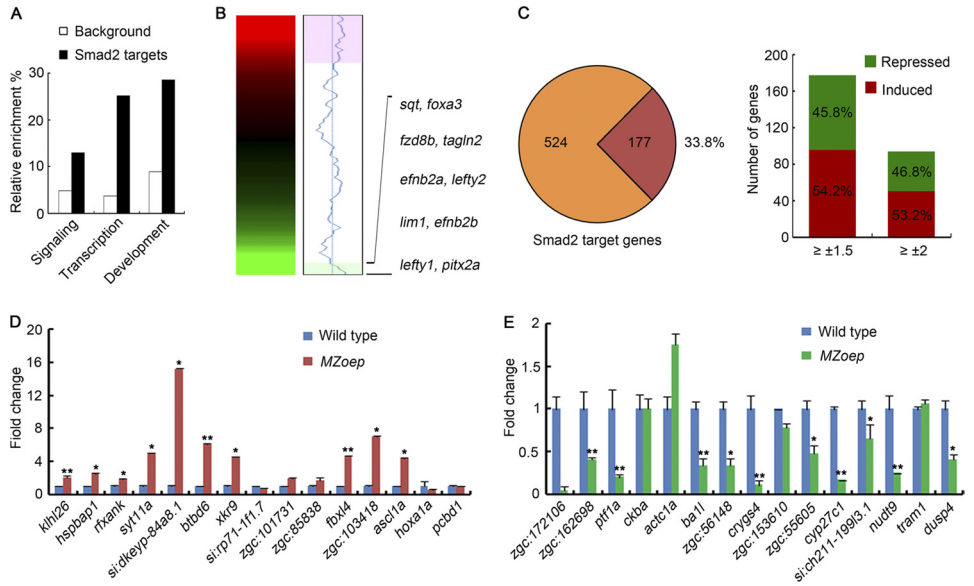


FIGURE 3. **Functional categories and nodal-dependent regulation of Smad2 target genes.** *A,* enrichment of three GO terms among the 294 comprehensive annotated Smad2 target genes. See supplemental Fig. S1 for the function categories among Smad2-targeted developmental regulators and see supplemental Fig. S2 for the putative regulatory network based on the Smad2-associated transcription factors. *B,* expression profiling of 524 genes, which were covered by both promoter and expression microarrays, in *MZoep* gastrulas using significance analysis of microarray analysis. The genes were sorted by the average expression ratio and means were centered. At *right* was the enrichment log ratio of probes associated with Smad2-binding sites. Ten highly down-regulated, previously known Nodal targets were indicated. *C,* distribution of differentially expressed Smad2 targets in *MZoep* mutants compared with wild-type embryos. The *amber circle* represented Smad2 target genes, and the *brown* area indicates those with a fold change of ≥1.5 in *MZoep* embryos. *Bar plots* show number and fractions of transcripts induced (*red*) or repressed (*green*) at different fold change cutoffs. See supplemental Table S2 for detail. *D* and *E,* expression levels of the top 15 *MZoep*-up (*D*) or *MZoep*-down (*E*) genes in wild-type or *MZoep* at the 50% epiboly stage were individually examined by quantitative RT-PCR. *Asterisks* indicated statistical significance of difference (*, $p < 0.01$; **, $p < 0.001$; $n = 3$). *Error bars* indicated S.D.

tated genes in the promoter array. Further analysis indicated that the enriched developmental regulators might participate in mesendoderm induction, axis formation, gastrulation, and nervous system development (supplemental Fig. S1), which are fully consistent with the developmental functions of Nodal signaling (11, 45, 46).

The third enriched group was transcription factors, which represented 25.2% of the comprehensively annotated Smad2 targets and were enriched nearly 7-fold over background levels (Fig. 3*A*). These transcription factors and their previously identified targets could form a large regulatory network of transcription factors (supplemental Fig. S2). The majority of the transcription factors within the regulatory network have been found to be involved at different developmental stages in mesendoderm specification, embryonic development, and tissue and organ formation. Importantly, the regulatory network seemed to converge on certain molecules such as Pitx2, Pax6, Gata6, Hif1a, FoxH1, and Dlx5, most of which are known to be important in the interrelated developmental progressions (supplemental Fig. S2). Taken together, these results imply that Nodal/Smad2 signaling exerts its developmental functions through initiating a cascade of regulatory events by controlling the expression of transcription factors, signaling molecules, and developmental regulators.

We next tested whether the transcription of these targets is truly subjected to Nodal regulation in zebrafish embryos at a matched developmental stage by microarray analysis. RNAs extracted from 50% epiboly stage wild-type or *MZoep* mutant embryos, which were defective in Nodal signaling (47, 48), were hybridized against the Agilent Zebrafish (V2) expression microarray representing over 40,000 transcripts. Among 679 identified Smad2 target genes, 524 (77.2%) were found to be present in the expression microarray, and a large proportion of these genes displayed an altered expression level in *MZoep* embryos (Fig. 3, *B* and *C*). Interestingly, 10 of 14 previously known Nodal-regulated genes were enriched in the group that was highly down-regulated in *MZoep* embryos, and their promoters were more enriched in Smad2-binding sites (Fig. 3*B*), which indicated that the expression of Nodal-regulated genes correlated with the Smad2 occupancy of their promoters. By using the median false discovery rate $p < 0.001$ and the fold change ≥1.5 as the criteria for significance, 177 (33.8%) of these Smad2 targets showed a remarkable change of expression in *MZoep* embryos (Fig. 3*C* and supplemental Table S2), which falls into reported ranges of regulated genes by other transcription factors (49, 50). We further validated by quantitative RT-PCR analysis the expression of the top 15 highly up- or down-regulated genes (those 10 previously well known Nodal regulated genes were excluded) in wild-type or *MZoep* embryos. As shown in Fig. 3, *D* and *E*, the transcription of these genes was up- or down-regulated in *MZoep* embryos in a way generally consistent with the expression microarray results. Taken together, these data suggest that the identified Smad2 targets are subjected to regulation by Nodal/Smad2 signaling during zebrafish embryogenesis.

*Smad2 Is Crucial for the Mesendodermal Expression of a Subset of Target Genes in Zebrafish Embryos*—Nodal and Smad2/3 signals play a pivotal role in mesendoderm induction in zebrafish (4, 19, 46, 51). We asked whether alterations of Smad2 activity would actually influence the expression of the identified Smad2 targets in the mesendoderm precursors located in the

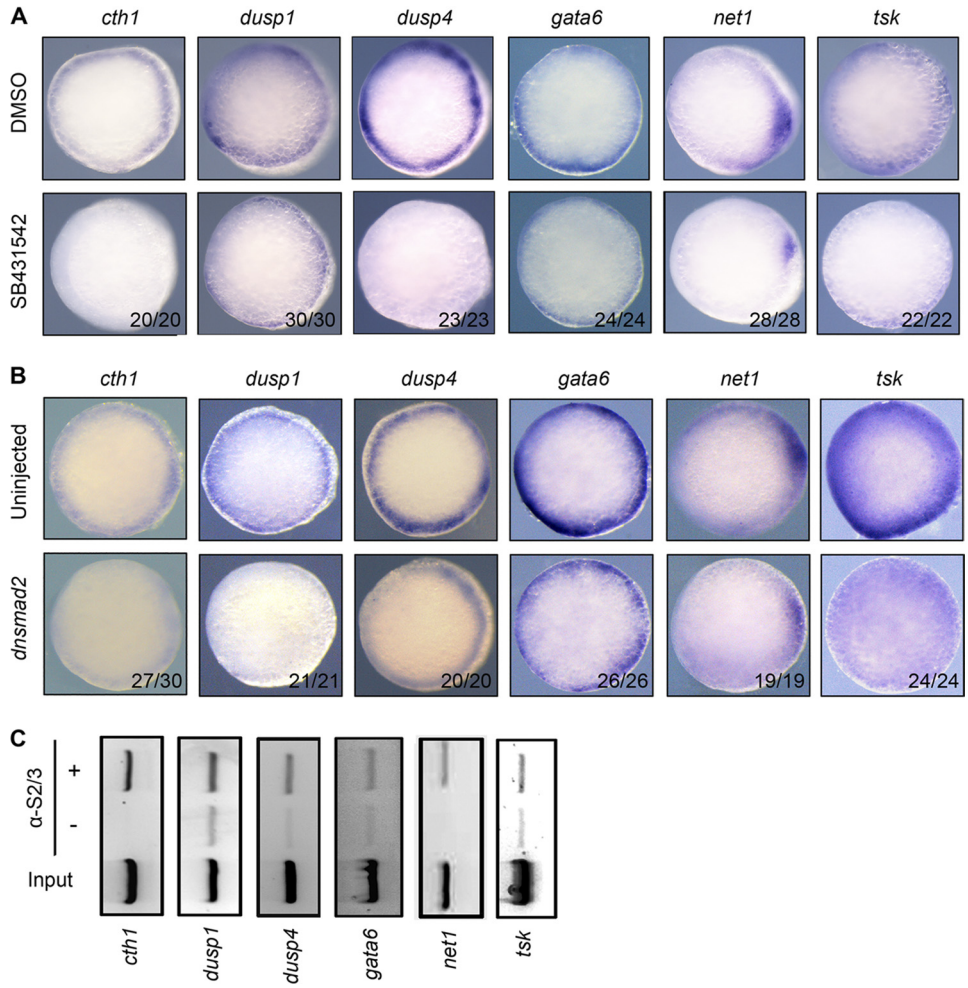


FIGURE 4. **Nodal/Smad2/3 signals regulate the mesendodermal expression of a subset of Smad2 target genes.** *A,* embryos were treated with DMSO or 75 $\mu$M SB431542 at the 16-cell stage and harvested at the 50% epiboly stage for *in situ* hybridization with the indicted antisense probes. The ratios of affected embryos are indicated. *B,* embryos injected with 475 pg of *dnsmad2* mRNA at the one-cell stage and harvested at the 50% epiboly stage for *in situ* hybridization. Embryos in *A* and *B* were animal-pole views with dorsal to the right. *C,* enrichment of SBRs in the promoters of indicated genes was detected in chromatin immunoprecipitates using the Smad2/3 antibody. 1/5× titration of input chromatin DNA was amplified by PCR as a positive control.

blastodermal margins. To address this question, we studied six genes, including *cth1, dusp1, dusp4, gata6, net1,* and *tsk,* which were selected from the identified Smad2 targets for their expression in mesendodermal precursors during early gastrulation. We demonstrated that down-regulation of Nodal signaling by treating 16-cell embryos with 75 $\mu$M of SB431542, an inhibitor of Nodal receptor (52), led to dramatic reduction of *cth1*, *dusp4*, *net1,* and *tsk* expression and moderate reduction of *dusp1* and *gata6* in the blastodermal margins at the 50% epiboly stage (Fig. 4*A*). Similarly, interfering with endogenous Smad2/3 activity by injecting dnsmad2 mRNA coding for a dominant negative form of Smad2 (19) decreased the expression of all of these genes in the blastodermal margins at varying degrees (Fig. 4*B*). We confirmed that the promoter regions of all of these genes were bound by endogenous Smad2 as detected by ChIP site-specific PCR analysis (Fig. 4*C*). These results suggest that Smad2 directly regulates the expression of a subset of genes in mesendoderm precursors at the onset of gastrulation.

*Spatiotemporal Regulation of efnb2b Is Directly Controlled by Smad2/4-binding Motif in the SBRs*—To exemplify the roles of SBRs in regulating gene expression, we fully investigated the transcription control of zebrafish *ephrin B2b* (*efnb2b*) (53), which has been previously identified as a Nodal-regulated gene in a microarray analysis (22) and shown to be a direct Smad2 target gene in this study. The SBR representing *efnb2b* possesses one CAGACCAGAC motif consisting of two successive CAGAC boxes and one upstream adjacent FoxH1-binding motif (CAATACACAA) and is located 2032 bp upstream of the transcription start site (Fig. 5*A*). Then a 3060-bp promoter region upstream of the *efnb2b* translation start site was amplified and fused to green fluorescent protein (GFP) to make the transgenic construct efnb2b-GFP. The embryos injected with efnb2b-GFP DNA showed mosaic GFP expression obviously in blastodermal margins and the embryonic shield at the shield stage, the presomitic mesoderm at the bud stage, and head at 24 h post-fertilization (Fig. 5, *B* and *D*), which were largely similar to the endogenous expression pattern of *efnb2b* (Fig. 5*C*) (53). Importantly, the reporter expression was obviously enhanced by co-injection of *sqt* or *casmad2* mRNA but almost eliminated by injection of *lefty1* (antagonist of Nodal signals) or *dnsmad2* mRNA (Fig. 5, *D* and *E*), which suggest that the 3060-kb *efnb2b* promoter contains Nodal/Smad2-responsible elements.

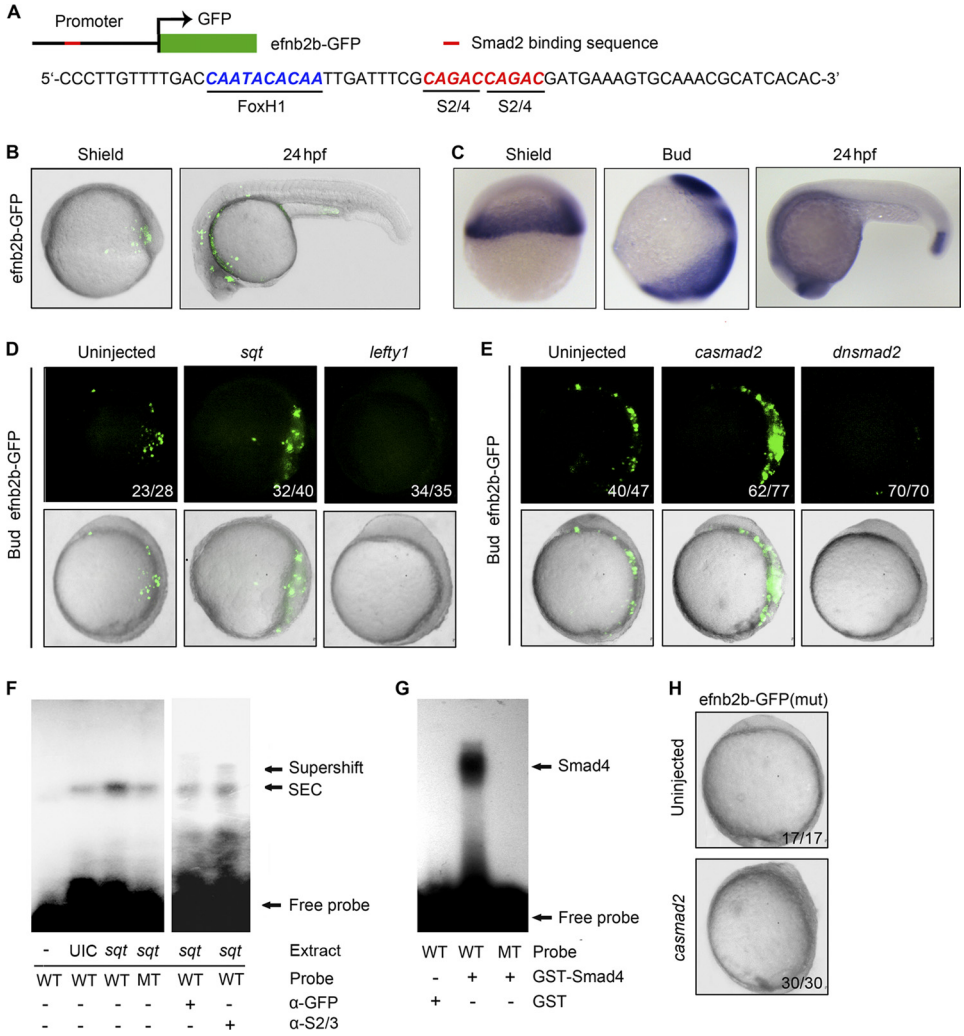


FIGURE 5. **Smad2 regulates the spatiotemporal expression of *efnb2b* gene.** *A,* illustration of efnb2b-GFP construct. The putative FoxH1- and Smad2/Smad4-binding motifs are located in the SBR of the *efnb2b* promoter. *B,* superimposed fluorescent images of living-shield stage and 24-h post-fertilization (*24hpf*) embryos injected with 100 pg of efnb2b-GFP construct at the one-cell stage. *C,* expression pattern of *efnb2b* at indicated stages, detected by *in situ* hybridization. *D* and *E,* dark field (*upper panel*) or superimposed (*lower panel*) fluorescent images of living bud-stage embryos injected with 100 pg of efnb2b-GFP construct at the one-cell stage following injection of 0.5 pg *sqt,* 90 pg *lefty1*, 110 pg *casmad2,* or 475 pg *dnsmad2* mRNA. The ratios of embryos with fluorescence similar to the shown picture are indicated. *F,* gel shift image showing the protein complex formation on the labeled Smad2/4-binding sequence in the *efnb2b* promoter. The $^{32}$P-radiolabeled probes were incubated with cell lysates of zebrafish 50% epiboly embryos that were uninjected or injected with 0.5 pg of *sqt* mRNA at the one-cell stage. Note the existence of *sqt*-enhanced complexes (*SEC*) and of a supershift band following addition of Smad2/3 antibody. *UIC*, uninjected embryo control. *G,* gel shift image showing that purified GST-Smad4 specifically bound to the labeled wild-type probe. *H,* superimposed fluorescent image of living bud-stage embryos injected with 100 pg of efnb2b(mut)-GFP construct at the one-cell stage. No fluorescence was detected even if *casmad2* mRNA was co-injected.

Next, we tested whether Smad2 could form complexes on the *efnb2b* promoter region consisting of the putative FoxH1- and Smad2/4-binding motifs. For electrophoretic mobility shift assays (EMSA), synthetic double-stranded 45-bp oligonucleotides (5′-GTTTTGAC**CAATACACAA**TTGATTTCGCAGACCAGACGATGAAAG-3′, boldface is for the FoxH1 site and underline is for Smad2/4 motif), which was derived from the *efnb2b* promoter region, as well as its mutant form (5′-GTTTTGACCAATACACAATTGATTTCGCAAACCAAACGATGAAAG-3′, two substituted bases are underlined), were radiolabeled and incubated with protein extracts from the 50% epiboly stage wild-type or *sqt*-injected embryos. The native gel electrophoresis revealed the existence of a band representing *sqt*-enhanced complexes during probe migration, which was strongly enhanced by *sqt* overexpression (Fig. 5*F*). Importantly, the addition of the Smad2/3 antibody to the incubation mixture resulted in the appearance of a supershift band, indicating that Smad2 was associated with the wild-type probe. Using purified GST-Smad4 protein for EMSA, we found that Smad4 also bound to the wild-type but not mutant probe (Fig. 5*G*). Thus, both Smad2 and Smad4 are components of the transcription complex forming on the CAGACCAGAC motif of the *efnb2b* promoter.

To further clarify the importance of the CAGACCAGAC motif of the *efnb2b* promoter for transcription, we generated an efnb2b(mut)-GFP construct by mutating the CAGACCAGAC motif (to CAAACCAAAC) of the construct efnb2b-GFP. Injection of efnb2b(mut) into zebrafish embryos failed to confer GFP expression even if *casmad2* mRNA was co-injected (Fig. 5*H*). This suggests that the identified CAGACCAGAC motif is essential for the *efnb2b* promoter activity.

**TABLE 1**

**Identification of other transcription factor binding motifs co-enriched with Smad2/4-binding motif in SBRs**

DNA sequences of 100 bp far from the Smad-binding motifs were analyzed, and the 200-bp sequences upstream of the TSS of zebrafish Refseq mRNA were chosen as the control set. The enriched motifs with top emergence frequency (≥20%, $Z \geq 10$, and $p < 0.01$) are shown. The factors that are previously known to interact with Smad proteins are written in boldface. All the enriched motifs ($Z \geq 10$, $p < 0.01$) are shown in supplemental Table S3.

| Factors | Emergence frequency | Factors | Emergence frequency |
|---|---|---|---|
| | % | | % |
| **FAC1** | 42.09 | Gata2 | 23.28 |
| Oct1 | 36.69 | Gfi1 | 23.09 |
| Crx | 31.66 | **Gata3** | 23.09 |
| **Sox9** | 30.17 | Pitx2 | 23.09 |
| **Foxh1** | 30.17 | Nkx2.2 | 23.09 |
| **Lef1** | 29.61 | DBP | 23.09 |
| **Evi1** | 27.37 | **Foxp3** | 22.91 |
| **Pax6** | 26.63 | AIRE | 22.91 |
| **Gata1** | 26.26 | MRF2 | 22.53 |
| **Pdx1** | 26.26 | **NF-AT** | 21.97 |
| **AP3** | 25.33 | **Fos** | 21.04 |
| **Tlx-2** | 24.95 | Helios A | 20.67 |
| **BRCA1** | 24.39 | Pbx1 | 20.30 |
| Gata6 | 23.46 | **YY1** | 20.11 |

*Smad2/4-binding Motif Clusters with Other Important Regulatory Sites*—The transcription complexes forming on a gene promoter usually consist of dozens of factors to initiate and maintain gene transcription. We suspected that the identified SBRs might contain, apart from the consensus Smad2/4-binding motif CAG(A/C)C, the binding motifs of other transcription factors that might function together with Smad2 in development. By searching TRANSFAC and JASPAR, we found 95 enriched binding motifs ($Z \geq 10$, $p < 0.01$) of diverse transcription factors around the consensus Smad2/4-binding motifs (supplemental Table S3), and the identified enriched motifs with top emergence frequency (emergence frequency ≥20%) in the 200-bp peaks containing the Smad2/4-binding motif are shown in Table 1. One of enriched motifs with top emergence frequency within 100 bp from the Smad2/4-binding motif was the FoxH1/Fast1-binding motif AAT(C/A)(A/C)ACA (38, 54), which normally interacts with Smad complexes to transduce Activin/Nodal signals during early embryonic development (43, 55). For example, the first intron of *sqt* or *lim1/lhx1a* holds an SBR containing one or two adjacent FoxH1- and Smad2/4-binding motifs, respectively (Fig. 2, *B* and *C*); the identified SBR of *efnb2b* also contained FoxH1- and Smad2/4-binding motifs in tandem (Fig. 5*A*). The other enriched binding sites with top emergence frequency included those for FAC1, T-cell factor/lymphoid enhancing factor (Tcf/Lef), Sox9, Evi1, Pax6, Gata1, and so on (Table 1), which have also been found to cooperate with Smad signaling to regulate gene expression (32, 56, 57). Taken together, we propose that cooperative control of gene expression by Smad2 and other transcription factors can be accomplished by gathering their binding sites in the gene promoters.

*Smad2 Cooperates with Maternal β-Catenin to Regulate Dorsal Gene Expression*—Tcf/Lef factors function as a cofactor of β-catenin in the nucleus (58), and maternal β-catenin plays a central role in the formation of the dorsal organizer in vertebrate embryos (59–62). Given that the Tcf/Lef1-binding motif and the Smad2/4-binding motif are clustered in a subset of the identified Smad2 targets, we asked if the expression of the Smad2 targets in the embryonic shield (zebrafish dorsal organizer) was coordinately regulated by β-catenin and Smad2. We tested two of the identified Smad2 target genes, *arl4l* and *lefty1*, which are expressed in the dorsal margin during gastrulation. The promoters of *arl4l* and *lefty1* contain a Smad2/4-binding motif and a nearby Tcf/Lef1-binding motif, which are conserved in three fish species (Fig. 6*A*). The corresponding promoter regions of both genes in early gastrulas could be immunoprecipitated using the anti-Smad2/3 antibody (Fig. 1*E*). Furthermore, an *in vitro* pulldown assay demonstrated that Smad2 could be pulled down by these putative Smad2/4-binding motifs, which was enhanced by co-expression of Smad4 (Fig. 6*B*). Furthermore, the corresponding promoter regions of both genes could also be immunoprecipitated by an anti-β-catenin antibody (Fig. 6*C*), suggesting that Smad2 and β-catenin/Tcf/Lef could reside in the same transcription factor complexes. The overexpression of Δ*N*-β-catenin mRNA, which encodes an activated form of β-catenin (63), or *casmad2* mRNA led to enhanced or even ectopic dorsal expression of *arl4l* and *lefty1* at the 30% epiboly and the shield stages (Fig. 6, *D* and *E*). When Δ*N*-β-catenin/*dnsmad2* or *casmad2*/Δ*N-tcf3* (encoding a dominant negative form of Tcf3) were co-injected or Δ*N*-β-catenin injection was combined with SB431542 treatment, however, the expression of *arl4l* and *lefty1* in the dorsal margin was inhibited (Fig. 6, *D* and *E*), suggesting an interdependence of Tcf/Lef factors and Smad2. These results indicate that Smad2 and β-catenin can coordinately regulate the dorsal expression of the target genes by binding to the neighboring Smad2- and Tcf/Lef-binding motifs in the promoters during embryogenesis.

*Oct1 and Gata6 Affect Mesendoderm Induction by Interacting with Smad2*—We noted a high emergence frequency for binding motifs of the transcription factors Oct1/Pou2f1, Crx, BRCA1, and Gata6 (Table 1). Because these factors have not been found previously to interact with Smad2/3 in transcriptional regulation, we decided to investigate their interaction with Smad2. Because zebrafish *crx* is not expressed before mid-segmentation periods (64) and the zebrafish *brca1* gene has not been identified, we paid attention to *oct1*, which is ubiquitously expressed during early embryogenesis (65), and *gata6,* which is expressed in the blastodermal margin and the shield (66). We first inspected physical interactions of Oct1 or Gata6 with Smad2 by GST pulldown assay. FLAG-tagged zebrafish Oct1 (FLAG-Oct1) or Gata6 (FLAG-Gata6) was detected in GST-captured zebrafish Smad2 (GST-Smad2) co-expressed in HEK293T cells (Fig. 7*A*). We then overexpressed FLAG-tagged zebrafish Oct1 or Gata6 by injecting their mRNAs into embryos. The endogenous Smad2 was also able to be co-immunoprecipitated by the overexpressed Oct1 and Gata6 (Fig. 7*B*). All these data suggest a physical interaction of Oct1 or Gata6 with Smad2. We next investigated the effect of Oct1 and Gata6 on Smad2 transcriptional activity using the TGF-β-responsive activin-response element (ARE)-luciferase reporter (67). Transfection of Oct1 into HaCat cells significantly promoted TGF-β-induced expression of the ARE-luciferase reporter, and co-transfection of Oct1 and Smad2 further enhanced the reporter expression (Fig. 7*C*). It is apparent that Oct1 binds to

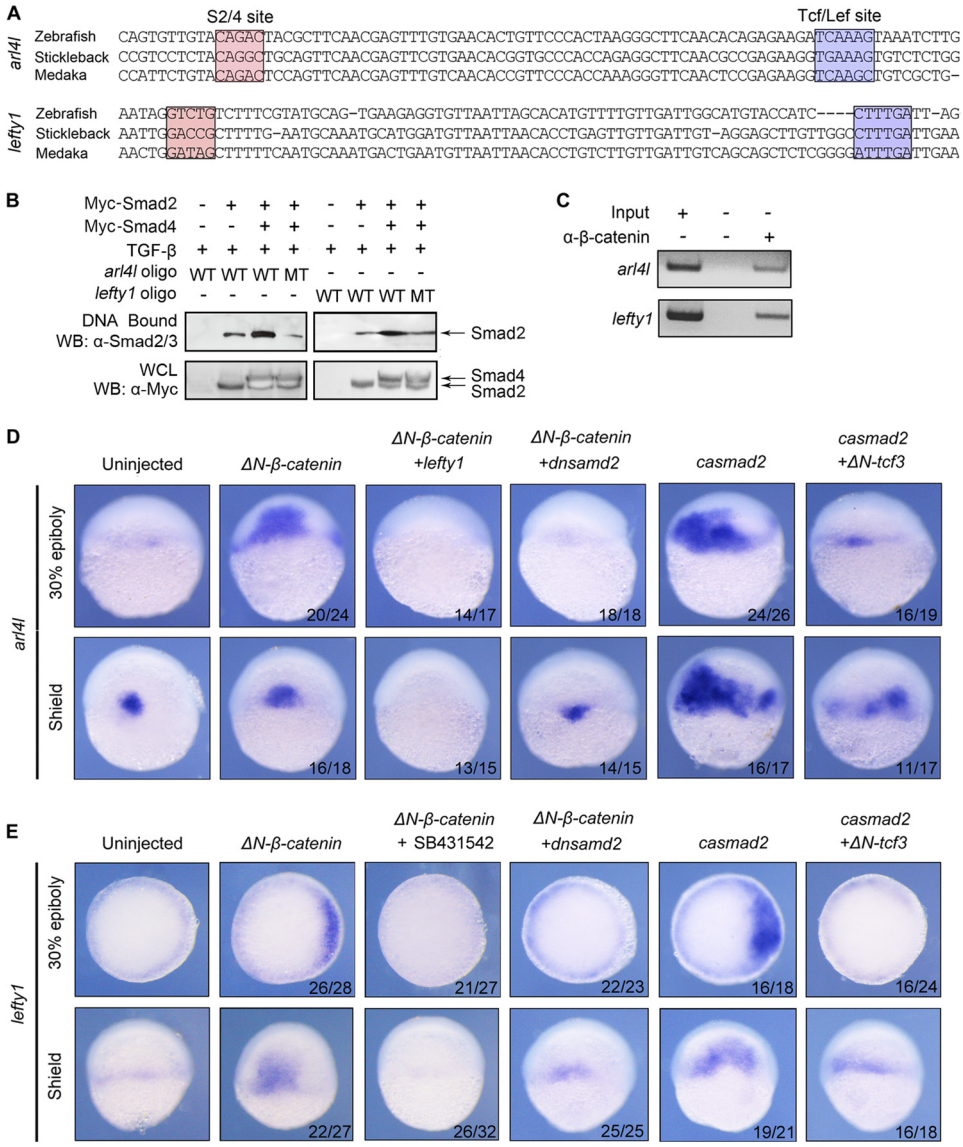


FIGURE 6. **Smad2/4- and Tcf/Lef-binding motifs in SBRs co-regulate expression of *arl4l* and *lefty1* in early gastrulas.** *A,* sequence alignments of putative Smad2/4- and Tcf/Lef-binding motifs of *arl4l* and *lefty1* from different species. *B,* Smad2 binds to the Smad2/4-binding motifs via Smad4. HEK293 cells were transfected with expression plasmids as indicated and treated with TGF-β1 (5 ng/ml) for 2 h. The cell lysates were harvested for oligonucleotide precipitation assays with the biotin-labeled double-stranded oligonucleotides corresponding to the Smad2/4-binding motifs of *arl4l* and *lefty1* promoters. The oligonucleotide (*oligo*) sequences are described under "Experimental Procedures." DNA-bound Smad2 (*top*) and total protein levels (*bottom*) were analyzed using anti-Smad2/3 and anti-Myc immunoblotting, respectively. *WB*, Western blot; *WCL*, whole cell lysate. *C,* sequences containing the putative Smad2/4- and Tcf/Lef-binding motifs in the *arl4l* and *lefty1* promoters were enriched in chromatin immunoprecipitates with anti-β-catenin antibody. 1/5× titration of input chromatin DNA was PCR-amplified as a positive control. For the affinity and specificity of the anti-β-catenin antibody, see supplemental Fig. S4. *D* and *E,* embryos injected with indicated mRNAs at the one-cell stage and harvested at the 30% epiboly or shield stage for *in situ* hybridization to detect alteration of *arl4l* (*D*) and *lefty1* (*E*) expression. The indicated embryos were treated with 75 μM SB431542 at the 16-cell stage. Injection doses are as follows: Δ*N-β-catenin*, 120 pg; *lefty1*, 90 pg; *dnsmad2*, 475 pg; *casmad2*, 110 pg; Δ*N-tcf3*, 100 pg. *arl4l,* dorsal views with animal pole to the top; *lefty1, top panel*, animal-pole view with dorsal to the *right*, and *lower panel*, dorsal views with animal pole to the top.

and enhances the transcriptional activity of Smad2. In contrast, Gata6 was found to attenuate TGF-β and/or Smad2-stimulated expression of the ARE-luciferase reporter (Fig. 7*D*), suggesting an inhibitory effect on Smad2 transcriptional activity.

Next, we briefly examined the potential cooperation of Oct1, Gata6, and Smad2 in mesendoderm induction of zebrafish embryos. The overexpression of *oct1* enhanced *casmad2*-induced expression of the mesoderm marker *ntl* and the endoderm marker *sox32* at shield stage, although oct1 overexpression alone did not obviously increase these marker expressions (Fig. 7, *E* and *F*). In contrast, *gata6* overexpression alone inhibited *ntl* and *sox32* expression, and its co-overexpression with *casmad2* compromised *casmad2*-induced *ntl* and *sox32* expression (Fig. 7, *E* and *F*). These results for the first time indicate that Oct1 cooperates with Smad2 to promote mesendodermal induction, whereas Gata6 inhibits mesendodermal induction by Smad2 signaling.

## DISCUSSION

Nodal signals are essential for mesendoderm induction during early development of vertebrate embryos. Smad2 and Smad3 are well known effectors of Nodal signals. Although the

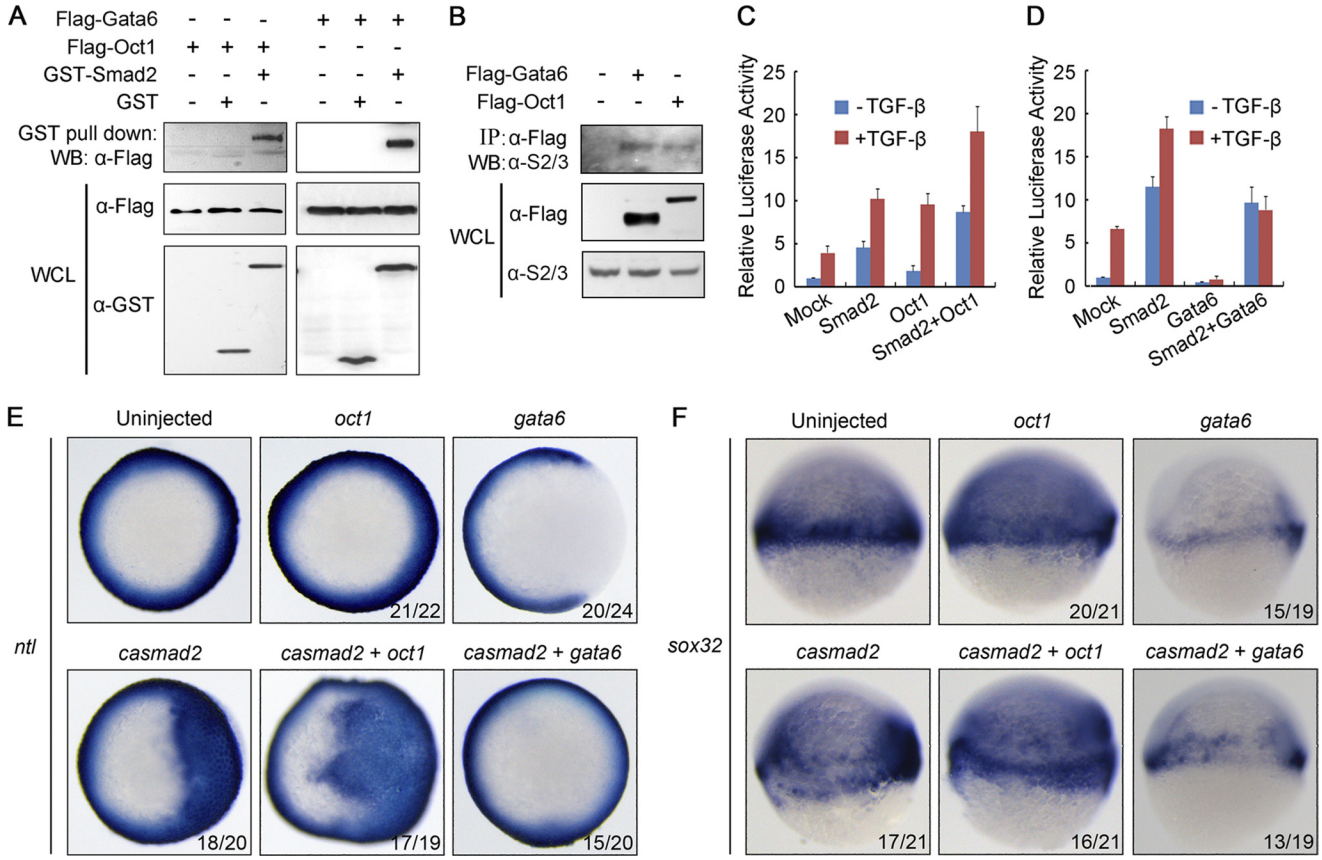


FIGURE 7. **Oct1 and Gata6 associate with and regulate transcriptional activity and function of Smad2.** *A,* physical interaction of Oct1 or Gata6 with Smad2. GST-Smad2 or GST plasmid along with FLAG-Oct1 or FLAG-Gata6 expression plasmids were transfected into HEK293T cells. At 48 h post-transfection, the cells were harvested for GST affinity purification. Smad2-associated Oct1 or Gata6 was detected by Western blotting using anti-FLAG antibody. The amount of individual proteins in the whole cell lysates (*WCL*) was assessed by Western blotting (*WB*) using corresponding antibodies. *B,* endogenous Smad2 co-immunoprecipitated with overexpressed Gata6 and Oct1. Embryos injected with FLAG-Oct1 or FLAG-Gata6 mRNA were collected at the 50% epiboly stage for co-immunoprecipitation using anti-FLAG antibody. FLAG-Oct1- or FLAG-Gata6-associated endogenous Smad2 was detected by Western blotting. *C* and *D,* effect of Oct1 (*C*) or Gata6 (*D*) on ARE-luciferase reporter expression. HaCat cells were co-transfected with ARE-luciferase and various expression plasmids as indicated and were treated or not treated with TGF-β1 (5 ng/ml) for 16 h before harvesting for luciferase assay. *Error bars* indicated S.D. *E* and *F, oct1* or *gata6* regulation of the mesodermal marker *ntl* (*E*) and the endodermal marker *sox32* (*F*). Embryos were injected individually with *casmad2*, *oct1*, or *gata6* mRNA at a dose of 110, 130, or 200 pg, respectively, or injected in combination, at the one-cell stage, and harvested at shield stage for *in situ* hybridization. Embryos in *E* were animal-pole view with dorsal to the *left*; embryos in *F* were lateral views with dorsal to the *right* and animal pole at *top*.

genome-wide analyses of Smad2/3-binding sites in HaCaT keratinocytes and mouse ES cells are previously reported (68, 69), the mechanisms of Nodal signaling acting through Smad2/3 targets during early embryonic development remain largely unclear. Upon the availability of an antibody recognizing zebrafish Smad2, in this study we successfully performed ChIP-chip assay in zebrafish gastrulas to identify direct targets of Nodal/Smad2 signals at the genome-wide level. We identified 679 Nodal/Smad2 targets whose promoter sequences are occupied by Smad2 at the 50% epiboly stage. These Smad2 targets are enriched in transcription factors, signaling molecules, and developmental regulators. Within the Smad2-bound regions of the target genes, the Smad2/4-binding motif CAG(A/C)C frequently clusters with the binding sites of other transcription factors, allowing cooperative regulation of mesendoderm induction by Smad2 and other factors. Gene expression profiling using microarray analysis at the matched developmental stage revealed that a subset of Smad2 targets is obviously up- or down-regulated by Nodal signaling. Thus, we provide a reservoir of Nodal/Smad2 targets for further elucidating mechanisms of Nodal/Smad2/3 functions in early development of vertebrate embryos.

Our ChIP-chip analysis initially identified 556 SBRs in the genome of zebrafish 50% epiboly embryos. Some of these SBRs are located in the intergenic regions, implying that one SBR may regulate the transcription of both genes located on either side. Therefore, the identified SBRs are subsequently mapped to the promoter regions of 679 genes, and these genes are believed to be Smad2 targets. Most of SBRs contain at least one consensus Smad2/4-binding motif. The Smad2/4-binding motifs present in the target gene promoters are frequently found to cluster with the binding sites of other transcription factors such as FoxH1 and Tcf/Lef, which have been shown to cooperate with the Smad2/3 proteins in regulating the transcription of target genes. More importantly, the genome-wide identification of SBRs opens up the possible roles of other regulatory factors in Nodal/Smad2-regulated transcription and developmental processes. As demonstrated in this study, the binding motifs for Oct1 and Gata6 are significantly enriched in SBRs, which predicts their potential implications in Nodal/

Smad2/3-regulated developmental processes. Our brief analyses demonstrate that Oct1 indeed physically associates with Smad2 and promotes Smad transcription activity and function in mesendodermal induction. In contrast, Gata6 interacts with Smad2 and inhibits transcription activity and mesendoderm induction activity of Smad2.

Many of the identified Smad2 targets in this study are transcription factors, signaling molecules, and developmental regulators. This fact agrees with the roles of Nodal signals in mesendoderm induction and patterning during late blastulation and early gastrulation (11, 45, 46). The identified Smad2 targets may act to relay Nodal/Smad2/3 signals to control the cell fates in early embryos. The functions of individual targets in embryonic development can be investigated in the future. Those targets can also be used as markers to investigate the timing and magnitude of Nodal/Smad2/3 signaling during embryonic development. For example, our previous study has demonstrated that maternal TGF-$\beta$/Nodal signals may be activated immediately after fertilization in zebrafish embryos (52), which conflicts with the classic view that zygotic gene transcription occurs at the midblastula transition (70). The timing of Smad2 occupancy on the promoters of the identified Smad2 target genes would provide evidence for the timing of Nodal/Smad2 signal activation.

Expression microarray analysis revealed that 177/524 of Smad2 targets showed an altered expression level in the 50% epiboly *MZoep* mutant embryos that are defective in Nodal signaling (47, 48). The expression level of the other Smad2 targets is not significantly changed in *MZoep* mutants probably because of the following: (*a*) their subtle changes in spatial pattern without alteration of overall transcript amount are unable to be detected by expression microarray analysis; (*b*) their expression changes might be more obvious at certain developmental stages. Other studies also note that only a fraction of targets identified by ChIP-chip is regulated by the corresponding factors as assayed by microassay expression profiling (49, 50). Nevertheless, the identified Smad2 targets could be useful subjects for other studies related to TGF-$\beta$ signaling and functions.

*Acknowledgments—We are grateful to members of the Anming Meng Laboratory for assistance and discussion.*

## REFERENCES


1. Conlon, F. L., Lyons, K. M., Takaesu, N., Barth, K. S., Kispert, A., Herrmann, B., and Robertson, E. J. (1994) *Development* **120,** 1919–1928
2. Iannaccone, P. M., Zhou, X., Khokha, M., Boucher, D., and Kuehn, M. R. (1992) *Dev. Dyn.* **194,** 198–208
3. Zhou, X., Sasaki, H., Lowe, L., Hogan, B. L., and Kuehn, M. R. (1993) *Nature* **361,** 543–547
4. Feldman, B., Gates, M. A., Egan, E. S., Dougan, S. T., Rennebeck, G., Sirotkin, H. I., Schier, A. F., and Talbot, W. S. (1998) *Nature* **395,** 181–185
5. Sampath, K., Rubinstein, A. L., Cheng, A. M., Liang, J. O., Fekany, K., Solnica-Krezel, L., Korzh, V., Halpern, M. E., and Wright, C. V. (1998) *Nature* **395,** 185–189
6. Agius, E., Oelgeschläger, M., Wessely, O., Kemp, C., and De Robertis, E. M. (2000) *Development* **127,** 1173–1183
7. Hirokawa, N., Tanaka, Y., Okada, Y., and Takeda, S. (2006) *Cell* **125,** 33–45
8. Raya, A., and Izpisúa Belmonte, J. C. (2006) *Nat. Rev. Genet.* **7,** 283–293
9. De Robertis, E. M., and Kuroda, H. (2004) *Annu. Rev. Cell Dev. Biol.* **20,** 285–308
10. Little, S. C., and Mullins, M. C. (2006) *Birth Defects Res. C Embryo Today* **78,** 224–242
11. Schier, A. F., and Talbot, W. S. (2005) *Annu. Rev. Genet.* **39,** 561–613
12. Moustakas, A., and Heldin, C. H. (2009) *Development* **136,** 3699–3714
13. Shi, Y., and Massagué, J. (2003) *Cell* **113,** 685–700
14. Nomura, M., and Li, E. (1998) *Nature* **393,** 786–790
15. Weinstein, M., Yang, X., Li, C., Xu, X., Gotay, J., and Deng, C. X. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95,** 9378–9383
16. Yang, X., Letterio, J. J., Lechleider, R. J., Chen, L., Hayman, R., Gu, H., Roberts, A. B., and Deng, C. (1999) *EMBO J.* **18,** 1280–1291
17. Zhu, Y., Richardson, J. A., Parada, L. F., and Graff, J. M. (1998) *Cell* **94,** 703–714
18. Dick, A., Mayr, T., Bauer, H., Meier, A., and Hammerschmidt, M. (2000) *Gene* **246,** 69–80
19. Jia, S., Ren, Z., Li, X., Zheng, Y., and Meng, A. (2008) *J. Biol. Chem.* **283,** 2418–2426
20. Jia, S., Wu, D., Xing, C., and Meng, A. (2009) *Dev. Biol.* **333,** 273–284
21. Sun, Z. H., and Meng, A. M. (2007) *Prog. Biochem. Biophys.* **34,** 595–603
22. Bennett, J. T., Joubin, K., Cheng, S., Aanstad, P., Herwig, R., Clark, M., Lehrach, H., and Schier, A. F. (2007) *Dev. Biol.* **304,** 525–540
23. Birch-Machin, I., Gao, S., Huen, D., McGirr, R., White, R. A., and Russell, S. (2005) *Genome Biol.* **6,** R63
24. Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., and Furlong, E. E. (2007) *Genes Dev.* **21,** 436–449
25. Zeitlinger, J., Zinzen, R. P., Stark, A., Kellis, M., Zhang, H., Young, R. A., and Levine, M. (2007) *Genes Dev.* **21,** 385–390
26. Akkers, R. C., van Heeringen, S. J., Jacobi, U. G., Janssen-Megens, E. M., Françoijs, K. J., Stunnenberg, H. G., and Veenstra, G. J. (2009) *Dev. Cell* **17,** 425–434
27. Kidder, B. L., and Palmer, S. (2010) *Genome Res.* **20,** 458–472
28. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., and Schilling, T. F. (1995) *Dev. Dyn.* **203,** 253–310
29. 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., and Smith, J. C. (2006) *Genome Biol.* **7,** R71
30. Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002) *Nucleic Acids Res.* **30,** e15
31. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) *Nat. Genet.* **25,** 25–29
32. Feng, X. H., and Derynck, R. (2005) *Annu. Rev. Cell Dev. Biol.* **21,** 659–693
33. Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S., Kinzler, K. W., Vogelstein, B., and Kern, S. E. (1998) *Mol. Cell* **1,** 611–617
34. Ho Sui, S. J., Mortimer, J. R., Arenillas, D. J., Brumm, J., Walsh, C. J., Kennedy, B. P., and Wasserman, W. W. (2005) *Nucleic Acids Res.* **33,** 3154–3164
35. Marstrand, T. T., Frellsen, J., Moltke, I., Thiim, M., Valen, E., Retelska, D., and Krogh, A. (2008) *PLoS One* **3,** e1623
36. Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A. E., and Wingender, E. (2006) *Nucleic Acids Res.* **34,** D108–D110
37. Bryne, J. C., Valen, E., Tang, M. H., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., and Sandelin, A. (2008) *Nucleic Acids Res.* **36,** D102–D106
38. Silvestri, C., Narimatsu, M., von Both, I., Liu, Y., Tan, N. B., Izzi, L., McCaffery, P., Wrana, J. L., and Attisano, L. (2008) *Dev. Cell* **14,** 411–423
39. Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003) *Bioinformatics* **19,** 185–193
40. Ma, J., Wang, Q., Fei, T., Han, J. D., and Chen, Y. G. (2007) *Blood* **109,** 987–994
41. Zhang, S., Fei, T., Zhang, L., Zhang, R., Chen, F., Ning, Y., Han, Y., Feng, X. H., Meng, A., and Chen, Y. G. (2007) *Mol. Cell. Biol.* **27,** 4488–4499

42. Fan, X., Hagos, E. G., Xu, B., Sias, C., Kawakami, K., Burdine, R. D., and Dougan, S. T. (2007) *Dev. Biol.* **310,** 363–378
43. Watanabe, M., Rebbert, M. L., Andreazzoli, M., Takahashi, N., Toyama, R., Zimmerman, S., Whitman, M., and Dawid, I. B. (2002) *Dev. Dyn.* **225,** 448–456
44. Whitman, M. (1998) *Genes Dev.* **12,** 2445–2462
45. Shen, M. M. (2007) *Development* **134,** 1023–1034
46. Tian, T., and Meng, A. M. (2006) *Cell. Mol. Life Sci.* **63,** 672–685
47. Gritsman, K., Zhang, J., Cheng, S., Heckscher, E., Talbot, W. S., and Schier, A. F. (1999) *Cell* **97,** 121–132
48. Zhang, J., Talbot, W. S., and Schier, A. F. (1998) *Cell* **92,** 241–251
49. Li, Z., Van Calcar, S., Qu, C., Cavenee, W. K., Zhang, M. Q., and Ren, B. (2003) *Proc. Natl. Acad. Sci. U.S.A.* **100,** 8164–8169
50. Marson, A., Kretschmer, K., Frampton, G. M., Jacobsen, E. S., Polansky, J. K., MacIsaac, K. D., Levine, S. S., Fraenkel, E., von Boehmer, H., and Young, R. A. (2007) *Nature* **445,** 931–935
51. Dougan, S. T., Warga, R. M., Kane, D. A., Schier, A. F., and Talbot, W. S. (2003) *Development* **130,** 1837–1851
52. Sun, Z., Jin, P., Tian, T., Gu, Y., Chen, Y. G., and Meng, A. (2006) *Biochem. Biophys. Res. Commun.* **345,** 694–703
53. Chan, J., Mably, J. D., Serluca, F. C., Chen, J. N., Goldstein, N. B., Thomas, M. C., Cleary, J. A., Brennan, C., Fishman, M. C., and Roberts, T. M. (2001) *Dev. Biol.* **234,** 470–482
54. Zhou, S., Zawel, L., Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) *Mol. Cell* **2,** 121–127
55. Whitman, M. (2001) *Dev. Cell* **1,** 605–617
56. Landry, J., Sharov, A. A., Piao, Y., Sharova, L. V., Xiao, H., Southon, E., Matta, J., Tessarollo, L., Zhang, Y. E., Ko, M. S., Kuehn, M. R., Yamaguchi, T. P., and Wu, C. (2008) *PLoS Genet.* **4,** e1000241
57. Itoh, S., Itoh, F., Goumans, M. J., and Ten Dijke, P. (2000) *Eur. J. Biochem.* **267,** 6954–6967
58. Korinek, V., Barker, N., Morin, P. J., van Wichen, D., de Weger, R., Kinzler, K. W., Vogelstein, B., and Clevers, H. (1997) *Science* **275,** 1784–1787
59. De Robertis, E. M., Larraín, J., Oelgeschläger, M., and Wessely, O. (2000) *Nat. Rev. Genet.* **1,** 171–181
60. Heasman, J., Crawford, A., Goldstone, K., Garner-Hamrick, P., Gumbiner, B., McCrea, P., Kintner, C., Noro, C. Y., and Wylie, C. (1994) *Cell* **79,** 791–803
61. Larabell, C. A., Torres, M., Rowning, B. A., Yost, C., Miller, J. R., Wu, M., Kimelman, D., and Moon, R. T. (1997) *J. Cell Biol.* **136,** 1123–1136
62. Schneider, S., Steinbeisser, H., Warga, R. M., and Hausen, P. (1996) *Mech. Dev.* **57,** 191–198
63. Xiong, B., Rui, Y., Zhang, M., Shi, K., Jia, S., Tian, T., Yin, K., Huang, H., Lin, S., Zhao, X., Chen, Y., Chen, Y. G., Lin, S. C., and Meng, A. (2006) *Dev. Cell* **11,** 225–238
64. Liu, Y., Shen, Y., Rest, J. S., Raymond, P. A., and Zack, D. J. (2001) *Invest. Ophthalmol. Vis. Sci.* **42,** 481–487
65. Li, M., Cao, Y., Zhao, Z. X., Lin, S., and Meng, A. M. (2001) *Chin. Sci. Bull.* **46,** 1523–1527
66. Seiliez, I., Thisse, B., and Thisse, C. (2006) *Dev. Biol.* **290,** 152–163
67. Chen, X., Weisberg, E., Fridmacher, V., Watanabe, M., Naco, G., and Whitman, M. (1997) *Nature* **389,** 85–89
68. Koinuma, D., Tsutsumi, S., Kamimura, N., Taniguchi, H., Miyazawa, K., Sunamura, M., Imamura, T., Miyazono, K., and Aburatani, H. (2009) *Mol. Cell. Biol.* **29,** 172–186
69. Fei, T., Zhu, S., Xia, K., Zhang, J., Li, Z., Han, J. D., and Chen, Y. G. (2010) *Cell Res.* **20,** 1306–1318
70. Kane, D. A., and Kimmel, C. B. (1993) *Development* **119,** 447–456