nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-020-1165-y

Check for updates

Genetic basis for divergence in developmental gene expression in two closely related sea urchins

Lingyu Wang[1], Jennifer W. Israel[1], Allison Edgar[1], Rudolf A. Raff[2], Elizabeth C. Raff[2], Maria Byrne[3,4] and Gregory A. Wray[1,5]✉

The genetic basis for divergence in developmental gene expression among species is poorly understood, despite growing evidence that such changes underlie many interesting traits. Here we quantify transcription in hybrids of *Heliocidaris tuberculata* and *Heliocidaris erythrogramma*, two closely related sea urchins with highly divergent developmental gene expression and life histories. We find that most expression differences between species result from genetic influences that affect one stage of development, indicating limited pleiotropic consequences for most mutations that contribute to divergence in gene expression. Activation of zygotic transcription is broadly delayed in *H. erythrogramma*, the species with the derived life history, despite its overall faster premetamorphic development. Altered expression of several terminal differentiation genes associated with the derived larval morphology of *H. erythrogramma* is based largely on differences in the expression or function of their upstream regulators, providing insights into the genetic basis for the evolution of key life history traits.

The expression level of most genes changes across the life cycle, among cell types, and in response to physiological and environmental conditions. A central question regarding the genetics of adaptation is the degree to which a specific aspect of a gene's expression profile can be altered without affecting other aspects of its expression. At one extreme, a mutation might have similar effects (magnitude and sign) across all regulatory states; at the other, it might have distinct effects among even similar regulatory states. Because a high degree of correlation across traits can constrain adaptation[1], understanding whether mutations affecting gene expression in natural populations typically alter limited aspects of an overall expression profile or instead have broadly pleiotropic effects is an important unsolved question in evolutionary genetics[2–4].

To date, the most extensive information about correlations across distinct regulatory states comes from differentiated adult cell types[5–8]. In humans and mice, most expression quantitative trait loci and cases of allele-specific expression appear to be largely cell type-specific in their effects[7,9–11]. In addition, most distal enhancers in these species appear to be cell type-specific in activity, with a much smaller fraction active in many cell types even for genes that are broadly expressed[12,13]. These findings suggest that gene expression level in one cell type can respond to selection somewhat independently of expression level in a different cell type. At least in principle, this would reduce pleiotropy, allowing independent adaptive tuning of the same gene's expression in different cell types.

Whether the same is true for genes that are expressed during different phases of development is currently unknown[14]. To explore this question, we examined two closely related sea urchins in the genus *Heliocidaris* that display strikingly divergent life history modes and early development (Fig. 1a): *Heliocidaris tuberculata* closely resembles most other camarodont sea urchins in having highly indirect development via a planktotrophic (feeding) larva while *Heliocidaris erythrogramma* has a derived condition of highly abbreviated development with a lecithotrophic (non-feeding) larva[15,16]. Developmental gene expression differs extensively between the two species[17–22] despite only approximately 5 million years of divergence[23]; several lines of evidence indicate that evolutionary changes in the expression of many different genes contributed to the derived life history of *H. erythrogramma*[17–22].

In this study, we evaluate the genetic basis for these gene expression differences by quantifying total and allele-specific transcript abundance with RNA sequencing (RNA-seq) at three developmental stages in both species and in their F_1 hybrids. We then infer mode of inheritance and *cis*/*trans* contributions to expression divergence between species using a well-developed statistical framework[2,5–11]. The results allow us to measure, for the first time, how often genetic influences on gene expression are limited to a specific phase of the life cycle. Our results also provide the first genetic analysis of evolutionary changes in transcription during the maternal-to-zygotic transition. Finally, we place these results within the context of a well-defined developmental gene regulatory network to examine the genetic basis for divergence in larval morphology between the two *Heliocidaris* species.

Results

Hybrids reveal gene regulatory divergence between species. We crossed female *H. erythrogramma* and male *H. tuberculata* to generate F_1 hybrids[24] and used gametes from the same parents to generate same-species crosses. Hybrids were generated in one direction only because the reciprocal cross arrests as gastrulae[24]. From the hybrid and same-species crosses, we collected three biological replicates at three developmental stages (blastula, gastrula, larva) and prepared libraries for RNA-seq. One sample contained insufficient RNA (same-species *H. erythrogramma* blastula stage); the remaining 26 samples served as the basis for all subsequent analyses (Fig. 1a).

We aligned reads from the same-species crosses to corresponding reference transcriptomes (Methods) and from them identified genetic differences in coding sequences within and between species.

[1]Department of Biology, Duke University, Durham, NC, USA. [2]Department of Biology, Indiana University, Bloomington, IN, USA. [3]School of Medical Science, The University of Sydney, Sydney, New South Wales, Australia. [4]School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia. [5]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. ✉e-mail: gwray@duke.edu

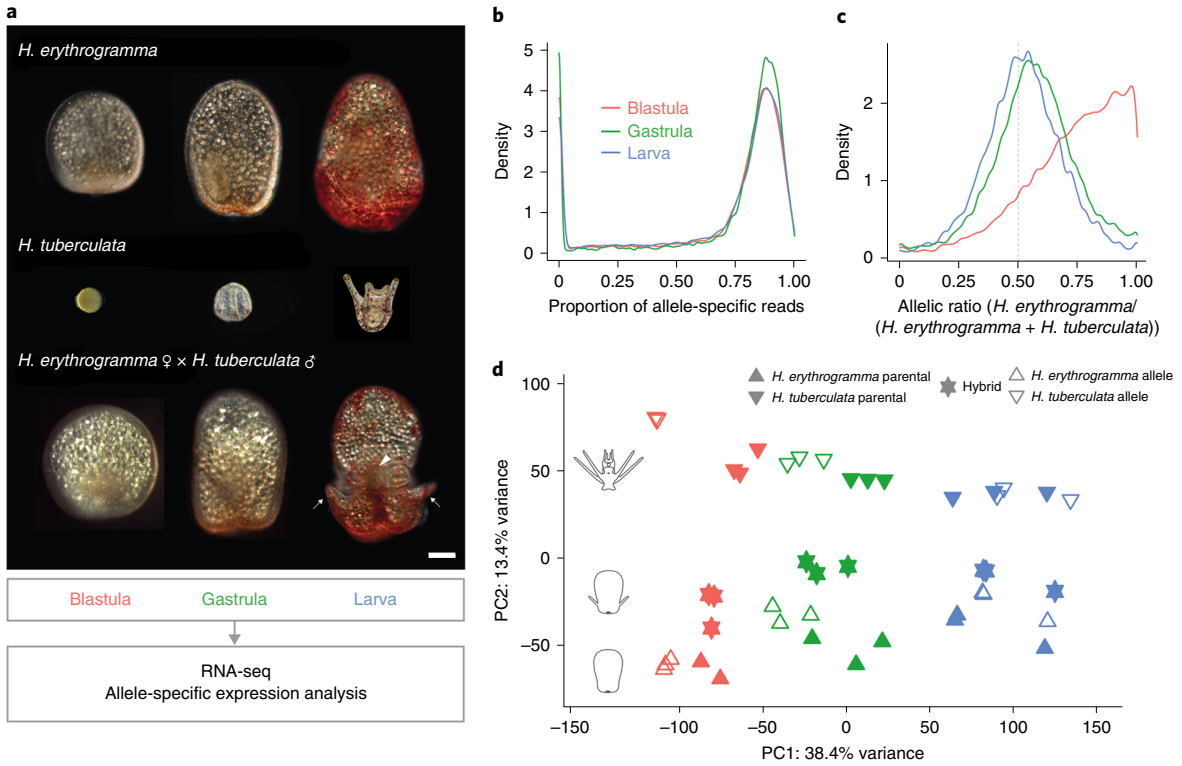


Fig. 1 | Development and gene expression in parents and hybrids. a, Light micrographs of live embryos and larvae at the time points analysed. Hybrid larvae contain rescued arms (arrows) and mouth (not visible; location indicated by the arrowhead). Scale bar, 100 µm. We collected mRNA from hybrids and parental crosses at the three time points shown, then carried out RNA-seq. For the hybrid samples, we used fixed genetic differences between species to assign reads to the *H. erythrogramma* or *H. tuberculata* genome to measure allele-specific expression. **b**, Proportion of allele-specific reads in hybrids. At all three stages, most reads in coding sequences could be assigned to one parental genome or the other based on species-informative sites. **c**, Frequency spectrum of allelic ratios by gene. RNA-seq reads derived from maternal alleles outnumber those from paternal alleles for most genes at the blastula stage, probably a consequence of maternal mRNA. The subsequent reduction in allelic bias at later stages probably reflects activation of the zygotic genome and turnover of maternal transcripts. **d**, PCA of transcriptomes from parental crosses, hybrids and alleles within hybrids. The first principal component separates sequential developmental stages and the second separates parental species with hybrids in between. Alleles in hybrids cluster near the inferred parental transcriptomes, indicating correct parent-of-origin inference for most reads.

Coding sequences contained an average of 2.34% variation within species and 5.49% divergence between species (Supplementary Table 1). We identified >230,000 species-informative sites (12.9% of all single-nucleotide differences) at each stage (Supplementary Table 2); on this basis, we could confidently assign 81.9% of reads in hybrid transcriptomes as originating from one or the other parental genome using HyLiTE[25] (Fig. 1b, Supplementary Table 2 and Supplementary Fig. 1; see Methods for details). Next, we used the aligned reads to measure the level of gene expression. We detected transcripts from thousands of genes from both parental alleles in hybrids at all stages examined (Supplementary Fig. 2), which is consistent with previous observations[26].

To uncover the primary drivers of variation among samples, we performed principal component analysis (PCA). We normalized total RNA-seq read counts for the parental crosses and hybrids; to assess allele-specific expression levels in the hybrids, we used the unnormalized data to retain allele-specific quantification. The first principal component explains 38.4% of the variance and separates the three developmental stages in chronological order (Fig. 1d), while the second explains 13.4% of the variance and separates genotypes, with hybrids in the centre and flanked by the two parentals. Alleles from hybrids plotted near the corresponding stage of the respective parental species (Fig. 1d). Thus, this plot reflects the most fundamental biological properties of the samples, demonstrates that variation among biological replicates is smaller than stage and species effects, and indicates generally high accuracy in inferring parent-of-origin for transcripts in hybrids.

Maternal effects dominate transcription in early embryos then rapidly decline. Studies examining adult tissues have reported relatively few genes with allelic imbalance consistently biased towards the male or female parent[5–8]. We expected to find a different situation because early development is initially dominated by maternal transcripts and proteins stored in the egg before fertilization. Consistent with this prediction, most transcripts in hybrids at the blastula stage were assigned to the maternal genome (Fig. 1c). The two subsequent stages contained nearly equal proportions of paternally and maternally derived transcripts (Fig. 1c), indicating that most messenger RNA molecules originally stored in the egg have turned over and the transcriptome is composed almost entirely of zygotically produced molecules.

We next inferred the mode of inheritance (dominance effects) for each divergently expressed gene by comparing expression levels among parental genotypes and their hybrids as outlined previously by Coolon et al.[27] (Supplementary Table 3). For example, if a gene's expression level in hybrids is close to that of *H. tuberculata* and different from that of *H. erythrogramma*, it would be classified as *H. tuberculata* dominant (Extended Data Fig. 1). We considered genes with low read counts in parental or allele genotypes to be uninformative and classified the rest based on the outcomes of

statistical tests as conserved, maternal-dominant, paternal-dominant, additive (codominant; expression near the average of the two parents), overdominant (expression higher than either parent), underdominant (expression lower than either parent), or ambiguous (not conserved and not statistically supported as one of the previous classifications) (Fig. 2a,d,g and Extended Data Fig. 1).

We observed a dramatic shift in mode of inheritance over developmental time in hybrids, with 33.9% of unambiguous cases classified as maternal-dominant in blastulae and dropping to 12.8% in gastrulae and 12.4% in larvae (Fig. 3a and Supplementary Table 5). At the two later stages, maternal-dominant and paternal-dominant modes of inheritance were almost equal. The most straightforward interpretation of these results is that mRNA and protein molecules originally stored in the egg strongly influence transcription during early development but that these maternal effects decline rapidly, with maternal and paternal influences becoming nearly equal in larvae.

***cis* contributions moderately exceed *trans* contributions in expression divergence.** The observation that alleles in hybrids cluster with their respective parents following dimensional reduction (Fig. 1d) suggests that the hybrid *trans* environment has less influence on transcript abundance than the *cis* genetic differences between the two species. If this were not the case, both alleles in hybrids would show similar expression levels and cluster with the average in hybrids rather than with their respective adults. To investigate these influences more formally, we next classified expression differences between species according to regulatory mode, namely *cis* and *trans* genetic contributions[2,5–11]. Allelic imbalance indicates that a genetic difference between the parents influencing transcription is in linkage disequilibrium with the gene in question[2,5–11]. Because linkage disequilibrium generally spans a tiny fraction of the genome, this is classified as a *cis* (nearby) genetic effect. A *trans* (distant) genetic effect is indicated by no allelic imbalance in hybrids. Mixed *cis* and *trans* effects are also possible and can have reinforcing or opposing influences on transcript abundance (Extended Data Fig. 2). We first eliminated from consideration genes with low read counts in parental or allele genotypes as uninformative. The rest were then classified based on the outcomes of statistical tests following the criteria outlined by Coolon et al.[27] (Supplementary Table 4 and Extended Data Fig. 3) as conserved, all *cis* (*cis* effects alone explain the observed expression difference between parental crosses), all *trans* (*trans* effects alone), *cis*+*trans* (*cis* and *trans* effects that are additive), *cis* × *trans* (*cis* and *trans* effects that partially cancel out), compensatory (*cis* and *trans* effects of equal and opposite sign) or ambiguous (not conserved and not statistically supported as one of the previous classifications) (Fig. 2b,e,h and Supplementary Table 6).

As reported by several previous analyses that examined expression divergence between species in adults[14], more expression differences during development were classified as all *cis* than all *trans* (Fig. 3). This was the case at all three developmental stages examined. In addition, the magnitude of median expression differences between species for genes classified as all *cis* was larger than that of genes classified as all *trans* at both blastula and gastrula stages (2.01 versus 1.83-fold and 2.20 versus 1.89-fold, respectively; Wilcoxon's rank-sum test, $P=4.6\times10^{-7}$ and $P=8.5\times10^{-5}$, respectively) (Fig. 3e). In larvae, we observed the opposite situation: the effect size for genes classified as all *trans* increased substantially while that of the all *cis* set barely changed (2.14 versus 2.60-fold; Wilcoxon's rank-sum test, $P=2.7\times10^{-5}$) (Fig. 3e).

Together, these results indicate a substantial role for both in *cis* and *trans* genetic contributions to expression divergence between the two *Heliocidaris* species, with *cis* effects moderately more extensive both quantitatively and qualitatively. These results indicate that no single metric captures the relative contribution of these two broad classes of genetic effects on the evolution of gene expression, since the number, proportion and magnitude of *cis* and *trans* influences can change across the life cycle (Fig. 3).

Very early transcription of the paternal genome influences the transcriptome. In other camarodont sea urchins, transcription of the zygotic genome begins for a few genes at the 16-cell stage and progressively expands to thousands of genes over the next several hours[22,28–30]. Our results are consistent with this finding. At the blastula stage in hybrids, 242 transcripts were classified as paternal-dominant and 211 were additive (Fig. 2a,c, blue dots and red dots, respectively and Supplementary Table 5). Since the sperm brings few, if any, transcripts to the zygote, this result indicates that transcription of 453 genes is influenced by zygotic expression of paternal alleles of other genes. This conclusion is supported by the measurements of *cis* and *trans* contributions to expression divergence: 41 genes differentially expressed between species that were classified as having an all *trans* or partial *trans* genetic contribution at the blastula stage also showed paternal dominance (41 = 16 + 21 + 4; Fig. 2c). The simplest mechanistic interpretation is that transcription of the paternal genome before the blastula stage in hybrids exerts a rapid influence on the expression of a specific set of target genes (Extended Data Fig. 4).

Activation of the zygotic genome is broadly delayed in *H. erythrogramma*. Gene expression in hybrids exhibited three striking asymmetries in regulatory mode at the blastula stage, the earliest stage we examined. First, most all *trans* cases (92.8%) involve genes with lower expression in *H. erythrogramma* (Fig. 2b, dark green dots, and Supplementary Table 9). Of the 499 all *trans* cases that also have an unambiguous mode of inheritance classification, 92.0% showed maternal dominance (Fig. 2c). These cases of all *trans*+reduced expression+maternal dominance in hybrids are consistent with evolutionary changes in the level or activity of a transcriptional activator in the *H. erythrogramma* embryo (or increased expression or activity of a repressor) (Extended Data Fig. 5). Second, most genes classified as compensatory (98.3%) involved higher expression of the *H. erythrogramma* allele (Fig. 2b, purple dots, and Supplementary Table 9). Compensatory cases involve no difference in transcript abundance between species but a divergence in genetic basis involving equal *cis* and *trans* effects of the opposite sign[14,27]. Such cases are tacitly attributed to gene activity within the individual from which samples were derived[2,5–10,14]. However, when analysing the results from early embryos a *trans* contribution could originate instead from changes in the level of maternal mRNA or protein deposited into the egg (Extended Data Fig. 6). This is a mechanism of compensatory evolution of gene expression that has not been considered previously[3,14] but may be important in the evolution of embryonic gene expression. Third, most all *cis* and *cis* × *trans* cases (73.1 and 96.8%, respectively) involve genes with higher expression of the *H. erythrogramma* allele (Fig. 2b, red and blue dots, and Supplementary Table 9). For reasons already discussed, this third asymmetry also probably reflects transcripts stored in the *H. erythrogramma* egg rather than zygotic transcription.

The first two asymmetries signal a delay in zygotic transcription of the maternal genome in hybrids during early development. Together, 2,379 genes are involved (*H. erythrogramma* allele high for compensatory and *H. tuberculata* allele high for all *trans*; Supplementary Table 9), indicating that this is a broad phenomenon. Particularly striking are the 1,786 genes whose expression does not differ between the two species in pure parental crosses but where it is likely that higher maternal loading during oogenesis in *H. erythrogramma* makes up for delayed zygotic transcription (the compensatory cases discussed earlier). All three asymmetries almost completely disappeared at the later stages of development in hybrids (Fig. 2e,h), indicating that genes in the maternal and paternal

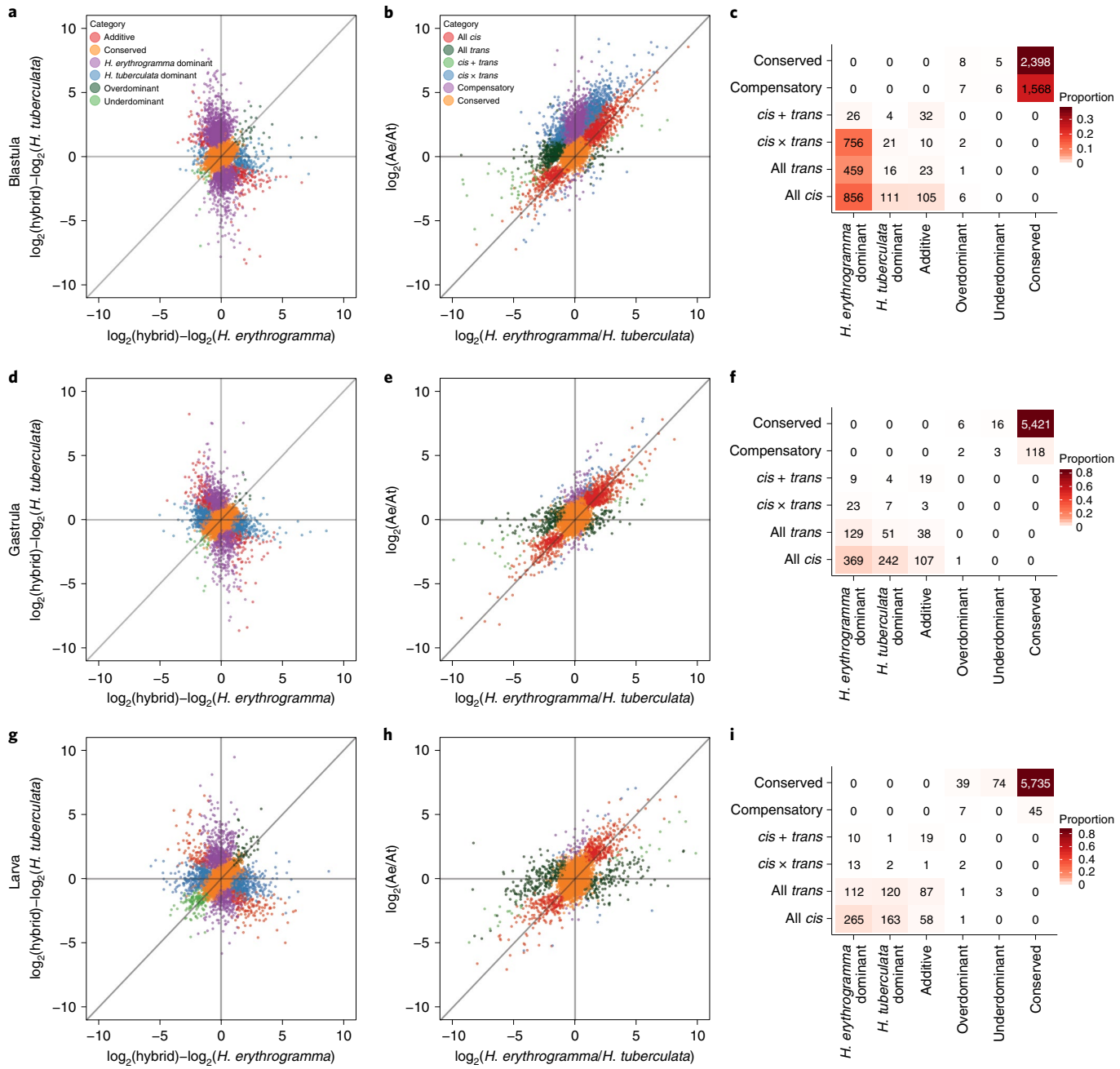


Fig. 2 | Inferred genetic basis for expression divergence among species. a,**d**,**g**, Inheritance mode (dominance). The scatterplots show differences in total minus maternal expression versus total minus paternal expression in hybrids at three time points (blastula (**a**), gastrula (**d**), larva (**g**)). The dots represent individual genes and are colour-coded according to the inferred parental effects (Supplementary Tables 3 and 5; for interpretation, see Supplementary Fig. 3). **b**,**e**,**h**, Regulatory mode (*cis*/*trans*). The scatterplots show differences in expression between species versus between maternal and paternal alleles in hybrids at three time points (blastula (**b**), gastrula (**e**), larva (**h**)). The dots represent individual genes and are colour-coded according to the inferred regulatory mode (Supplementary Tables 4 and 6; for interpretation, see Supplementary Fig. 5). Note that at the blastula stage, some classifications of both inheritance and regulatory mode are prominently asymmetric regarding sign. For example, most all-*trans* expression differences involve lower expression in *H. erythrogramma* (**b**). These asymmetries largely disappear at the later stages. Numbers are presented in Supplementary Table 10 and models for the two asymmetries in Supplementary Figs. 7 and 8. Ae, *H. erythrogramma* allele; At, *H. tuberculata* allele. **c**,**f**,**i**, Intersection of classifications. Heatmaps tally the number of genes with the same pair of classifications at three developmental time points (blastula (**c**), gastrula (**f**), larva (**i**)), with inheritance mode on the *x* axis and regulatory mode on the *y* axis. Only genes for which both modes could be inferred as informative (see text) at each stage are plotted. Note that the number of genes with conserved inheritance but compensatory regulatory mode is high at the blastula stage (1,568, upper right) but drops to low levels at the later stages of development.

genome are being transcribed at similar levels and that their influences on the expression of other genes is approximately equal overall (Extended Data Figs. 5 and 6).

Collectively, these results point to the evolution of a widespread delay in the activation of transcription during early development in *H. erythrogramma*, the species with the derived life history. This

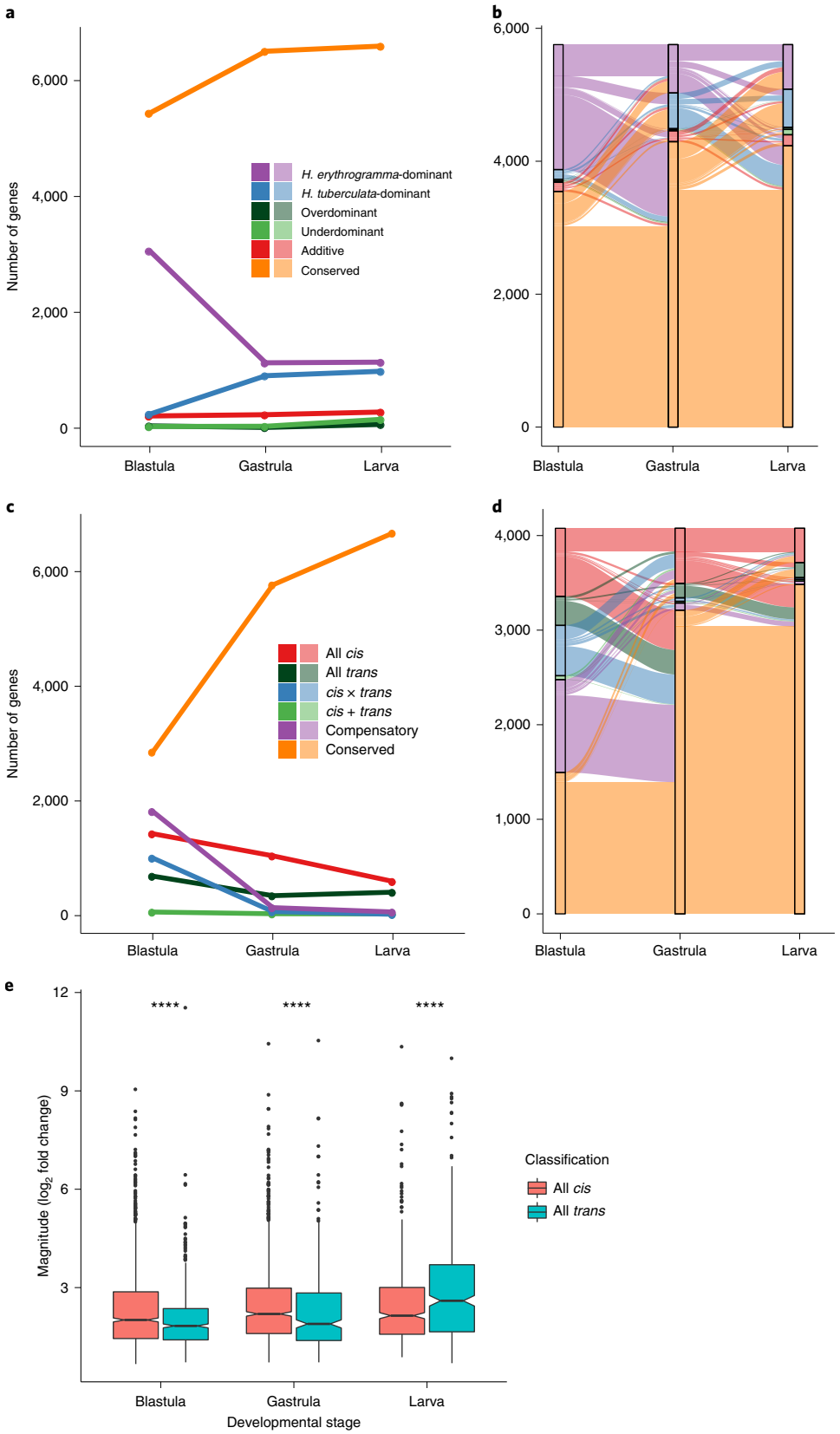


Fig. 3 | Developmental changes in genetic basis for expression divergence. a,**b**, Line (**a**) and alluvial (**b**) plots showing changes in the numbers of genes classified by inheritance mode across the three stages. Tallies include all genes for which we could detect expression, including cases where expression does not differ between species (conserved), and excluding genes that could not be confidently classified (ambiguous, uninformative). **c**,**d**, As in **a**,**b** (respectively) but for regulatory mode. Note that many genes change classification during development, both for inheritance and regulatory modes, and there is an overall trend towards more similar gene expression (increasing orange bars over time). **e**, Magnitude of expression difference between species for genes classified as all *cis* and all *trans* at three developmental stages. Differences at each stage were significant by Wilcoxon rank-sum and signed-rank tests (****$P < 0.0001$).

finding is consistent with previous reports that cell fate specification and differentiation are delayed in *H. erythrogramma* relative to the ancestral condition[17,22,31,32]. The scope of the delay is unexpected, however, given that premetamorphic development is enormously abbreviated in the lecithotroph[15,16].

The genetic architecture underlying expression divergence changes across development. In an earlier study, we found that the overall transcriptomes the two *Heliocidaris* species progressively converge during development (Supplementary Figs. 1 and 2 of Israel et al.[22]). Consistent with that observation, we found that the proportion of genes whose expression is classified as conserved in hybrids rises sharply during development (Fig. 3a,c). This result demonstrates that the extensive divergence of early-stage transcriptomes is primarily based in genetics rather than an indirect consequence of a non-genetic factor, such as the highly derived metabolism of the *H. erythrogramma* embryo[33,34].

This result prompted us to ask whether the genetic basis for expression divergence changes during development for a given gene. Strikingly, of the 308 genes that were differentially expressed at all three stages, 71.6% showed a change in regulatory mode classification (*cis*/*trans* contributions) across stages (Fig. 3d and Supplementary Table 8). Genes that were differentially expressed at just two stages changed regulatory mode classification in similar proportions (Supplementary Table 11). The most common situation involved switching from a difference in gene expression between species at one stage to conserved at another rather than the other way around (Fig. 3d), which contributed to the overall convergence in transcriptomes. Altogether, 88.7% of genes showed some form of change in regulatory mode classification between blastula and 74.6% between gastrula and larva (Supplementary Tables 8 and 11). These observations indicate that genetic contributions to expression divergence at one stage do not exert that influence at a different stage. The most probable explanation is that different transcription factors influence expression among developmental stages (Extended Data Fig. 7).

Different regulatory classifications showed distinct dynamics across development. Nearly every expression difference classified as all *trans* at one stage of development switched to conserved expression at the next stage; the reverse was also true, with most all *trans* deriving from conserved expression at the previous stage for larva and, to a lesser extent, at the gastrula stage (Fig. 3d). This pattern probably reflects the rapid changes in transcription factor expression that unfold during early development in sea urchins[35]. In contrast, approximately one-third of expression differences classified as all *cis* remained so at the subsequent stage, with most of the rest switching to conserved expression. This suggests that most *cis*-regulatory mutations do not exert an influence on transcript abundance at all stages. A tiny fraction of differentially expressed genes showed a switch between all *cis* and all *trans*, whereas those classified as having contributions from both often switch to or from only one influence (for example, *cis*×*trans* to all *cis*). Finally, compensatory cases dropped dramatically after the blastula stage. As argued earlier, this probably reflects the turnover of transcripts originally deposited into the egg rather than extensive compensatory divergence in the usual sense (that is, equal *cis* and *trans* effects of opposite sign) uniquely at the blastula stage.

Changes in gene expression may underlie the evolution of derived life history traits. A striking feature of development in hybrids is the restoration of morphological traits lost during the evolution of *H. erythrogramma*'s derived life history[24]. In particular, the larval arms and mouth that are used to capture and consume phytoplankton in the ancestral life history are absent in the non-feeding larvae of *H. erythrogramma* but are rescued in hybrids (Fig. 1a). This observation suggests that transcription of the paternal genome exerts a powerful influence on morphology in hybrids. The well-characterized ancestral developmental gene regulatory network (GRN) of camarodont sea urchins[35–38] includes several regulatory and structural genes that directly contribute to these structures[39–44] (Fig. 4a). We used this developmental GRN as a framework to investigate how paternal genetic influences contribute to the rescue of larval structures in hybrids.

We first examined the larval skeleton, which supports the larval arms in the ancestral state but is present only as an internal vestigial structure in *H. erythrogramma*[45,46]. Previously, we showed that transcription is delayed for terminal differentiation genes of the skeleton in *H. erythrogramma* relative to the ancestral condition[22]. This observation is consistent with previous studies, which noted delayed specification, differentiation and morphogenesis of skeletogenic cells in *H. erythrogramma*[17,22,31,32,46]. In this study, we found that the genetic basis for expression divergence of many terminal differentiation genes at the blastula and gastrula stages is all or partly *trans* (Fig. 4 and Supplementary Fig. 3). This suggests that changes earlier in the GRN are major contributors to the derived expression of genes that build the reduced larval skeleton of *H. erythrogramma*. Perturbation experiments demonstrated that ancestral regulatory interactions between *Alx1* and *Ets1/2* and downstream differentiation genes are conserved in *H. erythrogramma* (Extended Data Fig. 8). Remarkably, transcription of the maternal allele of *Alx1* is rescued at the blastula stage and of *Tbr* at the gastrula stage in hybrids (Fig. 4 and Supplementary Fig. 3), indicating that their respective *cis*-regulatory elements can still respond to appropriate *trans*-acting factors. In contrast, transcription of *Ets1/2* from the paternal genome is suppressed at the blastula stage in hybrids (Supplementary Fig. 3), suggesting that the maternal *trans* environment either lacks an activator or contains a repressor of this gene.

Finally, we examined genes involved in the development of the larval mouth. A previous study reported that EctoV, a marker of the oral ectoderm in the ancestral GRN, is rescued in hybrids and that the configuration of the ciliated band surrounding the larval mouth is partially restored[26]. In this study, we extend this phenomenon to *lefty* and *nodal*, which encode transcription factors upstream of the oral ectoderm specification in the ancestral GRN[37,42]. Expression of both genes is rescued in hybrids at the gastrula stage (Supplementary Fig. 4). Maternal and paternal alleles produce similar levels of transcripts in hybrids, again suggesting that the maternal genome retains functional *cis*-regulatory elements capable of responding to appropriate *trans*-acting factors. The genetic architecture is *H. tuberculata*-dominant in all three cases, suggesting that transcription of the paternal genome is responsible for restored maternal expression in hybrids. In the ancestral GRN, several genes are subsequently expressed during the larval stage within the stomodeum (*foxa, gsc*), the developing oral field (*not*, *ets4*, *sip1*) and the surrounding ciliated band (*hnf6*)[37,47–51]. The expression of all these genes is rescued in hybrids at the larval stage; the genetic architecture is again *H. tuberculata*-dominant and for most of these genes it is all *trans*.

Taken together, the evidence suggests that the genetic and mechanistic basis for loss of larval arms and mouth in *H. erythrogramma* does not lie primarily in the function of the *cis*-regulatory elements of terminal differentiation genes nor in the proteins they encode, but rather in altered expression of their upstream activators. In particular, the rescued expression of many maternal genes involved in the development of the larval arms and mouth in hybrids indicates that their *cis*-regulatory elements can still interact with the appropriate transcription factors. The predominance of *trans* and paternal genetic influences on the expression of terminal differentiation genes differs from the transcriptome as whole, where *cis* effects are more common and paternal effects are limited to a few hundred transcripts (Figs. 2 and 3).

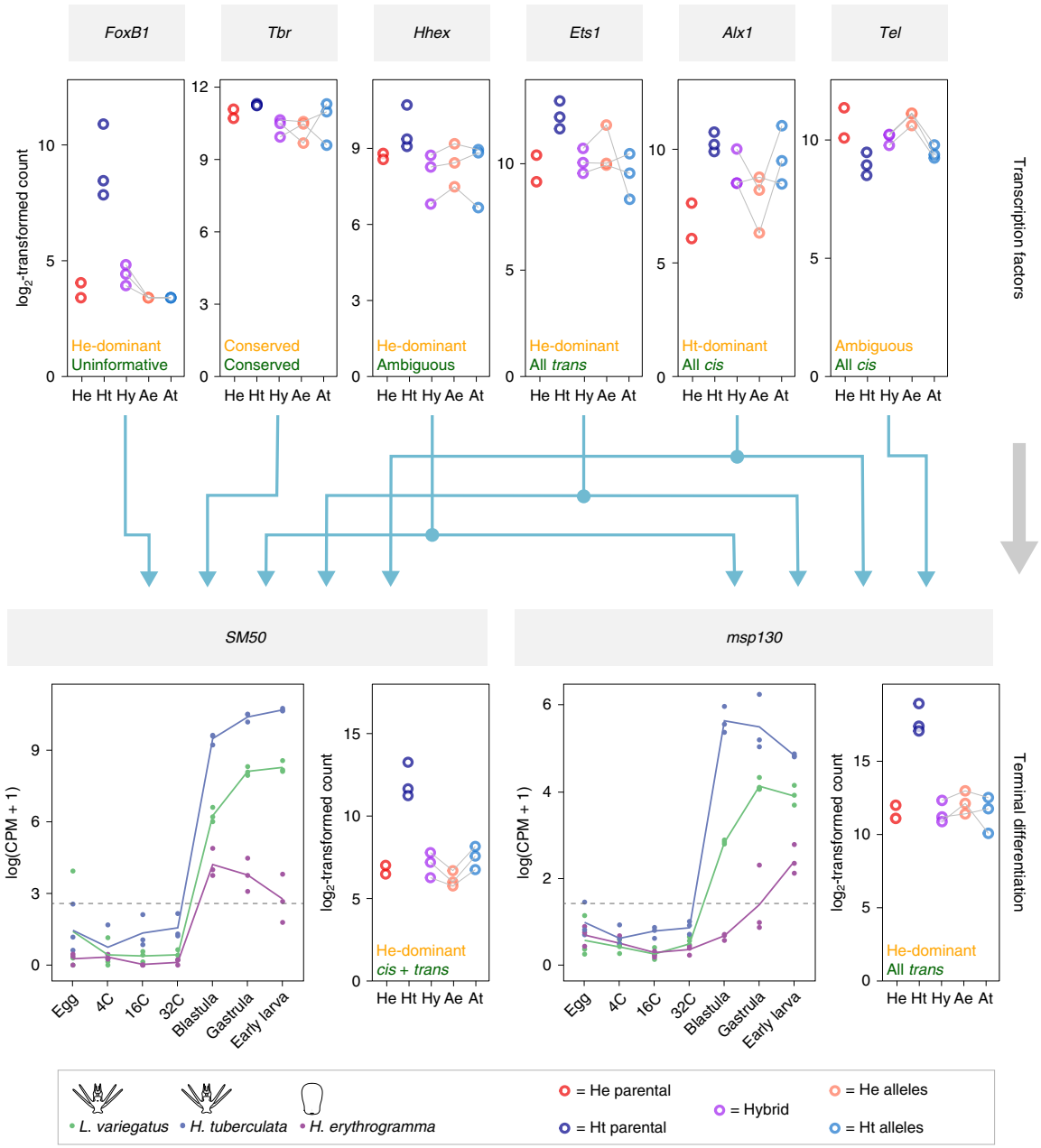


Fig. 4 | Inferred genetic basis for expression of genes that interact during development. Six genes encoding transcription factors (top) directly activate the transcription of two terminal differentiation genes (bottom)[41]. The aqua lines indicate interactions among genes, which are a small subset of the overall developmental GRN. For all eight genes, the plots show expression level at the gastrula stage with inferred genetic basis: red, *H. erythrogramma*; blue, *H. tuberculata*; purple, hybrids; pink, *H. erythrogramma* allele in hybrids; light blue, *H. tuberculata* allele in hybrids. The lines connect measurements from the same sample; inheritance mode, orange text; regulatory mode, green text. For the two terminal differentiation genes at the bottom, the plots on the left show transcript abundance (log_2 scale) in the two *Heliocidaris* species (*H. erythrogramma* and *H. tuberculata*) and an outgroup (*Lytechinus variegatus*) during development (egg; 4C, four-cell stage; 16C, 16-cell stage; 32C, 32-cell stage; blastula; gastrula; early larva; data from Israel et al.[22]). Note that the two terminal differentiation genes (*SM50* and *msp130*), as well as most other terminal differentiation genes for the skeleton and mouth (Supplementary Figs. 3 and 4), are classified as all *trans* or partly *trans*. This result suggests that evolutionary changes in the larval skeleton and mouth in *H. erythrogramma* are caused in part by genes that act earlier in the developmental GRN rather than entirely through the terminal differentiation genes themselves. In contrast, expression differences throughout the transcriptome are more commonly based in *cis* (Fig. 3). Experimental evidence for a conserved GRN linkage between Est1, one of the eight transcription factors, and terminal differentiation genes of the skeleton are presented in Extended Data Fig. 8. He, *H. erythrogramma*; Ht, *H. tuberculata*; Hy, hybrids.

Discussion

It has become clear during the past few years that many mutations that influence gene expression in adults only do so in a limited set of cell types or in response to specific physiological conditions[7,9–11]. In this study, we report a parallel finding for development: most genetic influences on gene expression are limited to a subset of stages. We took advantage of the ability to hybridize[24] two closely related species of sea urchins with extensive divergence in embryonic gene expression[22] to investigate the genetic basis for evolutionary changes in transcription using a robust statistical framework[5–8]. Strikingly,

the genetic architecture underlying expression divergence between species changes during development for 64% of all transcripts measured. This phenomenon is probably a direct consequence of the extensive changes in regulatory states that unfold during development, where the shifting composition of the machinery regulating transcription alters gene expression, patterning the embryo and driving differentiation of distinct cell types. Our observations indicate that the effects of genetic differences influencing developmental gene expression are frequently masked by these changes in regulatory states.

For several reasons, we probably underestimated the proportion of cases where a genetic influence on gene expression affects only a portion of the overall gene expression profile. First, we only examined three points in the life cycle. Analysis of post-metamorphic stages would probably reveal additional cases where the genetic architecture underlying gene expression divergence differs between life history stages. Second, in approximately 12% of cases, we were underpowered to classify the genetic architecture, so some changes among stages were probably missed. Finally, even when the classification of genetic architecture did not change between stages, partially non-overlapping sets of mutations could be responsible. For these reasons, the genetic correlation structure among different aspects of a gene expression profile is probably even weaker than our data indicate.

In principle, a low correlation structure across gene expression traits should allow natural selection to operate on a specific aspect of a gene expression profile without altering others[1]. Consistent with this possibility, extensive trait changes between the two *Heliocidaris* species are concentrated in the early phases of the life cycle: the morphology of embryos and larvae is highly dissimilar while that of post-metamorphic juveniles and adults is very similar[52,53]. We found that many of the extensive differences in gene expression during early development that have evolved within *Heliocidaris* are based largely in *cis* and thus probably influence only nearby genes. Furthermore, *cis* effects are most extensive during embryonic development, when patterning and cell fate decisions are made, while the number and magnitude of *trans* contributions gradually increase over development. Based on these observations, we hypothesize that the genetic basis for the highly derived life history and developmental mechanisms[18,32,52,54] of *H. erythrogramma* is highly polygenic and that many of the causal mutations influence transcriptional regulation primarily during early development, with limited consequences for gene expression later in the life cycle.

In a recent review, Signor and Nuzhdin[14] argued that understanding the evolution of gene expression will require placing results within the context of developmental GRNs. In this study, we drew on the well-defined developmental GRN of camarodont sea urchins[35–39] to investigate the genetic and mechanistic bases for the evolution of derived features of larval morphology in *H. erythrogramma*. Remarkably, several genes that contribute to producing the larval arms and mouth in the ancestral GRN retain functional *cis*-regulatory elements in *H. erythrogramma* that respond appropriately when their cognate *trans*-acting factors are activated in hybrids. This rescue of maternal expression in hybrids suggests that the evolutionary loss of larval arms and mouth in *H. erythrogramma* (Fig. 1a) resulted from changes not in the effector genes that build these structures, but instead in the regulation or function of genes that lie further upstream in the developmental GRN. Together, these findings indicate that the recent and dramatic evolution of a derived life history within *Heliocidaris* resulted from mutations that changed the expression of many individual genes during specific phases of development, often without altering their expression at other points in the life cycle. Furthermore, changes in key regulatory interactions in the early embryo altered the expression of effector genes in terminally differentiated cells later in development, contributing to derived morphological traits in larvae that are associated with a non-feeding life history.

Methods

Animals, crosses and staging. *H. erythrogramma* and *H. tuberculata* adults were collected near Sydney, Australia and maintained in aquaria at 22 °C with circulating natural seawater. Shedding of eggs and sperm were induced by intracoelomic injection of 0.55 M of KCl. Same-species crosses were generated as described previously[22]. Hybrid crosses were generated by treating *H. erythrogramma* eggs with exposure to pH 5.0 seawater for 45 s to remove the jelly coat, followed by rinsing with normal seawater and fertilization with *H. tuberculata* sperm[24]. Hybrid and same-species crosses were cultured at 22–24 °C with daily changes of Millipore-filtered seawater and samples collected as described previously[22,24] at three stages: blastula, late gastrula and late larva (12, 24 and 64 h post-fertilization, respectively).

Library preparation and data production. RNA was extracted using the QIAGEN RNeasy Mini Kit, quantity measured on a NanoDrop 1000 (Thermo Fisher Scientific), and quality-assessed with a BioAnalyzer (Agilent Technologies). One blastula same-species *H. erythrogramma* sample contained insufficient RNA for library construction. From the remaining 26 samples, approximately 2 µg of total RNA was used as input for the TruSeq Library Preparation Kit v2 (Illumina); libraries were prepared according to the manufacturer's instructions. Libraries were sequenced on an HiSeq 2000 (Illumina) as 101-base pair (bp) single-end reads.

Quality control, trimming and alignment of reads. A total of 2.05 billion clean reads were generated from the 26 libraries, with an average depth of 78.9 million reads per sample (range = 57.5–129.5 million). After examining the quality of raw reads using FastQC v.0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), the first 12 nucleotides of raw reads were clipped using seqtk v.1.2-r95-dirty (https://github.com/lh3/seqtk). Reads were further trimmed to remove low-quality bases using Trimmomatic v.0.36 (ref. [55]), with the parameters 'trimmomatic SE -threads 8 -phred33 SLIDINGWINDOW:4:5 TRAILING:5'. Because of the high level of sequence variation and divergence in our samples, we used the variation-tolerant aligner stampy v.1.0.28 (ref. [56]) to map reads, setting 'substitutionrate=0.100 -t 12' and all other parameters to their default values. HyLiTE v.1.6.3 (ref. [25]), which we used to measure allele-specific expression, requires mapping all reads to a single reference transcriptome or genome. We used our previously published *H. erythrogramma* transcriptome[22] as the reference because it was assembled from a larger number of stages and samples and is thus more complete than the *H. tuberculata* transcriptome. Mapping to the *H. erythrogramma* and *H. tuberculata* reference transcriptomes produced similar results (Supplementary Table 10). Because the reference transcriptomes consist only of coding sequences, any reads largely or entirely in untranslated regions will not map. In the genome of the purple sea urchin *Strongylocentrotus purpuratus*, the average coding sequence and mRNA lengths are 1,393 bp and 3,461 bp, respectively[57]. Assuming the two *Heliocidaris* species have a similar ratio of coding sequence:mRNA, a maximum of approximately 40% of reads have a possibility of aligning to the reference transcriptomes. The observed mapping rate averaged across samples was 32.26% from *H. erythrogramma* parental crosses, 35.52% from *H. tuberculata* parental crosses and 34.08% from hybrid crosses, indicating that approximately 85% of the estimated maximum possible reads aligned. To assess our ability to detect the extensive differences in gene expression known to exist in the early development of the two *Heliocidaris* species[22], we used DESeq2 1.20.0 to test for differences in transcript abundance in the same-species crosses (default settings with $P < 0.05$, adjusted $P < 0.1$). Fold change expression level (MA) plots (Supplementary Figs. 5 and 7) and histograms of P values (Supplementary Figs. 6 and 8) indicate that tests for differential expression were robust. The distribution of unadjusted P values showed the expected, strongly left-skewed distribution with a flat tail signifying a suitable null hypothesis.

Measuring allele-specific expression. We next assigned reads in hybrids to parent-of-origin based on the species-informative markers described earlier using HyLiTE v1.6.3 (ref. [25]). As explained earlier, we used stampy v.1.0.28 (ref. [56]) to align reads rather than Bowtie2, the default aligner called by HyLiTE. We called HyLiTE with the following command: 'HyLiTE -v -S -f sam_protocol_file.txt -r he.ref.jwi.fasta -n my_third_HyLiTE', where the sam_protocol_file.txt file provides output SAM files from stampy and he.ref.jwi.fasta provides the reference transcriptome. Samples from the three different stages were analysed separately, allocating 65 GB of RAM to each job. Some genes were expressed at a very low level at one or more stages in one species, providing very few species-informative sites (an example is shown in Supplementary Fig. 9). We carried out a second run of HyLiTE using the merged SAM files from all three stages using Picard Tools v.2.4.1 to increase the rate of read assignment to each species. Therefore, we analysed the hybrid transcriptomes at each stage with reference to all three stages in the same-parent crosses to identify and use more species-informative substitutions; this resulted in approximately 1.7% of genes moving from an ambiguous to an informative classification, depending on the stage (Supplementary Table 10).

Data cleaning and filtering, statistical tests and downstream analyses. Next, we identified cases of differential gene expression between species based on the same-species parental crosses and compared levels of expression in hybrids with

the parental crosses to classify each gene's mode of inheritance (dominance effects) and regulation (*cis*/*trans* genetic contributions to expression divergence) according to formal criteria. All test statistics were performed in R v.3.5.1. DESeq2 1.20.0 was used for normalization, differential expression tests, classification of inheritance and regulatory mode, and all plots were generated using custom R code (see Code availability). After the HyLiTE analysis, read count of *H. erythrogramma*, *H. tuberculata*, hybrids, *H. erythrogramma*/*H. tuberculata* alleles in hybrids from all replicates and stages were combined into one count table as the input for DESeq2. The count data were transformed by variance stabilizing transformation[58] for PCA. We followed the criteria of Coolon et al.[27] to classify genes by regulatory class (Supplementary Table 3), with modifications in the statistical methods as described. Classifying a gene requires three comparisons: (1) expression levels of parental genotypes (*H. erythrogramma* versus *H. tuberculata*), expression levels of *H. erythrogramma*/*H. tuberculata* alleles in hybrids and the ratio of expression from parental genotypes to the ratio of expression from allele genotypes (*H. erythrogramma*/*H. tuberculata* versus *H. erythrogramma*/*H. tuberculata* alleles in hybrids). Coolon et al.[27] used Fisher's exact test, binomial exact test and Fisher's exact test for these three comparisons, respectively (Supplementary Table 5 of Coolon et al. [27]) since their data did not include biological replicates. We altered their approach to account for biological replicates. The *H. erythrogramma* versus *H. tuberculata* and *H. erythrogramma* allele versus *H. tuberculata* allele in hybrids comparisons with replicates are straightforward. Negative binomial generalized linear model fitting and Wald statistical tests from standard DESeq2 are appropriate. To compare *H. erythrogramma*/*H. tuberculata* versus *H. erythrogramma*/*H. tuberculata* alleles in hybrids, we used the transTest with the following formula: 'design < - model.matrix (~0 + Gen * Ori)', where Gen denotes whether the reads were from *H. erythrogramma*/*H. tuberculata* or *H. tuberculata*/*H. tuberculata* alleles in hybrids and Ori denotes whether reads were from the parental species or hybrids[59]. Genes with low read counts in parental genotypes or allele genotypes were classified as 'uninformative'. Based on the results of the negative binomial generalized linear model fitting and Wald statistical tests, each remaining gene was classified as 'ambiguous', 'conserved', 'compensatory', 'all *cis*', 'all *trans*', '*cis* + *trans*' or '*cis* × *trans*' using the criteria in Supplementary Table 4. Note that the method used avoids an artefact identified by Fraser[60] that can inflate inference of compensatory genetic basis when considering *trans* contributions to equal parental minus *cis* contributions. Following Coolon et al.[27], we used a statistical test to measure *trans* contributions that are independent of inferred *cis* contributions, specifically by comparing $\log_2$(*H. erythrogramma*/*H. tuberculata* alleles in hybrids) versus $\log_2$(*H. erythrogramma*/*H. tuberculata*) (Supplementary Table 4). We followed the criteria set by Coolon et al.[27] to classify inheritance mode. Since we had biological replicates, we used DESeq2 to perform statistical tests to classify each gene according to mode of inheritance using the criteria described in Supplementary Table 3. To compare the magnitude of mean expression differences between species for genes classified as all *cis* versus all *trans*, the $\log_2$-transformed fold changes were first tested for normality using the Shapiro–Wilk test of normality, followed by Wilcoxon rank-sum and signed-rank tests.

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

Data availability

The sequence reads for this project are available from the Sequence Read Archive under accession no. SRP229522. The count tables are available from the project's GitHub repository (https://github.com/Wray-Group-at-Duke/HybridsProject_NEE).

Code availability

The code used in this project is available from the project's GitHub repository (https://github.com/Wray-Group-at-Duke/HybridsProject_NEE).

Received: 22 June 2019; Accepted: 3 March 2020;
Published online: 13 April 2020

References

1. Kirkpatrick, M. Patterns of quantitative genetic variation in multiple dimensions. *Genetica* **136**, 271–284 (2009).
2. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nat. Rev. Genet.* **7**, 862–872 (2006).
3. McGuigan, K., Collet, J. M., Allen, S. L., Chenoweth, S. F. & Blows, M. W. Pleiotropic mutations are subject to strong stabilizing selection. *Genetics* **197**, 1051–1062 (2014).
4. Hine, E., Runcie, D. E., McGuigan, K. & Blows, M. W. Uneven distribution of mutational variance across the transcriptome of *Drosophila serrata* revealed by high-dimensional analysis of gene expression. *Genetics* **209**, 1319–1328 (2018).
5. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Independent effects of *cis*- and *trans*-regulatory variation on gene expression in *Drosophila melanogaster*. *Genetics* **178**, 1831–1835 (2008).
6. McManus, C. J. et al. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* **20**, 816–825 (2010).
7. Crowley, J. J. et al. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat. Genet.* **47**, 353–360 (2015).
8. Metzger, B. P. H. et al. Contrasting frequencies and effects of *cis*- and *trans*-regulatory mutations affecting gene expression. *Mol. Biol. Evol.* **33**, 1131–1146 (2016).
9. Pirinen, M. et al. Assessing allele-specific expression across multiple tissues from RNA-seq read data. *Bioinformatics* **31**, 2497–2504 (2015).
10. Aguet, F. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
11. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
12. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
13. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
14. Signor, S. A. & Nuzhdin, S. V. The evolution of gene expression in *cis* and *trans*. *Trends Genet.* **34**, 532–544 (2018).
15. Williams, D. H. C. & Anderson, D.T. The reproductive system, embryonic development, larval development and metamorphosis of the sea urchin *Heliocidaris erythrogramma* (Val.) (Echinoidea: Echinometridae). *Aust. J. Zool.* **23**, 371–403 (1975).
16. Raff, R. A. Direct-developing sea urchins and the evolutionary reorganization of early development. *Bioessays* **14**, 211–218 (1992).
17. Parks, A. L., Parr, B. A., Chin, J.-E., Leaf, D. S. & Raff, R. A. Molecular analysis of heterochronic changes in the evolution of direct developing sea urchins. *J. Evol. Biol.* **1**, 27–44 (1988).
18. 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. B Mol. Dev. Evol.* **310B**, 609–622 (2008).
19. Wilson, K. A., Andrews, M. E., Rudolf Turner, F. & Raff, R. A. Major regulatory factors in the evolution of development: the roles of *goosecoid* and *Msx* in the evolution of the direct-developing sea urchin *Heliocidaris erythrogramma*. *Evol. Dev.* **7**, 416–428 (2005).
20. 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).
21. Zhou, N., Wilson, K. A., Andrews, M. E., Kauffman, J. S. & Raff, R. A. Evolution of OTP-independent larval skeleton patterning in the direct-developing sea urchin, *Heliocidaris erythrogramma*. *J. Exp. Zool. B Mol. Dev. Evol.* **300**, 58–71 (2003).
22. 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).
23. Smith, M. J., Boom, J. D. & Raff, R. A. Single-copy DNA distance between two congeneric sea urchin species exhibiting radically different modes of development. *Mol. Biol. Evol.* **7**, 315–326 (1990).
24. Raff, E. C. et al. A novel ontogenetic pathway in hybrid embryos between species with different modes of development. *Development* **126**, 1937–1945 (1999).
25. Duchemin, W., Dupont, P.-Y., Campbell, M. A., Ganley, A. R. & Cox, M. P. HyLiTE: accurate and flexible analysis of gene expression in hybrid and allopolyploid species. *BMC Bioinformatics* **16**, 8 (2015).
26. Nielsen, M. G., Wilson, K. A., Raff, E. C. & Raff, R. A. Novel gene expression patterns in hybrid embryos between species with different modes of development. *Evol. Dev.* **2**, 133–144 (2000).
27. Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R. & Wittkopp, P. J. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* **24**, 797–808 (2014).
28. Hough-Evans, B. R., Wold, B. J., Ernst, S. G., Britten, R. J. & Davidson, E. H. Appearance and persistence of maternal RNA sequences in sea urchin development. *Dev. Biol.* **60**, 258–277 (1977).
29. Hough-Evans, B. R., Ernst, S. G., Britten, R. J. & Davidson, E. H. RNA complexity in developing sea urchin oocytes. *Dev. Biol.* **69**, 258–269 (1979).
30. Ernst, S. G., Hough-Evans, B. R., Britten, R. J. & Davidson, E. H. Limited complexity of the RNA in micromeres of sixteen-cell sea urchin embryos. *Dev. Biol.* **79**, 119–127 (1980).
31. 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).
32. 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).

33. Byrne, M. et al. Maternal factors and the evolution of developmental mode: evolution of oogenesis in *Heliocidaris erythrogramma*. *Dev. Genes Evol.* **209**, 275–283 (1999).
34. Davidson, P. L. et al. A comparative analysis of egg provisioning using mass spectrometry during rapid life history evolution in sea urchins. *Evol. Dev.* **21**, 188–204 (2019).
35. Davidson, E. H. et al. A genomic regulatory network for development. *Science* **295**, 1669–1678 (2002).
36. Erkenbrack, E. M. & Davidson, E. H. & Peter, I. S. Conserved regulatory state expression controlled by divergent developmental gene regulatory networks in echinoids. *Development* **145**, dev167288 (2018).
37. Saudemont, A. et al. Ancestral regulatory circuits governing ectoderm patterning downstream of Nodal and BMP2/4 revealed by gene regulatory network analysis in an echinoderm. *PLoS Genet.* **6**, e1001259 (2010).
38. Oliveri, P., Carrick, D. M. & Davidson, E. H. A regulatory gene network that directs micromere specification in the sea urchin embryo. *Dev. Biol.* **246**, 209–228 (2002).
39. Shashikant, T., Khor, J. M. & Ettensohn, C. A. From genome to anatomy: the architecture and evolution of the skeletogenic gene regulatory network of sea urchins and other echinoderms. *Genesis* **56**, e23253 (2018).
40. Rafiq, K., Shashikant, T., McManus, C. J. & Ettensohn, C. A. Genome-wide analysis of the skeletogenic gene regulatory network of sea urchins. *Development* **141**, 950–961 (2014).
41. Rafiq, K., Cheers, M. S. & Ettensohn, C. A. The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Development* **139**, 579–590 (2012).
42. Duboc, V., Röttinger, E., Besnardeau, L. & Lepage, T. Nodal and BMP2/4 signaling organizes the oral-aboral axis of the sea urchin embryo. *Dev. Cell* **6**, 397–410 (2004).
43. Range, R. & Lepage, T. Maternal Oct1/2 is required for Nodal and vg1/univin expression during dorsal–ventral axis specification in the sea urchin embryo. *Dev. Biol.* **357**, 440–449 (2011).
44. 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).
45. Wray, G. A. Parallel evolution of nonfeeding larvae in echinoids. *Syst. Biol.* **45**, 308–322 (1996).
46. 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).
47. de-Leon, S. B.-T. & Davidson, E. H. Information processing at the foxa node of the sea urchin endomesoderm specification network. *Proc. Natl Acad. Sci. USA* **107**, 10103–10108 (2010).
48. Su, Y.-H. et al. A perturbation model of the gene regulatory network for oral and aboral ectoderm specification in the sea urchin embryo. *Dev. Biol.* **329**, 410–421 (2009).
49. Li, E., Materna, S. C. & Davidson, E. H. New regulatory circuit controlling spatial and temporal gene expression in the sea urchin embryo oral ectoderm GRN. *Dev. Biol.* **382**, 268–279 (2013).
50. Angerer, L. M. et al. Sea urchin goosecoid function links fate specification along the animal-vegetal and oral-aboral embryonic axes. *Development* **128**, 4393–4404 (2001).
51. Otim, O., Amore, G., Minokawa, T., McClay, D. R. & Davidson, E. H. SpHnf6, a transcription factor that executes multiple functions in sea urchin embryogenesis. *Dev. Biol.* **273**, 226–243 (2004).
52. Raff, R. A. Constraint, flexibility, and phylogenetic history in the evolution of direct development in sea urchins. *Dev. Biol.* **119**, 6–19 (1987).
53. McNamara, K. J., Pawson, D. L., Miskelly, A. D. & Byrne, M. in *Australian Echinoderms: Biology, Ecology and Evolution* (eds Byrne, M. & O'Hara, T. D.) 350–445 (CSIRO Publishing, 2017).
54. Henry, J. J., Wray, G. A. & Raff, R. A. The dorsoventral axis is specified prior to first cleavage in the direct developing sea urchin *Heliocidaris erythrogramma*. *Development* **110**, 875–884 (1990).
55. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
56. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
57. Sodergren, E. et al. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**, 941–952 (2006).
58. 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).
59. Davidson, J. H. & Balakrishnan, C. N. Gene regulatory evolution during speciation in a songbird. *G3 (Bethesda)* **6**, 1357–1364 (2016).
60. Fraser, H. B. Improving estimates of compensatory *cis-trans* regulatory divergence. *Trends Genet.* **35**, 88 (2019).

Acknowledgements

D. McClay, P. Davidson, H. Devens, A. Massri and S. Makohon-Moore provided advice and comments. J. Coolon (Wesleyan University) generously shared unpublished code and advice. This study was supported by an Australian Research Council grant to M.B. and National Science Foundation grant nos. IOS1457305 and IOS1929934 to G.A.W.

Author contributions

L.W., J.W.I. and G.A.W. designed the study. L.W., J.W.I. and A.E. generated the data. L.W., J.W.I., A.E., M.B. and G.A.W. contributed the analysis tools. All authors participated in analysing the results. G.A.W. wrote the paper with input from L.W., J.W.I., A.E., R.A.R., E.C.R. and M.B.

Competing interests

The authors declare no competing interests.

Additional information

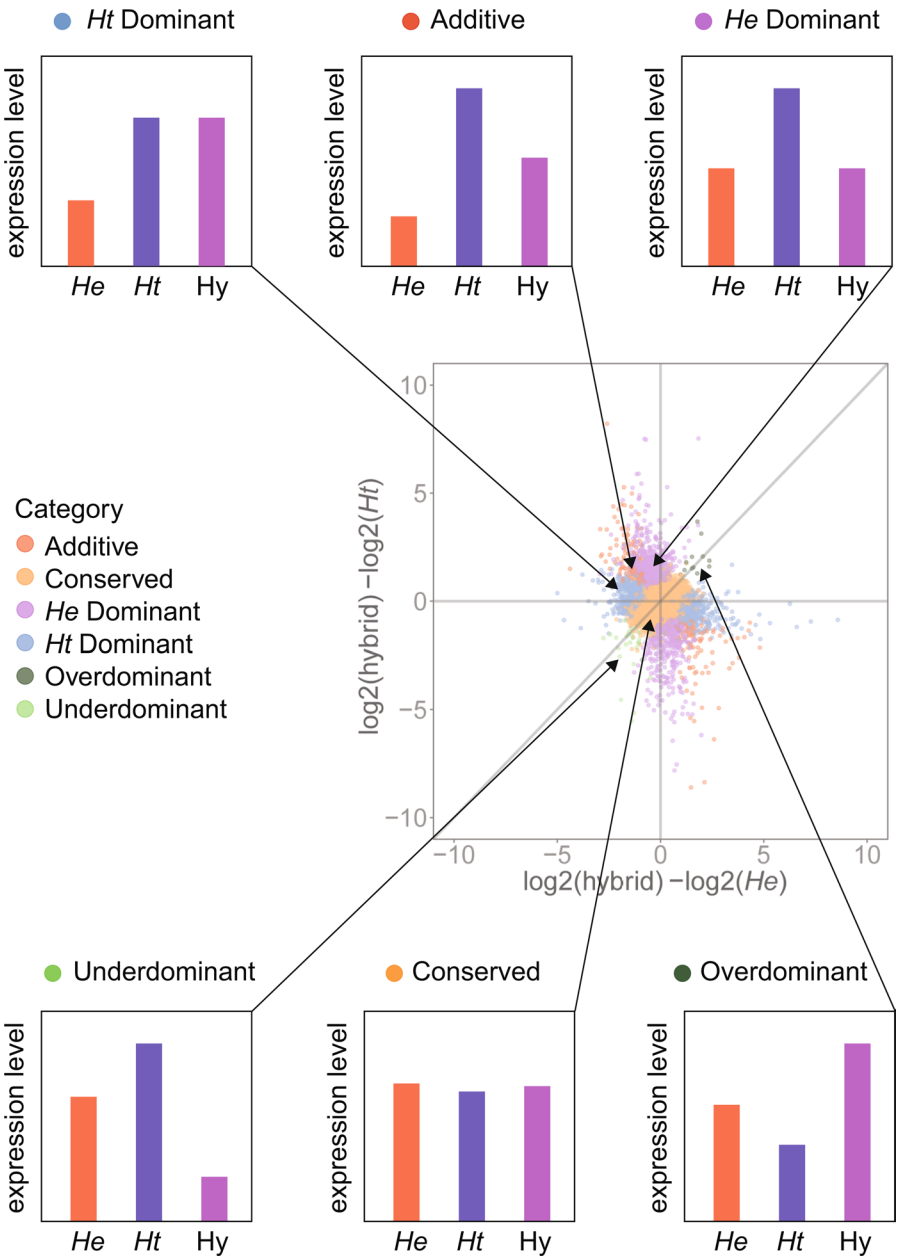
Extended data is available for this paper at https://doi.org/10.1038/s41559-020-1165-y.

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-020-1165-y.

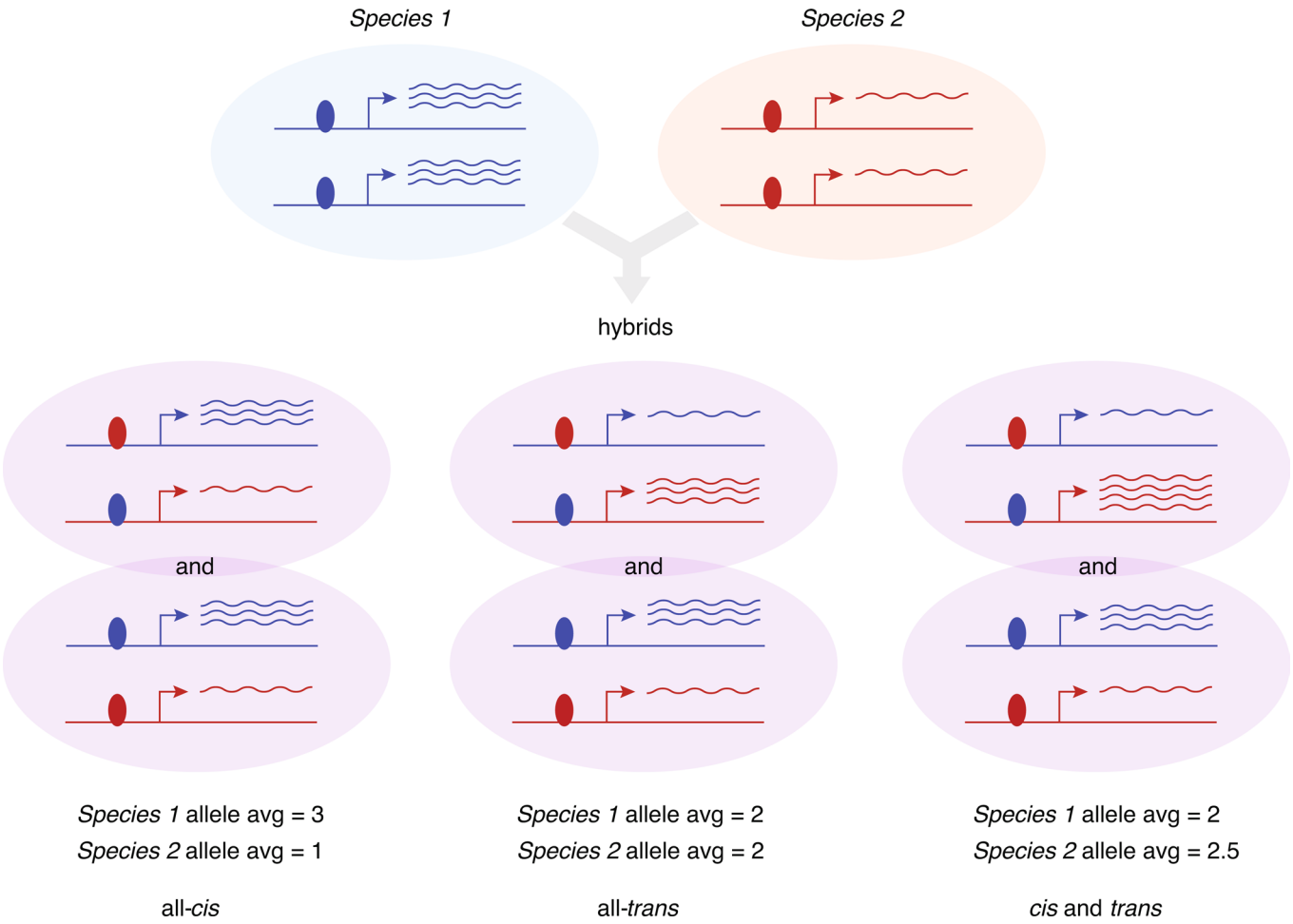
Correspondence and requests for materials should be addressed to G.A.W.

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

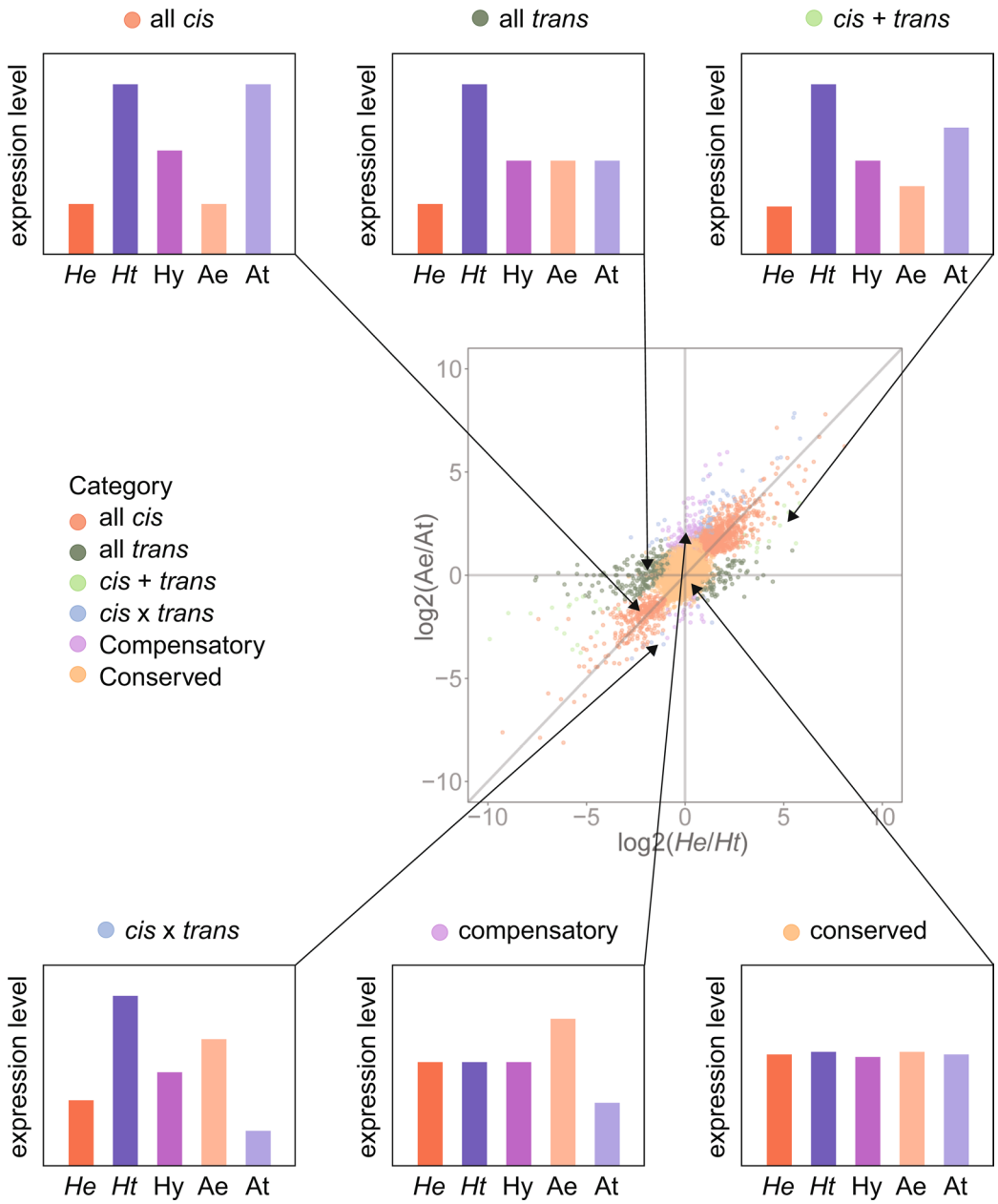
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



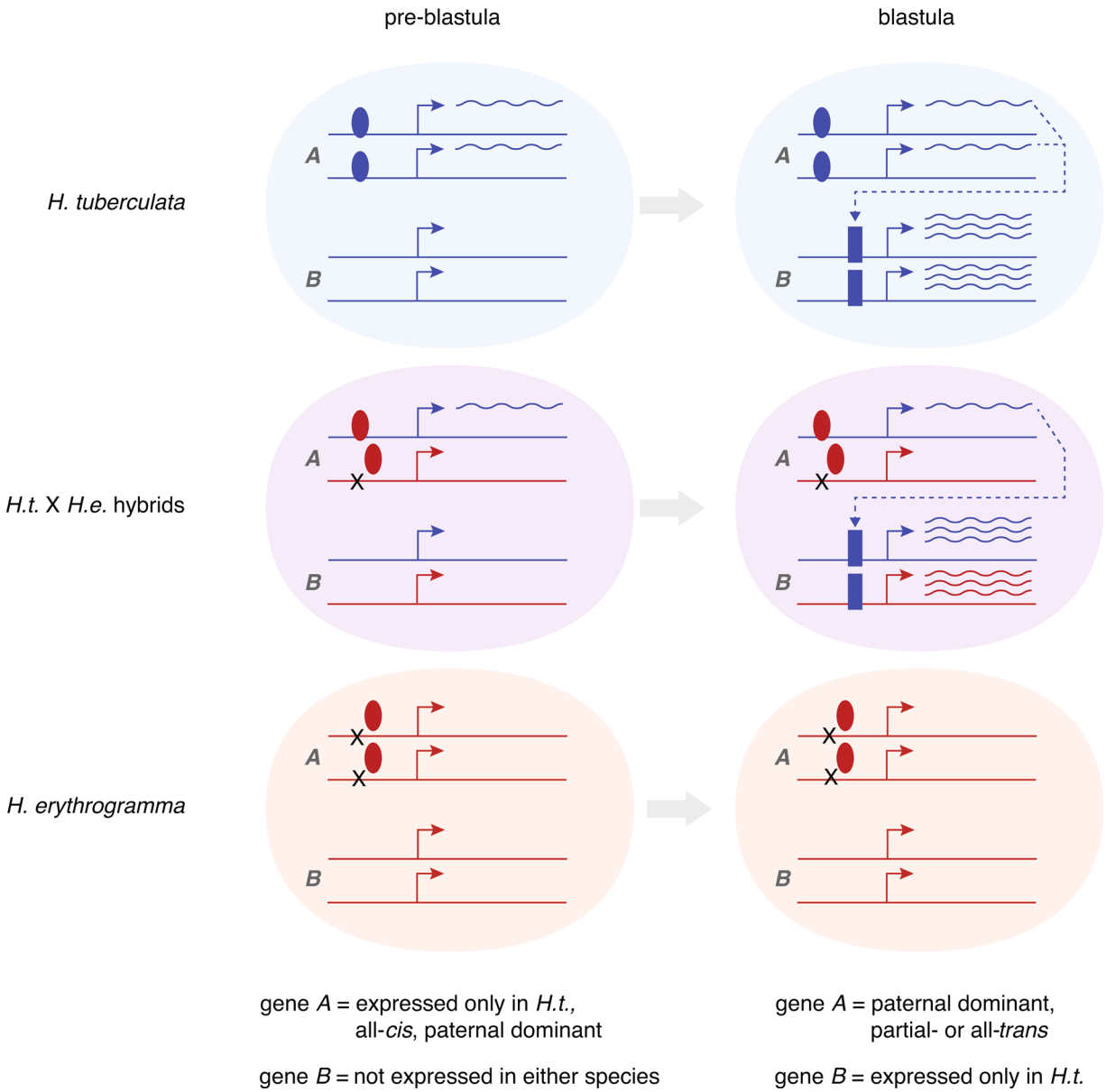
Extended Data Fig. 1 | Interpretation of inheritance mode (dominance) classification. The scatterplot in the center corresponds to Fig. 2d of the main text. Each dot corresponds to a single gene. The x-axis is the difference in expression level in hybrids and *H. erythrogramma* (*He*; maternal), while the y-axis is the difference in expression level in hybrids and *H. tuberculata* (*Ht*; paternal), both $\log_2$ scale for visualization of the full dynamic range. The six bar plots surrounding the scatterplot illustrate hypothetical cases of expression for individual genes that would be classified as different modes of inheritance: expression levels are shown for both parental crosses (*He* and *Ht*) and for hybrids (Hy). If expression in hybrids is *not* statistically distinguishable from *He* but *is* statistically distinguishable from *Ht*, mode of inheritance is classified as *He* dominant. The reverse case is classified as *Ht* dominant. If expression in hybrids is intermediate between *He* and *Ht*, the case is classified as additive (codominant). If expression in hybrids is higher than both parents, the case is classified as overdominant, while if lower than either parent it is classified as underdominant. Genes not differentially expressed between the parental crosses and are classified as conserved. The formal criteria for these classifications are presented in Supplementary Table 3. Note that although the arrows point to only one cloud of points, He dominant, Ht dominant, and additive each has a counterpart distributed symmetrically around the origin; underdominant and overdominant are symmetrical around the origin along the diagonal.



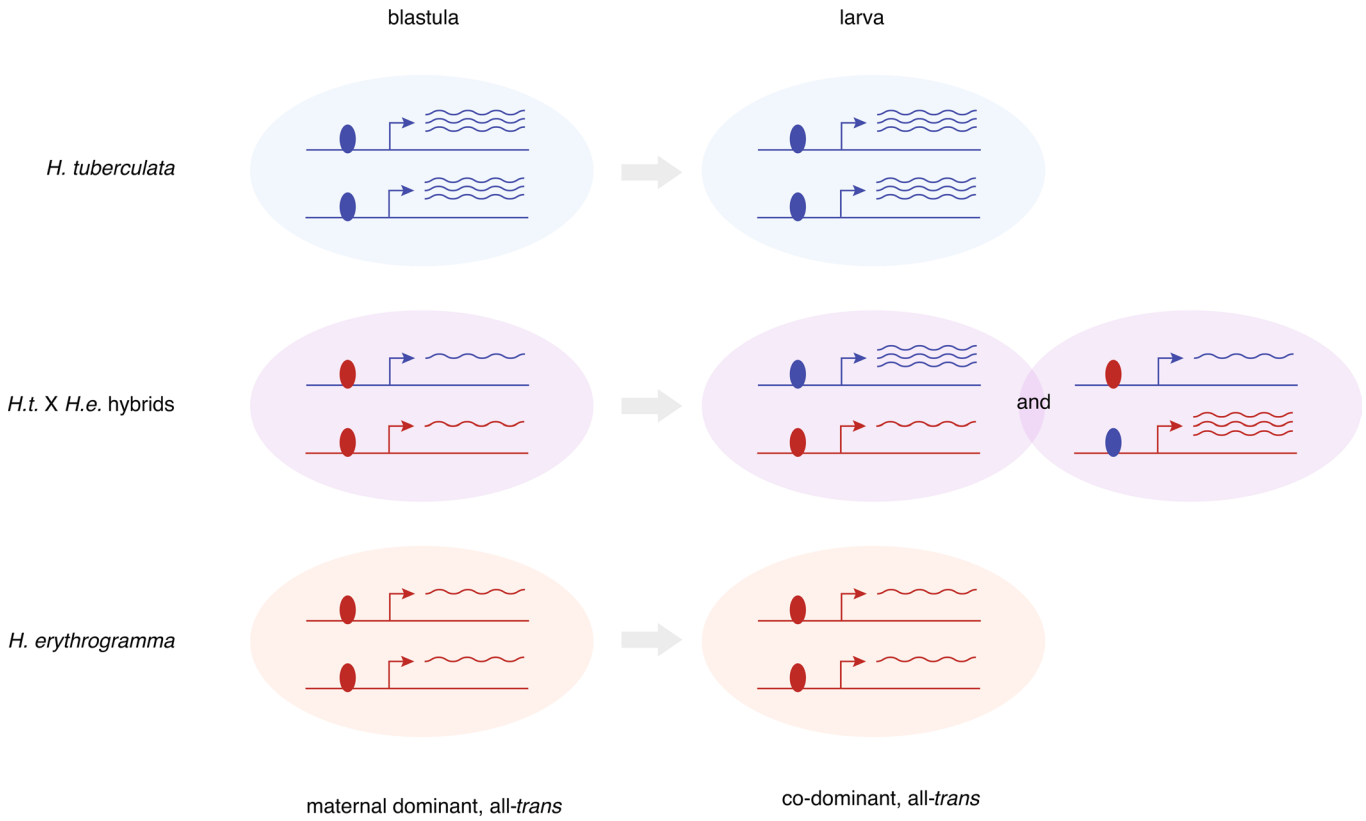
Extended Data Fig. 2 | *cis* and *trans* genetic influences on transcript abundance. Cells in hybrids contain regulatory machinery from both parents (blue and red ovals, representing orthologous transcription factors). Because transcription factors diffuse, they interact with both genomes in hybrids. Genetic differences between species allow assignment of reads in hybrids to parent of origin and indicates whether they are present at equal levels or not (allelic balance or imbalance, respectively). Lower left: allelic imbalance in hybrids indicates a genetic influence on an expression difference, since both alleles are in the same *trans* environment. Further, the genetic influence must be in linkage disequilibrium with the gene in question or it would not produce a consistent influence due to recombination and independent assortment. Lower center: allelic balance in hybrids indicates the absence of *cis* genetic effects, because the genetic basis is not in linkage disequilibrium with the gene. If the gene is also consistently expressed at different levels in the two parental species, however, it must have a genetic basis which must therefore be in *trans*. In such cases, something in the biochemical environment influences expression level independent of any differences in *cis*-regulatory elements between species. Possibilities include differences binding specificity or expression level of a transcription factor or presence of a co-factor that influences the expression of the gene. Comparing expression in hybrids that in the parental species provides additional information. If expression in hybrids is intermediate between that in the two parental species and there is no allelic imbalance, the case is classified as all *trans*. Lower right: other cases imply a mix of *cis* and *trans* influences, one of which is shown here. The *cis* and *trans* influences can reinforce or counteract and may not be of the same magnitude. For additional discussion, see references[2,5–9].



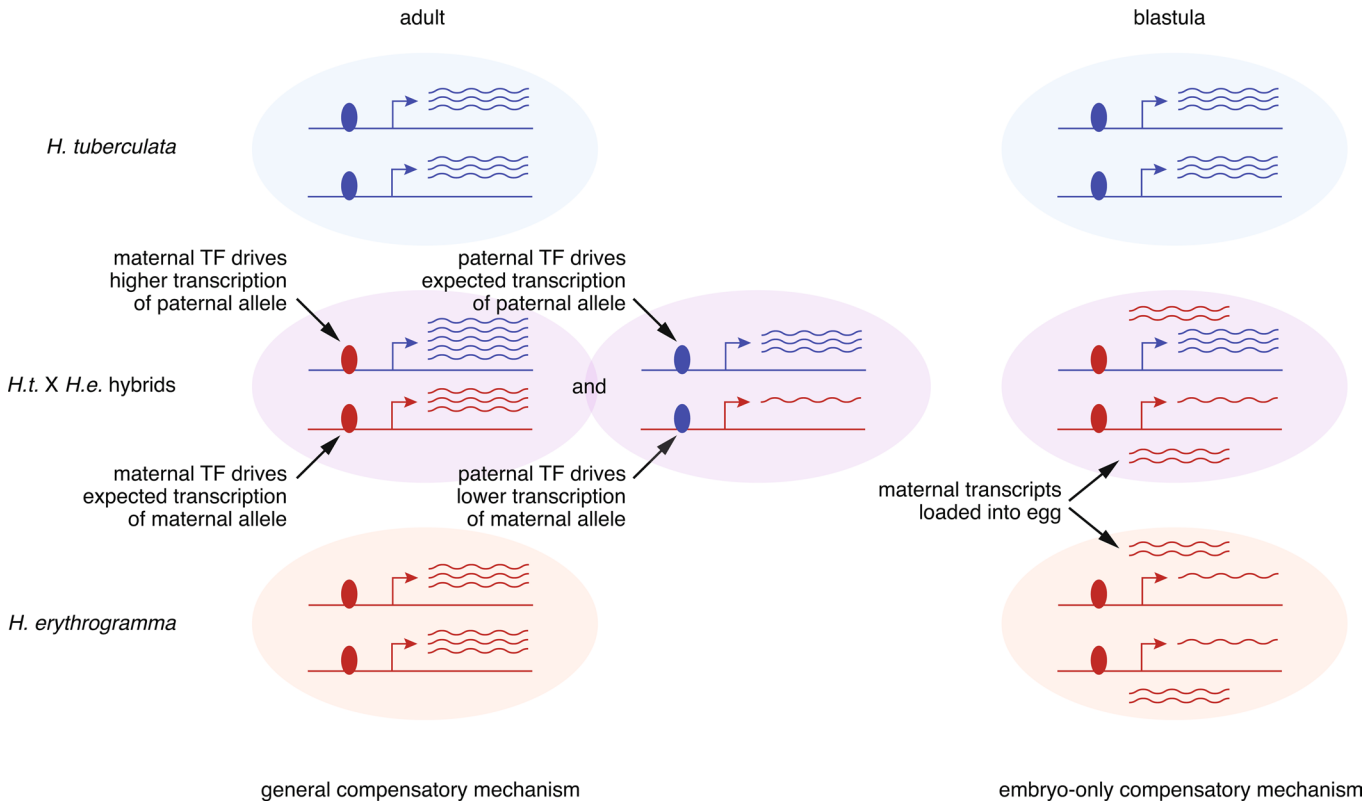
Extended Data Fig. 3 | Interpretation of regulatory mode (*cis/trans*) classification. The scatterplot in the center corresponds to Fig. 2e (main text), with each dot representing a gene. The x-axis is the ratio of expression levels in the two species, while the y-axis is the ratio of expression levels from the two alleles in hybrids (Ae = *H. erythrogramma* allele, At = *H. tuberculata* allele); note $\log_2$ scales. Bar plots show hypothetical cases illustrating classification of genes by regulatory mode: expression levels are shown for parental crosses (He = *H. erythrogramma*; Ht = *H. tuberculata*), hybrids (Hy), and alleles within hybrids (Ae = He allele; At = Ht allele). Cases where allelic expression in hybrids is not statistically different from the parental crosses are classified as all *cis*. If expression in hybrids is the average of parental crosses and the two alleles are expressed at the same level, the case is classified as all *trans*. When expression in hybrids is the average of the two parents but that of at least one allele lies closer to the same-species cross, the case is classified as *cis* + *trans*. When the case is similar to the previous but alleles in hybrids are closer to the opposite parental cross, it is classified as *cis* x *trans*. Compensatory cases involve no difference between same species crosses but alleles in hybrids show differences of similar magnitude and opposite sign. Note that although the arrows point to only one cloud of points, each has a counterpart distributed symmetrically around the origin; in addition, *cis* x *trans* has four clouds of points. For each of the two or four clouds of points, one or more of the bars in the hypothetical expression level plots will change (see formal criteria in Supplementary Table 4 for details).



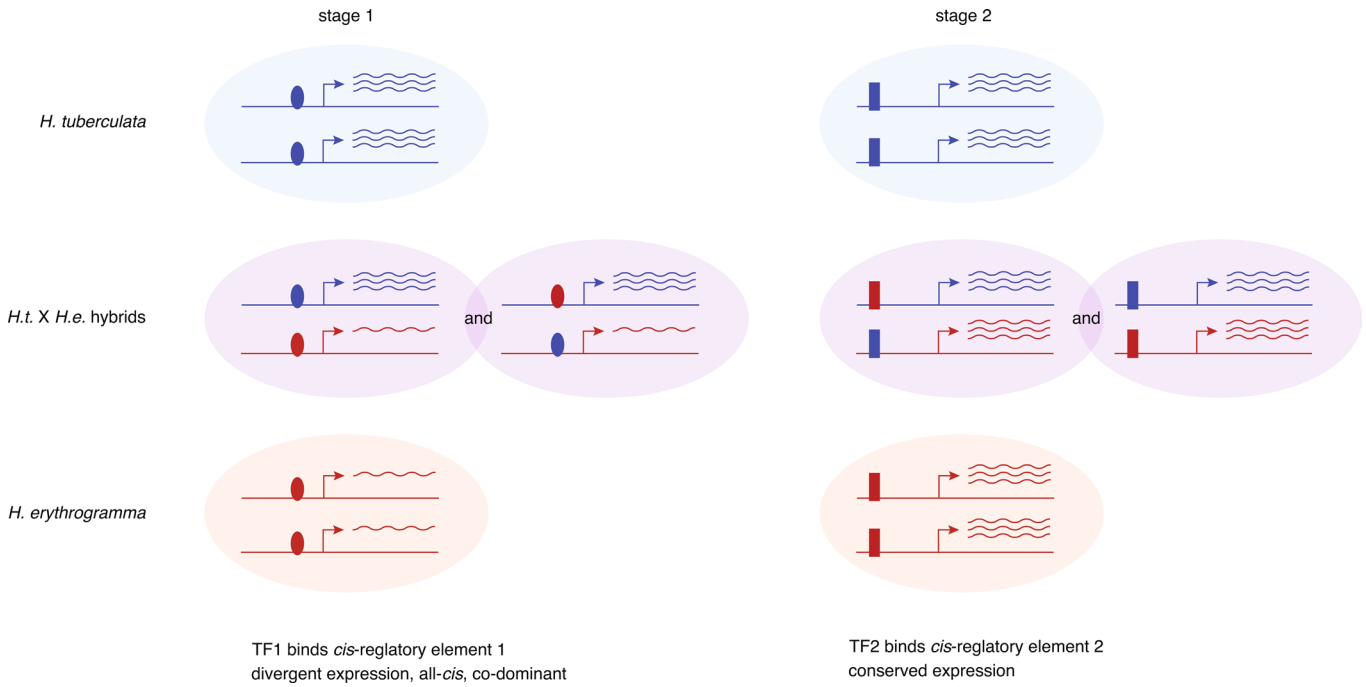
Extended Data Fig. 4 | Inference of early paternal influences on transcription. At the blastula stage, 42 genes in hybrids are classified as having paternal dominant expression with all or partly *trans* influences. This combination of classifications implies activity of the paternal allele of another gene prior to the blastula stage. Several plausible molecular mechanisms could produce this result. An example is sketched out here involving two genes: *gene B*, classified as paternal dominant and all or part *trans* at blastula stage, and *gene A*, which encodes a transcription factor (rectangles) that activates transcription of *gene B* at blastula stage. In this scenario, a different transcription factor is present in the pre-blastula embryo (ovals). It binds to a *cis*-regulatory element to activate transcription of *gene A* in *H. tuberculata* (blue ovals), but the orthologous transcription factor in *H. erythrogramma* (red ovals) cannot bind to the *cis*-regulatory element due to a mutation (marked X). In hybrids, the transcription factor binds to the *cis*-regulatory element of the paternal allele and activates its transcription, but still cannot bind to the *cis*-regulatory element of the maternal allele. As a consequence, the transcription factor produced from *gene A* is present in blastula stage embryos of *H. tuberculata* and hybrids (blue rectangles), but not those of *H. erythrogramma*. The transcription factor from *gene A* (rectangles) goes on to activate transcription of *gene B* in embryos of *H. tuberculata* and hybrids, but not embryos of *H. erythrogramma*. This transcription factor must interact with the *cis*-regulatory elements of both the maternal and paternal allele in hybrids, as this case was classified as all or part *trans*.



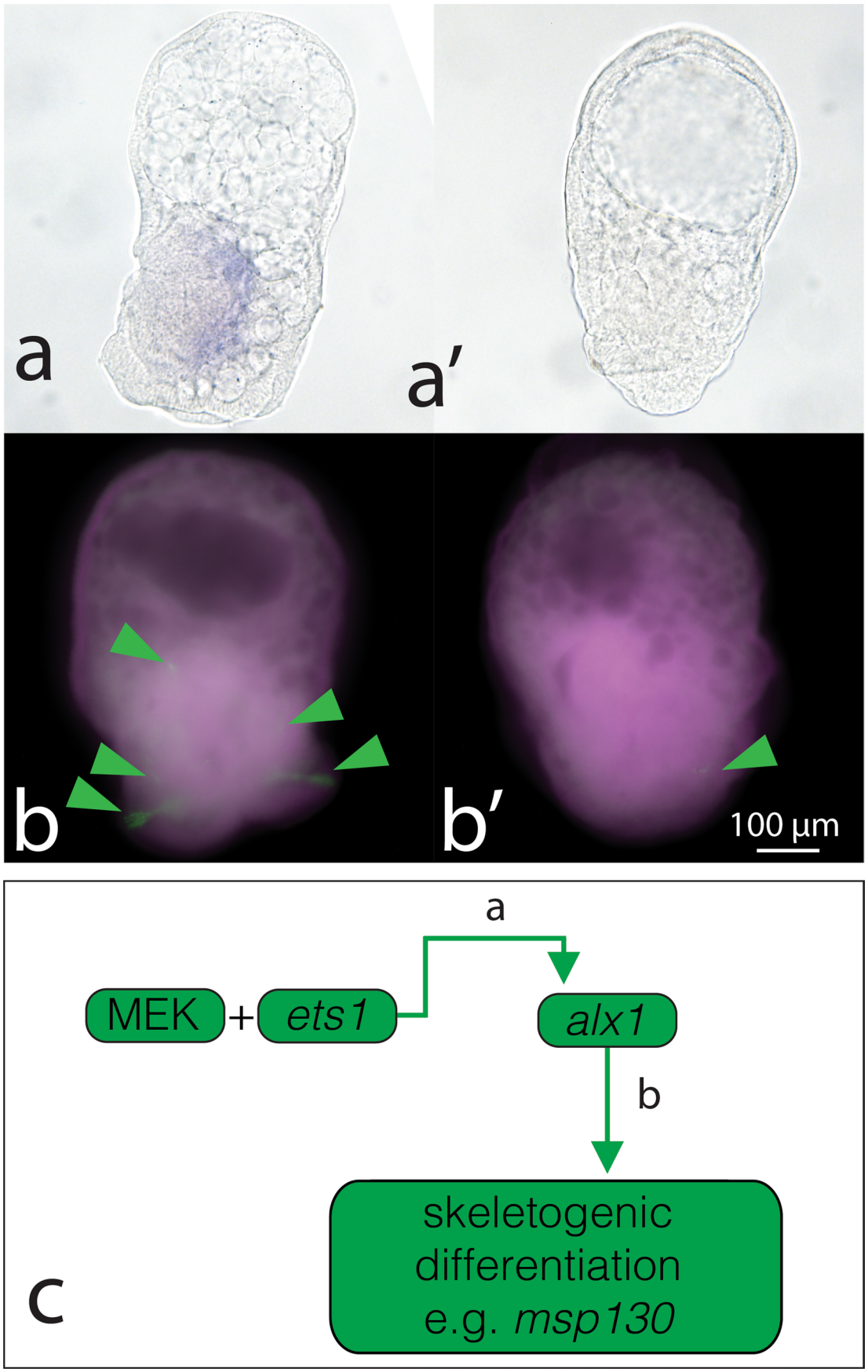
Extended Data Fig. 5 | Model for transient early embryonic asymmetry of *trans* effects. Several situations can produce a *trans* genetic effect. One, shown here, involves mutations in a transcription factor that alters the level of expression of a target gene. Possibilities include changes in binding affinity, changes in protein:protein interactions that stabilize its interaction with the binding site, and a variety of indirect effects mediated through co-factors or chromatin state. The essential point is that binding of orthologous transcription factors produces different levels of transcription. Since most of the transcriptional machinery in early embryos is maternally derived, the paternal orthologue will likely not be present in early hybrid embryos because the sperm brings few if any transcription factors to the zygote (middle left). Thus, even though the *cis*-regulatory elements of both species can bind the transcription factor, expression in hybrids is low from both alleles because the only orthologue of the transcription factor present is the maternal version, which activates expression more weakly. Since *trans*-acting factors such as transcription factors often exert an influence on the expression of many target genes, a few cases like the one outlined here (or scenarios with similar effects such as presence of a more potent repressor in eggs of *H. erythrogramma*) could readily account for the strong bias towards reduced expression for genes showing *trans*-only genetic influences in hybrids. By gastrula stage, the paternal genome is being widely transcribed and both orthologues of the transcription factor are present in the cytoplasm. Assuming the simplest and most common case that the gene encoding the transcription factor itself shows allelic balance and the protein has a similar DNA binding characteristics, the two orthologues will bind each *cis*-regulatory element allele with equal frequency, erasing the asymmetry (right). Again, more complex scenarios are possible.



Extended Data Fig. 6 | Model for asymmetry of compensatory cases based on a proposed novel underlying mechanism. Left: prior studies have generally interpreted the underlying mechanism in compensatory cases as involving equal and opposite *cis* and *trans* effects. In one simple scenario (shown here), both orthologues of the transcription factor bind to both orthologues of the *cis*-regulatory element, but with different kinetics or different consequences for transcriptional activation due to mutations in the transcription factor coding sequence and in the *cis*-regulatory element. The net level of transcripts remains the same in hybrids and both parents. A variety of specific combinations of *cis* and *trans* effects could result in no net difference in transcript abundance. In the scenario shown, the altered regulatory element and altered transcription factor interact with each other but this need not be the case. Right: a different underlying mechanism is possible when considering allelic imbalance in early embryos, due to the possibility of evolutionary changes in the presence or level of maternal transcripts that are loaded into the egg. If zygotic expression of a gene is reduced but maternal transcripts make up the difference, the net result will be conserved gene expression. In the scenario shown here, maternal mRNA is not loaded into the eggs of *H. tuberculata* but it is loaded in *H. erythrogramma*. This effect is necessarily asymmetric because the egg contributes vast quantities of transcripts to the zygote while the sperm contributes next to none. This mechanism would generally be possible only in early embryos, as it is a direct effect of evolutionary changes in maternal gene expression and maternal transcripts typically do not persist into later development.



Extended Data Fig. 7 | Model for differences among developmental stages in the genetic basis for the evolution of expression of the same gene. During development, the *trans* environment (primarily transcription factors and co-factors) changes but the *cis*-regulatory elements remain the same. Note that transcription factors diffuse and thus can interact with regulatory elements in both genomes ('and' indicates both possible states). Often, different transcription factors influence gene expression at different stages. Given that different transcription factors typically interact with different binding sites in the genome or even entirely different *cis*-regulatory elements, a mutation in one transcription factor binding site may only affect the interaction with one transcription factor at one stage of development. Later, this same mutation has no influence because that particular transcription factor is no longer present and the ones that are present bind to other sites within the same or a nearby *cis*-regulatory element. The net result is that the mutation only affects development at one stage. This becomes important for thinking about adaptation because a mutation that affects gene expression at one stage of development need not affect other stages – meaning its effects are not highly pleiotropic. Note that the precise scenario sketched out above is just one of many different possible mechanisms for producing an evolutionary difference at one stage of development but not at another. Protein modifications to transcription factors such as phosphorylation, changes in the presence of co-factors, competition with another transcription factor for the same binding site, and local modifications to chromatin are some of these alternative mechanisms. In all cases, the consequences of the mutation for expression of the gene in question is context-dependent, with the context changing during development.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Experimental evidence for evolutionarily conserved linkages within the skeletogenic GRN. The MEK-ERK pathway, via the transcription factor Ets1/2, is required for expression of the skeletogenic gene *alx1*. This locus encodes the transcription factor Alx1, which is required for normal expression of the gene *msp130*, which is a terminal differentiation gene that encodes a structural protein of the biomineral matrix. **a** and **a':** Localization of *alx1* transcripts by in situ hybridization in vehicle control-treated (**a**) and MEK-ERK inhibitor-treated (a') *H. erythrogramma* larvae reveal stained cells in the juvenile rudiment (in a but not a'). **b** and **b':** Indirect immunofluorescence localization of the Msp130 protein product in *H. erythrogramma* larva injected with a standard control morpholino (**b**) or a translation-blocking morpholino targeting *alx1* (b'). **c**. Diagram showing two key gene regulatory connections that are shared between *H. erythrogramma* and the ancestral GRN.

nature research

Corresponding author(s): Gregory A Wray

Last updated by author(s): Jul 2, 2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	No software was used for data collection
Data analysis	FastQC v0.11.5, seqtk, Trimmomatic v0.36, Stampy v1.0.28, HyLiTE, RStudio, R, DESeq2, limma, vsn, genefilter, ggplot2, reshape2, dplyr, ggalluvial, pheatmap, ggpubr and other packages listed in the supplementary codes were used for data analysis

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence reads for this project are available from SRA as accession SRP229522. The count tables are available from the project's GitHub repository (https://github.com/Wray-Group-at-Duke/HybridsProject_NEE).

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	Blastula, gastrula and larva are three key developmental stages where gene regulation plays important role for cell fate specification and morphogenesis. So, we collected samples from these 3 key stages to study the genetic basis for gene expression divergence between the two species of sea urchins.
Data exclusions	We designed to have 27 RNA-seq libraries. Unfortunately 1 sample didn't yield enough RNA for sequencing so we excluded that sample. The remaining 26 samples were used for this manuscript and meet the requirement for statistical tests.
Replication	To use DESeq2 for DE analysis, 3 biological replicates were recommended and 2 biological replicates were the minimal requirement. We have 3 biological replicates for all samples except for one sample where only 2 biological replicates were able to obtained.
Randomization	After the sequencing libraries were prepared, samples were loaded to the sequencing flow cells randomly to avoid batch effects related to sequencing chips.
Blinding	Blind is not relevant to our study because we need to know the exact sample information to group the samples for statistical tests.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Antibodies

Antibodies used	1D5 monoclonal mouse IgM antibody developed for sea urchin by the McClay lab.
Validation	1D5 has been used as a skeletogenic mesoderm marker in multiple sea urchin species: McClay, DR.; Cannon, GW.; Wessel, GM.; Fink, RD.; Marchase, RB. Patterns of antigenic expression in early sea urchin development. In: Jeffrey, WR.; Raff, RA., editors. Time, space, and pattern in embryonic development. Vol. 69. Alan R. Liss, Inc; New York: 1983. p. 157-169. microRNA-31 modulates skeletal patterning in the sea urchin embryo. Development (2015) 142, 3769-3780 doi:10.1242/dev.127969 Sub-circuits of a gene regulatory network control a developmental epithelial-mesenchymal transition. Development (2014) 141, 1503-1513 doi:10.1242/dev.101436 Frizzled1/2/7 signaling directs β-catenin nuclearisation and initiates endoderm specification in macromeres during sea urchin embryogenesis. Development 139, 816-825 (2012) doi:10.1242/dev.072215 Skeletogenesis by transfated secondary mesenchyme cells is dependent on extracellular matrix-ectoderm interactions in Paracentrotus lividus sea urchin embryos. Develop. Growth Differ. (2007) 49, 731–741

Localization and expression of msp130, a primary mesenchyme lineage- specific cell surface protein of the sea urchin embryo. Development 101, 255-265 (1987)