



Resource

Quantitative developmental transcriptomes of the sea urchin *Strongylocentrotus purpuratus*



Qiang Tu, R. Andrew Cameron, Eric H. Davidson*

Division of Biology, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

ARTICLE INFO

Article history:

Received 15 October 2013

Received in revised form

13 November 2013

Accepted 18 November 2013

Available online 26 November 2013

Keywords:

Sea urchin

Developmental transcriptome

RNA sequencing

ABSTRACT

Development depends on the precise control of gene expression in time and space. A critical step towards understanding the global gene regulatory networks underlying development is to obtain comprehensive information on gene expression. In this study, we measured expression profiles for the entire expressed gene set during sea urchin embryonic development. We confirmed the reliability of these profiles by comparison with NanoString measurements for a subset of genes and with literature values. The data show that ~16,500 genes have been activated by the end of embryogenesis, and for half of them the transcript abundance changes more than 10-fold during development. From this genome scale expression survey, we show that complex patterns of expression by many genes underlie embryonic development, particularly during the early stages before gastrulation. An intuitive web application for data query and visualization is presented to facilitate use of this large dataset.

© 2013 Elsevier Inc. All rights reserved.

Introduction

Differential gene expression in time and space is the essence of the developmental process. With the availability of material on the scale of liters of eggs, sea urchin embryos, in particular those of *Strongylocentrotus purpuratus*, have long been used as an experimental model for study of gene expression. A rich earlier history of studies of cell specification in this species complements examinations of differential gene expression (Davidson, 1986). Current studies have led to one of the most complete gene regulatory networks (GRNs) for early development (Oliveri et al., 2008; Peter and Davidson, 2009; Peter et al., 2012). Genomically encoded cis-regulatory modules constitute the nodes of the GRN. Various inputs are integrated at these modules to control gene expression. The regulatory state of a particular embryonic cell is the sum of the various regulatory gene expressions in that cell. Subsequent regulatory states in descendant daughter cells follow through the activity of the regulatory state in the ancestor.

Hints of these regulatory processes had been revealed in mass culture experiments before the 1990s. First maternal RNA transcripts, then later zygotic ones coding for regulatory molecules, install the progressively more complex sets of regulatory states which accompany the increasing number of embryonic cells. The messenger RNA is stored in the unfertilized egg as a complex mixture of poly(A)- and non-poly(A)-RNA. It was estimated to amount to 30 pg per egg in mass and approximately 8500 message species of 2 kb length (Davidson, 1986). Transcription increases in the early divisions after

fertilization, reaching a maximal rate by 4–5th cleavage (Wilt, 1970). Through nearly equivalent rates of synthesis and degradation, steady state mRNA content remains the same throughout embryonic development. From the kinetics of polysome assembly (Goustin and Wilt, 1981) and solution hybridization of genomic DNA with cytoplasmic RNA samples it is inferred that the loading of maternal RNA on polysomes to synthesize maternally encoded proteins is completed by approximately the 8-cell stage. Thereafter newly synthesized mRNAs are loaded.

Since these early mass measurements, our views on gene expression and developmental process have been refined through numerous measurements of single or small sets of genes in the embryo. New technologies make it possible to envision gene regulatory networks controlling the total of all genes expressed in development. Thus, a temporal profile of all of the genes expressed in the embryo would contribute an important piece of this puzzle. Using a largely complete gene set produced from deep sequencing of 22 stages and tissues of the purple sea urchin (Tu et al., 2012), we are now in a position to enumerate the temporal patterns of gene expression for the entire 21,000 genes at 10 time points in the embryonic development of this species. Estimates of temporal expression patterns are particularly informative during embryonic stages since the total mass of RNA remains approximately the same throughout this period.

Results and discussion

Quantification of developmental transcriptomes

We have analyzed transcriptomes throughout embryonic development of the *S. purpuratus* using the RNA-seq method. A set of gene

* Corresponding author.

E-mail addresses: jrigg@caltech.edu, davidson@caltech.edu (E.H. Davidson).

models was generated from RNA-seq data in an earlier study (Tu et al., 2012). The transcript abundance was estimated for each gene at each of the 10 embryonic stages (Table S1) in FPKM (Fragments Per Kilobase of transcript per Million mapped reads) using Cufflinks (Trapnell et al., 2010). A set of internal standards comprising mRNAs present in known numbers of molecules was added to the sample (Mortazavi et al., 2008). The mRNA mass of each sequenced sample before adding the internal standards was 100 ng. Using the precise values for mRNA in the sea urchin embryo previously obtained, we calculated that this is equivalent to ~3300 embryos. Thus the relative FPKM values can be translated to the absolute abundance number in terms of transcripts per embryo. For example, an FPKM value of ~5 is equal to 300 transcripts per embryo.

We refined the measurement of transcript abundance in several ways. A key point is that we used the coding sequence (CDS) instead of the full-length gene models to calculate the number of reads mapped. As its name indicates, an FPKM value is the number of reads mapped to the given gene model normalized by the length of the model and by the total number of reads mapped to the genome. Since the full-length gene models were generated from 22 samples, including embryonic stages, larval stages, and adult tissues, they include the longest untranslated regions (UTRs). The UTRs of transcripts in individual samples might vary significantly. When the full-length gene models are used in the FPKM calculation, transcripts with shorter UTRs will still be normalized to the longest lengths, thus distorting the FPKM values. However, CDSs are the same across all samples. Thus we used CDS length for the abundance calculation, and we found that counting reads mapped only to the CDSs produces the most accurate FPKM values.

In order to confirm the reliability of the quantitation method, for 9 stages the abundance of a set of 173 regulatory genes in the same samples was measured using a NanoString, an instrument which counts mRNA directly without use of enzymatic reactions (Geiss et al., 2008). These NanoString measurements were used as a gold standard for validation of the values obtained by the RNA-seq method. The FPKM values derived from the RNA-seq data were compared with the NanoString counts of the control set of genes (Fig. 1A) and correlation coefficients for the 173 pairs of profiles were also calculated (Fig. 1B and C). The results from two methods matched very well. The median of the correlation coefficients is 0.923. A typical pair of expression profiles is shown in Fig. 1B, and all individual comparisons are shown in Fig. S1A.

General quantitative aspects of developmental transcriptomes

There are ~23,000 genes predicted to exist in the *S. purpuratus* genome (Sea Urchin Genome Sequencing Consortium et al., 2006)

and the earlier transcriptome study compiled 21,092 gene models based on 22 transcriptomes (Tu et al., 2012). A cutoff of 300 transcripts per embryo (FPKM ~5) was chosen as the lower limit for a functionally meaningful level of transcript representation. In an embryo at the mesenchyme blastula stage this criterion is equivalent to less than one transcript per cell for a ubiquitously expressed gene, or more for a cell-type-specifically expressed gene. Thus for example, even productively transcribed regulatory genes which produce low abundance mRNAs are expressed above this level, and technical variations in the quantification by the RNA-seq method are relatively low at this cutoff (Tu et al., 2012). Calculated in this way, ~16,500 genes (72% of the estimated total gene number) have been activated in at least one of the embryonic stages surveyed. These genes can be defined as an embryonic gene repertoire. They are shown as dots above the dashed ordinate cutoff line in Fig. 2A, as their highest transcript abundances during embryogenesis are >300. Similarly, embryonic housekeeping genes that are continuously expressed throughout embryogenesis can be identified if the lowest transcript abundance is over this cutoff (Fig. 2A, dots to the right of the dotted abscissa cutoff line). This class constitutes ~5700 genes (35% of the repertoire). A large-scale developmental transcriptome study of *Drosophila melanogaster* reported a similar percentage: 40% of expressed genes are constitutively expressed in all 27 stages studied, including embryonic, larval, and adult stages (Graveley et al., 2011). Considering the relative expression profiles, the transcript abundance levels of ~7900 genes (48% of the embryonic gene repertoire) undergo changes greater than 10-fold (Fig. 2A, red dots), while only ~1300 genes (8% of the embryonic gene repertoire) are expressed at relatively constant levels (less than 3-fold variation) (Fig. 2A, blue dots). These statistics reflect the highly dynamic use of the embryonic transcriptome, which is characteristic of the majority of genes used during this time.

At each stage of sea urchin embryogenesis, on average ~11,500 genes are actively transcribed, producing about 39 million transcripts per embryo. The histograms with linear and log scales show that the transcript abundance follows a log-normal distribution (Fig. 2B, C), and the cumulative histogram shows that 1% of genes contribute 39% of transcripts, 20% of genes contribute 85% of transcripts, and 50% of those genes that are considered as active genes contribute 97% of transcripts (Fig. 2D). From these plots, it is apparent that there is no definitive boundary between so called “active” and “non-active” genes; thus an empirical threshold activity (>300 transcripts per embryo) was used, as indicated above. However, defined in such a way the active genes produce the overwhelming majority of transcripts.

Complexity of an mRNA population is defined as the length of single copy DNA sequence represented in the RNA population.

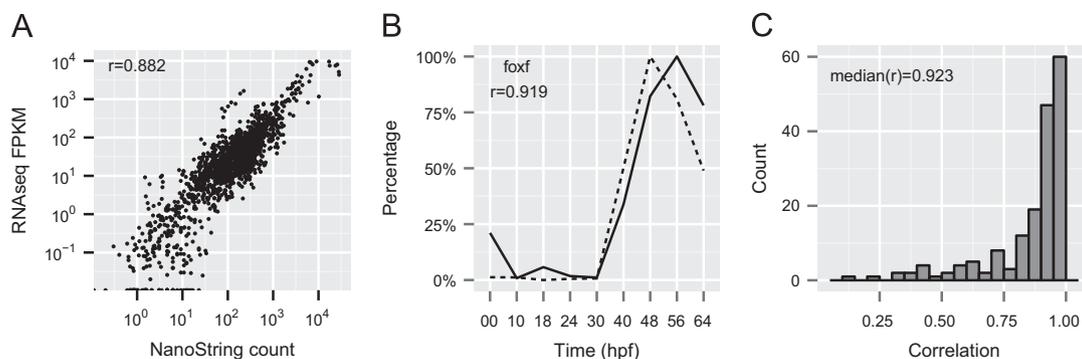


Fig. 1. Comparison between measurements by NanoString and RNA-seq. (A) Measurements of 173 genes in 9 samples of different embryonic developmental stages. (B) An example (FoxF) of time course profiles measured by NanoString and RNA-seq. The correlation coefficient between the two profiles is 0.919. Solid line: NanoString; dashed line: RNA-seq. (C) Distribution of correlation coefficients of the time course profiles measured by NanoString and RNA-seq for each gene. The median value is 0.923. See also Fig. S1.

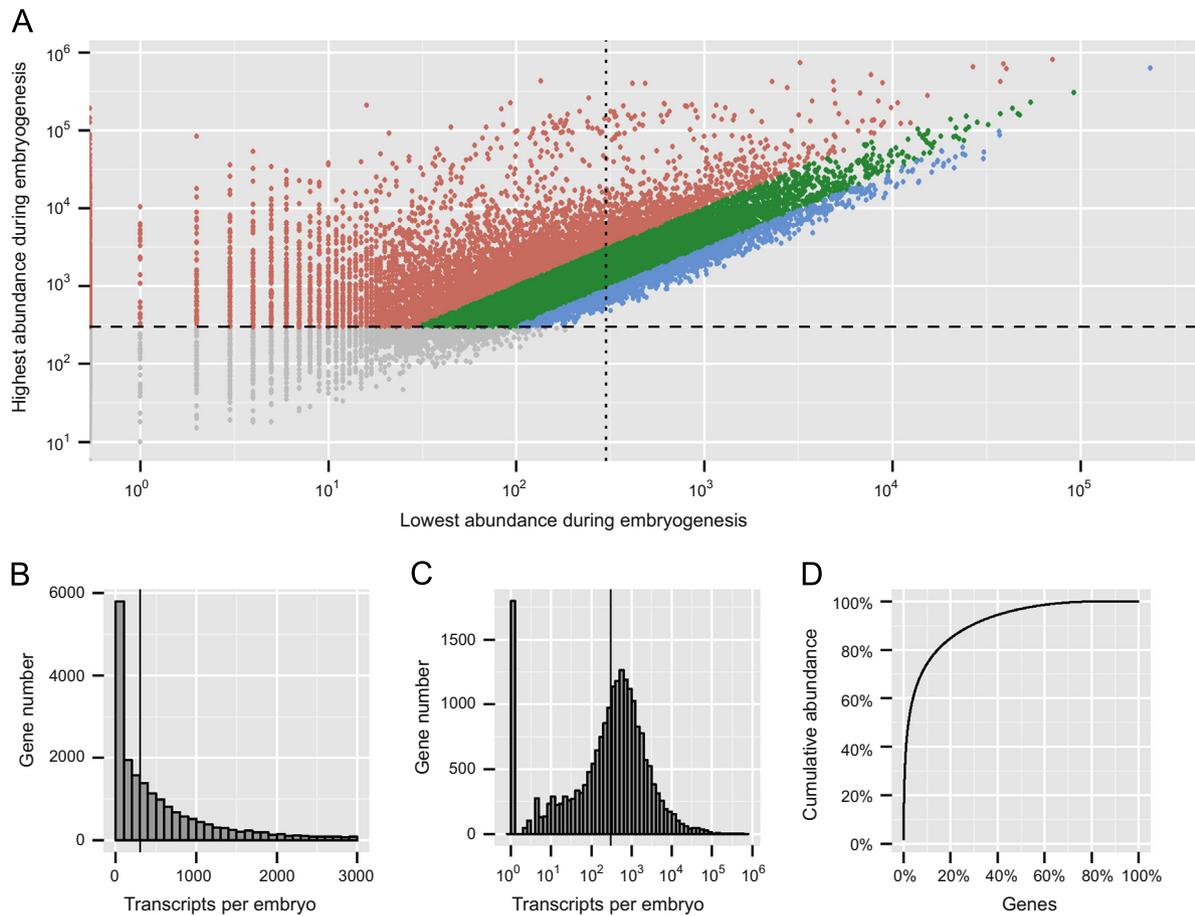


Fig. 2. Various distributions of transcript abundance. (A) Highest and lowest transcript abundance for each gene during embryogenesis. The ordinate is the highest and the abscissa is the lowest transcript abundance. The ordinate cutoff (> 300 transcripts per embryo, horizontal dashed line) defines whether a gene has been expressed: the upper colored dots are the embryonic gene repertoire; the lower gray dots are genes which are below the threshold. The abscissa cutoff (vertical dotted line) defines whether a gene has been continuously expressed: the dots on the right side represent the embryonic house-keeping genes, and the dots on the left side represent the genes whose transcript levels are very low at one or more time points. The diagonal color layers indicate the ratio of the highest to lowest transcript abundance: the blue dots represent genes expressed at constant levels with the fold less than 3; the red dots represent genes expressed at dynamic levels varying more than 10 fold. (B)–(D) Distributions of transcript abundance in the 24 hpf embryo. (B) Linear histogram of transcript abundance showing between 0 and 3000 transcripts per embryo. The vertical line represents 300 transcripts per embryo. (C) Log histogram of transcript abundance. All values were shifted up by 1 so that non-expressed genes (the bar at 10^0) can be made visible on the log plot. The vertical line represents 300 transcripts per embryo. (D) Cumulative abundance.

It describes the minimum amount of unique sequence content in a population of sequences. Traditionally it was measured by RNA/DNA reassociation kinetics. Here we can calculate it directly from the RNA-seq data. But the complexity is not the total length of gene models of active genes. As discussed before, the gene models generated in the previous study are based on the pool of 22 samples, and all combined UTRs could result in a gene model length greater than that of any one stage-specific transcript. To accurately estimate the complexity of each transcriptome, reads from each sample were assembled into sets of stage specific gene models which were in turn used to calculate the complexity. The average complexity is 24.6 Mb for the transcriptome and 2.1 kb for each gene. Solution hybridization methods estimated the complexity of the *S. purpuratus* blastula–gastrula transcriptome at 16.5 Mb. This lower estimate is comparable since the hybridization methods underestimate very similar gene sequences and very rare transcripts (Davidson, 1986).

Dynamics of expression

To more clearly represent the $\sim 16,500$ embryonic gene expression profiles, we performed a cluster analysis. As there are no well-defined boundaries between clusters of temporal profiles, we choose a fuzzy clustering algorithm mfuzz (Futschik and Carlisle,

2005) which can assign genes gradual degrees of membership to a cluster. Each expression profile was normalized by the peak value of the given gene during the development, because different genes execute their functions at very different levels, and the peak values are biologically meaningful. These relative expression profiles of the embryonic genes were grouped into 100 clusters (Fig. S2A). The distribution of the gene number per cluster is shown in Fig. S2B and C. For simplicity in the following discussion, each gene was assigned to the cluster of its maximum membership even though it could be included to a lesser extent in other clusters. However the fuzzy clustering details are retained in our web query application, so that users can decide the assignments manually (see below).

These clusters can be further merged into four major groups based on overall dynamics: “turn-off”, “turn-on”, “transient”, and “other” (Fig. 3). Each cluster was manually assigned to one of these groups. Because these expression profile clusters have no well-separated boundaries, the assignment is somewhat arbitrary. Nevertheless, an overview obtained from the analysis is useful.

The “turn-off” type constitutes 34% of the genes. The transcripts of these genes are highly abundant in the unfertilized egg, and then they start to decrease within 10–18 h post-fertilization (hpf). Only a small percentage of these maternal genes, the bottom four clusters (012, 013, 024, and 070) in Fig. 3, remain quiescent during

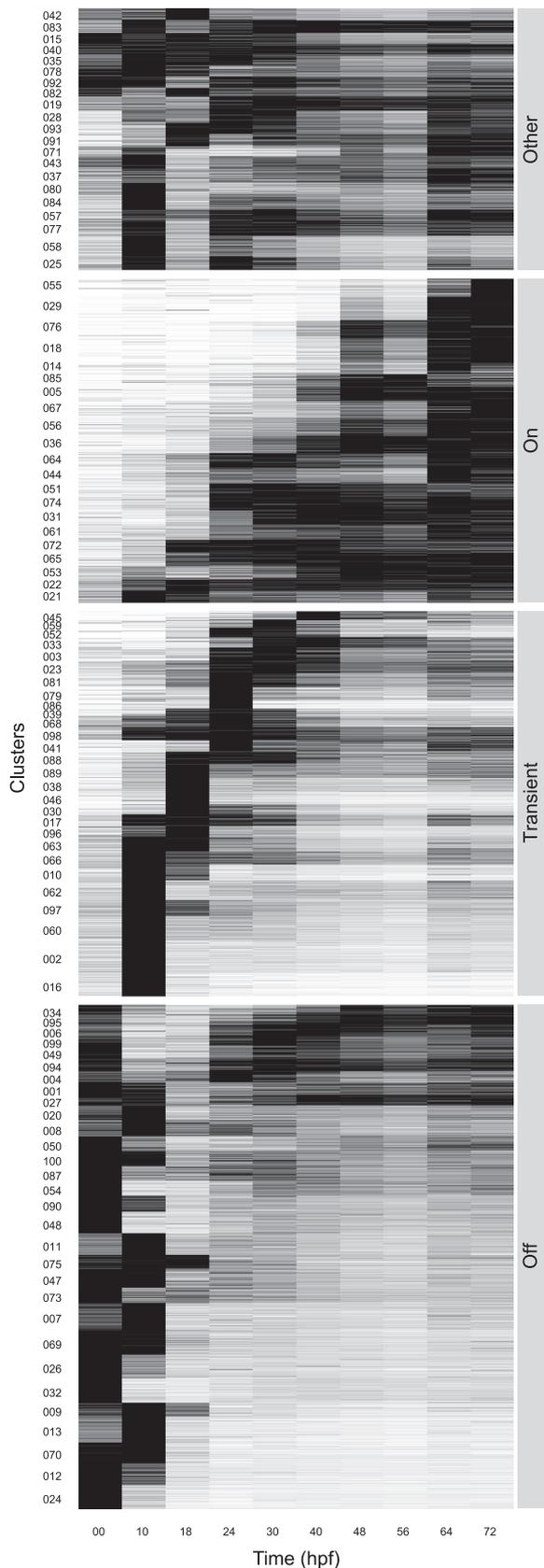


Fig. 3. Heatmap of relative expression profiles of embryonic genes. The transcript abundance of each gene is shown as a horizontal line and the lines are ordered by clusters. The ordinate shows IDs of clusters. The clusters are divided into four major groups. The abscissa gives developmental stages in terms of hours post-fertilization. The density indicates the abundance relative to the maximum value for the given gene during embryonic development. Black is maximum and white is zero. See also Figs. S2 and S3.

the whole of embryogenesis, while most of the maternal genes are expressed again at later stages (Flytzanis et al., 1982).

The “transient” type is typified by 26% of the genes. They are not maternally expressed, but reach a peak during development, and then return to much lower levels in late stages. Interestingly more than half attain a peak during cleavage (10 hpf) and early blastula stages (18 hpf). Almost all of them reach their peak values before early gastrulation (30 hpf).

The “turn-on” type constitutes 22% of the genes. These genes begin to be transcribed at various early or late stages and remain abundant throughout the remaining stages of embryonic development. The activity of these genes may increase at any developmental stage.

The remaining 18% of the expression profiles either remain constant throughout development or display more complex dynamic changes and cannot be easily grouped with the major types discussed above.

From this overview, it is apparent that there is no transcriptional quiescence in sea urchin embryos (Davidson, 1986). Most sea urchin maternal transcripts decrease quickly after fertilization and a large portion of the zygotic genome is activated during cleavage stages. Actually transcription is observed even in the unfertilized sea urchin egg, and continues in the zygote nucleus and throughout cleavage. This is in contrast to *Drosophila* and *Xenopus*, in which the genome remains quiescent until the blastoderm and blastula stages respectively (Davidson, 1986).

It is noteworthy that the activities of many genes undergo marked changes during the early phases of development before gastrulation (30–40 hpf). This includes almost all genes of the “turn-off” and “transient” types, and about half of the “turn-on” type. Altogether these genes account for 70% of the embryonic repertoire of genes. These varied profiles accompany drastic morphological changes: logarithmic increases in cell number, ingression of skeletogenic mesenchyme cells, and the establishment of many new regulatory states. After the early phases, gene expressions tend to remain at much more stable levels.

Gene usage in different functional classes

Previously we established a functional classification system using manually curated classes (Tu et al., 2012). Here we combined this classification system with a restricted group of GO terms derived from Blast2GO (Götz et al., 2008). In all, 39% of the genes fall into these 27 classes. The number of actively transcribed genes of each class at each developmental stage was compiled (Fig. 4). Many classes do not show a dramatic change in terms of active gene number. However, genes from classes of several specialized functions, e.g. biomineralization, defense, immunity, and nervous system, become significantly active only later in development, particularly after gastrulation begins (30 hpf).

Although knowledge of function and expression of many gene categories in the sea urchin is not extensive, some classification groups are well enough understood to allow comparisons between timing of expression and function to be drawn. For example in an earlier study the utilization during embryogenesis of every gene encoding a transcription factor predicted in the *S. purpuratus* genome was reported, and it was found that about 80% of the entire regulatory repertoire is expressed by the end of embryogenesis (Howard-Ashby et al., 2006). This is similar to the fraction of all genes expressed as reported in this work, a reasonable result since the expressed genes require the inputs provided by the expressed regulatory genes.

Turning to effector gene repertoires, we see that the pattern of utilization of the genes in the biomineralization functional group is not uniform (Fig. S3A). Transcripts of genes that code for spicule matrix proteins accumulate at about 24 hpf when the cells that

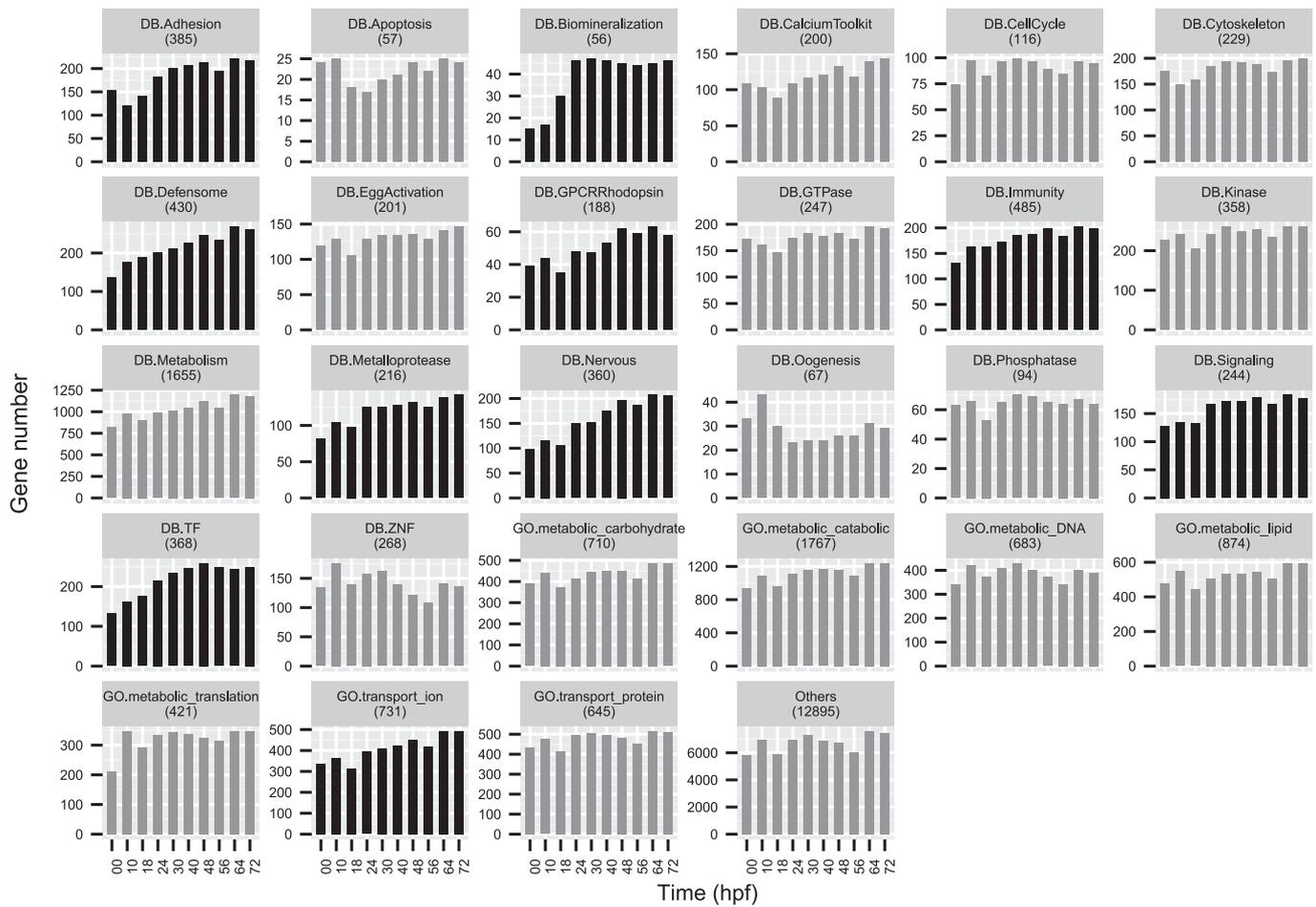


Fig. 4. Change of number of actively transcribed genes of each functional class. Classes compiled in our earlier study (Tu et al., 2012) are annotated with the prefix “DB”. Classes compiled using Blast2GO in this study are annotated with the prefix “GO”. Highlighted (black bars) classes show an increasing trend.

express these genes have ingressed into the blastocoel and are taking up their position before spiculogenesis begins. On the contrary, many cyclophilins started expression much earlier. Cyclophilins play multiple roles in cell biology. Their exact functions in sea urchin embryonic skeletogenesis are not clear, but indicated by their early expression profiles, they may play a role during specification of skeletogenic mesenchyme cells rather than in spiculogenesis directly.

Another example is the genes that encode for cell cycle proteins. The early cell cycles occur very rapidly and are devoid of the gap phases of the cycle. Cdk1, together with Cyclins A and B, drives the transition from G2 to M phase. Cdk2, together with Cyclins A and E, drives the G1 to S phase transition (Fernandez-Guerra et al., 2006). Transcripts for these five genes have highly similar profiles: they are very abundant in the egg as well as in cleavage stage embryos, and then decrease sharply. In fact, Cyclins A and B are amongst the most prominent maternal transcripts observed. Many other cell cycle kinases display distinctly different profiles: for instance those whose expression increases as cell cycles occur over longer intervals and gap phases emerge (Fig. S3B).

The requirement for DNA replication machinery is expected to increase with cell number during early stages. Replication forks are made active by the recruitment of the DNA helicase MCM2-7 proteins among others. It is interesting that transcripts for MCM3, MCM5 and MCM7 are very abundant in the egg but then decrease substantially by 10 hpf (Fig. S3C). The number of these transcripts should reflect the rates of cell division which are very high for the

first 12 h and then begin to slow as different embryo lineages drop out of the cell cycle.

Maternal transcripts

These data afford a detailed assessment of exactly what transcripts constitute the maternal mRNA stockpile with which the early embryo is equipped. There are ~9800 species of transcripts in the unfertilized egg above the threshold of >300 copies per egg. To examine the functional significance of these maternally expressed genes, we used a binomial test to compare the observed number of active genes with the expected number for each functional class, assuming random representation in the maternal class (Fig. 5). We found that maternal genes related to the cytoskeleton, GTPases, kinases, and protein transportation are the most over-represented classes ($p < 1e-5$). Furthermore, genes involved in more specialized functions, e.g. defensome, immunity, and nervous system related genes, are the most under-represented.

Despite these over- and under-represented classes, transcripts encoding proteins of almost every functional class that we know are presented in the unfertilized egg, which means basic requirements for life and for all cells are covered by the maternal transcripts. Spatially specific transcripts are then provided by the zygotic genome.

We define a gene as exclusively maternal if its expression level peaks in the egg sample (before fertilization), then falls to less than 25% of the peak after mesenchyme blastula stage (24 hpf), and its lowest expression level during embryogenesis is less than

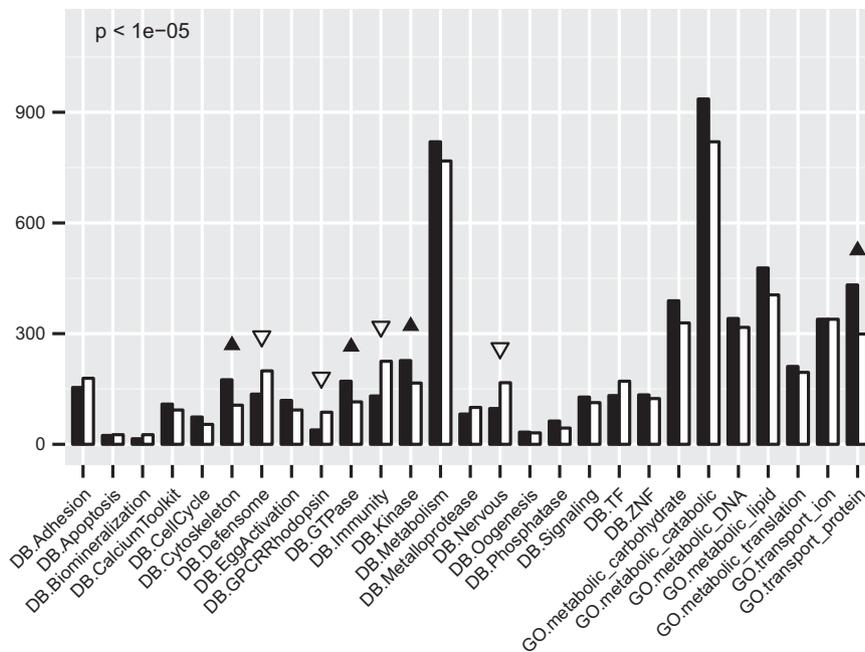


Fig. 5. Gene numbers for different functional classes in the maternal transcriptome. Black bars represent the gene number of the given class observed; white bars represent the gene number of the given class expected according to its percentage of the whole gene set. Black triangles indicate over-representation with significant p values when comparing the observed and expected gene numbers by a binomial test. White triangles indicate under-representation. See also Fig. S4.

300 transcripts per embryo. Only 489 genes belong to this exclusively maternal class. As discovered earlier (Davidson, 1986) most maternally transcribed genes are re-used during the subsequent stages of development. Among the exclusively maternal genes, those related to DNA metabolism (DNA repair, replication, catabolic, etc.) are significantly overrepresented ($p < 0.01$; Fig. S4). This suggests that some special functions related to DNA metabolism need to be executed only in early embryos.

Most of the highly abundant ($> 30,000$ transcripts per embryo) maternal transcription factors (Ets1/2, Otx, Soxb1, Soxb2, Dac, and Hmg2) are well studied. However, a newly identified, un-annotated gene (SPU_018056) encoding a homeobox protein also falls into this group.

Most mRNAs are degraded by exonucleases, and this type of decay pathway consists of four ribonucleolytic activities: decapping, 5'-to-3' exonucleolytic decay, deadenylation, and 3'-to-5' exonucleolytic decay. For some mRNAs, decay is initiated by endonuclease cleavage (Schoenberg and Maquat, 2012). We checked the major mRNA decay components and found that many of them are highly active at very early stages (Table S2, Fig. S3D), suggesting that some of these genes are involved in the rapid clearance of maternal polyadenylated RNAs.

Abundant transcripts

The most abundant transcripts did not escape the attention of molecular developmental biologists working in the time before RNA-seq analysis of transcription was available. Defined as transcripts present at over 200,000 copies per embryo at any single stage (Table S3), there are 32 genes including cyclins, sea urchin specific replication independent histones, translation apparatus components, cytoskeletal, adhesion, metal binding proteins, and so forth. These genes provide an additional perspective on the transcriptional activity of early development. In what follows, we elaborate on several such examples.

The gene *Rrm2* encodes the small chain of ribonucleotide reductase. Since the endogenous dNTP pool is sufficient to support DNA synthesis for only two to three cell divisions (Nemer, 1962),

ribonucleotide reductase is required to maintain DNA synthesis throughout the rapid cleavage stage. The expression profile of *Rrm2* observed in this study is consistent with the enzyme activity previously reported, which increases sharply after fertilization and reaches a peak at the cleavage stage (de Petrocellis and Rossi, 1976).

Three genes, *ClvhH2a*, *ClvhH2b*, and *H2a.Z*, encode for replication independent histones unique to sea urchins and which unlike all major histone mRNAs contain introns and are polyadenylated (Mandl et al., 1997; Marzluff et al., 2006). Transcripts for *ClvhH2a*, *ClvhH2b* are highly abundant in the egg as well as in cleavage stage embryos, while *H2a.Z* peaks at mesenchyme blastula stage, again consistent with what had previously been reported (Ernst et al., 1987).

A set of genes encoding translation apparatus components is very active and these genes show a gradual increasing expression profile. They include six ribosomal protein genes and one elongation factor EF1a. The expression levels of these genes increase concomitantly with the increase in the cell number of the embryo.

Of the four genes encoding metal binding proteins in the list, one encodes a heavy metal binding protein, Metallothionein A (MTA), another one, Ferritin heavy chain (Fth1), binds iron and the last two, *Spec1* and *Spec2a/c*, bind calcium ions. Metallothioneins bind various heavy metals, playing a major role in chemical defense and metal homeostasis. MTA reaches a high level at early blastula stage (24 hpf), when there are about 200,000 transcripts per embryo (Nemer et al., 1991). Ferritin is an iron storage protein, and it is under multiple regulation mechanisms. Ferritin protein is present in the unfertilized egg, but no ferritin transcripts are there. It starts to accumulate from early blastula stage (Infante et al., 1993). The calcium-binding proteins, *Spec1* and *Spec2*, are expressed in the embryonic ectoderm (Carpenter et al., 1984). The transcripts of these genes are present at very low levels in the unfertilized egg, begin to accumulate at the late cleavage to early blastula stage, and at 48 hpf reach levels 100-fold over those in the egg (Bruskin et al., 1981; Tomlinson and Klein, 1990).

Na^+/K^+ ATPase activity increases sharply at blastula stage and beyond (Leong and Manahan, 1997) and this is congruent with the expression profile for Na^+/K^+ ATPase 3 (*Atp1a3*) which reaches a

peak at 18–40 hpf. It is interesting to note that Na⁺/K⁺ ATPase contributes to a major portion of the metabolic activity of the embryo and up to 77% of the metabolism of the larval stage (Leong and Manahan, 1997).

Data access

This dataset has been used in several projects directed at elucidation of gene regulatory networks in our lab, and has proven to be very useful. To facilitate public use of this large dataset, a web application (<http://www.spbase.org:2000/quantdev/>) for data query and visualization was developed using the R/Shiny package. It can be used in answering questions such as the following: What are the expression profiles of the given genes or of all genes in a given functional class? What genes have a similar expression profile to that of a given gene? In which functional class are given genes annotated? What is the detailed annotation information for a given gene in SpBase (the sea urchin genome database)? What are the mRNA, predicted CDS, and predicted peptide sequences of given genes? And what are the genomic structure and mapped reads of given genes?

Users can query by gene names, IDs, function classes and expression profiles. The web application returns results with the annotation information, a link to the corresponding records in SpBase (Cameron et al., 2009), the link to the loci in the genome browser, quantitative measurements for each stage, and sequences, including mRNA, predicted CDS and protein. All these data can be downloaded. Heatmaps and customizable line plots representing the quantitative measurements are generated as well. The web application is intuitive and returns the results in real time. It provides some convenient functions, for example ID auto-detection, so that users can simply cut some text containing gene IDs from other sources and paste directly into the search box. The web application will extract IDs automatically without users' editing.

Users can take advantage of the fuzzy clustering algorithm: start with a given cluster but increase the fuzziness to obtain more profiles that to some degree are similar to the given cluster. As discussed above, commonly used hard clustering algorithms assign a profile to only a single cluster, thus preventing searching for temporal profiles that are less similar. Fuzzy clustering is a unique feature of this application that provides a flexible way to search gene expression profiles.

Combining these features together, the web application provides a powerful interactive tool to explore the dataset. One scenario could be this: a user can initiate a search from a known gene, obtain its expression profile cluster information, then perform another search for that profile cluster with relaxed fuzziness, and finally select transcription factors from the results using the annotation information. Thus by a few simple searches, the user gets all transcription factors with the expression profiles similar to those of the given gene.

Together, this dataset illustrates complex and dynamic gene expression profiles during sea urchin embryonic development. The precise control of gene expression resulting from the complex control circuits of gene regulatory networks that utilize many transcription factors and signal systems explains some of this complexity. However, temporal profiles alone cannot provide satisfactory descriptions in an embryo where the majority of the embryonic gene expression is restricted in space as well as time. A genome scale study of gene expression spatial patterns will be the next challenge on the road towards building a global gene regulatory network.

Materials and methods

Computational analysis

The analysis was done based on the sequencing dataset and gene models assembled in an early study (Tu et al., 2012).

Abundance was estimated by Cufflinks (Trapnell et al., 2010). Various versions and parameters were tested and version 1.3.0 with “frag-bias-correct”, “multi-read-correct”, and “compatible-hits-norm” gave the best result using ORFs as the gene model reference. FPKM values were then converted to the unit of transcripts per embryo as described above. For clustering, the abundance values were normalized by the maximum value of the given gene during the embryonic development. Then the relative expression profiles were analyzed by R package mfuzz (2.19) (Futschik and Carlisle, 2005). GO annotation was done by Blast2GO (Götz et al., 2008). All visualization was done in R with the ggplot2 (0.9.3) package (Wickham, 2009). The website was built in R with ggplot2 and shiny (0.1.0) package (<http://www.rstudio.com/shiny/>).

NanoString nCounter assay

The abundance of a set of transcripts was measured using the NanoString nCounter following manufacturer's instructions and a previous study (Materna et al., 2010). For each sample, 5 μ l containing 100 ng of total RNA were used.

Acknowledgments

We thank Dr. Julius C. Barsi for valuable comments on the manuscript. This work was supported by NIH (P40OD010959, P41HD071837, HD037105).

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.ydbio.2013.11.019>.

References

- Bruskin, A.M., Tyner, A.L., Wells, D.E., Showman, R.M., Klein, W.H., 1981. Accumulation in embryogenesis of five mRNAs enriched in the ectoderm of the sea urchin pluteus. *Dev. Biol.* 87, 308–318.
- Cameron, R.A., Samanta, M., Yuan, A., He, D., Davidson, E.H., 2009. SpBase: the sea urchin genome database and web site. *Nucleic Acids Res.* 37, D750–4.
- Carpenter, C.D., Bruskin, A.M., Hardin, P.E., Keast, M.J., Anstrom, J., Tyner, A.L., Brandhorst, B.P., Klein, W.H., 1984. Novel proteins belonging to the troponin C superfamily are encoded by a set of mRNAs in sea urchin embryos. *Cell* 36, 663–671.
- Davidson, E.H., 1986. *Gene Activity in Early Development*, 3rd ed. Academic Press, Orlando, FL.
- de Petrocellis, B., Rossi, M., 1976. Enzymes of DNA biosynthesis in developing sea urchins. Changes in ribonucleotide reductase, thymidine, and thymidylate kinase activities. *Dev. Biol.* 48, 250–257.
- Ernst, S.G., Miller, H., Brenner, C.A., Nocente-McGrath, C., Francis, S., McIsaac, R., 1987. Characterization of a cDNA clone coding for a sea urchin histone H2A variant related to the H2A.F/Z histone protein in vertebrates. *Nucleic Acids Res.* 15, 4629–4644.
- Fernandez-Guerra, A., Aze, A., Morales, J., Mulner-Lorillon, O., Cosson, B., Cormier, P., Bradham, C., Adams, N., Robertson, A.J., Marzluff, W.F., Coffman, J.A., Genevière, A.-M., 2006. The genomic repertoire for cell cycle control and DNA metabolism in *S. purpuratus*. *Dev. Biol.* 300, 238–251.
- Flytzanis, C.N., Brandhorst, B.P., Britten, R.J., Davidson, E.H., 1982. Developmental patterns of cytoplasmic transcript prevalence in sea urchin embryos. *Dev. Biol.* 91, 27–35.
- Futschik, M.E., Carlisle, B., 2005. Noise-robust soft clustering of gene expression time-course data. *J. Bioinform. Comput. Biol.* 3, 965–988.
- Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H. P., Ferree, S., George, R.D., Grogan, T., James, J.J., Maysuria, M., Mitton, J.D., Oliveri, P., Osborn, J.L., Peng, T., Ratcliffe, A.L., Webster, P.J., Davidson, E.H., Hood, L., Dimitrov, K., 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* 26, 317–325.
- Goustin, A.S., Wilt, F.H., 1981. Protein synthesis, polyribosomes, and peptide elongation in early development of *Strongylocentrotus purpuratus*. *Dev. Biol.* 82, 32–40.
- Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., Conesa, A., 2008. High-throughput functional

- annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435.
- Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., Brown, J.B., Cherbas, L., Davis, C.A., Dobin, A., Li, R., Lin, W., Malone, J.H., Mattiuzzo, N.R., Miller, D., Sturgill, D., Tuch, B.B., Zaleski, C., Zhang, D., Blanchette, M., Dudoit, S., Eads, B., Green, R.E., Hammonds, A., Jiang, L., Kapranov, P., Langton, L., Perrimon, N., Sandler, J.E., Wan, K.H., Willingham, A., Zhang, Y., Zou, Y., Andrews, J., Bickel, P.J., Brenner, S. E., Brent, M.R., Cherbas, P., Gingeras, T.R., Hoskins, R.A., Kaufman, T.C., Oliver, B., Celniker, S.E., 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471, 473–479.
- Howard-Ashby, M.L., Materna, S.C., Brown, C.T., Tu, Q., Oliveri, P., Cameron, R.A., Davidson, E.H., 2006. High regulatory gene use in sea urchin embryogenesis: implications for bilaterian development and evolution. *Dev. Biol.* 300, 27–34.
- Infante, A.A., Infante, D., Rimland, J., 1993. Ferritin gene expression is developmentally regulated and induced by heat shock in sea urchin embryos. *Dev. Genet.* 14, 58–68.
- Leong, P., Manahan, D., 1997. Metabolic importance of Na⁺/K⁺-ATPase activity during sea urchin development. *J. Exp. Biol.* 200, 2881–2892.
- Mandl, B., Brandt, W.F., Superti-Furga, G., Graninger, P.G., Birnstiel, M.L., Busslinger, M., 1997. The five cleavage-stage (CS) histones of the sea urchin are encoded by a maternally expressed family of replacement histone genes: functional equivalence of the CS H1 and frog H1M (B4) proteins. *Mol. Cell Biol.* 17, 1189–1200.
- Marzluff, W.F., Sakallah, S., Kelkar, H., 2006. The sea urchin histone gene complement. *Dev. Biol.* 300, 308–320.
- Materna, S.C., Nam, J., Davidson, E.H., 2010. High accuracy, high-resolution prevalence measurement for the majority of locally expressed regulatory genes in early sea urchin development. *Gene Expr. Patterns* 10, 177–184.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628.
- Nemer, M., 1962. Characteristics of the utilization of nucleosides by embryos of *Paracentrotus lividus*. *J. Biol. Chem.* 237, 143–149.
- Nemer, M., Thornton, R.D., Stuebing, E.W., Harlow, P., 1991. Structure, spatial, and temporal expression of two sea urchin metallothionein genes, SpMTB1 and SpMTA. *J. Biol. Chem.* 266, 6586–6593.
- Oliveri, P., Tu, Q., Davidson, E.H., 2008. Global regulatory logic for specification of an embryonic cell lineage. *Proc. Natl. Acad. Sci. USA* 105, 5955–5962.
- Peter, I.S., Davidson, E.H., 2009. Modularity and design principles in the sea urchin embryo gene regulatory network. *FEBS Lett.* 583, 3948–3958.
- Peter, I.S., Faure, E., Davidson, E.H., 2012. Predictive computation of genomic logic processing functions in embryonic development. *Proc. Natl. Acad. Sci. USA* 109, 16434–16442.
- Schoenberg, D.R., Maquat, L.E., 2012. Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.* 13, 246–259.
- Sea Urchin Genome Sequencing Consortium, Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., Burke, R.D., Coffman, J.A., Dean, M., Elphick, M.R., Etensohn, C.A., Foltz, K.R., Hamdoun, A., Hynes, R.O., Klein, W.H., Marzluff, W., McClay, D.R., Morris, R.L., Mushegian, A., Rast, J.P., Smith, L.C., Thorndyke, M.C., Vacquier, V.D., Wessel, G.M., Wray, G., Zhang, L., Elsik, C.G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M.J., Mackey, A.J., Maglott, D., Panopoulou, G., Poustka, A.J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C.A., Worley, K., Durbin, K.J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M.L., Jackson, A.R., Milosavljevic, A., Tong, M., Killian, C.E., Livingston, B.T., Wilt, F.H., Adams, N., Bellé, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Genevière, A.-M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A.J., Goldstone, J.V., Cole, B., Epel, D., Gold, B., Hahn, M.E., Howard-Ashby, M.L., Scally, M., Stegeman, J.J., Allgood, E.L., Cool, J., Judkins, K.M., McCafferty, S.S., Musante, A.M., Obar, R.A., Rawson, A.P., Rossetti, B.J., Gibbons, I.R., Hoffman, M.P., Leone, A., Istrail, S., Materna, S.C., Samanta, M.P., Stolc, V., Tongprasit, W., Tu, Q., Bergeron, K.-F., Brandhorst, B.P., Whittle, J., Berney, K., Bottjer, D.J., Calestani, C., Peterson, K., Chow, E., Yuan, Q.A., Elhaik, E., Graur, D., Reese, J.T., Bosdet, I., Heesun, S., Marra, M.A., Schein, J., Anderson, M.K., Brockton, V., Buckley, K.M., Cohen, A.H., Fugmann, S.D., Hibino, T., Loza-Coll, M., Majeske, A.J., Messier, C., Nair, S.V., Pancer, Z., Terwilliger, D.P., Agca, C., Arboleda, E., Chen, N., Churcher, A.M., Hallböök, F., Humphrey, G.W., Idris, M.M., Kiyama, T., Liang, S., Mellott, D., Mu, X., Murray, G., Olinski, R.P., Raible, F., Rowe, M., Taylor, J.S., Tessmar-Raible, K., Wang, D., Wilson, K.H., Yaguchi, S., Gaasterland, T., Galindo, B.E., Gunaratne, H.J., Juliano, C., Kinukawa, M., Moy, G.W., Neill, A.T., Nomura, M., Raisch, M., Reade, A., Roux, M.M., Song, J.L., Su, Y.-H., Townley, I.K., Voronina, E., Wong, J.L., Amore, G., Branno, M., Brown, E.R., Cavalieri, V., Duboc, V., Duloquin, L., Flytzanis, C., Gache, C., Lapraz, F., Lepage, T., Locascio, A., Martinez, P., Matassi, G., Matranga, V., Range, R., Rizzo, F., Röttinger, E., Beane, W., Bradham, C., Byrum, C., Glenn, T., Hussain, S., Manning, G., Miranda, E., Thomason, R., Walton, K., Wikramanayke, A., Wu, S.-Y., Xu, R., Brown, C.T., Chen, L., Gray, R.F., Lee, P.Y., Nam, J., Oliveri, P., Smith, J., Muzny, D., Bell, S., Chacko, J., Cree, A., Curry, S., Davis, C., Dinh, H., Dugan-Rocha, S., Fowler, J., Gill, R., Hamilton, C., Hernandez, J., Hines, S., Hume, J., Jackson, L., Jolivet, A., Kovar, C., Lee, S., Lewis, L., Miner, G., Morgan, M., Nazareth, L.V., Okwuonu, G., Parker, D., Pu, L.-L., Thorn, R., Wright, R., 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314, 941–952.
- Tomlinson, C.R., Klein, W.H., 1990. Temporal and spatial transcriptional regulation of the aboral ectoderm-specific *Spec* genes during sea urchin embryogenesis. *Mol. Reprod. Dev.* 25, 328–338.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
- Tu, Q., Cameron, R.A., Worley, K.C., Gibbs, R.A., Davidson, E.H., 2012. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res.* 22, 2079–2087.
- Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*, 1st ed. Springer, New York.
- Wilt, F.H., 1970. The acceleration of ribonucleic acid synthesis in cleaving sea urchin embryos. *Dev. Biol.* 23, 444–455.