Temporal regulation of the muscle gene cascade by Macho1 and Tbx6 transcription factors in Ciona intestinalis.

Jamie E. Kugler¹, Stefan Gazdoiu¹, Izumi Oda-Ishii¹, Yale J. Passamaneck¹, Albert J. Erives² and Anna Di Gregorio¹,*

¹Department of Cell and Developmental Biology
Weill Medical College of Cornell University
1300 York Avenue, Box 60
New York, NY 10065

²Department of Biology
Dartmouth College
Hanover, NH 03755

*Corresponding Author. E-mail: and2015@med.cornell.edu, Fax: 212-746-5596
Running title: Muscle gene regulation in Ciona.
Abbreviations: bp: base pair(s); bHLH: basic helix-loop-helix; cDNA: complementary DNA; CNS: Central Nervous System; eGFP: enhanced green fluorescent protein; EMSA: electrophoretic mobility shift assay(s); EST: Expressed Sequence Tag; IPTG: Isopropyl β-D-1-thiogalactopyranoside; kb: kilobase(s), or 1000 base pairs; ORF: open reading frame; PCR: Polymerase Chain Reaction; RT: reverse transcriptase; UTR: untranslated; wt: wild-type; X-Gal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside.
Key words: Ciona intestinalis, ascidian, Brachyury, Tbx6, collagen, Macho, muscle, enhancer, T-box, transcription.
SUMMARY

For over a century, muscle formation in the ascidian embryo has been representative of “mosaic” development. The molecular basis of muscle fate predetermination has been partly elucidated with the discovery of Macho-1, a maternal zinc-finger transcription factor necessary and sufficient for primary muscle development, and of its transcriptional intermediaries Tbx6b and Tbx6c. However, the molecular mechanisms by which the maternal information is decoded by cis-regulatory modules (CRMs) associated with muscle transcription factor and structural genes, and the ways by which a seamless transition from maternal to zygotic transcription is ensured, are still mostly unclear. By combining misexpression assays with CRM analyses, we have begun to uncover the mechanisms through which Ciona Macho1 (Ci-Macho1) initiates expression of Ci-Tbx6b and Ci-Tbx6c, and we have unveiled the cross-regulatory interactions between these transcription factors. Knowledge acquired from the analysis of the Ci-Tbx6b CRM facilitated both the identification of a related CRM in the Ci-Tbx6c locus and the characterization of two CRMs associated with the structural muscle gene fibrillar collagen 1 (Ci-FColl). We use these representative examples to reconstruct how compact CRMs orchestrate the muscle developmental program from pre-localized ooplastic determinants to differentiated larval muscle in ascidian embryos.
INTRODUCTION

Ascidians, or tunicates, provide an essential genetic and evolutionary reference for studies of chordate development, due to their well-characterized modes of development, their experimental amenability, their phylogenetic proximity to vertebrates and the organization of their larval body plan (Davidson and Christiaen, 2006; Delsuc et al., 2006; Passamaneck and Di Gregorio, 2005). Ascidian larvae are characterized by the presence of an axial notochord flanked by two rows of muscle cells, classified as primary (B-lineage) and secondary (A- and b-lineage; Meedel et al., 1987), and by a rudimentary dorsal nervous system (Satoh, 1994). As shown by classical embryological studies, the development of the primary larval muscle cells in solitary ascidians proceeds cell-autonomously from the first cleavages (Meedel et al., 1987; Nishida, 1992; Ortolani, 1955). Around the neurula stage, muscle precursors begin to position themselves paraxially (e.g., Rhee et al., 2005) and continue to divide synchronously until the early tailbud stage; after this point, they elongate ~4-fold, in the absence of cell division and evident positional reorganization, to allow tail extension (Passamaneck et al., 2007).

Recent molecular investigations have shown that primary muscle development is initially orchestrated by a maternally-deposited muscle determinant, the zinc-finger transcription factor Macho-1 (Nishida, 2002; Nishida and Sawada, 2001). In Ciona intestinalis, overexpression of Ci-macho1 and its morpholino-mediated knock-down cause ectopic expression and silencing, respectively, of the T-box transcription factor genes Ci-Tbx6b and Ci-Tbx6c, indicating that Ci-Macho1 somehow activates their early expression (Yagi et al., 2004a). In turn, Ci-Tbx6b and Ci-Tbx6c, which are considered the result of a recent lineage-specific duplication event (Dehal et al., 2002), function as
mediators of Ci-Macho1, since their morpholino-induced knock-downs result in the down-regulation of muscle-specific structural genes such as those encoding muscle actin, myosin chains and creatine kinase, among others (Yagi et al., 2005). Similarly, in Halocynthia roretzi, another solitary ascidian distantly related to Ciona, Hr-Macho1 has been shown to activate the expression of Hr-Tbx6, as well as that of structural muscle genes, such as Hr-muscle actin (Sawada et al., 2005). In addition to controlling primary muscle formation, Ci-Macho1 acts cooperatively with β-catenin to induce the formation of the heart field by activating Ci-Mesp (Christiaen et al., 2009; Davidson et al., 2005; Satou et al., 2004); the function of Ci-Macho1 in heart specification is also mediated by the Tbx6-related transcription factors (Christiaen et al., 2009; Davidson et al., 2005).

Despite the wealth of information on the gene regulatory network that initiates and sustains muscle development in Ciona and other ascidian embryos (e.g., Hudson and Yasuo, 2008; Imai et al., 2006; Meedel et al., 2007), the cis-regulatory mechanisms that integrate maternal and zygotic information along this complex gene cascade, from egg to swimming larva, are still unexplored. For example, it is still unclear how Ci-Macho1 activates its transcriptional intermediaries, and information on the structure and function of the cis-regulatory modules (CRMs) controlling expression of direct Ci-Macho-1 targets, and of their downstream structural muscle genes, is limited and fragmentary.

By combining misexpression assays and CRM analyses, we have begun to address these points and to gain a mechanistic understanding of how the activity of maternal and zygotic transcription factors is coordinated at the cis-regulatory level. Through the misexpression assays, we have probed the transcriptional plasticity of the Ciona mesoderm to assess the ability of temporally and spatially misexpressed Ci-
Macho1 to ectopically activate muscle genes in notochord cells, and we have identified a cross-regulatory interaction between Ci-Tbx6b and Ci-Tbx6c. Through a combination of sequence inspection, point-mutation analyses and electrophoretic mobility assays, we have characterized CRMs from the upstream genomic regions of these two genes, which occupy important nodes in the muscle gene regulatory network, as well as two muscle CRMs from a representative structural gene, *Fibrillar Collagen-1 (CiFColl)*, which is robustly expressed in muscle cells from late gastrulation until the late tailbud stages.

We present evidence that the expression patterns of *Ci-Tbx6b* and of *CiFColl* are recapitulated by compact bipartite CRMs, consisting of “early” and “late” moieties which are essential for initiation and maintenance of gene expression, respectively, and that the coordinated activity of these CRMs ensures continuity in the expression of genes necessary for muscle development and differentiation.
MATERIALS AND METHODS

Ascidians and electroporation

Adult *Ciona intestinalis* were purchased from Marine Research and Educational Products (M-REP; Carlsbad, CA, USA). Fertilization, dechorionation, electroporation and X-Gal staining were carried out as described in Corbo et al. (1997). Whenever necessary, embryos were fixed in 0.2% glutaraldehyde and stained at 37°C for 2-12 hours to enhance the signal. Each construct was tested on several different batches of embryos; graphs and error bars were obtained as previously described (Dunn and Di Gregorio, 2009).

Plasmid construction

The *Bra>macho* fusion was constructed by removing the coding region of *Ciona Brachyury (Ci-Bra)* from the -3.5 kb *Ci-Bra>eGFP* plasmid (Corbo et al., 1997) and by replacing *eGFP* with the coding region and 3’-UTR from the *Ci-macho1* cDNA. The *Ci-Bra* promoter was amplified from *Ci-Bra>eGFP* using the primers:

5’-ATAAGAATTCGGCTTATGACGAAATAATGT-3’ and
5’-ATATGCGGCCGCTATAGGTTTGTAACTCGCACT-3’

and the resulting PCR product was digested with *Eco*RI and *Not*I. The *Ci-macho1* coding sequence and 3’-UTR were amplified from a cDNA clone kindly provided by Dr. Yutaka Satou (Kyoto University), using primers:

5’-GTAGGCGGCGCATGGCCTTTACTGGTACGATGGGA-3’ and
5’-CAGGTGGCTCAGCTAATACGACTCACTATAGGGCG-3’
and the resulting PCR product was digested with \textit{Not}I and \textit{Blp}I. The two PCR products were cloned into \textit{Ci-Bra>eGFP} digested with \textit{Eco}RI and \textit{Blp}I through a triple ligation. To construct the \textit{Bra>Tbx6b} fusion plasmid, the \textit{Ci-Tbx6b} coding region was excised from the \textit{Ci-Tbx6b} cDNA and cloned into the \textit{Not}I and \textit{Kpn}I sites of \textit{Ci-Bra>eGFP}. To create the \textit{Bra>Tbx6c} plasmid, the full-length \textit{Ci-Tbx6c} ORF was reconstructed from two separate PCR-amplified fragments. The 5’-most fragment, containing the region encoding amino acids 9-269 of \textit{Ci-Tbx6c}, was PCR-amplified using as a template the published GST-Tbx6c fusion construct (Yagi et al., 2005; kindly provided by Dr. M. Levine), with primer Tbx6c-top:

5’-TGAAGCGGCCGCATGGCGACAGACATGAGAAGCCCAACCTTTGAACCGAAAGTTCTCATTCCAGG-3’,

which adds to the GST-Tbx6c construct the first eight codons of \textit{Ci-Tbx6c}, and primer:

5’-CAGTTTAGTGATCTGTCCGTTTTGGC-3’.

The remainder of the ORF was amplified using as a template clone GC43g03 from the \textit{Ciona} cDNA collection release 1 (Satou et al., 2002b), with primer:

5’-GGCACGAGGCCAAAACGGACAGATC-3’ and primer Tbx6c-ORF-bot:

5’-TTCCAAGCTAAGCTTTTATTCACTATAGGACACAATTACTAAC-3’.

The two PCR products resulting from these reactions were then used as templates for a final PCR in the presence of primers Tbx6c-top and Tbx6c-ORF-bot. The resulting product, encoding the full-length \textit{Ci-Tbx6c} protein, which spans 388 amino acids, was digested with \textit{Not}I and \textit{Blp}I and ligated into the p\textit{Ci-Bra-Linker} vector (Dunn and Di Gregorio, 2009).
The *Ciona Tbx6b* (*Ci-Tbx6b*) 5’-flanking region was PCR-amplified from *Ciona intestinalis* genomic DNA using the following primers:

5’-AATGTAGCGTGCCTTCACAAACCAGTCG-3’ and
5’-GAGACTCGTTTTCGATGCCACTTTG-3’.

The resulting ~2.4 kb fragment was cloned into the pFBΔSP6 vector (Oda-Ishii and Di Gregorio, 2007).

The *Ci-Tbx6b* basal promoter was PCR-amplified using primers:

5’-CGAGCCATGGGGCATCGAAAACGAGTCTCGC-3’ and
5’-ACTAGCGGCCGCCCATAGTCTTGTCTGGTGTTCAA-3’,

which yielded a 154-bp fragment encompassing the TATA box and nearby start of the longest *Ci-Tbx6b* EST, which we refer to as the transcription start site. The PCR-amplified fragment was cloned into the NcoI and NotI sites of the pFBΔSP6 vector.

Mutations in the core sequence (CCC->TTT) of the three Ci-Macho1 binding sites of the 2.4-kb *Ci-Tbx6b* CRM were introduced sequentially by PCR amplification.

The *Ciona Fibrillar Collagen-1* (*CiFCol1*) CRMs were identified through the analysis of a 2.2-kb fragment originally cloned from a *Ciona* genomic library.

The *Ci-Tbx6c* CRM was PCR-amplified from *Ciona* genomic DNA using the primers:

5’-ATCTCGAGGATTCTTTAAGATATTTTTGGATAATG-3’ and
5’-CATCTAGACGTAACGATCGATCAAAGTTAAATTAAAC-3’.

In all the PCR amplifications, ~125 ng of genomic DNA extracted from the sperm of a single *Ciona intestinalis* adult were used as a template, in the presence of either Hi-Fi Taq (Invitrogen, Carlsbad, CA, USA) or Turbo *Pfu* (Stratagene, La Jolla, CA, USA).
DNA polymerase. All plasmids were checked for accuracy either manually (Sambrook et al., 1989) or through automated sequencing facilities.

Whole-mount in situ hybridizations (WMISH)

WMISH were performed as previously described (Oda-Ishii and Di Gregorio, 2007) using digoxigenin-labeled antisense RNA probes. The CiFColl probe was prepared from EST GC01e23, the Ci-macho1 probe from EST GC18h07, the Ci-Tbx6b probe from EST GC09i19, and the Ci-Tbx6c probe from EST GC43g03 (Satou et al., 2002b). The LacZ probe was kindly provided by Dr. Y. Nibu.

Quantitative RT-PCR (qRT-PCR)

Ciona zygotes from a single round of fertilization were divided into two groups and were either electroporated with Ci-Bra>eGFP or co-electroporated with Ci-Bra>eGFP and Bra>macho constructs. Once they reached the 110-cell stage, only the embryos showing fluorescence in the notochord were selected for RNA extraction using an epifluorescence microscope. RNA was extracted using the RNeasy Protect Mini kit (Qiagen, Valencia, CA, USA), and treated with RNase-free DNase (Qiagen) using the on-column digestion protocol to remove possible genomic DNA contamination. cDNA was synthesized using the Superscript III kit (Invitrogen, Carlsbad, CA, USA). This process was repeated twice on different batches of embryos to generate two independent samples of Bra>macho cDNA and corresponding wild-type control cDNA.

qRT-PCR samples were prepared using 2x SYBR Green Hotstart Mix (Applied Biosystems, Foster City, CA, USA) according to the manufacturer’s instructions. The
resulting data were analyzed using the SDS2.3 software (Applied Biosystems). Both sets of cDNAs were loaded with the appropriate primers on a single plate, to eliminate variation between PCR amplifications. Samples were run in triplicate, using Ci-GAPDH as a control. Error bars represent the variation seen between the two sets of cDNAs. The relative standard curve method was used to calculate the fold change in gene expression, as described by the manufacturer (http://www3.appliedbiosystems.com/AB_Home/index.htm). We considered the 1.6-fold increase in gene expression relevant in consideration of the low number of notochord precursors at the 110-cell stage (ten), the electroporation efficiency, and the high variability in the number of cells that efficiently incorporate the construct(s) within each transgenic embryo.

**Recombinant proteins synthesis and purification**

The full-length Ci-Macho1 coding region was PCR-amplified from the Bra>macho plasmid described above and cloned into the BamHI and SacI sites of the pRSET-B vector (Invitrogen, Carlsbad, CA, USA). Full-length Ci-Tbx6b and Ci-Tbx6c coding regions were PCR-amplified and cloned into the pGEX-KG vector (GE Healthcare, Piscataway, NJ, USA) in frame with the GST tag. The resulting plasmids were transformed into E. coli BL21 (DE3) cells harboring the pJY2 plasmid (Affiniti Research Products, UK) and the fusion proteins were purified as previously described (Gazdoiu et al., 2005). The 6xHis-tagged Ci-Macho1 protein was expressed at 15°C in the presence of 0.1 mM IPTG.
**Electrophoretic mobility shift assays (EMSA)**

The following double-stranded oligonucleotides were used (only the 5’-3’ strand is reported, mutations are underlined):

Macho site-A-wt: CAAACTATGCATCGGGTGTCAGGGAGACA
Macho site-A-MUT: CAAACTATGCATCTttTGTCAGGGAGACA
Macho site-B-wt: GACAACGATCTATGGGAGCCGTGAGGATA
Macho site-B-MUT: GACAACGATCTATttAGCCGTGAGGATA
Macho site-C-wt: AAGTAATGAGGACCCCGCTGCGGTAACCT
Macho site-C-MUT: AAGTAATGAGGACTttGCTGCGGTAACCT

T-box site wt: 5’-CGGTATGCGTCACACTGAGTTTTG-3’
T-box site MUT: 5’-CGGTATGCGTCAacaTGAGTTTTG-3’

Radioactive labeling and subsequent procedures were performed as previously described (Dunn and Di Gregorio, 2009).
RESULTS

Effects of the ectopic expression of Ci-Macho1 on the transcription of Ci-Tbx6b and Ci-Tbx6c.

To gain insights on the mechanisms employed by Ci-Macho1 to control expression of Ci-Tbx6b and Ci-Tbx6c, we misexpressed this transcription factor spatially and temporally in notochord cells using the Ciona Brachyury (Ci-Bra) promoter region (Corbo et al., 1997). Ci-macho1 is normally expressed in germ cell precursors in 110-cell stage embryos, and in sensory vesicle, nerve cord and germ cells in mid-tailbud embryos (Fig. 1A,C,D; Satou et al., 2002a); no detectable expression in notochord cells has been seen at any stage. For this reason, the Ci-Bra promoter, which is active predominantly in notochord cells and mesenchyme, and only sporadically in a few muscle cells, was chosen for the misexpression experiments. When the misexpression experiments were attempted using the Ci-FoxA-a promoter, which encompasses a wider expression territory (Di Gregorio et al., 2001), the effects on embryogenesis were too extensive to allow meaningful interpretation of the results (data not shown).

Transgenic embryos carrying the Bra>macho construct (Fig. 1B) efficiently express Ci-macho1 in notochord cells and their precursors (Fig. 1E,F) and display a short, stubby tail (Fig. 1B,F,L). 110-cell stage transgenic Bra>macho embryos hybridized with a probe for Ci-Tbx6b (Fig. 1G), which is normally expressed in muscle precursors (Fig. 1I; Takatori et al., 2004), show ectopic expression of this gene in notochord precursors (Fig. 1K); however, at the early tailbud stage (Fig. 1H), when Ci-Tbx6b is normally undetectable (Fig. 1J), a consistent ectopic signal is observed only in a subpopulation of trunk mesenchyme cells (Fig. 1H,L). This result can be explained by previous
observations of the activity of the Ci-Bra promoter region in mesenchyme cells (Corbo et al., 1997).

In contrast to Ci-Tbx6b, Ci-Tbx6c (Fig. 1M-O), which is normally expressed only in a subset of muscle precursors (Fig. 1N), was ectopically activated in transgenic Bra>macho embryos in notochord precursors only at the 110-cell stage (Fig. 1O), and was not detected at later stages (data not shown). The transcriptional changes induced by the misexpression of Ci-macho1 in 110-cell transgenic embryos were validated and quantified by qRT-PCR (Fig. 1P).

Together, these results suggest that notochord precursors are competent to respond to Ci-Macho1, although only at early stages of their development.

**Cross-regulatory interactions between Ci-Tbx6b and Ci-Tbx6c.**

We next sought to assess whether Ci-Tbx6b and Ci-Tbx6c, which possess identical binding affinities in vitro (Yagi et al., 2005), are capable of activating each other’s transcription. To test this hypothesis, we misexpressed Ci-Tbx6b and Ci-Tbx6c in the notochord by cloning the respective cDNAs downstream of the Ci-Bra promoter. As a first step, we ascertained whether these genes were efficiently transcribed in the notochord of transgenic embryos; a representative control experiment is presented in Fig. S1. Subsequently, we hybridized embryos carrying the Bra>Tbx6b transgene with the Ci-Tbx6c probe, and embryos carrying the Bra>Tbx6c transgene with the Ci-Tbx6b probe (Fig. 1Q). We found that in embryos bearing the Bra>Tbx6b transgene, ectopic expression of Ci-Tbx6c is seen in notochord precursors (Fig. 1Q, left panel; compare to Fig. 1N) and similarly, in embryos carrying the Bra>Ci-Tbx6c transgene, transcription of
Ci-Tbx6b is ectopically activated in these cells (Fig. 1Q, right panel; compare to Fig. 1I). Under the experimental conditions that we employed, no cross-hybridization between the two probes was observed. These results indicate that these related transcription factors possess the potential for reciprocal activation.

A bipartite CRM recapitulates spatial and temporal muscle expression of Ci-Tbx6b.

Beginning at the 16-cell stage, Ci-Tbx6b is specifically expressed in the precursors of the larval muscles derived from the B5.1 blastomere pair (Takatori et al., 2004). In 110-cell embryos, Ci-Tbx6b expression has expanded to include numerous muscle precursors (Fig. 1I), while by the neurula stage it becomes confined to the posterior-most muscle cells (Takatori et al., 2004). Ci-Tbx6b (gene model ci0100144249; Dehal et al., 2002) is clustered in a tandem arrangement with Ci-Tbx6c (ci0100144293) on sequence scaffold 126 (JGI v2.0; http://genome.jgi-psf.org/Cioin2/Cioin2.home.html Dehal et al., 2002) (Fig. 2A,S and Fig. S2A). A genomic fragment spanning 2.4 kb was PCR-amplified from the 5′-flanking region of Ci-Tbx6b and its various truncations were tested in vivo by parallel electroporations (e.g. Di Gregorio and Levine, 2002) to identify the minimal sequences necessary for its function. Truncated versions of the 2.4-kb fragment were cloned upstream of both the endogenous Ci-Tbx6b promoter and the Ci-FoxA-a basal promoter (Fig. S2B). The 2.4-kb CRM directed strong expression in muscle and mesenchyme cells (Fig. 2B-D, Fig. S2D-F and Table 1); occasional staining was also observed in trunk ventral cells, the heart precursors (data not shown). The use of the Ci-FoxA-a basal promoter (Oda-Ishii and Di Gregorio, 2007) allowed the identification of a
minimal 112-bp fragment sufficient for activity in a heterologous context (shaded area in Fig. S2B).

Sequence analysis of the 2.4-kb Ci-Tbx6b CRM identified three bona fide binding sites for Ci-Macho1 (green squares in Fig. 2A), all partially matching the consensus sequence previously identified (Yagi et al., 2004a; Fig. 2A, bottom). We performed EMSA to verify the functionality of these sites, and found that all three sequences are bound in vitro by a Histidine-tagged Ci-Macho1 fusion protein (Fig. S3A), as well as by a GST-tagged Ci-Macho1 protein (data not shown), with affinities that reflect their different homologies with the published consensus sequence (Fig. 2A, bottom). The three Ci-Macho1 binding sites are found within a 463-bp region. To verify the involvement of the Ci-Macho1 binding sites in the temporal control of the CRM activity, we compared the developmental windows of transcriptional activity of the three Ci-Tbx6b constructs shown in Fig. 2A. The first construct contains the 2.4-kb wild-type Ci-Tbx6b CRM fused to LacZ, the second construct contains an 862-bp 5′-truncated version of the previous fragment, which lacks all Ci-Macho1 binding sites, and the third construct consists of the 2.4-kb Ci-Tbx6b CRM carrying mutations in all three Ci-Macho1 binding sites. To accurately determine the respective windows of activity, one-cell stage Ciona embryos were electroporated separately with each construct, cultured until the 32-cell (Fig. 2B,E,H), 110-cell (Fig. 2C,F,I) and early neurula stages (Fig. 2D,G,J), then fixed and hybridized in situ with a LacZ RNA probe. In embryos electroporated with the 2.4-kb wild-type Ci-Tbx6b CRM, a strong hybridization signal was detected in the B6.4 pair of muscle precursors starting from the 32-cell stage (Fig. 2B). At the 110-cell stage, LacZ expression was detected in descendants of B6.4 blastomeres and in the B8.7 and B8.8
pairs, which derive from the B6.2 pair (Fig. 2C). In early neurulae, LacZ expression was detected only in the posterior-most muscle cells (Fig. 2D). This pattern faithfully recapitulates the expression of Ci-Tbx6b (Takatori et al., 2004). Embryos electroporated with the 862-bp CRM, which does not contain the Ci-Macho1 binding sites, showed no detectable signal at either the 32-cell stage (Fig. 2E) or at the 110-cell stage (Fig. 2F); LacZ expression became detectable by the early neurula stage in only the posterior-most muscle cells (Fig. 2G; note that the asymmetric expression is due to mosaic incorporation of the transgene). Finally, the vast majority of embryos carrying the 2.4-kb CRM with mutations in all Ci-Macho1 binding sites showed no expression of the reporter at either the 32-cell (Fig. 2H) or the 110-cell stage (Fig. 2I), and displayed active transcription in the posterior-most muscle cells only at the neurula stage (Fig. 2J). These results indicate that mutations in the Ci-Macho1 sites do not affect the spatial (i.e., lineage-specific) activity of the 2.4-kb Ci-Tbx6b CRM, but are sufficient to cause a considerable delay in the onset of transcription driven by this fragment.

Together, these observations suggest that: a) the temporal information required for the early activity of the 2.4-kb Ci-Tbx6b CRM is stored in its distal region; b) this temporal information is encoded by the Ci-Macho1 binding sites; c) the 862-bp proximal region contains the cis-regulatory elements responsible for the late muscle activity observed in neurulae.

The proximal region of the Ci-Tbx6b CRM recapitulates the late muscle expression of Ci-Tbx6b.
The proximal 112-bp region of the *Ci-Tbx6b* CRM (shaded area in Fig. S2B), which was first identified using the *Ci-FoxA-a* basal promoter, was further tested using the endogenous 154-bp *Ci-Tbx6b* promoter. The resultant 266-bp construct (Fig. 2K; dark orange area in the schematics in Fig. 2A) was used for *LacZ* time-course experiments. The majority of the embryos electroporated with the 266-bp construct did not show staining at the 32-cell (Fig. 2L) or at the 110-cell stage (Fig. 2M); however, by the early neurula stage, 87% of the embryos showed strong staining in muscle cells (Fig. 2N). These results suggest that this 266-bp region is sufficient to recapitulate the late activity of the 862-bp fragment previously analyzed (Fig. 2H-J).

Sequence inspection identified at least three putative binding sites within this region: a distal sequence showing an incomplete match with the CREB binding sites previously identified in other *Ciona* muscle CRMs (TGACG core; blue diamond) (Kusakabe et al., 2004; Brown et al., 2007), a T-box binding site matching the core consensus sequence identified for *Ci-Tbx6b* and *Ci-Tbx6c* via SELEX assays (7 out of 10 matches; red oval) (Yagi et al., 2005), and an ‘AC’-core E-box (pink triangle) (Erives et al., 1998). In vivo analysis of progressive truncations and specific point mutations showed that removal of the imperfect CREB binding site did not noticeably affect either the intensity of the muscle staining or the percentage of embryos showing activity (data not shown). On the other hand, when the T-box binding site was mutated, a considerable reduction was observed not only in the intensity of the muscle staining but also in the percentage of embryos showing activation of the reporter gene (Fig. 2O,P,R). Finally, a mutation of the ‘AC’-core E-box left the muscle staining unaffected (Fig. 2Q,R). The results of three independent experiments are quantified in the graph in Fig. 2R.
Comparable results were obtained when transgenic embryos were fixed at the neurula stage (insets in Fig. 2O-Q).

Together, these findings suggest that the proximal region of the \( Ci-Tbx6b \) CRM is mainly controlled by a T-box transcription factor. To further investigate this point, we carried out EMSA using bacterially expressed Ci-Tbx6b and Ci-Tbx6c proteins and a radiolabeled oligonucleotide probe containing the putative T-box binding site (Fig. S3B). For these experiments we synthesized full-length proteins in order to identify possible differences in binding affinities (see Methods); nevertheless, we found that this sequence is bound with similar intensity by both GST-Ci-Tbx6b and GST-Ci-Tbx6c fusion proteins (Fig. S3B). These results suggest that the late-acting \( Ci-Tbx6b \) CRM might function as an autoregulatory and/or a cross-regulatory enhancer sequence.

Lastly, we used the information gathered from the analysis of the \( Ci-Tbx6b \) CRM to predict the location of a related muscle CRM in the \( Ci-Tbx6c \) locus. \( Ci-Tbx6c \) is clustered with \( Ci-Tbx6b \) within a 8-kb genomic region (Fig. 2S and Fig. S2A). Within the 2-kb sequence directly upstream of \( Ci-Tbx6c \), we identified a region spanning 293 bp which contained bona fide binding sites for Ci-Macho1 and Ci-Tbx6b/c, as well as an ‘AC’-core E-box (Fig. 2S). When we tested this fragment in vivo we found that, as expected, it was able to direct expression in muscle cells (Fig. 2T), although in a territory wider than that covered by \( Ci-Tbx6c \). This result suggests that the muscle activity of the highly related \( Ciona Tbx6 \) genes is controlled by the same basic set of structural cis-regulatory elements.
The composite cis-regulatory region of *Ciona Fibrillar Collagen-1* harbors two muscle CRMs with different temporal onsets.

We next sought to investigate whether the structural and functional criteria that we had identified for the Tbx6-related transcription factor genes applied to the cis-regulatory region of a structural muscle gene. The *Ciona Fibrillar Collagen-1* gene (*CiFCol1*; JGI gene model ci0100150759) (Wada et al., 2006) encodes a member of the fibrillar collagen family related to the vertebrate clade A *collagen* genes (Wada et al., 2006). During *Ciona* early embryogenesis, this gene is first expressed mostly in muscle precursors and by the neural plate stage its expression has expanded to mesenchyme, notochord, endoderm and CNS (Satou et al., 2001; our unpublished results). Expression of *CiFCol1* is maintained in all these tissues at very high levels throughout the mid-tailbud stage (Fig. 3A) until hatching (Satou et al., 2001; our unpublished results). The sustained expression of *CiFCol1* in muscle cells suggested that transcription of this gene might rely upon a combination of early and late activators. Sequence inspection of its 5’-flanking region indicated that *CiFCol1* might be controlled, at least at early developmental stages, by Ci-Macho1 and/or its intermediaries. To verify this hypothesis, we first employed the misexpression assays described in Figure 1. Since *CiFCol1* is normally expressed in notochord precursors starting from the late gastrula stage, and is not present at detectable levels in these cells in early gastrulae (Fig. 3B), we harvested embryos at this developmental stage which had been electroporated with either *Bra>macho*, *Bra>Tbx6b* or *Bra>Tbx6c* and monitored the effects on the activation of *CiFCol1* transcription in notochord precursors by WMISH (Fig. 3C-E). These
experiments show that Ci-Macho1, Ci-Tbx6b and Ci-Tbx6c are all able to induce precocious onset of CiFCol1 expression in the notochord lineage.

We then investigated the molecular mechanisms mediating the response of this structural gene to the misexpression of the three muscle transcription factors by characterizing the CiFCol1 cis-regulatory region. We found that a 2.2 kb fragment from the 5′-flanking region of CiFCol1 (yellow rectangle in Fig. 3F) is able to fully recapitulate its expression pattern (inset in Fig. 3F). Within this main fragment, 3 different CRMs were identified, including one 400-bp distal muscle CRM (blue rectangle in Fig. 3F), and two adjacent proximal CRMs, one directing expression in notochord (data not shown) and the other, 230 bp long, which is active in muscle cells (orange rectangle in Fig. 3F). The distal muscle CRM (Fig. 3G) contains two generic E-boxes (pink triangles) and directs expression in muscle cells (orange arrowheads; see lineage map in Fig. S4A) and in the two rows of lateral ependymal cells of the nerve cord (blue arrowheads). The 230-bp proximal muscle CRM (Fig. 3H) contains four putative T-box binding sites (generic sequence: TNNCAC; red ovals) and one putative low-affinity Ci-Macho1 binding site (Yagi et al., 2004a; green rectangle) and directs expression in most of the primary muscle cells and in trunk mesenchyme (Fig. 3H).

To accurately define the window of activity of each muscle CRM, we performed time-course experiments using the LacZ probe (Fig. 3I,J,L-O,Q,R). In addition, we monitored the accumulation of β-galactosidase in the muscle cells where each CRM was active by performing X-Gal staining (Fig. 3K,P). Embryos carrying the distal 400-bp muscle CRM began accumulating LacZ transcripts between the 32-cell and the 110-cell stage (Fig. 3I,J). In the majority of the embryos analyzed, high levels of LacZ transcripts
were predominantly detected in only one pair of muscle precursors (Fig. S4A), although a few embryos also showed a faint signal in additional muscle precursors (orange arrowheads in Fig. 3J), including the mixed-lineage A8.16 blastomeres (orange/blue arrowhead in Fig. 3J). Consistent with this early pattern, at the early neurula stage β-galactosidase accumulation was detected only in a small subset of muscle and mesenchyme precursors (Fig. 3K). By the early tailbud stage (Fig. 3L), the activity of the CRM had considerably expanded, to encompass virtually all muscle cells (18 per side; Fig. S4B), as well as part of the nerve cord and mesenchyme cells. The CRM remained active in most muscle cells through the mid-tailbud stage (Fig. 3M).

Starting at the 32-cell stage (Fig. 3N), over one third of the embryos electroporated with the proximal 230-bp CiFColl muscle CRM showed LacZ expression in primary muscle precursors; by the 110-cell stage, high levels of LacZ transcripts were observed in most pairs of muscle precursors, except for the A8.16 pair (Fig. S4A). At the early neurula stage, this pattern translated into a homogeneous accumulation of β-galactosidase in the majority of muscle cells (Fig. 3P). At the early tailbud stage, this CRM was still active in muscle and mesenchyme (Fig. 3Q). However, at the mid-tailbud stage, transcriptional activity had faded almost completely from the muscle in the majority of the embryos, while a residual signal was occasionally seen in mesenchyme cells (Fig. 3R). The activity of each CRM in muscle cells is plotted in the graph in Fig. S4C and compared to the number of muscle cells per side of the embryo at the stages analyzed.

From these observations, we conclude that of the two muscle CRMs that we have identified in the CiFColl 5′-flanking region, the proximal one is activated earlier,
beginning at the 32-cell stage, and ceases its function around the mid-tailbud stage, while the distal one is activated 2-3 cell divisions later, but persists beyond the mid-tailbud stage.

Selective activation of the CiFCol1 CRMs by ectopically expressed Tbx6 transcription factors.

To test whether the muscle CRMs identified within the CiFCol1 locus were mediating the response of the endogenous gene to the ectopic expression of Ci-Tbx6b and/or Ci-Tbx6c, we co-electroporated each CRM with either Bra>Tbx6b or Bra>Tbx6c. As shown in Figure 4, the misexpression of Ci-Tbx6b is sufficient to induce ectopic expression of the CiFCol1 proximal 230-bp muscle CRM in notochord cells (Fig. 4A,B), consistent with the possibility that Ci-Tbx6b activates this enhancer. While performing controls for these experiments, we noticed that in some batches of embryos the pFBΔSP6 vector was itself slightly responsive to the Bra>Tbx6b construct; however, the ectopic activation observed in the notochord when the 230-bp CiFCol1 CRM was used in these co-electroporations instead of the empty vector was at least 3.5 times higher (Fig. 4D). No ectopic activation in the notochord was observed when the 230-bp CiFCol1 CRM was co-electroporated with Bra>Tbx6c (Fig. 4C,D). In addition, we tested the response of these CRMs to another Tbx6-related transcription factor, Ci-Tbx6a (Takatori et al., 2004), and to a more distantly related T-box factor, Ci-Tbx15/18/22 (Erives and Levine, 2000; Takatori et al., 2004), and we found no response above background to either factor (data not shown).
Lastly, no ectopic activation was observed when the distal CiFCol1 CRM was co-electroporated with either \( \text{Bra} > Tbx6b \) or \( \text{Bra} > Tbx6c \) (data not shown); this result is consistent with the lack of evident Tbx6b/c binding sites in this sequence (Fig. 3G).

The flow of maternal and zygotic developmental information in the muscle cells of \textit{Ciona}.

Our findings are summarized by the model in Figure 5, which reconstructs part of the \textit{cis}-regulatory hierarchy underlying the cell-autonomous developmental program that characterizes primary muscle formation in \textit{Ciona}. Arrows indicate the interactions between transcription factors and \textit{cis}-regulatory sequences, either demonstrated by the work presented here (black) or inferred (grey).

After fertilization, binding of maternally encoded Ci-Macho1 protein to the early Ci-Tbx6b CRM activates transcription of this gene in muscle precursors. By the 16-cell stage, when the maternal Ci-macho1 transcripts start declining (Satou et al., 2002a), transcription of Ci-Tbx6b begins. At later stages, it is likely that the Ci-Tbx6b early CRM is progressively vacated and is first aided and then progressively replaced by the late CRM, which is likely activated by Ci-Tbx6b and/or Ci-Tbx6c. Given the similarities that we have identified in the \( \text{Ci-Tbx6b} \) and \( \text{Ci-Tbx6c} \) \textit{cis}-regulatory sequences, it is plausible that similar mechanisms activate the \( \text{Ci-Tbx6c} \) CRM. Ci-Tbx6b in turn activates the proximal CRM of CiFCol1, which recapitulates the early transcription of this gene. By the neurula stage, as expression of \( \text{Ci-Tbx6b} \) fades, different muscle activators bind the distal, late-acting CiFCol1 CRM and allow transcription of this gene to proceed without interruptions until the late tailbud stage.
DISCUSSION

Spatial and temporal heterogeneity of the embryonic territories responsive to Ci-Macho1.

For over a century, it has been known that upon fertilization and first cleavages of the ascidian egg, maternally-loaded cytoplasmic determinants are differentially segregated into the resulting blastomeres, thus giving rise to the early determination and invariance which characterize their developmental fates (Conklin, 1905; Deno et al., 1984; Ortolani, 1955). While recent molecular evidence has highlighted the major role of inductive processes in the formation of most embryonic tissues (reviewed by Lemaire, 2009), the cell-autonomous development of the primary muscle of the ascidian larva is largely attributable to the maternal determinant Macho-1 (Nishida and Sawada, 2001). macho-1 postplasmic mRNA is relocated after fertilization by the cortical centrosome-attracting body (CAB) (Sardet et al., 2003). As cleavage proceeds, in both Halocynthia (Nishida and Sawada, 2001) and Ciona (Satou et al., 2002a) macho-1 mRNA becomes progressively restricted to a narrow region of the embryo, the B7.6 blastomeres; however, the Macho-1 protein is generally believed to persist in an unlocalized form, and to be distributed to all descendants of the B4.1 cells (Kondoh et al., 2003). Studies in Halocynthia show that for the proper formation of other lineages that also derive from the B4.1 cells, such as mesenchyme and endoderm, the function of Macho-1 needs to be actively suppressed by FGF and BMP signaling pathways (Kim et al., 2000; Kim and Nishida, 1999; Kondoh et al., 2003). Similar mechanisms are also likely responsible for the functional suppression of zygotically expressed Ci-Macho-1 in the Ciona CNS,
considering that \textit{Ci-FGF16/19/20} is expressed in the \textit{Ciona} CNS through tailbud stages and is required for neural development (Imai et al., 2004; Bertrand et al., 2004).

The misexpression experiments described here suggest that no such restraining mechanism is present in notochord cells before the early tailbud stage. In fact, at early developmental stages the ectopic activation of both \textit{Ci-Tbx6b} and \textit{Ci-Tbx6c} was seen in notochord precursors of both lineages in \textit{Bra>macho} embryos, while at the mid-tailbud stage only the ectopic activation of \textit{Ci-Tbx6b} was observed, and it was confined to a subset of mesenchyme cells. These cells are most likely descendants of the B7.3 blastomere, a 64-cell stage precursor of both secondary notochord and mesenchyme cells (Satoh, 1994). The differential competence of the notochord to respond to Ci-Macho1 might be explained by the requirement for temporally and spatially localized co-factors and/or transcriptional intermediaries. Alternatively, as in the case of the CNS, Ci-Macho1 might be functionally suppressed in the notochord of tailbud embryos by the activation of the FGF signaling pathway, as suggested by the observation that \textit{Ci-FGFR} is expressed in the notochord beginning at the early tailbud stage (Imai et al., 2004; Shi et al., 2009). These mechanisms might also account for the relatively mild phenotype that we observed in embryos carrying the \textit{Bra>macho} transgene, whereby the notochord is still able to form, even in transgenic embryos where mosaic incorporation is minimal.

The spatio-temporal expression pattern of \textit{Ci-Tbx6b} is recapitulated by early and late \textit{cis}-regulatory sequences.

Using in vivo transient transgenic assays, we have identified a 2.4-kb CRM upstream of \textit{Ci-Tbx6b} that is able to faithfully recapitulate the muscle expression of this
gene. The temporal muscle activity of the 2.4-kb CRM represents the composite read-out of early- and late-acting cis-regulatory sequences, which interpret maternal and zygotic information. The Ci-Tbx6b CRM contains a distal region which functions as the repository of the temporal information necessary to recapitulate the early expression pattern previously reported for Ci-Tbx6b (Takatori et al., 2004). When this distal region is deleted, muscle activity is not lost, but its onset is considerably delayed. Sequence inspection and point-mutation analyses suggested that this early-acting distal region might be controlled by maternal Ci-Macho1, since three putative binding sites for this factor are present in this sequence. We found that these sites are bound in vitro by Ci-Macho1 and that their concomitant mutation is sufficient to cause the same delay in the onset of transcriptional activity that was observed when the entire fragment encompassing them was deleted. Together, these observations provide a mechanistic cis-regulatory explanation to the results of our misexpression assay, as well as to previous results showing that overexpression of Ci-Macho1 is sufficient to induce ectopic expression of Ci-Tbx6b (Yagi et al., 2004a) and that, likewise, Hr-Macho-1 is able to ectopically induce Hr-Tbx6, among other muscle genes (Sawada et al., 2005). It is noteworthy that in Ciona, Ci-ZicL cooperates with Ci-Macho1 to promote muscle development (Yagi et al., 2005); this zygotic zinc-finger transcription factor is related to Ci-Macho1 and recognizes a similar consensus binding site in vitro (Yagi et al., 2004b). Interestingly, one of the three Ci-Macho1 binding sites that we have characterized in the Ci-Tbx6b CRM, namely site “C”, contains permutations of the published Ci-ZicL consensus site that are compatible with binding in vitro (Yagi et al., 2004b). If this site is bound in vivo by either transcription factor, then this would explain the observation that
*Ci-Tbx6b* is still weakly expressed in Ci-Macho1 morphant embryos, while its expression is no longer detectable in Ci-Macho1/Ci-ZicL double-morphants (Yagi et al., 2005).

Within the 2.4-kb CRM, a 266-bp proximal region is able to direct transcription only from neurulation on, thus acting as a late muscle enhancer. Sequence analysis of this region revealed the presence of an imperfect CREB binding site, a T-box binding site (generic sequence: TNNCAC) partly matching the core consensus sequence previously reported for Ci-Tbx6b/c (Yagi et al., 2005), and an ‘AC’-core E-box. Both CREB binding sites and AC core E-boxes have been previously shown to be necessary for muscle activity of other muscle CRMs (Brown et al., 2007; Erives et al., 1998; Kusakabe et al., 2004); however, in this case, only the T-box site substantially contributes to the muscle activity, qualitatively and quantitatively. Through EMSA, we have shown that this T-box site is bound in vitro by both Ci-Tbx6b and Ci-Tbx6c.

**Structure and spatial range of action of the *Ci-Tbx6c* CRM.**

We have employed the structural knowledge gained from the analysis of the *Ci-Tbx6b* regulatory region to identify a similarly organized muscle CRM in the *Ci-Tbx6c* locus. The *Ci-Tbx6c* muscle CRM that we have identified contains putative binding sites for Ci-Macho1 and Ci-Tbx6b/c, as well as an ‘AC’-core E-box. Unlike the *Ci-Tbx6c* gene, which is expressed exclusively in a subset of muscle cells, the *Ci-Tbx6c* CRM encompasses a broader territory. This observation suggests that the activity of the *Ci-Tbx6c* CRM and the ability of Ci-Tbx6b to activate *Ci-Tbx6c* expression, which we have uncovered through the misexpression experiments, are efficiently counteracted in vivo, likely by a transcriptional repressor(s). A possible candidate for this role is Ci-Hairy-a,
one of the *Ciona* orthologs of a well-characterized transcriptional repressor (e.g., Nibu et al., 2001), which is expressed in blastomeres B8.7 and B8.8, but not in blastomere B8.15 (Imai et al., 2004), in a pattern that complements that of *Ci-Tbx6c*. Supporting this possibility, two putative Hairy binding sites (tggCACGTGcca; Ohsako et al., 1994) are present upstream of *Ci-Tbx6c* (data not shown) and are not included in the 293-bp *Ci-Tbx6c* CRM.

**Modular organization and temporal properties of the cis-regulatory region of *CiFCol1*.**

Originally isolated in a subtractive screen aimed to identify genes downstream of Ci-Bra (Takahashi et al., 1999), the *CiFCol1* gene attracted our interest because of its sustained muscle expression, which begins around mid-gastrulation, and because its upstream region is enriched in T-box binding sites.

Dissection of a 2.2-kb genomic fragment located upstream of the transcription start site of *CiFCol1* revealed the presence of discrete CRMs active in all the tissues where *CiFCol1* is expressed. In particular, this 2.2-kb fragment harbors two distinct muscle CRMs: a distal CRM containing two generic E-boxes and depleted of T-box binding sites and Ci-Macho1 binding sites, and a proximal CRM containing four clustered T-box binding sites, some of which are bound weakly in vitro by the Ci-Tbx6b protein (data not shown), and a low-affinity Ci-Macho1 binding site. The heterogeneity of these sequences is reflected by the temporal activity of the two CRMs, since the distal one, which does not contain any apparent T-box binding sites, is activated later than the proximal, which is enriched in these motifs. In particular, the distal *CiFCol1* muscle
CRM is active in a small subset of muscle precursors from the 110-cell to the neurula stage, and only by the early tailbud stage does its territory expand to encompass all muscle cells. Afterwards, it remains active in the majority of muscle cells. Therefore, the spatial range of action of this CRM in the muscle seems to be controlled by an activator(s) functioning from neurulation on. The presence of two E-boxes in this sequence prompted us to investigate the possible involvement of transcription factors of the bHLH family in the regulation of this CRM. We found that neither the mutation of the E-boxes nor the misexpression, individual or combined, of two bHLH transcription factors, Ci-MRF (Meedel et al., 2007) and Ci-paraxis (Erives, 2009; Imai et al., 2004) had any detectable effect (data not shown), thus leaving the identification of the late activator(s) to future investigations.

Conversely, the proximal CiFCol1 muscle CRM is ignited early in most muscle cell precursors, starting from the 32-cell stage, but its activity fades by the mid-tailbud stage. We conclude that the additive activity of the two CRMs is likely responsible for the sustained expression of CiFCol1 in muscle cells.

Interestingly, misexpression of Ci-Macho1, Ci-Tbx6b or Ci-Tbx6c in notochord cells all result in ectopic activation of CiFCol1 in this territory. While we cannot rule out that this might be attributable to the low-affinity Ci-Macho1 binding site in the CiFCol1 early CRM, given the late onset of CiFCol1 muscle expression it seems more likely that Ci-Macho1 activates expression of CiFCol1 indirectly, through Ci-Tbx6b. To test this hypothesis we monitored the response of the CiFCol1 proximal muscle CRM to the misexpression of Ci-Tbx6b in notochord cells. We found that misexpression of Ci-Tbx6b caused the ectopic activation of the CiFCol1 proximal muscle CRM in the notochord,
while misexpression of \textit{Ci-Tbx6c} did not have any effect. We conclude that the ectopic activation of \textit{CiFCol1} seen in notochord cells of embryos carrying the \textit{Bra>Tbx6c} construct might occur indirectly, via the activation of \textit{Ci-Tbx6b} expression by Ci-Tbx6c.

Finally, no ectopic activation was observed when the distal \textit{CiFCol1} muscle CRM was co-electroporated with either construct (data not shown), consistent with the lack of Tbx6b/c binding sites in its sequence.

\textbf{Muscle gene regulation in ascidians versus vertebrates: lineage-specific innovations and evolutionarily conserved mechanisms.}

By analyzing the \textit{cis}-regulatory sequences that mediate the response to Ci-Macho1 and its mediators, this study has begun to provide sharper insights into the molecular mechanisms controlling cell-autonomous muscle development in the ascidian embryo. Given the large number of genes that respond to Ci-Tbx6b and Ci-Tbx6c (Yagi et al., 2005; this study), it is conceivable that the mechanisms of transcriptional regulation that control the CRMs presented here might be shared by several other muscle genes. This hypothesis is supported by the abundance of putative Tbx6b/c binding sites in muscle CRMs identified by other groups (Brown et al., 2007; Kusakabe et al., 2004) and by our lab (data not shown).

While the early cell-fate determination mediated by Macho-like proteins in muscle cells has been described so far as an ascidian-specific mechanism, transcription factors of the Zic family, of which Macho, ZicL and related proteins represent a diverged branch (Aruga et al., 2006), are known to be required for shaping the body plan of widely different animals (Merzdorf, 2007). In addition, Tbx6-related proteins in \textit{Ciona} appear to
be part of an evolutionarily conserved kernel that is employed for the specification and
differentiation of paraxial mesoderm in several other chordates, including mouse (Chapman and Papaioannou, 1998; White et al., 2003), *Xenopus* (Tazumi et al., 2008; Uchiyama et al., 2001) and zebrafish (Goering et al., 2003). Hence, the elucidation of the *cis*-regulatory mechanisms employed by these transcription factors to modulate expression of their target genes should provide insights on the inner workings of other model systems in which *cis*-regulatory elements are less tractable, including higher chordates.

ACKNOWLEDGEMENTS

We thank Dr. Yutaka Nibu and the members of the Di Gregorio and Nibu labs for valuable comments on the manuscript and Dr. Hiroki Takahashi (NIBB, Japan) for the *Ci-Thx6b* cDNA clone. We are grateful to Mami Takeda and Gary Esses for technical help. This work was supported by grant 1-FY08-430 from the March of Dimes Birth Defects Foundation and by NIH/NICHD grant R01HD050704 to A.D.G. J.E.K. was supported in part by a Jacques Cohenca pre-doctoral fellowship from the Weill Graduate School of Medical Sciences; I.O.-I. was supported in part by a post-doctoral fellowship from the Uehara Memorial Foundation (Japan). A.D.G. is an Irma T. Hirschl Scholar.
REFERENCES


FIGURE LEGENDS

Figure 1. Transcriptional changes in response to ectopically expressed Ci-Macho1, Ci-Tbx6b and Ci-Tbx6c.

Microphotographs of control and transgenic Ciona intestinalis embryos hybridized in situ with digoxigenin-labeled antisense RNA probes. Note that in some cases, due to mosaic incorporation of the transgene(s), not all the cells of the same lineage show staining. The percentages of transgenic embryos displaying the patterns shown, and the total number of embryos scored, are indicated at the bottom of each panel. (A, B, D, F, H, J, L) Mid-tailbud embryos; individual embryos are oriented with anterior to the left, dorsal up. (A, B, G, H, M) Low-magnification microphotographs of embryos from representative experiments. (C, E, G, I, K, M-O, Q) 110-cell stage embryos, vegetal views. (A-P) Effects of the misexpression of Ci-macho1 on Ci-Tbx6b and Ci-Tbx6c expression. (A, C, D) Control embryos expressing zygotic Ci-macho1 in germ-line precursors (yellow arrowheads), sensory vesicle and nerve cord (blue arrowheads). (B, E, F) Bra>macho transgenics efficiently misexpresses Ci-macho1 in notochord (red arrowheads) and mesenchyme cells (purple arrowheads) at early (E) and late stages (F). (G, H) Bra>macho transgenics hybridized with the Ci-Tbx6b probe. (I, J) Ci-Tbx6b is expressed at the 110-cell stage in muscle precursors (I; orange arrowheads) but is no longer expressed by the early tailbud stage (J; see also Fig. S1A-C). (K) Ectopic expression of Ci-Tbx6b is detected in the notochord of 110-cell stage Bra>macho transgenics (red arrowhead). (L) Mid-tailbud Bra>macho embryos show ectopic expression of Ci-Tbx6b only in mesenchyme (purple arrowhead). (M) Bra>macho transgenics hybridized with the Ci-Tbx6c probe. (N) Expression of Ci-Tbx6c in control
embryos is detected in a subset of muscle precursors (orange arrowheads). (O) In 110-cell Bra>macho embryos, ectopic expression of Ci-Tbx6c is detected in notochord precursors (red arrowheads). (P) Changes in gene expression in 110-cell stage Bra>macho embryos versus control embryos, as monitored by qRT-PCR. (Q) Left panel: 110-cell stage Bra>Tbx6b embryo hybridized with the Ci-Tbx6c probe. Right panel: 110-cell stage Bra>Tbx6c embryo hybridized with the Ci-Tbx6b probe. Ectopic notochord staining is indicated by red arrowheads. Scale bars = 50 μm.

Figure 2. Identification of the early and late Ci-Tbx6b CRMs and prediction of a related muscle CRM in the Ci-Tbx6c locus.

(A) Schematic representation of the Ci-Tbx6b locus as shown in the JGI Ciona intestinalis genome browser (Dehal et al., 2002; http://genome.jgi-psf.org/Cioin2/Cioin2.home.html). Dark blue boxes and dashed lines indicate exons and introns, respectively. A yellow rectangle represents the 2.4-kb fragment that was analyzed. Green boxes represent putative Ci-Macho1 binding sites; black arrows indicate their orientations. An orange rectangle depicts the proximal region of the 2.4-kb fragment. Bottom: alignment of the Ci-Macho1 sites with the consensus sequence identified by Yagi et al. (2004). Highly conserved nucleotides are boxed in red. (B-J) Embryos electroporated with either the 2.4-kb Ci-Tbx6b CRM (B-D), the 862-bp fragment (E-G) or the 2.4-kb CRM carrying mutations in all Ci-Macho1 binding sites (H-J) were fixed at the 32-cell (B, E, H), 110-cell (C, F, I), or early neurula (D, G, J) stage and hybridized in situ with a LacZ probe. Stained blastomeres are labeled only on one side of the embryos (Conklin, 1905; Meinertzhagen and Okamura, 2001). (K) The 266-
bp proximal region of the *Ci-Tbx6b* CRM (orange in panel A) contains putative binding sites for CREB (blue diamond), T-box (red oval) and bHLH (“AC” core E-box; pink triangle) transcription factors. (L-N) Embryos carrying the 266-bp *Ci-Tbx6b* transgene in (K), hybridized in situ with a LacZ probe. (O-Q) Mutation analysis of the *Ci-Tbx6b* proximal CRM. A 112-bp subfragment of the 266-bp CRM, containing the 3 putative binding sites shown in (K), was cloned upstream of the *Ci-FoxA-a* basal promoter (pink box) and subjected to truncation (data not shown) and mutation analysis. (O) Low-magnification image of mid-tailbud embryos electroporated with the wild-type 112-bp *Ci-Tbx6b* CRM. (P) Mid-tailbud embryos electroporated with the 112-bp *Ci-Tbx6b* CRM carrying a mutation in the Tbx6b/c core binding site (TCACAC->TCAaca). (Q) Mid-tailbud embryos electroporated with the 112-bp *Ci-Tbx6b* CRM carrying a mutation in the E-box sequence (CAGTTG->actggt). Insets in (O-Q) show transgenic embryos fixed at the neurula stage. (R) Graph showing the percentage of stained mid-tailbud embryos in (O-Q); blue columns: muscle staining; purple: mesenchyme. The number of embryos scored for each construct is shown in white in (O-Q). (S) Organization of the *Ci-Tbx6c/Ci-Tbx6b* genomic locus. The predicted *Ci-Tbx6c* CRM (blue rectangle) contains one Ci-Machol binding site, one T-box site and an AC-core E-box. (T) Late-tailbud embryo electroporated with the *Ci-Tbx6c* CRM, showing activity in muscle (orange arrowhead) and mesenchyme (purple arrowhead). Panels B, E, H, L: vegetal views; C, F, I, M: dorsal/vegetal views; D, G, J, N: dorsal views, with anterior to the top. Scale bar = 50 μm.

**Figure 3.** The *Ciona Fibrillar Collagen-1 (CiFCol1)* gene and its CRMs.
(A) Mid-tailbud Ciona embryo hybridized in situ with the CiFColl probe. Arrowhead color code: blue, CNS; purple, mesenchyme; red, notochord; orange, muscle; yellow: endoderm. Expression at other stages is available online: http://ghost.zool.kyoto-u.ac.jp/cgi-bin3/photoget2.cgi?CLSTR00093 (Satou et al., 2001). (B-E) 110-cell stage embryos hybridized with the CiFColl probe. (B) Control embryo. (C-E) Embryos carrying the transgenes indicated on the top right. Ectopic notochord staining is highlighted by red arrowheads. (F) Partial representation of the CiFColl gene locus. The yellow box depicts the 2.2-kb cis-regulatory region that fully recapitulates the CiFColl expression pattern; a blue and an orange box indicate the distal and proximal muscle CRMs, respectively. Inset: mid-tailbud embryo electroporated with the 2.2-kb region. (G,H): Microphotographs of X-Gal-stained Ciona embryos electroporated with the 400-bp and 230-bp regions shown in (F), cloned upstream of the Ci-FoxA-a basal promoter (pink box in the schematics). The distal region (G) contains 2 E-boxes (pink triangles); the proximal region (H) contains 4 putative Tbx6b/c binding sites (red ovals) and one putative Ci-Macho1 site (green square). (I-R): Embryos electroporated with either the distal (I-M), or the proximal muscle CRM (N-R), fixed at the 32-cell (I,N) 110-cell (J,O), early tailbud (L,Q) and mid-tailbud (M,R) stage and hybridized with a LacZ probe. (K,P): Early neurula embryos electroporated with either the distal (K) or the proximal CRM (P), stained with X-Gal.

Figure 4. Effects of the ectopic expression of Ci-Tbx6b and Ci-Tbx6c on the CiFColl muscle CRMs.
(A-C) *Ciona* embryos electroporated with the *CiFColl* proximal CRM, alone (A) or in combination with either *Bra>Tbx6b* (B) or *Bra>Tbx6c* (C). Arrowhead color-code: red, notochord; orange, muscle; purple, mesenchyme. (D): Graph showing the percentages of embryos displaying staining in muscle (blue columns) and in notochord (red columns) in the experiments shown in (A-C), as well as the baseline staining attributable to the vector. (+) and (-) signs indicate the presence or absence, respectively, of either misexpression construct. Note that the height of the error bars on the red columns reflects the sporadic occurrence of the activation of the empty vector in notochord cells, which is observed only in some batches of embryos.

**Figure 5. A working model for the muscle cis-regulatory hierarchy in Ciona.**

The regulatory interactions identified by this study are summarized and plotted against the developmental time-course of *Ciona* embryogenesis at 18°C (black vertical bar on the left side). A horizontal green bar represents the maternally-deposited *Ci-macho1* transcripts, green squares depict the Ci-Macho1 protein. Horizontal bars symbolize the cis-regulatory regions of *Ci-Tbx6b* and *Ci-Tbx6c* (yellow) and *CiFColl* (orange); transcription start sites are indicated by horizontal arrows. Transcripts are schematized as vertical bars on the right. Transcription factors are depicted as described in Figure 2.

**Figure S1.** *Ci-Tbx6b* is efficiently expressed in notochord cells by the *Ci-Bra* promoter region.
(A-C) Whole-mount control *Ciona* embryos fixed at the stages indicated above each panel and hybridized in situ with a *Ci-Tbx6b* digoxigenin-labeled antisense RNA probe. (D-F) Embryos electroporated at the 1-cell stage with the *Bra>Tbx6b* construct, then fixed at the stages indicated above each panel and hybridized in situ with the *Ci-Tbx6b* probe. Wild-type (wt) embryos show expression of *Ci-Tbx6b* in muscle precursors at the 110-cell and gastrula stage (A,B), but not at the tailbud stage (C), consistent with what was previously reported (Takatori et al., 2004). Embryos electroporated with the *Bra>Tbx6b* construct expressed *Ci-Tbx6b* in both the tail muscle (the endogenous gene) and the notochord (the transgene) at the 110-cell and at the gastrula stage (D, E), while at the tailbud stage only the notochord expression directed by the transgene was detected (F). In all panels, the notochord and its precursors are outlined by a dashed red line. Note that not all the notochord cells show staining, due to mosaic incorporation of the transgene. Embryos were photographed after addition of small amounts of 80% glycerol.

Figure S2. Identification of the minimal sequences required for the activity of the *Ci-Tbx6b* CRM in muscle cells.

(A) Schematic representation of the *Ci-Tbx6c/Ci-Tbx6b* genomic locus according to the JGI *Ciona intestinalis* genome browser (Dehal et al., 2002; [http://genome.jgi-psf.org/Cioin2/Cioin2.home.html](http://genome.jgi-psf.org/Cioin2/Cioin2.home.html)). Blue boxes and dashed lines indicate exons and introns, respectively. The yellow bar depicts the 2.4-kb genomic fragment that was cloned and analyzed. (B) Truncation analysis of the 2.4-kb fragment. The shaded region indicates the minimal sequence sufficient to direct expression in muscle cells in a heterologous context (i.e., upstream of the *Ci-FoxA-a* basal promoter, pink bar). Selected
constructs were also attached to the endogenous *Ci-Tbx6b* basal promoter (yellow small bar) and assayed for activity. The presence and intensity of the muscle staining produced by each construct are noted to the right (see explanation below). (C) A representative mid-tailbud embryo which was electroporated with the 2.4-kb *Ci-Tbx6b* fragment at the 1-cell stage, then fixed and stained with X-Gal at the mid-tailbud stage. Orange and purple arrowheads indicate muscle and mesenchyme cells, respectively. Anterior is on the left, dorsal on top. (D, E) Low-magnification microphotographs of *Ciona intestinalis* embryos electroporated with either the 2.4-kb *Ci-Tbx6b* CRM (D) or the *Ci-Tbx6b* basal promoter (E). Embryos representative of our empirical “+” and “-” ranking system are highlighted. (F) Graph of a representative experiment reporting the percentage of embryos showing staining in muscle (orange columns) and in mesenchyme (grey columns) when electroporated with either the 2.4-kb *Ci-Tbx6b* CRM (left), the *Ci-Tbx6b* basal promoter (center), and the empty pFBΔSP6 vector, containing the *Ci-FoxA-a* basal promoter (right).

**Figure S3. In vitro binding of the sequences required for early and late muscle activity of the *Ci-Tbx6b* CRM by Ci-Macho1, Ci-Tbx6b and Ci-Tbx6c.**

(A) Electrophoretic mobility shift assays (EMSA) with radiolabeled probes containing the three putative Ci-Macho1 binding sites (see Fig. 2A) and a His-Ci-Macho1 fusion protein. Lanes contained: (1) free radiolabeled probes containing each of the three putative Ci-Macho1 binding sites; (2) radiolabeled probes incubated with the His-Ci-Macho1 fusion protein; (3) the complexes described in (2), incubated with the unlabeled oligonucleotides at 5 µM final concentration (competitors). (B) EMSA with a
radiolabeled probe containing the T-box binding site incubated with either the GST-Ci-
Tbx6b (lanes 2-4) or the GST-Ci-Tbx6c (lanes 5-7) fusion protein. Lanes contained: (1)
free radiolabeled probe containing the T-box site; (2) radiolabeled probe incubated with
the GST-Ci-Tbx6b fusion protein; (3-4) the complex described in (2) incubated with
unlabeled wild-type (wt) (3) and a mutant competitor where the T-box site has been
ablated (4), at 5 µM final concentration. (5) radiolabeled probe incubated with the GST-
Ci-Tbx6c fusion protein; (6-7) the complex described in (5) incubated with unlabeled
wild-type (wt) (6) and mutant competitor (7).

Figure S4. Differential temporal activity of the *CiFCol1* muscle CRMs.

(A, B) Simplified schematic representations of the muscle precursors in a 110-cell
*Ciona* embryo (A, dorsal/vegetal view) and of the 18 muscle cells per side that they give
rise to in an early tailbud embryo (B, lateral view). In (A), the blastomeres that will give
rise also to cells of the CNS are colored in blue and orange (A8.16), and those that will
give rise also to mesenchyme cells and trunk ventral cells are colored in purple and
orange (B7.5 pair). In (B) only the muscle cells of the tailbud are shown. Adapted from
Satoh, 1994. (C) Activity of the distal (blue columns) and proximal (orange columns)
*CiFCol1* CRMs in muscle cells was determined at consecutive developmental stages by
counting the cells stained in each positive embryo, taking into account mosaic
incorporation of the plasmids. The average number of positive cells found in at least
50% of the scored embryos was plotted against the number of muscle cells (beige
columns), which is invariant.
**Figure 2**

(A) Genetic map showing the location of Machol sites A, B, and C, with LacZ expression patterns in 32-cell, 110-cell, and early neurula stages.

(B) Expression pattern of Ci-Tbx6b in 32-cell stage, with percentages of B6.4, B8.8, B8.7, and B8.16 indicated.

(C, D) Expression patterns in 110-cell and early neurula stages, respectively, with percentages for each condition.

(E) Expression pattern in neurula stage, with a dark spot indicating high expression.

(F, G) Additional neurula expression patterns, showing variation in expression levels.

(H) Expression pattern in 32-cell stage, with text indicating Machol consensus sequence.

(I) Expression pattern in 110-cell stage, with text indicating Machol site sequences.

(J) Expression pattern in early neurula stage, with text indicating Machol site sequences.

(K) Genetic map showing Tbx6b/c consensus sequence and expression pattern in 32-cell, 110-cell, and early neurula stages.

(L, M) Expression patterns in 32-cell and 110-cell stages, respectively, with percentages for each condition.

(N) Expression pattern in early neurula stage, with text indicating Tbx6b CRM T-box sequence.

(O) Neurula stage expression patterns, with text indicating percentages.

(P, Q) Additional neurula expression patterns, showing variation in expression levels.

(R) Bar graph showing percentage of stained embryos for Tbx6b WT, Tbx6b T-boxM, and Tbx6b E-boxM.

(S) Genetic map showing Machol site and T-box site sequences, with LacZ expression patterns in 293 bp region.

(T) Neurula stage expression pattern with arrow indicating expression location.