© 2019. Published by The Company of Biologists Ltd | Development (2019) 146, dev180653. doi:10.1242/dev.180653

RESEARCH ARTICLE

# Genome-wide identification of binding sites and gene targets of Alx1, a pivotal regulator of echinoderm skeletogenesis

**Jian Ming Khor, Jennifer Guerrero-Santoro and Charles A. Ettensohn***

## ABSTRACT

Alx1 is a conserved regulator of skeletogenesis in echinoderms and evolutionary changes in Alx1 sequence and expression have played a pivotal role in modifying programs of skeletogenesis within the phylum. Alx1 regulates a large suite of effector genes that control the morphogenetic behaviors and biomineral-forming activities of skeletogenic cells. To better understand the gene regulatory control of skeletogenesis by Alx1, we used genome-wide ChIP-seq to identify Alx1-binding sites and direct gene targets. Our analysis revealed that many terminal differentiation genes receive direct transcriptional inputs from Alx1. In addition, we found that intermediate transcription factors previously shown to be downstream of Alx1 all receive direct inputs from Alx1. Thus, Alx1 appears to regulate effector genes by indirect, as well as direct, mechanisms. We tested 23 high-confidence ChIP-seq peaks using GFP reporters and identified 18 active cis-regulatory modules (CRMs); this represents a high success rate for CRM discovery. Detailed analysis of a representative CRM confirmed that a conserved, palindromic Alx1-binding site was essential for expression. Our work significantly advances our understanding of the gene regulatory circuitry that controls skeletogenesis in sea urchins and provides a framework for evolutionary studies.

**KEY WORDS: Alx1 transcription factor, ChIP-seq, Gene regulatory network, Sea urchin skeletogenesis, Biomineralization, Primary mesenchyme cells (PMCs)**

## INTRODUCTION

A central challenge of biology is to explain how morphology is encoded in the genome. The specification of distinct cell types and the subsequent organization of these cells into discrete anatomical structures are controlled by differential gene expression during embryonic development. As part of this process, networks of interacting regulatory (i.e. transcription factor-encoding) genes and signaling pathways, organized as modular gene regulatory networks (GRNs), control programs of gene expression in embryonic cells. Furthermore, there is considerable evidence that evolutionary changes in developmental GRNs that control anatomy have played a crucial role in morphological evolution (McGregor et al., 2007; Rebeiz and Tsiantis, 2017). Currently, a major goal is to determine how different combinations of regulatory genes, functioning within a hierarchical network, regulate downstream effector genes to drive the morphogenesis of specific anatomical features during development.

Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA.

*Author for correspondence (ettensohn@cmu.edu)

J.M.K., 0000-0002-1428-6770; C.A.E., 0000-0002-3625-0955

**Received 18 May 2019; Accepted 9 July 2019**

The GRN that governs the development of the endoskeleton of the sea urchin embryo is one of the best-characterized developmental GRNs (reviewed by Shashikant et al., 2018a). It is a valuable experimental model for dissecting the fine-structure of a GRN that controls the formation of a prominent anatomical structure and for elucidating the changes in a GRN that have contributed to morphological evolution. The embryonic skeleton of euechinoids is formed by a specialized group of skeletogenic cells known as primary mesenchyme cells (PMCs). These cells originate from large micromeres, four cells that arise near the vegetal pole during cleavage. The PMC GRN is activated in the large micromere territory through the combined action of maternal factors and unequal cell division (Oliveri et al., 2008; Shashikant et al., 2018a). These maternal inputs activate early regulatory genes such as *Alx1*, *Ets1* and *Tbr* (Ettensohn et al., 2003; Fuchikami et al., 2002; Kurokawa et al., 1999; Oliveri et al., 2002), which engage downstream intermediate regulatory genes and subsequently several hundred terminal differentiation genes that govern PMC behavior and skeletal morphogenesis (Barsi et al., 2014; Oliveri et al., 2008; Rafiq et al., 2012, 2014).

The Paired-class homeodomain protein Alx1 plays a crucially important and evolutionarily conserved role in biomineralization throughout the phylum. Regardless of their larval forms, all adult echinoderms possess an extensive, calcitic endoskeleton. Owing to similarities between the GRNs of skeletogenic cells in the embryo and adult, it is widely thought that the embryonic skeleton arose via co-option of the adult skeletogenic program (Czarkwiani et al., 2013; Gao and Davidson, 2008; Gao et al., 2015; Killian et al., 2010; Richardson et al., 1989). During euechinoid embryogenesis, expression of *Alx1* can be detected as early as the 56-cell stage and is entirely restricted to the large micromere-PMC lineage (Ettensohn et al., 2003). Perturbation of *Alx1* function using antisense morpholinos dramatically inhibits PMC specification and skeletal morphogenesis (Ettensohn et al., 2003). Conversely, over-expression of *Alx1* results in ectopic activation of the skeletogenic program in other mesodermal cell lineages (Ettensohn et al., 2007). Alx1 has positive inputs into almost half of the ∼400 genes known to be differentially expressed by the PMCs, highlighting its important role in establishing the unique identity of these cells (Rafiq et al., 2014). In other echinoderm clades that form embryonic skeletons, *Alx1* is expressed specifically by the skeletogenic lineage, where it is essential for biomineralization (Dylus et al., 2016; Erkenbrack and Davidson, 2015; Koga et al., 2016; McCauley et al., 2012; Rubinstein and de Souza, 2013). Alx1 is also selectively expressed in the adult skeletogenic centers of sea urchins, brittle stars and sea stars (Czarkwiani et al., 2013; Gao and Davidson, 2008; Gao et al., 2015). Ectopic expression of sea urchin or sea star Alx1 in sea star embryos, which lack an embryonic skeleton, is sufficient to activate several skeletogenic genes that have been reported to be downstream of Alx1 in sea urchins (Koga et al., 2016). Recently, it was shown that the evolution of echinoderm biomineralization was associated with a gene duplication event that

Development (2019) 146, dev180653. doi:10.1242/dev.180653

allowed Alx1 to evolve functions distinct from its paralog (Alx4/Calx) through exonization of a novel 46-amino-acid motif (Khor and Ettensohn, 2017). The neofunctionalization of Alx1 may have been associated with the evolution of regulatory linkages to new target genes with functions related to biomineralization.

In the present study, we performed genomic chromatin immunoprecipitation (ChIP-seq) to identify regions bound by Alx1 in euechinoid (*Strongylocentrotus purpuratus*) embryos. Our ChIP-seq data reveal several thousand Alx1-binding sites throughout the genome. Our data show enrichment in Alx1-bound regions for binding sites associated with several potential co-regulators, including Ets1, Irx, Fos and Jun. Coupled with previously published ATAC-seq, DNase-seq and RNA-seq data (Rafiq et al., 2014; Shashikant et al., 2018b), we determined that a large fraction of sea urchin skeletogenic terminal differentiation genes receive inputs from Alx1. We also found that many of the intermediate transcription factors differentially expressed by PMCs (e.g. Alx4, Dri, Fos, FoxB, Nfkbil1L and Nk7) receive direct inputs from Alx1. These findings demonstrate that Alx1 regulates many effector genes through direct, positive inputs, but also suggest that this direct regulation might operate in concert with inputs from intermediary transcription factors in a feed-forward fashion. Using GFP reporter assays, we examined 23 high-confidence ChIP-seq peaks and identified 18 active cis-regulatory modules (CRMs), 15 of which selectively drive GFP expression in PMCs. Detailed analysis of one representative CRM located in an intron of *Sp-EMI/TM*, a gene that encodes a novel, PMC-specific protein, revealed a conserved, palindromic Alx1-binding site that we found to be essential for expression. Electrophoretic mobility shift assay (EMSA) studies confirmed that recombinant Alx1 protein bound to this site. Taken as a whole, our study has identified hundreds of direct targets of Alx1 and revealed important features of the genetic network downstream of this pivotal regulator of echinoderm skeletogenesis. More generally, this work extends our understanding of an important, model developmental GRN and enhances its utility for evolutionary studies.

## RESULTS

### Sp-Alx1 antibody validation

Although Alx1 has an essential, evolutionarily conserved role in echinoderm skeletogenesis, there has been no genome-wide assessment of Alx1-binding sites or identification of direct targets of Alx1. Hence, we performed ChIP-seq to identify Sp-Alx1-binding sites using a custom, affinity-purified rabbit polyclonal antibody raised against a peptide contained within the D2 domain of Sp-Alx1 (Khor and Ettensohn, 2017). We first validated the Sp-Alx1 antibody by western blotting using bacterially expressed, recombinant Alx1 (rAlx1) (Fig. S1A,B). The antibody specifically recognized rAlx1 in induced bacterial cultures but not proteins in uninduced cultures and not rAlx4, a closely related homeodomain protein that lacks the D2 domain. The antibody also effectively immunoprecipitated rAlx1 from bacterial lysates (Fig. S1C). We further validated the Sp-Alx1 antibody using whole-mount immunofluorescence. As expected, the nuclei of PMCs were selectively labeled (Fig. S1D-F). Monoclonal antibody (mAb) 6a9, which recognizes PMC-specific cell surface proteins of the MSP130 family (Ettensohn and McClay, 1988; Illies et al., 2002), was used as a marker for this cell type. In contrast, Sp-Alx1 morphant embryos, which lack PMCs, exhibited no Sp-Alx1 or mAb 6a9 staining (Fig. S1G,H).

### Sp-Alx1 chromatin immunoprecipitation

ChIP-seq was performed using the validated Sp-Alx1 antibody and, for mock ChIP, normal rabbit IgG antibodies (Fig. S2A). We isolated crosslinked chromatin from mesenchyme blastula-stage embryos [~24 hours post-fertilization (hpf)] from three independent fertilizations (see supplementary Materials and Methods). We chose this particular developmental stage because terminal differentiation genes downstream of Alx1 in the PMC GRN are expressed maximally at this stage, as determined by RNA-seq (Rafiq et al., 2014). In addition, by using mesenchyme blastula-stage embryos it was possible to directly correlate our data with existing ATAC-seq and DNase-seq datasets derived from purified PMCs (Shashikant et al., 2018b). Following immunoprecipitation, three Sp-Alx1 ChIP and three mock ChIP preparations were pooled separately and each was sequenced to a depth of approximately 20 million reads. After read quality control, alignment, and redundancy filtering, peaks that were enriched in the Sp-Alx1 ChIP sample compared with the non-specific, mock ChIP control were identified from the remaining 8 million uniquely mapping reads from each sample (Fig. S2B, see supplementary Materials and Methods). Using MACS2 with an mfold of [5, 50] and a *P*-value cutoff of 0.005, we identified 2906 peaks with an average length of 250 bp that were used for further analysis (Fig. S2C, Table S1).

### ChIP-seq peak analysis

To remove potential false positives, the 2906 ChIP-seq peaks were first filtered to identify those that intersected by at least 1 nt with regions of accessible chromatin, which we defined as the union of the previously reported ATAC-seq and DNase-seq reference peak sets (RPSs) identified at the same developmental stage (Shashikant et al., 2018b). The great majority of the Sp-Alx1 ChIP-seq peaks (2353/2906 peaks, or 81%) overlapped with regions of accessible chromatin, and this peak set was used for all further analyses. To corroborate the significance of this peak set, we first investigated whether such peaks were more likely to intersect with regions of chromatin that are differentially accessible in PMCs relative to non-PMC cells ('ATAC-seq or DNA-seq differential peaks'), regions that were previously identified as likely PMC enhancers (Shashikant et al., 2018b) (Fig. 1A). We found that Sp-Alx1 ChIP-seq peaks were much more likely to overlap with ATAC-seq differential peaks than with ATAC-seq peaks as a whole (14.5-fold enrichment, hypergeometric *P*-value=$1.14e^{-146}$). Similarly, Sp-Alx1 ChIP-seq peaks were far more likely to overlap with DNase-seq differential peaks than with DNase-seq peaks as a whole (14.8-fold enrichment, hypergeometric *P*-value=$4.07e^{-313}$). Moreover, when considering the overlap between ATAC-seq and DNase-seq differential peaks, which consists of 168 peaks, these were much more likely to overlap with Sp-Alx1 ChIP-seq peaks than were other open regions of chromatin (32.0-fold enrichment, hypergeometric *P*-value=$1.56e^{-84}$). Taken together, these findings confirmed that Sp-Alx1-binding sites were enriched in regions of accessible chromatin and showed that they were particularly enriched in regions selectively open in PMCs, supporting the view that these regions represent Alx1-bound regulatory elements active in PMCs.

### Analysis of nearby genes

To associate Sp-Alx1 ChIP-seq peaks with putative gene targets, a peak-to-gene distance cutoff was required. We determined that most of the 2353 peaks (85%) were located within 20 kb of annotated genes (Fig. 2A). Furthermore, analysis of the location of the peaks relative to the closest genes revealed that 65% of the peaks were either in promoter regions (≤2 kb upstream) or within gene bodies (Fig. 2B). Moreover, 43% of peaks (1011 peaks) were found to be enriched near the 5′ ends of annotated genes, namely in the

Development (2019) 146, dev180653. doi:10.1242/dev.180653

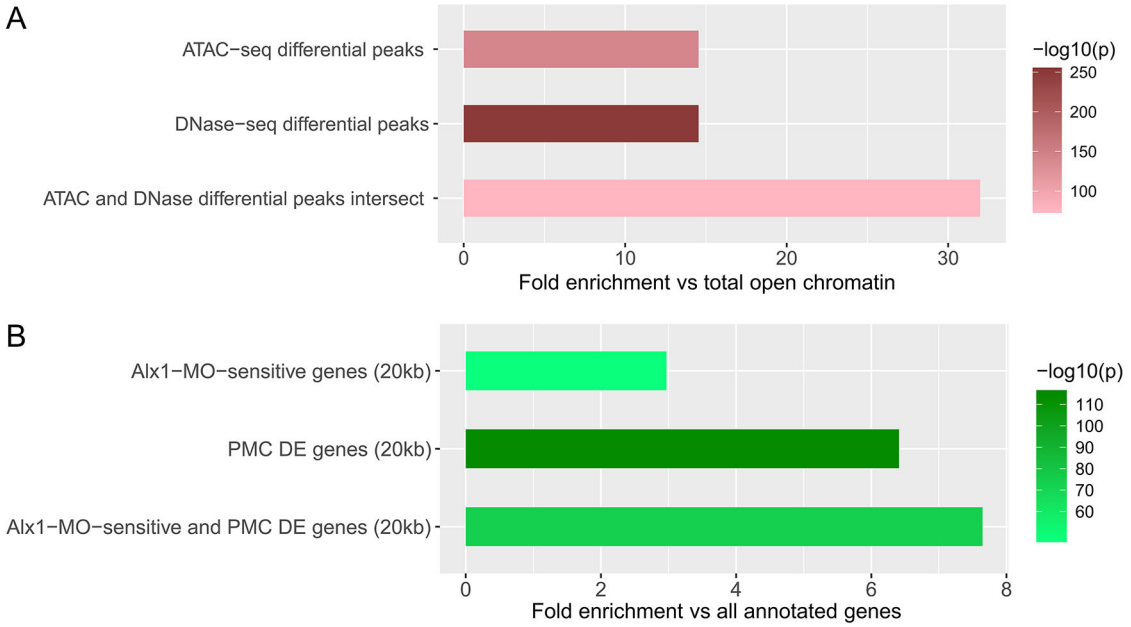



**Fig. 1. Sp-Alx1 ChIP-seq peak enrichment analysis.** (A) Proportion of peaks that intersect with differential peaks (by at least 1 nt) were compared with the proportion of peaks that intersect with reference peaks. Intensity of the bars corresponds to the significance of the enrichment, expressed as $-\log_{10}$(*P*-value). (B) Proportion of Alx1-MO-sensitive and/or PMC DE genes that have at least one Sp-Alx1 ChIP-seq peak within 20 kb were compared with the proportion of all annotated genes that have at least one Sp-Alx1 ChIP-seq peak within 20 kb.

promoter regions and first introns. Closer inspection of the peaks located near transcription start sites (TSSs) showed that they were concentrated in the 5′ untranslated regions (5′ UTRs) and upstream of the TSSs (Fig. 2C,D).

To evaluate further our choice of peak-to-gene distance, we investigated whether genes within 20 kb of Sp-Alx1 ChIP-seq peaks were enriched for genes that are sensitive to Alx1 morpholino (MO) and/or genes that are differentially expressed in PMCs (Shashikant et al., 2018b), in line with our goal of associating peaks with putative direct target genes (Fig. 1B). We found that Alx1-MO-sensitive genes were 3.0-fold (hypergeometric *P*-value=$7.92e^{-51}$) more likely to have an Sp-Alx1 ChIP-seq peak located within 20 kb (22.3%), than were other annotated genes (7.5%). Similarly, genes differentially expressed by PMCs (which we refer to as PMC DE genes) were 6.4-fold (hypergeometric *P*-value=$1.01e^{-115}$) more likely to have an Sp-Alx1 ChIP-seq peak located within 20 kb (48.1%), than were other annotated genes (7.5%). We observed that genes that met both criteria, i.e. DE genes that were also sensitive to Alx1 knockdown (which we refer to as Alx1 ‘functional targets’) were 7.6-fold (hypergeometric *P*-value=$3.27e^{-76}$) more likely to have an Sp-Alx1 ChIP-seq peak located within 20 kb (57.4%), than were other annotated genes (7.5%). Based on these observations, we decided on a 20 kb distance cutoff, which is well below the average sea urchin intergenic distance of ∼30 kb (Nam et al., 2010). Although this cutoff is relatively conservative, we reasoned that it would minimize false positives and increase our confidence in associating Sp-Alx1 ChIP-seq peaks with direct target genes.

Gene Ontology (GO) term analysis of 1604 genes located within 20 kb of Sp-Alx1 ChIP-seq peaks revealed a substantial enrichment of genes with metalloendopeptidase inhibitor activity, DNA-binding transcription factor activity and GTPase-binding activity (Fig. S3A). When using custom sea urchin-specific functional categories (Tu et al., 2014), we found a significant enrichment of genes associated with biomineralization, transcription factors and GTPase genes (Fig. S3B).

### ChIP-seq peak motif enrichment analysis

To identify putative transcription factor-binding sites, including Alx1-binding sites, in our ChIP-seq peaks, we performed *de novo* motif enrichment analysis using DREME. An advantage of DREME is its ability to discover co-regulatory motifs in addition to the primary motif. Using our set of 2019 ChIP-seq peaks located within 20 kb of annotated genes, we found the most enriched motif to be one matching the Ets1 consensus sequence, followed by Alx1, Irx and Fos::Jun motifs (Fig. 3A). Ets1 is a crucially important transcription factor in the PMC GRN that, together with Alx1, co-regulates a large fraction of genes differentially expressed by PMCs (Kurokawa et al., 1999; Rafiq et al., 2014). Fos and Jun are both expressed selectively by PMCs (Rafiq et al., 2012), although their role in the network is not known. The developmental function of Irx has not been studied but this gene is expressed predominantly in the ectoderm (Chen et al., 2011; Howard-Ashby et al., 2006). Using position weight matrices (PWMs) obtained from our DREME analysis, we performed a local enrichment analysis using CentriMo. We also included an Alx1 palindromic motif in our analysis. We observed that Ets1, Alx1 half-site and palindromic motifs tended to be located very close to peak summits, with the Alx1 motifs exhibiting the narrowest distribution (Fig. 3A,B). In contrast, Fos::Jun heterodimer motifs displayed a broader region of enrichment and the Irx motif was not significantly enriched near peak summits (Fig. 3A,C).

To identify a high-confidence set of genes that are direct targets of Sp-Alx1, we integrated published gene expression and gene knockdown data with our ChIP-seq results (Fig. 4, Fig. S4A) and found that 193 Sp-Alx1 ChIP-seq peaks were located within 20 kb of the 197 functional targets of Alx1 (Rafiq et al., 2014). As genes can have multiple peaks around them, this corresponded to 114 genes, which we consider to be Alx1 direct targets (Table S2). Many of these direct targets exist in clusters within the *S. purpuratus* genome and have several Sp-Alx1 ChIP-seq peaks and ATAC-seq/DNase-seq differential peaks located nearby (Fig. 5).

Development (2019) 146, dev180653. doi:10.1242/dev.180653

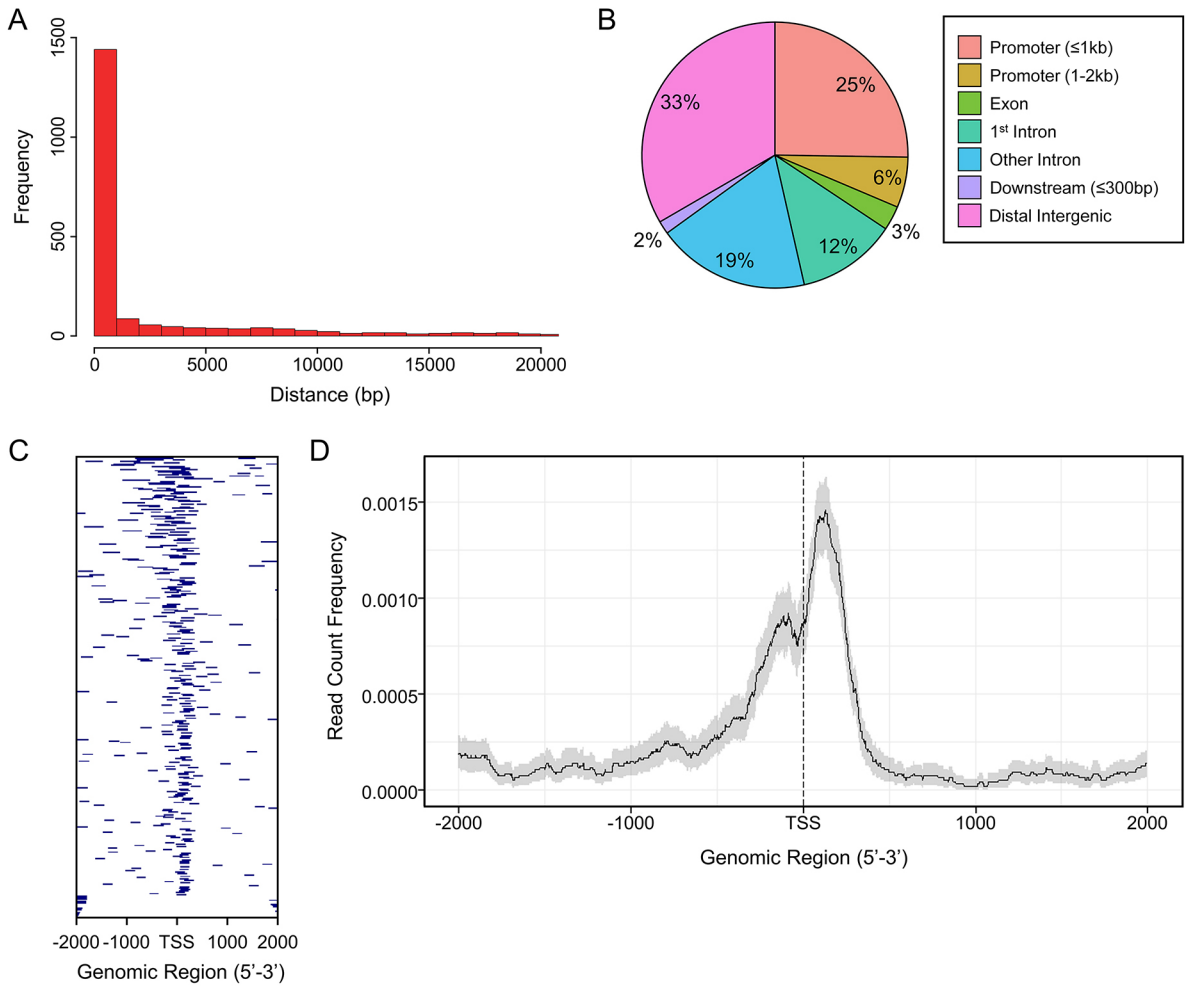



**Fig. 2. Annotation and analysis of Sp-Alx1 ChIP-seq peaks that map to open chromatin.** (A) Frequency histogram showing peak-to-gene distances, with each bar representing 1000 bp. (B) Pie chart illustrating peak location relative to the nearest annotated gene. (C) Heatmap of peaks found ±2000 bp around annotated gene TSSs. (D) Average profile of peak reads found ±2000 bp around TSSs. Confidence interval was estimated by the bootstrap method (resample=1000).

MO knockdown studies have identified several regulatory genes that are downstream of Alx1 (*Dri*, *Nfkbil1L*, *Fos*, *Alx4*, *Nk7* and *FoxB*) (Fig. 6A) (Oliveri et al., 2008; Rafiq et al., 2014). Although *Sp-Dri* and *Sp-Fos* were not identified as PMC DE genes in prior RNA-seq analysis (Rafiq et al., 2014), whole-mount *in situ* hybridization studies showed that both genes are highly expressed in the PMCs at the mesenchyme blastula stage (~24 hpf) (Amore et al., 2003; Rafiq et al., 2012). Remarkably, we found that all six regulatory genes have Sp-Alx1 ChIP-seq peaks within 20 kb, suggesting that they are regulated by Alx1 directly. To examine further the putative CRMs containing the Sp-Alx1 ChIP-seq peaks, we cloned these peaks and flanking non-coding sequences into reporter constructs (see Table S3) and injected them into fertilized eggs. We observed that CRMs associated with *Sp-Alx4*, *Sp-Fos*, *Sp-FoxB* and *Sp-Nk7* were active in driving GFP expression (Fig. 6B). We observed variability among these CRMs with respect to both their levels of GFP expression and patterns of expression (Table 1).

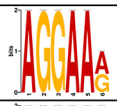
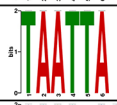
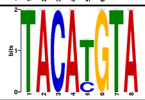
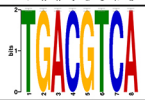
Many of our Sp-Alx1 ChIP-seq peaks were associated with genes that were not previously characterized as Alx1 functional targets (i.e. genes differentially expressed in PMCs and sensitive to Alx1 knockdown) (Rafiq et al., 2014). In an attempt to identify Alx1 direct targets that may not have met this relatively stringent threshold, we focused on ChIP-seq peaks that overlapped (by at least 1 nt) regions of chromatin that were differentially accessible in PMCs relative to non-PMC cells (Fig. S2C). Using this criterion, we identified 43 Sp-Alx1 ChIP-seq peaks that had not been flagged in our previous analyses (Fig. S4B). Surprisingly, among these we found Sp-Alx1 ChIP-seq peaks within the gene bodies of four regulatory genes (*Sp-Ets1*, *Sp-Erg*, *Sp-Jun* and *Sp-Smad1/5/8*) that are known to be expressed in the PMCs but have not been shown to be downstream of Alx1.

### Validation of Alx1-binding regions and their utility for CRM discovery

To select a set of Sp-Alx1 ChIP-seq peaks to be tested by reporter gene assays, we focused on those that were associated with direct target genes, but added the criterion that peaks overlapped (by at least 1 nt) with regions of chromatin that were previously determined to be differentially accessible in PMCs relative to non-PMC cells (Fig. S4B, Table 2). All 25 such peaks were found to have putative Alx1 half-sites, of which many (16/25) are part of Alx1 palindromic sites. Most (20/25) also contained Ets1-binding sites (Fig. S4D). We cloned these peaks and surrounding non-coding sequences into GFP reporter constructs (see Materials and Methods and Table S4). Remarkably, of the 23 putative CRMs tested (two of which contained a pair of adjacent Sp-Alx1 ChIP-seq peaks), 18 were active and drove

Development (2019) 146, dev180653. doi:10.1242/dev.180653

A

| Motif | Motif enrichment p-value | Transcription factor | Motif enriched region width centered around peak summit (bp) |
|---|---|---|---|
|  | $2.0e^{-51}$ | Ets1 | 196 |
|  | $5.2e^{-29}$ | Alx1 (half site) | 148 (half site)<br>83 (palindromic site: TAATTANATTA) |
|  | $3.1e^{-23}$ | Irx | N/A |
|  | $5.9e^{-13}$ | Fos::Jun (1) | 280 |
|  | $1.1e^{-9}$ | Fos::Jun (2) | 290 |

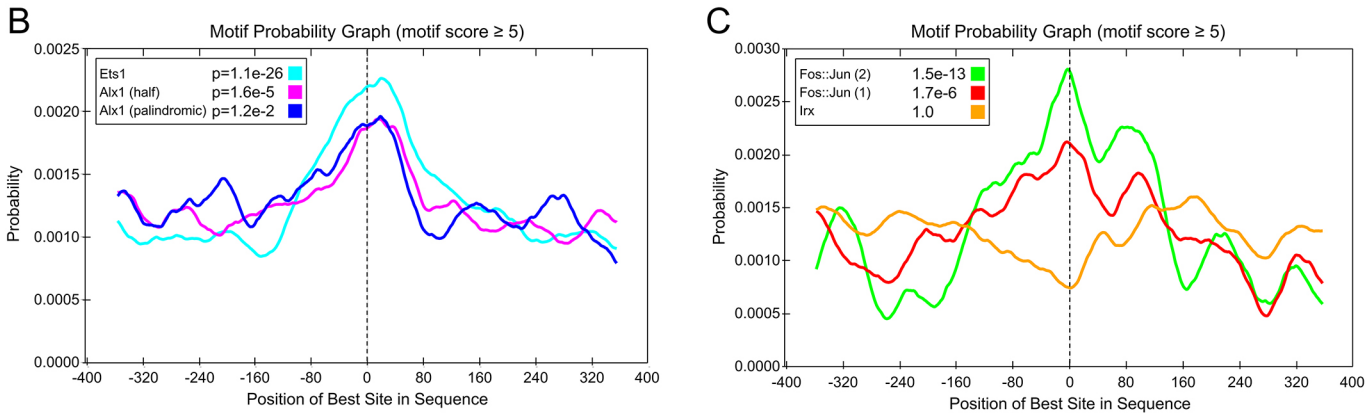



**Fig. 3. Peak motif enrichment analysis.** (A) Summary of peak *de novo* motif enrichment analysis using DREME and the width of the most enriched region for each motif determined by CentriMo. Local motif enrichment analysis using CentriMo on peak summits flanked by 400 bp sequences. (B) Motif probability curve for Ets1, Alx1 half site and Alx1 palindromic site, showing the probability of the best match to the given motifs occurring at a given position in the input sequences. (C) Motif probability curve for Fos::Jun (1), Fos::Jun (2) and Irx. Motif probability curves were smoothed according to weighted moving average of 80 bp.

GFP expression in PMCs (Fig. 7). These CRMs varied with respect to level of GFP expression and the extent of ectopic (i.e. non-PMC) expression, but most (15/18) CRMs drove GFP expression selectively in PMCs (>64% PMC only) (Table 3). As sea urchin genes are generally controlled by multiple CRMs, our reporter constructs may not have included CRMs that contribute to the temporal and spatial expression patterns of these genes. Nevertheless, our *in vivo* reporter gene studies (1) strongly support the reliability of the Sp-Alx1 ChIP-seq data, (2) highlight the power of combining ChIP-seq, chromatin accessibility, and gene expression/knockdown data for CRM discovery, and (3) identify a large number of previously uncharacterized PMC CRMs that can now be characterized in detail.

### Mutational and biochemical validation of Alx1-binding sites in a representative PMC CRM

To validate our Sp-Alx1 ChIP-seq data further, we carried out both mutational and EMSA analyses of a representative CRM identified through ChIP-seq. We chose an active, 522 bp CRM located near a novel PMC DE gene we call *Sp-EMI/TM* (WHL22.691495) (Fig. 8). Alignment of the *Sp-EMI/TM* CRM with *Lytechinus variegatus EMI/TM* intronic sequences revealed that this region is highly conserved across >50 million years of evolution (Fig. S5A). Moreover, when injected into fertilized *L. variegatus* eggs, the *Sp-EMI/TM* CRM was able to drive PMC-specific GFP expression (Fig. S5B). Through deletion studies we found that a 139 bp fragment of the *S. purpuratus* CRM was sufficient to drive PMC-specific expression of GFP in *S. purpuratus* (Figs 8 and 9). The active 139 bp fragment contained one palindromic Alx1 site that was perfectly conserved in the two sea urchin species. Mutation of the conserved, palindromic site completely abolished GFP expression in transgenic embryos (Fig. 9A, Table 4). In addition, EMSA experiments confirmed that recombinant rAlx1 protein bound to a 30 bp, double-stranded DNA probe that included the wild-type palindromic site (Fig. 9B). The binding of rAlx1 was abolished, however, when mutations were introduced into the palindromic site. Thus, our in *in vivo* reporter and EMSA studies demonstrate that Alx1 binds directly to the palindromic site and that binding is required for the transcriptional activity of the 139 bp *Sp-EMI/TM* CRM.

## DISCUSSION

### Architecture of the skeletogenic GRN

It has been proposed that sea urchins, nematodes, ascidians and several other animal groups develop by a ‘Type I’ mechanism

Development (2019) 146, dev180653. doi:10.1242/dev.180653

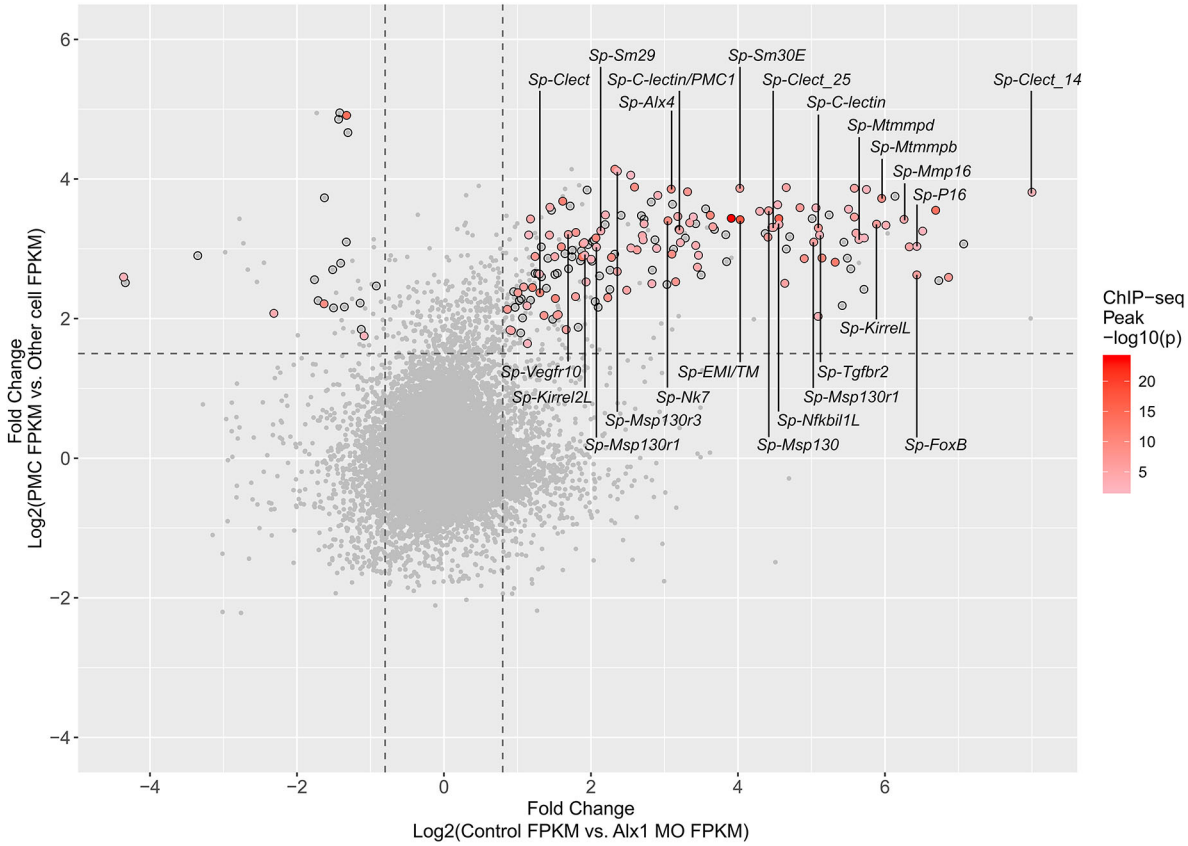



**Fig. 4. Integration of ChIP-seq and RNA-seq data to identify direct Sp-Alx1 targets.** Scatterplot shows fold change of transcripts from PMCs versus other cells plotted against fold change of transcripts from control versus Sp-Alx1 morphants, as reported by Rafiq et al. (2014). Points with black outline represent Sp-Alx1 'functional target genes' i.e. genes that are Alx1-MO-sensitive and differentially expressed in PMCs (PMC DE genes). Points with black outline that are colored red represent genes with at least one Sp-Alx1 ChIP-seq peak within 20 kb. Intensity of the red color corresponds to the significance of the associated peak, expressed as $-\log_{10}(P\text{-value})$. Several genes of interest are labeled.

(Davidson, 2006). According to this view, Type I development is characterized by the early (pre-gastrula stage) embryonic expression of terminal effector genes. A cardinal prediction that follows from these expression patterns is that the developmental GRNs that control the expression of differentiation genes in Type I embryos are relatively shallow; i.e. there are relatively few regulatory layers interposed between cell specification and effector gene activation. In euechinoid sea urchins, Alx1 is activated in the founder cells of the PMC lineage (the large micromeres) in the first cell cycle after they are born, through the activity of maternal factors and zygotically expressed Pmar1 (Ettensohn et al., 2003; Oliveri et al., 2003). Our findings reveal that Alx1 provides positive transcriptional inputs into many effector genes, revealing a direct linkage between cell specification and effector gene expression. At the same time, however, Alx1 provides direct inputs into several regulatory genes, as described below. The transcription factors encoded by these regulatory genes may control targets that are completely distinct from those of Alx1, or they may cooperate with Alx1 to regulate common effector genes by a feed-forward mechanism. To clarify the topology of the network, it will be necessary to identify the targets of the regulatory genes downstream of Alx1 and analyze the cis-regulatory control of those effector genes.

## Linking a GRN to morphogenesis

Morphogenesis is the product of hundreds of specialized proteins that directly regulate cell movement, cell adhesion, cell proliferation, and other cell behaviors that shape embryonic tissues. Developmental GRNs, including the PMC GRN, determine which genes are active or inactive in each spatiotemporal domain of the embryo. Linking developmental GRNs to effector genes that control morphogenesis is therefore essential for elucidating the connection between genotype and phenotype. Our current understanding of the architecture of the sea urchin skeletogenic GRN has been deduced primarily (although not exclusively) from gene knockdown and gene expression studies, which provide important information concerning functional gene regulatory interactions but do not reveal whether such interactions are direct or indirect. To link the skeletogenic GRN (or other GRNs) to morphogenesis, there is a need to identify the cis-regulatory elements of genes that control developmental anatomy and elucidate their direct transcriptional inputs (Wang et al., 2019).

## Intermediate network architecture

To improve our understanding of intermediate layers of regulatory control within the PMC GRN hierarchy, we searched for regulatory genes that are likely to be directly controlled by Alx1. There are 404 *S. purpuratus* genes that have been annotated as transcription factors and a significant subset of these (59/404, or 15%) have Alx1-binding sites nearby (<20 kb). Six regulatory genes (*Alx4*, *Dri*, *Fos*, *FoxB*, *Nfkbi1L* and *Nk7*) were previously shown to receive positive inputs from Alx1 (Rafiq et al., 2014) and we found that all six genes have Alx1-binding sites nearby, consistent with the view that Alx1 regulates these genes directly. Validation of these ChIP-seq peaks revealed active CRMs near *Sp-Alx4*, *Sp-Fos*, *Sp-FoxB* and *Sp-Nk7* (Fig. 6). Only the *Sp-Nk7* CRM, however, drove GFP expression selectively in PMCs. This is not unexpected, as regulatory genes

Development (2019) 146, dev180653. doi:10.1242/dev.180653

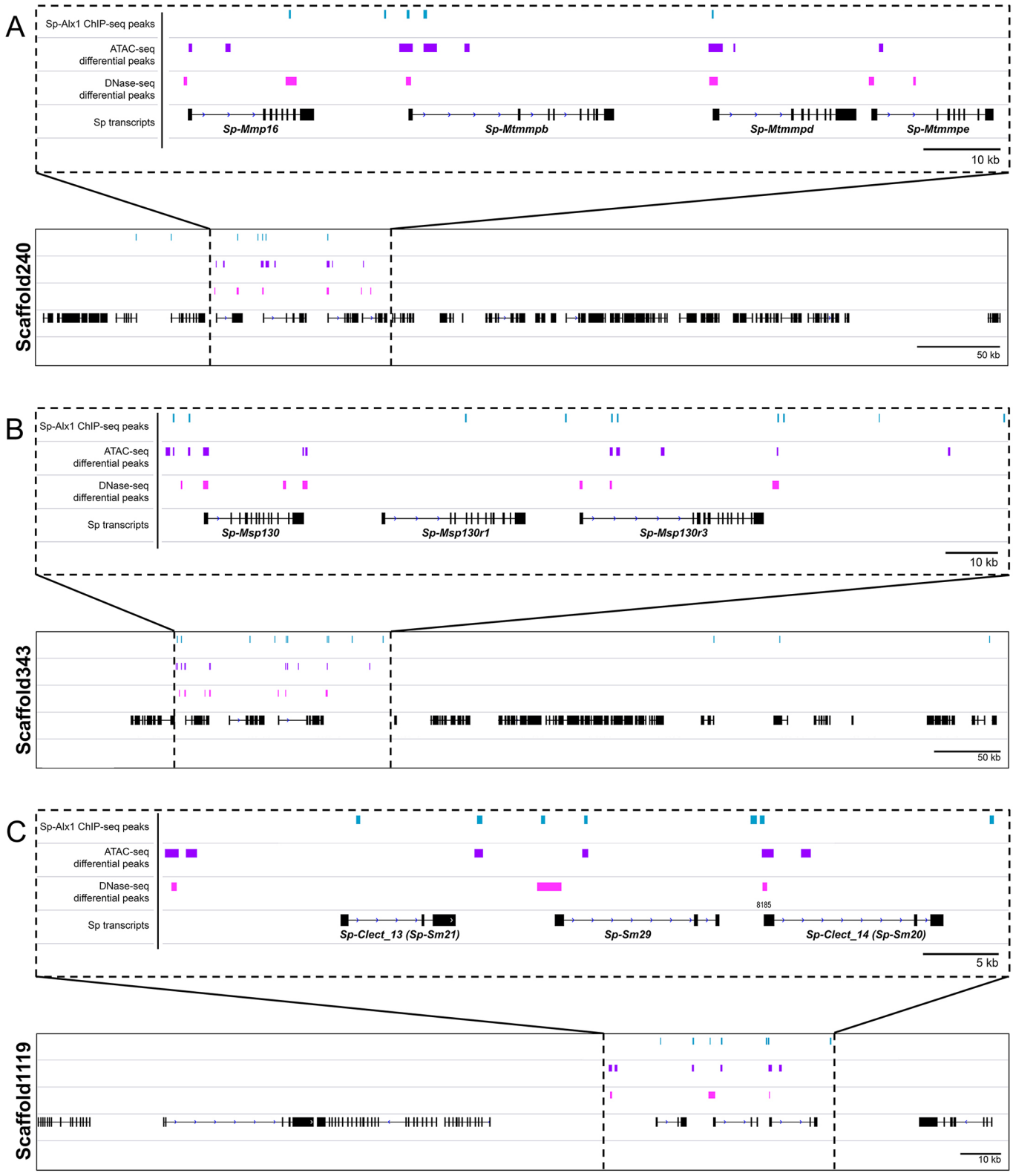



**Fig. 5. Clusters of PMC-enriched genes showing location of nearby Sp-Alx1 ChIP-seq peaks (from this study) and accessible chromatin as determined by ATAC-seq and DNase-seq (Shashikant et al., 2018b).** (A) The metalloprotease gene cluster consisting of *Sp-Mmp16*, *Sp-Mtmmpb*, *Sp-Mtmmpd* and *Sp-Mtmmpe*. (B) The MSP130 gene cluster consisting of *Sp-Msp130*, *Sp-Msp130r1* and *Sp-Msp130r3*. (C) The spicule matrix/C-lectin domain gene cluster containing *Sp-Clect_13* (*Sp-Sm21*), *Sp-Sm29* and *Sp-Clect_14* (*Sp-Sm20*).

often have complex expression patterns and cis-regulatory architectures. The Alx1-bound CRMs we identified may lack binding sites for tissue-specific repressors or contain binding sites for activators that normally drive expression in non-skeletogenic tissues. For example, Sp-FoxB is strongly expressed in PMCs at 24 hpf but at later stages is expressed in the oral ectoderm and oral

Development (2019) 146, dev180653. doi:10.1242/dev.180653

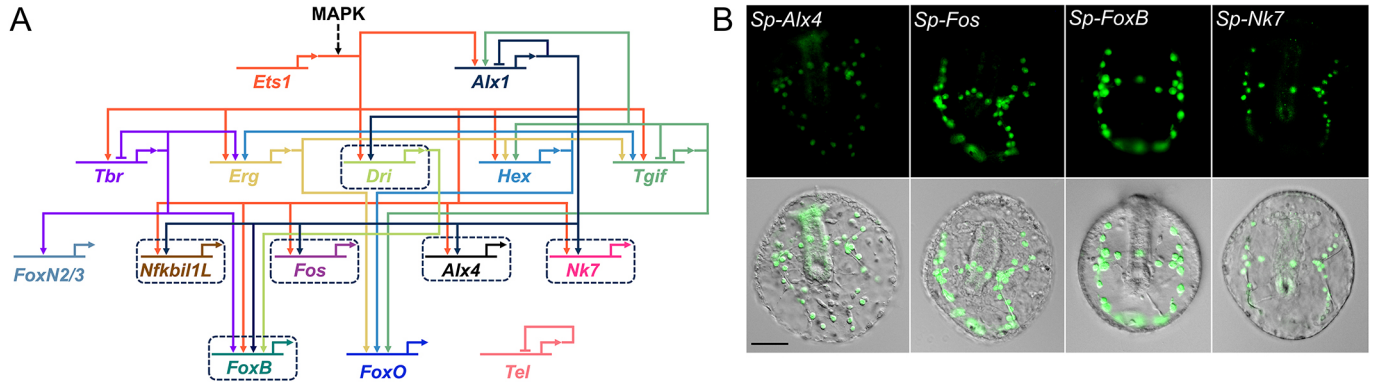



**Fig. 6. Experimental validation of Sp-Alx1 ChIP-seq peaks near regulatory genes that have been linked to the sea urchin skeletogenic GRN.** (A) Diagram of regulatory gene interactions (adapted from Shashikant et al., 2018b). All six regulatory genes that have been previously characterized as downstream of Alx1 are likely to be directly regulated by Alx1. (B) Embryos (~48 hpf) showing GFP expression in PMCs driven by CRMs containing ChIP-seq peaks. Of the eight Sp-Alx1 ChIP-seq peaks cloned into reporter constructs, four were observed to be active. Scale bar: 50 µm.

endoderm (Minokawa et al., 2004). At the stage we assessed GFP expression (48 hpf), the pattern of reporter gene expression closely resembled the endogenous expression pattern of *Sp-FoxB*.

Surprisingly, we also found evidence of direct Alx1 inputs into several regulatory genes that have not previously been considered Alx1 targets. These include known components of the skeletogenic GRN: *Sp-Ets1*, *Sp-Erg*, *Sp-Jun* and *Sp-Smad1/5/8*. A possible input from Alx1 into *Ets1* is of special interest as the reverse regulatory relationship has been proposed, i.e. it has been suggested that zygotic Ets1 is the primary activator of *Alx1* (Damle and Davidson, 2011). We hypothesize that, instead, Alx1 activates *Ets1* and establishes a positive-feedback loop whereby Ets1 maintains *Alx1* expression, consistent with data reported by Oliveri et al. (2008) and Sharma and Ettensohn (2010). The placement of Alx1 upstream of Ets1 and other regulatory genes is consistent with high-resolution NanoString data showing that zygotic expression of *Sp-Alx1* precedes that of *Sp-Ets1* and *Sp-Erg* by several hours (K. Rafiq and C.A.E., unpublished observations). In previous studies (Oliveri et al., 2008; Rafiq et al., 2014), decreases in the levels of *Ets1*, *Erg* and *Smad1/5/8* mRNAs in PMCs following Alx1 knockdown may have been obscured by the presence of these mRNAs in cell types other than the PMCs (Lapraz et al., 2009; Rizzo et al., 2006). Furthermore, building upon previous observations of repression between competing GRNs (Oliveri et al., 2008; and see below) we hypothesize that, in Alx1 morphants, activation of an alternative, non-skeletogenic mesoderm (NSM)-like GRN in the large micromere-PMC territory may have elevated expression of genes like *Ets1* and *Erg* by an Alx1-independent pathway. In summary, our findings point to several new regulatory relationships within the PMC GRN that can now be explored in detail.

### Terminal effector gene targets of Alx1

A previous study showed that Alx1 positively regulates ~50% of effector genes selectively expressed by PMCs, and an even larger fraction of the effector genes that are expressed at high levels (Rafiq et al., 2014). Surprisingly, we have found that most of these gene regulatory interactions (114/194 or 58%) appear to involve direct inputs from Alx1. Genes that are regulated directly by Alx1 have a diverse repertoire of functions, including matrix remodeling, PMC fusion, and biomineralization. For example, there are 12 sea urchin tissue inhibitor of metalloproteinases (TIMP) genes (Angerer et al., 2006). Ten of these TIMP genes are tandemly arranged in a single cluster and seven have Sp-Alx1 ChIP-seq peaks within 20 kb of the gene. Although their function in the sea urchin embryo is not known, TIMPs have been implicated in regulation of the function of matrix metalloproteases (MMPs) in other systems (reviewed by Brew and Nagase, 2010). There are nearly 240 metalloprotease genes in the sea urchin genome (Angerer et al., 2006) and metalloprotease inhibitors reversibly block spiculogenesis by PMCs *in vivo* and *in vitro* (Ingersoll and Wilt, 1996; Roe et al., 1989). We found that that 20 of the metalloprotease genes have Sp-Alx1 ChIP-seq peaks nearby, of which five had been shown previously to be downstream of Alx1 (*Sp-Anpep1*, *Sp-CbpdEL*, *Sp-Mmp16*, *Sp-Mtmmpb* and *Sp-Mtmmpd*) (Rafiq et al., 2014). Four of the Alx1-regulated

**Table 1. Experimental validation of Sp-Alx1 ChIP-seq peaks near regulatory genes by reporter gene assays**

| Gene name | Cufflinks ID | Peak number | Peak location | Embryos scored | Number of embryos expressing GFP (%) | Percentage of embryos with PMC expression only | Percentage of embryos with ectopic and PMC expression | Percentage of embryos with ectopic expression only |
|---|---|---|---|---|---|---|---|---|
| *Sp-Alx4* | WHL22.731149 | 2761 | 18,030 bp downstream | 146 | 98 (67.1) | 5.1 | 17.3 | 77.6 |
| | | 2762 | 10,502 bp downstream | – | – | – | – | – |
| | | 2763 | 2nd intron | – | – | – | – | – |
| | | 2764 | 5173 bp upstream | – | – | – | – | – |
| *Sp-Dri* | WHL22.544150 | 1993 | 9325 bp upstream | – | – | – | – | – |
| *Sp-Fos* | WHL22.538480 | 1985 | 1st intron | 363 | 172 (47.4) | 2.3 | 5.8 | 91.9 |
| *Sp-FoxB* | WHL22.743430 | 2807 | 5′ UTR | 143 | 95 (66.4) | 28.4 | 20.0 | 51.6 |
| *Sp-Nk7* | WHL22.567485 | 2131 | 8994 bp upstream | 313 | 35 (11.2) | 94.3 | 2.9 | 2.9 |
| *Sp-Nfkbil1L* | WHL22.761963 | 2877 | 4th intron | – | – | – | – | – |

–, not applicable/no expression

Development (2019) 146, dev180653. doi:10.1242/dev.180653

**Table 2. High-confidence Sp-Alx1 ChIP-seq peaks analyzed by reporter gene assays**

| Gene name | Cufflinks ID | SPU ID | Functional category | Peak number | Peak location |
|---|---|---|---|---|---|
| None | WHL22.452609 | SPU_000152 | Unclassified | 1675 | 5898 bp upstream |
| None (*Sp-EMI/TM*) | WHL22.691495 | None | Novel | 2621 | 1st intron |
| *Sp-Anpep_1* | WHL22.119959 | SPU_023693 | Metalloprotease | 477 | 744 bp upstream |
| *Sp-Anpep_1* | WHL22.119959 | SPU_023693 | Metalloprotease | 476 | 1501 bp upstream |
| *Sp-CbpdEL* | WHL22.363067 | SPU_007682 | Metalloprotease | 1354 | 1st intron |
| *Sp-Clect_14* (*Sp-Sm20*) | WHL22.39473 | SPU_005991 | Biomineralization | 186 | 5′ UTR |
| *Sp-Clect_25* | WHL22.411845 | SPU_011163 | Lectin | 1519 | 5′ UTR |
| *Sp-C-lectin/PMC1* (*Sp-Sm49*) | WHL22.411802 | SPU_027906 | Biomineralization | 1515 | 5′ UTR |
| *Sp-Fbn2* | WHL22.314476 | SPU_028567 | Unclassified | 1176 | 5512 bp upstream |
| *Sp-Hypp_2998* | WHL22.239379 | SPU_018406 | Biomineralization | 897 | 2230 bp downstream |
| *Sp-Hypp_3018* | WHL22.279611 | SPU_012929 | Unclassified | 1050 | 2053 bp upstream |
| *Sp-Hypp_313* | WHL22.722797 | SPU_000752 | Unclassified | 2736 | 1st intron |
| *Sp-Hypp_3152* (*Sp-P16r1*) | WHL22.239320 | SPU_018403 | Biomineralization | 894 | 1063 bp downstream |
| *Sp-Hypp_3153* (*Sp-P16r2*) | WHL22.239481 | SPU_018408 | Biomineralization | 901 | 329 bp downstream |
| *Sp-KirrelL* | WHL22.699052 | SPU_024995 | Unclassified (PMC cell fusion) | 2647 | 5233 bp upstream |
| *Sp-Msp130r2* | WHL22.451280 | SPU_016506 | Biomineralization | 1667 | 5′ UTR |
| *Sp-Msp130r3* | WHL22.438994 | SPU_013823 | Biomineralization | 1612 | 1st intron |
| *Sp-Msp130r3_1* | WHL22.438997 | SPU_006387 | Biomineralization | 1614 | 2543 bp downstream |
| *Sp-Mtmmpb* | WHL22.312057 | SPU_028749 | Metalloprotease | 1166 | 5′ UTR |
| *Sp-Mtmmpd* | WHL22.312130 | SPU_013669 | Metalloprotease | 1168 | 5′ UTR |
| *Sp-P16* | WHL22.239394 | SPU_018408 | Biomineralization | 898 | 844 bp upstream |
| *Sp-SerpL4* | WHL22.121357 | SPU_013378 | Immunity | 489 | 13,528 bp upstream |
| *Sp-Slc26a5_1* | WHL22.228208 | SPU_016155 | Metabolism | 871 | 6th intron |
| *Sp-Slc26a5_1* | WHL22.228208 | SPU_016155 | Metabolism | 867 | 7169 bp upstream |
| *Sp-Slc26a5_1* | WHL22.228208 | SPU_016155 | Metabolism | 866 | 7653 bp upstream |

metalloprotease genes (*Sp-Anpep1*, *Sp-CbpdEL*, *Sp-Mtmmpb* and *Sp-Mtmmpd*) are associated with what we define as high-confidence Sp-Alx1 ChIP-seq peaks (peaks that overlap regions of chromatin that are differentially open in PMCs relative to other cell types), strongly suggesting that the interactions are direct.

Other terminal differentiation genes directly regulated by Alx1 include members of several well-characterized families of biomineralization-related genes. These include the MSP130 gene family, the spicule matrix/C-lectin domain gene family and the P16 gene family. Specifically, likely direct targets of Alx1 include five MSP130 genes (*Sp-Msp130*, *Sp-Msp130r1*, *Sp-Msp130r2*, *Sp-Msp130r3* and *Sp-Msp130r3_1*), eight spicule matrix/C-lectin domain genes (*Sp-Clect*, *Sp-Clect_13*/*Sp-Sm21*, *Sp-Clect_14*/*Sp-Sm20*, *Sp-Clect_25*, *Sp-C-lectin*, *Sp-C-lectin/PMC1*/*Sp-Sm49*,

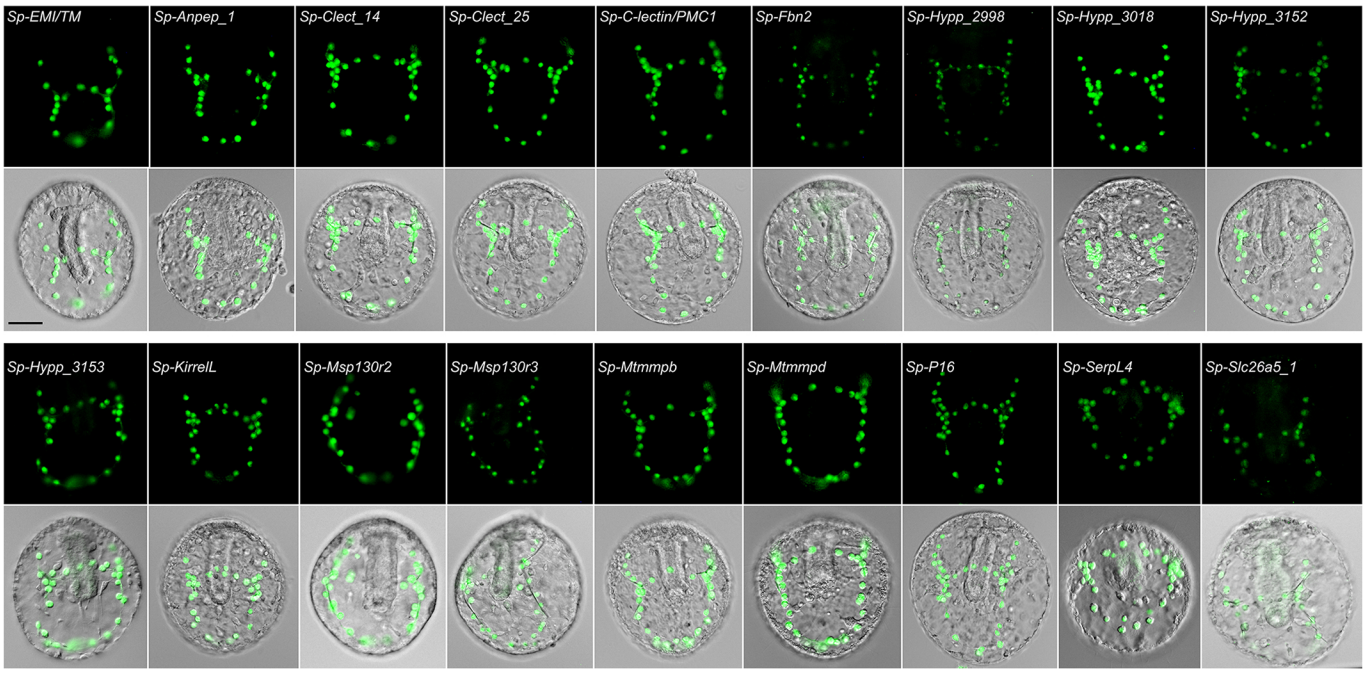



**Fig. 7. Experimental validation of high-confidence Sp-Alx1 ChIP-seq peaks near Sp-Alx1 targets.** Twenty-three putative CRMs containing 25 Sp-Alx1 ChIP-seq peaks were cloned into reporter constructs (two CRMs contain peaks adjacent to one another). Embryos (~48 hpf) showing GFP expression in PMCs driven by CRMs containing high-confidence ChIP-seq peaks are shown. Out of 18 active CRMs, 15 drove GFP expression selectively in PMCs (>64% PMC only). Scale bar: 50 µm.

Development (2019) 146, dev180653. doi:10.1242/dev.180653

**Table 3. Experimental validation of high-confidence Sp-Alx1 ChIP-seq peaks by reporter gene assays**

| Gene name | Cufflinks ID | Peak number | Number of embryos scored | Number of embryos expressing GFP (%) | Percentage of embryos with PMC expression only | Percentage of embryos with ectopic and PMC expression | Percentage of embryos with ectopic expression only |
|---|---|---|---|---|---|---|---|
| None | WHL22.452609 | 1675 | – | – | – | – | – |
| None (*Sp-EMI/TM*) | WHL22.691495 | 2621 | 171 | 62 (36.3) | 87.1 | 9.7 | 3.2 |
| *Sp-Anpep_1* | WHL22.119959 | 477<br>476 | 219 | 136 (62.1) | 72.1 | 21.3 | 6.6 |
| *Sp-CbpdEL* | WHL22.363067 | 1354 | – | – | – | – | – |
| *Sp-Clect_14* (*Sp-Sm20*) | WHL22.39473 | 186 | 208 | 117 (56.3) | 93.2 | 4.3 | 2.6 |
| *Sp-Clect_25* | WHL22.411845 | 1519 | 241 | 171 (71.0) | 89.5 | 8.8 | 1.8 |
| *Sp-C-lectin/PMC1* (*Sp-Sm49*) | WHL22.411802 | 1515 | 231 | 52 (22.5) | 78.8 | 7.7 | 13.5 |
| *Sp-Fbn2* | WHL22.314476 | 1176 | 147 | 55 (37.4) | 25.5 | 18.2 | 56.4 |
| *Sp-Hypp_2998* | WHL22.239379 | 897 | 142 | 31 (21.8) | 64.5 | 16.1 | 19.4 |
| *Sp-Hypp_3018* | WHL22.279611 | 1050 | 254 | 157 (61.8) | 75.2 | 14.6 | 10.2 |
| *Sp-Hypp_313* | WHL22.722797 | 2736 | – | – | – | – | – |
| *Sp-Hypp_3152 (Sp-P16r1)* | WHL22.239320 | 894 | 263 | 201 (76.4) | 96.5 | 3.0 | 0.5 |
| *Sp-Hypp_3153 (Sp-P16r2)* | WHL22.239394 | 898 | 321 | 123 (38.3) | 85.4 | 8.9 | 5.7 |
| *Sp-KirrelL* | WHL22.699052 | 2647 | 137 | 94 (68.6) | 69.1 | 28.7 | 2.1 |
| *Sp-Msp130r2* | WHL22.451280 | 1667 | 157 | 91 (58.0) | 79.1 | 15.4 | 5.5 |
| *Sp-Msp130r3* | WHL22.438994 | 1612 | 189 | 40 (21.2) | 80.0 | 5.0 | 15.0 |
| *Sp-Msp130r3_1* | WHL22.438997 | 1614 | – | – | – | – | – |
| *Sp-Mtmmpb* | WHL22.312057 | 1166 | 211 | 73 (34.6) | 72.6 | 12.3 | 15.1 |
| *Sp-Mtmmpd* | WHL22.312130 | 1168 | 201 | 76 (37.8) | 47.4 | 38.2 | 14.5 |
| *Sp-P16* | WHL22.239481 | 901 | 217 | 4 (1.8) | 75.0 | 25.0 | 0.0 |
| *Sp-SerpL4* | WHL22.121357 | 489 | 383 | 163 (42.6) | 16.6 | 30.7 | 52.8 |
| *Sp-Slc26a5_1* | WHL22.228208 | 871 | 211 | 37 (17.5) | 91.9 | 2.7 | 5.4 |
| | | 867<br>866 | – | – | – | – | – |

–, not applicable/no expression

*Sp-Sm29* and *Sp-Sm30E*) and three P16 genes (*Sp-P16*, *Sp-P16r1* and *Sp-P16r2*). We also determined that Alx1 directly targets two Ig-domain genes, *SpKirrelL* and *Sp-Kirrel2L*, one of which (*Sp-KirrelL*) is required for PMC fusion (Ettensohn and Dey, 2017). Remarkably, we found that Alx1 targets also include important, PMC-specific signaling receptors that play crucial roles in PMC guidance, migration and patterning. For example, we found that Alx1 provides direct transcriptional inputs into a VEGF receptor (*Sp-Vefgr-Ig10*) (Adomako-Ankomah and Ettensohn, 2013; Duloquin et al., 2007), an FGF receptor (*Sp-Fgfr2_1*) (Lapraz et al., 2006) and a TGFβ receptor (*Sp-Tgfbr2*) (Sun and Ettensohn, 2017). Taken together, direct Alx1 targets define a genetic subcircuit that impinges on almost all aspects of PMC behavior, including migration, fusion and biomineralization.

### Co-regulation of effector genes by Alx1 and Ets1

*Ets1* knockdown or the over-expression of a dominant-negative form of Ets1 inhibit PMC ingression and specification, effects that are also seen following *Alx1* knockdown (Ettensohn et al., 2003; Kurokawa et al., 1999). Out of 170 genes regulated by Ets1, 85% showed significant decreases in expression in *Alx1* morphants (Rafiq et al., 2014). This striking, overlapping regulatory control over downstream effector genes by Alx1 and Ets1 has not been fully elucidated. It has been proposed that Ets1 and Alx1 might regulate effector genes via a feed-forward mechanism, whereby Ets1 positively regulates *Alx1* and both regulatory inputs are required to drive the expression of downstream genes (Ets1→Alx1, Ets1+Alx1→effector) (Oliveri et al., 2008). Our observation that consensus Ets1-binding sites are highly enriched in Alx1 ChIP-seq peaks provides strong evidence in support of this hypothesis and suggests that such a mechanism operates to regulate the expression of a large fraction of effector genes. We found that 136/192 (70.8%) peaks near Alx1 direct targets and 20/25 (80.0%) high-confidence peaks had consensus Ets1-binding sites, in addition to Alx1 half or palindromic sites, consistent with previous studies showing enrichment of both Alx1 and Ets1 sites in regions differentially open in the PMCs (Fig. S4C,D) (Shashikant et al., 2018b).

### Competition between GRNs: the role of Alx1 in excluding alternative regulatory states

Oliveri et al. (2008) found that, in addition to its role as an activator of the PMC regulatory program, Alx1 prevents the deployment of the pigment cell GRN in the large micromere territory. Rafiq et al. (2014) provided evidence that this exclusionary function of Alx1 extends to other non-skeletogenic, mesodermal gene regulatory programs. The molecular mechanisms that might account for such an exclusionary function, and for competition between GRNs more generally, are of great interest but are presently unknown. In our analysis, we identified Sp-Alx1 ChIP-seq peaks near regulatory genes that are upregulated in Alx1 morphant embryos at 28-30 hpf (*Sp-Scl*, *Sp-Irf4*, *Sp-Z166* and *Sp-Nr4a*) (Rafiq et al., 2014). *Sp-Scl* is specifically expressed in oral non-skeletogenic mesoderm cells at this stage, although expression is detected in PMCs at 48 hpf (Solek et al., 2013). Notably, we identified an intronic ChIP-seq peak within *Sp-Scl* at a location previously shown to be differentially accessible in PMCs, both in ATAC-seq and DNase-seq data (Shashikant et al., 2018b). At present, we cannot distinguish whether the function of this Alx1 input is to repress *Sp-Scl* in PMCs at the mesenchyme blastula stage or to prime the gene for activation at later developmental stages. *Sp-Z166* and *Sp-Irf4*, in contrast, are expressed specifically by presumptive pigment cells (Materna et al., 2013) and blastocoelar cells (J.M.K. and C.A.E., unpublished observations), respectively. Our findings support the view that Alx1 provides direct, negative inputs into these NSM regulatory genes that prevent their deployment in presumptive PMCs.

Development (2019) 146, dev180653. doi:10.1242/dev.180653

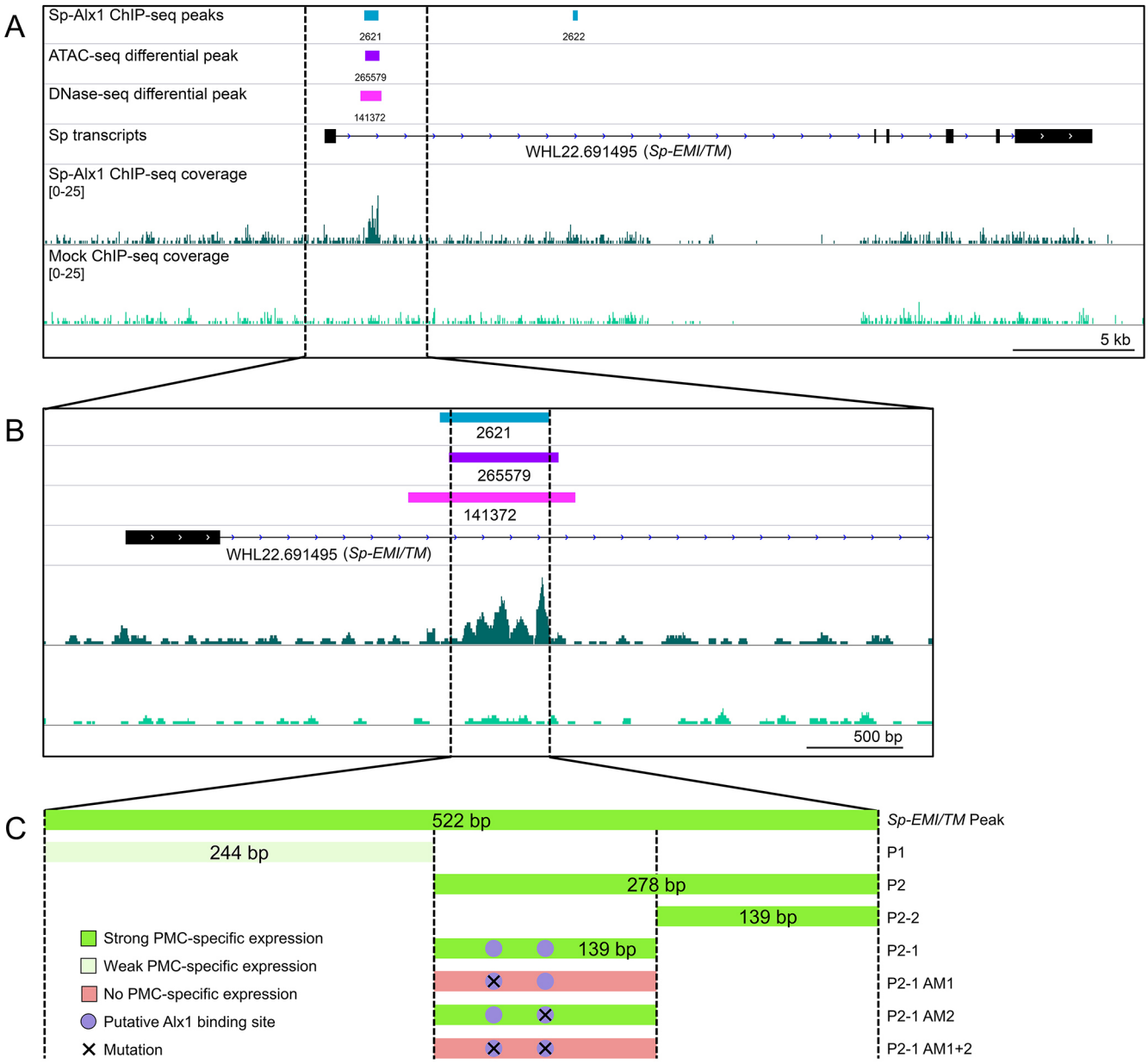



**Fig. 8. Detailed dissection and mutational analysis of Sp-EMI/TM CRM.** (A) Genome tracks showing the location of the Sp-Alx1 ChIP-seq peak within the first intron of the *Sp-EMI/TM* transcript. (B) The region of overlap between the ChIP-seq peak, ATAC-seq differential peak, and DNase-seq differential peak was cloned into a reporter construct for detailed analysis. (C) Schematic of the 522 bp CRM. Position of putative Alx1-binding sites are represented by purple circles. Crosses indicate mutated binding sites. Constructs in dark green represent active CRMs that supported strong, PMC-specific GFP expression. Constructs in light green represent CRMs that supported weak, PMC-specific GFP expression. Constructs in pink represent CRMs that were not active.

### Alx1 and the evolution of morphological novelty

Our previous study provided evidence of a trans-regulatory change, specifically a gene duplication event that permitted the functional specialization of the Alx1 protein through changes in the exon-intron organization of *Alx1* (Khor and Ettensohn, 2017). We found that the gain of a small, novel domain (the D2 domain) imparted new functions to Alx1 and supported the evolution of skeletogenesis in echinoderms. In our present study, we provide evidence that a large part of the sea urchin skeletogenic GRN is directly controlled by Alx1, including terminal differentiation genes that are expressed in adult skeletogenic centers, such as the spicule matrix and MSP130 genes. Hence, we hypothesize that evolutionary changes in the amino acid sequence of Alx1 that allowed it to acquire new targets preceded cis-regulatory changes in the *Alx1* gene that resulted in its embryonic expression. Consistent with observations by Koga et al. (2016), a heterochronic shift in *Alx1* expression from adult skeletogenic centers to PMCs may have been sufficient to directly activate a large cohort of biomineralization genes and transfer skeletogenesis into the embryo.

## MATERIALS AND METHODS

### Animals

Adult *Strongylocentrotus purpuratus* were acquired from Patrick Leahy (California Institute of Technology, USA). Gamete release was induced by intracoelomic injection of 0.5 M KCl and fertilized embryos were cultured in artificial seawater at 15°C in temperature-controlled incubators.

Development (2019) 146, dev180653. doi:10.1242/dev.180653

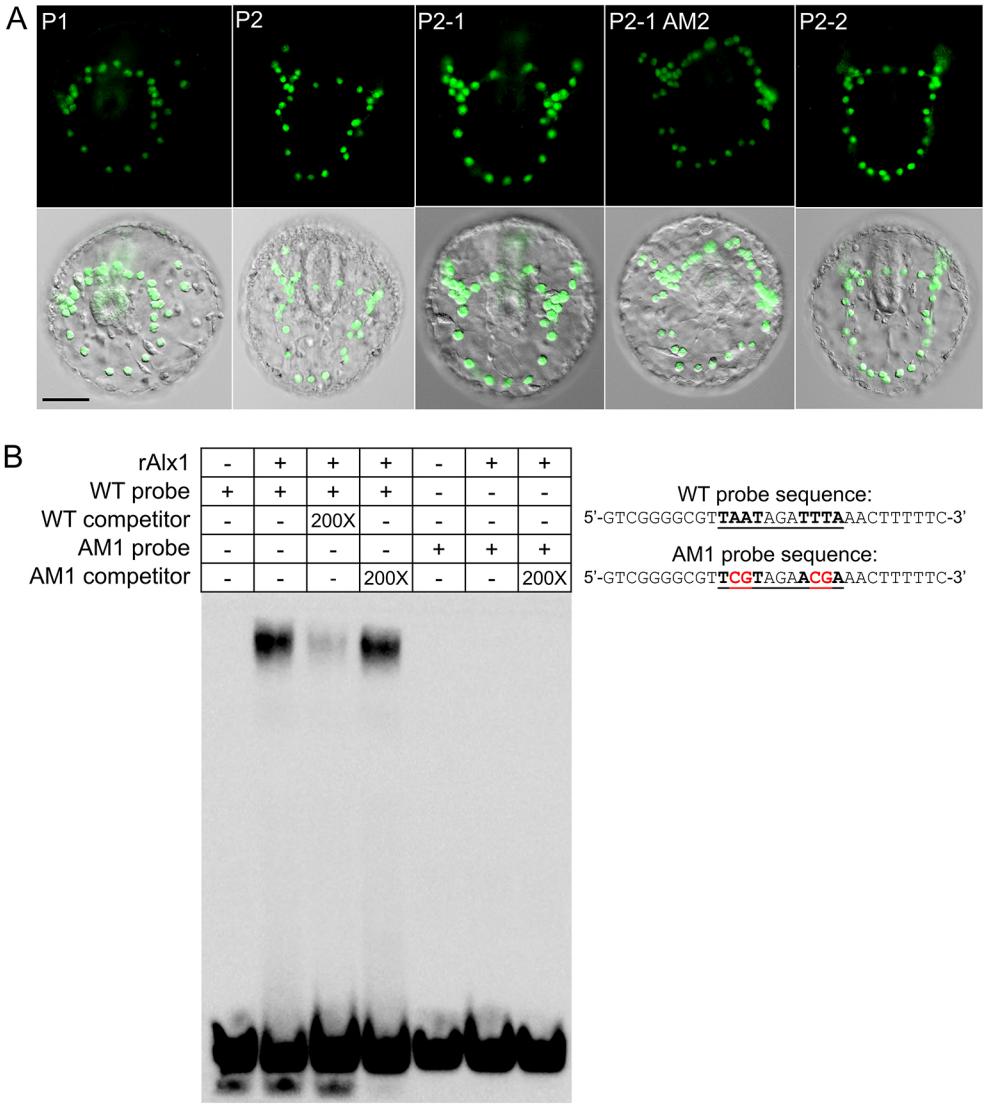



**Fig. 9. Experimental validation of Alx1 palindromic binding to the *Sp-EMI/TM* CRM.** (A) GFP expression in PMCs, driven by different truncations and site-specific mutations of the *Sp-EMI/TM* CRM (see Fig. 8). (B) EMSA showing association of rAlx1 with double-stranded DNA containing the wild-type Alx1 palindromic binding site, which was abolished when the binding site was mutated. Scale bar: 50 µm.

### Morpholino injections

Microinjection of MOs (Gene Tools) into fertilized eggs was performed as described (Cheers and Ettensohn, 2004). The translation-blocking Sp-Alx1 MO was complementary to the 5′ UTR and had the sequence 5′-TATTGAGTTAAGTCTCGGCACGACA-3′. This MO has been characterized in previous studies (Ettensohn et al., 2003; Rafiq et al., 2014). MO injection solutions contained 4 mM Sp-Alx1 MO, 20% glycerol (vol/vol) and 0.1% rhodamine dextran (wt/vol).

### Sp-Alx1 antibody immunofluorescence staining

Custom affinity-purified rabbit polyclonal Sp-Alx1 antibody was produced by Biomatik (Cambridge, Ontario, Canada) using the following chemically synthesized peptide as immunogen: QPPAPVEGAMLRICRNLQNLRRE-FDSRK. This peptide sequence is contained within a region unique to Alx1 known as the D2 domain (Khor and Ettensohn, 2017). The antibody was validated in control embryos and Sp-Alx1 morphants by double immunofluorescence staining with a monoclonal antibody (mAb) 6a9, which recognizes PMC-specific cell surface proteins of the MSP130 family (Ettensohn and McClay, 1988). Control embryos and Sp-Alx1 morphants were collected and transferred to round-bottom 96-well plates for fixation and staining, as described (Khor and Ettensohn, 2017). The embryos were double-stained with mAb 6a9 tissue culture supernatant and Sp-Alx1 antibody and counterstained with Hoechst dye (see supplementary Materials and Methods). Mounted embryos were imaged using a Zeiss LSM 880 confocal microscope.

**Table 4. Detailed dissection and mutational analysis of Sp-EMI/TM CRM by reporter gene assays**

| Construct | Number of embryos scored | Number of embryos expressing GFP (%) | Percentage of embryos with PMC expression only | Percentage of embryos with ectopic and PMC expression | Percentage of embryos with ectopic expression only |
|---|---|---|---|---|---|
| *Sp-EMI/TM* peak | 171 | 62 (36.3) | 87.1 | 9.7 | 3.2 |
| P1 | 377 | 50 (13.3) | 60.0 | 10.0 | 30.0 |
| P2 | 178 | 72 (40.4) | 84.7 | 8.3 | 6.9 |
| P2-1 | 449 | 104 (23.2) | 72.1 | 8.7 | 19.2 |
| P2-1 AM1 | 274 | 23 (8.4) | 8.7 | 4.3 | 87.0 |
| P2-1 AM2 | 264 | 57 (21.6) | 47.4 | 14.0 | 38.6 |
| P2-1 AM1+2 | 265 | 33 (12.5) | 3.0 | 3.0 | 93.9 |
| P2-2 | 676 | 122 (34.2) | 75.3 | 6.8 | 17.9 |

Development (2019) 146, dev180653. doi:10.1242/dev.180653

### Recombinant Alx1 immunoprecipitation

Recombinant *Lytechinus variegatus* Alx1 (rAlx1) and Alx4 (rAlx4) were expressed using the pETDuet-1 vector (Novagen, 71146) (see supplementary Materials and Methods). The bacterial cultures were pelleted and lysed in RIPA buffer (1× PBS, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with Roche cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, 11836153001). For IP of rAlx1, bacterial lysates were diluted in RIPA buffer and incubated with 5 µg Sp-Alx1 antibody at 4°C overnight. The IP immunoblot was probed with 1:1000 of 2.5 µg/µl Sp-Alx1 antibody and 1:5000 mouse monoclonal [SB62a] anti-rabbit IgG light chain (HRP) secondary antibody (Abcam, ab99697).

### Chromatin immunoprecipitation (ChIP)

ChIP was performed as previously described (Cary et al., 2017; Cheatle Jarvela et al., 2014; Mortazavi et al., 2006), with some modifications (see supplementary Materials and Methods). Three independent cultures of *S. purpuratus* were collected and processed at the mesenchyme blastula stage (24 hpf) (Fig. S2A). For each culture, ChIP was performed using 5 µg of Sp-Alx1 antibody and 5 µg of normal rabbit IgG (Sigma-Aldrich, 12370) was used for mock IP. Immunoprecipitated chromatin from the three independent ChIP experiments was pooled to prepare sequencing libraries from Sp-Alx1 immunoprecipitated DNA and mock IP DNA. ChIP-seq library construction and Illumina-based sequencing (Illumina HiSeq 2500 SE50) were carried out by Novogene Corporation.

### ChIP-seq bioinformatic pipeline

The bioinformatic workflow for processing the raw ChIP-seq data is summarized in Fig. S2B (see supplementary Materials and Methods). The initial set of 2906 peaks called by MACS2 was first filtered using Bedtools (v2.27) (Quinlan and Hall, 2010) to identify peaks that intersect with regions of open chromatin, defined as the union of ATAC-seq and DNase-seq RPSs from 24 hpf *S. purpuratus* embryos (Shashikant et al., 2018b). The resulting 2353 peaks were then used for peak annotation and visualization using ChIPseeker (v3.8) (Yu et al., 2015). Subsequently, the peaks were filtered to identify those located within 20 kb of annotated genes, resulting in 2019 peaks which were used for motif discovery and motif enrichment analysis using DREME (Bailey, 2011) and CentriMo (Bailey and MacHanick, 2012), part of the MEME suite (v5.0.2) of motif-based sequence analysis tools (http://meme-suite.org/). For CentriMo analysis, peaks of uniform lengths were generated by retrieving 400 bp sequences flanking MACS2-defined peak summits. GO term enrichment analysis was performed on genes that were found within 20 kb of at least one of the 2019 Sp-Alx1 ChIP-seq peaks using the webtool and GO annotations available at http://geneontology.org (The Gene Ontology Consortium, 2019; The Gene Ontology Consortium et al., 2011). Sea urchin-specific functional category enrichment analysis was also performed on the same gene set using functional assignments based on manual annotation (Sea Urchin Genome Sequencing Consortium, 2006) and on Blast2GO-derived GO terms (Tu et al., 2012). Statistical significance of term enrichment was assessed using a hypergeometric test and the hypergeometric *P*-values were corrected for multiple comparisons using the Bonferroni method. High-confidence Sp-Alx1 ChIP-seq peaks were scanned for Alx1 and Ets1 motifs (output from DREME) using FIMO (Find Individual Motif Occurrences) (Grant et al., 2011), one of the MEME suite tools.

### GFP reporter assay

Putative CRMs containing Sp-Alx1 ChIP-seq peaks were amplified from *S. purpuratus* genomic DNA and cloned upstream of the basal *Sp-Endo16* promoter in *EpGFPII* vector (see Tables S3 and S4 and supplementary Materials and Methods). ChIP-seq peaks that were adjacent to one another were typically amplified together and cloned as one DNA fragment. Overlap extension PCR was used to mutate putative transcription factor-binding sites, as previously described (Khor and Ettensohn, 2017). Microinjection of reporter constructs was performed following established protocols (Arnone et al., 2004) (see supplementary Materials and Methods). GFP expression in injected embryos were assayed at the late gastrula stage (48 hpf) by fluorescence microscopy. The total number of injected embryos (indicated by the presence of Texas Red dextran), the number of embryos showing PMC-specific GFP expression, the number of embryos showing PMC and ectopic GFP expression, and the number of embryos with only ectopic GFP expression were scored.

### Electrophoretic mobility shift assay

Recombinant double affinity-tagged rAlx1 was expressed in bacteria using the pTXB1 plasmid (New England Biolabs, N6707S). The protein was sequentially purified using two tags, first with the N-terminal His tag and then the C-terminal Intein tag (see supplementary Materials and Methods). The Intein tag was cleaved off while bound to the chitin column during purification. His-rAlx1 was processed according to previous published methods to achieve maximum solubility (Pullara et al., 2013). The protein was eluted, concentrated and desalted in 300 mM NaCl, 50 mM Tris pH 6.8, 0.1% Triton X-100 and Roche cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, 11836153001). For EMSA, 200 ng of purified rAlx1 was used per reaction with 20 fmol of biotinylated probes and 4 pmol of non-biotinylated probe as competitor when applicable. The reactions were run on 8% polyacrylamide gel and visualized using the LightShift Chemiluminescent EMSA kit (Thermo Fisher Scientific, 20148) (see supplementary Materials and Methods).

#### Competing interests

The authors declare no competing or financial interests.

#### Author contributions

Conceptualization: J.M.K., C.A.E.; Methodology: J.M.K., J.G.-S.; Validation: J.M.K.; Formal analysis: J.M.K.; Investigation: J.M.K.; Data curation: J.M.K.; Writing - original draft: J.M.K., C.A.E.; Writing - review & editing: J.M.K., C.A.E.; Visualization: J.M.K.; Supervision: C.A.E.; Project administration: C.A.E.; Funding acquisition: C.A.E.

#### Funding

This work was supported by National Science Foundation (IOS-1354973 to C.A.E.).

#### Data availability

Raw read and processed files for ChIP-seq data are available in Gene Expression Omnibus under accession number GSE131370.

#### Supplementary information

Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.180653.supplemental

#### References

**Adomako-Ankomah, A. and Ettensohn, C. A.** (2013). Growth factor-mediated mesodermal cell guidance and skeletogenesis during sea urchin gastrulation. *Development* **140**, 4214-4225. doi:10.1242/dev.100479

**Amore, G., Yavrouian, R. G., Peterson, K. J., Ransick, A., McClay, D. R. and Davidson, E. H.** (2003). Spdeadringer, a sea urchin embryo gene required separately in skeletogenic and oral ectoderm gene regulatory networks. *Dev. Biol.* **261**, 55-81. doi:10.1016/S0012-1606(03)00278-1

**Angerer, L., Hussain, S., Wei, Z. and Livingston, B. T.** (2006). Sea urchin metalloproteases: a genomic survey of the BMP-1/tolloid-like, MMP and ADAM families. *Dev. Biol.* **300**, 267-281. doi:10.1016/j.ydbio.2006.07.046

**Arnone, M. I., Dmochowski, I. J. and 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

**Bailey, T. L.** (2011). DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653-1659. doi:10.1093/bioinformatics/btr261

**Bailey, T. L. and MacHanick, P.** (2012). Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res.* **40**, e128. doi:10.1093/nar/gks433

**Barsi, J. C., Tu, Q. and Davidson, E. H.** (2014). General approach for in vivo recovery of cell type-specific effector gene sets. *Genome Res.* **24**, 860-868. doi:10.1101/gr.167668.113

**Bolger, A. M., Lohse, M. and Usadel, B.** (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120. doi:10.1093/bioinformatics/btu170

**Brew, K. and Nagase, H.** (2010). The tissue inhibitors of metalloproteinases (TIMPs): An ancient family with structural and functional diversity. *Biochim. Biophys. Acta Mol. Cell Res.* **1803**, 55-71. doi:10.1016/j.bbamcr.2010.01.003

**Cary, G. A., Cheatle Jarvela, A. M., Francolini, R. D. and Hinman, V. F.** (2017). Genome-wide use of high- and low-affinity Tbrain transcription factor binding sites during echinoderm development. *Proc. Natl. Acad. Sci. USA* **114**, 5854-5861. doi:10.1073/pnas.1610611114

**Cheatle Jarvela, A. M., Brubaker, L., Vedenko, A., Gupta, A., Armitage, B. A., Bulyk, M. L. and Hinman, V. F.** (2014). Modular evolution of DNA-binding preference of a tbrain transcription factor provides a mechanism for modifying gene regulatory networks. *Mol. Biol. Evol.* **31**, 2672-2688. doi:10.1093/molbev/msu213

**Cheers, M. S. and Ettensohn, C. A.** (2004). Rapid microinjection of fertilized eggs. In *Methods in Cell Biology* (ed. C. A. Ettensohn, G. A. Wray, and G. M. Wessel), pp. 287-310. Elsevier Academic Press.

**Chen, J.-H., Luo, Y.-J. and Su, Y.-H.** (2011). The dynamic gene expression patterns of transcription factors constituting the sea urchin aboral ectoderm gene regulatory network. *Dev. Dyn.* **240**, 250-260. doi:10.1002/dvdy.22514

**Czarkwiani, A., Dylus, D. V. and Oliveri, P.** (2013). Expression of skeletogenic genes during arm regeneration in the brittle star Amphiura filiformis. *Gene Expr. Patterns* **13**, 464-472. doi:10.1016/j.gep.2013.09.002

**Damle, S. and Davidson, E. H.** (2011). Precise cis-regulatory control of spatial and temporal expression of the alx-1 gene in the skeletogenic lineage of S. purpuratus. *Dev. Biol.* **357**, 505–517. doi:10.1016/j.ydbio.2011.06.016

**Davidson, E. H.** (2006). *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*, New edn. Oxford: Elsevier/Academic Press.

**Duloquin, L., Lhomond, G. and Gache, C.** (2007). Localized VEGF signaling from ectoderm to mesenchyme cells controls morphogenesis of the sea urchin embryo skeleton. *Development* **134**, 2293-2302. doi:10.1242/dev.005108

**Dylus, D. V., Czarkwiani, A., Stångberg, J., Ortega-Martinez, O., Dupont, S. and Oliveri, P.** (2016). Large-scale gene expression study in the ophiuroid Amphiura filiformis provides insights into evolution of gene regulatory networks. *Evodevo* **7**, 2. doi:10.1186/s13227-015-0039-x

**Erkenbrack, E. M. and Davidson, E. H.** (2015). Evolutionary rewiring of gene regulatory network linkages at divergence of the echinoid subclasses. *Proc. Natl. Acad. Sci. USA* **112**, E4075-E4084. doi:10.1073/pnas.1509845112

**Ettensohn, C. A. and Dey, D.** (2017). KirrelL, a member of the Ig-domain superfamily of adhesion proteins, is essential for fusion of primary mesenchyme cells in the sea urchin embryo. *Dev. Biol.* **421**, 258-270. doi:10.1016/j.ydbio.2016.11.006

**Ettensohn, C. A. and McClay, D. R.** (1988). Cell lineage conversion in the sea urchin embryo. *Dev. Biol.* **125**, 396-409. doi:10.1016/0012-1606(88)90220-5

**Ettensohn, C. A., Illies, M. R., Oliveri, P. and De Jong, D. L.** (2003). Alx1, a member of the Cart1/Alx3/Alx4 subfamily of Paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo. *Development* **130**, 2917-2928. doi:10.1242/dev.00511

**Ettensohn, C. A., Kitazawa, C., Cheers, M. S., Leonard, J. D. and Sharma, T.** (2007). Gene regulatory networks and developmental plasticity in the early sea urchin embryo: alternative deployment of the skeletogenic gene regulatory network. *Development* **134**, 3077-3087. doi:10.1242/dev.009092

**Fuchikami, T., Mitsunaga-Nakatsubo, K., Amemiya, S., Hosomi, T., Watanabe, T., Kurokawa, D., Kataoka, M., Harada, Y., Satoh, N., Kusunoki, S. et al.** (2002). T-brain homologue (HpTb) is involved in the archenteron induction signals of micromere descendant cells in the sea urchin embryo. *Development* **129**, 5205-5216.

**Gao, F. and Davidson, E. H.** (2008). Transfer of a large gene regulatory apparatus to a new developmental address in echinoid evolution. *Proc. Natl. Acad. Sci. USA* **105**, 6091-6096. doi:10.1073/pnas.0801201105

**Gao, F., Thompson, J. R., Petsios, E., Erkenbrack, E., Moats, R. A., Bottjer, D. J. and Davidson, E. H.** (2015). Juvenile skeletogenesis in anciently diverged sea urchin clades. *Dev. Biol.* **400**, 148-158. doi:10.1016/j.ydbio.2015.01.017

**Grant, C. E., Bailey, T. L. and Noble, W. S.** (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018. doi:10.1093/bioinformatics/btr064

**Howard-Ashby, M., Materna, S. C., Brown, C. T., Chen, L., Cameron, R. A. and Davidson, E. H.** (2006). Identification and characterization of homeobox transcription factor genes in Strongylocentrotus purpuratus, and their expression in embryonic development. *Dev. Biol.* **300**, 74-89. doi:10.1016/j.ydbio.2006.08.039

**Illies, M. R., Peeler, M. T., Dechtiaruk, A. M. and Ettensohn, C. A.** (2002). Identification and developmental expression of new biomineralization proteins in the sea urchin Strongylocentrotus purpuratus. *Dev. Genes Evol.* **212**, 419-431. doi:10.1007/s00427-002-0261-0

**Ingersoll, E. P. and Wilt, F. H.** (1996). Metalloproteinase inhibitors block spicule formation by primary mesenchyme cells in the sea urchin embryo. *Mol. Biol. Cell* **7**, 704. doi:10.1006/dbio.1998.8857

**Khor, J. M. and Ettensohn, C. A.** (2017). Functional divergence of paralogous transcription factors supported the evolution of biomineralization in echinoderms. *Elife* **6**, 1-21. doi:10.7554/eLife.32728

**Killian, C. E., Croker, L. and Wilt, F. H.** (2010). SpSM30 gene family expression patterns in embryonic and adult biomineralized tissues of the sea urchin, Strongylocentrotus purpuratus. *Gene Expr. Patterns* **10**, 135-139. doi:10.1016/j.gep.2010.01.002

**Koga, H., Fujitani, H., Morino, Y., Miyamoto, N., Tsuchimoto, J., Shibata, T. F., Nozawa, M., Shigenobu, S., Ogura, A., Tachibana, K. et al.** (2016). Experimental approach reveals the role of alx1 in the evolution of the echinoderm larval skeleton. *PLoS ONE* **11**, 1-19. doi:10.1371/journal.pone.0149067

**Kurokawa, D., Kitajima, T., Mitsunaga-Nakatsubo, K., Amemiya, S., Shimada, H. and Akasaka, K.** (1999). HpEts, an ets-related transcription factor implicated in primary mesenchyme cell differentiation in the sea urchin embryo. *Mech. Dev.* **80**, 41-52. doi:10.1016/S0925-4773(98)00192-0

**Langmead, B. and Salzberg, S. L.** (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359. doi:10.1038/nmeth.1923

**Lapraz, F., Röttinger, E., Duboc, V., Range, R., Duloquin, L., Walton, K., Wu, S.-Y., Bradham, C., Loza, M. A., Hibino, T. et al.** (2006). RTK and TGF-β signaling pathways genes in the sea urchin genome. *Dev. Biol.* **300**, 132-152. doi:10.1016/j.ydbio.2006.08.048

**Lapraz, F., Besnardeau, L. and Lepage, T.** (2009). Patterning of the dorsal-ventral axis in echinoderms: Insights into the evolution of the BMP-chordin signaling network. *PLoS Biol.* **7**, e1000248. doi:10.1371/journal.pbio.1000248

**Li, H., Handsaker, B., Wysoker, A., Fennell, T., Marth, G., Abecasis, G., Ruan, J., Li, H., Durbin, R., Homer, N. et al.** (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079. doi:10.1093/bioinformatics/btp352

**Materna, S. C., Ransick, A., Li, E. and Davidson, E. H.** (2013). Diversification of oral and aboral mesodermal regulatory states in pregastrular sea urchin embryos. *Dev. Biol.* **375**, 92-104. doi:10.1016/j.ydbio.2012.11.033

**McCauley, B. S., Wright, E. P., Exner, C., Kitazawa, C. and Hinman, V. F.** (2012). Development of an embryonic skeletogenic mesenchyme lineage in a sea cucumber reveals the trajectory of change for the evolution of novel structures in echinoderms. *Evodevo* **3**, 17. doi:10.1186/2041-9139-3-17

**McGregor, A. P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D. G., Payre, F. and Stern, D. L.** (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* **448**, 587-590. doi:10.1038/nature05988

**Minokawa, T., Rast, J. P., Arenas-Mena, C., Franco, C. B. and Davidson, E. H.** (2004). Expression patterns of four different regulatory genes that function during sea urchin development. *Gene Expr. Patterns* **4**, 449-456. doi:10.1016/j.modgep.2004.01.009

**Mortazavi, A., Thompson, E. C. L., Garcia, S. T., Myers, R. M. and Wold, B.** (2006). Comparative genomics modeling of the NRSF/REST repressor network: from single conserved sites to genome-wide repertoire. *Genome Res.* **16**, 1208-1221. doi:10.1101/gr.4997306

**Nam, J., Dong, P., Tarpine, R., Istrail, S. and Davidson, E. H.** (2010). Functional cis-regulatory genomics for systems biology. *Proc. Natl. Acad. Sci. USA* **107**, 3930-3935. doi:10.1073/pnas.1000147107

**Oliveri, P., Carrick, D. M. and Davidson, E. H.** (2002). A regulatory gene network that directs micromere specification in the sea urchin embryo. *Dev. Biol.* **246**, 209-228. doi:10.1006/dbio.2002.0627

**Oliveri, P., Davidson, E. H. and McClay, D. R.** (2003). Activation of pmar1 controls specification of micromeres in the sea urchin embryo. *Dev. Biol.* **258**, 32-43. doi:10.1016/S0012-1606(03)00108-8

**Oliveri, P., Tu, Q. and Davidson, E. H.** (2008). Global regulatory logic for specification of an embryonic cell lineage. *Proc. Natl. Acad. Sci. USA* **105**, 5955-5962. doi:10.1073/pnas.0711220105

**Pullara, F., Guerrero-Santoro, J., Calero, M., Zhang, Q., Peng, Y., Spåhr, H., Kornberg, G. L., Cusimano, A., Stevenson, H. P., Santamaria-Suarez, H. et al.** (2013). A general path for large-scale solubilization of cellular proteins: from membrane receptors to multiprotein complexes. *Protein Expr. Purif.* **87**, 111-119. doi:10.1016/j.pep.2012.10.007

**Quinlan, A. R. and Hall, I. M.** (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842. doi:10.1093/bioinformatics/btq033

**Rafiq, K., Cheers, M. S. and Ettensohn, C. A.** (2012). The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Development* **139**, 579-590. doi:10.1242/dev.073049

**Rafiq, K., Shashikant, T., McManus, C. J. and Ettensohn, C. A.** (2014). Genome-wide analysis of the skeletogenic gene regulatory network of sea urchins. *Development* **141**, 2542-2542. doi:10.1242/dev.112763

**Rebeiz, M. and Tsiantis, M.** (2017). Enhancer evolution and the origins of morphological novelty. *Curr. Opin. Genet. Dev.* **45**, 115-123. doi:10.1016/j.gde.2017.04.006

**Richardson, W., Kitajima, T., Wilt, F. and Benson, S.** (1989). Expression of an embryonic spicule matrix gene in calcified tissues of adult sea urchins. *Dev. Biol.* **132**, 266-269. doi:10.1016/0012-1606(89)90222-4

**Rizzo, F., Fernandez-Serra, M., Squarzoni, P., Archimandritis, A. and Arnone, M. I.** (2006). Identification and developmental expression of the ets gene family in the sea urchin (Strongylocentrotus purpuratus). *Dev. Biol.* **300**, 35-48.

**Roe, J. L., Park, H. R., Strittmatter, W. J. and Lennarz, W. J.** (1989). Inhibitors of metalloendoproteases block spiculogenesis in sea urchin primary mesenchyme cells. *Exp. Cell Res.* **181**, 542-550. doi:10.1016/0014-4827(89)90110-9

**Rubinstein, M. and de Souza, F. S. J.** (2013). Evolution of transcriptional enhancers and animal diversity. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368**, 20130017. doi:10.1098/rstb.2013.0017

**Sea Urchin Genome Sequencing Consortium** (2006). The Genome of the Sea Urchin. *Science* **314**, 941-952. doi:10.1126/science.1133609

**Sharma, T. and Ettensohn, C. A.** (2010). Activation of the skeletogenic gene regulatory network in the early sea urchin embryo. *Development* **137**, 1149-1157. doi:10.1242/dev.048652

**Shashikant, T., Khor, J. M. and Ettensohn, C. A.** (2018a). From genome to anatomy: the architecture and evolution of the skeletogenic gene regulatory network of sea urchins and other echinoderms. *Genesis* **56**, e23253. doi:10.1002/dvg.23253

**Shashikant, T., Khor, J. M. and Ettensohn, C. A.** (2018b). Global analysis of primary mesenchyme cell cis-regulatory modules by chromatin accessibility profiling. *BMC Genomics* **19**, 1-18. doi:10.1186/s12864-018-4542-z

**Solek, C. M., Oliveri, P., Loza-Coll, M., Schrankel, C. S., Ho, E. C. H., Wang, G. and Rast, J. P.** (2013). An ancient role for Gata-1/2/3 and Scl transcription factor homologs in the development of immunocytes. *Dev. Biol.* **382**, 280-292. doi:10.1016/j.ydbio.2013.06.019

**Sun, Z. and Ettensohn, C. A.** (2017). TGF-β sensu stricto signaling regulates skeletal morphogenesis in the sea urchin embryo. *Dev. Biol.* **421**, 149-160. doi:10.1016/j.ydbio.2016.12.007

**The Gene Ontology Consortium** (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330-D338. doi:10.1093/nar/gky1055

**The Gene Ontology Consortium** (2011). Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25-29. doi:10.1038/75556

**Tu, Q., Cameron, R. A., Worley, K. C., Gibbs, R. A. and Davidson, E. H.** (2012). Gene structure in the sea urchin Strongylocentrotus purpuratus based on transcriptome analysis. *Genome Res.* **22**, 2079-2087. doi:10.1101/gr.139170.112

**Tu, Q., Cameron, R. A. and Davidson, E. H.** (2014). Quantitative developmental transcriptomes of the sea urchin Strongylocentrotus purpuratus. *Dev. Biol.* **385**, 160-167. doi:10.1016/j.ydbio.2013.11.019

**Wang, L., Koppitch, K., Cutting, A., Dong, P., Kudtarkar, P., Zheng, J., Cameron, R. A. and Davidson, E. H.** (2019). Developmental effector gene regulation: Multiplexed strategies for functional analysis. *Dev. Biol.* **445**, 68-79. doi:10.1016/j.ydbio.2018.10.018

**Yu, G., Wang, L.-G. and He, Q.-Y.** (2015). ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383. doi:10.1093/bioinformatics/btv145

**Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W. et al.** (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. doi:10.1186/gb-2008-9-9-r137

# Supplementary information

## Supplementary Materials and Methods

### Sp-Alx1 antibody immunofluorescence staining

Embryos were fixed in 2% paraformaldehyde (PFA) in ASW for 1 hour, rinsed with ASW and permeabilized with 100% methanol at -20°C for 10 minutes. The fixed embryos were washed three times in phosphate-buffered saline (PBS), blocked in 5% goat serum in PBS (5% GS-PBS) overnight at 4°C and again incubated overnight at 4°C in a mixture of primary antibodies (2.5 µg/µL α-Sp-Alx1 diluted 1:1000 in full-strength 6a9 tissue culture supernatant, final concentration of 2.5 ng/µL). The embryos were washed five times in PBS with 0.1% Tween-20 (PBST), once with PBS, and once with 5% GS-PBS (5 mins/wash). They were then incubated for 2 hours at room temperature in a mixture of two secondary antibodies, Alexa 488 goat anti-mouse IgG and IgM (Jackson ImmunoResearch) and Dylight 594 goat anti-rabbit IgG (Jackson ImmunoResearch), both at a final concentration of 1:500 in 5% GS-PBS. They were washed three times in PBST and cell nuclei were stained by incubating embryos in 1 µg/mL Hoechst 33342 in PBS for 10 minutes at room temperature. Stained embryos were washed five times with PBST and once with PBS (5 mins/wash), then mounted on slides in anti-fade solution (DABCO) for examination.

### Recombinant Alx1 (rAlx1) immunoprecipitation

For α-Sp-Alx1 antibody validation, pETDuet-1 expression construct containing recombinant Alx1 was transformed into Rosetta 2 cells (Novagen). Bacterial cells were culture at 37°C and expression was induced at OD595 of 0.6000 with 0.5 mM IPTG. The temperature was lowered to 18°C and cells were allowed to grow for an additional 3 hours. The bacterial culture was then pelleted through centrifugation at 4000 rpm for 15 minutes. The pellet was lysed in RIPA buffer (1X PBS, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with Roche cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, Cat. No. 11836153001). For immunoprecipitation of rAlx1, bacterial lysates were diluted in RIPA buffer and incubated with α-Sp-Alx1 antibody at

Development • Supplementary information

4°C overnight. The mixture was then incubated with Dynabeads Protein A (Invitrogen, Cat. No. 10001D) for at least 2 hours. The beads were washed five times with RIPA buffer and bound protein was eluted with 1X Laemlli loading buffer (Bio-Rad, Cat. No. 1610747).

**Chromatin immunoprecipitation (ChIP)**

Embryos were collected at the mesenchyme blastula stage (24 hpf) and fixed with 1% formaldehyde in ASW for 10 minutes at room temperature. 0.125 M glycine was then added to stop the crosslinking reaction. Embryos were pelleted gently, washed once in fresh ASW with 0.125 M glycine and washed again in fresh ASW. Fixed embryos were pelleted and resuspended in ice-cold Farnham lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40) supplemented with Roche cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, Cat. No. 11836153001). After a 10-minute incubation on ice, embryos were passed through a 25-gauge needle 20 times to rupture cells while keeping nuclei intact. The crude nuclear preparation was pelleted (3,000 RPM for 15 minutes at 4°C) and resuspended in fresh RIPA buffer (1X PBS, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease inhibitor cocktail. After incubating for 15 minutes on ice, the crude nuclear preparation was sonicated using Bioruptor Pico (Diagenode) for 10 minutes (30 seconds ON, 30 seconds OFF) at 4°C. The sonicated chromatin was clarified via centrifugation (14,000 RPM for 15 minutes at 4°C) and the DNA concentration was measured using a NanoDrop 2000 Spectrophotometer (Thermo Scientific). 100 µg of chromatin was pre-cleared using Dynabeads Protein A (Invitrogen, Cat. No. 10001D) for at least 2 hours. The pre-cleared chromatin was subsequently incubated with 5 µg antibody (Sp-Alx1 or normal rabbit IgG at a final concentration of 0.5 µg/µL) overnight at 4°C with rotation. The following day, protein A beads were blocked with 5% BSA-RIPA for 2 hours at 4°C with rotation. After removing the blocking solution, the chromatin-antibody mix was added to the beads and incubated for at least 4 hours at 4°C with rotation. To remove unbound and non-specific chromatin, beads were then washed four times with RIPA buffer, five times with LiCL wash buffer (100 mM Tris pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate) supplemented with protease inhibitor cocktail and once with TE buffer (10 mM Tris pH 7.5, 0.1 mM EDTA). Next, immunoprecipitated chromatin was eluted by incubating beads in 150 µL IP elution buffer

Development • Supplementary information

(1% SDS, 0.1 M $NaHCO_3$) for 1 hours at 65°C. The supernatant was transferred to a fresh tube and incubated at 65°C overnight to reverse-crosslink the bound chromatin. The following day, 3 µL of RNase A (10 mg/mL) was added to the eluted DNA and incubated at 37°C for an hour. Subsequently, 3 µL of Proteinase K (20 mg/mL) was added to the immunoprecipitated DNA and incubated at 55°C for an hour. Finally, the DNA was purified using GeneJET PCR purification kit (ThermoFisher Scientific) and eluted with 50 µL of elution buffer from the kit.

**ChIP-seq bioinformatic pipeline**

Low-quality raw ChIP-seq reads based on positional information (i.e. reads from low-quality areas of the flowcell) were first removed using FilterByTile tool from BBMap package (available at https://sourceforge.net/projects/bbmap/). Next, the ChIP-seq reads were trimmed using Trimmomatic v0.38 to remove leading low quality bases (Bolger et al., 2014). The ChIP-seq reads were then mapped to the *S. pupuratus* genome (v3.1) using Bowtie2 (v2.3.4.3) (very sensitive setting) (Langmead and Salzberg, 2012). The v3.1 genome assembly (available at http://www.echinobase.org/Echinobase/) is 826 Mb in size and contains 32,008 scaffolds with an N50 of 401.6 kb. The alignment rate for both samples were approximately 80% each. Next, Samtools (v1.3) (Li et al., 2009) was to remove redundant reads and poorly aligned reads (i.e. MAPQ score <10). Non-redundant, uniquely-mapping reads were then used for peak detection using MACS2 (v2.1.2) (Zhang et al., 2008) with an mfold of [5,50] and p-value cutoff of 0.005.

**GFP reporter assay**

For mutation of specific Alx1 binding sites, overlap extension PCR was used, two sets of primers were used. Each pair was designed to amplify a different half of the insert and created a product with an overhanging sequence that contained the desired point mutations and an overhang sequence with a unique restriction site. A second round of PCR was carried out to generate a full-length insert, which was then closed into an *EpGFPII* vector. Next, the linear plasmid was digested with the corresponding restriction enzyme and self-ligated. Prior to injection, reporter constructs were linearized and mixed with carrier DNA that was prepared by overnight HindIII digestion of *S. purpuratus* genomic DNA. Injection solutions contained 200 ng/µL linearized plasmid DNA, 500 ng/µL

Development • Supplementary information

carrier DNA, 0.12 M KCl, 20% glycerol, 0.1% Texas Red dextran in DNAse-free, sterile water. Microinjections into fertilized eggs were carried out as described (Cheers and Ettensohn, 2004) embryos were allowed to develop for 48 hpf before imaging using an Olympus BX60 epifluorescence microscope equipped with an Olympus DP71 color CCD camera.

**Electrophoretic mobility shift assay (EMSA)**

The cDNA for *L. variegatus* Alx1 was amplified via PCR to introduce restriction sites and sequences for 6 histidines (6X His-tag) at the N-terminus. Recombinant His-Alx1 was then cloned into pTXB1 vector (New English Biolabs, Cat. No. N6707S) containing an Intein tag at the C-terminus. The double-tagged recombinant Alx1 (His-rAlx1-Intein) was transformed into Rosetta 2 cells (Novagen, Cat. No. 71400). Bacterial cells were cultured at 37°C and induced at $OD_{595}$ of 0.600 with 0.5 mM IPTG. The temperature was lowered to 18°C and cells were allowed to grow for an additional 3 hours. The bacterial cell pellet was lysed in buffer containing 300 mM NaCl, 50 mM Tris pH 6.8, 0.5% Triton X-100, 20 mM Sarcosine, 2 mM 2-mercaptoethanol, and protease inhibitor cocktail. The mixture was sonicated and then cleared through centrifugation. The lysate was diluted to lower the concentration of detergents to a final concentration of 3.3 mM Sarcosine and 0.16% Triton X-100.

Pre-equilibrated His-Select nickel beads (Sigma, Cat. No. H0537) were incubated with the diluted lysate overnight with gentle rocking at 4°C. Protein bound nickel beads were then washed with buffer containing 300 mM NaCl, 50 mM Tris pH 6.8, 0.1% Triton X-100, 2 mM 2-mercaptoethanol, and protease inhibitor cocktail. The protein was eluted with 800 mM of imidazole. The eluted His-rAlx1-Intein was then incubated for 2 hours with pre-equilibrated chitin beads (New England Biolabs, Cat. No. S6651S) on a rocker at 4°C. The mixture was then loaded onto a column and the beads were washed with buffer containing 300 mM NaCl, 50 mM Tris pH 6.8, 0.1% Tris pH 6.8, 0.1% Triton X-100, 2 mM 2-mercaptoethanol, and protease inhibitor cocktail. To elute the protein, the column was sealed and incubated in the wash buffer with 100 mM 2-mercaptoethanol for 40 hours at 4°C. His-Alx1 was then eluted, concentrated and desalted.

Development • Supplementary information

The binding conditions for the gel shift reactions were 75 mM NaCl, 15 mM Tris pH 7.6, 7.5% glycerol, 2 mM MgCl2, 1.5 mM EDTA, 0.1% NP-40, 40 mM DTT, 50 µg BSA and 1 µg poly(dI-dC) per 20 µL reaction. All DNA probes were synthesized, biotinylated (when applicable), and purified (either through gel or HPLC). Wild type double stranded DNA probe with Alx1 binding sites (5'-Biotin- GTCGGGGCGTTAATAGATTTAAACTTTTTC-3') and a mutant double stranded DNA probe (5'-Biotin-GTCGGGGCGTTCGTA GAACGAAACTTTTTC-3') were used. The reactions were incubated with the binding buffer and 200 ng protein on ice for 15 minutes and probes were then added and incubated for an additional 30 minutes at room temperature. The free probes and protein-DNA complexes were separated on an 8% polyacrylamide gel and visualized using the LightShift Chemiluminescent EMSA kit (ThermoScientific, Cat. No. 20148).

Development • Supplementary information

## Supplementary figures

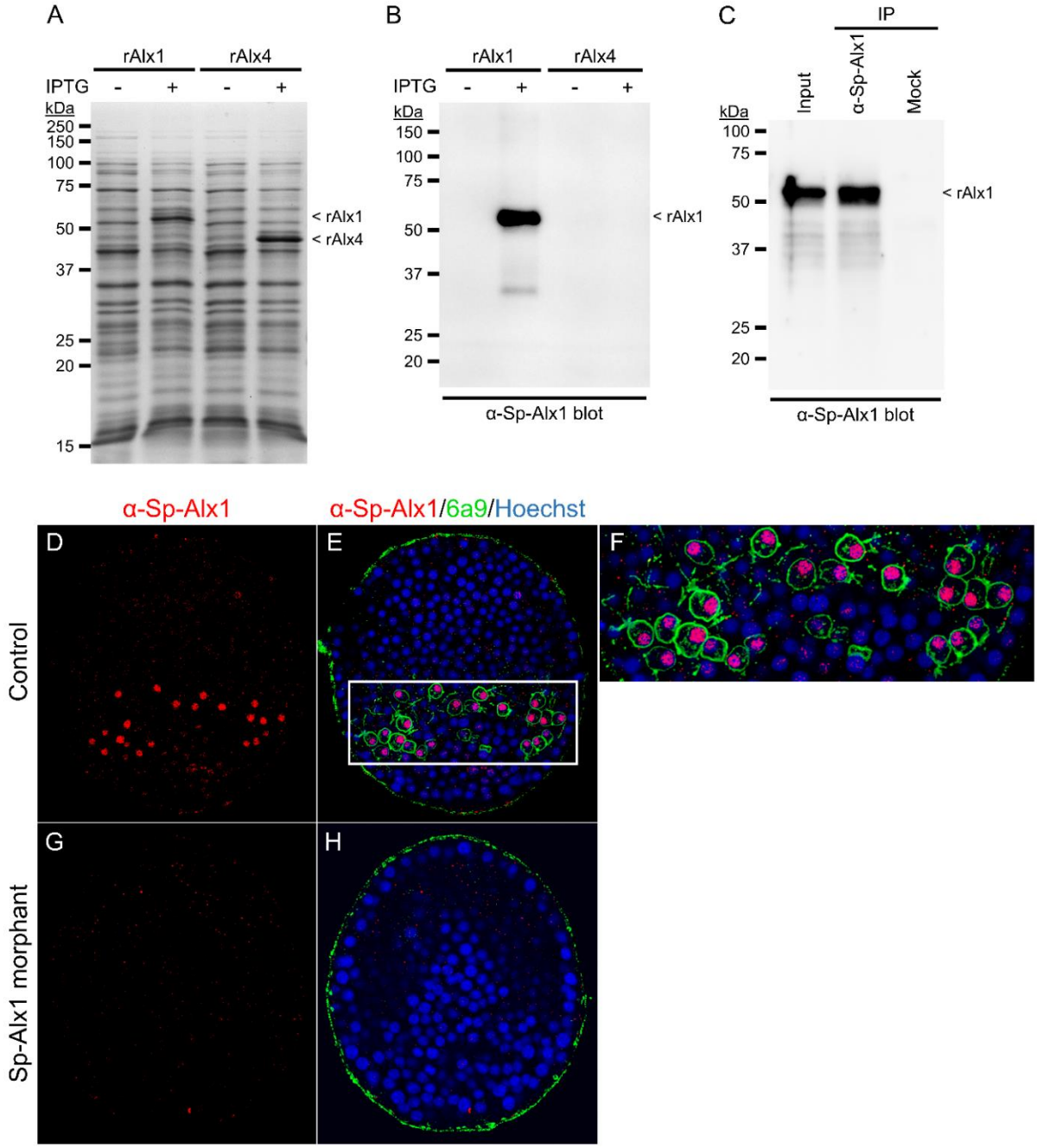



Figure S1: α-Sp-Alx1 polyclonal antibody validation. A custom, affinity-purified rabbit polyclonal antibody was raised against a peptide corresponding to the D2 domain of Sp-Alx1 (Khor and Ettensohn, 2017), the sequence of which is completely conserved in *Lytechinus variegatus* Alx1. (A) Coomassie staining of lysates from bacterial cultures that

Development • Supplementary information

were transformed with IPTG-inducible recombinant *L. variegatus* Alx1 (rAlx1) and Alx4 (rAlx4) expression constructs. Lysates from IPTG-induced cultures showed presence of bands corresponding to the predicted sizes of rAlx1 (~50 kDa) and rAlx4 (~40 kDa). (B) Immunoblot of the same bacterial lysates. The antibody specifically recognized rAlx1 in the induced culture but not the closely related homeodomain protein, rAlx4, which lacks the D2 domain. (C) The α-Sp-Alx1 antibody can effectively immunoprecipitate rAlx1. Bands corresponding to rAlx1 were detected in the input and α-Sp-Alx1 IP eluent but not in the mock (no primary antibody) IP eluent. (D) The nuclei of PMCs are selectively labeled by the α-Sp-Alx1 antibody (red). (E, F) Labelled Sp-Alx1 (red) co-localized with monoclonal antibody 6a9 (green), which recognizes PMC-specific cell surface proteins of the MSP130 family. (G, H) Sp-Alx1 morphant that lacked PMCs showing no observable α-Sp-Alx1 or mAb 6a9 staining.

Development • Supplementary information

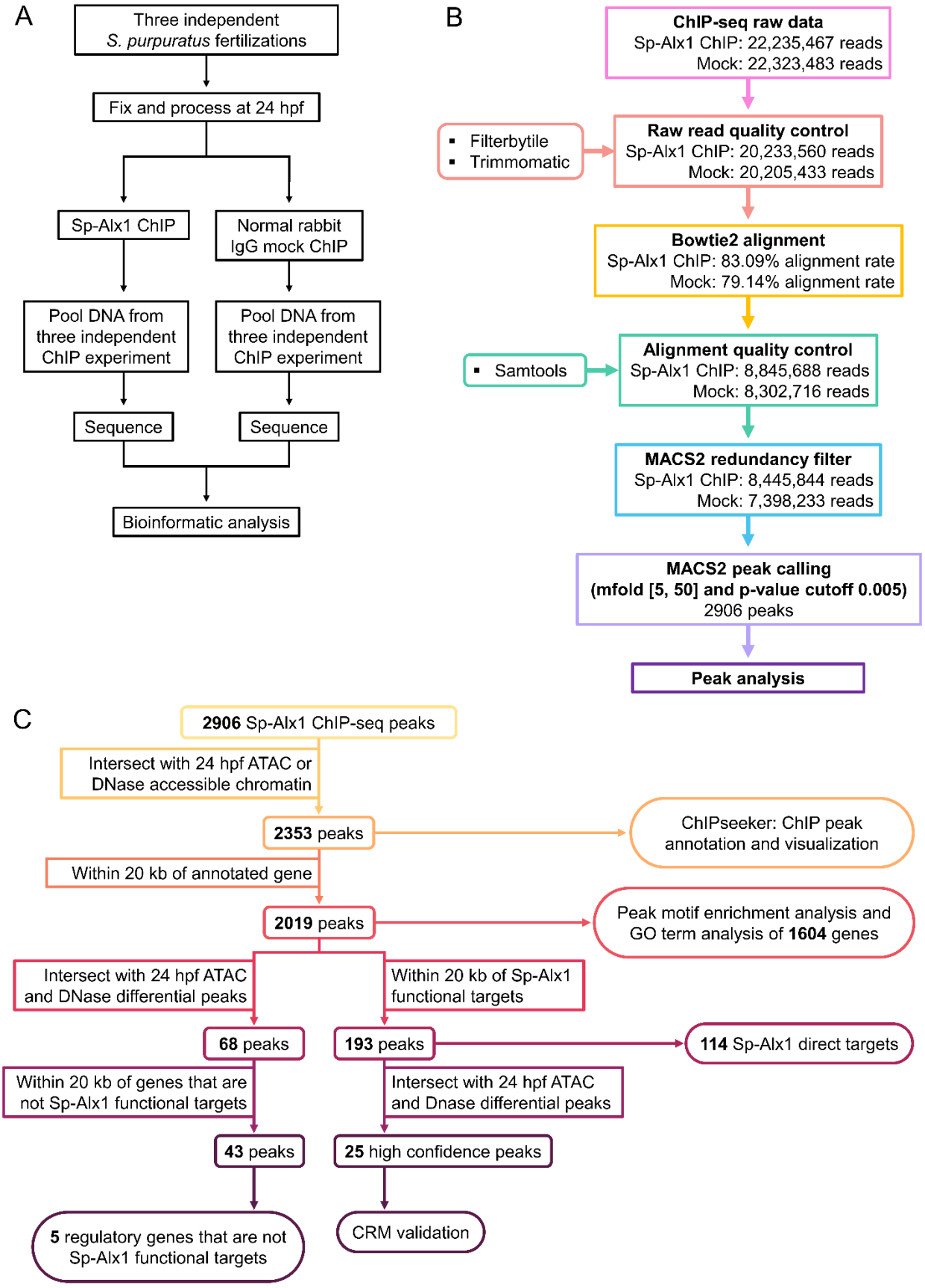

A
Three independent S. purpuratus fertilizations
Fix and process at 24 hpf
Sp-Alx1 ChIP
Normal rabbit IgG mock ChIP
Pool DNA from three independent ChIP experiment
Pool DNA from three independent ChIP experiment
Sequence
Sequence
Bioinformatic analysis
B
ChIP-seq raw data
Sp-Alx1 ChIP: 22,235,467 reads
Mock: 22,323,483 reads
Filterbytile
Trimmomatic
Raw read quality control
Sp-Alx1 ChIP: 20,233,560 reads
Mock: 20,205,433 reads
Bowtie2 alignment
Sp-Alx1 ChIP: 83.09% alignment rate
Mock: 79.14% alignment rate
Samtools
Alignment quality control
Sp-Alx1 ChIP: 8,845,688 reads
Mock: 8,302,716 reads
MACS2 redundancy filter
Sp-Alx1 ChIP: 8,445,844 reads
Mock: 7,398,233 reads
MACS2 peak calling
(mfold [5, 50] and p-value cutoff 0.005)
2906 peaks
Peak analysis
C
2906 Sp-Alx1 ChIP-seq peaks
Intersect with 24 hpf ATAC or DNase accessible chromatin
2353 peaks
ChIPseeker: ChIP peak annotation and visualization
Within 20 kb of annotated gene
2019 peaks
Peak motif enrichment analysis and GO term analysis of 1604 genes
Intersect with 24 hpf ATAC and DNase differential peaks
Within 20 kb of Sp-Alx1 functional targets
68 peaks
193 peaks
114 Sp-Alx1 direct targets
Within 20 kb of genes that are not Sp-Alx1 functional targets
Intersect with 24 hpf ATAC and Dnase differential peaks
43 peaks
25 high confidence peaks
5 regulatory genes that are not Sp-Alx1 functional targets
CRM validation

**Figure S2:** Sp-Alx1 ChIP-seq pipeline (see Materials and Methods and Supplementary Materials and Methods). (A) Flowchart of ChIP-seq protocol. (B) ChIP-seq bioinformatic analysis pipeline. (C) Summary of ChIP-seq peak filtering and analysis.

Development • Supplementary information

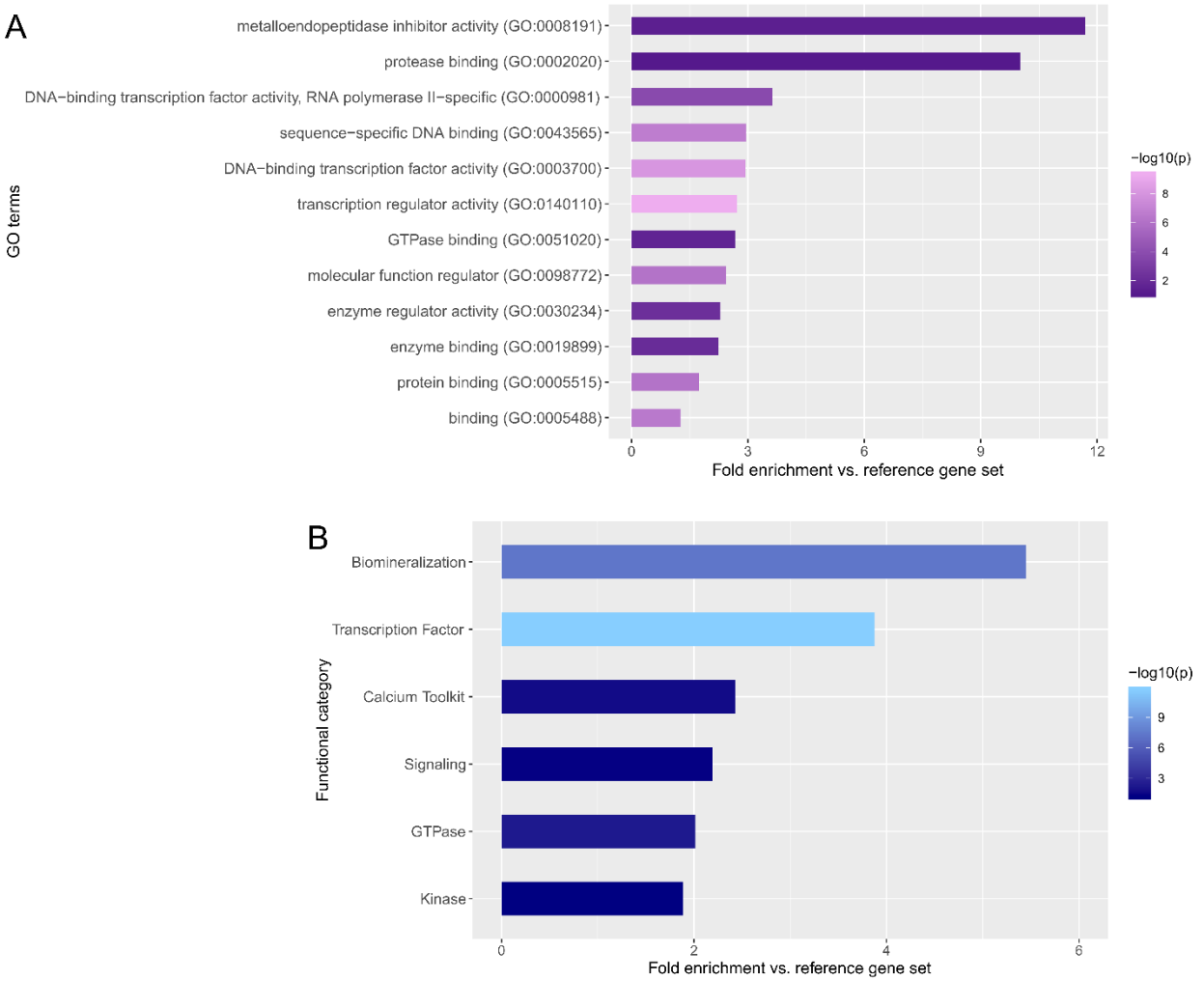



**Figure S3:** Gene ontology (GO) term and functional category enrichment analysis. (A) GO term enrichment analysis of genes within 20 kb of Sp-Alx1 ChIP-seq peaks. (B) Sea urchin-specific functional category enrichment analysis of the same peak set (categories developed by Tu et al., 2014). Intensity of the bars corresponds to the significance of the enrichment, expressed as -$\log_{10}$(p-value).

Development • Supplementary information

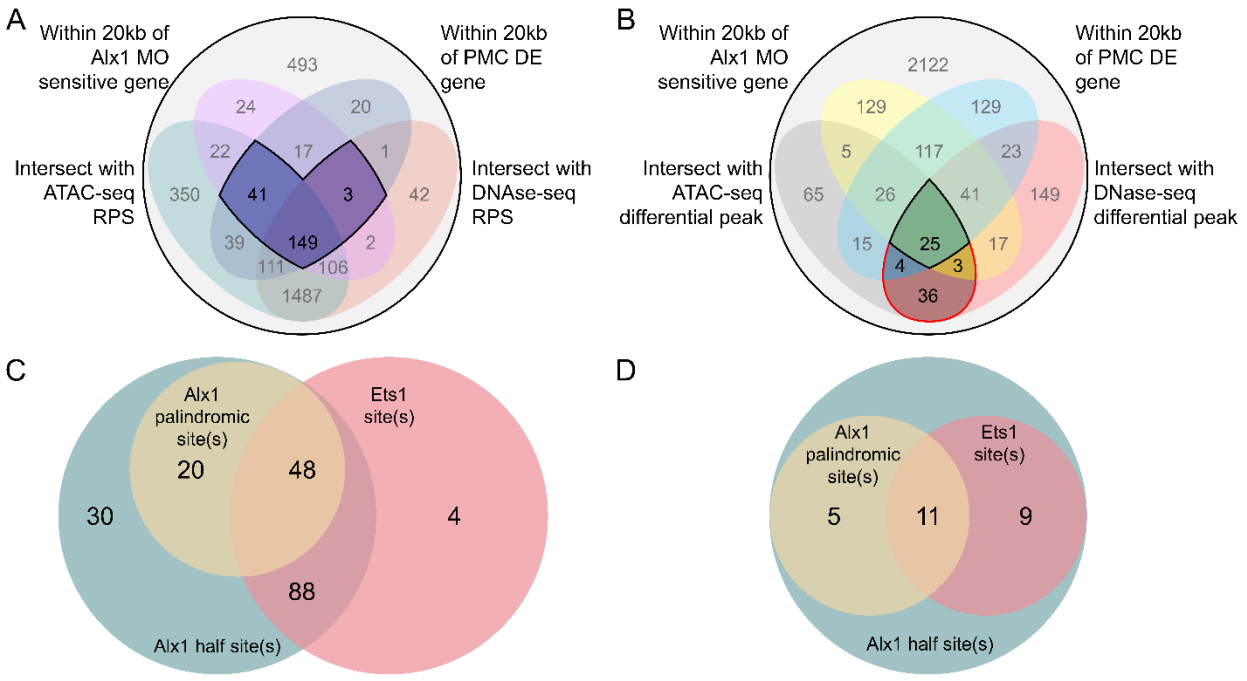


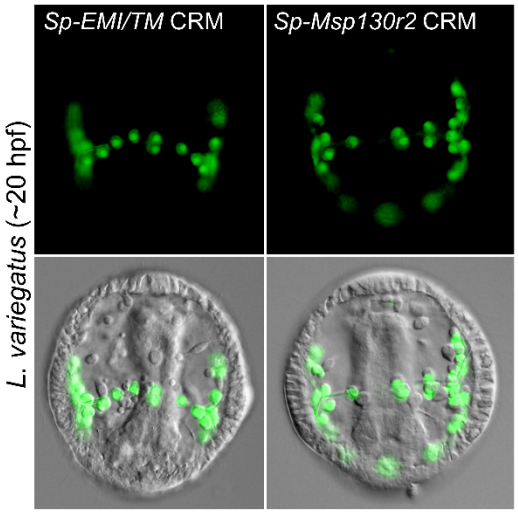

**Figure S4:** Sp-Alx1 ChIP-seq peak filtering. (A) Selection criteria used to identify a set of Sp-Alx1 ChIP-seq peaks near putative Alx1 direct targets (central region outlined in black). (B) Selection criteria used to identify a set of high confidence ChIP-seq peaks (region outlined in black) and ChIP-seq peaks near non-Alx1 functional targets (region outlined in red). (C) Analysis of Alx1 half sites (TAATNN), Alx1 palindromic sites (TAATNNNATTA), and Ets1 sites (AGGAAR) found in the 193 ChIP-seq peaks near Alx1 direct targets. (D) Analysis of Alx1 half sites, Alx1 palindromic sites, and Ets1 sites found in the 25 high-confidence ChIP-seq peaks.

Development • Supplementary information

A

```
Sp-EMI/TM       CTCATGTTTTGACATACAGTCATAAATTAAACCGTTAATGAAAACATTTCAAAACAAACT 60
Lv-EMI/TM       GTCATGTATAACCT--CATAAATGATATTTAAGAAGAAGTCAGCATTTCAAAAACATACT 58
                 ****** *   *   **   ** *  *  *     **   *    **  ****** ***

Sp-EMI/TM       TCCCCTCCAGCATGGCCTTAAACCATCCTTCTCAGAGGAATAAATTCATTTTTCTCCAAA 120
Lv-EMI/TM       TTCCTTTCAGCATTGCC-TCAAACATTTCTCTCAAAGGAAGCTATTCATTTTATTCCAAA 117
                * ** * ****** *** * ** ***   ***** *****   *********  ******

Sp-EMI/TM       ATAGGAATTAGAGATTCCCTTGAATGCTGCGGCAACGC-AGGGGATGTTGGCTAATTGAC 179
Lv-EMI/TM       ACAGGAATTCAATGTTGCTTTGAATGCTGTGGCATCATGTGTGGATGTTGGATAATTGAC 177
                * *******  *  ** * ********** **** *    * ********* ********

Sp-EMI/TM       ATAATTTAGCAATTATGTTGAACTGCCACGGTACAAGT----------ATATATTCTCGT 229
Lv-EMI/TM       ACAATTTAGTACTTATGTTCAACTGCAACCTGCGTTGCAAGTATATTGTCATGTTTACGT 237
                * ******* * ******* ****** **       *             ** **  ***

                                                AM1 Mutation: TCGTAGATCGA
Sp-EMI/TM       GCGAAACAACTAAGTGCAATAAGGCGTTTGTTATTTCGT-CGGGGCGTTAATAGATTTAA 288
Lv-EMI/TM       GCGAAAGAAGCAAGTATCACAAGGCGTTTGTTATTTCTTCTGGGGCGTTAATAGATTTAA 297
                ****** **  ****   * ***************** *  ******************

                                AM2 Mutation: TCGCAAAACGA
Sp-EMI/TM       ACTTTTTCGTCCGCCTGTAAAGTGATAACAAAATTAACAGGTAGACAACATACTCCGAGA 348
Lv-EMI/TM       ACGTTGCTATCCGCTTGTAATGTGATAACCAAAATTACCAGTTAAGCTCAGAGGTTAACC 357
                ** **    ***** ***** ******** *** * **  **  *   ** *     *

Sp-EMI/TM       GACTAACCGTCATTTCAAGAGAATGATACAAGAAGTTATTATCACCATGCACATGTATAA 408
Lv-EMI/TM       ATGTTAACCATCACTAAGGAGAAAATTAGAATTGGTTATTATCGCAGCGAACATGTCTCA 417
                   * * *      * * *****   ** **   ********* *   * ****** * *

Sp-EMI/TM       TTTTAGCTGATAAGTGAAAGAAGAAAGTGATAATTAATGTGGTGTCACTTCCTCGATTTT 468
Lv-EMI/TM       TTCAAACTGATAAA----TGGAAAAAGCGATAATTCTTGTGGTGTGACTTCCTCGATTTT 473
                **  * *******      * * **** *******  ******** ************

Sp-EMI/TM       GACGTTGTGCTTGA--CGGAAATATGGAAGGCATTGGTTGATAAACCAACCACAAC 522
Lv-EMI/TM       GACATTGTGCTTTTGACGGAAATATGAAAGGCATTGGTTATATAAACAAACCAAC- 528
                *** ********    ********** ************    ** *** *  *
```

B




Development • Supplementary information

**Figure S5:** Examples of CRMs that are conserved across >50 million years of evolution (*Sp-EMI/TM* and *Sp-Msp130r2*). (A) Clustal Omega alignment of *Sp-EMI/TM* Sp-Alx1 ChIP-seq peak and *L. variegatus* EMI/TM intronic sequence. The truncated CRM designated P2-1 is highlighted in grey. Red boxes indicate two putative Alx1 palidromic binding sites (AM1 and AM2) that were mutated, only one of which is conserved between the two species. (B) *Sp-EMI/TM* and *Sp-Msp130r2* CRMs injected into *L. variegatus* fertilized eggs were observed to drive PMC-specific GFP expression.

Development • Supplementary information

## Supplementary tables

**Table S1**: Sp-Alx1 ChIP-seq peaks generated by MACS2. Column 1: Peak location in the *S. purpuratus* genome (version 3.1) scaffold. Column 2: Peak start coordinate. Column 3: Peak end coordinate. Column 4: Length of peak region. Column 5: Absolute peak summit position. Column 6: Pileup height at peak summit. Column 7: $-\log_{10}$(p-value) for peak summit. Column 8: Fold enrichment for peak summit against random Poisson distribution with local lambda. Column 9: $-\log_{10}$(q-value) at peak summit. Column 10: Designated name for peak.

Click here to Download Table S1

**Table S2**: Alx1 direct targets and their corresponding peaks. Column 1: Name of the closest gene. Column 2: WHL model ID for corresponding gene transcript. Column 3: SPU gene ID. Column 4: Sea urchin-specific functional category for the corresponding gene. Column 5: Additional detail for functional category. Column 6: Peak name for Sp-Alx1 ChIP-seq peak. Column 7: Peak location in the *S. purpuratus* genome (version 3.1) scaffold. Column 8: Peak start coordinate. Column 9: Peak end coordinate. Column 10: Peak distance from annotated gene (bp); '0' represents overlap with gene body. Column 11: Presence of Alx1 half site(s) (TAATNN) in the corresponding peak. Column 12: Presence of Alx1 palindromic site(s) (TAATNNNATTA) in the corresponding peak. Column 13: Presence of Ets1 site(s) (AGGAAR) in the corresponding peak.

Click here to Download Table S2

Development • Supplementary information

**Table S3**: Sp-Alx1 peaks for regulatory genes downstream of Alx1 in the PMC GRN. Column 1: Name of the closest regulatory gene. Column 2: WHL model ID for the corresponding gene transcript. Column 3: SPU gene ID. Column 4: Peak name for Sp-Alx1 ChIP-seq peak. Column 5: Peak location relative to closest gene. Column 6: Forward primer used to clone the corresponding CRM. Column 7: Reverse primer used to clone the corresponding CRM. Column 8: Size of the insert cloned.

Click here to Download Table S3

**Table S4**: High-confidence Sp-Alx1 ChIP-seq peaks and their corresponding closest genes. Column 1: Name of the closest gene. Column 2: WHL model ID for the corresponding gene transcript. Column 3: SPU gene ID. Column 4: Sea urchin-specific functional category for the corresponding gene. Column 5: Peak name for Sp-Alx1 ChIP-seq peak. Column 6: Peak location relative to the closest gene. Column 7: Forward primer used to clone the corresponding CRM. Column 8: Reverse primer used to clone the corresponding CRM. Column 9: Size of the CRM insert. Column 10: Presence of Alx1 half site(s) (TAATNN) in the corresponding peak. Column 11: Presence of Alx1 palindromic site(s) (TAATNNNATTA) in the corresponding peak. Column 12: Presence of Ets1 site(s) (AGGAAR) in the corresponding peak.

Click here to Download Table S4

Development • Supplementary information