Gata is ubiquitously required for the earliest zygotic gene transcription in the ascidian embryo

Kaoru S. Imai[1,*], Kenji Kobayashi[2], Willi Kari[3], Ute Rothbächer[3], Naoki Ookubo[1], Izumi Oda-Ishii[2] and Yutaka Satou[2]

[1] Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka 560-0043, Japan.

[2] Department of Zoology, Graduate School of Science, Kyoto University, Kyoto, 606-8502, Japan

[3] Department of Evolution and Developmental Biology, Zoological Institute, University Innsbruck, Innsbruck, Austria

*Author for correspondence:

Kaoru S. Imai, imai@bio.sci.osaka-u.ac.jp; +81-6-6850-6760

Keywords: ascidian, Tcf7, Gata, β-catenin, zygotic gene expression

Abstract

In ascidian embryos, the earliest transcription from the zygotic genome begins between the 8-cell and 16-cell stages. Gata.a, a maternally expressed Gata transcription factor, activates target genes specifically in the animal hemisphere, whereas the complex of β-catenin and Tcf7 antagonizes the activity of Gata.a and activates target genes specifically in the vegetal hemisphere. Here, we show that genes zygotically expressed at the 16-cell stage have significantly more Gata motifs in their upstream regions. These genes included not only genes with animal hemisphere-specific expression but also genes with vegetal hemisphere-specific expression. On the basis of this finding, we performed knockdown experiments for *Gata.a* and reporter assays, and found that Gata.a is required for the expression of not only genes with animal hemisphere-specific expression, but also genes with vegetal hemisphere-specific expression. Our data indicated that weak Gata.a activity that cannot induce animal hemisphere-specific expression can allow β-catenin/Tcf7 targets to be expressed in the vegetal cells. Because genes zygotically expressed at the 32-cell stage also had significantly more Gata motifs in their upstream regions, Gata.a function may not be limited to the genes expressed specifically in the animal or vegetal hemispheres at the 16-cell stage, and Gata.a may play an important role in the earliest transcription of the zygotic genome.

Introduction

Maternal products initiate the zygotic developmental program of animal embryos. Such maternal factors are often specifically localized to set up pre-patterns. In an invertebrate chordate, *Ciona intestinalis* (Type A; or *Ciona robusta*), the earliest zygotic gene expression begins between the 8- and 16-cell stages. β-catenin is translocated into the nuclei of the vegetal hemisphere cells, weakly at the 8-cell stage and more strongly at the 16-cell stage, and functions together with the transcription factor Tcf7 to activate expression of target genes in the vegetal hemisphere (Hudson et al., 2013; Imai et al., 2000; Oda-Ishii et al., 2018; Oda-Ishii et al., 2016). The transcription factor Gata.a is provided maternally, and its level rapidly increases until the 4-cell stage, probably through translation of maternal mRNA (Oda-Ishii et al., 2018). Although this protein is present in all cells of the 16-cell embryo, its DNA binding activity is weakened by nuclear β-catenin in the vegetal hemisphere, and Gata.a activates its targets only in the animal hemisphere at this stage (Bertrand et al., 2003; Oda-Ishii et al., 2016; Rothbächer et al., 2007). At the 32-cell stage, β-catenin is observed in the nuclei of three pairs of vegetal cells around the vegetal pole, but excluded from the nuclei of marginal vegetal cells (Hudson et al., 2013). Due to this exclusion, Gata.a activity is not weakened in the marginal vegetal cells of 32-cell embryos and Gata.a targets are activated (Hudson et al., 2016; Imai et al., 2016). Thus, Gata.a activates target genes in cells without nuclear β-catenin at the 16- and 32-cell stages. Interestingly, in the present study, we found that Gata.a also bound to the upstream region of β-catenin/Tcf7 targets, and contributed to the activation of these target genes in vegetal cells, in which β-catenin is present in their nuclei.

Results

Genes activated in early embryos have more Gata.a binding sites

For the purpose to identify maternal transcription factors essential for the earliest zygotic transcription, we first performed total RNA sequencing (RNA-seq) of 16-cell and 32-cell embryos to identify genes zygotically expressed at the 16- and 32-cell stages. As in previous studies (Ameur et al., 2011; Madsen et al., 2015), we counted reads mapped onto introns in order to distinguish zygotic nascent transcripts from maternal transcripts, as zygotic transcription of genes expressed maternally cannot be easily identified by counting reads mapped onto exons. We categorized genes with introns of a total length ≥ 500 bases into three classes using the RNA-seq data: genes that had introns with abundant mapped reads were regarded as “abundantly transcribed”; those that did not have introns with mapped reads were regarded as “non-transcribed”; and genes with properties between these two categories were regarded as “rarely transcribed” (see the Materials and Methods section for details). We identified 95 and 165 abundantly transcribed genes at the 16- and 32-cell stages, respectively. The former was a subset of the latter, and hence, 70 (=165-95) genes began to be abundantly transcribed at the 32-cell stage. We also identified 2,498 genes that were not transcribed at the 16- or 32-cell stages, as well as 7,565 rarely transcribed genes. The genes identified as abundantly transcribed at the 16-cell stage included *Tfap2-r.b*, *Efna.a*, *Efna.d, Fgf9/16/20*, *Foxtun1*, *Sox1/2/3*, *Fog*, *Dusp6/9*, and *Fzd4*, which are known to be transcribed at the 16-cell stage (Hamaguchi et al., 2007; Imai et al., 2004; Lamy et al., 2006; Rothbächer et al., 2007). A small number of genes (including *Hes.a*, *Lefty* and *Admp)* that are known to be expressed at the 16-cell and 32-cell stages were identified as rarely transcribed genes. The introns of *Foxd* and *Tbx6-r.b*, which are expressed at the 16-cell stage (Imai et al., 2002; Takatori et al., 2004), were not long

enough; consequently, these genes were not included in this analysis. For conservative comparisons, we used only the abundantly transcribed and non-transcribed classes of genes in the following analyses, because the intermediate class may contain noisy data.

This method allowed us to identify zygotic transcription of genes maternally expressed. Indeed, we found that *Tbx21* (also called *maternal T*) was included among the abundantly transcribed genes (Fig. 1A). A previous study reported that this gene is abundantly expressed maternally, but it did not show zygotic expression in early embryos (Takatori et al., 2004). The data in the present study indicated that its abundant maternal transcript masked zygotic signals in this previous study. To further confirm that our method successfully identified zygotic transcription in 16-cell embryos, we examined whether *Tbx21* is indeed expressed zygotically at the 16-cell stage. Because zygotic transcripts, which appear in nuclei, can be discriminated from maternal transcripts in cytoplasm more clearly by fluorescence *in situ* hybridization, we examined *Tbx21* expression using this method, and found that its transcript was detected in the nuclei of all animal and vegetal cells, except for transcriptionally quiescent cells with a germ line fate (Fig. 1B). This indicated that *Tbx21* was indeed transcribed from the zygotic genome at the 16-cell stage. Thus, our method successfully identified zygotically transcribed genes.

Next, we compared the 1-kb upstream regions between the abundantly transcribed and non-transcribed classes of genes with the DREME software (Bailey, 2011). The following five motifs were predicted to be enriched in the abundantly transcribed class (E-value <0.05): CGCGAAAR, CGATAAGY, RGCGCS, AAATDGCG, and GCCAATYA (R=A/G, Y=A/T, S=G/C, D=AGT). As an independent test, we comprehensively counted all possible hexamers in the 1-kb upstream regions of the genes identified above. Eleven hexamers were overrepresented in the upstream regions of the genes

94 transcribed at the 16- and 32-cell stages [p<2.4E-5=0.05/2080; Chi-square tests with the Bonferroni
95 adjustment; the number of random hexamers is 2080 (=4^6/2+64/2; note that there are 64 palindromic
96 hexamers)] (Fig. 1C). The motif search program, Tomtom (Gupta et al., 2007), showed that the top two
97 hexamers (AGATAA and GATAAC) were similar to the GATA binding motif (Fig. 1D).

98 Because 'GATAA' was also included in the motifs identified with the DREME software, and
99 because Gata.a is involved in the earliest zygotic gene expression of ascidian embryos (Oda-Ishii et al.,
100 2018; Rothbächer et al., 2007), we further analyzed Gata.a binding sites in the upstream regions.
101 Specifically, we searched for Gata.a-binding sites using the Patser program (Hertz and Stormo, 1999)
102 with the position weight matrix for *Ciona* Gata.a protein in these upstream regions (Fig. 1E; see the
103 Materials and Methods section for details). On average, we found 2.9 and 2.7 Gata.a sites in the
104 upstream regions of genes that were transcribed in 16- and 32-cell embryos, respectively (Fig. 1F). On
105 the other hand, only 2.2 Gata.a sites were present in the upstream regions of non-transcribed genes.
106 Wilcoxon rank-sum tests indicated that these differences were significant. In other words, Gata.a sites
107 were more abundant in the upstream regions of genes expressed in early embryos, and therefore this
108 difference may be biologically important.

109 Using data obtained in a previously published chromatin-immunoprecipitation (ChIP) assay
110 for Gata.a (Oda-Ishii et al., 2016), we identified the highest Gata.a ChIP-seq signal within each of the 1-
111 kb upstream regions described. On average, the highest peak intensities for the transcribed genes at the
112 16- and 32-cell stages were 4.2 and 3.9, respectively, which were significantly higher than those for non-
113 transcribed genes (Fig. 1G). Similarly, mean fold change values of the ChIP-seq signal over the 1 kb
114 upstream regions were also significantly higher than that for non-transcribed genes (Fig. 1H). Therefore,

genes expressed at the 16- and 32-cell stages tended to bind Gata.a with higher affinity.

Chromatin-immunoprecipitation peaks for Gata.a and Tcf7 overlap significantly

The above results implied that Gata.a plays a role in activating genes in early embryos. As we mentioned above, previous studies have shown that genes expressed in the animal hemisphere of 16-cell and 32-cell embryos and genes expressed in marginal vegetal cells of 32-cell embryos are indeed regulated directly by Gata.a (Bertrand et al., 2003; Imai et al., 2016; Oda-Ishii et al., 2016; Rothbächer et al., 2007). On the other hand, *Foxd*, *Fgf9/16/20*, *Tbx6-r.b*, *Lefty, Admp*, and *Wnttun5* are expressed in the vegetal hemisphere at the 16-cell stage under the control of β-catenin/Tcf7 (Imai, 2003; Imai et al., 2004; Oda-Ishii et al., 2016). Therefore, we confirmed that the upstream regions of these genes expressed in the vegetal hemisphere indeed bound Gata.a using the ChIP data, which were previously published (Oda-Ishii et al., 2016) (Fig. 2A). Because there are two copies of *Foxd* in the genome, and the critical upstream regions are highly conserved, we were unable to map ChIP data confidently to these regions. Meanwhile, the 219-bp upstream region of *Fgf9/16/20* contains Tcf7-binding sites critical for expression in vegetal cells at the 16-cell stage (Oda-Ishii et al., 2016). A peak region in the Gata.a ChIP assay overlapped with this critical region (Fig. 2A). ChIP peaks for Gata.a were also observed in the upstream regions of *Lefty*, *Admp*, and *Wnttun5* (Fig. 2A). In the upstream region of *Tbx6-r.b*, a weak peak indicating Gata.a binding was visible, although the software failed to identify it as a significant peak (Fig. 2A). These peak regions included or overlapped the regulatory regions which had previously been identified as being necessary for expression in 16-cell embryos and indeed bound Tcf7 (Oda-Ishii et al., 2016). These observations indicated that Gata.a is involved in regulating not only genes expressed in the animal hemisphere but also genes expressed in the vegetal hemisphere under the control of β-

136 catenin/Tcf7 at the 16-cell stage.

137 We further re-examined ChIP data for Gata.a and Tcf7, which were previously published (Oda-
138 Ishii et al., 2016). Among peak regions identified by ChIP for Tcf7, 38% overlapped with peak regions
139 identified by ChIP for Gata.a (Fig. 2B). As a comparison, we randomly selected the same number of
140 non-overlapping genomic fragments with the same size distribution as the Tcf7 peak regions. We
141 repeated this random sampling 100 times and found only 8 % (mean=674 peaks, s.d.=22) overlapped
142 with the peak regions identified by ChIP for Gata.a. Consistently, the Gata.a ChIP signal was indeed
143 strong near each of those Tcf7 ChIP peaks (Fig. 2C). This observation supported the hypothesis that
144 Gata.a is involved in regulating the target genes of Tcf7.

145 **Gata.a binding sites are required for gene expression in the vegetal hemisphere**

146 Because whole embryos were used for the ChIP assay, it was not clear whether Gata.a bound
147 to these upstream regions only in the animal hemisphere in early embryos or in both of the animal and
148 vegetal hemispheres. Therefore, we next examined whether and how Gata.a is required for specific
149 expression of genes in the vegetal hemisphere. Among the aforementioned genes expressed in the
150 vegetal hemisphere at the 16-cell stage, the upstream regulatory regions for *Fgf9/16/20*, *Foxd*, and *Tbx6-*
151 *r.b* have been analyzed extensively (Oda-Ishii et al., 2016). Therefore, we examined whether binding of
152 Gata.a to the upstream regions of these genes with vegetal hemisphere-specific expression (Fig. 3) is
153 necessary for their expression. A reporter construct containing the 219-bp upstream region of *Fgf9/16/20*
154 was expressed specifically in the vegetal hemisphere at the 16-cell stage (Fig. 4A) (Oda-Ishii et al.,
155 2016). This region contained three putative Gata.a-binding sites, and two of them were highly conserved
156 in the upstream region of *Fgf9/16/20* of a closely related *Ciona savignyi* (Fig. 3A). Indeed, mutations

157 introduced into these three sites abolished reporter expression (Fig. 4B). Specifically, while the reporter
158 with intact Gata.a-binding sites was expressed in one or more cells of 63% of embryos, the reporter with
159 mutated Gata.a was rarely expressed. Similarly, a reporter construct that contained the 1255-bp upstream
160 region of *Foxd* was expressed specifically in the vegetal hemisphere at the 16-cell stage (Fig. 4C). A
161 previous study showed that the region between nucleotides -1012 and -1255 is essential for the
162 expression at this stage (Oda-Ishii et al., 2016). This region contained three putative Gata.a-binding sites,
163 and two of them were highly conserved in the upstream region of *C. savignyi Foxd* (Fig. 3B). We
164 introduced mutations into these three sites. These mutations abolished reporter expression (Fig. 4D).
165 Finally, we examined the upstream region of *Tbx6-r.b*. A reporter construct containing the 263-bp
166 upstream sequence reproduced endogenous expression, as reported previously (Oda-Ishii et al., 2016),
167 whereas a construct with mutations in two putative Gata sites, one of which was highly conserved in the
168 *C. savignyi* genome (Fig. 3C), was not expressed (Fig. 4E,F). Thus, Gata.a-binding sites were indeed
169 required for expression of the reporters for genes expressed in the vegetal hemisphere.

170 In the upstream region of *Foxd*, there was an additional 'GATA', which was not conserved in
171 *C. savignyi Foxd* (underlined in Fig. 3B; the third site in Fig. 5), in addition to the three Gata.a-binding
172 sites examined in Figure 4D. Three of these four sites (the second, third, and fourth sites in Fig. 5) were
173 flanked by 'TTT' (underlined in Fig. 5). Because this triplet was highly conserved in the Tcf7-binding
174 motif, which is typically represented CTTTGAT, we examined whether these TTT sequences were
175 important for *Foxd* expression. Similarly to the above-mentioned reporter construct, mutations
176 introduced into the four Gata.a binding sites and the three flanking regions abolished reporter expression
177 (Fig. 5). On the other hand, mutations introduced into only the flanking regions did not affect reporter
178 expression. Therefore, it is likely that these flanking 'TTT' triplets do not mediate the Tcf7 activity for

179 *Foxd* expression in the vegetal hemisphere. This observation reinforces the importance of the Gata.a
180 binding sites.

181 **Gata.a is required not only for genes expressed in the animal hemisphere but also for genes**
182 **expressed in the vegetal hemisphere**

183 Because the above reporter analysis cannot completely rule out the possibility that the
184 mutations introduced eliminated the binding of unknown necessary transcription factors, we examined
185 whether Gata.a was required for expression of these genes in the vegetal hemisphere by knocking down
186 *Gata.a* with a specific morpholino oligonucleotide (MO). Using reverse-transcription and quantitative
187 PCR (RT-qPCR), we compared expression levels between *Gata.a* morphants and embryos injected with
188 a control MO. We used *Zic-r.a* (*Macho-1*) as an endogenous control for relative quantification, because it
189 is expressed exclusively maternally and therefore its expression level is not expected to be affected by
190 injection of the MO. We used another maternally expressed gene, *Pou2f*, as an additional control, and
191 confirmed that its expression was not greatly changed (Fig. 6A). On the other hand, the expression levels
192 of two known Gata.a targets, *Tfap2-r.b* and *Efna.d*, which are expressed in the animal hemisphere, were
193 greatly reduced (Fig. 6A), as previously reported (Oda-Ishii et al., 2016). Similarly, although the effects
194 were not as dramatic, the expression levels of *Fgf9/16/20*, *Foxd*, and *Tbx6-r.b* were significantly reduced
195 (Fig. 6A). Therefore, Gata.a is required for these Tcf7/β-catenin target genes to be expressed properly.

196 Although several research groups including ours have used this *Gata.a* MO (Bertrand et al.,
197 2003; Imai et al., 2016; Oda-Ishii et al., 2016; Rothbächer et al., 2007), we further confirm the
198 specificity and efficiency of *Gata.a* knockdown as follows. First, to confirm the effects of knockdown of
199 *Gata.a* on gene expression in early embryos, we injected a MO against the *Gata.a* of *C. savignyi* into

200 *C.savignyi* eggs, and confirmed that the expression levels of *Efna.d*, *Fgf9/16/20*, *Foxd*, and *Tbx6-r.b* was
201 similarly reduced (Fig. 6B; Note that no clear ortholog for *Tfap2-r.b* was found in the published version
202 of the *C. savignyi* genome sequence and therefore it was not examined). In other words, the same
203 phenotype was obtained in *Gata.a* morphants between these two closely related species. This
204 observation strongly indicated that the effects we observed were specific. Second, we confirmed that the
205 MO for *C. intestinalis Gata.a* effectively reduced the expression level of Gata.a protein using a specific
206 antibody we prepared in a previous study (Oda-Ishii et al., 2018). However, because the expression level
207 of maternal Gata.a protein is approximately 60% of that of Gata.a protein at the 16-cell stage (Oda-Ishii
208 et al., 2018), we estimated that the expression level of Gata.a protein in Gata.a morphants was reduced
209 up to 60% compared to that in control embryos at the 16-cell stage. Because this level of reduction was
210 difficult to confirm, we examined whether this MO could effectively knock down *Gata.a* at the tailbud
211 stage. While *Gata.a* mRNA was detectable in endodermal cells of normal and *Gata.a* morphant embryos
212 (Fig. 6C,D), Gata.a protein was never detected in *Gata.a* morphant embryos (Fig. 6E,F). In this
213 experiment, we injected the MO after fertilization to minimize effects on the maternal *Gata.a* mRNA.
214 This observation indicated that this MO can effectively reduce the level of Gata.a protein. The above two
215 experiments strongly indicated that knock-down of maternal *Gata.a* specifically and effectively
216 downregulated the expression of *Efna.d*, *Tfap2-r.b, Fgf9/16/20*, *Foxd*, and *Tbx6-r.b* in *C. intestinalis*
217 embryos.

218 **Specific expression patterns can be reproduced with specific combinations of Gata.a sites and Tcf7**
219 **sites**

220 The above analyses demonstrated that Gata.a was required for β-catenin/Tcf7 targets to be

221 expressed. On the other hand, previous studies have shown that Gata.a acts as an activator for genes
222 specifically expressed in the animal hemisphere (Oda-Ishii et al., 2016; Rothbächer et al., 2007).
223 Therefore, using reporters with various numbers of Gata.a and Tcf7-binding sites shown in Figure 7, we
224 examined whether Gata.a could contribute to activation of β-catenin/Tcf7 targets in the vegetal
225 hemisphere without activating them in the animal hemisphere. A previous work showed that an upstream
226 sequence containing twelve Gata-binding sites induced reporter expression in the animal hemisphere
227 (Rothbächer et al., 2007). A similar reporter construct containing twelve Gata.a-binding sites was
228 expressed specifically in the animal hemisphere (Fig. 8A, B). On the other hand, an upstream sequence
229 containing two Tcf7-binding sites did not induce reporter expression at the 16-cell stage (Fig. 8A, C).
230 However, a combination of two Tcf7-binding sites and twelve Gata.a-binding sites induced reporter
231 expression in the vegetal hemisphere and in the animal hemisphere (Fig. 8A, D). Similarly, while an
232 upstream sequence containing four Tcf7-binding sites did not induce reporter expression at the 16-cell
233 stage (Fig. 8A, E), a combination of four Tcf7-binding sites and twelve Gata.a-binding sites induced
234 reporter expression in the vegetal hemisphere more efficiently (Fig. 8A, F).

235 For further confirmation, we tested two constructs. The first one contained four mutated Tcf7-
236 binding sites and twelve intact Gata.a-binding sites. This construct was expressed only in the animal
237 hemisphere (Fig. 9A, B, D), as the construct containing only twelve Gata.a-binding sites (Fig. 8A, B).
238 The second construct contained four intact Tcf7-binding sites and twelve mutated Gata.a-binding sites.
239 This construct was not expressed (Fig. 9A, C, E), as the construct containing only four Tcf7-binding sites
240 (Fig. 8A, E). Thus, the Gata.a-binding sites act together with the Tcf7-binding sites for gene expression
241 in the vegetal hemisphere, where nuclear β-catenin is present.

Because the above constructs containing Gata.a-binding sites and Tcf7-binidng sites were expressed in both of the animal and vegetal hemispheres (Fig. 8A, D, F), we next tested whether vegetal hemisphere specific expression pattern can be reproduced with specific combinations of Gata.a sites and Tcf7 sites. As we mentioned above, four Tcf7-binding sites did not induce gene expression (Fig. 8A, E). Likewise, two Gata.a-binding sites alone did not induce gene expression (Fig. 8A, G). However, a combination of four Tcf7-binding sites and two Gata.a-binding sites induced reporter expression in the vegetal hemisphere in a small number of embryos (Fig. 8A, H). A construct with four Tcf7-binding sites and four Gata.a-binding sites induced reporter gene expression specifically in the vegetal hemisphere more efficiently (Fig. 8A, I). Thus, two or four Gata.a binding sites were sufficient for Tcf7-binding sites to function in the vegetal hemisphere. This observation showed that Gata.a can contribute to specific activation of β-catenin/Tcf7 targets in the vegetal hemisphere. In this way, vegetal hemisphere specific expression like *Fgf9/16/20* expression was reproduced with a combination of four Gata.a binding sites and four Tcf7 binding sites, and expression in the animal and vegetal hemispheres like *Tbx21* expression was reproduced with twelve Gata.a binding sites and four Tcf7 binding sites. In addition, animal hemisphere specific expression like *Efna.d* expression was reproduced with twelve Gata.a binding sites.

Discussion

Gata.a has been identified as a transcription factor that activates genes expressed in the animal hemisphere at the 16- and 32-cell stages (Bertrand et al., 2003; Rothbächer et al., 2007). It is also known that Gata.a activity is weakened by nuclear β-catenin (Oda-Ishii et al., 2016; Rothbächer et al., 2007). Because nuclear β-catenin is not present in animal hemisphere cells at the 16-cell stage or in marginal cells of the vegetal hemisphere at the 32-cell stage, Gata.a can activate its targets in these cells (Hudson et al., 2013; Imai et al., 2016; Oda-Ishii et al., 2016). The present results suggest that Gata.a has an additional role in activating genes in early embryos; Gata.a is necessary for β-catenin/Tcf7 to efficiently activate its targets in the vegetal hemisphere, where nuclear β-catenin is present.

It is unlikely that genes activated in the animal hemisphere by Gata.a activates *Fgf9/16/20*, *Foxd*, and *Tbx6-r.b* indirectly through cell-cell interactions, because, as we showed above, Gata.a binding sites are required for expression of the reporters for *Fgf9/16/20*, *Foxd*, and *Tbx6-r.b*. In addition, these genes are expressed in continuously dissociated embryonic cells, in which cell-cell interactions are not expected to occur (Noda et al., 2009). Therefore, the above observations indicated that Gata.a regulates genes with vegetal hemisphere-specific expression cell-autonomously.

It is likely that a regulatory region with high affinity for Gata.a is not required for gene expression in the vegetal hemisphere, because two Gata.a-binding sites can promote Tcf7-binding sites to function but cannot induce expression in the animal hemisphere in our synthetic constructs. Although we changed only the number of Gata.a-binding sites in the present study, their quality and spacing may also be important as indicated previously (Farley et al., 2015). Because nuclear β-catenin does not completely abolish, but weakens, Gata.a binding activity, both Gata.a and Tcf7 may simultaneously bind

278 to the upstream region of this gene in vegetal cells at the 16-cell stage. Thus, Gata.a and β-catenin/Tcf7
279 may combinatorially activate vegetal genes. Meanwhile, it is also possible that Gata.a binding is required
280 for priming early enhancer regions before the 16-cell stage, and that this binding facilitates β-
281 catenin/Tcf7 binding at the 16-cell stage. Indeed, while nuclear β-catenin is rarely detectable before the
282 16-cell stage, Gata.a is found in nuclei even at the 2-cell stage (Oda-Ishii et al., 2018). This hypothesis is
283 reminiscent of the function of Zelda in *Drosophila* embryos (Liang et al., 2008) and those of Nanog,
284 Pou5f1, and SoxB1 in vertebrate embryos (Foygel et al., 2008; Lee et al., 2013; Leichsenring et al.,
285 2013; Pan and Schultz, 2011). These factors, like Gata proteins, are all considered to be pioneer
286 transcription factors (Iwafuchi-Doi and Zaret, 2014), and binding of pioneer factors to genes that are
287 expressed in early embryos may be a widely conserved mechanism among animals. The above two
288 hypotheses are not mutually exclusive, and may not be distinguishable in ascidian embryos. In addition,
289 we found that several sequence motifs other than the Gata binding motif were also more abundant in the
290 upstream regions of genes expressed in early embryos (Fig. 1CD). These sites might also play a role in
291 activating genes in early embryos, although this possibility has not been tested.

292 It is possible that Tcf7 also represses *Fgf9/16/20*, *Foxd*, and *Tbx6-r.b* in the animal hemisphere,
293 where β-catenin is not translocated into the nucleus. However, because mutations introduced into the
294 Tcf7-binding sites did not induce ectopic expression of reporters for *Fgf9/16/20, Foxd,* or *Tbx6-r.b* in the
295 animal hemisphere (Oda-Ishii et al., 2016), it is unlikely that Tcf7 sites are essential for repressing
296 *Fgf9/16/20*, *Foxd*, and *Tbx6-r.b* in the animal hemisphere. This also suggests that Gata.a binding sites in
297 the upstream regulatory regions of these genes are not strong enough for activation in the animal
298 hemisphere. Nevertheless, because the expression of synthetic reporter constructs with twelve Gata sites
299 was slightly reduced in the animal hemisphere by addition of two or four Tcf7 sites (Fig. 8), Tcf7 sites

300 may negatively, but weakly, regulate these genes in the animal hemisphere, where β-catenin is not
301 translocated to the nucleus. Thus, the balance between the binding sites of Gata.a and Tcf7 is important
302 for activating these genes specifically in the animal and vegetal hemispheres.

Materials and methods

Animals and cDNAs

Adult *C. intestinalis* (type A; also called *C. robusta*) were obtained from the National Bio-Resource Project for *Ciona intestinalis* in Japan. cDNA clones were obtained from our EST clone collection (Satou et al., 2005). Identifiers (Satou et al., 2008; Stolfi et al., 2015) for genes examined in the present study are as follows: CG.KH2012.S654.3 for *Tbx6-r.b*, CG.KH2012.C2.125 for *Fgf9/16/20*, CG.KH2012.C1.1271 for *Tfap2-r.b*, CG.KH2012.C3.762 for *Efna.d*, CG.KH2012.C8.890/396 for *Foxd*, CG.KH2012.C3.773 for *Tbx21*, CG.KH2012.L20.1 for *Gata.a*, CG.KH2012.C1.727 for *Zic-r.a*, CG.KH2012.C3.411 for *Lefty*, CG.KH2012.C2.421 for *Admp*, CG.KH2012.C9.257 for *Wnttun5*, and CG.KH2012.C4.85 for *Pou2f*.

RNA sequencing (RNA-seq)

For RNA-seq experiments, 100 embryos were collected at both the 16- and 32-cell stages. RNA was extracted using the Quick-RNA kit (Zymo Research) and ribosomal RNAs were removed using the Low Input RiboMinus Eukaryote System v2 (Thermo Fischer Scientific). Libraries were constructed using the Ion Total RNA-Seq kit ver 2 (Thermo Fischer Scientific), and sequenced with an Ion PGM instrument (Thermo Fischer Scientific) (SRA accession number: DRA007136). Biological duplicates were examined at each stage. Sequence reads were mapped onto the KH-version of the genome sequence (Satou et al., 2008) using the HISAT2 program (Kim et al., 2015) and analyzed using the Samtools program (Li et al., 2009). The KH Gene model set (Satou et al., 2008) was used to annotate introns. Analyses were limited to genes with introns that were equal to or longer than 500 bp in total and

323 that did not overlap with exons of other genes. When sequence reads covered 500 bp or more of introns,
324 and coverage length was equal to or more than one fifth of the total intron length of a given gene, the
325 gene was regarded as “abundantly transcribed”. When no sequence reads were mapped onto introns of a
326 given gene, it was regarded as “non-transcribed”. Genes that were not categorized into either of these
327 two groups were regarded as “rarely transcribed”. Gata sites were identified by the Patser program
328 (Hertz and Stormo, 1999) with a position weight matrix (PWM) determined by *in vitro* selection using
329 the *Ciona* Gata.a protein (Brozovic et al., 2018; Nitta et al., 2019; Vincentelli et al., 2011) with
330 parameters “-li -c -A a:t 65 g:c 35”. Scores shown in Fig. 3 were calculated with parameters “-ls 1 -c -A
331 a:t 65 g:c 35.”

332 **Gene knockdown assays and reporter assays**

333 The MO (Gene Tools, LLC) against *Gata.a*, which blocked translation, has been used
334 previously (5′ -GGGTTAGGCATATACATTCTTTGGA-3′) (Bertrand et al., 2003; Imai et al., 2016;
335 Oda-Ishii et al., 2016; Rothbächer et al., 2007). We used the standard control MO, also purchased from
336 Gene Tools. The MOs were introduced into unfertilized eggs by microinjection, unless otherwise noted.

337 The upstream sequences used to generate the reporter constructs used in experiments shown in
338 Figures 4 and 8 are provided in Figures 3 and 7, respectively. Reporter constructs were introduced by
339 electroporation. *Gfp* was used as a reporter, and reporter expression was examined by *in situ*
340 hybridization in experiments shown in Figures 4 and 8. In the experiment shown in Figure 5, *lacZ* was
341 used as a reporter and reporter expression was examined by histochemical staining of β-galactosidase
342 activity. Gata sites were mutated by replacing the core GATA with GCGA in the experiments shown in
343 Figure 4. Mutations introduced in the constructs used in the experiment shown in Figure 5 are shown

within the figure. The minimal promoter region of *Brachyury* (KhS1404:6203-6259; Fig. 7I) was used to generate the synthetic constructs shown in Figure 7. Two independent experiments were performed.

Whole-mount *in situ* hybridization and RT-qPCR

In situ hybridization was performed as described previously (Ikuta and Saiga, 2007; Satou et al., 1995). RT-qPCR to quantify gene expression was performed using the Cell-to-Ct kit (Thermo Fisher Scientific) with the TaqMan chemistry. For relative quantification, *Zic-r.a* (*Macho-1*) was used as an internal control. *Gata.a* morphants were compared with embryos injected with the control MO. Probes and primers were as follows: *Tfap2-r.b*, 5'-FAM-TACACCAGCTATTTGCGCTGCGATGA-TAMRA-3', 5'-CCAACGACCTCTTACACATTTCAG-3', 5'-GATAACGCAGCATCTCCGTTAAGT-3'; *Efna.d*, 5'-FAM-TTGTCGCTGTACCACGCAACGGAA-TAMRA-3', 5'-CGGATTTCGTTTCCAGTATTGC-3', 5'-GCCGCTCTGTTTGCCTCTT-3'; *Fgf9/16/20*, 5'-FAM-TTGCCAGGTAGAGACCACTTGCGACACC-TAMRA-3', 5'-ACCCAAGAAAGCCACAATCAATACG-3', 5'-TCCGAAGCATACAATCTTCCTTTGC-3'; *Foxd*, 5'-FAM-TCATTATCGTCACCAGCAACCCTTGTACG-TAMRA-3', 5'-AACTCAACATTCAGCTTTGAACGA-3', 5'-ATTTCGGCAACCAGTTTTGG-3'; *Tbx6-r.b*, 5'-FAM-CCATTGTTGCCCGCTGCAAGGTGAGT-TAMRA-3', 5'-AACCCCAAGTTCCGCAGAGA-3', 5'-CATGGAGTGTATGAGGAACTTTCCA-3'; *Sox1/2/3*, 5'-FAM-ATTTATGGTGTGGTCTCGCGGGCAA-TAMRA-3', 5'-CAAAGTACCACAAGAGCAGAGAGTGA-3', 5'-GGTTGTCCTGTGCCATCTTTCT-3'; *Pou2f*, 5'-VIC-TGGTCCAGCCAAATCACTCACGCCTA-TAMRA-3', 5'-TACCACAGCATACACTGGACAACA-3', 5'-GGCGCTGAGGTAATGCTTTG-3'; *Zic-r.a,* 5'-VIC-ACGGTCACTTTAGCACCTCCACCA-TAMRA-3', 5'-

CCCAGTATGCACCAAATTCAGA-3', 5'-TGGTGAGAAAACGGGTGAAAC-3'.

RT-qPCR to quantify gene expression in *C. savignyi* embryos was performed using the SYBR-green chemistry. Primers were as follows: *Fgf9/16/20*, 5'-CAATGAACAGCAAAGGAAGATTGTA-3', 5'-TGCTTATTTTCCAACACAGATTCGT-3'; *Foxd*, 5'-ATAGAATGCATGAGGAACGCAGTT-3', 5'-CGTCACGTCTACGTAGGTGTCGT-3'; *Tbx6-r.b*, 5'-CGGAAGGCGAATGTTCCCT-3', 5'-CCCGGACATTTTAACTCGGTAAC-3'; *Pou2f*, 5'-TAGCTCAGGCCAAAACCTCTTAA-3', 5'-CCGCTTGATGATTGTTGTCTTG-3'; *Zic-r.a,* 5'-GCTCGGAGCGAGAATTTGAA-3', 5'-ACCAGAAAACTCGCAACAAAATG-3'.

Acknowledgements

We would like to thank Dr. Reiko Yoshida, Dr. Satoe Aratake, Dr. Manabu Yoshida, and other technical staff working under the National Bio-Resource Project (Japan Agency for Medical Research and Development) for providing the experimental animals.

Funding

This research was supported by the Japan Science and Technology Agency [grant number JPMJCR13W6], the Japan Society for the Promotion of Science [grant numbers 17KT0020 to Y.S and 18H02376 to KSI], Austrian Academy of Sciences [grant number ÖAW DOC 23593], Tiroler Wissenschaftsfonds, Nachwuchsförderung 14/2 Universität Innsbruck, and the University Innsbruck Sonderförderung [grant number 214947].

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J Mol Biol 215, 403-410.

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., Feuk, L., 2011. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat Struct Mol Biol 18, 1435-1440.

Bailey, T.L., 2011. DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics 27, 1653-1659.

Bertrand, V., Hudson, C., Caillol, D., Popovici, C., Lemaire, P., 2003. Neural tissue in ascidian embryos is induced by FGF9/16/20, acting via a combination of maternal GATA and Ets transcription factors. Cell 115, 615-627.

Brozovic, M., Dantec, C., Dardaillon, J., Dauga, D., Faure, E., Gineste, M., Louis, A., Naville, M., Nitta, K.R., Piette, J., Reeves, W., Scornavacca, C., Simion, P., Vincentelli, R., Bellec, M., Ben Aicha, S., Fagotto, M., Gueroult-Bellone, M., Haeussler, M., Jacox, E., Lowe, E.K., Mendez, M., Roberge, A., Stolfi, A., Yokomori, R., Brown, C.T., Cambillau, C., Christiaen, L., Delsuc, F., Douzery, E., Dumollard, R., Kusakabe, T., Nakai, K., Nishida, H., Satou, Y., Swalla, B., Veeman, M., Volff, J.N., Lemaire, P., 2018. ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. Nucleic Acids Research 46, D718-D725.

Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., Levine, M.S., 2015. Suboptimization of developmental enhancers. Science 350, 325-328.

Foygel, K., Choi, B., Jun, S., Leong, D.E., Lee, A., Wong, C.C., Zuo, E., Eckart, M., Pera, R.A.R., Wong, W.H., Yao, M.W.M., 2008. A Novel and Critical Role for Oct4 as a Regulator of the Maternal-Embryonic Transition. Plos One 3, e4109.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., Noble, W.S., 2007. Quantifying similarity between motifs. Genome Biol 8, R24.

Hamaguchi, M., Fujie, M., Noda, T., Satoh, N., 2007. Microarray analysis of zygotic expression of transcription factor genes and cell signaling molecule genes in early Ciona intestinalis embryos. Dev Growth Differ 49, 27-37.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol

415 Cell 38, 576-589.
416 Hertz, G.Z., Stormo, G.D., 1999. Identifying DNA and protein patterns with statistically
417 significant alignments of multiple sequences. Bioinformatics 15, 563-577.
418 Hudson, C., Kawai, N., Negishi, T., Yasuo, H., 2013. β-catenin-driven binary fate specification
419 segregates germ layers in ascidian embryos. Curr Biol 23, 491-495.
420 Hudson, C., Sirour, C., Yasuo, H., 2016. Co-expression of Foxa.a, Foxd and Fgf9/16/20 defines a
421 transient mesendoderm regulatory state in ascidian embryos. Elife 5, e14692.
422 Ikuta, T., Saiga, H., 2007. Dynamic change in the expression of developmental genes in the
423 ascidian central nervous system: revisit to the tripartite model and the origin of the
424 midbrain-hindbrain boundary region. Dev Biol 312, 631-643.
425 Imai, K., Takada, N., Satoh, N., Satou, Y., 2000. β-catenin mediates the specification of
426 endoderm cells in ascidian embryos. Development 127, 3009-3020.
427 Imai, K.S., 2003. Isolation and characterization of β-catenin downstream genes in early
428 embryos of the ascidian *Ciona savignyi*. Differentiation 71, 346-360.
429 Imai, K.S., Hino, K., Yagi, K., Satoh, N., Satou, Y., 2004. Gene expression profiles of
430 transcription factors and signaling molecules in the ascidian embryo: towards a
431 comprehensive understanding of gene networks. Development 131, 4047-4058.
432 Imai, K.S., Hudson, C., Oda-Ishii, I., Yasuo, H., Satou, Y., 2016. Antagonism between β-catenin
433 and Gata.a sequentially segregates the germ layers of ascidian embryos. Development 143,
434 4167-4172.
435 Imai, K.S., Satoh, N., Satou, Y., 2002. An essential role of a FoxD gene in notochord induction
436 in Ciona embryos. Development 129, 3441-3453.
437 Iwafuchi-Doi, M., Zaret, K.S., 2014. Pioneer transcription factors in cell reprogramming. Genes
438 Dev 28, 2679-2692.
439 Kim, D., Landmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory
440 requirements. Nature methods 12, 357-U121.
441 Lamy, C., Rothbacher, U., Caillol, D., Lemaire, P., 2006. Ci-FoxA-a is the earliest zygotic
442 determinant of the ascidian anterior ectoderm and directly activates Ci-sFRP1/5.
443 Development 133, 2835-2844.
444 Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., DiVito, K.R., Fleming, E.S., Giraldez,
445 A.J., 2013. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-
446 to-zygotic transition. Nature 503, 360-364.
447 Leichsenring, M., Maes, J., Mossner, R., Driever, W., Onichtchouk, D., 2013. Pou5f1

448 transcription factor controls zygotic gene activation in vertebrates. Science 341, 1005-1009.
449 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
450 Durbin, R., Genome Project Data Processing, S., 2009. The Sequence Alignment/Map format
451 and SAMtools. Bioinformatics 25, 2078-2079.
452 Liang, H.L., Nien, C.Y., Liu, H.Y., Metzstein, M.M., Kirov, N., Rushlow, C., 2008. The zinc-
453 finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature
454 456, 400-403.
455 Madsen, J.G., Schmidt, S.F., Larsen, B.D., Loft, A., Nielsen, R., Mandrup, S., 2015. iRNA-seq:
456 computational method for genome-wide assessment of acute transcriptional regulation from
457 total RNA-seq data. Nucleic Acids Res 43, e40.
458 Nitta, K.R., Vincentelli, R., Jacox, E., Cimino, A., Ohtsuka, Y., Sobral, D., Satou, Y., Cambillau,
459 C., Lemaire, P., 2019. High-Throughput Protein Production Combined with High-
460 Throughput SELEX Identifies an Extensive Atlas of Ciona robusta Transcription Factor
461 DNA-Binding Specificities. Methods Mol Biol 2025, 487-517.
462 Noda, T., Hamada, M., Hamaguchi, M., Fujie, M., Satoh, N., 2009. Early zygotic expression of
463 transcription factors and signal molecules in fully dissociated embryonic cells of Ciona
464 intestinalis: A microarray analysis. Development Growth & Differentiation 51, 639-655.
465 Oda-Ishii, I., Abe, T., Satou, Y., 2018. Dynamics of two key maternal factors that initiate zygotic
466 regulatory programs in ascidian embryos. Dev Biol 437, 50-59.
467 Oda-Ishii, I., Kubo, A., Kari, W., Suzuki, N., Rothbacher, U., Satou, Y., 2016. A maternal system
468 initiating the zygotic developmental program through combinatorial repression in the
469 ascidian embryo. PLoS genetics 12, e1006045.
470 Pan, H., Schultz, R.M., 2011. SOX2 Modulates Reprogramming of Gene Expression in Two-Cell
471 Mouse Embryos. Biol Reprod 85, 409-416.
472 Rothbächer, U., Bertrand, V., Lamy, C., Lemaire, P., 2007. A combinatorial code of maternal
473 GATA, Ets and β-catenin-TCF transcription factors specifies and patterns the early ascidian
474 ectoderm. Development 134, 4023-4032.
475 Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A., Satoh, N., 2005. An integrated database
476 of the ascidian, Ciona intestinalis: Towards functional genomics. Zool Sci 22, 837-843.
477 Satou, Y., Kusakabe, T., Araki, S., Satoh, N., 1995. Timing of Initiation of Muscle-Specific Gene-
478 Expression in the Ascidian Embryo Precedes That of Developmental Fate Restriction in
479 Lineage Cells. Dev Growth Differ 37, 319-327.
480 Satou, Y., Mineta, K., Ogasawara, M., Sasakura, Y., Shoguchi, E., Ueno, K., Yamada, L.,

Matsumoto, J., Wasserscheid, J., Dewar, K., Wiley, G.B., Macmil, S.L., Roe, B.A., Zeller, R.W., Hastings, K.E.M., Lemaire, P., Lindquist, E., Endo, T., Hotta, K., Inaba, K., 2008. Improved genome assembly and evidence-based global gene model set for the chordate Ciona intestinalis: new insight into intron and operon populations. Genome Biol 9, R152.

Small, K.S., Brudno, M., Hill, M.M., Sidow, A., 2007. A haplome alignment and reference sequence of the highly polymorphic Ciona savignyi genome. Genome Biol 8, R41.

Stolfi, A., Sasakura, Y., Chalopin, D., Satou, Y., Christiaen, L., Dantec, C., Endo, T., Naville, M., Nishida, H., Swalla, B.J., Volff, J.N., Voskoboynik, A., Dauga, D., Lemaire, P., 2015. Guidelines for the nomenclature of genetic elements in tunicate genomes. Genesis 53, 1-14.

Takatori, N., Hotta, K., Mochizuki, Y., Satoh, G., Mitani, Y., Satoh, N., Satou, Y., Takahashi, H., 2004. T-box genes in the ascidian Ciona intestinalis: Characterization of cDNAs and spatial expression. Dev Dyn 230, 743-753.

UniProt, C., 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47, D506-D515.

Vincentelli, R., Cimino, A., Geerlof, A., Kubo, A., Satou, Y., Cambillau, C., 2011. High-throughput protein expression screening and purification in Escherichia coli. Methods 55, 65-72.

499 **Figure Legends**

500 **Figure 1. Genes activated in early embryos have more Gata.a binding sites** (A) Mapping of RNA-
501 seq reads obtained from 16-cell and 32-cell embryos onto the genomic region spanning *Tbx21*. Reads
502 mapped onto introns are shown in magenta, which indicates zygotic transcription. (B, B') Expression of
503 *Tbx21* revealed by fluorescence *in situ* hybridization. Nuclear signals, indicating zygotic expression, are
504 clearly present in all cells except the most posterior germ line cells (arrowheads). (B') A high
505 magnification view of b5.4 of an embryo different from the one shown in (B). Two distinct signals for
506 zygotic transcription are seen in the nucleus (arrowheads). (C) Numbers of individual hexamers per kilo
507 base in 1-kb upstream regions, compared between genes transcribed at the 16- and 32-cell embryos and
508 genes that are not transcribed at any of these stages. Magenta dots indicate hexamers overrepresented in
509 genes expressed in the former gene group. (D) Similarities of the overrepresented hexamers to known
510 transcription binding motifs examined by the Tomtom program (Gupta et al., 2007). (E) Sequence logo
511 representing the position weight matrix for *Ciona* Gata.a. (F) Numbers of Gata.a sites, (G) highest
512 Gata.a ChIP-seq signal values within 1-kb upstream regions, and (H) mean fold change values of the
513 Gata.a ChIP signal over the upstream 1-kb regions, compared between genes transcribed at the 16- and
514 32-cell stages and genes that are not transcribed in early embryos. Differences were assessed by
515 Wilcoxon rank–sum tests, and p-values are shown. The whiskers represent the smallest and largest
516 values in the 1.5× interquartile range. Outliers are considered in statistical tests, but not shown in graphs.

517 **Figure 2. Tcf7 and Gata.a extensively share target sites.** (A) Mapping of Gata.a and Tcf7 ChIP-seq
518 data onto the genomic region consisting of the first exon (a black box) and upstream regions of
519 *Fgf9/16/20*, *Lefty*, *Admp*, *Wnttun5*, and *Tbx6-r.b*. The upstream region of *Pou2f* is also shown as a

520 control. The y-axes show fold enrichment. Upstream regions sufficient for driving a reporter, revealed
521 previously (Oda-Ishii et al., 2016), are indicated by double-headed arrows with their relative nucleotide
522 positions. Peak regions identified by the Homer program (Heinz et al., 2010) are shown as colored
523 boxes. ChIP-seq data are from a previous study (Oda-Ishii et al., 2016). (B) Venn diagram showing
524 overlaps of ChIP-seq peaks between Tcf7 and Gata.a. (C) Heatmap of Gata.a ChIP-seq signals over 6-kb
525 regions around the 4458 Tcf7 peaks (position 0) that overlapped with Gata.a peaks. Red, enrichment;
526 white, no enrichment.

527 **Figure 3. Alignments of the upstream nucleotide sequences of *Fgf9/16/20*, *Foxd*, and *Tbx6-r.b* of**
528 **two *Ciona species*.** The core "GATA" sequences mutated in reporter assays in Figure 4 are shown in
529 magenta in the upstream region of genes of *Ciona intestinalis* (type A, or *C. robusta*). "GATA"
530 sequences found in similar positions within the corresponding upstream regions of genes of *C. savignyi*
531 are also shown in magenta. Scores calculated by the Patser program (Hertz and Stormo, 1999) using the
532 position weight matrix for *Ciona* Gata.a protein (Brozovic et al., 2018; Vincentelli et al., 2011) are
533 shown for each. Nucleotide positions in the KH-version of the *C. intestinalis* genome sequence (Satou et
534 al., 2008) and the reftig-version of the *C. savignyi* genome (Small et al., 2007) are shown. Asterisks
535 indicate conserved nucleotides. The "GATA" sequence underscored in (B) indicates that an additional
536 site with a low score (0.25) flanked with "TTT".

537 **Figure 4. Gata.a sites are required for expression of β-catenin/Tcf7 target genes at the 16-cell stage.**
538 Expression of reporter constructs revealed by *in situ* hybridization. Reporter constructs for (A, B)
539 *Fgf9/16/20*, (C, D) *Foxd*, and (E, F) *Tbx6-r.b* with (A, C, E) intact Gata sites or (B, D, F) mutated Gata
540 sites were analyzed. Numbers of embryos examined and percentages of embryos that expressed the

reporter are shown within each panel. Numbers of cells that expressed the reporter in individual embryos are shown in the graphs. Two independent experiments are represented in white and gray in each bar. Note that endogenous *Fgf9/16/20* and *Foxd* are expressed in six cells of 16-cell embryos, and *Tbx6-r.b* is in two cells. Differences were assessed by Wilcoxon rank–sum tests, and p-values are shown Scale bar in (B) represents 50 μm.

Figure 5. Three TTT sequences flanking Gata.a sites are not essential for *Foxd* expression. The critical upstream regions of reporter constructs for *Foxd* with or without mutations in four Gata.a sites and/or three flanking TTT sequences (boxes) are depicted in the left. Mutated nucleotides are shown by magenta. Two independent *lacZ* assays were performed, and were represented by different bars. Numbers of embryos examined in *lacZ* assays at the early gastrula stage are shown on the right of the bars. Percentages of embryos with reporter expression in the two experiments are shown by arrows.

Figure 6. Gata.a is required for gene expression at the 16-cell stage. (A) Relative mRNA abundance of *Tfap2-r.b*, *Efna.d*, *Fgf9/16/20*, *Foxd*, and *Tbx6-r.b* in *Gata.a* morphants compared against embryos injected with the control MO. *Pou2f* is expressed maternally but not zygotically, and was therefore included as a control. Another maternally expressed mRNA, *Zic-r.a*, was used as an internal reference. Six independent sets of experiments (biological replicates) were performed, and are represented by different bars. Differences in expression levels between embryos injected with the *Gata.a* and control MOs were analyzed by paired *t*-tests. (B–F) Experiments to confirm that the MO for *C. intestinalis Gata.a* acts specifically. (B) Relative mRNA abundance of *Efna.d*, *Fgf9/16/20*, *Foxd*, and *Tbx6-r.b* in *Gata.a* morphants of *C. savignyi* compared against embryos injected with the control MO. *Pou2f* is expressed maternally but not zygotically, and was therefore included as a control. Another maternally

562 expressed mRNA, *Zic-r.a*, was used as an internal reference. Six independent sets of experiments
563 (biological replicates) were performed, and are represented by different bars. Differences in expression
564 levels between embryos injected with the *Gata.a* and control MOs were analyzed by paired *t*-tests. (C–F)
565 An experiment to show that the *Gata.a* MO of *C. intestinalis* inhibits translation of Gata.a protein. (C,
566 D) *In situ* hybridization of *Gata.a* mRNA at the early tailbud stage (C) in a normal embryo and (D) in a
567 *Gata.a* morphant embryo. *Gata.a* mRNA is expressed in endodermal cells. Numbers of embryos
568 examined and percentages of embryos that expressed *Gata.a* are shown within each panel. (E, F)
569 Immunostaining of Gata.a with the specific antibody at the middle tailbud stage. Signals for Gata.a
570 protein are visible in normal embryos (E) but not in morphant embryos (F). Numbers of embryos
571 examined and percentages of embryos that expressed Gata.a are shown within each panel. (E', F')
572 Higher magnification views of the trunk region of the embryos shown in (E) and (F). (E'', F'') Bright
573 field images for the same regions shown in (E') and (F'). Arrowheads in (E) and (E') indicate Gata.a
574 expression. Scale bars represent 100 μm. Note that *Gata.a* mRNA was detectable in endodermal cells of
575 normal and *Gata.a* morphant embryos, and that Gata.a protein was not detected in *Gata.a* morphant
576 embryos. In other words, the *Gata.a* MO blocked translation of *Gata.a* mRNA. In this experiment, we
577 injected the MO after fertilization to minimize effects on the maternal *Gata.a* mRNA.

578 **Figure 7. Nucleotide sequences of synthetic reporter constructs with various numbers of Gata.a**
579 **and Tcf7-binding sites.** To test whether specific expression patterns can be reproduced with specific
580 combinations of Gata.a sites and Tcf7 sites, we prepared eight reporter constructs with various numbers
581 of Gata.a and Tcf7-binding sites, which are shown in (A–H). The core 'GATA' sequences of Gata.a-
582 binding sites are enclosed by boxes, and the core 'CAAAG' sequences of Tcf7-binding sites are shown
583 in magenta. These sequences are connected to the minimal *Brachyury* promoter shown in (I).

584 **Figure 8. Expression of synthetic reporter constructs at the 16-cell stage.** (A) Numbers of embryos
585 examined and percentages of embryos with reporter expression in the animal hemisphere and vegetal
586 hemisphere. Two independent reporter assays were performed, and were represented by different bars.
587 Mean percentages of embryos with reporter expression in the two experiments are shown. (B–I)
588 Photographs of embryos examined for reporter expression by *in situ* hybridization. Left photograph
589 includes multiple embryos. Slightly higher-magnification views of the animal and vegetal hemispheres
590 are shown in the upper and lower right positions of each panel. Scale bars in (B) represent 50 μm.

591 **Figure 9. Gata.a-binding sites act together with the Tcf7-binding sites for gene expression in the**
592 **vegetal hemisphere.** (A) Numbers of embryos examined and percentages of embryos with reporter
593 expression in the animal hemisphere and vegetal hemisphere. Two independent reporter assays were
594 performed, and were represented by different bars. Mean percentages of embryos with reporter
595 expression in the two experiments are shown. (B, C) Photographs of embryos examined for reporter
596 expression by *in situ* hybridization. Left photograph includes multiple embryos. Slightly higher-
597 magnification views of the animal and vegetal hemispheres are shown in the upper and lower right
598 positions of each panel. Scale bars in (B) represent 50 μm. (D, E) Nucleotide sequences of the synthetic
599 reporter constructs with various numbers of Gata.a and Tcf7-binding sites. The core sequences of intact
600 and mutated Gata.a-binding sites are enclosed by boxes, and the core sequences of intact and mutated
601 Tcf7-binding sites are shown in magenta. Mutated nucleotides are shown in lower letters.

602

603 **Table 1. Genes that are abundantly transcribed in the 16-cell embryo.**

Gene	*Ciona* gene name	The best hit protein in the human proteome# Protein name (accession number; Blast E-value)
KH.C1.578		prolylcarboxypeptidase (P42785; 1.33E-152)
KH.C1.787		tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 2 (Q9HCD6; 0)
KH.C1.99	*Sox1/2/3*	SRY-box 2 (P48431; 8.9E-51)
KH.C10.536		tRNA-yW synthesizing protein 1 homolog B (Q6NUM6; 0.092)
KH.C10.574	*Zf143* (*Fog*)	zinc finger protein 226 (Q9NYT6; 0.0000000000795)
KH.C11.149		ras homolog family member U (Q7L0Q8; 1.34E-61)
KH.C11.257		solute carrier organic anion transporter family member 4A1 (Q96BD0; 2.49E-133)
KH.C11.334		chromatin licensing and DNA replication factor 1 (Q9H211; 8.38E-97)
KH.C11.378		claudin 7 (O95471; 2.5E-31)
KH.C11.532		SprT-like N-terminal domain (Q9H040; 1.79E-74)
KH.C11.548		cingulin like 1 (Q0VF96; 0.42)
KH.C11.571		mediator of DNA damage checkpoint 1 (Q14676; 1.12E-37)
KH.C11.697		hormonally up-regulated Neu-associated kinase (P57058; 5.34E-105)
KH.C12.444		structural maintenance of chromosomes 6 (Q96SB8; 9.25E-84)
KH.C12.445		ribosomal protein S6 kinase C1 (Q96S38; 4.26E-54)
KH.C12.91		RNA binding motif protein 19 (Q9Y4C8; 5.69E-79)
KH.C13.152		beta-1,3-galactosyltransferase 5 (Q9Y2C3; 1.16E-25)
KH.C13.165		serpin family B member 6 (P35237; 1.75E-83)
KH.C13.22	*Zf220*	Sp9 transcription factor (P0CG40; 7.08E-45)
KH.C13.28		pleckstrin homology like domain family B member 2 (Q86SQ0; 0.23)
KH.C13.61		REV3 like, DNA directed polymerase zeta catalytic subunit (O60673; 1)
KH.C14.312		PDZ and LIM domain 3 (Q53GG5; 2.48E-36)
KH.C14.356		no hit
KH.C14.414		Pim-1 proto-oncogene, serine/threonine kinase (P11309; 1.15E-87)

KH.C14.51		receptor interacting serine/threonine kinase 3 (Q9Y572; 0.000000352)
KH.C14.520	*Foxtun1*	forkhead box I2 (Q6ZQN5; 0.00000000000655)
KH.C2.1026		DLC1 Rho GTPase activating protein (Q96QB1; 3.59E-158)
KH.C2.1104		protocadherin 9 (Q9HC56; 3.46E-123)
KH.C2.1123		ADAM metallopeptidase with thrombospondin type 1 motif 3 (O15072; 0.0000000372)
KH.C2.1129		mitochondrial ribosomal protein S26 (Q9BYN8; 0.0000000623)
KH.C2.125	*Fgf9/16/20*	fibroblast growth factor 9 (P31371; 1.35E-44)
KH.C2.27		neuropilin 2 (O60462; 5.08E-22)
KH.C2.573	*Smad1/5/9*	SMAD family member 1 (Q15797; 0)
KH.C2.992		spondin 1 (Q9HCB6; 0.0000519)
KH.C2.994		ring finger protein 149 (Q8NC42; 0.00000000000973)
KH.C3.164		terminal nucleotidyltransferase 4B (Q8NDF8; 7.78E-121)
KH.C3.22		neural precursor cell expressed, developmentally down-regulated 9 (Q14511; 6.77E-31)
KH.C3.272		catenin alpha 1 (P35221; 5.38E-141)
KH.C3.30		Snf2 related CREBBP activator protein (Q6ZRS2; 0)
KH.C3.470		transducin beta like 3 (Q12788; 3.69E-135)
KH.C3.705		gonadotropin releasing hormone receptor (P30968; 2.28E-41)
KH.C3.716	*Efna.d*	ephrin A5 (P52803; 6.45E-18)
KH.C3.738		KIAA0586 (Q9BVV6; 0.68)
KH.C3.762	*Efna.a*	ephrin A1 (P20827; 0.000000000000138)
KH.C3.773	*Tbx21*	T-box 4 (P57082; 1.96E-44)
KH.C4.260		pleckstrin homology domain containing A4 (Q9H4M7; 0.000000379)
KH.C4.299		growth arrest and DNA damage inducible alpha (P24522; 0.26)
KH.C4.359		cell adhesion molecule 3 (Q8N126; 0.000000015)
KH.C4.391		FERM domain containing 6 (Q96NE9; 4.46E-69)
KH.C4.600		GINS complex subunit 1 (Q14691; 6.04E-83)
KH.C5.12		CD2 associated protein (Q9Y5K6; 0.25)
KH.C5.142		anti-silencing function 1A histone chaperone (Q9Y294; 5.36E-87)
KH.C5.143	*Dusp6/9*	dual specificity phosphatase 7 (Q16829; 3.7E-100)
KH.C5.520		nuclear factor, interleukin 3 regulated (Q16649; 0.000000146)

KH.C5.597		adhesion G protein-coupled receptor L2 (O95490; 0)
KH.C6.162	*Fzd4*	frizzled class receptor 4 (Q9ULV1; 0)
KH.C6.248		anoctamin 6 (Q4KMQ2; 0.27)
KH.C7.212		methylcrotonoyl-CoA carboxylase 1 (Q96RQ3; 0)
KH.C7.327		DNA meiotic recombinase 1 (Q14565; 0.13)
KH.C7.36		ASXL transcriptional regulator 1 (Q8IXJ9; 3.79E-34)
KH.C7.43	*Tfap2-r.b*	transcription factor AP-2 alpha (P05549; 2.52E-115)
KH.C7.47		UFM1 specific peptidase 2 (Q9NUQ7; 7.34E-39)
KH.C7.540		vesicle associated membrane protein 4 (O75379; 0.048)
KH.C7.787		selectin P (P16109; 0.0000483)
KH.C8.101		receptor tyrosine kinase like orphan receptor 2 (Q01974; 1.09E-138)
KH.C8.316		prickle planar cell polarity protein 2 (Q7Z3G6; 3.51E-156)
KH.C8.678		bone morphogenetic protein 1 (P13497; 2.54E-51)
KH.C8.737		receptor tyrosine kinase like orphan receptor 2 (Q01974; 2.63E-165)
KH.C9.161		solute carrier family 22 member 16 (Q86VW1; 3)
KH.C9.189		RNA polymerase II subunit A (P24928; 0.25)
KH.C9.397		ankyrin repeat and LEM domain containing 1 (Q8NAG6; 1.03E-48)
KH.C9.40		glutathione S-transferase omega 1 (P78417; 0.0000000748)
KH.C9.608		phosphodiesterase 9A (O76083; 9.2E-87)
KH.C9.695		protein phosphatase 1 regulatory subunit 10 (Q96QC0; 5.65E-62)
KH.C9.737		BRCA1 DNA repair associated (P38398; 0.000000000000178)
KH.L109.2		ubiquitin specific peptidase 34 (Q70CQ2; 1.3)
KH.L152.12		growth arrest specific 2 (O43903; 3.01E-19)
KH.L154.10		smoothelin (P53814; 2.53E-79)
KH.L170.29		protocadherin alpha 4 (Q9UN74; 1.2)
KH.L170.64		centrosomal protein 128 (Q6ZU80; 0.02)
KH.L172.5		programmed cell death 11 (Q14690; 3.11E-79)
KH.L24.7		ribonucleotide reductase regulatory subunit M2 (P31350; 6.33E-97)
KH.L28.6		insulin receptor substrate 1 (P35568; 9.63E-48)
KH.L34.21		lysosomal associated membrane protein 1 (P11279; 0.00000000000000134)
KH.L37.48		vinculin (P18206; 2.1E-96)

KH.L37.5		DIX domain containing 1 (Q155Q3; 0.0000381)
KH.L4.17	*Zf221*	spalt like transcription factor 1 (Q9NSC2; 1.16E-31)
KH.L41.2		spermatogenesis associated 7 (Q9P0W8; 0.00000000000432)
KH.L50.8		serine carboxypeptidase 1 (Q9HB40; 1.2)
KH.L83.4		oligodendrocyte transcription factor 3 (Q7RTU3; 1.3)
KH.L9.5		mitochondrial ribosomal protein L23 (Q16540; 1.8E-20)
KH.L94.1		DEP domain containing 7 (Q96QD5; 0.0000000000000193)
KH.L96.75		ATP binding cassette subfamily F member 1 (Q8NE71; 1.31E-111)
KH.S1011.2		RAS like proto-oncogene A (P11233; 1.62E-94)
KH.S494.5		alkB homolog 1, histone H2A dioxygenase (Q13686; 3.37E-58)

To find the best hit protein in the human proteome, we searched the human uniprotKB (UniProt, 2019) database with the Blastp program (Altschul et al., 1990).

607 **Table 2. Genes that are abundantly transcribed in the 32-cell embryo.**

Gene	*Ciona* gene name	The best hit protein in the human proteome# Protein name (accession number; Blast E-value)
KH.C1.142		Ras association domain family member 9 (O75901; 0.0000000000000226)
KH.C1.32		zinc finger protein 609 (O15014; 1E-55)
KH.C1.351		CDC42 binding protein kinase alpha (Q5VT25; 0.35)
KH.C1.432		choline kinase alpha (P35790; 4.74E-98)
KH.C1.637		discs large MAGUK scaffold protein 2 (Q15700; 1.48E-166)
KH.C1.972		5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase (P31939; 1)
KH.C10.461		TOPBP1 interacting checkpoint and replication regulator (Q7Z2Z1; 0.16)
KH.C11.215		GDNF family receptor alpha 2 (O00451; 1.5)
KH.C11.218		sperm antigen with calponin homology and coiled-coil domains 1 like (Q69YQ0; 5.11E-142)
KH.C11.482		SMG6 nonsense mediated mRNA decay factor (Q86US8; 1.2)
KH.C11.556		fumarate hydratase (P07954; 0.65)
KH.C12.14		cadherin 18 (Q13634; 7.24E-135)
KH.C12.229		methionine adenosyltransferase 2A (P31153; 0)
KH.C12.317		heterogeneous nuclear ribonucleoprotein A/B (Q99729; 4.34E-54)
KH.C12.491	*Bmp3*	bone morphogenetic protein 3 (P12645; 3.99E-41)
KH.C13.144		chromosome 10 open reading frame 55 (Q5SWW7; 1.8)
KH.C14.33		family with sequence similarity 167 member A (Q96KS9; 6.56E-18)
KH.C14.52		eukaryotic translation elongation factor 1 alpha 1 (P68104; 0)
KH.C2.469	*Sfrp2*	secreted frizzled related protein 2 (Q96HF1; 1.1E-68)
KH.C2.69		major intrinsic protein of lens fiber (P30301; 4.05E-32)
KH.C3.181		TPX2 microtubule nucleation factor (Q9ULW0; 1.1E-58)
KH.C3.347		pleckstrin homology and RhoGEF domain containing G4B (Q96PX9; 4.2E-132)

KH.C3.348	*Tcf3 (E12/E47)*	transcription factor 12 (Q99081; 2.68E-45)
KH.C3.426		kizuna centrosomal protein (Q2M2Z5; 0.00000000413)
KH.C3.52	*Efna.c*	ephrin B1 (P98172; 0.000000000000834)
KH.C3.596		NKD inhibitor of WNT signaling pathway 2 (Q969F2; 0.002)
KH.C3.751	*Snai (Snail)*	snail family transcriptional repressor 2 (O43623; 8.05E-59)
KH.C4.249		coiled-coil domain containing 88B (A6NC98; 2.4)
KH.C4.326		switching B cell complex subunit SWAP70 (Q9UH65; 2.85E-20)
KH.C4.410		Aly/REF export factor (Q86V81; 1.57E-38)
KH.C4.579		CDC like kinase 2 (P49760; 0)
KH.C4.760		cyclin dependent kinase 5 regulatory subunit 1 (Q15078; 3.69E-70)
KH.C4.84	*Otx*	orthodenticle homeobox 1 (P32242; 9.92E-47)
KH.C4.97		glucose-6-phosphatase catalytic subunit 2 (Q9NQR9; 0.54)
KH.C5.96		SH3 domain binding protein 4 (Q9P0V3; 3.32E-38)
KH.C6.103		oxysterol binding protein like 9 (Q96SU4; 0.0000953)
KH.C7.190		tumor protein, translationally-controlled 1 (P13693; 9.75E-17)
KH.C7.61		ninein like (Q9Y2I6; 0.00000000604)
KH.C8.163		MAP kinase interacting serine/threonine kinase 1 (Q9BUB5; 2.41E-85)
KH.C8.210		leucine rich adaptor protein 1 (Q96LR2; 0.000313)
KH.C8.247	*Elk*	ETS transcription factor ELK1 (P19419; 4.66E-43)
KH.C8.350		PDZ domain containing ring finger 4 (Q6ZMN7; 0.0000000172)
KH.C8.356		cyclin G associated kinase (O14976; 0)
KH.C8.510		hemicentin 1 (Q96RW7; 3.05E-52)
KH.C8.826		WDFY family member 4 (Q6ZS81; 0.016)
KH.C9.121		transferrin receptor (P02786; 0.43)
KH.C9.257	*Wnttun5*	Wnt family member 6 (Q9Y6F9; 3.77E-57)
KH.C9.589		bromodomain testis associated (Q58F21; 3.3E-155)
KH.C9.894		G protein-coupled receptor 101 (Q96P66; 3.9)
KH.L102.1		SH3 binding domain protein 5 like (Q7L8J4; 4.43E-57)
KH.L102.3		phytanoyl-CoA dioxygenase domain containing 1 (Q5SRE7; 4.09E-90)
KH.L119.8		sorting nexin 11 (Q9Y5W9; 0.00000397)

KH.L139.17		cyclin B3 (Q8WWL7; 5.58E-74)
KH.L152.28		protein tyrosine phosphatase receptor type R (Q15256; 1.19E-92)
KH.L152.45	*Wnt5*	Wnt family member 5A (P41221; 8.24E-132)
KH.L169.2		heparan sulfate proteoglycan 2 (P98160; 0.00000113)
KH.L170.10		repulsive guidance molecule BMP co-receptor a (Q96B86; 3.82E-59)
KH.L170.3		brevican (Q96GW7; 3.9)
KH.L170.89		TraB domain containing 2B (A6NFA1; 1.13E-126)
KH.L171.5	*Sfrp1/5*	secreted frizzled related protein 5 (Q5T4F7; 1.85E-60)
KH.L20.29	*Dkk*	dickkopf WNT signaling pathway inhibitor 4 (Q9UBT3; 0.00000000000823)
KH.L36.8		hemicentin 1 (Q96RW7; 1.73E-49)
KH.L4.50		RAN binding protein 10 (Q6VN20; 1.8)
KH.L65.9		alpha-L-fucosidase 2 (Q9BTY2; 7.59E-168)
KH.L73.1		cubilin (O60494; 2.15E-41)
KH.L96.3		Rho/Rac guanine nucleotide exchange factor 18 (Q6ZSZ5; 6.14E-93)
KH.S115.2		eva-1 homolog C (P58658; 0.039)
KH.S1725.1		SCO-spondin (A2VEC9; 0.000123)
KH.S215.4	*Lhx3/4*	LIM homeobox 3 (Q9UBR4; 1.03E-110)
KH.S390.2		RNA binding motif single stranded interacting protein 1 (P29558; 2.61E-88)

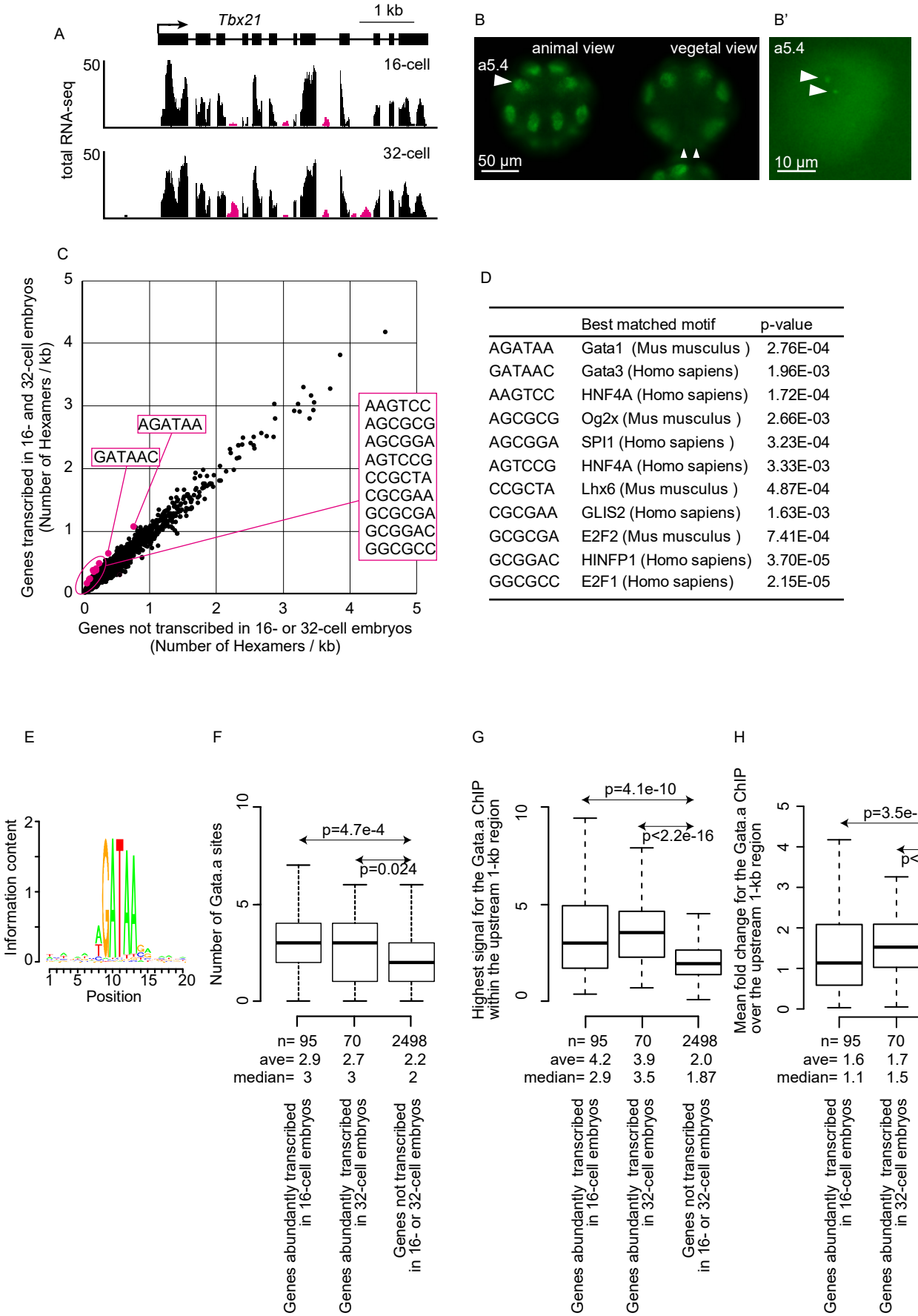
608 # To find the best hit protein in the human proteome, we searched the human uniprotKB (UniProt,
609 2019) database with the Blastp program (Altschul et al., 1990).

610

611

612

Figure 1



	Best matched motif	p-value
AGATAA	Gata1 (Mus musculus)	2.76E-04
GATAAC	Gata3 (Homo sapiens)	1.96E-03
AAGTCC	HNF4A (Homo sapiens)	1.72E-04
AGCGCG	Og2x (Mus musculus)	2.66E-03
AGCGGA	SPI1 (Homo sapiens)	3.23E-04
AGTCCG	HNF4A (Homo sapiens)	3.33E-03
CCGCTA	Lhx6 (Mus musculus)	4.87E-04
CGCGAA	GLIS2 (Homo sapiens)	1.63E-03
GCGCGA	E2F2 (Mus musculus)	7.41E-04
GCGGAC	HINFP1 (Homo sapiens)	3.70E-05
GGCGCC	E2F1 (Homo sapiens)	2.15E-05

Figure 2

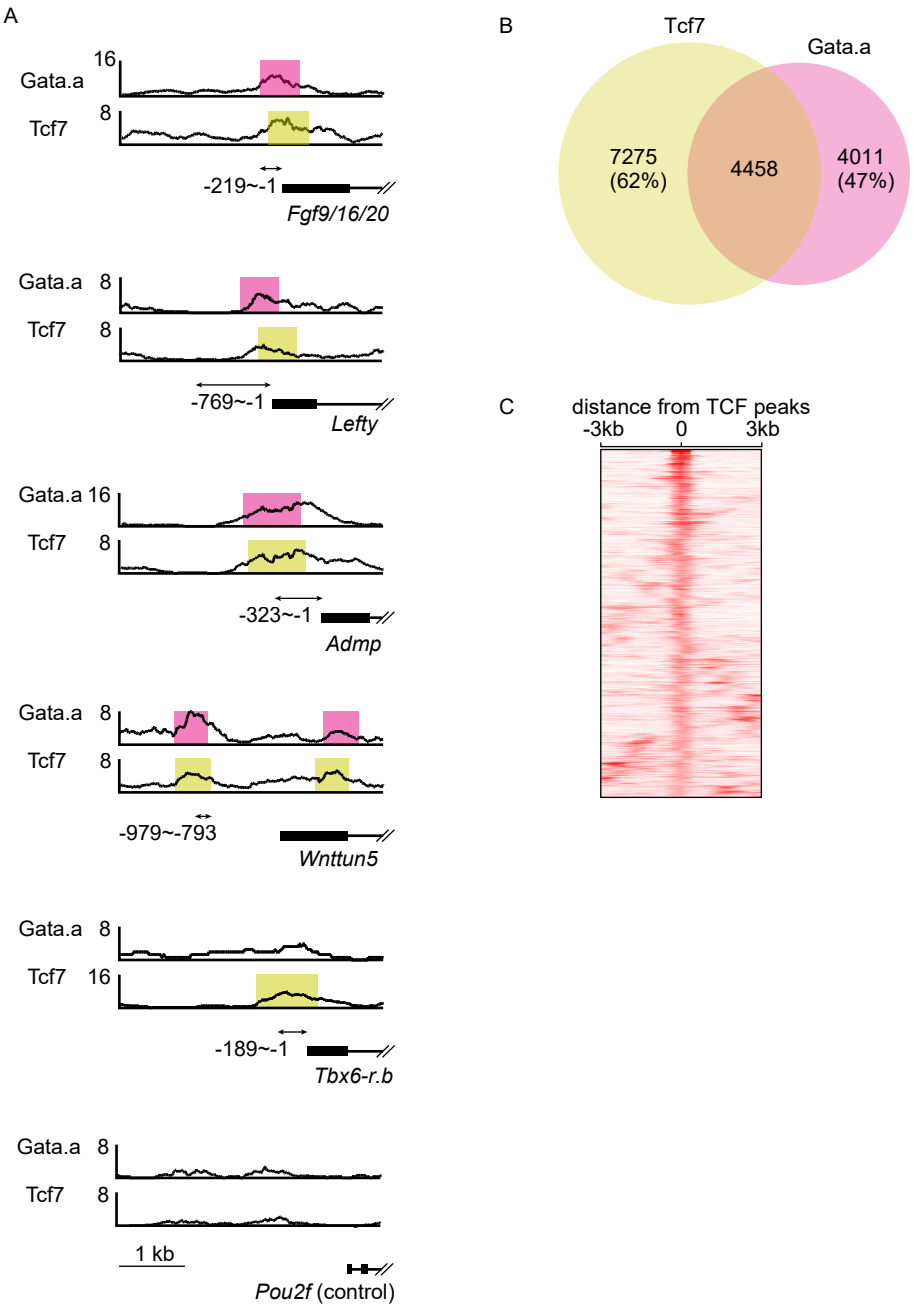

A
Gata.a
Tcf7
16
8
-219~-1
Fgf9/16/20
-769~-1
Lefty
-323~-1
Admp
-979~-793
Wnttun5
-189~-1
Tbx6-r.b
1 kb
Pou2f (control)
B
Tcf7
Gata.a
7275
(62%)
4458
4011
(47%)
C
distance from TCF peaks
-3kb
0
3kb

Figure 3

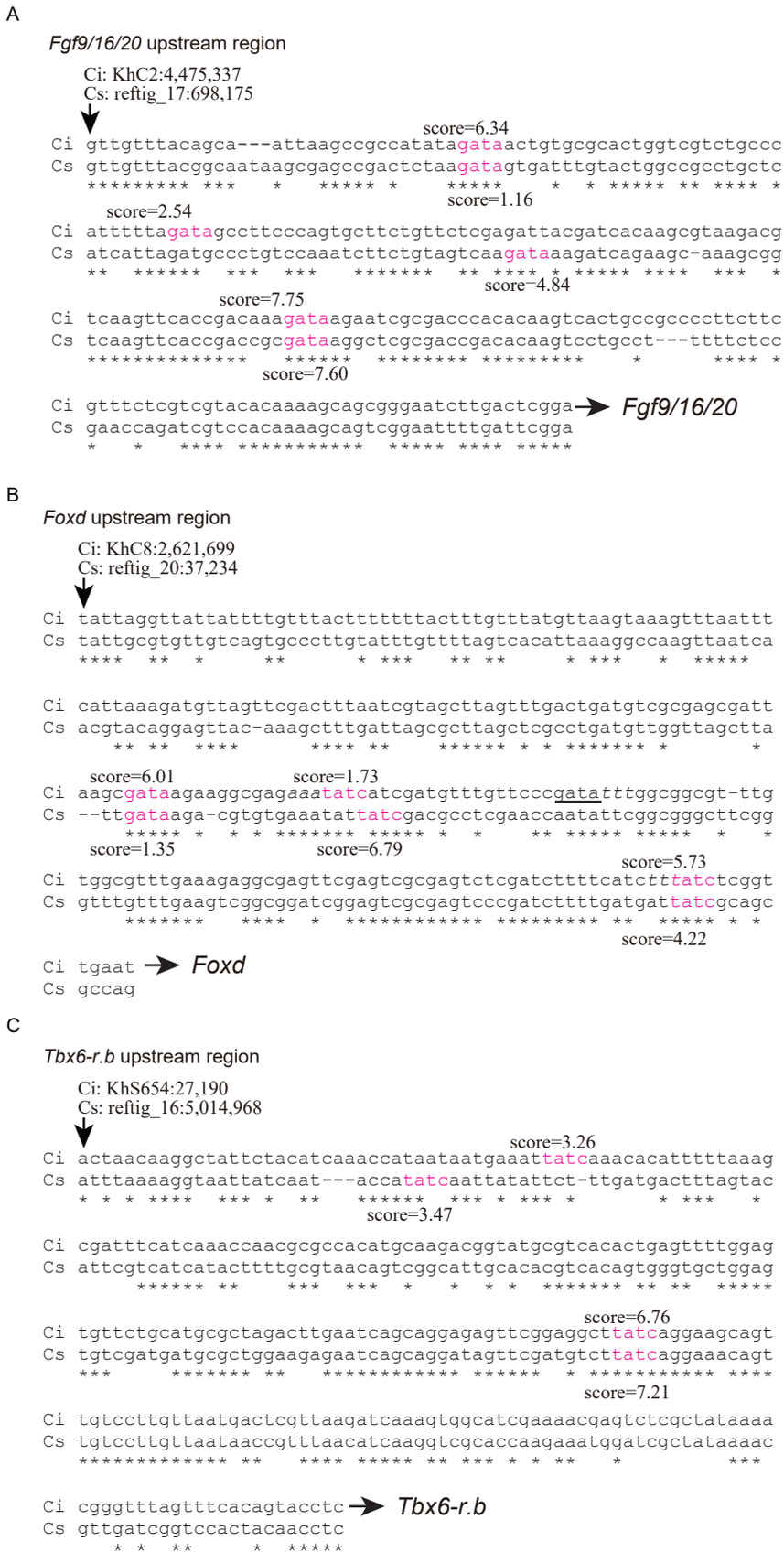

A
Fgf9/16/20 upstream region
Ci: KhC2:4,475,337
Cs: reftig_17:698,175
score=6.34
Ci gttgtttacagca---attaagccgccatatagataactgtgcgcactggtcgtctgccc
Cs gttgtttacggcaataagcgagccgactctaagatagtgatttgtactggccgcctgctc
score=1.16
score=2.54
Ci atttttagatagccttcccagtgcttctgttctcgagattacgatcacaagcgtaagacg
Cs atcattagatgccctgtccaaatcttctgtagtcaagataagatcagaagc-aaagcgg
score=4.84
score=7.75
Ci tcaagttcaccgacaaagataagaatcgcgacccacacaagtcactgccgccccttcttc
Cs tcaagttcaccgaccgcgataaggctcgcgaccgacacaagtcctgcct---ttttctcc
score=7.60
Ci gtttctcgtcgtacacaaaagcagcgggaatcttgactcgga
Fgf9/16/20
Cs gaaccagatcgtccacaaaagcagtcggaattttgattcgga
B
Foxd upstream region
Ci: KhC8:2,621,699
Cs: reftig_20:37,234
Ci tattaggttattattttgtttacttttttttactttgtttatgttaagtaaagtttaattt
Cs tattgcgtgttgtcagtgcccttgtatttgttttagtcacattaaaggccaagttaatca
Ci cattaaagatgttagttcgactttaatcgtagcttagtttgactgatgtcgcgagcgatt
Cs acgtacaggagttac-aaagctttgattagcgcttagctcgcctgatgttggttagctta
score=6.01
score=1.73
Ci aagcgataagaaggcgagaaatatcatcgatgtttgttcccgatattggcggcgt-ttg
Cs --ttgataaga-cgtgtgaaatattatcgacgcctcgaaccaatattcggcgggcttcgg
score=1.35
score=6.79
score=5.73
Ci tggcgtttgaaagaggcgagttcgagtcgcgagtctcgatcttttcatcttttatctcggt
Cs gtttgtttgaagtcggcggatcggagtcgcgagtcccgatcttttgatgattatcgcagc
score=4.22
Ci tgaat
Foxd
Cs gccag
C
Tbx6-r.b upstream region
Ci: KhS654:27,190
Cs: reftig_16:5,014,968
score=3.26
Ci actaacaaggctattctacatcaaaccataataatgaaattatcaaacacatttttaaag
Cs atttaaaaggtaattatcaat---accatatcaattatattct-ttgatgactttagtac
score=3.47
Ci cgatttcatcaaaccaacgcgccacatgcaagacggtatgcgtcacactgagttttggag
Cs attcgtcatcatacttttgcgtaacagtcggcattgcacacgtcacagtgggtgctggag
score=6.76
Ci tgttctgcatgcgctagacttgaatcagcaggagagttcggaggcttatcaggaagcagt
Cs tgtcgatgatgcgctggaagagaatcagcaggatagttcgatgtcttatcaggaaacagt
score=7.21
Ci tgtccttgttaatgactcgttaagatcaaagtggcatcgaaaacgagtctcgctataaaa
Cs tgtccttgttaataaccgtttaacatcaaggtcgcaccaagaaatggatcgctataaaac
Ci cgggtttagtttcacagtacctc
Tbx6-r.b
Cs gttgatcggtccactacaacctc

Figure 4

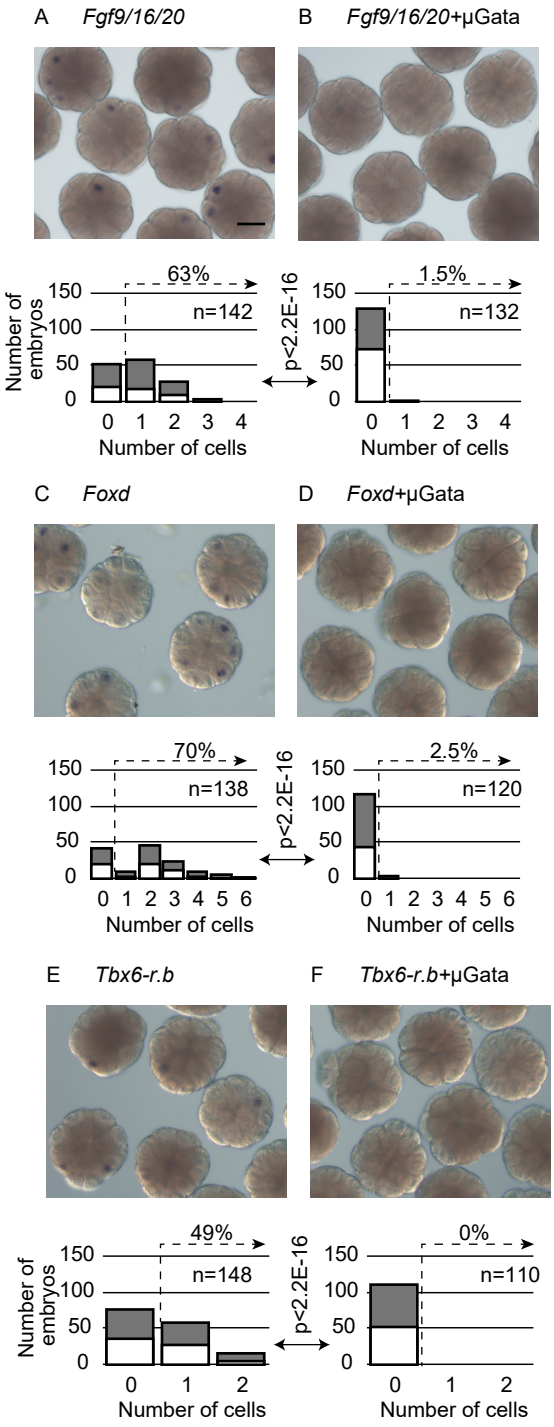

A Fgf9/16/20
B Fgf9/16/20+µGata
63%
n=142
1.5%
n=132
p<2.2E-16
Number of embryos
Number of cells
C Foxd
D Foxd+µGata
70%
n=138
2.5%
n=120
p<2.2E-16
Number of cells
E Tbx6-r.b
F Tbx6-r.b+µGata
49%
n=148
0%
n=110
p<2.2E-16
Number of embryos
Number of cells

Figure 5

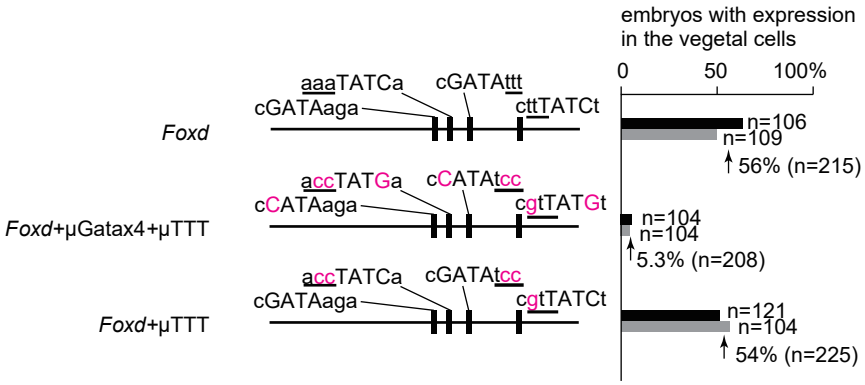

embryos with expression
in the vegetal cells
0
50
100%
aaaTATCa
cGATAttt
cGATAaga
cttTATCt
Foxd
n=106
n=109
56% (n=215)
accTATGa
cCATAtcc
cCATAaga
cgtTATGt
Foxd+μGatax4+μTTT
n=104
n=104
5.3% (n=208)
accTATCa
cGATAtcc
cGATAaga
cgtTATCt
Foxd+μTTT
n=121
n=104
54% (n=225)

Figure 6

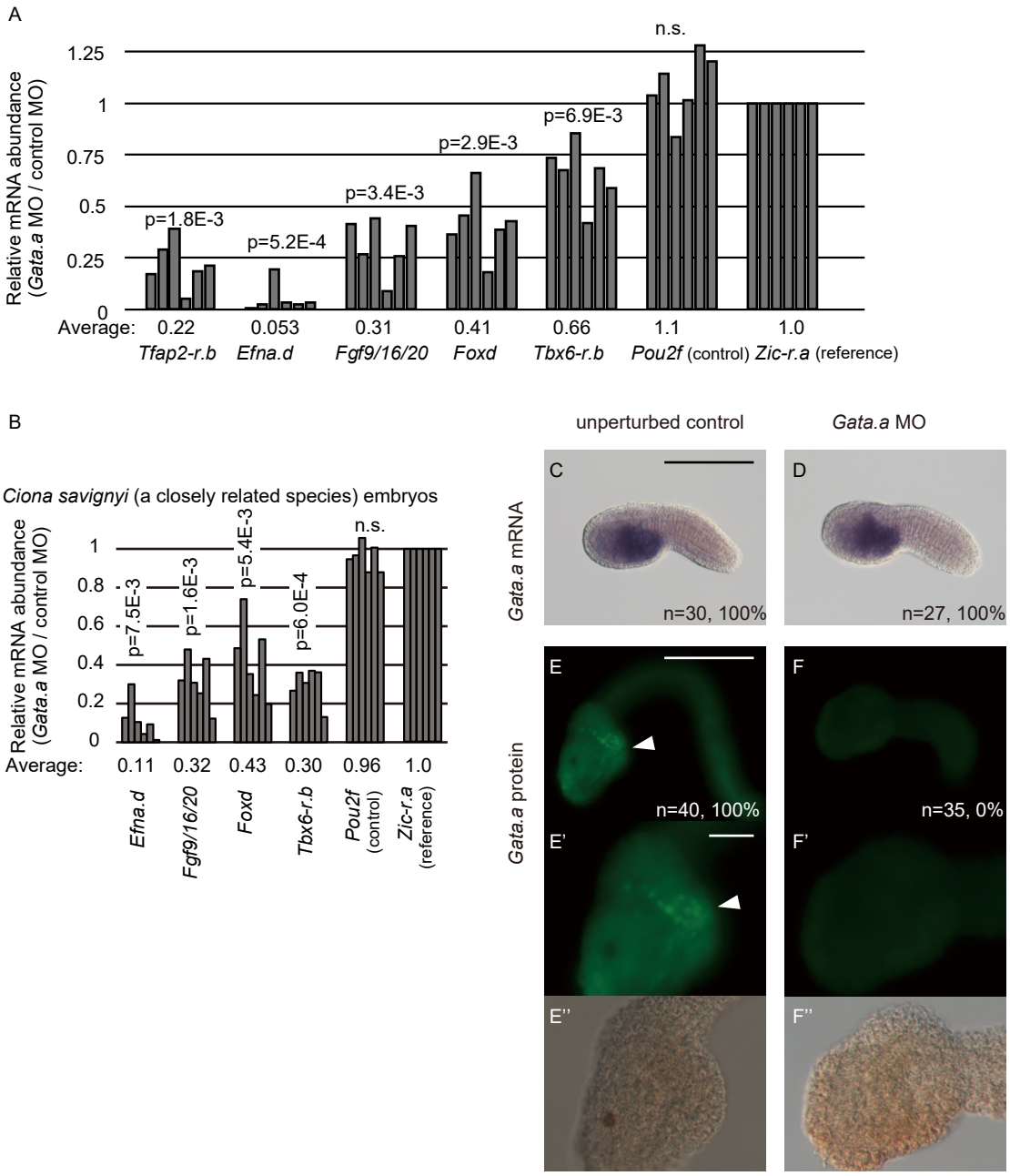

A
Relative mRNA abundance (Gata.a MO / control MO)
1.25
1
0.75
0.5
0.25
0
n.s.
p=6.9E-3
p=2.9E-3
p=3.4E-3
p=1.8E-3
p=5.2E-4
Average: 0.22 0.053 0.31 0.41 0.66 1.1 1.0
Tfap2-r.b Efna.d Fgf9/16/20 Foxd Tbx6-r.b Pou2f (control) Zic-r.a (reference)
B
Ciona savignyi (a closely related species) embryos
Relative mRNA abundance (Gata.a MO / control MO)
1
0.8
0.6
0.4
0.2
0
p=7.5E-3
p=1.6E-3
p=5.4E-3
p=6.0E-4
n.s.
Average: 0.11 0.32 0.43 0.30 0.96 1.0
Efna.d Fgf9/16/20 Foxd Tbx6-r.b Pou2f (control) Zic-r.a (reference)
unperturbed control
Gata.a MO
Gata.a mRNA
C
n=30, 100%
D
n=27, 100%
Gata.a protein
E
n=40, 100%
E'
E''
F
n=35, 0%
F'
F''

Figure 7

A Gata x12

TCCTATCTTAGATATTGAGAGATAACGTCCTATCTTAGATATTGAGAGATAACGCGTTATCTC
TCAATATCTAAGATAGGATCCTATCTTAGATATTGAGAGATAACG

B Tcf7 x2

TAAGATCAAAGGGGGTAAGATCAAAGGAGGAGCTTCAGTCGACG

C Tcf7x2 – Gata x12

TAAGATCAAAGGGGGTAAGATCAAAGGAGGAGCTTCAGTCGACGCTATTACGCCATCCTATCT
TAGATATTGAGAGATAACGTCCTATCTTAGATATTGAGAGATAACGCGTTATCTCTCAATATC
TAAGATAGGATCCTATCTTAGATATTGAGAGATAACG

D Tcf7 x4

AGCTATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGAGGAGC
TTCAGTCGACGT

E Gata x2

AAGATAAGACTCATAAAAGATAAGAAACTACTTGAGTAA

F Tcf7 x4 – Gata x2

AGCTATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGAGGAGC
TCAGTCGACGAAGATAAGACATAAAAGATAAG

G Tcf7 x4 – Gata x4

AAGATAAGACTCATAAAAGATAAGAAACTACTTGAGTAAGAGCTCTTACGCGTGCTAGAGTCG
ACCTGCAGCCCAAGCTATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGGGGTAAGAT
CAAAGGAGGAGCTTCAGTCGACGAAGATAAGACATAAAAGATAAG

H Tcf7 x 4 – Gata x12

AGCTATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGAGGAGC
TTCAGTCGACGCTATTACGCCATCCTATCTTAGATATTGAGAGATAACGTCCTATCTTAGATA
TTGAGAGATAACGCGTTATCTCTCAATATCTAAGATAGGATCCTATCTTAGATATTGAGAGAT
AACG

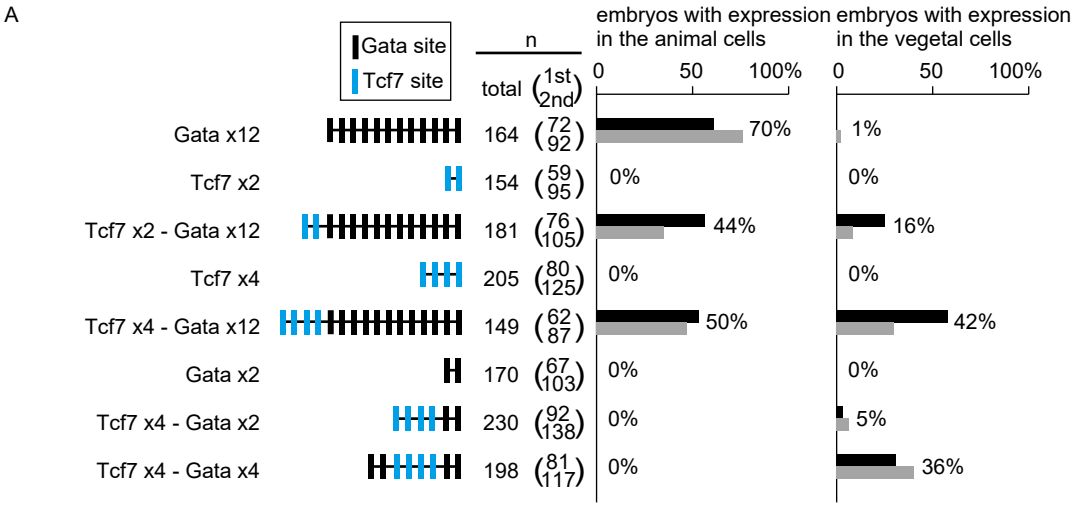
I minimal promoter of *Brachyury* (KhS1404:6203..6259)

ttatgacgtcacaatcctgtataaacttgcacccgagtgtgatttggagcgcagaatg

translation

transcription

Figure 8


A
Gata site
Tcf7 site
n
total (1st 2nd)
embryos with expression in the animal cells
embryos with expression in the vegetal cells
0
50
100%
Gata x12
164 (72 92)
70%
1%
Tcf7 x2
154 (59 95)
0%
0%
Tcf7 x2 - Gata x12
181 (76 105)
44%
16%
Tcf7 x4
205 (80 125)
0%
0%
Tcf7 x4 - Gata x12
149 (62 87)
50%
42%
Gata x2
170 (67 103)
0%
0%
Tcf7 x4 - Gata x2
230 (92 138)
0%
5%
Tcf7 x4 - Gata x4
198 (81 117)
0%
36%

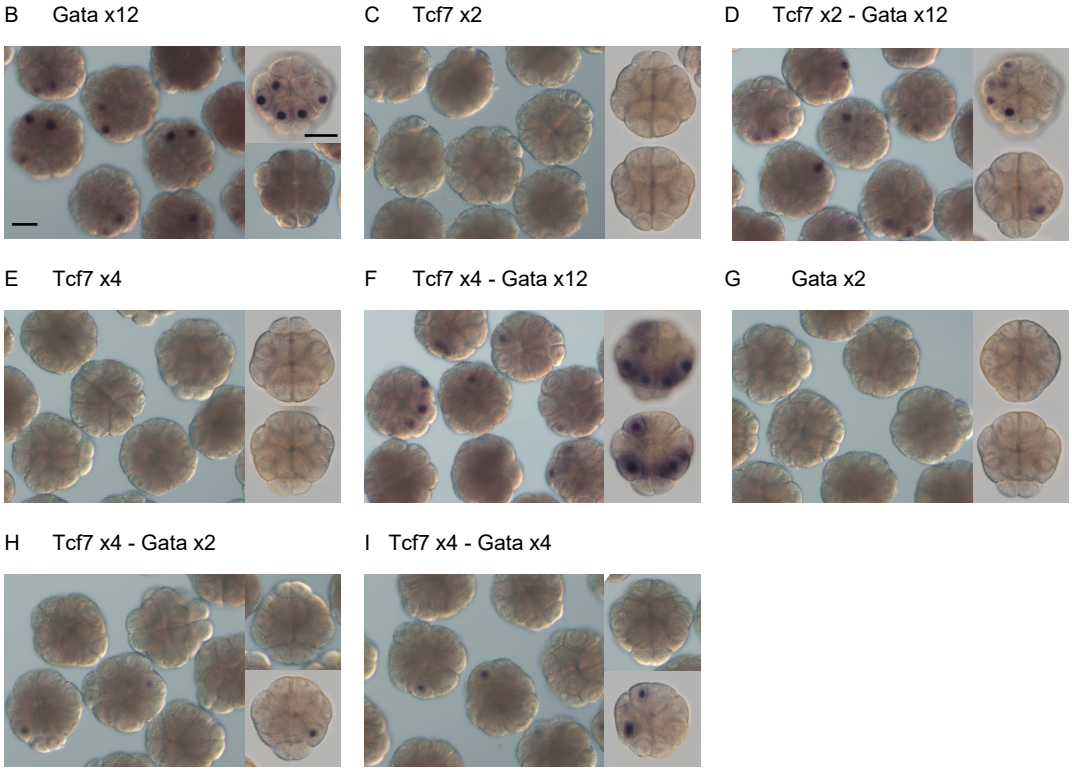
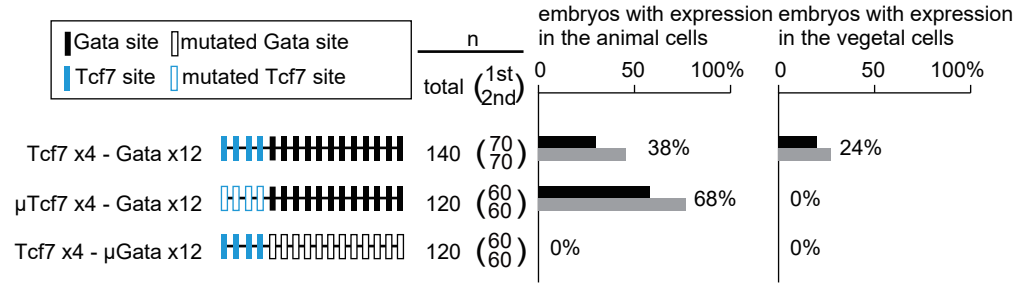
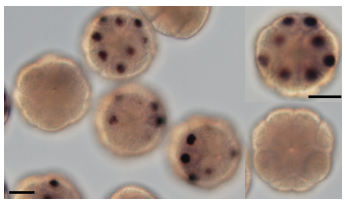

B Gata x12
C Tcf7 x2
D Tcf7 x2 - Gata x12
E Tcf7 x4
F Tcf7 x4 - Gata x12
G Gata x2
H Tcf7 x4 - Gata x2
I Tcf7 x4 - Gata x4

Figure 9

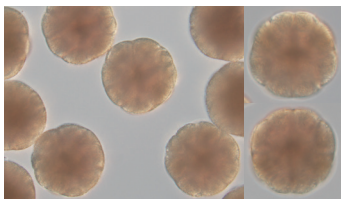
A



B μTcf7 x4 - Gata x12



C Tcf7 x4 - μGata x12



D μTcf7 x 4 – Gata x12

AGCTATaccctGGGGTAAGATaccctGGGGTAAGATCccctGGGGTAAGATCccctGAGGAGC
TTCAGTCGACGCTATTACGCCATCCTATCTTAGATATTGAGAGATAACGTCCTATCTTAGATA
TTGAGAGATAACGCGTTATCTCTCAATATCTAAGATAGGATCCTATCTTAGATATTGAGAGAT
AACG

E Tcf7 x 4 – μGata x12

AGCTATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGGGGTAAGATCAAAGGAGGAGC
TTCAGTCGACGCTATTACGCCATCCtcgcTTAgcgaTTGAGAgcgaACGTCCtcgcTTAgcga
TTGAGAgcgaACGCGTtcgcTCTCAAtcgcTAAgcgaGGATCCtcgcTTAgcgaTTGAGAgcg
aACG