

## TECHNIQUES AND RESOURCES

## RESEARCH ARTICLE

# A single cell RNA sequencing resource for early sea urchin development

Stephany Foster, Nathalie Oulhen and Gary Wessel\*

## ABSTRACT

Identifying cell states during development from their mRNA profiles provides insight into their gene regulatory network. Here, we leverage the sea urchin embryo for its well-established gene regulatory network to interrogate the embryo using single cell RNA sequencing. We tested eight developmental stages in *Strongylocentrotus purpuratus*, from the eight-cell stage to late in gastrulation. We used these datasets to parse out 22 major cell states of the embryo, focusing on key transition stages for cell type specification of each germ layer. Subclustering of these major embryonic domains revealed over 50 cell states with distinct transcript profiles. Furthermore, we identified the transcript profile of two cell states expressing germ cell factors, one we conclude represents the primordial germ cells and the other state is transiently present during gastrulation. We hypothesize that these cells of the Veg2 tier of the early embryo represent a lineage that converts to the germ line when the primordial germ cells are deleted. This broad resource will hopefully enable the community to identify other cell states and genes of interest to expose the underpinning of developmental mechanisms.

**KEY WORDS:** Germ cells, Nanos, Sea urchin, Single cell mRNA-seq

## INTRODUCTION

How cells of a developing organism change relative to each other is key to understanding embryogenesis. Important regulatory steps for developmental change may include various intersections with the central dogma of molecular biology, so testing diverse organisms is essential for revealing conserved principles in development, and nodes in which evolutionary change has occurred. Although the functional end product of gene expression is the driving force of change, whether it is a protein, metabolite or other biochemistry, it is usually very difficult to see such specific activities with current technologies. Instead, investigators focus on DNA and RNA polymers because they can be analyzed specifically through their sequences and amplified for remarkable sensitivity.

Single cell RNA sequencing (scRNA-seq) technology combines advances in microfluidics and nucleic acid biochemistry to identify genes expressed within cells on a cell-by-cell basis (Klein et al., 2015; Macosko et al., 2015). By capturing mRNAs through oligo-dT annealing on a bead within an oil droplet, one can rapidly construct cDNA with individualized barcoding so that upon mixing the many beads, each representing a single cell, the investigator can

back-calculate the sequenced mRNAs and their relative abundance from each cell. By then comparing the overall content and abundance of mRNAs in each cell, clusters are parsed out into more or less similarity, as distinguished by their steady-state transcript accumulation from which gene expression is concluded. The strengths of such technology include readouts of mRNAs from potentially thousands of cells of a single embryo for direct, side-by-side comparison of the constituents. Shortcomings of the approach are that spatial interpretation requires known markers by *in situ* hybridization, and less depth in sequencing than in traditional bulk RNA-seq.

Here, we take advantage of the ability to dissociate sea urchin embryos into single cells to parse out cell states based on patterns of mRNA accumulation using scRNA-seq technologies. With this approach, one can identify distinctions in genes expressed in each cell state, thus enabling greater depth in analysis of gene regulatory networks. Echinoderms are the lone invertebrate deuterostome and their members are an important model for establishing gene regulatory networks through experimental templates, to understand the drive in development and differences between cells. The ready dissociation of echinoderm embryos into single cells has been used in many applications to identify unique gene expression (Bruskin et al., 1981; McClay and Marchase, 1979), specific cell adhesion mechanisms (Fink and McClay, 1985), cell fate changes (McClay et al., 1977) and many other important principles of development. Recently, a proof of concept experiment was reported using sea urchin embryos analyzed by single cell mRNA sequencing (scmRNA-seq) (Foster et al., 2019). Two sibling populations of dissociated cells were analyzed at one time point (early gastrula), the only difference being the presence or absence of an inhibitor (DAPT) of the Delta/Notch pathway during development and prior to dissociation. Remarkably, all cell clusters were overlapped in these two populations except for one lineage, the Veg2 lineage that contributes to endomesoderm. This lineage is known to be a target of the Delta/Notch pathway during development. The reproducibility of these results encouraged us to perform a broad and inclusive analysis of cells during development from cleavage stages to late in gastrulation. These developmental time points have been most intensively studied and form the basis for the gene regulatory network in this animal [<http://www.echinobase.org/endomes/>].

Of particular interest to us are the primordial germ cells (PGCs), the stem cells that give rise to eggs and sperm. These cells are formed in the sea urchin at the 32-cell stage, after two sequential asymmetric divisions. These presumed PGCs promptly differ significantly from their sibling somatic cells. For example, although the somatic cells continue to divide rapidly, the four PGCs become mitotically quiescent shortly after they are formed and divide only once until after gastrulation, resulting in an embryo of eight PGCs versus over 1000 somatic cells. In addition to downregulation of their cell cycle, the PGCs also reduce their

Department of Molecular and Cellular Biology, Division of BioMedicine, Brown University, Providence, RI 02912, USA.

\*Author for correspondence (rhet@brown.edu)

ID G.W., 0000-0002-1210-9279

Handling Editor: Susan Strome  
Received 14 April 2020; Accepted 31 July 2020

transcription, translation and mitochondrial activity to less than 10% of their somatic counterparts. We discovered that this dramatic quiescence is temporary; the PGCs restart their activities after gastrulation. These transitions serve mightily as a rapid, predictable and transient quiescent phenotype although their transcript profile during this period is not known. In the sea urchin, the RNA-binding protein, Nanos2, is essential for maintaining PGC quiescence and survival, and the turnover of Nanos2 in these cells is correlated with a return to full metabolic activity (Fujii et al., 2009; Juliano et al., 2010; Oulhen et al., 2017).

Here, we report on the changes in cell state from the 8-cell stage to late gastrulae using scmRNA-seq technology in the embryo of *Strongylocentrotus purpuratus*. We use the term ‘cell states’ here instead of ‘cell types’ for the compelling reason that technology has exceeded observational distinctions (Klein, 2017). Distinctions in transcript profiles outnumber the classic definition of cell types, which is necessarily limited when one considers lineage variations and boundaries between lineages. Instead, ‘cell state’ refers to populations of cells that show distinctions in transcript accumulation, which may or may not be reflected in the functional endpoint of the cell. The term ‘cell state’ is also distinct from ‘regulatory state’, the term used for the profile of transcription factors in each cell. We find 22 major cell states in late gastrulae and report on cell states in each of the three primary germ layers and in the cells that we posit form the germ line.

## RESULTS

### Single cell mRNA sequencing of developing sea urchin embryos

Sea urchin embryos were cultured and collected for scRNA-seq analysis via DropSeq using the 10 $\times$  Genomics platform. Our analysis included eight time points spanning early development, from 3h post-fertilization to late gastrula stage (Table S1), selected on the basis of dynamic features of cell fate changes anticipated by candidate-based studies. The datasets from all eight time points were integrated and clustered to identify cell states and cell state markers present across early development using Seurat (Butler et al., 2018; Stuart et al., 2019). In total, the transcriptomes of 60,399 cells were included in our analysis. We identified cell states of the sea urchin arising from the three germ layers: ectoderm, endoderm and mesoderm. Upon integration of the datasets and visualization with t-SNE (t-distributed stochastic neighbor embedding) dimensionality reduction, we identified 22 major cell states, including the germline (Fig. 1A), based on known marker gene expression (Table S2). These 22 major clusters were further segregated to reveal regions within the digestive system, distinct ectoderm cell states and diverse blastocoelar populations (Fig. 1B; Fig. S8). The identification of 22 clusters was a direct result of the resolution parameter used in the clustering step of the analysis. The resolution of 0.5 is a conservative setting, whereas increased resolution resulted in 58 separable cell states (Fig. 1B). We characterized each of the major 22 clusters, which are well annotated cell states of the sea urchin embryo, and because no resolution was optimal for each cell type, we show results from subclustering of each of the major cell populations (Fig. S26).

We first tested consistency in the single cell sequencing approach for this animal. Although the cost of DropSeq runs and sequencing is very high, we were able to test the same species at the same developmental time in two experiments separated by one year. Our first scRNA-seq experiment showed lineage-specific effects of cell signaling by treating the embryo with small molecule inhibitors (Foster et al., 2019). The control in that experiment was replicated here, one year later, and was closely correlated in this new dataset as

shown by cluster positioning, gene markers of each cluster and regression analysis of the transcripts identified in each experiment (Fig. S1). We saw strong correlation between both datasets across the computationally derived clusters, which led us to conclude that this approach is highly replicable. The consistency of output appears to be of high fidelity even with the variations in numbers of cells and depth of sequencing, as well as differences in developmental timing between these wild-type embryos.

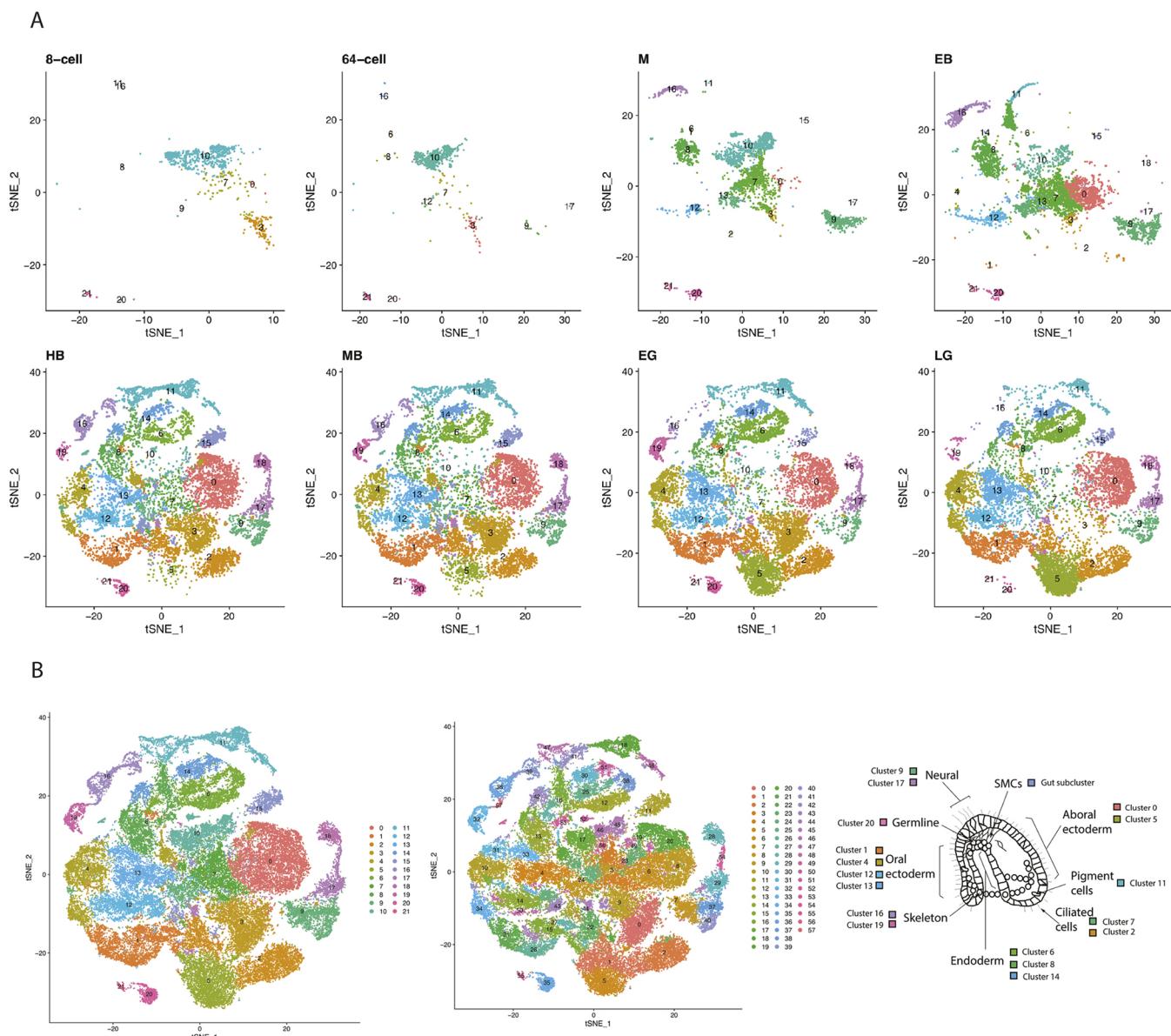
In analyzing the eight developmental datasets, differences between early and late time points became apparent. We detected a range of 2000 to 7000 genes per cell of embryos collected between the 8- and 64-cell stages. In contrast, we detected on average fewer than 2000 genes per cell in blastula and gastrula stage cells (Fig. S2). Overall, we observed a downward trend in the number of unique transcripts detected per cell as development progressed (Fig. S3). This change could be explained by (1) a progressive decrease in mRNA content (cell volume) with embryogenesis, (2) differing depth of sequencing for each time point, (3) an overall decrease in maternal mRNAs present in the early embryo (Swartz et al., 2014), and/or (4) the reduced potency of cells during development and acquired specialization in their steady-state mRNA expression. This later type of mRNA change can be quantitated and, indeed, appears to reflect the change in potency of cells in development (Gulati et al., 2020). Maternal mRNAs can persist in the embryo sometimes until the mesenchyme blastula stage and our data supports the model that developmental progression yields specialization and decreased mRNA diversity. Overall, the number of genes detected per cell by this approach was consistent with other reports using a variety of cell types (Macosko et al., 2015; Siebert et al., 2019). At each stage of sea urchin (*S. purpuratus*) embryogenesis, on average ~11,500 genes are actively transcribed, producing about 39 million transcripts per embryo (Tu et al., 2014). Depending on the stage, each of the many different cell states of the embryo necessarily expresses less than this overall maximum so that the yield of mRNA information from each cell state could range as high as 60%. Thus, although this scmRNA-seq approach does not yield a complete mRNA profile, it is sufficient for distinguishing diverse cell states not otherwise achievable.

### Identification of 22 major cell populations through development to late gastrulae

Cluster visualization by t-SNE plots showed 22 major cell states in late gastrulae. As a test of the stringency of cell state identification, we first identified marker gene expression in the three germ layers: ectoderm, endoderm and mesoderm. The reader is directed to further analysis in the supplementary material (see also Fig. S4–S27) and to the datasets deposited in GEO [GSE149221].

### Ectoderm representatives

We first assessed expression of a broad ectodermal marker *SoxB2*. *SoxB2* expression was present throughout the sea urchin ectoderm and is known to play a role in neurogenesis (Anishchenko et al., 2018; Garner et al., 2016). Expression of *SoxB2* was prominent in our dataset across several clusters at early time points, first in one cluster (10) at the 8-cell stage, later expanding to three more clusters (7, 8, 9). Beginning at the hatched blastula stage, we noted a decrease in the number of cells in clusters 7–10 expressing *SoxB2* (Fig. S4). The sea urchin nervous system consists of neurons arising from the animal pole domain (APD) and the ciliary band, both specialized regions of the ectoderm. Expression of the transcription factor *Foxq2* marks the animal cap and the animal pole domain in the sea urchin embryo (Tu et al., 2006; Wei et al., 2009; Yaguchi et al., 2008). *Foxq2* expression was detected quite early in our



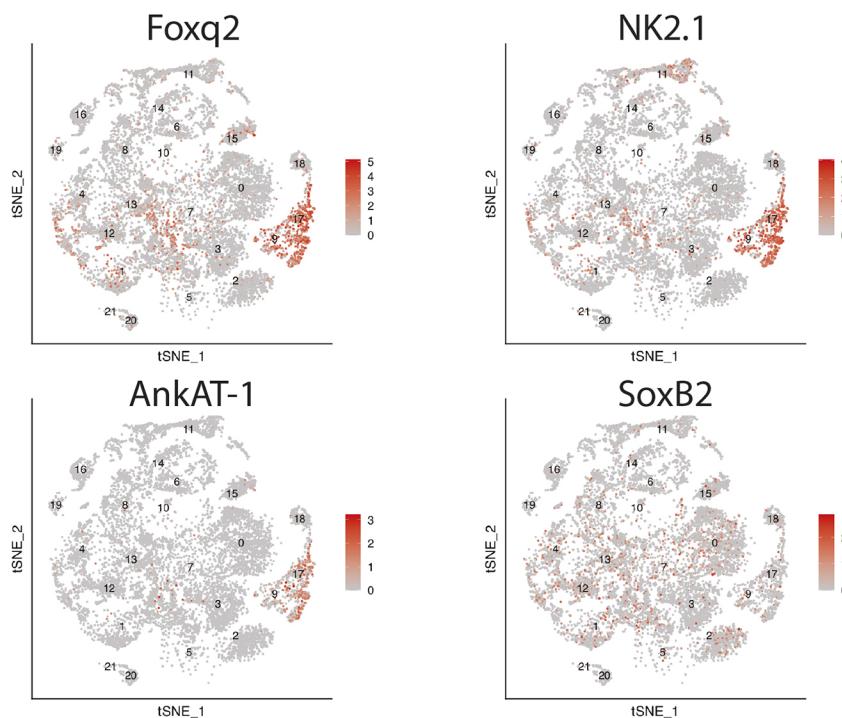
**Fig. 1. t-SNE plots of major developmental stages.** (A) The 8-cell, 64-cell, morula, early blastula, hatched blastula, mesenchyme blastula, early gastrula and late gastrula stages after alignment. A total of 60,399 single cells are colored by cluster identity. Twenty-two cell populations are detected across the eight time points. (B) Higher resolution enables visualization of more cell states present. t-SNE plot of integrated dataset with clustering performed at a resolution of 0.5 (left) shows 22 cell states and a resolution of 3 reveals 58 cell states (center). Cells are colored by cluster identity. Right: Drawing of late gastrula stage sea urchin mapping cell states identified by scRNA-seq to the embryo. Colors match cell states seen at resolution 0.5.

dataset. Significant and enriched expression of *Foxq2* was localized to one cluster (9) at morula stage and this pattern was maintained until late gastrula, with a secondary cluster (17) arising at hatched blastula (Fig. S5). *NK2.1* expression is known to be restricted to the apical ectoderm at blastula stage (Takacs et al., 2004), and its expression was present in the same two clusters where *Foxq2* was enriched. Expression of the *Foxq2* target gene, *AnkAT-1*, overlapped *Foxq2* and *NK2.1* (Fig. 2). *AnkAT-1* was expressed in the animal plate at the blastula stage (Yaguchi et al., 2010). We defined clusters 9 and 17 as neural cell states based on coexpression of *Foxq2*, *NK2.1* and *AnkAT-1* at late gastrula.

#### Endoderm representatives

The endoderm is subject to a number of regulatory steps, which result in specialized regions of the gut. *FoxA*, whose function is

required for invagination of the gut and endoderm specification, is expressed vegetally at the blastula stage and becomes localized to the gut at the gastrula stage (Oliveri et al., 2006). *Endo16*, which is enriched in the endoderm, is first detected via *in situ* hybridization in the vegetal pole of mesenchyme blastula embryos. At the gastrula stage, its expression is explicitly in the gut (Ransick et al., 1993) and serves as an excellent cell-type marker for endodermal cells of the mid- and hindgut. *FoxA* expression is concentrated in one cluster (8) at the early blastula stage and its expression in this cluster remains high through later stages. Beginning at hatched blastula, three clusters maintained high *FoxA* expression (6, 8, 14) with overlapping *Endo16* expression starting at early blastula stage (Figs S6 and S7). In gastrula stages, *FoxA* expression was more dynamic, being present in cell states that did not express *Endo16* (Fig. 3). *FoxA* expression is known to extend to the foregut, whereas



**Fig. 2. Feature plots showing coexpression of neural precursor genes at hatched blastula stage.** Expression of apical ectoderm markers *NK2.1*, *Foxq2* and *AnkAT-1* is enriched in clusters 9 and 17. Neural precursor gene *SoxB2*, known to be expressed throughout the ectoderm, is more broadly expressed. Cells are colored by relative expression level.

*Endo16* does not (Annunziata et al., 2014). Moreover, even though *Endo16* was mostly expressed throughout the mid- and hindguts of the endoderm, these regions are actually composed of different cell states distinguishable by transcript fingerprints. *Endo16*-positive cells at late gastrula (clusters 6, 8, 11, 14) subcluster into seven cell populations. After subclustering, we can distinguish the hind and midgut, the foregut, secondary mesenchyme cells, pigment cells, skeleton cells and neuronal cells (Fig. S8).

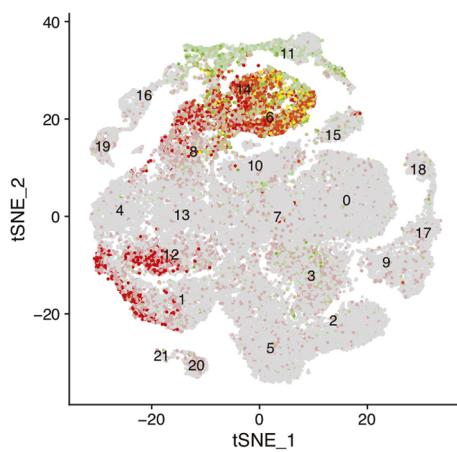
#### Mesoderm representatives

The primary mesenchyme cells (PMCs) are a mesodermally derived cell type that gives rise to the larval skeleton. The PMC lineage starts with asymmetric division of the four micromeres in which the smaller cells that result become the small micromeres (the presumed primordial germ cells) and the larger cells become the PMCs. They divide a limited and consistent number of times, ingress into the

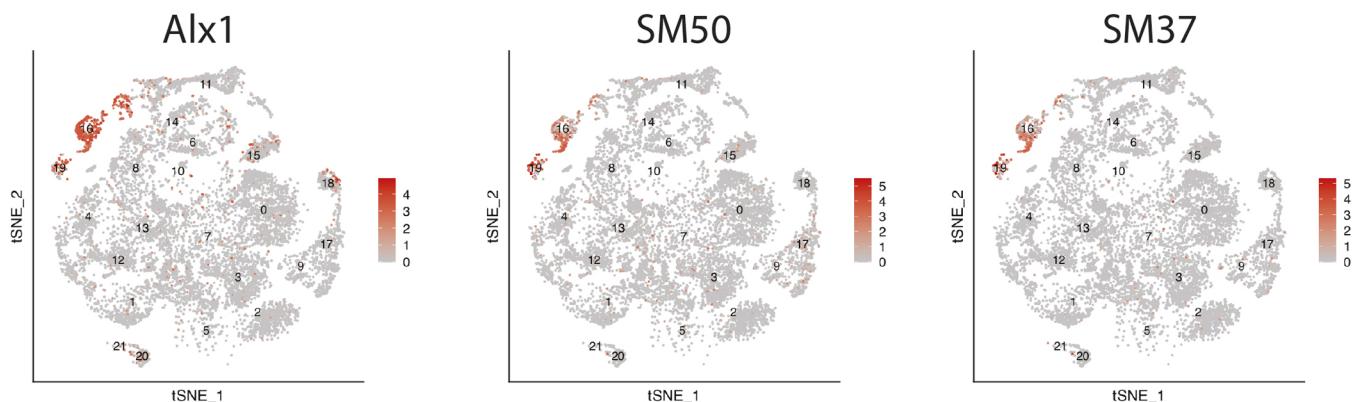
blastocoel, migrate to specific cluster sites, initiate mineralization and then fuse and expand with the growing skeleton. The PMCs were identified by expression of *Alx1* and the spicule matrix protein genes *SM50* and *SM37*. *Alx1* is a homeobox-containing gene known to be expressed in the large micromeres and essential for activation of the many genes responsible for the spicule matrix of the larval skeleton (Ettensohn et al., 2003). *SM37* and *SM50* are transcriptional targets of *Alx1* and are selectively expressed in the PMCs (George et al., 1991; Urry et al., 2000). Low level expression of *Alx1* was first detected by scRNA-seq at the 64-cell stage in cluster 21. At the morula stage, *Alx1* expression was enriched in cluster 16 at a high level. Expression of *Alx1* in this cluster persisted until late gastrula. A second cluster (19) also showed persistent *Alx1* expression from hatched blastula to late gastrula. Enrichment of *SM50* and *SM37* in the same cell clusters as *Alx1* at blastula and later stages suggests that these clusters represent the PMC cells (16, 19) (Fig. 4; Figs S9, S10, S11). At late gastrula, the percentage of cells expressing the PMC markers was greatly reduced. For example, at mesenchyme blastula, clusters 16 and 19 represented 2.6 and 1.9% of total cells, respectively, whereas at the late gastrula stage, these numbers decreased to 0.11 and 0.66%, respectively. We interpret this result as the reduced cell cycle in the PMCs with development relative to other somatic cells, so that the percentage of cells seen by scRNA-seq was decreased. Correlatively, we observed a decreasing level of *Cyclin B* mRNA in both of these clusters over time (Fig. S27).

#### Identification of distinct cell populations expressing the germ cell marker *Nanos2*

The primordial germ cells (PGCs) of the sea urchin are thought to be derived from the small micromeres resulting from asymmetric cell division of the micromeres at the 16- to 32-cell stage transition. *Nanos2* is first expressed in the germline and can be detected by RNA *in situ* hybridization at the 64-cell stage. Later in development, when the embryos start to gastrulate, *Nanos2* mRNA can also be detected in somatic cells derived from the Veg2 lineage. Despite the



**Fig. 3. Feature plot showing expression of FoxA and Endo16 across the dataset.** Coexpression (yellow) of FoxA (red) and Endo16 (green) in clusters 6 and 14.



**Fig. 4. Clusters 16 and 19 are marked by expression of PMC marker genes *Alx1*, *SM50* and *SM37* at the hatched blastula stage.** Cells are colored by relative expression level.

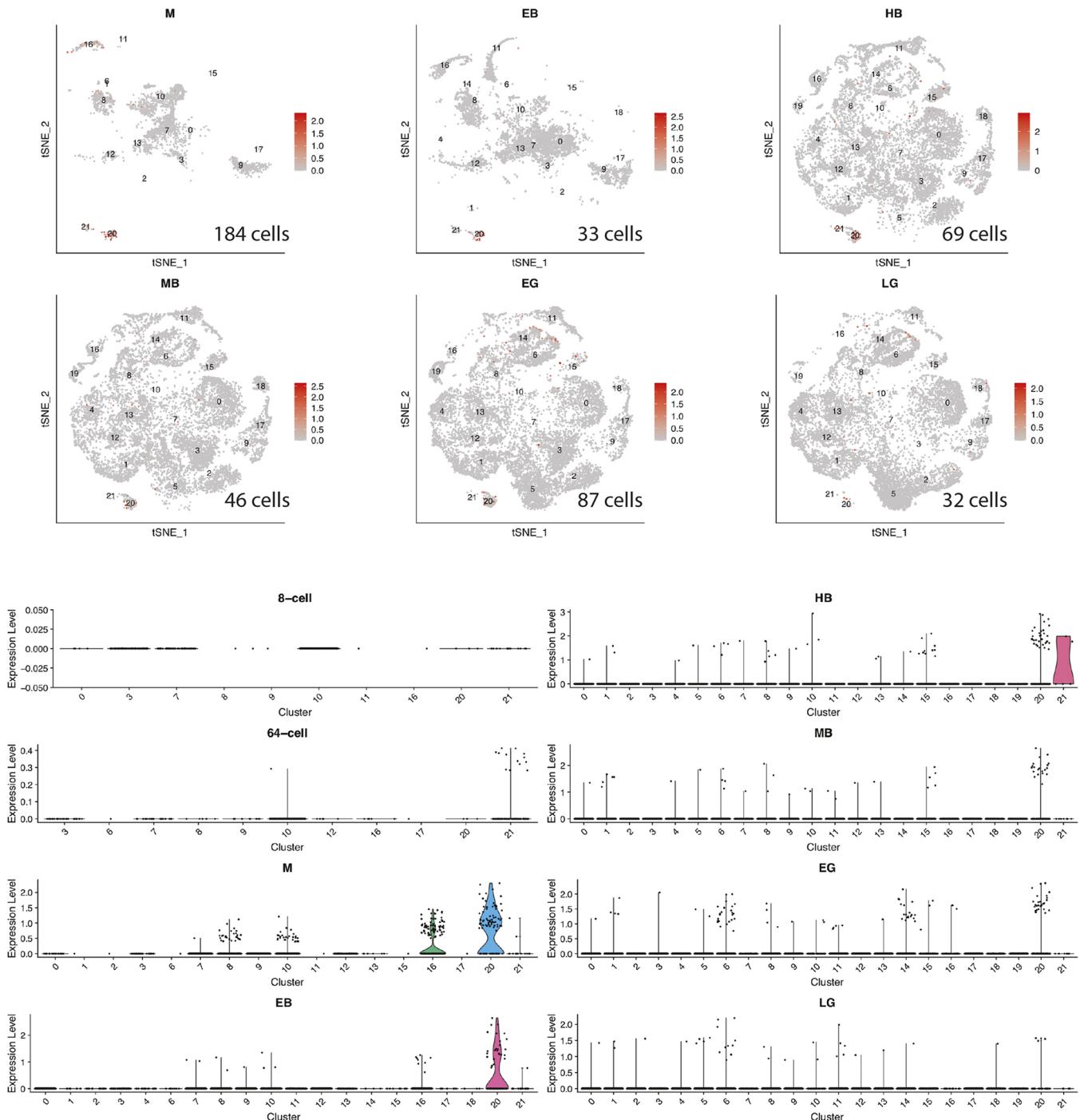
rarity of these two cell types in the embryos, this single cell analysis approach detected distinct *Nanos2*-expressing cells (Fig. 5). *Nanos2* expression was first detected at the 64-cell stage in 14 cells at a low level. At the morula stage, 184 *Nanos2*-expressing cells were captured (0.49% of the population). The number of *Nanos2*-expressing cells detected varied depending on the developmental stage. For example, at late gastrula that number dropped to 32 (0.25% of population). This drop in *Nanos2*-positive cells is expected, because while the embryo grows and develops with rapid cell cycling, the quiescent *Nanos2*-expressing cells (including the germ cells) do not cycle significantly and instead become a rare cell type of the embryo.

*Nanos2* was first detected in cluster 20 at the 64-cell stage and its expression was maintained in this cluster up to the early gastrula stage. This expression of *Nanos2* corresponds to the small micromere/PGC lineage, therefore we identified cluster 20 as the germline. Surprisingly, at the morula stage, *Nanos2* was transiently detected in an additional cluster (16). These *Nanos2*-positive cells in both clusters 16 and 20 coexpress two additional germline markers, *Vasa* and *Seawi* (Fig. 6B). *Nanos2* expression is known to be very restricted in the embryo, whereas *Vasa* and *Seawi* are more widespread, and that pattern is reflected in the plot (Fig. 6A). Later in development, cells of cluster 16 started expressing genes such as *Alx1* and *SM50* that are involved in skeleton formation. At the 16-cell stage, the micromeres divide asymmetrically to give rise to the large micromeres (PMCs) and the small micromeres (PGCs). The transient *Nanos2* expression in cluster 16 (PMCs) suggests that the micromeres themselves might be expressing *Nanos2* before giving rise to the germline. *Nanos2* was then promptly lost in the PMCs and retained only in the germline. This analysis is the first evidence of two distinct cell states expressing the germline marker *Nanos2* early in development and we are currently investigating this function. We first observed differentially expressed transcripts between clusters 16 and 20 at the morula stage to identify what makes them unique. Although both clusters are quite similar in terms of gene expression, high levels of *SoxC* expression were seen in cluster 16 compared with cluster 20 (Fig. 6C). *SoxC* is a transcription factor expressed in the ectodermal neural progenitors of the animal pole domain at blastula stage (Garner et al., 2016; Poustka et al., 2007) as well as in the Veg2 lineage (Peter and Davidson, 2010).

In gastrulae, two additional *Nanos2* cell states appeared (clusters 6 and 14). This secondary *Nanos2* expression corresponds to the time in development in which the somatic Veg2 cells start to express *Nanos2*. *Nanos2* expression is regulated by *FoxY* in the somatic

cells and *FoxY* mRNA is expressed in both the germ cells and these Veg2 cells (Andrikou et al., 2013; Materna et al., 2013; Oulhen et al., 2019b; Song and Wessel, 2012). To further test the identity of these *Nanos2*-positive cells during early gastrulation, we examined *FoxY* expression. As predicted, *FoxY* was found enriched in clusters 6, 14 and 20 (Fig. S12), suggesting that the *Nanos2*-expressing cells in cluster 6 and 14 represent the Veg2 lineage. Only a few transcripts were differentially expressed between cluster 6 and cluster 14, suggesting that they are either two closely related cell populations, or that it is one cell population in which some Veg2 cells are just starting to express new markers and/or are at different stages in their cell cycle. We then searched for the genes differentially expressed between these Veg2 clusters together (6 and 14) versus the germline cluster (20) (Tables S3, S4 and S5). Genes such as *Endo16* and *Blimp1/Krox* were found enriched in the Veg2 lineage. In contrast, histone-related genes, such as *H2A.2.1*, were the most highly enriched genes found in the germline (Fig. S13). This may reflect a particular chromatin difference between the germ line and the somatic cells. *H2A.2.1* was already highly expressed in the earlier stages of development and became specifically retained in the PGCs (20), suggesting the maternal origin of this transcript. *Vasa* and *Seawi* were also more enriched in the germline compared to the Veg2 lineage (Figs S14 and S15). Thus, the Veg2-*Nanos2* population of cells has characteristics of germ line definition but is distinct from the PGC lineage derived from the small micromeres.

Overall, we learned that *Nanos2* is expressed not only in the PGCs but also in distinct somatic cell populations in both morulae and gastrulae. We identified genes that were differentially expressed between these *Nanos2*-positive cell states, making them distinct. These analyses revealed genes that were always coexpressed with *Nanos2* throughout development, independently of the time point and the identity of the cell states (Table S6). Among the gene sets, we found uncharacterized transcripts such as LOC590448 (Homeodomain-containing transcription factor), LOC100888091 (sec independent translocase domain) and LOC582810 (Tudor domain) (Fig. S16). Interestingly, genes such as *Krl* (*Krueppel like*), *Odz*, *Staufen* and *Maelstrom* were also found in this analysis (Figs S17-S20). Krüppel-like factors are a family of zinc-finger transcription factors that are essential for maintaining pluripotency (Bialkowska et al., 2017). *Odz* is a pair rule gene required for somatic gonad formation in *Caenorhabditis elegans* (Drabikowski et al., 2005). *Staufen* is an RNA binding protein required for the localization and translational repression of mRNAs in the *Drosophila* oocyte for example (Roegiers and Jan, 2000;

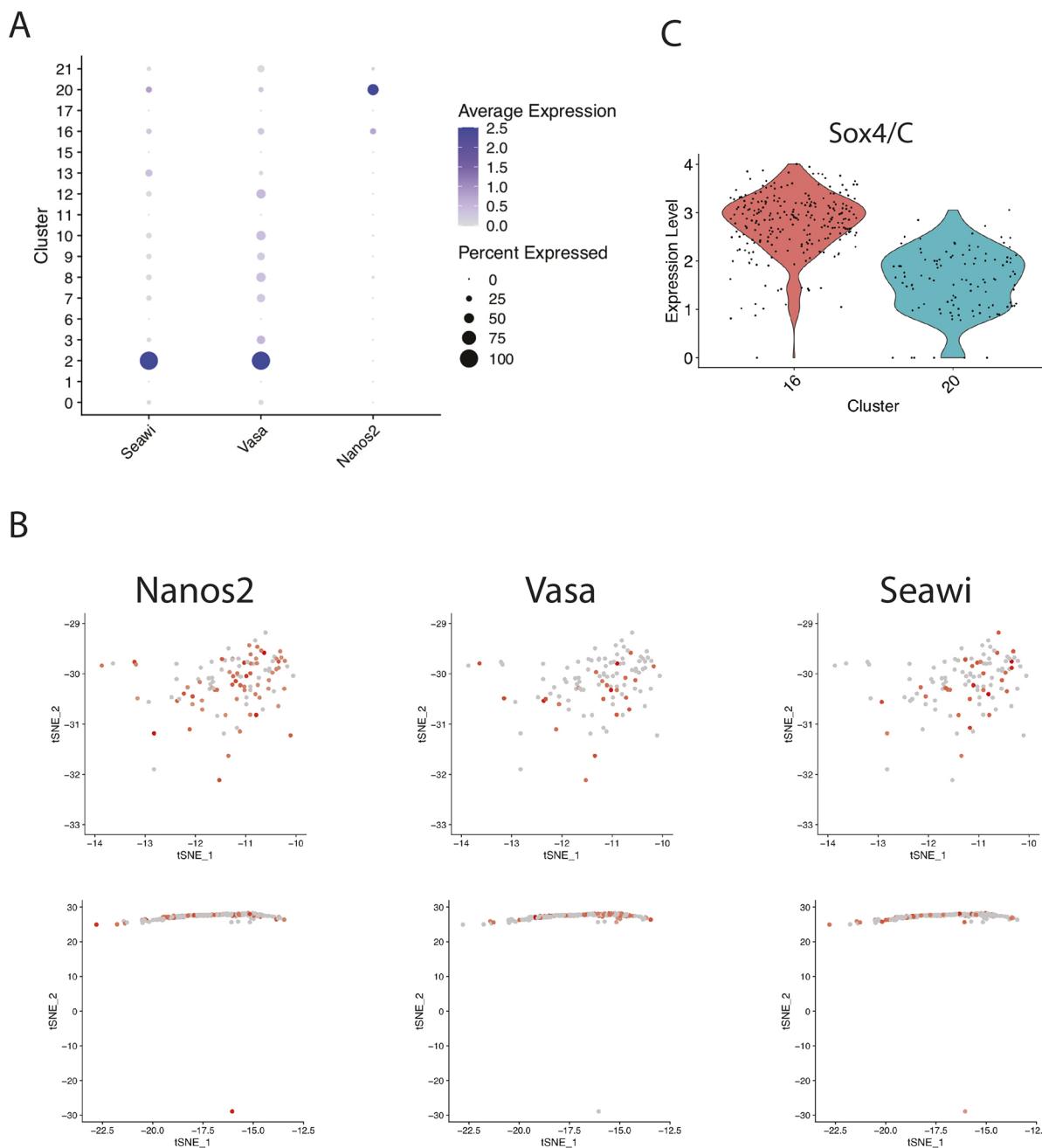


**Fig. 5. Feature and violin plots of relative *Nanos2* expression per cluster (cell state) at different developmental time points.** Number of *Nanos2*-positive cells detected per time point included in feature plots. Cells are included based on normalized expression level, with a lower threshold set to 1, except for 64-cell stage in which it was set to 0.2.

St Johnston et al., 1991). In mice, staufen is expressed in the germ cells in both males and females (Saunders et al., 2000). Maelstrom protein has been suggested to play multiple roles in *Drosophila* oogenesis (Clegg et al., 2001; Findley et al., 2003) and mouse spermatogenesis (Soper et al., 2008). These germ line gene sets may help identify functional kernels in gene regulatory networks for a variety of germ line cells, and in cells that may more readily transition into a germ line lineage (e.g. during regeneration, iPSC proclivities).

#### Genes expressed in the PGCs throughout development

Early in development, the somatic cells divide rapidly but the primordial germ cells reduce their overall activities: transcription, RNA degradation, translation, mitochondria and cell cycle. The PGCs restart these activities after gastrulation, demonstrating a rapid, predictable and transient quiescent activity. We previously identified some of the molecular mechanisms involved in regulating PGC quiescence: Cyclin B (cell cycle), eEF1A (translation), ADP/ATP translocase 1 (mitochondria) and CNOT6 (RNA degradation)



**Fig. 6. Coexpression of germ line genes in clusters 16 and 20 at the morula stage.** (A) Dot plot showing expression of germ line markers *Nanos2*, *Vasa* and *Seawi*. Dot size indicates the percentage of cells expressing the gene of interest; dot color conveys expression level. (B) Expression of *Nanos2*, *Vasa* and *Seawi* in clusters 20 (above) and 16 (below) in morulae. (C) Violin plot showing enrichment of *Sox4/C* in cluster 16 at morula stage. Normalized gene expression values at log scale are shown.

(Oulhen et al., 2017; Swartz et al., 2014). Quiescence is a common phenotype of stem cells and this scRNA-seq data set is essential for a better understanding of the molecular mechanisms required to induce, maintain and exit cell quiescence. We first obtained the genes specifically enriched in the germ cells (cluster 20) for each developmental time point (Table S7). Interestingly, most of the transcripts found enriched in the PGCs are not unique to the germline. These mRNAs are also found in many other clusters, but their expression is enriched in the germline. As expected, *Cyclin B* and *CNOT6* abundance decreased in the germline from the 64-cell stage to mesenchyme blastula. Markers of cell differentiation, such

as *Alx1* or *Blimp1/krox*, also show reduced expression in this cluster when the embryos reach mesenchyme blastula. *Cyclin A* and *Geminin* were highly enriched in the early stage PGCs compared with those in mesenchyme blastulae. In *Drosophila*, the degradation of Cyclin A is essential for the maintenance of the germline stem cells (Chen et al., 2009). Geminin has been associated with proliferation-differentiation decisions through balanced interactions with multiple binding partners (Patmanidi et al., 2017; Wohlschlegel et al., 2002). In contrast, the transcripts coding for ribosomal proteins become highly enriched in the germline in mesenchyme blastula, compared with the 64-cell stage. The

transcript *Maelstrom* mentioned above is also in this category of genes enriched in PGCs at the mesenchyme blastula stage. The ability to interrogate even rare cells in this dataset opens analysis to beyond the candidate genes and provides new avenues for testing germ cell gene functionality through diverse germ-line properties.

## DISCUSSION

We found distinct molecular profiles for at least 22 cell states and two distinct lineages for the germ line in gastrulae of the sea urchin, *S. purpuratus*. Although these cell states are not identified by *in vivo* lineage analysis, each has consistent features of marker genes to provide location for each cell state. These results enable more in-depth analysis for GRN interrogation using this discovery-based experimentation. Of surprise in the results is the fact that many genes thought to be specific to a cell type by *in situ* hybridization do show enrichment in one or another cell state, but then lesser and more broad distribution throughout many other cell states. *Alx1*, for example, is seen in many cell clusters, albeit at much lower levels than present in the PMCs. Are those broadly expressed transcripts functional in the alternative cell states? We cannot predict this outcome simply on the basis of these wild-type embryos, but the results are sufficiently broad, consistent and statistically significant that future experimentation of these gene products should be sensitive to more widespread phenotypes. We have explored whether this result could be artefactual. The broad distribution is most apparent in the violin plots in which each cell with a gene marker can be visualized, versus the feature plots, a more common figure format in the literature of scRNA-seq in which the eye is directed to clusters of the greatest abundance (satijalab.org/seurat). Furthermore, we found clusters lacking the breadth of expression, suggesting that the mRNA was not a broad contaminant of the experiment. In previous results (Zazueta-Novoa and Wessel, 2014) (Fig. S21), the selectivity of an mRNA for a cell type is dependent on the length of substrate development. Short periods of signal development can show enrichment or even specificity in a cell type, but additional signal exposure shows much more uniform accumulation. Because the scRNA-seq is analyzed with definitive sequence information, we have confidence in these identification calls. We provide a paradigm wherein the conclusion is based on the technique used. *Gustavus* (*Gus*) mRNA is reported to accumulate specifically in cells of the vegetal pole in blastulae, surrounding the primordial germ cells. This is functionally relevant because *Gus* is an E3 ubiquitin ligase responsible for Vasa protein degradation, and during this selective expression Vasa protein becomes progressively restricted to the PGCs from a broad somatic presence (Gustafson et al., 2011). Here, by scRNA-seq, we found *Gus* transcripts enriched in cells of the vegetal pole, but broadly detectable in all cells of the embryo (Fig. S21). When, however, the *in situ* hybridization signal reaction was extended, *Gus* mRNA was broadly present throughout the embryo (Fig. S21) (Zazueta-Novoa and Wessel, 2014) consistent with the scRNA-seq data. This phenomenon does not change the interpretation of what these genes are doing, only that the specificity may be less than previously appreciated. This resonates with the results of the Human and Mouse ENCODE projects (<https://www.encodeproject.org/>) in which vastly more transcripts from throughout the genome were detected than otherwise anticipated. Perhaps the transcripts of ‘cell-specific’ genes present outside of the cell of interest are not actually functional, or short lived or not translated by a variety of regulatory mechanisms. The consistency of this phenomenon from widely disparate cells and organisms [(Briggs et al., 2018; Siebert et al., 2019); results from sea urchin here] suggests that the finding is of biological relevance.

The phenomenon of broad but low expression of ‘cell-type specific’ markers has been reported in other organisms as well (Briggs et al., 2018; Macosko et al., 2015; Siebert et al., 2019) although the feature plots shown make such a result less apparent than the violin plots. We note that some transcripts were present at low level in all cell clusters, except for the PGC cluster. It is not yet clear whether these transcripts are repressed from transcription or are short lived in the germ line. It is noteworthy that maternal transcripts, even ectopic transcripts, are usually retained in the germ line more effectively than in the soma (Gustafson and Wessel, 2010; Oulhen and Wessel, 2013; Swartz et al., 2014). Thus, these low-level transcripts broadly expressed outside of their ‘enriched’ cell cluster appear to violate that selectivity in both ways – they are selectively not retained in the germ line (if in fact they are transcribed in that cell type), and they are selectively retained outside of the germ line. Because the 3'-UTR of some mRNAs in the sea urchin are known to be regulatory in such retention and turnover functions, we will compare the UTRs of these transcripts in the future to determine whether conserved sequence signatures are present. *Nanos2* is an RNA binding protein; it functions through its interaction with Pumilio, which binds RNAs containing a conserved motif in their 3'-UTR, the Pumilio Response Element (PRE). These target mRNAs are then translationally repressed and/or degraded. Interestingly, some PMC markers such as *SM37* and *Alx1* have one or more predicted PRE in their 3'-UTRs.

We previously reported mRNAs enriched in the small micromeres, based on RNA-seq analysis of FACS-sorted cells (Swartz et al., 2014). Several of the genes reported to be enriched in the small micromeres by FACS isolation are present in our list of differentially expressed genes between *Nanos2*-positive and *Nanos2*-negative cells at the matching time point of early blastula and across the entire dataset. This includes Sp-z62 (LOC580113) and Sp-Ctdspl2 (LOC583286). Single cell sequencing revealed that the expression of these common genes is not specific to the germline, as is *Nanos2* expression, but rather the genes are present in many cell clusters with slightly higher expression in the germline. These differences may be explained by significant differences in the two experimental approaches, one being bulk RNA-seq of the germ line cells (calcein-positive cells) against all the other cells of the embryo (calcein-negative cells) whereas the current analysis is focused on single cell sequencing analyzed and compared individually to other cell states. A sharp gradient of mRNA accumulation highest in the vegetal pole would appear ‘specific’ to the vegetal pole cells by bulk sequencing of the mRNA, comparing isolated vegetal pole cells to the rest of the embryo, whereas scRNA-seq would show this same transcript broadly expressed in clusters throughout the vegetal hemisphere of the embryo.

*Nanos* is an RNA-binding protein and was first described as a translational repressor in *Drosophila* (Cho et al., 2006; Irish et al., 1989). *Nanos* orthologs have since been found in the germ line of all animals tested, including *C. elegans* (Kraemer et al., 1999), *Xenopus* (Lai et al., 2011) and planarians (Wang et al., 2007). We recently discovered that *Nanos2* in the sea urchin is not only expressed in the germline but is also present in the somatic Veg2 mesoderm lineage during gastrulation. This is significant because when the precursor cells of the PGCs are removed from the embryo, the resultant adult still makes eggs and sperm. A replacement, or a second germ line lineage, must be present prior to adulthood. It was learned that loss of the micromeres, the PGC precursor cells, results in dramatic upregulation of Vasa protein throughout the embryo. Later, Vasa becomes restricted to subpopulations of cells, including Veg2 descendants (Voronina et al., 2008). Thus, it was thought that

a secondary site of germ line potential is present and may be within the Veg2 tier. This Veg2 mesoderm is derived from the Veg2 endomesoderm lineage, the distinction between endoderm and mesoderm specification in Veg2 depends on the Delta/Notch signaling system at the blastula stage. The Veg2 mesoderm gives rise to multiple cell types, such as pigment cells, blastocoelar cells, esophageal muscle cells and cells of the coelomic pouches (Cameron et al., 1991; Davidson et al., 1998; Peter and Davidson, 2010). The early expression of *Nanos2* in PGCs requires the maternal Wnt pathway for expression, whereas somatic *Nanos2* expression requires Delta/Notch signaling through the forkhead family member FoxY (Oulhen et al., 2019b). The expression of germ-line determinant genes is highly regulated in many animals. Indeed, ectopic expression of these genes can induce cell cycle and developmental defects (Luo et al., 2011; Wu and Ruvkun, 2010) and Nanos is thought to be ‘toxic’ outside of its normal domain (Lai and King, 2013). With the results presented here, we conclude that a population of Veg2 cells contains a unique repertoire of cell-type factors, including those normally responsible for a germ line fate. This cell lineage also has other, partial germ line characteristics that support this conclusion. These include partial quiescence of protein synthesis and other metabolic features (Oulhen et al., 2017; Oulhen et al., 2019b) that reflect a lower level of *Nanos2* translational repression. Future experimentation will attempt to analyze scRNA-seq data in this embryo following micromere depletion to determine how the Veg2 lineage changes in such instances in loss of PGC precursors. As some other animal embryos also appear to compensate somehow for a lost germ line, this molecular mechanism may serve as a paradigm for germ line conversion or regeneration mechanisms (Wessel et al., 2020).

Overall, we find the scRNA-seq approach used herein to be highly productive in identifying changes in cell states and gene expression throughout early development of the sea urchin. By use of an estimated 300 embryos after gastrulation, one can acquire sufficient numbers of cells to analyze with this technique, opening a protocol for MASO or CRISPR/Cas9 manipulation of the embryo and subsequent testing of impact on the 22 cell clusters revealed herein. The breadth of developmental perturbations is further magnified by common pharmacological approaches, which have already revealed remarkable selectivity (Foster et al., 2019), further increasing the confidence level of valid interpretations from such experimental approaches. Coupled with the reproducibility shown here, this animal and this technology are a wonderful marriage for deep biomedical research.

## MATERIALS AND METHODS

### Embryo cultures and dissociations

Eggs and sperm of *S. purpuratus* were spawned by injection of 0.5 M KCl into the adult coelomic cavity. Fertilization was accomplished in sea water containing 10 mM *p*-aminobenzoic acid to reduce cross-linking of the fertilization envelope, and which was washed out after 30 min. Embryos were cultured in filtered (0.2 µm) sea water collected at the Marine Biological laboratories in Woods Hole MA, until the appropriate stage for dissociation. All embryos used in the study resulted from mating of one male and one female. Multiple fertilizations were initiated in this study and timed such that the appropriate stages of embryonic development were reached at a common endpoint. The embryos were then collected and washed twice with calcium-free sea water, and then suspended in hyalin-extraction medium (HEM) for 10–15 min, depending on the stage of dissociation. When cells were beginning to dissociate, the embryos were collected and washed in 0.5 M NaCl, gently sheared with a pipette, run through a 40 µm Nitex mesh, counted on a hemocytometer, and diluted to reach the appropriate concentration for the scRNA-seq protocol. Equal numbers of embryos were used for each time point and at no time were cells or embryos pelleted in a centrifuge (Oulhen et al., 2019a).

### Single cell RNA sequencing

Single cell encapsulation was performed using the Chromium Single Cell Chip B kit on the 10× Genomics Chromium Controller. Single cell cDNA and libraries were prepared using the Chromium Single Cell 3' Reagent Kit v3 Chemistry. Libraries were sequenced by Genewiz on the Illumina Hiseq (2×150 bp paired-end runs). Single cell unique molecular identifier (UMI) counting (counting of unique barcodes given to individual transcript molecules), was performed using Cell Ranger Single Cell Software Suite 3.0.2 from 10× Genomics. The custom transcriptome reference was generated from assembly Spur-4.2 using CellRanger. Duplicate blastula and gastrula stage libraries were aggregated using the CellRanger aggr function. CellRanger gene expression matrices were further analyzed using the R package Seurat v 3.1.4 (Butler et al., 2018; Stuart et al., 2019). Cells of 8-cell to early blastula stage with at least 1500 and at most 7000 genes (features), and cells with at least 400 and at most 2500 genes in hatched blastula to late gastrula stage were included in downstream analysis. Individual datasets were normalized by scaling gene expression in each cell by total gene expression and then log transformed. The top 2000 highly variable genes across the datasets were then used to integrate the datasets. Individual time point datasets were integrated (following the Seurat v3 pipeline) to identify conserved cell populations across the datasets. This technique involves pairwise comparison of individual cells across multiple datasets followed by hierarchical clustering. The t-SNE (t-distributed stochastic neighbor embedding is a machine learning algorithm for visualizations) projection and clustering analysis for visualization of the integrated data was conducted using 15 parameter dimensions and resolution of 0.5. The use of 15 dimensions (i.e. inclusion of 15 principle components) was made with consideration of principal component heatmaps, which show heterogeneity sources in a dataset, and the ElbowPlot function, which depicts the number of principle components that include the variance present in the data. The dataset was also visualized at a resolution of 3 to provide an example of how additional cell states may be revealed, including subtypes of the states seen and identified at resolution 0.5. No one resolution setting is optimal for all clusters but these disparate settings are intended to assist the reader in data interpretation and identification of candidate genes. The resolution parameter of the FindClusters function can be modulated to show greater or fewer clusters and a series of different resolutions can be tested before choosing a value that is appropriate for the biological context of an experiment. Subclustering of the endoderm representatives was performed using the Subset function to capture clusters of interest, followed by finding variable features and rescaling. The clustering parameters used were 10 dimensions and resolution 1.0. Cluster markers were found using FindConservedMarkers and FindMarkers functions. Comparison of the Foster et al. (2019) dataset and the corresponding time point from this dataset was performed following the same pipeline described above and visualized with parameters of 20 dimensions and resolution 0.5. Average expression was measured as count data normalized to library size and log transformed (Stuart et al., 2019).

### Acknowledgements

Part of this research was conducted using resources and services at the Computational Biology Core and Center for Computation and Visualization, Brown University. A special thanks to August Guang and Ashok Ragavendran of the Computational Biology Core. We thank Shane Evans for helpful conversations regarding analysis of the data.

### Competing interests

The authors declare no competing or financial interests.

### Author contributions

Conceptualization: S.F., N.O., G.W.; Methodology: S.F., N.O., G.W.; Software: S.F., N.O.; Validation: S.F., N.O.; Formal analysis: S.F., N.O.; Investigation: S.F., N.O., G.W.; Resources: S.F., N.O.; Data curation: S.F., N.O.; Writing - original draft: S.F., N.O., G.W.; Writing - review & editing: S.F., N.O., G.W.; Visualization: S.F., N.O.; Supervision: G.W.; Project administration: G.W.; Funding acquisition: N.O., G.W.

### Funding

We are grateful for support from the National Institutes of Health (9R01GM125071, 1R01GM132222 and 1P20GM119943). Deposited in PMC for release after 12 months.

**Data availability**

The sequencing files and gene expression matrices for the single-cell RNA-seq analysis presented here have been deposited in GEO under accession number GSE149221.

**Supplementary information**

Supplementary information available online at  
<https://dev.biologists.org/lookup/doi/10.1242/dev.191528.supplemental>

**References**

- Andrikou, C., Iovene, E., Rizzo, F., Oliveri, P. and Arnone, M. I.** (2013). Myogenesis in the sea urchin embryo: the molecular fingerprint of the myoblast precursors. *EvoDevo* **4**, 33. doi:10.1186/2041-9139-4-33
- Anishchenko, E., Arnone, M. I. and D'Aniello, S.** (2018). SoxB2 in sea urchin development: implications in neurogenesis, ciliogenesis and skeletal patterning. *EvoDevo* **9**, 5. doi:10.1186/s13227-018-0094-1
- Annunziata, R., Perillo, M., Andrikou, C., Cole, A. G., Martinez, P. and Arnone, M. I.** (2014). Pattern and process during sea urchin gut morphogenesis: the regulatory landscape. *Genesis* **52**, 251-268. doi:10.1002/dvg.22738
- Bialkowska, A. B., Yang, V. W. and Mallipattu, S. K.** (2017). Krüppel-like factors in mammalian stem cells and development. *Development* **144**, 737-754. doi:10.1242/dev.145441
- Briggs, J. A., Weinreb, C., Wagner, D. E., Megason, S., Peshkin, L., Kirschner, M. W. and Klein, A. M.** (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780. doi:10.1126/science.aar5780
- Bruskin, A. M., Tyner, A. L., Wells, D. E., Showman, R. M. and Klein, W. H.** (1981). Accumulation in embryogenesis of five mRNAs enriched in the ectoderm of the sea urchin pluteus. *Dev. Biol.* **87**, 308-318. doi:10.1016/0012-1606(81)90154-8
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R.** (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411-420. doi:10.1038/nbt.4096
- Cameron, R. A., Fraser, S. E., Britten, R. J. and Davidson, E. H.** (1991). Macromere cell fates during sea urchin development. *Development* **113**, 1085-1091.
- Chen, D., Wang, Q., Huang, H., Xia, L., Jiang, X., Kan, L., Sun, Q. and Chen, D.** (2009). Effete-mediated degradation of Cyclin A is essential for the maintenance of germline stem cells in Drosophila. *Development* **136**, 4133-4142. doi:10.1242/dev.039032
- Cho, P. F., Gamberi, C., Cho-Park, Y. A., Cho-Park, I. B., Lasko, P. and Sonenberg, N.** (2006). Cap-dependent translational inhibition establishes two opposing morphogen gradients in Drosophila embryos. *Curr. Biol.* **16**, 2035-2041. doi:10.1016/j.cub.2006.08.093
- Clegg, N. J., Findley, S. D., Mahowald, A. P. and Ruohola-Baker, H.** (2001). Maelstrom is required to position the MTOC in stage 2-6 Drosophila oocytes. *Dev. Genes Evol.* **211**, 44-48. doi:10.1007/s004270000114
- Davidson, E. H., Cameron, R. A. and Ransick, A.** (1998). Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms. *Development* **125**, 3269-3290.
- Drabikowski, K., Trzebiatowska, A. and Chiquet-Ehrismann, R.** (2005). ten-1, an essential gene for germ cell development, epidermal morphogenesis, gonad migration, and neuronal pathfinding in *Caenorhabditis elegans*. *Dev. Biol.* **282**, 27-38. doi:10.1016/j.ydbio.2005.02.017
- Ettensohn, C. A., Illies, M. R., Oliveri, P. and De Jong, D. L.** (2003). Alx1, a member of the Cart1/Alx3/Alx4 subfamily of Paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo. *Development* **130**, 2917-2928. doi:10.1242/dev.00511
- Findley, S. D., Tamanaha, M., Clegg, N. J. and Ruohola-Baker, H.** (2003). Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. *Development* **130**, 859-871. doi:10.1242/dev.00310
- Fink, R. D. and McClay, D. R.** (1985). Three cell recognition changes accompany the ingressation of sea urchin primary mesenchyme cells. *Dev. Biol.* **107**, 66-74. doi:10.1016/0012-1606(85)90376-8
- Foster, S., Teo, Y. V., Neretti, N., Oulhen, N. and Wessel, G. M.** (2019). Single cell RNA-seq in the sea urchin embryo show marked cell-type specificity in the Delta/Notch pathway. *Mol. Reprod. Dev.* **86**, 931-934. doi:10.1002/mrd.23181
- Fujii, T., Sakamoto, N., Ochiai, H., Fujita, K., Okamitsu, Y., Sumiyoshi, N., Minokawa, T. and Yamamoto, T.** (2009). Role of the nanos homolog during sea urchin development. *Dev. Dyn.* **238**, 2511-2521. doi:10.1002/dvdy.22074
- Garner, S., Zysk, I., Byrne, G., Kramer, M., Moller, D., Taylor, V. and Burke, R. D.** (2016). Neurogenesis in sea urchin embryos and the diversity of deuterostome neurogenic mechanisms. *Development* **143**, 286-297. doi:10.1242/dev.124503
- George, N. C., Killian, C. E. and Wilt, F. H.** (1991). Characterization and expression of a gene encoding a 30.6-kDa Strongylocentrotus purpuratus spicule matrix protein. *Dev. Biol.* **147**, 334-342. doi:10.1016/0012-1606(91)90291-A
- Gulati, G. S., Sikandar, S. S., Wesche, D. J., Marjunath, A., Bharadwaj, A., Berger, M. J., Ilagan, F., Kuo, A. H., Hsieh, R. W., Cai, S. et al.** (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405-411. doi:10.1126/science.aax0249
- Gustafson, E. A. and Wessel, G. M.** (2010). Exogenous RNA is selectively retained in the small micromeres during sea urchin embryogenesis. *Mol. Reprod. Dev.* **77**, 836. doi:10.1002/mrd.21241
- Gustafson, E. A., Yajima, M., Juliano, C. E. and Wessel, G. M.** (2011). Post-translational regulation by gustavus contributes to selective Vasa protein accumulation in multipotent cells during embryogenesis. *Dev. Biol.* **349**, 440-450. doi:10.1016/j.ydbio.2010.10.031
- Irish, V., Lehmann, R. and Akam, M.** (1989). The *Drosophila* posterior-group gene *nanos* functions by repressing *hunchback* activity. *Nature* **338**, 646-648. doi:10.1038/338646a0
- Juliano, C. E., Yajima, M. and Wessel, G. M.** (2010). *Nanos* functions to maintain the fate of the small micromere lineage in the sea urchin embryo. *Dev. Biol.* **337**, 220-232. doi:10.1016/j.ydbio.2009.10.030
- Klein, A. M.** (2017). What is your conceptual definition of "Cell type" in the context of a mature organism? *Cell Syst.* **4**, 256.
- Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A. and Kirschner, M. W.** (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201. doi:10.1016/j.cell.2015.04.044
- Kraemer, B., Crittenden, S., Gallegos, M., Moulder, G., Barstead, R., Kimble, J. and Wickens, M.** (1999). NANOS-3 and FBF proteins physically interact to control the sperm-oocyte switch in *Caenorhabditis elegans*. *Curr. Biol.* **9**, 1009-1018. doi:10.1016/S0960-9822(99)80449-7
- Lai, F. and King, M. L.** (2013). Repressive translational control in germ cells. *Mol. Reprod. Dev.* **80**, 665-676. doi:10.1002/mrd.22161
- Lai, F., Zhou, Y., Luo, X., Fox, J. and King, M. L.** (2011). Nanos1 functions as a translational repressor in the Xenopus germline. *Mech. Dev.* **128**, 153-163. doi:10.1016/j.mod.2010.12.001
- Luo, X., Nerlick, S., An, W. and King, M. L.** (2011). Xenopus germline nanos1 is translationally repressed by a novel structure-based mechanism. *Development* **138**, 589-598. doi:10.1242/dev.056705
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M. et al.** (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202-1214. doi:10.1016/j.cell.2015.05.002
- Materna, S. C., Swartz, S. Z. and Smith, J.** (2013). Notch and Nodal control forkhead factor expression in the specification of multipotent progenitors in sea urchin. *Development* **140**, 1796-1806. doi:10.1242/dev.091157
- McClay, D. R. and Marchase, R. B.** (1979). Separation of ectoderm and endoderm from sea urchin pluteus larvae and demonstration of germ layer-specific antigens. *Dev. Biol.* **71**, 289-296. doi:10.1016/0012-1606(79)90170-2
- McClay, D. R., Chambers, A. F. and Warren, R. H.** (1977). Specificity of cell-cell interactions in sea urchin embryos. Appearance of new cell-surface determinants at gastrulation. *Dev. Biol.* **56**, 343-355. doi:10.1016/0012-1606(77)90275-5
- Oliveri, P., Walton, K. D., Davidson, E. H. and McClay, D. R.** (2006). Repression of mesodermal fate by foxa, a key endoderm regulator of the sea urchin embryo. *Development* **133**, 4173-4181. doi:10.1242/dev.02577
- Oulhen, N. and Wessel, G. M.** (2013). Retention of exogenous mRNAs selectively in the germ cells of the sea urchin requires only a 5'-cap and a 3'-UTR. *Mol. Reprod. Dev.* **80**, 561-569. doi:10.1002/mrd.22193
- Oulhen, N., Swartz, S. Z., Laird, J., Mascaro, A. and Wessel, G. M.** (2017). Transient translational quiescence in primordial germ cells. *Development* **144**, 1201-1210. doi:10.1242/dev.144170
- Oulhen, N., Foster, S., Wray, G. and Wessel, G. M.** (2019a). Identifying gene expression from single cells to single genes. *Methods Cell Biol.* **151**, 127-158. doi:10.1016/bs.mcb.2018.11.018
- Oulhen, N., Swartz, S. Z., Wang, L., Wikramanayake, A. and Wessel, G. M.** (2019b). Distinct transcriptional regulation of Nanos2 in the germ line and soma by the Wnt and delta/notch pathways. *Dev. Biol.* **452**, 34-42. doi:10.1016/j.ydbio.2019.04.010
- Patmanidi, A. L., Champeris Tsaniras, S., Karamitros, D., Kyrousi, C., Lygerou, Z. and Taraviras, S.** (2017). Concise review: geminin-a tale of two tails: DNA replication and transcriptional/epigenetic regulation in stem cells. *Stem Cells* **35**, 299-310. doi:10.1002/stem.2529
- Peter, I. S. and Davidson, E. H.** (2010). The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Dev. Biol.* **340**, 188-199. doi:10.1016/j.ydbio.2009.10.037
- Poustka, A. J., Kühn, A., Groth, D., Weisse, V., Yaguchi, S., Burke, R. D., Herwig, R., Lehrach, H. and Panopoulou, G.** (2007). A global view of gene expression in lithium and zinc treated sea urchin embryos: new components of gene regulatory networks. *Genome Biol.* **8**, R85. doi:10.1186/gb-2007-8-5-r85
- Ransick, A., Ernst, S., Britten, R. J. and Davidson, E. H.** (1993). Whole mount *in situ* hybridization shows Endo 16 to be a marker for the vegetal plate territory in sea urchin embryos. *Mech. Dev.* **42**, 117-124. doi:10.1016/0925-4773(93)90001-E

- Roegiers, F. and Jan, Y. N.** (2000). Staufen: a common component of mRNA transport in oocytes and neurons? *Trends Cell Biol.* **10**, 220-224. doi:10.1016/S0962-8924(00)01767-0
- Saunders, P. T. K., Pathirana, S., Maguire, S. M., Doyle, M., Wood, T. and Bownes, M.** (2000). Mouse staufen genes are expressed in germ cells during oogenesis and spermatogenesis. *Mol. Hum. Reprod.* **6**, 983-991. doi:10.1093/molehr/6.11.983
- Siebert, S., Farrell, J. A., Cazet, J. F., Abeykoon, Y., Primack, A. S., Schnitzler, C. E. and Juliano, C. E.** (2019). Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science* **365**, eaav9314. doi:10.1126/science.aav9314
- Song, J. L. and Wessel, G. M.** (2012). The forkhead transcription factor FoxY regulates Nanos. *Mol. Reprod. Dev.* **79**, 680-688. doi:10.1002/mrd.22073
- Soper, S. F., van der Heijden, G. W., Hardiman, T. C., Goodheart, M., Martin, S. L., de Boer, P. and Bortvin, A.** (2008). Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev. Cell* **15**, 285-297. doi:10.1016/j.devcel.2008.05.015
- St Johnston, D., Beuchle, D. and Nüsslein-Volhard, C.** (1991). Staufen, a gene required to localize maternal RNAs in the Drosophila egg. *Cell* **66**, 51-63. doi:10.1016/0092-8674(91)90138-O
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., III, Hao, Y., Stoeckius, M., Smibert, P. and Satija, R.** (2019). Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e1821. doi:10.1016/j.cell.2019.05.031
- Swartz, S. Z., Reich, A. M., Oulhen, N., Raz, T., Milos, P. M., Campanale, J. P., Hamdoun, A. and Wessel, G. M.** (2014). Deadenylylase depletion protects inherited mRNAs in primordial germ cells. *Development* **141**, 3134-3142. doi:10.1242/dev.110395
- Takacs, C. M., Amore, G., Oliveri, P., Poustka, A. J., Wang, D., Burke, R. D. and Peterson, K. J.** (2004). Expression of an NK2 homeodomain gene in the apical ectoderm defines a new territory in the early sea urchin embryo. *Dev. Biol.* **269**, 152-164. doi:10.1016/j.ydbio.2004.01.023
- Tu, Q., Brown, C. T., Davidson, E. H. and Oliveri, P.** (2006). Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev. Biol.* **300**, 49-62. doi:10.1016/j.ydbio.2006.09.031
- 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
- Urry, L. A., Hamilton, P. C., Kilian, C. E. and Wilt, F. H.** (2000). Expression of spicule matrix proteins in the sea urchin embryo during normal and experimentally altered spiculogenesis. *Dev. Biol.* **225**, 201-213. doi:10.1006/dbio.2000.9828
- Veronina, E., Lopez, M., Juliano, C. E., Gustafson, E., Song, J. L., Extavour, C., George, S., Oliveri, P., McClay, D. and Wessel, G.** (2008). Vasa protein expression is restricted to the small micromeres of the sea urchin, but is inducible in other lineages early in development. *Dev. Biol.* **314**, 276-286. doi:10.1016/j.ydbio.2007.11.039
- Wang, Y., Zayas, R. M., Guo, T. and Newmark, P. A.** (2007). nanos function is essential for development and regeneration of planarian germ cells. *Proc. Natl. Acad. Sci. USA* **104**, 5901-5906. doi:10.1073/pnas.0609708104
- Wei, Z., Yaguchi, J., Yaguchi, S., Angerer, R. C. and Angerer, L. M.** (2009). The sea urchin animal pole domain is a Six3-dependent neurogenic patterning center. *Development* **136**, 1179-1189. doi:10.1242/dev.032300
- Wessel, G. M., Morita, S. and Oulhen, N.** (2020). Somatic cell conversion to a germ cell lineage: a violation or a revelation? *J. Exp. Zool. B Mol. Dev. Evol.* doi:10.1002/jez.b.22952
- Wohlschlegel, J. A., Kutok, J. L., Weng, A. P. and Dutta, A.** (2002). Expression of geminin as a marker of cell proliferation in normal tissues and malignancies. *Am. J. Pathol.* **161**, 267-273. doi:10.1016/S0002-9440(10)64178-8
- Wu, X. and Ruvkun, G.** (2010). Germ cell genes and cancer. *Science* **330**, 1761-1762. doi:10.1126/science.1200772
- Yaguchi, S., Yaguchi, J., Angerer, R. C. and Angerer, L. M.** (2008). A Wnt-FoxQ2 nodal pathway links primary and secondary axis specification in sea urchin embryos. *Dev. Cell* **14**, 97-107. doi:10.1016/j.devcel.2007.10.012
- Yaguchi, S., Yaguchi, J., Wei, Z., Shiba, K., Angerer, L. M. and Inaba, K.** (2010). ankAT-1 is a novel gene mediating the apical tuft formation in the sea urchin embryo. *Dev. Biol.* **348**, 67-75. doi:10.1016/j.ydbio.2010.09.011
- Zazueta-Novoa, V. and Wessel, G. M.** (2014). Protein degradation machinery is present broadly during early development in the sea urchin. *Gene Expr. Patterns* **15**, 135-141. doi:10.1016/j.gep.2014.06.002