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

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

(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

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.

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).

Supplementary figures

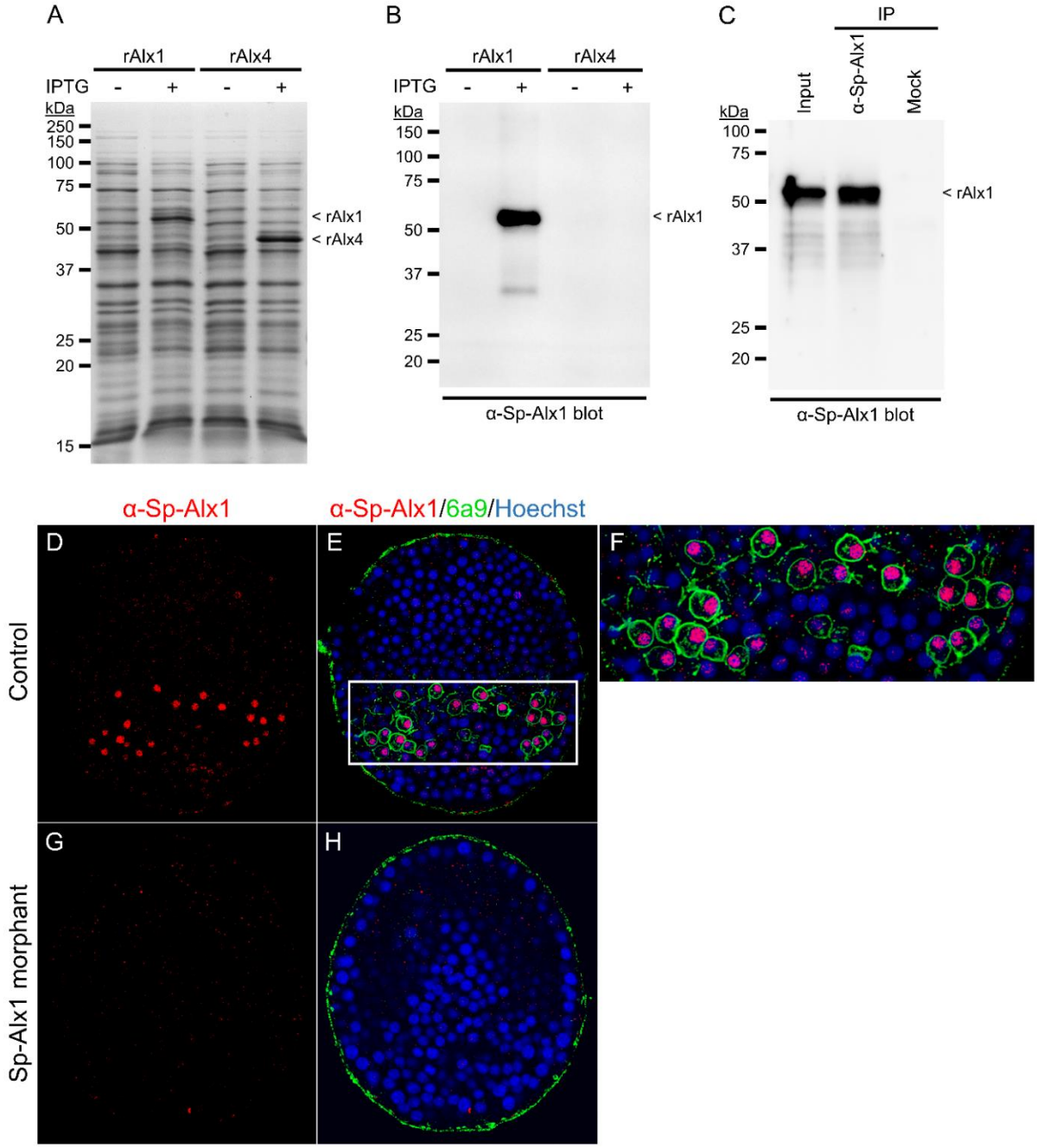

Figure S1: α-Sp-Alx1 polyclonal antibody validation. A custom, affinity-purified rabbit polyclonal antibody was raised against a peptide corresponding to the D2 domain of Sp-Alx1 (Khor and Ettensohn, 2017), the sequence of which is completely conserved in *Lytechinus variegatus* Alx1. (A) Coomassie staining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

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.

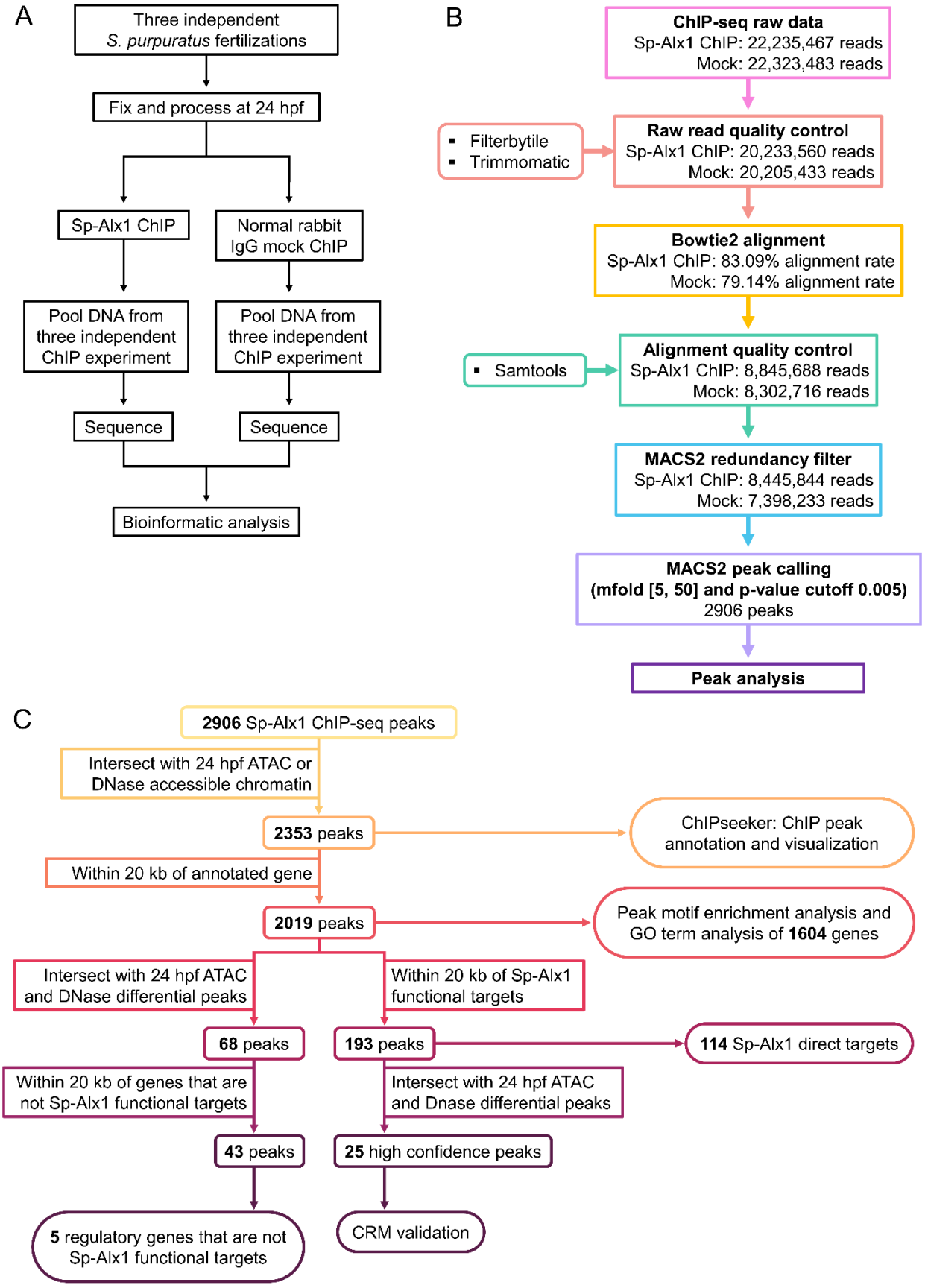
A
Three independent S. purpuratus fertilizations
Fix and process at 24 hpf
Sp-Alx1 ChIP
Normal rabbit IgG mock ChIP
Pool DNA from three independent ChIP experiment
Pool DNA from three independent ChIP experiment
Sequence
Sequence
Bioinformatic analysis
B
ChIP-seq raw data
Sp-Alx1 ChIP: 22,235,467 reads
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.

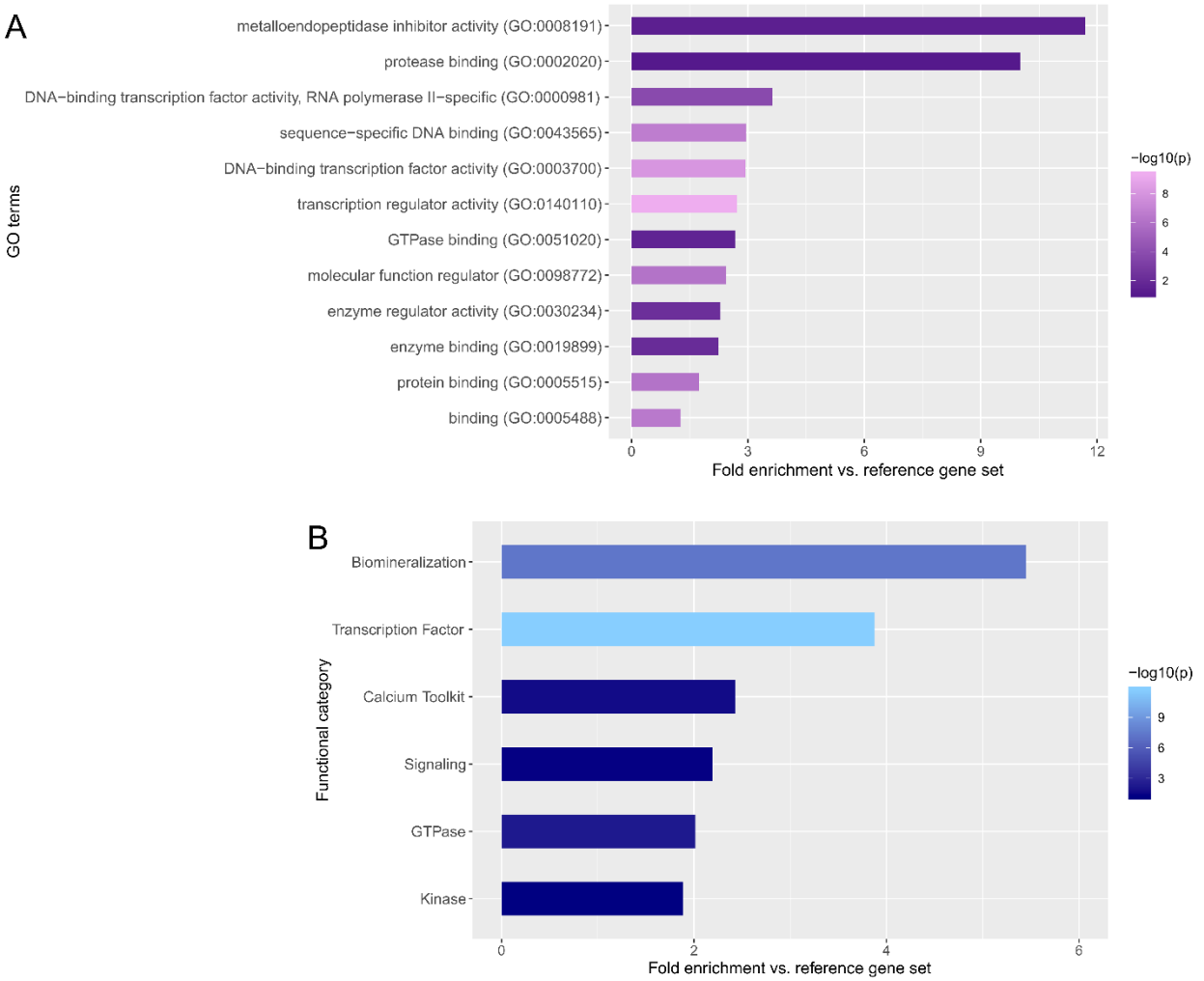

Figure S3: Gene ontology (GO) term and functional category enrichment analysis. (A) GO term enrichment analysis of genes within 20 kb of Sp-Alx1 ChIP-seq peaks. (B) Sea urchin-specific functional category enrichment analysis of the same peak set (categories developed by Tu et al., 2014). Intensity of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

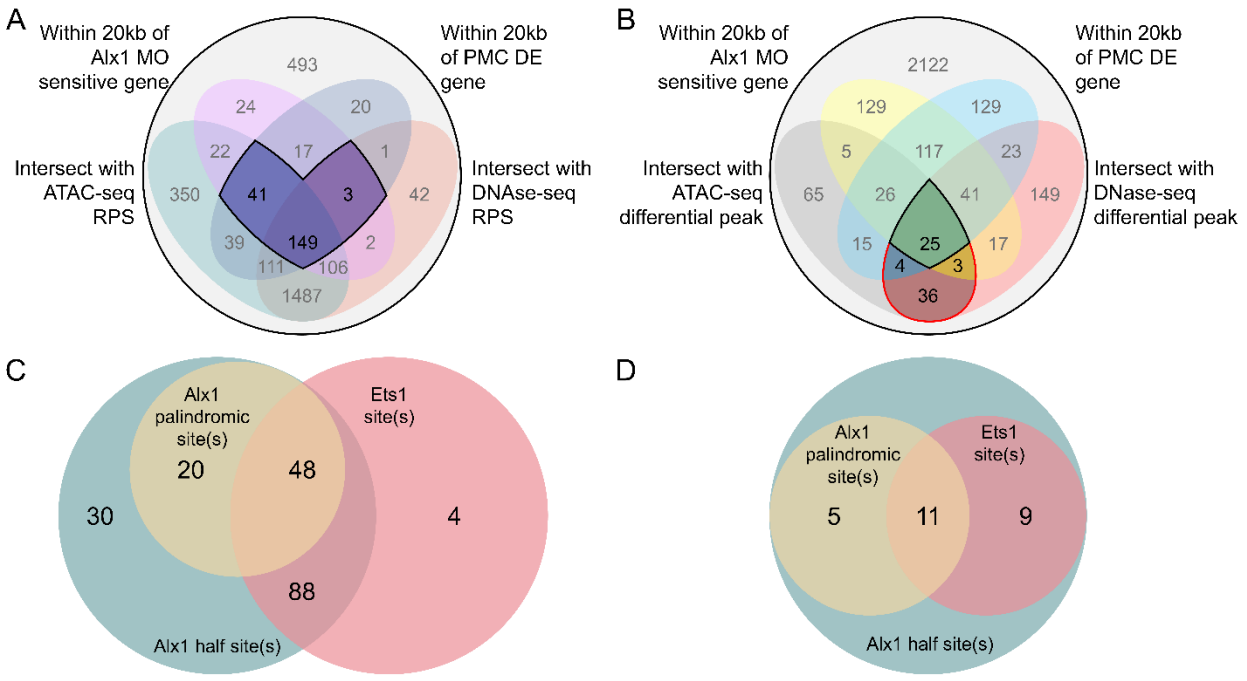

Figure S4: Sp-Alx1 ChIP-seq peak filtering. (A) Selection criteria used to identify a set of Sp-Alx1 ChIP-seq peaks near putative Alx1 direct targets (central region outlined in black). (B) Selection criteria used to identify a set of high confidence ChIP-seq peaks (region outlined in black) and ChIP-seq 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.

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

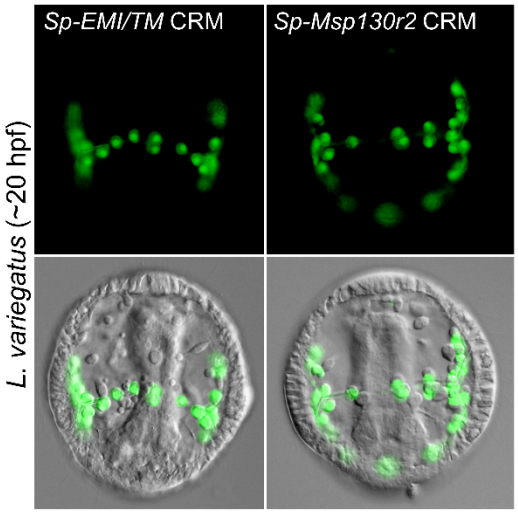

Figure S5: Examples of CRMs that are conserved across >50 million years of evolution (*Sp-EMI/TM* and *Sp-Msp130r2*). (A) Clustal Omega alignment of *Sp-EMI/TM* Sp-Alx1 ChIP-seq peak and *L. variegatus* EMI/TM intronic sequence. The truncated CRM designated P2-1 is highlighted in grey. Red boxes indicate 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.

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

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