

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

Christopher J. Johnson¹, Zheng Zhang^{2,3}, Haifeng Zhang³, Renjie Shang^{2,3}, Katarzyna M. Piekarz¹, Pengpeng Bi^{2,3}, Alberto Stolfi¹

1. School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA
2. Department of Genetics, University of Georgia, Athens, GA, USA
3. Center for Molecular Medicine, University of Georgia, Athens, GA, USA.

Abstract

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

Introduction

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

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

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

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

Results

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

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

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

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

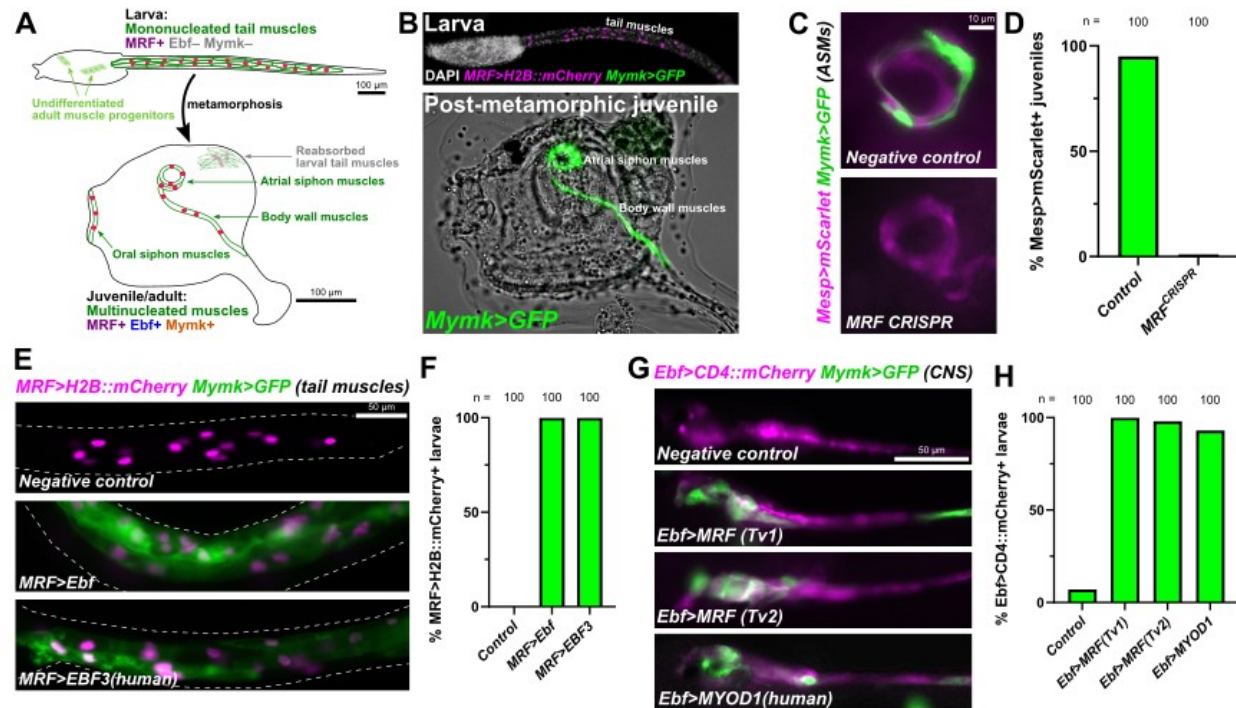
promoter (Davidson et al., 2005) to drive Cas9 expression in this lineage. To target *MRF*, we used a combination of two sgRNAs (*U6>MRF.2* and *U6>MRF.3*) that had been previously designed and validated (Gandhi et al., 2017). We allowed animals to develop into metamorphosing juveniles at 46 hours post-fertilization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

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

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

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

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



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.

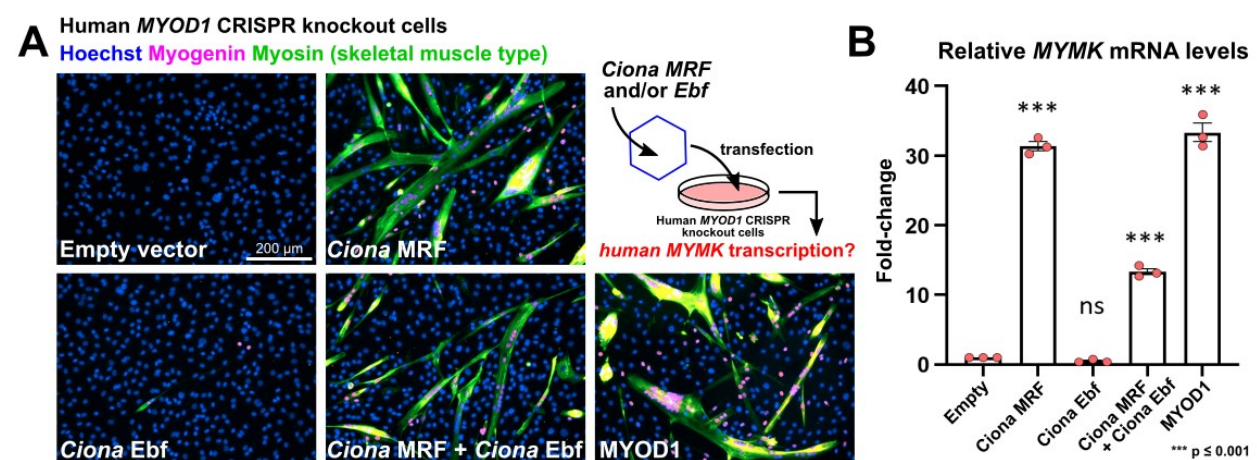


Figure 2. *Ciona* MRF alone can activate transcription of *MYMK* in human cells

(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.

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**).

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

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

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

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

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

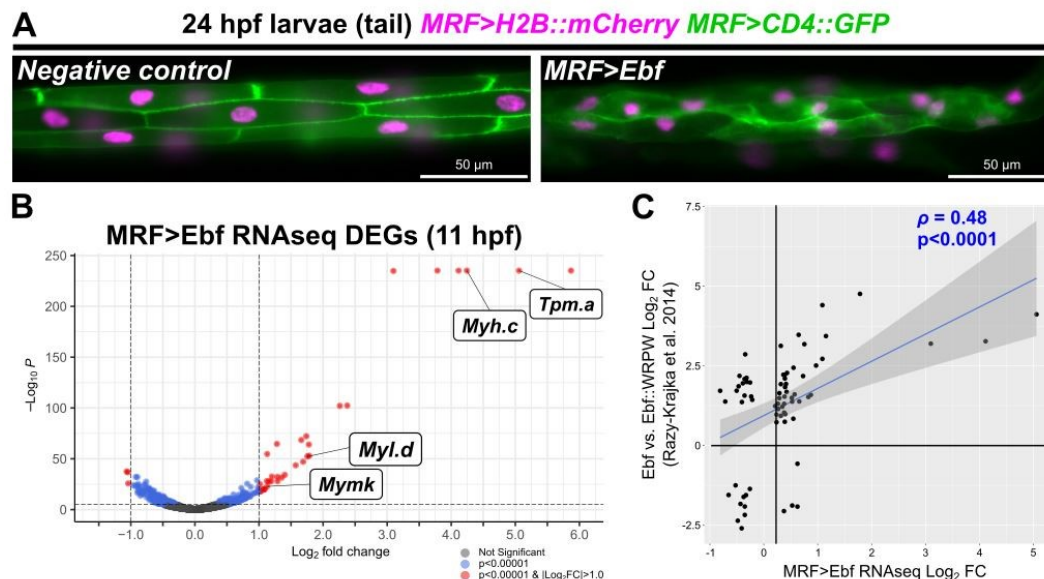


Figure 3. RNAseq analysis of MRF-Ebf targets

(A) Morphological phenotype of larval tail muscles altered by overexpression of Ebf (right panel). Note loss of clearly delineated polygonal cell shapes visualized by membrane-bound CD4::GFP (green), compared to the negative control electroporated with the reporter constructs only. (B) Volcano plot showing differentially expressed genes (DEGs) detected by bulk RNAseq of whole embryos at 11 hours post-fertilization (hpf), comparing Ebf overexpression (MRF>Ebf) to a negative control condition (MRF>CD4::GFP). *Mymk* and other confirmed post-metamorphic muscle-expressed genes indicated by white boxes: *Tropomyosin.a* (*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 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 data to microarray analysis of DEGs between Foxf>Ebf and Foxf>Ebf:WRPW conditions in FACS-isolated cardiopharyngeal progenitors (CPPs), published in Razy-Krajka et al. 2014 (see reference for original experimental details). Only genes with $p < 0.05$ in both datasets were compared. Rho (ρ) indicates Pearson's correlation. Dark grey area indicates 95% confidence interval.

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

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

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

Predicted HES binding site represses *Mymk* activation

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

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

Adding an additional high-quality MRF binding site abolishes the need for MRF-Ebf cooperativity

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

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

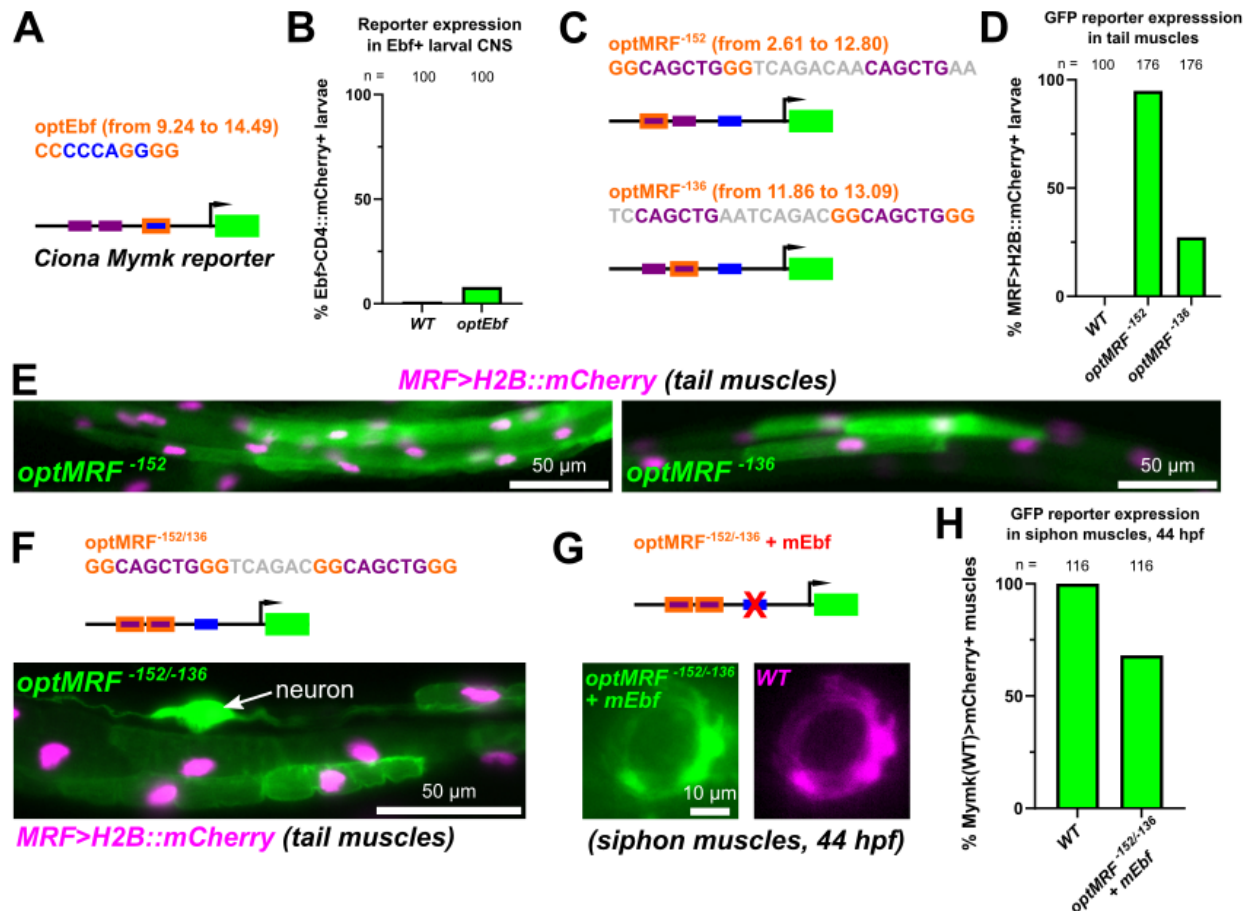


Figure 5. Altered regulatory logic unlocks *Mymk* reporter expression in larval muscles

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

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

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

Discussion

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

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

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

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

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

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

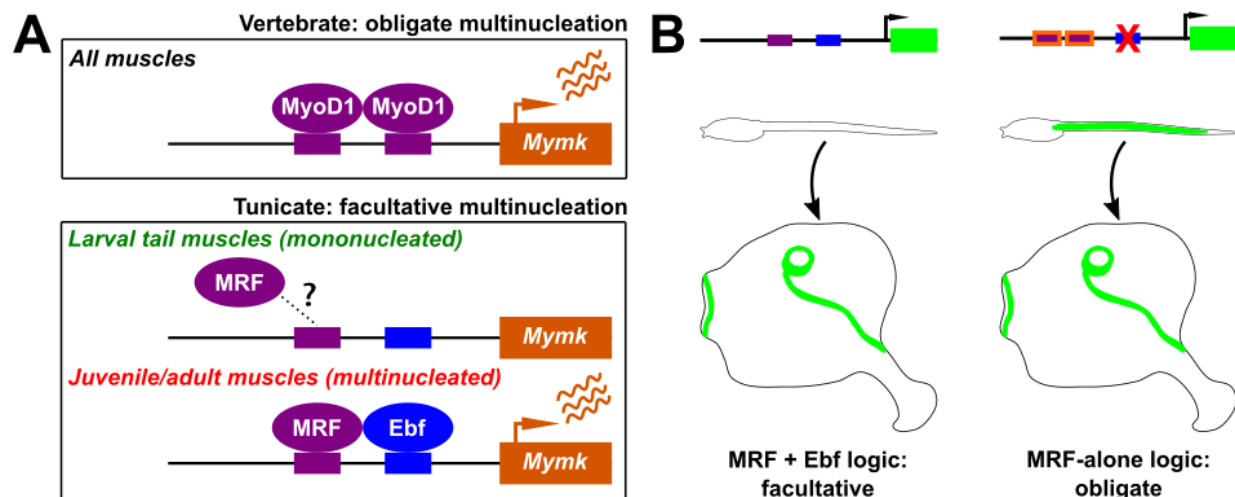


Figure 6. Proposed models for *Mymk* regulation and myoblast fusion in chordates

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

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

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

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

Materials and Methods

Ciona handling, electroporation, fixing, staining, imaging and scoring

Ciona robusta (intestinalis Type A) specimens were obtained and shipped from San Diego, California, USA (M-REP). The eggs were fertilized, dechorionated, and subjected to electroporation using established methods as detailed in published protocols (Christiaen et al., 2009a, b). The embryos were then raised at a temperature of 20°C. At various stages, including embryos, larvae, and juveniles, the specimens were fixed using MEM-FA solution (composed of 3.7% formaldehyde, 0.1 M MOPS at pH 7.4, 0.5 M NaCl, 1 mM EGTA, 2 mM MgSO₄, and 0.1% Triton-X100), followed by rinsing in 1X PBS with 0.4% Triton-X100 and 50 mM NH₄Cl to quench autofluorescence, and one final wash in 1X PBS with 0.1% Triton-X100.

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

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

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

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

RNA sequencing and analysis

Total RNA was extracted at 11 hours post-fertilization (Stage 23, late tailbud) from two independent replicates each of electroporated larvae that were transfected either with 50 g *MRF>CD4::GFP* (Negative control) or 50 g *MRF>Ebf transcript variant 1* (Ebf overexpression). Library preparation was at the Georgia Tech Molecular Evolution Core Facility as previously described (Johnson et al., 2023), and sequenced on the Illumina NovaSeq 6000 with an SP PE100bp run. Reads were processed and differential gene expression analysis was performed using DESeq2 in Galaxy as previously described (Johnson et al., 2023). KY21 gene model ID numbers (Satou et al., 2022) were matched to KH gene model ID numbers (Satou et al., 2008) using the Ciona Gene Model Converter application: https://github.com/katarzynampiekarz/ciona_gene_model_converter (Johnson et al., 2023). Our RNAseq analysis was also compared to published microarray analysis of Ebf perturbations in FACS-isolated cardiopharyngeal lineage cells (Razy-Krajka et al., 2014). Volcano plots and comparative transcriptome plots were constructed using R studio and Bioconductor (Huber et al., 2015) with packages EnhancedVolcano (Blighe et al., 2018) and ggplot2 (Wickham, 2016). Raw sequencing reads are archived under NCBI BioProject accession number PRJNA1068599.

Acknowledgments

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

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

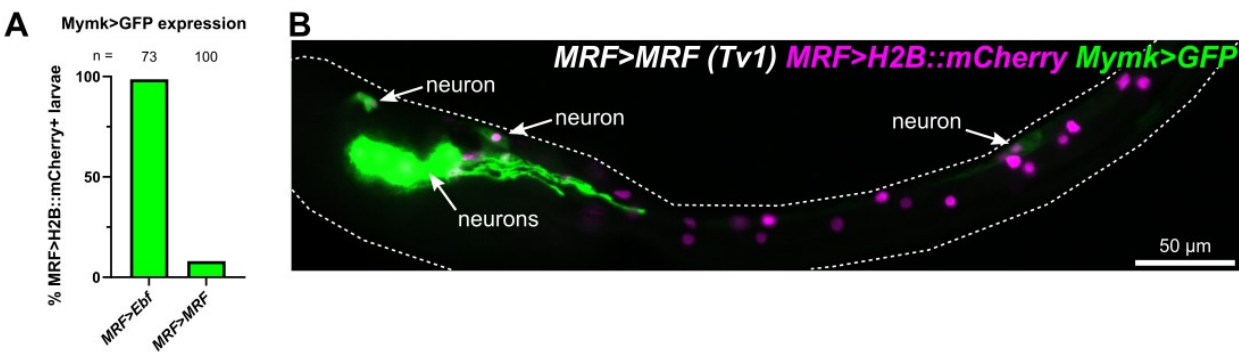


Figure S1. Increased MRF dose in larval tail muscles does not adequately replace Ebf
(A) Scoring of larvae at 21 hours post-fertilization (hpf) comparing effect of Ebf or MRF overexpression on ectopic *Mymk>GFP* expression in larval tail muscles ($p < 0.0001$). (B) Most of the effect of *MRF>MRF* on activating ectopic *Mymk>GFP* was limited to the nervous system, likely due to leaky activity of the *MRF* promoter in Ebf+ neurons. See **Table S1** for statistical test details.

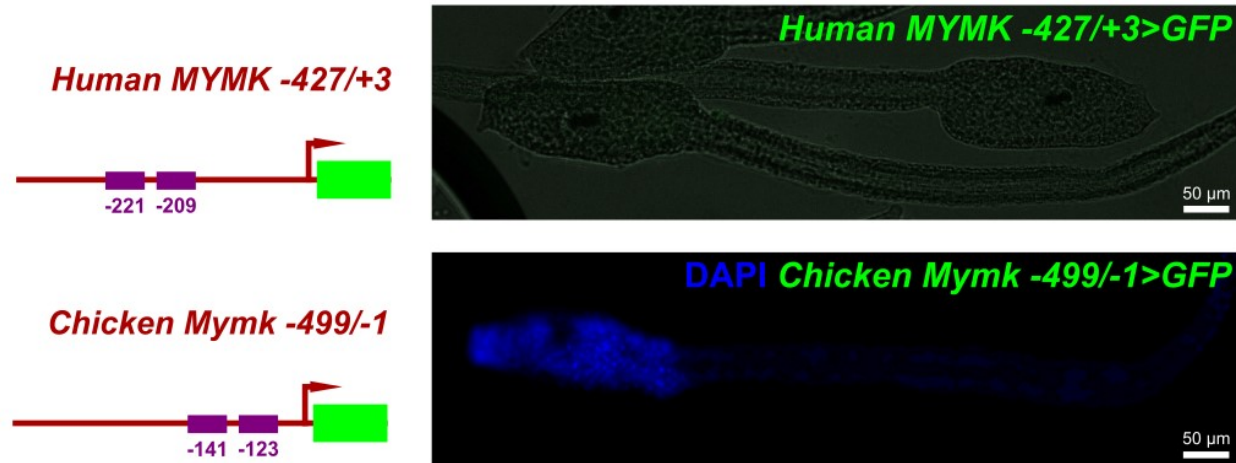


Figure S2. Human *MYMK* and chicken *Mymk* reporter plasmids are not active in *Ciona*

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

MRF binding site optimization

wild type sites			"optimized" sites		
JASPAR PWM	MRF ⁻¹⁵²	MRF ⁻¹³⁶	JASPAR PWM	MRF ⁻¹⁵²	MRF ⁻¹³⁶
MA0499.1.Myod1	-0.37	13.07	MA0499.1.Myod1	13.56	13.86
MA0499.2.MYOD1	2.79	9.57	MA0499.2.MYOD1	10.50	10.25
MA0500.1.Myog	-0.47	13.65	MA0500.1.Myog	13.59	14.24
MA0500.2.MYOG	8.47	11.15	MA0500.2.MYOG	13.56	13.99
Average	2.61	11.86	Average	12.80	13.09

TCCAGCTGAATCAGACAACAGCTGAA

-152-136

GGCAGCTGGGTCAGACGGCAGCTGGG

-152-136

Ebf binding site optimization

wild type site		"optimized" site	
JASPAR PWM	Ebf ¹¹⁶	JASPAR PWM	Ebf ¹¹⁶
MA0154.3.EBF1	11.49	MA0154.3.EBF1	15.03
MA1604.1.Ebf2	8.32	MA1604.1.Ebf2	14.16
MA1637.1.EBF3	7.90	MA1637.1.EBF3	14.29
Average	9.24	Average	14.49

TTCCCAAGAT

CCCCCAGGGG

Figure S3. JASPAR scores before and after “optimization” of putative bindings sites
Predicted JASPAR scores for affinity of MRF (top) or Ebf (bottom) to their respective putative binding sites, before and after point mutations to “optimize” them, or rather increase their predicted JASPAR scores. Predictions based on individual human transcription factor PWMs and their averages shown.

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

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

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

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

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

References

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

Izzi, S.A., Colantuono, B.J., Sullivan, K., Khare, P., Meedel, T.H., 2013. Functional studies of the *Ciona intestinalis* myogenic regulatory factor reveal conserved features of chordate myogenesis. *Developmental biology* 376, 213-223.

Jin, S., Kim, J., Willert, T., Klein-Rodewald, T., Garcia-Dominguez, M., Mosqueira, M., Fink, R., Esposito, I., Hofbauer, L.C., Charnay, P., 2014. Ebf factors and MyoD cooperate to regulate muscle relaxation via Atp2a1. *Nature communications* 5.

Jindal, G.A., Bantle, A.T., Solvason, J.J., Grudzien, J.L., D'Antonio-Chronowska, A., Lim, F., Le, S.H., Song, B.P., Ragsac, M.F., Klie, A., 2023. Single-nucleotide variants within heart enhancers increase binding affinity and disrupt heart development. *Developmental Cell* 58, 2206-2216.

Johnson, C.J., Razy-Krajka, F., Zeng, F., Piekarz, K.M., Biliya, S., Rothbacher, U., Stolfi, A., 2023. Specification of distinct cell types in a sensory-adhesive organ for metamorphosis in the *Ciona* larva. *bioRxiv*.

Kaplan, N., Razy-Krajka, F., Christiaen, L., 2015. Regulation and evolution of cardiopharyngeal cell identity and behavior: insights from simple chordates. *Current opinion in genetics & development* 32, 119-128.

Karaiskou, A., Swalla, B.J., Sasakura, Y., Chambon, J.P., 2015. Metamorphosis in solitary ascidians. *Genesis* 53, 34-47.

Kim, K., Gibboney, S., Razy-Krajka, F., Lowe, E., Wang, W., Stolfi, A., 2020. Regulation of neurogenesis by FGF signaling and Neurogenin in the invertebrate chordate *Ciona*. *Frontiers in Cell and Developmental Biology* 8, 477.

Kratsios, P., Stolfi, A., Levine, M., Hobert, O., 2012. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nature neuroscience* 15, 205.

Lowe, E.K., Stolfi, A., 2018. Developmental system drift in motor ganglion patterning between distantly related tunicates. *EvoDevo* 9, 18.

Luo, W., Li, E., Nie, Q., Zhang, X., 2015. Myomaker, regulated by MYOD, MYOG and miR-140-3p, promotes chicken myoblast fusion. *International journal of molecular sciences* 16, 26186-26201.

Meedel, T.H., Chang, P., Yasuo, H., 2007. Muscle development in *Ciona intestinalis* requires the b-HLH myogenic regulatory factor gene *Ci-MRF*. *Developmental biology* 302, 333-344.

Meedel, T.H., Farmer, S.C., Lee, J.J., 1997. The single MyoD family gene of *Ciona intestinalis* encodes two differentially expressed proteins: implications for the evolution of chordate muscle gene regulation. *Development* 124, 1711-1721.

Meedel, T.H., Lee, J.J., Whittaker, J., 2002. Muscle Development and Lineage-Specific Expression of *CiMDF*, the MyoD-Family Gene of *Ciona intestinalis*. *Developmental biology* 241, 238-246.

Millay, D.P., O'Rourke, J.R., Sutherland, L.B., Bezprozvannaya, S., Shelton, J.M., Bassel-Duby, R., Olson, E.N., 2013. Myomaker is a membrane activator of myoblast fusion and muscle formation. *Nature* 499, 301-305.

Nishiyama, A., Fujiwara, S., 2008. RNA interference by expressing short hairpin RNA in the *Ciona intestinalis* embryo. *Development, growth & differentiation* 50, 521-529.

Passamaneck, Y.J., Hadjantonakis, A.-K., Di Gregorio, A., 2007. Dynamic and polarized muscle cell behaviors accompany tail morphogenesis in the ascidian *Ciona intestinalis*. *PLoS One* 2, e714.

Popsuj, S., Stolfi, A., 2021. Ebf Activates Expression of a Cholinergic Locus in a Multipolar Motor Ganglion Interneuron Subtype in *Ciona*. *Frontiers in Neuroscience* 15, 784649.

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453, 1064-1071.

Razy-Krajka, F., Gravez, B., Kaplan, N., Racioppi, C., Wang, W., Christiaen, L., 2018. An FGF-driven feed-forward circuit patterns the cardiopharyngeal mesoderm in space and time. *eLife* 7, e29656.

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

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 Biology* 9, R152.

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

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

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

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

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

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

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

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

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

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

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

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

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