nature ecology & evolution

Article

https://doi.org/10.1038/s41559-022-01906-9

Recent reconfiguration of an ancient developmental gene regulatory network in *Heliocidaris* sea urchins

Received: 20 January 2022

Accepted: 12 September 2022

Check for updates

Phillip L. Davidson[1], Haobing Guo[2,3], Jane S. Swart[1], Abdull J. Massri[1], Allison Edgar[1], Lingyu Wang[1], Alejandro Berrio[1], Hannah R. Devens[1], Demian Koop[4], Paula Cisternas[4], He Zhang[3], Yaolei Zhang[2,5], Maria Byrne[4], Guangyi Fan[2,3] and Gregory A. Wray[1,6]

Changes in developmental gene regulatory networks (dGRNs) underlie much of the diversity of life, but the evolutionary mechanisms that operate on regulatory interactions remain poorly understood. Closely related species with extreme phenotypic divergence provide a valuable window into the genetic and molecular basis for changes in dGRNs and their relationship to adaptive changes in organismal traits. Here we analyse genomes, epigenomes and transcriptomes during early development in two *Heliocidaris* sea urchin species that exhibit highly divergent life histories and in an outgroup species. Positive selection and chromatin accessibility modifications within putative regulatory elements are enriched on the branch leading to the derived life history, particularly near dGRN genes. Single-cell transcriptomes reveal a dramatic delay in cell fate specification in the derived state, which also has far fewer open chromatin regions, especially near conserved cell fate specification genes. Experimentally perturbing key transcription factors reveals profound evolutionary changes to early embryonic patterning events, disrupting regulatory interactions previously conserved for ~225 million years. These results demonstrate that natural selection can rapidly reshape developmental gene expression on a broad scale when selective regimes abruptly change. More broadly, even highly conserved dGRNs and patterning mechanisms in the early embryo remain evolvable under appropriate ecological circumstances.

The well-defined developmental gene regulatory network (dGRN) of early development in sea urchins[1,2] provides a powerful framework for investigating the evolution of embryonic patterning mechanisms. Interactions between genes encoding transcription factors and their target genes within this dGRN are almost completely conserved among species that diverged ~30–40 million years (my) ago[3,4], with some interactions conserved for ~225 my (ref. [5]), ~275 my (ref. [6]) or even ~480 my (refs. [7,8]). One possible explanation for this observation is developmental constraints, such that early developmental processes are largely immutable given their critical roles in body plan organization and tissue

[1]Department of Biology, Duke University, Durham, NC, USA. [2]BGI-Qingdao, BGI-Shenzhen, Qingdao, China. [3]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen, China. [4]School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia. [5]Qingdao-Europe Advanced Institute for Life Sciences, BGI-Qingdao, BGI-Shenzhen, Qingdao, China. [6]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. e-mail: gwray@duke.edu

specification[9,10]. Under this scenario, any change in a critical interaction during early development would have widespread effects on later processes, which would almost always be deleterious. Still, an important confound remains untested: the species with deeply conserved developmental mechanisms all share the same life history mode, involving low maternal provisioning and an extended feeding larval phase. Species with derived life histories involving massive maternal provisioning and highly abbreviated, non-feeding pre-metamorphic development have evolved on multiple occasions within sea urchins[11–13], possibly in response to lower or more unpredictable food availability[14]. These species can reveal how conserved regulatory interactions and patterning mechanisms respond to major shifts in selective regimes.

The Australian sea urchin genus *Heliocidaris* includes two recently diverged species: *H. tuberculata*, representing the ancestral life history, and *H. erythrogramma*, the derived state[15] (Fig. 1a). The shift to non-feeding development radically alters natural selection on development: with feeding no longer necessary, high mortality rates in the plankton[16] impose strong selection to decrease time to metamorphosis[12]. Numerous anatomical features and gene expression profiles of early development that are broadly conserved among sea urchins differ markedly between these closely related species[17–20](Fig. 1b). In this Article, we sought to learn whether these recently evolved differences are merely superficial and mask deeply conserved developmental mechanisms, or whether they are the product of substantive evolutionary changes in early cell fate specification and dGRN organization. Evidence for the former would suggest that developmental constraints play an important role in limiting genetic and regulatory composition of the ancestral gene regulatory network (GRN), whereas support for the latter would point to a flexible morphogenetic system derived from an embryonic programme conserved at least in part by stabilizing selection that is adaptable to alternative developmental life histories.

Results

Natural selection has sculpted the regulatory landscape

We took advantage of the recent (~5 my) divergence between the two *Heliocidaris* species to carry out detailed analyses of orthologous coding and non-coding regions of the genome, focusing on the transcription factors and regulatory elements that constitute the backbone of the dGRN and underlie cell fate specification mechanisms (for list of 192 dGRN genes, see Supplementary Data 1). Genomes of *H. erythrogramma* and *H. tuberculata* were each sequenced, assembled into 21 full-length chromosomes and annotated (Fig. 1c). Genome sequences were then aligned to one another and to that of *Lytechinus variegatus*[21], an outgroup representing the ancestral life history condition (Fig. 1a).

To understand how natural selection altered the genomes of *Heliocidaris* during the evolution of non-feeding development, we began by testing for evidence of branch-specific positive selection within single-copy protein coding regions[22]. At a genome-wide scale, we found statistical support for modest enrichment of positive selection along the *H. erythrogramma* branch, but not the *H. tuberculata* branch, when considering the full set of genes (Fisher exact test, two-sided: $P < 1.33 \times 10^{-3}$). Of note, coding sequences of dGRN genes showed no enrichment of positive selection on either branch (Fig. 1d,e). This result provides little support for the idea that changes in transcription factor structure and function are primarily responsible for the extensive modifications in development and life history in *H. erythrogramma*. Scant evidence of positive selection in the coding sequences of dGRN genes probably reflects pleiotropic constraints imposed by the multiple functions that their encoded transcription factors execute during cell type specification and differentiation.

Therefore, we hypothesized that functional changes in regulatory elements are instead largely responsible for these trait changes. We carried out Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) on the two *Heliocidaris* species and *L. variegatus* to identify open chromatin regions (OCRs) representing putative regulatory elements in early (hatched) blastula stage embryos, by which time initial cell fates have been specified in the ancestral dGRN. Our analyses are based on all OCRs present in at least one species and located within a genomic region with 1:1:1 orthology among all three species. We then tested for branch-specific positive selection within these OCRs using an approach analogous to that described above for protein-coding regions. At a global scale, these putative enhancer and promoter regions are enriched for evidence of positive selection on the branch leading to *H. erythrogramma* (545) relative to *H. tuberculata* (347) (Fisher exact test, two-sided: $P < 1.33 \times 10^{-11}$) (Fig. 1f,g). This higher incidence in signatures of positive selection specifically within OCRs on the *H. erythrogramma* branch is indicative of positive selection on regulatory element function that is remarkably widespread within its genome and is consistent with our earlier finding that many expression differences between the two *Heliocidaris* species are genetically based in *cis*[23].

Strikingly, signals of *H. erythrogramma*-specific positive selection are even more enriched when considering only OCRs near dGRN genes (Fig. 1f,g; difference in median zeta: 0.182; Fisher exact test, two-sided: $P < 5.33 \times 10^{-5}$). In all, 26 putative regulatory elements located near 23 distinct dGRN genes exhibit evidence of positive selection on the *H. erythrogramma* branch, as opposed to just 4 on the *H. tuberculata* branch (Fig. 1c, g). These 23 genes represent 17.0% of the total within the defined dGRN with a nearby OCR, a marked enrichment compared with the remainder of the genome, where positive selection is detected in OCRs near just 5.7% of genes (Fisher exact test, two-sided: $P < 4.92 \times 10^{-4}$; Fig. 2c).

Two regulatory mechanisms underlie transcriptomic divergence

While the accessibility of most OCRs and expression[19,23] of most genes are conserved between species (for examples, see Fig. 2b and Extended Data Fig. 1), we observed a striking decrease in chromatin accessibility of many putative regulatory elements throughout the *H. erythrogramma* genome relative to both species representing the ancestral life history (Fig. 2a). Of 2,625 orthologous, differentially accessible OCRs between developmental modes, 1,795 sites (68.4%) are significantly less accessible in *H. erythrogramma* (for example, Fig. 2b; *hesC*). As decreased chromatin accessibility can limit transcription factor access to regulatory elements and because most regulatory interactions in the early sea urchin embryo involve activation of transcription[24], widespread evolutionary reduction in chromatin accessibility throughout the genome in *H. erythrogramma* embryos suggests an important role for evolutionary changes in chromatin configuration for divergence in gene expression, in this case associated with generally decreased or delayed zygotic transcription for many genes. This interpretation is consistent with indications of a broad delay in embryonic cell fate specification in this species[19,25–29].

In a previous study[19] we analysed changes in temporal gene expression profiles during early development within *Heliocidaris* and found that the largest changes are concentrated on the branch leading to *H. erythrogramma* and are enriched for developmental regulatory genes generally and dGRN genes specifically. Results reported here suggest that these derived expression profiles are the product of two distinct molecular mechanisms that alter transcription factor binding: changes in nucleotide sequence and changes in chromatin configuration (Fig. 2c,d). The former may alter protein–DNA binding, while the latter may alter protein access to regulatory elements. Both modes of regulatory evolution are concentrated near dGRN genes relative to the rest of the transcriptome in *H. erythrogramma* (Fig. 2c). Notably, accelerated sequence evolution or altered chromatin state (or both) is present in an OCR near a differentially expressed dGRN gene approximately three times more frequently than the rest of the transcriptome in *H. erythrogramma*, while no such relationship is evident in *H. tuberculata*.

https://doi.org/10.1038/s41559-022-01906-9

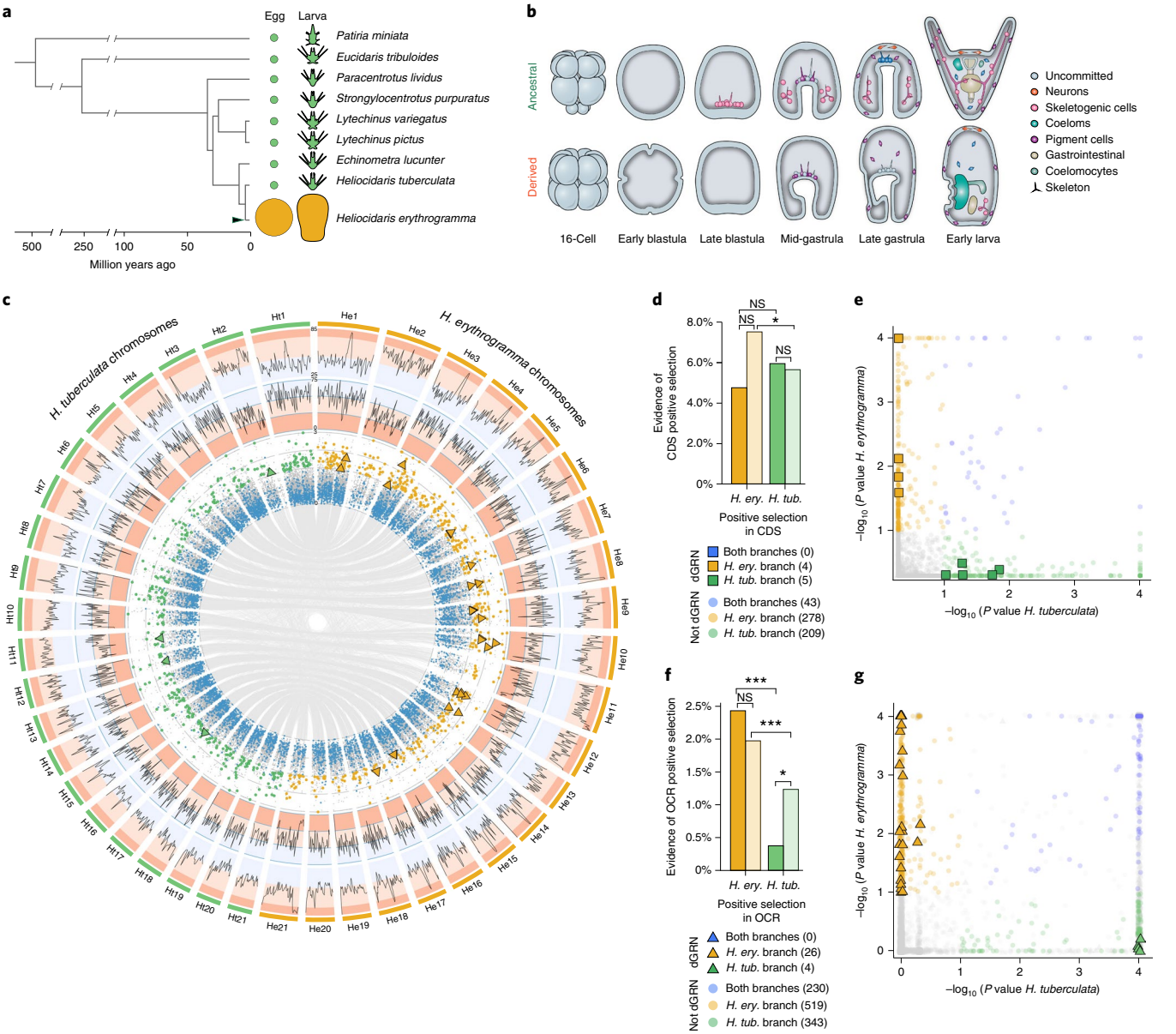


Fig. 1 | Evolution of life history and genomes. a, Feeding larval development (planktotrophy: green) represents the ancient and ancestral life history in sea urchins (sea star *Patiria miniata* represents the outgroup)[11]. Non-feeding larval development (lecithotrophy: orange) has evolved on multiple occasions, including recently within the genus *Heliocidaris* (arrowhead)[9]. **b**, Evolution of non-feeding development in *H. erythrogramma* (bottom) included dramatic modifications to otherwise broadly conserved developmental mechanisms, including changes in cleavage geometry, cell fate specification and morphogenesis[16,18]. **c**, Chromosome-scale genome assemblies of *H. tuberculata* (green) and *H. erythrogramma* (orange). Outer ring: repetitive element content; middle ring: gene content, inner ring: zeta values in OCRs from the selection analyses. Coloured points indicate statistically elevated zeta values (indicative of positive selection) within a single OCR on the branch leading to *H. tuberculata* (green) or *H. erythrogramma* (orange). Blue points indicate highly conserved OCRs (top 10% of phastCon scores). Triangles denote OCRs with signature of branch-specific positive selection located near dGRN genes. Synteny lines between chromosomes denote locations of 1–1 orthologous OCRs between *Heliocidaris* species. **d**, Signatures of positive selection in protein CDS of 84 dGRN and 3,832 non-dGRN single-copy orthologues. Evidence of selection is slightly enriched on the *H. erythrogramma* branch, but dGRN genes show no difference between branches. **e**, *P* values of likelihood ratio test for positive selection in CDS on the branch leading to each species (colour indicates significant *P* values; squares indicate dGRN genes). NS, not significant. **f**, Signatures of positive selection within single-copy OCRs near dGRN ($n = 1{,}069$) and non-dGRN ($n = 26{,}253$) genes are overall much higher on the *H. erythrogramma* branch (a single gene can have multiple OCRs associated with it). For OCRs near dGRN genes, this difference is notably amplified: signatures of positive selection are depleted relative to non-dGRN genes on the *H. tuberculata* branch but substantially elevated on the *H. erythrogramma* branch. **g**, *P* values of likelihood ratio test for positive selection in OCRs on the branch leading to each species. Fisher exact test, two-sided: *P value $< 5 \times 10^{-2}$; ***P value $< 5 \times 10^{-4}$. CDS, coding sequence.

The distribution of both mechanisms of regulatory evolution is highly non-random within the genome (enriched near differentially expressed genes and near developmental regulatory genes) and phylogenetically (enriched on the *H. erythrogramma* branch). These departures from the null expectation of random distribution (that is, resulting from genetic drift) suggest that many of the specific changes are adaptive. Adults of the two *Heliocidaris* species occupy overlapping habitats and ranges[30], making the suite of derived life history traits that evolved on the *H. erythrogramma* branch the most plausible driver for many of the extensive gene regulatory changes.

https://doi.org/10.1038/s41559-022-01906-9

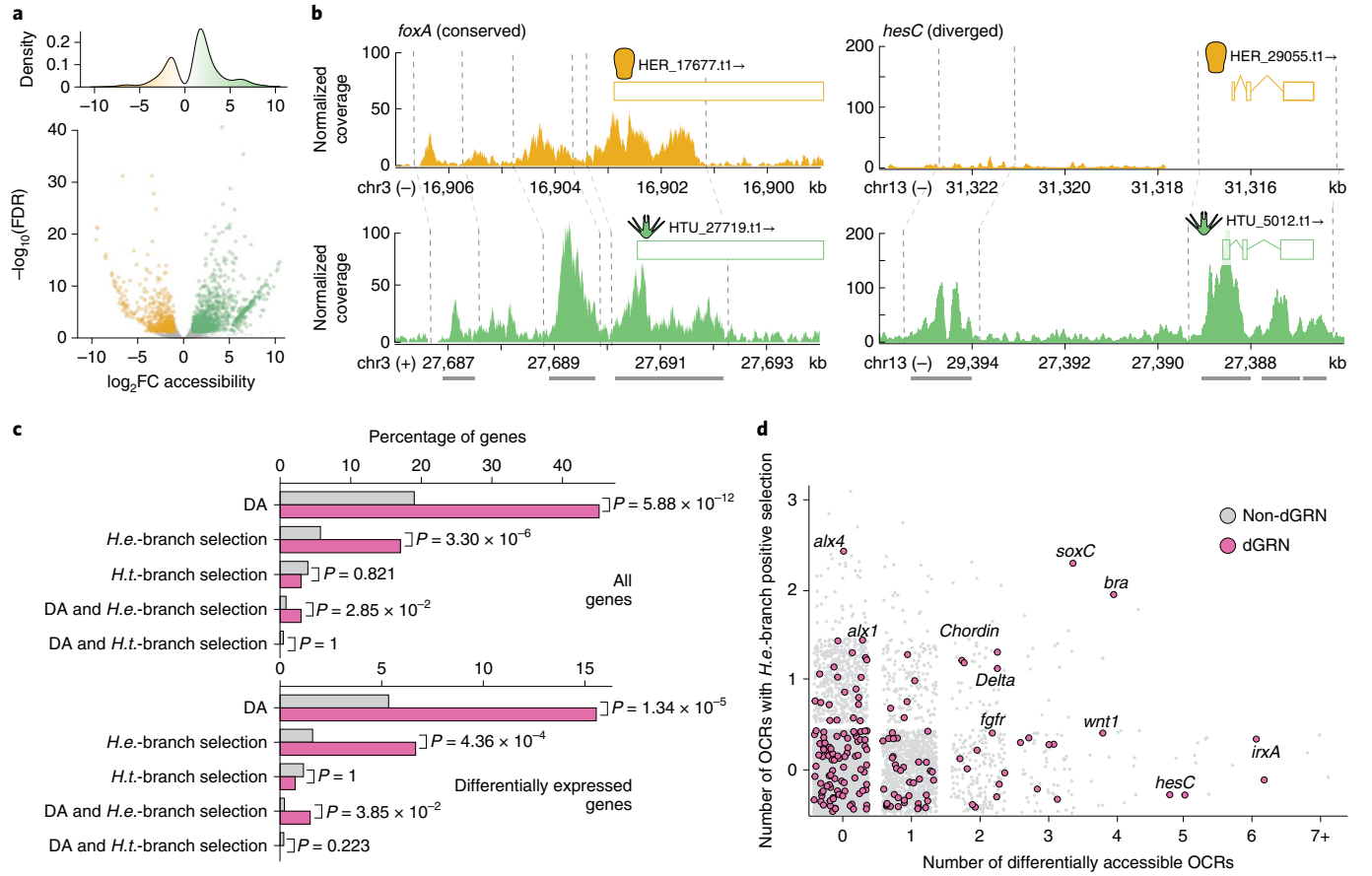


Fig. 2 | Evolution of open chromatin landscape. a, Density and volcano plots of significantly differentially accessible orthologous OCRs between developmental life histories (green: more open in *H. tuberculata* and *L. variegatus*; orange: more open in *H. erythrogramma*). **b**, Examples of conservation (*foxA*) and change (*hesC*) in chromatin accessibility landscape near dGRN genes. For an additional example of a conserved chromatin landscape nearby *foxN2/3* and corresponding expression domains at blastula stage of both *Heliocidaris* species, see Extended Data Fig. 1. **c**, Relationship between chromatin, positive selection and gene expression. Percentage of all genes (top) and differentially expressed genes (bottom) with at least one OCR nearby that is differentially accessible (DA), has evidence of positive selection on the *H. erythrogramma* (*H.e.*) or *H. tuberculata* (*H.t.*) branch, or is both DA and has evidence of positive selection on the *H.e.* or *H.t.* branch. All genes, $n = 8{,}483$; dGRN genes, $n = 135$. Fisher exact test, two-sided. **d**, Number of differentially accessible OCRs and OCRs with evidence of positive selection in *H. erythrogramma* for a given gene.

Early cell fate specification is delayed in *H. erythrogramma*

To understand how changes in the regulation of gene expression in the *H. erythrogramma* embryo influenced developmental mechanisms and life history traits, we leveraged information about the ancestral dGRN to examine embryonic cell fate specification. The earliest zygotic patterning event in the ancestral state involves specification of skeletogenic and germ cell fates following two successive unequal cleavages of vegetal blastomeres[31]. We focus here on the well-characterized skeletogenic cell lineage, which rapidly establishes a distinct transcriptional state[32] and, within 24 h after fertilization, undergoes an epithelial-to-mesenchymal transition (EMT), fully differentiates and begins to synthesize a complex larval endoskeleton (Fig. 1b). Specification and maintenance of the skeletogenic cell fate is regulated by interactions between ~11 transcription factors[33]. These developmental events and most of the underlying dGRN interactions are conserved across >225 my of sea urchin evolution[6].

Morphological development of the skeletogenic cell lineage in *H. erythrogramma* differs in several regards from this ancestral state: cleavage divisions are all equal, no cells undergo EMT before gastrulation, and the larval skeleton is delayed and reduced[20] (Fig. 1b). To understand whether underlying developmental mechanisms are conserved despite these overt morphological differences, we carried out single-cell RNA sequencing (scRNA-seq) of early blastula stage *H. erythrogramma* embryos and compared the results with our published scRNA-seq data from *L. variegatus*[34] at the same early blastula stage (before EMT). We chose this stage because many major cell fates have been specified by early blastula in the ancestral condition[2,31–33]. This result is clearly reflected in the uniform manifold approximation and projection (UMAP) of *L. variegatus*, which contains seven cell clusters (Fig. 3a), each expressing a distinct suite of regulatory proteins predicted by the dGRN, with skeletogenic cells exhibiting a particularly disparate transcriptional state (Fig. 3c and Supplementary Fig. 1). These indications of early cell fate specification and rapid divergence in transcriptional states are also apparent in scRNA-seq data from *Strongylocentrotus purpuratus*[35], another sea urchin representing the ancestral condition (Fig. 1a), suggesting early embryonic regulatory interactions are conserved among planktotrophic species and detectable by scRNA-seq.

In *H. erythrogramma*, only three cell clusters are apparent at the same resolution and developmental stage (Fig. 3b). The observation of fewer distinct transcriptional states in *H. erythrogramma* along with less localized expression of known GRN tissue marker genes (Supplementary Fig. 1) suggests a delayed establishment of distinct transcriptional states in the early embryo of this species—a conclusion not attributable technical factors such as analytical parameterization (Methods), cell number (Extended Data Fig. 2) or genes/unique

https://doi.org/10.1038/s41559-022-01906-9

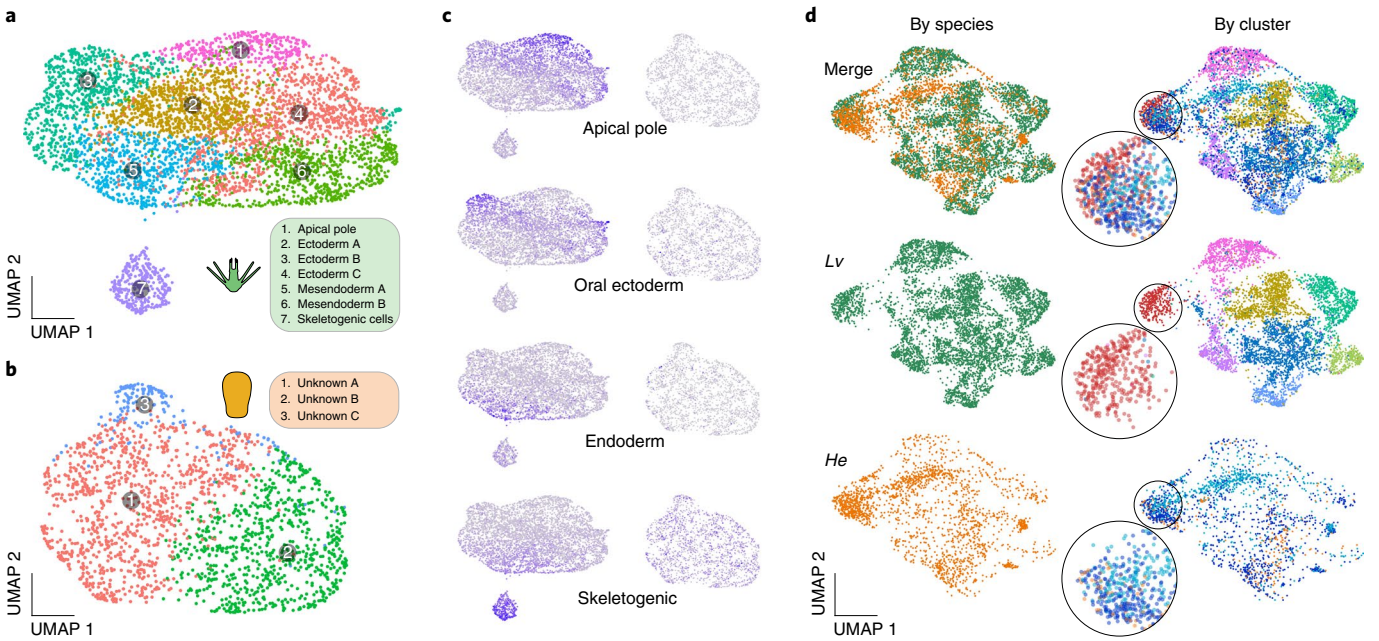


Fig. 3 | Evolution of transcriptomes. a,**b**, scRNA-seq of pre-mesenchyme blastula stage embryos from *L. variegatus* (**a**) and *H. erythrogramma* (**b**) with cells plotted into a UMAP space before integration of expression data across species. Colours indicate inferred cell lineages based on expression of marker genes (Supplementary Fig. 1). Note that *H. erythrogramma* shows fewer distinct transcriptional states than *L. variegatus*. **c**, Composite expression scores for four embryonic territories based on combined expression of multiple known marker genes in the ancestral GRN (Supplementary Data 11). These composite expression markers are localized to discrete domains in *L. variegatus* but not in *H. erythrogramma*, suggesting that major embryonic cell lineages have yet to differentiate transcriptionally in this species at this stage of development. **d**, Integrated UMAP of cells from both species, colour-coded by species (left) and by cluster (right), and showing both species (top) and separated by species (middle and bottom). *L. variegatus* clusters remain well separated (centre right) but *H. erythrogramma* clusters broadly overlap with each other (lower right). As shown in the 2× magnifications, *L. variegatus* clusters contain primarily cells of one colour, while cells from all three *H. erythrogramma* clusters are present in appreciable numbers in the same UMAP space (transparency is set to 40% in the insets to circumvent masking; example inset clusters are *L. variegatus* skeletogenic cells, but the same holds for other *L. variegatus* clusters).

molecular identifiers per cell (Supplementary Fig. 2). Given the limited number of cells in this dataset (2,065 post-filtering) and representation from a single stage, future work examining later developmental stages will be necessary to fully resolve the timing of cell fate specification in *H. erythrogramma* and associated expression of developmental regulatory factors.

The presence of more numerous transcriptional states in *L. variegatus* is even clearer in an integrated projection of expression for 7,671 one-to-one orthologous genes, where *L. variegatus* clusters remain separated but those of *H. erythrogramma* overlap broadly in UMAP space, suggesting cells of this species have more homogeneous transcriptional profiles relative to cells in *L. variegatus* at the same developmental stage (Fig. 3d, compare centre right and lower right panels as well as insets). Three independent methods of scRNA-seq data integration are consistent in showing fewer clusters and greater degree of overlap among cells from different clusters in *H. erythrogramma* (Extended Data Fig. 3; see Methods). These findings are consistent with the ATAC-seq results presented above and our earlier lineage-tracing and bulk RNA-seq studies[19,23,25,26], all of which indirectly point to a delay in fate specification in *H. erythrogramma*.

Further, clusters in the early *H. erythrogramma* embryo do not express similar suites of transcription factors to those in the ancestral state, and none corresponds to the distinctive skeletogenic cell lineage of *L. variegatus* and *S. purpuratus* that is established earlier in their development (Fig. 3c,d and Supplementary Figs. 1 and 3). For instance, *delta* and *alx1*, which encode critical early regulatory proteins, are expressed exclusively in the skeletogenic cell precursors at the blastula stage in the ancestral state[36,37]. In *H. erythrogramma*, localized transcription of neither *delta* nor *alx1* has commenced (Supplementary Fig. 1). Furthermore, a composite of skeletogenic cell marker gene expression is localized to a discrete cell population in the ancestral state but is not detected in the *H. erythrogramma* embryo (Fig. 3c). Together, these expression differences suggest that the roles of key regulators of the skeletogenic cell fate have evolved during the life history shift.

Some ancient dGRN interactions are lost in *H. erythrogramma*

To investigate how these roles might differ in the derived developmental mode, we first experimentally perturbed the function of Alx1, which is both necessary and sufficient for skeletogenic cell fate specification in the ancestral state[36] (Fig. 4a). Knocking down Alx1 protein with a translation-blocking morpholino anti-sense oligonucleotide (MASO) in *H. erythrogramma* eliminates both larval and adult skeleton (Fig. 4b and Supplementary Table 1), phenocopying the results of previous experiments in sea urchins representing the ancestral condition[36]. This concordance suggests that the function of Alx1 in skeletogenic cell fate specification is conserved. Zygotic transcription of *alx1* (ref. [19]) and skeletogenesis[38] are both markedly delayed in *H. erythrogramma* relative to the ancestral state, but this shift in timing does not by itself indicate a substantive change to the organization of the dGRN.

We next examined HesC, a transcription factor that acts even earlier in the dGRN, repressing transcription of *alx1* outside of the skeletogenic cell lineage (Extended Data Fig. 4). Experimentally eliminating HesC protein in the ancestral dGRN produces a dramatic phenotype, with most cells differentiating as skeletogenic because, in the absence of *hesC* repression, *alx1* is broadly transcribed[39]. In *H. erythrogramma*, however, we found that embryos develop normally following HesC knockdown (Extended Data Fig. 4e,f), suggesting that it no longer acts as a repressor of *alx1* transcription. This interpretation is consistent with restricted spatial expression of *hesC* in *H. erythrogramma* (Extended Data Fig. 4a–d) that would seem to preclude a

broad repressive function for HesC outside of the skeletogenic lineage. However, these experiments cannot rule out co-option of additional developmental functions by *hesC* in *H. erythrogramma* as the assay presented in this study measured only this transcription factor's effect on skeletal mesenchyme differentiation. Future work aimed at validating the loss-of-function phenotype will confirm whether the function of *hesC* is completely lost or acquires novel regulatory roles during *H. erythrogramma* development. Still, altered expression of *hesC* and lack of an overt knockdown phenotype hint at a more profound evolution change within the dGRN.

We therefore turned to Pmar1, another transcriptional repressor that interacts with *hesC* to form a double-negative logic gate within the dGRN[39]: throughout most of the embryo HesC directly represses transcription of *alx1* and other genes encoding positive regulators of the skeletogenic cell fate, permitting differentiation of other cell types; in the vegetal-most cells of the embryo, however, *pmar1* is transiently expressed beginning the 16-cell stage where it represses *hesC*, allowing *alx1* transcription and thus specification of the skeletogenic cell fate[40] (Fig. 4a).

Pmar1 is encoded by a cluster of tandem genes in sea urchins[41]. We identified 10 and 20 closely linked *pmar1* paralogues in *L. variegatus* and *H. tuberculata*, respectively (Supplementary Table 2). The homeodomain, nuclear localization signal and two EH1 protein–protein interaction domains are typically well conserved, although a few likely pseudogenes are present in each species (Fig. 4c and Supplementary Fig. 4). In *H. erythrogramma* we identified 11 *pmar1* paralogues (Supplementary Table 2). Surprisingly, all of these copies contain numerous substitutions, deletions and/or frameshifts, in many cases altering or eliminating over half of the residues within the homeodomain (Fig. 4c), and their expression is barely detectable at the 16-cell and 32-cell stages (Supplementary Fig. 5). In contrast, likely functional orthologues in the other two species differ by 0–3 amino acids out of 60 within the homeodomain. Furthermore, pairwise similarity between *pmar1* orthologues within a species averages greater than 88% for the entire peptide and 93% for the homeodomain in the ancestral state, while *H. erythrogramma* averages just 71.0% and 45.3%, respectively (Fig. 4d). These sequence comparisons indicate that the integrity of the *pmar1* gene family has dramatically decayed in *H. erythrogramma*, raising the question whether these genes with a crucial role in early embryonic patterning have maintained their function in the derived developmental mode.

Previous studies demonstrate that microinjecting *pmar1* messenger RNA into eggs produces a dramatic phenotype, with the resulting widespread overexpression of Pmar1 protein converting most of the embryo to skeletogenic cells[39,42]. Here we utilized this assay to test the repressive function of specific *pmar1* paralogues. We separately microinjected into *L. variegatus* embryos mRNA encoding one *pmar1* paralogue from *L. variegatus* and two from *H. tuberculata*. As expected, these treatments replicated the published phenotype, inducing extensive conversion to the skeletogenic cell fate, confirmed by widespread expression of the larval spicule matrix protein MSP130 (Fig. 4e and Supplementary Table 3). We then separately tested the three most intact paralogues of *pmar1* from *H. erythrogramma*. At the same and higher concentrations, none was able to produce the specific or any other discernible phenotype (Fig. 4e, Supplementary Figs. 6 and 7, and Supplementary Table 3). These results indicate that the repressive role of Pmar1 is retained in *H. tuberculata* but has been lost in *H. erythrogramma*.

Together, these perturbation experiments and sequence comparisons indicate that both components of the double-negative gate near the very top of the dGRN that specifies the skeletogenic cell fate do not function in *H. erythrogramma* as they do in species with the ancestral life history. Remarkably, this excision of a critical early regulatory interaction does not abort either the specification or subsequent function of skeletogenic cells: the role of Alx1, the component of the skeletogenic subcircuit immediately following the double-negative gate, remains intact (Fig. 4b), structural genes characteristic of differentiated skeletogenic cells are transcribed[19] and a simplified larval skeleton is synthesized[38]. This finding is all the more remarkable given that many other transcription factors appear to have conserved roles in *H. erythrogramma*, based on similar expression profiles and, in some cases, experimental perturbation[19,23,43,44]. Taken together, these results reveal *H. erythrogramma* embryos as a mosaic of conserved and altered features that have evolved as a product of its derived life history and associated changes in selective regimes.

Discussion

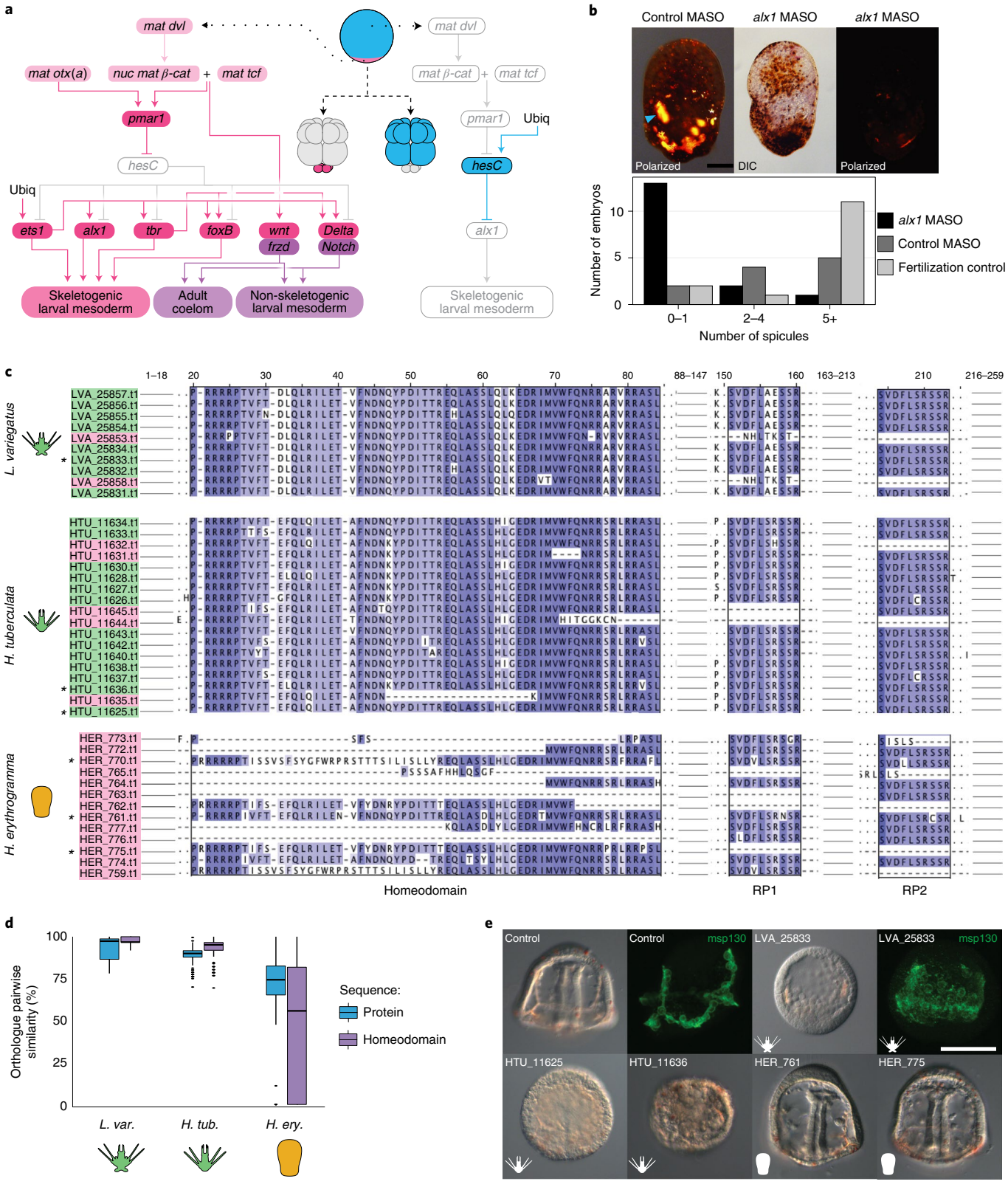
Previous work showed that the evolution of non-feeding development in *Heliocidaris* was accompanied by overt changes in oogenesis, cleavage geometry, morphogenesis and larval morphology, with extensive underlying changes in gene expression[17–20,23]. Whole genome sequence analysis presented here demonstrates that these changes are not merely superficial consequences of amplified maternal provisioning. Although we find evidence for adaptive changes within some coding regions, these are dwarfed by the sheer number and widespread distribution of apparently adaptive changes in the sequences of putative regulatory elements and in the regulation of their chromatin states during early development (Figs. 1c–g and 2). Both types of molecular change are strikingly enriched on the branch where non-feeding development evolved and are over-represented among differentially expressed genes and especially among dGRN genes (Figs. 1f,g and 2c,d). While the potential for natural selection to influence trait evolution through changes in gene regulation by altering regulatory element sequence and chromatin accessibility is widely appreciated, we are aware of few cases that illustrate the influence of both so extensively at a genomic scale and during such a short interval.

Focusing on transcriptional regulation that patterns the early embryo provides a test of the idea that evolutionary conservation of early development is the product of intrinsic constraints. We examined

Fig. 4 | Evolutionary change at the 'top' of a conserved developmental dGRN. **a**, Schematic of the ancestral dGRN that specifies skeletogenic cell fate: HesC suppresses this fate in most of the embryo (blue), but Pmar1 suppresses *hesC* in the precursors of the skeletogenic cells (magenta), where Alx1 then activates a differentiation programme. **b**, Images of control and MASO knockdown of Alx1 in *H. erythrogramma* (early larva; scale bar, 100 µm). Polarized light illuminates skeletal elements (most are anlage of the adult skeleton, with a longer larval 'arm' element out of focus on the left (blue arrow)). Bar chart of *alx1* injection summary statistics in *H. erythrogramma* (for replication details, see also Supplementary Table 1). Knockdown of *alx1* expression eliminates skeleton formation in *H. erythrogramma*, as in the ancestral dGRN[36]. **c**, Alignment of homeodomain (DNA binding) and RP domains (protein–protein interaction) from *pmar1* paralogues. Green: likely functional copy; red: predicted non-functional copy; asterisks: paralogues whose function was experimentally validated. **d**, Within-species pairwise sequence similarity of *pmar1* paralogues. Note rapid sequence divergence among paralogues in *H. erythrogramma*, and particularly within the homeodomain. *He*: $n = 13$ *pmar1* orthologues, 78 comparisons; *Ht*: $n = 18$ *pmar1* orthologues, 153 comparisons; *Lv*: $n = 10$ *pmar1* orthologues, 45 comparisons. Boxes depict the upper and lower quartiles (the 75th and 25th percentiles) of *pmar1* orthologue pairwise BLAST percent identity levels, with the centre line showing the median, and the whiskers extending from the box to the largest and lowest value no further than 1.5× the interquartile range. **e**, Overexpression assays of control and *pmar1* mRNA (prism stage; skeletogenic cells labelled with antibody that recognizes cell surface protein MSP130). DIC and fluorescent images demonstrate that mRNA of *pmar1* paralogues from *L. variegatus* and *H. tuberculata* convert most of the embryo to skeletogenic cells, whereas even the most intact *H. erythrogramma* paralogues show no such phenotype, indicating loss of function (for additional antibody staining and replication details, see Supplementary Fig. 7 and Supplementary Table 3, respectively). RP, repeated peptides.

https://doi.org/10.1038/s41559-022-01906-9

the earliest zygotic patterning event in the sea urchin embryo, where three transcription factors interact to specify two distinct cell fates and simultaneously establish the primary signalling centre of the embryo. There is arguably no set of interactions within the dGRN that is more fundamental to patterning the early sea urchin embryo, and they are conserved among sea urchins that diverged ~225 my ago[5]. Remarkably, however, Pmar1 and HesC, which interact to form a crucial double-negative logic gate[39], have lost their early patterning roles in *H. erythrogramma* (Fig. 4b–e and Extended Data Fig. 4). The case of *pmar1* is particularly striking, as it is present as a tandem array of genes; uniquely in the genome of *H. erythrogramma*, numerous deletions and point mutations alter about half of the homeodomain in each of ten the paralogues, rendering their proteins non-functional (Fig.4c–e and Supplementary Fig. 7).



The magnitude and extent of modifications to the earliest regulatory interactions within the sea urchin dGRN in *H. erythrogramma* demonstrate that some deeply conserved embryonic patterning mechanisms remain evolvable during substantial shifts in selective regimes. More broadly, conservation of gene network architecture does not necessarily imply developmental constraint, but may instead reflect long-term stabilizing selection for performance relative to a particular environment or life history. Abrupt shifts in natural selection provide valuable natural 'perturbation experiments' that can reveal in detail how evolutionary mechanisms shape conservation and change in gene regulation and dGRN organization across the tree of life.

Methods

Genome sequencing and assembly

Tissue collection. *Heliocidaris erythrogramma* (*He*) and *H. tuberculata* (*Ht*) specimens were collected near Sydney Harbor in Sydney, New South Wales, Australia and housed in natural sea water at the Sydney Institute of Marine Science in Mosman, New South Wales, Australia. The inter-pyramidal muscle of Aristotle's lantern (the sea urchin's feeding apparatus), tube feet and ovarian tissue were dissected from a single female individual, flash-frozen in liquid nitrogen and stored at −80 °C until DNA extraction and sequencing.

Genomic DNA sequencing. For each species, a third-generation DNA library was sequenced on a PacBio sequel II CLR platform, generating 90.01 (*He*) and 89.47 (*Ht*) Gb of data with an N50 read length of 17.24 (*He*) and 23.70 (*Ht*) kb. DNA from the same individual for each species was also used to construct 10x Genomics linked-reads and Hi-C libraries, which were sequenced on a BGI-SEQ 500 platform, generating 194.11 (*He*) and 199.11 (*Ht*) Gb and 130.85 (*He*) and 229.03 (*Ht*) Gb of data, respectively. Jellyfish v2.2.6 (ref. [45]) and GenomeScope v1.0.0 (ref. [46]) were deployed to conduct a *k*-mer-based survey of genome composition using linked-read sequencing data based on 17-mer frequency distribution to estimate the genome size and heterozygosity of both *He* and *Ht* (Supplementary Fig. 4a,b).

Genome assembly. PacBio sequencing data was employed to assemble a de novo contig-level genome assembly using Canu v1.8 (minReadLength 1,200, minOverlapLength 1,000) (ref. [47]). Subsequently, HaploMerger2 v3.6 (ref. [48]) was used to create breakpoints in the contigs where potential misjoins have occurred by aligning allelic contigs via Lastz v1.02.00 (ref. [49]). From these fragmented contigs, the longest of each allelic pair was identified and selected using Redundans v0.14a[50], resulting in a near-haploid level genome assembly. The output of this pipeline was polished using Pilon v1.23 (ref. [51]) with 10x sequencing data to improve assembly quality and accuracy at single base resolution. Lastly, contigs were assembled into scaffolds by mapping Hi-C read pairs to the polished assembly with HiC-Pro[52], resulting in approximately 21.95% (*He*) and 32.10% (*Ht*) valid Hi-C reads pairs. Juicer v1.5 (ref. [53]) and 3D-DNA v180419 (ref. [54]) were used to correct and finalize the construction of chromosome-length scaffolds for each species. For Hi-C contact maps, see Supplementary Fig. 8.

Repeat identification and classification. Genomic repetitive elements were identified with RepeatModeler v2.01 (ref. [55]) to generate species-specific repeat element libraries for each. Repeat families were filtered via BLASTn v2.3.0 (ref. [56]) for significant hits to gene models of the well-studied sea urchin *S. purpuratus* (www.echinobase.org) to prevent unintentional masking of genic regions. Repeats were masked from the genome of each species with RepeatMasker v4.1.1 using the most sensitive setting (-s) to identify the location of repetitive elements. Long-terminal repeats were also secondarily identified with LTR_Finder v1.0.7. Outputs of both RepeatMasker and LTR_Finder were then input into RepeatCraft v1.0 (ref. [57]) under default parameters to improve repeat element annotation and identification, resulting in a final genome annotation of repetitive elements. Lastly, repeats were broadly classified into functional categories described by their mode of transposition using TEclass[58].

Gene annotation and prediction strategy. Previously published paired-end RNA-seq reads from six developmental stages for each *Heliocidaris* species[19] were trimmed using Trimmomatic v0.39 (ref. [59]) (TruSeq3-PE.fa: 2:30:10; leading: 3; trailing: 3; slidingwindow: 4:15; minlen: 36), and properly paired reads were mapped to their respective genomes using STAR v2.7.2. For each species, these RNA-seq alignments as well as protein models of the *S. purpuratus* v5.0 genome[60] were input into BRAKER2 (ref. [61]) (–etpmode). This program utilizes a number of additional software as a part of its pipeline including Augustus[62], Genmark-EP+[63], Genemark-ET[64], DIAMOND[65] and SAMtools[66]. Gene models from the BRAKER output for filtered for transposable elements by aligning to a combined database of transposable element sequences from the MAKER gene annotations pipeline[67] and the Dfam v3.3 transposable element database[68] using BLAST-P[56]. Lastly, gene models were improved using the PASA pipeline[69] by supplementing pre-existing gene models with a de novo transcriptome retrieved from ref. [19]. These gene models were annotated by aligning peptide sequences to three separate databases using BLAST-P v2.3.0 (ref. [56]): (1) *S. purpuratus* v4.2 gene models; (2) UniProt KnowledgeBase SwissProt protein models[70]; (3) RefSeq invertebrate protein models with *S. purpuratus* excluded (e-value cut-off: 1×10^{-5}) (ref. [71]) (Supplementary Data 8). The list of sea urchin GRN genes is provided in Supplementary Data 1, retrieved from the Institute of Systems Biology (www.biotapestry.org: accessed 27 June 2017).

Whole genome alignment

Before whole genome alignment, each genome was soft-masked for repetitive elements using each species repeat element library. An optimal scoring matrix for whole genome alignment between each set of species was inferred using the *lastz_D_Wrapper.pl* script of HaploMerger2 v3.6 (ref. [48]). Next, whole genome alignment between each species pair was performed in both directions following UCSC guidelines outlined in the runLastzChain.sh and doBlastzChainNet.pl (https://github.com/ucscGenomeBrowser/kent) to produce .psl, .lav, .chain and finally liftOver files for each whole genome alignment. In addition, .maf files were generated for *H. erythrogramma*, *H. tuberculata* and *L. variegatus* for each chromosome using *H. erythrogramma* as the reference genome using Multiz and TBA[72]. *L. variegatus* was chosen as an outgroup species in this study because its genome was assembled and annotated using an identical sequencing and bioinformatic strategy[21] as the two *Heliocidaris* species presented here, thereby minimizing technical bias in this regard.

ATAC-seq

Sample preparation. For each sea urchin species (*H. erythrogramma*, *H. tuberculata* and *L. variegatus*), adult animals were induced to spawn via injection of 0.5 M KCl solution into the coelom. For each species, three unique male–female pairs were crossed to produce three biologically independent replicates of sea urchin embryos. Each culture was reared in large glass dishes supplied with 20 mm filtered sea water (FSW) that was changed every 6 h. As these species exhibit different developmental rates, a conspicuous developmental milestone, shedding of the fertilization envelope at hatched blastula stage, was selected to maximize developmental synchrony within cultures and across species for comparison. Once a culture reached the blastula stage, live embryos were collected and processed immediately for nucleus preparation and transposase treatment as a part of the ATAC-seq protocol.

ATAC-seq protocol and sequencing. ATAC sample preparation was carried out according to the Omni-ATAC-seq protocol[73]. For each replicate, embryos were washed once in 1 mm FSW and lysed, then 50,000

nuclei were isolated for the transposition reaction as described in the Omni-ATAC-seq protocol using the Illumina TDE1 enzyme and tagmentation (TD) buffer (catalogue numbers 20034197 and 20034198). Sequencing libraries for each replicate were generated via qPCR, and sequencing libraries were purified and size selected using Ampure XP Beads at a 1.8:1 bead volume:library volume (Beckman Coulter). Library quality and transposition efficiency was accessed using a Fragment Analyzer and PROSize 2.0 (Agilent). *H. erythrogramma* and *L. variegatus* libraries were sequenced on an Illumina HiSeq 4000 instrument using 50 bp single-end sequencing at an average of 41.9 million and 37.3 million reads per sample, respectively. *H. tuberculata* libraries were sequenced on an Illumina NovaSeq 6000 instrument using 50 bp paired-end sequencing (only single-end was used for data analysis) at an average of 31.4 million reads per sample.

ATAC-seq data analysis. Raw ATAC-seq reads were trimmed for quality and sequencing adapters using cutadapt[74] v2.3 with the following parameters: -a CTGTCTCTTATACACATCT -q 20–trim-n -m 40. Trimmed reads were then aligned to each species' respective genome using stampy[75] v.1.0.28 using the '−sensitive' set of parameters. ATAC-seq alignments were filtered for mitochondrial sequences and required an alignment quality score of at least 5 using SAMtools v1.9 (ref. [66]).

In this study, we aimed to compare the evolution of orthologous non-coding sites. To accomplish this, we performed a series of liftOvers[76] to convert ATAC-seq alignments between genomic coordinates of each sea urchin species (for description of genome alignments, see previous section). We took an iterative, reciprocal liftOver strategy described below to minimize possible reference bias associated with converting between genome assemblies: (1) *H. erythrogramma*: *He* → *Lv* → *He*; (2) *H. tuberculata*: *Ht* → *Lv* → *He*; (3) *L. variegatus*: *Lv* → *He Lv* → *He*. After filtering and coordinate conversion, all ATAC-seq alignments were referenced to the *H. erythrogramma* genome with an average of 5.9 million alignments per sample to orthologous genomic loci.

Following filtering and coordinate conversion, peaks were called from these alignments using the MACS2 v2.1.2 (ref. [77]) callpeak function (parameters: –nomodel, –keep-dup=auto, –shift 100, –extsize 200) for each species separately. Peak coordinates were merged using the bedtools[78] v2.25 merge function requiring a peak overlap of at least 200 bp to be merged into a single peak. Lastly, for each sample, accessibility of each peak was measured with the bedtools[78] v2.25 multiBamCov function.

Tests for positive selection within OCRs. To test for evidence of positive selection, a neutral genomic reference across all species was assembled. To do this, the genome was first masked for repetitive elements, coding sequence, untranslated genic sequence, non-coding RNAs (including microRNAs, ribosomal RNAs, small nuclear RNAs and transfer RNAs) and ATAC-seq OCRs (see below) in the genome. The remaining, putatively neutrally evolving genome was then divided into 300 bp windows, orthologous regions retrieved from each species' genome, and filtered using the filtering.py and pruning.py scripts of the 'adaptiphy'[79] program (https://github.com/wodanaz/adaptiPhy). Next, branch lengths of each of these neutral sites was estimated using phyloFit[80] (–subst-mod HKY85), highly conserved sites were removed (Supplementary Fig. 5), relative branch lengths were calculated, and sites falling within the middle 50% of relative branch lengths in the *H. erythrogramma* genome were selected as the neutral reference (88,004 sites; Supplementary Data 2 and Supplementary Fig. 9).

To measure branch-specific signatures of positive selection in the non-coding genome, the adaptiPhy[79] pipeline (https://github.com/wodanaz/adaptiPhy) for global tests of natural selection was followed. First, orthologous sequences for non-coding sites of interest were selected from each species' genome into FASTA format. Sequences were trimmed to include only contiguous DNA sequence using the prunning.py script and filtered using the filtering.py script, requiring a minimum alignment length of 75 bases. These trimmed and filtered alignments serve as 'query' sequences of tests for selection. To generate a neutral reference for comparison, ten neutral sites were randomly selected (see above) and concatenated into a single neutral reference sequence. In addition, for each OCR, tests for positive selection were repeated ten times against a unique putatively neutral reference. For each query site replicate, substitution rates of both the query and randomly concatenated neutral reference were estimated using phyloFit[80], and the zeta score was calculated as the ratio of the query substitution rate to the neutral reference substitution rate. In addition, *P* values of likelihood ratio tests for significant levels of branch-specific positive selection were calculated with adaptiPhy[79] pipeline using HyPhy[81]. *P* values and substitution rates for all query and neutral sites were then imported in R v4.0.2 for analysis (Supplementary Data 3).

ATAC-seq peak filtering. After accessibility and rates of selection were calculated for each ATAC-seq peak, herein referred to as an OCR, a series of filtering and quality control metrics were carried out to ensure only high-confidence and quality peaks were compared between species. These filtering steps are as follows: (1) each OCR is required to have at least 75 bp of contiguous, single copy sequence (see Section 3.4) for accurate estimations of selection; (2) for each species, a local composition complexity[82] value of 1.9 or more was required for the OCR to remove repetitive or other low-complexity sequences that may generate inaccurate estimations of selection (module: biopython.org/docs/1.75/api/Bio.SeqUtils.lcc.html); (3) a counts per million value of 3 or more was required in at least two (of the nine) samples to remove OCRs with extremely low accessibility; (4) the midpoint of the OCR must lie within 25 kb (in either direction) of the translational start site of a gene model; (5) the gene nearest to an OCR must be the same gene in each of the species' genomes—in other words, for each OCR and its nearest gene in the *H. erythrogramma*, the orthologous region in the *H. tuberculata* and *L. variegatus* genome must also be closest to a gene model that is orthologous (determined by annotation) to the same gene in the *H. erythrogramma* genome. Given there is nearly no prior knowledge on the *cis*-regulatory landscape for these sea urchin species, these stringent filtering methods were carried out to maximize confidence in comparisons of non-coding sequence evolution and function. This method resulted in a final set of 27,322 high-confidence OCRs for cross-species analysis (Supplementary Data 4).

ATAC-seq statistical analysis. Raw counts of the filtered OCRs were loaded into DESeq2 (ref. [83]) v1.30 to calculate differential accessibility between sample groups. For life history strategy comparisons, *H. tuberculata* and *L. variegatus* were treated as a single group. Differentially accessible sites were classified as having a two-fold accessibility difference between sample groups and supported by a false discovery rate (FDR) of 10%. Significant levels of positive selection were classified as having a median zeta value greater than 1.5 and supported by a median FDR less than 10% across ten replicates for each query site. Branch-specific evidence of positive selection met these criteria for one species, but failed to meet these criteria in the other, as evidenced by a zeta score < 1.5.

Coding selection analyses

To make tests for positive selection in coding sequences analogous to non-coding sequences, only single-copy orthogroups were considered in these analyses. Single-copy orthologues between *H. erythrogramma*, *H. tuberculata*, *L. variegatus* and *Echinometra lucunter* were identified using OrthoFinder[84]. Evidence of episodic positive selection was queried on both the *H. erythrogramma* branch and *H. tuberculata* branch under default parameters using BUSTED[22], by specifying either branch as the 'foreground' branch. *P* values from these analyses are available in Supplementary Data 5. Genes with significant evidence

of episodic positive selection were supported by a P value ≤ 0.10 by likelihood ratio test.

Bulk RNA-sequencing analysis

Raw RNA-seq reads from blastula stage embryos of *He*, *Ht* and *Lv* were retrieved from ref. [19], trimmed and filtered for low-quality bases and reads with Trimmomatic[59], and aligned to each species respective genomes and gene models with STAR[85]. From these alignments, mRNA expression was estimated with Salmon[86] and loaded to R for statistical analysis. Read counts for summed to each gene's best match to the *S. purpuratus* v4.2 gene models to generate a common reference for expression comparisons between species as described in ref. [19]. Differentially expressed genes between life histories were called as having a fold change (FC) in expression >2 and supported by a FDR of 10% or less between *He* and both planktotrophic species (in the same direction), and not DE between *Ht* and *Lv* (Supplementary Data 6).

scRNA-seq

***H. erythrogramma* embryo culturing.** Female *H. erythrogramma* individuals were spawned via intra-coelomic injection of 0.5 ml of 0.5 M KCl. Unfertilized eggs were washed three times in 100 µm FSW. Eggs were fertilized by 2 µl of concentrated sperm in 0.02 g *para*-amino benzoic acid/100 ml FSW. Following fertilization, eggs were washed three additional times in FSW to remove residual sperm and *para*-amino benzoic acid. Fertilized embryos were then cultured at 22–23 °C. At 6 h post-fertilization (hpf) embryos were sampled for microscopy and dissociation, then fixed for scRNA-seq.

Embryo dissociation and fixation. Once embryos developed to the early blastula stage (pre-skeletogenic cell ingression), a portion of the co-culture was taken and washed one time in calcium-free artificial seawater. After washing embryos with calcium-free artificial seawater, 3 ml of embryos was added to 7 ml of dissociation buffer made (1.0 M glycine and 0.25 mM EDTA, pH 8.0) at 4 °C and gently rocked on a rocker for 10 min. Following incubation, embryos were gently triturated 15–20 times to increase disassociation, then 10 ml ice cold 100% methanol was added, and cells were incubated for 10 min and on a rocker. Following incubation, cells were triturated again 15–20 times, and then another 30 ml of ice cold 100% methanol was added to bring the suspension to a final concentration of 80% methanol. This 80% methanol suspension of cells was incubated at 4 °C for 1 h. Following this last fixation step, cells were stored at −20 °C until library preparation.

Rehydration of cells, library preparation and sequencing. Cells were centrifuged at 50*g*, supernatant was discarded, and fixed cells were washed twice and rehydrated in a Sigma 3× saline sodium citrate buffer (SKU SRE0068) before cell count and library preparation. Cell concentration was estimated with a haemocytometer, and volume was adjusted to a final concentration of ~300 cells µl^{-1}. Single cell libraries were prepared using the 10x Genomics 3′ v3 gene expression kit and the 10x Chromium platform to encapsulate single cells within droplets. Library quality was verified using the Agilent 2100 Bioanalyzer. In total, ~3,960 cells were loaded onto the 10x instrument. For the single library preparation, 2,500 cells were targeted, of which 2,066 were successfully captured for sequencing. Libraries were sequenced by the Duke Genomics and Computational Biology Core facility on two NovaSeq6000 S1 flow cells with 28 × 8 × 91 bp sequencing performed.

FastQ generation, indexing and quantification of scRNA-seq. Following sequencing, we used Cellranger v3.1.0 to convert Illumina-generated BCL files to fastq files using the Cellranger 'mkfastq' command. scRNA-seq data for early blastula stage embryos (pre-skeletogenic cell ingression) of *L. variegatus* were retrieved from a published scRNA-seq developmental time course of the species[34]. We then applied the 'mkref' command to index the most recent Lv3.0 genome[21] (for the *Lytechinus* data) and the *H. erythrogramma* genome assembled in this study. The 'count' command was used to demultiplex and count reads mapping to the respective reference *Lv* (53.3% mapping rate) or *He* (94.1% mapping rate) genome. The 'mat2csv' command was used to obtain comma-separated value (CSV) RNA count matrix files for each sample for further downstream analysis. Orthofinder v2.3.12 (ref. [84]) was implemented to identify putative 1–1 orthologous gene models between *Lv* and *He* (Supplementary Data 10).

scRNA computational analyses. We employed a dual strategy for comparing scRNA-seq expression between *He* and *Lv*: (1) a non-integrated analysis in which scRNA-seq from each species was quantified against its own gene models (Figs. 3a) and (2) an integrated analysis in which orthologous anchor genes were used to identify cell types with overlapping expression profiles between each species (Fig. 3d). CSV RNA count matrix files were uploaded to R, and a Seurat object was generated for (1) each species separately quantified against their own gene models and (2) orthologous genes between *H. erythrogramma* and *L. variegatus* (Supplementary Data 7). Each dataset was filtered to remove low-quality cells with nFeature_RNA >200, nFeature_RNA <5,500, and nCount_RNA <7,500. To ensure differences in input cell number did not bias detection of different cell types between samples, the *Lv* dataset was separately subsampled to 2,065 randomly selected cells and clustering analyses were repeated (Extended Data Fig. 2). Furthermore, both datasets had comparable distributions of genes per cell and unique molecular identifiers per cell numbers (Supplementary Fig. 2).

For the non-integrated analyses, 'SCTransform', a regularized negative binomial regression method that stabilizes variance across samples, was applied to perform normalization and removal of technical variation[87], while preserving biological variation. We next performed principal component analysis on the SCTransformed Seurat object file of raw gene expression counts and found the nearest neighbours using ten principal component dimensions of variable gene space. UMAP[88] was applied to multi-dimensional scRNA-seq data to visualize the cells in a two-dimensional space. Finally, clustering was performed using graph-based Louvain clustering with resolution 0.5, resulting in seven clusters in *L. variegatus* and three clusters in *H. erythrogramma*. The clusters were putatively annotated using dGRN genes and published in situ hybridization (ISH) patterns as markers (Supplementary Fig. 1 and Supplementary Data 11), and ambiguous cluster identities are conservatively denoted as broad embryological territories (for example, endomesoderm and ectoderm A–C) or as 'unknown'.

To perform the integrated analyses of scRNA-seq data between species, we carried out three independent methods for scRNA-seq data integration to compare expression of the same orthologous genes between *H. erythrogramma* and *L. variegatus* and identify putative overlapping cell types: (1) canonical correlation analysis[89], (2) reciprocal principal component analysis (satijalab.org/seurat/articles/integration_rpca.html) and (3) Harmony[90] (Fig. 3d and Extended Data Fig. 3). For each scRNA-seq strategy (non-integrated and canonical correlation analysis), counts for all gene models and orthologous sets of genes are included as Supplementary Data 7, and top marker genes for each cluster are available in Supplementary Data 9.

H. erythrogramma microinjection and ISH

MASO design and microinjection. MASOs were constructed to target the translation start site of *alx1* (ATCAATTCGGAGTTAAGTCTCGGCA) and *hesC* (ATCCAGATGTGTTAAGCATGGTTGC) and synthesized by Gene Tools (Philomath). Control morpholinos included a standard negative control morpholino recommended by the manufacturer (CCTCTTACCTCAGTTACAATTTATA) and a scrambled morpholino for HesC (ATCGACATCTGTTAACCATCGTTGC). Fertilized eggs of *H. erythrogramma* were injected as described in ref. [91] at a concentration of 100 µM and 200 µM for *alx1* and 500 µM for *hesC*, then reared

at 22 °C in pasteurized, 0.22 µm filtered seawater + penicillin (100 units ml^{-1}) and streptomycin sulfate (0.1 mg ml^{-1}) (Sigma P4333A). Injected embryos were checked for developmental abnormalities and mortality every 6 h until fixation.

Fixation and ISH. *H. erythrogramma* embryos were fixed for ISH for ~16 h overnight at 4 °C in 4% paraformaldehyde (Sigma, 158127) + 20 mM EPPS (Sigma, E1894) in FSW, washed three times in FSW, and dehydrated step-wise into 100% MeOH and stored at −20 °C. The full-length *HesC* coding sequence was synthesized in vitro by GenScript (Piscataway) and subcloned (NCBI insert number MK749159) and used as template to make anti-sense RNA probes for ISH. ISH of *H. erythrogramma* was performed according to previously published methods[44]. Hybridizations were carried out at 65 °C and stringency washed at 0.1% saline sodium citrate.

Imaging. Fixed *H. erythrogramma* embryos were washed with 100% EtOH, cleared and mounted in 2:1 (v/v) benzyl benzoate:benzyl alcohol. Differential interference contrast (DIC) or polarized light (PL) micrographs were taken on Olympus BX60 upright microscope with an Olympus DP73 camera. ISH images were taken on a Zeiss Upright AxioImager with a Zeiss MRm camera using ZEN Pro 2012 software.

Pmar1 mRNA overexpression assays

mRNA synthesis. Sequences of *pmar1* orthologues were retrieved from each species' respective genome annotations (*Lv*[21], *He* and *Ht*, this study) (Supplementary Table 2). One *Lv* orthologue (LVA_25833.t1) was selected for overexpression assays as it represents the orthologue tested in previous overexpression assays[42], while two *Ht* (HTU_11636.t1 and HTU_11625.t1) and three *He* (HER_770.t1, HER_761.t1 and HER_775.t1) were selected for overexpression assays as they represent orthologues with the highest identity to the species' consensus sequence and therefore predicted as genes most likely to be functional. Construct templates for each orthologue were ordered from Twist Biosciences, and mRNA was synthesized from these constructs with a ThermoFisher MEGAshortscript T7 Transcript Kit (AM1354).

Overexpression experiments. Female *L. variegatus* individuals were spawned via intra-coelomic injection of 0.5 M KCl, washed in FSW and fertilized with 1 µl of concentrated sperm in 100 ml FSW. For each construct, at least four rounds of microinjections were conducted, with each round including 30–50 healthy embryos. *Lv* constructs were injected at a concentration of 250 ng $µl^{-1}$. *Ht* constructs were injected at a concentration of 1,200 ng $µl^{-1}$, and *He* constructs were injected at 1,500 ng $µl^{-1}$. Higher concentrations of *Heliocidaris* constructs were used to reproducibly obtain the skeletal cell conversion phenotype, and reduced sensitivity of these assays may be attributable to less optimal cross-species interactions of Pmar1 in regulating the *L. variegatus* genome. Following injection, embryos were incubated at 23 °C and imaged at 24 hpf.

Imaging and immunostaining. At 24 hpf, live embryos from each experiment were mounted on slides and imaged using DIC microscopy. Embryos were imaged on a Zeiss Axioplan II upright microscope controlled by Zen software. Also at 24 hpf, selected embryos were fixed in 100% ice-cold methanol. Immunostaining was carried out as described in ref. [92] to mark expression of Msp130 protein. Blocking and incubation of the secondary antibody was increased to 1 h. Incubation of the primary antibody was set to 48 h. A Zeiss 880 inverted confocal Airyscan microscope controlled by Zen software was used to take *Z*-stack images of stained embryos. Pmar1 overexpression results are summarized in Supplementary Table 3.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data Availability

Genomes. Sequencing reads used to assemble the *Heliocidaris* genomes and the genome assemblies themselves are available on the Chinese National GeneBank (CNP0002233) and NCBI (PRJNA869508). Genome assemblies of *Heliocidaris erythrogramma* and *Heliocidaris tuberculata* are also available on NCBI (PRJNA827916 and PRJNA827769, respectively). Genome assembly of *Lytechinus variegatus* is previously published[21] and available on NCBI (PRJNA657258). Genome annotations and files associated with whole genome alignments between species are available on Dryad (https://doi.org/10.5061/dryad.sj3tx966v). ATAC-seq. Raw sequencing reads for the ATAC-seq dataset are available on NCBI (PRJNA828607). Alignment files are available on Dryad (https://doi.org/10.5061/dryad.sj3tx966v). Result files associated with the ATAC-seq analyses are available as Supplementary Data 2–4. Bulk RNA-seq. Bulk RNA-seq data were retrieved from ref. [19]. scRNA-seq. Raw sequencing reads for the *Heliocidaris erythrogramma* single-cell ATAC-seq dataset are available on NCBI (PRJNA833141). Sequencing reads from the *Lytechinus variegatus* scRNA-seq dataset were retrieved from ref. [34] and are available on NCBI (PRJNA765003). Results files associated with the scRNA-seq analyses are available as Supplementary Data 7 and 9.

Code Availability

Code and analyses associated with these result figures are available on GitHub at https://github.com/phillipdavidson/heliocidaris_analyses.

References

1. Erwin, D. H. & Davidson, E. H. The evolution of hierarchical gene regulatory networks. *Nat. Rev. Genet.* **10**, 141–148 (2009).
2. Oliveri, P. & Davidson, E. H. Gene regulatory network controlling embryonic specification in the sea urchin. *Curr. Opin. Genet Dev.* **14**, 351–360 (2004).
3. McClay, D. R. Evolutionary crossroads in developmental biology: sea urchins. *Development* **138**, 2639–2648 (2011).
4. Gildor, T. & Ben-Tabou de-Leon, S. Comparative study of regulatory circuits in two sea urchin species reveals tight control of timing and high conservation of expression dynamics. *PLoS Genet.* **11**, e1005435 (2015).
5. Yamazaki, A. & Minokawa, T. Expression patterns of mesenchyme specification genes in two distantly related echinoids, *Glyptocidaris crenularis* and *Echinocardium cordatum*. *Gene Expr. Patterns* **17**, 87–97 (2015).
6. Erkenbrack, E. M., Davidson, E. H. & Peter, I. S. Conserved regulatory state expression controlled by divergent developmental gene regulatory networks in echinoids. *Development* https://doi.org/10.1242/dev.167288 (2018).
7. Hinman, V. F., Nguyen, A. T., Cameron, R. A. & Davidson, E. H. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc. Natl Acad. Sci. USA* **100**, 13356–13361 (2003).
8. Koch, N. M. et al. Phylogenomic analyses of echinoid diversification prompt a re-evaluation of their fossil record. *eLife* **11**, e72460 (2022).
9. Maynard Smith, J. et al. Developmental constraints and evolution: a perspective from the Mountain Lake conference on development and evolution. *Q. Rev. Biol.* **60.3**, 265–287 (1985).
10. Arthur, W. *A Theory of the Evolution of Development* (John Wiley & Sons Inc., 1988).
11. Raff, R. A. Constraint, flexibility, and phylogenetic history in the evolution of direct development in sea urchins. *Dev. Biol.* **119**, 6–19 (1987).
12. Wray, G. A. Parallel evolution of nonfeeding larvae in echinoids. *Syst. Biol.* **45**, 308–322 (1996).
13. McEdward, L. R. & Miner, B. G. Larval and life-cycle patterns in echinoderms. *Can. J. Zool.* **79**, 1125–1170 (2001).

14. Marshall, D. J., Krug, P. J., Kupriyanova, E. K., Byrne, M. & Emlet, R. B. The biogeography of marine invertebrate life histories. *Annu Rev. Ecol. Evol. S* **43**, 97 (2012).
15. Laegdsgaard, P., Byrne, M. & Anderson, D. T. Reproduction of sympatric populations of *Heliocidaris erythrogramma* and *H. tuberculata* (Echinoidea) in New South Wales. *Mar. Biol.* **110**, 359–374 (1991).
16. Morgan, S. G. Life And Death in the Plankton: Larval Mortality and Adaptation. in *Ecology of Marine Invertebrate Larvae* (ed. McEdward, L.R.) 279-321 (CRC Press, 2020).
17. Raff, R. A. Direct-developing sea urchins and the evolutionary reorganization of early development. *Bioessays* **14**, 211–218 (1992).
18. Raff, R. A. & Byrne, M. The active evolutionary lives of echinoderm larvae. *Heredity* **97**, 244–252 (2006).
19. Israel, J. W. et al. Comparative developmental transcriptomics reveals rewiring of a highly conserved gene regulatory network during a major life history switch in the sea urchin genus *Heliocidaris*. *PLoS Biol.* **14**, e1002391 (2016).
20. Wray, G. A. Extreme phenotypic divergence and the evolution of development. *Curr. Top. Developmental Biol.* **146**, 79–112 (2021).
21. Davidson, P. L. et al. Chromosomal-level genome assembly of the sea urchin *Lytechinus variegatus* substantially improves functional genomic analyses. *Genome Biol. Evol.* **12**, 1080–1086 (2020).
22. Murrell, B. et al. Gene-wide identification of episodic selection. *Mol. Biol. Evol.* **32**, 1365–1371 (2015).
23. Wang, L. et al. Genetic basis for divergence in developmental gene expression in two closely related sea urchins. *Nat. Ecol. Evol.* **4**, 831–840 (2020).
24. Davidson, P. L., Byrne, M. & Wray, G. A. Evolutionary changes in the chromatin landscape contribute to reorganization of a developmental gene network during rapid life history evolution in sea urchins. *Mol. Biol. Evol.* **39**, msac172 (2022).
25. Wray, G. A. & Raff, R. A. Evolutionary modification of cell lineage in the direct-developing sea urchin *Heliocidaris erythrogramma*. *Dev. Biol.* **132**, 458–470 (1989).
26. Wray, G. A. & Raff, R. A. Novel origins of lineage founder cells in the direct-developing sea urchin *Heliocidaris erythrogramma*. *Dev. Biol.* **141**, 41–54 (1990).
27. Smith, M. S., Turner, F. R. & Raff, R. A. Nodal expression and heterochrony in the evolution of dorsal-ventral and left-right axes formation in the direct developing sea urchin *Heliocidaris erythrogramma*. *J. Exp. Zool. Part B* **310b**, 609–622 (2008).
28. Wilson, K. A., Andrews, M. E. & Raff, R. A. Dissociation of expression patterns of homeodomain transcription factors in the evolution of developmental mode in the sea urchins *Heliocidaris tuberculata* and *H. erythrogramma*. *Evol. Dev.* **7**, 401–415 (2005).
29. Love, A. C. & Raff, R. A. Larval ectoderm, organizational homology, and the origins of evolutionary novelty. *J. Exp. Zool. Part B* **306**, 18–34 (2006).
30. O'Hara, T. & Byrne, M. *Australian Echinoderms: Biology, Ecology and Evolution* (CSIRO Publishing, 2017).
31. Cameron, R. A., Hough-Evans, B. R., Britten, R. J. & Davidson, E. H. Lineage and fate of each blastomere of the eight-cell sea urchin embryo. *Genes Dev.* **1**, 75–85 (1987).
32. Davidson, E. H. Lineage-specific gene expression and the regulative capacities of the sea urchin embryo: a proposed mechanism. *Development* **105**, 421–445 (1989).
33. Oliveri, P., Tu, Q. & Davidson, E. H. Global regulatory logic for specification of an embryonic cell lineage. *Proc. Natl Acad. Sci. USA* **105**, 5955–5962 (2008).
34. Massri, A. J. et al. Developmental single-cell transcriptomics in the *Lytechinus variegatus* sea urchin embryo. *Development* https://doi.org/10.1242/dev.198614 (2021).
35. Davidson, E. H., Cameron, R. A. & Ransick, A. Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms. *Development* **125**, 3269–3290 (1998).
36. Ettensohn, C. A., Illies, M. R., Oliveri, P. & De Jong, D. L. 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 (2003).
37. Croce, J. C. & McClay, D. R. Dynamics of Delta/Notch signaling on endomesoderm segregation in the sea urchin embryo. *Development* **137**, 83–91 (2010).
38. Emlet, R. B. Larval spicules, cilia, and symmetry as remnants of indirect development in the direct developing sea urchin *Heliocidaris erythrogramma*. *Dev. Biol.* **167**, 405–415 (1995).
39. Revilla-i-Domingo, R., Oliveri, P. & Davidson, E. H. A missing link in the sea urchin embryo gene regulatory network: hesC and the double-negative specification of micromeres. *Proc. Natl Acad. Sci. USA* **104**, 12383–12388 (2007).
40. Oliveri, P., Davidson, E. H. & McClay, D. R. Activation of pmar1 controls specification of micromeres in the sea urchin embryo. *Dev. Biol.* **258**, 32–43 (2003).
41. Yamazaki, A. et al. pmar1/phb homeobox genes and the evolution of the double-negative gate for endomesoderm specification in echinoderms. *Development* https://doi.org/10.1242/dev.182139 (2020).
42. Wu, S. Y. & McClay, D. R. The Snail repressor is required for PMC ingression in the sea urchin embryo. *Development* **134**, 1061–1070 (2007).
43. Smith, M. S., Turner, F. R. & Raff, R. A. Nodal expression and heterochrony in the evolution of dorsal–ventral and left–right axes formation in the direct-developing sea urchin Heliocidaris erythrogramma. *J. Exp. Zool. Part B* **310.8**, 609–622 (2008).
44. Koop, D. et al. Nodal and BMP expression during the transition to pentamery in the sea urchin Heliocidaris erythrogramma: insights into patterning the enigmatic echinoderm body plan. *BMC Dev. Biol.* **17**, 4 (2017).
45. Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
46. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
47. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722–736 (2017).
48. Huang, S., Kang, M. & Xu, A. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics* **33**, 2577–2579 (2017).
49. Harris, R. S. *Improved Pairwise Alignment of Genomic DNA* (Pennsylvania State Univ., 2007).
50. Pryszcz, L. P. & Gabaldon, T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **44**, e113 (2016).
51. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
52. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
53. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
54. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
55. Flynn, J. M., Hubley, R., Goubert, C. & Smit, A. F. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).

56. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
57. Wong, W. Y. & Simakov, O. RepeatCraft: a meta-pipeline for repetitive element de-fragmentation and annotation. *Bioinformatics* **35**, 1051–1052 (2019).
58. Abrusan, G., Grundmann, N., DeMester, L. & Makalowski, W. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* **25**, 1329–1330 (2009).
59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
60. Arshinoff, B. I. et al. Echinobase: leveraging an extant model organism database to build a knowledgebase supporting research on the genomics and biology of echinoderms. *Nucleic Acids Res.* **50.D1**, D970–D979 (2022).
61. Bruna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, lqaa108 (2021).
62. Stanke, M., Schoffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).
63. Bruna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom. Bioinform.* **2**, lqaa026 (2020).
64. Lomsadze, A., Burns, P. D. & Borodovsky, M. Integration of mapped RNA-seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42**, e119 (2014).
65. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
66. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
67. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinforma.* **48**, 4 11 11–39 (2014).
68. Storer, J., Hubley, R., Rosen, J., Wheeler, T. J. & Smit, A. F. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob. DNA* **12**, 2 (2021).
69. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
70. Bateman, A. Uniprot: a universal hub of protein knowledge. *Protein Sci.* **28**, 32–32 (2019).
71. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
72. Blanchette, M. et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
73. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
74. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
75. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
76. Hinrichs, A. S. et al. The UCSC genome browser database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
77. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
78. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
79. Berrio, A., Haygood, R. & Wray, G. A. Identifying branch-specific positive selection throughout the regulatory genome using an appropriate proxy neutral. *BMC Genomics* **21**, 359 (2020).
80. Siepel, A. & Haussler, D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* **21**, 468–488 (2004).
81. Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
82. Konopka, A. K. Sequence complexity and composition. *eLS* http://www.els.net/WileyCDA/ElsArticle/refId-a0005260.html (2005).
83. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
84. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
85. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
86. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
87. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
88. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* https://doi.org/10.1038/nbt.4314 (2018).
89. Butler, A. et al. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36.5**, 411–420 (2018).
90. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
91. Edgar, A., Byrne, M. & Wray, G. A. Embryo microinjection of the lecithotrophic sea urchin *Heliocidaris erythrogramma*. *J. Biol. Methods* **6**, e119 (2019).
92. Rho, H. K. & McClay, D. R. The control of foxN2/3 expression in sea urchin embryos and its function in the skeletogenic gene regulatory network. *Development* **138**, 937–945 (2011).

Acknowledgements

We thank the Sydney Institute of Marine Science (SIMS) for facilities, as well as the SIMS staff for their assistance. This work was supported by the National Science Foundation Division of Integrative Organismal Systems (award no. 1929934 to G.A.W.), National Science Foundation Graduate Research Fellowships to H.R.D. and A.J.M., and an Australian Research Council Discovery Grant (award no. DP120102849 to M.B.).

Author contributions

L.W., M.B., G.F. and G.A.W. conceived and designed the study. L.W., D.K. and P.C. collected tissues for genomic sequencing. Y.Z. performed genomic DNA extraction and sequencing library preparations. P.L.D., H.G. and H.Z. assembled the genomes. P.L.D. performed genome annotation, alignments and data analysis. P.L.D. and A.B. performed selection analyses. P.L.D. and H.R.D. collected, prepared and analysed ATAC-seq libraries. A.J.M. collected, prepared and analysed scRNA-seq libraries. J.S.S. and A.E. performed embryonic expression and injection assays. P.L.D. and G.A.W. wrote the manuscript, and all authors contributed to manuscript revisions.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-022-01906-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01906-9.

Correspondence and requests for materials should be addressed to Gregory A. Wray.

Peer review information *Nature Ecology & Evolution* thanks Ina Arnone, Veronica Hinman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

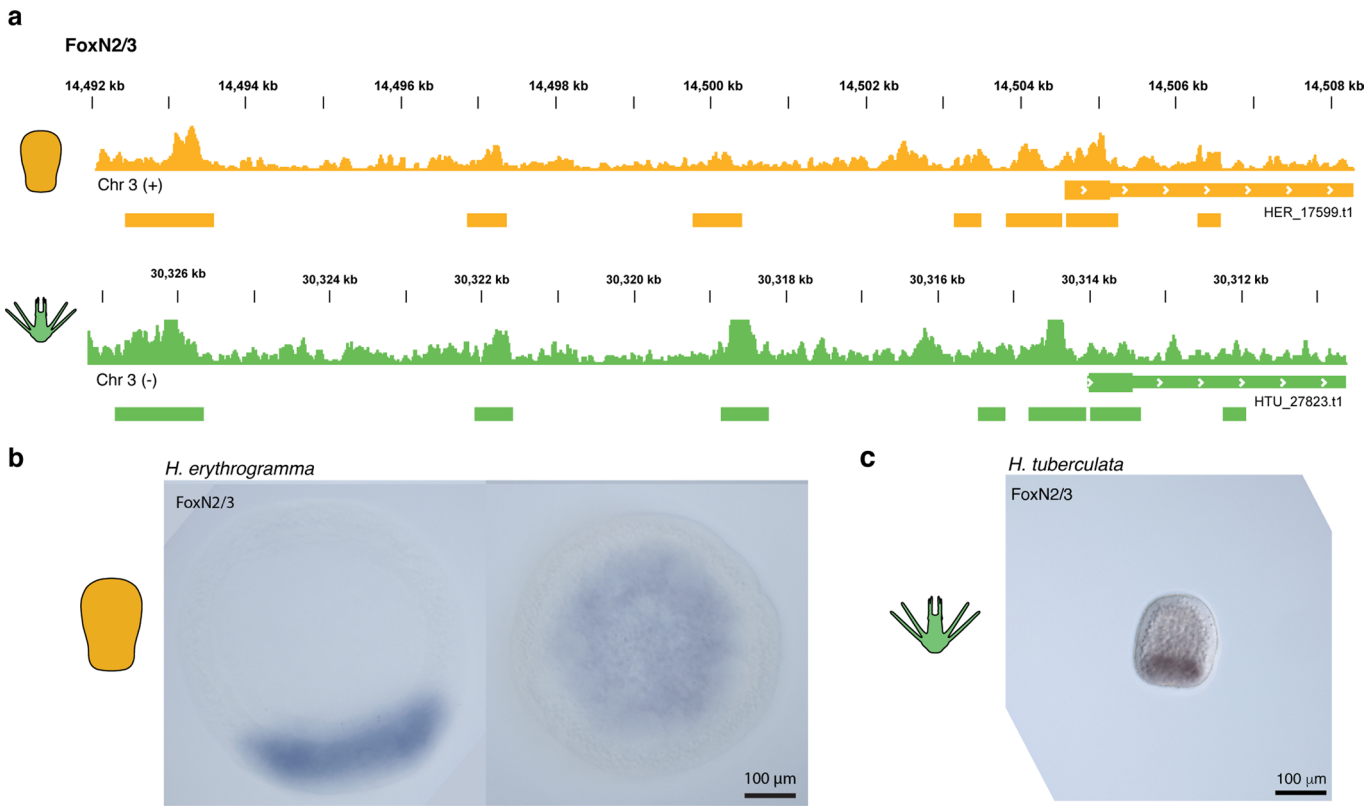
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

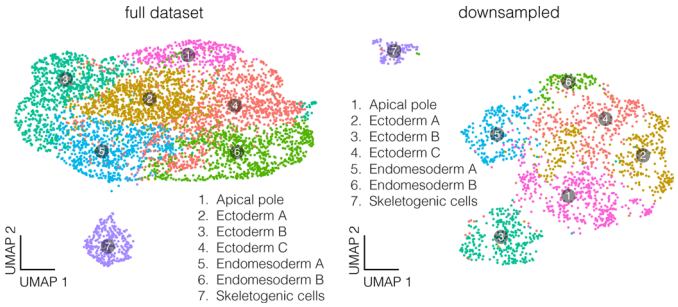
© The Author(s), under exclusive licence to Springer Nature Limited 2022

https://doi.org/10.1038/s41559-022-01906-9



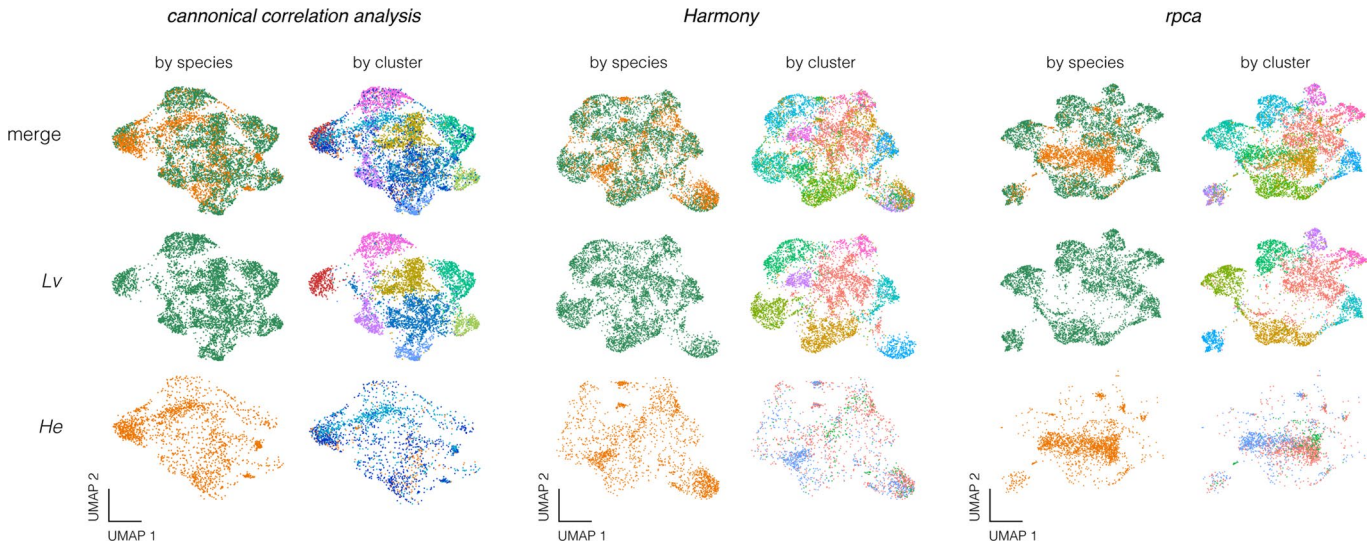
Extended Data Fig. 1 | Chromatin landscape and expression domain of *foxN2/3* is conserved. Chromatin accessibility nearby *foxN2/3* in *H. erythrogramma* (top, orange) and *H. tuberculata* (bottom, green), including seven open chromatin regions (OCRs). *In-situ* hybridization of *foxN2/3* in blastula-stage **b**, *H. erythrogramma* and **c**, *H. tuberculata* embryos. Micrographs derive from a single round of *in-situ* experiments for each species.

https://doi.org/10.1038/s41559-022-01906-9



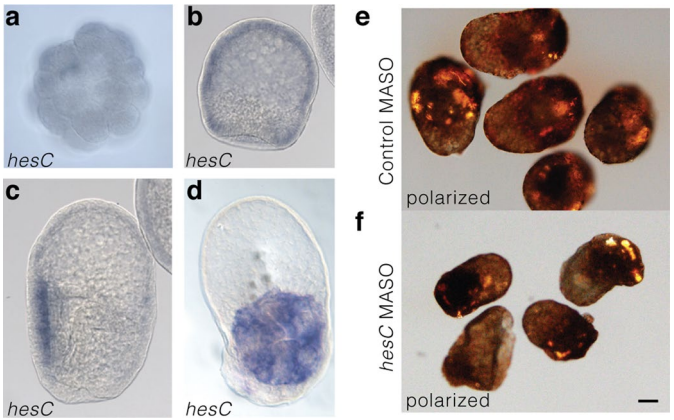
Extended Data Fig. 2 | Scaling down number of *L. variegatus* cells does not substantially affect clustering. UMAP of non-integrated single cell RNA-seq data from *L. variegatus*, in which the data has been randomly subsampled to 2065 cells so that the cell number is equivalent to the *H. erythrogramma* dataset. This subsampling does not change the number and general spatial relationship among clusters in *L. variegatus* (that is distinct cluster of skeletogenic cells separate from remainder of cells in the embryo).

https://doi.org/10.1038/s41559-022-01906-9



Extended Data Fig. 3 | Three independent methods of integrating single cell RNA-seq data recover similar clustering relationships among *L. variegatus* and *H. erythrogramma* cells. Canonical correlation analysis (CCA), reciprocal principal component analysis (RPCA), and Harmony each recover more numerous, distinct transcriptional states in *L. variegatus* relative to *H. erythrogramma* following integration of single cell expression data. All three methods reveal a consistent number of clusters in both species. In particular, *H. erythrogramma* contains fewer clusters there is more extensive overlap of cells among clusters. These results suggesting that differentiation of discrete cell populations is delayed in the *H. erythrogramma* embryo.

https://doi.org/10.1038/s41559-022-01906-9



Extended Data Fig. 4 | *hesC* appears to have lost its ancestral role of repressing larval skeletal cell specification in *H. erythrogramma*. Derived expression patterns of *hesC* in *H. erythrogramma* at **a**, cleavage; **b**, blastula; **and c**, **d**, larva stage embryos. Micrographs derive from a single round of *in-situ* experiments. **e**, **f**, Control and MASO knock-down of HesC in *H. erythrogramma* (early larva; scale bar 100 µm). Polarized light illuminates skeletal elements. HesC knockdown appears to show no phenotype, a dramatic change from the ancestral dGRN. Injection experiments were replicated twice.

nature portfolio

Corresponding author(s): Gregory Wray

Last updated by author(s): Aug 19, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Details of custom code or analyses are provided on GitHub at https://github.com/phillipdavidson/heliocidaris_analyses . Otherwise, only pulbished software was used and described in the Methods.
Data analysis	Details of custom code or analyses are provided on GitHub at https://github.com/phillipdavidson/heliocidaris_analyses . Otherwise, only pulbished software was used and described in the Methods.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Genomes
Sequencing reads used to assemble the Heliocidaris genomes and the genome assemblies themselves are available on the Chinese National GeneBank

(CNP0002233) and NCBI (PRJNA869508). Genome assemblies of Heliocidaris erythrogramma and Heliocidaris tuberculata are also available on NCBI (PRJNA827916 and PRJNA827769, respectively). Genome assembly of Lytechinus variegatus is previously published21 and available on NCBI (PRJNA657258). Genome annotations and files associated with whole genome alignments between species are available on Dryad (https://doi.org/10.5061/dryad.sj3tx966v).

ATAC-seq

Raw sequencing reads for the ATAC-seq dataset are available on NCBI (PRJNA828607). Alignment files are available on Dryad (https://doi.org/10.5061/dryad.sj3tx966v). Result files associated with the ATAC-seq analyses are available as Supplementary Data 2-4.

Bulk RNA-seq

Bulk RNA-seq data was retrieved from Israel et al, 201619.

Single cell RNA-seq

Raw sequencing reads for the Heliocidaris erythrogramma single cell ATAC-seq dataset are available on NCBI (PRJNA833141). Sequencing reads from the Lytechinus variegatus single cell RNA-seq dataset were retrieved from Massri et al, 202134 and available on NCBI (PRJNA765003). Results files associated with the single cell RNA-seq analyses are available as Supplementary Data 7,9.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For ATAC-seq, 3 biological replicates for each species for a total of 9 samples. For scRNA-seq, eggs from one cross were used to make a single library.
Data exclusions	No data excluded from analysis.
Replication	ATAC-seq sample groups include 3 biological replicates. Developmental phenotypes were replicated via injection of 1-3 uniques mRNA constructs per species. 2-7 biological replicates were used in microinjection experiments.
Randomization	For microinjection experiments, eggs were randomly selected from each biologically independent culture.
Blinding	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	1d5 mouse antibody (anti-MSP130); Cy2-conjugated secondary antibodies (Jackson Laboratories)
Validation	Rho, H. K. & McClay, D. R. The control of foxN2/3 expression in sea urchin embryos and its function in the skeletogenic gene regulatory network. Development 138, 937-945, doi:10.1242/dev.058396 (2011).

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	NA
Wild animals	Heliocidaris erythrogramma and Heliocidaris tuberculata individuals were collected in Sydney Harbor, Sydney, Australia. Lytechinus variegatus individuals were collected in Beaufort, North Carolina, USA.
Reporting on sex	NA
Field-collected samples	Heliocidaris animals were housed in filtered natural sea water at the Sydney Institute of Marine Science at 22 degrees Celsius under a 12hr/12hr photoperiod. Lytechinus individuals were housed in aquaria at Duke University at 23 degrees Celsius on a 24 hr photoperiod.
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.