© 2021. Published by The Company of Biologists Ltd | Development (2021) 148, dev199538. doi:10.1242/dev.199538

RESEARCH ARTICLE

Two distinct motifs for Zic-r.a drive specific gene expression in two cell lineages

Izumi Oda-Ishii, Deli Yu and Yutaka Satou*

ABSTRACT

Zic-r.a, a maternal transcription factor, specifies posterior fate in ascidian embryos. However, its direct target, *Tbx6-r.b*, does not contain typical Zic-r.a-binding sites in its regulatory region. Using an *in vitro* selection assay, we found that Zic-r.a binds to sites dissimilar to the canonical motif, by which it activates *Tbx6-r.b* in a sub-lineage of muscle cells. These sites with non-canonical motifs have weak affinity for Zic-r.a; therefore, it activates *Tbx6-r.b* only in cells expressing Zic-r.a abundantly. Meanwhile, we found that Zic-r.a expressed zygotically in late embryos activates neural genes through canonical sites. Because different zinc-finger domains of Zic-r.a are important for driving reporters with canonical and non-canonical sites, it is likely that the non-canonical motif is not a divergent version of the canonical motif. In other words, our data indicate that the non-canonical motif represents a motif distinct from the canonical motif. Thus, Zic-r.a recognizes two distinct motifs to activate two sets of genes at two timepoints in development.

This article has an associated 'The people behind the papers' interview.

KEY WORDS: Ascidians, Transcription factor, Non-canonical-motif sites, Zic, *Tbx6*

INTRODUCTION

In ascidians, *Zic-r.a* (also called *Macho-1*), which encodes a Zic transcription factor, is expressed maternally and is essential for specification of muscle cells (Nishida and Sawada, 2001). *Zic-r.a* mRNA is localized in the most posterior cells, which have the potential to adopt a germ cell fate and are transcriptionally quiescent (magenta cells in Fig. 1A) (Kumano et al., 2011; Shirae-Kurabayashi et al., 2011). Therefore, Zic-r.a does not act in these most posterior cells, but instead acts in their sister cells (B5.1 and B6.4 in Fig. 1A), which contribute to muscle cells and are transcriptionally active. In these cells, Zic-r.a activates *Tbx6-r.b*, which encodes a key transcription factor for muscle cell specification (Yagi et al., 2005; Yu et al., 2019) at the 16-cell and 32-cell stages (Oda-Ishii et al., 2016; Yagi et al., 2004a).

At the 16-cell stage, *Tbx6-r.b* is expressed in B5.1 (brown cells in Fig. 1A) through the cooperative action of Zic-r.a, β-catenin and Tcf7 (Oda-Ishii et al., 2016). At the 32-cell stage, *Tbx6-r.b* is hardly detectable with *in situ* hybridization in daughter cells of B5.1, and it is again activated in B6.4 (green cells in Fig. 1A). Although *Zic-r.a* is similarly required for *Tbx6-r.b* expression in B6.4 (Yagi et al., 2004a), it has been suggested that β-catenin does not activate its target in these cells (Hudson et al., 2013). Thus, collectively, these studies suggest that *Tbx6-r.b* is regulated differently in these two muscle lineages. In the present study, by addressing the issue of this differential regulation, we found that *Tbx6-r.b* is activated in B6.4 of the 32-cell embryo through Zic-r.a binding to sites with non-canonical motifs, the nucleotide sequences of which are dissimilar to the canonical motif we previously determined using an *in vitro* selection assay (Yagi et al., 2004a).

Many studies have shown that low-affinity binding sites for transcription factors are important for driving temporally or spatially controlled gene expression. These include a study of a temporal control by Pha-4 in nematodes (Gaudet and Mango, 2002), spatial control of *Otx* in ascidians (Farley et al., 2015), and temporal and spatial control of *Svb* in flies (Crocker et al., 2015; Fuqua et al., 2020). In these studies, each binding site that transcription factors recognize is represented by a single motif, and low-affinity sites are regarded as its divergent versions. On the other hand, it has been postulated that some transcription factors recognize multiple motifs (Badis et al., 2009; Franco-Zorrilla et al., 2014), although the biological importance of this property has not been well clarified. For example, it was reported that mammalian Zic3 transcription factor recognizes two motifs, one of which resembles the canonical motif of ascidian Zic-r.a (Badis et al., 2009). In the present study, to clarify the biological importance of this property, we show that Zic-r.a has two recognition motifs that are not divergent versions of a single motif, and that these two distinct binding motifs are used for driving different sets of genes in early and late embryos.

RESULTS

Tbx6-r.b is regulated differently in the B5.1 and B6.4 lineages

Tbx6-r.b is activated by *Zic-r.a* in concert with β-catenin and Tcf7 in the B5.1 lineage at the 16-cell stage (Oda-Ishii et al., 2016). Therefore, we first examined whether *Tbx6-r.b* was similarly regulated in the B6.4 lineage (B6.4) (Fig. 1B-D). Previously, we showed that *Tbx6-r.b* expression is lost in the B6.4 lineage of 32-cell embryos injected with a morpholino antisense oligonucleotide (MO) against *Zic-r.a* (*Zic-r.a* morphants) (Yagi et al., 2004a) (Fig. 1C). On the other hand, in the present study, we discovered that *Tbx6-r.b* expression was not lost in B6.4 of *β-catenin* morphants at the 32-cell stage (Fig. 1D). Therefore, in contrast to *Tbx6-r.b* expression in the B5.1 lineage of 16-cell embryos, *Tbx6-r.b* expression in the B6.4 lineage of 32-cell embryos is regulated independently of β-catenin.

This result was further confirmed using reporter constructs. A construct containing the 189 bp upstream sequence of *Tbx6-r.b* (Fig. S1A) was expressed in the B5.1 lineage at the 16-cell stage, and a construct with mutations in two binding sites for Tcf7, which forms a complex with β-catenin, was not expressed, as reported previously (Oda-Ishii et al., 2016) (Fig. 1E,F). However, in the B6.4

Department of Zoology, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan.

*Author for correspondence (yutaka@ascidian.zool.kyoto-u.ac.jp)

Y.S., 0000-0001-5193-0708

Handling Editor: Cassandra Extavour
Received 16 February 2021; Accepted 26 April 2021

Development (2021) 148, dev199538. doi:10.1242/dev.199538

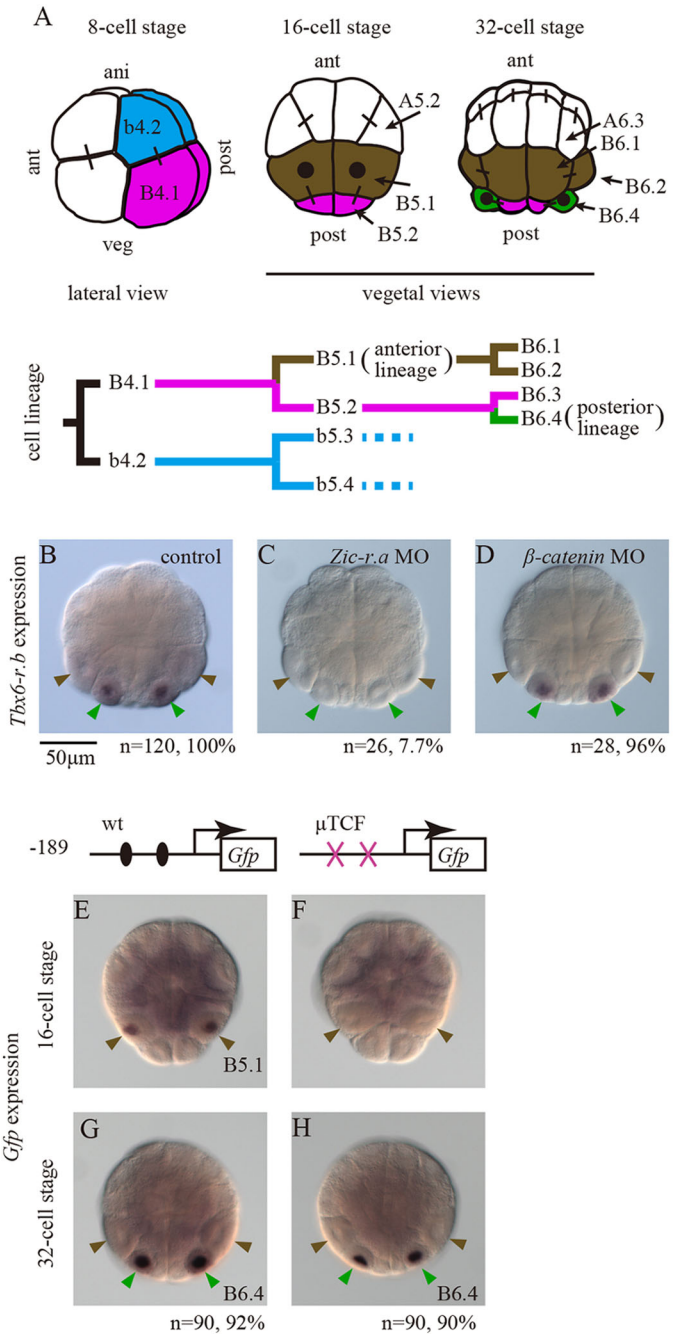

Fig. 1. *Tbx6-r.b* is activated differently in the B5.1 and B6.4 lineages. (A) Schematic illustrations of 8- to 32-cell embryos. Sister cell relationships are shown by short black lines. The most posterior cells, which can adopt a germ cell fate, are colored in magenta. Sister cells of the most posterior cells, which arose at the 8-, 16- and 32-cell stages are colored cyan, brown and green, respectively. Cells with *Tbx6-r.b* expression are indicated by black dots. Their cell lineages are shown below. (B-D) *Tbx6-r.b* expression in (B) control, (C) *Zic-r.a* morphant and (D) *β-catenin* morphant embryos at the 32-cell stage. Numbers of embryos examined and percentages of embryos that express *Tbx6-r.b* in B6.4 at the 32-cell stage are shown below. (E-H) Expression of reporter constructs containing the *Tbx6-r.b* upstream sequence with (E,G) intact and (F,H) mutated Tcf7-binding sites at the (E,F) 16- and (G,H) 32-cell stages. Reporter expression was examined using *in situ* hybridization. β-Catenin and Tcf7 form a complex to activate their targets. Numbers of embryos examined and percentages of embryos that expressed the reporter gene in B6.4 at the 32-cell stage are shown below G and H. Brown and green arrowheads indicate B5.1 lineage cells (B5.1 and B6.2) and B6.4 cells, respectively.

lineage of 32-cell embryos, both of these constructs were expressed (Fig. 1G,H). Thus, *Tbx6-r.b* expression in the B6.4 lineage is regulated independently of β-catenin/Tcf7 at the 32-cell stage.

The *Tbx6-r.b* upstream regulatory region contains putative Zic-r.a-binding sites similar to a non-canonical motif

Previously, we suggested that direct binding of Zic-r.a to DNA might be unnecessary for *Tbx6* expression in the B5.1 lineage at the 16-cell stage, because Zic-r.a can bind to DNA indirectly by interacting with Tcf7 (Oda-Ishii et al., 2016). However, our reporter analysis in the present study showed that Tcf7 binding sites are not necessary for *Tbx6-r.b* expression in the B6.4 lineage at the 32-cell stage. Therefore, we considered the possibility that *Zic-r.a* directly binds to the upstream region of *Tbx6-r.b* in the B6.4 lineage at the 32-cell stage.

Although Zic-r.a preferentially binds to sequences similar to 5′-GCAGCGGGGGG-3′ (Fig. 2A) and potential binding sites are identified between −1095 and −1547 nucleotide positions from the transcription start site of *Tbx6-r.b* (Kugler et al., 2010; Yagi et al., 2004a), no similar sequences are found in the 189 bp upstream sequence of *Tbx6-r.b*, using the computer program Patser (Hertz and Stormo, 1999) and the *Ciona* position weight matrix (Yagi et al., 2004a). Meanwhile, a study of mouse Zic proteins has shown that Zic proteins bind to two different sequences, with consensus sequences of ‘CCCCnnGGGG’ and ‘CnCAGCAGG’ (Fig. 2A) (Badis et al., 2009). Using the same computer program, we found no significant hits in the *Ciona Tbx6-r.b* upstream region with the matrix for the mouse primary canonical motif, which is more similar to the *Ciona* motif. However, it identified one significant hit with the matrix for the mouse secondary, non-canonical motif (Fig. 2B). Additionally, another sequence similar to the non-canonical motif was identified in a nearby region, although this was not significant (Fig. 2B). Hereafter, we call the distal and proximal sites Zic-d and Zic-p, respectively.

We examined whether these two sites could bind Zic-r.a *in vitro* using a gel-shift assay. We observed shifted bands, indicating binding of Zic-r.a to these two sites (Fig. 2C). Next, to compare Zic-r.a binding activity between canonical and non-canonical motifs, we added them to the reaction as competitors. The shifted band was weakened by incubation with a 20 or 50 molar excess of the non-canonical motif competitor, and almost completely disappeared when the non-canonical motif competitor was increased to a 100 or 200 molar excess (Fig. 2D). On the other hand, the shifted band disappeared in the presence of 5 molar excess of the canonical motif competitor, which we have used in our previous study (Yagi et al., 2004a) (Fig. 2D). These results show that Zic-r.a does bind to the non-canonical motif, but that its affinity for this motif is much weaker than for the canonical motif.

Finally, we introduced mutations into the Zic-r.a binding sites, Zic-d and Zic-p, in the reporter construct containing the 189 bp upstream sequence of *Tbx6-r.b*. These mutations reduced reporter expression in the B6.4 lineage of 32-cell embryos (Fig. 2E), although 35% of embryos still expressed the reporter; therefore, the reporter may contain additional unrecognized Zic-r.a-binding sites. Our observation indicates that Zic-r.a binds to the Zic-d- and Zic-p-binding sites to activate *Tbx6-r.b* in the B6.4 lineage at the 32-cell stage. In addition, these mutations also reduced reporter expression in B5.1 of 16-cell embryos (Fig. S2), so these sites contribute to expression in B5.1 (see Discussion).

The difference in Zic-r.a concentration in the two muscle sub-lineages is important for specific activation of the enhancer

We have previously demonstrated that Zic-r.a protein is detected in nuclei of B5.1 and B5.2 and in the posterior pole of B5.2 at the

Development (2021) 148, dev199538. doi:10.1242/dev.199538

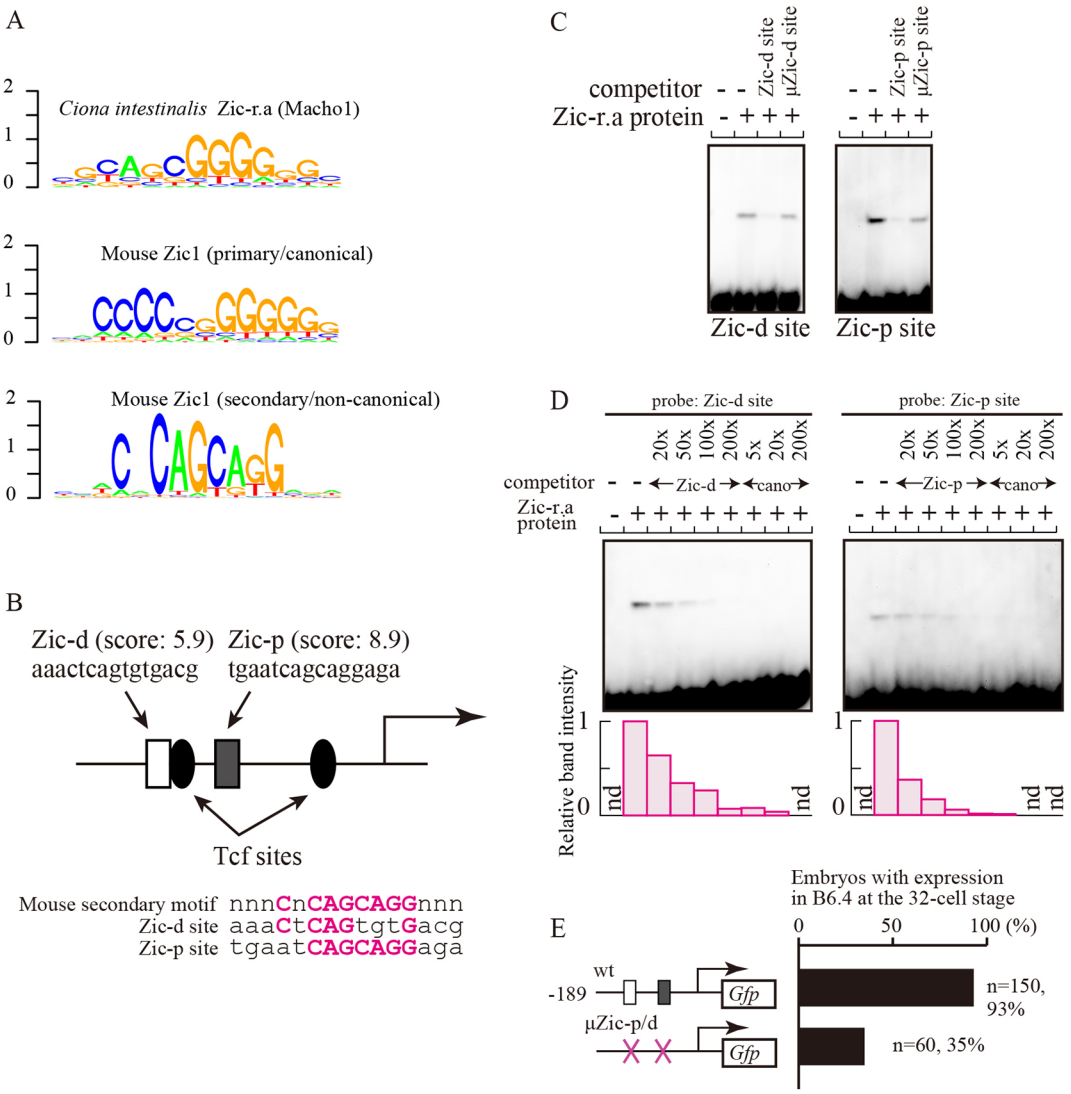

Fig. 2. The upstream sequence of *Tbx6-r.b* contains two Zic-r.a-binding sites with the non-canonical motif. (A) Sequence logos representing the binding motif of *Ciona* Zic-r.a, and the primary canonical and secondary non-canonical motifs of mouse Zic1. Sequence logos (Schneider and Stephens, 1990) were reproduced from data downloaded from the Uniprobe database (accession number UP00102) (Hume et al., 2015) and from previously published data (Yagi et al., 2004a). (B) A depiction of the *Tbx6-r.b* upstream region showing two possible Zic-r.a-binding sites similar to the mouse secondary non-canonical motif. Nucleotide sequences and scores obtained with the Patser program (Hertz and Stormo, 1999) for these two sites are indicated. (C) Gel-shift analysis showing that these two sites bind Zic-r.a *in vitro*. The shifted bands disappeared after incubation with specific competitors, but not with competitors for a mutant Zic-r.a-binding site. (D) Gel-shift analysis showing that the canonical motif binds Zic-r.a more strongly than does the non-canonical motif. Left and right panels show gel-shift assays for Zic-d and Zic-p sites, respectively. Two competitors for the labeled probes shown at the top were added in various amounts. Quantification of the shifted bands is shown in bar graphs. (E) Expression of a reporter construct containing the *Tbx6-r.b* upstream sequence with mutated Zic-r.a-binding sites in B6.4 of 32-cell embryos.

16-cell stage, using immunostaining with an anti–Zic-r.a antibody (Oda-Ishii et al., 2016). Immunostaining of 32-cell embryos with the same antibody showed that the Zic-r.a protein was barely detectable in the descendants of B5.1 (B6.1 and B6.2), and that it was detected in nuclei of B6.4 and B6.3, and in the posterior pole (Fig. 3A). This observation was confirmed by quantification of fluorescent signals in nuclei (Fig. 3B). Specifically, Zic-r.a is expressed strongly in cells in which *Zic-r.a* mRNA is localized (B5.2 in 16-cell embryos and B6.3 in 32-cell embryos) and in their sister cells (B5.1 in 16-cell embryos and B6.4 in 32-cell embryos). Although signal strengths in B6.4 were almost the same as in B6.3, where *Zic-r.a* mRNA is localized, they were significantly stronger than those in B5.1 of 16-cell embryos. These observations support the hypothesis that *Tbx6-r.b* is activated through direct binding of Zic-r.a to the Zic-p and Zic-d sites in the B6.4 lineage of 32-cell embryos, because a higher concentration of Zic-r.a is apparently necessary for sufficient binding to low-affinity sites. On the other hand, the absence of *Tbx6-r.b* mRNA in B6.3 is consistent with the earlier finding that the B6.3 cell pair is transcriptionally silent (Kumano et al., 2011; Shirae-Kurabayashi et al., 2011).

Overexpression of *Zic-r.a* evoked *Tbx6-r.b* ectopic expression at the 16-cell stage, even in the animal hemisphere, where no nuclear β-catenin is expected (Fig. 3C,C′). Our result was consistent with the hypothesis that a high concentration of Zic-r.a can evoke *Tbx6-r.b* expression independently of β-catenin, although we cannot completely rule out the possibility that *Zic-r.a* overexpression induced nuclear translocation of β-catenin ectopically in the animal hemisphere.

Development (2021) 148, dev199538. doi:10.1242/dev.199538

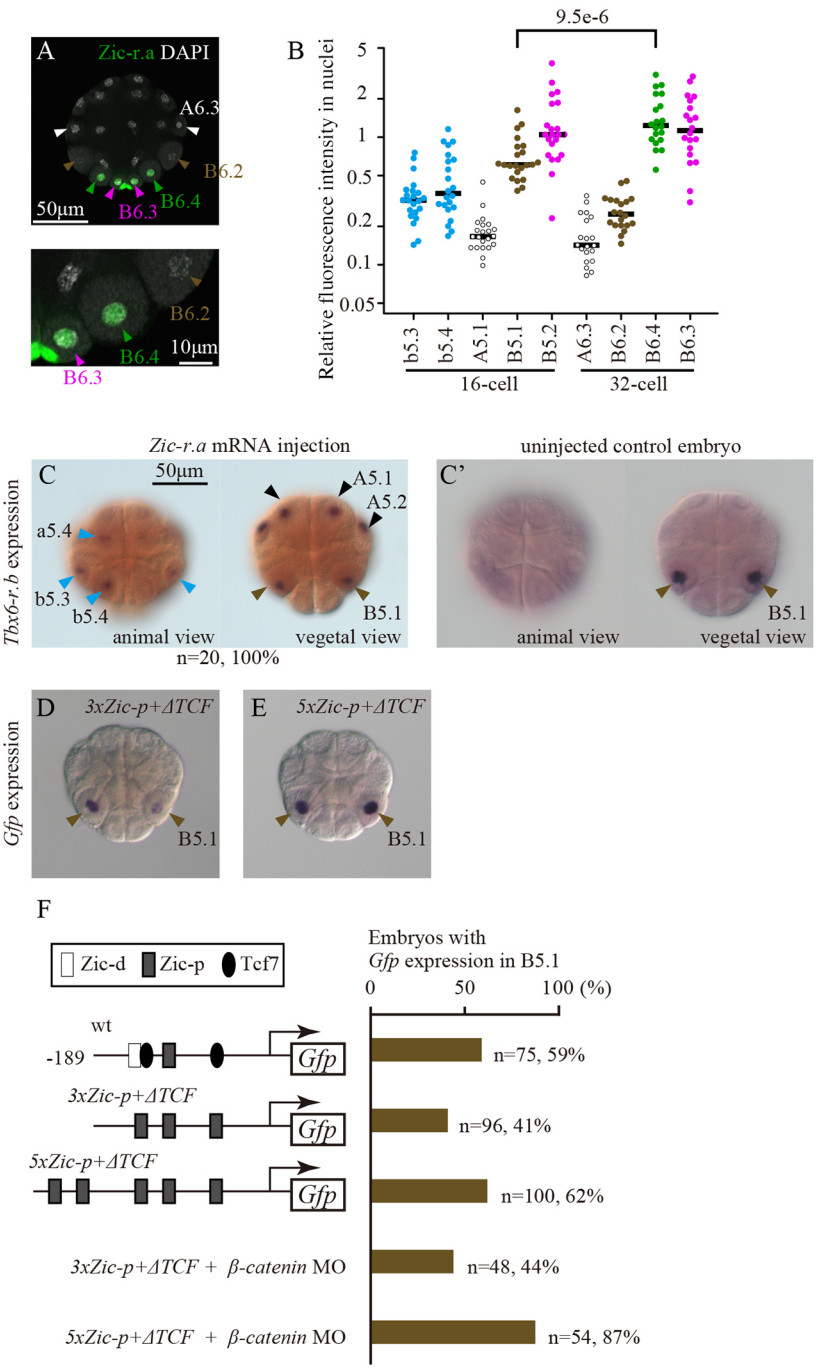

Fig. 3. Higher Zic-r.a concentration directly activates *Tbx6-r.b*. (A) Immunostaining signals for Zic-r.a at the 32-cell stage are green. Nuclei are shown using DAPI staining (gray). The lower panel is a higher magnification view of the right posterior part of the embryo. Photographs are *z*-projected image stacks and are overlaid in pseudo-color. (B) Quantification of fluorescence intensity in nuclei. Relative intensities compared with DAPI signals are shown on the *y*-axis. Individual measurements are shown with dots. Averages are shown by horizontal black lines. The difference in signal strengths between B5.1 and B6.4 was examined with the Wilcoxon rank sum test (P=9.5×10^{-6}). Data for the most posterior cells, which can adopt a germ cell fate, are colored in magenta. Data for sister cells of the most posterior cells, which arose at the 8-, 16- and 32-cell stages are colored cyan, brown and green, respectively. Data for cells in other lineages are shown by white dots. (C) Overexpression of *Zic-r.a* evokes ectopic expression of *Tbx6-r.b* (cyan and black arrowheads) in addition to expression in B5.1 (brown arrowheads) at the 16-cell stage. The number of embryos examined and the percentage of embryos that expressed *Tbx6-r.b* ectopically are shown below. (C′) Expression of *Tbx6-r.b* in control embryos is shown. (D-F) Mutant constructs, which lack Tcf7 binding sites and contain (D) three or (E) five Zic-p sites, are expressed in B5.1 at the 16-cell stage, as observed by *in situ* hybridization (see Fig. 1E for expression of the wild-type construct). (F) Percentages of embryos that expressed the reporter in B5.1 are shown in the graph.

To further test this hypothesis, we next used modified *Tbx6-r.b* reporter constructs, because it is expected that more binding sites recruit more Zic-r.a (Fig. S1BC). In the first construct (*3xZic-p+ΔTCF>Gfp*), the two Tcf-binding sites and the Zic-d site were replaced with Zic-p sites (three Zic-p sites in total). The second construct (*5xZic-p+ΔTCF>Gfp*) contains two additional Zic-p sites (five Zic-p sites in total). In spite of the lack of Tcf7 sites, *3xZic-p+ΔTCF>Gfp* was expressed in B5.1 at the 16-cell stage (Fig. 3D,F; compare it with Fig. 1F), and *5xZic-p+ΔTCF>Gfp* drove the reporter more frequently than did *3xZic-p+ΔTCF>Gfp* (Fig. 3E,F) (Fisher's exact test P=1.8e-08). These constructs indicate that the enhancer drives gene expression independently of β-catenin/Tcf7 when it binds enough Zic-r.a. Indeed, injection of the β-catenin MO rarely affected reporter expression (Fig. 3F).

Taken together, our data indicate that Zic-r.a binding to the *Tbx6-r.b* enhancer in normal 16-cell embryos is insufficient to activate *Tbx6-r.b*. This explains why β-catenin/Tcf7 is required at the 16-cell stage. Our data also indicate that the enhancer binds Zic-r.b sufficiently to activate *Tbx6-r.b* independently of β-catenin/Tcf7 in B6.4 cells of 32-cell embryos, but not in any other cells of 16-cell or 32-cell embryos.

Canonical-motif sites induce ectopic expression in cells with less Zic-r.a

To understand functional differences between the canonical and non-canonical-motif sites, we changed the Zic-d and Zic-p site sequences to the primary canonical motif sequence (*2xcano>Gfp*, in which 'ATCAGCAGGAGAG' (Zic-p) and 'CTCAGTGTGACGC'

Development (2021) 148, dev199538. doi:10.1242/dev.199538

(Zic-d) were changed to 'ATCAGCGGGGGGC' and 'CTCAGC GGGGGGC') (Fig. S1D). Although the control non-mutated upstream sequence specifically drove reporter expression only in B5.1 at the 16-cell stage (*wt>Gfp*), the mutated upstream sequence drove the reporter ectopically (Fig. 4A,B). Ectopic expression was found mainly in b-line cells, which are derived from cells where *Zic-r.a* mRNA is localized at the four-cell stage (see Fig. 1A). Among the cells of 16-cell embryos, these cells ranked third and fourth, after the germline cells (where transcription is quiescent) and B5.1, in the amount of Zic-r.a they contained (Fig. 3B).

In addition to maternal expression, *Zic-r.a* is zygotically expressed in neural cells at the tailbud stage (Satou et al., 2002). Judging from its mRNA expression pattern (Satou et al., 2002) and its protein distribution pattern (Oda-Ishii et al., 2016), it is likely that these neural cells are derived from the a-, b- and A-lineages. Concomitant with this zygotic expression, the above mutant construct was expressed ectopically in neural cells, in which *Zic-r.a* is expressed. However, it was also expressed ectopically at the gastrula stage in nerve cord cells in which Zic-r.a is not expressed. It is possible that a paralog Zic-r.b, which is expressed in nerve cord cells (Imai et al., 2002) and binds to a sequence similar to the canonical site (Yagi et al., 2004b), activates the reporter. Therefore, we used the 3.4 kbp upstream sequence, which was used by a previous study (Kugler et al., 2010). The construct containing the intact 3.4 kb upstream region was expressed specifically in B5.1 at the 16-cell stage (Fig. S3A) and was expressed only in muscle cell lineages at the tailbud stage (Fig. 4C), indicating that the 189 bp upstream sequence lacked elements responsible for repression at the gastrula stage. On the other hand, the mutant construct, which contained two canonical-motif sites instead of the non-canonical-motif sites, induced ectopic expression in neural cells at the tailbud stage (Fig. 4D), as well as ectopic expression in the animal hemisphere at the 16-cell stage (Fig. S3B). Thus, canonical-motif sites drove reporter expression in neural cells at the tailbud stage.

Genes expressed in neural cells of late embryos under the control of Zic-r.a have canonical-motif sites

The above observation raised the possibility that Zic-r.a, which is derived from zygotically expressed mRNA, regulates gene expression through canonical binding-motif sites in neural cells at the tailbud stage. To identify such genes, we obtained candidates from embryos in which neural fate was induced by upregulation of Zic-r.b activity and downregulation of activities of Foxa.a, Foxd, Neurog and Erk signaling (Kobayashi et al., 2018). Specifically, we compared transcriptomes of such embryos and embryos into which the *Zic-r.a* MO was also injected (Fig. 5A). Among greatly downregulated genes, we chose 20 genes (NOIseqsim $P>0.99$; reads per million values >50), and then examined whether they were expressed in cells with *Zic-r.a* expression, using published single-cell transcriptome data for middle and late tailbud embryos (Cao et al., 2019). We found that two genes were expressed in *Zic-r.a*-expressing cells at the middle and late tailbud stages. One (KY.Chr10.963) encodes Claudin,; the other (KY.Chr7.686) did not show a significant similarity to known proteins. Specifically, *Claudin* was expressed in a subset of cells with *Zic-r.a* expression, whereas the latter was expressed not only in cells with *Zic-r.a* expression, but also in other cell populations (Fig. 5B; Fig. S4).

To confirm whether zygotically expressed *Zic-r.a* controls these genes, we mutated *Zic-r.a* by injecting a pair of TALEN mRNAs whose products were designed to target the third zinc-finger domain. Because this injection cannot affect maternal *Zic-r.a* mRNA, we were able to examine effects of zygotically expressed *Zic-r.a*. To confirm that injection of these TALEN mRNAs successfully mutated *Zic-r.a*, we extracted genomic DNA and amplified this genomic site by PCR. PCR products were cloned into a plasmid vector and nucleotide sequences of 24 randomly selected clones were determined. All clones contained an insertion or deletion in this site, suggesting high efficiency of these TALEN mRNAs (Fig. S5). We next measured expression of the putative targets by reverse transcription, followed by quantitative PCR (RT-qPCR). We found that mRNA levels of *Claudin* and *KY.Chr7.686* were significantly decreased, while mRNA levels of ubiquitously expressed *EF1α* or a gene specifically expressed in notochord (*Noto1*) were not significantly changed (Fig. 5C). The effect upon the expression level of *KY.Chr7.686* was weaker than that of *Claudin*. This is probably because *KY.Chr7.686* is expressed not only in cells with zygotic *Zic-r.a* expression, but also in other cell populations (Fig. 5B; Fig. S4).

In the 1 kb upstream regions of *Claudin* and *KY.Chr7.686*, two and one canonical-motif sites were recognized by Patser (Hertz and Stormo, 1999) and the *Ciona* position weight matrix (Yagi et al.,

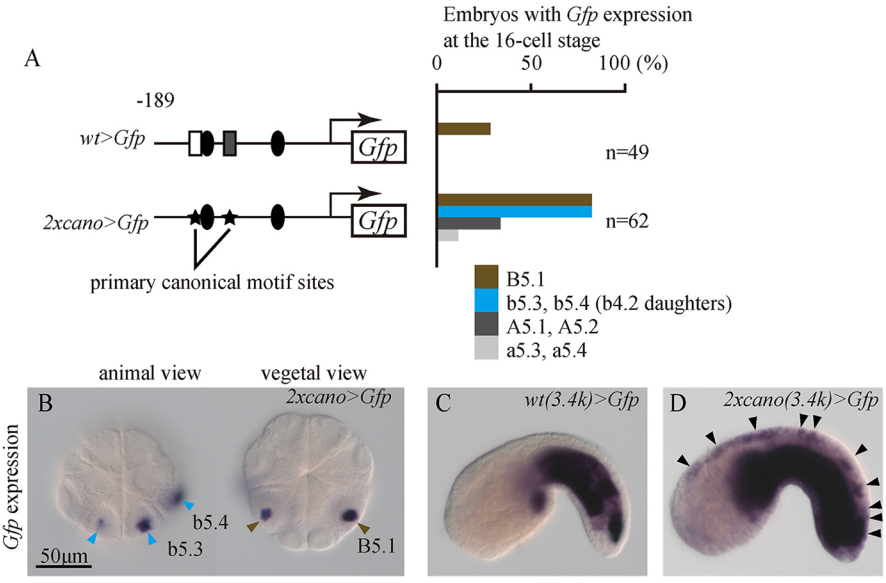

Fig. 4. A strong enhancer with the canonical-motif sequences evokes ectopic gene expression. (A) Percentages of embryos that expressed reporter constructs with intact Zic-r.a-binding sites (*wt>Gfp*) and with canonical-motif Zic-r.a-binding sites (*2xcano>Gfp*). Expression in four lineages was counted individually and is indicated with bars of different colors. (B) *In situ* hybridization examining expression of *2xcano>Gfp* at the 16-cell stage. Ectopic expression is indicated by cyan arrowheads (see Fig. 1E for expression of the wild-type construct). Brown arrowheads indicate expression in B5.1. (C,D) *In situ* hybridization examining expression of (C) the construct containing the intact 3.4 kb upstream region and (D) the construct containing two canonical-motif sites at the middle tailbud stage. Black arrowheads indicate ectopic expression in neural cells.

Development (2021) 148, dev199538. doi:10.1242/dev.199538

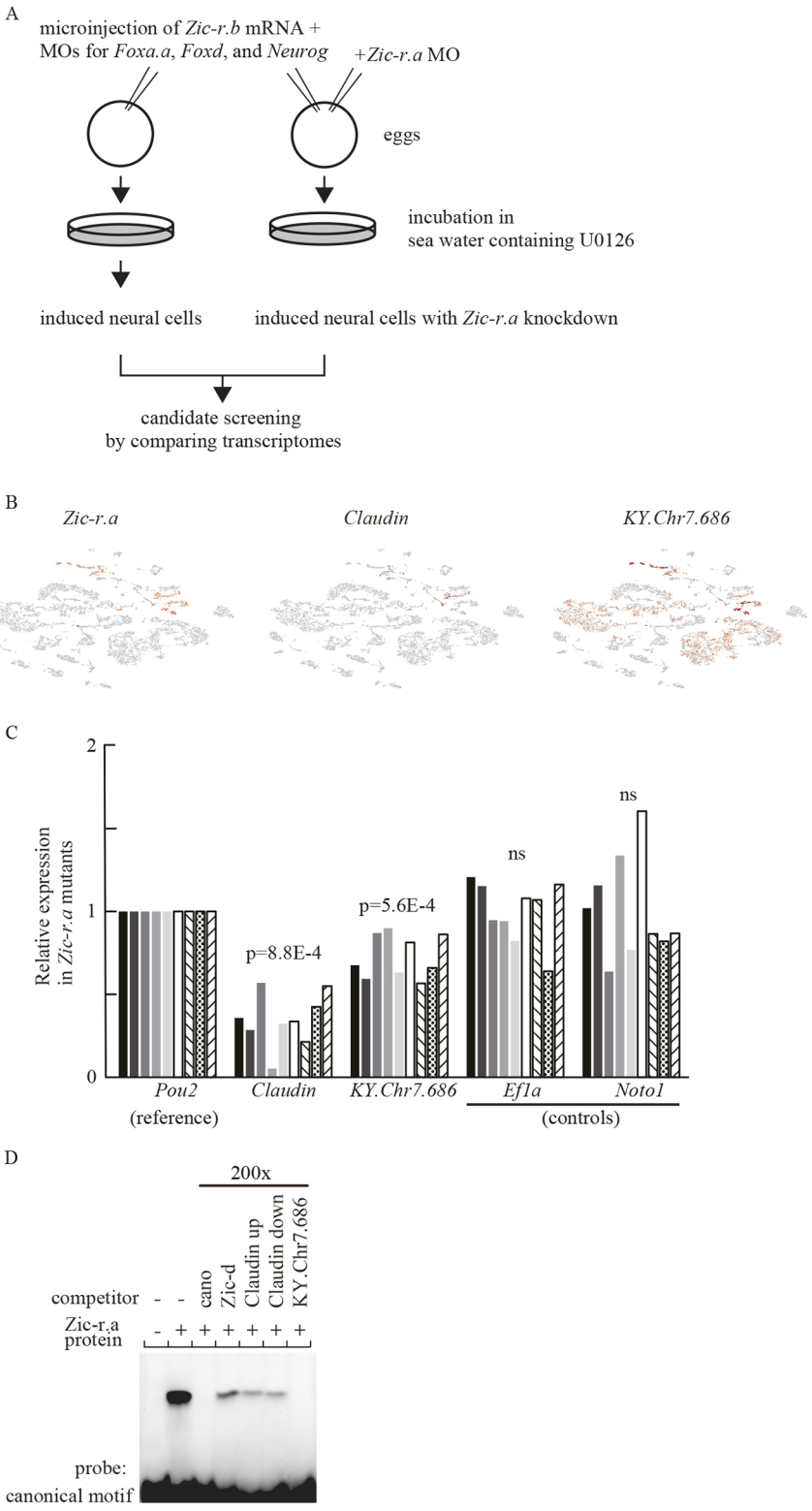

Fig. 5. Genes expressed under control of *Zic-r.a* at the tailbud stage have canonical-motif sites. (A) An experiment designed to screen candidates for *Zic-r.a* targets. (B) t-SNE plots of single-cell transcriptome data of middle tailbud embryos. Cells that express (left) *Zic-r.a*, (middle) *Claudin* and (right) *KY.Chr7.686* are colored. Color intensity indicates expression levels. Sequence data derived from a previous study (Cao et al., 2019) (SRA accession numbers: SRR9050997, SRR9050998 and SRR9050999-SRR9051004) are reanalyzed for the latest version of the gene model set (Satou et al., 2019). (C) mRNA levels that were measured using delta-delta Ct methods of RT-qPCR with maternally expressed *Pou2* as the internal reference. *Ef1α* and *Noto1* are negative controls. *P*-values were calculated with paired *t*-tests for delta Ct values; ns, no significant difference ($P>0.05$). (D) Gel-shift analysis showing that three sites found as canonical-motif sites in the 1 kb upstream regions of *Claudin* and *KY.Chr7.686* bind Zic-r.a with high affinity *in vitro*. The upstream region of *Claudin* contains two sites. Nucleotide sequences and positions of these canonical-motif sites are shown in Fig. S6.

2004a) (Fig. S6), although no non-canonical-motif sites were recognized with the position weight matrix for the mouse secondary motif. A gel-shift assay showed that these sites competed more strongly than non-canonical-motif sites, which indicates that these sites act as high-affinity sites for Zic-r.a (Fig. 5D). These data support the hypothesis that *Claudin* and *KY.Chr.7.686* are regulated through Zic-r.a binding to primary canonical-motif sites in their upstream regions.

Different zinc-finger domains are primarily responsible for recognizing canonical and non-canonical motifs

To understand how Zic-r.a recognized canonical and non-canonical motifs, we examined whether Zic-r.a proteins with mutations in one of its five zinc fingers could drive gene expression. We changed two residues that are crucial for DNA recognition (Wolfe et al., 2000) to glycine in each zinc-finger domain (Fig. 6A). Injection of mRNA encoding intact Zic-r.a protein induced ectopic expression of the

Development (2021) 148, dev199538. doi:10.1242/dev.199538

reporter genes, *wt>Gfp* and *2xcano>Gfp*, in a and A lineages of cells (Fig. 6B,C). The mutation introduced into the first zinc-finger domain (μZF1) reduced the proportion of embryos with ectopic expression of *2xcano>Gfp*, but not of *wt>Gfp*, while the mutation introduced into the third zinc-finger domain reduced the proportion of embryos with ectopic expression of *wt>Gfp*, but not of *2xcano>Gfp* (Fig. 6B,C). In addition, the mutation introduced into the fourth zinc-finger domain slightly reduced the proportion of embryos with ectopic expression of *2xcano>Gfp*. Thus, zinc-finger domains that are primarily responsible for binding to canonical- and non-canonical-motif sites appear to differ (Fig. 6D,E). This observation indicates that the non-canonical motif is not a divergent version of the canonical motif. In other words, these two motifs are recognized in different ways by Zic-r.a.

DISCUSSION

Zic-r.a acts in three modes

In the present study, we showed that *Tbx6-r.b* expression is regulated differently in B6.4 of 32-cell embryos from B5.1 of 16-cell embryos (Fig. 7). While maternal Zic-r.a activates *Tbx6-r.b* cooperatively with β-catenin and Tcf7 in the B5.1 lineage of 16-cell embryos (Oda-Ishii et al., 2016), Zic-r.a alone activates *Tbx6-r.b* in the B6.4 lineage of 32-cell embryos. In B6.4, the limited number of weak non-canonical-motif sites (Zic-d and Zic-p) activate *Tbx6-r.b* in response to a high Zic-r.a concentration. The B5.1 lineage contains less Zic-r.a than does the B6.4 lineage; therefore, it is likely that the Zic-d- and Zic-p-binding sites cannot respond sufficiently at the 16-cell stage. On the other hand, because the Zic-d site abuts the Tcf7-binding site (Fig. S1), and because Zic-r.a and Tcf7 can interact (Oda-Ishii et al., 2016), Zic-r.a may bind to these non-canonical sites with the help of Tcf7 in 16-cell embryos. Thus, the weak non-canonical motif is used by reducing and tuning the activity of the enhancer, and it induces *Tbx6-r.b* expression in the B6.4 lineage without the help of β-catenin/Tcf7.

In ascidian embryos, zygotic transcription begins between the 8-cell and 16-cell stages in most somatic cells (Oda-Ishii et al., 2018). However, the most posterior cells, which have the potential to give rise to germ cells, are transcriptionally silent (Kumano et al., 2011; Shirae-Kurabayashi et al., 2011), and B6.4 cells in the 32-cell embryo are somatic daughters of the most posterior cell pair of the 16-cell embryo. Therefore, transcription in the B6.4 lineage begins at the 32-cell stage. Because nuclear β-catenin is unavailable in this lineage (Hudson et al., 2013), the regulatory mechanism that activates *Tbx6-r.b* at the 16-cell stage does not act in the B6.4 lineage at the 32-cell stage. This is probably the reason that the ascidian embryo has evolved a distinct regulatory mechanism activating *Tbx6-r.b* in the B6.4 lineage.

Similarly, another Zic gene, *Zic-r.b*, and *Snail* are expressed in the B5.1 and B6.4 lineages at the 32-cell stage, and the regulatory mechanisms regulating these genes differ in these two lineages: *Zic-r.b* is activated in the daughter cells of B5.1 by maternal Gata.a, while the expression in B6.4 is not lost in Gata.a morphants (Imai et al., 2016); *Snail* is activated in the daughter cells of B5.1 by *Tbx6-r.b*, and in B6.4 by constitutively active form of Raf (Tokuoka et al., 2018). Intriguingly, the regulatory mechanisms activating these three genes in the B6.4 lineage also differ, suggesting that the ascidian embryo has independently evolved these mechanisms under a common selective pressure.

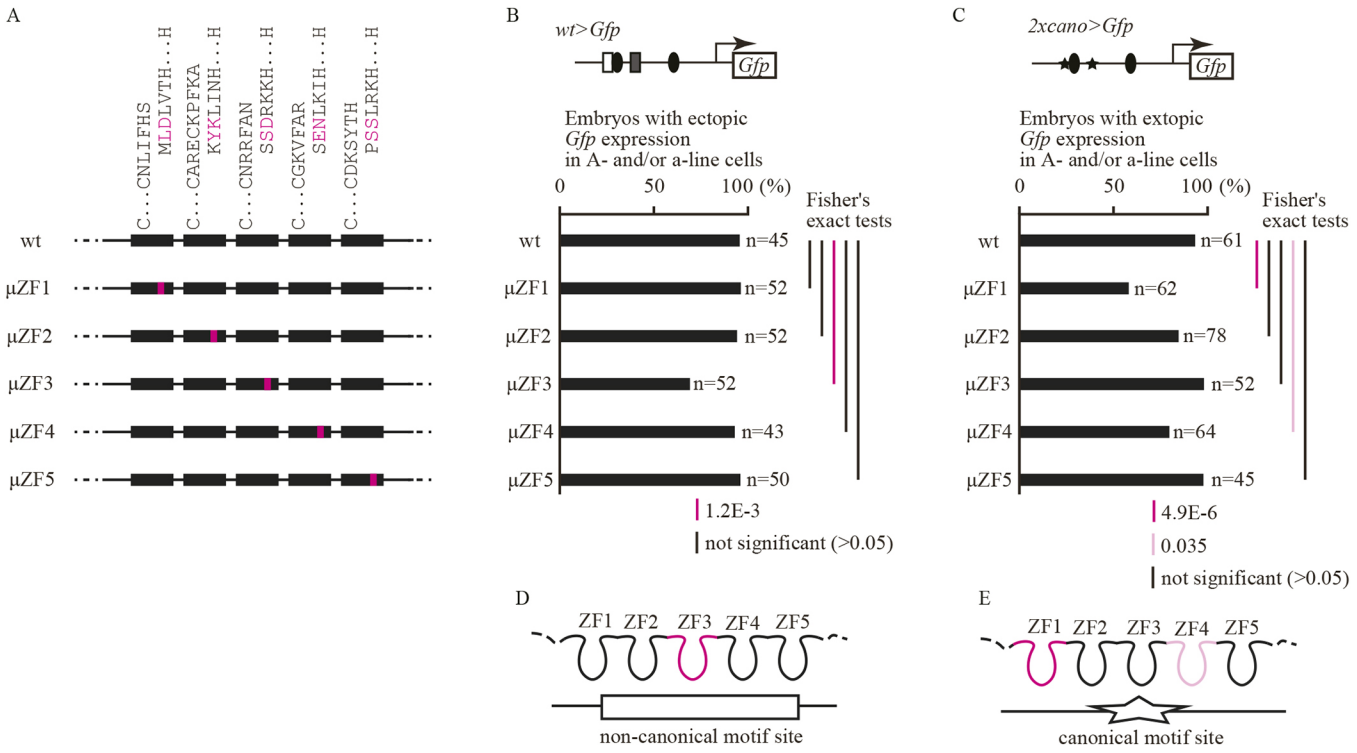

Fig. 6. Zinc-finger domains of Zic-r.a that are crucial for motif recognition differ between canonical- and noncanonical-motif sites. (A) Schematics for zinc-finger domains of wild-type and mutant Zic-r.a. μZF1 to μZF5 each contain two mutated amino acids (amino acids indicated in magenta were mutated to glycine) in one of the zinc-finger domains. (B,C) Percentages of embryos that expressed (B) *wt>Gfp* and (C) *2xcano>Gfp* ectopically in A- and/or a-line cells. Vertical lines on the right represent *P*-values of Fisher's exact tests. (D,E) A model to explain how Zic-r.a recognizes two different motif sites. Zinc-finger domains that are primarily responsible for binding to non-canonical-motif sites (D) and canonical-motif sites (E) are colored in dark and light magenta, respectively.

Development (2021) 148, dev199538. doi:10.1242/dev.199538

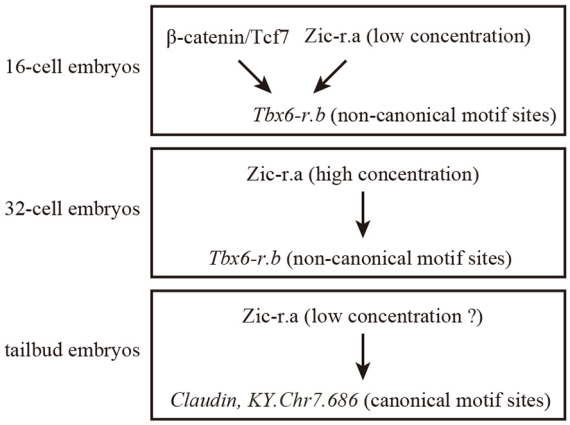

Fig. 7. Zic-r.a acts in three modes. Zic-r.a cannot activate *Tbx6-r.b*, which contains non-canonical-motif sites in its upstream regulatory region, in B5.1 at the 16-cell stage without the help of β-catenin and Tcf7, because B5.1 contains a relatively small amount of Zic-r.a protein. On the other hand, Zic-r.a activates *Tbx6-r.b* in B6.4 at the 32-cell stage without the help of β-catenin and Tcf7, because Zic-r.a is relatively abundant in B6.4. In neural cells of tailbud embryos, Zic-r.a activates *Claudin* and *KY.Chr7.686*, which contain canonical-motif sites in their upstream regulatory regions.

The present and previous studies have established that Zic-r.a acts in three modes (Fig. 7). First, at the 16-cell stage, it acts by interacting with a complex of Tcf7 and β-catenin to activate *Tbx6-r.b* (Oda-Ishii et al., 2016). Second, in B6.4 of the 32-cell embryo, a high level of Zic-r.a expression allows it to bind to non-canonical-motif sites in the upstream regulatory region of *Tbx6-r.b*. Third, in late embryos, Zic-r.a, which is zygotically expressed in the nervous system, activates neural genes through canonical-motif sites.

One unsolved problem is why neural genes that are controlled by Zic-r.a through canonical sites are not expressed in early embryos. It is possible that a higher order of transcriptional regulation, including epigenetic regulation, is responsible, although this remains to be tested. It is also possible that Zic-r.a activates neural genes combinatorially with additional factors specific to neural cells. On the other hand, our experiments indicate the possibility that *Tbx6-r.b* is not activated in neural cells of late embryos because the amount of Zic-r.a is insufficient. This possibility has not been tested either, because comparisons of Zic-r.a concentrations in early and late embryos are technically difficult.

Zic-r.a recognizes two distinct motifs

We showed that the first and third zinc-finger domains are important to drive gene expression through canonical and noncanonical binding sites. This observation indicates that these two sites (motifs) are recognized differently by Zic-r.a. That is, the non-canonical-binding motif is not a variant of the canonical motif. Therefore, Zic-r.a indeed uses two sites with different motifs to induce expression of specific genes in early and late embryos, respectively.

On the other hand, the non-canonical-motif sites are low-affinity sites. Low-affinity sites are often used to control gene expression temporally and spatially (Crocker et al., 2015; Farley et al., 2015; Fuqua et al., 2020; Gaudet and Mango, 2002). Although most low-affinity sites are thought to be variants of high-affinity sites, the non-canonical motif recognized by Zic-r.a is not. Several studies have suggested that some transcription factors can recognize two or more motifs (Badis et al., 2009; Franco-Zorrilla et al., 2014), although this remains controversial (Jolma et al., 2013; Morris et al., 2011; Zhao and Stormo, 2011). In ascidian embryos, Zic-r.a recognizes two motifs with different affinities, and activates its target genes in a specific cell pair through non-canonical-motif sites. Zic-r.a is a transcription factor with five zinc-finger domains. Our data indicate that this modular nature enables this transcription factor to recognize multiple motifs.

Zic-r.a is composed of three exons and the five zinc-finger domains are encoded by these three exons. While expressed sequence tags have been obtained from 34 cDNAs of seven cDNA libraries (Satou et al., 2005; Tassy et al., 2010), none of them indicates alternative splicing. Therefore, it is not likely that different isoforms recognize different motifs, but it is more likely that a single isoform recognizes the canonical and non-canonical motif sites.

Gene regulatory networks for embryonic fate specification have been studied extensively. Knockout, knockdown and overexpression assays to identify network connections have been complemented by reporter assays and chromatin immunoprecipitation assays to identify direct interactions. However, it is common that not all peaks identified by chromatin immunoprecipitation assays contain recognizable binding sites. Non-canonical sites may be present within such peak regions.

In summary, the non-canonical motif of Zic-r.a is functional and important for ascidian embryogenesis. Multiple motifs that some transcription factors recognize may be used for tuning enhancer activity to evoke expression of specific targets.

MATERIALS AND METHODS

Animals, whole-mount *in situ* hybridization and gene identifiers

Adult specimens of *Ciona intestinalis* (type A; also called *Ciona robusta*) were obtained from the National BioResource Project for *Ciona*. cDNA clones were obtained from our EST clone collection (Satou et al., 2005). Whole-mount *in situ* hybridization was performed as described previously (Satou et al., 1995). Identifiers for genes examined in this study are: KY.Chr11.468-470 for *Tbx6-r.b*, KY.Chr1.1698 for *Zic-r.a*, KY.C9.48 for *β-catenin*, KY.Chr11.1167 for *Foxa.a*, KY.Chr8.660/661 for *Foxd*, KY.Chr6.427 for *Neurog*, KY.Chr10.963 for *Claudin*, KY.Chr4.359 for *Pou2*, KY.Chr6.606 for *Noto1* and KY.Chr14.194 for *Ef1a*.

Gene knockdown, knockout, overexpression and reporter assays

MOs for *Zic-r.a*, *β-catenin*, *Foxa.a*, *Foxd* and *Neurog*, which block translation, have been used previously and their specificity has been evaluated (Imai et al., 2004; Kobayashi et al., 2018; Yagi et al., 2004a). These MOs were microinjected under a microscope. For overexpression, the coding sequence of *Zic-r.a* was cloned into pBluscript RN3 (Lemaire et al., 1995). Injected RNAs were transcribed using a mMESSAGE mMACHINE T3 Transcription Kit (ThermoFisher Scientific).

To mutate *Zic-r.a*, we used TALEN (transcription activator-like effector nuclease) technology. N- and C-terminal domains of TALE and FokI nuclease domain were taken from the Platinum Gate TALEN kit (Sakuma et al., 2013). Protein products derived from a pair of synthetic mRNAs were designed to cleave the first exon of *Zic-r.a* (recognition sequences 5′-CTTTGCTCGAAGCGAAAACC-3′ and 5′-GTCTCTTACCAGTGT GAGTT-3′). These mRNAs were introduced by microinjection into fertilized eggs. To confirm successful cleavage, we performed PCR using the following primers: 5′-GATCTATTTCTGGTTTTACCTTG-3′ and 5′-CTGGTACGATGGGATATGAATC-3′ .

Reporter constructs were introduced into fertilized eggs by electroporation or microinjection. Chromosomal positions of upstream sequences for reporter constructs and mutated sequences are indicated in Fig. S1. We randomly chose embryos with introduced reporter constructs to examine reporter construct expression using *in situ* hybridization. These experiments were performed at least twice with different batches of embryos.

The patser program was run with the following parameters: '-c -A a:t 65 c: g 35 -li'.

Development (2021) 148, dev199538. doi:10.1242/dev.199538

Gel-shift assay

Recombinant Zic-r.a protein was produced as a His-tag-fusion protein in *E. coli* and purified with Ni-NTA agarose (QIAGEN). After annealing two complementary oligonucleotides (Zic-p, 5′-aaaACACGAATCAGCAGGAGAGTTC-3′ and 5′-aaaGAACTCTCCTGCTGATTCGTGT-3′; Zic-d, 5′-aaaATGCGTCACACTGAGTTTTGGA-3′ and 5′-aaaTCCAAAACTCAGTGTGACGCAT-3′; cano, 5′-aaaACTAGTGCCCCCCGCTGCGCGG-3′ and 5′-aaaCCGCGCAGCGGGGGGCACTAGT-3′), the protruding ends of the double-stranded oligonucleotides were filled with biotin-11-dUTP. The resultant biotin-labeled oligonucleotides were used as probes. Unlabeled double-stranded oligonucleotides with the same sequences were used as specific competitors. Other competitors were similarly made from two oligonucleotides (Claudin up, 5′-aaaTGTCTTCCCCCACCCTACTATACAT-3′ and 5′-aaaATGTATAGTAGGGTGGGGGAAGACA-3′; Claudin down, 5′-aaaATAGTAGGGTGGGGGGAGATGGAAC-3′ and 5′-aaaGTTCCATCTCCCCCCACCCTACTAT-3′; KY.Chr7.686, 5′-aaaTTGCGTCGCAGGGGGCTGACACGGC-3′ and 5′-aaaGCCGTGTCAGCCCCCTGCGACGCAA-3′). Mutant competitors were produced from two complementary oligonucleotides (μZic-p, 5′-aaaACACGAATTCATTGGAGAGTTC-3′ and 5′-aaaGAACTCTCCAATGAATTCGTGT-3′; μZic-d, 5′-aaaATGCGTCATGTCAAGTTTTGGA-3′ and 5′-aaaTCCAAAACTTGACATGACGCAT-3′). Proteins and biotin-labeled probes were mixed in 25 mM Hepes (pH 7.5), 100 mM KCl, 1 mM DTT, 50 ng/μl poly(dIdC), 2.5% glycerol, 0.05% NP40 and 50 μM $ZnSO_4$ with or without competitor double-stranded DNAs. Protein concentrations were empirically determined. Protein-DNA complexes were detected using a Chemiluminescent Nucleic Acid Detection Module Kit (ThermoFisher Scientific). Bands were quantified as arbitrary units using an imager (ChemiDoc XRS, BioRad) and Quantity-One software (BioRad).

Immunostaining and quantification of fluorescence intensity

Immunostaining with anti-Zic-r.a antibody was performed as described previously (Oda-Ishii et al., 2016). ImageJ was used to quantify fluorescence intensity. All photographs were taken in the same conditions and DAPI signal intensity was used for reference.

Identification of Zic-r.a targets in tailbud embryos

To induce neural fate in single-cell syncytium embryos, we injected *Zic-r.b* mRNA and MOs for *Foxa.a*, *Foxd* and *Neurog*, and incubated injected embryos in sea water containing the MEK inhibitor U0126 and cytochalasin B, as described previously (Kobayashi et al., 2018). We also prepared embryos injected with the *Zic-r.a* MO. We next performed RNA-sequencing for these two specimens using Ion Total RNA-Seq kit ver 2 (ThermoFisher Scientific) and an Ion PGM instrument (ThermoFisher Scientific), as described previously (Tokuhiro et al., 2017) (DRA accession number DRA011314). We did not take duplicates, because we used this experiment for screening; the obtained result was confirmed with other methods. NOISeq software (Tarazona et al., 2011) was used to identify differentially expressed genes. Single-cell transcriptome data (Cao et al., 2019) were mapped to a gene model set recently developed (Satou et al., 2019) and analyzed using Cell Ranger (10×Genomics).

For RT-qPCR, we extracted RNA and converted RNA to cDNA using a Cells-to-Ct kit (ThermoFisher Scientific). cDNA samples were analyzed by quantitative PCR with the SYBRG method. Primers used were: *Claudin*, 5′-CGATGTCTTACAGGGTGGTTCT-3′ and 5′-CCGATGGTATAACCAAGTCCAG-3′; *KY.Chr7.686*, 5′-TGTACCAGACATGAGAGCGAAA-3′ and 5′-CGTAAGCCGCCTTCAGTTC-3′; *Pou2*, 5’-AAGATGGTTGCTGGATGCTAATAAT-3′ and 5’-TTGGATTGGAGTGGGAATAACAA-3′; *Ef1α*, 5’-CTCCCGGTCACAGAGATTTCA-3′ and 5′-CAATAAGCACGGCACAATCG-3′; and *Noto1*, 5′-GGCTTGCCTGCGAATGG-3′ and 5′-GAGCACACGACTGCATCGTAA-3′.

Acknowledgements

We thank Reiko Yoshida (Kyoto University), Manabu Yoshida (University of Tokyo) and other members working under the National BioResource Project for *Ciona* (MEXT, Japan) at Kyoto University and the University of Tokyo for providing experimental animals. We thank Steven D. Aird for editing a draft of the manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: Y.S.; Investigation: I.O.-I., D.Y.; Writing - original draft: Y.S.; Writing - review & editing: I.O.-I., D.Y.; Visualization: I.O.-I., Y.S.; Funding acquisition: I.O.-I., Y.S.

Funding

This research was supported by grants from the Japan Society for the Promotion of Science (17KT0020 and 21H02486 to Y.S., and 19J40136 to I.O.-I.).

Data availability

RNA-seq data are available in the DRA/ERA/SRA database under the accession number DRA011314.

Peer review history

The peer review history is available online at https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.199538

References

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E. T., Metzler, G., Vedenko, A., Chen, X. et al. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720-1723. doi:10.1126/science.1162327

Cao, C., Lemaire, L. A., Wang, W., Yoon, P. H., Choi, Y. A., Parsons, L. R., Matese, J. C., Wang, W., Levine, M. and Chen, K. (2019). Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349-354. doi:10.1038/s41586-019-1385-y

Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F. et al. (2015). Low affinity binding site clusters confer Hox specificity and regulatory robustness. *Cell* **160**, 191-203. doi:10.1016/j.cell.2014.11.041

Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S. and Levine, M. S. (2015). Suboptimization of developmental enhancers. *Science* **350**, 325-328. doi:10.1126/science.aac6948

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P. and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci. USA* **111**, 2367-2372. doi:10.1073/pnas.1316278111

Fuqua, T., Jordan, J., van Breugel, M. E., Halavatyi, A., Tischer, C., Polidoro, P., Abe, N., Tsai, A., Mann, R. S., Stern, D. L. et al. (2020). Dense and pleiotropic regulatory information in a developmental enhancer. *Nature* **587**, 235-239. doi:10.1038/s41586-020-2816-5

Gaudet, J. and Mango, S. E. (2002). Regulation of organogenesis by the *Caenorhabditis elegans*, FoxA protein PHA-41. *Science* **295**, 821-825. doi:10.1126/science.1065175

Hertz, G. Z. and Stormo, G. D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15**, 563-577. doi:10.1093/bioinformatics/15.7.563

Hudson, C., Kawai, N., Negishi, T. and Yasuo, H. (2013). β-catenin-driven binary fate specification segregates germ layers in ascidian embryos. *Curr. Biol.* **23**, 491-495. doi:10.1016/j.cub.2013.02.005

Hume, M. A., Barrera, L. A., Gisselbrecht, S. S. and Bulyk, M. L. (2015). UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* **43**, D117-D122. doi:10.1093/nar/gku1045

Imai, K. S., Satou, Y. and Satoh, N. (2002). Multiple functions of a Zic-like gene in the differentiation of notochord, central nervous system and muscle in Ciona savignyi embryos. *Development* **129**, 2723-2732. doi:10.1242/dev.129.11.2723

Imai, K. S., Hino, K., Yagi, K., Satoh, N. and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development* **131**, 4047-4058. doi:10.1242/dev.01270

Imai, K. S., Hudson, C., Oda-Ishii, I., Yasuo, H. and Satou, Y. (2016). Antagonism between β-catenin and Gata.a sequentially segregates the germ layers of ascidian embryos. *Development* **143**, 4167-4172. doi:10.1242/dev.141481

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G. et al. (2013). DNA-binding specificities of human transcription factors. *Cell* **152**, 327-339. doi:10.1016/j.cell.2012.12.009

Kobayashi, K., Maeda, K., Tokuoka, M., Mochizuki, A. and Satou, Y. (2018). Controlling cell fate specification system by key genes determined from network structure. *iScience* **4**, 281-293. doi:10.1016/j.isci.2018.05.004

Kugler, J. E., Gazdoiu, S., Oda-Ishii, I., Passamaneck, Y. J., Erives, A. J. and Di Gregorio, A. (2010). Temporal regulation of the muscle gene cascade by Macho1 and Tbx6 transcription factors in Ciona intestinalis. *J. Cell Sci.* **123**, 2453-2463. doi:10.1242/jcs.066910

Kumano, G., Takatori, N., Negishi, T., Takada, T. and Nishida, H. (2011). A maternal factor unique to ascidians silences the germline via binding to P-TEFb and RNAP II regulation. *Curr. Biol.* **21**, 1308-1313. doi:10.1016/j.cub.2011.06.050

Lemaire, P., Garrett, N. and Gurdon, J. B. (1995). Expression cloning of *Siamois*, a Xenopus homeobox gene expressed in dorsal-vegetal cells of blastulae and able to induce a complete secondary axis. *Cell* **81**, 85-94. doi:10.1016/0092-8674(95)90373-9

Morris, Q., Bulyk, M. L. and Hughes, T. R. (2011). Jury remains out on simple models of transcription factor specificity. *Nat. Biotechnol.* **29**, 483-484. doi:10.1038/nbt.1892

Nishida, H. and Sawada, K. (2001). macho-1 encodes a localized mRNA in ascidian eggs that specifies muscle fate during embryogenesis. *Nature* **409**, 724-729. doi:10.1038/35055568

Oda-Ishii, I., Kubo, A., Kari, W., Suzuki, N., Rothbächer, U. and Satou, Y. (2016). A maternal system initiating the zygotic developmental program through combinatorial repression in the ascidian embryo. *PLoS Genet.* **12**, e1006045. doi:10.1371/journal.pgen.1006045

Oda-Ishii, I., Abe, T. and Satou, Y. (2018). Dynamics of two key maternal factors that initiate zygotic regulatory programs in ascidian embryos. *Dev. Biol.* **437**, 50-59. doi:10.1016/j.ydbio.2018.03.009

Sakuma, T., Ochiai, H., Kaneko, T., Mashimo, T., Tokumasu, D., Sakane, Y., Suzuki, K.-i., Miyamoto, T., Sakamoto, N., Matsuura, S. et al. (2013). Repeating pattern of non-RVD variations in DNA-binding modules enhances TALEN activity. *Sci. Rep.* **3**, 3379. doi:10.1038/srep03379

Satou, Y., Kusakabe, T., Araki, L. and Satoh, N. (1995). Timing of initiation of muscle-specific gene-expression in the ascidian embryo precedes that of developmental fate restriction in lineage cells. *Dev. Growth Differ.* **37**, 319-327. doi:10.1046/j.1440-169X.1995.t01-2-00010.x

Satou, Y., Yagi, K., Imai, K. S., Yamada, L., Nishida, H. and Satoh, N. (2002). macho-1-related genes in Ciona embryos. *Dev. Genes Evol.* **212**, 87-92. doi:10.1007/s00427-002-0218-3

Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A. and Satoh, N. (2005). An integrated database of the ascidian, *Ciona intestinalis*: towards functional genomics. *Zool. Sci.* **22**, 837-843. doi:10.2108/zsj.22.837

Satou, Y., Nakamura, R., Yu, D., Yoshida, R., Hamada, M., Fujie, M., Hisata, K., Takeda, H. and Satoh, N. (2019). A nearly complete genome of *Ciona intestinalis* type A (*C. robusta*) reveals the contribution of inversion to chromosomal evolution in the genus *Ciona*. *Genome Biol. Evol.* **11**, 3144-3157. doi:10.1093/gbe/evz228

Schneider, T. D. and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097-6100. doi:10.1093/nar/18.20.6097

Shirae-Kurabayashi, M., Matsuda, K. and Nakamura, A. (2011). Ci-Pem-1 localizes to the nucleus and represses somatic gene transcription in the germline of *Ciona intestinalis* embryos. *Development* **138**, 2871-2881. doi:10.1242/dev.058131

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. and Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. *Genome Res.* **21**, 2213-2223. doi:10.1101/gr.124321.111

Tassy, O., Dauga, D., Daian, F., Sobral, D., Robin, F., Khoueiry, P., Salgado, D., Fox, V., Caillol, D., Schiappa, R. et al. (2010). The ANISEED database: digital representation, formalization, and elucidation of a chordate developmental program. *Genome Res.* **20**, 1459-1468. doi:10.1101/gr.108175.110

Tokuhiro, S.-i., Tokuoka, M., Kobayashi, K., Kubo, A., Oda-Ishii, I. and Satou, Y. (2017). Differential gene expression along the animal-vegetal axis in the ascidian embryo is maintained by a dual functional protein Foxd. *PLoS Genet.* **13**, e1006741. doi:10.1371/journal.pgen.1006741

Tokuoka, M., Kobayashi, K. and Satou, Y. (2018). Distinct regulation of *Snail* in two muscle lineages of the ascidian embryo achieves temporal coordination of muscle development. *Development* **145**, dev163915. doi:10.1242/dev.163915

Wolfe, S. A., Ramm, E. I. and Pabo, C. O. (2000). Combining structure-based design with phage display to create new Cys_2His_2 zinc finger dimers. *Structure* **8**, 739-750. doi:10.1016/S0969-2126(00)00161-1

Yagi, K., Satoh, N. and Satou, Y. (2004a). Identification of downstream genes of the ascidian muscle determinant gene *Ci-macho1*. *Dev. Biol.* **274**, 478-489. doi:10.1016/j.ydbio.2004.07.013

Yagi, K., Satou, Y. and Satoh, N. (2004b). A zinc finger transcription factor, ZicL, is a direct activator of Brachyury in the notochord specification of Ciona intestinalis. *Development* **131**, 1279-1288. doi:10.1242/dev.01011

Yagi, K., Takatori, N., Satou, Y. and Satoh, N. (2005). Ci-Tbx6b and Ci-Tbx6c are key mediators of the maternal effect gene Ci-macho1 in muscle cell differentiation in Ciona intestinalis embryos. *Dev. Biol.* **282**, 535-549. doi:10.1016/j.ydbio.2005.03.029

Yu, D., Oda-Ishii, I., Kubo, A. and Satou, Y. (2019). The regulatory pathway from genes directly activated by maternal factors to muscle structural genes in ascidian embryos. *Development* **146**, dev173104. doi:10.1242/dev.173104

Zhao, Y. and Stormo, G. D. (2011). Quantitative analysis demonstrates most transcription factors require only simple models of specificity. *Nat. Biotechnol.* **29**, 480-483. doi:10.1038/nbt.1893