Genome-wide use of high- and low-affinity Tbrain transcription factor binding sites during echinoderm development

Gregory A. Cary[a], Alys M. Cheatle Jarvela[a,1], Rene D. Francolini[a], and Veronica F. Hinman[a,2]

[a]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA 15213

Sea stars and sea urchins are model systems for interrogating the types of deep evolutionary changes that have restructured developmental gene regulatory networks (GRNs). Although *cis*-regulatory DNA evolution is likely the predominant mechanism of change, it was recently shown that Tbrain, a Tbox transcription factor protein, has evolved a changed preference for a low-affinity, secondary binding motif. The primary, high-affinity motif is conserved. To date, however, no genome-wide comparisons have been performed to provide an unbiased assessment of the evolution of GRNs between these taxa, and no study has attempted to determine the interplay between transcription factor binding motif evolution and GRN topology. The study here measures genome-wide binding of Tbrain orthologs by using ChIP-sequencing and associates these orthologs with putative target genes to assess global function. Targets of both factors are enriched for other regulatory genes, although nonoverlapping sets of functional enrichments in the two datasets suggest a much diverged function. The number of low-affinity binding motifs is significantly depressed in sea urchins compared with sea star, but both motif types are associated with genes from a range of functional categories. Only a small fraction (∼10%) of genes are predicted to be orthologous targets. Collectively, these data indicate that Tbr has evolved significantly different developmental roles in these echinoderms and that the targets and the binding motifs in associated *cis*-regulatory sequences are dispersed throughout the hierarchy of the GRN, rather than being biased toward terminal process or discrete functional blocks, which suggests extensive evolutionary tinkering.

Tbrain | echinoderm | binding site affinity | ChIP-seq

O ne of the most striking revelations that arose from the emergence of the field of "EvoDevo" in the late 1980s was the concept of functional equivalence [i.e., that orthologous transcription factors (TFs) from vastly different species will functionally compensate for each other when experimentally substituted in vivo]. This concept implies that the function of these orthologous proteins has remained essentially unchanged over hundreds of millions of years. The experimental observations that demonstrated this concept (e.g., refs. 1 and 2) fit neatly with theoretical predictions that TFs, which regulate many target genes in multiple spatiotemporal contexts, will be highly constrained, and therefore any change in function will have wide-sweeping changes that are unlikely to pass the filter of natural selection (3). Thus, it is argued, changes in developmental gene regulatory networks (GRNs) must arise from alterations to the DNA of the *cis*-regulatory targets bound by these factors. In other words, the biochemical function of a TF (i.e., the preferred sequence motif to which the factor binds) remains conserved, whereas its developmental function, the genes that are regulated, evolves more rapidly through *cis*-DNA evolution of the target genes.

Modern and genome-wide assessments have largely borne out the results of these earlier experiments and theoretical predictions (4, 5). However, there has been a recent and growing interest in understanding how TFs might evolve changed

biochemical functions (6). Conceptually, this investigation must entail understanding how these proteins evolve changes that limit the dramatic effects of pleiotropy and can include evolving differences with small phenotypic effects and independently changing individual subsets, or modules, of functions (7).

A recently recognized, and surprising, source of modularity has been DNA binding itself. The potential prevalence of this source of variation was realized only after protein-binding microarray technology, which universally assesses DNA-binding (8, 9), provided a high-resolution description of binding-site preferences. Studies using this technology have shown that many TF binding motif preferences can be described by multiple position weight matrices (PWMs). These PWMs are commonly called primary and secondary motifs, where the primary is the most preferred and often higher-affinity motif. An increasing number of studies are showing that low-affinity secondary sites can execute particular functions, including providing specificity, controlling the timing of gene expression, and even mediating activation vs. repressive regulation (10–12). Recent work from our laboratory has shown that Tbrain (Tbr), a T-box TF, has evolved modular preferences for binding sequence recognition (13). We showed that the primary motif recognized by the sea star (*Patiria miniata*) and sea urchin (*Strongylocentrotus purpuratus*) Tbr proteins were extraordinarily similar, and this similarity extended even to the mouse ortholog Eomesodermin. Sea stars and sea urchins are representatives of two classes of echinoderms, which last shared an ancestor >450 million years ago in the early Ordovician (14), whereas the common ancestor of echinoderms and mice sits at the base of all deuterostomes. Our finding, therefore, is in keeping with the theoretical prediction that highly pleiotropic factors are constrained from evolving changed motif preferences over even immense evolutionary distances. Importantly, however, we also showed that the factors from mouse and sea star recognized different secondary motifs, whereas the sea urchin protein showed no

preference for any secondary motif. These results therefore demonstrate that these proteins have evolved a modular change of preference for a low-affinity binding motif.

The next challenge is to understand the developmental consequences of this modular use of evolvable binding motifs in these taxa. Sea urchin and sea star *tbr* genes have been partially functionally characterized (15–17), but we have no genome-wide understanding of their function. Therefore, we first determined the developmental function of Tbr in these taxa and asked whether these factors are indeed highly pleiotropic, whether they share developmental functions, and how these functions might fit within a developmental GRN. As a measure of developmental function, we assessed the binding of Tbr factors genome-wide in both the sea star and sea urchin at comparable developmental stages and identified genes likely to be regulated by Tbr at these sites. We then sought to determine the distribution of primary and secondary motif use associated with these taxa and the functions of identified genes. This approach allowed us to assess the evolution of the developmental role of Tbr in these taxa and to understand how this developmental function then interfaces with the evolution of the biochemical function of motif use.

Results

Tbr function in sea urchins is reasonably well understood, because the factor acts within the GRN for specification of the primary mesenchyme lineage, which is considered the most well-described GRN for the specification of any cell type (15). In the sea urchin, *tbr* is localized exclusively to this lineage (ref. 18 and Fig. 1). However, there has been no systematic assessment of Tbr binding sites, and there is no *cis*-regulatory module (CRM) yet known to be directly bound by Tbr. These data are needed to understand the intersection of binding site and developmental function of Tbr. Therefore, we used ChIP-sequencing (ChIP-seq) to identify *S. purpuratus* Tbr (SpTbr) binding sites genome-wide. Whole-mount immunofluorescence (IF) using our specific antibody (13) revealed that SpTbr is localized to primary mesenchyme cells (PMC) at 24 h (Fig. 1B), which is the same cell population that expresses the *Sp-tbr* transcript (18), although the transcripts are known to be maternally deposited. ChIP-seq was performed by using this antibody to specifically enrich SpTbr-bound chromatin at this stage. Resulting reads were mapped to the genome, and peaks were called by MACS2 (*Materials and Methods*; see Fig. S1 for experimental workflow). In total, we found 3,149 peak regions that were enriched by SpTbr ChIP across the genome compared with the chromatin-only input control.

Targets of Tbr in sea urchin have been detected by performing quantitative PCR and whole-mount in situ hybridization (WMISH) to analyze the effects of *Sp-tbr* morpholino antisense oligonucleotide (MASO) knockdown (15, 16). Known targets include *Sp-Nrl*, *Sp-FoxN2/3*, *Sp-FoxB*, *Sp-Erg*, and *Sp-Msp130*, which are curated within the PMC-GRN, as well as *Sp-Nebnph*, *Sp-Lisp1*, *Sp-C-lectin/PMC*, *SPU_018403*, and *Sp-Hypp_2998*. For 8 of these 11 targets, SpTbr ChIP peaks are detected within 75 kb of the gene (Table 1). In several cases, multiple genes that are clustered in the genome share a single peak. Peaks associated with several genes also overlap with assay for transposase-accessible chromatin-sequencing (ATAC-seq) chromatin accessibility peaks from similar-stage embryos (Table 1, Fig. 1 C and D, and Fig. S2) (see below). Although there is a MACS called peak on the same genomic scaffold as *Sp-FoxB*, it is 281 kb away from the body of the gene and, therefore, is less likely to represent a functional interaction between the binding site and the gene. This finding supports an indirect effect of SpTbr on *Sp-FoxB* expression, as suggested by the data supporting the GRN model. The only gene from the developmental GRN without an SpTbr ChIP-seq peak detected on the same scaffold is *Sp-Erg*. The identification of significantly enriched peaks near known SpTbr target genes supports the conclusion that the ChIP-seq protocol is capable of detecting genomic targets of SpTbr binding.

To reduce potential false-positive peaks from our analysis, we additionally restricted our peak set to those associated with an annotated gene. It is not uncommon, in the sea urchin genome, for characterized CRMs to be 10 kb or more from the coding gene, both upstream and downstream of the transcription start (e.g., refs. 19 and 20). Given that the majority of predicted SpTbr targets for which a ChIP peak is detected are observed within 75 kb of that peak (Table 1), we elected to further filter the SpTbr ChIP peaks to those within 75 kb of a gene annotation. This distance corresponds to approximately three times the average intergenic distance predicted for the sea urchin genome (21). This method further filtered the set of SpTbr peaks to 1,952 (61.9% of all peaks) (Fig. S1). This approach may also eliminate genuinely functional peaks, but because our goal is to associate peaks with putative function, those not obviously associated with a gene are not required in our later analyses.

To further corroborate these peaks, we used unpublished, but publicly accessible, ATAC-seq chromatin accessibility data. The ATAC assay identifies open chromatin based on the propensity for in vitro transposase integration genome-wide. These data are available as JBrowse tracks at Echinobase.org (22), representing MACS2 called peaks of ATAC-based chromatin accessibility data from three replicate measurements of embryos at 24 hours post fertilization (hpf). We used the regions common to all three replicate ATAC samples and filtered the detected 3,149 peaks for only those that overlap the consensus region of ATAC sensitivity. A total of 1,542 SpTbr peaks (49.0%) overlap ATAC peaks. This fraction of ChIP overlap with chromatin accessibility is comparable to what has been reported in the literature for
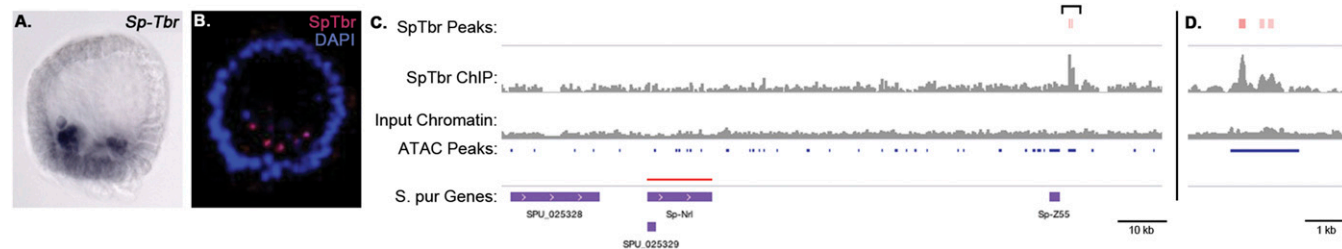

Fig. 1. Anti-SpTbr antibody validation. (A) WMISH for *Sp-tbr* in *S. purpuratus* at 24 hpf showing expression localized to the ingressed primary mesenchyme cell population. (B) IF localization of SpTbr in an *S. purpuratus* embryo; staining is also observed in the primary mesenchyme cells. (C) Example data from the SpTbr ChIP-seq experiments showing the *Sp-Nrl* locus. The top track shows regions called as SpTbr ChIP peaks by MACS2 analysis, where the color intensity of the box corresponds to peak fold enrichment. The two tracks in gray are bedgraph data output from MACS2 showing sequence tag pileups from ChIP and local lambda from input chromatin datasets. The blue bed track represents consensus ATAC-seq peaks for three replicate 24-hpf datasets (hosted at Echinobase.org), and the bottom track is the annotated *S. purpuratus* gene models. The target gene is highlighted by a red line. (Scale bar: 10 kb.) The area surrounding the detected ChIP peaks, indicated by a square bracket, is expanded in *D* to show finer resolution of ChIP and ATAC intersection. (Scale bar: 1 kb.)

Table 1. ChIP-seq peaks near predicted targets of SpTbr

Glean ID	Gene name	Peak_ID	ATAC overlap	Distance, kb
SPU_013698*	*Sp-Nrl*	tbr_macs_peak_2546	Yes	72
SPU_015243*	*Sp-FoxN2/3*	tbr_macs_peak_2415	Yes	63
SPU_002088*	*Sp-Msp130*	tbr_macs_peak_1696	No[†]	36
SPU_013821*	*Sp-Msp130_1*			34
SPU_004551*	*Sp-FoxB*	(tbr_macs_peak_3011)	(No)	(281)
SPU_018483*	*Sp-Erg*	—	—	—
SPU_009123[‡]	*Sp-Nebnph*	tbr_macs_peak_2337	No	48
SPU_009124[‡]	*Sp-Lisp1*			28
SPU_018403[‡]	None	tbr_macs_peak_977	Yes	18
SPU_018407[‡]	*Sp-Hypp_2998*			35
SPU_027906[‡]	*Sp-C-lectin/PMC1*	—	—	—

Genes predicted to be targets of SpTbr either from the curated sea urchin developmental GRN dataset or from Rafiq et al. (16) were surveyed to determine whether a SpTbr ChIP peak was detected on the same scaffold. If a detected SpTbr ChIP peak was detected, the presence of an overlapping ATAC-seq peak was scored, and the distance from the peak to the body of the gene is reported. In several cases, multiple genes shared a single peak (e.g., *Sp-Msp130* and *Sp-Msp130_1* or *Sp-Nebnph* and *Sp-Lisp1*). These genes are grouped together by like shading; the peak ID is reported once, and the distance from the peak to each gene is indicated. Genes in this set for which no peak is detected are marked with —. The peak associated with *Sp-FoxB* is indicated in parentheses because its extreme distance to the body of the gene—281 kb—makes it unlikely that this peak is involved in the direct regulation of *Sp-FoxB*.
*SpTbr targets predicted from PMC-GRN.
[†]Although there is no ATAC-seq peak overlapping the SpTbr peak at the *Sp-Msp130* locus, there is an ATAC peak immediately adjacent, as shown in Fig. S2.
[‡]SpTbr targets predicted by Rafiq et al. (16).

other organisms (e.g., ref. 23). In total, 1,492 SpTbr peaks both overlap an ATAC peak and are positioned within 75 kb of an annotated gene (Fig. S1). These peaks now represent the highest-stringency set that we use for further analyses.

To predict the function of SpTbr genome-wide, we next performed Gene Ontology (GO) term enrichment analysis based on the annotation of the 1,052 genes that are closest to the 1,492 SpTbr peaks. The most significantly enriched terms among these genes were found to be "nucleic acid binding TF activity" (63 genes, $P = 2.3e^{-22}$), "MAP kinase phosphatase activity" (5 genes, $P = 1.3e^{-4}$), "proteoglycan binding" (3 genes, $P = 0.02$), and "regulation of signal transduction" (10 genes, $P = 0.04$) (Fig. 2 and Dataset S1). These enrichments highlight the known roles of SpTbr in regulating both gene expression and cell signaling, both central mechanisms of information flow through GRNs. Indeed, the genes identified are significantly enriched for genes present in the current GRN model (49 genes, hypergeometric $P = 2.3e^{-45}$). The MAPK pathway, in particular, is known to be critical for PMC ingression and patterning during the stage sampled here (24, 25), although to our knowledge Tbr has not been previously been implicated as a direct regulator of this pathway. It has been demonstrated, however, that Tbr is required for the epithelial–mesenchymal transition associated with PMC ingression (26). These data suggest the potential for regulatory feedback between MAPK signaling and PMC-GRN regulatory genes, where, for example, SpTbr may regulate aspects of the MAPK pathway that, in turn, have been shown to phosphorylate and regulate SpEts1 at the top of the network and act as a regulatory input to SpTbr in PMCs (25, 26).

Having defined a set of high-confidence SpTbr binding sites genome-wide and associations with adjacent genes in sea urchins, we next shifted our focus to ascertaining the targets of sea star Tbr (*P. miniata* Tbr; PmTbr). Gastrulation in the sea star begins with the invagination of the vegetal plate (there is no ingressing primary mesenchyme) at ~30 hpf (27). At this stage, PmTbr is localized throughout endomesoderm, as well as at lower levels throughout the ectoderm (13), and therefore the ortholog is distributed far more broadly within the sea star than the sea urchin embryo.

The sea star anti-Tbr antibody was generated against a peptide specific to PmTbr, as described in ref. 13. Our initial analysis of PmTbr ChIP-seq peaks revealed a statistical overrepresentation of predicted repetitive elements (details in *SI Materials and Methods*). Sequencing reads that aligned to predicted RTE-2–like elements, which conspicuously and significantly overlapped with the initial peak set, were removed before subsequent peak detection. Ultimately, 13,977 peaks were called by MACS2 using the filtered read set, of which 9,164 (65.6%) were within 75 kb of an annotated gene. We detected more peaks in the PmTbr ChIP-seq than in the SpTbr ChIP-seq (9,164 vs. 1,952), although a similar percentage were associated with a gene (65.6% vs. 62.0%). This observation is consistent with the larger domain of PmTbr in sea stars, throughout the endomesoderm and ectoderm, compared with the highly restricted localization of SpTbr to the small number of PMCs in sea urchins.

The only previously known direct target of PmTbr is *Pm-Otx* (28). The previously functionally characterized CRM is completely overlapped by a significant MACS-called ChIP-seq peak (Fig. 3A). Furthermore, *Pm-Delta* has previously been shown to be sensitive to PmTbr morpholino knockdown, and in the ChIP-seq data, we found a significant MACS-called PmTbr peak 11 kb away from the *Pm-Delta* gene (Fig. 3B). Although there are fewer known targets for PmTbr in sea star compared with sea urchin, we found ChIP peaks proximal to both known targets, indicating that the PmTbr ChIP-seq dataset is capable of detecting genomic binding sites of PmTbr.

In the absence of available sea star ATAC-seq data to corroborate peaks, and given the limited known functions of sea star Tbr, we performed RNA-sequencing (RNA-seq) on *Pm-tbr* knockdowns to enable further filtering of PmTbr peaks for functionality. We used RNA collected at early gastrula from Tbr antisense morpholino injected embryos (anti-Tbr MASO) compared with control morpholino injected (control). Genes significantly differentially expressed [false discovery rate (FDR) < 0.05] between control and anti-Tbr MASOs in biological triplicate were detected. In total 2,562 genes (9.3% of all expressed genes) were found to be significantly differentially expressed and 1,165 of these (45.5%) were found to be within 75 kb of a PmTbr ChIP-seq peak (1,105
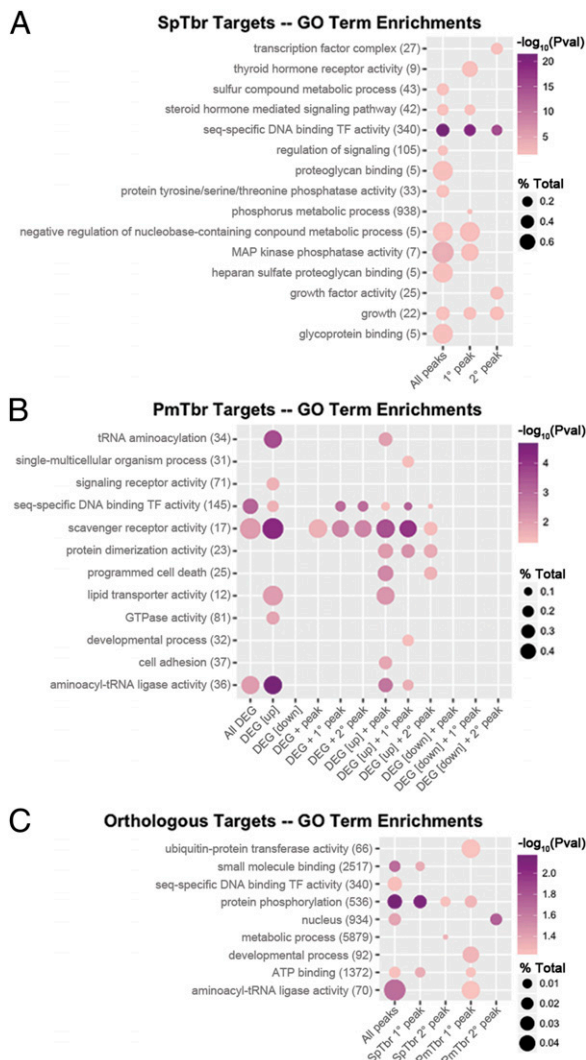
A   SpTbr Targets -- GO Term Enrichments

B   PmTbr Targets -- GO Term Enrichments

C   Orthologous Targets -- GO Term Enrichments

Fig. 2. GO term enrichments for Tbr target genes. The significantly enriched GO terms for each set of targets is reported as a point where the color intensity corresponds to the corrected hypergeometric P value for that enrichment, and the area corresponds to the percent of all genes annotated with that term identified among the target set. (*A*) GO term enrichments for targets of SpTbr peaks, including all peaks as well as the subsets of peaks containing primary motifs (1° peak) or secondary motifs (2° peak). (*B*) GO term enrichments for PmTbr targets, including genes that are differentially expressed after Tbr knockdown (All DEG), DEGs that are up- or down-regulated (DEG [up] and DEG [down], respectively), DEGs that have a PmTbr peak assigned (DEG + peak), and DEGs with assigned peaks that contain primary motifs (1° peak) or secondary motifs (2° peak). (*C*) GO term enrichments for the 108 orthologous target genes found to be regulated by Tbr in both sea urchin and sea star datasets. These are subsetted to indicate whether the assigned peak in each dataset contains primary motifs (SpTbr 1° peak and PmTbr 1° peak) or secondary motifs (SpTbr 2° peak and PmTbr 2° peak).

peaks) (Fig. 4 and Fig. S1). This fraction of differentially expressed genes (DEGs) associated with ChIP peaks is within the range of what has been observed in other model systems (29, 30). Importantly, when we considered the union of DEGs with an associated peak we found both up- and down-regulated genes; 676 of 1,521 (44.4%) significantly up-regulated genes and 489 of 1,041 (47.0%) significantly down-regulated genes had an associated ChIP peak. This finding is strong evidence that PmTbr has the potential to act as both a direct repressor and an activator. This activity is consistent with known dual repressor/activator functions in related Tbox genes from other taxa (31, 32).

We examined the previously characterized targets of PmTbr (i.e., *Pm-Otx* and -*Delta*) and found that neither were significantly differentially expressed. *Pm-Otx* was found to be 1.3-fold down relative to the control, and *Pm-Delta* was found to be 1.7-fold down. The directionality of both of these genes was consistent with previous data, even though the changes were not statistically significant by RNA-seq. Thus, our RNA-seq analysis at this time point may miss some target genes we expect are regulated by PmTbr, but we chose to use the DEGs to help filter detected peaks because, again, we sought to minimize the potential for false positives among identified peaks.

To assess GO term enrichment to explore the function of PmTbr genome-wide, GO annotations were mapped from sea urchin genes to predicted sea star orthologs. First, likely sea star orthologs of sea urchin genes were predicted by using a reciprocal best blast hits (BBH) method. We examined term enrichments among the sea urchin orthologs to genes of interest from the sea star dataset—both the 2,562 genes that significantly changed after *Pm-tbr* knockdown, as well as the subset of 1,165 DEGs that also have a PmTbr ChIP peak detected within 75 kb. In general, GO term enrichments from the sea star dataset were less robust, owing to the fact that only a fraction of the genes examined (35–41%) have a reciprocal BBH mapping that informs the annotation of the sea star genes. Nonetheless, these enrichments suggest both conserved and divergent functions for Tbr in these two species. For example, the set of DEGs proximal to PmTbr ChIP peaks were enriched for some of the same ontology terms as was found in the sea urchin set—including "sequence-specific DNA binding TF activity" (17 genes, $P = 0.069$). The MAP kinase-associated terms were not found to be enriched in the sea star set, and there is no evidence to suggest that the MAP kinase pathway is used during sea star gastrulation. Additionally, there were several other terms enriched in the sea star set, not found in the sea urchin set, including "scavenger receptor activity" (six genes, $P = 0.023$), "apoptotic process" (six genes, $P = 0.14$), "cell adhesion" (six genes, $P = 0.35$), and "aminoacyl-tRNA ligase activity" (seven genes, $P = 0.069$) (Fig. 2 and Dataset S1). These terms are all more significantly enriched (all $P < 0.05$) among genes that were up-regulated by *Pm-tbr* knockdown (i.e., those genes that PmTbr would be predicted to repress), whereas there were no significant term enrichments for genes down-regulated by the *Pm-tbr* MASO knockdown. Thus, genes up-regulated by *Pm-tbr* knockdown are more functionally coherent, further supporting the hypothesis that PmTbr functions as a repressor.

We next compared binding site motif utilization in peaks from both species. We have previously defined the PWMs for Tbr primary and secondary site motifs in vitro. Although the sea star and sea urchin primary site PWM are quite similar, only sea star Tbr has an enhanced preference and affinity for the defined secondary site motif. We used the Finding Individual Motif Occurrences (FIMO) tool from the MEME-suite to scan both peak sets for the presence of the primary and secondary site motifs, and overlapping motif occurrences were filtered such that only the most significant motif was reported for each position. This filtering step is important because the primary and secondary site PWMs are partially overlapping, and frequently we would find both primary and secondary sites called at one position within a peak, although one with a much higher score and associated P value. Each peak was then classified based on the presence or absence of each motif. Although the SpTbr protein has a very low affinity for the secondary motif in vitro, the protein is nonetheless apparently capable of binding to sites with this motif in vivo, because we found several hundred SpTbr peaks in which only this motif was detected (Fig. S6). However, the proportion of sea urchin peaks with a secondary motif-containing site was significantly lower than the proportion of sea star peaks with a secondary site ($\chi^2\ P = 1.8e^{-4}$) (Fig. S6). The proportion of peaks with primary sites, conversely, was not significantly different
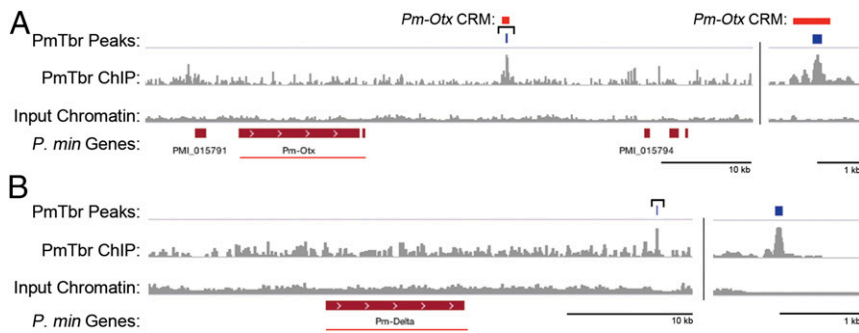
Fig. 3. Known PmTbr targets have proximal ChIP peaks. For both the *Pm-Otx* locus (*A*) and *Pm-Delta* locus (*B*), the top track shows regions called by MACS2 as PmTbr ChIP peaks, and corresponding ChIP sequence tag pileup and local lambda pileup (input) are shown as bedgraph tracks in gray. The bottom track shows annotated genes in the locus with the known target of PmTbr indicated by a red line. For each locus, an expansion of the area around the detected peak, indicated by a square bracket, is shown to the right of the vertical black line. The position of the previously described *Pm-Otx* CRM is indicated by the red box.

between the two datasets (χ^2: $P = 0.069$). Therefore, even though the secondary motif was present in the sea urchin dataset, it was present in significantly fewer of the detected peaks, suggesting a bias against use of this lower-affinity motif. We found minimal differences among the GO term enrichments for peaks with primary vs. secondary site motifs. In the sea urchin, primary site-containing peaks were near genes enriched in the MAPK pathway and thyroid hormone receptor activity, whereas secondary site peaks were near genes enriched for growth factor activity. Sea star primary site-containing peaks were near genes enriched for developmental processes and tRNA ligase activity, whereas targets near secondary site peaks were enriched for cell death (Fig. 2 and Dataset S1). There was no association between motif use and up- or down-regulation among the sea star target genes.

We were next interested to understand which, if any, orthologous genes and processes were regulated by Tbr in both sea urchin and sea star. Although multiple studies have compared gene function by assessing specific differential target gene expression in knockdowns vs. controls (e.g., refs. 33–36), this study provides the

opportunity to perform an unbiased genome-wide survey, as well as to compare potential direct regulation through comparison of our ChIP-seq datasets. There is no previous knowledge to provide an expectation of conservation and divergence between these echinoderms. Estimates of binding site conservation across other taxa show a wide range of target conservation. Experiments in yeast (37) and mammals (5, 38–40) reported low levels (e.g., 5–40%) of target conservation for several TFs examined, whereas studies of the mesodermal regulator *Twist* in *Drosophila* species revealed as high as 60–80% target conservation across the six species tested (41). Thus, the level of target conservation may be factor- and clade-dependent.

Using the reciprocal BBH mappings, we sought to identify orthologous genes that have proximal peaks in both sea urchin and sea star datasets. Given the differences in genome assembly completeness, and our preference toward eliminating false positives over maintaining data, we cannot make strong conclusions about missing associations (i.e., those genes that have a peak in one dataset but are missing a peak in the other) because there is a likelihood
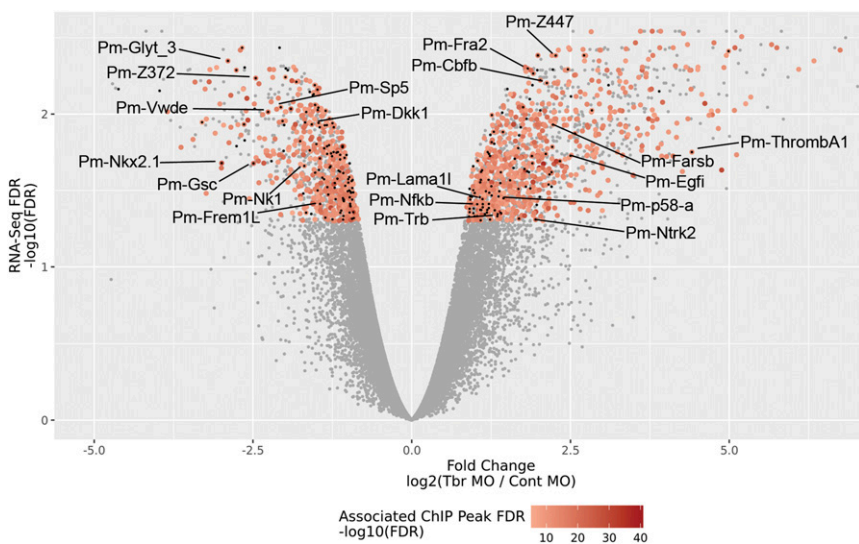

Fig. 4. Identification of targets of PmTbr by integration of ChIP-seq and RNA-seq data. Each point represents a gene detected in the *Pm-Tbr* MASO RNA-seq experiment, and the fold change for each gene is plotted against the calculated FDR. Significantly, DEGs are further annotated to indicate whether a corresponding ChIP peak was detected in the PmTbr ChIP-seq dataset (red points), where the intensity of the point color corresponds to the FDR of the associated peak; color scale is indicated on bottom key. Finally, any significant DEG that has a sea urchin ortholog with a peak detected within 75 kb is indicated as a black point. Points where both the sea star and sea urchin ortholog have an associated peak are colored red with a black point in the center. Several genes of interest from this category are labeled.

for false-negative associations in the sea star dataset, where the genome is assembled into generally smaller scaffolds (20) (Fig. S5). Therefore, we restrict the following analyses to associations between genes and peaks found in both datasets. Of the 4,444 sea urchin genes detected within 75 kb of a SpTbr peak, 995 had an orthologous sea star gene that also had a proximal PmTbr peak. A total of 108 of these genes (10.9%) were also differentially expressed in the *Pm-Tbr* knockdown RNA-seq dataset (Dataset S2). The GO terms enriched in the set of 108 genes with orthologous genes in both species that had peaks included "protein phosphorylation" (eight genes, $P = 0.007$), aminoacyl-tRNA ligase activity (three genes, $P = 0.02$), and "TF activity, sequence-specific DNA binding" (five genes, $P = 0.057$) (Fig. 2 and Dataset S1).

Finally, we asked whether the gene targets that are regulated in common (108 genes) (Fig. 4 and Dataset S2) had any notable binding site motif attributes. We found a similar decrease in the number of secondary motifs present in SpTbr peaks associated with the 108 genes in the overlapping ortholog set, compared with the total peak dataset (Fig. S6). The proportion of secondary sites in the sea urchin peaks in this subset was significantly lower (χ^2: $P = 0.014$), whereas there was no significant difference in the proportion of primary site peaks (χ^2: $P = 0.34$). A total of 17 of these genes were associated with peaks that had the same combination of primary and secondary motifs in both organisms (Dataset S2). Here, again, we found limited functional differences for genes associated with either primary vs. secondary site motif-containing peaks (Fig. 2 and Dataset S1). Therefore, primary and secondary motif use is not biased toward any of the functional categories tested here.

Discussion

Echinoderms are a powerful model system for understanding deep divergence in developmental GRNs and the consequences that these changes have for the evolution of morphology. Despite extensive comparative studies (33, 34, 36, 42), there has been no direct and unbiased whole-genome comparison of the role of orthologous factors in these taxa. Put simply, we are as yet completely ignorant of the scale of conservation and change that might occur among these classes. Indeed, although there are many studies among classes of chordates (5), such comparisons are essentially unknown in any other animal taxa at this level of evolutionary comparison. Echinoderms diverged into separate classes ~450–500 Mya in the Ordovician, but the radiation is considered to have occurred extremely rapidly, with five separate classes emerging in as few as 5 million years (14, 43, 44). During this rapid radiation, the classes underwent dramatic morphological change, and yet their body plans, both as adults and larvae, have remained remarkably stable since. This stability contrasts with vertebrate classes, which have undergone multiple waves of morphological radiations. Recently, this deep evolutionary separation of echinoderm classes has also been shown to present an opportunity to understand how TF proteins might evolve changed biochemical functions (13). Specifically, our previous study showed that the TF Tbr has evolved changed preferences for a low-affinity motif among echinoderm classes, and these secondary motifs were more sensitive to changing levels of Tbr. However, the role that such differences in motif utilization might have for the structure and function of the GRN is unknown in any taxa.

ChIP-seq provides the only method to assess direct binding and to reveal CRMs genome-wide, and it is therefore the only current technology that can provide a holistic comparison of direct regulatory connections and motif utilization. This approach has only recently become feasible in echinoderms because of the improvements made to the assembled genomes for these species (21). The high number of scaffolds in these genomes, however, still limits our ability to collect all associations between factor binding and target genes. In using these genomic techniques, a key goal has been to minimize false positives by stringently filtering the data in

various ways. These aspects of our analysis restrict our ability to make conclusions about missing associations because they may be missed due to the limitations of the genome or by the application of stringent filters.

Accepting the nature of these limitations, our study has shown that Tbr in these species is highly pleiotropic, and especially so in sea stars, where PmTbr has 9,164 predicted binding sites compared with only 1,952 in sea urchins. The expression of *tbr* has been well characterized in many species from different groups of echinoderms. Whereas in euechinoid sea urchins (e.g., *S. purpuratus*), *tbr* expression is restricted to skeletogenic mesoderm, orthologs in other echinoderm groups are expressed more broadly throughout the endomesoderm [e.g., as has been shown in sea stars (13), brittle stars (45), cidaroid sea urchins (46), and sea cucumbers (34)]. Parsimony, therefore, suggests that the euechinoid sea urchins have relatively recently lost this broader domain of *tbr* expression. The reduced number of Tbr binding sites identified in the sea urchin supports this hypothesis. This finding also suggests that our dataset should identify processes associated with a dramatic loss of target genes in sea urchin. Indeed, our data highlight functions found to be significantly enriched among sea star target genes that are lost in the sea urchin set, specifically those related to scavenger receptor activity and apoptotic processes suggesting a specific loss of these functions. Intriguingly, sea star Tbr appears to act as a repressor of these genes, implying that sea urchin Tbr has specifically lost repressor activities relative to the sea star. Both the sea urchin and sea star factors have predicted roles in regulating genes at all levels of the GRN hierarchy (i.e., other regulatory genes as well as genes involved in differentiation, cellular processes, and morphogenesis, which are found at the termini of the GRN). Therefore, alterations in function have not occurred solely through loss of terminal processes, but also through changes to regulation of other TFs and, hence, we predict, to GRN topology.

Even though the sea urchin and sea star proteins have very different domains of expression, and show a marked difference in numbers and types of regulated genes, they nonetheless share ~10% of their targets. Because we have only two taxa for comparison, we cannot know to what extent these are evolutionarily maintained and thus genuinely homologous, or are converged upon independently. However, this study reveals the overlap in targets that may occur in these taxa and is within the low range of similar comparisons of yeast and vertebrate lineages (5, 37–40). GO analyses and close inspection of the target genes in common indicate that protein kinases and other regulatory genes are commonly regulated, again pointing to a role for Tbr within the body of the GRN topology. This finding implies that GRN structure can change dramatically, while maintaining some regulatory connections. A future direction will be to more carefully dissect the GRN circuitry surrounding these conserved nodes to understand the types of motifs and network functions that can be maintained in the face of such dramatic rewiring.

We found peaks containing secondary binding motifs in sea urchin as well as sea star, which shows that, although it has very low affinity for these motifs in vitro, in vivo SpTbr is able to bind to these sites. This finding suggests that some combination of CRM characteristics must act to stabilize binding to these low-affinity sites in vivo. For example, CRMs may contain multiple Tbr binding sites or potential cofactor binding sites that could mitigate the effect of a single low-affinity site. Importantly, we show that secondary motifs are significantly underrepresented in sea urchins compared with sea stars, which implies that there are indeed fewer permissive contexts, or an increase in the requirements for functional utilization, for these motifs in sea urchin CRMs. This finding has important implications for how changes to TF binding affinity can restructure GRNs during evolution. Genes regulated by CRMs containing only a single low-affinity binding site (i.e., a secondary motif) will be prone to loss of regulation

as the TF evolves a further reduced affinity, such as we found for sea urchin Tbr. Many CRMs, however, are likely to be regulated by combinations of binding sites, including multiple low-affinity sites, as well as high-affinity motifs and comotifs, and are therefore expected to be relatively insulated from changes to TF binding affinity. Based on the described functions of low-affinity motifs (47), the CRMs most sensitive to a reduction in TF affinity may be those that regulate genes expressed at low levels, those that provide highly specific expression (for example in the presence of multiple related TFs), those that mediate repressive vs. activator functions, or possibly those that affect precise developmental expression timing.

It has been shown for many species, including in echinoderms, that there is a high rate of turnover in TF binding sites (48). Thus, there will be a gradual turnover that results in the loss of low-affinity motifs, a switching from high to low affinity, and loss of comotifs. Changes in affinity of these motifs can result from just one SNP, and hence can be quite frequent. Such turnover can lead to scenarios in which a gene is now regulated by a single, or few, low-affinity motifs. In the sea star, where the TF can readily bind these secondary motifs, the gene will maintain TF regulation. The sea urchin, however, may no longer express these genes in response to TF input. At this point, the CRM is no longer under selection for this TF binding and could rapidly acquire further mutations that are deleterious in this context. For example, loss of Tbr regulation might lead to a change in the timing of the regulation of the target, thereby relaxing other features of the CRM that direct function at this stage. Conversely the sea star CRM, being under maintained functional selection, may in the future acquire additional motifs and/or switch back from a low- to a high-affinity site. Thus, the relatively higher-affinity preference for a secondary motif in sea stars should provide an increased robustness to natural turnover in CRM sequences. The sea star has a greater range of functional binding site variants, and thus a larger sequence space for maintained selection. This hypothesis leads to the following predictions: that genes regulated by one or few low-affinity motifs will be more sensitive to TF levels, and that the sea star should have a greater frequency of CRMs under the control of single or few numbers of secondary motifs, whereas the frequency of these should be suppressed in sea urchins. Additionally, we predict greater population-level variation among Tbr-regulated CRMs in sea stars than in sea urchins. Ultimately, it will be essential to consider the consequences of binding site turnover and CRM loss in the context of the broader GRN and recognize that not all losses will manifest equivalent developmental phenotypes.

Materials and Methods

ChIP-Seq Analysis of SpTbr and PmTbr. ChIP was performed as described (13) (details in *SI Materials and Methods*). Custom rabbit polyclonal antibodies were produced by Thermo-Fisher against peptides from the nonconserved N terminus of each protein (PmTbr: EQGERYTVSHHGATEDTR; SpTbr: KFQKTTEPEESDKVYEDENLDRD) to ensure high specificity for Tbr over other t-box TFs present in the genome (e.g., Tbx2/3 and Brachyury). Embryos were cultured until mesenchyme blastula stage (*S. purpuratus*; ~24 hpf) or hatched blastula stage (*P. miniata*; ~30 hpf), at which point they were collected and processed as described (ref. 13 and *SI Materials and Methods*). One biological replicate each, prepared by pooling chromatin from two or three independently fertilized cultures before immunoprecipitation, was used to prepare sequencing libraries from total (input) and immunoprecipitated chromatin, and Illumina HiSeq 2500 76-bp SR sequencing runs were performed (Yale Center for Genome Analysis). Analysis of ChIP-seq data are described (*SI Materials and Methods*). Raw and processed sequencing reads have been deposited into the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus database (accession no. GSE89862), and analysis scripts are available upon request.

RNA-Seq Analysis of PmTbr Knockdown. *P. miniata* zygotes were injected with 600 μM Tbr MASO or standard control MASO (Gene Tools) as described (1, 2, 49, 50). RNA was extracted by using the GenElute Mammalian Total RNA Kit (Sigma-Aldrich). Illumina TruSeq library preparation and HiSeq 2500 50-bp SR sequencing were performed (University of Southern California Epigenome Center). RNA from biological triplicate paired sibling control- and Tbr-MASO injected embryos was analyzed. Analysis of RNA-seq data is described in *SI Materials and Methods*. Raw and processed sequencing reads have been deposited into the NCBI Gene Expression Omnibus database (accession no. GSE89863), and analysis scripts are available upon request.

Ortholog Mapping, GO Set Enrichment, and Motif Detection. Orthologous genes between the sea star and sea urchin transcriptome were detected by using a reciprocal BBH script generated in-house (*SI Materials and Methods*). GO term annotations for sea urchin genes were extracted from annotation files available through Echinobase. Term enrichment testing was performed by using the GOstats Bioconductor package, with statistical assessment determined by hypergeometric test, and reported *P* values were adjusted to compensate for multiple testing errors using the Benjamini–Hochberg correction (*SI Materials and Methods*). Motif occurrences within peaks was detected by using the FIMO program (*SI Materials and Methods*).

1. McGinnis N, Kuziora MA, McGinnis W (1990) Human Hox-4.2 and *Drosophila* deformed encode similar regulatory specificities in *Drosophila* embryos and larvae. *Cell* 63(5):969–976.
2. Grens A, Mason E, Marsh JL, Bode HR (1995) Evolutionary conservation of a cell fate specification gene: The Hydra achaete-scute homolog has proneural activity in Drosophila. *Development* 121(12):4027–4035.
3. Wray GA (2007) The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8(3):206–216.
4. Meader S, Ponting CP, Lunter G (2010) Massive turnover of functional sequence in human and other mammalian genomes. *Genome Res* 20(10):1335–1343.
5. Schmidt D, et al. (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328(5981):1036–1040.
6. Lynch VJ, Wagner GP (2008) Resurrecting the role of transcription factor change in developmental evolution. *Evolution* 62(9):2131–2154.
7. Cheatle Jarvela AM, Hinman VF (2015) Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. *Evodevo* 6(1):3.
8. Berger MF, Bulyk ML (2009) Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc* 4(3):393–411.
9. Berger MF, et al. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol* 24(11):1429–1435.
10. Ramos AI, Barolo S (2013) Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philos Trans R Soc Lond B Biol Sci* 368(1632):20130018.
11. Rowan S, et al. (2010) Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes Dev* 24(10):980–985.
12. Crocker J, et al. (2015) Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* 160(1-2):191–203.
13. Cheatle Jarvela AM, et al. (2014) Modular evolution of DNA-binding preference of a Tbrain transcription factor provides a mechanism for modifying gene regulatory networks. *Mol Biol Evol* 31(10):2672–2688.
14. Erwin DH, et al. (2011) The Cambrian conundrum: Early divergence and later ecological success in the early history of animals. *Science* 334(6059):1091–1097.
15. Oliveri P, Tu Q, Davidson EH (2008) Global regulatory logic for specification of an embryonic cell lineage. *Proc Natl Acad Sci USA* 105(16):5955–5962.
16. Rafiq K, Cheers MS, Ettensohn CA (2012) The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Development* 139(3):579–590.
17. Hinman VF, Nguyen AT, Cameron RA, Davidson EH (2003) Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc Natl Acad Sci USA* 100(23):13356–13361.
18. Croce J, Lhomond G, Lozano JC, Gache C (2001) ske-T, a T-box gene expressed in the skeletogenic mesenchyme lineage of the sea urchin embryo. *Mech Dev* 107(1-2):159–162.
19. Li X-Y, et al. (2011) The role of chromatin accessibility in directing the widespread, overlapping patterns of *Drosophila* transcription factor binding. *Genome Biol* 12(4):R34.
20. Revilla-i-Domingo R, Minokawa T, Davidson EH (2004) R11: A cis-regulatory node of the sea urchin embryo gene network that controls early expression of SpDelta in micromeres. *Dev Biol* 274(2):438–451.
21. Nam J, Dong P, Tarpine R, Istrail S, Davidson EH (2010) Functional cis-regulatory genomics for systems biology. *Proc Natl Acad Sci USA* 107(8):3930–3935.
22. Cameron RA, Kudtarkar P, Gordon SM, Worley KC, Gibbs RA (2015) Do echinoderm genomes measure up? *Mar Genomics* 22:1–9.

Cary et al.

23. Cameron RA, Samanta M, Yuan A, He D, Davidson E (2009) SpBase: The sea urchin genome database and web site. *Nucleic Acids Res* 37(Database issue):D750–D754.

24. Sun Z, Ettensohn CA (2014) Signal-dependent regulation of the sea urchin skeletogenic gene regulatory network. *Gene Expr Patterns* 16(2):93–103.

25. Röttinger E, Besnardeau L, Lepage T (2004) A Raf/MEK/ERK signaling pathway is required for development of the sea urchin embryo micromere lineage through phosphorylation of the transcription factor Ets. *Development* 131(5):1075–1087.

26. Saunders LR, McClay DR (2014) Sub-circuits of a gene regulatory network control a developmental epithelial-mesenchymal transition. *Development* 141(7):1503–1513.

27. Hinman VF, Nguyen AT, Davidson EH (2003) Expression and function of a starfish Otx ortholog, AmOtx: A conserved role for Otx proteins in endoderm development that predates divergence of the eleutherozoa. *Mech Dev* 120(10):1165–1176.

28. Hinman VF, Nguyen A, Davidson EH (2007) Caught in the evolutionary act: Precise cis-regulatory basis of difference in the organization of gene networks of sea stars and sea urchins. *Dev Biol* 312(2):584–595.

29. Tong A-J, et al. (2016) A stringent systems approach uncovers gene-specific mechanisms regulating inflammation. *Cell* 165(1):165–179.

30. Cusanovich DA, Pavlovic B, Pritchard JK, Gilad Y (2014) The functional consequences of variation in transcription factor binding. *PLoS Genet* 10(3):e1004226.

31. Sakabe NJ, et al. (2012) Dual transcriptional activator and repressor roles of TBX20 regulate adult cardiac structure and function. *Hum Mol Genet* 21(10):2194–2204.

32. Ouimette J-F, Jolin ML, L'honoré A, Gifuni A, Drouin J (2010) Divergent transcriptional activities determine limb identity. *Nat Commun* 1:35.

33. McCauley BS, Akyar E, Saad HR, Hinman VF (2015) Dose-dependent nuclear β-catenin response segregates endomesoderm along the sea star primary axis. *Development* 142(1):207–217.

34. McCauley BS, Weideman EP, Hinman VF (2010) A conserved gene regulatory network subcircuit drives different developmental fates in the vegetal pole of highly divergent echinoderm embryos. *Dev Biol* 340(2):200–208.

35. Koga H, et al. (2016) Experimental approach reveals the role of alx1 in the evolution of the echinoderm larval skeleton. *PLoS One* 11(2):e0149067.

36. Yankura KA, Koechlein CS, Cryan AF, Cheatle A, Hinman VF (2013) Gene regulatory network for neurogenesis in a sea star embryo connects broad neural specification and localized patterning. *Proc Natl Acad Sci USA* 110(21):8591–8596.

37. Borneman AR, et al. (2007) Divergence of transcription factor binding sites across related yeast species. *Science* 317(5839):815–819.

38. Conboy CM, et al. (2007) Cell cycle genes are the evolutionarily conserved targets of the E2F4 transcription factor. *PLoS One* 2(10):e1061.

39. Kunarso G, et al. (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* 42(7):631–634.

40. Odom DT, et al. (2007) Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat Genet* 39(6):730–732.

41. He Q, et al. (2011) High conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species. *Nat Genet* 43(5):414–420.

42. Cheatle Jarvela AM, Yankura KA, Hinman VF (2016) A gene regulatory network for apical organ neurogenesis and its spatial control in sea star embryos. *Development* 143(22):4214–4223.

43. Smith AB, et al. (2006) Testing the molecular clock: Molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). *Mol Biol Evol* 23(10):1832–1851.

44. Smith AB, Zamora S, Álvaro JJ (2013) The oldest echinoderm faunas from Gondwana show that echinoderm body plan diversification was rapid. *Nat Commun* 4:1385.

45. Dylus DV, et al. (2016) Large-scale gene expression study in the ophiuroid *Amphiura filiformis* provides insights into evolution of gene regulatory networks. *Evodevo* 7:2.

46. Yamazaki A, Kidachi Y, Yamaguchi M, Minokawa T (2014) Larval mesenchyme cell specification in the primitive echinoid occurs independently of the double-negative gate. *Development* 141(13):2669–2679.

47. Crocker J, Noon EP-B, Stern DL (2016) The soft touch: Low-affinity transcription factor binding sites in development and evolution. *Curr Top Dev Biol* 117:455–469.

48. Garfield D, Haygood R, Nielsen WJ, Wray GA (2012) Population genetics of cis-regulatory sequences that operate during embryonic development in the sea urchin *Strongylocentrotus purpuratus. Evol Dev* 14(2):152–167.

49. Cheatle Jarvela AM, Hinman V (2014) A method for microinjection of Patiria miniata zygotes. *J Vis Exp* (91):e51913.

50. Hinman VF, Davidson EH (2007) Evolutionary plasticity of developmental gene regulatory network architecture. *Proc Natl Acad Sci USA* 104(49):19404–19409.

51. Mortazavi A, Leeper Thompson EC, Garcia ST, Myers RM, Wold B (2006) Comparative genomics modeling of the NRSF/REST repressor network: From single conserved sites to genome-wide repertoire. *Genome Res* 16(10):1208–1221.

52. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.

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

54. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.

55. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

56. Kim D, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.

57. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.

58. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140.

59. Lund SP, Nettleton D, McCarthy DJ, Smyth GK (2012) Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat Appl Genet Mol Biol* 11(5):8.

60. Camacho C, et al. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421.

61. Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* 23(2):257–258.

62. Benjamini Y (1995) Hochberg Y controlling the false discovery rate: A practical and powerful approach to multiple testing. *JR Stat Soc* 57(1):289–300.

63. Grant CE, Bailey TL, Noble WS (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017–1018.

64. Bailey TL, et al. (2009) MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res* 37(Web Server issue):W202–8.

DEVELOPMENTAL BIOLOGY