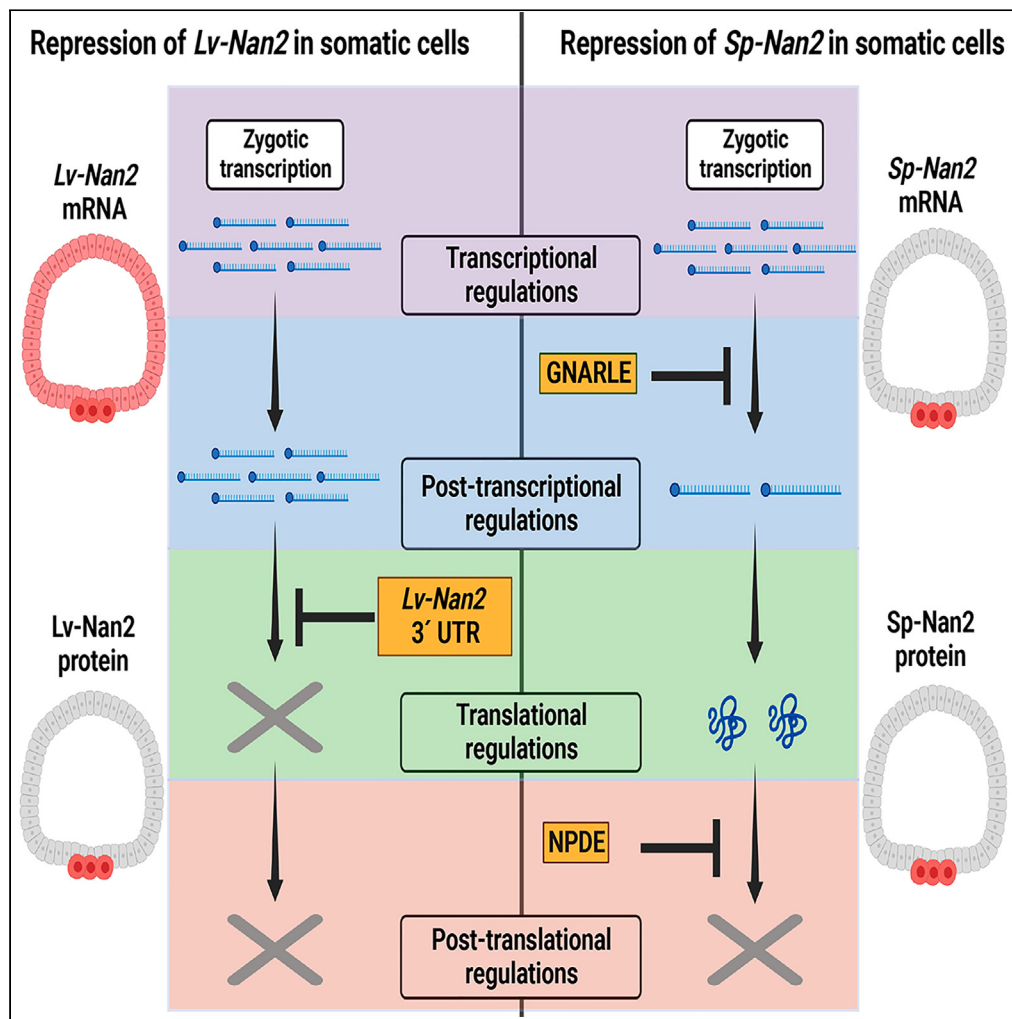


Article

Elements of divergence in germline determination in closely related species



Shumpei Morita,
Nathalie Oulhen,
Stephany Foster,
Gary M. Wessel

rhet@brown.edu

Highlights

Two echinoderm species were compared by stage-matched scRNA-seq datasets

Expression of the germline factor, Nan2, varied significantly between the species

We found that Nan2 mRNA and protein are regulated differentially in these two species

Post-transcriptional and post-translation regulation dominates Nan2 expression

Morita et al., iScience 26, 106402
April 21, 2023 © 2023 The Author(s).
<https://doi.org/10.1016/j.isci.2023.106402>

Article

Elements of divergence in germline determination in closely related species

Shumpei Morita,^{1,2,3} Nathalie Oulhen,^{1,3} Stephany Foster,¹ and Gary M. Wessel^{1,4,*}

SUMMARY

Evolutionary transitions are particularly important in development of the germ line, cells which directly impact sexual reproduction. Differences in the primordial germ cells (PGCs) of two sea urchin species were examined here by stage-matched, integrated, single cell RNA-seq (scRNA-seq) datasets. Even though both species rely on inherited mechanisms to specify their germ line, this analysis revealed a variety of differences in germline gene expression, including a broader expression of the germline factor *Nanos2* (*Nan2*) in *Lytechinus variegatus* (*Lv*) compared to *Strongylocentrotus purpuratus* (*Sp*). In *Sp*, *Nan2* mRNA expression is highly restricted to the PGCs by a lability element in its 3'UTR, which is lacking in the mRNA of *Lv-Nan2*, thus explaining the difference. We discovered that the *Lv-Nan2* 3'UTR instead leads to its specific translation in the PGCs. The results emphasize that regulatory mechanisms resulting in germline specification rely greatly on post-transcriptional restrictions of key gene products.

INTRODUCTION

The germ line is one of the few cell lineages shared by nearly all animals; it generates the eggs and sperm necessary for sexual reproduction. However, the germ line is formed by highly diverse mechanisms, sometimes even between closely related taxa.^{1,2} The marked variations in these mechanisms is likely a node of evolutionary manipulation, and selected for by success in fitness.³ Such a direct impact on fitness is likely a common feature of germ cell diversification, but how such transitions in mechanism have occurred is only beginning to be appreciated.

The primordial germ cells (PGCs) in most animals known are specified during embryogenesis and subsequently give rise to the germ line; Nanos is an RNA binding protein essential for the maintenance and survival of the PGCs. Its function has been tested in multiple species.^{2,4–7} Together with its partner Pumilio, Nanos binds to a specific element in the 3'UTR of its target mRNAs, the PRE (Pumilio response element). This binding induces the translational inhibition and/or the degradation of the targeted mRNAs. So far, only a few targets of the Nanos/Pumilio complex have been identified and include hunchback, cyclin B, hid, VegT, CNOT6, eEF1a.^{8–16} Nanos is rigidly regulated and its ectopic expression leads to embryonic lethality.¹⁷ However, a variety of methods are employed by animals to restrict Nanos protein expression to the PGCs.

In the sea urchin, the PGCs arise from two asymmetric cell divisions, resulting in the four small micromeres at the fifth cell division (32 cell stage). Shortly after their formation, their cell cycle is downregulated, they divide only once more by the end of gastrulation, and they show a transient downregulation of their transcriptional, translational, RNA degradation, and mitochondrial activities.^{9,18} In the sea urchin *Strongylocentrotus purpuratus* (*Sp*), *Nanos2* (*Nan2*) is essential to maintain this transient quiescence. Multiple levels of regulation restrict *Sp-Nan2* mRNA and protein to the PGCs early in development: It is transcribed broadly in the early embryo through the Wnt pathway^{19,20} but its mRNA only accumulates in the PGCs. *Sp-Nan2* contains an element named Global Nanos Associated RNA Lability Element (GNARLE) in its 3'UTR that leads to its degradation in the somatic cells and its retention in the PGCs.^{21,22} Finally, it is not possible to overexpress this protein in the somatic cells because the protein itself possesses regulatory elements that leads to its degradation in somatic cells and its retention in the PGCs.²³

More recently, we used single-cell RNA-seq (scRNA-seq) to identify the transcriptome of the PGCs throughout the development of *Sp* embryos. As expected, the transcript coding for *Sp-Nan2* was highly

¹Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, 02912, USA

²Present Address: Asamushi Research Center for Marine Biology, Graduate School of Life Sciences, Tohoku University, Aomori, Aomori, 039-3501, Japan

³These authors contributed equally

⁴Lead contact

*Correspondence: rhet@brown.edu

<https://doi.org/10.1016/j.isci.2023.106402>



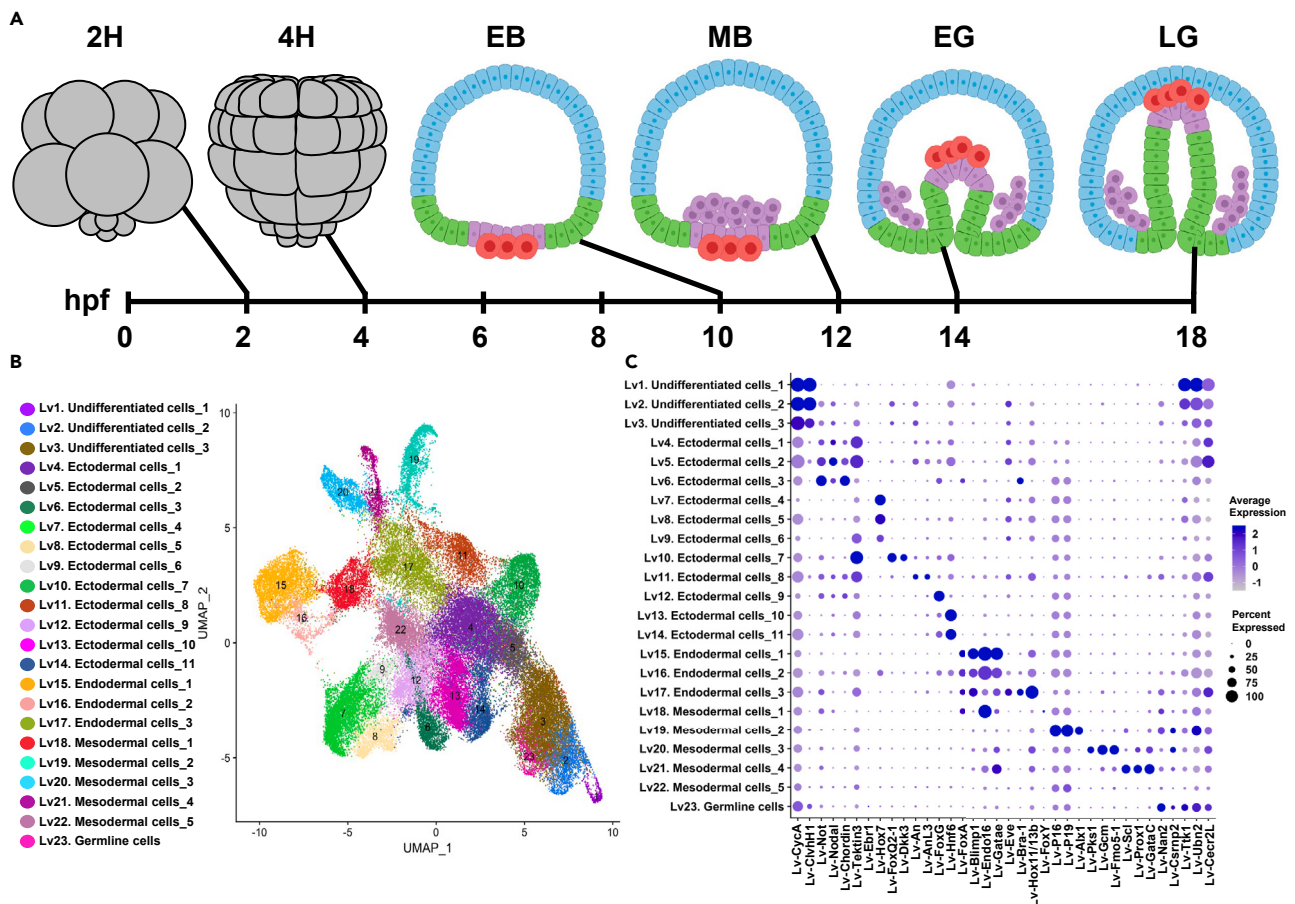


Figure 1. Identification of 23 cell-populations in *Lytechinus variegatus* embryos

(A) Diagram of sea urchin embryogenesis. We collected embryos at eight developmental time points from 2 h postfertilization (hpf) to 18 hpf. The diagrams of embryos at 2 hpf (2H), 4 hpf (4H), 10 hpf (EB), 12 hpf (MB), 14 hpf (EG) and 18 hpf (LG) are illustrated. Undifferentiated cells (gray), ectodermal cells (blue), endodermal cells (green), mesodermal cells (purple) and germline cells (red) are shown. This figure was created with [BioRender.com](https://app.biorender.com/) (<https://app.biorender.com/>).

(B) Identification of 23 cell-populations using UMAP visualization for integrated dataset of 8 developmental time points.

(C) Dot plot represents marker gene expression characterizing cell types in the Lv dataset. Dot size and dot color indicate the percentage of cells expressing the gene and the average expression level, respectively.

restricted to the PGC cluster.²⁴ Here we use scRNA-seq analysis to compare the developmental profile of two sea urchin species with a last common ancestor of ~50 myr, *S. purpuratus* (Sp) and *Lytechinus variegatus* (Lv). Although both species appear to specify their PGCs by inherited mechanisms, they are highly divergent in detail. This scRNA-seq analysis suggests that *Nan2* expression is one of the most differently regulated genes between these sea urchin species. *Sp-Nan2* mRNA is tightly restricted to the PGCs, whereas *Lv-Nan2* mRNA is more broadly expressed throughout the embryo. We show that this accumulation is a result of GNARLE present in the 3' UTR of *Sp*, but not in *Lv*.

RESULTS

Single-cell RNA-seq analysis of *Lv* embryos

A previous study reported the scRNA-seq analysis of fixed cells from sea urchin embryos of *Lytechinus variegatus*²⁵; the authors analyzed 18 timepoints across the first 24 h of development, capturing 50,935 total cells. Here, for the purpose of integrating and comparing scRNA-seq datasets from *S. purpuratus* and *L. variegatus*, we generated a new *Lv* scRNA-seq dataset using live cells with the same protocol and morphological stages as described for *Sp*.²⁴ We cultured *Lv* embryos and processed eight time points from 2 h postfertilization (2H) to late gastrula (LG) stage (Figure 1A). In total, the

transcriptome profiles of 63,930 cells were analyzed in this dataset (Table S1). The datasets of the eight time points were integrated into a single dataset (Lv-dataset) and cells were grouped into clusters by Uni-form Manifold Approximation and Projection (UMAP)^{26,27} (Figure S1A). However, the cells were clustered by developmental time points rather than by transcriptional profiles probably because of technical batch-effects among samples (Figures S1A and S1B). To overcome this issue, we performed batch-effect correction to align all developmental time points by using Harmony (Figure S1C).²⁸ In the Harmony-normalized Lv-dataset, cell populations derived from the eight developmental time points overlapped each other with high fidelity (Figures S1C and S1D). Thus, the batch-effect correction procedure enabled us to compare transcriptional profiles between stages and characterize cell types throughout embryogenesis.

In our Lv-dataset, we identified 23 cell states (Figures 1B and 1C). Among them, three cell states (Lv1–Lv3, Lv stands for *L. variegatus*) were identified as undifferentiated cells because these were predominantly observed in the cleavage stages (2H–8H) and did not show characteristic marker gene expression as described below (Figure S2). We also identified eleven ectodermal clusters (Lv4–Lv14), three endodermal clusters (Cluster Lv15–Lv17), five mesodermal cell states (Cluster Lv18–Lv22) and one germline cluster (Lv23). These cell states were identified according to the expression of well-described markers previously published.^{25,29–36} Markers for each of these clusters are presented in Table S2 and supplemental document 1.

A previous Lv-scrRNA seq dataset was published with 18 developmental time points analyzed to explore cell trajectories using the Waddington-OT.²⁵ Even though the normalization, the method, and the goals were different from the dataset presented here, similar cluster markers were obtained. For example, as expected, both datasets showed an enrichment of *Alx1*, *Gcm*, *Pks1*, *Scl* in the mesodermal cell states, *Endo16*, *FoxA* and *Blimp1* in the endodermal cell states, *Chordin*, *Nodal*, *FoxQ2_1* in the ectodermal cell states.

Transcriptional profile of the PGCs in Lv embryos

Our previous study revealed that *Lv-Nan2* is transcribed during early embryogenesis and first accumulated in cells in the vegetal region at the 128-cell stage.³⁷ In this current study, we employed scrRNA-seq analysis to investigate *Lv-Nan2* expression during embryogenesis. Consistent with the previous study, no or low expression of *Lv-Nan2* was detected at 2H and 4H post-fertilization but it is then upregulated at 6H and 8H post-fertilization (Figures 2A–2D). However, the expression was not enriched in a specific cell population (Figures 2C and 2D). Notably, in early (EB) and mesenchyme (MB) blastula, endodermal and mesodermal cell populations (Lv16–Lv22) tended to express higher levels of *Lv-Nan2* mRNA compared to ectodermal cell populations (Lv4–Lv15) (Figures 2E and 2F), suggesting that *Lv-Nan2* mRNA is expressed primarily in the vegetal plate, which is composed of endo-mesodermal cells and PGCs (Figure 1A). *Lv-Nan2* expression gradually became prominent in Lv23 from the EB stage onward (Figures 1C and 2E–2H), and Lv23 was thus identified as the germline. Somatic cells, though, exhibited significant *Lv-Nan2* expression, especially the undifferentiated cells (Lv2 and Lv3), animal ectoderm (Lv10 and Lv11) and mesodermal cells (Lv18 and Lv21) in the early and late gastrula stages (Figures 2G and 2H).

Because *Nan2* is one of the most extensively studied genes governing germline development, we hypothesized that the *Lv-Nan2*-expressing somatic cells have overlapped marker gene expression with the germ cells. To test this hypothesis, we analyzed how many marker genes are shared between germ cells and *Lv-Nan2*-expressing somatic cells. First, germ cells were compared with undifferentiated cell populations (Lv1–Lv3). We identified 394 marker genes enriched in the germ cells (Table S2). Among them, 141 genes (36%) were shared with all of three undifferentiated cell clusters, and 237 genes (60%) were shared with at least one cluster (Figure 2I). Next, the germ cells were compared with *Lv-Nan2*-expressing ectodermal cell populations (Lv10 and Lv11) and mesodermal cell populations (Lv18 and Lv21). In contrast to the undifferentiated cells, none, or as few as six marker genes were shared among ectodermal cell and mesodermal cell populations, respectively (Figures 2J and 2K). These results show that even though *Lv-Nan2* expression was observed in several somatic cell populations, the transcriptional profiles were clearly different from that of the germ cells. The transcriptional profile of the germ cells instead was more similar to that of the undifferentiated cell populations containing abundant maternal transcripts (Figure 2I). The cluster markers and cluster annotations for each time point are detailed in supplemental document 1.

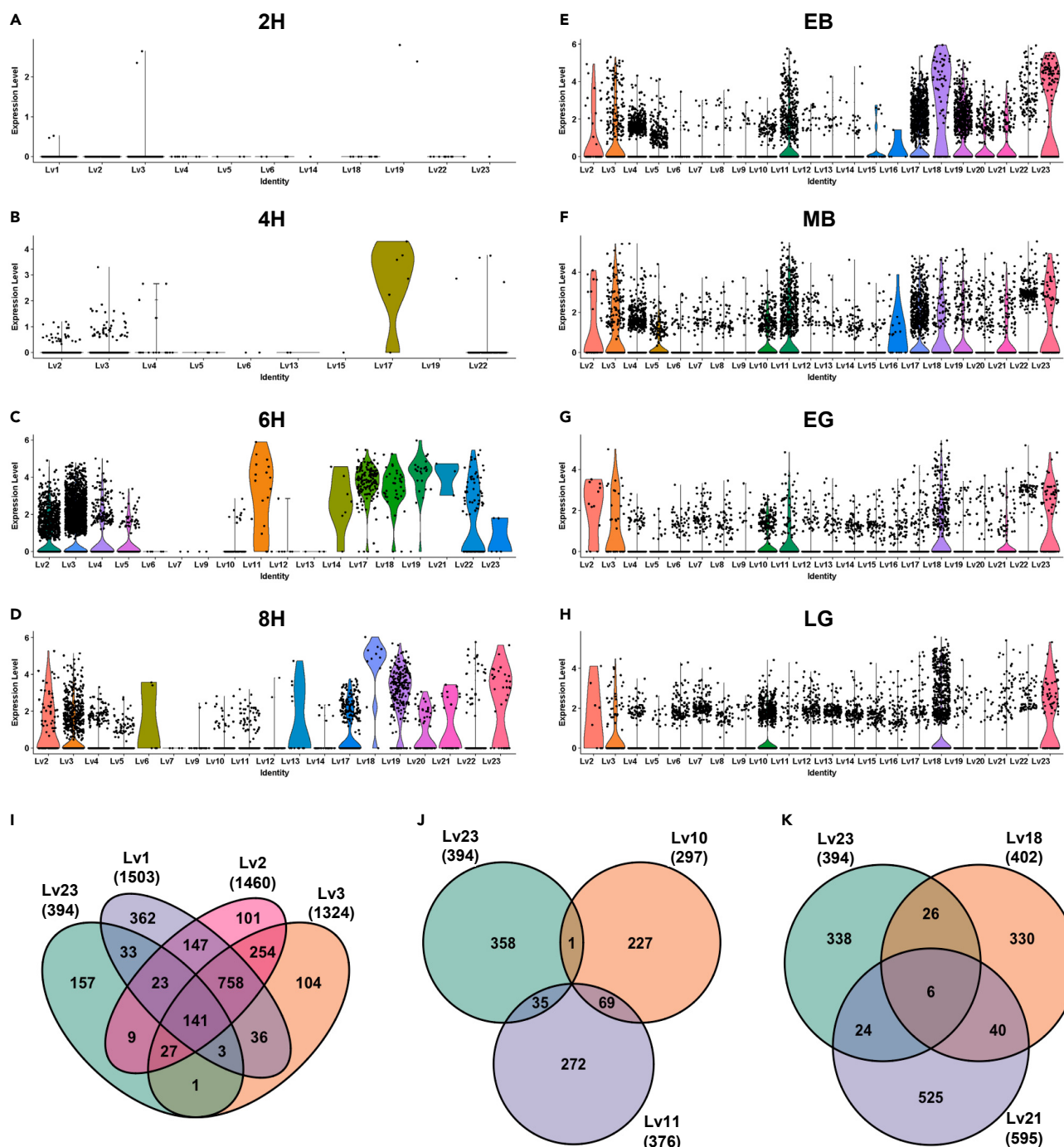


Figure 2. Transcriptional profile of *Nan2*-expressing cells in *Lv* embryos

(A–H) Violin plots showing *Lv-Nan2* expression levels of each cluster (Lv1–Lv23) in 2H (A), 4H (B), 6H (C), 8H (D), EB (E), MB (F), EG (G) and LG (H) embryo. Dots represent expression level of *Lv-Nan2* in individual cells.

(I–K) Venn diagrams showing the number of shared marker gene between germline cells (Lv23) and undifferentiated cells (I; Lv1–Lv3), ectodermal cells (J; Lv10 and Lv11) and endo/mesodermal cells (K; Lv18 and Lv21).

Three PGC subclusters were identified in *Lv* embryos

The germ cell cluster Lv23 can be subclustered into three subpopulations (Germline_Lv1–3, Lv stands for *L. variegatus*) (Figures 3A and 3A'). Germline_Lv1 appeared in the 6H embryos as the sole PGC population but its proportion eventually decreased as the embryo developed (Figures 3A' and 3B). Germline_Lv2

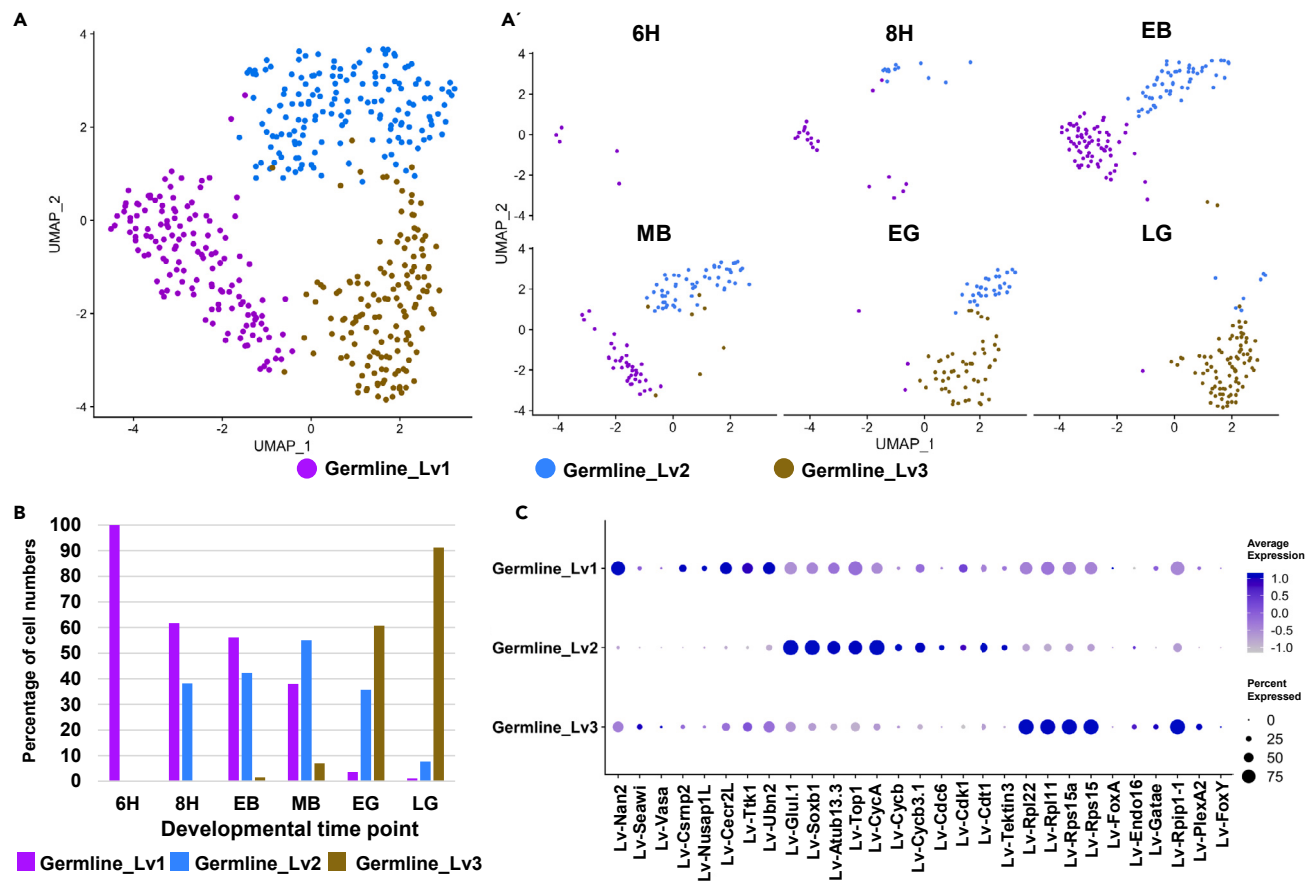


Figure 3. Gene expression changes during germline development in *Lv* embryos

(A and A') UMAP plot generated by subclustering of the germ cells (Lv23) extracted from the *Lv*-dataset (A) and UMAP plots split by developmental stages (A') are presented.

(B) The graph shows Germline_Lv1 (purple), Germline_Lv2 (blue) and Germline_Lv3 (brown) cells as a percentage of total cell number analyzed for each developmental time point.

(C) Dot plot showing marker gene expression in the germline subpopulations (Germline_Lv1–Lv3). Dot size and dot color indicate the percentage of cells expressing the gene and the average expression level, respectively.

emerged at 8H post-fertilization (Figures 3A' and 3B), its proportion peaked at mesenchyme blastula and decreased by about 7% at the late gastrula stage. The third subpopulation was first detected in the early blastula embryo. Although the proportion was less than 7% in both the EB and MB stages, it was elevated by about 60% at the EG stage, when the proportion of Germline_Lv1 was drastically decreased (Figure 3B). The proportion of Germline_Lv3 reached over 90% at the LG stage (Figure 3B).

We then analyzed differential gene expression among the three germline subpopulations. We tested well-conserved germline genes (*Lv-Nan2*, *Lv-Seawi* and *Lv-Vasa*) and germline-enriched genes (*Lv-Ttk1*, *Lv-Cecr2L*, *Lv-Csrnp2*, *Lv-Nusap1L* and *Lv-Ubn2*), which were identified as the marker genes of the germline cell population (Lv23) in this *Lv*-dataset (Figure 1C). Almost all of these genes except for *Lv-Seawi* and *Lv-Vasa* were significantly enriched in Germline_Lv1 (Figures 3C and S3 and Table S7). By contrast, Germline_Lv2 did not show significant germline gene expression (Figures 3C and S3 and Table S7). The top-5 marker genes of Germline_Lv2 included *Lv-Gli1*, *Lv-Soxb1*, *Lv-Atub13.3*, *Lv-Top1* and *Lv-CycA*, all of which were most likely to be maternally deposited transcripts, because of its enriched expression in undifferentiated cell populations (Figures 1C and 3C; Tables S2 and S7). Furthermore, cell cycle-related genes including *Lv-CycA*, *Lv-CycB*, *Lv-CycB3*, *Lv-Cdc6*, *Lv-Cdk1* and *Lv-Cdt1* were highly expressed in Germline_Lv2 (Figure 3C). Although germline genes were not statistically enriched in Germline_Lv3 when compared with other germline populations, expression levels of germline-enriched genes such as *Lv-Nan2*, *Lv-Ttk1* and *Lv-Cecr2L* were comparable with that of Germline_Lv1 (Figures 3C and S3). These

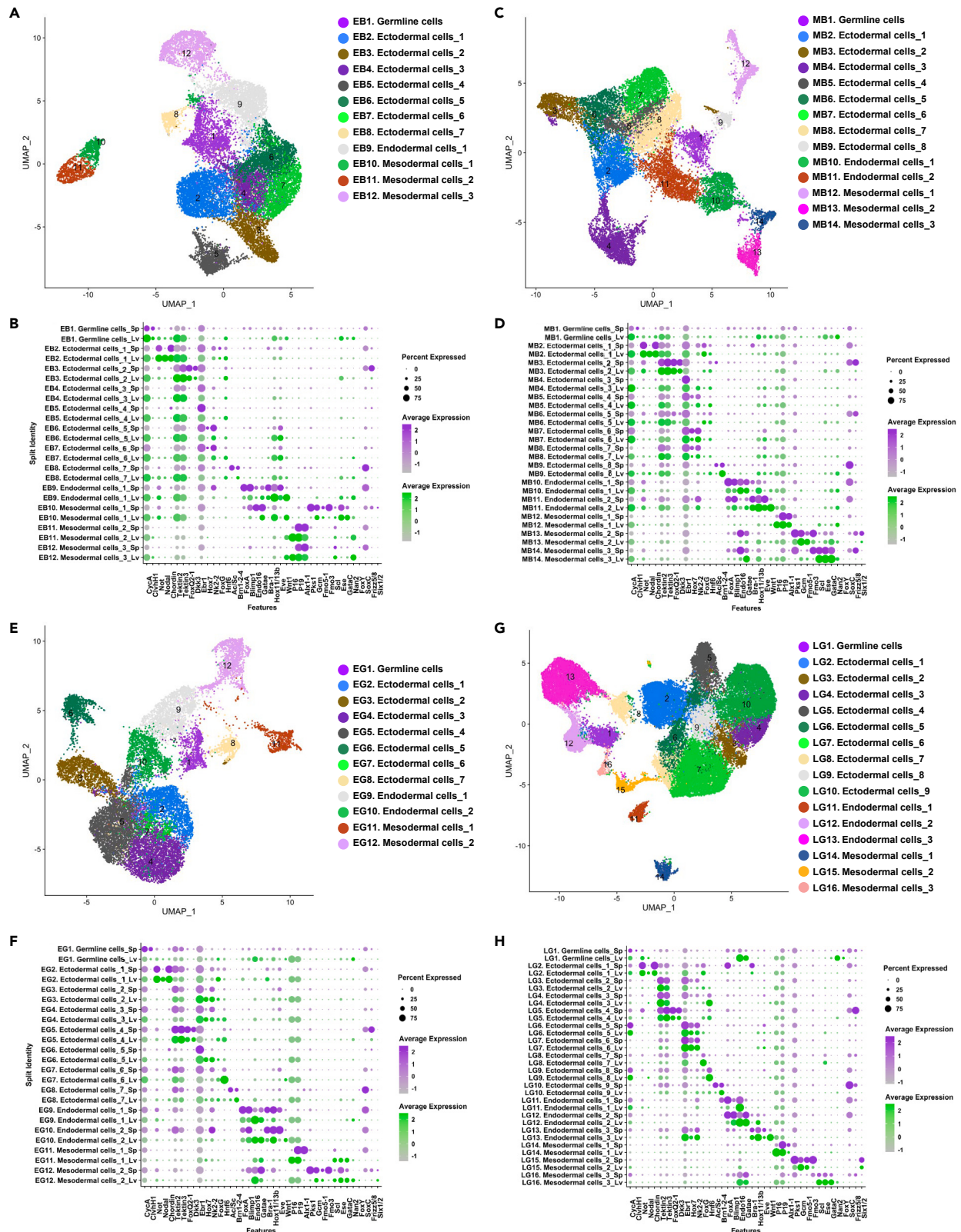


Figure 4. Marker gene expression in the *Lv-Sp* datasets

(A–H) UMAP plots (A, C, E and G) and dot plots (B, D, F and H) performed using *Lv-Sp* datasets. EB (A and B), MB (C and D), EG (E and F) and LG (G and H) embryos are analyzed individually. Dot plot represents marker gene expression in *Lv*-derived cells (green) and *Sp*-derived cells (purple) separately. Dot size and dot color indicate the percentage of cells expressing the gene and the average expression level, respectively.

results strongly suggest that Germline_Lv1 and Germline_Lv3 were *bona fide* germline cell populations. Upregulated ribosomal gene expression was observed in Germline_Lv3 (Figure 3C). Indeed, the top5 marker genes include *Lv-Rpl22*, *Lv-Rpl11*, *Lv-Rps15a*, *Lv-Rpl28* and *Lv-Rps15* (Table S7). Furthermore, the top-50 marker genes include 39 ribosomal genes (Table S7). However, because the upregulation of ribosomal genes was observed throughout the embryo (Figure S4), the upregulation is a general phenomenon observed throughout the *Lv* embryo rather than a germline-specific phenomenon. In addition, endodermal gene expression was also upregulated in the Germline_Lv3 compared to other germline cell populations (Figures 3C and S5). These results suggest that the PGC population in these embryos is dynamic in a way that may differentially influence their future fate.

Integration of both scRNA-seq datasets from *Sp* and *Lv* embryos

We employed the batch-effect correction to enable a direct comparison between this new *Lv*-dataset and the *Sp*-dataset.²⁴ We integrated the *Lv*-dataset with the *Sp*-dataset using four developmental time points individually (EB–LG; *Lv-Sp* dataset) (Figures S6A–S6D) and then normalized by Liger, another batch-effect correction tool (Figures S6I–S6L).³⁸ Harmony did not correct the batch effects between *Lv* and *Sp* datasets effectively (Figures S6E–S6H). The cluster markers and cluster annotations for each time point are detailed in supplemental document 2. In the *Lv-Sp* datasets, we identified seven to nine ectodermal cell populations (Clusters EB2–EB8, MB2–9, EG2–8 and LG2–10; Figures 4A–4H). In addition, we showed the expression patterns of *Ebr1*, *Ac/Sc* and *Brn1-2-4* as ectodermal marker genes (Figures 4B, 4D, 4F, and 4H). *Ebr1* was enriched in dorsal ectoderm in the *Lv*-dataset (Cluster Lv7–9) (Figure 1C, Table S2). *Ac/Sc* and *Brn1-2-4* are known to be expressed in proneural cells during embryogenesis.³⁹ Endodermal cell populations were identified by enriched expression of *FoxA*, *Blimp1*, *Endo16*, *Gatae*, *Bra-1*, *Hox11/13b* or *Eve* (Figures 4B, 4D, 4F, and 4H). The mesodermal populations were identified with markers such as *Alx1*, *P16*, *P19*, *Pks1*, *Gcm*, *Fmo3*, *Gatac* (Figures 4B, 4D, 4F and 4H; Tables S3–S6).

PGC comparison between *Lv* and *Sp*

In the *Lv-Sp* dataset, *Nan2* expression was observed not only in the germ cells but also in somatic cells of the *Lv* embryo (Figure S7). Therefore, *Nan2*-positive cells were extracted from the *Lv-Sp* datasets and re-analyzed in detail (Figure 5). These *Nan2*-positive cells were grouped into seven subpopulations by UMAP plots (*Nan2pos*_1–7, *Nan2pos* stands for *Nan2*-positive) (Figures 5A–5C) and were further separated into two datasets containing either *Lv*-derived (Figure 5B) or *Sp*-derived (Figure 5C) cells.

Among the seven subpopulations, *Nan2pos*_1 is the most likely to be the PGCs for the following reasons: (1) *Nan2pos*_1 showed enriched expression of *Csrnp2*, *Ttk1* and *Nusap1L*, all of which were expressed in the Germline_Lv1 of the *Lv*-dataset. *Nan2pos*_1 was detected throughout embryogenesis (Figures S9A and S9B). *Nan2pos*_5 and Germline_Lv2 were expressing common marker genes, *Soxb1*, *CycA*, *Atub13* and *CDC6* (Figures 3C and 5F; Tables S7 and S8). These results show that *Nan2pos*_1 and *Nan2pos*_5 correspond to Germline_Lv1 and Germline_Lv2, respectively. In the *Lv*-dataset, Germline_Lv3 was found as the germ cell population in EG and LG embryos (Figures 3A–3C) and *Nan2pos*_2 shows similar marker gene expression with Germline_Lv3. *Endo16*, *Gatae* and *FoxY* were enriched in *Nan2pos*_2, suggesting that it represents the Veg2 mesoderm. Importantly, this *Nan2pos*_2 population is specific to *Lv* (Figure 5B). *Nan2pos*_7 contains similar marker genes such as (*Endo16*, *Gatae*, *FoxY*) but this population is specific to *Sp*. We conclude that *Nan2pos*_2 and *Nan2pos*_7 represent the Veg2 mesoderm in *Lv* and *Sp* respectively.

In addition, somatic cells were identified based on well-studied marker gene expression as follows (Figure 5D): *Nan2pos*_3 (Secondary mesenchyme cells (SMCs): *Ese*), *Nan2pos*_4 (Veg1 lineage: *Eve*), *Nan2pos*_6 (Primary mesenchyme cells (PMCs): *Alx1*, *P19*, *P16*). Most of these *Nan2* positive somatic cell populations were abundantly detected in *Lv* (especially *Nan2pos*_3, *Nan2pos*_5, *Nan2pos*_6) but were barely detected in *Sp*, supporting that *Nan2* is predominantly expressed in the PGCs in the *Sp* embryos (Figure S7).²⁴

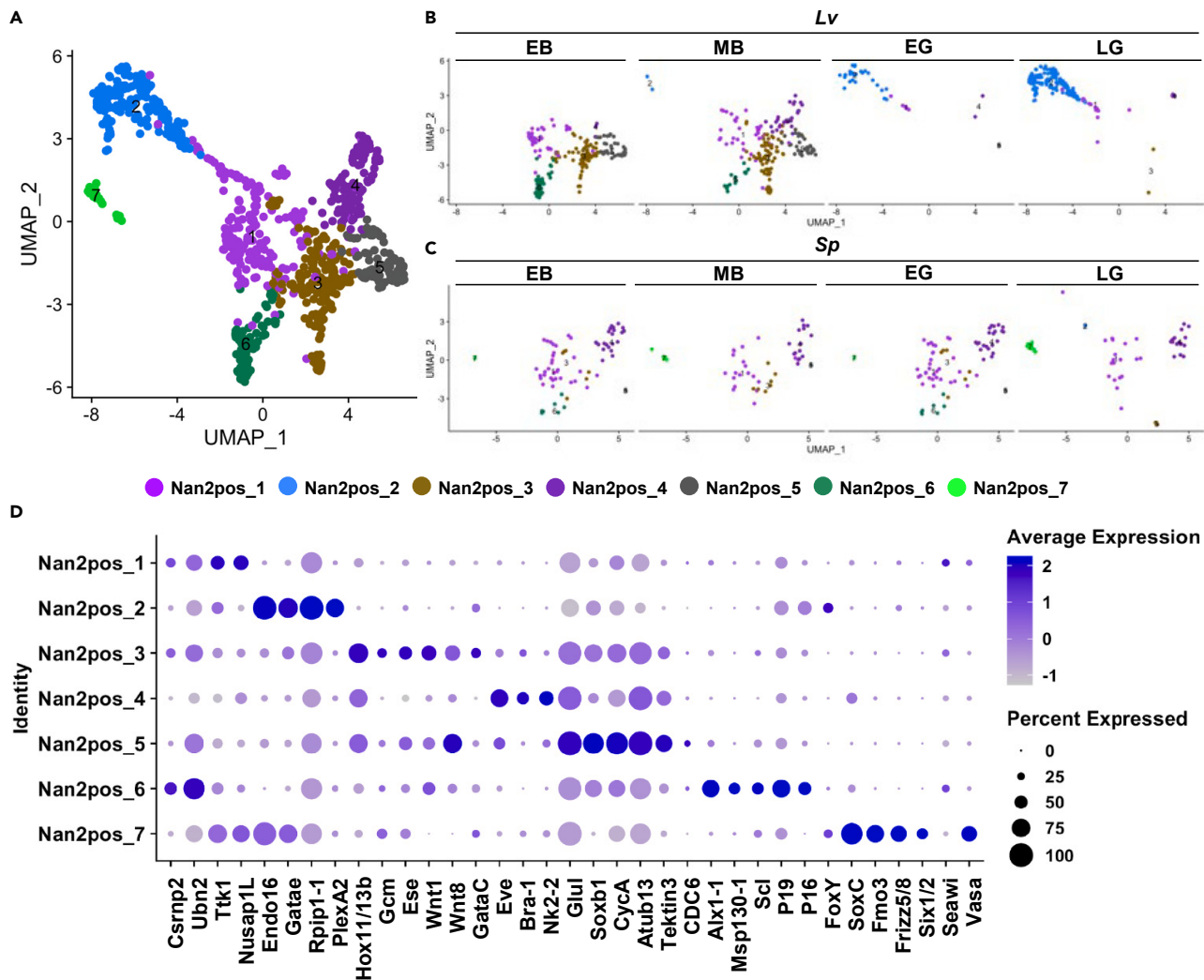


Figure 5. Germline subpopulations in the *Lv-Sp* dataset

(A) UMAP plots for Nan2-positive cells extracted from the *Lv-Sp* datasets. UMAP plots of *Lv*-derived cells (B) and *Sp*-derived cells (C) in each developmental time point (EB–LG) are presented.

(D) Dot plot showing marker gene expression in Nan2-positive cells (Nan2pos_1–7).

NPDE-mediated post-transcriptional regulation of *Nanos2*

Considering that *Nan2* mRNA was widely expressed in *Lv* embryos compared to *Sp* embryos (Figures 2A–2H and S7A–S7H), we hypothesized that *Lv-Nan2* is most likely to be regulated translationally and/or post-translationally to establish a germline-specific *Nan2* protein expression. In *Sp* embryos, *Sp-Nan2* expression is post-translationally regulated by the Nanos Protein Degradation Element (NPDE).²³ Therefore, we aimed to investigate whether the NPDE sequence is conserved among sea urchin species. For this purpose, *Nan2* protein sequences were compared between *Lv*, *Sp* and *Hemicentrotus pulcherimus* (*Hp*), a sea urchin species closely related to *Sp* (Figure S10). Although *Nan2* is highly divergent in sequence between animals, the NPDE sequence is highly conserved between *Sp-Nan2* and *Hp-Nan2* (Figure S10). Although the *Lv-Nan2* protein contains indels in the NPDE, the C-terminal region is completely conserved (Figure S10, aa 36–46; ITELKVMRG). These results led us to hypothesize that the well-conserved C-terminal region is functionally important among these species.

To test this hypothesis, we divided the *Sp-Nan2* NPDE element into eight fragments each consisting of 13–22 amino acids (Figure 6A) and the corresponding nucleotide sequences were fused to the GFP ORF and *Sp-Nan2* 3'UTR without the GNARLE sequence (Δ GNARLE). The GNARLE is required to restrict mRNAs to

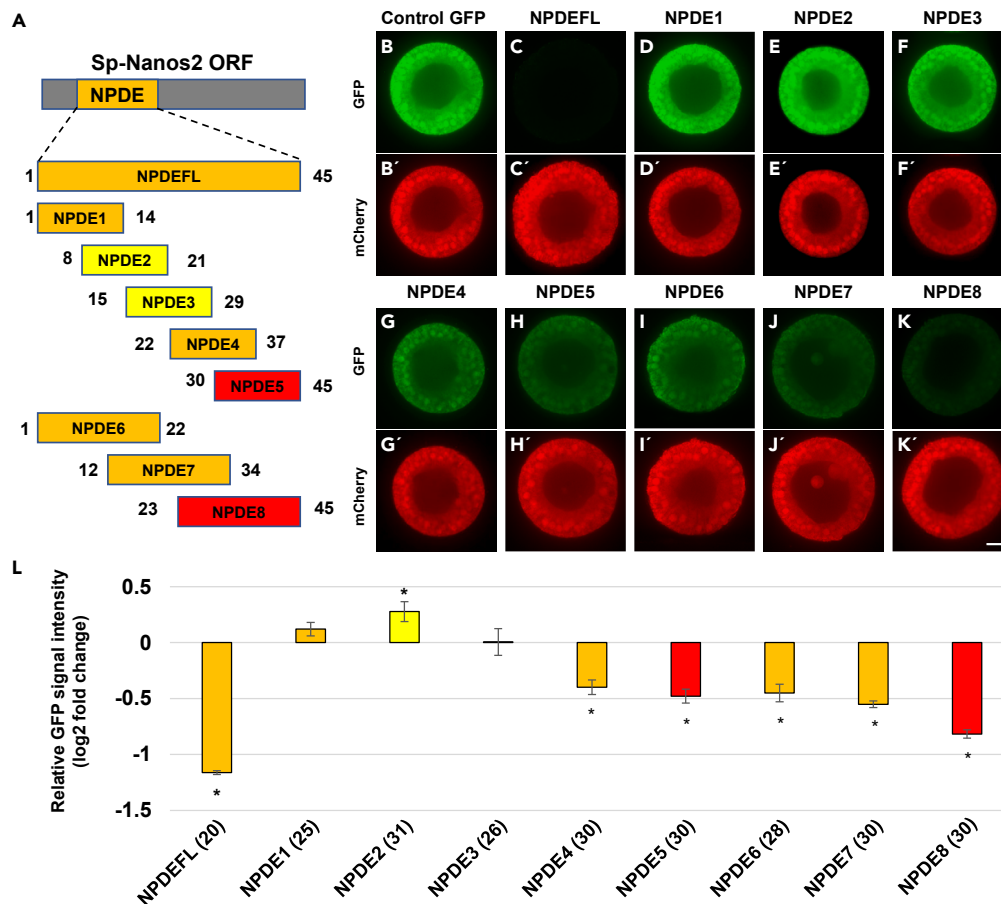


Figure 6. NPDE-mediated protein degradation in *Sp* embryos

(A) NPDE located at the N-terminal side of *Sp-Nan2* ORF (gray box) was divided into eight fragments containing 14–16 amino acids (aa) (NPDE1–5) and 22–23 aa (NPDE6–8). Amino acid position within the NPDE is shown at both side of each fragment. The fragments are color-coded by the percentage of amino acid residue conserved between *Sp-Nan2* and *Lv-Nan2* (red: > 60%; orange: 30–60%; yellow: < 30%).

(B–K and B'–K') *Sp* embryos were injected with Control GFP or GFP-NPDE fusion mRNAs and mCherry mRNA.

Fluorescence of GFP (B–K; green) and mCherry (B'–K'; red) were observed in the EB stage of *Sp* embryos. Scale bar: 20 μ m.

(L) Relative signal intensity against the Control GFP is presented. GFP signal intensity was normalized with mCherry signal intensity within the same embryo. The normalized GFP signal intensities of GFP-NPDE fusion proteins were compared with that of Control GFP. Graph is color-coded by the percentage of amino acid residue conserved between *Sp-Nan2* and *Lv-Nan2* as described above. At least 50–100 embryos were injected in each experiment, and at least 20 embryos were used for each quantification. Significance was calculated between Control GFP and GFP-NPDE fusion proteins (NPDEFL and NPDE1–8) by Student's *t* test (*: *p* < 0.05). Error bars indicate standard errors of samples. The number of embryos (n) examined are shown in parentheses.

the PGCs.²¹ The mRNAs containing *Sp-Nan2* 3' UTR Δ G_{NARLE} are not regulated post-transcriptionally, and in turn are distributed in both somatic and germ cells. These *in vitro* synthesized mRNAs were injected into *Sp* embryos, together with mCherry mRNA. The expression of these fused proteins (GFP-NPDEFL and GFP-NPDE1–8) were observed in blastula stage embryos and then normalized by the signal intensity of mCherry protein. The normalized signal intensity of GFP-NPDE fusion proteins were compared to GFP without the NPDE fragment (Control GFP) (Figures 6B, 6B', and 6L). As reported previously, expression of the full length NPDE (GFP-NPDEFL) was barely detectable (Figures 6C and 6C').²³ In addition, fusion with NPDE4–8 significantly decreased the GFP signal intensity compared to the Control GFP, whereas fusion with NPDE1–3 did not attenuate GFP signal intensity (Figures 6D–6L and 6D'–6K'). Notably, fragments including the highly conserved C-terminal region (GFP-NPDE5 and 8) exhibited a significant decrease in GFP intensity but were never comparable to the full length NPDE (Figures 6C, 6C', 6K, and

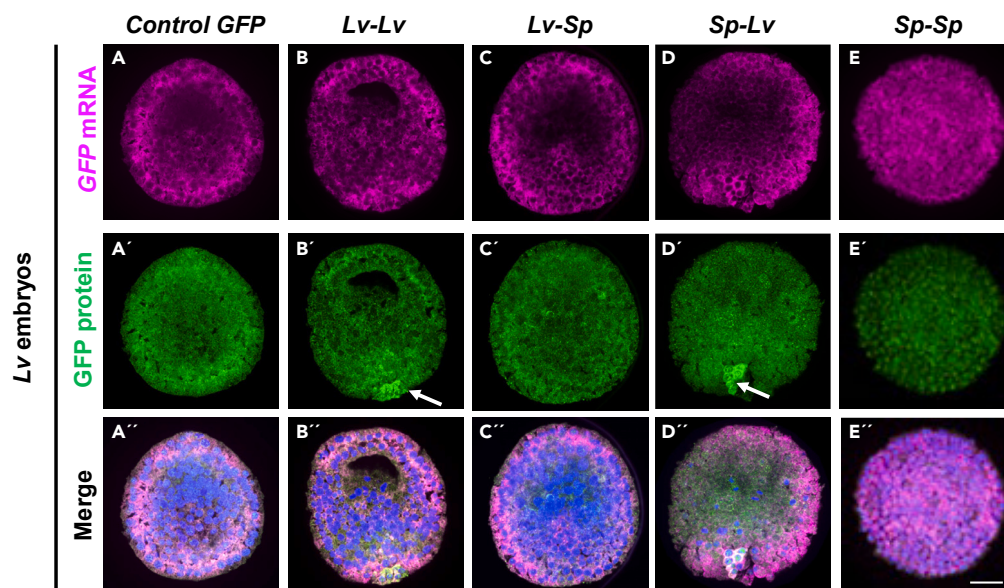


Figure 7. *Lv-Nan2* 3'UTR regulates its protein expression

Lv embryos injected with *Control GFP* mRNA (A), *Lv-Lv* mRNA (B), *Lv-Sp* mRNA (C), *Sp-Lv* mRNA (D) and *Sp-Sp* mRNA (E) were labeled by whole mount *in situ* hybridization (magenta; A–E) and immunostaining for GFP protein (green; A'–E'). DNA staining (blue) merged with *in situ* hybridization and immunostaining are presented (A''–E''). White arrows in B' and D' indicate the specific translation of the constructs in the germ cells. Scale bar: 20 μ m.

6K'). These results suggest that the C terminal region of the NPDE (Figure S10; amino acids 36 to 45) is important but not sufficient to regulate Nan2 protein expression.

Regulation of *Lv-Nan2* expression

To determine whether the *cis*- and *trans*-regulatory mechanisms of *Sp-Nan2* mRNA are conserved in *Lv*, we injected GFP mRNAs fused with *Sp-Nan2* ORF and *Sp-Nan2* 3' UTR (*Sp-Sp*), *Lv-Nan2* ORF and *Lv-Nan2* 3' UTR (*Lv-Lv*), and the hybrid mRNAs (*Sp-Lv* and *Lv-Sp*) into *Lv* embryos. GFP ORF fused with β -globin UTRs (*Control GFP*) was used as the control. We observed the GFP mRNA and protein expression at blastula stage (Figure 7). *Control GFP*, *Lv-Sp* and *Sp-Sp* did not show enriched GFP mRNA or protein in the *Lv* embryos (Figures 7A–7A'', 7C–7C'', and 7E–7E''). These results suggest that the regulatory elements of *Sp-Nan2* are not functional in *Lv* embryos. In addition, and as expected by the low overall conservation of the NPDE of *Lv-Nan2*, we discovered that the *Lv-Nan2* ORF is not sufficient for its protein enrichment in the PGCs (*Lv-Sp*). Of interest, when mRNAs containing *Lv-Nan2* 3'UTR (*Lv-Lv* and *Sp-Lv*) were injected into these *Lv* embryos, the translated protein was enriched in the PGCs, whereas the corresponding RNAs were not (Figures 7B–7B'' and 7D–7D''). This result shows that the *Lv-Nan2* 3' UTR is sufficient to establish germline-specific expression by translational regulation in the *Lv* embryos.

DISCUSSION

Germline gene expression in *Lv* embryos

We found that the transcriptional profile of the PGCs in *Lv* was quite similar with that of undifferentiated cells of the early embryo, and they contained no prominent marker genes of germ cells except for the *Lv-Nan2*. This may be explained by previous studies in the closely related sea urchin species, *S. purpuratus*. In *Sp*, CNOT6 transcripts encoding a deadenylase are depleted from the PGCs which creates a stable environment for maternally inherited mRNAs.⁸ This stable environment is likely to be established in both *Lv* and *Sp* germ cells because maternal mRNAs are also retained in *Lv* germ cells. In *Sp*, transcriptional, RNA degradation and translational activities are globally repressed in the PGCs.^{8,9} Quiescence is a widely conserved feature of germ cells, perhaps to protect them from somatic differentiation by repressing somatic gene expression. Considering that *Sp-Nan2* is the key regulator to establish quiescence, *Lv-Nan2* may be responsible for the transcriptional profile of the germ cells.

Our results indicate the presence of two germline subclusters throughout development, an early Germline_Lv1 and a later Germline_Lv3. In the *Sp* embryos, four PGCs are formed early during development, their cell cycle is downregulated until gastrulation when the level of *Sp-Nan2* starts to decrease and that the PGCs finally divide once to give rise to 8 cells. These 2 subclusters could represent the quiescent PGCs (Germline_Lv1), becoming more active during gastrulation (Germline_Lv3). Comparing the differential gene expression between these two subclusters overtime could reveal new mechanisms of how PGCs exit quiescence during gastrulation.

Comparison of *Nan2* regulation between *Sp* and *Lv* sea urchins

In this study, we found that *Lv-Nan2* mRNA is expressed not only in the PGCs but also in the somatic cells including the three germ layers. In the sea urchin *Sp*, *Nan2* mRNA and protein are highly restricted to the PGCs at blastula stage. *Sp-Nan2* is transcribed broadly by the Wnt pathway,¹⁹ but the resulting mRNA is quickly degraded outside of the PGCs and retained only in the PGCs. The presence of an element, GNARLE, on its 3'UTR is required for this regulation.²¹ In contrast, *Lv-Nan2* mRNA is not restricted to the PGCs and its 3'UTR does not have a conserved GNARLE. *Sp-Nan2* expression is also regulated at the protein level by the NPDE. This element is not highly conserved and is not functional in *Lv* embryos (Figure S11). Sea urchins like *Sp* and *Lv* rely on inherited mechanisms to specify their germline. In contrast, sea stars use inductive mechanisms.^{40,41} In the sea star *Patiria miniata* (*Pm*), *Nanos3* is enriched in the germline whereas *Nanos2* is only transiently expressed in the vegetal pole of the embryo during gastrulation. The strategy used by the sea star in *Nanos* regulation is markedly distinct from those used in the sea urchin. In the sea star, the *Nanos* that will become specific to the germline - *PmNanos3* - is expressed broadly in the endomesoderm during gastrulation, and progressively becomes restricted to the site of germ line formation by *PmNanos3* mRNA degradation and transcriptional inhibition induced by Nodal signaling. The GNARLE and the NPDE found in *Sp* *Nanos 2* are not present in *Pm* *Nanos3*, illustrating the vast diversity of mechanisms leading to *Nanos* expression in the germ cells throughout Echinoderms.

Here, we discovered that instead of relying on RNA and/or protein stability, the germline expression of *Lv-Nan2* mostly depends on translational regulation. The presence of the *Lv-Nan2* 3'UTR on an exogenous RNA leads to a significant restriction of the corresponding protein in the PGCs. We conclude that whereas its mRNA is present broadly in the embryos, it is predominantly translated in the PGCs. *Nanos* has previously been shown in other animals to also be regulated at the translational level: in *Xenopus*, *Caenorhabditis elegans*, and *Drosophila*.^{5,17,42-47} Taken together, translational regulation is most likely to be the ancestral mechanism widely conserved between animals including *Lv*. By contrast, post-transcriptional and post-translational regulations play a key role in *Sp*. The selection for these diverse regulatory mechanisms is still unknown.

Somatic cell gene expression

Nan2 expression was not the only difference observed between both datasets. Several somatic genes were also differentially expressed between *Sp* and *Lv*. For example, endodermal cell populations were identified by enriched expression of *FoxA*, *Blimp1*, *Endo16*, *Gatae*, *Bra-1*, *Hox11/13b* and *Eve* in the *Lv-Sp* dataset (Figure 4). At the early gastrula stage, *Endo16* is more abundant in *Lv* compared to *Sp* (cluster EG9). Similar results were observed in the ectodermal clusters. An obvious example is *Hnf6* (cluster EG7), that is also found more abundantly in *Lv* compared to *Sp*. These are just a few examples suggesting that different mechanisms also could be used by these two sea urchins to regulate the expression of their somatic genes through development.

Nan2 mRNA was detected in the Veg2 mesoderm

In the *Lv-Sp* dataset, seven subpopulations of *Nan2*-positive cells have been characterized (Table S8). Importantly, *Nan2pos_7* is only detected in the *Sp* sea urchin. *Nan2pos_2* is only detected in the *Lv* sea urchin. Both of these populations reached their peak of abundance at the late gastrula stage. They both express markers such as *FoxY* and *Endo16* (Figures 5A–5D), suggesting that both of these populations represent the Veg2 mesoderm. In *Sp* embryos of which micromeres, parent cells of PGCs, are surgically removed, new germline cells are regenerated during embryogenesis and then most of the resulting adults generate functional gametes.⁴⁸ Our previous data also suggested that in *Sp*, the Veg2 mesoderm could be important to form new germline cells after micromere removal.^{20,49} Micromere-deleted *Lv* embryos can also regenerate their germline but use distinct mechanisms.⁴⁹ For example, in contrast to *Lv*, micromere-deleted *Sp* embryos show significant upregulation of *Vasa* protein expression. Our data here suggest

that these two sea urchins respond differently to micromere removal because of their differential gene expression in the Veg2 mesoderm that do not overlap in the UMAP: *Sp* Nan2pos_7 and *Lv* Nan2pos_2 (Figures 5A–5C). *Nan2* expression may be optimized for germline regeneration in *Sp* embryos.

Evolution of germline gene regulation

We showed that *Nan2* is differentially expressed between both urchins. It is highly enriched in the germline cluster in *Sp*, but it is more broadly expressed in *Lv* (Figures S7A–S7Q). However, this observation is not a common characteristic of all the germline genes. For example, the expression of *Vasa* and *Seawi* mRNAs do not follow the same pattern; instead, their expression is similar in both sea urchins (Figure S12). *Nanos* ectopic expression in embryos is toxic,¹⁷ so animals have found various methods to restrict its expression to the germ cells. In both sea urchins, the 3'UTR of *Nan2* is essential but the regulatory mechanisms are different. In *Sp*, this 3'UTR is essential to degrade *Nan2* mRNA in the somatic cells, but in *Lv*, this 3'UTR instead leads to its specific translation in the germ cells. How these regulations happen is still unknown. Further studies will include the identification of the proteins associated with each of these 3'UTRs.

Limitations of the study

Direct integration of scRNA-seq datasets derived from *Lythechinus variegatus* and *S. purpuratus* enabled us to compare where and when orthologous genes were expressed between closely related sea urchins. However, we did not perform statistical analyses of gene expression levels between these species. The expression level, including Transcript Per Million (TPM), is highly sensitive to both transcript length and quality of gene models, which are significantly different between species. Furthermore, we did not analyze the molecular mechanisms by which *Lv-Nan2* establishes specific germline protein expression. Future studies should seek to identify the regulatory elements resident in the *Lv-Nan2* 3' UTR for this mechanism.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Animals
- METHOD DETAILS
 - *Lv* culture and dissociation for scRNA-seq
 - Genome-guided *de novo* transcriptome assembly
 - Prediction of UTR information
 - Prediction of conventional gene names
 - Single cell RNA sequencing
 - Comparison between *Lv* and *Sp* datasets
 - Plasmid construction
 - Microinjections
 - Whole mount *in situ* hybridization
 - Immunostaining
 - Sequence comparison between species
 - Statistical analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.106402>.

ACKNOWLEDGMENTS

We are grateful to the NIH (1R35GM140897, GMW and 1P20GM119943, NO), the NSF (IOS-1923445, GMW) and the Toyobo Biotechnology Foundation long term research grant (to SM) for the support to conduct this research. We dedicate this work to the memory of our friend, colleague, and scientific spiritual leader, Andy

Cameron. Many moons ago, he graciously helped us begin this pathway of friendship with the germ line. May his spirit remain comfortably on E23.

AUTHOR CONTRIBUTIONS

All authors have approved this manuscript submission.

Conceptualization, S.M., N.O., S.F., and G.M.W.; Methodology, S.M., N.O., S.F., and G.M.W.; Investigation, S.M., N.O., and S.F.; Writing – Original Draft, S.M., N.O., and G.M.W.; Writing – Review and Editing, S.M., N.O., S.F., and G.M.W.; Funding Acquisition, N.O. and G.M.W.; Resources, N.O. and G.M.W.; Supervision, G.M.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper received support from a program designed to increase minority representation in their field of research.

Received: December 5, 2022

Revised: January 20, 2023

Accepted: March 10, 2023

Published: March 15, 2023

REFERENCES

1. Extavour, C.G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130, 5869–5884. <https://doi.org/10.1242/dev.00804>.
2. 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. <https://doi.org/10.1016/j.ydbio.2009.10.030>.
3. Johnson, A.D., Richardson, E., Bachvarova, R.F., and Crother, B.I. (2011). Evolution of the germ line-soma relationship in vertebrate embryos. *Reproduction* 141, 291–300. <https://doi.org/10.1530/REP-10-0474>.
4. Kobayashi, S., Yamada, M., Asaoka, M., and Kitamura, T. (1996). Essential role of the posterior morphogen nanos for germline development in *Drosophila*. *Nature* 380, 708–711. <https://doi.org/10.1038/380708a0>.
5. Subramaniam, K., and Seydoux, G. (1999). nos-1 and nos-2, two genes related to *Drosophila* nanos, regulate primordial germ cell development and survival in *Caenorhabditis elegans*. *Development* 126, 4861–4871. <https://doi.org/10.1242/dev.126.21.4861>.
6. Köprunner, M., Thisse, C., Thisse, B., and Raz, E. (2001). A zebrafish nanos-related gene is essential for the development of primordial germ cells. *Genes Dev.* 15, 2877–2885. <https://doi.org/10.1101/gad.212401>.
7. Tsuda, M., Sasaoka, Y., Kiso, M., Abe, K., Haraguchi, S., Kobayashi, S., and Saga, Y. (2003). Conserved role of nanos proteins in germ cell development. *Science* 301, 1239–1241. <https://doi.org/10.1126/science.1085222>.
8. Swartz, S.Z., Reich, A.M., Oulhen, N., Raz, T., Milos, P.M., Campanale, J.P., Hamdoun, A., and Wessel, G.M. (2014). Deadenylase depletion protects inherited mRNAs in primordial germ cells. *Development* 141, 3134–3142. <https://doi.org/10.1242/dev.110395>.
9. 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. <https://doi.org/10.1242/dev.144170>.
10. Lai, F., Singh, A., and King, M.L. (2012). *Xenopus* Nanos1 is required to prevent endoderm gene expression and apoptosis in primordial germ cells. *Development* 139, 1476–1486. <https://doi.org/10.1242/dev.079608>.
11. Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K., and Kobayashi, S. (1999). Maternal Pumilio acts together with Nanos in germline development in *Drosophila* embryos. *Nat. Cell Biol.* 1, 431–437. <https://doi.org/10.1038/15666>.
12. Dalby, B., and Glover, D.M. (1993). Discrete sequence elements control posterior pole accumulation and translational repression of maternal cyclin B RNA in *Drosophila*. *EMBO J.* 12, 1219–1227. <https://doi.org/10.1002/j.1460-2075.1993.tb05763.x>.
13. Kadyrova, L.Y., Habara, Y., Lee, T.H., and Wharton, R.P. (2007). Translational control of maternal Cyclin B mRNA by Nanos in the *Drosophila* germline. *Development* 134, 1519–1527. <https://doi.org/10.1242/dev.002212>.
14. Murata, Y., and Wharton, R.P. (1995). Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in *Drosophila* embryos. *Cell* 80, 747–756. [https://doi.org/10.1016/0092-8674\(95\)90353-4](https://doi.org/10.1016/0092-8674(95)90353-4).
15. Sato, K., Hayashi, Y., Ninomiya, Y., Shigenobu, S., Arita, K., Mukai, M., and Kobayashi, S. (2007). Maternal Nanos represses hid/skl-dependent apoptosis to maintain the germ line in *Drosophila* embryos. *Proc. Natl. Acad. Sci. USA* 104, 7455–7460. <https://doi.org/10.1073/pnas.0610052104>.
16. Wreden, C., Verrotti, A.C., Schisa, J.A., Lieberfarb, M.E., and Strickland, S. (1997). Nanos and pumilio establish embryonic polarity in *Drosophila* by promoting posterior deadenylation of hunchback mRNA. *Development* 124, 3015–3023. <https://doi.org/10.1242/dev.124.15.3015>.
17. 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. <https://doi.org/10.1242/dev.056705>.
18. Oulhen, N., and Wessel, G. (2017). A quiet space during rush hour: quiescence in primordial germ cells. *Stem Cell Res.* 25, 296–299. <https://doi.org/10.1016/j.scr.2017.11.001>.

19. Pieplow, A., Dastaw, M., Sakuma, T., Sakamoto, N., Yamamoto, T., Yajima, M., Oulhen, N., and Wessel, G.M. (2021). CRISPR-Cas9 editing of non-coding genomic loci as a means of controlling gene expression in the sea urchin. *Dev. Biol.* 472, 85–97. <https://doi.org/10.1016/j.ydbio.2021.01.003>.
20. Oulhen, N., Swartz, S.Z., Wang, L., Wikramanayake, A., and Wessel, G.M. (2019). Distinct transcriptional regulation of Nanos2 in the germ line and soma by the Wnt and delta/notch pathways. *Dev. Biol.* 452, 34–42. <https://doi.org/10.1016/j.ydbio.2019.04.010>.
21. Oulhen, N., Yoshida, T., Yajima, M., Song, J.L., Sakuma, T., Sakamoto, N., Yamamoto, T., and Wessel, G.M. (2013). The 3'UTR of nanos2 directs enrichment in the germ cell lineage of the sea urchin. *Dev. Biol.* 377, 275–283. <https://doi.org/10.1016/j.ydbio.2013.01.019>.
22. Oulhen, N., and Wessel, G.M. (2014). Every which way—nanos gene regulation in echinoderms. *Genesis* 52, 279–286. <https://doi.org/10.1002/dvg.22737>.
23. Oulhen, N., and Wessel, G.M. (2016). Differential Nanos 2 protein stability results in selective germ cell accumulation in the sea urchin. *Dev. Biol.* 418, 146–156. <https://doi.org/10.1016/j.ydbio.2016.07.007>.
24. Foster, S., Oulhen, N., and Wessel, G. (2020). Asingle cell RNA sequencing resource for early sea urchin development. *Development* 147. <https://doi.org/10.1242/dev.191528>.
25. Massri, A.J., Greenstreet, L., Afanassiev, A., Berrio, A., Wray, G.A., Schiebinger, G., and McClay, D.R. (2021). Developmental single-cell transcriptomics in the lytechinus variegatus sea urchin embryo. *Development* 148. <https://doi.org/10.1242/dev.198614>.
26. Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37, 38–44. <https://doi.org/10.1038/nbt.4314>.
27. McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold approximation and projection. *J. Open Source Softw.* 3, 861.
28. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16, 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>.
29. Duboc, V., Röttinger, E., Lapraz, F., Besnardeau, L., and Lepage, T. (2005). Left-right asymmetry in the sea urchin embryo is regulated by nodal signaling on the right side. *Dev. Cell* 9, 147–158. <https://doi.org/10.1016/j.devcel.2005.05.008>.
30. Angerer, L.M., Dolecki, G.J., Gagnon, M.L., Lum, R., Wang, G., Yang, Q., Humphreys, T., and Angerer, R.C. (1989). Progressively restricted expression of a homeo box gene within the aboral ectoderm of developing sea urchin embryos. *Genes Dev.* 3, 370–383. <https://doi.org/10.1101/gad.3.3.370>.
31. Angerer, L., Hussain, S., Wei, Z., and Livingston, B.T. (2006). Sea urchin metalloproteases: a genomic survey of the BMP-1/toll-like, MMP and ADAM families. *Dev. Biol.* 300, 267–281. <https://doi.org/10.1016/j.ydbio.2006.07.046>.
32. 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. <https://doi.org/10.1016/j.ydbio.2006.09.031>.
33. Otim, O., Amore, G., Minokawa, T., McClay, D.R., and Davidson, E.H. (2004). SpHnf6, a transcription factor that executes multiple functions in sea urchin embryogenesis. *Dev. Biol.* 273, 226–243. <https://doi.org/10.1016/j.ydbio.2004.05.033>.
34. 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. <https://doi.org/10.1242/dev.02577>.
35. Etensohn, 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. <https://doi.org/10.1242/dev.00511>.
36. Cheers, M.S., and Etensohn, C.A. (2005). P16 is an essential regulator of skeletogenesis in the sea urchin embryo. *Dev. Biol.* 283, 384–396. <https://doi.org/10.1016/j.ydbio.2005.02.037>.
37. Fresques, T., Swartz, S.Z., Juliano, C., Morino, Y., Kikuchi, M., Akasaka, K., Wada, H., Yajima, M., and Wessel, G.M. (2016). The diversity of nanos expression in echinoderm embryos supports different mechanisms in germ cell specification. *Evol. Dev.* 18, 267–278. <https://doi.org/10.1111/ede.12197>.
38. Welch, J., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E. (2018). Integrative inference of brain cell similarities and differences from single-cell genomics. Preprint at bioRxiv. <https://doi.org/10.1101/459891>.
39. Slota, L.A., and McClay, D.R. (2018). Identification of neural transcription factors required for the differentiation of three neuronal subtypes in the sea urchin embryo. *Dev. Biol.* 435, 138–149. <https://doi.org/10.1016/j.ydbio.2017.12.015>.
40. Fresques, T.M., and Wessel, G.M. (2018). Nodal induces sequential restriction of germ cell factors during primordial germ cell specification. *Development* 145. <https://doi.org/10.1242/dev.155663>.
41. Foster, S., Oulhen, N., Fresques, T., Zaki, H., and Wessel, G. (2022). Single-cell RNA-sequencing analysis of early sea star development. *Development* 149. <https://doi.org/10.1242/dev.200982>.
42. D'Agostino, I., Merritt, C., Chen, P.L., Seydoux, G., and Subramaniam, K. (2006). Translational repression restricts expression of the C. elegans Nanos homolog NOS-2 to the embryonic germline. *Dev. Biol.* 292, 244–252. <https://doi.org/10.1016/j.ydbio.2005.11.046>.
43. Forrest, K.M., and Gavis, E.R. (2003). Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. *Curr. Biol.* 13, 1159–1168. [https://doi.org/10.1016/s0960-9822\(03\)00451-2](https://doi.org/10.1016/s0960-9822(03)00451-2).
44. Forrest, K.M., Clark, I.E., Jain, R.A., and Gavis, E.R. (2004). Temporal complexity within a translational control element in the nanos mRNA. *Development* 131, 5849–5857. <https://doi.org/10.1242/dev.01460>.
45. Gavis, E.R., and Lehmann, R. (1992). Localization of nanos RNA controls embryonic polarity. *Cell* 71, 301–313. [https://doi.org/10.1016/0092-8674\(92\)90358-j](https://doi.org/10.1016/0092-8674(92)90358-j).
46. Jadhav, S., Rana, M., and Subramaniam, K. (2008). Multiple maternal proteins coordinate to restrict the translation of C. elegans nanos-2 to primordial germ cells. *Development* 135, 1803–1812. <https://doi.org/10.1242/dev.013656>.
47. Kalifa, Y., Huang, T., Rosen, L.N., Chatterjee, S., and Gavis, E.R. (2006). Glorund, a Drosophila hnRNP F/H homolog, is an ovarian repressor of nanos translation. *Dev. Cell* 10, 291–301. <https://doi.org/10.1016/j.devcel.2006.01.001>.
48. Yajima, M., and Wessel, G.M. (2011). Small micromeres contribute to the germline in the sea urchin. *Development* 138, 237–243. <https://doi.org/10.1242/dev.054940>.
49. Voronina, 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. <https://doi.org/10.1016/j.ydbio.2007.11.039>.
50. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
51. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
52. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>.
53. Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the Arabidopsis

- p>genome annotation using maximal transcript alignment assemblies.
- Nucleic Acids Res.*
- 31, 5654–5666.
- <https://doi.org/10.1093/nar/gkg770>
- .
54. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>.
 55. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. <https://doi.org/10.1038/nmeth.2019>.
 56. McClay, D.R. (2004). Methods for embryo dissociation and analysis of cell adhesion. *Methods Cell Biol.* 74, 311–329. [https://doi.org/10.1016/s0091-679x\(04\)74014-5](https://doi.org/10.1016/s0091-679x(04)74014-5).
 57. Oulhen, N., Foster, S., Wray, G., and Wessel, G. (2019). Identifying gene expression from single cells to single genes. *Methods Cell Biol.* 151, 127–158. <https://doi.org/10.1016/bs.mcb.2018.11.018>.
 58. Li, Y., Omori, A., Flores, R.L., Satterfield, S., Nguyen, C., Ota, T., Tsurugaya, T., Ikuta, T., Ikeo, K., Kikuchi, M., et al. (2020). Genomic insights of body plan transitions from bilateral to pentameral symmetry in Echinoderms. *Commun. Biol.* 3, 371. <https://doi.org/10.1038/s42003-020-1091-1>.
 59. 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. <https://doi.org/10.1038/nbt.4096>.
 60. Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217. <https://doi.org/10.1006/jmbi.2000.4042>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-GFP	Abcam	Cat# ab6556
anti-Digoxigenin-POD	Roche	Cat# 11207733910
Alexa Fluor 488-conjugated goat anti-rabbit	Molecular Probe	Cat# A-11034
Streptavidin-Alexa Fluor 546-conjugated	Invitrogen	Cat# S11225
Biological samples		
<i>Lytechinus variegatus</i>	Pelagic Corporation and Duke University marine lab	N/A
Deposited data		
Raw data of single-cell RNA-seq analysis	This paper	GSE208709
Recombinant DNA		
GFP-NPDEFL	This paper	N/A
GFP-NPDE1-8	This paper	N/A
Lv-Lv	This paper	N/A
Lv-Sp	This paper	N/A
Sp-Lv	This paper	N/A
Sp-Sp	This paper	N/A
Software and algorithms		
Trimmomatic	Bolger et al., 2014 ⁵⁰	http://www.usadellab.org/cms/?page=trimmomatic
Hisat2	Kim et al., 2019 ⁵¹	http://daehwankimlab.github.io/hisat2/
Trinity	Grabherr et al., 2011 ⁵²	https://github.com/trinityrnaseq/trinityrnaseq
PASA	Haas et al., 2003 ⁵³	https://github.com/PASApipeline/PASApipeline
Seurat	Stuart et al., 2019 ⁵⁴	https://satijalab.org/seurat/index.html
Harmony	Korsunsky et al., 2019 ²⁸	https://github.com/immunogenomics/harmony
Liger	Welch et al., 2018 ³⁸	https://github.com/welch-lab/liger
Fiji	Schindelin et al., 2012 ⁵⁵	https://ImageJ.net/Fiji

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Gary M. Wessel (rhet@brown.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The sequencing files and gene expression matrices for the single-cell RNA-seq analysis presented here have been deposited at NCBI GEO accession: GSE208709.
- This paper does not report original code.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

Adult *Strongylocentrotus purpuratus* were obtained from Pete Halmay of Pt. Loma Marine Invertebrate Lab (Lakeside, CA, e-mail: peterhalmay@gmail.com) and housed in aquaria with artificial seawater (ASW) at 16°C (Coral Life Scientific Grade Marine Salt; Carson, CA). Adult *Lytechinus variegatus* were obtained from the Pelagic Corporation (pelagiccorp@bellsouth.net) and from the Duke University Marine Laboratory, Beaufort NC (joshua.osterberg@duke.edu). They were housed in aquaria with ASW at room temperature (Coral Life Scientific Grade Marine Salt; Carson, CA). Eggs and sperm of *S.purpuratus* or *L.variegatus* were spawned by injection of 0.5M KCl into the adult coelomic cavity. Fertilization was accomplished in sea water containing 1 mM 3-Amino-1,2,4-triazole to reduce cross-linking of the fertilization envelope, and which was washed out after 30 min. Embryos were cultured in filtered (0.2 micron) sea water collected at the Marine Biological laboratories in Woods Hole MA, until the appropriate stage. Lv embryos were cultured at room temperature, Sp embryos were cultured at 15°C.

METHOD DETAILS

Lv culture and dissociation for scRNA-seq

All embryos used in the study to obtain the Lv scRNA-seq dataset 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 seawater (For 1 L: NaCl 26.5 gm; KCl 0.7 gm; MgSO₄·7H₂O 11.9 gm; NaHCO₃ 0.5 gm; pH to 8.0, salinity should be 34 ppt; McClay, 2004), and then suspended in hyalin-extraction media (HEM, for 1 L: NaCl 18.5 gm; KCl 0.7 gm; MgSO₄·7H₂O 11.9 gm; glycine 22.5 gm; Tris base 1.21 gm; EGTA 0.76 gm; pH to 8.0;⁵⁶) for 10-15 min, depending on the stage of dissociation. When cells were beginning to dissociate, the embryos were collected and washed in 0.5M NaCl, gently sheared with a pipette, run through a 40 micron Nitex mesh, counted on a hemocytometer, and diluted to reach the appropriate concentration for the scRNA-seq protocol. Equal numbers of embryos were used in each time point and at no time were cells or embryos pelleted in a centrifuge.⁵⁷

Genome-guided de novo transcriptome assembly

For de novo transcriptome assembly, whole embryo RNA-seq dataset⁵⁸ were downloaded from NCBI (<https://www.ncbi.nlm.nih.gov>): SRR9673381, SRR9673382, SRR9673383, SRR9673384, SRR9673385, SRR9673386, SRR9673387, SRR9673388, SRR9673403, SRR9673404, SRR9673405, SRR9673406, SRR9673407, SRR9673408, SRR9673409 and SRR9673410. The raw reads were preprocessed by Trimmomatic⁵⁰ and then mapped to the Lv genome sequence from Li et al., (2020) by using Hisat2.⁵¹ The output SAM files were converted to BAM files and merged into a single BAM file. To obtain transcriptome data, the BAM file was assembled de novo using Trinity with default settings.⁵²

Prediction of UTR information

We used a gene annotation file from Li et al., (2020) with some modification. Since the annotation file contains only ORF regions, the UTR regions were predicted by Program to Assemble Spliced Alignments (PASA) based on the transcriptome information described in "genome-guided de novo transcriptome assembly".⁵³ In order to maximize the incorporation of UTR information into gene annotation, we performed PASA-pipeline two times. The PASA-updated annotation GFF3 file was converted into GTF file format by gffread.

Prediction of conventional gene names

To identify the conventional gene names of contigs in PASA-updated annotation files, we built a BLASTP database with a Sp peptide sequences downloaded from echinobase (*S. purpuratus* v5.0). The Lv peptide sequences from Li et al., (2020) was used for the query sequence for BLASTP analyses and best hit genes were identified as SPU_IDs. The SPU-IDs were converted into conventional gene names by using a table from echinobase (GenePageGeneralInfo_AllGenes.txt). The gene names in PASA-updated annotation file were replaced with the identified conventional gene names.

Single cell RNA sequencing

Single cell encapsulation was performed using the Chromium Single Cell Chip B kit on the 10x 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 counting was performed using Cell Ranger Single Cell Software Suite 3.0.2 from 10X Genomics. 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.^{54,59} A Seurat object was created by simply combining the individual datasets and then normalized by scaling gene expression in each cell by total gene expression. The top 2000 highly variable genes were then used for downstream analysis. The batch effects between samples were corrected by Harmony²⁸ with default setting. UMAP analysis was performed with the following parameters: dims = 1:20, resolution = 1. For Figures 1, 3, 4, and 5, cell cluster marker genes for single cell RNA sequencing datasets were extracted using the FindMarkers function in Seurat (using R) to obtain the adjusted p values for each gene using Bonferroni correction based on the total number genes in the corresponding dataset.

Comparison between *Lv* and *Sp* datasets

To compare transcriptional profiles between *Lv* and *Sp*, 993 genes from the *Lv* dataset and 1218 genes from the *Sp* dataset were selected with adjusted p value <1.00E-200. In addition, about 1000 genes that may play important roles during development were added manually. Next, the homologous genes between *Lv* and *Sp* were identified by checking the reciprocal best hits of BLASTP analysis. Genes without reciprocal best hit were not used for the downstream analysis because the genes are likely to be species-specific and not suitable for inter-species comparison. Finally, 886 genes were identified as the homologous genes between *Lv* and *Sp*. The homologous gene information was extracted from *Lv* annotation file⁵⁸ and *Sp* annotation file (*S. purpuratus* v5.0) and then the gene names were unified with exactly the same conventional gene names in both annotation files. Sequenced reads from *Lv* and *Sp* embryos were mapped to corresponding genome sequence and then counted based on the newly made annotation files containing 886 homologous genes. The resulting datasets were combined between species and then normalized by scaling gene expression in each cell by total gene expression. The batch effects between species were corrected by Liger with the following settings: k = 20, lambda = 5. UMAP analysis was performed with the following parameters: dims = 1:20, resolution = 0.5 (Welch et al., 2019). These inter-species comparisons were performed using EB, MB, EG and LG embryos separately.

Plasmid construction

For construction of GFP-NPDEFL and GFP-NPDE1–8 plasmids, *Sp-Nan2* 5' UTR, *Sp-Nan2* 3' UTR ΔGNARLE, GFP ORF were amplified from Nanos2-GFP ΔGNARLE.²³ NPDE fragments were artificially synthesized as single-strand DNA oligonucleotides and annealed them to give rise to double strand NPDE fragments. These fragments were ligated with pGEM-T-Easy vector using NEBuilder HiFi DNA Assembly kit (New England Biolabs, Cat#E2621). For construction of plasmids containing *Sp Nanos2* ORF and *Sp-Nan2* 3' UTR (*Sp-Sp*), *Lv-Nan2* ORF and *Lv-Nan2* 3' UTR (*Lv-Lv*), and the hybrid (*Sp-Lv* and *Lv-Sp*), total RNAs were isolated from *Lv* and *Sp* embryos using TRIzol, and cDNAs were synthesized using Super-Script III Reverse Transcriptase (Thermo Fisher Scientific, Cat#18080093). DNA fragments containing *Lv-Nan2* ORF and 3' UTR, and *Sp-Nan2* 3' UTR were amplified from the cDNAs. *Sp-Nan2* ORF, *Sp-Nan2* 5' UTR and GFP ORF were amplified from Nanos2-GFP ΔGNARLE²³ as described above.

Microinjections

Sea urchin eggs were dejellied by washing for 10 min in pH5.0 seawater (*Sp*), or by passing the eggs through a 100μm mesh (*Lv*). Eggs were then rowed on protamine sulfate coated Petri dishes. Zygotes were injected with 2 pL of injection solution as described below. The injected zygotes were cultured at 16°C for *Sp*, or at room temperature for *Lv*. To quantify expression levels of GFP-NPDE fusion protein, zygotes were injected with 2 pL of injection solution containing 3 pmol of GFP-NPDEFL or GFP-NPDE1–8 and *mCherry* mRNA by constant pressure and in the presence of 1 mM 3-AT (Sigma). As the control, GFP mRNA without NPDE fragment was injected. The GFP signal intensity was observed in the EB embryos and normalized with that of *mCherry* to standardize the injection volumes between embryos. The relative signal intensities of fusion proteins were calculated against the signal intensity of GFP without NPDE fragment. To observe the expression patterns of *Lv-Lv*, *Sp-Sp*, *Lv-Sp* and *Sp-Lv* mRNAs, these mRNAs were injected with Dextran Alexa Fluor 647 to screen injected and non-injected embryos. As the control, mRNA encoding GFP ORF, flanked by *Xenopus β-globin* 5' and -3'UTRs (Control GFP) was injected. The injected embryos were mounted on a slide glass and observed on a Nikon CSW-1 Scanning Disk confocal microscope. The signal intensity of GFP fusion protein was measured using the Fiji software.⁵⁵ At least 50-100 embryos were injected in each experiment, and at least 20 embryos were used for each quantification.

Significance was calculated between Control GFP and GFP-NPDE fusion proteins (NPDEFL and NPDE1–8) by Student's *t* test (*: $p < 0.05$).

Whole mount *in situ* hybridization

To synthesize Digoxigenin (DIG)-labeled RNA probe for GFP mRNAs, full-length of GFP ORF was amplified from Nanos2-GFP Δ GNARLE plasmid²³ using a reverse primer tailed with the T7 promoter sequence. DIG-labeled antisense probe was transcribed using the Roche DIG RNA labeling kit according to the manufacturer's instructions. WMISH was performed as described previously.⁸ Embryos injected with *Lv-Lv*, *Sp-Sp*, *Lv-Sp*, *Sp-Lv* or *Control GFP* mRNA were fixed with MOPS buffer containing 4% paraformaldehyde overnight and hybridized at 50°C with hybridization buffer (50% formamide, 0.1 M MOPS buffer, 0.5 M NaCl, 0.1% Tween 20, 1 mg/mL BSA and 1 ng/ μ L probe). Signal was detected using anti-DIG-POD antibody (Roche, Cat#11633716001) and amplified using TSA Plus Biotin Kit (Akoya Biosciences, Cat#NEL749A001KT) and Streptavidin Alexa Fluor 546 conjugated. Samples were mounted in Pro-Long Diamond Antifade Mountant (Thermo Fisher Scientific, Cat#P36961) and observed on a Nikon CSW-1 Scanning Disk confocal microscope.

Immunostaining

The fixed embryos were incubated in a blocking solution [2% BSA in PBST (0.1% Tween 20 in PBS)] for 30 min. After blocking, the embryos were incubated overnight at 4°C in a blocking solution containing rabbit anti-GFP antibody at 1:500 dilution. Then, the embryos were washed with PBST three times for 20 min each. For detection of anti-GFP antibody, the embryos were incubated in a blocking solution containing Alexa Fluor 488-conjugated goat anti-rabbit at 1:1000 dilution. The embryos were washed with PBST three times for 20 min each, and then mounted in Pro-Long Diamond Antifade Mountant (Thermo Fisher Scientific, Cat#P36961). The embryos were observed by a Nikon CSW-1 Scanning Disk confocal microscope.

Sequence comparison between species

To compare the NPDE and GNARLE sequences between *Lv*, *Sp* and *Hp*, nucleotide sequence and peptide sequence of *Sp*-Nan2 and *Hp*-Nan2 were obtained from Echinobase (<https://www.echinobase.org/entry/>). Sequences of *Lv*-Nan2 were extracted from the dataset used for scRNA-seq analysis. These sequences were aligned with T-coffee program with default setting⁶⁰; <https://tcoffee.crg.eu/apps/tcoffee/index.html> and then the results were visualized by Boxshade (http://www.ch.embnet.org/software/BOX_form.html).

Statistical analyses

Each experimental analysis is documented in the main text, in the Figure legends, and throughout this [method details](#). For example, at least 50–100 embryos were injected in each experiment (e.g. legend [Figure 6](#)), and at least 20 embryos were used for each quantification. Significance was calculated by Student's *t* test (*: $p < 0.05$). Error bars indicate standard errors of samples. The number of embryos (*n*) examined are shown in parentheses. Additional analyses are described, for example, in the section of [comparison between *lv* and *sp* datasets](#), e.g. transcriptional profiles were selected with adjusted *p* value $< 1.00E-200$.