

1 **A change in *cis*-regulatory logic underlying obligate versus facultative muscle**  
2 **multinucleation in chordates**

3 Christopher J. Johnson<sup>1</sup>, Zheng Zhang<sup>2,3</sup>, Haifeng Zhang<sup>3</sup>, Renjie Shang<sup>2,3</sup>, Katarzyna M.  
4 Piekarz<sup>1</sup>, Pengpeng Bi<sup>2,3</sup>, Alberto Stolfi<sup>1</sup>

5  
6 1. School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

7 2. Department of Genetics, University of Georgia, Athens, GA, USA

8 3. Center for Molecular Medicine, University of Georgia, Athens, GA, USA.

9

10

11

12

13

14

15

16

17

18

19 **Abstract**

20 Vertebrates and tunicates are sister groups that share a common fusogenic factor, Myomaker  
21 (*Mymk*), that drives myoblast fusion and muscle multinucleation. Yet they are divergent in when  
22 and where they express *Mymk*. In vertebrates, all developing skeletal muscles express *Mymk*  
23 and are obligately multinucleated. In tunicates, *Mymk* is only expressed in post-metamorphic  
24 multinucleated muscles, but is absent from mononucleated larval muscles. In this study, we  
25 demonstrate that *cis*-regulatory sequence differences in the promoter region of *Mymk* underlie  
26 the different spatiotemporal patterns of its transcriptional activation in tunicates and vertebrates.  
27 While in vertebrates Myogenic Regulatory Factors (MRFs) like MyoD1 alone are required and  
28 sufficient for *Mymk* transcription in all skeletal muscles, we show that transcription of *Mymk* in  
29 post-metamorphic muscles of the tunicate *Ciona* requires the combinatorial activity of  
30 MRF/MyoD and Early B-Cell Factor (Ebf). This macroevolutionary difference appears to be  
31 encoded in *cis*, likely due to the presence of a putative Ebf binding site adjacent to predicted  
32 MRF binding sites in the *Ciona Mymk* promoter. We further discuss how *Mymk* and myoblast  
33 fusion might have been regulated in the last common ancestor of tunicates and vertebrates, for  
34 which we propose two models.

## 35 Introduction

36

37 In vertebrates, multinucleated myofibers are formed through fusion of mononucleated  
38 myoblasts. Myomaker (*Mymk*) is a transmembrane protein required for myoblast fusion and  
39 muscle multinucleation (Millay et al., 2013). In tunicates, the sister group to vertebrates (Delsuc  
40 et al., 2006; Putnam et al., 2008), *Mymk* is also required for myoblast fusion and muscle  
41 multinucleation (Zhang et al., 2022). *Mymk* from tunicate species such as *Ciona robusta* can  
42 rescue cell fusion in *Mymk* CRISPR knockout myoblasts in diverse vertebrate species,  
43 suggesting highly conserved function (Zhang et al., 2022). Phylogenomic analyses indicate that  
44 *Mymk* (previously referred to as *Tmem8c*) arose in the last common ancestor of tunicates and  
45 vertebrates through duplication of an ancestral *Tmem8* gene, and is not found in other  
46 invertebrates including cephalochordates (Zhang et al., 2022).

47

48 However, unlike mammalian *Mymk*, which is expressed in all skeletal muscles, *Ciona Mymk* is  
49 exclusively expressed in the differentiating precursors of multinucleated, post-metamorphic (i.e.  
50 juvenile/adult) muscles and not in those of mononucleated larval tail muscles (Zhang et al.,  
51 2022). Transcription of *Mymk* in mammalian skeletal myoblasts is carried out by Myogenic  
52 Regulatory Factor (MRF) family members, especially MyoD1, which function as the molecular  
53 switch for muscle specification and differentiation (Zhang et al., 2020). Most tunicates, including  
54 *Ciona*, have a biphasic life cycle transitioning from a swimming larval phase to a sessile filter-  
55 feeding adult phase (Karaïskou et al., 2015). Their larvae have muscles in their tail that are  
56 specified by the *Ciona* MRF ortholog (Meedel et al., 2007). However, unlike the post-  
57 metamorphic muscles of the adult body wall and siphons, they do not express *Mymk* and do not  
58 undergo cell fusion or multinucleation (Zhang et al., 2022). We therefore sought to understand  
59 the molecular mechanism underlying this muscle subtype- and life cycle stage-specific  
60 activation of *Mymk* and myoblast fusion in tunicates.

61

62 Here, we describe the *cis*- and *trans*-regulatory bases of *Mymk* expression specifically in the  
63 post-metamorphic muscles of *Ciona*. First, we show that, like in vertebrates, MRF is required for  
64 *Mymk* activation in *Ciona*. However, in *Ciona* the transcription factor Early B-Cell Factor (Ebf,  
65 also known as *Collier/Olf1/EBF* or *COE*) is required in combination with MRF to activate *Mymk*  
66 transcription. Mis-expressing Ebf together with MRF in the larval tail and other tissues is  
67 sufficient to activate ectopic *Mymk* transcription. We show that these effects are recapitulated  
68 even by using human MYOD1 and EBF3 in *Ciona*, while *Ciona* MRF alone is sufficient to

69 activate human *MYMK* in cultured myoblasts. Finally, we identify the likely binding sites for MRF  
70 and Ebf in the *Ciona Mymk* promoter, suggesting that differences in *cis* (promoter sequences),  
71 not in *trans* (transcription factor protein-coding sequences), are the primary drivers of  
72 evolutionary change between facultative and obligate *Mymk* expression and myoblast fusion in  
73 the chordates.

74

## 75 **Results**

76 Transcription of *Mymk* in mammalian skeletal myoblasts is carried out by bHLH transcription  
77 factors of the MRF family (Zhang et al., 2020), which comprise the major molecular switch for  
78 muscle specification and differentiation (Hernández-Hernández et al., 2017). In *Ciona*, MRF is  
79 the sole ortholog of human MRF family members MYOD1, MYF5, MYOG, and MRF4, and is  
80 necessary and sufficient for myoblast specification in the early embryo (Hernández-Hernández  
81 et al., 2017; Meedel et al., 2007; Meedel et al., 1997; Meedel et al., 2002). In tunicates, *MRF* is  
82 expressed in larval tail muscles and in the atrial and oral siphon muscles (ASMs and OSMs,  
83 respectively) of the post-metamorphic juvenile/adult (Razy-Krajka et al., 2014). Yet *Mymk* is only  
84 expressed in the ASMs/OSMs (Zhang et al., 2022), suggesting the regulation of *Mymk* in  
85 tunicates is different than that of vertebrates as MRF alone is not sufficient to activate *Mymk* in  
86 larval tail muscle (**Figure 1A,B**).

87

88 Comparing multinucleated post-metamorphic and mononucleated larval muscles, one key  
89 molecular difference between the two that we hypothesized might determine the selective  
90 regulation of *Mymk* is the expression of Ebf in the former, but not the latter (Stolfi et al., 2010)  
91 (**Figure 1A**). Ebf orthologs have been frequently associated with myogenic activity throughout  
92 animals. In models such as *Drosophila* and *Xenopus*, Ebf orthologs are upstream of or in  
93 parallel to MRF in muscle development (Dubois et al., 2007; Enriquez et al., 2012; Green and  
94 Vetter, 2011). In *Ciona*, Ebf specifies post-metamorphic muscle fate (Stolfi et al., 2010; Tolkin  
95 and Christiaen, 2016) and activates both *MRF* and ASM-specific gene expression (Razy-Krajka  
96 et al., 2014). Therefore, MRF and Ebf were the prime candidates for post-metamorphic muscle-  
97 specific activation of *Mymk*.

98

## 99 **CRISPR/Cas9-mediated disruption of *MRF* shows it is necessary for *Mymk* expression**

100 To test whether MRF is necessary for *Mymk* expression in post-metamorphic *Ciona* muscles,  
101 we targeted the *MRF* locus using tissue-specific CRISPR/Cas9-mediated mutagenesis (Stolfi et  
102 al., 2014). We specifically targeted the B7.5 lineage that gives rise to ASMs using the *Mesp*

103 promoter (Davidson et al., 2005) to drive Cas9 expression in this lineage. To target *MRF*, we  
104 used a combination of two sgRNAs (*U6>MRF.2* and *U6>MRF.3*) that had been previously  
105 designed and validated (Gandhi et al., 2017). We allowed animals to develop into  
106 metamorphosing juveniles at 46 hours post-fertilization (hpf) and scored the expression of a  
107 previously published *Mymk>GFP* reporter plasmid in *Mesp>mScarlet+* ASMs (**Figure 1C**). In  
108 *MRF* CRISPR juveniles, *Mymk>GFP* expression is nearly extinguished, as we observed only  
109 1% of mScarlet+ juveniles showing GFP expression in the *MRF* CRISPR juveniles compared to  
110 95% in the negative control condition (**Figure 1D**). These data strongly suggest that MRF is  
111 necessary for *Mymk* transcription in *Ciona* post-metamorphic muscles, just like in vertebrate  
112 skeletal muscles.

113

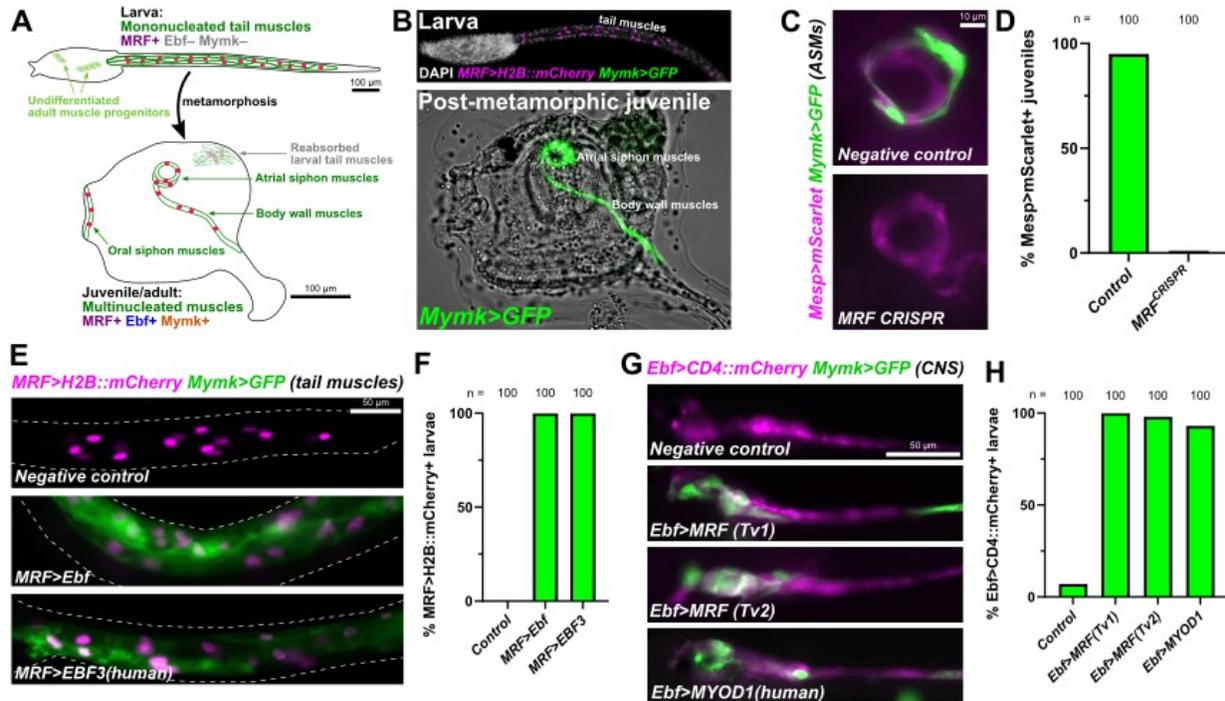
### 114 **Forced co-expression of MRF and Ebf activates ectopic *Mymk* expression**

115 We were not able to utilize the same approach to test the requirement of *Ebf* for *Mymk*  
116 expression, as *Ebf* is required for *MRF* activation in post-metamorphic muscle progenitors  
117 (Razy-Krajka et al., 2014), and CRISPR/Cas9-mediated disruption of *Ebf* results in loss of *MRF*  
118 expression, converting siphon muscles to heart cell fate (Stolfi et al., 2014). Instead, to test  
119 whether the combination of *Ebf* and MRF is sufficient to activate *Mymk* expression, we forced  
120 their combined expression in different larval cells. Outside of being expressed in the progenitors  
121 of multinucleated post-metamorphic muscles, *Ebf* is also expressed in the central nervous  
122 system of the larva, where it is important for cholinergic gene expression and motor neuron  
123 development (Kratsios et al., 2012; Popsuj and Stolfi, 2021). In contrast, *MRF* is expressed in  
124 the tail muscles of the larvae, where *Ebf* is not expressed. Therefore we sought to ectopically  
125 express *Ebf* in MRF+ tail muscles, and MRF in *Ebf*+ neural progenitors.

126

127 We first drove ectopic *Ebf* expression in the larval tail muscles using the extended *MRF*  
128 promoter (Zhang et al., 2022). Ectopic expression of EBF in larval tail muscles with *MRF>Ebf*  
129 resulted in transcription of the *Mymk* reporter in 100% of successfully transfected larvae  
130 compared to 0% in negative control larvae (**Figure 1E,F**). Similarly, we overexpressed MRF in  
131 the larval central nervous system using an *Ebf* promoter (Stolfi and Levine, 2011). In this case  
132 we tested two different transcript variants (Tv) of MRF, *MRF-Tv1* and *MRF-Tv2*, which differ in  
133 the length of the encoded C termini with slightly different functional properties (Izzi et al., 2013).  
134 Strong ectopic *Mymk* reporter expression was seen in the larval nervous system with either  
135 variant of MRF (*MRF-Tv1*: 100% *Mymk>GFP+*; *MRF-Tv2*: 98% *Mymk>GFP+*) (**Figure 1G,H**). In  
136 negative control larvae, scattered weak *Mymk>GFP* expression was visible in the nervous

137 system in only 7% of transfected larva. In contrast, overexpression of MRF in the larval tail  
 138 muscles using an *MRF>MRF-Tv1* construct resulted in *Mymk>GFP* expression in only 8% of  
 139 larvae, mostly comprising strong expression in certain neurons (presumably Ebf+) that also had  
 140 leaky *MRF* promoter activity (**Figure S1**) We conclude that increased MRF dose does not  
 141 adequately replace the function of Ebf. Taken together, our results suggest that MRF and Ebf  
 142 co-expression is sufficient for *Mymk* activation in *Ciona*.  
 143



144

145 **Figure 1. The combination of MRF and Ebf activates *Mymk* expression in *Ciona***

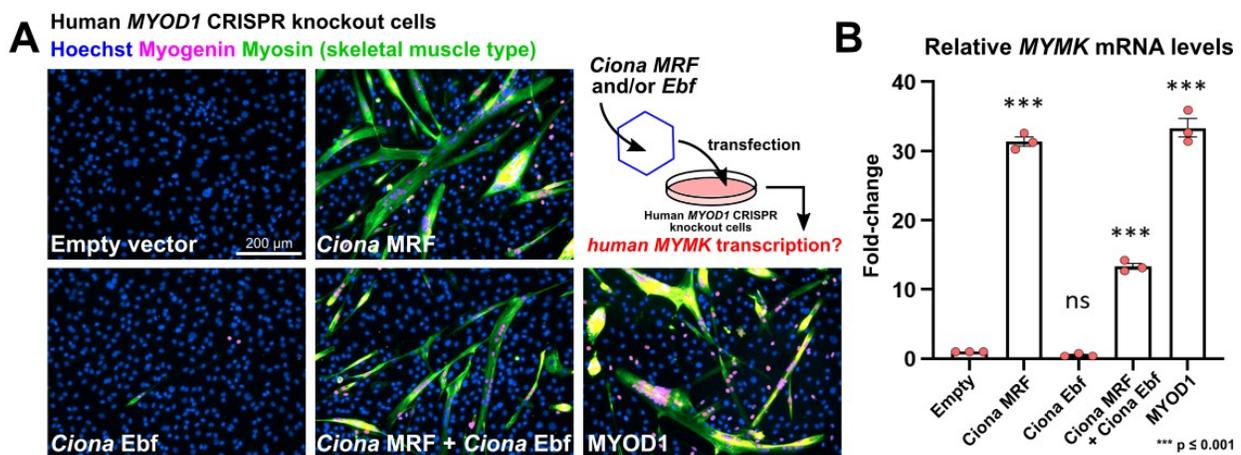
146 (A) Diagram depicting larval or post-metamorphic muscles in the biphasic lifecycle of the tunicate *Ciona robusta*. Based primarily on  
 147 Razy-Krajka et al. 2014 and Zhang et al. 2022. (B) A GFP reporter plasmid containing the entire intergenic region upstream of the  
 148 *Ciona Mymk* gene (-508/-1 immediately preceding the start codon) is visibly expressed in juvenile/adult muscles at 60 hours post-  
 149 fertilization (hpf, lower panel) but not in larval tail muscles (21 hpf, upper panel). Juvenile figure re-processed from raw images  
 150 previously published in Zhang et al. 2022. (C) B7.5 lineage-specific CRISPR/Cas9-mediated disruption (using *Mesp>Cas9*) of the  
 151 *MRF* gene results in loss of *Mymk>GFP* reporter expression ( $p < 0.0001$ ) in atrial siphon muscles (ASMs). Metamorphosing  
 152 juveniles fixed and imaged at 46 hpf. Untagged mScarlet reporter used. Negative control juveniles electroporated with *U6>Control*  
 153 sgRNA vector instead. (D) Scoring of data represented in previous panel. (E) Ectopically expressing Ebf in MRF+ larval tail muscles  
 154 (using the *MRF* promoter) results in ectopic activation of *Ciona Mymk* reporter in larvae imaged at 21 hpf. Ectopically expressing  
 155 human EBF3 in the same cells produces a comparable result. Negative control electroporated with reporter plasmids only. (F)  
 156 Scoring of data represented in previous panel ( $p < 0.0001$  for both Ebf and EBF3). (G) Using the *Ebf* promoter to ectopically express  
 157 either isoform of *Ciona MRF* (Tv1 or Tv2) or human MYOD1 in Ebf+ neural cells results in ectopic activation of *Mymk>GFP* at 16  
 158 hpf. Negative control electroporated with *Ebf>lacZ* instead. (H) Scoring of data represented in previous panel ( $p < 0.0001$  for all  
 159 experimental conditions). See text for all experimental details. See **Table S1** for all statistical test details.

160

161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173

## Human EBF3 and MYOD1 can replace their *Ciona* homologs

To test whether this MRF-Ebf cooperativity we observe is unique to the *Ciona* proteins, we replaced *Ciona* Ebf and MRF in the above experiments with their human homologs EBF3 and MYOD1. We then assayed whether they can cooperate with endogenous *Ciona* MRF or Ebf and activate the *Ciona Mymk* reporter in the larval tail muscles or central nervous system. Remarkably, both replacements resulted in strong ectopic *Mymk* reporter expression (**Figure 1E-H**) This implies that the species origin of the proteins themselves does not seem to matter as long as an MRF ortholog and an Ebf ortholog is expressed in the same cell. This suggested that the difference in obligate versus facultative *Mymk* expression is likely due to changes in *cis* (i.e. different binding sites in the *Mymk* promoter) rather than in *trans* (i.e. change to MRF or EBF family proteins) between tunicates and vertebrates.



174  
175  
176  
177  
178  
179  
180  
181  
182  
183

## Figure 2. *Ciona* MRF alone can activate transcription of *MYMK* in human cells

(A) Representative images of differentiating myoblasts in culture, under different rescue conditions following CRISPR/Cas9-mediated knockout of *MYOD1*. *Ciona* MRF (Tv2) and Ebf were compared to human *MYOD1* for their ability to activate *MYMK*, in combination or solo. Cell nuclei stained with Hoechst (blue), and muscle specification and differentiation visualized with immunostaining for Myogenin (magenta) and skeletal muscle myosin (green). Diagram of rescue experiment in top right panel, with viral vector represented as a hexagon. (B) Quantification of *MYMK* mRNA in conditions depicted in previous panel, by qPCR. Experiment performed in triplicate, with statistical significance tested by one-way ANOVA with multiple comparisons to the empty vector condition. See text for experimental details.

184  
185  
186  
187

## *Ciona* MRF alone can activate *MYMK* in human *MYOD1*-CRISPR knockout cells

Using a similar logic from the previous experiment, we introduced *Ciona* MRF (Tv2) and/or Ebf into CRISPR-generated, *MYOD1*-deficient human myoblasts to see whether they (alone or in combination) are sufficient to activate the expression of human *MYMK* (**Figure 2A**).

188 Remarkably, transfected *Ciona* MRF alone resulted in nearly identical levels of *MYMK* mRNA  
189 expression as human MYOD1 (**Figure 2B**). In contrast, when *Ciona* Ebf was expressed alone,  
190 no significant *MYMK* mRNA expression was detected, and Ebf in combination with MRF  
191 resulted in a significant reduction of MRF efficacy, appearing to hamper its activation of *MYMK*.  
192 These data further support the idea that changes in *cis*, and not in *trans*, underlie the differential  
193 requirement of Ebf in activating *Ciona Mymk*.

194

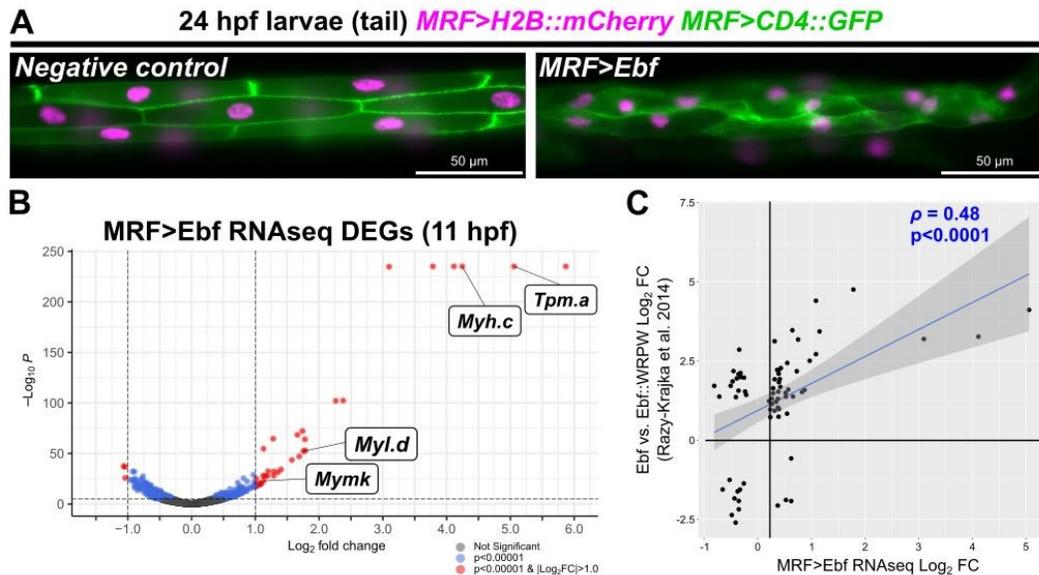
### 195 **RNAseq confirms upregulation of *Mymk* and other post-metamorphic muscle-specific** 196 **genes by combinatorial activity of MRF and Ebf**

197 Ebf has been established as an important regulator of post-metamorphic muscle fate in *Ciona*  
198 (Kaplan et al., 2015; Razy-Krajka et al., 2018; Razy-Krajka et al., 2014; Stolfi et al., 2010; Tolkin  
199 and Christiaen, 2016). When we ectopically expressed Ebf in larval tail muscle cells, we  
200 observed a striking change in their morphology (**Figure 3A**). Typical larval tail muscles are  
201 mononucleated with defined polygonal shapes and cell-cell junctions (Passamaneck et al.,  
202 2007), but Ebf expression suppressed these features. Instead, the tail muscle cells became  
203 more reminiscent of post-metamorphic siphon and body wall muscles, becoming elongated and  
204 myofiber-like, losing their characteristic polygonal shape, and gaining more and smaller-sized  
205 nuclei. Although we could not detect any clear instances of tail muscle cells fusing, these  
206 observations suggest that the combination of MRF and Ebf might be activating the expression of  
207 additional determinants of post-metamorphic muscle-specific morphogenesis.

208

209 To identify other putative targets of MRF-Ebf cooperativity, we performed bulk RNAseq to  
210 compare the transcriptomes of wildtype embryos to embryos in which Ebf was overexpressed in  
211 the tail muscles. We extracted RNA at 11 hpf from whole embryos that were either transfected  
212 with *MRF>Ebf* or, *MRF>CD4::GFP* as a negative control. Using DESeq2 to analyze the  
213 resulting RNAseq data, we observed significant upregulation of many genes in the *MRF>Ebf*  
214 condition. Of the 16,433 genes detected, 163 were significantly upregulated ( $p < 0.00001$ ,  
215  $\log_{2}FC > 0$ ), of which 33 showed a  $\log_{2}FC$  greater than 1.0 (**Figure 3B, Table S2**). In contrast, 155  
216 genes were significantly downregulated ( $p < 0.00001$ ,  $\log_{2}FC < 0$ ). Interestingly, high-ranking genes  
217 that had both a significant p-value and log fold change  $\geq 1$  include *Tropomyosin.a* (*Tpm.a*),  
218 *Col24a-related*, *myosin heavy chain.c* (*Myh.c*), and *myosin light chain.d* (*Myl.d*), which have all  
219 been confirmed as upregulated specifically in the ASMs by *in-situ* hybridization (Razy-Krajka et  
220 al., 2014). *Mymk* was #25 in this list, confirming that endogenous *Mymk* (and not just the  
221 *Mymk>GFP* reporter) is ectopically activated in tail muscles upon Ebf overexpression. This

222 suggests that Ebf is sufficient to partially convert larval tail muscle cells into post-metamorphic,  
 223 atrial siphon-like muscles.



224

### 225 **Figure 3. RNAseq analysis of MRF-Ebf targets**

226 (A) Morphological phenotype of larval tail muscles altered by overexpression of Ebf (right panel). Note loss of clearly delineated  
 227 polygonal cell shapes visualized by membrane-bound CD4::GFP (green), compared to the negative control electroporated with the  
 228 reporter constructs only. (B) Volcano plot showing differentially expressed genes (DEGs) detected by bulk RNAseq of whole  
 229 embryos at 11 hours post-fertilization (hpf), comparing Ebf overexpression (MRF>Ebf) to a negative control condition  
 230 (MRF>CD4::GFP). *Mymk* and other confirmed post-metamorphic muscle-expressed genes indicated by white boxes: *Tropomyosin.a*  
 231 (*Tpm.a*), *Myosin heavy chain.c* (*Myh.c*, also known as *MHC3*), *Myosin light chain.d* (*Myl.d*). Note the top six genes are flattened at  
 232 the limit of p-value calculation by the algorithm (see **Table S2** for details and full list of genes). (C) Plot comparing our bulk RNAseq  
 233 data to microarray analysis of DEGs between *Foxf>Ebf* and *Foxf>Ebf::WRPW* conditions in FACS-isolated cardiopharyngeal  
 234 progenitors (CPPs), published in Razy-Krajka et al. 2014 (see reference for original experimental details). Only genes with  $p < 0.05$  in  
 235 both datasets were compared. Rho ( $\rho$ ) indicates Pearson's correlation. Dark grey area indicates 95% confidence interval.

236

237 To further investigate this muscle subtype fate change, we compared our EBF overexpression  
 238 bulk RNAseq results to a published microarray analysis of Ebf overexpression in the Trunk  
 239 Ventral Cells (TVCs) that give rise to both heart and ASM progenitors (Razy-Krajka et al., 2014).  
 240 Indeed, when comparing both data sets, many of the top genes in our list were also significantly  
 241 upregulated by Ebf overexpression in the TVCs, resulting in a Pearson correlation coefficient ( $\rho$ )  
 242 of 0.48 (**Figure 3C, Table S3**). Although there are several genes that show discrepant changes  
 243 in expression between the two datasets, this may reflect differences in the timing of RNA  
 244 extraction and territory of Ebf overexpression (tail muscles at 11 hpf vs. TVCs at 21 hpf). Taken  
 245 together, these data suggest that *Mymk* is just one of several genes that might be preferentially  
 246 activated in post-metamorphic, multinucleated muscles by a similar MRF-Ebf combinatorial logic  
 247 in *Ciona*.

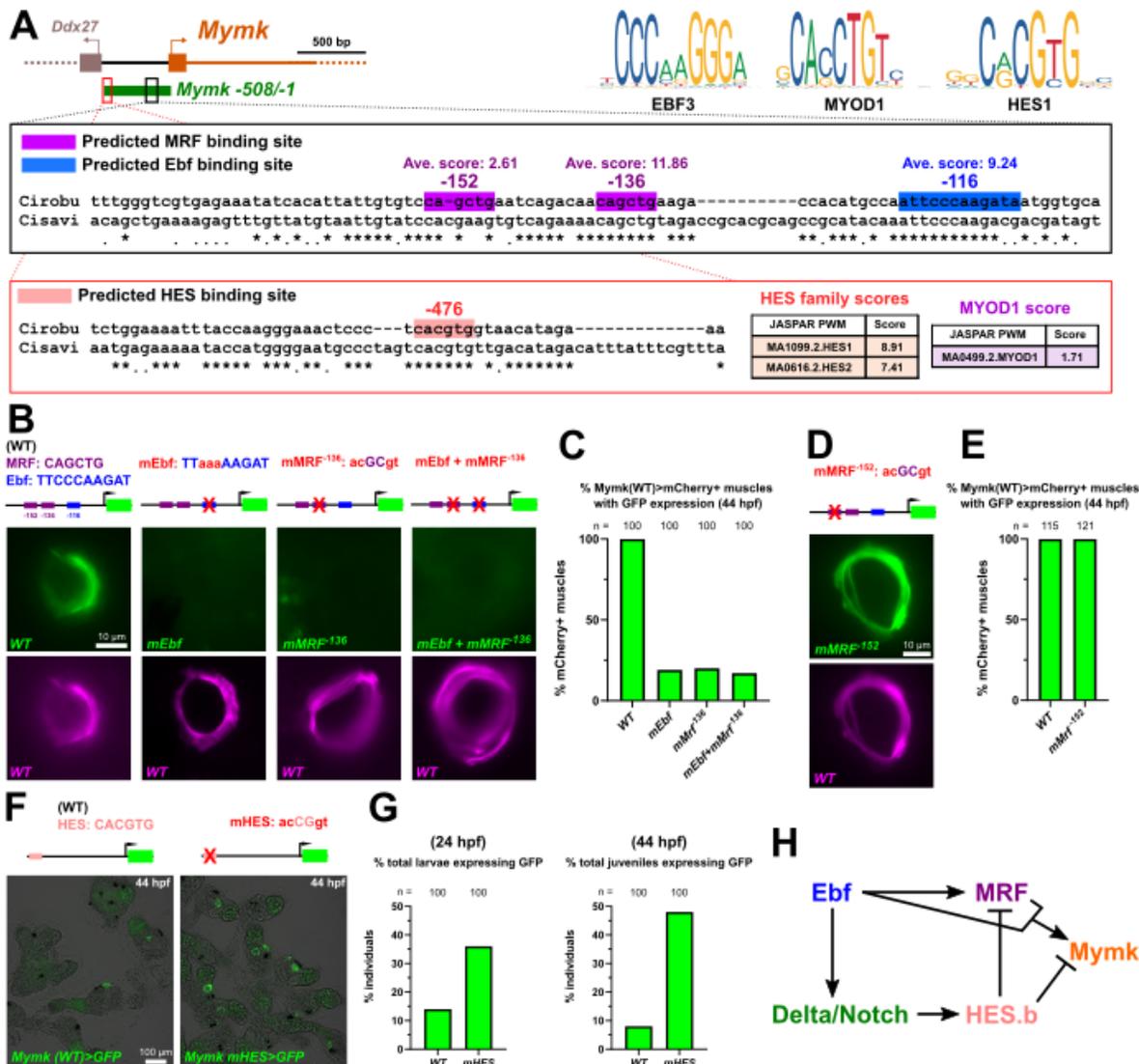
## 248 **Analyzing candidate MRF and Ebf binding sites in the *Mymk* promoter**

249 Because we suspected vertebrate-tunicate differences in *Mymk* activation to be due primarily to  
250 differences in *cis*, we aligned the *Ciona robusta Mymk* promoter to the homologous sequence  
251 from the related species *Ciona savignyi* (Satou et al., 2008; Satou et al., 2019; Satou et al.,  
252 2022; Vinson et al., 2005) to identify potentially conserved transcription factor binding sites  
253 (**Figure 4A**). We also utilized JASPAR (Castro-Mondragon et al., 2022), a predictive binding site  
254 search algorithm, to look for putative MRF and Ebf sites (Chaudhary and Skinner, 1999; Treiber  
255 et al., 2010). This led us to identifying MRF<sup>-136</sup> and Ebf<sup>-116</sup> as conserved, high-scoring candidate  
256 binding sites to test (**Figure 4A, Table S4**). We made mutations predicted to disrupt MRF or Ebf  
257 binding to these putative sites in the *Mymk>GFP* reporter plasmid, and scored the co-  
258 expression of these mutant reporters with a wild type (“WT”) *Mymk>mCherry* reporter. When  
259 observing juveniles electroporated with the *Mymk>GFP* reporter bearing the MRF<sup>-136</sup> mutation, it  
260 was clear that its activity was significantly reduced, with only 19% of *Mymk(WT)>mCherry+*  
261 siphon muscles faintly expressing GFP as well (**Figure 4B,C**). As expected, we also observed  
262 dramatic reporter expression loss with the Ebf<sup>-116</sup> mutation (20% GFP expression, although  
263 mutating both MRF<sup>-136</sup> and Ebf<sup>-116</sup> did not further abolish the residual GFP expression (**Figure**  
264 **4B,C**). In contrast, 100% of juveniles co-expressed wild type GFP and mCherry reporters.  
265 Similarly, mutation of a nearby poorly conserved, low-scoring predicted MRF site (MRF<sup>-152</sup>) did  
266 not significantly reduce reporter activity, suggesting it is not required for activation and likely not  
267 a functional MRF binding site (**Figure 4D,E**). Taken together, these data suggests that *Ciona*  
268 *Mymk* activation is dependent on closely spaced, conserved MRF and Ebf predicted binding  
269 sites in its proximal promoter region.

270

## 271 **Predicted HES binding site represses *Mymk* activation**

272 When examining the *Mymk* promoter for potential transcription factor binding sites, we noticed a  
273 conserved Ebox sequence further upstream in the *Mymk* promoter (**Figure 4A**). We initially  
274 thought could be an MRF binding site, but JASPAR predictions revealed a much higher score  
275 for binding by Hairy Enhancer of Split (HES) transcriptional repressor family members (**Figure**  
276 **4A**). In *Ciona*, HES has been shown to mediate Delta/Notch-dependent repression of *MRF*  
277 expression and myogenic differentiation in the inner ASM precursor cells, prolonging their  
278 undifferentiated, proliferative state (Razy-Krajka et al., 2014). In vertebrates, Delta/Notch  
279 signaling also represses MyoD expression and muscle differentiation (Delfini et al., 2000). In  
280 chick, HEYL (a HES homolog) binds to the *Mymk* promoter and inhibits its transcription, hinting  
281 at a deeply conserved strategy for restricting the onset of *Mymk* expression and fusion in



282

## 283 Figure 4. Mutational analysis of predicted binding sites in the *Ciona Mymk* promoter

284 (A) Diagram of *Ciona Mymk* genomic region with predicted binding sites highlighted in insets. Coordinates given as relative to *Ciona Mymk*  
285 translational start codon, as transcription start sites are generally unavailable for *Ciona* genes. Conserved basepairs indicated by  
286 asterisks under alignment between *C. robusta* and *C. savignyi* orthologous sequences. Top right: position-weight matrices (PWMs)  
287 for human orthologs of the major candidate transcription factors analyzed in this study. Bottom inset: Far upstream Ebox (-476)  
288 shows greater predicted affinity for HES-family repressors than for MYOD1/MRF activators. Predicted scores obtained from  
289 JASPAR. (B) Disruptions to predicted binding sites in the *Mymk>GFP* reporter results in significant loss of activity in post-  
290 metamorphic siphon muscles, imaged at 44 hours post-fertilization (hpf). All GFP reporters co-electroporated with wild type  
291 *Mymk>mCherry* reporter. (C) Scoring of data represented in previous panel ( $p < 0.0001$  for all). (D) Disrupting the low-scoring, non-  
292 conserved MRF<sup>-152</sup> site does not significantly reduce reporter expression ( $p > 0.9999$ ), as quantified in (E). (F) Mutating the  
293 predicted HES site at position -476 results in higher frequency of reporter expression. (G) Scoring of data represented in previous  
294 panel and similarly electroporated larvae at 24 hpf. Total individuals were assayed for GFP reporter expression. Normally, only ~5-  
295 15% of all individuals show *Mymk>GFP* expression, likely due to mosaic uptake/retention of electroporated plasmids. Mutating the  
296 HES site boosts this to ~30-45% ( $p < 0.0005$  at 24 hpf,  $p < 0.0001$  at 44 hpf). (H) Gene regulatory network diagram showing  
297 proposed regulation of *Ciona Mymk* by Ebf, MRF, and Notch-dependent HES. Regulatory connections between Ebf, MRF,  
298 Delta/Notch, and HES.b based on Razy-Krajka et al. 2014. See text for experimental details and **Table S1** for statistical test details.

299 developing myoblasts (Esteves de Lima et al., 2022). When we tested a *Mymk* GFP reporter  
300 plasmid carrying a mutation to disrupt this upstream Ebox, we observed a significant increase in  
301 frequency of GFP expression compared to the wild type reporter (**Figure 4F,G**). Increased GFP  
302 expression suggests that this site is most likely bound by a repressor. Our results suggest that  
303 the direct repression of *Mymk* transcription by HES repressors (**Figure 4H**) may have been an  
304 ancestral trait present in the last common ancestor of tunicates and vertebrates.

305

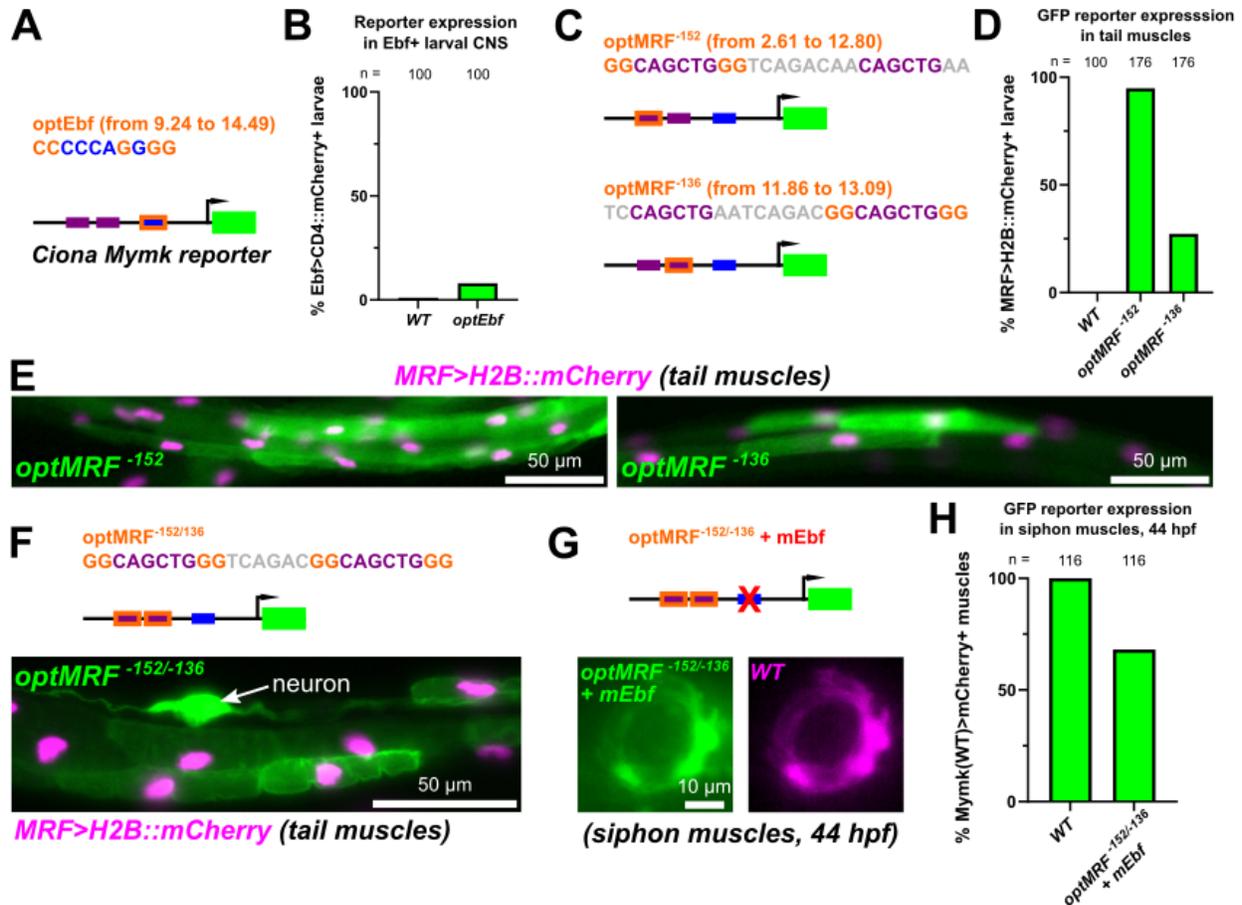
### 306 **Adding an additional high-quality MRF binding site abolishes the need for MRF-Ebf** 307 **cooperativity**

308 What might be the exact *cis*-regulatory change that result in the difference observed between  
309 tunicate and vertebrate *Mymk* regulation? Unfortunately, reporter constructs made using  
310 published human *MYMK* (Zhang et al., 2020) or chicken *Mymk* (Luo et al., 2015) promoters  
311 were not expressed at all in *Ciona* tail muscles (**Figure S2**). This was not entirely surprising,  
312 given that orthologous promoters are frequently incompatible (i.e. “unintelligible”) even between  
313 different tunicate species due to developmental system drift (Lowe and Stolfi, 2018). We  
314 therefore focused instead on testing different point mutations in the *Ciona Mymk* promoter that  
315 might result in ectopic activation in larval tail muscles.

316

317 *Cis*-regulatory logic can be complex with subtle changes in promoter sequences resulting in  
318 drastically different activation patterns (Spitz and Furlong, 2012). In tunicates, it has been  
319 shown that by making changes to the sequences flanking a given transcription factor binding  
320 site one can increase its predicted binding affinity, resulting in higher expression levels or  
321 ectopic activation (Farley et al., 2015; Farley et al., 2016; Jindal et al., 2023). To test whether  
322 such “optimized” MRF and Ebf binding sites might result in activation of the *Ciona Mymk*  
323 reporter by MRF or Ebf alone (without the need for MRF-Ebf cooperation), we manipulated  
324 flanking sequences of putative MRF or Ebf binding sites, resulting in higher binding affinity  
325 scores predicted by JASPAR (**Figure 5A,C, Figure S3**).

326



327

328 **Figure 5. Altered regulatory logic unlocks *Mymk* reporter expression in larval muscles**

329 (A) Diagram indicating basepair changes to “optimize” the putative Ebf binding site (optEbf) in the *Ciona Myomaker* promoter by  
 330 increasing its predicted JASPAR score. (B) Optimizing the putative Ebf site (optEbf) resulted in a small but statistically significant (p  
 331 < 0.0349) effect on activating a *Myomaker* reporter in the absence of MRF in Ebf+ central nervous system (CNS) cells. Larvae fixed  
 332 at 22 hpf. (C) Diagram showing the optimization of either the predicted MRF<sup>-152</sup> or MRF<sup>-136</sup> sites. (D) Scoring of ectopic reporter  
 333 expression in larval tail muscles with the optimized putative MRF sites (optMRF), assayed at 17 hpf (p < 0.0001 for both). More  
 334 frequent ectopic expression was observed with optMRF<sup>-152</sup> than optMRF<sup>-136</sup>. (E) Representative images of larval tail muscles  
 335 assayed in the previous panel. (F) Optimization of both putative MRF<sup>-152</sup> and MRF<sup>-136</sup> sites in combination resulted in similar ectopic  
 336 expression in tail muscles, but also in ectopic expression in neurons in 44% of larvae assayed at 22 hpf. (G) Combining optimized  
 337 MRF<sup>-152</sup> and MRF<sup>-136</sup> sites together partially rescues reporter expression even with the putative Ebf binding site disrupted (mEbf). (H)  
 338 Scoring of siphon muscle expression depicted in the previous panel. See text for experimental details, and **Table S1** for statistical  
 339 test details.

340

341 Optimization of the conserved Ebf<sup>-116</sup> site did not significantly increase *Mymk>GFP* activation in  
 342 the central nervous system (**Figure 5A,B**). This suggested that either the Ebf site is already  
 343 “optimal”, or that Ebf binding affinity is not rate-limiting in this context. However, optimization of  
 344 the conserved, indispensable MRF<sup>-137</sup> site and/or the non-conserved, dispensable MRF<sup>-152</sup> site  
 345 resulted in significant *Mymk>GFP* expression in tail muscles (**Figure 5C-F**). Interestingly,  
 346 optimization of MRF<sup>-152</sup> resulted in visible GFP expression in 95% of electroporated larval tails,

347 while optimization of MRF<sup>-136</sup> resulted in GFP expression in a more modest 27% of tails (**Figure**  
348 **5D**). Because the increase in average predicted JASPAR score was most pronounced between  
349 the wild type MRF<sup>-152</sup> (JASPAR score 2.61) and its “optimized” counterpart (JASPAR score  
350 12.8, **Figure S3**), this suggested that creating an additional high-scoring MRF binding site is  
351 particularly effective for switching a combinatorial MRF-Ebf transcriptional logic to an MRF-  
352 alone one. This switch in logic was confirmed when we observed *Mymk* reporter expression in  
353 post-metamorphic muscles even when combining optimized MRF sites with a disrupted Ebf site  
354 (**Figure 5G,H**). Interestingly, combining both optimized MRF<sup>-152</sup> and MRF<sup>-136</sup> sites resulted in  
355 ectopic reporter activation in neurons, in addition to tail muscles, in 44% of larvae (**Figure 5F**).  
356 This expression might be due to greater affinity for proneural transcription factors that also bind  
357 Ebox sequences, such as Neurogenin (Kim et al., 2020). This suggests that the exact  
358 sequences flanking each site might also be under purifying selection, minimizing ectopic  
359 activation of *Mymk* in tissues where its expression might be detrimental.

360

## 361 **Discussion**

362 In this study, we have investigated the *cis*-regulatory logic of muscle subtype-specific *Mymk*  
363 expression in *Ciona*. We have identified two essential transcriptional regulators, MRF and Ebf,  
364 that together activate the transcription of *Mymk*, which encodes a transmembrane protein that  
365 drives myoblast fusion and muscle multinucleation in tunicates and vertebrates (Zhang et al.,  
366 2022). This is in stark contrast to human *MYMK* expression, which only requires the activity of  
367 MRFs (Zhang et al., 2020). We have also revealed a potentially conserved repressive input into  
368 *Ciona Mymk* transcription, in which direct binding and repression by HES factors might restrict  
369 the spatiotemporal window of *Mymk* expression and, consequently, of myoblast fusion. This  
370 repression, likely mediated through Delta-Notch signaling, might predate the divergence of  
371 tunicates and vertebrates.

372

373 We propose that the difference between “MRF+Ebf” and “MRF-alone” logic is responsible for  
374 the difference between the pan-skeletal muscle expression of vertebrate *Mymk* and the more  
375 selective, post-metamorphic muscle-specific expression in *Ciona* (**Figure 6A**). This in turn might  
376 underlie the difference between obligate (vertebrate) versus facultative (tunicate) muscle  
377 multinucleation. Although *Mymk* overexpression is not sufficient to drive the fusion of *Ciona*  
378 larval tail muscle cells, *Ciona Mymk* is sufficient to induce human myoblast fusion (Zhang et al.,  
379 2022). Our RNAseq results show that this same MRF-Ebf logic is regulating a larger suite of

380 post-metamorphic muscle-specific genes, some of which might encode additional factors  
381 required for myoblast fusion.

382

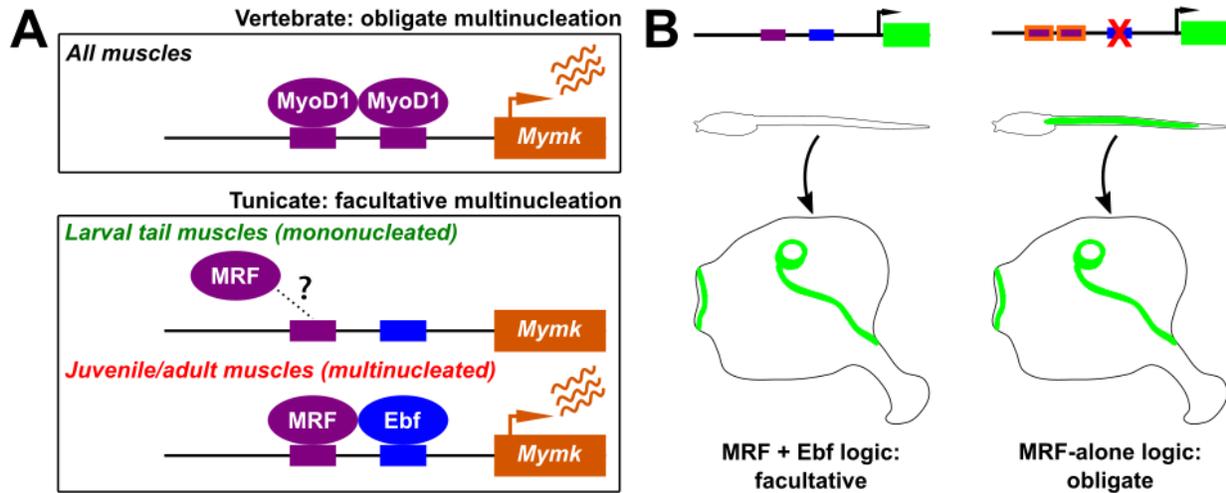
383 Although we have largely revealed the basis of *Mymk* regulation in *Ciona*, there are still details  
384 that have yet to be elucidated. For instance, what is the mechanism of MRF-Ebf cooperative  
385 activity? MRF-Ebf synergy in transcription of muscle subtype-specific genes has been reported  
386 in mammals, for instance the activation of the *Atp2a1* gene by MyoD1 and Ebf3 specifically in  
387 mouse diaphragm muscles (Jin et al., 2014). Myod1 and Ebf3 and its homologs alone have the  
388 ability to activate *Atp2a1*, but this expression was substantially higher when both MyoD1 and  
389 Ebf3 were present. However, there was no evidence of the two transcription factors directly  
390 contacting one another to drive cooperative binding. This may be similar to the mechanism of  
391 activation of *Mymk* by MRF and Ebf in *Ciona*. Optimization of the predicted Ebf site in the *Mymk*  
392 promoter did not significantly increase reporter plasmid expression in Ebf+ larval neurons. This  
393 suggests that MRF and Ebf might act on different steps of *Mymk* activation in *Ciona*. In other  
394 words, the MRF-Ebf combinatorial logic we have revealed might not be dependent on  
395 cooperative binding, as is the case for other examples of cooperativity (Zeitlinger, 2020).

396

397 Given the difference between vertebrates (obligate) and tunicate (facultative) activation of  
398 *Mymk*, we present two hypotheses for how regulation of *Mymk* might have been controlled in  
399 the last common ancestor of tunicates and vertebrates. In the first scenario, the ancestor would  
400 have been more like tunicates, in which the combination of MRF and Ebf would have  
401 cooperatively activated *Mymk* only in a subset of muscles. In the second scenario, the ancestor  
402 would have been more like vertebrates, in which MRF alone would have activated *Mymk* in all  
403 muscles.

404

405 In the first scenario, the last common ancestor would have had both mononucleated and  
406 multinucleated muscles, like we see in most tunicates. It is unclear if the common ancestor had  
407 a biphasic life cycle or not, but it is likely they had separate lineages for the trunk and  
408 pharyngeal muscles as seen in both vertebrates and tunicates (Razy-Krajka et al., 2014). The  
409 ancestor may have had specialized pharyngeal muscles homologous to the siphon muscles of  
410 tunicates. One key feature of tunicate siphon muscles is that they are formed by a series of



411

412 **Figure 6. Proposed models for *Mymk* regulation and myoblast fusion in chordates**

413 (A) Proposed regulatory models for transcriptional activation of *Mymk* in vertebrates compared to *Ciona* (tunicates). Question mark  
 414 and dashed lines indicate uncertainty whether MRF can still bind to the *Mymk* promoter in tunicate larval tail muscles, or whether the  
 415 co-requirement for Ebf acts on other steps independently of MRF binding. (B) Summary diagram showing the switch from a  
 416 combinatorial MRF + Ebf logic to MRF-alone logic for *Ciona Mymk* regulation, obtained in this study through optimization of putative  
 417 MRF binding sites, with or without disrupting the putative Ebf site.

418

419 concentric circular myotubes. It is possible that the ancestor had a similar set of circular  
 420 muscles around the openings of a pharyngeal atrium, and the process of *Mymk*-driven myoblast  
 421 fusion might have evolved to allow for the formation of such muscles. After splitting from  
 422 tunicates, vertebrates would have lost the requirement of Ebf for *Mymk* expression, and MRF  
 423 would have become the sole activator of *Mymk*, allowing all muscle cells to become  
 424 multinucleated. This may have been advantageous for their survival, perhaps permitting larger  
 425 myofibers throughout the body and advanced movement capabilities.

426

427 Alternatively, the last common ancestor might have only had multinucleated muscles under the  
 428 regulation of MRF alone, as in extant vertebrates. Later, vertebrates would have kept this mode  
 429 of regulation, while tunicates would have recruited Ebf to activate *Mymk* only in post-  
 430 metamorphic muscles, as an adaptation specifically tied to their biphasic life cycle. As it stands,  
 431 we do not have enough evidence to conclusively favor one evolutionary scenario over the other.  
 432 On the vertebrate side, there are no reports of muscle subtype-specific fusion as far as we can  
 433 tell. On the tunicate side, with the exception of groups that have generally lost the larval phase  
 434 (e.g. salps and pyrosomes), there are no reports of obligate myoblast fusion. However, we have  
 435 shown that a switch to pan-muscle expression of *Mymk* is possible through “optimization” of  
 436 putative MRF binding sites in its promoter, or by creating an additional high-scoring predicted

437 MRF site (**Figure 6B**). Whether this is actually a recapitulation of what happened in evolution or  
438 not, we may never know.

439

## 440 **Materials and Methods**

441

### 442 ***Ciona* handling, electroporation, fixing, staining, imaging and scoring**

443 *Ciona robusta* (*intestinalis* Type A) specimens were obtained and shipped from San Diego,  
444 California, USA (M-REP). The eggs were fertilized, dechorionated, and subjected to  
445 electroporation using established methods as detailed in published protocols (Christiaen et al.,  
446 2009a, b). The embryos were then raised at a temperature of 20°C. At various stages, including  
447 embryos, larvae, and juveniles, the specimens were fixed using MEM-FA solution (composed of  
448 3.7% formaldehyde, 0.1 M MOPS at pH 7.4, 0.5 M NaCl, 1 mM EGTA, 2 mM MgSO<sub>4</sub>, and 0.1%  
449 Triton-X100), followed by rinsing in 1X PBS with 0.4% Triton-X100 and 50 mM NH<sub>4</sub>Cl to quench  
450 autofluorescence, and one final wash in 1X PBS with 0.1% Triton-X100.

451

452 Imaging of the specimens was carried out using either a Leica DMI8 or DMIL LED inverted  
453 epifluorescence microscope. Scoring was carried out only on mCherry+ individuals as to  
454 exclude potentially unelectroporated animals, unless otherwise noted in the figure legends. To  
455 carry out CRISPR/Cas9-mediated mutagenesis of *MRF* in the B7.5 lineage we used  
456 *Mesp>Cas9* to restrict Cas9 expression to this lineage (Stolfi et al., 2014), together with  
457 previously validated *MRF*-targeting sgRNA plasmids *U6>MRF.2* and *U6>MRF.3* (Gandhi et al.,  
458 2017). For the negative control, previously published *U6>Control* sgRNA vector was used,  
459 which expresses an sgRNA that is predicted to not target any sequence in the *C. robusta*  
460 genome (Stolfi et al., 2014). The sgRNAs are expressed *in vivo* from plasmids using the  
461 ubiquitous RNA polymerase III-transcribed U6 small RNA promoter (Nishiyama and Fujiwara,  
462 2008). Mutations to disrupt or optimize putative binding sites were all generated through *de*  
463 *novo* synthesis and custom cloning by Twist Bioscience. All GFP or mCherry sequences fused  
464 to the N-terminal Unc-76 extranuclear localization tag (Dynes and Ngai, 1998), unless otherwise  
465 specified. All plasmid and sgRNA sequences can be found in the **Supplemental Sequences**  
466 **File**. All statistical tests summarized in **Table S1**.

467

### 468 **Ectopic expression of MRF orthologs and *Ciona* Ebf in human *MYOD1*-knockout cells**

469 Human *MYOD1*-knockout myoblasts were generated by CRISPR-Cas9 mediated gene editing  
470 and cultured as described previously (Zhang et al., 2020). Retroviral expression vector pMXs-

471 Puro (Cell Biolabs, RTV-012) was used for cloning and the expression of the human *MYOD1*,  
472 *Ciona MRF* (Transcript Variant 2), and *Ciona Ebf*. The DNA sequences were verified by Sanger  
473 sequencing. For the myogenic rescue experiments, the sgRNA-insensitive version of human  
474 MYOD1 open reading frame was used. Retrovirus was produced through transfection of HEK293  
475 cells using FuGENE 6 (Promega, E2692). Two days after transfection, virus medium was  
476 collected, filtered and used to infect human myoblasts assisted by polybrene (Sigma-Aldrich, TR-  
477 1003-G). When the culture reached 80-90% confluency, cells were induced for myogenic  
478 differentiation by switching to myoblast differentiation medium (2% horse serum in DMEM with  
479 1% penicillin/streptomycin). Human myoblasts were differentiated for three days and used for  
480 immunostaining and RNA extraction. For immunostaining, the primary antibody for Myosin  
481 (Developmental Studies Hybridoma Bank, MF20) and the primary antibody for Myogenin  
482 (Developmental Studies Hybridoma Bank, F5D) were used. The qPCR primers for measurements  
483 of human *MYMK* and 18S expression are provided in the **Supplemental Sequences File**.

484

#### 485 **RNA sequencing and analysis**

486 Total RNA was extracted at 11 hours post-fertilization (Stage 23, late tailbud) from two  
487 independent replicates each of electroporated larvae that were transfected either with 50 g  
488 *MRF>CD4::GFP* (Negative control) or 50 g *MRF>Ebf transcript variant 1* (Ebf overexpression).  
489 Library preparation was at the Georgia Tech Molecular Evolution Core Facility as previously  
490 described (Johnson et al., 2023), and sequenced on the Illumina NovaSeq 6000 with an SP  
491 PE100bp run. Reads were processed and differential gene expression analysis was performed  
492 using DESeq2 in Galaxy as previously described (Johnson et al., 2023). KY21 gene model ID  
493 numbers (Satou et al., 2022) were matched to KH gene model ID numbers (Satou et al., 2008)  
494 using the Ciona Gene Model Converter application:

495 [https://github.com/katarzynampiekarz/ciona\\_gene\\_model\\_converter](https://github.com/katarzynampiekarz/ciona_gene_model_converter) (Johnson et al., 2023). Our

496 RNAseq analysis was also compared to published microarray analysis of Ebf perturbations in  
497 FACS-isolated cardiopharyngeal lineage cells (Razy-Krajka et al., 2014). Volcano plots and  
498 comparative transcriptome plots were constructed using R studio and Bioconductor (Huber et  
499 al., 2015) with packages EnhancedVolcano (Blighe et al., 2018) and ggplot2 (Wickham, 2016).  
500 Raw sequencing reads are archived under NCBI BioProject accession number PRJNA1068599.

501

#### 502 **Acknowledgments**

503 The authors are indebted to Dr. Florian Razy-Krajka for thoughtful discussion on Ebf function in  
504 specifying post-metamorphic muscle identity in *Ciona*. We thank Lindsey Cohen for technical

505 assistance and Shweta Biliya for help with RNAseq library preparation and sequencing. We  
506 thank Brian Hammer, Annalise Paaby, and Will Ratcliff for critical reading and helpful  
507 suggestions. This work funded by NIH grant GM143326 to AS, NIH grant GM147209 to PB, and  
508 an NSF graduate fellowship to CJJ.

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

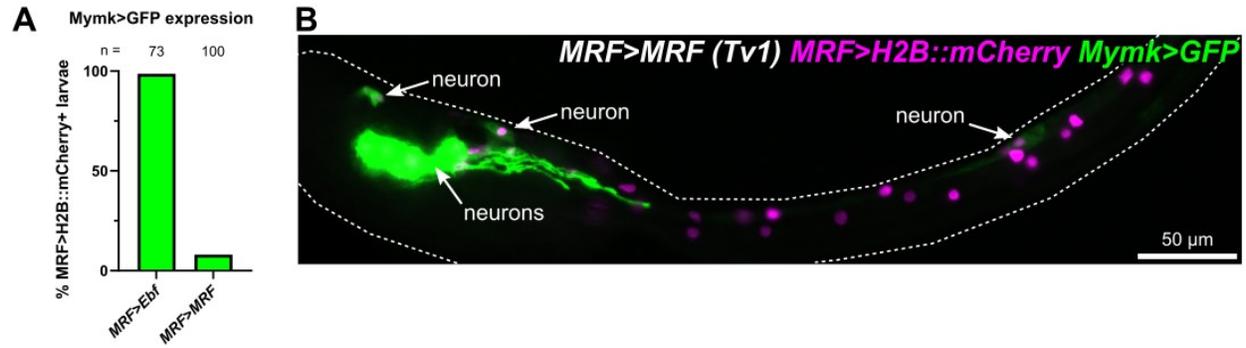
528

529

530

531

532



533

534 **Figure S1. Increased MRF dose in larval tail muscles does not adequately replace Ebf**

535 (A) Scoring of larvae at 21 hours post-fertilization (hpf) comparing effect of Ebf or MRF

536 overexpression on ectopic *Mymk>GFP* expression in larval tail muscles ( $p < 0.0001$ ). (B) Most

537 of the effect of *MRF>MRF* on activating ectopic *Mymk>GFP* was limited to the nervous system,

538 likely due to leaky activity of the *MRF* promoter in Ebf+ neurons. See **Table S1** for statistical test

539 details.

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

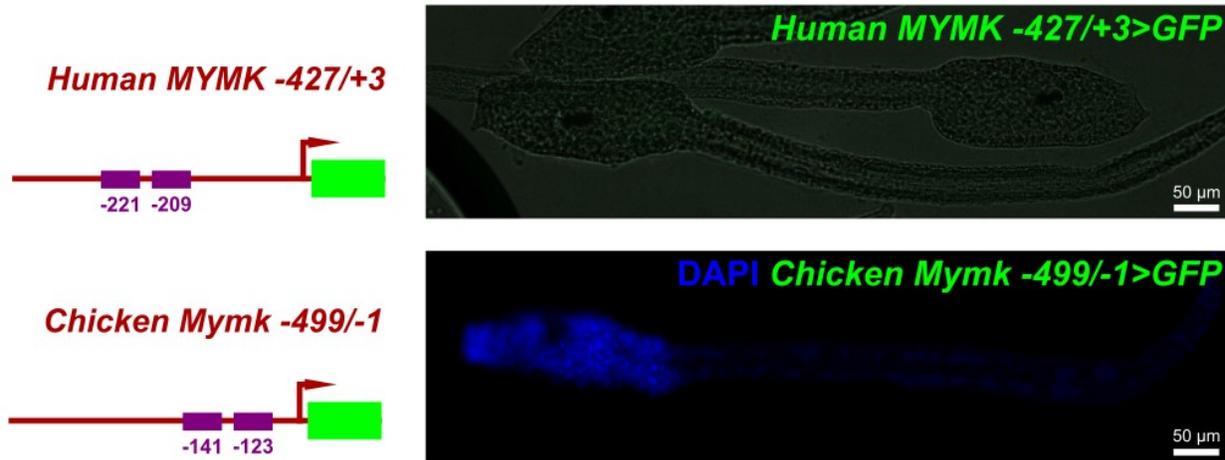
556

557

558

559

560



561

562 **Figure S2. Human MYMK and chicken Mymk reporter plasmids are not active in *Ciona***

563 Left: diagrams describing human MYMK and chicken Mymk reporter plasmids using previously  
564 published sequences (Luo et al., 2015; Zhang et al., 2020) and their predicted MRF family  
565 member binding sites (purple boxes). Size/spacing of sites is not to exact scale. See  
566 **Supplemental Sequences File** for detailed sequences. Right: images of larvae at 17 hours  
567 post-fertilization (hpf) electroporated with the reporter plasmids at left, showing no expression in  
568 larval tail muscles.

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

## MRF binding site optimization

| wild type sites |                     |                     |
|-----------------|---------------------|---------------------|
| JASPAR PWM      | MRF <sup>-152</sup> | MRF <sup>-136</sup> |
| MA0499.1.Myod1  | -0.37               | 13.07               |
| MA0499.2.MYOD1  | 2.79                | 9.57                |
| MA0500.1.Myog   | -0.47               | 13.65               |
| MA0500.2.MYOG   | 8.47                | 11.15               |
| Average         | 2.61                | 11.86               |

| "optimized" sites |                     |                     |
|-------------------|---------------------|---------------------|
| JASPAR PWM        | MRF <sup>-152</sup> | MRF <sup>-136</sup> |
| MA0499.1.Myod1    | 13.56               | 13.86               |
| MA0499.2.MYOD1    | 10.50               | 10.25               |
| MA0500.1.Myog     | 13.59               | 14.24               |
| MA0500.2.MYOG     | 13.56               | 13.99               |
| Average           | 12.80               | 13.09               |

TCCAGCTGAATCAGACAACAGCTGAA  
 -152 -136

GGCAGCTGGGTCAGACGGCAGCTGGG  
 -152 -136

## Ebf binding site optimization

| wild type site |                    |
|----------------|--------------------|
| JASPAR PWM     | Ebf <sup>116</sup> |
| MA0154.3.EBF1  | 11.49              |
| MA1604.1.Ebf2  | 8.32               |
| MA1637.1.EBF3  | 7.90               |
| Average        | 9.24               |

| "optimized" site |                    |
|------------------|--------------------|
| JASPAR PWM       | Ebf <sup>116</sup> |
| MA0154.3.EBF1    | 15.03              |
| MA1604.1.Ebf2    | 14.16              |
| MA1637.1.EBF3    | 14.29              |
| Average          | 14.49              |

TTCCCAAGAT

CCCCCAGGGG

586

587 **Figure S3. JASPAR scores before and after "optimization" of putative bindings sites**

588 Predicted JASPAR scores for affinity of MRF (top) or Ebf (bottom) to their respective putative  
 589 binding sites, before and after point mutations to "optimize" them, or rather increase their  
 590 predicted JASPAR scores. Predictions based on individual human transcription factor PWMs  
 591 and their averages shown.

592

593

594

595

596

597

598

599

600

601

602

603 **Table S1. Scoring data for *Ciona* experiments and statistical test details**

604

605 **Table S2. DESeq2 analysis of differential gene expression of Ebf overexpression in**  
606 **developing larval tail muscles measured by Illumina bulk RNAseq**

607

608 **Table S3. Comparison of genes significantly up- or down-regulated in the RNAseq**  
609 **analysis in the current study and the microarray study of Razy-Krajka et al. 2014**

610

611 **Table S4. Predicted JASPAR affinity scores for putative MRF and Ebf sites in the *Ciona***  
612 ***Mymk* promoter using various human ortholog position weight matrices**

613

614 **Supplemental Sequences File. All relevant DNA and protein sequences (reporters,**  
615 **perturbation constructs, primers, etc.) used in this study, including *Ciona* electroporation**  
616 **mix recipes.**

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637 **References**

- 638 Blighe, K., Rana, S., Lewis, M., 2018. EnhancedVolcano: Publication-Ready Volcano Plots with Enhanced  
639 Colouring and Labeling.
- 640 Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Berhanu Lemma, R., Turchi, L., Blanc-  
641 Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., 2022. JASPAR 2022: the 9th release of  
642 the open-access database of transcription factor binding profiles. *Nucleic acids research* 50, D165-D173.
- 643 Chaudhary, J., Skinner, M.K., 1999. Basic helix-loop-helix proteins can act at the E-box within the serum  
644 response element of the c-fos promoter to influence hormone-induced promoter activation in Sertoli  
645 cells. *Molecular endocrinology* 13, 774-786.
- 646 Christiaen, L., Wagner, E., Shi, W., Levine, M., 2009a. Electroporation of transgenic DNAs in the sea  
647 squirt *Ciona*. *Cold Spring Harbor Protocols* 2009, pdb. prot5345.
- 648 Christiaen, L., Wagner, E., Shi, W., Levine, M., 2009b. Isolation of sea squirt (*Ciona*) gametes,  
649 fertilization, dechoriation, and development. *Cold Spring Harbor Protocols* 2009, pdb. prot5344.
- 650 Davidson, B., Shi, W., Levine, M., 2005. Uncoupling heart cell specification and migration in the simple  
651 chordate *Ciona intestinalis*. *Development* 132, 4811-4818.
- 652 Delfini, M.-C., Hirsinger, E., Pourquié, O., Duprez, D., 2000. Delta 1-activated notch inhibits muscle  
653 differentiation without affecting Myf5 and Pax3 expression in chick limb myogenesis. *Development* 127,  
654 5213-5224.
- 655 Delsuc, F., Brinkmann, H., Chourrout, D., Philippe, H., 2006. Tunicates and not cephalochordates are the  
656 closest living relatives of vertebrates. *Nature* 439, 965-968.
- 657 Dubois, L., Enriquez, J., Daburon, V., Crozet, F., Lebreton, G., Crozatier, M., Vincent, A., 2007. Collier  
658 transcription in a single *Drosophila* muscle lineage: the combinatorial control of muscle identity.  
659 *Development* 134, 4347-4355.
- 660 Dynes, J.L., Ngai, J., 1998. Pathfinding of olfactory neuron axons to stereotyped glomerular targets  
661 revealed by dynamic imaging in living zebrafish embryos. *Neuron* 20, 1081-1091.
- 662 Enriquez, J., de Taffin, M., Crozatier, M., Vincent, A., Dubois, L., 2012. Combinatorial coding of  
663 *Drosophila* muscle shape by Collier and Nautilus. *Developmental biology* 363, 27-39.
- 664 Esteves de Lima, J., Blavet, C., Bonnin, M.-A., Hirsinger, E., Havis, E., Relaix, F., Duprez, D., 2022.  
665 TMEM8C-mediated fusion is regionalized and regulated by NOTCH signalling during foetal myogenesis.  
666 *Development* 149, dev199928.
- 667 Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., Levine, M.S., 2015. Suboptimization of  
668 developmental enhancers. *Science* 350, 325-328.
- 669 Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S., Levine, M.S., 2016. Syntax compensates for poor  
670 binding sites to encode tissue specificity of developmental enhancers. *Proceedings of the National*  
671 *Academy of Sciences* 113, 6508-6513.
- 672 Gandhi, S., Haeussler, M., Razy-Krajka, F., Christiaen, L., Stolfi, A., 2017. Evaluation and rational design of  
673 guide RNAs for efficient CRISPR/Cas9-mediated mutagenesis in *Ciona*. *Developmental biology* 425, 8-20.
- 674 Green, Y.S., Vetter, M.L., 2011. EBF proteins participate in transcriptional regulation of *Xenopus* muscle  
675 development. *Developmental biology* 358, 240-250.
- 676 Hernández-Hernández, J.M., García-González, E.G., Brun, C.E., Rudnicki, M.A., 2017. The myogenic  
677 regulatory factors, determinants of muscle development, cell identity and regeneration. *72*, 10-18.
- 678 Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto,  
679 L., Girke, T., Gottardo, R., Hahne, F., Hansen, K.D., Irizarry, R.A., Lawrence, M., Love, M.I., MacDonald, J.,  
680 Obenchain, V., Oleś, A.K., Pagès, H., Reyes, A., Shannon, P., Smyth, G.K., Tenenbaum, D., Waldron, L.,  
681 Morgan, M., 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*  
682 12, 115-121.

- 683 Izzi, S.A., Colantuono, B.J., Sullivan, K., Khare, P., Meedel, T.H., 2013. Functional studies of the *Ciona*  
684 *intestinalis* myogenic regulatory factor reveal conserved features of chordate myogenesis.  
685 *Developmental biology* 376, 213-223.
- 686 Jin, S., Kim, J., Willert, T., Klein-Rodewald, T., Garcia-Dominguez, M., Mosqueira, M., Fink, R., Esposito, I.,  
687 Hofbauer, L.C., Charnay, P., 2014. Ebf factors and MyoD cooperate to regulate muscle relaxation via  
688 Atp2a1. *Nature communications* 5.
- 689 Jindal, G.A., Bantle, A.T., Solvason, J.J., Grudzien, J.L., D'Antonio-Chronowska, A., Lim, F., Le, S.H., Song,  
690 B.P., Ragsac, M.F., Klie, A., 2023. Single-nucleotide variants within heart enhancers increase binding  
691 affinity and disrupt heart development. *Developmental Cell* 58, 2206-2216.
- 692 Johnson, C.J., Razy-Krajka, F., Zeng, F., Piekarczyk, K.M., Biliya, S., Rothbacher, U., Stolfi, A., 2023.  
693 Specification of distinct cell types in a sensory-adhesive organ for metamorphosis in the *Ciona* larva.  
694 bioRxiv.
- 695 Kaplan, N., Razy-Krajka, F., Christiaen, L., 2015. Regulation and evolution of cardiopharyngeal cell  
696 identity and behavior: insights from simple chordates. *Current opinion in genetics & development* 32,  
697 119-128.
- 698 Karaiskou, A., Swalla, B.J., Sasakura, Y., Chambon, J.P., 2015. Metamorphosis in solitary ascidians.  
699 *Genesis* 53, 34-47.
- 700 Kim, K., Gibboney, S., Razy-Krajka, F., Lowe, E., Wang, W., Stolfi, A., 2020. Regulation of neurogenesis by  
701 FGF signaling and Neurogenin in the invertebrate chordate *Ciona*. *Frontiers in Cell and Developmental*  
702 *Biology* 8, 477.
- 703 Kratsios, P., Stolfi, A., Levine, M., Hobert, O., 2012. Coordinated regulation of cholinergic motor neuron  
704 traits through a conserved terminal selector gene. *Nature neuroscience* 15, 205.
- 705 Lowe, E.K., Stolfi, A., 2018. Developmental system drift in motor ganglion patterning between distantly  
706 related tunicates. *EvoDevo* 9, 18.
- 707 Luo, W., Li, E., Nie, Q., Zhang, X., 2015. Myomaker, regulated by MYOD, MYOG and miR-140-3p,  
708 promotes chicken myoblast fusion. *International journal of molecular sciences* 16, 26186-26201.
- 709 Meedel, T.H., Chang, P., Yasuo, H., 2007. Muscle development in *Ciona intestinalis* requires the b-HLH  
710 myogenic regulatory factor gene *Ci-MRF*. *Developmental biology* 302, 333-344.
- 711 Meedel, T.H., Farmer, S.C., Lee, J.J., 1997. The single MyoD family gene of *Ciona intestinalis* encodes two  
712 differentially expressed proteins: implications for the evolution of chordate muscle gene regulation.  
713 *Development* 124, 1711-1721.
- 714 Meedel, T.H., Lee, J.J., Whittaker, J., 2002. Muscle Development and Lineage-Specific Expression of  
715 *CiMDF*, the MyoD-Family Gene of *Ciona intestinalis*. *Developmental biology* 241, 238-246.
- 716 Millay, D.P., O'Rourke, J.R., Sutherland, L.B., Bezprozvannaya, S., Shelton, J.M., Bassel-Duby, R., Olson,  
717 E.N., 2013. Myomaker is a membrane activator of myoblast fusion and muscle formation. *Nature* 499,  
718 301-305.
- 719 Nishiyama, A., Fujiwara, S., 2008. RNA interference by expressing short hairpin RNA in the *Ciona*  
720 *intestinalis* embryo. *Development, growth & differentiation* 50, 521-529.
- 721 Passamanek, Y.J., Hadjantonakis, A.-K., Di Gregorio, A., 2007. Dynamic and polarized muscle cell  
722 behaviors accompany tail morphogenesis in the ascidian *Ciona intestinalis*. *PLoS One* 2, e714.
- 723 Popsuj, S., Stolfi, A., 2021. Ebf Activates Expression of a Cholinergic Locus in a Multipolar Motor Ganglion  
724 Interneuron Subtype in *Ciona*. *Frontiers in Neuroscience* 15, 784649.
- 725 Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M.,  
726 Shoguchi, E., Terry, A., Yu, J.-K., 2008. The amphioxus genome and the evolution of the chordate  
727 karyotype. *Nature* 453, 1064-1071.
- 728 Razy-Krajka, F., Gravez, B., Kaplan, N., Racioppi, C., Wang, W., Christiaen, L., 2018. An FGF-driven feed-  
729 forward circuit patterns the cardiopharyngeal mesoderm in space and time. *eLife* 7, e29656.

730 Razy-Krajka, F., Lam, K., Wang, W., Stolfi, A., Joly, M., Bonneau, R., Christiaen, L., 2014. Collier/OLF/EBF-  
731 dependent transcriptional dynamics control pharyngeal muscle specification from primed  
732 cardiopharyngeal progenitors. *Developmental cell* 29, 263-276.

733 Satou, Y., Mineta, K., Ogasawara, M., Sasakura, Y., Shoguchi, E., Ueno, K., Yamada, L., Matsumoto, J.,  
734 Wasserscheid, J., Dewar, K., Wiley, G.B., Macmil, S.L., Roe, B.A., Zeller, R.W., Hastings, K.E.M., Lemaire,  
735 P., Lindquist, E., Endo, T., Hotta, K., Inaba, K., 2008. Improved genome assembly and evidence-based  
736 global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon  
737 populations. *Genome Biology* 9, R152.

738 Satou, Y., Nakamura, R., Yu, D., Yoshida, R., Hamada, M., Fujie, M., Hisata, K., Takeda, H., Satoh, N.,  
739 2019. A nearly complete genome of *Ciona intestinalis* Type A (*C. robusta*) reveals the contribution of  
740 inversion to chromosomal evolution in the genus *Ciona*. *Genome biology and evolution* 11, 3144-3157.

741 Satou, Y., Tokuoka, M., Oda-Ishii, I., Tokuhiro, S., Ishida, T., Liu, B., Iwamura, Y., 2022. A manually  
742 curated gene model set for an ascidian, *Ciona robusta* (*Ciona intestinalis* type A). *Zoological Science* 39.

743 Spitz, F., Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control.  
744 *Nature reviews genetics* 13, 613-626.

745 Stolfi, A., Gainous, T.B., Young, J.J., Mori, A., Levine, M., Christiaen, L., 2010. Early chordate origins of the  
746 vertebrate second heart field. *Science* 329, 565.

747 Stolfi, A., Gandhi, S., Salek, F., Christiaen, L., 2014. Tissue-specific genome editing in *Ciona* embryos by  
748 CRISPR/Cas9. *Development* 141, 4115-4120.

749 Stolfi, A., Levine, M., 2011. Neuronal subtype specification in the spinal cord of a protovertebrate.  
750 *Development* 138, 995-1004.

751 Tolkin, T., Christiaen, L., 2016. Rewiring of an ancestral *Tbx1/10-Ebf-Mrf* network for pharyngeal muscle  
752 specification in distinct embryonic lineages. *Development* 143, 3852-3862.

753 Treiber, N., Treiber, T., Zocher, G., Grosschedl, R., 2010. Structure of an *Ebf1*: DNA complex reveals  
754 unusual DNA recognition and structural homology with *Rel* proteins. *Genes & development* 24, 2270-  
755 2275.

756 Vinson, J.P., Jaffe, D.B., O'Neill, K., Karlsson, E.K., Stange-Thomann, N., Anderson, S., Mesirov, J.P., Satoh,  
757 N., Satou, Y., Nusbaum, C., 2005. Assembly of polymorphic genomes: algorithms and application to  
758 *Ciona savignyi*. *Genome research* 15, 1127-1135.

759 Wickham, H., 2016. *ggplot2: elegant graphics for data analysis*. Springer-Verlag, New York.

760 Zeitlinger, J., 2020. Seven myths of how transcription factors read the cis-regulatory code. *Current*  
761 *Opinion in Systems Biology* 23, 22-31.

762 Zhang, H., Shang, R., Kim, K., Zheng, W., Johnson, C.J., Sun, L., Niu, X., Liu, L., Zhou, J., Liu, L., 2022.  
763 Evolution of a chordate-specific mechanism for myoblast fusion. *Science Advances* 8, eadd2696.

764 Zhang, H., Wen, J., Bigot, A., Chen, J., Shang, R., Mouly, V., Bi, P., 2020. Human myotube formation is  
765 determined by *MyoD-Myomixer/Myomaker* axis. *Science Advances* 6, eabc4062.

766