

## Non-Homologous Structured CRMs from the *Ciona* Genome

ALBERT ERIVES

### ABSTRACT

The internal functional organization of *cis*-regulatory modules (CRMs) lies at the heart of our understanding the mode and tempo of gene regulatory evolution as well as practical efforts at deciphering and annotating genomic sequences. In an open-ended search for loose clusters of known mesodermal enhancer motifs in the *Ciona intestinalis* genome, I discovered the existence of a class of highly organized CRMs in otherwise unrelated genes expressed early in development. Each such CRM is composed of distinct motifs located at specific positions along ~160 bp of DNA sequence, and is able to drive expression in distinct mesodermal compartments descended from the B4.1 blastomere. The majority of the loci bearing these B4.1-specific modules encode important early mesodermal transcription factors at the *snail*, *paraxis*, and *tbx6* orthologous loci of this invertebrate chordate system. These unrelated genes encode members of the C2H2 zinc-finger, bHLH, and T-box transcription factor families, and likely serve as a chordate-specific *trans*-code for paraxial mesoderm. One other similarly organized enhancer was discovered in the *TNC3* muscle structural locus. These results suggest that organization of binding sites over the length of the enhancer sequence is a critical aspect of gene regulatory biology. The extent to which this is a general principle will facilitate our ability to identify, decipher, and categorize the regulatory functions contained in whole genome assemblies.

**Key words:** *Ciona*, CRMs, genome annotation, paraxial mesoderm, regulatory genomics, transcriptional enhancers.

### 1. INTRODUCTION

**C***is*-REGULATORY MODULES (CRMs), including the large class known as transcriptional enhancers, are important DNA sequences that control one context-specific aspect of a gene's overall transcriptional program (Arnone et al., 1997; Gluzman et al., 1983). Each gene in a genome may be driven by multiple, independently acting CRMs, which collectively drive complex and robust patterns of gene expression. As such CRMs are critical control points determining the transcriptional states of a cell.

The internal functional organization of CRMs lies at the heart of our understanding the mode and tempo of gene regulatory evolution as well as practical efforts at deciphering and annotating genomic sequences.

Each CRM is typically composed of multiple binding sites clustered over a distance of ~300 bp. According to an organizational hypothesis of enhancer structure, the nature and arrangement of these binding sites determine how specific transcription factors will dynamically interact on these sequences and drive a highly nuanced program of gene expression. An extreme model of a sterically constrained “enhancesome” protein complex whose formation is directed by a DNA template of organized elements has been didactically contrasted to a billboard model of “information” display in which elements are loosely organized and read individually by the basal transcriptional machinery (Armosti et al., 2005). Theoretical and experimental analyses of these hypothetical models depend heavily on access to full genome sequence plus an ability to conduct transgenic reporter analyses of regulatory DNA sequences. One such system is the simple ascidian chordate, *Ciona intestinalis*. Ascidiates have long held a key position in embryology as a model system for studying the developmental determination of cell lineages. Part of the reason for this is the distinctive, cell-type-specific cell morphologies, such as those of the tail muscle cells of the tadpole stage as well as their mosaic mode of development (Chabry, 1887; Conklin, 1905). Another reason is that ascidiates are invertebrate chordates and their study has been expected to yield what Darwin called “a clue to the source whence the Vertebrata have been derived” (Darwin, 1871). For these reasons, sequencing efforts have included a complete determination of the genome of *Ciona intestinalis*, a cosmopolitan ascidian species located in marinas around the world (Dehal et al. 2002).

We have previously used the *Ciona intestinalis* system to study the regulation of the *snail* ortholog, *Ci-sna*, which is expressed in the mesendodermal compartment that is subsequently fated to become mesoderm (Erives et al. 1998, 2000). *Snail* expression precedes gastrulation across animals and is likely to be a critical component of the early gene regulatory network that specifies the embryonic germ layers of ectoderm, endoderm, and mesoderm. As part of these studies, we have focused on defining functional *cis*-regulatory DNA elements with the understanding that such work can facilitate the identification of early embryonic regulons via whole-genome queries. Additionally, such work can provide an example corpus for testing computational methods used in the annotation of the regulatory components of genomes.

The first *Ci-sna* study identified a cluster of specific E-box motifs that defined an early B4.1 enhancer immediately upstream of the *Ci-sna* basal promoter (Erives et al., 1998). This enhancer imparts similar activity to reporter genes fused to heterologous promoters. A follow-up *Ci-sna* study identified a second set of sites in the same enhancer that resembled T-box domain binding sites and identified one potential *trans* factor with maternal expression (Erives et al., 2000). This mode of regulation in the *Ciona*, an ascidian of the order Enterogona, has key differences with the site composition and organization of muscle structural genes documented in *Halocynthia roretzi*, an ascidian in the distantly related order Pleurogona (Hikosaka et al., 1994; Erives et al., 2001). Current work by different labs are working on which factors, such as bHLH heterodimers for the E-boxes, and which of many potential maternal and zygotic T-box containing factors and other types of factors, operate at these sites throughout embryogenesis.

Here I use the recently sequenced genome to computationally identify other B4.1 enhancers utilizing combinations of the two *cis*-elements found in the *Ci-sna* enhancer: AC-core E-boxes and T-box domain sites. Using a transient transgenic reporter assay in developing *Ciona* embryos, I find that the only functional modules with these sites correspond to three other genomic sites with an almost identical organization of elements despite an absence of homology in the sequences intervening these sites. Naturally occurring clusters of these same sites that lack this characteristic organization were not found to possess any activity in the same assays. These results show that a rigid organization of elements is necessary for function in at least one set of co-regulated developmental enhancers.

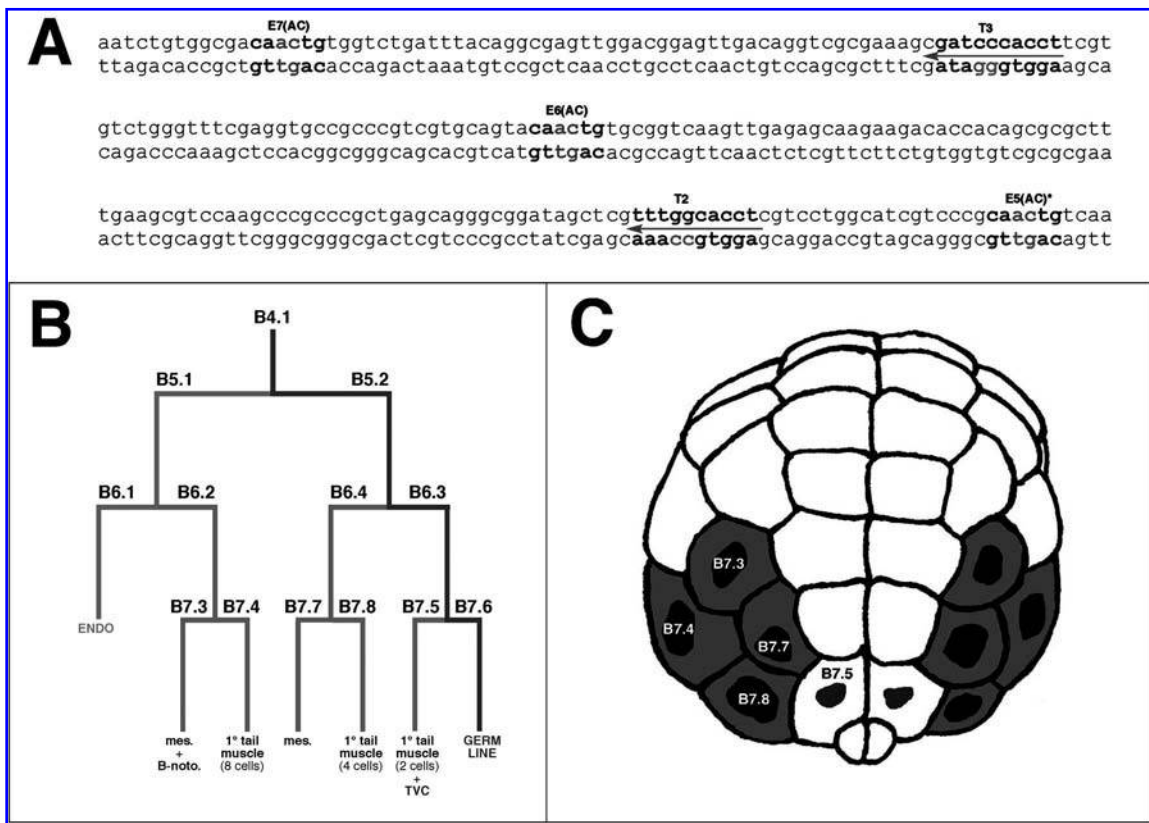
## 2. RESULTS

The *Ci-sna* B4.1 enhancer is active beginning around the 32-cell stage in B4.1 descendants destined to give rise to primary tail muscle cells, trunk mesenchyme, and trunk ventral cells. This enhancer has three AC-core E-boxes (5'-CAACTG), herein called E(AC) motifs, although one of these sites is polymorphic (data not shown). Mutation of the core E-box sequence to other types abolishes activity in the primary tail muscle cells, unless the core is mutated to a GC-core (5'-CAGCTG), in which case expression is augmented (Erives, 1999; Erives et al., 1998). A GC-core E-box corresponds to the binding site for the muscle terminal differentiation gene *MyoD* (Davis et al., 1990; Ma et al., 1994; Weintraub et al., 1990).

This enhancer also contains two T-box binding sites (5'-AGGTGNNNA) and mutations of these also abolish expression (Erives et al. 2000). These sites are co-clustered in a length of DNA spanning 240 bp (Fig. 1) and a similar sized fragment of 262 bp containing these sites recapitulates the early expression of *Ci-snail* when placed in front of a heterologous promoter (Fig. 1).

I therefore searched the genomic sequence for unstructured co-clusters of these two motifs (Table 1). I looked for two to three AC-core E-boxes co-clustered with two to three T-box binding sites of either low (5'-AGGTGNNAA) or high (5'-GGTGNNNA) levels of degeneracy. I looked for these sites in any arrangement that fit within either 240 bp or a wider window of 320 bp. While the 240 bp co-cluster of E(AC) and low-degeneracy T-box sites produced the shortest or most specific list of genes, this list included none of the many genes known to be expressed in the tail muscle cells. The other more degenerate searches using either a wider window of DNA sequence (320 bp) or a more degenerate T-box binding site (5'-GGTGNNNA) were not significantly more sensitive at targeting muscle loci and yet they were still less specific by hitting up two ~1900 loci.

Nonetheless, I carefully examined these hits and identified 20 clusters that were within 20 kb of putative and/or known muscle specific loci. These were individually fused to a *Ci-fkhd-lacZ* reporter gene and used to electroporate fertilized, dechorionated, *Ciona intestinalis* eggs as previously described (Erives et al., 1998). The *Ci-fkh* (FoxA) basal promoter fused to *lacZ* possesses no activity on its own. The electroporated single cell zygotes were then grown overnight to at least the early tailbud embryonic stage, fixed, and X-gal stained for  $\beta$ -gal activity. Each construct was tested at least three times using different batches of ascidian embryos collected from both the East Coast (Woods Hole MBL) as well as the West



**FIG. 1.** The B4.1 mesodermal enhancer of the *Ci-snail* locus. (A) This 240-bp enhancer contains two types of sites, AC-core E-boxes or E(AC) motifs (5'-CAACTG) and T-box domain binding motifs (5'-AGGTGNNNA) co-clustered over a distance of ~240 bp. Both of these sites are necessary for function (Erives et al., 1998; Erives and Levine, 2000). This enhancer drives expression in mesodermal derivatives of the B4.1 lineage (B) but only after dividing away from presumptive germ line blastomere B7.6. (B) Lineage descendents of the B4.1 blastomere. (C) Diagram of a 32-cell *Ciona intestinalis* embryo depicting early *Ci-snail* enhancer driven *lacZ* expression for blastomeres depicted in B.

TABLE 1. GENOME QUERY PATTERNS AND MATCHES

<i>Genome query pattern</i>	<i>Loci</i>
UNSTRUCT: 2-3(CAACTG) + 2-3(GGTGNNNA) in 240 bp	1014 loci
UNSTRUCT: 2-3(CAACTG) + 2-3(AGGTGNNAA) in 240 bp	174 loci
UNSTRUCT: 2-3(CAACTG) + 2-3(GGTGNNNA) in 320 bp	1918 loci
UNSTRUCT: 2-3(CAACTG) + 2-3(AGGTGNNAA) in 320 bp	332 loci
CAACTG 79–82 bp TTGRACCK 39–44 bp CTCACSTG	<i>Ci-sna</i> , <i>Ci-pxs</i>
CAACTG 79–83 bp TTRRCACMK 39–44 bp CTYASBTG	<i>Ci-sna</i> , <i>Ci-pxs</i> , <i>Ci-Tbx6b</i>
CAACTG 79–83 bp TTRDCACMK 39–48 bp YTYASBWS	<i>Ci-sna</i> , <i>Ci-pxs</i> , <i>Ci-Tbx6b</i> , <i>Ci-TNC3</i>

This is a list of the different patterns that I used to search the *Ciona* genome sequence and the number of matches found for each pattern. “UNSTRUCT” refers to unstructured queries in which the indicated motifs were allowed to occur in any position and in any orientation within the indicated sequence window. The last three structured queries involved searching for the indicated motifs in the indicated orders, orientations, and spacings. Only a very small number of loci are identified with the structured queries. All such matching loci are listed by name.

Coast (California). Except for one construct described below, all of these initial clusters failed to display any activity (data not shown). To help rule out the possibility of an unknown incompatibility with the former core promoter, a few of these site clusters were secondarily cloned in front of another heterologous core promoter (*Ci-bra*). However, changing core promoters failed to allow expression of activity in such clusters that previously failed to drive *lacZ* activity.

I found that one such cluster, located at the *Ci-paraxis* locus is expressed in a B4.1 descendent pattern and interestingly shares a similar configuration with the *Ci-sna* enhancer as explained below. This reporter drives expression a little later than *Ci-sna* B4.1 enhancer and is exclusive to the primary tail muscle cells without any expression in the trunk mesenchyme cells (Fig. 2A). *Ci-paraxis* encodes the *Ciona* ortholog of a well-known paraxial mesodermal bHLH factor that is chordate specific, and which in vertebrates is expressed in unsegmented paraxial mesoderm prior to somitogenesis (Burgess et al. 1995, 1996).

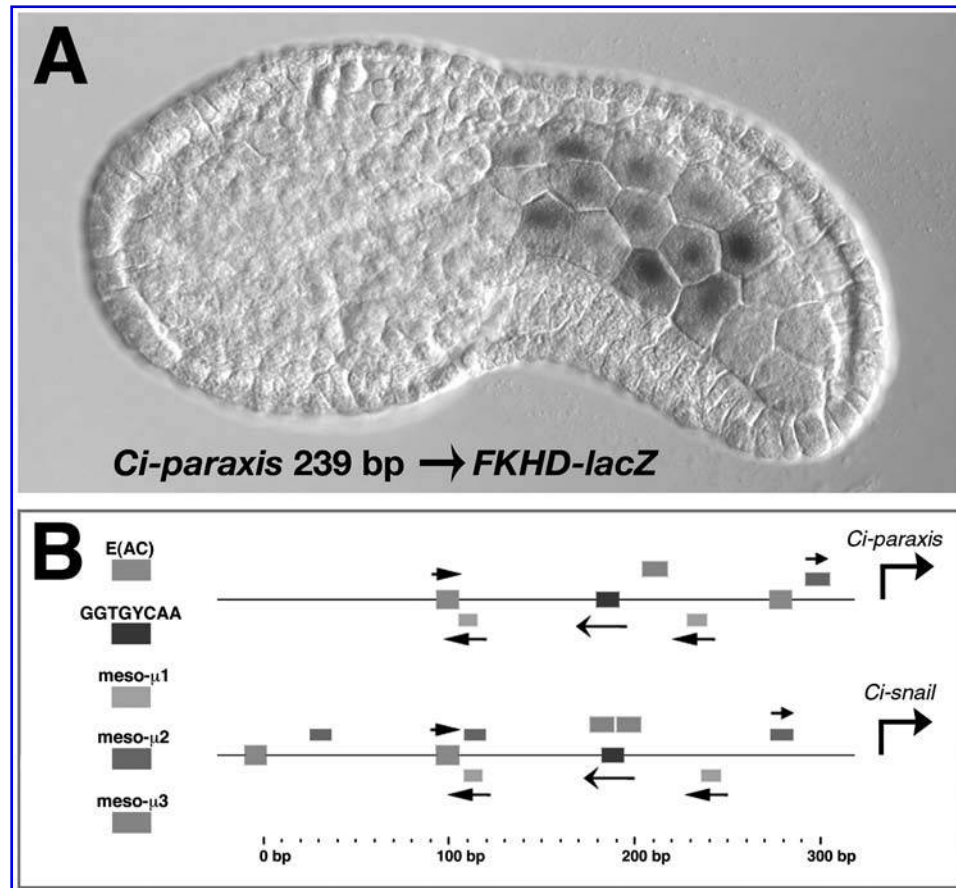
Because the *Ci-paraxis* enhancer activity was rather striking in light of the negative results with all of the other clusters, a thorough sequence analysis was conducted for any additional unknown motifs that could explain the difference between false-positive clusters lacking activity and the “true-positive” clusters at the *Ci-sna* and *Ci-paraxis* locus. Such sites were discovered and classed into families (meso- $\mu$ 1, meso- $\mu$ 2 etc.). Because none of the other false-positive clusters contained these sites, a second search was conducted for triple clusters containing E(AC), a T-box binding site, and one of the new motifs. One such cluster was identified, cloned and fused to the *lacZ* reporter but it also failed to drive expression (data not shown).

Having exhausted the search for simple compact elements less than 10 bp, I next turned to look at whether the *Ci-sna* and *Ci-paraxis* enhancers shared any similar organizations of their constituent *cis*-elements. In fact, these two promoters shared alignment of both known *cis*-elements and some of the new motifs, although the intervening sequences were non-alignable (Fig. 2B).

This result suggested that the organization was critically important for function. I therefore conducted new genomic queries for organized sites while allowing the sites to be increasingly degenerate. In this manner, I was able to identify two more such clusters, also present in the extended upstream promoter region, in the *Ci-Tbx6b* and *Ci-TNC3* loci. The former locus is already known to be a tail muscle specific marker (A. Di Gregorio, manuscript in preparation), while the latter locus codes a troponin-related, muscle-specific structural product. This latter cluster drives expression in the trunk mesenchyme stage beginning in the early tailbud stage (Fig. 3A) and also drives expression in the mid-tailbud stage in primary tail muscle cells (Fig. 3B).

All four muscle-specific enhancers are found in the immediate upstream regions of classic TATAA-box core promoters (Fig. 3C). These enhancer sequences do not share sequence homology or alignment in the intervening sequences between the elements described here (Fig. 4). It is therefore likely that these un-related loci evolved their collection of sites in parallel or at least in an ancient chordate ancestor. Indeed, *paraxis* and *tbx6* loci are co-expressed and involved in patterning unsegmented paraxial mesoderm in vertebrates (Burgess et al., 1995, 1996; Chapman et al., 1996).

Increasing the degeneracy of these sites while maintaining the organization leads to many more query hits than could be tested (data not shown). Furthermore, these degenerate queries indicate an absence of

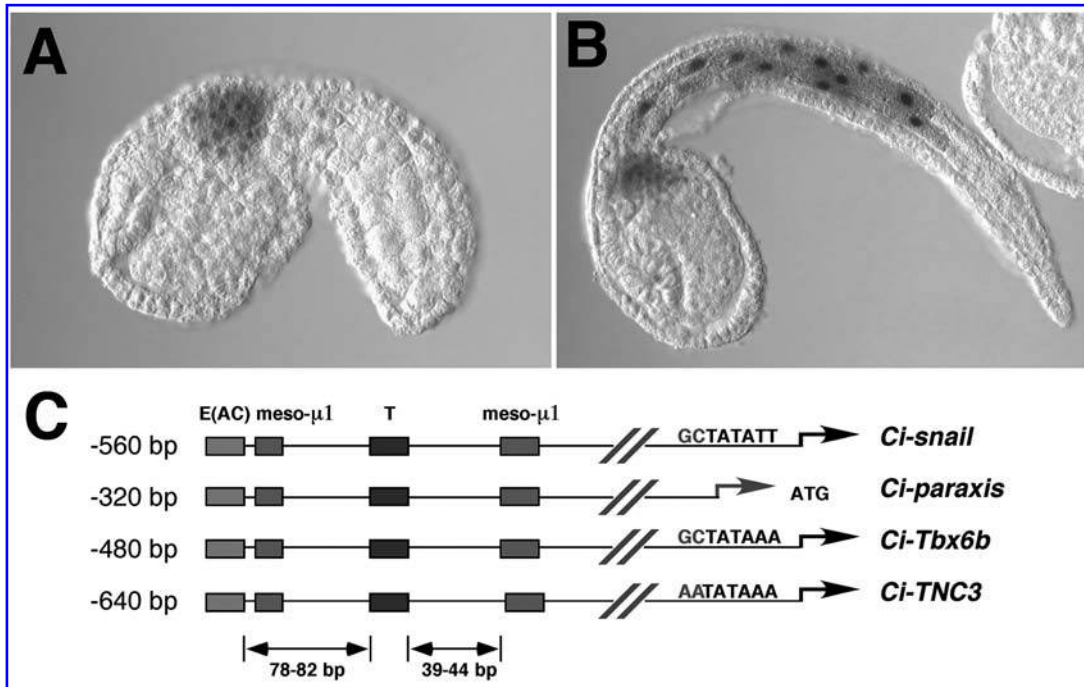


**FIG. 2.** A new B4.1 mesodermal enhancer at the *Ci-paraxis* locus. **(A)** A 239-bp fragment containing a similar cluster of E(AC) and T-box binding motifs is sufficient for driving expression of a *lacZ* transgene in primary tail muscle cells of the early tailbud stage embryo. **(B)** This fragment shares both additional sites and a common organization of these sites with the previously characterized *Ci-snail* enhancer.

sensitivity for mesodermal or tail muscle type genes. Thus, by searching the genome for similarly structured enhancers, I have been able to identify only the four enhancers. These appear to be the only loci in the genome regulated by this configuration of elements: (1) *snail*, which encodes a transcriptional repressor of endodermal and/or ectodermal cell fate; (2) *paraxis* and *tbx6b*, both of which encode transcriptional activators of mesodermal cell fates from the bHLH and T-box families, respectively; and (3) *TNC3*, whose presence in this august list may indicate that it encodes an early muscle structural gene that must be incorporated early in the physical construction of tail muscle cells.

### 3. DISCUSSION

In this study I show the existence of a class of highly-organized CRMs in otherwise unrelated genes expressed early in development. Each such CRM is composed of distinct motifs located at specific positions along ~160 bp of DNA sequence, and is able to drive expression in distinct mesodermal compartments descended from the B4.1 blastomere of *Ciona intestinalis* embryos. The majority of the loci bearing these B4.1-specific modules encode important early mesodermal transcription factors at the *snail*, *paraxis*, and *tbx6* orthologous loci of this invertebrate chordate system. These unrelated genes encode members of the C2H2 zinc-finger, bHLH, and T-box transcription factor families, respectively, and likely serve as a chordate-specific *trans*-code for paraxial mesoderm. One other similarly organized enhancer was discovered in the *TNC3* muscle structural locus. These results provide an example of a functionally equivalent class

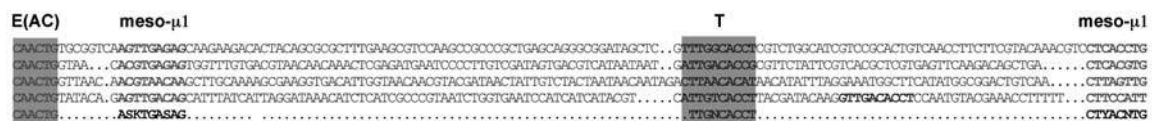


**FIG. 3.** Genomic search for organized elements yields diverse B4.1-descendant enhancers. New B4.1-descendant enhancers at the *Ci-Tbx6b* and *Ci-TNC3* loci were discovered on the basis of shared organization of motifs seen in the *Ci-snail* enhancer. The *Ci-TNC3* enhancer drives expression in trunk mesenchyme cells of the early tailbud embryonic stage (A), and this expression expands to the primary tail muscle cells during the mid-tailbud embryonic stage (B). All of these enhancers are present as upstream extensions of TATAAA-box containing core promoters (C). All of these modules are also pointing in the same direction relative to transcription.

of *cis*-regulatory modules defined by similar patterns of expression, as well as by the similar organization of a common set of binding sites. The extent to which this is true for other equivalence classes of CRMs will critically inform our understanding of gene regulatory sequence evolution as well as the annotation of genomic sequences according to the native gene regulatory logic.

This study thus addresses the very interesting issue of the syntax of clustered *cis*-regulatory elements in the context of the developmental enhancers of a simple chordate. The conceptual models of structured CRMs that work as organizational scaffolds for enhancer complexes versus flexible clusters of binding sites has been useful for generating specific questions on enhancer biology (Arnosti et al., 2005). While this study may not reveal a property common to all CRM classes, it does provide a specific example of a class of highly structured enhancers, in which the organization appears to play a functional role.

Most of these highly structured B4.1 enhancers do not share multiple sites as found in the *Ci-snail* enhancer, which suggests diverse interpretations. For example, these additional sites may be the result of site turnover that have been replaced by compensatory organized sites elsewhere in the enhancer (Ludwig et al., 1998, 2000; Moses et al., 2006). Indeed, with additional sequence sampling, I have seen that one of the E(AC) sites is polymorphic in wild-type populations of *Ciona* (data not shown). A dynamic history of



**FIG. 4.** Highly organized B4.1 descendant enhancer sequences. The sequences for *Ci-snail*, *Ci-paraxis*, *Ci-Tbx6b*, and *Ci-TNC3* B4.1-descendant enhancer sequences are aligned. E(AC) and T-box domain binding motifs are highlighted as indicated as well as meso- $\mu$ 1 sites. Note the stereotypical linkage between these sites despite the lack of sequence homology in the spacer sequences.

such turn-over could therefore be the cause of much of the site clustering seen in many enhancers and could work to obscure the identification of an organized collection of motifs maintained in a subset of these sites. For example, low-affinity Dorsal sites in the organized neuroectodermal enhancers of *Drosophila melanogaster* have served to obscure the presence of a semi-organized collection of binding sites involving a critical Dorsal site linked to a CA-core Twist binding E-box, as well as a Su(H) motif overlapping a second Dorsal site (Erives et al., 2004). In a more recent study, we have shown that selection can adapt enhancer activity through precise changes in this basic organization (Crocker et al., 2008).

One recent study failed to find shared structural organization in enhancers by analyzing motif distribution in nine *Ciona* muscle structural genes (Brown et al., 2007). However, it is worth considering that this conclusion is based only on the absence of evidence for shared organization on an extremely small set of enhancers, which may or may not be co-regulated. In contrast, the study presented here constitutes concrete evidence that shared organization can be identified in some regulons. Such results hold much promise for the rapid computational annotation efforts in regulatory genomics.

From a regulatory genomics perspective, the existence of structured enhancers is not unexpected for two simple reasons. First, without selecting for additional functional sites, gene regulation would not be very specific to the loci that need to be regulated, and would not result in the phenomenon of differential gene regulation, which is the fundamental cornerstone of developmental biology. Thus, organization of binding sites is a useful parameter to block haphazard and maladaptive gene induction events at other loci. Such maladaptively regulated loci might be continuously produced via sequence drift creating functional sites that drive leaky expression. Second, the highly organized structure of DNA binding domains, together with the bulk of the non-DNA binding moiety, is likely to impart to these factors certain steric constraints in how they interact with other co-factors. These interactions are likely to be influenced by the organization of elements along the DNA strand. Because the spacing of sites across a typical enhancer or basal promoter is roughly on the same length scale as that of a nucleosome (160–240 bp), further work may reveal that site organization is also tied to nucleosomal positioning and/or re-modeling (Lowary et al., 1998; Segal et al., 2006; Thastrom et al., 1999, 2004; Wallrath et al., 1994; Widom, 2001). In one of the curious side notes of early gene regulatory studies, primarily conducted in tissue culture co-transfection assays, many animal activator binding sites were found to have very little or extremely modest activities on their own when cloned in front of a heterologous promoter. To achieve activation such sites were often multimerized in 5 to 10 copies. Such results could be explained if these early experiments lacked the critical organization that is documented here for one class of mesodermal enhancers.

Our future studies will also attempt to look at differences in activities amongst the B4.1 sub-lineages displayed by this set of structured explore enhancers. The *Ci-snail* enhancer can drive expression in early pre-gastrula B4.1 descendants destined to become either primary tail muscle or trunk mesenchyme. The *Ci-paraxis* enhancer can drive expression only beginning in early tailbud stage in the primary tail muscle stage. The *Ci-TNC3* can drive expression in early tailbud stage in the trunk mesenchyme, and in the primary tail muscle cells only beginning in mid tailbud stage. Thus, while the enhancers may share a basic organization of similar sites, there remains many interesting ways that evolution can customize these sequences through slight changes in spacing and/or the presence of tissue-specific repressors to further customize a specific mesodermal B4.1 expression pattern.

## 4. METHODS

### 4.1. Bioinformatics

A copy of the *Ciona* genome was obtained from Dan Rokshar and Mike Levine, and was used to conduct whole-genome searches of enhancer motifs using perl scripts to annotate a copy of the genome that was stripped of newline characters: “perl -pi -e ‘s/({320}REGEX1.{320})/\n REGEX2 \$1\n/g’ genome\_file,” where REGEX1 and REGEX2 correspond to regular expression pattern for a unique identifying flag. The UNIX grep tool was used to find all lines containing REGEX2 into a new file: “grep -E REGEX3 genome\_file > target\_file.” This procedure was repeated on the target file until a file was obtained with target sequences containing all of the desired collection of patterns. These sequences were then individually used to conduct BLAST searches to identify the corresponding genomic locus using the publicly-available

genome browser maintained by JGI (Joint Genome Institute). Identification of novel shared motifs was conducted using motif elicitation through expectation maximization (MEME) algorithms developed by Bailey and Elkan (Bailey et al., 1995, 2006). Analysis for shared organization of motifs identified by the MEME algorithm was conducted visually.

#### 4.2. Animal husbandry

Gravid animals were obtained from various commercial marinas throughout Southern and Northern California as well as from the Marine Biological Laboratory in Woods Hole, Massachusetts. Animals were housed in an aquarium containing circulating sea water refrigerated to 18°C and kept for no more than 4 weeks. Surgical dissections were used to obtain fresh eggs and sperm for fertilization in clean glass fingerbowls. These fingerbowls are maintained free of detergents and chemicals for their entire life and are washed only with copious amounts of sea water, followed by deionized water in between uses. Chemical dechorionations, and transgene electroporations were conducted exactly as previously described (Erives et al., 1998, 2000).

#### 4.3. DNA constructs

All DNA constructs were made by first amplifying the target sequence from genomic DNA and cloned into the pSP1.72-27 *lacZ* cloning vector containing either the *Ci-fkhd* or *Ci-Bra* core promoters as employed in previous experiments (Erives et al., 1998; Erives et al., 2000). All DNA constructs were obtained by growing 500-mL cultures in 2xYT media followed by purification by EtBr/CsCl<sub>2</sub> gradient band centrifugation. Special care was taken to obtain only the band corresponding to supercoiled, circularized DNA plasmid.

### ACKNOWLEDGMENTS

I would like to thank Paul Mineiro, who believed me that we could “grok” the human genome by first looking at the sea squirt genome. I thank Jody Biggs and the rest of the CodeGrok team. I thank Mike Levine for the use of his lab to clone out the new *Ciona* enhancers predicted at CodeGrok. I thank Anna Di Gregorio for discussing her work with the Tbx6 promoter region. I thank Justin Crocker for helpful comments on the manuscript. I thank the continued support of Dartmouth College—*Vox clamantis in deserto*.

### DISCLOSURE STATEMENT

No competing financial interests exist.

### REFERENCES

- Arnone, M.I., and Davidson, E.H. 1997. The hardwiring of development: organization and function of genomic regulatory systems. *Development* 124, 1851–1864.
- Arnosti, D.N., and Kulkarni, M.M. 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.* 94, 890–898.
- Bailey, T.L., and Elkan, C. 1995. Unsupervised learning of multiple motifs in biopolymers using EM. *Machine Learn.* 21, 51–80.
- Bailey, T.L., Williams, N., Mischel, C., et al. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373.
- Brown, C.D., Johnson, D.S., and Sidow, A. 2007. Functional architecture and evolution of transcriptional elements that drive gene coexpression. *Science* 317, 1557–1560.
- Burgess, R., Cserjesi, P., Ligon, K.L., et al. 1995. Paraxis: a basic helix-loop-helix protein expressed in paraxial mesoderm and developing somites. *Dev. Biol.* 168, 296–306.

- Burgess, R., Rawls, A., Brown, D., et al. 1996. Requirement of the paraxis gene for somite formation and musculoskeletal patterning. *Nature* 384, 570–573.
- Chabry, L.M. 1887. Contribution à l'embryologie normale et à la tératologie des ascidies simples. *J. Anat. Physiol. Norm. Pathol.* 23, 167–231.
- Chapman, D.L., Agulnik, I., Hancock, S., et al. 1996. Tbx6, a mouse T-Box gene implicated in paraxial mesoderm formation at gastrulation. *Dev. Biol.* 180, 534–542.
- Conklin, E.G. 1905. Mosaic development in ascidian eggs. *J. Exp. Zool.* 2, 146–223.
- Crocker, J., Tamori, Y., and Erives, A. 2008. Evolution acts on enhancer organization to fine-tune gradient threshold readouts. *PLoS Biol.* 6(11):e263.
- Darwin, C. 1871. *The Descent of Man, and Selection in Relation to Sex*. Princeton University Press Princeton, NJ.
- Davis, R.L., Cheng, P.F., Lassar, A.B., et al. 1990. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* 60, 733–746.
- Dehal, P., Satou, Y., Campbell, R.K., et al. 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298, 2157–2167.
- Erives, A. 1999. Appendix C: Transcriptional output of diverse E-box sequences. In: *Cis-Regulation of Two Gene Sets in the Ciona Embryo: The Tail Muscle Genes and the Notochord Genes*. University of California, Berkeley, CA.
- Erives, A., Corbo, J.C., and Levine, M. 1998. Lineage-specific regulation of the *Ciona* snail gene in the embryonic mesoderm and neuroectoderm. *Dev. Biol.* 194, 213–225.
- Erives, A., and Levine, M. 2000. Characterization of a maternal T-Box gene in *Ciona intestinalis*. *Dev. Biol.* 225, 169–178.
- Erives, A., and Levine, M. 2001. Cis-regulation of ascidian tail muscle genes. *Proc. First Int. Symp. Biol. Ascidians* 193–201.
- Erives, A., and Levine, M. 2004. Coordinate enhancers share common organizational features in the *Drosophila* genome. *Proc. Natl. Acad. Sci. USA* 101, 3851–3856.
- Gluzman, Y., and Shenk, T. 1983. *Enhancers and Eukaryotic Gene Expression*. Cold Spring Harbor Laboratory Cold Spring Harbor.
- Hikosaka, A., Kusakabe, T., and Satoh, N. 1994. Short upstream sequences associated with the muscle-specific expression of an actin gene in ascidian embryos. *Dev Biol* 166, 763–769.
- Lowary, P.T., and Widom, J. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* 276, 19–42.
- Ludwig, M.Z., Bergman, C., Patel, N.H., et al. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403, 564–567.
- Ludwig, M.Z., Patel, N.H., and Kreitman, M. 1998. Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* 125, 949–958.
- Ma, P.C., Rould, M.A., Weintraub, H., et al. 1994. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* 77, 451–459.
- Moses, A.M., Pollard, D.A., Nix, D.A., et al. 2006. Large-scale turnover of functional transcription factor binding sites in *Drosophila*. *PLoS Comput. Biol.* 2, e130.
- Segal, E., Fondufe-Mittendorf, Y., Chen, L., et al. 2006. A genomic code for nucleosome positioning. *Nature* 442, 772–778.
- Thastrom, A., Bingham, L.M., and Widom, J. 2004. Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning. *J. Mol. Biol.* 338, 695–709.
- Thastrom, A., Lowary, P.T., Widlund, H.R., et al. 1999. Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.* 288, 213–229.
- Wallrath, L.L., Lu, Q., Granok, H., et al. 1994. Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures. *Bioessays* 16, 165–170.
- Weintraub, H., Davis, R., Lockshon, D., et al. 1990. MyoD binds cooperatively to two sites in a target enhancer sequence: occupancy of two sites is required for activation. *Proc. Natl. Acad. Sci. USA* 87, 5623–5627.
- Widom, J. 2001. Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* 34, 269–324.

Address reprint requests to:

Dr. Albert Erives  
Department of Biological Sciences  
Dartmouth College  
Hanover, NH 03755

E-mail: Albert.Erives@Dartmouth.edu