

# Global patterns of enhancer activity during sea urchin embryogenesis assessed by eRNA profiling

Jian Ming Khor, Jennifer Guerrero-Santoro, William Douglas, and Charles A. Ettensohn

Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

We used capped analysis of gene expression with sequencing (CAGE-seq) to profile eRNA expression and enhancer activity during embryogenesis of a model echinoderm: the sea urchin, *Strongylocentrotus purpuratus*. We identified more than 18,000 enhancers that were active in mature oocytes and developing embryos and documented a burst of enhancer activation during cleavage and early blastula stages. We found that a large fraction (73.8%) of all enhancers active during the first 48 h of embryogenesis were hyperaccessible no later than the 128-cell stage and possibly even earlier. Most enhancers were located near gene bodies, and temporal patterns of eRNA expression tended to parallel those of nearby genes. Furthermore, enhancers near lineage-specific genes contained signatures of inputs from developmental gene regulatory networks deployed in those lineages. A large fraction (60%) of sea urchin enhancers previously shown to be active in transgenic reporter assays was associated with eRNA expression. Moreover, a large fraction (50%) of a representative subset of enhancers identified by eRNA profiling drove tissue-specific gene expression in isolation when tested by reporter assays. Our findings provide an atlas of developmental enhancers in a model sea urchin and support the utility of eRNA profiling as a tool for enhancer discovery and regulatory biology. The data generated in this study are available at Echinobase, the public database of information related to echinoderm genomics.

[Supplemental material is available for this article.]

Spatial and temporal control of gene expression is the hallmark of animal embryogenesis. Gene regulatory programs that drive development can be viewed as dynamic, lineage-restricted transcriptional networks, also known as developmental gene regulatory networks (GRNs) (Davidson and Levine 2008; Peter and Davidson 2015, 2016). The core components of developmental GRNs are the suites of transcription factors (TFs) present in a cell at any specific time, which define the transient regulatory state of that cell, and the noncoding, *cis*-regulatory DNA modules (CRMs) to which these TFs bind. A detailed understanding of the genetic circuitry that drives development therefore requires the identification of regulatory DNA elements and an analysis of their dynamic patterns of utilization during embryogenesis.

Enhancers, traditionally characterized by their ability to function over long distances and their insensitivity to orientation, play a central role in controlling spatial and temporal patterns of differential gene expression during development (Ong and Corces 2012; Farley et al. 2015; Long et al. 2016; Furlong and Levine 2018). The importance of enhancers has been shown, in part, by their ability to drive the transcription of reporter genes *in vivo* in temporal and spatial patterns that partially or fully recapitulate expression patterns of endogenous genes. Enhancers are themselves transcribed, a process that generates a class of noncoding RNAs known as enhancer RNAs (eRNAs) (De Santa et al. 2010; Kim et al. 2010; for review, see Arnold et al. 2020; Sartorelli and Lauberth 2020; Hou and Kraus 2021). Although their characteristics are still being actively investigated, eRNAs are typically short (<400 nt), capped, nonpolyadenylated, unspliced, nuclear RNAs (Andersson et al. 2014). They are typically bidirectionally transcribed, although there have also been reports of various levels of strand-biased transcription at enhancers (De Santa et al. 2010; Koch et al. 2011;

Mikhaylichenko et al. 2018), and the bidirectional transcription of enhancers observed at a cell population level reflects a mixture of highly strand-biased transcriptional patterns at a single-cell level (Kouno et al. 2019). Because eRNAs are typically nonpolyadenylated and are of low abundance, they are not readily detectable by most conventional, steady-state mRNA-sequencing methods, and their annotation is facilitated by specialized approaches such as cap analysis of gene expression (CAGE) (Murakawa et al. 2016; Sakaguchi et al. 2018; Hirabayashi et al. 2019; Kouno et al. 2019; Morioka et al. 2020). It remains somewhat unclear whether the process of transcription itself, the eRNA products of that transcription, or both are important for enhancer activity. A growing number of studies have shown that eRNAs have important molecular functions and can act to facilitate enhancer-promoter looping, recruit TFs, or release paused RNA polymerase II (Sartorelli and Lauberth 2020; Hou and Kraus 2021).

Predicting the location and activity of enhancers remains a significant challenge, although a variety of strategies are currently available (Calo and Wysocka 2013; Suryamohan and Halfon 2015; Catarino and Stark 2018). Recent studies have shown that eRNA expression is a highly reliable indicator of enhancer activity, and analyses of specific genetic loci have revealed that eRNA synthesis precedes or coincides with transcription from the associated locus (Andersson et al. 2014; Arner et al. 2015; Kim et al. 2015; Henriques et al. 2018; Mikhaylichenko et al. 2018; Tyssowski et al. 2018; Hirabayashi et al. 2019). Based on these findings, several studies have successfully leveraged eRNA expression to analyze specific sets of enhancers (Wang et al. 2011; Cauchy et al. 2016; Baillie et al. 2017; Denisenko et al. 2017). There is a growing view that eRNA expression is the most reliable marker of enhancer

**Corresponding author:** [ettensohn@cmu.edu](mailto:ettensohn@cmu.edu)

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.275684.121>.

© 2021 Khor et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

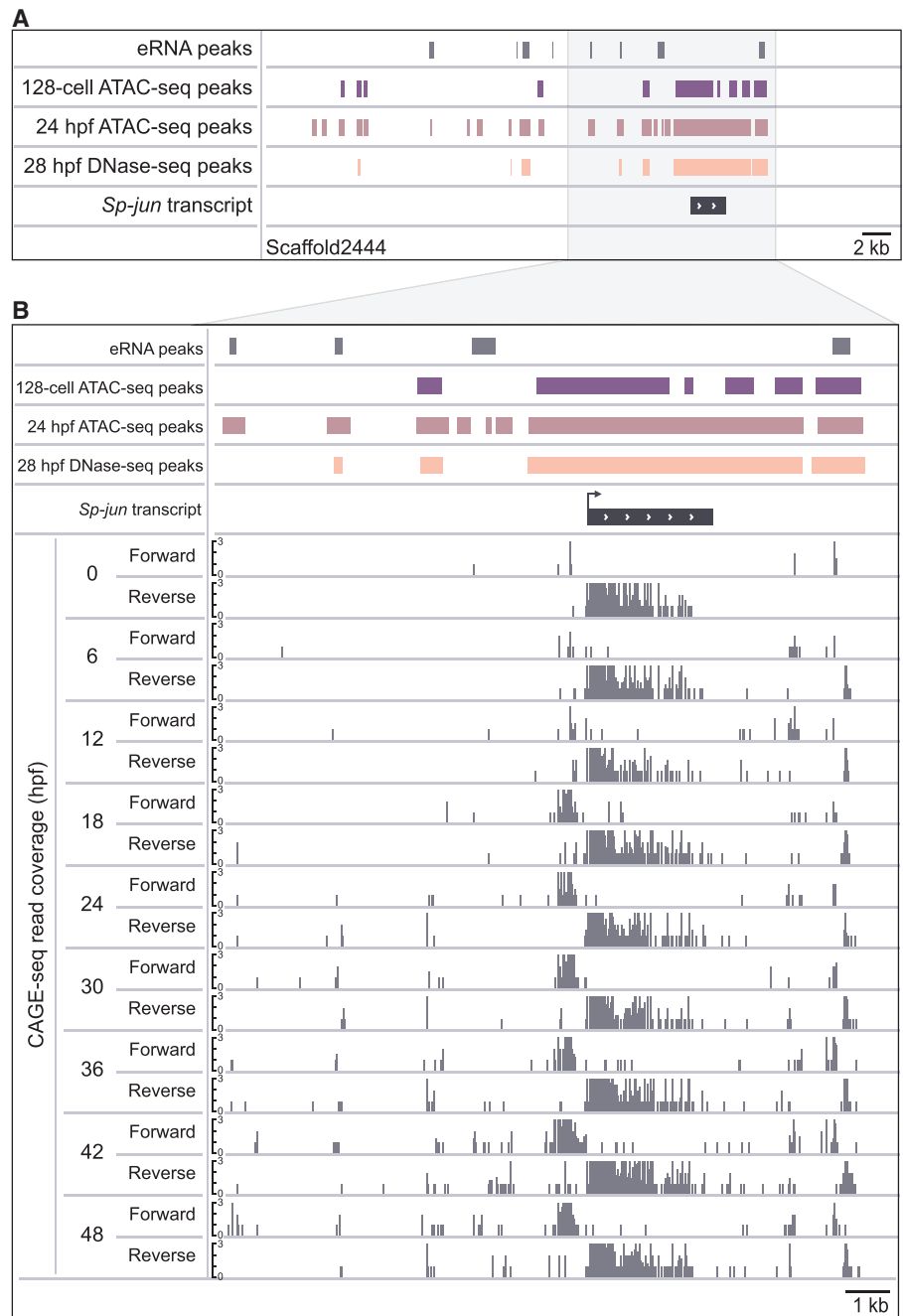
activity currently available (Arnold et al. 2020; Sartorelli and Lauberth 2020).

This work has two related goals: (1) to leverage eRNA expression to provide a global picture of enhancer activity during the early development of an animal embryo, and (2) to thereby provide a resource to the echinoderm research community to aid in *cis*-regulatory analysis and GRN biology. We profiled eRNA expression during early embryogenesis of the purple sea urchin, *Strongylocentrotus purpuratus*, an organism used widely for developmental studies, including GRN biology. Through deep-sequencing of CAGE libraries, we provide an atlas of more than 18,000 developmental enhancers and confirm the activity of a representative subset by transgenic reporter assays. We use eRNA expression data to analyze the patterns of activity of developmental enhancers and identify global trends as well as signatures of lineage-specific transcriptional programs. This work provides an important resource for future studies of gene regulatory processes that underlie the development of sea urchins and other echinoderms. The data generated in this study are available through Echinobase (<https://www.echinobase.org/>), the public database of genomic resources related to echinoderms.

## Results

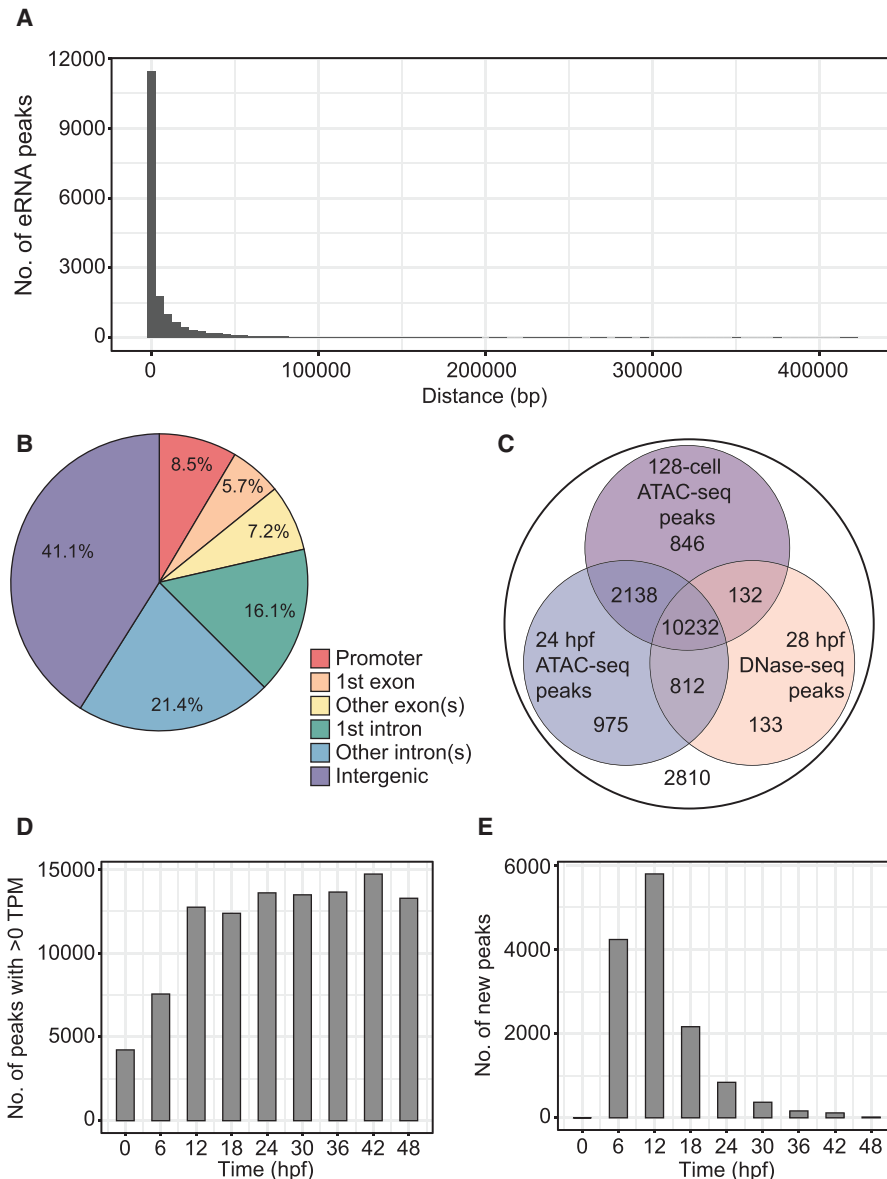
### General characterization of embryonic eRNAs

We collected samples of total RNA from *S. purpuratus* embryos immediately after fertilization (0 hpf) and at successive 6-h intervals until the end of gastrulation (6, 12, 18, 24, 30, 36, 42, and 48 hpf). CAGE libraries were sequenced at high depth (more than 100 million reads per sample), and eRNAs were identified through a bioinformatics pipeline described by Hirabayashi et al. (2019) (an example of eRNA signal near the *Sp-jun* gene is shown in Fig. 1). This analysis led to the identification of 18,078 distinct eRNAs expressed in diverse temporal patterns during early *S. purpuratus* embryogenesis. As has been well documented, the steady-state abundance of eRNAs was much lower on average than that of mRNAs. Of the genomic elements associated with these eRNAs, which represent putative enhancers, approximately half (50.4%) overlapped annotated genes by at least 1 nucleotide (nt). Most of the eRNA peaks that did not overlap annotated genes (the “intergenic” class in Fig. 2B) were located within 20 kb of the annotated genes, as ex-



**Figure 1.** Examples of eRNA peaks near *Sp-jun*. (A) Three of the four eRNA peaks shown (top track; gray bars) overlap with chromatin regions that are hyperaccessible at 24 hpf but not at the 128-cell stage, as determined by ATAC-seq and DNase-seq (Shashikant et al. 2018). (B) Temporal CAGE-seq coverage tracks illustrate the bidirectional expression of eRNAs and the very low abundance of eRNA reads relative to coding genes (scale: zero to three read counts). Sea urchin *jun* is expressed at high levels maternally, followed by a second peak of expression at gastrula stage. Spatially, *jun* is distributed homogeneously in unfertilized egg and during early cleavages and is restricted to the PMCs at later stages (Russo et al. 2014).

pected given that the average intergenic distance in *S. purpuratus* is 23.5 kb (Fig. 2A; Tu et al. 2012). Of the eRNA peaks that overlapped annotated genes, 63.7% were located in introns. In addition, 8.5% of all eRNA peaks were located within promoters (defined as a 2-kb region upstream of the transcriptional start site [TSS]) (Fig. 2B); 2951/18,078 eRNA peaks (or 16.3% of all peaks) were not located



**Figure 2.** Annotation and analysis of eRNA peaks. (A) Frequency histogram illustrating peak-to-gene distances, with each bar representing 5000 bp. (B) Pie chart showing the location of eRNA peaks relative to nearest annotated gene. (C) Venn diagram showing the distribution of eRNA peaks that overlap with regions of chromatin shown to be hyperaccessible at 128-cell stage and 24 hpf by ATAC-seq and at 28 hpf by DNase-seq (Shashikant et al. 2018). (D) Total number of eRNAs (greater than zero TPM expression) at each time point. (E) Number of newly-appearing eRNAs at each time point.

within 20 kb of any annotated gene. In some cases, however, these peaks were located near the ends of genomic scaffolds, and their positions relative to genes could not be accurately determined. The eRNA peaks that overlapped with or were located within 20 kb of annotated genes (83.7% of all eRNA peaks, or 15,127 peaks) were associated with 7528 genes. The great majority (14,422/18,078, or 79.8%) of eRNA peaks overlapped regions of chromatin that were previously shown by ATAC-seq and/or DNase-seq to be hyperaccessible at 24 hpf, a time point in the middle of the developmental period sampled in this study (0–48 hpf) (Fig. 2C; Shashikant et al. 2018). We also found that most eRNA peaks (13,348/18,078, or 73.8%) overlapped regions of chromatin that were

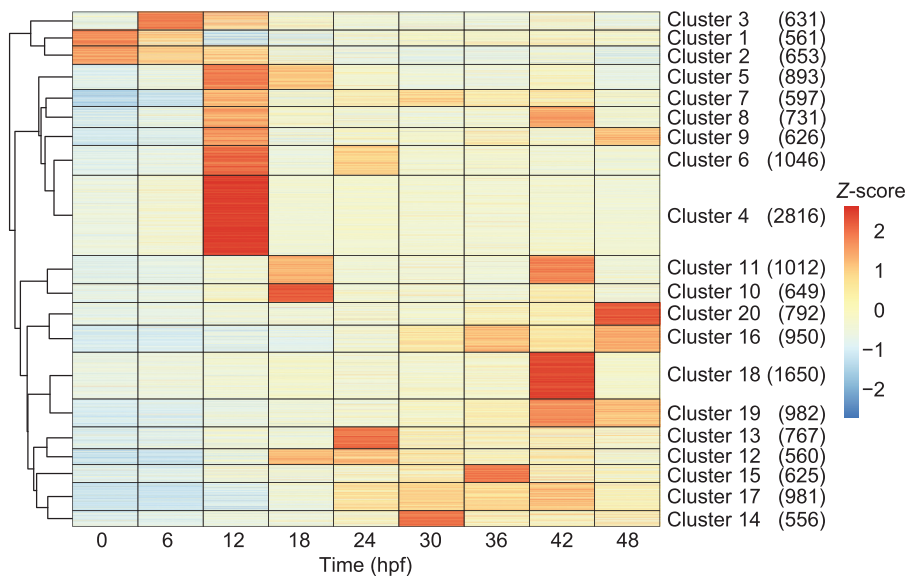
open at the 128-cell stage based on ATAC-seq data (Shashikant et al. 2018). We identified some eRNA peaks (1920/18,078, or 10.6%) that were inaccessible at the 128-cell stage and open at 24–28 hpf (some examples can be seen in Fig. 1B), but these were relatively uncommon (Fig. 2C).

### Temporal patterns of eRNA expression during early sea urchin development

Our developmental profile of eRNA expression identified approximately 4000 eRNAs in the fertilized egg (0 hpf) (Fig. 2D). These eRNAs mark enhancers that regulate the expression of the more than 9000 transcripts that are maternally provisioned in *S. purpuratus* (Tu et al. 2014) and may preferentially reveal enhancers that are active late in oogenesis, as we sampled mature oocytes. The great majority of eRNAs (13,706/18,078, or 75.8%) lacked detectable maternal expression and thus marked enhancers that were active only zygotically. We found that the diversity of the eRNA population increased during cleavage and early blastula stages, increasing by >300% between 0 hpf and 12 hpf, whereas at later stages of embryogenesis, the total number of eRNAs remained relatively constant (Fig. 2D). Early-zygotic deployment of enhancers was also evident when we focused on the number of newly-appearing eRNAs at each developmental stage, which was maximal at 6 and 12 hpf (Fig. 2E). Although eRNAs showed diverse patterns of expression during embryogenesis (see below), we observed that, once activated, zygotic enhancers tended to remain active through gastrulation. Thus, of the 13,706 eRNAs that appeared during early embryogenesis, 73.9% (10,122/13,706) were expressed at 48 hpf, the latest developmental stage we examined. We found that maternally provisioned eRNAs were, on average, located closer to gene bodies and more likely to be found in promoter regions (the 2-

kb region upstream of the TSS), although the biological significance of this is unclear (Supplemental Fig. S1). The average size of eRNA peaks showed little change during development (mean length = 320–360 at all stages examined).

We used *k*-means clustering to assess the temporal expression patterns of eRNAs during early development. We partitioned the expression patterns of eRNAs into 20 clusters (Fig. 3; Supplemental Fig. S2), a number that was chosen based on estimations of sum of squares of cluster tightness as a function of cluster number (Supplemental Fig. S3A). Our analysis revealed a relatively small subset of eRNAs (1214/18,078, or 6.7%) that were maximally expressed in the fertilized egg (clusters 1 and 2) (Supplemental Fig. S4). A large



**Figure 3.** Heatmap based on *k*-means clustering (*K* = 20) of eRNA temporal expression profiles. The number of eRNAs in each cluster is shown in parentheses.

fraction of eRNAs (7340/18,078, or 40.6%) showed peaks in expression at 6 or 12 hpf (clusters 3–9). Within this group, we detected a small set of transiently expressed eRNAs (cluster 3) that were maximally expressed at 6 hpf; these eRNAs mark the first enhancers to be zygotically activated in *S. purpuratus*. The remaining ~60% of embryonic eRNAs were expressed mostly at later developmental stages and included sets that peaked in expression at 18 hpf (Cluster 10), 24 hpf (cluster 13), 30 hpf (cluster 14), or 36 hpf (cluster 15). A substantial fraction of eRNAs (6367/18,087 or 35.2%) was maximally expressed late in gastrulation (42 hpf or 48 hpf; clusters 11, 16–20). A subset of zygotic eRNAs showed biphasic patterns of expression; that is, they were activated early in development, declined in expression, and then increased again later in development (e.g., clusters 6, 8, 9, and 11).

For each expression cluster, we examined Gene Ontology (GO) terms associated with the nearest annotated genes located within 20 kb of the eRNAs in that cluster. This analysis revealed largely nonoverlapping patterns of GO term enrichment (Supplemental Fig. S5). For example, genes associated with Wnt signaling were preferentially associated with cluster 5. Enhancers in this cluster showed maximal activity at 12 hpf, a time when canonical and noncanonical Wnt signaling patterns the early embryo along the animal–vegetal axis (Cui et al. 2014; Martínez-Bartolomé and Range 2019). Several clusters showed enrichment of genes associated with regulation of RNA polymerase II-mediated transcription, and cluster 19 (peak expression at 42 hpf) was preferentially associated with genes associated with growth regulation. De novo motif analysis of eRNA peaks using HOMER also revealed cluster-specific signatures of likely TF binding sites (Supplemental Table S1).

We next asked whether patterns of eRNA expression were related to the expression patterns of likely target genes. For this analysis, we used mRNA expression values from the same nine CAGE-seq libraries that were used for eRNA identification, and focused on the expression of (1) genes that had at least one eRNA peak located within 20 kb of the gene at any stage (5837 genes) and (2) eRNAs that were within 20 kb of any gene (13,056 eRNAs). The expression

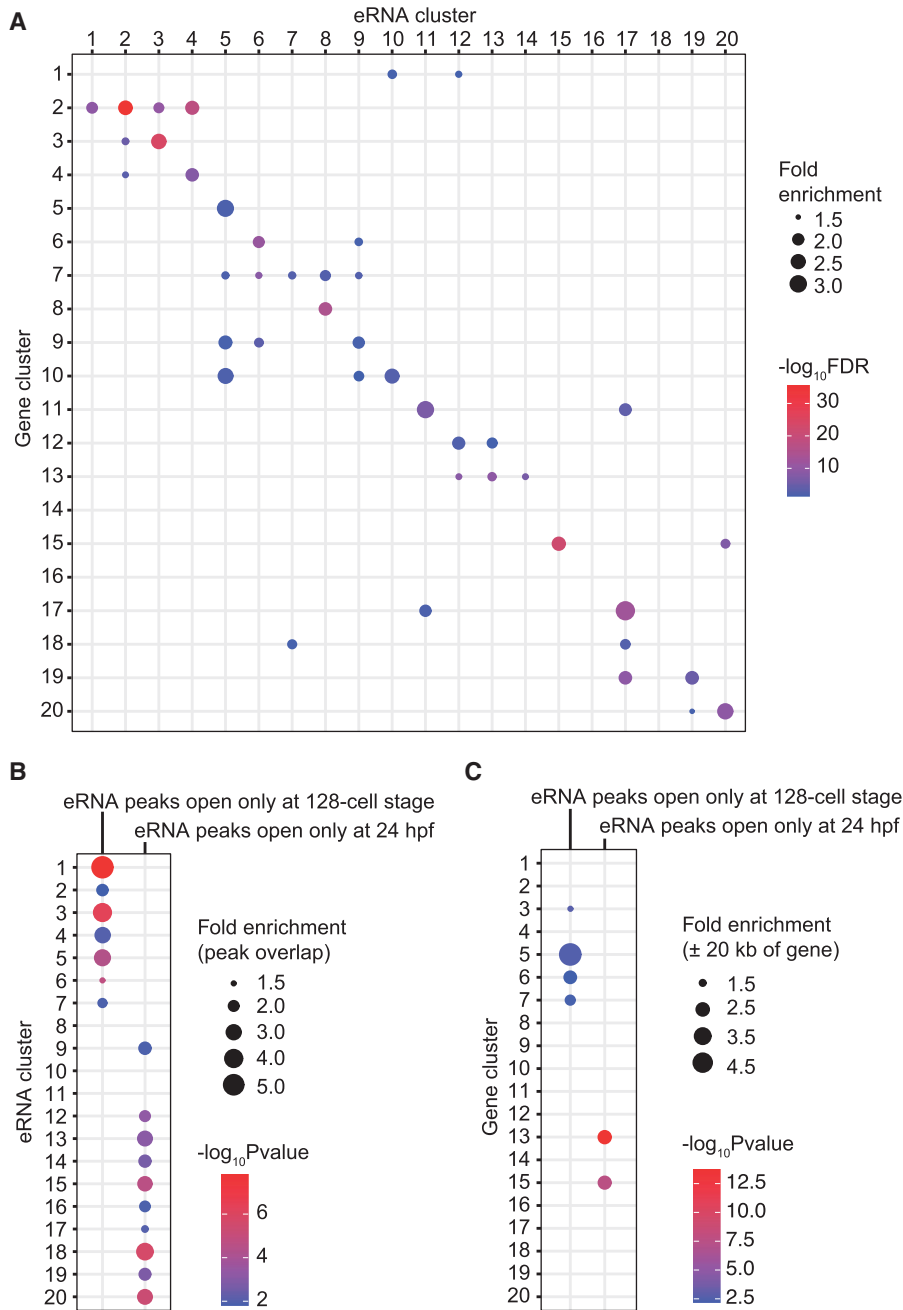
levels of all eRNAs and mRNAs that met these criteria were first normalized to their mean expression levels, thereby allowing the two different classes of temporal expression data to be pooled. *k*-Means clustering of the pooled eRNA/mRNA expression patterns led to the identification of 20 expression clusters, a number that was chosen based on estimations of sum of squares of cluster tightness as a function of cluster number, as described above (Supplemental Fig. S6). Analysis of these clusters showed that eRNAs in a given expression cluster were preferentially located near the genes in that same expression cluster or located near genes in other clusters with expression profiles that overlapped that of the “parental” cluster (Fig. 4A). Thus, the expression profiles of eRNAs tended to resemble those of nearby genes; namely, genes that are the most likely direct targets of the cognate enhancers.

As shown above (Fig. 2C), most eRNA peaks were accessible at both the 128-cell stage (~9 hpf) and 24 hpf. We looked more closely at the subsets of eRNA peaks that were open (1) only at the 128-cell stage or (2) only at 24 hpf. Peaks in the former set were found preferentially in early eRNA expression clusters, whereas those in the latter set were found preferentially in late eRNA expression clusters (Fig. 4B,C). Thus, although most eRNA peaks are open at both stages, some peaks reveal a relationship between enhancer accessibility and activity.

We also noted that a substantial fraction (~18%) of the 13,348 enhancers accessible at the 128-cell stage showed no detectable eRNA expression at 0, 6, or 12 hpf, and 40% showed below-mean levels of expression at all three developmental stages. These enhancers also showed a variety of later temporal patterns of eRNA expression, but several thousand fell into expression clusters with maximal expression many hours later in development. For example, approximately 2000 of the enhancers accessible at the 128-cell stage fell into expression clusters 18–20, with eRNA expression maxima at 42 or 48 hpf. These findings indicate that the accessibility of sea urchin enhancers can precede their activity, and suggest that many of these regulatory elements undergo a progressive developmental maturation after they become accessible.

### Signatures of lineage-specific GRNs

Several developmental GRNs have been characterized in sea urchins, including those deployed in skeletogenic primary mesenchyme cells (PMCs), pigment cells, and the ciliary band. We examined enhancers located within 20 kb of genes shown previously to be differentially expressed by PMCs (Rafiq et al. 2014), pigment cells (Calestani et al. 2003; Barsi et al. 2015), and ciliated band cells (Barsi et al. 2015). De novo motif analysis of these sets of enhancers using HOMER revealed distinct patterns of enriched motifs; more significantly, each set was enriched in motifs associated with cell type-specific TFs known to have important regulatory functions in the GRN deployed in that particular lineage (Table 1). Thus, eRNA peaks located near genes differentially expressed by PMCs were enriched in motifs that matched binding sites for Ets1



**Figure 4.** Expression pattern correlation analysis of eRNAs and nearby genes (cluster analysis summarized in Supplemental Fig. S6). (A) Fold enrichment of clustered eRNAs within 20 kb of genes in “parental” or other clusters (compared with total eRNAs in this analysis) is represented by the circles, and the color of the circles correspond to the significance of the enrichment, expressed as  $-\log_{10}(\text{FDR})$ . Only cluster pairs that show greater than 1.5-fold enrichment and  $\text{FDR} < 0.05$  are shown. (B) Analysis of eRNA peaks that are exclusively accessible at the 128-cell stage (128-cell eRNAs) or 24 hpf (24 hpf eRNAs) to overlap with eRNA peaks from different temporal expression clusters. (C) Analysis of eRNA peaks that are exclusively accessible at the 128-cell stage or 24 hpf to be within 20 kb of clustered genes. Fold enrichment is represented by the size of the circles, and the colors correspond to the significance of the enrichment, expressed as  $-\log_{10}(P\text{-value})$ . Only enrichments that show greater than 1.5-fold enrichment and  $P < 0.05$  are shown.

and Alx1, two key early TFs in the skeletogenic GRN (Kurokawa and Etensohn 1999; Etensohn et al. 2003; Khor and Etensohn 2020), whereas those located near genes differentially expressed

by pigment cells showed an enrichment of consensus binding sites for Gcm, a pivotal regulator of pigment cell specification (Ransick and Davidson 2006; Wessel et al. 2020). Similarly, eRNA peaks located near genes differentially expressed in the ciliated band were enriched in consensus binding sites for Hnf6 (Otim et al. 2004), a component of the ciliary band GRN.

ATAC-seq and DNase-seq have been used to identify regions of chromatin that are differentially open in PMCs relative to other cell types (Shashikant et al. 2018), and ChIP-seq data are available for Alx1, which is expressed specifically in PMCs (Khor et al. 2019). We examined the set of eRNA peaks that were located within 20 kb of genes differentially expressed by PMCs (a total of 734 eRNA peaks) and found that they were significantly more likely to overlap Alx1-bound regions and regions differentially accessible in PMCs than eRNAs as a whole (Supplemental Fig. S7A). Our analysis required only that at least 1 nt of the eRNA peak overlap the secondary genomic feature, but increasing the overlap requirement had little effect, highlighting the correspondence between these genomic features (Supplemental Table S7). *k*-Means clustering of the expression patterns of the 734 corresponding eRNAs identified six clusters (Supplemental Fig. S7B,C). Many of these eRNAs (36.4%; clusters 2 and 3) were activated zygotically with peak expression at 12 hpf or 18 hpf, time points that corresponded closely to the cell-autonomous, early-zygotic activation of many effector genes in the PMCs GRN (Rafiq et al. 2014). We found it noteworthy that another large subset of eRNAs (38.1%; clusters 5 and 6) showed peak expression at 42 hpf or 48 hpf; this may reflect the well-documented global shift in the regulation of PMC effector genes to a signal-dependent mode during gastrulation (Sun and Etensohn 2014).

**Experimental validation of predicted enhancers**

To assess the potential value of eRNAs for enhancer discovery and GRN biology, we took advantage of a relatively large set of CRMs from *S. purpuratus* that has been identified and analyzed to varying degrees in published studies, typically through the use of in vivo reporter assays. A comprehensive survey of the literature identified 92 such CRMs (Supplemental Table S2). We found that a large fraction of these elements (55/92, or 60%)

**Table 1.** De novo motif analysis of eRNAs near genes differentially expressed in specific embryonic territories

Embryonic territory	Motif rank	P-value	Targets (%)	Background (%)	Background (%)	Best match (JASPAR/HOMER)	Match score	Cell type-enriched homolog with similar motif	
PMC	1	1 x 10 <sup>-26</sup>	19.8	7.3	7.3	Ews:Erg-fusion (Ets)	0.98	Ets1	
	2	1 x 10 <sup>-16</sup>	23.1	11.9	11.9	JunD	0.96	Fosl, Jun	
	3	1 x 10 <sup>-13</sup>	10.8	4.1	4.1	Dbp	0.91		
	4	1 x 10 <sup>-13</sup>	25.1	14.6	14.6	Lhx2	0.94	Alx1, Alx11	
	5	1 x 10 <sup>-13</sup>	4.5	0.9	0.9	Pou3f4	0.77		
	Pigment cell	1	1 x 10 <sup>-102</sup>	9.4	2.1	2.1	Gcm1	0.86	Gcm1
		2	1 x 10 <sup>-38</sup>	12.1	6.0	6.0	Sox3	0.90	Sox5, Sox171
		3	1 x 10 <sup>-34</sup>	14.9	8.3	8.3	Nfil3	0.91	Nfil3
		4	1 x 10 <sup>-27</sup>	5.1	1.9	1.9	Erg (Ets)	0.94	Ets1
		5	1 x 10 <sup>-23</sup>	5.3	2.2	2.2	Crx	0.92	
		6	1 x 10 <sup>-18</sup>	11.8	7.4	7.4	Ahr::Amt	0.76	LOC575350, Egr3
		7	1 x 10 <sup>-16</sup>	10.1	6.3	6.3	Otx2	0.74	Otx2
		8	1 x 10 <sup>-16</sup>	6.4	3.4	3.4	Irx3_2	0.87	
9		1 x 10 <sup>-16</sup>	0.5	0.0	0.0	Hbp1_2	0.60		
10		1 x 10 <sup>-15</sup>	29.1	23.0	23.0	Cst6	0.77		
Ciliated band	11	1 x 10 <sup>-14</sup>	7.7	4.6	4.6	Jdp2	0.96		
	12	1 x 10 <sup>-14</sup>	1.5	0.4	0.4	FoxO6	0.81	Foxk1, Foxm1, Foxn3	
	13	1 x 10 <sup>-13</sup>	9.1	5.8	5.8	Sox10	0.82	Sox5, Sox171	
	14	1 x 10 <sup>-13</sup>	7.5	4.5	4.5	Gata10	0.77	Gata6	
	15	1 x 10 <sup>-13</sup>	3.2	1.4	1.4	Gata14	0.64	Gata6	
	16	1 x 10 <sup>-13</sup>	10.2	6.7	6.7	Nfya	0.84		
	17	1 x 10 <sup>-12</sup>	3.5	1.7	1.7	Abi3	0.63		
	18	1 x 10 <sup>-12</sup>	1.3	0.3	0.3	Nfkb2	0.84		
	1	1 x 10 <sup>-37</sup>	29.2	19.3	19.3	Sox15	0.91	LOC575231	
	2	1 x 10 <sup>-29</sup>	6.3	2.4	2.4	Dbp	0.93	Sox9, Sox171, Sox4	
	3	1 x 10 <sup>-24</sup>	6.0	2.5	2.5	Gcm1	0.86		
	4	1 x 10 <sup>-20</sup>	11.2	6.6	6.6	JunD	0.71		
	5	1 x 10 <sup>-18</sup>	3.1	1.1	1.1	Ets (Ets)	0.92		
6	1 x 10 <sup>-17</sup>	12.2	7.6	7.6	Nfya	0.79	Hnf6		
7	1 x 10 <sup>-17</sup>	1.6	0.3	0.3	Nfya	0.78	Hnf6		
8	1 x 10 <sup>-14</sup>	1.1	0.2	0.2	Ash1	0.80	Mcm8		
9	1 x 10 <sup>-13</sup>	0.8	0.1	0.1	Athb20	0.71			
10	1 x 10 <sup>-13</sup>	6.3	3.5	3.5	Yrm1	0.80			
11	1 x 10 <sup>-12</sup>	6.9	4.0	4.0	Pitx3	0.87	Pitx1, Pitx2		
12	1 x 10 <sup>-12</sup>	5.4	2.9	2.9	Dux	0.73	Hnf6		
13	1 x 10 <sup>-12</sup>	0.2	0.0	0.0	Zscan4_2	0.68			

Genes differentially expressed by PMCs (420 genes) (Rafiq et al. 2014), pigment cells (2425 genes that are greater than fivefold enriched compared with controls) (Barsi et al. 2015), and ciliated band cells (3036 genes that are greater than twofold enriched compared with controls) (Barsi et al. 2015) were used for de novo motif analysis.



overlapped eRNA peaks by at least 1 nt. To provide additional evidence of the utility of eRNAs for CRM discovery, we selected 20 eRNA peaks and cloned these regions into the EpGFPII reporter plasmid, a vector that has been widely used for the identification and analysis of CRMs in sea urchins (Cameron et al. 2004). One criterion was applied in selecting eRNAs; we required that the nearest gene have a known spatial pattern of expression based on prior WMISH analysis, so that any tissue-specific pattern of reporter expression we observed could be compared with the endogenous expression pattern of the gene most likely to be regulated by that enhancer. We tested 10 eRNA peaks located near genes expressed selectively by PMCs, seven peaks near genes expressed selectively by pigment cells, and three peaks near genes expressed by blastocoelar cells. Significantly, we found that a high fraction (10/20, or 50%) of the peaks showed enhancer activity in isolation, and all 10 active elements supported reporter gene expression in tissue-restricted patterns that resembled the expression patterns of the closest endogenous genes (Fig. 5A,B; Table 2; Supplemental Table S8). The fraction of embryos expressing GFP was consistently <100% because of the mosaic inheritance of transgenes during embryonic development (see Methods). We typically found that constructs with especially low numbers of GFP-expressing embryos also

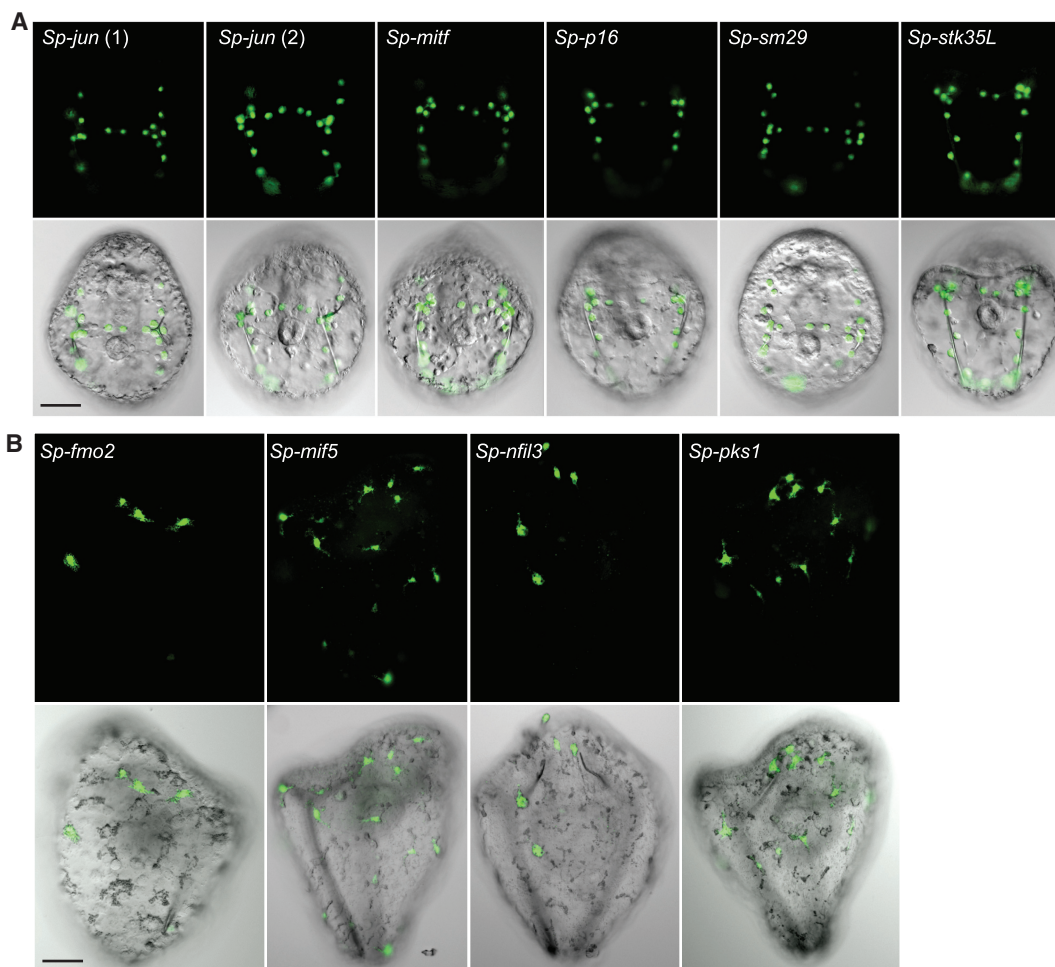
showed relatively faint GFP fluorescence (i.e., low expression), suggesting that some embryos with low levels of expression were probably not detected.

#### Public availability of *S. purpuratus* eRNA expression data

The eRNA expression data described in this study are publicly available. Tables containing the genomic coordinates of all eRNA peaks and expression data at the nine time points analyzed in this study are contained in the Supplemental Materials (Supplemental Tables S3–S6). In addition, JBrowse tracks indicating the locations of eRNA peaks at each developmental stage examined and a file containing the complete sequences of all eRNAs identified in this study are available at Echinobase (<https://www.echinobase.org/>), the public database of genomic resources related to sea urchins and other echinoderms.

#### Discussion

To our knowledge, this is the first study that leverages eRNA expression to profile enhancer activity during animal embryogenesis and the first to comprehensively profile developmental enhancer



**Figure 5.** Experimental validation of eRNA peaks using GFP reporter constructs. (A) Embryos (~48 hpf) injected with reporter constructs containing eRNA peaks near genes differentially expressed by PMCs. (B) Embryos (~72 hpf) injected with reporter constructs containing eRNA peaks near genes differentially expressed by pigment cells. Scale bar, 50  $\mu$ m.

**Table 2.** Summary of spatial expression patterns of GFP reporter constructs

eRNA peak	Closest gene	Embryos scored	GFP-expressing embryos (%)	PMC-only (%)	PMC and ectopic (%)	Ectopic-only (%)
Scaffold878:194028–194322	<i>Sp-clectin</i>			No expression		
Scaffold2444:124662–125218	<i>Sp-jun</i>	183	96 (52.5)	No expression	12 (12.5)	35 (36.5)
Scaffold2444:128312–128489		193	142 (73.5)	49 (51.0)	37 (26.1)	28 (19.7)
Scaffold2444:136194–136711		207	78 (37.7)	77 (34.2)	9 (11.5)	32 (41.0)
Scaffold743:206992–207569	<i>Sp-mitf</i>			No expression		
Scaffold520:512839–513272	<i>Sp-nk7</i>	222	63 (28.4)	No expression	4 (6.3)	8 (12.7)
Scaffold198:294374–294548	<i>Sp-p16</i>	185	13 (7.0)	12 (92.3)	0 (0.0)	1 (7.7)
Scaffold1119:169493–169889	<i>Sp-sm29</i>	246	62 (25.2)	58 (93.5)	3 (4.8)	1 (1.6)
Scaffold5796:28303–28866	<i>Sp-stk35L</i>			No expression		
Scaffold1985:181118–181460	<i>Sp-vegfr10</i>			No expression		
eRNA peak	Closest gene	Embryos scored	GFP-expressing embryos (%)	Pigment cell-only (%)	Pigment cell and ectopic (%)	Ectopic-only (%)
Scaffold856:135856–136256	<i>Sp-fmo2</i>	229	74 (32.3)	70 (94.6)	2 (2.7)	2 (2.7)
Scaffold775:9368–9884	<i>Sp-mif5</i>	234	42 (17.9)	38 (90.5)	1 (2.4)	3 (7.1)
Scaffold89:638821–639027	<i>Sp-nfil3</i>	174	45 (25.9)	40 (88.9)	2 (4.4)	3 (6.7)
Scaffold89:625203–625748				No expression		
Scaffold174:192881–193344	<i>Sp-pks1</i>	228	70 (30.7)	62 (88.6)	3 (4.3)	5 (7.1)
Scaffold174:190972–191109				No expression		
Scaffold500:293918–294295	<i>Sp-sult1c_2</i>			No expression		
eRNA peak	Closest gene	Embryos scored	GFP-expressing embryos (%)	Blastocoelar cell-only (%)	Blastocoelar cell and ectopic (%)	Ectopic-only (%)
Scaffold1386:69345–69725	<i>Sp-ese</i>			No expression		
Scaffold1386:69919–70182				No expression		
Scaffold1386:105544–106026				No expression		



activity in an organism other than *Drosophila* (Nègre et al. 2011; Thomas et al. 2011; Bonn et al. 2012; Kvon et al. 2014; Reddington et al. 2020). Our eRNA-based identification of approximately 18,000 developmental enhancers in *S. purpuratus* is very likely an underestimate given the challenges in detecting eRNAs, which are expressed at very low levels, the fragmented nature of the genome assembly we used, and the use of a single biological replicate. Moreover, in whole-embryo analyses, it may be difficult to detect eRNA signal from enhancers that are active in only a small fraction of embryonic cells. It is noteworthy in this regard, however, that we detected many enhancers associated with lineage-specific genes, including genes expressed specifically by skeletogenic PMCs, which represent only ~5% of the cells of the embryo.

Global patterns of eRNA expression during sea urchin embryogenesis parallel gene expression patterns in several respects. RNA-seq data indicate that approximately 16,500 genes are expressed at a level of at least 300 transcripts per embryo during all of *S. purpuratus* embryogenesis (0–72 hpf) (Tu et al. 2014). Our identification of around 18,078 enhancers yields an average value of 1.1 enhancers per gene, but the time frame is shorter in our analysis (0–48 hpf), and as noted above, there are reasons to believe that our estimate of the total number of enhancers is on the low side. Based on the CAGE-seq data generated in this study, we found that a total of 11,692 genes were expressed 0–48 hpf, a value that leads to an overall average of around 1.5 enhancers per gene. Of these expressed genes, 22% did not have an eRNA peak within 50 kb of the gene. The remaining 78% had at least one peak within 50 kb of the gene (57% of this subset had at least one peak within 20 kb of the gene). These percentages were presumably lowered by two factors: (1) the low expression levels of eRNAs, which meant that some were undoubtedly undetected, and (2) the incomplete nature of the *S. purpuratus* genome version 3.1 assembly (see Methods). We detected about 4000 maternally provisioned eRNAs, a number somewhat smaller than the approximately 9800 species of transcripts present in the unfertilized egg at levels greater than 300 copies per egg (Tu et al. 2014). Again, eRNA expression likely provides a conservative estimate of the actual number of developmental enhancers; in addition, eRNAs in the mature oocyte may be preferentially associated with enhancers that are active late in oogenesis. We detected 13,000–15,000 eRNAs at each embryonic stage (Fig. 2A), a number somewhat larger than the number of genes that, on average, are expressed at more than 300 copies per embryo in *S. purpuratus* at the same developmental stages (approximately 11,500) (Tu et al. 2014).

Our analysis revealed a burst of enhancer mobilization between fertilization and 12 hpf, a period that corresponds to cleavage and early blastula stages. This is consistent with considerable evidence that there is no period of transcriptional quiescence following fertilization in sea urchins. Instead, a large portion of the zygotic genome is activated during cleavage, leading to the rapid establishment of several, distinct embryonic territories by the early blastula stage (Davidson 1986; Davidson et al. 1998). Tu et al. (2014) found that early (pregastrula) development was a particularly dynamic phase when many genes showed marked changes in levels of expression, whereas at later developmental stages, patterns of gene expression remained more stable.

In general, the relationship between the accessibility of developmental enhancers and their activity is poorly understood. Our previous work showed that many PMC enhancers (79%) were open at the 128-cell stage, several hours before the expression of most PMC-specific transcripts (Shashikant et al. 2018). In this

study, using eRNA expression to globally catalog sea urchin enhancers, we found that a similar fraction (73.8%) of all enhancers active during the first 48 h of embryogenesis was open no later than the 128-cell stage, and possibly even earlier. There is evidence pointing to the early accessibility of developmental enhancers in other organisms. An analysis of chromatin accessibility during *Drosophila* development highlighted the dynamic nature of the chromatin landscape but also found that more than half of DNase I hypersensitive sites (DHSs) (which marked a variety of *cis*-regulatory elements, including enhancers) present in the late embryo were also open at the cellular blastoderm stage (Thomas et al. 2011). More recently, Reddington et al. (2020) also reported many examples of constitutively accessible DHSs across the same developmental window. They provided evidence that these were preferentially associated with ubiquitous chromatin remodelers and insulator proteins; in apparent contrast, when 54 midembryonic, neuronal enhancers were examined, very few were accessible in advance of activity. In our eRNA-based study, we identified several thousand developmental enhancers that were accessible at the 128-cell stage (~9 hpf) but that showed little or no detectable eRNA expression at this time (0–12 hpf). Hundreds of these enhancers showed maximal eRNA expression many hours later in development. Our findings indicate that the accessibility of sea urchin enhancers can precede their activity, and suggest that many of these regulatory elements undergo a progressive maturation during embryogenesis. It should be borne in mind that because our analysis is not lineage-restricted, some open eRNA peaks might mark ubiquitous, repressive complexes rather than lineage-specific complexes that mediate transcriptional activation.

The observations reported here further validate the utility of eRNA profiling for enhancer discovery and analysis. Our finding that the temporal expression profiles of eRNAs tended to resemble those of nearby genes, that is, genes that are the most likely direct targets of the cognate enhancers, supports the view that the presence of eRNA transcripts is a reliable indicator of enhancer function (Arnold et al. 2020; Sartorelli and Lauberth 2020) and further suggests that levels of eRNA expression and enhancer activity are correlated. In addition, we found that of the *S. purpuratus* CRMs that have been experimentally validated in previous studies, usually by *in vivo* reporter gene assays, the majority (60%) were associated with eRNA expression. Not all CRMs with activity in *in vivo* reporter assays expressed detectable levels of eRNAs, however, as has also been observed in human cells (Andersson et al. 2014) and *Drosophila* embryos (Mikhaylichenko et al. 2018). Two factors probably contribute to this effect. First, as noted above, the steady-state levels of many eRNAs are extremely low and, even at high sequencing depth, undoubtedly some fall below the limit of detection. Second, the activity of CRMs when tested in isolation sometimes exceeds their natural regulatory function (Peter and Davidson 2015), and this could explain an absence of detectable transcription at such sites. Our second approach, which involved testing the activity of eRNA peaks in transgenic embryos, showed that 50% of the elements examined (10/20) lacked detectable activity. It should be noted, however, that the reporter assay we used does not detect activity from CRMs that modulate the level or timing of gene expression, that require interactions with other CRMs, or that execute repressive functions. It is therefore to be expected that some bona fide regulatory elements will not show activity by this assay. Of those elements that were active in the reporter assay, 100% (10/10) drove GFP expression in tissue-specific

patterns that resembled those of endogenous genes near the enhancers.

Enhancers are traditionally characterized by their insensitivity to orientation and position relative to the core promoters they regulate. Other regulatory functions are executed by the proximal promoter, a region from a few hundred to several kilobases in size flanking the TSS that contains a high density of TF binding sites and *cis*-regulatory control elements (Hurst et al. 2014; Humi-niecki and Horbańczuk 2020). Proximal promoter elements are typically more sensitive to orientation and position than distal enhancers and are thought to influence transcription at the core promoter either directly or by tethering distal enhancers (Calhoun et al. 2002). Although the association of distal enhancers with eRNA expression is well documented, less attention has been given to the transcription of proximal regulatory elements. In our analysis, a substantial fraction of eRNAs (8.5%) mapped to proximal promoters (defined as a 2-kb region upstream of the TSS). We have not systematically tested the influence of orientation or position on the regulatory function of these transcribed, proximal promoter elements. Theoretically, eRNAs transcribed from such elements could have any of the molecular functions that have been ascribed to distal eRNAs (Hou and Kraus 2021), and the possible biological significance of the elevated levels of eRNA expression associated with proximal promoter eRNAs remains to be explored.

Models of developmental GRNs are particularly well developed in sea urchins (Davidson 2009; Peter and Davidson 2015; Martik et al. 2016; Lowe et al. 2017; Peter 2019). The construction of such models often begins with perturbations of regulatory (TF-encoding) genes and gene expression profiling, which together allow the deduction of functional interactions among genes and the reconstruction of network circuitry. To determine whether epistatic gene interactions revealed by these methods are direct or indirect, however, it is necessary to identify immediate inputs into the regulatory modules of genes in the network. Thus, a central component of GRN analysis is the identification and experimental dissection of enhancers. The present work provides an important resource for GRN biology by identifying developmental enhancers throughout the sea urchin genome and mapping the activity of those enhancers across early development. Moreover, we found that eRNA-marked enhancers located near lineage-specific genes contained sequence signatures of known regulatory inputs in those cell types. This suggests that experimental dissection of additional, predicted transcriptional inputs is likely to be informative and, when combined with temporal patterns of enhancer activity deduced from eRNA expression profiles, will aid in the construction of dynamic GRN models.

## Methods

### Embryo culture and RNA isolation

Adult *S. purpuratus* were obtained from Pat Leahy (Caltech). Gametes were collected by intra-coelomic injection of 0.5 M KCl. Fertilization and embryo culture were performed according to the method of Adams et al. (2019). Embryos were cultured at 15°C in 4-liter plastic beakers fitted with battery-powered stirrers. Embryos were harvested at 0, 6, 12, 18, 24, 30, 36, 42, and 48 hpf. To provide sufficient material for the analysis, two fertilizations were performed at different times on the same day using eggs from two different females and sperm from the same male. Culture 1 was used for seven time points, and culture 2 was used for two time points (18 hpf and 42 hpf). Cultures were not pooled

at any time point. For the 0-hpf time point, embryos were collected immediately after fertilization. Total RNA (75–100 µg per sample) was isolated using the RNeasy Plus Mini Kit (Qiagen 74134), and the quality of the preparations was confirmed using an RNA TapeStation (Agilent). RNA integrity numbers (RINs) for all samples were 9.8–10.0. RNA samples were stored at –80°C.

### CAGE-seq and eRNA identification

CAGE library preparation, sequencing, mapping, and eRNA identification were performed by DNAFORM. Thirty to 60 µg total RNA was used for the generation of each CAGE library (nine total). In brief, CAGE library construction involved the generation of single-stranded cDNA by reverse transcription using random primers, biotinylation of the 5′ RNA CAP structure followed by affinity purification of 5′-CAP-associated cDNA, ligation of bar-coded linkers, second-strand synthesis, and Illumina HiSeq-based sequencing (Murata et al. 2014). Sequence reads (more than 100 million reads per sample) were subjected to quality control (FastQC), filtered to remove small numbers of rRNA reads, and mapped to the *S. purpuratus* genome (v. 3.1) first with BWA (Li and Durbin 2010) and then using HISAT2 (Kim et al. 2019) to align reads with BWA MAPQ < 20. Although a newer genome assembly (v.5.0) recently became publicly available, we opted to perform our analysis with the v3.1 assembly in order to associate eRNA peaks with several previously published genomic data sets (ATAC-seq peaks, DNase-seq peaks, regions of chromatin differentially open in PMCs, and Alx1 ChIP-seq peaks), all of which were based on the v3.1 assembly. The v3.1 assembly is of very high quality in gene-rich regions, and the gene models themselves are mostly identical in the two assemblies; the principal improvement in the v5.0 assembly is that smaller scaffolds have been linked into much larger ones.

A total of 60–80 million uniquely mapped reads were obtained per sample and used for all subsequent analysis. De novo eRNA peak-calling was performed as described by Hirabayashi et al. (2019). Briefly, bidirectional enhancers were identified using the FANTOM5 pipeline (The FANTOM Consortium and the RIKEN PMI and CLST [DGT] 2014). TSSs were identified according to [http://fantom.gsc.riken.jp/5/ssstar/Protocols:HeliScopeCAGE\\_read\\_alignment](http://fantom.gsc.riken.jp/5/ssstar/Protocols:HeliScopeCAGE_read_alignment), and clustered using the decomposition-based peak identification (DPI) software ([https://github.com/hkawaji/dpi1/blob/master/identify\\_tss\\_peaks.sh](https://github.com/hkawaji/dpi1/blob/master/identify_tss_peaks.sh)). DPI was used with default parameters but without the decomposition parameter. Peaks with at least two supporting CAGE tags were retained and used as input to call bidirectional enhancers using a program available at GitHub ([https://github.com/anderssonrobin/enhancers/blob/master/scripts/bidir\\_enhancers](https://github.com/anderssonrobin/enhancers/blob/master/scripts/bidir_enhancers)).

### Additional bioinformatic analysis

Manipulation of peak and gene BED files was performed using BEDTools (v2.19.1) (Quinlan and Hall 2010). Distances between eRNA peaks and genes were determined using the “closest” and “intersect” utilities in BEDTools. GO term enrichment analyses were performed using the PANTHER webtool (Mi et al. 2019) and GO annotations available at <http://geneontology.org> (The Gene Ontology Consortium 2000,2021). To visualize peak positions relative to transcripts, 17,291 eRNA peaks located in scaffolds containing 8372 annotated genes were analyzed using ChIPseeker (1.22.1) (Yu et al. 2015). As some annotations overlap, ChIPseeker adopts the following priority in genomic annotation: Promoter > Exon > First Intron > Other Intron(s) > Intergenic. TF binding motifs for eRNA clusters were identified using the HOMER de novo motif discovery tool (Heinz et al. 2010). Enriched motifs were filtered for those with a *P*-value of  $< 1 \times 10^{-12}$ . HOMER randomly

selects background sequences from the genome and normalizes the GC content and sequencing biases to resemble the nucleotide distribution observed in the target sequences. Motifs are ranked based on *P*-value. The best-known JASPARHOMER motif that closely matches with the de novo motif is displayed in the “best match” column. Comparison of the de novo and best match motif probability matrices are reported in the “Mmatch score” column, where a score of 1.0 signifies complete similarity. The final column corresponds to sea urchin-specific TFs that are known or predicted to have similar motifs as the best-matched motifs. Additional information on the statistical analyses of HOMER can be found at <http://homer.ucsd.edu/homer/motif/>.

To determine whether the expression profiles of eRNAs were similar to those of nearby genes, *Z*-scores of 13,056 eRNAs within 20 kb of genes (a set of genes that numbered 5837) were obtained by first log-transforming the temporal expression data (TPMs) and then normalizing the transformed values to their respective means ( $\mu$ ) and standard deviations ( $\sigma$ ). *Z*-Scores were calculated using the following equation:  $Z = [\log_2(\text{TPM} + 1) - \mu] / \sigma$ . The *Z*-scores for both eRNAs and genes were pooled and clustered. We then calculated the fold enrichment within each expression cluster of eRNA peaks that were within 20 kb of genes in that same cluster, relative to all 13,056 eRNAs. Statistical significance of enrichments was assessed using a hypergeometric test followed by Benjamini–Hochberg correction to obtain the false-discovery rate. The hypergeometric test is used to determine whether subpopulations are over- or underrepresented in a sample. Enrichment of eRNA peaks that are hyperaccessible exclusively at the 128-cell stage (1018 peaks) or 24 hpf (1820 peaks) for different eRNA (Fig. 3) or eRNA/gene (Supplemental Fig. S6) clusters was examined using Fisher’s exact test. We examined lineage-specific enhancers by identifying eRNAs near genes differentially expressed by PMCs (420 genes) (Rafiq et al. 2014), pigment cells (2425 genes that are greater than fivefold enriched compared with controls) (Barsi et al. 2015), and ciliated band cells (3036 genes that are greater than twofold enriched compared with controls) (Barsi et al. 2015). TF binding motifs for eRNAs located within 20 kb of genes expressed specifically by PMCs (734 eRNA peaks), pigment cells (3274 eRNA peaks), and ciliated band cells (2941 eRNA peaks) were identified using HOMER, and enriched motifs were filtered for those with a *P*-value of  $< 1 \times 10^{-11}$ . To determine whether the 734 eRNA peaks located within 20 kb of 420 genes differentially expressed by PMCs were more likely to overlap with Alx1 ChIP-seq (Khor et al. 2019) and with ATAC-seq and DNase-seq differentially accessible regions (Shashikant et al. 2018), we compared the proportion of PMC eRNAs to the proportion of the total 18,078 eRNA peaks that overlapped with those data sets. Statistical significance of enrichments was assessed using a hypergeometric test.

### Transgenic reporter assays

GFP reporter gene constructs were generated by cloning individual, putative enhancers (Supplemental Table S7) into the EpGFPII plasmid, which contains the basal promoter of the *Sp-endo16* gene (Cameron et al. 2004). Putative enhancers were synthesized as gBlocks (Integrated DNA Technologies) that corresponded to eRNA peaks and were flanked by restriction sites for insertion upstream of the basal *Sp-endo16* promoter. Reporter plasmids were linearized and injected into *S. purpuratus* eggs following established protocols (Arnone et al. 2004). *S. purpuratus* eggs were fertilized in the presence of 0.1% (wt/vol) para-aminobenzoic acid to prevent hardening of the fertilization envelope. For each construct, 20  $\mu$ L of injection solution was prepared that contained 100 ng of the reporter plasmid, 500 ng HindIII-digested genomic *S. purpuratus* DNA, 0.12 M KCl, 20% glycerol, and 0.25%

Dextran, Texas Red. GFP expression was assayed by fluorescence microscopy at the late gastrula stage (48 hpf). Embryos were scored to determine total number of injected embryos (using Dextran, Texas Red as a marker), the number of embryos showing GFP expression, and the number of embryos with cell type-specific GFP expression. It should be noted that plasmids injected into fertilized sea urchin eggs rapidly form a concatemer that is randomly inherited by only one or a few cells during cleavage. This mosaic inheritance of transgenes during early development means that in some embryos the exogenous DNA is not incorporated into the lineage of cells in which the enhancer is normally active, and therefore, the reporter will not be expressed. For this reason, analysis of the expression patterns of reporter constructs requires the examination of relatively large populations of embryos.

### Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE169227. The table of *S. purpuratus* eRNAs (Supplemental Table S4) is accessible via Echinobase ([https://wiki.echinobase.org/echinowiki/index.php/ERNA\\_Table](https://wiki.echinobase.org/echinowiki/index.php/ERNA_Table)). The positions of *S. purpuratus* eRNAs on the v5.0 genomic scaffolds (Supplemental Data S1) can be viewed in JBrowse ([https://www.echinobase.org/common/displayJBrowse.do?data=data/sp5\\_0](https://www.echinobase.org/common/displayJBrowse.do?data=data/sp5_0)). “Available Tracks” are shown on the left. From the “Other” options, select “eRNAs–*Strongylocentrotus purpuratus* 3.1 liftover”.

### Competing interest statement

The authors declare no competing interests.

### Acknowledgments

We thank Dr. Yasuhiro Murakawa and Dr. Akiko Oguchi for the identification of eRNAs from CAGE-seq data. This work was supported by grants from the National Institutes of Health (R24-OD023046) and the National Science Foundation (IOS-2004952), both to C.A.E.

### References

- Adams NL, Heyland A, Rice LL, Foltz KR. 2019. Procuring animals and culturing of eggs and embryos. *Methods Cell Biol* **150**: 3–46. doi:10.1016/bs.mcb.2018.11.006
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* **507**: 455–461. doi:10.1038/nature12787
- Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drabløs F, Lennartsson A, Rönnerblad M, Hrydziszko O, Vitezic M, et al. 2015. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**: 1010–1014. doi:10.1126/science.1259418
- Arnold PR, Wells AD, Li XC. 2020. Diversity and emerging roles of enhancer RNA in regulation of gene expression and cell fate. *Front Cell Dev Biol* **7**: 377. doi:10.3389/fcell.2019.00377
- Arnone MI, Dmochowski IJ, Gache C. 2004. Using reporter genes to study *cis*-regulatory elements. *Methods Cell Biol* **74**: 621–652. doi:10.1016/S0091-679X(04)74025-X
- Baillie JK, Arner E, Daub C, De Hoon M, Itoh M, Kawaji H, Lassmann T, Carninci P, Forrest ARR, Hayashizaki Y, et al. 2017. Analysis of the human monocyte-derived macrophage transcriptome and response to lipopolysaccharide provides new insights into genetic aetiology of inflammatory bowel disease. *PLoS Genet* **13**: e1006641. doi:10.1371/journal.pgen.1006641

- Barsi JC, Tu Q, Calestani C, Davidson EH. 2015. Genome-wide assessment of differential effector gene use in embryogenesis. *Development* **142**: 3892–3901. doi:10.1242/dev.127746
- Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EE. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* **44**: 148–156. doi:10.1038/ng.1064
- Calestani C, Rast JP, Davidson EH. 2003. Isolation of pigment cell specific genes in the sea urchin embryo by differential microarray screening. *Development* **130**: 4587–4596. doi:10.1242/dev.00647
- Calhoun VC, Stathopoulos A, Levine M. 2002. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* antennapedia complex. *Proc Natl Acad Sci* **99**: 9243–9247. doi:10.1073/pnas.142291299
- Calo E, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**: 825–837. doi:10.1016/j.molcel.2013.01.038
- Cameron RA, Oliveri P, Wyllie J, Davidson EH. 2004. *cis*-Regulatory activity of randomly chosen genomic fragments from the sea urchin. *Gene Expr Patterns* **4**: 205–213. doi:10.1016/j.modgep.2003.08.007
- Catarino RR, Stark A. 2018. Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev* **32**: 202–223. doi:10.1101/gad.310367.117
- Cauchy P, Maqbool MA, Zacarias-Cabeza J, Vanhille L, Koch F, Fenouil R, Gut M, Gut I, Santana MA, Griffon A, et al. 2016. Dynamic recruitment of Ets1 to both nucleosome-occupied and -depleted enhancer regions mediates a transcriptional program switch during early T-cell differentiation. *Nucleic Acids Res* **44**: 3567–3585. doi:10.1093/nar/gkv1475
- Cui M, Siriwon N, Li E, Davidson EH, Peter IS. 2014. Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo. *Proc Natl Acad Sci* **111**: E5029–E5038. doi:10.1073/pnas.1419141111
- Davidson EH. 1986. *Gene activity in early development*, 3rd ed. Academic Press, New York.
- Davidson EH. 2009. Network design principles from the sea urchin embryo. *Curr Opin Genet Dev* **19**: 535–540. doi:10.1016/j.gde.2009.10.007
- Davidson EH, Levine MS. 2008. Properties of developmental gene regulatory networks. *Proc Natl Acad Sci* **105**: 20063–20066. doi:10.1073/pnas.0806007105
- Davidson EH, Cameron RA, Ransick A. 1998. Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms. *Development* **125**: 3269–3290. doi:10.1242/dev.125.17.3269
- Denisenko E, Guler R, Mhlanga MM, Suzuki H, Brombacher F, Schmeier S. 2017. Genome-wide profiling of transcribed enhancers during macrophage activation. *Epigenetics Chromatin* **10**: 50. doi:10.1186/s13072-017-0158-9
- De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei C-L, Natoli G, et al. 2010. A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol* **8**: e1000384. doi:10.1371/journal.pbio.1000384
- Ettensohn CA, Illies MR, Oliveri P, De Jong DL. 2003. Alx1, a member of the Cart1/Alx3/Alx4 subfamily of paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo. *Development* **130**: 2917–2928. doi:10.1242/dev.00511
- The FANTOM Consortium and the RIKEN PMI and CLST (DGT). 2014. A promoter-level mammalian expression atlas. *Nature* **507**: 462–470. doi:10.1038/nature13182
- Farley EK, Olson KM, Levine MS. 2015. Regulatory principles governing tissue specificity of developmental enhancers. *Cold Spring Harb Symp Quant Biol* **80**: 27–32. doi:10.1101/sqb.2015.80.027227
- Furlong EEM, Levine M. 2018. Developmental enhancers and chromosome topology. *Science* **361**: 1341–1345. doi:10.1126/science.aau0320
- The Gene Ontology Consortium. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* **25**: 25–29. doi:10.1038/75556
- The Gene Ontology Consortium. 2021. The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res* **49**: D325–D334. doi:10.1093/nar/gkaa1113
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576–589. doi:10.1016/j.molcel.2010.05.004
- Henriques T, Scruggs BS, Inouye MO, Muse GW, Williams LH, Burkholder AB, Lavender CA, Fargo DC, Adelman K. 2018. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev* **32**: 26–41. doi:10.1101/gad.309351.117
- Hirabayashi S, Bhagat S, Matsuki Y, Takegami Y, Uehata T, Kanemaru A, Itoh M, Shirakawa K, Takaori-Kondo A, Takeuchi O, et al. 2019. NET-CAGE characterizes the dynamics and topology of human transcribed *cis*-regulatory elements. *Nat Genet* **51**: 1369–1379. doi:10.1038/s41588-019-0485-9
- Hou TY, Kraus WL. 2021. Spirits in the material world: enhancer RNAs in transcriptional regulation. *Trends Biochem Sci* **46**: 138–153. doi:10.1016/j.tibs.2020.08.007
- Huminięcki Ł, Horbańczuk J. 2020. Can we predict gene expression by understanding proximal promoter architecture? *Trends Biotechnol* **38**: 463. doi:10.1016/j.tibtech.2019.12.003
- Hurst LD, Sachenkova O, Daub C, Forrest AR, Huminięcki L, FANTOM consortium. 2014. A simple metric of promoter architecture robustly predicts expression breadth of human genes suggesting that most transcription factors are positive regulators. *Genome Biol* **15**: 413. doi:10.1186/s13059-014-0413-3
- Khor JM, Ettensohn CA. 2020. Transcription factors of the Alx family: evolutionarily conserved regulators of deuterostome skeletogenesis. *Front Genet* **11**: 569314. doi:10.3389/fgene.2020.569314
- Khor JM, Guerrero-Santoro J, Ettensohn CA. 2019. Genome-wide identification of binding sites and gene targets of Alx1, a pivotal regulator of echinoderm skeletogenesis. *Development* **46**: dev180653. doi:10.1242/dev.180653
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182–187. doi:10.1038/nature09033
- Kim YW, Lee S, Yun J, Kim A. 2015. Chromatin looping and eRNA transcription precede the transcriptional activation of a gene in the  $\beta$ -globin locus. *Biosci Rep* **35**: e00179. doi:10.1042/BSR20140126
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**: 907–915. doi:10.1038/s41587-019-0201-4
- Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, Spicuglia S, de la Chapelle AL, Heidemann M, Hintermair C, et al. 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* **18**: 956–963. doi:10.1038/nsmb.2085
- Kouno T, Moody J, Kwon AT, Shibayama Y, Kato S, Huang Y, Böttcher M, Motakis E, Mendez M, Severin J, et al. 2019. C1 CAGE detects transcription start sites and enhancer activity at single-cell resolution. *Nat Commun* **10**: 360. doi:10.1038/s41467-018-08126-5
- Kurokawa D, Kitajima T, Mitsunaga-Nakatsubo K, Amemiya S, Shimada H, Akasaka K. 1999. HpEts, an ets-related transcription factor implicated in primary mesenchyme cell differentiation in the sea urchin embryo. *Mech Dev* **80**: 41–52. doi:10.1016/S0925-4773(98)00192-0
- Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, Dickson BJ, Stark A. 2014. Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature* **512**: 91–95. doi:10.1038/nature13395
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**: 589–595. doi:10.1093/bioinformatics/btp698
- Long HK, Prescott SL, Wysocka J. 2016. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**: 1170–1187. doi:10.1016/j.cell.2016.09.018
- Lowe EK, Cuomo C, Arnone MI. 2017. Omics approaches to study gene regulatory networks for development in echinoderms. *Brief Funct Genomics* **16**: 299–308. doi:10.1093/bfpg/elx012
- Martik ML, Lyons DC, McClay DR. 2016. Developmental gene regulatory networks in sea urchins and what we can learn from them. *F1000Res* **5**(F1000 Faculty Rev-203): 203. doi:10.12688/f1000research.7381.1
- Martínez-Bartolomé M, Range RC. 2019. A biphasic role of non-canonical Wnt16 signaling during early anterior-posterior patterning and morphogenesis of the sea urchin embryo. *Development* **146**: dev168799. doi:10.1242/dev.168799
- Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* **47**: D419–D426. doi:10.1093/nar/gky1038
- Mikhaylichenko O, Bondarenko V, Harnett D, Schor IE, Males M, Viales RR, Furlong EEM. 2018. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev* **32**: 42–57. doi:10.1101/gad.308619.117
- Morioka MS, Kawaji H, Nishiyori-Sueki H, Murata M, Kojima-Ishiyama M, Carninci P, Itoh M. 2020. Cap analysis of gene expression (CAGE): a quantitative and genome-wide assay of transcription start sites. *Methods Mol Biol* **2120**: 277–301. doi:10.1007/978-1-0716-0327-7\_20
- Murakawa Y, Yoshihara M, Kawaji H, Nishikawa M, Zayed H, Suzuki H, FANTOM Consortium, Hayashizaki Y. 2016. Enhanced identification of transcriptional enhancers provides mechanistic insights into diseases. *Trends Genet* **32**: 76–88. doi:10.1016/j.tig.2015.11.004

- Murata M, Nishiyori-Sueki H, Kojima-Ishiyama M, Carninci P, Hayashizaki Y, Itoh M. 2014. Detecting expressed genes using CAGE. *Methods Mol Biol* **1164**: 67–85. doi:10.1007/978-1-4939-0805-9\_7
- Nègre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R, et al. 2011. A cis-regulatory map of the *Drosophila* genome. *Nature* **471**: 527–531. doi:10.1038/nature09990
- Ong CT, Corces VG. 2012. Enhancers: emerging roles in cell fate specification. *EMBO Rep* **13**: 423–430. doi:10.1038/embor.2012.52
- Otim O, Amore G, Minokawa T, McClay DR, Davidson EH. 2004. *Sphnf6*, a transcription factor that executes multiple functions in sea urchin embryogenesis. *Dev Biol* **273**: 226–243. doi:10.1016/j.ydbio.2004.05.033
- Peter IS. 2019. Methods for the experimental and computational analysis of gene regulatory networks in sea urchins. *Methods Cell Biol* **151**: 89–113. doi:10.1016/bs.mcb.2018.10.003
- Peter IS, Davidson EH. 2015. *Genomic control process, development and evolution*. Academic Press, Oxford.
- Peter IS, Davidson EH. 2016. Implications of developmental gene regulatory networks inside and outside developmental biology. *Curr Top Dev Biol* **117**: 237–251. doi:10.1016/bs.ctdb.2015.12.014
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- Rafiq K, Shashikant T, McManus CJ, Etensohn CA. 2014. Genome-wide analysis of the skeletogenic gene regulatory network of sea urchins. *Development* **141**: 950–961. doi:10.1242/dev.105585
- Ransick A, Davidson EH. 2006. cis-Regulatory processing of Notch signaling input to the sea urchin glial cells missing gene during mesoderm specification. *Dev Biol* **297**: 587–602. doi:10.1016/j.ydbio.2006.05.037
- Reddington JP, Garfield DA, Sigalova OM, Karabacak Calviello A, Marco-Ferreres R, Girardot C, Viales RR, Degner JF, Ohler U, Furlong EEM. 2020. Lineage-resolved enhancer and promoter usage during a time course of embryogenesis. *Dev Cell* **55**: 648–664.e9. doi:10.1016/j.devcel.2020.10.009
- Russo R, Pinsino A, Costa C, Bonaventura R, Matranga V, Zito F. 2014. The newly characterized *Pl-jun* is specifically expressed in skeletogenic cells of the *Paracentrotus lividus* sea urchin embryo. *FEBS J* **281**: 3828–3843. doi:10.1111/febs.12911
- Sakaguchi Y, Nishikawa K, Seno S, Matsuda H, Takayanagi H, Ishii M. 2018. Roles of enhancer RNAs in RANKL-induced osteoclast differentiation identified by genome-wide Cap-analysis of gene expression using CRISPR/Cas9. *Sci Rep* **8**: 7504. doi:10.1038/s41598-018-25748-3
- Sartorelli V, Lauberth SM. 2020. Enhancer RNAs are an important regulatory layer of the epigenome. *Nat Struct Mol Biol* **27**: 521–528. doi:10.1038/s41594-020-0446-0
- Shashikant T, Khor JM, Etensohn CA. 2018. Global analysis of primary mesenchyme cell cis-regulatory modules by chromatin accessibility profiling. *BMC Genomics* **19**: 206. doi:10.1186/s12864-018-4542-z
- Sun Z, Etensohn CA. 2014. Signal-dependent regulation of the sea urchin skeletogenic gene regulatory network. *Gene Expr Patterns* **16**: 93–103. doi:10.1016/j.gep.2014.10.002
- Suryamohan K, Halfon MS. 2015. Identifying transcriptional cis-regulatory modules in animal genomes. *Wiley Interdiscip Rev Dev Biol* **4**: 59–84. doi:10.1002/wdev.168
- Thomas S, Li XY, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE, et al. 2011. Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol* **12**: R43. doi:10.1186/gb-2011-12-5-r43
- Tu Q, Cameron RA, Worley KC, Gibbs RA, Davidson EH. 2012. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res* **22**: 2079–2087. doi:10.1101/gr.139170.112
- Tu Q, Cameron RA, Davidson EH. 2014. Quantitative developmental transcriptomes of the sea urchin *Strongylocentrotus purpuratus*. *Dev Biol* **385**: 160–167. doi:10.1016/j.ydbio.2013.11.019
- Tyssowski KM, DeStefino NR, Cho J-H, Dunn CJ, Poston RG, Carty CE, Jones RD, Chang SM, Romeo P, Wurzelmann MK, et al. 2018. Different neuronal activity patterns induce different gene expression programs. *Neuron* **98**: 530–546.e11. doi:10.1016/j.neuron.2018.04.001
- Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, et al. 2011. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**: 390–394. doi:10.1038/nature10006
- Wessel GM, Kiyomoto M, Shen TL, Yajima M. 2020. Genetic manipulation of the pigment pathway in a sea urchin reveals distinct lineage commitment prior to metamorphosis in the bilateral to radial body plan transition. *Sci Rep* **10**: 1973. doi:10.1038/s41598-020-58584-5
- Yu G, Wang LG, He QY. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**: 2382–2383. doi:10.1093/bioinformatics/btv145

Received April 22, 2021; accepted in revised form July 28, 2021.