Transcriptional profiles of early stage red sea urchins (*Mesocentrotus franciscanus*) reveal differential regulation of gene expression across development

Juliet M. Wong [a,*], Juan D. Gaitán-Espitia [b], and Gretchen E. Hofmann [a]

[a] Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, CA 93106, USA

[b] The Swire Institute of Marine Science, School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong, SAR, People's Republic of China

[*] Corresponding author: julietmwong@ucsb.edu (J. Wong)

Email addresses: julietmwong@ucsb.edu (J. Wong), jdgaitan@hku.hk (J. Gaitán-Espitia), hofmann@ucsb.edu (G. Hofmann)

Abstract

The red sea urchin, *Mesocentrotus franciscanus*, is an ecologically important kelp forest species that also serves as a valuable fisheries resource. In this study, we have assembled and annotated a developmental transcriptome for *M. franciscanus* that represents eggs and six stages of early development (8- to 16-cell, morula, hatched blastula, early gastrula, prism and early pluteus). Characterization of the transcriptome revealed distinct patterns of gene expression that corresponded to major developmental and morphological processes. In addition, the period during which maternally-controlled transcription was terminated and the zygotic genome was activated, the maternal-to-zygotic transition (MZT), was found to begin during early cleavage and persist through the hatched blastula stage, an observation which is similar to the timing of the MZT in other sea urchin species. The presented developmental transcriptome will serve as a useful resource for investigating, in both an ecological and fisheries context, how the early developmental stages of this species respond to environmental stressors.

Keywords [4-6 words]

Red sea urchin, *Mesocentrotus franciscanus*, RNA-seq, *De novo* assembly, Early development

1. Introduction

The red sea urchin *Mesocentrotus franciscanus* (A. Agassiz, 1863) is found along the West Coast of North America, ranging from Baja California, Mexico to Kodiak, Alaska, USA (Ebert, et al. 1999). *M. franciscanus* is harvested for its gonads (i.e., roe), and in recent decades has suffered overfishing and exploitation as a high-demand wild fishery species (Andrew, et al. 2002, Keesing and Hall 1998). In terms of worth, the export value of roe in the United States was

estimated to be approximately $28.7 million in 2011 (Rogers-Bennett 2013). However, the price of red sea urchins harvested in the states of California, Oregon, and Washington has risen in recent years, increasing from $0.70 USD per pound in 2011 to $1.73 USD per pound in 2018 according to data reported by the Pacific Fisheries Information Network (PacFIN) (www.pacfin.psmfc.org; accessed 5 April 2019).

As a valuable wild fishery, red sea urchins are found in temperate rocky reefs in nearshore coastal regions. As benthic marine invertebrates in coastal marine ecosystems, they are threatened by ocean change (Sato, et al. 2018), including future ocean acidification in regions dominated by episodic upwelling (Chan, et al. 2017) and from marine heat waves (Gentemann, et al. 2017). For example, between 2013 to 2016, anomalous warming events associated with "the Blob" led to dramatically increased sea surface temperatures in the northeast Pacific Ocean (Bond, et al. 2015, Gentemann, et al. 2017, Hu, et al. 2017). While a direct cause-effect relationship between these warming events and red urchin harvests has not been formally established, in California alone, the weight of harvested *M. franciscanus* decreased from 6,320 to 2,669 metric tons per year between 2013 and 2016 (www.pacfin.psmfc.org; accessed 5 April 2019). Importantly, marine heat waves like "the Blob" are predicted to increase in frequency and intensity in the future (Oliver, et al. 2018); these events will likely have an impact on marine systems, including organism physiology, species abundances, and population biogeographic distributions (Frölicher and Laufkötter 2018, Leung, et al. 2017, Sanford, et al. 2019). Ecologically, *M. franciscanus* acts as an important ecosystem engineer by controlling algae populations, particularly in kelp forest ecosystems, and are capable of transforming algal communities into urchin barrens (Leighton, et al. 1966, Rogers-Bennett 2007). Given the high economic and ecological importance of *M. franciscanus*, as well as its potential vulnerability to

environmental change, genomic resources are very useful for studying and monitoring this species.

The introduction of next-generation sequencing and increasing affordability of associated technologies has expanded the molecular resources and knowledge available for non-model species, particularly those of high value to fisheries and aquaculture. Annotated and assembled *de novo* transcriptomes have been published across a variety of valuable fisheries and aquaculture species, including mollusks (Coppe, et al. 2012, De Wit and Palumbi 2012, Tian, et al. 2018, Zhao, et al. 2012), crustaceans (Ghaffari, et al. 2014, Lv, et al. 2014, Souza, et al. 2018), echinoderms (Gaitán-Espitia, et al. 2016, Gillard, et al. 2014, Jo, et al. 2016), and fishes (Carruthers, et al. 2018, Ji, et al. 2012, Liao, et al. 2013). These transcriptomes are useful for investigating parameters important to fisheries health and management, such as population dynamics, evolutionary processes, the effects of abiotic stress, disease susceptibility and resilience, and stock assessments (Valenzuela-Quiñonez 2016, Wenne, et al. 2007). For example, assembled transcriptomes have been used to investigate salinity stress in the Pacific oyster *Crassostrea gigas* (Zhao, et al. 2012), and viral infection in the Pacific whiteleg shrimp *Litopenaeus vannamei* (Chen, et al. 2013). Transcriptomic data have also been used to investigate patterns of gene flow and local adaptation in the red abalone *Haliotis rufescens* (De Wit and Palumbi 2012). Here, we used RNA sequencing (RNA-seq) to assemble and annotate a developmental transcriptome for the economically and ecologically important sea urchin, *M. franciscanus*, and assessed patterns of gene expression throughout early development. This tool could be used to assess the response of early stages to climate-changed related stressors.

Numerous studies have used transcriptomics to investigate how marine organisms respond to changes in their environment that are related to climate change, such as elevated

84 temperatures and lowered pH and oxygen concentrations (Ekblom and Galindo 2011, Franks and
85 Hoffmann 2012, Reusch and Wood 2007, Strader, et al. 2019). Describing the transcriptional
86 dynamics of *M. franciscanus* during its early development is particularly pertinent, as the early
87 stages of development are believed to be the most vulnerable times during the life history of
88 many marine organisms (Byrne 2011, Dupont and Thorndyke 2009, Gosselin and Qian 1997,
89 Kurihara 2008). A reduction in fitness at the embryological and larval stages leading to poor
90 recruitment could have devastating impacts on marine population dynamics. As rapid
91 environmental change continues, the early life stages may act as a bottleneck that dictates
92 whether a species will be successful in the future (Byrne 2012, Byrne and Przeslawski 2013,
93 Kurihara 2008).
94 Several recent studies have reported developmental transcriptomes for marine
95 invertebrates (Brekhman, et al. 2015, Gildor, et al. 2016, Heyland, et al. 2011, Lenz, et al. 2014,
96 Zeng, et al. 2011), and have described stage-specific expression of many transcription factors.
97 Such transcriptomes help to identify which genes are important for development as well as the
98 timing of expression of these genes. One group of processes that occur during development
99 involves the change of control from the maternal to zygotic genome, identified as the maternal-
100 to-zygotic transition (MZT), during which time there is a shift in expression from maternal to
101 zygotic transcripts (Shier 2007, Tadros and Lipshitz 2009). Understanding the timing of the
102 MZT is important for interpreting expression dynamics during early development. In general, the
103 generation and characterization of developmental transcriptomes provide useful genomic
104 resources that offer the opportunity to unveil mechanisms underlying developmental plasticity
105 and its role buffering different abiotic stressors across ontogeny.

106 Here, we present a developmental transcriptome for *M. franciscanus* that represents eggs
107 as well as embryos and larvae from six stages of development: 8- to 16-cell, morula (composed
108 of approximately 64 cells), hatched blastula, early gastrula, prism, and early pluteus. The
109 developmental transcriptome presented here will provide insight into the timing of the MZT and
110 will help identify genes and regulatory pathways that are important for successful development
111 at each stage. Overall, this is an important resource for transcriptomic analyses of this species,
112 including ecological or fisheries studies that use gene expression to assess the response of early
113 developmental stages to stress.
114
115 **2. Materials and methods**
116 *2.1. Animal collection and culturing*
117 Adult sea urchins were collected in May 2016 at Mohawk Reef, CA, USA (34° 23.606'
118 N, 199° 43.807' W) under California Scientific Collection permit SC-1223. Urchins were
119 immediately transported to the Marine Science Institute at the University of California, Santa
120 Barbara (UCSB) (Santa Barbara, CA), and maintained in flow-through seawater tanks for
121 approximately one week prior to spawning. Spawning was induced via intracoelomic injection of
122 0.53 M KCl. Egg samples (EG) were collected by gently transferring ~5,000 eggs into a 1.5 mL
123 microcentrifuge tube, quickly pelleting the sample by centrifugation, removing the excess
124 seawater, and flash freezing the sample using liquid nitrogen. All samples were stored at -80 °C.
125 Test fertilizations were performed to verify egg-sperm compatibility. The eggs from two females
126 were gently pooled together and were fertilized using sperm from a single male. To avoid
127 polyspermy, dilute sperm was slowly added to the eggs until at least 95% fertilization success

was reached. The newly fertilized embryos were then placed into each of three replicate culture vessels (total volume = 12 L) at a concentration of ~9 embryos per mL of seawater.

All *M. franciscanus* early developmental stage (EDS) cultures were raised in 0.35 µm filtered, UV-sterilized seawater (FSW). The EDS cultures were raised at ~15 °C and ~425 µatm pCO_2. These conditions were chosen to represent average, "normal" ambient abiotic conditions that populations of *M. fransciscanus* have been observed to experience *in situ* near Mohawk reef; these observations are made via sensor arrays deployed by the Santa Barbara Coastal LTER (Hofmann and Washburn 2015). During the EDS culturing, water temperature was controlled using a Delta Star® heat pump with a Nema 4x digital temperature controller (AquaLogic, San Diego, CA, USA), which maintained culturing temperatures at 15 °C. A flow-through CO_2-mixing system modified from Fangue, et al. (2010) was used to ensure stable carbonate chemistry conditions throughout development. The CO_2 system was used to establish a 5-gallon reservoir tank, in which water was treated to the target pCO_2 level prior to delivering the treated water to each culture vessel.

Each culture vessel was composed of two, nested 5-gallon buckets (12 L capacity). The inner bucket has a dozen holes 5.5 cm in diameter, each fitted with 64-micron mesh to prevent the loss of embryos or larvae while allowing for a flow-through of seawater. Seawater flow to each vessel was controlled using irrigation button drippers (DIG Corporation), which regulated the flow to a rate of 4 L/hr. Each vessel contained a 15 cm x 15 cm plastic paddle driven by a 12-V motor to allow for continuous, gentle mixing and to prevent early embryos from settling to the bottom of the bucket.

Embryos and larvae were sampled at six developmental stages: 8- to 16- cell (CL; ~4 hours post-fertilization (hpf)), morula (MA; ~7 hpf), hatched blastula (BL; ~16 hpf), early

151 gastrula (GA; ~29 hpf), prism (PR; ~44 hpf), and early pluteus (PL; ~64 hpf). While the
152 development of *M. franciscanus* in culture is generally synchronous, during early cellular
153 divisions, it is unlikely to collect a large batch of embryos that are exhibiting identical timing. As
154 such, the samples collected at the 8- to 16-cell stage (CL) were composed of a mixture of
155 embryos undergoing their third and fourth cleavage divisions. At the morula stage (MA), all
156 embryos were composed of approximately 64 or more cells. The blastula stage (BL) was
157 designated by the enzymatic digestion of the fertilization envelope and emergence of swimming
158 blastula. The early gastrula (GA) stage was designated by the formation of mesenchyme cells
159 and an archenteron extended to approximately one-half the body length. The prism stage (PR)
160 was identified by the formation of the pyramid-like prism shape, the archenteron becoming
161 tripartite, and the early development of skeletal rods. Lastly, the early pluteus state (PL) was
162 defined as having internal structures, including the mouth, esophagus, stomach and anus, as well
163 as anterolateral and postoral skeletal body rods and the early formation of feeding arms. Samples
164 from each of the three culture vessels were taken at each developmental stage. All samples were
165 preserved using the same methods for preserving the eggs.
166 Temperature, salinity, pH, and total alkalinity (TA) were recorded daily to monitor the
167 culturing conditions throughout development. Temperature was measured using a wire
168 thermocouple (Thermolyne PM 20700 / Series 1218), and salinity was measured using a
169 conductivity meter (YSI 3100). Daily pH measurements were conducted by following the
170 standard operating procedure (SOP) 6b (Dickson, et al. 2007), using a spectrophotometer (Bio
171 Spec-1601, Shimadzu) and *m*-cresol purple (Sigma-Aldrich) indicator dye. Water samples for
172 TA were poisoned with saturated 0.02% mercuric chloride. TA was estimated using SOP 3b
173 (Dickson, et al. 2007). Using the carbonic acid dissociation constants from Mehrbach, et al.

(1973) refit by Dickson and Millero (1987), parameters of pCO_2, Ω_{ara}, and Ω_{cal} were calculated using CO_2calc (Robbins, et al. 2010).

2.2. RNA extractions and sequencing

Total RNA was extracted using 500 µL of Trizol® reagent, following the manufacturer's instructions (Invitrogen). Briefly, each sample was homogenized in Trizol® reagent by passing the sample three times through decreasing sizes of needles (21-guage, 23-gauge, and then 25-gauge). A chloroform addition and centrifugation were used to isolate the RNA-containing upper aqueous phase. The RNA was precipitated in isopropyl alcohol, washed using ethanol, and resuspended in DEPC-treated water. RNA purity, quantity, and quality were verified using a NanoDrop® ND100, a Qubit® fluorometer, and a Tapestation 2200 system (Agilent Technologies).

Three libraries were generated from triplicate samples of eggs. For each developmental stage, one library was generated for each of the three replicate culture vessels. This resulted in a total of 21 libraries. Libraries were generated using high quality total RNA (RIN values > 9.1) using a TruSeq Stranded mRNA Library Preparation Kit (Illumina) following the manufacturer's instructions. The quantity and quality of each library was verified using a Qubit® fluorometer and a Tapestation 2200 system (Agilent). The libraries were submitted to the Genome Center at the University of California, Davis for sequencing on an Illumina HiSeq 4000 sequencer on two lanes with 150 base-pair (bp) paired-end reads.

2.3. De novo *transcriptome assembly*

196 Additional *M. franciscanus* raw sequence data from Gaitán-Espitia and Hofmann (2017)
197 were included with sequence data from our 21 libraries to generate the *de novo* transcriptome.
198 These data represented gastrula stage embryos (GenBank accession numbers SRS823202 and
199 SRS823216) and pluteus larvae (accession numbers SRS823218 and SRS82322) of Bioproject
200 PRJNA272924. Any potential adapter sequence contamination as well as any base pairs with
201 quality scores below 30 were removed from all raw sequence data using Trim Galore! (version
202 0.4.1) (Krueger 2015). Sequence quality was verified using FastQC (version 0.11.5) (Andrews
203 2010).
204 The transcriptome was assembled following a pipeline available from the National Center
205 for Genome Analysis Support (NCGAS) at Indiana University (https://github.com/NCGAS/de-
206 novo-transcriptome-assembly-pipeline). This workflow generates a combined *de novo* assembly
207 that uses multiple assemblers with multiple parameters. Prior to assembly, the data were
208 normalized using the in silico read normalization function in Trinity (version 2.6.6) (Grabherr, et
209 al. 2011). Multiple *de novo* assemblies were created using different assemblers with different
210 selections of kmer lengths: Trinity (version 2.6.6) (kmer = 25), SOAPdenovo-Trans (version
211 1.03) (Xie, et al. 2014) (kmers = 35, 45, 55, 65, 75, and 85), Velvet (version 1.2.10) (Zerbino and
212 Birney 2008) and Oases (version 0.2.09) (Schulz, et al. 2012) (kmers = 35, 45, 55, 65, 75, and
213 85), and Trans-ABySS (version 2.0.1) (Robertson, et al. 2010) (kmers = 35, 45, 55, 65, 75, and
214 85). These 19 assemblies were then combined using EvidentialGene (version 2013.07.27)
215 (Gilbert 2013), which removes perfect redundancy and fragments to reduce false transcripts
216 while predicting unique transcripts within the final assembly. Quast (version 5.0.0) (Gurevich, et
217 al. 2013) was used to generate basic quality metrics of the final assembly. BUSCO (version

218 3.0.2) (Simão, et al. 2015) was used to assess completeness of the final assembly using the
219 single-copy ortholog reference for metazoans.
220
221 *2.4 Gene prediction and functional annotation*
222 Gene models from the *de novo* assembled transcriptome were inferred and annotated
223 using the BLASTP (against the nr database), BLASTN (against the eukaryotic nt database) and
224 BLASTX (against the Uniprot database, Swiss-Prot and TrEMBL) algorithms with an e-value
225 cutoff of $1e^{-5}$. Annotated sequences were further searched for Gene Ontology (GO) terms using
226 Blast2GO software (version 5.2.5) according to the main categories of Gene Ontology (GO;
227 molecular functions, biological processes and cellular components) (Ashburner, et al. 2000).
228 Complementary annotations were done with the InterProScan v.5 software (Jones, et al. 2014).
229 Finally, the annotation results were further fine-tuned with the Annex and GO slim functions and
230 the enzyme code annotation tool of the Kyoto Encyclopedia of Genes and Genomes (KEGG)
231 (Kanehisa and Goto 2000) implemented in Blast2GO.
232
233 *2.5 Expression quantitation and differential expression analyses*
234 Trimmed sequence data from the 21 libraries were mapped onto the *de novo* reference
235 transcriptome and expression values were calculated using RSEM (version 1.3.0) (Li and Dewey
236 2011) and bowtie2 (version 2.3.2) (Langmead and Salzberg 2012). Using the LIMMA package
237 (Ritchie, et al. 2015) in R (version 3.4.4), the data were filtered to sequences that have more than
238 0.5 counts per million mapped reads across at least three of the 21 samples. A trimmed mean of
239 M-values (TMM) normalization method (Robinson and Oshlack 2010) was used to apply scale
240 normalization to the read counts. The data were voom-transformed using LIMMA to convert the

241 read counts to log-counts per million while accounting for sample-specific quality weights and
242 blocking design (i.e., technical replicates). The filtered, normalized and voom-transformed data
243 were used to perform a principal component analysis (PCA) using the prcomp function in R. An
244 unweighted pair group method with arithmetic mean (UPGMA) method using Euclidean
245 distances was implemented to perform hierarchical clustering on the principal components using
246 the HCPC function of the FactoMineR package (Le, et al. 2008) in R.
247 Using the WGCNA package (Langfelder and Horvath 2008) in R, a Weighted Gene Co-
248 Expression Network Analysis (WGCNA) was performed on the same filtered, normalized and
249 voom-transformed data to identify clusters of similarly expressed genes into modules, in which
250 each module contained a minimum of 30 genes. Modules with highly correlated eigengenes were
251 merged using a threshold of 0.27 (i.e., a height cut-off of 0.27 and a correlation of 0.73 for
252 merging). Eigengene expression was correlated with eggs and each developmental stage (i.e.,
253 EG, CL, MA, BL, GA, PR, and PL), and a heatmap was generated to visualize significant
254 correlations between each stage and module.
255 Functional enrichment analyses were performed on lists of genes within modules with
256 significant correlations to a developmental stage (r^2 correlation ≥ 0.50 and p-value ≤ 0.05).
257 Enrichment analyses were performed in Blast2GO (version 5.2.5) using a Fisher's Exact Test
258 with an FDR filter value of 0.05 to identify gene ontology (GO) terms within the GO categories:
259 biological process, molecular function, and cellular component.
260
261 *2.6. Survey of the maternal-to-zygotic transition*
262 To investigate the timing of the maternal-to-zygotic transition (MZT), two aspects of the
263 data were probed: 1) the loss of maternally-derived transcripts, and 2) the initiation of zygotic

transcription. To examine the loss of maternally-derived transcripts, putative genes related to the removal of maternal RNAs were identified. These included putative genes for the *microprocessor complex subunit DGCR8* (*dgcr8*), *endoribonuclease dicer* (*dicer*), *smaug* (*smaug1*), and *nonsense mediated mRNA decay* (*smg7*, *smg8*, and *smg9*) (Gildor, et al. 2016, Marlow 2010, Tadros and Lipshitz 2009) (Table 1). The expression levels of these transcripts were examined across the eggs and early development. Additionally, a heatmap was constructed using Euclidean distances to visualize the expression of maternal transcripts throughout development. Transcripts expressed in the eggs were considered to be maternal. The heatmap was constructed using the top 500 transcripts expressed in unfertilized eggs (EG).

Transcript ID	Name	Description
058424	*dgcr8*	microprocessor complex subunit DGCR8-like
103747	*dicer*	endoribonuclease Dicer
035070	*smaug*	protein Smaug homolog 1 isoform X1
089348	*smg7*	protein SMG7 isoform X2
052093	*smg8*	protein smg8
035665	*smg9*	protein SMG9-like
089934	*alx*	aristaless-like homeobox protein
082528	*bra*	transcription factor Brachyury
015822	*dri*	protein dead ringer homolog
036581	*gcm*	glial cells missing transcription factor
035045	*gsc*	homeobox protein goosecoid-like
047087	*hox11/13b*	transcription factor Hox11/13b
104463	*lefty2*	left-right determination factor 2-like
047266	*nodal*	nodal homolog 2-A-like
076942	*wnt8*	Wnt8

Table 1. Transcripts that play a functional role during the MZT.

To examine the initiation of zygotic transcription, transcripts associated with zygotic development were targeted. Putative genes for *aristaless-like homeobox* (*alx*), *brachyury* (*bra*), *dead ringer* (*dri*), *glial cells missing* (*gcm*), *goosecoid* (*gsc*), *homeobox 11/13b (hox11/13b)*, *left-*

278 *right determination factor 2* (*lefty2*), *nodal*, and *wnt8* were targeted within *M. franciscanus*
279 (Tadros and Lipshitz 2009, Wei, et al. 2006) (Table 1). The expression levels of these transcripts
280 were examined in eggs and as early development progressed. A heatmap was also constructed
281 using Euclidean distances to visualize the expression of transcripts that were not maternally
282 expressed (i.e., not expressed in unfertilized eggs (EG)). Transcripts with a negative expression
283 value of log2 counts per million reads (log2 CPM < 0) at the EG stage were selected, for a total
284 of 19,044 transcripts. Using these transcripts, a heatmap was constructed to visualize at what
285 stages throughout development these transcripts were expressed.
286
287 **3. Results and discussion**
288 *3.1 Embryological and larval development conditions*
289 The embryo and larval cultures developed normally with little to no mortality observed
290 throughout development. Temperature and seawater chemistry conditions were stable throughout
291 the ~64 hour culturing period. Across all replicate culture vessels, the temperature was 15.3 ± 0.1
292 °C, the salinity was 33.4 ± 0.04, the pH was 8.00 ± 0.03, the pCO_2 level was 438 ± 33.1 µatm,
293 and the TA was 2228.97 ± 3.19 µmol kg^{-1}.
294
295 *3.2 Summary statistics of the transcriptome assembly and annotation*
296 Sequencing of the 21 libraries yielded a total of 751,578,474 150-bp paired-end reads.
297 After trimming to remove any adapter contamination or low quality reads, an average of 35.4 ±
298 5.8 million reads remained per library. FastQC reports of the trimmed reads from all libraries
299 showed high sequence quality (scores >30) with limited adapter contamination or presence of
300 overrepresented sequences. The transcriptome generated by the NCGAS pipeline was 96.74

megabases (Mb) with 115,719 contigs, a N50 of 3,292 bp, and a GC content of 42.58% (Table 2). The BUSCO analysis that used metazoan as the single-copy ortholog reference showed high transcriptome completeness with a complete BUSCO score of 92.2% (Table 2). Therefore, this transcriptome should offer a suitable foundation for transcriptomic analyses of *M. franciscanus*.

Assembly statistic	Value
No. contigs	115,719
No. contigs > 1kb	19,511
Assembly size (Mb)	96.74
Mean contig length (bp)	836
Median contig length (bp)	341
Max contig length (bp)	132,566
GC content	42.58
N_{50} (bp)	3,292
L_{50} (bp)	6,526
BUSCO completeness (%)	92.2
BUSCO fragmented (%)	0.8
BUSCO missing (%)	7.0

Table 2. Statistics of *de novo* transcriptome assembly

The gene discovery and functional annotation analyses identified 35,632 contigs that blasted to known proteins in the public databases (Table S1). From these, 24,900 contigs were linked to GO classifications. Hypothetical or predicted proteins in these databases were excluded by discarding matches associated to “hypothetical”, “predicted”, “unknown” and “putative” categories. Over 95% of the annotated contigs hit against the genomes of the purple sea urchin, *Strongylocentrotus pupuratus*, followed by the sea star, *Acanthaster planci*, and the sea cucumber, *Apostichopus japonicus*. The functional annotation analysis retrieved 48,990 GO terms, with 23,053 linked to molecular function (mainly protein binding), 15,754 linked to biological process (mainly G protein-coupled receptor signaling pathway, oxidation-reduction process and transmembrane transport), and 10,183 linked to cellular component (mainly integral

316 component of membrane) (Table S2). Finally, the enzyme code annotation with KEGG mapping
317 identified 1,948 transcripts, which represented 433 enzymes in 122 unique pathways (Table 3,
318 S3). KEGG pathways included those related to purine metabolism, biosynthesis of antibiotics, T
319 cell receptor signaling pathway, Th1 and Th2 cell differentiation, and ether lipid metabolism.

Pathway	Pathway ID	No. transcripts	No. enzymes
Purine metabolism	map00230	825	47
Thiamine metabolism	map00730	727	6
Drug metabolism - other enzymes	map00983	233	16
Biosynthesis of antibiotics	map01130	185	100
T cell receptor signaling pathway	map04660	132	2
Th1 and Th2 cell differentiation	map04658	129	1
Glutathione metabolism	map00480	76	13
Ether lipid metabolism	map00565	69	7
Cysteine and methionine metabolism	map00270	60	23
Sphingolipid metabolism	map00600	58	14

Table 3. Top 10 KEGG pathways in the transcriptome.

320

321 *3.3 Gene expression patterns in eggs and throughout early development*

322 A total of 35,126 sequences remained after filtering to those with more than 0.5 counts
323 per million mapped reads across at least three of the 21 samples. A PCA of sample-to-sample
324 distances showed that gene expression profiles were more similar within stages (i.e., among
325 sample replicates) than across developmental stages (Fig. 1). The first and second dimensions
326 captured 62.94% and 13.09% of the variation, respectively, and revealed a clear separation
327 between egg/early embryonic stages and later developmental stages. Moreover, hierarchical
328 clustering revealed two primary clusters (Fig. 1): cluster 1 included cluster 1a, which contained
329 eggs (EG) and cluster 1b which included the 8- to 16-cell (CL) and morula (MO) stages; cluster
330 2 included cluster 2a, which contained blastula (BL) and gastrula (GA) stages, and cluster 2b,
331 which contained prism (PR) and pluteus stages (PL). In general, gene expression profiles

followed the progression of development during which time major developmental processes and alterations in morphology occurred. Similar separations in transcriptomic patterns between earlier (e.g. egg and cleavage) and later (e.g., blastula an onwards) developmental stages have been observed in several other urchin species, and have been attributed to the transition between maternal and zygotic transcription (i.e., the MZT) (Gildor, et al. 2016, Israel, et al. 2016). The timing of the MZT in *M. franciscanus* will be discussed in further detail later (section 3.4)

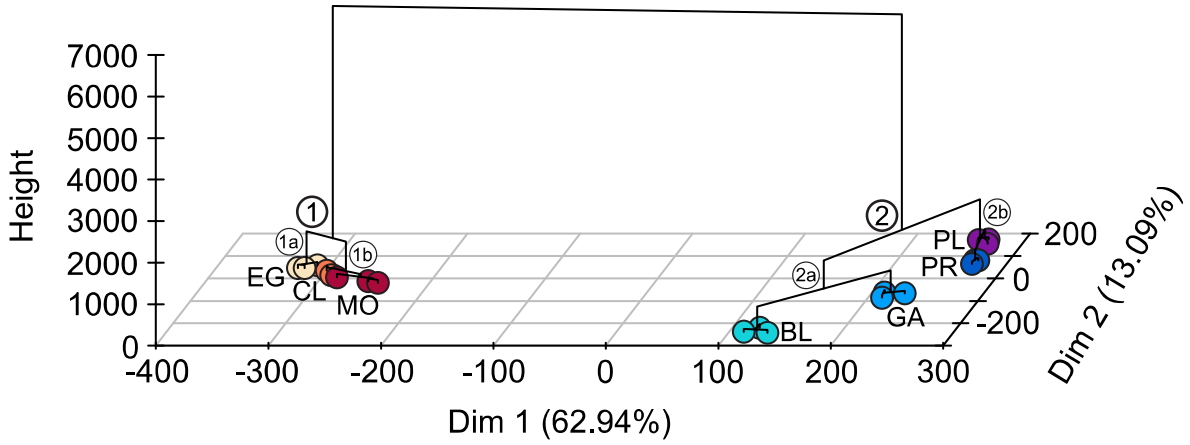


Fig. 1. PCA of *Mesocentrotus franciscanus* eggs and early developmental stages showing the first two dimensions and hierarchical clustering of the samples. Sample colors denote the different stages, which include: egg (EG), 8- to 16-cell (CL), morula (MO), blastula (BL), gastrula (GA), prism (PR), and pluteus (PL). Hierarchical clustering show two main clusters (1 and 2), which each contain two clusters (a and b).

WGCNA was used to highlight groups of genes that were co-expressed in eggs and each developmental stage. After filtering, normalizing and voom-transforming the data, the remaining 35,126 genes were assigned into module eigengenes containing similarly expressed genes. Only 86 genes remained unclustered and unassigned, and were grouped into the grey module (Fig. 2). All other genes were assigned into 15 different modules that were designated by color, and hierarchical clustering of the module eigengenes revealed three main clusters (Fig. 2). Each

module was related to developmental stage to generate eigengene networks with positive or negative correlation values ranging from 1 to -1 (Fig. 2).

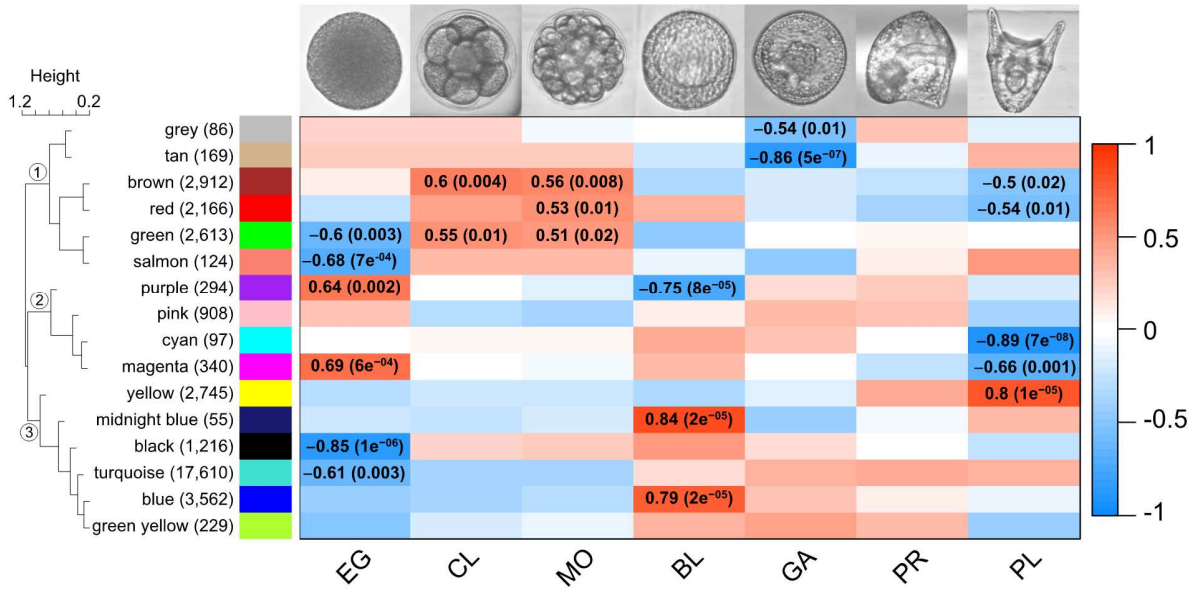


Fig. 2. WGCNA identified significant correlations between module eigengenes (rows) and stages (columns). The stages are: egg (EG), 8- to 16-cell (CL), morula (MO), blastula (BL), gastrula (GA), prism (PR), and pluteus (PL). The number of genes within each module eigengene is noted in parentheses following each color name. The red-blue color scale represents the strength of the correlation (1 to -1). Each correlation value (r^2) is followed by a *p*-value in parentheses. Hierarchical clustering of the module eigengenes revealed three primary clusters of gene expression (1-3).

Overall, WGCNA showed a mixed result for the 15 modules that were identified. Of the 15 module eigengenes, pink (908 genes) and green yellow (229 genes) were not significantly correlated to any stage (r^2 correlation ≤ 0.50, *p*-value ≥ 0.05). In addition, functional enrichment analyses did not identify any GO terms within the module eigengenes purple (294 genes), cyan (97 genes), magenta (340 genes), or midnight blue (55 genes). However, there were nine remaining module eigengenes that were significantly correlated to at least one stage (r^2 correlation ≥ 0.50 and *p*-value ≤ 0.05), and in which functional enrichment analyses successfully identified GO terms. These module eigengenes were tan (169 genes), brown (2,912

genes), red (2,166 genes), green (2,613 genes), salmon (124 genes), yellow (2,745 genes), black (1,216 genes), turquoise (17,610 genes), and blue (3,562).

The WGCNA analysis revealed that the unfertilized eggs (EG) were significantly correlated to six module eigengenes, the greatest number of any developmental stage (Fig. 2). A significant negative correlation of EG with module eigengenes green, salmon, black, and turquoise revealed that, relative to the measured developmental stages post-fertilization, the egg transcriptome was characterized by a down-regulation of genes related to metabolic processes and catalytic activity (Tables 4, S4). Unfortunately, while EG had a significant positive correlation with module eigengenes purple and magenta, functional enrichment analyses failed to reveal any GO terms within these modules. However, in other studies, there is evidence that active transcription and translation occurs at low levels in unfertilized sea urchin eggs (Chassé, et al. 2018, Ruderman and Schmidt 1981). In the purple urchin, *Strongylocentrotus purpuratus*, eggs possess transcripts of genes related to the cell cycle and DNA replication (Tu, et al. 2014). In the Mediterranean sea urchin, *Paracentrotus lividus,* genes related to the cell cycle and DNA replication are translated upon fertilization (i.e., at the one-cell embryo stage) (Chassé, et al. 2018).

Unlike the eggs analyses, functional enrichment analyses were able to identify GO terms within module eigengenes positively correlated with early embryo stages. In this study, early embryo cell divisions are represented by the 8- to 16-cell stage (CL) and the morula stage (MO), whose transcriptomes are highly similar to one another (Fig. 1). CL and MO were both positively correlated to module eigengenes brown and green (module eigengene cluster 1, Fig. 2). The MO stage also had a significant correlation with module eigengene red, which was within module eigengene cluster 1 (Fig. 2). Enrichment analyses of these modules revealed that during these

early cell divisions, the embryos contain transcripts encoding proteins related to metabolic processes, catalytic activity, and organelle and membrane formation (Tables 4, S4), which likely reflect various processes involved in cell proliferation. This finding is consistent with proteomic data from early embryos of *S. purpuratus* (Guo, et al. 2015). Additionally, GO terms related to signal transduction, G protein-coupled receptor signaling pathway, and transmembrane transport were identified. This result is similar to observations made in *P. lividus*, in which there was an enrichment of GO terms related to signaling pathways and cellular transport in the early embryological stages (Gildor, et al. 2016). Lastly, GO terms related to the cell cycle were identified in each of the brown, green, and red modules. This is comparable to *S. purpuratus*, in which genes encoding proteins related to the cell cycle are prominent in cleavage stage embryos (Tu, et al. 2014).

Upon progression to the blastula stage, extensive cell differentiation occurs in which the embryo forms a blastocoel, cilia, and enzymes required to digest the fertilization membrane during the hatching process (Barrett and Edwards 1976, Lepage and Gache 1989). With regard to stage-specific gene expression, the blastula stage (BL) was negatively correlated with module eigengene purple and positively correlated with module eigengenes midnight blue and blue (module eigengene cluster 3, Fig. 2). Although functional enrichment analyses failed to reveal any GO terms within module eigengenes purple or midnight blue, module eigengene blue contained GO terms related to RNA-directed DNA polymerase activity, protein binding, DNA integration, G protein-coupled receptor activity, and transmembrane transport (Tables 4, S4). The enrichment of these genes is in alignment with other studies in which genes related to DNA replication and energy production were expressed during the blastula stage in *S. purpuratus* (Gildor, et al. 2016, Tadros and Lipshitz 2009).

401 Gastrulation is a major and fundamental process of metazoan development (Wolpert
402 1992) that begins by invagination at the vegetal plate and the formation of the archenteron (Dan
403 and Okazaki 1956, Ettensohn 1984). Somewhat surprisingly, there were few correlations
404 between the gastrula stage (GA) and module eigengenes identified by WGCNA. GA was
405 negatively correlated with only the module eigengene tan (module eigengene cluster 1, Fig. 2).
406 GA was also negatively correlated with the grey module, which contained the unclustered and
407 unassigned genes. Functional enrichment analysis of module eigengene tan revealed 39 GO
408 terms, including oxidation-reduction process, integral component of membrane, ion
409 transmembrane transport, and ATPase activity (Tables 4, S4). Unfortunately, the WGCNA
410 analysis did not identify any modules with a significant positive correlation with GA. Previous
411 investigations of the *M. franciscanus* gastrula transcriptome, however, reported GO terms related
412 to cell differentiation and signal transduction involved in cell cycle checkpoints (Gaitán-Espitia
413 and Hofmann 2017). In *S. purpuratus*, there is an increase in expression of genes related to
414 biomineralization, the nervous system, immunity and the defensome once gastrulation begins
415 (Tu, et al. 2014).
416 The digestive tract and supporting skeletal rods are formed during the prism and early
417 pluteus stages, which are necessary for the planktotrophic feeding strategy of the urchin larvae
418 (Burke 1980, Ettensohn and Malinda 1993). The prism stage (PR) was not significantly
419 correlated to any module eigengene (Fig. 2). However, its expression patterns were similar to
420 that of the pluteus stage (PL) (Figs. 1, 2). PL was negatively correlated to module eigengenes
421 brown and red (Fig. 2), which include GO terms related to ATP binding, integral component of
422 membrane, transmembrane transport, signal transduction and the cell cycle (Tables 4, S4). PL
423 was also negatively correlated with module eigengenes cyan and magenta, although functional

enrichment analyses were unable to identify GO terms within these modules. Lastly, PL was positively correlated with one module eigenegene, yellow, which was within module eigengene cluster 3 (Fig. 2). Enrichment analysis identified GO terms within yellow that included those related to metallopeptidase activity, metabolism, adhesion, the cytoskeleton, and the immune system (Tables 4, S4). This observation was in agreement with our earlier work (Gaitán-Espitia and Hofmann 2017), in which genes related to these processes and structures were up-regulated in *M. franciscanus* pluteus larvae relative to gastrula stage embryos. The yellow module also included GO terms related to ATP binding, oxidation-reduction process, calcium ion binding, acetylcholine-gated cation-selective channel activity, and ion transmembrane transport (Tables 4, S4). This expression pattern likely reflects the energy production and biomineralization processes necessary to support gut and skeletal formation in the developing pluteus larvae.

GO ID	GO term name	GO category	FDR value	No. transcripts (% of ref)
tan: GA (r2 = -0.86)				
GO:0016021	integral component of membrane	Cellular Component	6.67E-06	26 (0.6)
GO:0055114	oxidation-reduction process	Biological Process	2.27E-05	12 (1.2)
GO:0016491	oxidoreductase activity	Molecular Function	2.82E-04	11 (1.1)
GO:0034220	ion transmembrane transport	Biological Process	3.13E-03	6 (2)
GO:0015267	channel activity	Molecular Function	5.97E-03	6 (1.7)
GO:0042623	ATPase activity, coupled	Molecular Function	5.97E-03	5 (2.5)
brown: CL (r2 = 0.6), MO (r2 = 0.56), PL (r2 = -0.5)				
GO:0016021	integral component of membrane	Cellular Component	4.81E-12	207 (4.6)
GO:0005524	ATP binding	Molecular Function	1.63E-10	79 (6.5)
GO:0055085	transmembrane transport	Biological Process	1.12E-05	57 (5.7)
GO:0007049	cell cycle	Biological Process	5.60E-04	20 (8.5)
GO:0035556	signal transduction	Biological Process	6.36E-04	24 (7.4)
GO:0140098	catalytic activity, acting on RNA	Molecular Function	3.75E-03	20 (7.2)
red: MO (r2 = 0.53), PL (r2 = -0.54)				
GO:0016021	integral component of membrane	Cellular Component	9.33E-14	172 (3.8)
GO:0005524	ATP binding	Molecular Function	1.18E-11	68 (5.6)
GO:0022402	cell cycle process	Biological Process	2.01E-05	16 (10.1)
GO:0007165	signal transduction	Biological Process	3.78E-05	80 (3.5)
GO:0055114	oxidation-reduction process	Biological Process	8.93E-04	39 (4.1)
GO:0055085	transmembrane transport	Biological Process	2.53E-02	36 (3.5)
green: EG (r2 = -0.6), CL (r2 = 0.55), MO (r2 = 0.02)				
GO:0005524	ATP binding	Molecular Function	2.79E-34	117 (10)
GO:0035556	intracellular signal transduction	Biological Process	2.35E-21	48 (16.1)

GO:0140098	catalytic activity, acting on RNA	Molecular Function	1.48E-10	31 (11.7)
GO:0016021	integral component of membrane	Cellular Component	4.84E-09	177 (3.9)
GO:0007049	cell cycle	Biological Process	1.24E-08	26 (11.4)
GO:0055114	oxidation-reduction process	Biological Process	5.06E-07	53 (5.7)
salmon: EG (r2 = -0.68)				
GO:0003824	catalytic activity	Molecular Function	1.81E-07	34 (0.4)
GO:0005515	protein binding	Molecular Function	9.29E-04	20 (0.4)
GO:0008152	metabolic process	Biological Process	5.40E-03	26 (0.3)
GO:0008168	methyltransferase activity	Molecular Function	5.40E-03	6 (1.7)
GO:0005543	phospholipid binding	Molecular Function	5.40E-03	4 (4.7)
GO:0016020	membrane	Cellular Component	9.21E-03	21 (0.3)
yellow: PL (r2 = 0.8)				
GO:0055114	oxidation-reduction process	Biological Process	8.55E-23	87 (9.7)
GO:0005509	calcium ion binding	Molecular Function	1.44E-16	77 (8.3)
GO:0022848	acetylcholine-gated cation-selective channel activity	Molecular Function	1.78E-07	9 (60)
GO:0034220	ion transmembrane transport	Biological Process	4.70E-07	27 (9.8)
GO:0005524	ATP binding	Molecular Function	3.95E-04	57 (4.6)
GO:0002376	immune system process	Biological Process	6.06E-04	18 (8.4)
black: EG (r2 = -0.85)				
GO:0005515	protein binding	Molecular Function	1.18E-11	111 (2.4)
GO:0003677	DNA binding	Molecular Function	8.07E-07	39 (3.3)
GO:0006464	cellular protein modification process	Biological Process	3.49E-05	38 (2.8)
GO:0046872	metal ion binding	Molecular Function	4.01E-05	69 (2.1)
GO:0006396	RNA processing	Biological Process	1.46E-04	16 (4.9)
GO:0016021	integral component of membrane	Cellular Component	1.59E-04	86 (1.8)
turquoise: EG (r2 = -0.61)				
GO:0005524	ATP binding	Molecular Function	1.50E-92	499 (63.2)
GO:0015074	DNA integration	Biological Process	5.19E-40	271 (53.3)
GO:0005509	calcium ion binding	Molecular Function	3.64E-36	315 (45.5)
GO:0003964	RNA-directed DNA polymerase activity	Molecular Function	3.09E-33	415 (38)
GO:0006278	RNA-dependent DNA biosynthetic process	Biological Process	4.22E-33	415 (37.9)
GO:0005525	GTP binding	Molecular Function	2.54E-19	136 (51.7)
blue: BL (r2 = 0.79)				
GO:0016021	integral component of membrane	Cellular Component	7.53E-28	296 (6.7)
GO:0003964	RNA-directed DNA polymerase activity	Molecular Function	5.84E-24	132 (9.6)
GO:0005515	protein binding	Molecular Function	5.01E-21	277 (6.1)
GO:0015074	DNA integration	Biological Process	1.28E-10	65 (9.1)
GO:0004672	protein kinase activity	Molecular Function	2.26E-09	48 (10.3)
GO:0055085	transmembrane transport	Biological Process	4.85E-06	67 (6.7)

Table 4. Select GO term results from functional enrichment analyses of WGCNA module eigengenes

3.4. The maternal-to-zygotic transition

To examine the timing of the MZT, we assessed: 1) the decline of maternally-derived transcripts, and 2) the increase of zygotic transcription. Upon targeting genes that play a role in

the degradation of maternal RNAs, one *DGCR8*-like gene (*dgcr8*), one *dicer* gene (*dicer*), one *smaug* homolog (*smaug1*) and three putative *smg* genes (*smg7*, *smg8*, and *smg9*) were identified within the *M. franciscanus* developmental transcriptome (Table 1). The expression levels of *dgcr8, dicer*, *smg7*, *smg8*, and *smg9* all peaked during the 8- to 16-cell (CL) and morula (MO) stages (Fig. 3A). The *dgcr8* gene plays a role in processing microRNAs that are required for degrading mRNAs in mammals (Marlow 2010, Wang, et al. 2007). The Mediterranean sea urchin, *P. lividius*, exhibited a similar pattern of expression of *dgcr8* as reported here, in which there was a peak in expression within 8- and 16-cell embryos (Gildor, et al. 2016). The authors attributed this observation to the role of *dgcr8* in degrading maternal mRNAs (Gildor, et al. 2016). *Dicer* is involved in clearing maternal messages in zebrafish and mice (Giraldez, et al. 2005, Marlow 2010), and mutations in the *dicer* gene are known to alter and arrest embryonic development in some species (Murchison, et al. 2007).

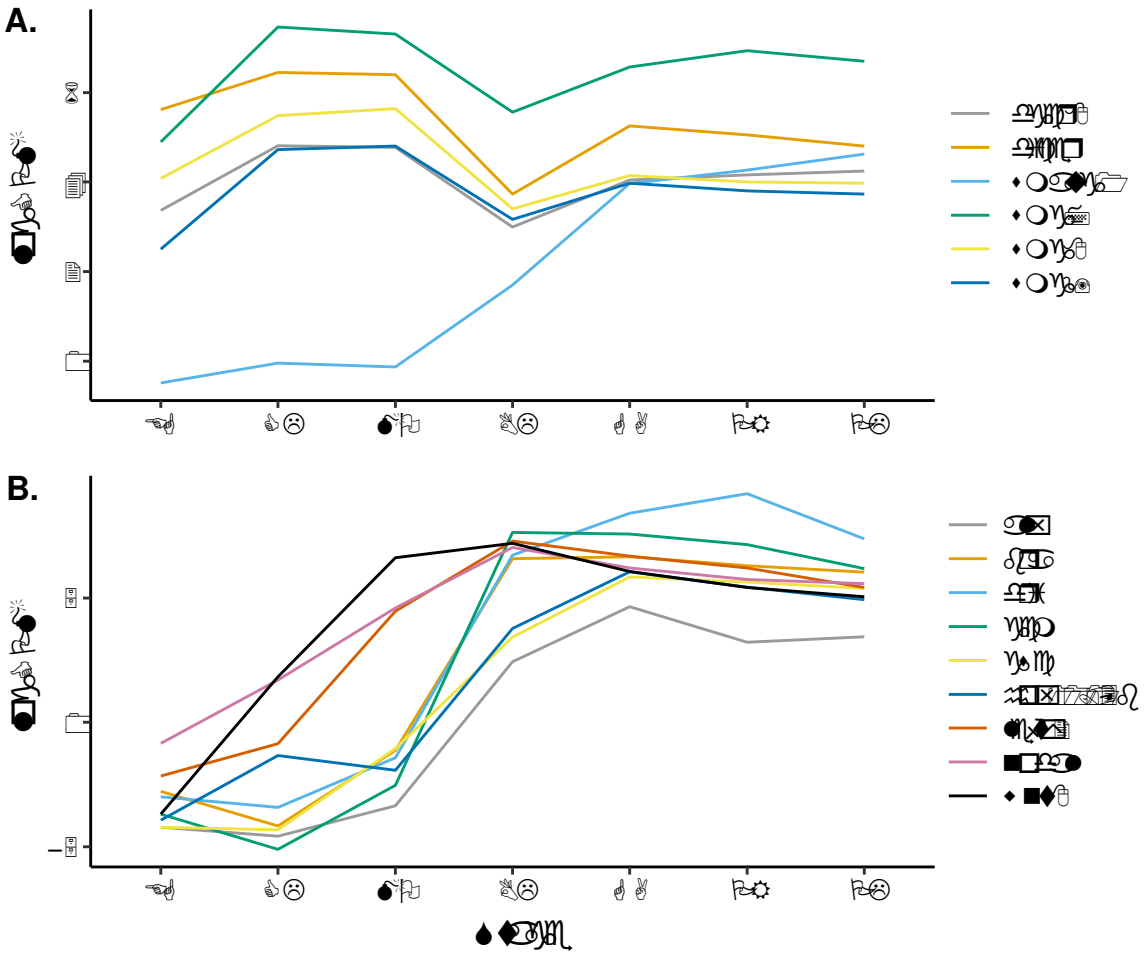


Fig. 3. The expression of putative genes that play a functional role during the MZT. These genes **A.** regulate the removal of mRNA, and **B.** regulate zygotic development. The data is in log2 counts per million reads (log CPM) expressed at each stage of development: egg (EG), 8- to 16-cell (CL), morula (MO), blastula (BL), gastrula (GA), prism (PR), and pluteus (PL).

The *smg* genes code for proteins that function in nonsense-mediated mRNA decay (NMD) in a variety of organisms (Okada-Katsuhata, et al. 2012, Pulak and Anderson 1993, Yamashita, et al. 2009). The NMD pathway detects and degrades mRNAs, and is often described as a surveillance pathway that serves as a quality-control mechanism to remove mRNAs with premature termination codons (Chang, et al. 2007, Hentze and Kulozik 1999). However, the NMD pathway also serves functional roles that shape gene expression and are important for differentiation and development (Lykke-Andersen and Jensen 2015). Additionally, the NMD pathway has been shown to selectively degrade mRNA transcripts with longer 3' UTRs, causing

460 a relative enrichment of shorter 3' UTR transcripts (Bao, et al. 2016). In zebrafish embryos, 3'
461 UTR length affects the stability of maternal mRNAs because longer 3' UTRs confer resistance to
462 codon-mediated deadenylation, the first step required for mRNA decay (Mishima and Tomari
463 2016). Therefore, the removal of long 3' UTR transcripts via the NMD pathway may increase
464 the relative proportion of short 3' UTR transcripts available for deadenylation and decay during
465 the CL and MO stages. Overall, the peak in expression of *dgcr8, dicer*, *smg7*, *smg8*, and *smg9*
466 during the CL and MO stages supports that maternal mRNAs are degraded during this period of
467 embryonic development.

468 In contrast to the expression of *dgcr8*, *dicer*, and *smg* genes, *smaug1* was not expressed
469 until the blastula stage (Fig. 3A). The *smaug* gene is a transcriptional regulator known to bind to
470 and target maternal RNAs for degradation in *Drosophila melanogaster* and is highly conserved
471 across taxa (Tadros, et al. 2007). It is therefore possible that degradation of maternal mRNAs is
472 still ongoing at the blastula stage. This differs from observations in *S. purpuratus*, in which
473 maternal degradation appears to end prior to the blastula stage (Tadros and Lipshitz 2009, Wei,
474 et al. 2006). With the exception of *smaug1* expression, the degradation of maternal transcripts
475 appears to primarily occur during the 8 - to 16-cell (CL) and morula (MO) stages. This is
476 additionally supported by the WGCNA analysis, which revealed genes related to catalytic
477 activity acting on RNA in module eigengenes brown and green, both of which share significant,
478 positive correlations with the CL and MO stages (Table 4).

479 Evidence of maternal transcript degradation was also reflected by a decrease in levels of
480 maternal transcripts, which were represented by those expressed in unfertilized eggs (EG). A
481 heatmap of the top 500 transcripts expressed in eggs revealed that expression of these transcripts
482 began to decline at the 8-to 16-cell and morula stages (Fig. 4). By the blastula stage, the overall

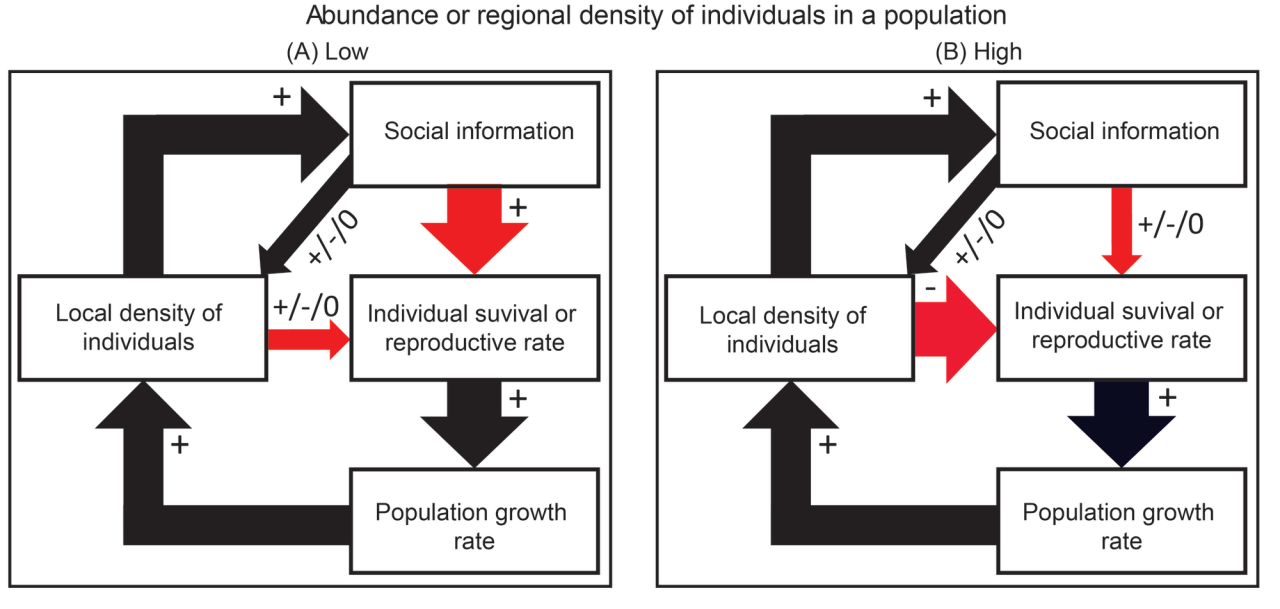
patterns of the maternal transcript levels had completely changed, with moderate retention of some maternal transcripts, although the majority had dramatically decreased. Most of these maternal transcripts continued to show low levels relative to the eggs during the remaining stages of development (i.e., gastrula through pluteus stages). Taken together, the degradation of maternal RNAs and the resulting reduction in expression of maternal transcripts begin as early as the 8-cell stage, although it is possible that the process begins even sooner after fertilization at a stage prior to what was examined in this study (e.g., at the 2-cell or 4-cell stage). This result is similar to the timing of maternal mRNA degradation in *S. purpuratus*, in which maternal transcripts are destabilized by early cleavage stages (Tadros and Lipshitz 2009, Tu, et al. 2014, Wei, et al. 2006).


(A)


(B)


(C)


(D)


Abundance or regional density of individuals in a population
(A) Low
Social information
+
+/-/0
+
Local density of individuals
+/-/0
Individual suvival or reproductive rate
+
+
Population growth rate
(B) High
Social information
+
+/-/0
+/-/0
-
Local density of individuals
Individual suvival or reproductive rate
+
+
Population growth rate