CrossMark

Ancestral state reconstruction by comparative analysis of a GRN kernel operating in echinoderms

Eric M. Erkenbrack[1] · Kayla Ako-Asare[2] · Emily Miller[2] · Saira Tekelenburg[2] ·
Jeffrey R. Thompson[3] · Laura Romano[2]

Abstract Diverse sampling of organisms across the five major classes in the phylum Echinodermata is beginning to reveal much about the structure and function of gene regulatory networks (GRNs) in development and evolution. Sea urchins are the most studied clade within this phylum, and recent work suggests there has been dramatic rewiring at the top of the skeletogenic GRN along the lineage leading to extant members of the euechinoid sea urchins. Such rewiring likely accounts for some of the observed developmental differences between the two major subclasses of sea urchins—cidaroids and euechinoids. To address effects of topmost rewiring on downstream GRN events, we cloned four downstream regulatory genes within the skeletogenic GRN and surveyed their spatiotemporal expression patterns in the cidaroid *Eucidaris tribuloides*. We performed phylogenetic analyses with homologs from other non-vertebrate deuterostomes and characterized their spatiotemporal expression by quantitative polymerase chain reaction (qPCR) and whole-mount in situ hybridization (WMISH). Our data suggest the *erg–hex–tgif* subcircuit, a putative GRN kernel, exhibits a mesoderm-specific expression pattern early in *Eucidaris* development that is directly downstream of the initial mesodermal GRN

circuitry. Comparative analysis of the expression of this subcircuit in four echinoderm taxa allowed robust ancestral state reconstruction, supporting hypotheses that its ancestral function was to stabilize the mesodermal regulatory state and that it has been co-opted and deployed as a unit in mesodermal subdomains in distantly diverged echinoderms. Importantly, our study supports the notion that GRN kernels exhibit structural and functional modularity, locking down and stabilizing clade-specific, embryonic regulatory states.

Introduction

Echinoids, or sea urchins, are constituents of the phylum Echinodermata and are comprised of two extant subclasses, the cidaroids (Cidaroidea) and the euechinoids (Euechinoidea). Fossil evidence suggests these two clades had already diverged by the middle of the Permian period at least 268 million years ago (mya) (Thompson et al. 2015). In addition to the conspicuous differences observed in the adult morphologies of these clades (Gao et al. 2015), embryological evidence also indicates numerous developmental differences between these two clades, suggesting extensive rewiring of the gene regulatory networks (GRNs) directing their development (Schroeder 1981). Two decades of research have parsed out the elaborate circuitry of the GRNs guiding early euechinoid development (Peter and Davidson 2011). Thus, with an abundance of euechinoid data in hand, an auspicious opportunity presents itself not only to enumerate the observed developmental differences between these two clades at the

✉ Eric M. Erkenbrack
erkenbra@caltech.edu

[1] Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

[2] Department of Biology, Denison University, Granville, OH 43023, USA

[3] Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089, USA

molecular level but also to utilize them to better understand the plasticity and role of GRNs in developmental evolution.

The most striking difference between the early development of cidaroids and euechinoids is exhibited in the skeletogenic mesenchyme. In contrast to euechinoids, which exhibit a precociously ingressing, pre-invagination skeletogenic primary mesenchyme (PMC) lineage, the skeletogenic mesenchyme of cidaroids ingresses from the tip of the archenteron well after gastrulation has begun (Wray and McClay 1988). In euechinoids, the micromere quartet gives rise to the PMC lineage and is fated from very early in development to become embryonic skeleton (Oliveri et al. 2008). Cidaroids, in juxtaposition, exhibit a variable number of micromeres in each embryo (Schroeder 1981), and yet in spite of this variance their micromeres were found to be homologous to euechinoid PMCs (Wray and McClay 1988).

The skeletogenic GRN of the purple sea urchin *Strongylocentrotus purpuratus* (*Sp*), a euechinoid, proffers a highly detailed account of the specification and differentiation of the PMC lineage (Oliveri et al. 2008). Multiple lines of experimental evidence support the hypothesis that this developmental GRN is highly conserved amongst euechinoids (Ettensohn 2013). Recently, similar investigations have been extended to cidaroids (Erkenbrack and Davidson 2015; Yamazaki et al. 2014). For example, a recent study investigated expression and function of genes at the top of the euechinoid PMC GRN in the Atlantic Basin-dwelling cidaroid *Eucidaris tribuloides* (*Et*) (Erkenbrack and Davidson 2015). Functional interrogation of these genes in *Et* revealed nine specific regulatory inputs that are absent and that are likely to be gain-of-function changes that occurred during the evolution of euechinoids. More specifically, the localization mechanism of skeletogenic factors to the euechinoid PMCs—the double-negative gate—was missing in this cidaroid, suggesting large-scale rewiring of this circuitry occurred in the lineage leading to extant euechinoids. Furthermore, the transcription factors *ets1/2* and *tbrain* were not spatially restricted to the *Et* skeletogenic cells. This was also the case in the Indo-Pacific-dwelling cidaroid *Prionocidaris baculosa* (*Pb*), suggesting a broader role in mesodermal specification for these genes in the cidaroid clade (Yamazaki et al. 2014). In euechinoids, *ets1/2* and *tbrain* are restricted to the PMCs and function as important early inputs into an *erg–tgif–hex–alx1* skeletogenic GRN circuit (Oliveri et al. 2008). An additional function of the micromeres in *Et* is to exclude the skeletogenic fate in the surrounding non-skeletogenic mesoderm (Erkenbrack and Davidson 2015). Many of these findings in *Et* are consistent with those in *Pb*, suggesting that many of these observations are conserved amongst cidaroids (Yamazaki et al. 2014). These results beg the

question of how manifold the changes to the downstream circuitry are between the cidaroid skeletogenic GRN and the euechinoid PMC GRN.

In examined euechinoids, *erg*, *hex*, and *tgif* form a recursively wired subcircuit downstream of *ets1/2* and *tbrain* that plays a role in stabilizing the regulatory state of the skeletogenic lineage (Oliveri et al. 2008). There is strong evidence that in the echinoderm clade this subcircuit serves as a highly conserved, recursively wired GRN kernel that can be deployed at alternative embryonic addresses to lock down and stabilize regulatory states (McCauley et al. 2010; Davidson and Erwin 2006). Importantly, it is clear that there has been significant GRN rewiring in the lineage leading to euechinoids, in which much of this circuitry has been restricted to the skeletogenic mesoderm as opposed to the broader endomesodermal roles seen in the sea star *Patiria miniata* (*Pm*) and mesodermal roles in the sea cucumber *Parastichopus parvimensis* (*Pp*) (McCauley et al. 2010; McCauley et al. 2012). Comparative analyses in sea stars and euechinoid sea urchins indicate that the initial onset of this circuit differs between these clades. In euechinoids, *ets1* and *tbrain* set off the *erg–hex–tgif* regulatory cascade (Oliveri et al. 2008). However, in asteroids, *tbrain* is the major driver of this subcircuit, suggesting that *ets1* is a derived input unique to euechinoids (McCauley et al. 2010). We hypothesized that in *Et* this conserved circuit would play roles mainly in mesoderm specification, as the drivers of this kernel—*ets1/2* and *tbrain*—are expressed throughout the mesoderm in cidaroids, as opposed to their more restricted roles in the PMCs in euechinoids. Thus, we predicted that in cidaroid sea urchins the spatial expression pattern of this subcircuit would mirror more closely that of asteroids and holothuroids than that of euechinoids. To address these queries, we cloned the genes from the conserved *erg–hex–tgif* subcircuit and investigated their spatiotemporal expression in *Et*. We also present here expression dynamics of the skeletogenic GRN gene *tel*, as similar data in other echinoderms is lacking.

Materials and methods

Cloning and phylogenetic analyses

RNA was extracted from embryos following in vitro fertilization of eggs obtained from adult *E. tribuloides* collected by Gulf Specimens Marine Lab (Panacea, FL) or Reeftopia (Sugarloaf Key, FL). Complementary DNA (cDNA) was prepared using the SMART RACE cDNA Amplification Kit. Full-length sequences of *erg*, *hex*, *tgif*, and *tel* were obtained through a combination of cloning and existing transcriptome data available in Echinobase (echinobase.org). DNA binding domains were predicted using ExPASy: Bioinformatics Resource Portal (expasy.org). Multiple sequence alignments

were performed using ClustalX 2.1. Phylogenetic reconstruction was carried out using maximum likelihood methods with bootstrap confidence intervals determined by using 1000 replicates. The output was viewed using FigTree 1.4.0.

The following sequences were used to construct the phylogenetic trees: SpErg (SPU_018483), LvErg (retrieved by BLAST in Echinobase), PmErg (GU_251975), CiErg (NM_001078474), SkErg (XM_006822711), BfErg (XM_002613065), Xl (AJ_224125), and GgErg (X_77159); SpHex (SPU_027215), LvHex (retrieved by BLAST in Echinobase), PmHex (GU_251972), CiHex (NM_001078262), SkHex (SQ_431047), BfHex (EU_296398), XlHex (NM_001085590), and GgHex (NM_205252); SpTel (SPU_028479), LvTel (retrieved by BLAST in Echinobase), BfTel (XP_002608583), XlTel (NM_001124423), and GgTel (NM_001199273); and SpTgif (SPU_018126), LvTgif (retrieved by BLAST in Echinobase), PmTgif (GU_251973), CiTgif (XP_002124000), SkTgif (NM_001164980), BfTgif (NP_001071803), XlTgif (NP_001080420), and GgTgif (NM_205379).

Whole-mount in situ hybridization

Probes were synthesized from cDNA using the DIG RNA Labeling Kit (SP6/T7) (Roche). Primers used for whole-mount in situ hybridization (WMISH) probe amplification were designed from full-length coding sequences (Table 1). Embryos were fixed in PFA-MAB (4 % PFA, 32.5 % MFSW, 32.5 mM maleic acid (pH 7), 162.5 mM NaCl) solution on ice and left overnight at 4 °C. Embryos were transferred into hybridization buffer (50 % formamide, 5× Denhardt's, 5× SSC, 1 mg/ml yeast tRNA, 100 mM NaCl, 0.1 % Tween-20, 50 μg/ml heparin) by series (10 %, 25 %, 50 %, 75 %, 100 %). Probe concentrations were 0.5–1.0 ng/μl. Probe hybridization and post-hybridization stringency washes were carried out at 63 °C. Antibody concentration was 0.25 μg/ml. Detailed procedures are described in Erkenbrack and Davidson (2015).

Real-time quantitative PCR

Eggs from two different females were used as the starting point for two timecourse cultures. From these, 100 embryos were counted at hourly intervals, gently centrifuged, and lysed with Buffer RLT. To allow quantification of messenger RNA (mRNA) transcripts, each timepoint was spiked with ~1000 transcripts of synthetic Xeno RNA (Cells-to-Ct Control Kit, Life Technologies) and harvested according to the manufacturer's protocol (RNeasy, Qiagen). Approximately one embryo per reaction was assayed in triplicate by quantitative polymerase chain reaction (qPCR) (SYBR Green, Life Technologies). qPCR primers used in this study are listed in Table 2.

Results and discussion

Cloning and phylogenetic analysis of skeletogenic genes

Homologs of *erg*, *hex*, *tgif*, and *tel* were isolated from *Et*. Erg and Tel belong to the ETS (E-twenty six) family of related transcription factors, which is characterized by a DNA binding domain known as the ETS domain with a "winged" helix–turn–helix motif (Seidel and Graves 2002). The pointed (PNT) domain occurs rarely in subfamilies and regulates phosphorylation by serving as a docking site for kinases such as Erk2 (Seidel and Graves 2002). In *Sp*, there are 11 members of the ETS family with at least one ortholog for each of the subfamilies that have been identified in vertebrates (with the exception of one mammal-specific subfamily) (Rizzo et al. 2006). Genes in this family interact with a diverse array of co-regulatory proteins and serve as both transcriptional activators and repressors (Mavrothalassitis and Ghysdael 2000). In *Et*, the full-length coding sequence of *erg* is 1785 base pairs (bp) and encodes a 594-amino acid (aa) protein that includes a PNT domain (161–246) and an ETS domain (425–505) . The full-length coding sequence of *tel* is 2121 bp and encodes a 706-aa protein that includes a PNT domain (42–126) and an ETS domain (473–554). The predicted aa sequences for Erg and Tel show 69.67 and 55.06 % similarity to their homologs in *Sp*, respectively.

Hex and TG-interacting factor (Tgif) belong to the homeobox (HOX) family of related transcription factors, characterized by the presence of a DNA binding homeodomain with a helix–turn–helix motif. There are 96 members belonging to various subfamilies in *Sp* (Howard-Ashby et al. 2006). A member of the NK subfamily, Hex (PRH) is known to act

Table 1 Sequences of WMISH primer sets used in this study

Gene	WMISH forward primer	WMISH reverse primer
erg	TCTCGGATGACCAGTCTATGT	CACGGCTCAGTTTATCGTAGTT
hex	GTCCCTGCCATCTTCGTTGTCTCCCTTGCC	CTCCAACGATCAGACGATGGAACTCACTCG
tel	GTGCTTGTTCTCCCATCGGATGTAGGGCCG	GGAGGAGTTCTCGCTGGACAGCGTGAATGC
tgif	GGCGGGCATCGACAAGAATGTGGAATGG	CATCACCACGGCGCGAGTAGTTCTGCTC

Table 2 Primer sequences used in this study for quantitative PCR

Gene	qPCR forward primer	qPCR reverse primer
erg	TTCGACGCCCGAGGAAC	CCACTGGACCCACTGTTGA
hex	CTCTACCCGTACTCTAGGAATGA	ATCGTTGGAGAACCTGACTTG
tel	AAATTCAGCATGAACGGGAAGGCG	TCGGTGTTCTCTGATTCCTGCTCT
tgif	GCGAGTAGTTCTGCTCCAAA	ATGGCGAATCTCACTCTCTTG

both as an activator and a repressor of transcription (Soufi and Jayaraman 2008), while Tgif is a member of the three-amino acid loop extension (TALE) subfamily, serving mostly to repress transcription, although it can also function as an activator. In *Et*, the full-length coding sequence of *hex* is 885 bp and encodes a 294-aa protein that includes a HOX (151–211) domain. The full-length coding sequence of *tgif* is 1083 bp and encodes a 360-aa protein that contains a HOX domain (44–107). The predicted aa sequences show 64.98 and 70.01 % similarity to their homologs in *Sp*, respectively.

To confirm the identity of each *Et* sequence, we carried out phylogenetic analyses in which the cephalochordate *Branchiostoma floridae* was always the outgroup. In each case, our analyses resulted in a phylogenetic tree that confirmed the orthology of our sequences and their placement as anciently diverged with respect to euechinoids within a large clade consisting of ambulacrians (i.e., echinoderms and the hemichordate *Saccoglossus kowalevskii*) (Fig. 1).

Characterization of spatiotemporal expression of *Et-erg*

Previous studies in *Sp* showed that *erg* is initially ubiquitous and subsequently restricted to the PMCs by 21 h post fertilization (hpf). Upon ingression of the PMCs, *Sp-erg* is activated in the oral non-skeletogenic mesenchyme (NSM), where it functions in blastocoelar cell fate (Materna et al. 2013; Solek et al. 2013). *Pm-erg* is initially expressed at blastula stage in

mesoderm precursors at the center of the vegetal pole (McCauley et al. 2010), and is localized during gastrulation to the bulb of the archenteron with some expression in ingressing mesenchymal cells. Similar expression is also observed in the holothurian *Pp* (McCauley et al. 2012). In *Et*, our data indicate that, as in euechinoids, *erg* is not maternally expressed. Both qPCR and WMISH data indicate that zygotic expression in *Et* begins specifically in a few cells at the center of the vegetal pole 4–5 hpf prior to gastrulation (Fig. 2(a, e, e′)). This localized expression is similar to that for *Sp-erg*, which begins to be expressed in the PMCs around 15 hpf (Rizzo et al. 2006). In *Et*, this restricted initial expression pattern is reminiscent of mesodermal genes at the top of the euechinoid PMC GRN that begin their expression in the micromere descendants, including that of *alx1*, *ets1/2*, and *tbrain*, which begin to be expressed at 4, 8, and 6 hpf prior to *erg*, respectively (Erkenbrack and Davidson 2015). *Et-erg* expression is seen in the entire mesoderm by 24 hpf; later, expression is observed in skeletogenic mesenchyme ingressing into the blastocoel (Fig. 2(f, f′)). Moreover, staining persists at 60 hpf not only at the tip of the archenteron and in all mesenchyme but also in three to five cells that have migrated to the ventral lateral clusters and which likely constitute the skeletogenic mesenchyme (Fig. 2(g′)). These data are consistent with the hypothesis that *Et-erg* potentially is a broad mesodermal regulator that is downstream of the first wave of mesodermal GRN circuitry (Erkenbrack and Davidson 2015). Though the initial input of *erg* is not known, it is highly
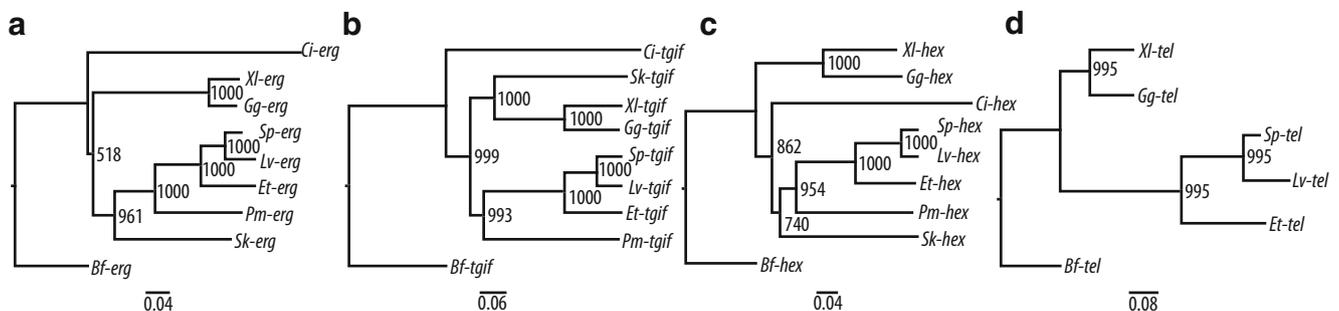

Fig. 1 Phylogenetic analyses of Erg, Tgif, Hex, and Tel amino acid sequences from selected deuterostomes to confirm homology with predicted amino acid sequences from the cidaroid *Eucidaris tribuloides*. **a** Erg, **b** Tgif, **c** Hex, **d** Tel. N–J trees were constructed in FigTree 1.4.0 following multiple sequence alignments in ClustalX 2.1. Bootstrap values are indicated at the *nodes*, and *scale bars* beneath the *trees* represent the average number of substitutions per site as a measure of evolutionary distance. Species abbreviations: *Et Eucidaris tribuloides*, *Sp Strongylocentrotus purpuratus*, *Lv Lytechinus variegatus*, *Pm Patiria miniata*, *Sk Saccoglossus kowalevskii*, *Ci Ciona intestinalis*, *Xl Xenopus laevis*, *Gg Gallus gallus*, *Bf Branchiostoma floridae* (outgroup)
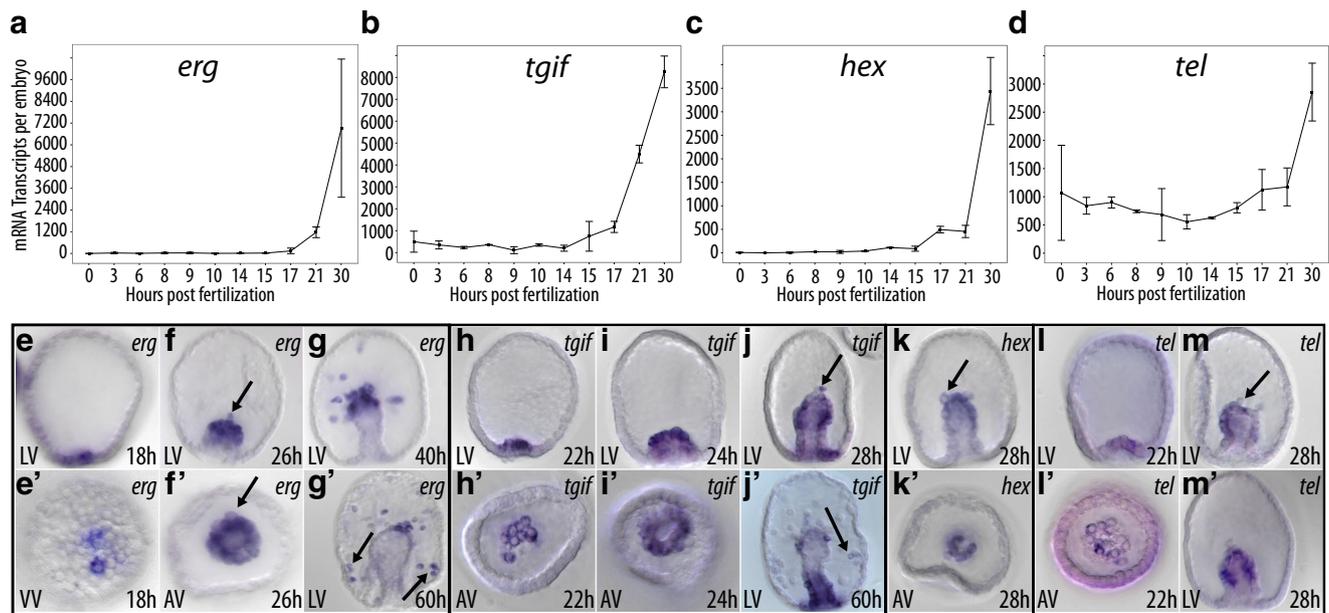
Fig. 2 Spatiotemporal expression of the *erg–hex–tgif* subcircuit and the euechinoid skeletogenic gene *tel* in the early development of the cidaroid *Eucidaris tribuloides*. **a–d** mRNA expression sampled approximately every 3 h over the course of the first 30 h post fertilization (hpf) of *E. tribuloides* development as revealed by quantitative PCR. **a** qPCR profile of *erg* showing it is first detected between 15 and 17 hpf. **b** qPCR profile of *tgif* showing it is maternal and is first detected between 8 and 14 hpf. **c** qPCR profile of *hex* showing it is first detected between 10 and 15 hpf. **d** qPCR profile of *tel* showing it is maternal and is first detected between 10 and 14 hpf. **e–m** Whole-mount in situ hybridization (WMISH) of *erg*, *tgif*, *hex*, and *tel* at late blastula, early gastrula, and mid-gastrula. Gene of interest, view of the embryo, and hours post fertilization (*h*) are indicated on each image. **e–g**, **e′–g′** WMISH of *erg* at four timepoints in development. **e**, **e′** At 18 hpf, *erg* is expressed in a few cells at the vegetal pole. **f**, **f′** At 26 hpf, *erg* is expressed in all invaginating cells (mesoderm) and is also expressed in skeletogenic mesenchyme (*arrow*). **g** At 40 hpf, *erg* is expressed throughout the mesoderm and ingressing mesenchyme. **g′** At 60 hpf,

erg expression is present at the tip of the archenteron and in migrating mesenchyme, including mesenchymal cells that have migrated to the ventral lateral cluster (skeletogenic cells, *arrows*). **h–j**, **h′–j′** Expression of *tgif* at four timepoints in development. **h**, **h′** At 22 hpf, *tgif* is expressed in a few cells at the vegetal pole. **i**, **i′** At 24 hpf, *tgif* is now seen throughout the mesoderm. **j** By 28 hpf, *tgif* is expressed throughout the mesoderm, including skeletogenic cells (*arrow*) and also in the endoderm. **j′** At 60 hpf, *tgif* is expressed at the tip of the archenteron, in the mid- and hindgut endoderm, and also in cells that have migrated to the ventral lateral clusters (*arrow*). *Tgif* is not, however, expressed in most migrating mesenchyme. **k**, **k′** At 28 hpf, *hex* is expressed throughout the mesodermal domain and is also present in ingressing skeletogenic mesenchyme (*arrow*). **l**, **l′**, **m**, **m′** *tel* expression at two different developmental timepoints. **l**, **l′** At 22 hpf, *tel* is expressed in isolated mesodermal cells. **m**, **m′** At 28 hpf, *tel* expression is seen in the majority of the mesodermal domain but is absent from ingressing skeletogenic cells (*arrow*). *h* hours post fertilization, *LV* lateral view, *VV* vegetal view, *AV* apical view

probable there is an intermediate regulator between the initial mesodermal regulators *ets1/2* and *tbrain*, since these genes are already expressed in many more cells than the number of cells in which *erg* begins to be expressed. Minimally, these data further substantiate the claim that *erg* played an ancestral role in mesoderm specification in eleutherozoans, and possibly all echinoderms, as it is expressed in this tissue in all taxa where spatial expression data is available and thus is a crucial cog in GRN circuitry of mesodermal specification (Table 3).

Characterization of spatiotemporal expression of *Et-tgif*

In *Sp*, *tgif* is a maternally deposited factor, the zygotic expression of which begins at 16 hpf, where it is initially expressed in both the PMCs and NSM (Howard-Ashby et al. 2006). After gastrulation begins, *Sp-tgif* is seen in NSM and midgut endoderm. In *Pm*, *tgif* expression is first observed broadly in the endomesoderm and by mid-

gastrula is observed in the endoderm, whereas in *Pp*, *tgif* is expressed first in mesodermal cells and later is seen in endoderm and non-ingressed mesoderm at the tip of the archenteron (McCauley et al. 2010; McCauley et al. 2012). Our qPCR data indicate that *Et-tgif* is maternally expressed (Fig. 2(b)). Zygotic expression begins by at least 14 hpf, indicating this is the first gene to be zygotically expressed in the *erg–hex–tgif* subcircuit. Up to gastrulation *Et-tgif* is restricted to a subset of cells in the mesoderm, after which point it comes to be expressed throughout the mesoderm by early gastrula and then subsequently in mesoderm and endoderm by mid-gastrula (Fig. 2(h, i, j)). By late gastrula, very faint expression can be seen in migrating mesenchyme cells, weak expression is seen in mesodermal cells at the tip of the archenteron, and strong expression occurs in the mid- and hindgut—while expression is conspicuously absent from the foregut (Fig. 2(j′)). These data indicate that, in all

Table 3 Spatial expression patterns of *erg*, *hex*, *tgif*, and *tel* in echinoderms. Data for the cidaroid lineage are from this study. All other data are from previously published work, for which the references are cited

Taxon/stage	Gene/spatial domain			
	Erg	*Tgif*	*Hex*	*Tel*
Asteroids				
Blastula	Mesodermal[a]	Endomesodermal[a]	Endomesodermal[a]	No data
Mid-gastrula	Mesodermal[a]	Endodermal[a]	Endomesodermal[a]	No data
Ophiuroids				
Blastula	No data	No data	No data	No data
Mid-gastrula	No data	No data	No data	No data
Holothuroids				
Blastula	Mesodermal[b]	Mesodermal[b]	No data	No data
Mid-gastrula	Mesodermal	Mesodermal (non-ingressed)	No data	No data
	Skeletogenic/NSM[b]	Endodermal[b]		
Cidaroids				
Blastula	Mesodermal[c]	Mesodermal[c]	No data	Mesodermal[c]
Mid-gastrula	Mesodermal Skeletogenic/NSM[c]	Mesodermal Endodermal[c]	Mesodermal[c]	Mesodermal[c]
Euechinoids				
Blastula	PMC[d]	PMC/NSM[f]	PMC[e]	PMC[d]
Mesenchyme blastula	PMC/NSM[g]	PMC/NSM[f]	PMC/NSM[g]	PMC[d]
Mid-gastrula	NSM, PMC, coelomic pouches[h]	NSM/endoderm[f]	PMC/NSM[e]	No data

[a] McCauley et al. (2010)

[b] McCauley et al. (2012)

[c] This work

[d] Rizzo et al. (2006)

[e] Poutska et al. (2007)

[f] Howard-Ashby et al. (2006)

[g] Materna et al. (2013)

[h] Solek et al. (2013)

eleutherozoans examined, *tgif* has an early, spatially restricted role in mesodermal specification and, later, a distinct role in endoderm specification. However, only in echinozoans is *tgif* restricted to a mesodermal lineage early in development. Given that the role of *tgif* in endodermal specification appears conserved in all eleutherozoans, these data suggest that following the echinozoan–asterozoan divergence (481 mya; Jell 2014), this endodermal activity was activated later in embryonic development in echinozoans and is a derived feature of this clade.

Characterization of spatiotemporal expression of *Et-hex*

Zygotic expression of *Sp-hex,* which is not maternal in *Sp,* begins prior to hatching in the early blastula and is observed in the PMCs by 20 hpf (Poutska et al. 2007). Similarly to *Sp-erg, Sp-hex* expression is also activated in the oral NSM upon ingression of the PMCs (Materna et al.

2013). By late gastrula stage, expression is observed in PMCs and SMCs but not endoderm. *Pm-hex* is initially expressed at blastula stage throughout the presumptive endomesoderm at the vegetal pole. During gastrulation, *Pm-hex* continues to be weakly expressed in both the endoderm and the mesoderm with the exception of some cells at the tip of the archenteron (McCauley et al. 2010). Our data suggest that, like in *Sp*, *hex* is not maternally expressed in *Et* (Fig. 2(c)). Zygotic expression of *Et-hex* begins concurrently with *Et-erg* after hatching in the late blastula by 16 hpf (Fig. 2(c)). By 28 hpf, spatial expression occurs broadly in mesodermal cells and also occurs in cells ingressing into the blastocoel, very likely overlapping in its expression with *Et-erg* (Fig. 2(k, k′)). The spatiotemporal expression patterns of *Et-hex* are indicative of the possibility that it shares inputs with *erg*. Unlike in sea stars, however, these data indicate that *Et-hex* is not expressed in the endoderm up to the time that

the skeletogenic mesenchyme ingresses at early gastrula. Indeed, this is also the case in euechinoids (Poutska et al. 2007). These data suggest that in early embryonic development of echinoids, *hex* is strictly mesodermal, whereas in asteroids it functions in both endoderm and mesoderm. Additionally, even though the spatial expression pattern of *hex* is not known in *Pp*, comparative analyses of spatiotemporal expression patterns from three taxa predict that the expression pattern of *hex* in this holothuroid will mirror that of *erg* and *tgif* in *Pp*.

Initiation and conserved wiring of *erg–hex–tgif* subcircuit in *Et*

Taken together, the spatial expressions of *Et-erg*, *Et-hex*, and *Et-tgif* are remarkable in their congruence, all broadly expressed in—and restricted to—the mesoderm at least until skeletogenic mesenchyme ingresses at 26 hpf. Furthermore, our data specifically show that *erg*, *hex*, and *tgif* are all expressed in the first mesenchyme to ingress in *Et* (Fig. 2(f, f′, j, k, k′)). These observations lend strong support to the supposition that micromere descendants in *Et* are homologous to the PMC lineage of euechinoids (Wray and McClay 1988). Additionally, these data suggest that many aspects of the downstream euechinoid PMC GRN circuitry were already in place at the divergence of the two extant echinoid clades. It is also important to note for *erg* and *tgif* that their initial activation in *Et* is restricted to a few cells at the pole of the vegetal plate. While we do not present the necessary experimental evidence to claim the *erg–hex–tgif* subcircuit initiates solely in *alx1*-positive cells, our data are very suggestive that its activation begins in micromere descendants and subsequently expands to the surrounding NSM rather than vice versa. However, this being the case, it can be said with certainty that the *erg–hex–tgif* subcircuit is running in *alx1*-positive cells at the time of skeletogenic mesenchyme ingression (28 hpf). Thus, by 22 hpf, *Et-tgif* and *Et-erg* have come on in cells that are very likely the micromere descendants. Given that *ets1/2* and *tbrain* are known to be upstream of *erg* and *tgif* in *Sp*, their activation by *ets1/2* and *tbrain* in *Et* must be addressed in future studies. However, the initial activation and spatial restriction of the *erg–hex–tgif* subcircuit to a few cells at the tip of the vegetal pole occur long after *ets1/2* and *tbrain* have been initiated in the mesoderm, suggesting there are other factors at play in their activation. With that said, as soon as *erg* and *tgif* begin to run in the micromere descendants, all of the inputs required to initiate the *erg–hex–tgif* subcircuit are in place: *ets1/2* and *tbrain* are running in the mesoderm; *erg* would feed into both *hex* and *tgif*; and *hex* would feed back into both *erg* and *tgif*. Lastly, it is worth noting that the repressive function of *erg* on *tbrain*, described in *Pm*

(McCauley et al. 2010), must be missing in cidaroids, as *erg* and *tbrain* are expressed in overlapping cells at least up to 40 hpf.

Characterization of spatiotemporal expression of *Et-tel*

While *tel* is not part of the *erg–hex–tgif* subcircuit, it was included in the study due to its role immediately downstream of the initial activators of the *Sp* PMC GRN. In *Sp*, *tel* is a maternal factor that begins to be zygotically expressed by 15 hpf in the PMCs, where it remains until at least mesenchyme blastula stage (Rizzo et al. 2006). In *Pm*, there is no evidence of a *tel* homolog. We found that *tel* is maternally expressed in *Et* and that zygotic expression begins much like *erg*, *hex*, and *tgif*, in a few cells at the center of the vegetal pole; thereafter, it expands to be broadly mesodermal (Fig. 2(d, l, m)). Interestingly, we found that ingressing skeletogenic mesenchyme cells of *Et* were not positive for *Et-tel* even though it is clearly visible in the surrounding NSM at 30 hpf (Fig. 2(m, m′)). In *Sp*, *tel* is an input into differentiation genes in the PMC GRN (Oliveri et al. 2008). In *Et*, our data indicate that, while *Et-tel* is expressed early on in skeletogenic cells, by the time the skeletogenic cells ingress into the blastocoel, *tel* expression is absent from those cells. These data suggest that the restricted, PMC-specific expression of *tel* is a euechinoid novelty and likely the result of another euechinoid co-option event.

Evolution of the *erg–hex–tgif* subcircuit in echinoderms

Comparative analysis of three or more taxa in a monophyletic clade is the gold standard for making claims regarding developmental character state evolution. This is because with only two taxa, it is impossible to establish polarity of characters and thus determine which character states are ancestral and which are derived. Our results buttress and broaden already published data on spatiotemporal expression patterns from three major taxa of echinoderms—asteroids, holothuroids, and echinoids—members of which last shared a common ancestor over 481 mya (Jell 2014). By comparing data from multiple taxa, we can minimally enumerate 12 statements about the embryos of the ancestors of these modern species (Table 4). These statements are incredibly significant, as they afford predictions regarding both timing and classification of evolutionary events that must have occurred in the various evolutionary lineages that led to extant taxa. Data from asteroids and holothuroids allow for establishment of character polarity with regard to the spatiotemporal deployment of the *erg–hex–tgif* subcircuit in echinoids. That the deployment of the *erg–hex–tgif* subcircuit appears to be limited to the mesoderm in modern cidaroids, as it is in holothuroids (McCauley et al. 2012), indicates

Table 4 Ancestral state reconstruction for embryos of ancestors of extant echinoderm clades by comparative analysis of spatial gene expression data from three or more taxa

At least 481 mya: in ancestral embryos prior to the asterozoan–echinozoan divergence

1. *erg* was a mesodermal driver at blastula stage and gastrula stage
2. *hex* was a mesodermal driver at blastula stage
3. *tgif* was a mesodermal driver at blastula stage and gastrula stage
4. *tgif* was an endodermal driver at mid-gastrula
5. *erg–hex–tgif* kernel operated in mesoderm
6. Prediction: *hex* is likely to be expressed in mesoderm of holothurians, but endodermal expression after blastula stage is unclear

At least 462 mya: in ancestral embryos prior to the holothuroid–echinoid divergence

7. *erg* and *tgif* were initiated in the mesoderm and *tgif* came to be expressed in the endoderm at a later time in development; whereas *erg* remained restricted to the mesoderm throughout early embryonic development to fulfill its ancestral function, *tgif* was expressed first in the mesoderm and then in the mesoderm and the endoderm
8. *tgif* mesoderm expression at mid-gastrula stage was either lost in asteroids or gained in the lineage leading to the last common ancestor of echinozoans
9. *erg* was expressed in the skeletogenic lineage at least as late in development as mid-gastrula stage
10. *hex* endodermal expression is acquired early in asteroid embryogenesis or lost in last common ancestor of extant echinozoans

At least 268 mya: in ancestral embryos at the cidaroid–euechinoid divergence, e.g., in *Archaeocidaris* embryos

11. *erg*, *hex*, and *tel* were initiated in a few cells at the center of the vegetal pole; later in the lineage leading to camaradont euechinoids following the cidaroid–euechinoid divergence, these three genes are restricted PMCs prior to PMC ingression
12. *tgif* remains expressed in mesodermal cells that ingressed into the blastocoel (*tgif* is not expressed in mesodermal cells that have ingressed in holothuroids)

Fig. 3 Co-option of the *erg–hex–tgif* subcircuit in the echinoderm clade. **a–c** Gene regulatory network (GRN) diagrams of the *erg–hex–tgif* subcircuit showing its wiring and embryonic domain of expression during development in three echinoderm clades. GRN diagrams were constructed using the BioTapestry online software suite (biotapestry. org) **a** GRN diagram of the *erg–hex–tgif* subcircuit in asteroids based on data presented in McCauley et al. (2010). In asteroids, this subcircuit is expressed in the endomesoderm in early blastulae and *tbrain* is its early activator. **b** Hypothetical GRN diagram of the *erg–hex–tgif* subcircuit in cidaroids, in which it is expressed exclusively in the mesoderm in pregastrular embryos. The *transparency* and *question mark* indicate that the *cis*-regulatory inputs between these genes have not been verified by perturbation analysis. The early activator of this GRN is not known. **c** GRN diagram of the *erg–hex–tgif* subcircuit in euechinoids, as presented in McCauley et al. (2010). In euechinoids, this circuit is expressed exclusively in the PMCs, a population of mesodermal cells that give rise to the embryonic skeleton, and its early activators are *tbrain* and *ets1/2*. **d** Phylogeny depicting the geological era of divergence and co-option events of the *erg–hex–tgif* subcircuit in three echinoderm clades in which this subcircuit has been investigated. Divergence estimates are taken from Thompson et al. (2015) for the euechinoid–cidaroid divergence and Jell (2014) for the asterozoan–echinozoan divergence. *Red bars* indicate a co-option event, and the embryonic address to which the subcircuit was co-opted is also indicated

that this subcircuit was likely deployed in the mesoderm of the last common ancestor of cidaroids and euechinoids, e.g., in the embryos of the taxon *Archaeocidaris*, which lived over 268 mya (Thompson et al. 2015). Furthermore, this rigorously demonstrates that the PMC and NSM restricted expression of the *erg–hex–tgif* subcircuit in euechinoids is a derived character state and must have arisen in stem-group or early crown-group euechinoids since the euechinoid–cidaroid divergence. Although these data indicate that the broader mesodermal utilization of the *erg–hex–tgif* subcircuit in cidaroids is basal with respect to the more specific deployment in euechinoids, this is but one set of characters, and phylogenetic analyses indicate that neither the cidaroids nor euechinoids are more ancestral than the other (Thompson et al. 2015).

Lastly, these results proffer a straightforward explanation as to the evolution of the spatial control of the *erg–hex–tgif* conserved GRN circuitry in echinoderms (Fig. 3). An interesting hypothesis is that *tbrain* is the ancestral regulator of the *erg–hex–tgif* subcircuit in echinoderms. This is consistent with the observations that the
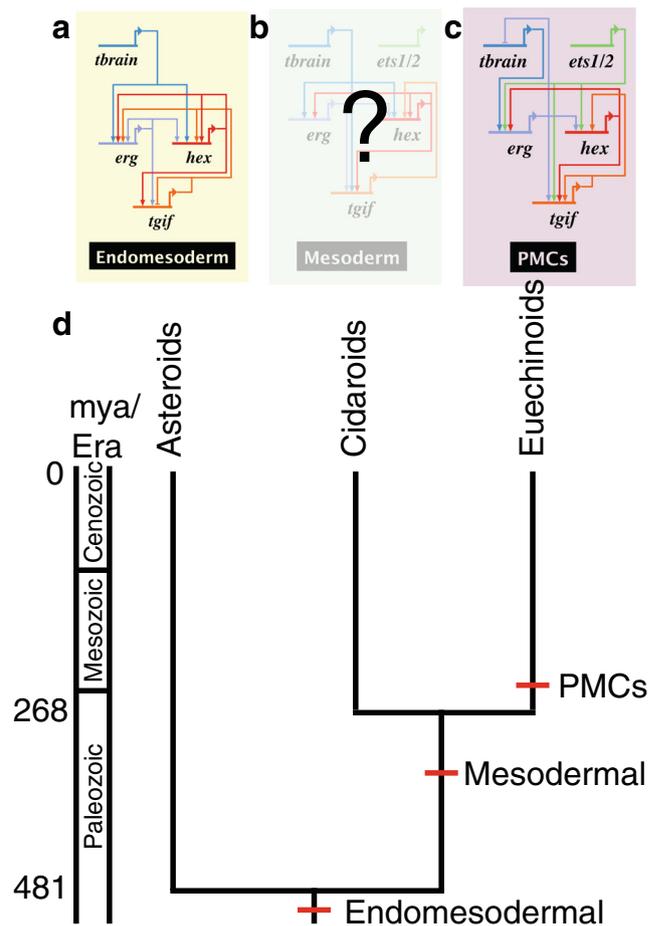
embryonic spatial expression of *tbrain* grades from PMC restricted in euechinoids (Ettensohn 2013), to broadly mesodermal in cidaroids and holothuroids (Erkenbrack and Davidson 2015; McCauley et al. 2012), and to

mesodermal and endodermal in asteroids (McCauley et al. 2010). That the *erg–hex–tgif* subcircuit also exhibits these expression patterns in the same clades suggests that all four of these genes may be recursively wired. Further, *tbrain* solely regulates this subcircuit in asteroids and partially regulates it in euechinoids in spite of the fact that *tbrain* is expressed at different development addresses in these clades. Taken together, these observations are consistent with the hypothesis that this subcircuit is downstream of *tbrain* in eleutherozoans. Developmentally, this suggests that, as proposed by McCauley et al. (2010), the ancestral function of the *erg–hex–tgif* subcircuit was to stabilize the mesodermal regulatory state.

Our data lend support to this hypothesis and also add weight to the hypothesis put forward by Gao and Davidson (2008), that whole apparatus of ancestral mesodermal GRN circuitry were loaded into the micromere embryonic address in the lineage leading to modern euechinoids following the euechinoid–cidaroid divergence—a truly remarkable case of evolutionary co-option (Fig. 3). Phenomena like those observed here are best explained by the observation that GRNs are fundamentally hierarchical and modular in nature (Davidson and Erwin 2006). The *erg–hex–tgif* kernel in the early embryogenesis of these echinoderms provides an extraordinary example of the modularity and clade-specific functions of GRNs in evolution and development. The correspondence of spatial expression of the *erg–hex–tgif* kernel to *ets1/2* and *tbrain* in these disparate clades, sea stars, sea cucumbers, and sea urchins, suggests that, even though these organisms last shared a common ancestor over 481 mya, the regulatory embrace they find themselves locked in is so difficult to genomically disentangle, that during evolution they are deployed differentially around the embryo as a parcel.

References

Davidson EH, Erwin DH (2006) Gene regulatory networks and the evolution of animal body plans. Science 311:796–800

Erkenbrack EM, Davidson EH (2015) Evolutionary rewiring of gene regulatory network linkages at divergence of the echinoid subclasses. Proc Natl Acad Sci U S A 112:E4075–E4084

Ettensohn CA (2013) Encoding anatomy: developmental gene regulatory networks and morphogenesis. Genesis 51:383–409

Gao F, Davidson EH (2008) Transfer of a large gene regulatory apparatus to a new developmental address in echinoid evolution. Proc Natl Acad Sci U S A 105:6091–6096

Gao F, Thompson JR, Petsios E, Erkenbrack EM, Moats RA, Bottjer DJ, Davidson EH (2015) Juvenile skeletogenesis in anciently diverged sea urchin clades. Dev Biol 400:148–158

Howard-Ashby M, Materna SC, Brown T, Chen L, Cameron RA, Davidson EH (2006) Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development. Dev Biol 300:74–89

Jell PA (2014) A Tremadocian asterozoan from Tasmania and a late Llandovery edrioasteroid from Victoria. Alcheringa 38:528–540

Materna SC, Ransick A, Li E, Davidson EH (2013) Diversification of oral and aboral mesodermal regulatory states in pregastrular sea urchin embryos. Dev Biol 375:92–104

Mavrothalassitis G, Ghysdael J (2000) Proteins of the ETS family with transcriptional repressor activity. Oncogene 19:6524–6532

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 340:200–208

McCauley BS, Wright EP, Exner C, Kitazawa C, Hinman VF (2012) Development of an embryonic skeletogenic mesenchyme lineage in a sea cucumber reveals the trajectory of change for the evolution of novel structures in echinoderms. EvoDevo 3:17

Oliveri P, Tu Q, Davidson EH (2008) Global regulatory logic for specification of an embryonic cell lineage. Proc Natl Acad Sci U S A 105: 5955–5962

Peter IS, Davidson EH (2011) A gene regulatory network controlling the embryonic specification of endoderm. Nature 474:635–639

Poutska AJ, Kuhn A, Groth D, Weise V, Yaguchi S, Burke RD, Herwig R, Lehrach H, Panopoulou G (2007) A global view of gene expression in lithium and zinc treated sea urchin embryos: new components of gene regulatory networks. Genome Biol 8:r85

Rizzo F, Fernandez-Serra M, Squarzoni P, Archimandritis A, Arnone MI (2006) Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*). Dev Biol 300:35–48

Schroeder T (1981) Development of a "primitive" sea urchin (*Eucidaris tribuloides*): irregularities in the hyaline layer, micromeres, and primary mesenchyme. Biol Bull 161:141–151

Seidel JJ, Graves BJ (2002) An ERK2 docking site in the pointed domain distinguishes a subset of ETS transcription factors. Genes Dev 16: 127–137

Solek CM, Oliveri P, Loza-Coll M, Schrankel CS, Ho ECH, Wang G, Rast JP (2013) An ancient role for Gata-1/2/3 and Scl transcription factor homologs in the development of immunocytes. Dev Biol 382: 280–292

Soufi A, Jayaraman P (2008) PRH/Hex: an oligomeric transcription factor and multifunctional regulator of cell fate. Biochem J 412:399–413

Thompson JR, Petsios E, Davidson EH, Gao F, Erkenbrack EM, Bottjer DJ (2015) Reorganization of sea urchin gene regulatory networks at least 268 million years ago as revealed by oldest fossil cidaroid echinoid. Sci Rep 5:15541

Wray GA, McClay DR (1988) The origin of spicule-forming cells in a 'primitive' sea urchin (Eucidaris tribuloides) which appears to lack primary mesenchyme cells. Development 103:305–315

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 141:2669–2679