# The temporal dynamics of the sea urchin regulome

Roberto Feuda*,‡

## ABSTRACT

In this work, we used Nanostring N-counter technology, to evaluate the mRNA expression level of more than 330 regulatory genes over 34 time points covering the first three days of development of the sea urchin larvae. The hierarchical clustering of the mRNAs expression levels has identified groups corresponding to the major developmental landmarks (e.g. maternal to zygotic transition and gastrulation). Furthermore, comparison with previous experiments indicates high reproducibility of mRNA level temporal dynamics across batches. Finally, we generated an online tool to visualise gene expression during sea urchin larval development. The site can be accessed at and https://www621.lamp.le.ac.uk/nanostring_app/nanostring/.

## INTRODUCTION

The larvae of sea urchin *Strongylocentrotus purpuratus* represent one of the best-studied organisms with a well-established gene regulatory networks (GRN) up to early gastrula embryo [30 h post fertilisation (hpf) see for example Li et al., 2014; Oliveri et al., 2008; Peter and Davidson, 2011, 2015]. However, several events crucial for the development of the larval body plan (e.g. the neuronal differentiation) happen in the post-gastrula stages where information on the mRNAs level is scant. There are about 280 transcription factors encoded in *Strongylocentrotus purpuratus* genome (Howard-Ashby et al., 2006) and previous works have focused only on a few genes or at few developmental time points. Materna and colleagues analysed the mRNA expression level for 138 spatially restricted regulatory genes in early sea urchin development from fertilised egg to 48 hpf (Materna et al., 2010). Subsequently, Tu and collaborators sequenced the developmental transcriptome using mRNA-seq (Tu et al., 2014) for 10 timepoints but only three were from stages after 40 hpf. A comprehensive, high-density description of mRNA expression for the encoded transcription factors is crucial to identify genes involved in the establishment of the larval morphology. Furthermore, the dynamics in the mRNAs level can be used to infer potential regulatory interactions and the overall developmental hierarchy, the first critical step in the identification of the GRNs.

In this work, we measured the level of expression of 335 regulatory genes from zygote to 3 days of development, with a 2-h interval for a total of 34 time points. We used Nanostring N-counter Technology that represents a middle-range technology where hundreds of genes can be measured simultaneously yet maintaining a high level of precision. Compared to other technologies, such as mRNA-seq and qPCR, the Nanostring N-counter Technology does not require a library or enzymes, reducing substantially the possible biases (Geiss et al., 2008). In brief, the technology is based on the hybridisation of two probes, a capture and reporter probe, to each target transcript and the number of hybridisation events is quantified using an automated fluorescent microscope (Geiss et al., 2008).

To study the mRNA temporal dynamics, we used hierarchical clustering and identified seven clusters that correlate with morphological events happening during sea urchin larval development). The data suggest that the level of expression of regulatory genes is sufficient to discriminate between developmental stages. Next, we compared our data with the previous observation from Materna et al., 2010, and identified a high level of reproducibility for the mRNA expression level. Finally, to allow accessibility to the data, we developed an online tool that can be used to visualise the mRNA level of expression. The developmental time course can be found in a searchable database that is accessible at https://www621.lamp.le.ac.uk/nanostring_app/nanostring/.

## RESULTS AND DISCUSSION
### Developmental time explains the majority of the observed variation

To quantify the variation between the two biological replicates, we performed a PCA using ClusVis (Metsalu and Vilo, 2015). The results suggest that the majority of variance (63.8%) is explained by developmental time (Fig. 1). In all 34 points, the two biological replicates are clustered together, suggesting a high level of similarity between replicates.

To clarify the temporal dynamics in gene expression, we performed a clustering analysis using ClusVis (Metsalu and Vilo, 2015). The results in Fig. 2 indicate the existence of multiple distinct clusters that correspond to well-defined developmental stages defined based on embryological features. The first cluster confirms the results of the PCA and identifies a cluster that includes all the samples from 0 to 8 hpf that are consistent with previous data from transcriptional kinetics and suggest that the activation of the zygotic genome happens after 5-6 hpf (Gildor et al., 2016). The remaining time points are organised in six distinct clusters that reflect the morphological events happening during sea urchin larval development. For example, one cluster comprises 10 to 18 hpf that encompasses the majority of the blastula stage. A similar pattern is observed for later stages where our clustering analysis identifies a sample cluster from 20 hpf to 26 hpf representing the mesenchyme blastula. These results indicate the expression of regulatory genes contains enough statistical power to identify well-defined developmental stages. Furthermore, these results indicate that classic morphological changes happening during sea urchin development are the results of the changes in underlying expression of regulatory genes.

Division of Biology and Biological Engineering, MC156-29, California Institute of Technology, Pasadena, CA 91125, USA.
*Present address: Department of Genetics and Genome Biology, University of Leicester, UK.

‡Author for correspondence (rf190@leicester.ac.uk)
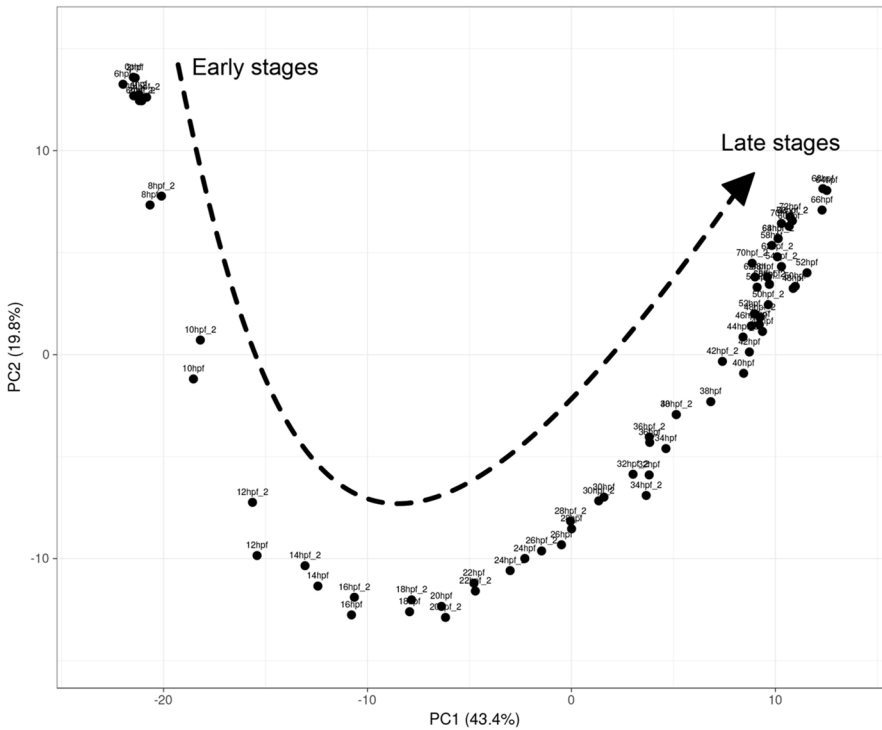
R.F., 0000-0003-0857-1732

**Fig. 1. Principal component analysis of the temporal expression data.** PCA analysis of Nanostring dataset from 0 to 72 hpf (*n*=2 replicates per time point). The 63.8% of variance is explained by the gene expression temporal dynamics.

## High reproducibility of temporal dynamics of regulatory gene expression

Finally, we tested the reproducibility of our developmental time course. To this aim, we compared our expression profile with Nanostring data obtained by Materna and collaborators (2010). In this case, the authors evaluated the level of expression to sea urchin development for 138 regulatory genes for the 48 h of development. Differently from this work, Materna and collaborators used spike-GFP and RFP to normalise the data. To compare the datasets, we selected eight genes that have a different level of expression, from
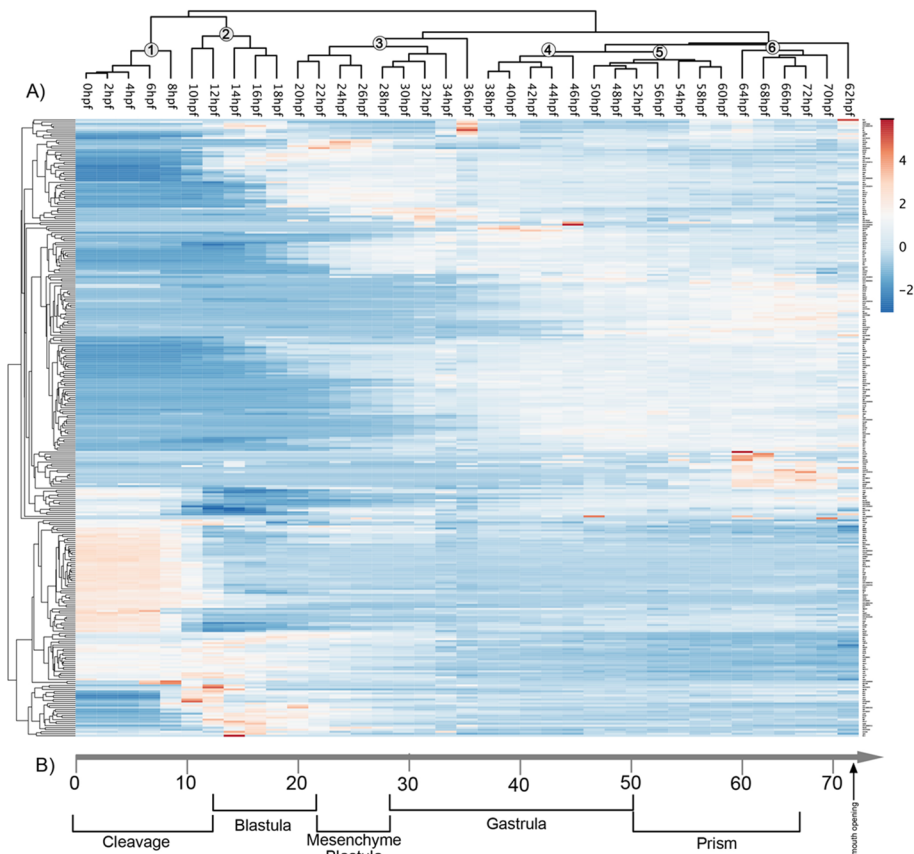


**Fig. 2. Heat map of gene expression.** (A) Heat map of scaled expression for average mRNA expression level for 335 genes averages obtained from two replicates from 0 to 72 hpf (34-time points) of sea urchin larval development. The results indicate the existence of the X clusters that reflect the developmental transition during sea urchin embryonic development (lower part of the figure). (B) Developmental stages of *S. purpuratus* embryos at 15°C. Numbers indicate the six clusters identified.

100 to 20,000 copies of mRNA per embryo. Comparison (Fig. 3) shows that results are highly consistent and reproducible, despite using different instruments, conditions, and animal batches. This indicates that the developmental progression of genome regulation is tightly controlled.

### Data accessibility

To visualise the mRNA expression for all the genes, we have created a visualisation tool available at https://www621.lamp.le.ac.uk/nanostring_app/nanostring/. A screenshot is shown in Fig. 4. With this tool, it is possible visualise the mRNA expression level for the 335 regulatory genes analysed in this work. A table with all expression data is available as supplementary data.

### Conclusions

By describing the developmental dynamics, the data presented in this work might inform future functional perturbations. The high-density sampling and the inclusion of potentially all the known transcription factors data can be used to reconstruct the temporal developmental hierarchy between the regulatory genes, and together with spatial data, inform the functional perturbation.
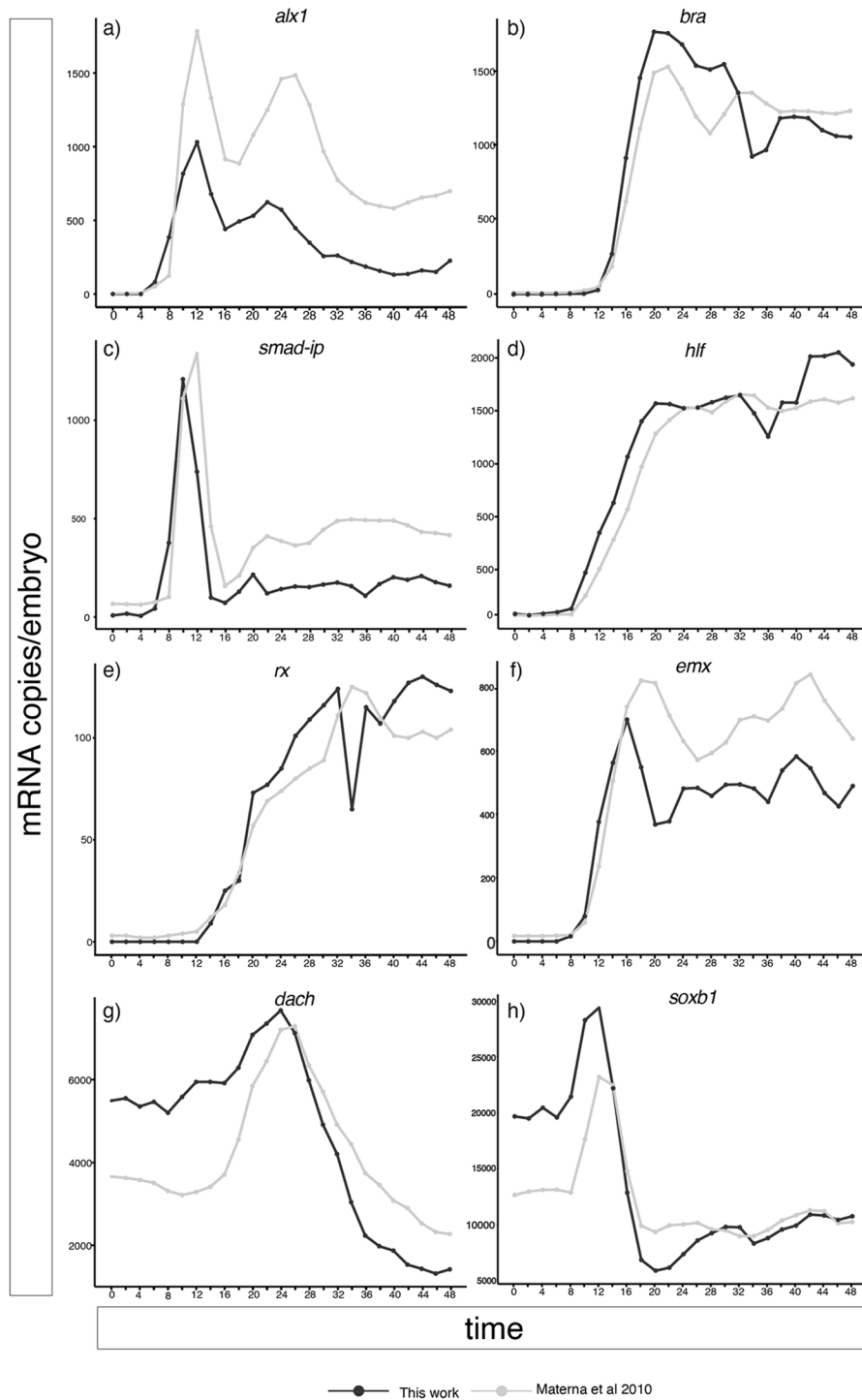


**Fig. 3. Reproducibility of mRNA measurements.** Line plots obtained comparing mRNA expression from this work and those from Materna and collaborators (2010). The results indicate a high level of reproducibility of Nanostring measurements as well as of the transcriptional hierarchy.
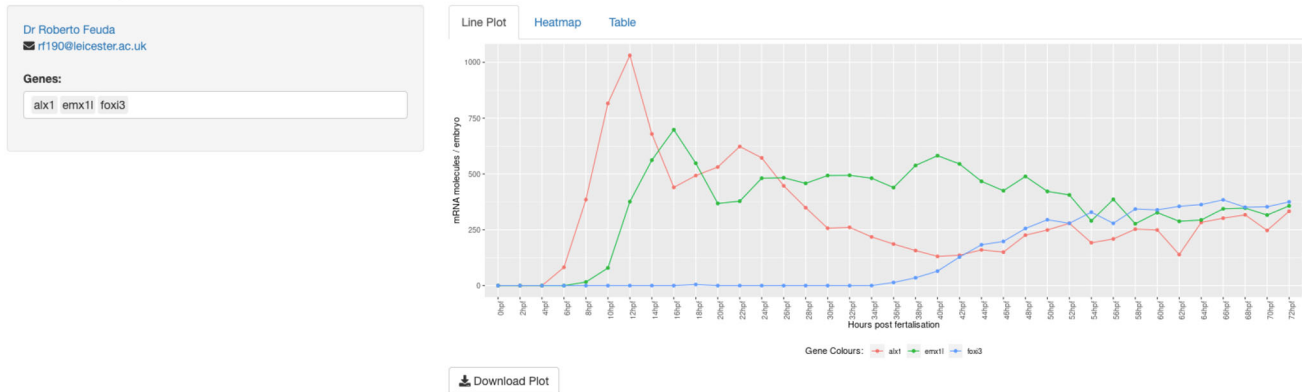
## Nanostring Data Visualisation



**Fig. 4. Snapshot of the expression data.** The webpage allows for the visualisation of up to 100 genes simultaneously both using heatmap or line plot.

## MATERIALS AND METHODS

### Code set
We designed a probe set containing 335 genes covering most of the regulatory genes expressed during the development of *Strongylocentrotus purpuratus* (see Table S1 for the sequences). This includes transcription factors and other modulators of gene expression such as WNT and FGF signaling and a transcriptional cofactor (see Table S2). The code set was designed using previous information from Materna et al. (2010) and using gene models predicted by Tu et al. (2014). The genes were classified using EggNog Mapper (Huerta-Cepas et al., 2019), and BLASTP against the Metazoan transcription factor database (Hu et al., 2019) (see Table S2 for the classification).

### Embryo culture and RNA extraction
Sea urchin embryos were fertilised in filtered seawater and cultured at 15°C. Every other hour ∼300 embryos were counted and lysed in 350 µl of a solution of RLT buffer and β-mercaptoethanol from the Qiagen RNeasy Micro Kit (Qiagen, Hilden, Germany). The lysates were immediately stored at −80°C until use. The RNA was extracted according to the manufacturer's instructions but, similarly to Materna et al. (2010), to maximise recovery, RNA was eluted with 100 µl nuclease-free water at 60°C. The samples were ethanol precipitated and resuspended in 7 µl nuclease-free water, five of which were used in the following NanoString hybridisation.

### Nanostring nCounter assays
For each individual time point the transcript count was measured using the NanoString nCounter. Hybridisation reactions were performed according to the manufacturer's instructions with 5 µl RNA solution. Care was taken to minimise the time after the addition of the capture probe set to minimise background due to non-specific interactions between detection probes and capture probes. All hybridisation reactions were incubated at 65°C for a minimum of 18 h. Hybridised probes were recovered with the NanoString Prep Station and immediately evaluated with the NanoString nCounter. For each reaction, 600 fields of view were counted.

### mRNA quantification
We performed two biological replicates, and the mRNA level was quantified as follows. First, for each sample, the background correction was performed individually by removing the total sum negative control using NanoStringNorm (Waggott et al., 2012). This step removes aspecific probe binding. Second, to account for differences in hybridisation efficiency, we estimated the mRNA counts using the Nanostring internal spikes. We took advantage of the six Nanostring exogenous mRNA spikes-in those covers from 0.125 femtomolar to 125 femtomolar. We converted this into mRNA molecules and estimated the mRNA level for each sample by comparing gene counts and mRNA spikes-in counts. While these steps ameliorate the differences in the hybridisation events, they do not correct the differences between samples (e.g. different number of embryos) that can affect the reconstruction of the time-course. We combined the correct counts

for all time points and used the total counts to account for the variation between time points. Finally, the average between the mRNA was used to combine the two replicates.

### Competing interests
The authors declare no competing or financial interests.

### Author contributions
Conceptualization: R.F.; Methodology: R.F.; Software: R.F.; Validation: R.F.; Formal analysis: R.F.; Investigation: R.F.; Resources: R.F.; Data curation: R.F.; Writing - original draft: R.F.; Writing - review & editing: R.F.; Visualization: R.F.; Supervision: R.F.; Project administration: R.F.; Funding acquisition: R.F.

### References
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. et al. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317-325. doi:10.1038/nbt1385

Gildor, T., Malik, A., Sher, N. and Ben-Tabou de-Leon, S. (2016). Mature maternal mRNAs are longer than zygotic ones and have complex degradation kinetics in sea urchin. *Dev. Biol.* **414**, 121-131. doi:10.1016/j.ydbio.2016.04.007

Howard-Ashby, M., Materna, S. C., Brown, C. T., Tu, Q., Oliveri, P., Cameron, R. A. and Davidson, E. H. (2006). High regulatory gene use in sea urchin embryogenesis: implications for bilaterian development and evolution. *Dev. Biol.* **300**, 27-34. doi:10.1016/j.ydbio.2006.10.016

Hu, H., Miao, Y.-R., Jia, L.-H., Yu, Q.-Y., Zhang, Q. and Guo, A.-Y. (2019). AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, D33-D38. doi:10.1093/nar/gky822

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J. et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309-D314. doi:10.1093/nar/gky1085

Li, E., Cui, M., Peter, I. S. and Davidson, E. H. (2014). Encoding regulatory state boundaries in the pregastrular oral ectoderm of the sea urchin embryo. *Proc. Natl. Acad. Sci. USA* **111**, E906-E913. doi:10.1073/pnas.1323105111

Materna, S. C., Nam, J. and 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. doi:10.1016/j.gep.2010.04.002

**Metsalu, T. and Vilo, J.** (2015). ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **43**, W566-W570. doi:10.1093/nar/gkv468

**Oliveri, P., Tu, Q. and Davidson, E. H.** (2008). Global regulatory logic for specification of an embryonic cell lineage. *Proc. Natl. Acad. Sci. USA* **105**, 5955-5962. doi:10.1073/pnas.0711220105

**Peter, I. S. and Davidson, E. H.** (2011). A gene regulatory network controlling the embryonic specification of endoderm. *Nature* **474**, 635-639. doi:10.1038/nature10100

**Peter, I. and Davidson, E. H.** (2015). *Genomic Control Process: Development and Evolution*. Academic Press.

**Tu, Q., Cameron, R. A. and Davidson, E. H.** (2014). Quantitative developmental transcriptomes of the sea urchin Strongylocentrotus purpuratus. *Dev. Biol.* **385**, 160-167. doi:10.1016/j.ydbio.2013.11.019

**Waggott, D., Chu, K., Yin, S., Wouters, B. G., Liu, F.-F. and Boutros, P. C.** (2012). NanoStringNorm: an extensible R package for the pre-processing of NanoString mRNA and miRNA data. *Bioinformatics* **28**, 1546-1548. doi:10.1093/bioinformatics/bts188