Genomes & Developmental Control

Cis-regulatory analysis of the sea urchin pigment cell gene *polyketide synthase*

Cristina Calestani *, David J. Rogers [1]

Department of Biology, University of Central Florida, Orlando, FL 32816, USA

ABSTRACT

The *Strongylocentrotus purpuratus* polyketide synthase gene (*SpPks*) encodes an enzyme required for the biosynthesis of the larval pigment echinochrome. *SpPks* is expressed exclusively in pigment cells and their precursors starting at blastula stage. The 7th–9th cleavage Delta–Notch signaling, required for pigment cell development, positively regulates *SpPks*. In previous studies, the transcription factors *glial cell missing* (*SpGcm*), *SpGatae* and *kruppel-like* (*SpKrl*/*z13*) have been shown to positively regulate *SpPks*. To uncover the structure of the Gene Regulatory Network (GRN) regulating the specification and differentiation processes of pigment cells, we experimentally analyzed the putative *SpPks cis*-regulatory region. We established that the −1.5 kb region is sufficient to recapitulate the correct spatial and temporal expression of *SpPks*. Predicted DNA-binding sites for SpGcm, SpGataE and SpKrl are located within this region. The mutagenesis of these DNA-binding sites indicated that SpGcm, SpGataE and SpKrl are direct positive regulators of *SpPks*. These results demonstrate that the sea urchin GRN for pigment cell development is quite shallow, which is typical of type I embryo development.

© 2010 Elsevier Inc. All rights reserved.

Introduction

The genetic program underlying embryo development is a tightly regulated process. Gene expression must be accurately controlled in space and time in order to develop a normal embryo. One of the most informative approaches to understand the genetic basis of development is the study, at the system level, of developmental genetic pathways and their regulatory interactions, i.e. Gene Regulatory Networks (GRNs; Levine and Davidson 2005; Davidson and Levine, 2008; Davidson, 2009). Currently the sea urchin GRN for endo-mesoderm specification is the most extensively characterized (Davidson et al., 2002a,b). However, the genetic circuits downstream of the sea urchin endo-mesoderm specification process, which lead to differentiated cell-types, are poorly understood. In order to understand from a system perspective the functionality of GRNs, together with their properties of plasticity and robustness, it is essential to integrate the knowledge on the genetic basis of cell specification with the terminal process of cell differentiation.

The pigment cells of the sea urchin larvae are of mesodermal origin, specifically they are one of four cell-types that develop from the Secondary Mesenchyme Cells (SMCs; Cameron et al., 1991; Ruffins and Ettensohn, 1996). A Delta (Dl)–Notch (N) signaling has been proven to be necessary for the differential specification of the presumptive SMC and endodermal territories (Sherwood and McClay, 1999; Sweet et al., 1999, 2002; McClay et al., 2000; Oliveri et al., 2002). Pigment cells are the first out of the four SMC types to be specified. They migrate into the blastocoel in the early gastrula stage and by the pluteus stage are embedded in the ectoderm (Gustafson and Wolpert, 1967; Gibson and Burke, 1985; Kominami et al., 2001). After the gastrula stage the pigment cells start producing a naphthoquinone compound called echinochrome, which in *S. purpuratus* produces the characteristic orange color (Griffiths, 1965). Pigment cells function has not yet been defined, but evidences suggest that they might have a role in the immune system of the sea urchin larvae. (Service and Wardlaw 1984; Gibson and Burke, 1987; Hibino et al., 2006; Castoe et al., 2007). The sea urchin polyketide synthase (*SpPks*) encodes an enzyme that is required for the biosynthesis of the echinochrome pigment (Calestani et al., 2003). The *SpPks* gene is exclusively expressed in pigment cells and their precursors (Calestani et al., 2003). The onset of expression occurs at early blastula, between 15 and 18 h in *S. purpuratus*. At blastula stage *SpPks* is expressed in a ring of about 20 SMC precursors surrounding the Primary Mesenchyme Cell (PMC) precursors. At gastrula stage *SpPks* expression is detected in cells just beneath or embedded in the

---

* Corresponding author. Department of Biology, University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816-2368, USA. Fax: +1 407 8235769.

E-mail addresses: calestac@mail.ucf.edu (C. Calestani), david.rogers@duke.edu (D.J. Rogers).

[1] Present address: Duke University, Developmental and Stem Cell Biology Program, Duke University Medical Center, DUMC 103855, Durham, NC 27710, USA.

ectoderm. Gene expression is maintained throughout the pluteus stage in cells embedded in the aboral ectoderm, coincident with the distribution of pigment cells (Gibson and Burke, 1985; Cameron et al., 1991; Ruffins and Ettensohn, 1996).

A BAC containing *SpPks* was previously isolated and annotated (Davidson et al., 2002b; Castoe et al., 2007). The *SpPks* coding sequence is 7275 base pairs long and it is composed of 7 predicted exons. Introns are relatively short (intron I is 4.8 kb, introns II, III, IV and VI approximately 1 kb, and intron V is 0.4 kb). The coding sequence of a *laminin* gene is found at about 2 kb downstream of *SpPks* and the coding sequence of an unknown gene at about 5 kb upstream. Therefore, the *cis*-regulatory region of *SpPks* is likely to be relatively compact.

Previous studies have identified a set of *SpPks* upstream transcriptional regulators specifically expressed in pigment cells: *glial cells missing* (*SpGcm*; Ransick et al., 2002; Davidson et al., 2002b; Calestani et al., 2003); *SpGatae* (Pancer et al., 1999; Davidson et al., 2002b; Lee and Davidson, 2004)and *kruppel-like* (*SpKrl/z13*, Howard et al., 2001; Minokawa et al., 2004; Materna et al., 2006; http://sugp.caltech.edu/endomes/qpcr.html, L.Chen, unpublished).

SpGcm is a direct target of the 7th–9th cleavage Dl–N signaling (Ransick and Davidson, 2006) and it positively regulates the expression of *SpPks* (Davidson et al., 2002b). *SpGcm* begins to be expressed between 10 and 12 h in pigment cell precursors, a few hours before the onset of expression of *SpPks* (Ransick et al., 2002; Calestani et al., 2003). Given the close timing of expression within the same cell type, it is hypothesized that *SpGcm* is a direct regulator of *SpPks*. *SpGatae* is expressed in the mesoderm between 15 h and 20 h, after the onset of *SpGcm* expression (Lee and Davidson, 2004). *SpGatae* does not in fact regulate *SpGcm* expression nor does *SpGcm* regulate *SpGatae* (Davidson et al., 2002b; http://sugp.caltech.edu/endomes/qpcr.html). This indicates that *SpGcm* and *SpGatae* belong to parallel pathways. Moreover, *SpKrl* is expressed in SMC precursors at the hatched blastula stage and by gastrulation it is expressed only in the endoderm (Howard et al., 2001; Minokawa et al., 2004). *SpKrl* appears to act independently and parallel to the Dl–N pathway (Yamazaki et al., 2008). In fact, *SpKrl* does not regulate *SpGatae* expression nor does it regulate *SpGcm* (http://sugp.caltech.edu/endomes/qpcr.html). In all probability the *SpKrl* regulatory input into *SpPks* is parallel to that of *SpGatae* and *SpGcm*.

In this work we studied the *cis*-regulatory architecture of the pigment cell differentiation gene *SpPks*, ultimately to start elucidating the structure of the upstream GRN, which is mostly uncharacterized. First, we studied the regulatory activity of the first 3kb upstream of the transcription start to identify the genomic sequence required to recapitulate the normal gene expression pattern. Second, we tested the hypothesis that *SpPks* is a direct target of *SpGcm*, *SpGatae* and *SpKrl*.

Material and methods

Computational analysis of putative cis-regulatory sequences

Gene annotation was performed previously on the isolated BAC clone 80H21 containing the *SpPks* gene (Davidson et al., 2002b; NCBI accession AC131453) using the Sea urchin Genome AnnotatoR software (SUGAR; Brown et al., 2002) and *SpPks* cDNA sequences previously obtained (Calestani et al., 2003; Castoe et al., 2007). Putative *cis*-regulatory sequences of other pigment cell differentiation genes were recovered from the sea urchin genome sequence available at NCBI (http://www.ncbi.nlm.nih.gov/; *SpFmo1 LOC586240, SpFmo2 LOC588371, SpFmo3 LOC589643* and *SpSult LOC592500*). The Family Relations (Brown et al., 2002) software was used to search for DNA-binding sites for the putative direct regulators, SpGcm, SpGataE and SpKrl. The DNA-binding site sequences for the sea urchin SpGcm, SpGataE and SpKrl are not known. For this reason DNA-binding

consensus sequences known from other organisms were used: the Gcm consensus sequence ATRCGGGY identified in *Drosophila* and mammals (Akiyama et al., 1996), the GataE consensus WGATAR identified in vertebrate (Evans et al., 1988), the Krl-like consensus CCNCNCCCN identified in *Drosophila* and vertebrates (Rosenberg et al., 1986; Klevit, 1991; Miller and Bieker, 1993).

Preparation of GFP reporter constructs

The genomic regions of interest (−3 kb, −2 kb, −1.5 kb and −1 kb) were PCR amplified from the BAC clone 80H21. PCR primers included restriction digestion sites for SacI (forward primer) and MluI (reverse primer) to facilitate directional cloning into the Green Fluorescent Protein reporter vector EpGFPII (Arnone et al., 1997). A list of all primers can be found in Supplementary Data (Table S1). The −3 kb sequence was amplified using the Expand High Fidelity PCR System method according to the manufacturer (Roche, Indianapolis, IN). The −2 kb, −1.5 kb and −1 kb sequences were amplified with Taq DNA Polymerase (Roche, Indianapolis, IN).

The mutagenesis of the predicted SpGcm DNA-binding site in the −2 kb sequence was generated by fusion PCR. PCR primers were designed to amplify two overlapping DNA fragments covering the –2 kb region and the binding site ACCCGCAT was changed to GTATTAGC (see Supplemental Table S1 for primer sequences). The −1.8 kb control sequence and the −1.8 kb with the mutated SpGataE or SpKrl sites were produced by *de novo* double-stranded oligonucleotide synthesis (Gen-Script USA Inc., Piscataway, NJ). The SpGataE predicted sites were changed to CTCGCA (forward sites) or ACGCTC (reverse sites). The genomic sequences containing multiple overlapping SpKrl sites were changed from AGGGTGTGGGGGGGGGGCGGGG to CTTTGTGTTTTTT-TTTTATTTT and from CCCCCCCCCCCCT to AAAAAAAAAAAAG.

The *cis*-regulatory DNA fragments, wild type and mutagenized, were cloned into the EpGFPII reporter vector between SacI and MluI. Each construct sequence was verified by restriction digestion and sequencing. Reporter constructs were linearized by SacI digestion and purified (QIAquick PCR purification kit, Qiagen, Valencia, CA) prior to embryo microinjection.

Embryo microinjection and observation of GFP expression

Embryo microinjection was performed as previously described (McMahon et al., 1985; Arnone et al., 2004). Injection solutions were prepared at a concentration of 1000 molecules/pl of linearized plasmid in 0.12 M KCl with the addition of 5× molar excess of restriction digested sea urchin genomic DNA (carrier DNA). Approximately 2 pl of injection solution was delivered to each embryo. Embryos injected with each GFP reporter construct were observed throughout development using fluorescence microscopy (Olympus BX60, Center Valley, PA).

Quantification of transcript accumulation by Real Time Quantitative PCR (QPCR)

Approximately 100 injected embryos were collected at different developmental stages (15 h, 21 h, 26 h, 40 h, 50 h and 72 h post-fertilization) and total RNA was isolated using the RNAeasy kit (Qiagen, Valencia, CA) following the manufacturer's procedure. The RNA was DNAse treated with DNA-Free (Applied Biosystems/Ambion, Austin, TX) for 15 min at 37 °C, followed by purification with the Qiagen RNAeasy kit. The cDNA was prepared from DNAse-treated RNA using the High Capacity cDNA Archive Kit (Applied Biosystems, Foster City, CA). QPCR was used to measure the relative amount of the endogenous *SpPks* mRNA, as well as the *gfp* mRNA in injected embryos (ABI PRISM 7000 Sequence Detection System, Applied Biosystems, Foster City, CA). QPCR reactions were set up in triplicates using ABI SYBR Green and 15 pmol of each forward and reverse primer for

SpPks, gfp, in addition to two endogenous control genes, *ubiquitin* and *SpZ12*. *SpZ12* was used as the internal standard to estimate the number of mRNA transcripts per embryo, as the number of *SpZ12* transcripts in various stages of embryo development are known from RNA titration experiments (Wang et al., 1995). The following number of mRNA molecules per embryo per stage were used: 1600 at 15 h, 1900 at 21 and 26 h, 1200 at 40 and 50 h, and 1600 at 72 h.

The mRNA levels were normalized to the *gfp* copy number per embryo as previously described (Revilla-i-Domingo et al., 2004). The *gfp* copy number per embryo was estimated using *SpPks* as internal standard (one copy per genome). The *gfp* and *SpPks* copy numbers were measured by QPCR (triplicate reactions) using an equivalent of three embryos per reaction.

Results

Map of predicted DNA-binding sites for SpPks transcriptional regulators

Multiple putative DNA-binding sites for the pigment cell transcription factors SpGcm, SpGataE and SpKrl, known to positively regulate *SpPks*, were computationally identified in the *SpPks* BAC clone. For the purpose of this study, we selected the subset of putative DNA-binding sites within the −3 kb of *SpPks*, since this region, in most genes, contains the functional regulatory elements that drive the correct pattern of gene expression. The positions and orientation of these predicted regulatory sequences are listed in Table 1. Within the −3 kb region, we identified one putative SpGcm binding site, nine SpGataE sites and three regions containing multiple overlapping SpKrl sites. The SpGcm site is located approximately 1.2 kb upstream of the transcription start. The SpGataE sites are distributed along the −3 kb sequence and the distance between contiguous sites ranges from 20 to 484 bp. The distance between the SpGcm site and the two most proximal SpGataE sites are 207 and 248 bp, respectively. Interestingly, the group of six SpKrl sites starting at −1138 is very close to the SpGcm site, just 41 bp downstream.

We obtained, by comparative genomics, further evidences supporting a regulatory function of the predicted SpGcm SpGataE and SpKrl DNA-binding sites in *SpPks*. A comparative analysis with *SpPks*

Table 1
Predicted transcription factor DNA-binding sites within the −3 kb *SpPks* sequence.

Transcription factor (consensus sequence)[a]	Predicted DNA-binding site	Orientation	Position
GCM (ATRCGGGY)	ACCCGCAT	Reverse	−1178
GATA E (WGATAR)	TGATAA	Reverse	−2605
	TGATAA	Reverse	−2585
	TGATAA	Reverse	−2542
	AGATAA	Reverse	−2329
	AGATAG	Reverse	−1845
	TGATAG	Forward	−1385
	TGATAA	Forward	−930
	AGATAA	Forward	−809
	TGATAA	Forward	−671
KRL-like (CCNCNCCCN)	CCTCTCCCT	Forward	−2107
	CCACACCCT	Reverse	−1138
	CCCCCCCCA	Reverse	−1131
	CCCCCCCCC	Reverse	−1130
	CCCCCCCCC	Reverse	−1129
	CCGCCCCCC	Reverse	−1126
	CCCCGCCCC	Reverse	−1124
	CCCCCCCCC	Forward	−287
	CCCCCCCCC	Forward	−286
	CCCCCCCCC	Forward	−285
	CCCCCCCCC	Forward	−284
	CCCCCCCCC	Forward	−283
	CCCCCCCCT	Forward	−282

[a] Consensus sequence of homologous gene DNA-binding site used to search the sea urchin genomic sequence.

orthologous sequences from other sea urchin species was not possible because of lack of sequence data in the genomic region of interest. Consequently, we extended the computational prediction of DNA-binding sites to other *S. purpuratus* pigment cell differentiation genes: *flavin-monooxygenase 1, 2* and *3* (*SpFmo*) and *sulfotransferase* (*SpSult*; Calestani et al., 2003). *SpPks*, *SpFmo* and *SpSult* are co-expressed (Calestani et al., 2003) and are likely to belong to the same differentiation gene battery based on gene expression data (Davidson et al., 2002b; http://sugp.caltech.edu/endomes/qpcr.html). The predicted DNA-binding sites identified within the putative *cis*-regulatory sequence of *SpFmo* and *SpSult* are listed in Table 2. SpGcm binding sites were predicted in *SpFmo3* and *SpSult* regulatory sequences. The absence of predicted SpGcm sites in *SpFmo1* could be due to incomplete sequence data in the 5′ upstream region (only 1.8 kb available). SpGataE and SpKrl sites were predicted in all the four genes.

Cis-regulatory region that recapitulates the SpPks expression

We first tested the functionality of the −3 kb region of *SpPks* by linking this putative *cis*-regulatory region to a *gfp* reporter gene (−*3KbSpPks-gfp*; Fig. 1). GFP fluorescence was clearly detected by blastula stage in cells at the vegetal plate and its expression was maintained throughout development with a spatial pattern coincident with the pigment cell distribution (Figs. 2A–C). Sixty-five percent of the injected embryos expressed GFP and no ectopic expression was observed (Table 3). Specifically, blastula stage GFP expression was seen in the ring of cells surrounding the PMCs. By late gastrula stage GFP was observed in cells that were delaminating from the archenteron and embedding in the aboral ectoderm. At pluteus stage, GFP was expressed in differentiated pigment cells embedded in the aboral ectoderm. In parallel, the −*3KbSpPks-gfp* activity was compared to the *D-E-Sp-P-gfp* construct activity, which recapitulates the expression of the transcription factor *SpGcm* (Ransick and Davidson, 2006). The *D-E-Sp-P* regulatory element drives *SpGcm* expression in SMCs from 12 h (onset of *SpGcm* transcription) to 24 h (Ransick and Davidson, 2006). In our study the GFP expression driven by the *D-E-Sp-P* element preceded the GFP expression of the −*3KbSpPks-gfp* construct by at least 3 h.

Subsequently, we performed serial deletions of the *SpPks* −3 kb region of approximately 1 to 0.5 kb to identify the 5′-end boundary of the regulatory region, while taking into account the positions of the predicted DNA-binding sites described above (Fig. 1). The following DNA constructs were produced and tested *in vivo*: −*2KbSpPks-gfp*, which lacks four SpGataE and one SpKrl binding sites; −*1.8KbSpPks-gfp*, omits one additional SpGataE site; −*1.5KbSpPks-gfp*; −*1kbSpPks-*

Table 2
Number of SpGcm, SpGataE and SpKrl predicted DNA-binding sites within the putative *cis*-regulatory region of *SpFmo* and *SpSult*.

Gene	Transcription factor	Entire *cis*-regulatory sequence[a]	5′-upstream sequence and first intron
SpFmo1	Gcm	0	0
	GataE	58	7
	Krl	102	6
SpFmo2	Gcm	0	0
	GataE	85	35
	Krl	38	0
SpFmo3	Gcm	3	1
	GataE	146	20
	Krl	132	10
SpSult	Gcm	1[b]	0
	GataE	76	43
	Krl	124	69

[a] The putative *cis*-regulatory sequence includes the introns and the intergenic regions 5′ upstream and 3′ downstream of the predicted mRNA sequence.
[b] Site found within the second intron.
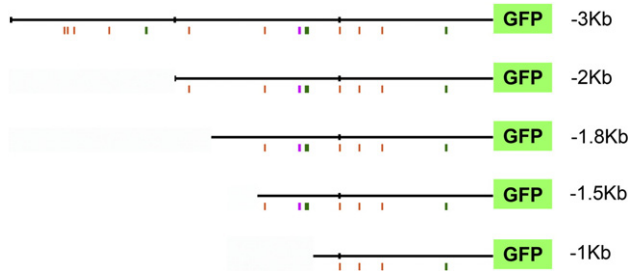
Fig. 1. Identification of the *cis*-regulatory sequence recapitulating *SpPks* expression. The schematic representation shows the DNA constructs used in this study. The first 3 kb upstream of the transcription start and the 5′-end deletions indicated were cloned into the EpGFPII vector and tested *in vivo*. The colored boxes below the genomic sequence (in black) indicate the positions of the predicted DNA-binding sites for SpGcm (in pink), SpGataE (in orange) and SpKrl (in green).

gfp, which lacks six out of nine SpGataE sites and the additional stretch of sequence containing six overlapping SpKrl sites and one SpGcm site.

The −*2KbSpPks-gfp*, −*1.8KbSpPks-gfp* and −*1.5KbSpPks-gfp* produced the same pigment cell specific pattern of GFP expression observed for the −*3KbSpPks-gfp*, with an average of GFP-positive embryos of 57%, 63% and 52% respectively (Table 3). The −*1kbSpPks-gfp* did not produce any GFP expression (Table 3).

It is known that the fluorescence of GFP can only be detected about 4h after the start of transcription. Furthermore, GFP is very stable in sea urchin embryos and fluorescence may continue to be observed for several hours after the actual transcription of *gfp* stops (Arnone et al., 1997). In order to obtain a more precise temporal pattern of gene expression, we measured the relative amount of *SpPks* and *gfp* transcripts during the course of development (Fig. 2D). We specifically tested the −*2KbSpPks-gfp* construct and overall it showed a similar trend of expression to the endogenous *SpPks* with one difference: the −*2KbSpPks-gfp* construct has a higher notable expression at 15 and 21 h

Table 3
Cis-regulatory activity of *SpPks-gfp* reporter constructs[a].

Construct (# replicates)[b]	% GFP-positive[c] (# scored embryos)[d]	% GFP-positive pigment cells only	% GFP-positive ectopically[c]
−3 kb (4)	65 (101)	100	0
−2 kb (3)	57 (77)	100	0
−1.8 kb (2)	63 (123)	100	
−1.5 kb (2)	52 (110)	100	0
−1 kb (2)	0 (76)	NA	NA
−2Kb-gcm-mut (3)	0 (145)	NA	NA
	0 (122)	NA	NA
−1.8Kb-gatae-mut (2)	52 (120)	100	0
−1.8Kb-krl-mut (2)	32 (114)	100	0

[a] Data for the −3Kb, −2Kb, −1.8Kb, −1.5Kb and −1Kb constructs were obtained from late gastrula embryos; data for the −2Kb-gcm-mut construct were obtained from late gastrula and pluteus stage embryos. For each construct, embryos were initially observed at different developmental time points (15 h, 26 h, 32 h, 45 h, and 72 h) to determine the GFP expression pattern.
[b] Each replicate experiment was carried out with embryos derived from a different set of parents.
[c] GFP positive were embryos with more than two fluorescent cells; GFP expression in only one or two cells was considered background due to the position of integration in the genome.
[d] The total number of scored embryos was obtained by combining all the replicate experiments.

(Fig. 2D). This is expected due to the multiple copies of reporter construct that are normally incorporated in the sea urchin genome (McMahon et al., 1985; Revilla-i-Domingo et al., 2004). Instead after 21 h (mesenchyme blastula) we observed a drop in *gfp* expression relative to *SpPks* of approximately four times compared to earlier stages.

SpPks transcription is directly regulated by SpGcm, SpGataE and SpKrl

The serial deletion of the *SpPks* −3 kb region led to the identification of 500 bp (between −1.5 and −1 kb) that should contain
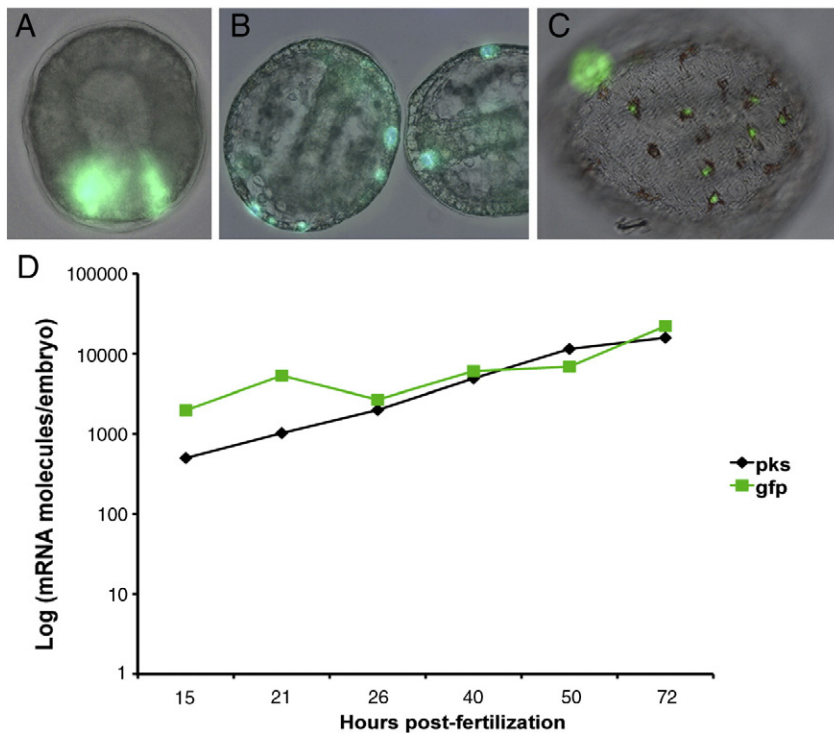
Fig. 2. Spatial and temporal pattern of expression of the reporter constructs. (A–C) Overlay of DIC and fluorescence images of embryos at blastula (A), late gastrula (B) and pluteus stage (C); shown is a representative example of the expression pattern observed for the −*3KbSpPks*, −*2KbSpPks* and −*1.5KbSpPks* constructs. GFP expression is exclusively restricted to pigment cells and their precursors, as indicated in Table 2. The −*1KbSpPks* construct did not show any GFP expression (data not shown). (D) Transcript levels of *SpPks* compared to the −*2KbSpPks-gfp* construct throughout development. Data are the estimated number of mRNAs per embryo in Log scale. Data shown are derived from one batch of embryos (one set of parents) and the same trend was replicated from a second independent batch of embryos (see text for results description).

transcription factor DNA-binding site/s required for the positive regulation of *SpPks*. This sequence includes at least one putative DNA-binding site for each of the three known upstream activators, SpGataE, SpGcm and SpKrl (Fig. 1; Table 1). We tested *in vivo* the functionality of the predicted SpGcm, SpGataE and SpKrl sites by mutagenesis. The mutagenized construct −2Kb-gcm-mut-gfp failed to produce any GFP fluorescence in the injected embryos, indicating that the predicted *SpGcm* DNA-binding site is functional and it is required for *SpPks* transcription (Table 3). The mutagenized constructs −1.8Kb-gatae-mut and −1.8Kb-krl-mut showed a drastic reduction of the *gfp* transcript accumulation measured at mesenchyme blastula stage (Fig. 3D). A recovery of pigment cell specific *gfp* expression was observed by late gastrula stage (Figs. 3A–C), with an average of GFP-positive embryos of 52% for the −1.8Kb-gatae-mut construct and of 32% for the −1.8Kb-krl-mut construct (Table 3). These data indicate that SpGataE and SpKrl are direct positive regulators of *SpPks* and they are required for transcription at least up to mesenchyme blastula stage.

Discussion

The *cis*-regulatory analysis of the pigment cell differentiation gene *SpPks* led us to identify a genomic region able to recapitulate the correct pattern of gene expression and to determine that *SpPks* is a direct target of the transcription factors *SpGcm*, *SpGataE* and *SpKrl*. This study brings new insights into the GRN architecture downstream of the Dl–N signaling required for SMC development and adds to the general knowledge of GRN structure and function during the differentiation process.

The qualitative and quantitative analysis of the −2 kb sequence regulatory activity (Fig. 2) shows that overall this genomic region recapitulates the trend of expression of the endogenous *SpPks*. However, this region might be missing DNA-binding sites for additional positive regulators indicated by the drop in reporter expression after 21 h p.f. (mesenchyme blastula). It needs to be noted that this developmental stage marks the beginning of PMC ingression into the blastocoel, which interrupts the Dl–N signaling from PMC to SMC. As a result of this morphogenetic event the set of transcription factors that regulates *SpPks* after mesenchyme blastula might change. We know of one regulatory component that certainly contributes to the maintenance of *SpPks* expression independently from the Dl–N signaling, which is the *SpGcm* positive regulatory feedback (Davidson et al., 2002b). Nevertheless, according to our data this auto-regulatory loop is not sufficient to produce the correct transcript level throughout development and additional regulatory inputs, acting outside the *SpPks* −2 kb region, must exist.

Further deletion of the −2 kb sequence showed that the genomic region within the first 1.5 kb upstream of *SpPks* transcription start includes the *cis*-regulatory system necessary for the correct temporal and spatial expression of *SpPks*. Our data indicated that the 5′-end boundary of *SpPks* regulatory sequence is located between −1.5 kb and −1 kb. Further studies are needed to refine the boundaries of the minimal regulatory sequence, although, our data suggest a relatively compact structure of the *SpPks* regulatory system. This property has been observed for other differentiation genes in sea urchins and in other organisms (Davidson, 2006). For example, the skeletogenic gene *cyclophilin*, has a minimal promoter of 218 bp, which recapitulates the relatively simple expression pattern restricted to one cell-type throughout time, similarly to *SpPks* (Amore and Davidson, 2006). The relatively small number of studies of transcriptional regulation of differentiation genes limits our ability to formulate general definitions of structure–function relationships of differentiation regulatory
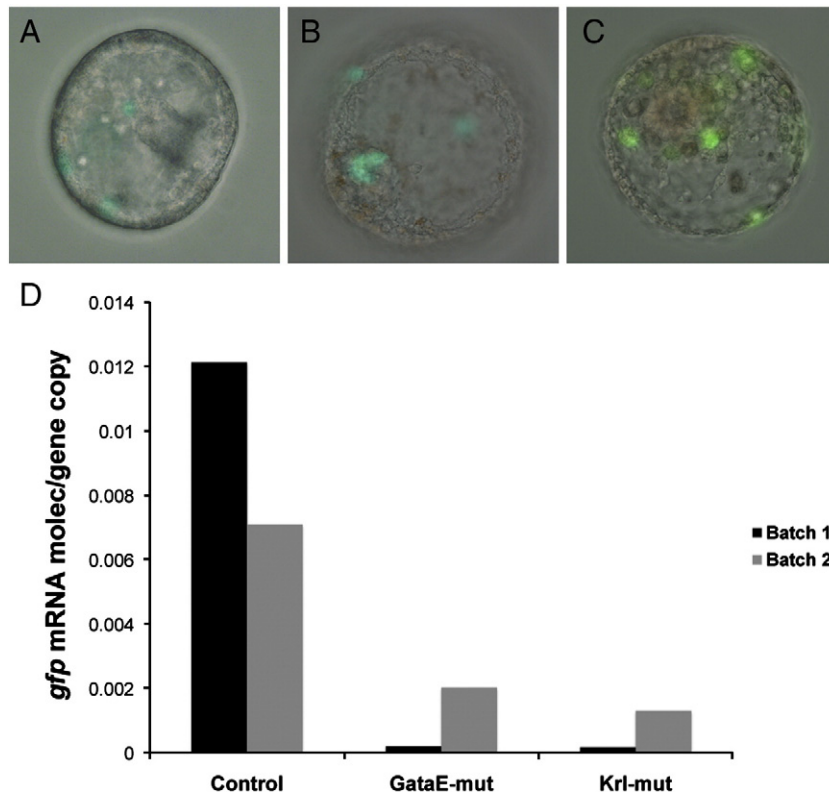
Fig. 3. Mutagenesis of putative SpGataE and SpKrl DNA-binding sites. (A–C) Overlay of DIC and fluorescence images of embryos injected with the −1.8Kb-gfp control construct (A), with the −1.8Kb-gatae-mut-gfp construct (B) and with the −1.8Kb-krl-mut-gfp construct (C); shown is a representative example of the GFP expression pattern observed at late gastrula stage. GFP expression was observed in both control and mutagenized constructs and it was exclusively restricted to pigment cells and their precursors, as indicated in Table 3. (D) *gfp* transcript levels of the −1.8Kb-gfp control, the −1.8Kb-gatae-mut-gfp and the −1.8Kb-krl-mut-gfp constructs in mesenchyme blastula stage embryos. The quantification of transcripts was done by QPCR. Data are the estimated number of mRNA molecules normalized to the *gfp* copy number per embryo. Data shown are derived from two batches of embryos (two sets of parents). The *gfp* reporter expression was drastically reduced in embryos injected with the mutagenized constructs as compared to the control.

systems. However, current knowledge suggests that a less complex and dynamic expression pattern does not necessarily correlate with a short minimal promoter. In fact the 450 bp module regulating the sea urchin actin gene *CyIIa* is sufficient to recapitulate a very dynamic expression pattern (Arnone et al., 1998).

Interestingly, none of the deletion constructs produced any ectopic expression of GFP. This suggests that *SpPks* regulation does not involve repression to restrict its expression to pigment cells. This is in accordance with previous GRN studies of cell differentiation. Generally, the earlier process of specification sets the boundaries of the differentiation gene batteries' domain of expression (Levine and Davidson, 2005; Oliveri et al., 2008). Yet, a higher resolution mapping of the regulatory module/s will be needed to ultimately exclude the role of repressors in defining the spatial domain of *SpPks* expression.

Our data indicate that SpGcm, SpGataE and SpKrl are direct positive regulators of *SpPks* (Table 3; Fig. 3). SpGcm is required for *SpPks* transcription throughout development, while SpGataE and SpKrl are most likely required only up to mesenchyme blastula–early gastrula stages. Our conclusion is supported by previous studies. *SpGatae* is expressed in SMC precursors only up the hatched blastula stage (20 h p.f; Lee and Davidson, 2004). *SpKrl* starts to be expressed in SMC precursors at the hatched blastula stage and by gastrulation is expressed only in the endoderm (Howard et al., 2001; Minokawa et al., 2004). Even though the protein turnovers of SpGataE and SpKrl are not known, most likely other transcription factors are responsible for the maintenance of *SpPks* expression after the mesenchyme blastula stage. Another supporting evidence is that SpGataE and SpKrl knock-down embryos produce the echinochrome pigment by the pluteus stage (Lee and Davidson, unpublished; Howard et al., 2001).

The sequence characteristics of SpGataE and SpKrl DNA-binding sites make them more likely to be computationally identified by chance (not functional DNA-binding sites). SpGataE site WGATAR (Evans et al., 1988) is short and relatively degenerate, while SpKrl CCNCNCCCN (Rosenberg et al., 1986; Klevit, 1991; Miller and Bieker, 1993) has a very low sequence complexity and can appear several times in the genome, particularly in GC-rich regions. Most likely, not all of the SpGataE and SpKrl sites predicted within the *SpPks* −1.8 kb are functional sites. The DNA-binding site position relative to the SpGcm site could suggest which are the most likely functional sites. The two SpGataE sites closer to the SpGcm site are 200–250 bp away, which is within the average length of a *cis*-regulatory module (Davidson, 2006). One set of the predicted SpKrl DNA-binding sites are closely clustered with the SpGcm site, just 41 bp apart. Another type of evidence is that this predicted SpKrl DNA-binding sequence CCACACCCT is identical to functional binding sites for Krl-like factors identified in several erythroid-specific genes in vertebrates (Miller and Bieker, 1993; Mantovani et al., 1988; Frampton et al., 1990).

We identified several putative DNA-binding sites for SpGcm, SpGataE and SpKrl within the *cis*-regulatory sequence of other pigment cell differentiation genes: the *SpFmo* gene family encoding flavin-monooxygenases (FMOs) and *SpSult*, which encodes a sulfo-transferase (Ransick et al., 2002; Calestani et al., 2003; Table 2). In fact, previous studies showed that *SpPks* and *SpSult* are positively regulated by SpGcm, SpGataE and SpKrl; *SpFmo1* is also positively regulated by SpGcm and SpGataE (regulation by SpKrl has not been tested; *SpFmo2* and *3* have not been tested; Davidson et al., 2002b; http://sugp.caltech.edu/endomes/qpcr.html). This suggests that *SpPks*, *SpFmo* and *SpSult* might belong to the same differentiation gene battery.

Our data demonstrates that the GRN for pigment cell development is shallow, as it involves just two steps of gene activation from the 7th–9th cleavage Dl–N signaling to the activation of a differentiation gene. The time interval between the activation of *SpGcm* and the target gene *SpPks* is in the range of 3 to 5 h (Ransick et al., 2002; Calestani et al., 2003). Similar results have been observed in PMCs, as described by Amore and Davidson (2006) and Oliveri et al. (2008). Our study further confirms the shallow GRN structure of Type I

embryo development, as also described in other organisms such as *C. elegans* and ascidians (Davidson, 2006).

In conclusion, the transcriptional regulation of the pigment cell differentiation gene *SpPks* involves at least three parallel positive inputs. Two are acting downstream of the 7th–9th cleavage Dl–N signaling through SpGcm and SpGataE and one is Dl–N independent acting through SpKrl. Parallel positive regulatory inputs into differentiation gene batteries have also been observed in another sea urchin cell-type, PMC (Amore and Davidson, 2006; Oliveri et al., 2008), suggesting that this is a common feature of the terminal differentiation phase of developmental GRN in sea urchin.

References

Akiyama, Y., Hosoya, T., Poole, A.M., Hotta, Y., 1996. The gcm-motif: a novel DNA-binding motif conserved in *Drosophila* and mammals. Proc. Natl. Acad. Sci. U. S. A. 93, 14912–14916.

Amore, G., Davidson, E.H., 2006. *cis*-Regulatory control of cyclophilin, a member of the ETS-DRI skeletogenic gene battery in the sea urchin embryo. Dev. Biol. 293, 555–564.

Arnone, M.I., Bogarad, L.D., Collazo, A., Kirchhamer, C.V., Cameron, R.A., Rast, J.P., Gregorians, A., 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.

Arnone, M.I., Martin, E.L., Davidson, E.H., 1998. *Cis*-regulation downstream of cell type specification: a single compact element controls the complex expression of the CyIIa gene in sea urchin embryos. Development 125, 1381–1395.

Arnone, M.I., Dmochowski, I.J., Gache, C., 2004. Using reporter genes to study *cis*-regulatory elements. Methods Cell Biol. 74, 621–652.

Brown, C.T., Rust, A.G., Clarke, P.J., Pan, Z., Schilstra, M.J., De Buysscher, T., Griffin, G., Wold, B.J., Cameron, R.A., Davidson, E.H., Bolouri, H., 2002. New computational approaches for analysis of *cis*-regulatory networks. Dev. Biol. 246, 86–102.

Calestani, C., Rast, J.P., Davidson, E.H., 2003. Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. Development 130, 4587–4596.

Cameron, R.A., Fraser, S.E., Britten, R.J., Davidson, E.H., 1991. Macromere cell fates during sea urchin development. Development 113, 1085–1091.

Castoe, T.A., Stephens, T., Noonan, B.P., Calestani, C., 2007. A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. Gene 392, 47–58.

Davidson, E.H., 2006. The Regulatory Genome: Gene Regulatory Networks in Development and Evolution. Academic Press, San Diego.

Davidson, E.H., 2009. Developmental biology at the systems level. Biochim. Biophys. Acta 1789, 248–249.

Davidson, E.H., Levine, M.S., 2008. Properties of developmental gene regulatory networks. Proc. Natl. Acad. Sci. U. S. A. 105, 20063–20066.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J., Rust, A.G., Pan, Z., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002a. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev. Biol. 246, 162–190.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002b. A genomic regulatory network for development. Science 295, 1669–1678.

Evans, T., Reitman, M., Felsenfeld, G., 1988. An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc. Natl. Acad. Sci. U. S. A. 85, 5976–5980.

Frampton, J., Walker, M., Plumb, M., Harrison, P.R., 1990. Synergy between the NF-E1 erythroid-specific transcription factor and the CACCC factor in the erythroid-specific promoter of the human porphobilinogen deaminase gene. Mol. Cell. Biol. 10, 3838–3842.

Gibson, A.W., Burke, R.D., 1985. The origin of pigment cells in embryos of the sea urchin *Strongylocentrotus purpuratus*. Dev. Biol. 107, 414–419.

Gibson, A.W., Burke, R.D., 1987. Migratory and invasive behavior of pigment cells in normal and animalized sea urchin embryos. Exp. Cell. Res. 173, 546–557.

Griffiths, M., 1965. A study of the synthesis of naphthaquinone pigments by the larvae of two species of sea urchins and their reciprocal hybrids. Dev. Biol. 11, 433–447.

Gustafson, T., Wolpert, L., 1967. Cellular movement and contact in sea urchin morphogenesis. Biol. Rev. Camb. Philos. Soc. 42, 442–498.

Hibino, T., Loza-Coll, M., Messier, C., Majeske, A.J., Cohen, A.H., Terwilliger, D.P., Buckley, K.M., Brockton, V., Nair, S.V., Berney, K., Fugmann, S.D., Anderson, M.K., Pancer, Z., Cameron, R.A., Smith, L.C., Rast, J.P., 2006. The immune gene repertoire encoded in the purple sea urchin genome. Dev. Biol. 300, 349–365.

Howard, E.W., Newman, L.A., Oleksyn, D.W., Angerer, R.C., Angerer, L.M., 2001. SpKrl: a direct target of beta-catenin regulation required for endoderm differentiation in sea urchin embryos. Development 128, 365–375.

Klevit, R.E., 1991. Recognition of DNA by Cys2, His2 zinc fingers. Science 253 (1367), 1393.

Kominami, T., Takata, H., Takaichi, M., 2001. Behavior of pigment cells in gastrula-stage embryos of *Hemicentrotus pulcherrimus* and *Scaphechinus mirabilis*. Dev. Growth Differ. 43, 699–707.

Lee, P.Y., Davidson, E.H., 2004. Expression of Spgatae, the *Strongylocentrotus purpuratus* ortholog of vertebrate GATA4/5/6 factors. Gene Expr. Patterns 5, 161–165.

Levine, M., Davidson, E.H., 2005. Gene regulatory networks for development. Proc. Natl. Acad. Sci. U. S. A. 102, 4936–4942.

Mantovani, R., Malgaretti, N., Nicolis, S., Giglioni, B., Comi, P., Cappellini, N., Bertero, M. T., Caligaris-Cappio, F., Ottolenghi, S., 1988. An erythroid specific nuclear factor binding to the proximal CACCC box of the beta-globin gene promoter. Nucleic Acids Res. 16, 4299–4313.

Materna, S.C., Howard-Ashby, M., Gray, R.F., Davidson, E.H., 2006. The C2H2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development. Dev. Biol. 300, 108–120.

McClay, D.R., Peterson, R.E., Range, R.C., Winter-Vann, A.M., Ferkowicz, M.J., 2000. A micromere induction signal is activated by beta-catenin and acts through notch to initiate specification of secondary mesenchyme cells in the sea urchin embryo. Development 127, 5113–5122.

McMahon, A.P., Flytzanis, C.N., Hough-Evans, B.R., Katula, K.S., Britten, R.J., Davidson, E.H., 1985. Introduction of cloned DNA into sea urchin egg cytoplasm: replication and persistence during embryogenesis. Dev. Biol. 108, 420–430.

Miller, I.J., Bieker, J.J., 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. Mol. Cell. Biol. 13, 2776–2786.

Minokawa, T., Rast, J.P., Arenas-Mena, C., Franco, C.B., Davidson, E.H., 2004. Expression patterns of four different regulatory genes that function during sea urchin development. Gene Expr. Patterns 4, 449–456.

Oliveri, P., Carrick, D.M., Davidson, E.H., 2002. A regulatory gene network that directs micromere specification in the sea urchin embryo. Dev. Biol. 246, 209–228.

Oliveri, P., Tu, Q., Davidson, E.H., 2008. Global regulatory logic for specification of an embryonic cell lineage. Proc. Natl. Acad. Sci. U. S. A. 105, 5955–5962.

Pancer, Z., Rast, J.P., Davidson, E.H., 1999. Origins of immunity: transcription factors and homologues of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes. Immunogenetics 49, 773–786.

Ransick, A., Davidson, E.H., 2006. *cis*-regulatory processing of Notch signaling input to the sea urchin glial cells missing gene during mesoderm specification. Dev. Biol. 297, 587–602.

Ransick, A., Rast, J.P., Minokawa, T., Calestani, C., Davidson, E.H., 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. Dev. Biol. 246, 132–147.

Revilla-i-Domingo, R., Minokawa, T., Davidson, E.H., 2004. R11: a *cis*-regulatory node of the sea urchin embryo gene network that controls early expression of SpDelta in micromeres. Dev. Biol. 274, 438–451.

Rosenberg, U.B., Schroder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I., Jackle, H., 1986. Structural homology of the product of the *drosophila* Kruppel gene with *Xenopus* transcription factor IIIA. Nature 319, 336–339.

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

Service, M., Wardlaw, A.C., 1984. Echinochrome-A as a bactericidal substance in the coelomic fluid of *Echinus esculentus* (L.). Comp. Biochem. 79B, 161–165.

Sherwood, D.R., McClay, D.R., 1999. LvNotch signaling mediates secondary mesen-chyme specification in the sea urchin embryo. Development 126, 1703–1713.

Sweet, H.C., Hodor, P.G., Ettensohn, C.A., 1999. The role of micromere signaling in Notch activation and mesoderm specification during sea urchin embryogenesis. Devel-opment 126, 5255–5265.

Sweet, H.C., Gehring, M., Ettensohn, C.A., 2002. LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. Development 129, 1945–1955.

Wang, D.G., Britten, R.J., Davidson, E.H., 1995. Maternal and embryonic provenance of a sea urchin embryo transcription factor, SpZ12-1. Mol. Mar. Biol. Biotechnol. 4, 148–153.

Yamazaki, A., Kawabata, R., Shiomi, K., Tsuchimoto, J., Kiyomoto, M., Amemiya, S., Yamaguchi, M., 2008. Kruppel-like is required for nonskeletogenic mesoderm specification in the sea urchin embryo. Dev. Biol. 314, 433–442.