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

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[5], Paula Cisternas[5], He Zhang[3], Yaolei Zhang[2,4], Maria Byrne[5], Guangyi Fan[2,3], Gregory A Wray[1,6]*

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. Qingdao-Europe Advanced Institute for Life Sciences, BGI-Qingdao, BGI-Shenzhen, Qingdao, China
5. School of Life and Environmental Science, The University of Sydney, Sydney, NSW, Australia
6. Center for Genomic and Computational Biology, Duke University, Durham, NC, USA

* Corresponding author: gwray@duke.edu

ABSTRACT

Changes in developmental gene regulatory networks (dGRNs) underlie much of the diversity of life[1], but the evolutionary mechanisms that operate on interactions with these networks 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 analyze genomes, epigenomes, and transcriptomes during early development in two sea urchin species in the genus *Heliocidaris* that exhibit highly divergent life histories and in an outgroup species. Signatures of positive selection and changes in chromatin status within putative gene regulatory elements are both enriched on the branch leading to the derived life history, and particularly so near dGRN genes; in contrast, positive selection within protein-coding regions have at most a modest enrichment in branch and function. 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 dGRN genes with conserved roles in cell fate specification. Experimentally perturbing the function of three key transcription factors reveals profound evolutionary changes in the earliest events that pattern the embryo, disrupting regulatory interactions previously conserved for ~225 million years. Together, these results demonstrate that natural selection can rapidly reshape developmental gene expression on a broad scale when selective regimes abruptly change and that even highly conserved dGRNs and patterning mechanisms in the early embryo remain evolvable under appropriate ecological circumstances.

39 **MAIN**

40 The well-defined dGRN of early development in sea urchins[2] provides a powerful
41 framework for investigating the evolution of embryonic patterning mechanisms. Interactions
42 between genes encoding transcription factors and their target genes within this dGRN are almost
43 completely conserved among species that diverged ~30-40 million years (my) ago[3,4], with some
44 interactions conserved for ~225 my[5], ~275 my[6], or even ~480 my[7,8]. One possible explanation
45 for this observation is developmental constraints, such that early developmental processes are
46 largely immutable given their critical roles in body plan organization and tissue specification[9,10].
47 Under this scenario, any change in a critical interaction during early development would have
48 widespread effects on later processes, which would almost always be deleterious. Still, an
49 important confound remains untested: the species with deeply conserved developmental
50 mechanisms all share the same life history mode, involving low maternal provisioning and an
51 extended feeding larval phase. Species with derived life histories involving massive maternal
52 provisioning and highly-abbreviated, nonfeeding pre-metamorphic development have evolved on
53 multiple occasions within sea urchins[11-13], possibly in response to lower or more unpredictable
54 food availability[14]. These species can reveal how conserved regulatory interactions and
55 patterning mechanisms respond to major shifts in selective regimes.

56 The Australian sea urchin genus *Heliocidaris* includes two recently diverged species: *H.*
57 *tuberculata* representing the ancestral life history and *H. erythrogramma* the derived state[15]
58 (Fig.1a). The shift to nonfeeding development radically alters natural selection on development:
59 with feeding no longer necessary, high mortality rates in the plankton[16] impose strong selection
60 to decrease time to metamorphosis[12]. Numerous anatomical features and gene expression profiles
61 of early development that are broadly conserved among sea urchins differ markedly between

62 these closely related species[17-20](Fig.1b). We sought to learn whether these recently evolved
63 differences are merely superficial and mask deeply conserved developmental mechanisms, or
64 whether they are the product of substantive evolutionary changes in early cell fate specification
65 and dGRN organization. Evidence for the former would suggest developmental constraints play
66 an important role in limiting genetic and regulatory composition of the ancestral GRN, whereas
67 support for the latter would point to a flexible morphogenetic system derived from an embryonic
68 program conserved at least in part by stabilizing selection that is adaptable to alternative
69 developmental life histories.

70 **RESULTS**

71 **Natural selection has sculpted the regulatory landscape of *H. erythrogramma***

72 We took advantage of the recent (~5 my) divergence between the two *Heliocidaris*
73 species to carry out detailed analyses of orthologous coding and noncoding regions of the
74 genome, focusing on the transcription factors and regulatory elements that constitute the
75 backbone of the dGRN and underlie cell fate specification mechanisms (see Supplementary Data
76 1 for list of 192 dGRN genes). Genomes of *H. erythrogramma* and *H. tuberculata* were each
77 sequenced, assembled into twenty-one full length chromosomes, and annotated (Fig.1c). Genome
78 sequences were then aligned to one another and to that of *Lytechinus variegatus*[21], an outgroup
79 representing the ancestral life history condition (Fig.1a).

80 To understand how natural selection altered the genomes of *Heliocidaris* during the
81 evolution of nonfeeding development, we began by testing for evidence of branch-specific
82 positive selection within single-copy protein coding regions[22]. At a genome-wide scale, we found
83 statistical support for modest enrichment of positive selection along the *H. erythrogramma*
84 branch, but not the *H. tuberculata* branch, when considering the full set of genes (Fisher exact

85 test, two-sided: $p<1.33x10^{-3}$). Of note, coding sequences of dGRN genes showed no enrichment
86 of positive selection on either branch (Fig.1d-e). This result provides little support for the idea
87 that changes in transcription factor structure and function are primarily responsible for the
88 extensive modifications in development and life history in *H. erythrogramma*. Scant evidence of
89 positive selection in the coding sequences of dGRN genes likely reflects pleiotropic constraints
90 imposed by the multiple functions that their encoded transcription factors execute during cell
91 type specification and differentiation.

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

107 Strikingly, signals of *H. erythrogramma*-specific positive selection are even more
108 enriched when only considering OCRs near dGRN genes (Fig.1f-g; difference in median zeta:
109 0.182; Fisher exact test, two-sided: p<5.33e-5). In all, 26 putative regulatory elements located
110 near 23 distinct dGRN genes exhibit evidence of positive selection on the *H. erythrogramma*
111 branch, as opposed to just four on the *H. tuberculata* branch (Fig.1c,g). These 23 genes represent
112 17.0% of the total within the defined dGRN with a nearby OCR, a marked enrichment compared
113 with the remainder of the genome, where positive selection is detected in OCRs near just 5.7% of
114 genes (Fisher exact test, two-sided: p<4.92e-4: Fig.2c).

115 **Two distinct regulatory mechanisms underlie divergence in transcriptomes**

116 While the accessibility of most OCRs and expression[19,23] of most genes are conserved
117 between species (see Fig. 2b and Extended Data Fig. 1 for examples), we observed a striking
118 decrease in chromatin accessibility of many putative regulatory elements throughout the *H.*
119 *erythrogramma* genome relative to both species representing the ancestral life history (Fig.2a).
120 Of 2,625 orthologous, differentially accessible OCRs between developmental modes, 1,795 sites
121 (68.4%) are significantly less accessible in *H. erythrogramma* (e.g. Fig.2b: *hesC*). Because
122 decreased chromatin accessibility can limit transcription factor access to regulatory elements and
123 because most regulatory interactions in the early sea urchin embryo involve activation of
124 transcription[24], widespread evolutionary reduction in chromatin accessibility throughout the
125 genome in *H. erythrogramma* embryos suggests an important role for evolutionary changes in
126 chromatin configuration for divergence in gene expression, in this case associated with generally
127 decreased or delayed zygotic transcription for many genes. This interpretation is consistent with
128 indications of a broad delay in embryonic cell fate specification in this species[19,25-29].

129 In a previous study[19] we analyzed changes in temporal gene expression profiles during
130 early development within *Heliocidaris* and found that the largest changes are concentrated on the
131 branch leading to *H. erythrogramma* and are enriched for developmental regulatory genes
132 generally and dGRN genes specifically. Results reported here suggest that these derived
133 expression profiles are the product of two distinct molecular mechanisms that alter transcription
134 factor binding: changes in nucleotide sequence and changes in chromatin configuration (Fig.2c-
135 d). The former may alter protein:DNA binding while the latter may alter protein access to
136 regulatory elements. Both modes of regulatory evolution are concentrated near dGRN genes
137 relative to the rest of the transcriptome in *H. erythrogramma* (Fig.2c). Notably, accelerated
138 sequence evolution or altered chromatin state (or both) is present in an OCR near a differentially
139 expressed dGRN gene ~3 times more frequently than the rest of the transcriptome in *H.*
140 *erythrogramma*, while no such relationship is evident in *H. tuberculata*.
141 The distribution of both mechanisms of regulatory evolution is highly nonrandom within
142 the genome (enriched near differentially expressed genes and near developmental regulatory
143 genes) and phylogenetically (enriched on the *H. erythrogramma* branch). These departures from
144 the null expectation of random distribution (i.e., resulting from genetic drift) suggest that many
145 of the specific changes are adaptive. Adults of the two *Heliocidaris* species occupy overlapping
146 habitats and ranges[30], making the suite of derived life history traits that evolved on the *H.*
147 *erythrogramma* branch the most plausible driver for many of the extensive gene regulatory
148 changes.
149 **The timing of cell fate specification has changed in *H. erythrogramma***
150 To understand how changes in the regulation of gene expression in the *H. erythrogramma*
151 embryo influenced developmental mechanisms and life history traits, we leveraged information

152 about the ancestral dGRN to examine embryonic cell fate specification. The earliest zygotic
153 patterning event in the ancestral state involves specification of skeletogenic and germ cell fates
154 following two successive unequal cleavages of vegetal blastomeres[31]. We focus here on the well-
155 characterized skeletogenic cell lineage, which rapidly establishes a distinct transcriptional state[32]
156 and, within 24 hours after fertilization, undergoes an epithelial-to-mesenchymal transition
157 (EMT), fully differentiates, and begins to synthesize a complex larval endoskeleton (Fig.1b).
158 Specification and maintenance of the skeletogenic cell fate is regulated by interactions between
159 ~11 transcription factors[33] (Fig.4a). These developmental events and most of the underlying
160 dGRN interactions are conserved across >225 my of sea urchin evolution[6].
161 Morphological development of the skeletogenic cell lineage in *H. erythrogramma* differs
162 in several regards from this ancestral state: cleavage divisions are all equal, no cells undergo
163 EMT before gastrulation, and the larval skeleton is delayed and reduced[20] (Fig.1b). To
164 understand whether underlying developmental mechanisms are conserved despite these overt
165 morphological differences, we carried out single-cell RNA-sequencing (scRNA-seq) of early
166 blastula stage *H. erythrogramma* embryos and compared the results with our published scRNA-
167 seq data from *L. variegatus*[34] at the same early blastula stage (prior to EMT). We chose this stage
168 because many major cell fates have been specified by early blastula in the ancestral condition[2,31-
169 33]. This result is clearly reflected in the UMAP of *L. variegatus*, which contains seven cell
170 clusters (Fig. 3a), each expressing a distinct suite of regulatory proteins predicted by the dGRN,
171 with skeletogenic cells exhibiting a particularly disparate transcriptional state (Fig.3c;
172 Supplementary Fig. 1). These indications of early cell fate specification and rapid divergence in
173 transcriptional states are also apparent in scRNA-seq data from *Strongylocentrotus purpuratus*[35],
174 another sea urchin representing the ancestral condition (Fig.1a), suggesting early embryonic

175 regulatory interactions are conserved among planktotrophic species and detectable by scRNA-
176 seq.

177 In *H. erythrogramma*, only three cell clusters are apparent at the same resolution and
178 developmental stage (Fig.3b). The observation of fewer distinct transcriptional states in *H.*
179 *erythrogramma* along with less localized expression of known GRN tissue marker genes
180 (Supplementary Fig. 1) suggests a delayed establishment of distinct transcriptional states in the
181 early embryo of this species – a conclusion not attributable technical factors such as analytical
182 parameterization (see Methods), cell number (Extended Data Fig. 2) or genes/UMIs per cell
183 (Supplementary Fig. 2). Given the limited number of cells in this dataset (2065 post-filtering)
184 and representation from a single stage, future work examining later developmental stages will be
185 necessary to fully resolve the timing of cell fate specification in *H. erythrogramma* and
186 associated expression of developmental regulatory factors.

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

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

209 **The ancient double negative gate within the dGRN is lost in *H. erythrogramma***

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

220 We next examined HesC, a transcription factor that acts even earlier in the dGRN,
221 repressing transcription of *alx1* outside of the skeletogenic cell lineage (Extended Data Fig. 4).
222 Experimentally eliminating HesC protein in the ancestral dGRN produces a dramatic phenotype,
223 with most cells differentiating as skeletogenic because in the absence of *hesC* repression, *alx1* is
224 broadly transcribed[39]. In *H. erythrogramma*, however, we found that embryos develop normally
225 following HesC knock-down (Extended Data Fig. 4e-f), suggesting that it no longer acts as a
226 repressor of *alx1* transcription. This interpretation is consistent with restricted spatial expression
227 of *hesC* in *H. erythrogramma* (Extended Data Fig. 4a-d) that would seem to preclude a broad
228 repressive function for HesC outside of the skeletogenic lineage. However, these experiments
229 cannot rule out co-option of additional developmental functions by *hesC* in *H. erythrogramma* as
230 the assay presented in this study only measured this transcription factor's effect on skeletal
231 mesenchyme differentiation. Future work aimed at validating the loss-of-function phenotype will
232 confirm whether the function of *hesC* is completely lost or acquired novel regulatory roles during
233 *H. erythrogramma* development. Still, altered expression of *hesC* and lack of an overt knock-
234 down phenotype hint at a more profound evolution change within the dGRN.
235 We therefore turned to Pmar1, another transcriptional repressor that interacts with *hesC*
236 to form a double-negative logic gate within the dGRN[39]: throughout most of the embryo HesC
237 directly represses transcription of *alx1* and other genes encoding positive regulators of the
238 skeletogenic cell fate, permitting differentiation of other cell types; in the vegetal-most cells of
239 the embryo, however, *pmar1* is transiently expressed beginning the 16-cell stage where it
240 represses *hesC*, allowing *alx1* transcription and thus specification of the skeletogenic cell fate[40]
241 (Fig.4a).

242 Pmar1 is encoded by a cluster of tandem genes in sea urchins[41]. We identified 10 and 20
243 closely linked *pmar1* paralogues in *L. variegatus* and *H. tuberculata* respectively
244 (Supplementary Table 2). The homeodomain, nuclear localization signal, and two EH1
245 protein:protein interaction domains are typically well conserved, although a few likely
246 pseudogenes are present in each species (Fig.4c, Supplementary Fig. 4). In *H. erythrogramma*
247 we identified 11 *pmar1* paralogues (Supplementary Table 2). Surprisingly, all of these copies
248 contain numerous substitutions, deletions, and/or frameshifts, in many cases altering or
249 eliminating over half of the residues within the homeodomain (Fig.4c), and their expression is
250 barely detectable at the 16-cell and 32-cell stages (Supplementary Fig. 5). In contrast, likely
251 functional orthologs in the other two species differ by 0-3 amino acids out of 60 within the
252 homeodomain. Furthermore, pairwise similarity between *pmar1* orthologs within a species
253 averages greater than 88% for the entire peptide and 93% for the homeodomain in the ancestral
254 state, while *H. erythrogramma* averages just 71.0 % and 45.3%, respectively (Fig.4d). These
255 sequence comparisons indicate that the integrity of the *pmar1* gene family has dramatically
256 decayed in *H. erythrogramma,* raising the question whether these genes with a crucial role in
257 early embryonic patterning have maintained their function in the derived developmental mode.
258 Previous studies demonstrate that microinjecting *pmar1* mRNA into eggs produces a
259 dramatic phenotype, with the resulting widespread overexpression of Pmar1 protein converting
260 most of the embryo to skeletogenic cells[39,42]. Here, we utilized this assay to test the repressive
261 function of specific *pmar1* paralogues. We separately microinjected into *L. variegatus* embryos
262 mRNA encoding one *pmar1* paralog from *L. variegatus* and two from *H. tuberculata*. As
263 expected, these treatments replicated the published phenotype, inducing extensive conversion to
264 the skeletogenic cell fate, confirmed by widespread expression of the larval spicule matrix

265 protein MSP130 (Fig.4e; Supplementary Fig. 7; Supplementary Table 3). We then separately
266 tested the three most intact paralogues of *pmar1* from *H. erythrogramma*. At the same and higher
267 concentrations, none was able to produce the specific or any other discernible phenotype (Fig.4e,
268 Supplementary Fig. 6-7; Supplementary Table 3). These results indicate that the repressive role
269 of Pmar1 is retained in *H. tuberculata* but has been lost in *H. erythrogramma*.

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

283 **DISCUSSION**

284 Prior work showed that the evolution of nonfeeding development in *Heliocidaris* was
285 accompanied by overt changes in oogenesis, cleavage geometry, morphogenesis, and larval
286 morphology, with extensive underlying changes in gene expression[17-20,23]. Whole genome
287 sequence analysis presented here demonstrates that these changes are not merely superficial

288 consequences of amplified maternal provisioning. Although we find evidence for adaptive
289 changes within some coding regions, these are dwarfed by the sheer number and widespread
290 distribution of apparently adaptive changes in the sequences of putative regulatory elements and
291 in the regulation of their chromatin states during early development (Fig.1c-g; 2). Both types of
292 molecular change are strikingly enriched on the branch where nonfeeding development evolved
293 and are over-represented among differentially expressed genes and especially among dGRN
294 genes (Fig.1f-g; 2c-d). While the potential for natural selection to influence trait evolution
295 through changes in gene regulation by altering regulatory element sequence and chromatin
296 accessibility is widely appreciated, we are aware of few cases that illustrate the influence of both
297 so extensively at a genomic scale and during such a short interval.

298 Focusing on transcriptional regulation that patterns the early embryo provides a test of
299 the idea that evolutionary conservation of early development is the product of intrinsic
300 constraints. We examined the earliest zygotic patterning event in the sea urchin embryo, where
301 three transcription factors interact to specify two distinct cell fates and simultaneously establish
302 the primary signaling center of the embryo. There is arguably no set of interactions within the
303 dGRN that is more fundamental to patterning the early sea urchin embryo, and they are
304 conserved among sea urchins that diverged ~225 my ago[5]. Remarkably, however, Pmar1 and
305 HesC, which interact to form a crucial double-negative logic gate[39], have lost their early
306 patterning roles in *H. erythrogramma* (Fig.4b-e, Extended Data Fig. 4). The case of *pmar1* is
307 particularly striking, as it is present as a tandem array of genes; uniquely in the genome of *H.*
308 *erythrogramma*, numerous deletions and point mutations alter about half of the homeodomain in
309 each of 10 the paralogues, rendering their proteins nonfunctional (Fig.4c-e, Supplementary
310 Fig.7).

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

319 **METHODS**

320 **1. Genome Sequencing and Assembly**

321 *1.1 Tissue Collection*

322 *Heliocidaris erythrogramma* (*He*) and *H. tuberculata* (*Ht*) specimens were collected near
323 Sydney Harbor in Sydney, NSW, Australia and housed in natural sea water at the Sydney
324 Institute of Marine Science in Mosman, NSW, AU. The interpyramidal muscle of Aristotle's
325 lantern (the sea urchin's feeding apparatus), tube feet, and the ovarian tissue were dissected from
326 a single female individual, flash-frozen in liquid nitrogen, and stored at -80C until DNA
327 extraction and sequencing.

328 *1.2 Genomic DNA Sequencing*

329 For each species, a 3rd-generation DNA library was sequenced on a PacBio sequel II
330 CLR platform, generating 90.01 (*He*) and 89.47 (*Ht*) Gb of data with an N50 read length of
331 17.24 (*He*) and 23.70 (*Ht*) Kb. DNA from the same individual for each species was also used to
332 construct 10x Genomics linked-reads and Hi-C libraries, which were sequenced on a BGI-SEQ
333 500 platform, generating 194.11 (*He*) and 199.11 (*Ht*) Gb and 130.85 (*He*) and 229.03 Gb (*Ht*)

334 of data, respectively. Jellyfish v2.2.6[45] and GenomeScope v1.0.0[46] were deployed to conduct a k-
335 mer based survey of genome composition using linked-read sequencing data based on 17-mer
336 frequency distribution to estimate the genome size and heterozygosity of both *He* and *Ht* (Figure
337 S4 a-b).

338 *1.3 Genome Assembly*

339 PacBio sequencing data was employed to assemble a *de novo* contig-level genome
340 assembly using Canu v1.8 (minReadLength=1200; minOverlapLength=1000)[47]. Subsequently,
341 HaploMerger2 v3.6[48] was used to create breakpoints in the contigs where potential misjoins have
342 occurred by aligning allelic contigs via Lastz v1.02.00[49]. From these fragmented contigs, the
343 longest of each allelic pair was identified and selected using Redundans v0.14a[50] resulting in a
344 near-haploid level genome assembly. The output of this pipeline was polished using Pilon
345 v1.23[51] with 10X sequencing data to improve assembly quality and accuracy at single base
346 resolution. Lastly, contigs were assembled into scaffolds by mapping Hi-C read pairs to the
347 polished assembly with HiC-Pro[52], resulting in approximately 21.95% (*He*) and 32.10% (*Ht*)
348 valid Hi-C reads pairs. Juicer v1.5[53] and 3D-DNA v180419[54] were used to correct and finalize
349 the construction of chromosome-length scaffolds for each species. See Supplementary Fig. 8 for
350 HiC contact maps.

351 *1.4 Repeat Identification and Classification*

352 Genomic repetitive elements were identified with RepeatModeler v2.01[55] to generate
353 species-specific repeat element libraries for each. Repeat families were filtered via BLASTn
354 v2.3.0[56] for significant hits to gene models of the well-studied sea urchin *Strongylocentrotus*
355 *purpuratus* (www.echinobase.org) to prevent unintentional masking of genic regions. Repeats
356 were masked from the genome of each species with RepeatMasker v4.1.1 using the most

357 sensitive setting (-s) to identify the location of repetitive elements. Long-terminal repeats were
358 also secondarily identified with LTR_Finder v1.0.7. Outputs of both RepeatMasker and
359 LTR_Finder were then input into RepeatCraft v1.0[57] under default parameters to improve repeat
360 element annotation and identification, resulting in a final genome annotation of repetitive
361 elements. Lastly, repeats were broadly classified into functional categories described by their
362 mode of transposition using TEclass[58].

363 *1.5 Gene Annotation and Prediction Strategy*

364 Previously published paired-end RNA-seq reads from six developmental stages for each
365 *Heliocidaris* species[19] were trimmed using Trimmomatic v0.39[59] (TruSeq3-PE.fa: 2:30:10;
366 leading: 3; trailing: 3; slidingwindow: 4:15; minlen: 36) and properly paired reads were mapped
367 to their respective genomes using STAR v2.7.2. For each species, these RNA-seq alignments as
368 well as protein models of the *S. purpuratus* v5.0 genome[60] were input into BRAKER2[61] (--
369 etpmode). This program utilizes a number of additional software as a part of its pipeline
370 including Augustus[62], Genmark-EP+[63], Genemark-ET[64], DIAMOND[65], and Samtools[66]. Gene
371 models from the BRAKER output for filtered for transposable elements by aligning to a
372 combined database of transposable element sequences from the MAKER gene annotations
373 pipeline[67] and the Dfam v3.3 transposable element database[68] using BLAST-P[56]. Lastly, gene
374 models were improved using the PASA pipeline[69] by supplementing pre-existing gene models
375 with a de-novo transcriptome retrieved from ref. 19. These gene models were annotated by
376 aligning peptide sequences to three separate databases using BLAST-P v2.3.0[56] : 1) *S.*
377 *purpuratus* v4.2 gene models; 2) UniProt KnowledgeBase SwissProt protein models[70]; 3)
378 RefSeq invertebrate protein models with *S. purpuratus* excluded (e-value cutoff: 1e-5)[71]
379 (Supplementary Data 8). The list of sea urchin gene regulatory network (GRN) genes are

380 provided in Supplementary Data 1, retrieved from the Institute of Systems Biology
381 (www.biotapestry.org: accessed June 27, 2017).

382 **2. Whole Genome Alignment**

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

395 **3. ATAC-seq**

396 *3.1 Sample Preparation*

397 For each sea urchin species (*H. erythrogramma, H. tuberculata*, and *Lytechinus*
398 *variegatus*), adult animals were induced to spawn via injection of 0.5 M KCl solution into the
399 coelom. For each species, three unique male-female pairs were crossed to produce three
400 biologically-independent replicates of sea urchin embryos. Each culture was reared in large glass
401 dishes supplied with 20 mm filtered sea water (FSW) that was changed every six hours. Because
402 these species exhibit different developmental rates, a conspicuous developmental milestone,

403 shedding of the fertilization envelope at hatched blastula-stage, was selected to maximize
404 developmental synchrony within cultures and across species for comparison. Once a culture
405 reached the blastula stage, live embryos were collected and processed immediately for nuclei
406 preparation and transposase treatment as a part of the ATAC-seq protocol.

407 *3.2 ATAC-seq Protocol and Sequencing*

408 ATAC sample preparation was carried out according to the Omni-ATAC-seq protocol[73].
409 For each replicate, embryos were washed once in 1 mm FSW, lysed, and 50,000 nuclei were
410 isolated for the transposition reaction as described in the Omni-ATAC-seq protocol using the
411 Illumina TDE1 enzyme and tagmentation (TD) buffer (Cat. No. 20034197 and 20034198) (San
412 Diego, CA, USA). Sequencing libraries for each replicate were generated via qPCR and
413 sequencing libraries were purified and size selected using Ampure XP Beads at a 1.8:1 bead
414 volume: library volume (Beckman Coulter, Brea, CA, USA). Library quality and transposition
415 efficiency was accessed using a Fragment Analyzer and PROSize 2.0 (Agilent). *H.*
416 *erythrogramma* and *L. variegatus* libraries were sequenced on an Illumina HiSeq 4000
417 instrument using 50 bp SE sequencing at an average of 41.9 million and 37.3 million reads per
418 sample, respectively. *H. tuberculata* libraries were sequenced on an Illumina NovaSeq 6000
419 instrument using 50 bp PE sequencing (only SE were used for data analysis) at an average of
420 31.4 million reads per sample.

421 *3.3 ATAC-seq Data Analysis*

422 Raw ATAC-seq reads were trimmed for quality and sequencing adapters using cutadapt[74]
423 v2.3 with the following parameters: -a CTGTCTCTTATACACATCT -q 20 --trim-n -m 40.
424 Trimmed reads were then aligned to each species' respective genome using stampy[75] v.1.0.28

425 using the "--sensitive" set of parameters. ATAC-seq alignments were filtered for mitochondrial
426 sequences and required an alignment quality score of at least 5 using samtools v1.9[66].
427 In this study, we aimed to compare the evolution of orthologous non-coding sites. To
428 accomplish this, we performed a series of liftOvers[76] to convert ATAC-seq alignments between
429 genomic coordinates of each sea urchin species (see previous section for description of genome
430 alignments). We took an iterative, reciprocal liftOver strategy described below to minimize
431 possible reference bias associated with converting between genome assemblies: 1) *H.*
432 *erythrogramma*: *He* → *Lv* → *He*; 2) *H. tuberculata*: *Ht* → *Lv* → *He*; 3) *L. variegatus*: *Lv* → *He*
433 *Lv* → *He*. After filtering and coordinate conversion, all ATAC-seq alignments were referenced
434 to the *H. erythrogramma* genome with an average of 5.9 million alignments per sample to
435 orthologous genomic loci.
436 Following filtering and coordinate conversion, peaks were called from these alignments
437 using the MACS2 v2.1.2[77] callpeak function (parameters: –nomodel, –keep-dup=auto, –shift
438 100, –extsize 200) for each species separately. Peak coordinates were merged using the
439 bedtools[78] v2.25 *merge* function requiring a peak overlap of at least 200 bp to be merged into a
440 single peak. Lastly, for each sample, accessibility of each peak was measured with the bedtools[78]
441 v2.25 *multiBamCov* function.
442 *3.4 Tests for positive selection within OCRs*
443 In order to test for evidence of positive selection, a neutral genomic reference across all
444 species was assembled. To do this, the genome was first masked repetitive elements, coding
445 sequence, untranslated genic sequence, non-coding RNAs (including microRNAs, ribosomal
446 RNAs, small nuclear RNAs, and transfer RNAs) and ATAC-seq open chromatin regions (OCRs)
447 (see below) in the genome. The remaining, putatively neutrally evolving genome was then

448 divided into 300 bp windows, orthologous regions retrieved from each species' genome, and
449 filtered using the *filtering.py* and *pruning.py* scripts of the "adaptiphy"[79] program
450 (https://github.com/wodanaz/adaptiPhy). Next, branch lengths of each of these neutral sites was
451 estimated using phyloFit[80] (--subst-mod HKY85), highly-conserved sites removed (Figure S5),
452 relative branch lengths calculated, and sites falling within the middle 50% of relative branch
453 lengths in the *H. erythrogramma* genome were selected as the neutral reference (88,004 sites;
454 Supplementary Data 2, Supplementary Fig. 9).

455 To measure branch-specific signatures of positive selection in the non-coding genome,
456 the adaptiPhy[79] pipeline (https://github.com/wodanaz/adaptiPhy) for global tests of natural
457 selection was followed. First, orthologous sequences for non-coding sites of interest were
458 selected from each species' genome into FASTA format. Sequences were trimmed to include
459 only contiguous DNA sequence using the *prunning.py* script and filtered using the *filtering.py*
460 script, requiring a minimum alignment length of 75 bases. These trimmed and filtered alignments
461 serve as "query" sequences of tests for selection. To generate a neutral reference for
462 comparison, ten neutral sites were randomly selected (see above) and concatenated into a single
463 neutral reference sequence. In addition, for each OCR, tests for positive selection were repeated
464 10 times against a unique putatively neutral reference. For each query site replicate, substitution
465 rates of both the query and randomly concatenated neutral reference were estimated using
466 phyloFit[80], and the zeta score was calculated as the ratio of the query substitution rate to the
467 neutral reference substitution rate. In addition, p-values of likelihood ratio tests for significant
468 levels of branch-specific positive selection were calculated with adaptiPhy[79] pipeline using
469 HyPhy[81]. P-values and substitution rates for all query and neutral sites were then imported in R
470 v4.0.2 for analysis (Supplementary Data 3).

471 *3.5 ATAC-seq Peak Filtering*

472 After accessibility and rates of selection were calculated for each ATAC-seq peak, herein
473 referred to as an "open chromatin region" (OCR), a series of filtering and quality control metrics
474 were carried out to ensure only high confidence and quality peaks were compared between
475 species. These filtering steps are as follows: 1) each OCR is required to have at least 75bp of
476 contiguous, single copy sequence (see *Section 3.4*) for accurate estimations of selection; 2) for
477 each species, a local composition complexity (LCC)[82] value of 1.9 or more was required for the
478 OCR to remove repetitive or other low-complexity sequences that may generate inaccurate
479 estimations of selection (module: biopython.org/docs/1.75/api/Bio.SeqUtils.lcc.html); 3) a CPM
480 of 3 or more was required in at least 2 (of the 9) samples to remove OCRs with extremely low
481 accessibility; 4) the midpoint of the OCR must lie within 25kb (in either direction) of the
482 translational start site of a gene model; 5) the gene nearest to an OCR must be the same gene in
483 each of the species' genomes —in other words, for each OCR and its nearest gene in the *H.*
484 *erythrogramma*, the orthologous region in the *H. tuberculata* and *L. variegatus* genome must
485 also be closest to a gene model that is orthologous (determined by annotation) to the same gene
486 in the *H. erythrogramma* genome. Given nearly no prior knowledge is known of the cis-
487 regulatory landscape for these sea urchin species, these stringent filtering methods were carried
488 out in order to maximize confidence in comparisons of non-coding sequence evolution and
489 function. This method resulted in a final set of 27,322 high-confidence OCRs for cross-species
490 analysis (Supplementary Data 4).

491 *3.6 ATAC-seq Statistical Analysis*

492 Raw counts of the filtered OCRs were loaded into DESeq2[83] v1.30 to calculate
493 differential accessibility between sample groups. For life history strategy comparisons, *H.*

494 *tuberculata* and *L. variegatus* were treated as a single group. Differentially accessible sites were
495 classified as having a 2-fold accessibility difference between sample groups and supported by an
496 FDR of 10%. Significant levels of positive selection were classified as having a median zeta
497 value greater than 1.5 and supported by a median false-discovery rate less than 10% across 10
498 replicates for each query site. Branch-specific evidence of positive selection met these criteria for
499 one species, but failed to meet these criteria in the other, as evidenced by a zeta score < 1.5.

500 **4. Coding Selection Analyses**

501 In order to make tests for positive selection in coding sequences analogous to non-coding
502 sequences, only single-copy orthogroups were considered in these analyses. Single-copy
503 orthologs between *H. erythrogramma*, *H. tuberculata*, *L. variegatus*, and *Echinometra lucunter*
504 were identified using OrthoFinder[84]. Evidence of episodic positive selection was queried on both
505 the *H. erythrogramma* branch and *H. tuberculata* branch under default parameters using
506 BUSTED[22], by specifying either branch as the "foreground" branch. P-values from these
507 analyses are available in Supplementary Data 5. Genes with significant evidence of episodic
508 positive selection were supported by a p-value <= 10% by the likelihood ratio test.

509 **5. Bulk RNA-sequencing Analysis**

510 Raw RNA-seq reads from blastula stage embryos of *He*, *Ht*, and *Lv* were retrieved from
511 ref 19, trimmed and filtered for low quality bases and reads with Trimmomatic[59], and aligned to
512 each species respective genomes and gene models with STAR[85]. From these alignments, mRNA
513 expression was estimated with Salmon[86] and loaded to R for statistical analysis. Read counts for
514 summed to each gene's best match to the *S. purpuratus* v4.2 gene models to generate a common
515 reference for expression comparisons between species as described in ref 19. Differentially
516 expressed genes between life histories were called as having a fold-change in expression > 2 and

517 supported by an FDR of 10% or less between *He* and both planktotrophic species (in the same
518 direction), and not DE between *Ht* and *Lv* (Supplementary Data 6).

519 **6. Single-cell RNA-Sequencing**

520 *6.1 H. erythrogramma embryo culturing*

521 Female H. erythrogramma individuals were spawned via intracoelomic injection of 0.5ml
522 of 0.5M KCl. Unfertilized eggs were washed three times in 100um filtered natural sea water
523 (FSW). Eggs were fertilized by 2ul of concentrated sperm in .02g Para-Amino Benzoic Acid
524 (PABA)/100 ml FSW. Following fertilization, eggs were washed three additional times in FSW
525 to remove residual sperm and PABA. Fertilized embryos were then cultured at 22-23 °C. At 6
526 hours post fertilization (hpf) embryos were sampled for microscopy and dissociation, then fixed
527 for scRNA-seq.

528 *6.2 Embryo Dissociation and Fixation*

529 Once embryos developed to the early blastula stage (pre-skeletogenic cell ingression), a
530 portion of the co-culture was taken, and washed one time in Calcium-Free Artificial SeaWater
531 (CFASW). After washing embryos with CFASW, 3ml of embryos were added to 7ml of
532 dissociation buffer made (1.0M Glycine and 0.25mM EDTA, pH 8.0) at 4 °C, and gently rocked
533 on a rocker for 10 minutes. Following incubation, embryos were gently triturated 15-20 times to
534 increase disassociation, then 10ml ice cold 100% methanol was added, and cells were incubated
535 for 10 minutes and on the rocker. Following incubation, cells were triturated again 15-20x times,
536 and then another 30ml of ice cold 100% methanol was added to bring the suspension to a final
537 concentration of 80% methanol. This 80% methanol suspension of cells was incubated at 4 °C
538 for 1 hour. Following this last fixation step, cells were stored at -20 °C until library preparation.

539 *6.3 Rehydration of Single Cells, Library Preparation, and Sequencing*

540 Cells were centrifuged at 50xg, supernatant was discarded, and fixed cells were washed
541 twice and rehydrated in a Sigma 3x Saline Sodium Citrate (SSC) buffer (SKU SRE0068) before
542 cell count and library preparation. Cell concentration was estimated with a hemocytometer and
543 volume was adjusted to a final concentration of ~300 cells/ uL. Single cell libraries were
544 prepared using the 10x Genomics 3' v3 gene expression kit and the 10x Chromium platform to
545 encapsulate single cells within droplets. Library quality was verified using the Agilent 2100
546 Bioanalyzer. In total, ~ 3960 cells were loaded onto the 10X instrument. For the single library
547 preparation, 2500 cells were targeted, of which 2066 were successfully captured for sequencing.
548 Libraries were sequenced by the Duke Genomics and Computational Biology Core facility on
549 two NovaSeq6000 S1 flow cells with 28 x 8 x 91 bp sequencing performed.
550 *6.4 FastQ Generation, Genome Indexing, and Quantification* of scRNA-seq
551 Following sequencing, we used Cellranger v3.1.0 to convert Illumina-generated BCL
552 files to fastq files using the Cellranger "mkfastq" command. scRNA-seq data for early blastula
553 stage embryos (pre-skeletogenic cell ingression) of L*ytechinus variegatus* were retrieved from a
554 published scRNA-seq developmental time course of the species[34]. We then applied the "mkref"
555 command to index the most recent Lv3.0 genome[21] (for the *Lytechinus* data) and the *H.*
556 *erythrogramma* genome assembled in this study. The "count" command was used to demultiplex
557 and count reads mapping to the respective reference *Lv* (53.3% mapping rate) or *He* (94.1%
558 mapping rate) genome. The "mat2csv" command was used to obtain CSV RNA count matrix
559 files for each sample for further downstream analysis. Orthofinder v2.3.12[84] was implemented to
560 identify putative 1-1 orthologous gene models between *Lv* and *He* (Supplementary Data 10).
561 *6.5 scRNA computational analyses*

562 We employed a dual strategy for comparing scRNA-seq expression between *He* and *Lv*:
563 1) A non-integrated analysis in which scRNA-seq from each species was quantified against its
564 own gene models (Fig. 3a) and 2) an integrated analysis in which orthologous anchor genes were
565 used to identify cell types with overlapping expression profiles between each species (Fig. 3d).
566 CSV RNA count matrix files were uploaded to R and a Seurat object was generated for 1) each
567 species separately quantified against their own gene models and 2) orthologous genes between
568 *H. erythrogramma* and *L. variegatus* (Supplementary Data 7). Each dataset was filtered to
569 remove low quality cells with nFeature_RNA > 200, nFeature_RNA < 5500, and nCount_RNA
570 < 7500. To ensure differences in input cell number did not bias detection of different cell types
571 between samples, the *Lv* dataset was separately subsampled to 2065 randomly-selected cells and
572 clustering analyses were repeated (see Extended Data Fig. 2). Furthermore, both datasets had
573 comparable distributions of genes per cell and UMIs per cell numbers (Supplementary Fig. 2).
574 For the non-integrated analyses, "SCTransform", a regularized negative binomial
575 regression method that stabilizes variance across samples, was applied to perform normalization
576 and removal of technical variation[87], while preserving biological variation. We next performed
577 Principal Component Analysis on the SCTransformed Seurat object file of raw gene expression
578 counts, and found the nearest neighbors using 10 PC dimensions of variable gene space. UMAP
579 (Uniform Manifold Approximation and Projection)[88] was applied to multi-dimensional scRNA-
580 seq data to visualize the cells in a two-dimensional space. Finally, clustering was performed
581 using graph-based Louvain Clustering with resolution, res=0.5, resulting in 7 clusters in *L.*
582 *variegatus* and 3 clusters in *H. erythrogramma*. The clusters were putatively annotated using
583 dGRN genes and published in situ hybridization patterns as markers (Supplementary Fig. 1,

584 Supplementary Data 11), and ambiguous cluster identities are conservatively denoted as broad
585 embryological territories (e.g. endomesoderm, ectoderm A-C) or as "unknown".
586 To perform the integrated analyses of scRNA-seq data between species, we carried out 3
587 independent methods for scRNA-seq data integration to compare expression of the same
588 orthologous genes between *H. erythrogramma* and *L. variegatus* and identify putative
589 overlapping cell types: 1) Canonical Correlation Analysis (CCA)[89], 2) Reciprocal Principal
590 Component Analysis (RPCA) (satijalab.org/seurat/articles/integration_rpca.html), and 3)
591 Harmony[90] (Fig. 3d; Extended Data Fig. 3). For each scRNA-seq strategy (non-integrated and
592 CCA), counts for all gene models and orthologous sets of genes are included as Supplementary
593 Data 7, and top marker genes for each cluster are available in Supplementary Data 9.
594 **7. *H. erythrogramma* microinjection and in-situ hybridization**
595 *7.1 MASO design and microinjection*
596 Morpholino antisense oligos (MASOs) were constructed to target the translation start site
597 of *alx1* (ATCAATTCGGAGTTAAGTCTCGGCA) and *hesC*
598 (ATCCAGATGTGTTAAGCATGGTTGC) and synthesized by Gene Tools (Philomath, OR,
599 USA). Control morpholinos included a standard negative control morpholino recommended by
600 the manufacturer (CCTCTTACCTCAGTTACAATTTATA) and a scrambled morpholino for
601 HesC (ATCGACATCTGTTAACCATCGTTGC). Fertilized eggs of *H. erythrogramma* were
602 injected as described in ref. 91 at a concentration of 100 μM and 200 μM for *alx1* and 500 μM
603 for *hesC*, then reared at 22 °C in pasteurized, 0.22-micron filtered seawater + penicillin (100
604 unit/mL) and streptomycin sulfate (0.1 mg/mL) (Sigma P4333A). Injected embryos were
605 checked for developmental abnormalities and mortality every six hours until fixation.
606 *7.2 Fixation and ISH*

607 *H. erythrogramma* embryos were fixed for in-situ hybridization (ISH) for ~16 hrs
608 overnight at 4 °C in 4% paraformaldehyde (Sigma 158127) + 20 mM EPPS (Sigma E1894) in
609 FSW, washed 3x in FSW, and dehydrated step-wise into 100% MeOH and stored at -20 °C. The
610 full-length *HesC* coding sequence was synthesized in vitro by GenScript (Piscataway, NJ, USA)
611 and subcloned (NCBI insert number: MK749159), and used as template to make antisense RNA
612 probes for ISH. ISH of *H. erythrogramma* was performed according to previously published
613 methods[44]. Hybridizations were carried out at 65 °C and stringency washed at 0.1% SSC.

614 *7.3 Imaging*

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

620 **8. Pmar1 mRNA overexpression assays**

621 *8.1 mRNA Synthesis*

622 Sequences of *pmar1* orthologs were retrieved from each species respective genome
623 annotations (*Lv*[21], *He* and *Ht*, this study) (Supplementary Table 2). One *Lv* ortholog
624 (LVA_25833.t1) was selected for overexpression assays as it represents the ortholog tested in
625 previous overexpression assays[42], while two *Ht* (HTU_11636.t1 and HTU_11625.t1) and three
626 *He* (HER_770.t1, HER_761.t1, and HER_775.t1) were selected for overexpression assays as
627 they represent orthologs with the highest identity to the species' consensus sequence, and
628 therefore predicted as genes most likely to be functional. Construct templates for each ortholog

629 were ordered from Twist Biosciences (San Francisco, CA, USA) and mRNA was synthesized
630 from these constructs with a ThermoFisher MEGAshortscript T7 Transcript Kit (AM1354).

631 *8.2 Overexpression experiments*

632 Female *L. variegatus* individuals were spawned via intracoelomic injection of 0.5M KCl,
633 washed in FSW, and fertilized with 1 uL of concentrated sperm in 100 ml FSW. For each
634 construct, at least 4 rounds of microinjections were conducted with each round including 30-50
635 healthy embryos. *Lv* constructs were injected at a concentration of 250ng/µL. *Ht* constructs were
636 injected at a concentration of 1200ng/µL, and *He* constructs were injected at 1500ng/µL. Higher
637 concentrations of *Heliocidaris* constructs were used to reproducibly obtain the skeletal cell
638 conversion phenotype, and reduced sensitivity of these assays may be attributable to less optimal
639 cross-species interactions of Pmar1 in regulating the *L. variegatus* genome. Following injection,
640 embryos were incubated at 23 °C and imaged at 24 hpf.

641 *8.3 Imaging and Immunostaining*

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

650 **9. Data Availability**

651 *9.1 Genomes*

652 Sequencing reads used to assemble the *Heliocidaris* genomes and the genome assemblies
653 themselves are available on the Chinese National GeneBank (CNP0002233). Genome assemblies
654 of *Heliocidaris erythrogramma* and *Heliocidaris tuberculata* are also available on NCBI
655 (PRJNA827916 and PRJNA827769, respectively). Genome assembly of *Lytechinus variegatus* is
656 previously published[21] and available on NCBI (PRJNA657258). Genome annotations and files
657 associated with whole genome alignments between species are available on Dryad
658 (https://doi.org/10.5061/dryad.sj3tx966v).

659 *9.2 ATAC-seq*

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

663 *9.3 Bulk RNA-seq*

664 Bulk RNA-seq data was retrieved from Israel et al, 2016[19].

665 *9.4 Single cell RNA-seq*

666 Raw sequencing reads for the *Heliocidaris erythrogramma* single cell ATAC-seq dataset are
667 available on NCBI (PRJNA833141). Sequencing reads from the *Lytechinus variegatus* single
668 cell RNA-seq dataset were retrieved from Massri et al, 2021[34] and available on NCBI
669 (PRJNA765003). Results files associated with the single cell RNA-seq analyses are available as
670 Supplementary Data 7,9.

671

672 **AUTHOR CONTRIBUTIONS**

673 LW, MB, GF, and GAW conceived and designed the study. LW, DK, and PC collected tissues
674 for genomic sequencing. YZ performed genomic DNA extraction and sequencing library

675 preparations. PLD, HG, and HZ assembled the genomes. PLD performed genome annotation,
676 alignments, and data analysis. PLD collected, prepared, and analyzed ATAC-seq libraries. AJM
677 collected, prepared, and analyzed single cell RNA-seq libraries. JSS and AE performed
678 embryonic expression and injection assays. PLD and GAW wrote the manuscript and all authors
679 contributed to manuscript revisions.

680 **COMPETING INTERESTS**

681 There are no competing interests to declare.

682 **ACKNOWLEDGEMENTS**

683 We thank the Sydney Institute of Marine Science (SIMS) for facilities, as well as the SIMS staff
684 for their assistance. This work was supported by the National Science Foundation Division of
685 Integrative Organismal Systems (award no. 1929934 to GAW) and National Science Foundation
686 Graduate Research Fellowships to HRD and AJM.

687 **MATERIALS & CORRESPONDENCE**.

688 Correspondence and material requests regarding this work should be addressed to Dr. Gregory
689 Wray at gwray@duke.edu .

690 **REFERENCES**

691 1 Erwin, D. H. & Davidson, E. H. The evolution of hierarchical gene regulatory networks.
692 *Nat Rev Genet* **10**, 141-148, doi:10.1038/nrg2499 (2009).
693 2 Oliveri, P. & Davidson, E. H. Gene regulatory network controlling embryonic
694 specification in the sea urchin. *Curr Opin Genet Dev* **14**, 351-360,
695 doi:10.1016/j.gde.2004.06.004 (2004).
696 3 McClay, D. R. Evolutionary crossroads in developmental biology: sea urchins.
697 *Development* **138**, 2639-2648, doi:10.1242/dev.048967 (2011).
698 4 Gildor, T. & Ben-Tabou de-Leon, S. Comparative study of regulatory circuits in two sea
699 urchin species reveals tight control of timing and high conservation of expression
700 dynamics. *PLoS Genet* **11**, e1005435, doi:10.1371/journal.pgen.1005435 (2015).
701 5 Yamazaki, A. & Minokawa, T. Expression patterns of mesenchyme specification genes
702 in two distantly related echinoids, Glyptocidaris crenularis and Echinocardium cordatum.
703 *Gene Expr Patterns* **17**, 87-97, doi:10.1016/j.gep.2015.03.003 (2015).
704 6 Erkenbrack, E. M., Davidson, E. H. & Peter, I. S. Conserved regulatory state expression

controlled by divergent developmental gene regulatory networks in echinoids. *Development* **145**, doi: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 U S A* **100**, 13356-13361, doi:10.1073/pnas.2235868100 (2003).

8 Koch, N. M. *et al.* Phylogenomic analyses of echinoid diversification prompt a re-evaluation of their fossil record. *bioRxiv*, 2021.2007.2019.453013, doi:10.1101/2021.07.19.453013 (2021).

9 Maynard Smith, J. *et al*. "Developmental constraints and evolution: a perspective from the Mountain Lake conference on development and evolution." The Quarterly Review of Biology 60.3: 265-287 (1985).

10 Arthur, W. A theory of the evolution of development. John Wiley & Sons Incorporated. (1988).

11 Raff, R. A. Constraint, flexibility, and phylogenetic history in the evolution of direct development in sea urchins. *Developmental Biology* **119**, 6-19, doi:Doi 10.1016/0012-1606(87)90201-6 (1987).

12 Wray, G. A. Parallel evolution of nonfeeding larvae in echinoids. *Syst Biol* **45**, 308-322, doi:Doi 10.2307/2413566 (1996).

13 McEdward, L. R. & Miner, B. G. Larval and life-cycle patterns in echinoderms. *Can J Zool* **79**, 1125-1170, doi:DOI 10.1139/cjz-79-7-1125 (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-+, doi:10.1146/annurev-ecolsys-102710-145004 (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, doi:10.1007/Bf01344355 (1991).

16 Morgan, S. G. Life and death in the plankton: larval mortality and adaptation. *Ecology of marine invertebrate larvae*, 279-321 (2020).

17 Raff, R. A. Direct-developing sea urchins and the evolutionary reorganization of early development. *Bioessays* **14**, 211-218, doi:10.1002/bies.950140403 (1992).

18 Raff, R. A. & Byrne, M. The active evolutionary lives of echinoderm larvae. *Heredity* **97**, 244-252, doi:10.1038/sj.hdy.6800866 (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, doi:10.1371/journal.pbio.1002391 (2016).

20 Wray, G. A. Extreme phenotypic divergence and the evolution of development. *Current Topics in Developmental Biology* 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, doi:10.1093/gbe/evaa101 (2020).

22 Murrell, B. *et al.* Gene-wide identification of episodic selection. *Mol Biol Evol* **32**, 1365-1371, doi:10.1093/molbev/msv035 (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, doi:10.1038/s41559-020-1165-y (2020).

24 Davidson, P. L., Byrne, M. & Wray, G. A. Evolutionary changes in the chromatin landscape reshape a developmental gene regulatory network during rapid life history divergence in sea urchins. *bioRxiv* 2022.02.22.481523 (2022).
25 Wray, G. A. & Raff, R. A. Evolutionary modification of cell lineage in the direct-developing sea urchin *Heliocidaris erythrogramma*. *Developmental Biology* **132**, 458-470, doi:10.1016/0012-1606(89)90242-X (1989).
26 Wray, G. A. & Raff, R. A. Novel origins of lineage founder cells in the direct-developing sea urchin *Heliocidaris erythrogramma*. *Developmental Biology* **141**, 41-54, doi:10.1016/0012-1606(90)90100-W (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, doi:10.1002/jez.b.21233 (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, doi:10.1111/j.1525-142X.2005.05045.x (2005).
29 Love, A. C. & Raff, R. A. Larval ectoderm, organizational homology, and the origins of evolutionary novelty. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **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 U S A* **105**, 5955-5962, doi:10.1073/pnas.0711220105 (2008).
34 Massri, A. J. *et al.* Developmental single-cell transcriptomics in the *Lytechinus variegatus* sea urchin embryo. *Development* **148**, doi: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, doi:10.1242/dev.044149 (2010).
38 Emlet, R. B. Larval spicules, cilia, and symmetry as remnants of indirect development in the direct developing sea urchin *Heliocidaris erythrogramma*. *Developmental Biology* **167**, 405-415, doi: 10.1006/dbio.1995.1037 (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. *Proceedings of the National Academy of Sciences* **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* **147**, doi: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, doi:10.1242/dev.02805 (2007).

43 Smith, M.S., Turner, F.R. and 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. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **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, doi:10.1186/s12861-017-0145-1 (2017).

45 Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-770, doi:10.1093/bioinformatics/btr011 (2011).

46 Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202-2204, doi:10.1093/bioinformatics/btx153 (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, doi:10.1101/gr.215087.116 (2017).

48 Huang, S., Kang, M. & Xu, A. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics* **33**, 2577-2579, doi:10.1093/bioinformatics/btx220 (2017).

49 Harris, R. S. Improved pairwise Alignment of genomic DNA. The Pennsyvania State University (2007).

50 Pryszcz, L. P. & Gabaldon, T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res* **44**, e113, doi:10.1093/nar/gkw294 (2016).

51 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963, doi:10.1371/journal.pone.0112963 (2014).

52 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (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, doi:10.1016/j.cels.2016.07.002 (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, doi:10.1126/science.aal3327 (2017).

55 Smit, A. F. & Hubley, R. RepeatModeler Open-1.0. (2010).

56 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421, doi:10.1186/1471-2105-10-421 (2009).

57 Wong, W. Y. & Simakov, O. RepeatCraft: a meta-pipeline for repetitive element de-fragmentation and annotation. *Bioinformatics* **35**, 1051-1052, doi:10.1093/bioinformatics/bty745 (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, doi:10.1093/bioinformatics/btp084 (2009).

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

60 Arshinoff, Bradley 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 research* 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, doi:10.1093/nargab/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, doi:10.1186/1471-2105-7-62 (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, doi:10.1093/nargab/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, doi:10.1093/nar/gku557 (2014).

65 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59-60, doi:10.1038/nmeth.3176 (2015).

66 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

67 Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome Annotation and Curation Using MAKER and MAKER-P. *Curr Protoc Bioinformatics* **48**, 4 11 11-39, doi:10.1002/0471250953.bi0411s48 (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, doi:10.1186/s13100-020-00230-y (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, doi:10.1186/gb-2008-9-1-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**, D733-745, doi:10.1093/nar/gkv1189 (2016).

72 Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* **14**, 708-715, doi:10.1101/gr.1933104 (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, doi:10.1038/nmeth.4396 (2017).

74 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10-12 (2011).

75 Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome research* **21**, 936-939, doi:10.1101/gr.111120.110 (2011).

76 Hinrichs, A. S. *et al.* The UCSC genome browser database: update 2006. *Nucleic acids research* **34**, D590-D598 (2006).

77 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).

78 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (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, doi:10.1186/s12864-020-6752-4 (2020).

80 Siepel, A. & Haussler, D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol Biol Evol* **21**, 468-488, doi:10.1093/molbev/msh039 (2004).

81 Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676-679, doi:10.1093/bioinformatics/bti079 (2005).

82 Konopka, A. K. Sequence complexity and composition. *eLS* (2005). http://www.els.net/WileyCDA/ElsArticle/refId-a0005260.html

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, doi:10.1186/s13059-014-0550-8 (2014).

84 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238, doi:10.1186/s13059-019-1832-y (2019).

85 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (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, doi:10.1038/nmeth.4197 (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, doi:10.1186/s13059-019-1874-1 (2019).

88 Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*, doi:10.1038/nbt.4314 (2018).

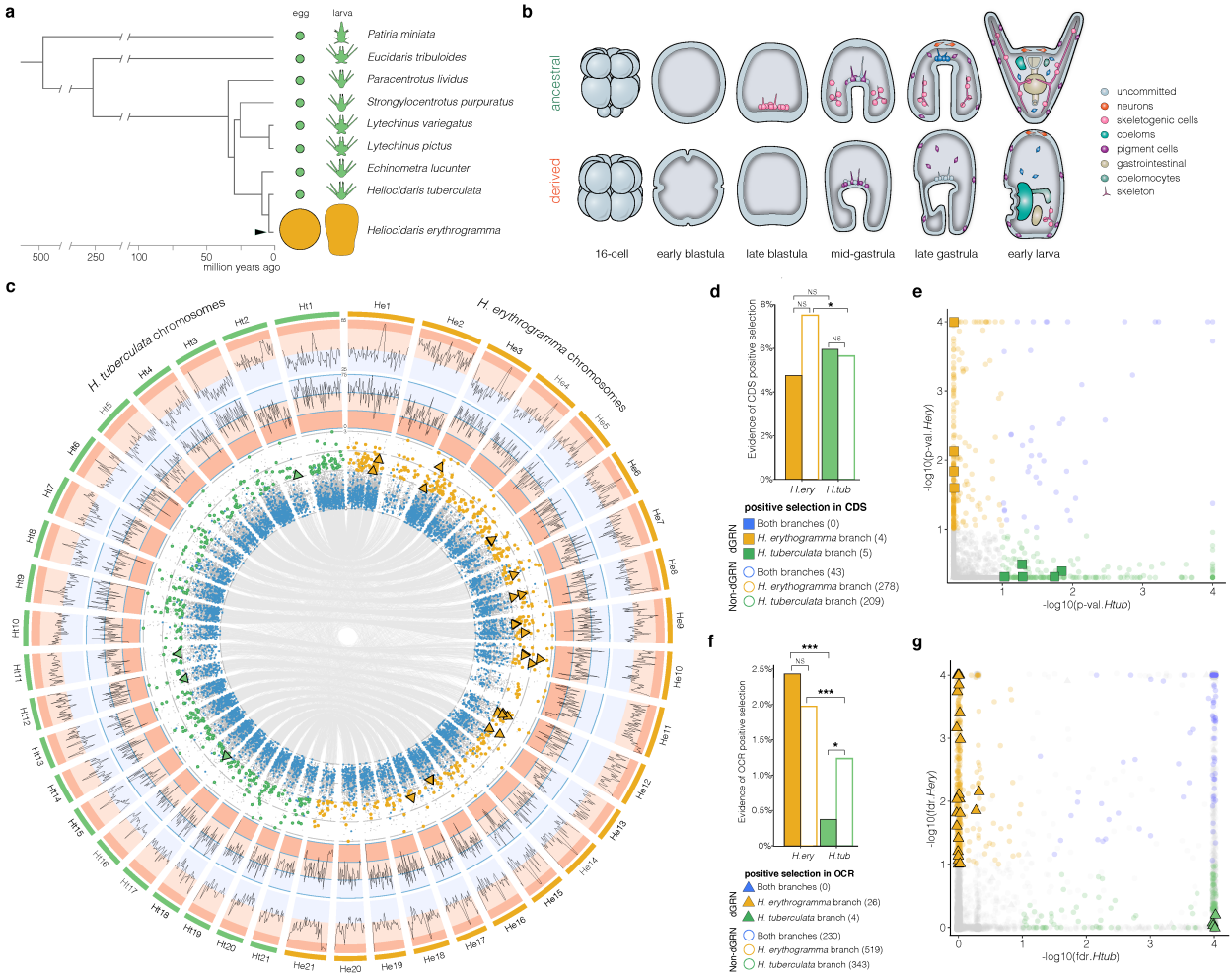
89 Butler, Andrew, *et al*. "Integrating single-cell transcriptomic data across different conditions, technologies, and species." *Nature Biotechnology* **36.5**: 411-420 (2018).

90 Korsunsky, I, *et al*. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature 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, doi:10.14440/jbm.2019.292 (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, doi:10.1242/dev.058396 (2011).

934 **FIGURE LEGENDS**

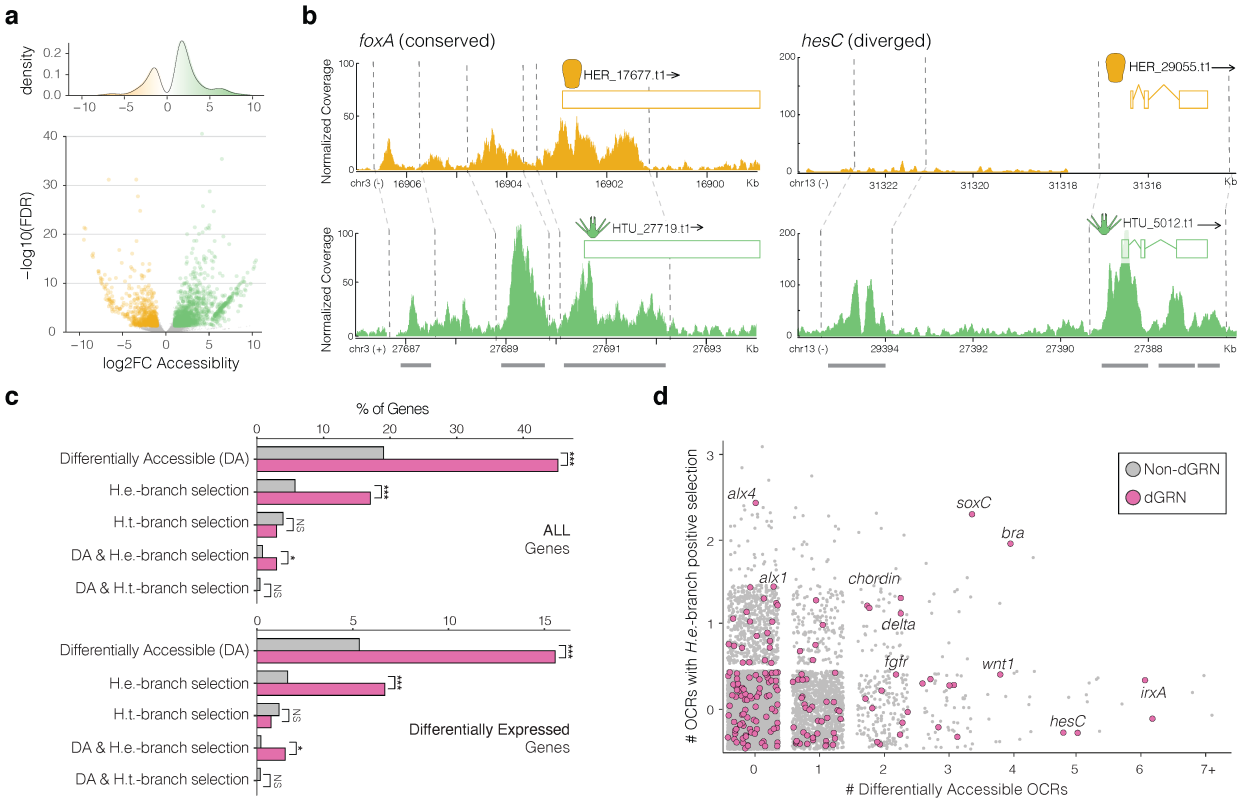


935

936 **Figure 1. Evolution of life history and genomes. a,** Feeding larval development
937 (planktotrophy: green) represents the ancient and ancestral life history in sea urchins (seastar *P.*
938 *miniata* represents the outgroup)[11]. Non-feeding larval development (lecithotrophy: orange) has
939 evolved on multiple occasions, including recently within the genus *Heliocidaris* (arrowhead)[9]. **b,**
940 Evolution of non-feeding development in *H. erythrogramma* (bottom) included dramatic
941 modifications to otherwise broadly conserved developmental mechanisms, including changes in
942 cleavage geometry, cell fate specification, and morphogenesis[16,18]. **c,** Chromosome-scale genome
943 assemblies of *H. tuberculata* (green) and *H. erythrogramma* (orange). Outer ring: repetitive

944 element content; middle ring: gene content, inner ring: zeta values in OCRs from the selection
945 analyses. Colored points indicate statistically elevated zeta values (indicative of positive
946 selection) within a single OCR on the branch leading to *H. tuberculata* (green) or *H.*
947 *erythrogramma* (orange). Blue points indicate highly conserved OCRs (top 10% of phastCon
948 scores). Triangles denote OCRs with signature of branch-specific positive selection located near
949 dGRN genes. Synteny lines between chromosomes denote locations of 1-1 orthologous OCRs
950 between *Heliocidaris* species. **d,** Signatures of positive selection in protein-coding sequences
951 (CDS) of 84 dGRN and 3,832 non-dGRN single-copy orthologs. Evidence of selection is slightly
952 enriched on the *H. erythrogramma* branch, but dGRN genes show no difference between
953 branches. **e,** P-values of likelihood ratio test for positive selection in CDS on the branch leading
954 to each species (color indicates significant p-values; squares indicate dGRN genes). **f,** Signatures
955 of positive selection within single-copy OCRs near dGRN (n = 1069) and non-dGRN (n =
956 26,253) genes are overall much higher on the *H. erythrogramma* branch (a single gene can have
957 multiple OCRs associated with it). For OCRs near dGRN genes, this difference is notably
958 amplified: signatures of positive selection are depleted relative to non-dGRN genes on the *H.*
959 *tuberculata* branch but substantially elevated on the *H. erythrogramma* branch. **g,** P-values of
960 likelihood ratio test for positive selection in OCRs on the branch leading to each species. Fisher
961 exact test, two-sided- *: p-value < 5e-2; ***: p-value < 5e-4. OCR: open chromatin region;
962 dGRN: developmental gene regulatory network; CDS: coding sequence.

963



964

965 **Figure 2. Evolution of open chromatin landscape. a,** Density and volcano plots of
966 significantly differentially accessible orthologous OCRs between developmental life histories
967 (green = more open in *H. tuberculata* and *L. variegatus*, orange = more open in *H.*
968 *erythrogramma*). **b,** Examples of conservation (*foxA*) and change (*hesC*) in chromatin
969 accessibility landscape near dGRN genes. See Extended Data Fig. 1 for an additional example of
970 a conserved chromatin landscape nearby *foxN2/3* and corresponding expression domains at
971 blastula stage of both *Heliocidaris* species. **c,** Relationship between chromatin, positive
972 selection, and gene expression. Percentage of all genes (top) and differentially expressed genes
973 (bottom) with at least one OCR nearby that is differentially accessible (DA), has evidence of
974 positive selection on the *H. erythrogramma* (*H.e.*) or *H. tuberculata* (*H.t.*)-branch, or is both DA
975 and has evidence of positive selection on the *H.e.* or *H.t.* branch. **d,** Number of differentially

accessible OCRs and OCRs with evidence of positive selection in *H. erythrogramma* for a given gene. Fisher exact test, two-sided-*: p-value < 5e-2; *** p-value < 5e-4. OCR: open chromatin region; dGRN: developmental gene regulatory network.

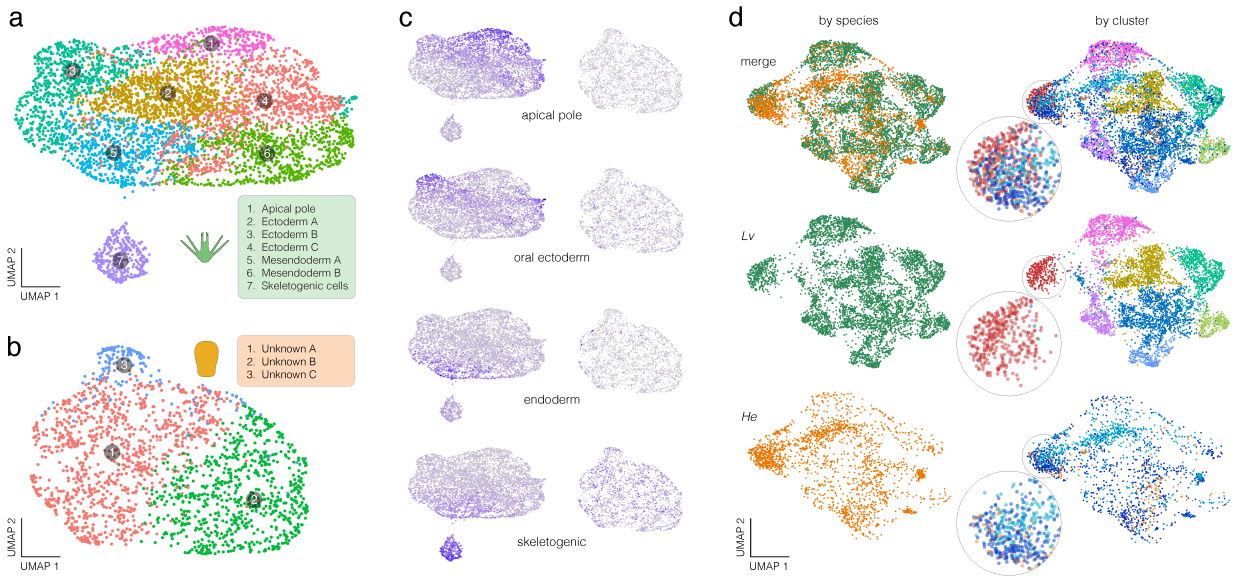


Figure 3. Evolution of transcriptomes. Single cell RNA sequencing of pre-mesenchyme blastula stage embryos from **a.** *L. variegatus* and **b.** *H. erythrogramma* with cells plotted into a UMAP space prior to integration of expression data across species. Colors indicate inferred cell lineages based 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, color-coded by species (left) and by cluster (right), and showing both species (top),

992 and separated by species (middle and bottom). *L. variegatus* clusters remain well separated
993 (center right) but *H. erythrogramma* clusters broadly overlap with each other (lower right). As
994 shown in the 2X magnifications, *L. variegatus* clusters contain primarily cells of one color, while
995 cells from all three *H. erythrogramma* clusters are present in appreciable numbers in the same
996 UMAP space (transparency is set to 40% in the insets to circumvent masking; example inset
997 clusters are *L. variegatus* skeletogenic cells, but the same holds for other *L. variegatus* clusters).

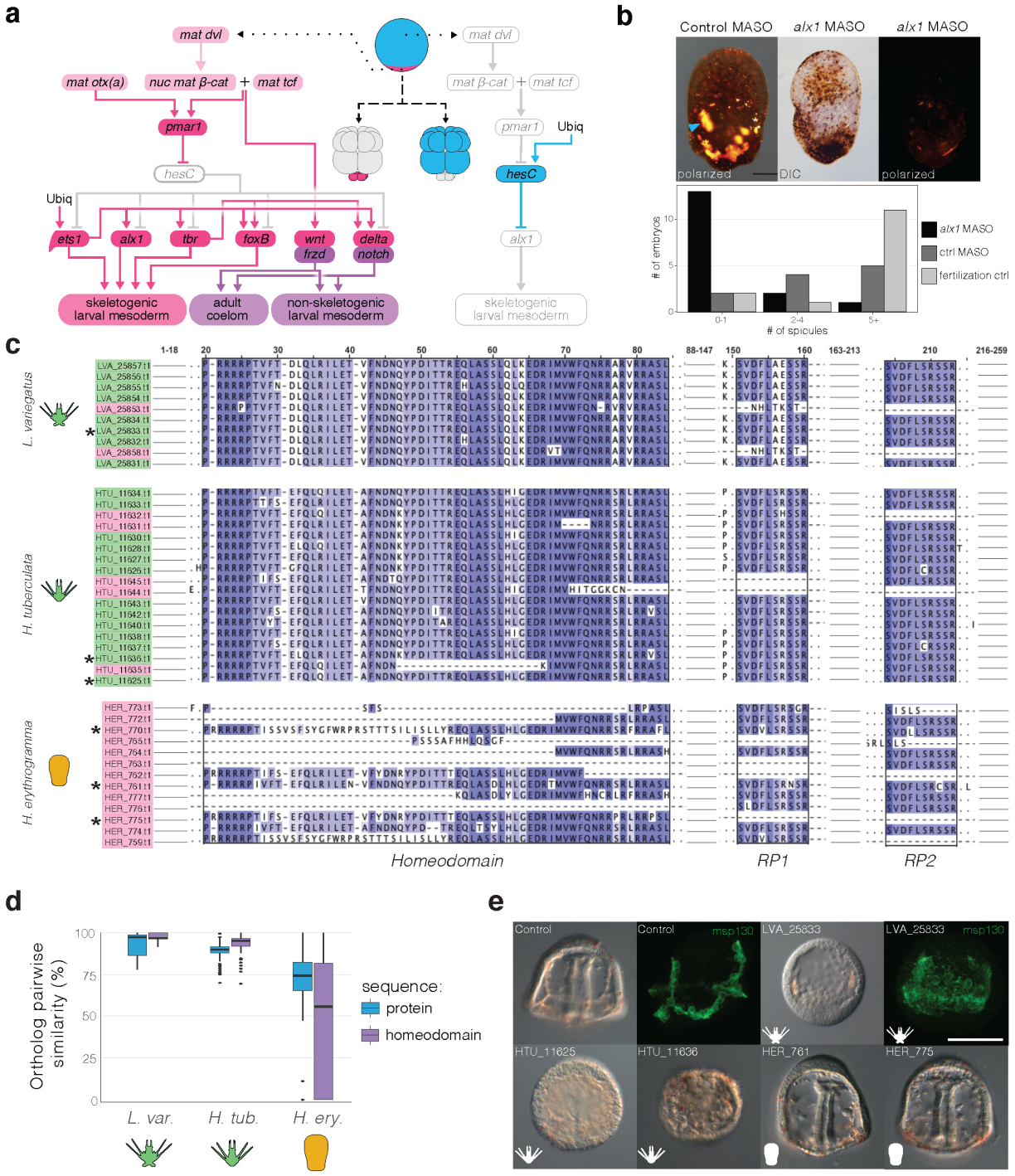
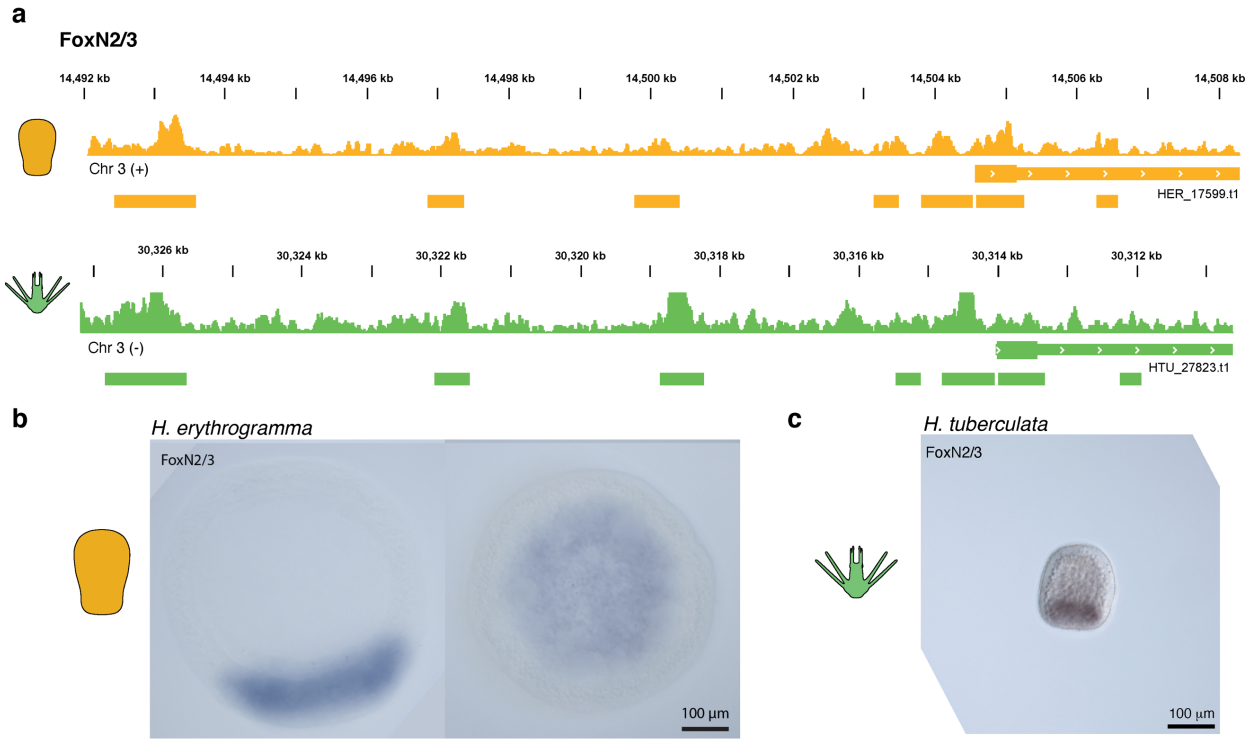


Figure 4. Evolutionary change at the "top" of a conserved developmental dGRN. a, Schematic of the ancestral dGRN that specifies skeletogenic cell fate: HesC suppresses this fate in most of the embryo (blue) but Pmar1 suppresses *hesC* in the precursors of the skeletogenic cells (magenta), where Alx1 then activates a differentiation program. **b,** Images of control and

1003 MASO knock-down of Alx1 in *H. erythrogramma* (early larva; scale bar 100µm). Polarized light
1004 (PL) illuminates skeletal elements (most are anlage of the adult skeleton, with a longer larval
1005 "arm" element out of focus on the left [blue arrow]). Bar chart of *alx1* injection summary
1006 statistics in *H. erythrogramma* (see also Supplementary Table 1). Knockdown of *alx1* expression
1007 eliminates skeleton formation in *H. erythrogramma*, as in the ancestral dGRN[36]. **c,** Alignment of
1008 homeodomain (DNA-binding) and RP domains (protein:protein interaction) from *pmar1*
1009 paralogues. Green = likely functional copy, red = predicted non-functional copy, asterisks =
1010 paralogues whose function was experimentally validated. **d,** Within-species pairwise sequence
1011 similarity of *pmar1* paralogues. Note rapid sequence divergence among paralogues in *H.*
1012 *erythrogramma*, and particularly within the homeodomain. Center line = median pairwise
1013 similarity; dots = outliers **e,** Overexpression assays of control and *pmar1* mRNA (prism stage;
1014 skeletogenic cells labeled with antibody that recognizes cell surface protein MSP130). DIC and
1015 fluorescent images demonstrate that mRNA of *pmar1* paralogues from *L. variegatus* and *H.*
1016 *tuberculata* convert most of the embryo to skeletogenic cells, whereas even the most intact *H.*
1017 *erythrogramma* paralogues show no such phenotype, indicating loss of function (see
1018 Supplementary Fig. 7 for additional antibody staining). dGRN: developmental gene regulatory
1019 network; MASO: morpholino antisense oligonucleotide; DIC: differential interference contrast;
1020 RP: repeated peptides.
1021
1022
1023
1024
1025

1026 **EXTENDED DATA FIGURE LEGENDS**



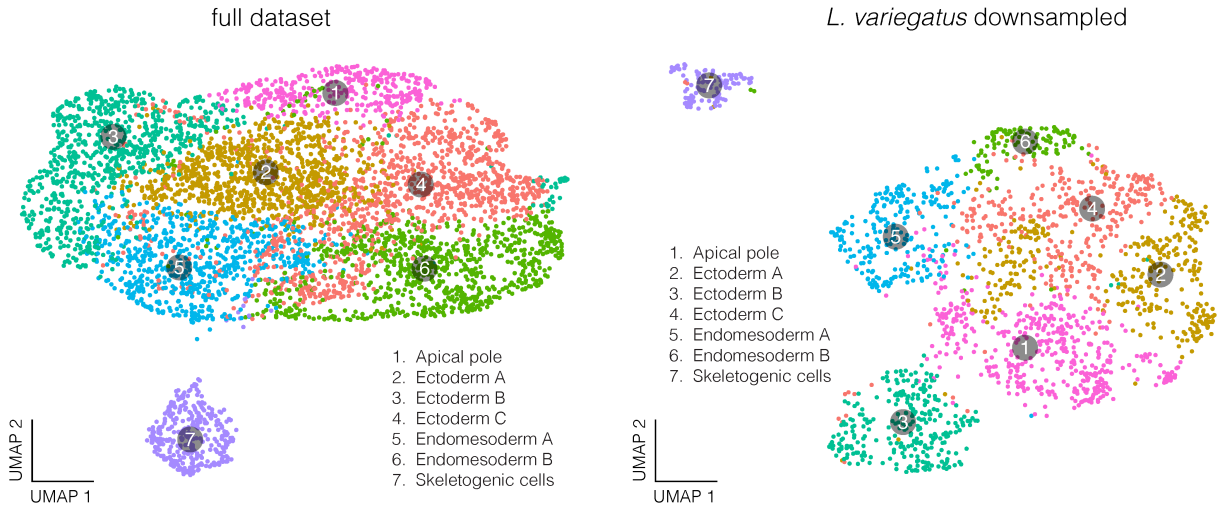
1027

1028 **Extended Data Figure 1. Chromatin landscape and expression domain of *foxN2/3* is**

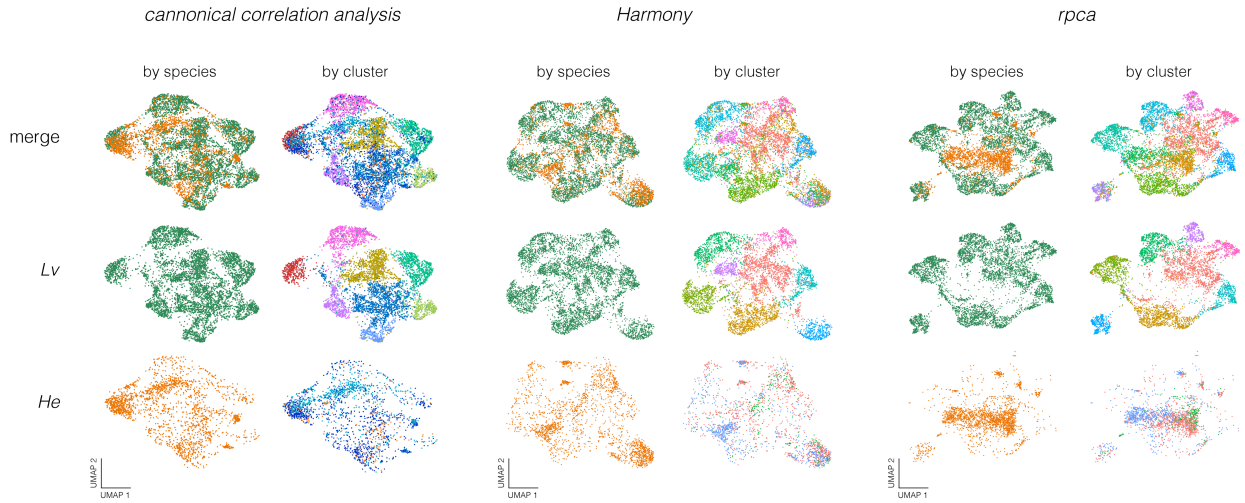
1029 **conserved. a,** Chromatin accessibility nearby *foxN2/3* in *H. erythrogramma* (top, orange) and *H.*

1030 *tuberculata* (bottom, green), including seven open chromatin regions (OCRs). *In-situ*

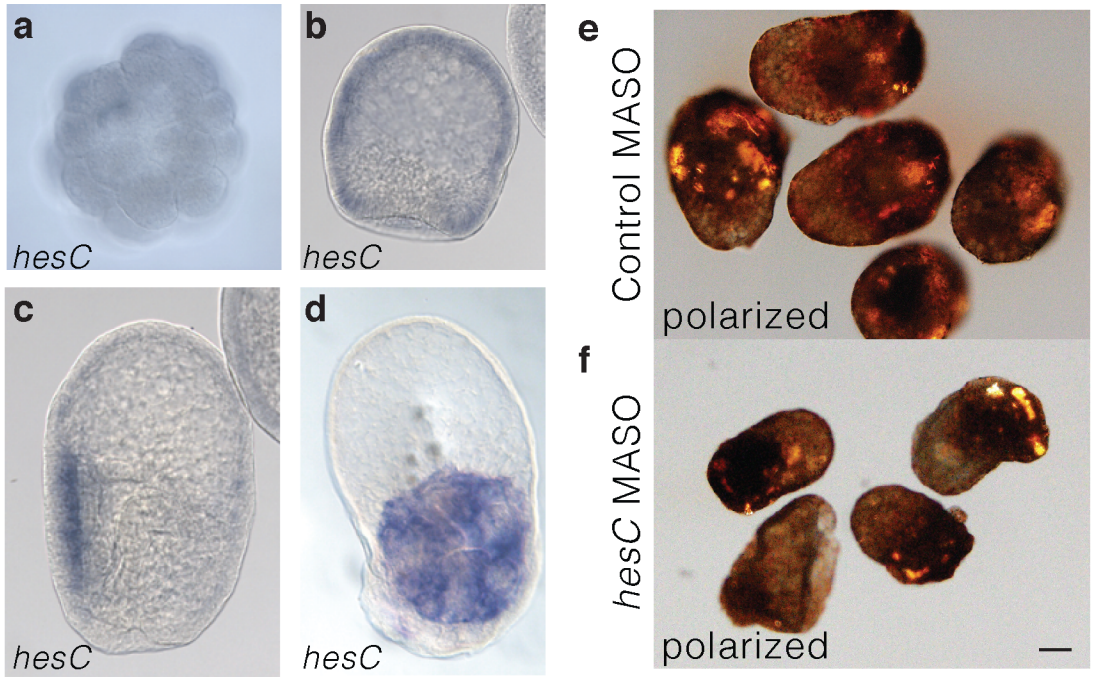
1031 hybridization of *foxN2/3* in blastula-stage **b,** *H. erythrogramma* and **c,** *H. tuberculata* embryos.



Extended Data Figure 2: Scaling down number of *L. variegatus* cells does not significantly 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* (i.e. distinct cluster of skeletogenic cells separate from remainder of cells in the embryo).



1041 **Extended Data Figure 3: Three independent methods of integrating single cell RNA-seq**
1042 **data recover similar clustering relationships among *L. variegatus* and *H. erythrogramma***
1043 **cells.** Canonical correlation analysis (CCA), reciprocal principal component analysis (RPCA),
1044 and Harmony each recover more numerous, distinct transcriptional states in *L. variegatus*
1045 relative to *H. erythrogramma* following integration of single cell expression data. All three
1046 methods reveal a consistent number of clusters in both species. In particular, *H. erythrogramma*
1047 contains fewer clusters there is more extensive overlap of cells among clusters. These results
1048 suggesting that differentiation of discrete cell populations is delayed in the *H. erythrogramma*
1049 embryo.



1050

1051 **Extended Data Figure 4: *hesC* appears to have lost its ancestral role of repressing larval**
1052 **skeletal cell specification in *H. erythrogramma*.** Derived expression patterns of *hesC* in *H.*
1053 *erythrogramma* at **a,** cleavage; **b,** blastula; **and c-d,** larva stage embryos. **e-f,** Control and MASO
1054 knock-down of HesC in *H. erythrogramma* (early larva; scale bar 100µm). Polarized light

1055 illuminates skeletal elements. HesC knockdown appears to show no phenotype, a dramatic

1056 change from the ancestral dGRN.