

In silico characterization of the neural alpha tubulin gene promoter of the sea urchin embryo *Paracentrotus lividus* by phylogenetic footprinting

Maria Antonietta Ragusa · Valeria Longo ·
Marco Emanuele · Salvatore Costa ·
Fabrizio Gianguzza

Received: 30 July 2010 / Accepted: 2 June 2011
© Springer Science+Business Media B.V. 2011

Abstract During *Paracentrotus lividus* sea urchin embryo development one alpha and one beta tubulin genes are expressed specifically in the neural cells and they are early end output of the gene regulatory network that specifies the neural commitment. In this paper we have used a comparative genomics approach to identify conserved regulatory elements in the *P. lividus* neural alpha tubulin gene. To this purpose, we have first isolated a genomic clone containing the entire gene plus 4.5 Kb of 5' upstream sequences. Then, we have shown by gene transfer experiments that its non-coding region drives the spatio-temporal gene expression corresponding substantially to that of the endogenous gene. In addition, we have identified by genome and EST sequence analysis the *S. purpuratus* alpha tubulin orthologous gene and we propose a revised annotation of some tubulin family members. Moreover, by computational techniques we delineate at least three putative regulatory regions located both in the upstream region and in the first intron containing putative binding sites for Forkhead and Nkx transcription factor families.

Keywords Sea urchin · Neural development · Gene expression · Phylogenetic footprint · Cis-regulatory analysis

Introduction

Microtubules constitute one of the major components of the cytoskeleton of eukaryotic cells and they are involved in many essential processes, including cell division, ciliary and flagellar motility and intracellular transport. The basic building block of the microtubule is the heterodimeric $\alpha\beta$ -tubulin protein that assembles in a head-to tail arrangement to form a linear protofilament. The subsequent formation of lateral interactions between protofilaments results in their assembly into the wall of the cylindrical microtubule.

The properties of specific tubulin isotypes have an enormous influence upon assembly of the central pair microtubules and associated structures. Although many eukaryotes possess multiple α - and β -tubulin genes, sequence analysis reveals that the aminoacidic sequences of these isotypic variants are generally well conserved, being their divergence limited to the C-terminal region. Many evidences suggest that this extreme C-terminal region is exposed on the outside surface of the microtubule and represents the major source of variation between α and β -tubulin isotypes; consequently it has been proposed that this domain determines the functional specificity of each tubulin isotype.

Moreover, interestingly, many tubulin post-translational modifications occur in the exposed acidic C-terminal domain of α and β -tubulin [1]. Polyglutamylation represents the major posttranslational modification of axonal tubulin in neural cells, where it appears to regulate the differential interaction between microtubules and microtubule-associated proteins

Electronic supplementary material The online version of this article (doi:10.1007/s11033-011-1016-7) contains supplementary material, which is available to authorized users.

M. A. Ragusa (✉) · V. Longo · M. Emanuele · S. Costa ·
F. Gianguzza (✉)
Dipartimento di Scienze e Tecnologie Molecolari e
Biomolecolari, Università degli Studi di Palermo,
Viale delle Scienze, Ed. 16, 90128 Palermo, Italy
e-mail: marag@unipa.it

F. Gianguzza
e-mail: gianfab@unipa.it

(MAPs). In particular Bonnet et al. [2] suggested that the differential binding of MAPs to polyglutamylated tubulin could facilitate their selective recruitment to distinct microtubule populations and thereby could modulate the functional properties of microtubules.

Sea urchins are used as a model for research into the molecular mechanisms of development. The developmental biology of sea urchin embryos is relatively well-known because of their many natural experimental advantages, which were exploited for many decades by embryologists. Molecular biologists have exploited other advantages: the ease of gene transfer into sea urchin eggs, which encourages high-throughput *cis*-regulatory analysis, and the tremendous fecundity of these animals, which permits picomole quantities of rare proteins (e.g., transcription factors) to be purified directly from embryo nuclei [3].

In *Paracentrotus lividus* sea urchin at least five alpha and five beta tubulin genes are differentially expressed during embryogenesis [4, 5], all coding for axonemal tubulin isotypes, one of which (alpha2) is a neural-specific alpha isotype.

The spatial expression pattern of PI-Talpha2 gene, revealed by WMISH, in fact shows that alpha2 transcript is initially expressed in the thickened epithelium of the apical tuft (at the swimming blastula stage) and later in the ciliary band as well as in the apical, oral and lateral ganglia at the pluteus stage [5]. This expression pattern correlates with the cellular commitment to a neural fate [6, 7].

The sea urchin larval nervous system is an array of neurons that control swimming and feeding. There are several clusters of neurons with associated neuropil that are thought to function as ganglia. The animal plate of the sea urchin embryo becomes the apical organ, a sensory structure of the larva; the neurogenic ciliated band becomes morphologically distinct after gastrulation between the oral and aboral ectoderm territories with lineage contributions from both. Duboc et al. [8] proposed that the initial specification of ectoderm is to a ciliary band phenotype and that oral and aboral domains arise later. The vegetal influences, which are interpreted as the effects of nuclearization of beta-catenin, appear to be specification of cell types that restrict the expansion of the animal plate region of ectoderm [9].

In all the morphogenetic process, sequence-specific binding of transcription factors to *cis*-regulatory elements is involved in appropriate spatio-temporal patterns of gene expressions at each specification step. *Cis*-regulatory system could be separated into several *cis*-regulatory modules. To establish a specific pattern of each gene expression, *cis*-regulatory modules are integrated in the promoter/enhancer. In the last years, models for gene regulatory networks (GRN) that are responsible for endomesoderm and oral-aboral ectoderm specification of sea urchin embryos have

been constructed, based on experimental perturbation analyses [10, 11]. Recently, GRN subcircuits controlling individual components of complex developmental processes have been revealed [12].

In order to study gene expression regulation, comparative sequence analysis of model organism genes and genomes has recently been emerging as a complementary approach to experimentation. Echinoderms provide excellent opportunities for comparative sequence analyses of regulatory systems. In fact the *Strongylocentrotus purpuratus* genome assembly [13] and the combined BAC-WGS assembly, released in July 2006 as version (v) 2.1 [14], provide a useful resource to find both orthologous genes and interspersed conserved regulative sequences (ICR).

Dramatic improvements in the specificity of transcription factor binding sites (TFBS) prediction are in fact attained by limiting the search space to regions of conserved non-coding sequences, using a comparative genomics approach. The use of orthologous sequences, also referred to as phylogenetic footprinting, introduces the filtering power of evolutionary constraint to identify putative regulatory regions that stand apart from the background sequence conservation.

The search for both known binding sites (pattern matching) and over-represented novel motifs (pattern detection) can be improved through the analysis of data sets containing orthologous or co-regulated genes (summarized in [15]).

Human and mouse provide a good example, having a common ancestor extant some 100 million years ago (100 Mya). Smith [16] has estimated the phylogenetic relationships for several camarodont species, from which a tentative date of 50 Mya can be derived as the time of divergence between *P. lividus* and *S. purpuratus*.

Previously, we reported that an upstream region of the PITalpha2 promoter, consisting of about 800 bp, is sufficient to drive the temporal expression of a reporter gene congruent with that of endogenous PITalpha2 gene [15]. Here we have identified a new genomic clone (named PI-Tuba1a) containing the entire gene plus few kilobases upstream. This clone differs of about 4 Kb in the 5' upstream region from what previously published. We also demonstrate that the region from -4.5 to +0.8 kb is able to drive the right spatio-temporal expression of GFP reporter gene during *P. lividus* development. Moreover, a combination of in silico methods has been employed to analyze this extended promoter region.

In this study we also propose a refined annotation of some tubulin family members in sea urchin *S. purpuratus* which has allowed the PI-Tuba1a orthologous gene identification.

Finally a comparative sequence analysis of the two genes has allowed the identification of putative *cis*-regulatory elements both in the upstream region and in the introns.

Methods

Genomic clone isolation and analysis

The PITuba1a genomic clone has been isolated from a genomic library in λ EMBL3 prepared as previously reported [17]. This library has been screened by PCR using a primer pair specific for sequences from -720 to -543 nt (E1 fragment) corresponding to the most upstream region of the previously identified PITalpha2 clone. Moreover, a second primer pair specific for the second exon has been used (a2-76 5'-CGTGAATGTATCTCAGTCC-3' and a2-126R 5'-AGGTGTTGAAGGAGTCATC-3'). After purification of PI-Tuba1a DNA, restriction enzyme mapping of this clone has been carried out and a 7.2 kb fragment (including the PITalpha2 clone) subcloned into pBluescript KS+ (Stratagene, La Jolla, CA) and totally sequenced (GenBank accession number: JF272003).

Generation of *cis*-regulatory reporter construct and microinjection

A PI-Tuba1a genomic fragment spanning from -4.5 to $+0.8$ kb has been cloned upstream and in frame to GFP coding region in pGREEN Lantern1 (GIBCO BRL) modified vector, at KpnI site, and verified by sequencing.

In particular, the genomic fragment has been PCR amplified by alphaKpnI-For (ACTCGGTACCGTTCGACA GATTTTCTAACC) and alphaKpnI-Rev (GCATGGTAC CCCATGATGATACATTATTCGAATTCTGAAGT) primers. The primers were designed to have KpnI sites at their 5' ends. To eliminate a stop codon in the polylinker, the pGREEN Lantern1 vector (GIBCO BRL) has been digested with SpeI and SalI restriction enzymes, blunted with Mung Bean nuclease and self-ligated to generate pGREEN Lantern1-modified vector.

KpnI digested PCR fragment has been inserted upstream of the GFP gene into KpnI site of the pGREEN Lantern1-modified plasmid.

High fidelity PCR has been used in the generation of the reporter construct. PCR product has been injected into the embryos after purification.

Unfertilized eggs have been injected with 2 μ l of a solution containing 5 ng/ μ l of reporter construct together with 5% Texas Red-conjugated dextran, following the microinjection and embryo culture procedures already described [17–19].

Injected embryos at the desired stage have been harvested, mounted on glass slides and examined under an epifluorescence Olympus BX50 microscope. DIC, bright-field, or fluorescence images have been captured with a Nikon digital camera and processed using the Nikon Nis Elements software.

Ortholog identification

The sequences of *S. purpuratus* alpha tubulins have been retrieved performing BLAST searches of the sea urchin genome assembly (<http://sugp.caltech.edu/SpBase/>—Database: Spur_v2.1.assembly.all.fa 114,222 sequences; 907,070,087 total letters) [20] from the Spbase database using as query *P. lividus* alpha2 mRNA sequence (GenBank: AY388462) or deduced protein sequence (452 letters). Preliminary alignments have been performed with Kalign (<http://www.ebi.ac.uk/Tools/kalign/> [21]). EST sequences have been retrieved performing BLAST searches of *S. purpuratus* EST database [22] at NCBI (<http://www.ncbi.nlm.nih.gov>); using tubulin coding and putative 3' UTR genomic sequences as queries. Assembled EST tubulin sequences have been aligned with corresponding genomic scaffolds with BLAT (<http://genome.ucsc.edu/cgi-bin/hgBlat>). Multiple sequence alignments of genomic coding sequences (nucleotide and amino acid sequences) have been created with ClustalW2 (<http://www.ebi.ac.uk/Tools/clustalw2/> [23]). Alignment between 5'UTR sequences of mRNAs has been performed also with MAP (<http://searchlauncher.bcm.tmc.edu/multi-align/Options/map.html>) [24]. Gene figures have been constructed using Genepalette [25].

Comparison of genomic sequences and TFBS analysis

The sequence of Sp-Tuba1a genomic locus (from 3380 to 9647 corresponding approximately to from -3000 to $+3300$ bp) has been retrieved from the Spbase database.

Comparison of the genomic sequences around the alpha tubulin genes of *P. lividus* and *S. purpuratus* have been performed with the VISTA platform (<http://genome.lbl.gov/vista/index.shtml> [26]). The window size used varied between 50 and 100 bp, with a window of 50 bp with 70% conservation.

The FamilyRelationsII software package [27] has been also used: window sizes used in the comparison ranged from 10 to 100 bp and the similarity values ranged from 70 to 100%. The pairwise view of the software has been used to identify conserved regions.

The CONREAL web server (<http://conreal.niob.knaw.nl/> [28]) has been used and vertebrate matrices from JASPAR as well as PWMs from public part of TransFac (v.6.0) have been retrieved.

Results and discussion

Genomic clone isolation and gene transfer experiments

To investigate the *cis*-regulatory elements that control the spatio-temporal pattern of the *P. lividus* neural alpha

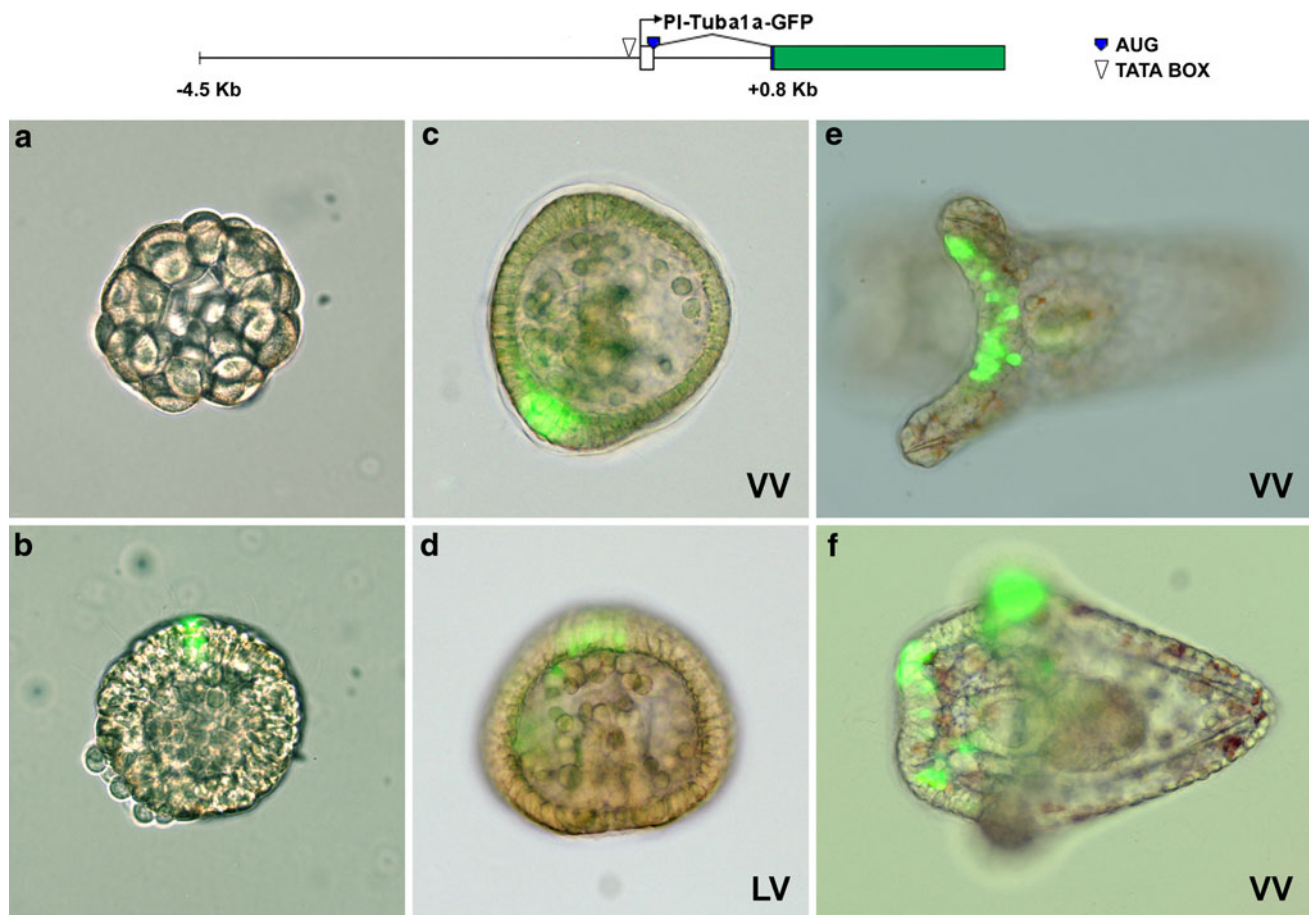


Fig. 1 Expression of the PI-Tuba1a-GFP transgene construct during embryonic development of *P. lividus*. At the top is a schematic drawing of the GFP transgene construct driven by the PI-Tuba1a cis-regulatory apparatus. An arrowhead denotes the TATA-box sequence. The bent arrow indicates the transcription start site. A grey pentagon represents the translation start site. Fluorescence and bright field merged images from transgene microinjected embryos are shown (a–f). Embryos orientation is indicated on the lower right corner.

a Morula stage embryo: at this stage the gene is not expressed. **b** Blastula stage embryo: the first cells show GFP expression at the animal pole. **c–d** Gastrula stage embryos express the transgene in the ciliary band (**c**) and at the apical organ (**d**). **e–f** Pluteus stage embryos: GFP is still localized in the ciliary band (neural cells), both in anal arms (**e**) and in oral arms (**f**). LV lateral view with the animal pole at the top; VV ventral view with the oral side to the left

tubulin gene expression, we isolated an approximately 7.2 kb genomic DNA fragment corresponding to this gene from a genomic library. This genomic DNA fragment spans from -4.5 to $+2.7$ kb and contains the upstream region through to the 3'UTR of the gene (GenBank accession number: JF272003). This genomic clone contains the previously analysed PITalpha2 clone (from -720 to $+2.7$ kb) [17] and 3.7 kb upstream sequences. This was named PI-Tuba1a, respecting the revised nomenclature of the vertebrate tubulin genes [29].

A reporter plasmid, PI-Tuba1a-GFP-pGREEN, in which a fusion protein of alpha1a tubulin and green fluorescent protein (GFP) expresses under the control of a 5.3 kb region of PI-Tuba1a gene (from -4.5 to $+0.8$), was constructed and then used for gene transfer experiments (see “Methods”).

By tracing alpha1a-GFP expression in the injected embryos up to the pluteus stage (Fig. 1), we show that this

genomic fragment is able to drive spatio-temporal regulation of the reporter gene overlapping the endogenous PI-Tuba1a gene. GFP is in fact expressed from blastula to pluteus stage and strictly in the neurogenic territory at the pluteus stage. Statistical analysis of pluteus stage expression pattern, where neural territories are easily identified, is reported in Table 1.

To discover control elements such as enhancers and promoters in genomic DNA sequences, evolutionary relationships can be used. It has been shown, in fact, that control elements between related but moderately divergent species are frequently similar, presumably to conserve a vital function.

Therefore, in order to find a PI-Tuba1a ortholog, we used Genome Assembly Version2.1 of the sea urchin *S. purpuratus* genomic sequence [14, 20]; this sea urchin species diverged from *P. lividus* approximately 50 Mya [16].

Table 1 Summary of the results of four independent microinjection experiments with the PI-Tuba1a-GFP construct

PI-Tuba1a-GFP Gene transfer results—Pluteus stage							
Experiment no.	Injected embryos	GFP-expressing embryos	Ciliary band	Oral ectoderm	Aboral ectoderm	Mesenchyme cells	Endoderm
1	420	327	318	39	68	59	32
2	285	251	232	30	54	58	26
3	304	285	261	35	49	55	24
4	267	238	221	28	41	49	19
Mean (%)		87	94	12	19	20	9
SD (%)		7	3	0	2	2	1

The search of *S. purpuratus* PI-Tuba1a ortholog (Sp-Tuba1a) has not been easy. A reason for this difficulty is inherent to the tubulin features: there is a high level of similarity between members of this gene family. For example, several human alpha tubulin proteins are more than 99% identical, while at the nucleotide level all members of the human alpha tubulin family are more than 72% identical to one another. Despite this extreme tubulin conservation, as already mentioned, the C-terminal region of $\alpha\beta$ -tubulin represents the major source of variation between $\alpha\beta$ -tubulin isotypes and it has often been proposed that this domain determines the functional specificity of tubulin isotypes. Therefore, in order to find Sp-Tuba1a it is particularly important to analyze this region. Moreover, assigning orthology has been more difficult because of some gaps in genome assembly.

Neural alpha tubulin ortholog identification

Initially, we based our analysis on annotated alpha tubulins. Searching alpha tubulin genes in Sp-base [20], we found 14 published items [30] and 7 extra (Table 1 in Online Resource). We collected coding nucleotide sequences and protein sequences and, using Kalign [21], we created preliminary multiple sequence alignments. Unfortunately, looking at the alignments we realized that many alpha tubulin sequences appeared to be annotated incorrectly in public databases (automated analysis-GLEAN prediction).

Thus a revision was needed even because, according to the biological implications of tubulin activity, it is important to provide a correct annotation of alpha tubulin genes in sea urchin DNA sequences.

Therefore, in order to identify the *S. purpuratus* neural alpha tubulin ortholog (Sp-Tuba1a) we used whole genome analysis, and BLAST searches of the sea urchin genome assembly were performed using as query *P. lividus* neural alpha tubulin mRNA sequence or deduced protein sequence. Using nucleotide sequence (blastn, 1520 nt) 18 scaffolds

produced significant alignments with E-value = 0, and using amino acid sequence (tblastn, 452 aa) we got twelve scaffolds with E-value = 0 (Table 2 in Online Resource).

We selected five loci with high percent similarity and the scaffolds were analyzed in Spbase. Many of these scaffolds corresponded to already annotated genes, but after the manual annotation process we adopted new names (Table 2). From these scaffolds we extracted putative coding sequences with unique 3'UTR and these sequences were used as query to verify them and to find correlate 5'UTR sequences. Performing BLAST search in NCBI (*S. purpuratus* EST database [22]) we selected EST sequences with high similarity ($\geq 98\%$) in coding sequence and in 3'UTR. In this way we were able not only to construct in silico mRNA (coding and untranslated regions) alpha tubulin sequences, but also to find for three of them (scaffold_v2_18119, scaffold_v2_86155, scaffold_v2_61948) 5'UTR sequences in the genome. The 5'UTR of tubulin genes can be far from the coding region due to the presence of an intron at codon 1. For scaffold_v2_49057 and scaffold_v2_43416 no EST sequences were found in database (so no 5'UTR sequence was identified), and we supposed they are not expressed at all or they are expressed at very low levels during embryogenesis, not consistent with PI-Tuba1a orthology.

Sp-Atub1_2 sequence is found on the minus strand of scaffold_v2_86155 and it is split up in two contigs (144934 and 173322). Contig173322 has 97% identity with scaffold_v2_18119 (from 8478 to 9402). Part of the third exon of SPU_007435 and SPU_024617 (coding aa 253 to Stop as well as 3'UTR) are almost entirely identical (99% identity).

Concerning scaffold_v2_61948, the tubulin sequence is incomplete, lacking in the first half part of the gene, but assembling EST sequences it was possible to complete the gene sequences founding 5'UTR and 3'UTR unique regions in the scaffold_v2_74218 (contig145029). Moreover, very similar sequences were found in the scaffold_v2_83232, but they are split in two contigs (contig98070 and

Table 2 Five *S. purpuratus* annotated genes (with high percent identity to Pl-Tuba1a) analyzed as ortholog candidates

<i>S. purpuratus</i> scaffold	Annotated gene	Common names	New annotations
scaffold_v2_86155	SPU_007435	Sp-Atub1	Sp-Atub1_2
scaffold_v2_49057	SPU_012679	Sp-Atub13	Sp-Atub13_2
scaffold_v2_61948	SPU_019990	Sp-Atub5	Sp-Atub5
scaffold_v2_74218			
scaffold_v2_83232			
scaffold_v2_43416	SPU_024615	Sp-Atub6	Sp-Atub6_2
scaffold_v2_18119	SPU_024617	Sp-Tuba1c_1	Sp-Tuba1a

Common names and new annotation names are reported

contig98067), and so were not considered. The scaffold_v2_61948 is contained in the scaffold_v2_74218 and closes the gap between contig145028 and contig145029. The comparison of this assembled sequence with annotated tubulin genes has showed it corresponds to Gene ID SPU_019990 (Sp-Atub5).

It is worth noting that the comparison between assembled Sp-Tuba1a mRNA and the corresponding scaffold_v2_18119 shows two cytosines in the coding region of the mRNA (and in all EST sequences examined) but three cytosines in genomic sequence. This further cytosine, if present, would cause a reading frame shift (therefore, it could be a DNA sequencing error).

Based on assembled mRNA sequences we propose a different annotation for SPU_007435 and SPU_024617 and a new annotation for SPU_019990 on a sequence built assembling scaffold_v2_61948 and scaffold_v2_74218 (Online Resource, Table 3).

We collected coding genomic nucleotide sequences and deduced aa sequences and using ClustalW2 [23], we created multiple sequence alignments. Pl-Tuba1a showed very high similarity with all *S. purpuratus* protein sequence selected but not with scaffold_v2_43416 (Sp-Atub6_2). When we performed the same work using nucleotide sequences or when we compared the last 32 aa (C-term domain), maximum score was obtained with scaffold_v2_18119 (Sp-Tuba1a).

The analysis performed using the protein coding nucleotide sequences of the alpha tubulin genes and using the COOH-term amino acid sequence of these proteins suggests an orthology across *P. lividus* and *S. purpuratus* genomes for Pl-Tuba1a and Sp-Tuba1a (scaffold_v2_18119) (Online Resource, Tables 4, 5).

When inference of homology is hampered by low or too high sequence similarity, or in distinguishing true orthologs from paralogs, conserved gene neighborhood (CGN) may assist in finding orthologs. So we also looked for the scaffold containing the homeobox transcription factor gene, PITX2 (GenBank accession number: AY714498, [31]), located upstream to Pl-Tuba1a. BLAST analysis

unfortunately showed that all *S. purpuratus* sequences similar to PITX2 are in different scaffolds from those containing alpha tubulin, probably because of their poor length.

Gene structure similarity

The *P. lividus* alpha1a tubulin gene is organized in three exons: the first one contains the 5'UTR and the open reading frame coding just for the first aa (Met), the second one (800 bp away from the first exon) encodes the aa 2–75 and the third exon (about 330 bp from the second exon) the aa 76–452 and 3'UTR [17].

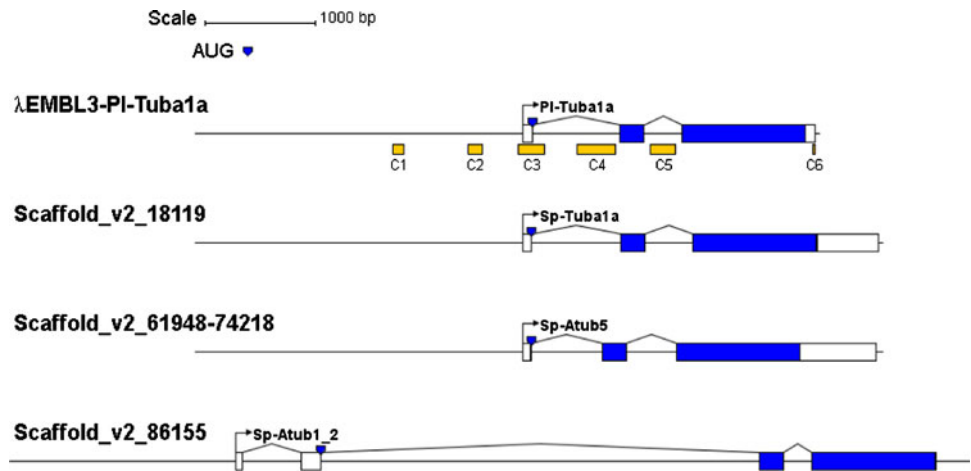
This tubulin gene structure appears more similar to vertebrate than invertebrate ones. In fact, vertebrate tubulin gene classes contain three or rarely four introns: the first two are at codon 1 and 76, the third is located between codon 125 and 126, the fourth at codon 352 [32, 33]. More different are the invertebrate gene structures. In *Drosophila*, for example, the coding regions of alphaTub84D does not contain intervening sequences; alphaTub84B and alphaTub67C contain instead one intervening sequence (between codon 1 and 2) whereas alphaTub85E contains two introns at codon 1 and 176 (FlyBase; [34]).

Every alpha tubulin coding sequence identified in our search on *S. purpuratus* genome can be distributed in two gene classes: in the first one the coding sequence is interrupted by two introns at codon 1 and 76 (like alpha1a), whereas in the second one it is interrupted by three introns at codon 1, 76 and 125.

All alpha tubulin loci, chosen for their over 90% of similarity to Pl-Tuba1a in coding sequence, have also a gene structure similar to that of Pl-Tuba1a (Fig. 2).

A comparison of the intron–exon structure of each gene shows that the most similar tubulin gene is Sp-Tuba1a which presents the same intron position and length as Pl-Tuba1a, although Sp-Atub5 is also quite similar. In contrast the gene structure of Sp-Atub1_2 is different from that of the other genes: the intron following ATG is much longer (4 kb) and it seems there is a further intron in the

Fig. 2 Schematic drawing of gene structure comparison of the PI-Tuba1a gene and the three *S. purpuratus* α -tubulin genes with high percent sequence identity. The *bent arrows* denote the transcription start sites. The *grey pentagons* indicate the translation start sites. *White boxes* represent 5' and 3' UTR, *filled boxes* indicate coding regions. Interspecifically conserved patches of the PI-Tuba1a gene are represented as *clear boxes* drawn under the gene structure scheme. Scale bar 1000 bp



leader (5'UTR). Also this gene structure analysis strongly suggests that PI-Tuba1a and Sp-Tuba1a are orthologous.

Finally, to confirm orthology, we performed an alignment between 5'UTR sequences of the three different mRNAs using both ClustalW2 and MAP [23, 24]. We pointed out a higher similarity between PI-Tuba1a and Sp-Tuba1a (score: 53) than that between PI-Tuba1a and Sp-Atub5 (score: 36).

ICRs identification

In order to identify the *cis*-regulatory regions responsible for the control of PI-Tuba1a expression, we used phylogenetic footprinting. Previous studies have shown that comparisons of genomic sequences between *L. variegatus* and *S. purpuratus* allow efficient identification of the highly conserved regions, expected to contain regulatory elements [35]. Here we scanned the genomic sequences of the two orthologous alpha tubulin genes of *P. lividus* (PI-Tuba1a) and *S. purpuratus* (Sp-Tuba1a) with several programs.

In particular we scanned the full genomic sequences of the two orthologous alpha tubulin genes for short conserved motifs using VISTA program [36] setting 70% identity in a 50 bp window parameters. In addition to regions corresponding to the coding sequences, this comparison identified three conserved non-coding regions. These consist of a 100 bp region (C1) centred at -1150 bp from the transcription start site, a 130 bp region from -510 to -380 bp (C2), a 240 bp one straddling the transcriptional start site (from -50 to $+190$ bp, C3) and also other less conserved regions within the first intron (starting from $+400$ to $+900$, C4), the second intron (C5) and the 3'UTR (C6).

The complete nucleotide sequence of the PI-Tuba1a genomic clone was also compared with the corresponding region of *S. purpuratus* gene by FamilyRelationsII

software [27] (Fig. 3). When a window size of 25 bp and a similarity of 80% were adopted, approximately the same conserved regions were found besides the exons.

Finally, to find candidates for *cis*-regulatory elements within each conserved region, we searched for TFBS that can be found within the 7.2 kb region of PI-Tuba1a gene and located also in the of Sp-Tuba1a gene.

To evaluate putative binding sites we used CONREAL (conserved regulatory elements anchored alignment) web server [28] that allows identification of TFBS that are conserved between two orthologous promoter sequences. The predictions were performed by four different methods (CONREAL-, LAGAN-, MAVID- and BLASTZ-based) and results were compared to each other. Because few DNA-binding site sequences for the sea urchin transcription factors are known, DNA-binding consensus sequences known from other organisms were used (vertebrates or *Drosophila*). By this search, we found a number of TFBS in the conserved regions previously detected and among them we chose those with a similarity score above a threshold of 7.5 (Table 3).

Among all the possible TFBS in silico identified, in particular our attention was focused on those binding sites corresponding to vertebrate nerve transcription factors as candidates for *cis*-regulatory elements and we referred also to sea urchin candidate neurogenic genes defined analyzing the *S. purpuratus* genome and their temporal and spatial patterns of expression during embryogenesis [37].

The nucleotide sequence of element C1 contains potential binding sites for interactions with POU1F1, Myb, and MIF-1. SpMyb is a member of the Myb family of transcription factors. SpMyb functions as an intramodular repressor to regulate the spatial expression of CyIIIa in the oral ectoderm and skeletogenic mesenchyme [38].

The four founder members of the POU family play a critical role in regulating gene expression, particularly in cells of the nervous system. Thus, for example, the Pit-1

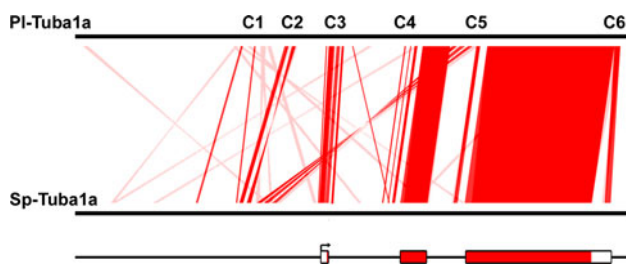


Fig. 3 View of a FamilyRelationsII comparison of *P. lividus* and *S. purpuratus* orthologous sequences in alpha1a tubulin gene. At the top interspecifically conserved regions of the PI-Tuba1a gene are highlighted. The ICRs correspond to positions: -1200 to -1000 (C1), -510 to -380 (C2), -50 to $+190$ (C3), $+400$ to $+900$ (C4), $+1180$ to 1460 (C5), $+2550$ to $+2650$ (C6). At the bottom is the gene structure map to highlight ICR positions respect to exons

protein plays an essential role in the development of the pituitary gland and its absence therefore, results in dwarfism in both mice and humans. Similarly, the *unc-86* mutation in the nematode results in a failure to form specific neuronal cell types, especially sensory neurons (for a review see [39]).

The fork head class TFBS (detected in C2 and in C3 modules) are the most interesting since many transcription factors of this class are expressed in the same domains of PI-Tuba1a tubulin, the neurogenic domains of the ectoderm [40]. Consensus sequences present in databases recognized by many members of this family are very similar, so it is difficult to discriminate them. Anyway, FoxG is the earliest Fox gene expressed in the oral ectoderm of the embryo and by the end of gastrulation it is expressed in the ciliated band. FoxQ2 is expressed in all mesomeres at cleavage stage and by blastula stage it begins to be localized in the apical plate cells until larval stage. FoxJ1 and FoxQ2 are both expressed exactly in the apical organ in a subset of cells derived from what is initially the animal pole of the embryo [40, 41 and Spur WISH database (<http://goblet.molgen.mpg.de/eugene/cgi/eugene.pl>)]. FoxD has a pattern of expression similar to FoxJ1, but is expressed at later stages of development and also in the endoderm.

In contrast, differently from FoxD and FoxJ1, Fox J2 is expressed at low level in secondary mesenchyme and therefore, also for its maternal expression, it is not a good candidate for PI-Tuba1a regulation.

Moreover, it is noteworthy that Fox G and FoxQ contain engrailed homology-1 motif (eh1) which interacts with Groucho and they belong to a class of genes known to function as inhibitors of TGFbeta signaling.

Finally it is remarkable that the expression profile of PI-Tuba1a seems to overlap the J1/Q2 + G sequential expression.

It is also noteworthy the presence in C3 region of a binding site for a homeodomain factor of deformed class (antennapedia family) even if *S. purpuratus* lacks a Hox4 gene [42]. In fact the 11 hox genes display a unique order in which one end begins with the anterior genes (Hox1 and Hox2) and Hox3 gene and it continues counting downward from the posterior Hox11/13c to the central group gene, Hox5. Thus, the anterior genes lie next to the posterior ones but in reverse order. This arrangement is unique for sea urchins and, given that it involves a break between Hox3 and Hox5, it was suggested that the breakage is responsible for the loss of Hox4 [42].

Several Hox genes have an evolutionary sister gene or paralogue, called ParaHox gene [43]. In the sea urchin ParaHox genes are dispersed in the genome and are used during embryogenesis [44]. There are three ParaHox genes: Gsx, the paralogue of the anterior Hox genes; Xlox, the paralogue of Hox3; and Cdx, the paralogue of the posterior Hox genes. Generally, in those animals in which the spatial expression has been determined, Gsx is expressed in or near the anterior regions of the embryo, Xlox is mostly expressed in central regions while Cdx is expressed at the posterior end [45]. However, there is a discordance in the tissues of expression because Gsx is usually expressed in the neural tissue while Xlox and Cdx are expressed in the endoderm.

Two of the three ParaHox genes discovered through the *S. purpuratus* genome project, Sp-lox and Sp-Cdx, are expressed in the developing gut with nested domains in a

Table 3 The presumptive TFBS detected in the conserved regions of PI-Tuba1a gene and Sp-Tuba1a

ICR	Presumptive TFBS (score)				
C1-Upstream ($-1200/-1100$ bp)	POU1F1 (9.3)	Myb (8.0)	MIF-1 (11.0)		
C2-Upstream ($-510/-380$ bp)	POU3F2/Brm-2 (9.4)	FOXJ1/D3/FoxJ2 (11.6)	FREAC-4 (8.2)		
C3-core promoter ($-50/+190$ bp)	TBP (12.0)	FOXJ1/D3/FoxJ2 (9.9)	NKX6B (8.0)	FREAC-4 (9.1)	OCT/POU2F1 (11.0)
	NCX (7.6)	HMG1 (8.7)	DFD (8.0)		
C4-I intron ($+400/+500$ bp)	POU3F2/MEF2 (10.7)	POUF1/PIT (10.9)	PAX8 (7.8)	FOXN1 (8.6)	
C5 II intron ($+1160/+1390$ bp)	TEF (9.2)	FOXJ1/D3 (7.5)	DEC1 (8.4)	TFE (8.3)	
C6-3' UTR ($+2650/+2670$ bp)	FOXG1/O1/O4 (9.1)	FOXJ1/D3 (9.2)			

spatially co-linear manner. However, transcripts of Sp-Gsx, although anterior of Sp-lox, are detected in the ectoderm and not in the gut [44].

Sp-Gsx is detectable from 24 h; it is clearly detected at gastrula stage (reaching a maximum accumulation of transcript at 48 h) until pluteus stage. Its expression domain is confined to two small patches of ectodermal cells, one in each side of the embryo, apparently located in the vegetal half of the embryo at gastrula stage and more clearly at the level of the midgut in plutei [44]. So, even though it was suggested that these cells could be asynchronously dividing neuroblasts, none of parahox genes seem to be Tuba1a regulators because they are expressed later in development and in different regions.

C3 region contains also putative Ncx and Nkx6 binding sites. CONREAL analysis predicts a putative binding site for Ncx, a member of Hox11 family proteins. In mice the Ncx gene (Enx, Hox11L1) is specifically expressed in neural crest derived tissues such as dorsal root ganglia, cranial ganglia, sympathetic ganglia, adrenal medulla and enteric ganglia [46]. Ncx-deficient mice develop megacolon with increased numbers of neuronal cells in enteric ganglia [47, 48].

Nkx6.1 is expressed early in the medial neural plate of both higher and lower vertebrates [49], suggesting that vertebrate Nkx6 genes may have an early role in neural patterning (earlier than Nkx2 genes). Both Nkx2.2 and Nkx6 genes are required for motor neuron development [50]. In *Drosophila*, the Nkx6 gene is also expressed in the midline cells and in a subset of ventral (medial) and intermediate column neuroblasts. In *Drosophila* and vertebrates, Nkx6 genes are closely associated with the development of medial column neurons and in vertebrate lineages they may perform some of the functions of vnd in *Drosophila*. For example, Nkx6.1 represses interneuron development from the intermediate column in zebrafish [48–50].

In sea urchin embryo the apical organ, derived from the animal plate, expresses the transcription factor SpNK2.1 in gastrulae and the serotonergic cells form at the interface of the SpNK2.1 expression domain and the aboral ectoderm [51]; Nkx2.2 instead is expressed in aboral ectoderm [52].

Discussion and conclusion

In conclusion it is clear that understanding the structure and composition of gene regulatory networks that specify neurons is critical for learning how the sea urchin nervous system develops. Although little is currently known about neural specification in either larvae or adults, echinoderms are in an informative phylogenetic position and have the potential to contribute to our understanding of neural

evolution. Here, our objective was to analyze the regulatory region of a neural gene like alpha1 tubulin to help to define neurogenic gene regulatory networks in the embryo.

For this purpose we isolated a genomic clone containing all the regulative sequences that are necessary for the correct gene expression, as shown by gene transfer experiments. It is worth noting that in a previous work we tested (by gene transfer experiment) a different construct lacking of the most upstream region and containing a different reporter gene. Above all we focused the attention on the expression until the gastrula stage. In this work we extended the analysis, focusing our interest on the expression at the pluteus stage, at which neural territories are better distinguishable, using a suitable reporter gene. In comparison with the previous results, these gene transfer experiments show a lower ectopic expression especially at gastrula stage. This is probably due to the presence of a wide upstream region (that contains the C1 patch) that could act synergistically with the other regulative sequences (that contain other modules).

We then used a phylogenetic approach, searching the alpha1 tubulin ortholog in *S. purpuratus* genome. In *S. purpuratus* a cDNA coding for an alpha tubulin has been isolated and the transcript has been detected, like Pl-Tuba1a, in the apical plate by WMISH [53].

On the other hand, it is known that a microtubule is composed by heterodimers of alpha and beta-tubulin and that microtubule behavior varies according to isotype composition. This suggests that each isotype may have properties necessary for specific cellular functions. Moreover, differences among isotypes are often highly conserved in evolution, suggesting that they have functional significance (for a review see [54]). We also have previously shown that in *P. lividus* one alpha and one beta transcript exist (Pl-Tuba1a and PITbeta3) and these only are expressed exclusively in the neural cells at the time of neural cell differentiation [55]. These transcripts code for tubulin isotypes showing the characteristics of neural specific isotypes (i.e., the COOH-term) [55]. Thus, the neural unique properties and spatio-temporal expression pattern of Pl-Tuba1a and Pl-Tbeta3 suggest that this α/β heterodimer could have a specific function for nervous system development. Therefore, the fact that in *S. purpuratus* a β -tubulin transcript has identical expression as Pl-Tbeta3 (Spur WISH database) and an α -tubulin transcript has identical expression as Pl-Tuba1a [53], strongly suggests the existence in *S. purpuratus* of an α - and a β -tubulin transcript, encoding a neural heterodimer structurally and functionally similar to Pl-Tuba1a/PITbeta3. So in our research, using several neural specific features (e.g., COOH-term specific sequence signature, 5' and 3'UTRs), we were able to identify Sp-Tuba1a gene as the real ortholog.

Finally, using the comparison between the two regulatory sequences we were able to find interspecific conserved regions and to detect putative transcription factors binding sites therein. Phylogenetic footprinting was performed utilizing powerful biocomputational tools allowing false positive elimination and conducted with stringent settings.

Searches of the *S. purpuratus* genome have identified homologues of a number of genes known to encode transcription factors that are expressed in neural tissues in other animals [37]. Q-PCR, in situ hybridization or microarray analyses indicate that many of these genes are expressed during early development [37 and references therein]. Expression of several of these genes begins before gastrulation, the time when neuroblasts are first detected. Good candidates for alpha1a gene activation could be Sp-Ac-Sc (achaete-scute, basic-HLH atonal class) and Sp-Hbn (homeobrain, paired-class), homologues of genes known to be involved in early neural specification in other systems. They are expressed exclusively in the neurogenic ectoderm at the animal pole and they begin to be expressed between the hatching blastula and early mesenchyme blastula stages, before the onset of morphogenesis. Other sea urchin genes expressed exclusively in animal plate ectoderm encode the proneural, atonal class factors, neurogenin and neuroD, that are expressed later in development, apparently exclusively in neural cells [37].

The best candidates could be those transcription factors encoded by genes that are expressed exactly in the apical organ. These are the transcription factors FoxJ, FoxQ2, Mox, glass, hpf4 (a zinc finger gene) etc. [40]. Mox is specific for serotonergic neurons, whereas glass localizes in cells adjacent to serotonergic cells. FoxQ2 is expressed in the neurogenic ectoderm.

Other transcription factors are expressed in cells of the ciliary band (Sp-Hnf6, FoxG), a strip of cells at the interface of oral and aboral ectoderm [41, 56] and/or later in a region of ectoderm bordering serotonergic neurons (SpNk2.1) where the oral ectoderm contacts aboral ectoderm. Although NK2.1 is not required for initial differentiation of embryonic neurons [51], its role in the nervous system development and in particular in neural alpha tubulin expression has been supported by MASO injections [53].

On the other hand, it is well-known that homeodomain containing Hox proteins (we found Hox4 and Hox11 binding sites) alone bind to very similar and frequently occurring sequences in the genome. This implies that proper spatio-temporal Hox target gene regulation is achieved by the combined action of multiple transcriptional regulators that interact with Hox proteins themselves [57].

Which of these *cis*-acting putative regulatory elements (and the factors which bind to them) are really involved in transcriptional regulation of PI-Tuba1a gene will come

from *P. lividus* microinjection functional experiments in which the expression of a reporter gene (GFP or Luc), driven by the conserved interspecific regions (or a combination of them), will be studied. Moreover, to reveal the inputs to each of these in silico identified *cis*-regulatory modules, gene transfer experiments will be also performed using site-directed mutants in PI-Tuba1a regulatory region.

Acknowledgments This study was supported by the University of Palermo, Italy, grant from MIUR “ex 60%” to F.G.

References

- Rosenbaum J (2000) Cytoskeleton: functions for tubulin modifications at last. *Curr Biol* 10:R801–R803. doi:[10.1016/S0960-9822\(00\)00767-3](https://doi.org/10.1016/S0960-9822(00)00767-3)
- Bonnet C, Boucher D, Lazereg S, Pedrotti B, Islam K, Denoulet P, Larcher JC (2001) Differential binding regulation of microtubule-associated proteins MAP1A, MAP1B, and MAP2 by tubulin polyglutamylation. *J Biol Chem* 276:12839–12848. doi:[10.1074/jbc.M011380200](https://doi.org/10.1074/jbc.M011380200)
- Calzone FJ, Höög C, Teplow DB, Cutting AE, Zeller RW, Britten RJ, Davidson EH (1991) Gene regulatory factors of the sea urchin embryo. I. Purification by affinity chromatography and cloning of P3A2, a novel DNA-binding protein. *Development* 112:335–350
- Gianguzza F, Di Bernardo MG, Sollazzo M, Palla F, Ciaccio M, Carra E, Spinelli G (1989) DNA sequence and pattern of expression of the sea urchin (*Paracentrotus lividus*) alpha-tubulin genes. *Mol Reprod Dev* 1:170–181. doi:[10.1002/mrd.1080010305](https://doi.org/10.1002/mrd.1080010305)
- Gianguzza F, Casano C, Ragusa M (1995) Alpha-tubulin marker gene of neural territory of sea urchin embryos detected by whole-mount in situ hybridization. *Int J Dev Biol* 39:477–483
- Sasaki H, Kominami T (2008) Specification process of animal plate in the sea urchin embryo. *Dev Growth Differ* 50:595–606. doi:[10.1111/j.1440-169X.2008.01057.x](https://doi.org/10.1111/j.1440-169X.2008.01057.x)
- Yaguchi S, Yaguchi J, Burke RD (2006) Specification of ectoderm restricts the size of the animal plate and patterns neurogenesis in sea urchin embryos. *Development* 133:2337–2346. doi:[10.1242/dev.02396](https://doi.org/10.1242/dev.02396)
- Duboc V, Röttinger E, Besnardeau L, Lepage T (2004) Nodal and BMP2/4 signaling organizes the oral-aboral axis of the sea urchin embryo. *Dev Cell* 6:397–410. doi:[10.1016/S1534-5807\(04\)00056-5](https://doi.org/10.1016/S1534-5807(04)00056-5)
- Yaguchi S, Yaguchi J, Burke RD (2007) Sp-Smad2/3 mediates patterning of neurogenic ectoderm by nodal in the sea urchin embryo. *Dev Biol* 302:494–503. doi:[10.1016/j.ydbio.2006.10.010](https://doi.org/10.1016/j.ydbio.2006.10.010)
- Ben-Tabou de-Leon S, Davidson EH (2006) Deciphering the underlying mechanism of specification and differentiation: the sea urchin gene regulatory network. *Science Signaling STKE*. pe47. doi: [10.1126/stke.3612006pe47](https://doi.org/10.1126/stke.3612006pe47)
- Su YH, Li E, Geiss GK, Longabaugh WJ, Krämer A, Davidson EH (2009) A perturbation model of the gene regulatory network for oral and aboral ectoderm specification in the sea urchin embryo. *Dev Biol* 329:410–421. doi:[10.1016/j.ydbio.2009.02.029](https://doi.org/10.1016/j.ydbio.2009.02.029)
- Davidson EH (2009) Network design principles from the sea urchin embryo. *Curr Opin Genet Dev* 19:535–540. doi:[10.1016/j.gde.2009.10.007](https://doi.org/10.1016/j.gde.2009.10.007)
- Cameron RA, Mahairas G, Rast JP et al (2000) A sea urchin genome project: sequence scan, virtual map, and additional resources. *Proc Natl Acad Sci USA* 97:9514–9518. doi:[10.1073/pnas.160261897](https://doi.org/10.1073/pnas.160261897)

14. Sea Urchin Genome Sequencing Consortium (2006) The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314(5801):941–952. doi:[10.1126/science.1133609](https://doi.org/10.1126/science.1133609)
15. Elnitski L, Jin VX, Farnham PJ, Jones SJ (2006) Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. *Genome Res* 16:1455–1464. doi:[10.1101/gr.4140006](https://doi.org/10.1101/gr.4140006)
16. Smith AB (1988) Phylogenetic relationship, divergence times, and rates of molecular evolution for camarodont sea urchins. *Mol Biol Evol* 5:345–365
17. Costa S, Ragusa MA, Drago G, Casano C, Alaimo G, Guida N, Gianguzza F (2004) Sea urchin neural alpha2 tubulin gene: isolation and promoter analysis. *Biochem Biophys Res Commun* 316:446–453. doi:[10.1016/j.bbrc.2004.02.070](https://doi.org/10.1016/j.bbrc.2004.02.070)
18. Arnone MI, Dmochowski IJ, Gache C (2004) Using reporter genes to study *cis*-regulatory elements. *Methods Cell Biol* 74:621–652. doi:[10.1016/S0091-679X\(04\)74025-X](https://doi.org/10.1016/S0091-679X(04)74025-X)
19. McMahon AP, Flytzanis CN, Hough-Evans BR, Katula KS, Britten RJ, Davidson EH (1985) Introduction of cloned DNA into sea urchin egg cytoplasm: replication and persistence during embryogenesis. *Dev Biol* 108:420–430. doi:[10.1016/0012-1606\(85\)90045-4](https://doi.org/10.1016/0012-1606(85)90045-4)
20. Cameron RA, Samanta M, Yuan A, He D, Davidson E (2009) SpBase: the sea urchin genome database and web site. *Nucleic Acids Res* 37:D750–D754. doi:[10.1093/nar/gkn887](https://doi.org/10.1093/nar/gkn887)
21. Lassmann T, Sonnhammer EL (2005) Kalign—an accurate and fast multiple sequence alignment algorithm. *BMC Bioinform* 6:298. doi:[10.1186/1471-2105-6-298](https://doi.org/10.1186/1471-2105-6-298)
22. Poustka AJ, Groth D, Hennig S, Thamm S, Cameron A, Beck A, Reinhardt R, Herwig R, Panopoulou G, Lehrach H (2003) Generation, annotation, evolutionary analysis, and database integration of 20,000 unique sea urchin EST clusters. *Genome Res* 13:2736–2746. doi:[10.1101/gr.1674103](https://doi.org/10.1101/gr.1674103)
23. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. doi:[10.1093/bioinformatics/btm404](https://doi.org/10.1093/bioinformatics/btm404)
24. Huang X (1994) On global sequence alignment. *Comput Appl Biosci* 10:227–235. doi:[10.1093/bioinformatics/10.3.227](https://doi.org/10.1093/bioinformatics/10.3.227)
25. Rebeiz M, Posakony JW (2004) GenePalette: a universal software tool for genome sequence visualization and analysis. *Dev Biol* 271:431–438. doi:[10.1016/j.ydbio.2004.04.011](https://doi.org/10.1016/j.ydbio.2004.04.011)
26. Dubchak I, Ryaboy DV (2006) VISTA family of computational tools for comparative analysis of DNA sequences and whole genomes. *Methods Mol Biol* 338:69–89. doi:[10.1385/1-59745-097-9:69](https://doi.org/10.1385/1-59745-097-9:69)
27. Brown CT, Xie Y, Davidson EH, Cameron RA (2005) Paircomp, FamilyRelationsII and Cartwheel: tools for interspecific sequence comparison. *BMC Bioinform* 6:70. doi:[10.1186/1471-2105-6-70](https://doi.org/10.1186/1471-2105-6-70)
28. Berezikov E, Guryev V, Plasterk RH, Cuppen E (2004) CONREAL: conserved regulatory elements anchored alignment algorithm for identification of transcription factor binding sites by phylogenetic footprinting. *Genome Res* 14:170–178. doi:[10.1101/gr.1642804](https://doi.org/10.1101/gr.1642804)
29. Khodiyar VK, Maltais LJ, Ruef BJ, Sneddon KM, Smith JR, Shimoyama M, Cabral F, Dumontet C, Dutcher SK, Harvey RJ, Lafanechère L, Murray JM, Nogales E, Piquemal D, Stanchi F, Povey S, Lovering RC (2007) A revised nomenclature for the human and rodent alpha-tubulin gene family. *Genomics* 90:285–289. doi:[10.1016/j.ygeno.2007.04.008](https://doi.org/10.1016/j.ygeno.2007.04.008)
30. Morris RL, Hoffman MP, Obar RA, McCafferty SS, Gibbons IR, Leone AD, Cool J, Allgood EL, Musante AM, Judkins KM, Rossetti BJ, Rawson AP, Burgess DR (2006) Analysis of cytoskeletal and motility proteins in the sea urchin genome assembly. *Dev Biol* 300:219–237. doi:[10.1016/j.ydbio.2006.08.052](https://doi.org/10.1016/j.ydbio.2006.08.052)
31. Duboc V, Röttinger E, Lapraz F, Besnardeau L, Lepage T (2005) Left-right asymmetry in the sea urchin embryo is regulated by nodal signaling on the right side. *Dev Cell* 9(1):147–158. doi:[10.1016/j.devcel.2005.05.008](https://doi.org/10.1016/j.devcel.2005.05.008)
32. Dibb NJ, Newman AJ (1989) Evidence that introns arose at proto-splice sites. *EMBO J* 8:2015–2021
33. Perumal BS, Sakharkar KR, Chow VT, Pandjassaram K, Sakharkar MK (2005) Intron position conservation across eukaryotic lineages in tubulin genes. *Front Biosci* 10:2412–2419. doi:[10.2741/1706](https://doi.org/10.2741/1706)
34. Grumblin G, Strelets V (2006) FlyBase: anatomical data, images and queries. *Nucleic Acids Res* 34:D484–D488. doi:[10.1093/nar/gkj068](https://doi.org/10.1093/nar/gkj068)
35. Yuh C, Brown CT, Livi CB, Rowen L, Clarke PJC, Davidson EH (2002) Patchy interspecific sequence similarities efficiently identify positive *cis*-regulatory elements in the sea urchin. *Dev Biol* 246(1):148–161. doi:[10.1006/dbio.2002.0618](https://doi.org/10.1006/dbio.2002.0618)
36. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32:W273–W279. doi:[10.1093/nar/gkh458](https://doi.org/10.1093/nar/gkh458)
37. Burke RD, Angerer LM, Elphick MR, Humphrey GW, Yaguchi S, Kiyama T, Liang S, Mu X, Agca C, Klein WH, Brandhorst BP, Rowe M, Wilson K, Churcher AM, Taylor JS, Chen N, Murray G, Wang D, Mellott D, Olinski R, Hallböök F, Thorndyke MC (2006) A genomic view of the sea urchin nervous system. *Dev Biol* 300:434–460. doi:[10.1016/j.ydbio.2006.08.007](https://doi.org/10.1016/j.ydbio.2006.08.007)
38. Coffman JA, Kirchner CV, Harrington MG, Davidson EH (1997) SpMyb functions as an intramodular repressor to regulate spatial expression of CyIIIa in sea urchin embryos. *Development* 124:4717–4727
39. Phillips K, Luisi B (2000) The virtuoso of versatility: POU proteins that flex to fit. *J Mol Biol* 302:1023–1039. doi:[10.1006/jmbi.2000.4107](https://doi.org/10.1006/jmbi.2000.4107)
40. Poustka AJ, Kühn A, Groth D, Weise V, Yaguchi S, Burke RD, Herwig R, Lehrach H, Panopoulou G (2007) A global view of gene expression in lithium and zinc treated sea urchin embryos: new components of gene regulatory networks. *Genome Biol* 8:R85. doi:[10.1186/gb-2007-8-5-r85](https://doi.org/10.1186/gb-2007-8-5-r85)
41. Tu Q, Brown CT, Davidson EH, Oliveri P (2006) Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev Biol* 300:49–62. doi:[10.1016/j.ydbio.2006.09.031](https://doi.org/10.1016/j.ydbio.2006.09.031)
42. Cameron RA, Rowen L, Nesbitt R, Bloom S, Rast JP, Berney K, Arenas-Mena C, Martinez P, Lucas S, Richardson PM, Davidson EH, Peterson KJ, Hood L (2006) Unusual gene order and organization of the sea urchin hox cluster. *J Exp Zool B Mol Dev Evol* 306:45–58. doi:[10.1002/jez.b.21070](https://doi.org/10.1002/jez.b.21070)
43. Garcia-Fernández J (2005) Hox, ParaHox, ProtoHox: facts and guesses. *Heredity* 94:145–152. doi:[10.1038/sj.hdy.6800621](https://doi.org/10.1038/sj.hdy.6800621)
44. Arnone MI, Rizzo F, Annunziata R, Cameron RA, Peterson KJ, Martínez P (2006) Genetic organization and embryonic expression of the ParaHox genes in the sea urchin *S. purpuratus*: insights into the relationship between clustering and colinearity. *Dev Biol* 300:63–73. doi:[10.1016/j.ydbio.2006.07.037](https://doi.org/10.1016/j.ydbio.2006.07.037)
45. Brooke NM, Garcia-Fernández J, Holland PW (1998) The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* 392:920–922. doi:[10.1038/31933](https://doi.org/10.1038/31933)
46. Hatano M, Iitsuka Y, Yamamoto H, Dezawa M, Yusa S, Kohno Y, Tokuhisa T (1997) Ncx, a Hox11 related gene, is expressed in a variety of tissues derived from neural crest cells. *Anat Embryol* 195:419–425. doi:[10.1007/s004290050061](https://doi.org/10.1007/s004290050061)
47. Shirasawa S, Yunker AMR, Roth KA, Brown GA, Horning S, Korsmeyer SJ (1997) Enx (Hox11L1)-deficient mice develop myenteric neuronal hyperplasia and megacolon. *Nat Med* 3:646–650. doi:[10.1038/mm0697-646](https://doi.org/10.1038/mm0697-646)
48. McMahon AP (2000) Neural patterning: the role of Nkx genes in the ventral spinal cord. *Genes Dev* 14:2261–2264. doi:[10.1101/gad.840800](https://doi.org/10.1101/gad.840800)

49. Cheesman SE, Layden MJ, Von Ohlen T, Doe CQ, Eisen JS (2004) Zebrafish and fly Nkx6 proteins have similar CNS expression patterns and regulate motoneuron formation. *Development* 131:5221–5232. doi:[10.1242/dev.01397](https://doi.org/10.1242/dev.01397)
50. Pattyn A, Vallstedt A, Dias JM, Sander M, Ericson J (2003) Complementary roles for Nkx6 and Nkx2 class proteins in the establishment of motoneuron identity in the hindbrain. *Development* 130:4149–4159. doi:[10.1242/dev.00641](https://doi.org/10.1242/dev.00641)
51. Takacs CM, Amore G, Oliveri P, Poustka AJ, Wang D, Burke RD, Peterson KJ (2004) Expression of an NK2 homeodomain gene in the apical ectoderm defines a new territory in the early sea urchin embryo. *Dev Biol* 269:152–164. doi:[10.1016/j.ydbio.2004.01.023](https://doi.org/10.1016/j.ydbio.2004.01.023)
52. Smolenicka Z, Pani F, Hwang KJ, Gruschus JM, Ferretti JA (2003) Sequence of a conserved region of a new sea urchin homeobox gene from the NK family. *Cell Biol Int* 27:81–87
53. Dunn EF, Moy VN, Angerer LM, Angerer RC, Morris RL, Peterson KJ (2007) Molecular paleoecology: using gene regulatory analysis to address the origins of complex life cycles in the late Precambrian. *Evol Dev* 9:10–24. doi:[10.1111/j.1525-142X.2006.00134.x](https://doi.org/10.1111/j.1525-142X.2006.00134.x)
54. Ludueña RF, Banerjee A (2008) The isotypes of tubulin. Distribution and functional significance. In: Fojo T (ed) *The role of microtubules in cell biology neurobiology and oncology*. Humana Press, Totowa, pp 123–175. doi:[10.1007/978-1-59745-336-3](https://doi.org/10.1007/978-1-59745-336-3)
55. Casano C, Ragusa M, Cutrera M, Costa S, Gianguzza F (1996) Spatial expression of alpha and beta tubulin genes in the late embryogenesis of the sea urchin *P. lividus*. *Int J Dev Biol* 40:1033–1041
56. Otim O, Amore G, Minokawa T, McClay DR, Davidson EH (2004) SpHnf6, a transcription factor that executes multiple functions in sea urchin embryogenesis. *Dev Biol* 273:226–243. doi:[10.1016/j.ydbio.2004.05.033](https://doi.org/10.1016/j.ydbio.2004.05.033)
57. Stöbe P, Stein MAS, Habring-Müller A, Bezdán D, Fuchs AL, Hueber SD, Wu H, Lohmann I (2009) Multifactorial regulation of a hox target gene. *PLoS Genet* 5:e1000412. doi:[10.1371/journal.pgen.1000412](https://doi.org/10.1371/journal.pgen.1000412)