Appendix C

Patchy Interspecific Sequence Similarities Efficiently Identify Positive *cis*-Regulatory Elements in the Sea Urchin

Chiou-Hwa Yuh, C. Titus Brown, **Carolina B. Livi**, Lee Rowen, Peter J. C. Clarke and Eric H. Davidson

(Published in 2002)

ABSTRACT

We demonstrate that interspecific sequence conservation can provide a systematic guide to the identification of functional *cis*-regulatory elements within a large expanse of genomic DNA. The test was carried out on the *otx* gene of *Strongylocentrotus purpuratus*. This gene plays a major role in the gene regulatory network that underlies endomesoderm specification in the embryo. The *cis*-regulatory organization of the *otx* gene is expected to be complex, because the gene has three different start sites (Li *et al*., Dev. Biol. 187, 253-266, 1997), and it is expressed in many different spatial domains of the embryo. BAC recombinants containing the *otx* gene were isolated from *Strongylocentrotus purpuratus* and *Lytechinus variegatus* libraries, and the ordered sequence of these BACs was obtained and annotated. Sixty kb of DNA flanking the gene, and included in the BAC sequence from both species, were scanned computationally for short conserved sequence elements. For this purpose we used a newly constructed software package assembled in our laboratory, "FamilyRelations." This tool allows detection of sequence similarities above a chosen criterion within sliding windows set at 20-50 bp. Seventeen partially conserved regions, most a few hundred bp long, were amplified from the *S. purpuratus* BAC DNA by PCR, inserted in an expression vector driving a CAT reporter, and tested for *cis*-regulatory activity by injection into fertilized *S. purpuratus* eggs. The regulatory activity of these constructs was assessed by whole mount *in situ* hybridization (WMISH) using a probe against CAT

mRNA. Of the 17 constructs, 11 constructs displayed spatially restricted regulatory activity, and six were in this test inactive. The domains within which the *cis*-regulatory constructs were expressed are approximately consistent with results from a WMISH study on *otx* expression in the embryo, in which we used probes specific for the mRNAs generated from each of the three transcription start sites. Four separate *cis*-regulatory elements that specifically produce endomesodermal expression were identified, as well as ubiquitously active elements, and ectoderm-specific elements. We confirm predictions from other work with respect to target sites for specific transcription factors within the elements that express in the endoderm.

***Key Words*: *otx* gene, FamilyRelations, *cis*-regulatory sequence, computational genomics, sea urchin embryo**

INTRODUCTION

The functional linkages of gene control networks terminate in the *cis*-regulatory elements of the network genes. Here is where the inferred architecture of such networks can be proved, disproved, or improved, i.e., by determining whether predicted sets of inputs into the important genes of the network exist, and if not what the inputs are. So, sooner or later, it is necessary to hunt down the relevant *cis*-regulatory elements in the expanse of genomic DNA that surrounds each gene. This is a major bottleneck in regulatory gene network analysis: the control sequences may occur anywhere within tens of kilobases (kb) of the transcription start site(s) upstream of the gene, downstream, or within its introns (reviewed by Davidson, 2001). Conventionally, this problem is approached by the tedious route of building and testing expression constructs that contain successive blind deletions of the DNA flanking the gene. Eventually this usually works, but sometimes it does not, as when the desired elements are jumbled together with other elements on a long fragment several kb in length, which then requires much further work to disentangle. Recently several labs including our own have begun to use more systematic, high throughput approaches to find *cis*-regulatory elements (e.g., Cameron *et al*., 2002; Sivasankaran *et al*., 2000). To this end an expression vector equipped with a promiscuous basal promoter that will service any (or most) *cis*-regulatory modules and will work in any domain of the embryo is used as a test system for regulatory activity. A scan for spatial expression can then be carried out, using either mapped genomic DNA fragments cloned into the expression vector or sufficient randomly generated fragments to provide coverage of the regions flanking the gene of interest. But this too is a great

deal of work, if only because of the large number of fragments that need to be tested in any blind procedure. On the scale required of a developmental gene network analysis, where some *cis*-regulatory knowledge of at least 10-20 key genes might be necessary, no blind procedure is really practical.

It is a well accepted view that *cis*-regulatory sequences are among the conserved sequence elements of the genome, a view supported both by logic and by numerous *ex post facto* observations. That is, once discovered, many regulatory elements have been shown to be conserved, with respect to their surrounding DNA sequences, in the genomes of species in which the gene that the element controls is used in a more or less similar way. Only recently has there begun to appear a literature on the use of interspecific sequence conservation as a means to find *cis*-regulatory elements in the first place, and this literature is so far mainly confined to the vertebrates. One kind of application is for the discovery of target sites for unknown transcription factors, within DNA fragments that have already been shown to harbor given regulatory modules. For example, a particular site that targets expression of the *foxa2* gene to the notochord was detected by comparison of several kb of chicken, mouse, and fish DNA (Nishizaki *et al*., 2001). But in principle a much more general and powerful application lies open. This is to use computational methods to compare long stretches of genomic DNA of one species with the orthologous region of another species' genome, and thereby simultaneously to detect all conserved patches of sequence that might contain *cis*-regulatory elements. This strategy was explored by Loots *et al*. (2000) in a comparative scan of over a megabase of orthologous human and mouse DNA containing interleukin genes. Ninety conserved sequence elements were noted, and 15 were examined further, though only by

computational means: most of them were also found to be conserved in the orthologous genomic regions of dogs, rabbits, cows, etc., and a few even in *Fugu*. One of the conserved sequence elements, 401 bp in length, was experimentally tested by transgenic analysis in mice. It was found to possess regulatory activity for three genes encoding Interleukin proteins, located within a 120 kb region that also includes the conserved sequence. *cis*-Regulatory elements of *hox* genes are also known to be conserved within vertebrates (e.g., Nonchev *et al*., 1996; Carr *et al*., 1998), and even to the non-vertebrate chordate amphioxus (Manzanares *et al*., 2000). In another example in which interspecific sequence conservation was used as a guide, a *cis*-regulatory module of the *hoxb11* gene was recovered by comparison of orthologous regions of *Fugu* and mouse genomes, and then demonstrated to work by gene transfer (Aparicio *et al*., 1995).

Unselected DNA sequence, i.e., introns or intergenic DNA not included in regulatory elements changes continuously in evolution, at a basal rate that is specific for each animal clade (Britten, 1986). But the various species constituting each animal clade, i.e., families, orders, and even classes, display very similar morphologies even though many millions of years have elapsed since their last common ancestor. Essentially this means that they have continued to use (most of) their developmental regulatory genes in much the same way as did the ancestor of the clade. The *cis*-regulatory sequences that control these developmentally important genes diverge at much less than the unselected basal rate, and hence their relative conservation. So in principle it should always be possible to find a species pair in which the contrast between the rate of *cis*-regulatory sequence divergence and the basal rate of unselected DNA sequence divergence affords a distinction sufficient to reveal the *cis*-regulatory elements, except perhaps for clades

populated by very few living species. But the choice of the species pair that will work is not obvious in advance, at least outside the vertebrates, and is likely to differ to some extent gene to gene. If the pair chosen is too close, the background "noise" will be too high, and if it is too distant the regulatory elements may have changed in function, and therefore also in sequence.

Here we show that for the *otx* gene, comparison of the orthologous *S. purpuratus* and *L. variegatus* genomic sequence around the gene reveals positive *cis*-regulatory elements with very remarkable efficiency. As did Loots *et al.* (2000) we used a small sliding window (20-50 bp in our case, 100 bp in theirs) and looked for patches of sequence conservation greater than a set threshold (>80% similarity within the window in our case; >70% in theirs). Our analysis was carried out with a new software package designed in our lab specifically for the purpose of identifying *cis*-regulatory elements (see Brown *et al*., 2002, this Issue). The object was to test in a systematic way the usefulness of interspecific sequence comparison for simultaneous identification of multiple *cis*-regulatory modules. We scanned a 60 kb region of orthologous sequences that includes the gene, its large introns and the flanking sequence, and we found 17 patches of conserved sequence that appeared possibly significant. Every one of these was then cloned into an expression vector and tested for *cis*-regulatory activity by gene transfer into *S. purpuratus* eggs.

MATERIALS AND METHODS

FamilyRelations Analysis

The FamilyRelations package for comparative sequence evaluation in described elsewhere in this Issue (Brown *et al.*, 2002), and a user's guide is available at **family.caltech.edu**. The *Strongylocentrotus purpuratus* and *Lytechinus variegatus* BAC clones were selected by screening the respective arrayed BAC libraries (Cameron *et al*., 2000) with *otx* cDNA clones (Li *et al*., 1997), and ordered sequence was obtained. The *S. purpuratus* BAC was sequenced at the Joint Genome Institute (DOE) at Walnut Creek, CA, and the *L. variegatus* BAC at the Institute for Systems Biology at Seattle, WA. Both sequences are available at **sea-urchin.caltech.edu/genome**. The BACs were annotated as shown in Results and the 60 kb *L. variegatus* BAC was thereby shown to overlap most of the relevant region around the *S. purpuratus otx* gene. These orthologous sequences were imported into FamilyRelations for the comparative analysis.

Whole mount in situ *hybridization (WMISH)*

The WMISH procedures used to detect the various *otx* gene transcripts were carried out with the minor modifications of the standard procedures that are described by Ransick *et al*. (2002).

Preparation of Expression Vectors

A universal *S. purpuratus* expression vector "CRETrap," or "*cis*-Regulatory Element Trap," designed specifically for assessment of *cis*-regulatory activity was

described by Cameron *et al*. (2002). It consists essentially of the enhanced basal promoter of the *endo16* gene (Yuh *et al*., 2001) linked to a reporter gene. Here we used as a reporter the bacterial CAT (chloramphenicol acetyltansferase) gene, the same as earlier used for the analysis of the *endo16* gene (Yuh *et al*., 1996, 1998, 2001; Yuh and Davidson, 1996) and the *cyIIIa* gene (Kirchhamer *et al*., 1996). This reporter is different from that used in the CRETrap of Cameron *et al*. (2002), which encodes GFP. In the present work the transcriptional output of the CAT reporter was monitored by whole mount *in situ* hybridization (Yuh and Davidson, 1996). CAT mRNA is relatively unstable in sea urchin embryos, and this is a desirable characteristic, as it affords a more sensitive indication of cellular expression in each given time frame than does GFP fluorescence. GFP is extremely stable in these embryos, so that the fluorescence at any given time is the sum of all prior episodes of expression (Arnone et al., 1997). Immediately upstream of the enhanced basal promoter in the CAT expression vector that we used here is a cloning site, consisting of a *Kpn*I target sequence for the 5' end of the insert followed by a *Sma*I site. The inserts of the constructs to be tested were produced by PCR amplification from the *S. purpuratus* BAC. A cell lysate was prepared from the bacterium carrying the BAC and the derived sequences were amplified directly from the lysate using the Expand High Fidelity PCR system (Roche). Primers were selected to amplify the elements indicated in the FamilyRelations analysis, and were equipped with *Kpn*I and *Sma*I anchors. The construct clones were grown, and DNA prepared, linearized, and injected into eggs together with carrier by the standard methods described in detail earlier (see Yuh and Davidson, 1996; Kirchhamer and Davidson, 1996; Livant *et al*., 1991).

RESULTS

Assessment of Zygotic otx *Expression by Whole Mount* In situ *Hybridization*

Whether thought of in terms of its functions in the embryo, its spatial expression, its alternative splicing pattern, or its genomic transcriptional apparatus, the *otx* gene of *S. purpuratus* can only be described as complex. The structure of the gene was worked out by Li *et al*. (1997) in a detailed analysis of *otx* cDNAs and of an overlapping series of λ genome recombinants that together included the whole of the gene. Their results showed that the gene has eight exons, uses three different transcriptional start sites, and produces four different mRNAs encoding various Otx proteins (a summary map is at the top of Fig. 5 of this paper). All include the sequences encoding the Otx homeodomain and the C-terminal portion of the protein, i.e., exons 7 and 8. The *αotx* mRNA is initiated at exon 6, and this is the only form that utilizes exon 6. There is a maternal *αotx* mRNA that is globally present, and this form is also expressed zygotically since the amount of *αotx* mRNA in the embryo rises progressively in the blastula stage (Li *et al*., 1997; Chuang *et al*., 1996). *In situ* hybridization on sections indicated *αotx* mRNA in oral ectoderm and gut by early gastrula stage (Li *et al*., 1997). Two other closely related *otx* mRNAs are transcribed from the second start site at exon 3. This is the "β1/2" start site, located 8.9 kb upstream of the *αotx* start site. The β1 transcript includes exons 3, 4, and 5, which are located close together, and exons 7 and 8, while the β2 transcript instead includes only exons 3, 5, 7 and 8. Li *et al*. (1997) showed that there are no maternal *β1/2otx* mRNAs. Expression of β1/2 is predominantly in oral ectoderm and endoderm at

blastula and early gastrula stages, according to their *in situ* hybridizations on sections. The third or "β3" start site at exon 1 is located about 17 kb upstream of the start at exon 3. The β3 mRNA consists of two small exons, 1 and 2, which are separated by a small intron, spliced to exons 5, 7, and 8 (Li *et al*., 1997). Nothing was known of *β3otx* expression.

To flesh out our knowledge of the spatial patterns of *otx* expression with respect to the gene organization, we carried out a whole mount *in situ* hybridization (WMISH) study. For this probes were prepared which according to Li *et al*. (1997) would uniquely identify the *αotx*, *β1/2otx* and *β3otx* transcripts; i.e., the probes represent the exons specific to each mRNA, as above. A sample of the results is shown in Fig. 1. In general we confirmed the results of Li *et al*. (1997), with some minor exceptions. Taking the two data sets together, the most certain aspects of the transcription patterns for the three start sites are as follows:

αotx transcripts are at first ubiquitous (these are probably largely maternal, as we show below); *αotx* transcripts then accumulate zygotically in the vegetal plate endomesoderm and according to Li *et al.* (1997) in the oral ectoderm. We could not see a an *αotx* signal in oral ectoderm at 24 h by WMISH, though as Fig. 1A shows, there is clearly *αotx* expression in endoderm at this time. After 24 h *αotx* transcripts decline sharply in prevalence, as was also indicated in a direct assessment of transcript concentration by quantitative PCR (data not shown). It follows from these kinetics that the Otx form that initiates the expression of the *endo16* gene is *αotx* (Yuh *et al*., 1998).

β1/2otx mRNA is localized to oral ectoderm during late blastula-gastrula stages, and is strongly expressed in the endoderm as well (Fig. 1B-D). At 72 h there is *β1/2otx* expression in the gut and hood region of the oral ectoderm (Fig. 1E-F).

The *β3otx* transcription unit begins to be expressed ubiquitously, during cleavage, and its products can be detected as early as 10 h (not shown). It is still being expressed ubiquitously at midblastula stage (17 h; Fig. 1G). A phase of intense vegetal plate expression then begins, as illustrated in Fig. 1H. As gastrulation is initiated *β3otx* transcripts are seen clearly in the invaginating archenteron. This is shown in Fig. 1I, which also illustrates the presence of *β3otx* transcripts in the ectoderm on one side, probably the oral ectoderm. The *β3otx* transcription apparatus turns off soon after the stage shown in Fig. 1I, however, and no *β3otx* RNAs are detectable in prism stage and later embryos.

For our present purposes, this information provides a standard of expectation against which to match the results of *cis*-regulatory analysis. We see, for example, that all three of the transcription units are expected to answer to endodermal or endomesodermal control elements; that the *β3otx* regulatory system should include ubiquitously acting elements; and that expression in skeletogenic mesenchyme cells is not expected to be driven off the *αotx* or the *β1/2otx cis*-regulatory system. But not even the complex pattern of *otx* gene expression revealed in Fig. 1 predicted the multiplicity of *cis*-regulatory elements that we demonstrate below.

Genomic Sequence of BACs Containing the* otx *Gene

The annotated features of a 161 kb *S. purpuratus otx* BAC sequence are shown in Fig. 2A, and a shorter *L. variegatus* BAC chosen to overlap the *otx* gene region is similarly displayed in Fig. 2B. The annotation was performed with the SUGAR software package, which is briefly described by Brown *et al*. (2002; this Issue). In addition to known genes recognized by BLAST comparisons against the public data bases, SUGAR also co-plots the occurrence of known *S. purpuratus* repetitive sequences, cDNAs, BAC-end sequences (Cameron *et al*., 2000) and the results of several exon prediction analyses (see legend). The *otx* gene is oriented left to right (Fig. 2A), so the upstream region extends leftward. The 60 kb region of orthology between the *Lytechinus* and the *Strongylocentrotus* sequences is indicated by the "comparative" (i.e., comparative BLAST) feature. Shortly to the left of this overlap region in the *S. purpuratus* BAC are two other genes, the first a predicted gene (at about 65-72 kb) similar to an unknown human gene in the data bases (KIAA0903), and the second, the *spectrin* gene (at about 40-55 kb). The domain in which *otx* regulatory elements might reside is unlikely to extend beyond these other genes; i.e., most of the region that we needed to scan for *otx* *cis*-regulatory elements is included within the 60 kb region of overlap with the *L. variegatus* BAC sequence. Because of the large size of its introns, the *otx* gene itself occupies 43 kb of this overlap region, from about 106 kb (exon 1) to 148 kb (exon 8). Just within the intron that follows exon 2 is an insertion element that includes a homology with reverse transcriptase. The *S. purpuratus* BAC also extends about 13 kb downstream of exon 8, but this region was not included in the present studies because the *Lytechinus* BAC sequence does not extend this far.

FamilyRelations Analysis of Shared Genomic Sequence Elements

A series of FamilyRelations comparisons is shown in Fig. 3 (see Materials and Methods). These comparisons extend across the whole of the 60 kb overlap region that includes the *otx* gene and its upstream sequence, most of the way to the KIAA0903 homology domain. The different panels in Fig. 3 represent comparisons done at different criteria. For the comparisons in Fig. 3A-C, a 20 bp sliding window was applied, and those in Fig. 3D-F used a 50 bp sliding window. The most stringent threshold criteria are shown at the top of each series, i.e., 100% similarity over 20 bp in Fig. 3A and 90% over 50 bp in Fig. 3D, and the least stringent criteria for each window size are shown at the bottom of each series, as indicated. Exons are marked in red and numbered on the *S. purpuratus* sequence in each panel. The predominant feature of the comparisons is occurrence of patches of similar sequence, outside of the conserved exons. These patches occur in the same order in the two genomes. This is shown by the approximately parallel lines connecting them which do not cross one another. Minor deviations from parallel indicate insertions/deletions. Particularly at the lower criteria we also see some similarities that occur out of parallel register. Where the lines connecting these emerge from a single source in one genome and extend in a fan-like manner to the other, they indicate a sequence element that occurs repetitively within this region in one genome but not in the other. As a thumbnail guide to which comparison criterion to apply, we have gravitated toward the lowest at which a majority of the out of register and fan-like structures have disappeared from the plot. In this case we chose the criterion of Fig. 3E, i.e., 80% similarity in a 50 bp window. A few of the features in Fig. 3E would have been

missed at the criterion in Fig. 3D, while there appeared to be too much noise at the two lowest criteria, shown in Fig. 3C and F. The result in Fig. 3A is essentially similar to that in Fig. 3E. In selecting the sequence elements to be examined experimentally, we usually exclude all that are very short, i.e., <100 bp; all exons; any remaining out of register elements; and any that consist only of simple sequence (e.g., microsatellites or regions consisting almost completely of A's and T's).

Seventeen conserved regions were chosen for direct experimental test of their *cis*-regulatory activity. These are marked with green boxes in Fig. 3E.

cis-*Regulatory Activity of Conserved Sequence Patches*

The boxed sequence elements were amplified from the *S. purpuratus* BAC by means of PCR. FamilyRelations has a zoom feature which enables the operator to see the sequence of the conserved elements and their flanking regions, so that appropriate primers could be designed. The amplified sequences were cloned into a universal *S. purpuratus* expression vector, the same as the *cis*-Regulatory Element Trap vector of Cameron *et al*. (2002), except that it contains a CAT rather than GFP reporter (see Materials and Methods). *cis*-Regulatory activity was detected by WMISH assay for CAT mRNA, in 24 h and 48 h embryos grown from batches of eggs into each of which a given construct had been injected. The results for all 17 constructs are listed in detail in Table 1, and examples of the WMISH displays for many of the most important constructs can be seen in Fig. 4. An overall summary is provided in Fig. 5.

The main import of these experiments is simple, and perhaps astonishing: no less than 11 of the 17 constructs proved to be clearly active. "Active" here means that clones

of >2 cells per embryo (usually much more, i.e., 8-16 cells on the average) display CAT mRNA in an appreciable fraction of embryos. Background for this vector is about 2-3% of embryos displaying a few randomly positioned, single, active cells (Yuh *et al*., 1996; Yuh and Davidson, 1996; Cameron *et al*., 2002). Each of these 11 active constructs generated strong activity in at least some domains of the embryo at one stage or another.

Construct 17, just upstream of the *αotx* start site at exon 6, contains a powerful, early acting endomesodermal enhancer, and also expresses well in ectoderm at 24 h (Table 1, and Fig. 4G). Construct 16 expresses strongly in ectoderm at 24 h. Later, construct 17 is expressed in some secondary mesoderm cells as well as in the gut, though the fraction of embryos and number of cells expressing declines, as seen in Table 1. Construct 17 is also expressed weakly in both oral and aboral ectoderm, and construct 16 continues to be expressed strongly in oral and aboral ectoderm and in skeletogenic mesenchyme. Together, constructs 16 and 17 are expressed more widely than is the endogenous *αotx* transcription unit and further into development. For example, though at 48 h construct 16 expresses strongly in both ectoderm territories, as well as skeletogenic mesenchyme, and construct 17 weakly in the ectoderm territories, in life the *αotx* transcript is much diminished at this stage; there is no expression in skeletogenic mesenchyme at all; and the ectodermal expression that persists is oral (Li *et al*., 1997). The direct implication is that there is a repressor of aboral ectoderm and mesenchyme expression in the native *cis*-regulatory system governing *αotx* expression. However, this repression cannot be detected using the individual constructs tested. There may be a general late-acting repressor of *αotx* which is not detected either.

Construct 11, upstream of the *β1/2otx* start site, expresses exclusively in endoderm at blastula stage (Fig. 4D; Table 1); and constructs 14 and 15 also express strongly in the endoderm (Fig. 4E, F; Table 1). At 48 h both construct 14 and construct 11 continue to express strongly in endoderm and secondary mesenchyme. These results are in concord with the WMISH results of Fig. 1B-E, which display clearly the endomesodermal and endodermal expression of the *β1/2otx* transcript throughout this period. However, the strong oral ectoderm expression of this transcript that can be seen in Fig. 1B-F is again not represented in the ectodermal pattern of expression generated by construct 15, which is active in aboral as well as oral ectoderm (Fig. 4F; Table 1). There is no endogenous mesodermal *β1/2otx* expression (Fig. 1B), but constructs 11 and 14 express significantly in secondary mesenchyme, though not in skeletogenic mesenchyme (Table 1).

Finally, we see that constructs 3 and 4 at 24 and 48 h, and construct 8 at 24 h, promote ubiquitous expression (Fig. 4A, Table 1). These constructs lie well upstream of the *β3otx* start site, and this result is consistent with the ubiquitous presence of the *β3otx* transcript *in vivo* (Fig. 1G). However, the ubiquitous expression of the endogenous *β3otx* regulatory system occurs only early in development, and all endogenous *β3otx* transcripts have disappeared from the embryo by 48 h, while in the expression experiments the CAT mRNA signal remains ubiquitously distributed at this stage, though weaker than at 24 h (Table 1). Element 3 includes an endomesoderm enhancer (Table 1), and this could be responsible for the endomesodermal expression of the endogenous *β3otx* transcription unit (Fig. 1H, I). Elements 5, 6 and 7 express strongly in ectoderm (Fig. 4B, C; Table 1) again consistent with the expression pattern of the endogenous *β3otx* transcription unit,

except that like construct 16, these constructs lack the discrimination between oral and aboral ectoderm that the native system displays (Fig. 1I).

To summarize, though the picture is not complete or consistent in every detail, the positive *cis*-regulatory elements unearthed in this exploration seem to account for the major features of spatial expression of the *αotx, β1/2otx*, and *β3otx* transcripts, particularly the endodermal expression of all three of the transcription units. In spatial terms the major discrepancy is the predominantly oral ectoderm expression of *αotx*, *β1/2otx*, and probably *β3otx*, in contrast to expression of constructs 3-7 and 15-17 in both oral and aboral ectoderm. Clearly the endogenous control systems include a means of repressing oral ectoderm expression, and this feature is lacking in the constructs tested, or could only have been observed had several been combined. Temporally, our WMISH results and QPCR analyses show that both the endogenous *αotx* and *β3otx* regulatory systems shut down after gastrulation begins, *β3otx* more abruptly than *αotx*. In contrast, constructs 3-7 and 16-17 remain strongly active at 48 h. The late phase of expression of construct 16 also includes appreciable activity in skeletogenic mesenchyme cells, while no endogenous *αotx* expression occurs in these cells (Fig. 1A). Evidently these constructs are all also missing response to a late acting repression system that controls the activity of the endogenous *αotx* and *β3otx* transcription units. Of course it is impossible to assume with any confidence that each promoter is affected only by the elements that lie upstream of it and not by elements that lie upstream of the adjacent promoter. For example, the strong oral ectodermal expression of the *β1/2otx* transcript might in fact be due to the influence of the ectodermal modules that are upstream of the *β3otx* start site,

so we cannot unequivocally relate the activity of given modules to the endogenous patterns of expression of each of the three transcription units.

Our main objective, in any case, was not to obtain a *cis*-regulatory analysis of the *otx* gene, but rather to ask whether the use of interspecific sequence conservation could lead us to the *cis*-regulatory elements of this gene. Clearly it has done exactly that, though the discrepancies noted above show that some negatively acting elements are yet to be recovered. Modules the entire function of which is to cause repression of other modules (e.g., regions F-C of the *endo16* gene; Yuh and Davidson, 1996; Yuh *et al*., 1996) will of course fail to display activity in the test used in this paper. Such elements could well be located in the conserved sequences that we scored as "inactive" in Table 1.

DISCUSSION

The cis-*Regulatory System of the* otx *Gene*

As can be seen in the summary map of Fig. 5, the 11 spatial expression modules of the *otx* gene revealed in this work are distributed in a particular fashion. The distal region upstream of the *β3* promoter contains the only ubiquitously active modules (i.e., modules 3, 4 and 8). The strongest exclusively ectodermal modules are also in this same region (modules 5, 6 and 7), except for that in module 16 (at 24 h), upstream of the *αotx* start site. Elements that generate strong endomesoderm- and endoderm-specific expression are found in all the regions: upstream of the *β3otx* start site (module 3); upstream of the *β1/2otx* start site (modules 11 and 14 and the element included in construct 15); and immediately upstream of the *αotx* start site (module 17). But there is a

mystery to be seen in Fig. 5: why are there three different ubiquitously active modules near one another (modules 3, 4 and 8); why are there three strong, exclusively ectodermal expression modules, separated by several kb but likewise in the same general region with respect to the transcription map (modules 5, 6 and 7); and why are there three separate endoderm modules (11, 14 and 15) in the domain upstream of the *β1/2otx* start? Traditional methods of *cis*-regulatory analysis inexorably funnel experimental effort towards "minimal regulatory elements," i.e., the smallest single DNA fragments that can be shown to execute a given regulatory function in gene transfer experiments. For instance, had we isolated only module 11, and considered it a "minimal endomesoderm regulatory element," the interesting question of why there are several separate endoderm elements near module 11 would not have arisen. So once again the use of an unbiased, general approach, here looking at a whole set of conserved sequence patches, produces something unexpected. According to the crudely qualitative assays of Fig. 4 and Table 1 the multiple elements of each class function similarly. Are they in fact redundant, or do they perform subtly different functions so that each has an individual, selectively valuable role?

The last possibility cannot be addressed except by means of thorough *cis*-regulatory analyses. However, we can determine whether the individual endomesodermal or endodermal modules are in fact redundant with respect to target sites for factors that are likely regulators of the *otx* gene in this domain of the embryo (Davidson *et al*., 2002, this Issue). Whichever *cis*-regulatory elements drive expression of the *β1/2otx* transcription unit in the endomesoderm should, according to our other evidence, include sites at which the Otx factor itself binds; at which the Tcf factor binds;

and at which a Gata factor binds. Table 2 shows that all three of the relevant modules (11, 14 and 15) indeed contain Tcf and Gata sites, though only module 14 looks capable of autoactivation. So in this sense the three modules are at least to some extent indeed redundant with respect to one another, whatever other particular features each may have.

Redundancy carries the implication of lack of functional usefulness, but where formation of transcription factor complexes is the issue this is probably a false argument. During specification processes, the concentrations and activities of transcription factors that provide inputs to the key *cis*-regulatory elements driving developmental choice are by definition in a state of change. Other things being equal, a *cis*-regulatory system that affords multiple opportunities for formation of active complexes will produce a more robust, and also an earlier response. The regionality of the apparent modular redundancy that we see in Fig. 5 is consistent with this idea. In life, that is in genomes undergoing the evolutionary process, as opposed to in the laboratory where Ockham's Razor prevails, the tendency is toward the elaboration of maximal not minimal regulatory systems.

The sites listed in Table 2 are entirely in accord with prediction. Thus our analysis indicates that *αotx* transcription is not under Tcf control, since it is not affected by introduction of cadherin mRNA, while *β1/2otx* transcripts are severely decreased in embryos bearing cadherin mRNA (Davidson *et al*., 2002): indeed module 17 lacks a Tcf site, while the three endomesodermal modules that presumably service the *β1/2otx* start site (11, 14 and 15) all have these sites. But none of the three ectoderm-specific modules in Table 2 have Tcf sites. Gata (hGATAr) sites are present in both ectodermal and endodermal modules (a couple of these sites would be expected to occur in 2 kb of random sequence anyway). Table 2 shows that the density of these sites is about six

times higher than random expectation in the *β1/2otx* endoderm modules, and twice as high in the ectoderm modules. With respect to autoregulation, the *otx* gene is active in the ectoderm, as we have seen, and so it is reasonable that ectodermal modules such as that included in construct 15 as well as endodermal modules such as construct 14 include Otx target sites.

FamilyRelations and Network Analysis

In this work we scanned 60 kb of DNA sequence and tested 17 separate sequence elements for regulatory activity within a few weeks. The efficiency with which active positive elements were revealed by interspecific comparison is spectacular, relative to conventional approaches. One interesting statistic is the fraction of randomly chosen genomic DNA fragments that yield evidence of *cis*-regulatory activity when tested in the same way. Cameron *et al*. (2002) showed that this number is about 10%, for 2-3 kb fragments. With the exception of construct 12, which turned out to be inactive, the inserts in the present work average 590 bp, about a third of the average length used by Cameron *et al*. (2002). So for these inserts, had they been randomly chosen fragments, we might have expected about a 3% success rate. Instead we had a 65% success rate. Of course some of the difference is due to the fact that the study was carried out on 60 kb of DNA known to lie within and extend upstream of a gene active in the embryo; but on the other hand the average intergenic distance on either side of a given gene in *S. purpuratus* is just about 30 kb (Cameron *et al*., 2002). Had we randomly chopped up and subcloned the same 60 kb sequence into the expression vector we would have eventually found the same active elements, but this would have required many times more constructs,

injections and assays. Furthermore, the results would have been much less satisfactory: FamilyRelations provides immediate information on the boundaries of the elements, as well as marking their location in the genome. We ended up scanning a total of only about 20% of the 60 kb length all told, and finding 11 active regulatory elements therein.

For most bilaterian organisms that one might wish to study, there will be another species at the right evolutionary distance so that the same approach should always be useful. The right distance is one at which the two species develop the same way, so that they use (most of) their regulatory genes similarly; and at which the non-selected DNA sequence has diverged sufficiently so that a FamilyRelations signal over background can be obtained for conserved patches of non-coding sequence. The catch is that the best species pair may be different for different genes. In the case of the sea urchin genome we had the advantage of earlier measurements on overall single copy DNA sequence divergence among various species, which showed that under standard criterion conditions only about 15% of the *Strongylocentrotus* single-copy DNA cross-reacts with *Lytechinus* DNA (Angerer *et al*., 1976). The simplest and most general solution is that for each reference species, a set of BAC genomic libraries for surrounding species at different phylogenetic distances should be available. This would generally make it possible to find *cis*-regulatory elements for any desired gene, quickly and efficiently. Therein is likely to lie the solution to the major problem confronting experimental verification of regulatory gene network architecture: how to get one's hands on the *cis*-regulatory elements of a large number of different genes at minimum cost in effort, so that testing such networks becomes experimentally accessible.

ACKNOWLEDGMENTS

We are grateful to Prof. Ellen Rothenberg of Caltech for a thoughtful and careful critique of this manuscript. This work was supported by NIH grants GM-61005 and HD-37105 and by the Caltech Beckman Institute. CTB is a participant in the Initiative in Computational Molecular Biology, which is funded by an award from the Burroughs Wellcome Fund Interfaces program. CL was supported by the Gordon Ross Fellowship and the Walter and Sylvia Treadway Foundation.

REFERENCES

Angerer, R. C., Davidson, E. H. and Britten, R. J. (1976). Single copy DNA and structural gene sequence relationships among four sea urchin species. *Chromosoma* **56**, 213-226.

Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R. and Brenner, S. (1995). Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc. Natl. Acad. Sci. USA* **92**, 1684-1688.

Arnone, M. I., Bogarad, L. D., Collazo, A., Kirchhamer, C. V., Cameron, R. A., Rast, J. P., Gregorians, A. and Davidson, E. H. (1997). Green fluorescent protein in the sea urchin: New experimental approaches to transcriptional regulatory analysis in embryos and larvae. *Development* **124**, 4649-4659.

Britten, R. J. (1986). Rates of DNA sequence evolution differ between taxonomic groups. *Science* **231**, 1393-1398.

Brown, C. T., Rust, A. G., Clarke, P. J. C., Pan, Z., Schilstra, M. J., De Buysscher, T., Griffin, G., Wold, B. J., Cameron, R. A., Davidson, E. H. and Bolouri, H. (2002). New computational approaches for analysis of *cis*-regulatory networks. *Dev. Biol*., in press.

Cameron, R. A., Mahairas, G., Rast, J. P., Martinez, P., Biondi, T. R., Swartzell, S., Wallace, J. C., Poustka, A. J., Livingston, B. T., Wray, G. A., Ettensohn, C. A., Lehrach, H., Britten, R. J., Davidson, E. H. and Hood, L. (2000). A sea urchin

genome project: Sequence scan, virtual map, and additional resources. *Proc. Natl. Acad. Sci. USA* **97**, 9514-9518.

Cameron, R. A., Oliveri, P., Wyllie, J. and Davidson, E. H. (2002). *cis*-Regulatory activity of randomly chosen genomic fragments from the sea urchin. *Proc. Natl. Acad. Sci. USA*. Submitted.

Carr, J. L., Shashikant, C. S., Bailey, W. J. and Ruddle, F. H. (1998). Molecular evolution of *Hox* gene regulation: Cloning and transgenic analysis of the lamprey *HoxQ8* gene. *J. Exp. Zool*. **280**, 73-85.

Chuang, C.-K., Wikramanayake, A. H., Mao, C.-A., Li, X. and Klein, W. H. (1996). Transient appearance of *Strongylocentrotus purpuratus* Otx in micromere nuclei: Cytoplasmic retention of SpOtx possibly mediated through an α-actinin interaction. *Dev. Genet*. **19**, 231-237.

Davidson, E. H. (2001). "Genomic Regulatory Systems. Development and Evolution." Academic Press, San Diego, CA.

Davidson, E. H., Rast, J. P., Oliveri, P., Cameron, R. A., Ransick A., Yuh, C.-H., Calestani, C., Arenas-Mena, C., Otim, O., Minokawa, T., Brown, C. T., Lee, P. Y., Livi, C., Revilla, R., Dong, P., Wyllie, J., Yun, M., Yun, C. H., Clarke, P. J. C., Hood, L. E., Rowen, L. and Bolouri, H. (2002). A large, provisional gene regulatory network for endomesodermal specification in the sea urchin embryo. In preparation.

Kirchhamer, C. V. and Davidson, E. H. (1996). Spatial and temporal information processing in the sea urchin embryo: Modular and intramodular organization of the *CyIIIa* gene *cis*-regulatory system. *Development* **122**, 333-348.

Kirchhamer, C. V., Yuh, C.-H. and Davidson, E. H. (1996). Modular *cis*-regulatory organization of developmentally expressed genes: Two genes transcribed territorially in the sea urchin embryo, and additional examples. *Proc. Natl. Acad. Sci. USA* **93**, 9322-9328.

Li, X., Chuang, C.-K., Mao, C.-A., Angerer, L. M. and Klein, W. H. (1997). Two Otx proteins generated from multiple transcripts of a single gene in *Strongylocentrotus purpuratus*. *Dev. Biol*. **187**, 253-266.

Livant, D. L., Hough-Evans, B. R., Moore, J. G., Britten, R. J. and Davidson, E. H. (1991). Differential stability of expression of similarly specified endogenous and exogenous genes in the sea urchin embryo. *Development* **113**, 385-398.

Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M. and Frazer, K. A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**, 136-140.

Manzanares, M., Wada, H., Itasaki, N., Trainor, P. A., Krumlauf, R. and Holland, P. W. H. (2000). Conservation and elaboration of *Hox* gene regulation during evolution of the vertebrate head. *Nature* **408**, 854-857.

Nishizaki, Y., Shimazu, K., Kondoh, H. and Sasaki, H. (2001). Identification of essential sequence motifs in the node/notochord enhancer of *Foxa2* (*Hnf3β*) gene that are conserved across vertebrate species. *Mech. Dev*. **102**, 57-66.

Nonchev. S., Maconochie, M., Vesque, C., Aparicio, S., Ariza-McNaughton, L., Mansanares, M., Maruthainar, K., Kuroiwa, A., Brenner, S., Charnay, P. and Krumlauf, R. (1996). The conserved role of *Krox-20* in directing *Hox* gene

expression during vertebrate hindbrain segmentation. *Proc. Natl. Acad. Sci. USA* **93**, 9339-9345.

Ransick, A., Rast, J. P., Minokawa, T., Calestani, C. and Davidson, E. H. (2002). New early zygotic regulators of endomesoderm specification in sea urchin embryos discovered by differential array hybridization. *Dev. Biol*., in press.

Ruffins, S. W. and Ettensohn, C. A. (1996). A fate map of the vegetal plate of the sea urchin (*Lytechinus variegatus*) mesenchyme blastula. *Development* **122**, 253-263.

Sivasankaran, R., Vigano, M. A., Müller, B., Affolter, M. and Basler, K. (2000). Direct transcriptional control of the Dpp target *omb* by the DNA binding protein Brinker. *EMBO J*. **19**, 6162-6172.

Springer, M. S., Davidson, E. H. and Britten, R. J. (1991). Retroviral-like element in a marine invertebrate. *Proc. Natl. Acad. Sci. USA* **88**, 8401-8404.

Yuh, C.-H. and Davidson, E. H. (1996). Modular *cis*-regulatory organization of *Endo16*, a gut-specific gene of the sea urchin embryo. *Development* **122**, 1069-1082.

Yuh, C.-H., Moore, J. G. and Davidson, E. H. (1996). Quantitative functional interrelations within the *cis*-regulatory system of the *S. purpuratus Endo16* gene. *Development* **122**, 4045-4056.

Yuh, C.-H., Bolouri, H. and Davidson, E. H. (1998). Genomic *cis*-regulatory logic: Functional analysis and computational model of a sea urchin gene control system. *Science* **279**, 1896-1902.

Yuh, C.-H., Bolouri, H. and Davidson, E. H. (2001). *cis*-Regulatory logic in the *endo16* gene: Switching from a specification to a differentiation mode of control. *Development* **128**, 617-628.

TABLE 1. Whole Mount *In Situ* Hybridization Assays of *otx* Expression Constructs

Stage	Constructs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Location in BAC sequence	88945 89124	90382 91049	92916 93284	94045 94936	95513 96156	99585 100120	101161 101514	103965 104580	105266 105622	107136 107458	112545 113472	114783 118248	119555 120797	122639 123313	126557 127252	131164 131641	135944 136525
24 h	Non-stained	106	53	27	8	28	11	9	41	92	86	4	75	68	10	16	16	6
	Total stained embryos	11	16	91	124	80	98	73	89	9	13	25	11	9	61	89	58	119
	(% stained embryos)	(9%)	(23%)	(77%)	(94%)	(74%)	(90%)	(89%)	(68%)	(8.9%)	(13%)	(86%)	(13%)	(12%)	(86%)	(85%)	(67%)	(95%)
	Vegetal plate endoderm	|	|	31%	30%	0	0	0	19%	|	|	100%	|	|	72%	65%	0	88%
	(***Stained cells/embryo***)	|	|	***10***	***3***				***3***	|	|	***10***	|	|	***5***	***10***		***10***
	Vegetal plate mesoderm	|	|	16%	16%	0	0	0	13%	|	|	8%	|	|	13%	4.5%	0	32%
	(***Stained cells/embryo***)	|	|	***9***	***3***				***3***	|	|	***3***	|	|	***5***	***4***		***8***
	Skeletogenic mesenchyme (***Stained cells/embryo***)	|	|	23% ***9***	19% *7*	14% ***3***	5.6% ***3***	12% ***3***	21% ***4***	|	|	0	|	|	13% ***3***	11% ***3***	14% ***3***	13% ***3***
	Ectoderm	|	|	54%	36%	90%	98%	100%	52%	|	|	0	|	|	18%	30%	86%	60%
	(***Stained cells/embryo***)	|	|	***8***	***8***	***10***	***16***	***10***	***8***	|	|		|	|	***4***	***8***	***8***	***8***
48 h	Non-stained	50	83	20	5	95	5	14	37	122	127	7	38	26	10	8	4	1
	Total stained embryos	14	26	112	32	50	68	132	36	15	21	38	11	7	75	50	40	32
	(% stained embryos)	(22%)	(21%)	(85%)	(87%)	(34%)	(93%)	(90%)	(49%)	(11%)	(14%)	(84%)	(22%)	(21%)	(88%)	(86%)	(91%)	(97%)
	Gut	|	|	27%	25%	0	0	3%	28%	|	|	100%	|	|	65%	64%	0	50%
	(***Stained cells/embryo***)	|	|	***3***	***3***			***3***	***3***	|	|	***10***	|	|	***6***	***10***		***8***
	Secondary mesenchyme	|	|	17%	19%	0	0	2%	8.3%	|	|	44%	|	|	39%	2.0%	0	25%
	(***Stained cells/embryo***)	|	|	***3***	***3***			***3***	***3***	|	|	***4***	|	|	***6***	***4***		***6***
	Skeletogenic mesenchyme cell (***Stained cells/embryo***)	|	|	25% ***9***	19% *7*	8% ***3***	0	9.8% ***3***	11% ***3***	|	|	0	|	|	8.0% ***3***	14% ***3***	33% ***3***	13% ***3***
	Aboral ectoderm	|	|	16%	19%	54%	53%	50%	31%	|	|	2.6%	|	|	5.3%	14%	45%	13%
	(***Stained cells/embryo***)	|	|	***9***	***9***	***5***	***10***	***12***	***3***	|	|	***3***	|	|	***3***	***5***	***8***	***5***
	Oral ectoderm	|	|	15%	19%	44%	53%	47%	22%	|	|	2.6%	|	|	5.3%	12%	38%	17%
	(***Stained cells/embryo***)	|	|	***9***	*7*	***5***	***10***	***12***	***3***	|	|	***3***	|	|	***3***	***5***	***8***	***5***

Note to Table 1.

Embryos were injected with expression constructs and assessed for mRNA generated from the CAT reporter as described in Materials and Methods. The first two rows of data for each time point give the number of normally developing embryos scored, and the percent of total normally developing embryos examined which display activity, and the remainder give the percent of those which are actively expressing at each location within the embryo. Data are provided for 24 h mesenchyme blastula and 48 h late gastrula-stage embryos. At 48 h the oral and aboral ectoderm can easily be distinguished by the proximity of the archenteron to the oral side when the embryo is viewed laterally, by the location of the skeletogenic mesenchyme cell clusters on the oral side, and by the general shape of the embryo. The average number of cells active per embryo is also given (bold italics) for each location. Inactive constructs (vertical dashed lines) displayed insignificant numbers of embryos expressing in any given domain, compared to background in the same experiment, normally 0-2 isolated stained cells per embryo. The exact location of the starting and ending position of each construct is given in the BAC sequence under the name of the construct.

Table 2

TABLE 2
Tcf, Gata, and Otx Factor Target Sites in *otx cis*-Regulatory Elements

Domain	Transcript	Module	L (bp)	Tcf sites (tTCAAAGg)	Gata sites (hGATAr)	Otx sites (TAATCY)
Endodermal	*β1/2-otx*	11	927	1	4	0
		14	674	1 (442^{-1a})	6 (169^{-1a})	1
		15	697	3	3	0
	α-otx	17	575	0	1	1
Ectodermal	*β3-otx*	5	643	0	3	1
		6	535	0	3 (251^{-1a})	0
		7	353	0	1	0
	α-otx	16	477	0	1	0

[a] Density of sites, i.e., sites per *x* base pairs.

FIGURE LEGENDS

FIG. 1. Localization of transcripts deriving from the three *otx* start sites, displayed by whole mount *in situ* hybridization. Probes were produced to exons that uniquely define the products of the *αotx, β1/2otx and β3otx* transcription units, according to the analysis of Li *et al*. (1997; see text and transcript map in Fig. 5 of this paper). The identity of the probe is shown at the lower left in each panel, and the embryonic stage at lower right: (A), *αotx*; (B-F), *β1/2otx*; (G-I), *β3otx*.

Figure 1

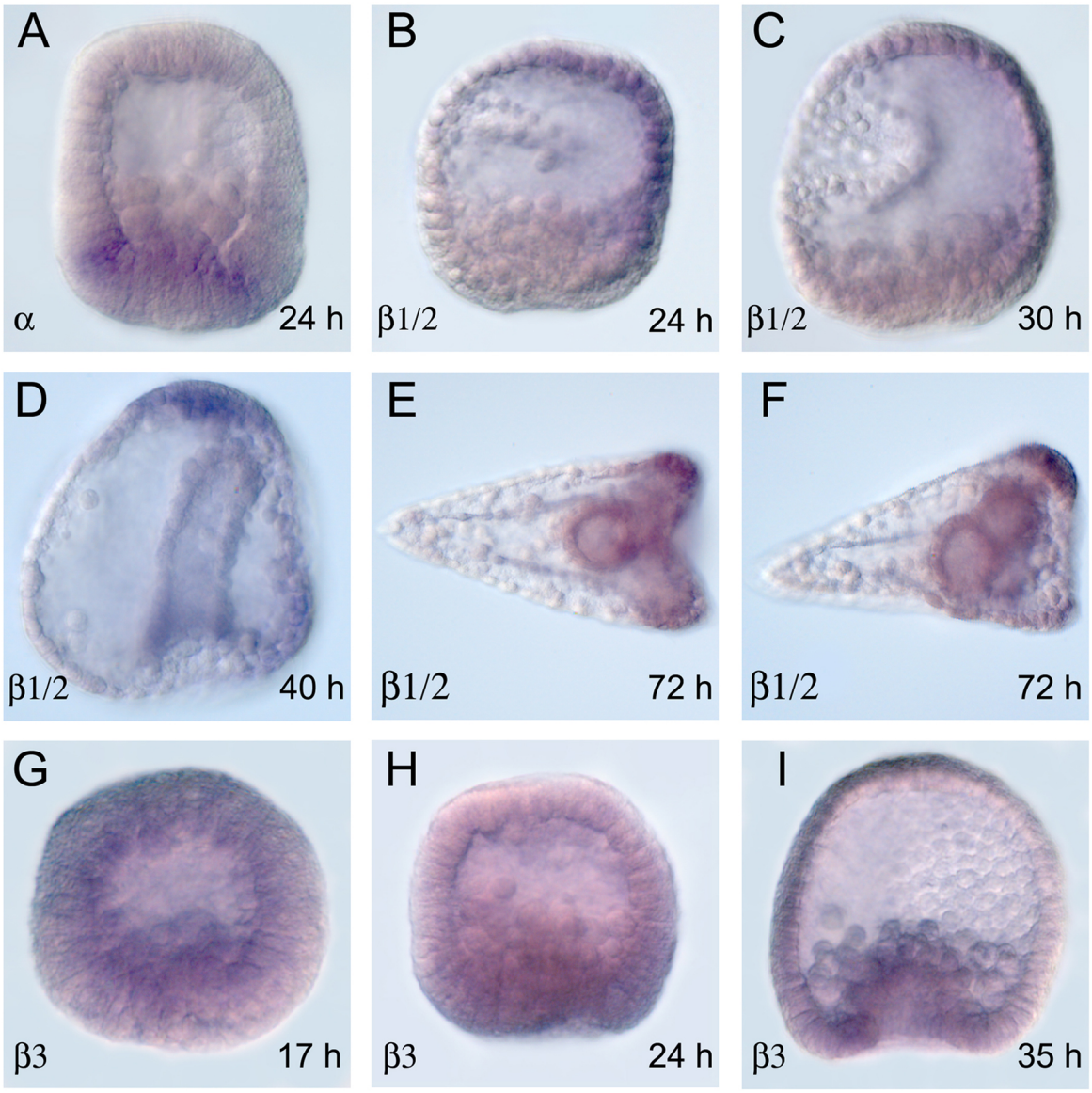

A
α
24 h
B
β1/2
24 h
C
β1/2
30 h
D
β1/2
40 h
E
β1/2
72 h
F
β1/2
72 h
G
β3
17 h
H
β3
24 h
I
β3
35 h

FIG. 2. Annotated BAC sequence including the *S. purpuratus* and *L. variegatus otx* genes. (A) *S. purpuratus* BAC sequence 161,475 bp in length. The sequence was annotated using the SUGAR software package (see Brown *et al*., 2002 this Issue for details, the provenance of program components, and additional features). The contig bar in the middle indicates the seven assembled contigs from which the sequence scaffold was assembled. Between the contigs (blue and yellow bars) are short gaps in the sequence. Below the contig bar are exons predicted by three algorithms, HMMGene, Genscan, and Geneid; exons oriented 5' to left are shown in blue and 5' to right in green. Above the contig bar are various features of the sequence recognized in BLAST comparisons. At the top are protein coding sequence elements identified (after six open reading frame translation) at BLAST criteria e^{-12} and e^{-6}, by comparison to the Swiss Prot data base; 5' to left in red and 5' to right in green. From left to right in the *S. purpuratus* BAC (006F13) these are sequences encoding a transposon reverse transcriptase (RT; Springer *et al*., 1991), a gene encoding Spectrin, and the *otx* gene. The first intron of the *otx* gene includes another reverse transcriptase element. Note that the reverse transcriptases and most of the *otx* and spectrin exons are also predicted by the exon finders at the bottom. In addition all three of these exon finders agree in predicting another gene, which at lower BLAST score resembles an unidentified human gene, KIAA0903. The cDNA comparison (dark blue) refers to the *αotx* cDNA, which consists of exons 6, 7 and 8 (see text). Upstream, as shown in the Swiss Prot comparison, are exons 1 and 2, seen as a single red block at about 106 kb (see length scale on bottom) and exons 3, 4 and 5, seen at this scale as a single red block at about 128 kb. The comparative BLAST line (light blue) indicates regions of sequence that are similar in the

Lytechinus and *Strongylocentrotus* BACs, i.e., at BLAST criterion e^{-12}. These regions include the exons, except exon 8 which lies off the end of the *Lytechinus* sequence. The red bar labeled "hot region" extends from KIAA0903 to the end of the BAC and indicates the domain within which *cis*-regulatory elements of the *otx* gene might be sought. The *otx* gene is also represented in some *S. purpuratus* EST sequences (purple blocks). Repetitive sequences identified by comparison to a canconical table of repeat sequences, or to the BAC-end repetitive sequence collection (Cameron *et al*., 2000), are shown in red, and in ochre and black, respectively; black indicates a sequence that occurs >3 times in the 76,000 BAC-end sequence library and ochre a sequence that occurs 1, 2 or 3 times. (B) Similar annotation for a 61,782 bp *Lytechinus* BAC sequence that includes the *otx* gene. Note the difference in scale from A.

Figure 2


A
Sea Urchin Genome Viewer v0.42b: 006F13
File
Display
Start 0
End 0
Length 0
Annotation
Sp otx
Hot region
SwissProt e-12
SwissProt e-6
cDNA e-6
Comparative e-12
S.p. ESTs e-70
S.p. Repeat Table e-6
BAC ends e-10
Contigs
HMMGene +
Genscan +
geneid +
geneid -
HMMGene -
Genscan -
kb
Zoom
B
Sea Urchin Genome Viewer v0.42b: 229L52
File
Display
Start 3032
End 16250
Length 13219
Annotation
Lv otx
otx
SwissProt e-12
SwissProt e-6
cDNA e-6
Comparative e-12
S.p. ESTs e-70
S.p. Repeat Table e-6
BAC ends e-10
Contigs
HMMGene +
Genscan +
geneid +
geneid -
HMMGene -
Genscan -
kb
Zoom

FIG. 3. FamilyRelations comparisons of *L. variegatus* and *S. purpuratus* sequence, at various criteria. For the regions compared, see Fig. 2; the sequence considered extends from base pair 77064 to the end of the BAC at 161475. The complete *L. variegatus* BAC sequence (bottom) is included within this *S. purpuratus* sequence (top), and is used in its entirety in the scan. Coordinate positions in the BACs are indicated, and each tic on the horizontal black lines representing these sequences demarcates 1 kb from the previous tic. Red blocks on the *S. purpuratus* sequence indicate the positions of the *otx* exons, which are identified by the red numerals. Exons are not marked on the *Lytechinus* sequence because the intron/exon boundaries of the gene are not known for this species, and there are indications that some are not identical with those in *S. purpuratus*. The blue lines connecting the two BAC sequences indicate interspecific sequence similarities at the criterion indicated. (A-C) 20 bp sliding window; (D-F) 50 bp sliding window. (A) 100% sequence similarity required within this window; (B) 95% similarity; (C) 90% similarity; (D) 90% similarity; (E) 80% similarity; (F) 70% similarity. In E, the criterion chosen as a guide to experimental examination, the sequence patches used to produce expression constructs are indicated as numbered green boxes.

Figure 3

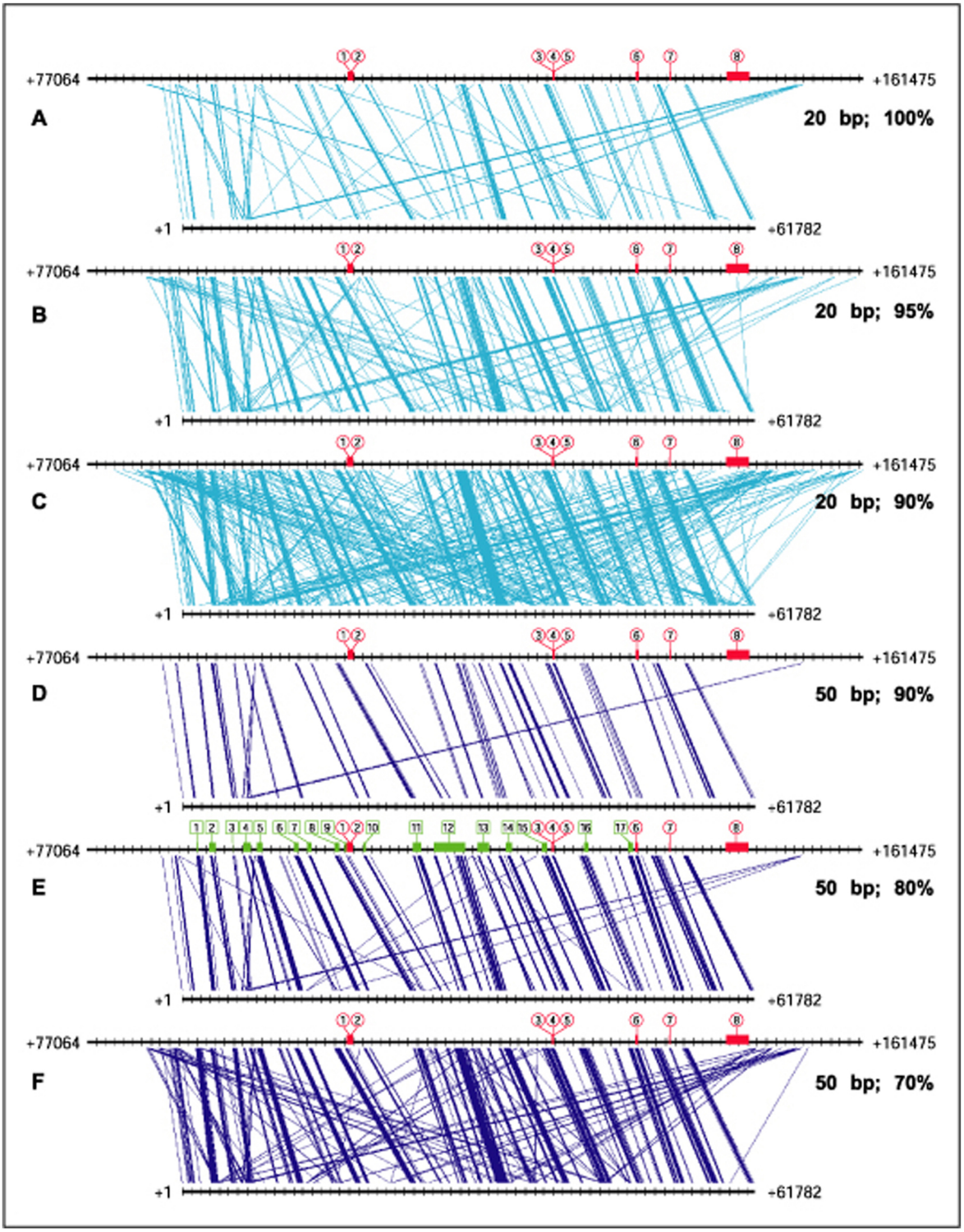

+77064
+161475
A
20 bp; 100%
+1
+61782
+77064
+161475
B
20 bp; 95%
+1
+61782
+77064
+161475
C
20 bp; 90%
+1
+61782
+77064
+161475
D
50 bp; 90%
+1
+61782
+77064
+161475
E
50 bp; 80%
+1
+61782
+77064
+161475
F
50 bp; 70%
+1
+61782

FIG. 4. Examples of *otx cis*-regulatory activity. Four typical embryos are shown for each of the seven different constructs indicated in the left margin. The constructs were injected into fertilized eggs, and the presence of CAT mRNA produced by the reporter gene was assessed by WMISH, as indicated in Materials and Methods. Quantitative data giving the fraction of embryos displaying expression in each of the indicated spatial domains, and the number of cells active per embryo, are given in Table 1. In ectoderm the clones of active cells generally remain contiguous because these cells are not motile, while in the mesodermal cell types and the gut endoderm cells are motile during and after the onset of gastrulation, so descendants of the same initially transgenic blastomere are often scattered rather than clustered. In each panel the location of stained cells is indicated by a one-letter code. For the 24 h mesenchyme blastula-stage embryos in the left two panels of B-G the code is: E, vegetal plate endoderm; M, vegetal plate mesoderm territory (Ruffins and Ettensohn, 1996); S, skeletogenic mesenchyme; C, ectoderm. For the 48 h late gastrulae shown in A, and in the right two panels of B-G, the code is: G, gut; M, secondary mesenchyme, i.e., all mesodermal cell types other than skeletogenic mesenchyme; S, skeletogenic mesenchyme; A, aboral ectoderm; O, oral ectoderm.

Figure 4

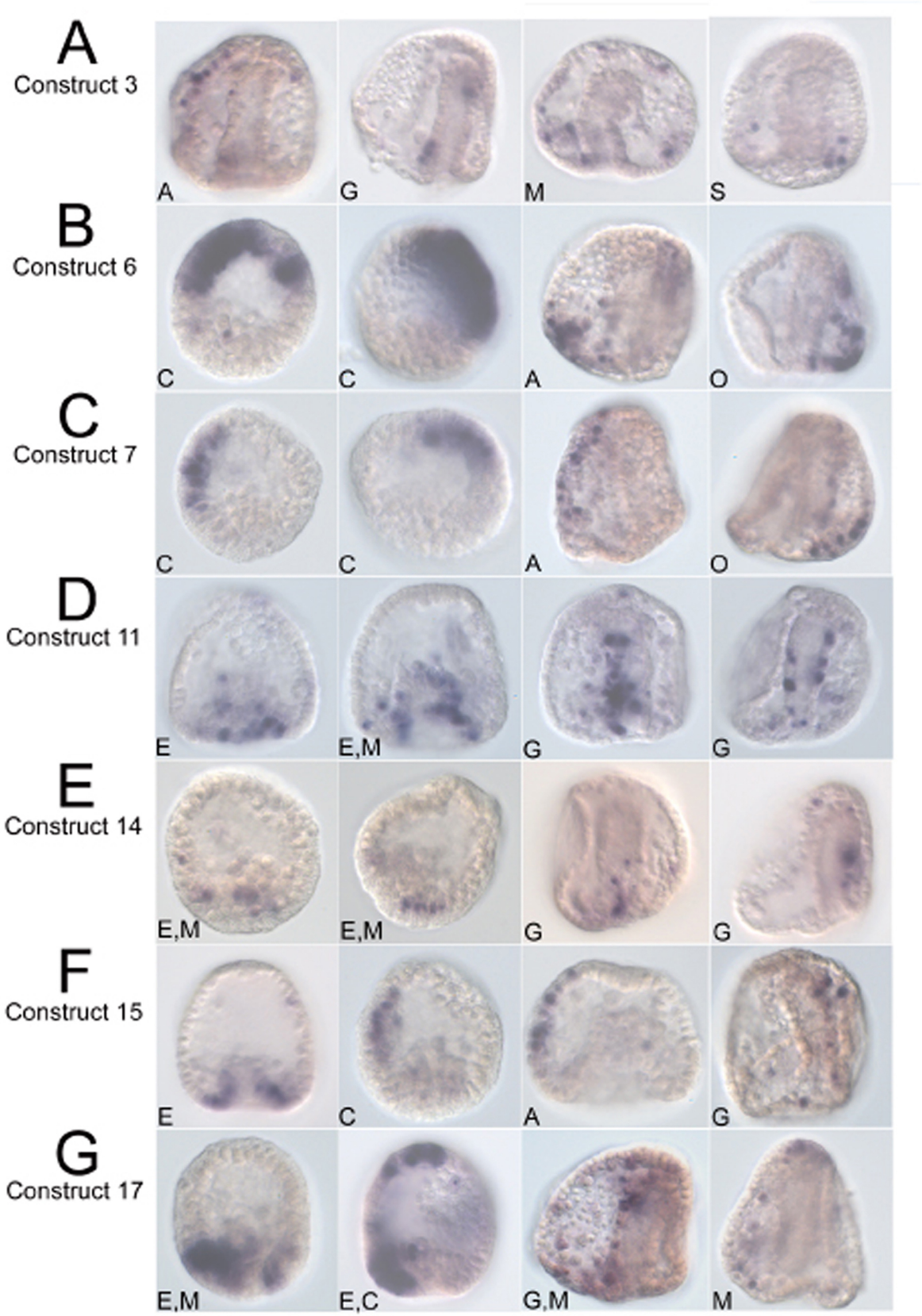
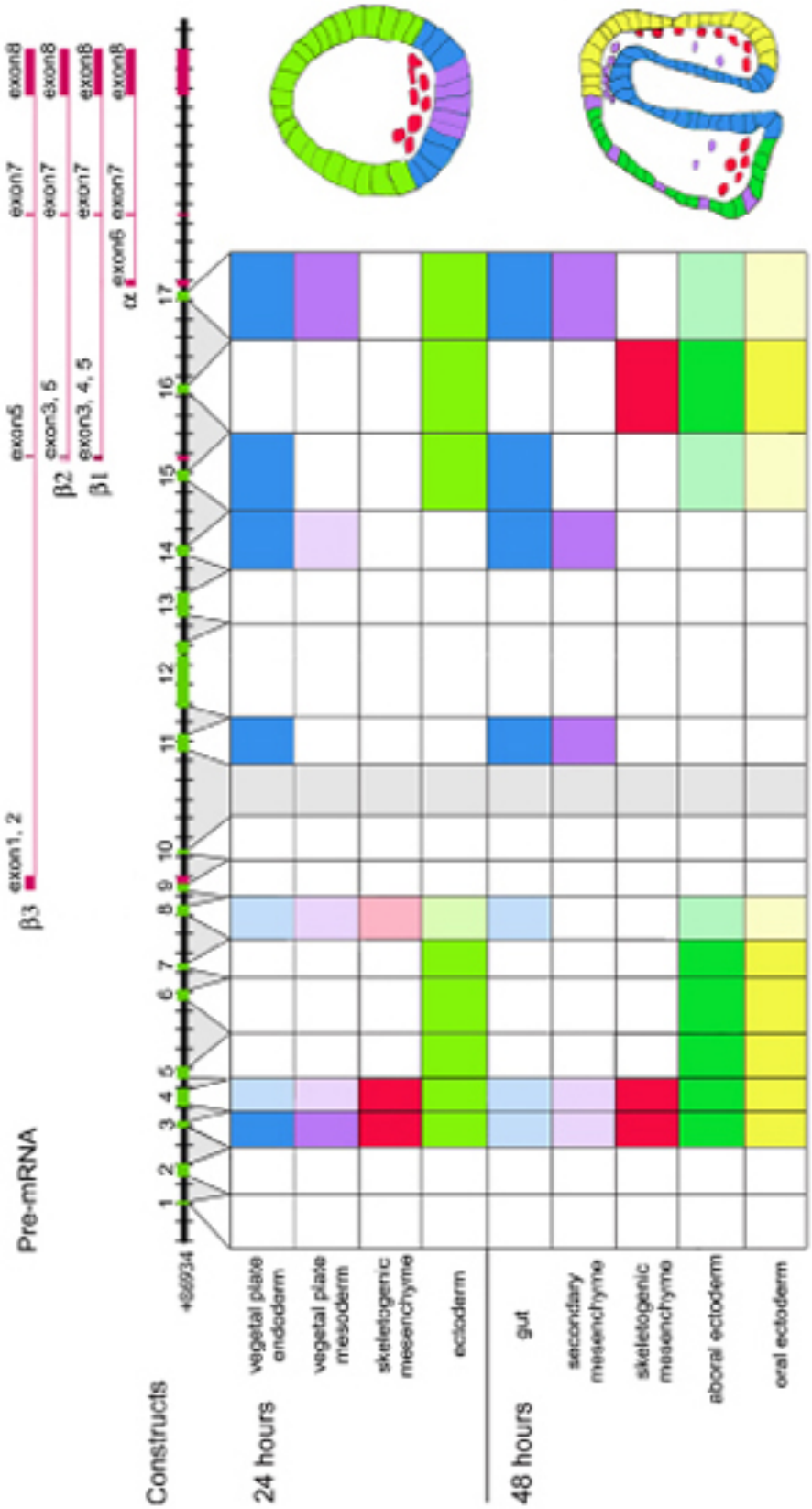

A
Construct 3
A
G
M
S
B
Construct 6
C
C
A
O
C
Construct 7
C
C
A
O
D
Construct 11
E
E,M
G
G
E
Construct 14
E,M
E,M
G
G
F
Construct 15
E
C
A
G
G
Construct 17
E,M
E,C
G,M
M

FIG. 5. Summary of *cis*-regulatory activity of *otx* constructs. At the top of the Figure the transcription map of the *otx* gene is shown, after Li *et al*. (1997). Beneath is the 60 kb region of the *S. purpuratus* BAC 006F13 (see Fig 2) that was used for the FamilyRelations analysis of Fig. 3. The 17 constructs chosen on the basis of this analysis are indicated as green blocks, as in Fig. 3. A color code is used to indicate construct activity, the key to which is given in the diagrams of the 24 h and 48 h embryos at right (the cells indicated in these cartoons are merely presentation aides and are not meant to provide data as to cell number in any particular domain of the embryo in the optical sections represented, nor are the boundaries of those domains intended to be exact representations). In the cartoon of the 48 h embryo the purple (mesodermal) cells embedded in the ectoderm are pigment cells. Dark colors in the central chart indicate strong expression, and light shades weak expression. Data are from Table 1; as can be seen there, strong expression means a high fraction of embryos expressing in a given domain of the embryo and eight or more cells expressing the construct therein (except for a few cases where there are less cells expressing but a very large fraction of embryos display expression compared to the maximum that could be expected from the pattern of mosaic incorporation; Livant *et al*., 1991; Yuh and Davidson, 1996). Weak expression denotes smaller fractions of embryos expressing in the given domain and only 3-5 cells per embryo active in that domain. Here failure of detection (i.e., lower numbers of cells expressing) means activity of insufficient intensity to be scored by WMISH.

Figure 5


Pre-mRNA
β3
exon1, 2
exon5
β2
exon3, 5
exon7
exon8
β1
exon3, 4, 5
exon7
exon8
α
exon6
exon7
exon8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
+86934
Constructs
24 hours
vegetal plate endoderm
vegetal plate mesoderm
skeletogenic mesenchyme
ectoderm
48 hours
gut
secondary mesenchyme
skeletogenic mesenchyme
aboral ectoderm
oral ectoderm